Risk Analysis:
Measures of concordance,
their compatibility and capital allocation

by

Takaaki Koike

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Steven Vanduffel Professor Faculty of Economics Vrije Universiteit Brussel (VUB)
Supervisor:	Marius Hofert Associate Professor Statistics and Actuarial Science
Internal Members:	Alexander Schied Professor Statistics and Actuarial Science
	Ruodu Wang Associate Professor Statistics and Actuarial Science
Internal-External Member:	Alexandru Nica Professor Pure Mathematics

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Takaaki Koike was the sole author for Chapters 1 and 7 which were written under the supervision of Dr. Marius Hofert and were not written for publication. This thesis consists in part of five manuscripts written for publication. Exceptions to sole authorship of material are as follows:

Research presented in Chapters 2, 3, 5 and 6:

This research was conducted at the University of Waterloo by Takaaki Koike under the supervision of Dr. Marius Hofert. Takaaki Koike designed the study with assistance from Dr. Marius Hofert. Takaaki Koike drafted the manuscript and Dr. Marius Hofert provided intellectual input on manuscript drafts.

Citations:

Chapter 2: Marius Hofert and Takaaki Koike (2019). Compatibility and attainability of matrices of correlation-based measures of concordance. ASTIN Bulletin: *The Journal of the IAA*, **49**(3): 885-918.

Chapter 3: Takaaki Koike and Marius Hofert (2020). Estimation and comparison of correlation-based measures of concordance. *arXiv preprint arXiv:2006.13975*.

Chapter 5: Takaaki Koike and Marius Hofert (2020). Markov Chain Monte Carlo methods for estimating systemic risk allocations. *Risks*, **8**(1):6.

Chapter 6: Takaaki Koike and Marius Hofert (2020). Modality for scenario analysis and maximum likelihood allocation. *arXiv preprint arXiv:2005.02950*.

Research presented in Chapter 4:

This research was conducted partially at Keio University and completed at the University of Waterloo by Takaaki Koike under the supervision of Dr. Mihoko Minami. Takaaki Koike designed the study with assistance from Dr. Mihoko Minami. Takaaki Koike drafted the manuscript and Dr. Mihoko Minami provided intellectual input on manuscript drafts.

Citation:

Chapter 4: Takaaki Koike and Mihoko Minami. (2019). Estimation of risk contributions with MCMC. *Quantitative Finance*, **19**(9), 1579–1597.

As lead author of these five chapters, I was responsible for contributing to conceptualizing study design, carrying out data collection and analysis, and drafting and submitting manuscripts. My coauthors provided guidance during each step of the research and provided feedback on draft manuscripts.

Abstract

This thesis addresses various topics in the field of probability theory and statistics with applications in quantitative risk management.

The first topic concerns matrix compatibility and attainability problems for measures of concordance. We characterize a class of bivariate measures of concordance arising as Pearson's correlations of random variables transformed by a so-called concordance-inducing function. This class of transformed rank correlations includes Spearman's rho, Blomqvist's beta and van der Waerden's coefficient as special cases by taking uniform, Bernoulli and normal distributions as concordance-inducing functions, respectively. For multivariate random vectors, the correlation-based measures are extended as square matrices with entries given by the bivariate measures. We study compatibility and attainability problems for such measures, which ask whether a given square matrix can be realized as a matrix of pairwise measures of concordance for some random vector, and how such a random vector can be constructed. Dimension reduction of compatibility and attainability for block matrices is also studied.

The second topic of this thesis is estimating and comparing transformed rank correlation coefficients. We propose a novel framework for comparing transformed rank correlations in terms of the asymptotic variance of their canonical estimators. A general criterion derived from this framework is that concordance-inducing functions with smaller variances of squared random variables are more preferable. In particular, we show that Blomqvist's beta attains the optimal asymptotic variance and Spearman's rho outperforms van der Waerden's coefficient. We also find that the optimal bounds of the asymptotic variance are attained by Kendall's tau.

The third topic of this thesis is to efficiently estimate risk allocations, which is known to be challenging due to their rare-event nature. We first focus on the problem of estimating Value-at-Risk (VaR) contributions derived by the Euler principle, and propose a novel framework of their estimation by using Markov chain Monte Carlo (MCMC) methods. We prove consistency and asymptotic normality of the proposed estimators under certain assumptions on the underlying marginal and copula densities. The framework of estimating VaR contributions with MCMC methods is then extended to the estimation of wider class of the systemic risk measures and risk allocations whose jth component can be written as a risk measure of the jth conditional marginal loss distribution given the so-called crisis event. Improved sample efficiency of our MCMC estimators is expected since they consist of samples from the conditional loss distribution given the rare event of interest whereas existing estimators are constructed by first simulating the unconditional loss distribution and then extracting the samples satisfying the rare event condition. In a series of numerical experiments, we demonstrate that biases and mean squared errors for our MCMC estimators are reduced in comparison to existing estimators.

The last topic of this thesis is to investigate the conditional distribution of a loss random vector given that the aggregate loss equals an exogenously provided capital. This conditional distribution serves as a building block for calculating risk allocations such as such as VaR contributions. A superlevel set of this conditional distribution can be interpreted as a set of severe and plausible stress scenarios the given capital is supposed to cover. We show that various distributional properties of this conditional distribution are inherited from those of the underlying joint loss distribution. Among these properties,

we find that modality of the conditional distribution is an important feature in risk profile related to the number of risky scenarios likely to occur in a stressed situation. Under unimodality, we study a novel risk allocation method called maximum likelihood allocation (MLA), defined as the mode of the conditional distribution given the total capital. Under multimodality, a single vector of allocations can be less sound. To overcome this issue, we investigate the so-called multimodalty adjustment to increasing the soundness of risk allocations. Properties of the conditional distribution, MLA and multimodality adjustment are demonstrated in numerical experiments. In particular, we observe that negative dependence among losses typically leads to multimodality, and thus to multiple risky scenarios and higher multimodality adjustment.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Marius Hofert, for his guidance, passion for research and generous mind to respect my research. Despite my limited experience and ability as a researcher, he always respected my research interests and supported my work, by virtue of which I have enjoyed my research under his supervision at University of Waterloo. I thank my lucky stars to have met him at ETH Zürich, and it is a great honor to be one of his students. I wish I could requite his favor in the future.

I would also like to thank the members of my thesis examiners Professor Alexandru Nica, Alexander Schied, Ruodu Wang and Steven Vanduffel for kindly serving on my committee in spite of the severe and uncertain situation of the ongoing COVID-19 pandemic. I want to specifically thank Professor Ruodu Wang for all his support and inspiring conversations in which I always get encouraged by his passion for research and from which I learned how to cope with academic challenges.

I am deeply grateful to my former supervisor, Professor Mihoko Minami at Keio University, for her earnest guidance, thanks to which I got to know the joy of research and decided to pursue an academic career. I also thank people at Keio University, Professor Hiroshi Shiraishi, Kei Kobayashi, Kenichi Hayashi and Tomoshige Nakamura for fruitful discussions and warmly welcoming me when I visited Keio University. I would also like to give thanks to Professor Cathy Chen and her student Li Jie at Feng Chia University for their heartful welcome and assistance when I visited Feng Chia University, which was the first time of going abroad for me and was one of the best experiences in my life. I am also indebted to Professor Paul Embrechts at ETH Zürich for all his support concerning my visit of ETH Zürich, which broadened my horizon. I want to thank Professor Junichi Imai at Keio University, Kengo Kamatani at Osaka University, Klaus Herrmann at Université de Sherbrooke, Sebastian Fuchs at Technische Universität Dortmund and Tsukahara Hideatsu at Seijo University for research discussions and invaluable comments on my research.

Amongst many people in the Department of Statistics and Actuarial Science who have continuously supported my work, I would first like to thank Mary Lou Dufton for her generous support. I would also like to show my gratitude to all the colleagues making my life at Waterloo brighter, and I wish them the best for their future. I am also obliged to all the fund providers for their financial support so that I could concentrate on my research in Canada. In the end, I would like to express my gratitude to my parents and girlfriend for their patience and countless warmth.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

This thesis addresses various topics in the field of probability theory, statistics and their applications in quantitative risk management. This chapter serves as an introduction, four main chapters follow, and the last chapter is devoted to conclusion.

Chapter 2 is dedicated to measures of concordance and related compatibility and attainability problems. A review of measures of concordance is provided in Section 1.2. After motivating this topic in Section 2.1, we characterize a class of bivariate measures of concordance arising from Pearson's linear correlation coefficient of transformed random variables (Section 2.2). Compatibility and attainability problems of this class of measures of concordance are studied in Section 2.3. Dimension reduction of these problems for block matrices is investigated in Section 2.4. Section 2.5 concludes with discussions on related open problems.

Chapter 3 addresses the problem of estimating and comparing transformed rank correlation coefficients defined in Chapter 2. After a brief introduction of the problem in Section 3.1, we introduce a framework for comparing transformed rank correlations in terms of their asymptotic variances in Section 3.2. A general criterion derived from this framework is that concordance-inducing functions with smaller variances of squared random variables are more preferable. In particular, we show that Blomqvist's beta attains the optimal asymptotic variance and Spearman's rho outperforms van der Waerden's coefficient. In Section 3.3, we compare transformed rank correlations and Kendall's tau, and show that the optimal bounds of the asymptotic variance are attained by Kendall's tau. In Section 3.4, a simulation study is conducted to compare asymptotic variances for various parametric copulas and concordance-inducing fuctions. Section 3.5 concludes this work with discussions about directions for future research.

Chapter 4 addresses the problem of estimating Value-at-Risk (VaR) contributions with Markov chain Monte Carlo (MCMC) methods. Section 1.3 and 1.4 serve as preliminaries for briefly introducing the problem of risk allocation and MCMC methods. An overview and key ideas of MCMC estimators are

presented in Section 4.1. After reviewing existing estimators of VaR contributions in Section 4.2, we propose MCMC (MH) estimators that utilize the Metropolis-Hastings (MH) algorithm for estimating VaR contributions (Section 4.3). Consistency and asymptotic normality of MH estimators are studied in Section 4.4. Numerical studies are conducted in Section 4.5. Finally, Section 4.6 concludes with potential future work.

In Chapter 5, the framework of estimating VaR contributions with MCMC methods is extended to a more general class of allocations called systemic risk allocations. After a brief introduction in Section 5.1, the class of systemic risk allocations and their crude estimation methods are presented in Section 5.2. In Section 5.3, we explain how advanced MCMC methods such as Hamiltonian Monte Carlo and Gibbs samplers are utilized to estimate systemic risk allocations. Numerical experiments are conducted in Section 5.4 and Section 5.5 concludes with remarks on further research.

In Chapter 6, we analyze dependence, tail behavior and multimodality of the conditional distribution of a loss random vector given that the aggregate loss equals an exogenously provided capital. After a brief introduction in Section 6.1, Section 6.2 provides motivations for studying this conditional distribution from the viewpoint of scenario analysis and stress testing of risk allocations. Section 6.3 is devoted to the investigation of the distributional properties of such conditional distributions, including dependence, tail behaviors and unimodality. In Section 6.4, we propose a novel risk allocation method called the maximum likelihood allocation and study its properties and related adjustment under multimodality. Section 6.5 is dedicated to numerical experiments for demonstrating modality of the conditional distribution and for comparing Euler and maximum likelihood allocations. Section 6.6 concludes with potential future work.

In the remainder of this chapter, we review basic concepts used in subsequent chapters. In Section 1.2, we briefly review the axioms of measures of concordance proposed by Scarsini (1984). In Section 1.3, we introduce the mathematical setting of the capital allocation problem and explain challenges when estimating VaR contributions with existing estimators. A brief review of MCMC methods is provided in Section 1.4. Throughout the thesis, we assume all appearing random variables to be defined on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For a random vector $\boldsymbol{X}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with probability density function $f_{\boldsymbol{X}}$, denote by $\operatorname{supp}(f_{\boldsymbol{X}})$ or $\operatorname{supp}(\boldsymbol{X})$ the support of $\boldsymbol{X}$, that is, $\operatorname{supp}(\boldsymbol{X}) = \{\boldsymbol{x} : f_{\boldsymbol{X}}(\boldsymbol{x}) > 0\}$. Furthermore let $\mathcal{M}_+^{d\times d}$, $d \in \mathbb{N}$, denote the set of all positive definite $d \times d$ matrices. In addition, all numerical experiments are run on a MacBook Air with 1.4 GHz Intel Core i5 processor and 4 GB 1600 MHz of DDR3 RAM.

1.2 Measures of concordance

Pearson's linear correlation is a widely used measure of dependence, that is, a single number to quantify the dependence between two random variables. However, as indicated by Embrechts et al. (2002), linear correlation has a number of limitations as a measure of dependence and a *measure of concordance* (also called a *rank correlation*) is known to be more suitable than linear correlation for quantifying the dependence of two random variables independently of their marginal distributions. In this section, we define a measure of concordance by seven axioms. We then briefly explain the notions of compatibility and attainability which will be studied in Chapter 2.

Let $\mathcal{H}$ be the set of all bivariate distributions with the same continuous margins, that is, $\lim_{x\to\infty} H(x,y) = \lim_{x\to\infty} H'(x,y)$ and $\lim_{y\to\infty} H(x,y) = \lim_{y\to\infty} H'(x,y)$ for all $x, y \in \mathbb{R}$ and $H, H' \in \mathcal{H}$. We call $H' \in \mathcal{H}$ *more concordant* than $H \in \mathcal{H}$, denoted by $H \preceq H'$, if $H(x,y) \leq H'(x,y)$ for all $(x,y) \in \mathbb{R}^2$. A map $\kappa : \mathcal{H} \to \mathbb{R}$ is called a *measure of concordance* if it satisfies the following seven axioms. Note that we write $(X,Y) \sim H$ when a random vector (X,Y) has a bivariate distribution $H \in \mathcal{H}$, and we identify $\kappa(X,Y)$ for $(X,Y) \sim H \in \mathcal{H}$ with $\kappa(H)$ for $\kappa : \mathcal{H} \to \mathbb{R}$.

Definition 1.2.1 (Scarsini (1984))**.** A map $\kappa : \mathcal{H} \to \mathbb{R}$ is called a *measure of concordance* if it satisfies the following seven axioms.

1. *Domain:* $\kappa(X,Y)$ is defined for any $(X,Y) \sim H \in \mathcal{H}$.
2. *Symmetry:* $\kappa(X,Y) = \kappa(Y,X)$ for any $(X,Y) \sim H \in \mathcal{H}$.
3. *Coherence:* If $H \preceq H'$ for $H, H' \in \mathcal{H}$, then $\kappa(H) \leq \kappa(H')$.
4. *Range:* $-1 \leq \kappa(X,Y) \leq 1$ for any $(X,Y) \sim H \in \mathcal{H}$ and the bounds are attainable.
5. *Independence:* $\kappa(X,Y) = 0$ if X and Y are independent.
6. *Change of sign:* $\kappa(X,-Y) = -\kappa(X,Y)$ for any $(X,Y) \sim H \in \mathcal{H}$.
7. *Continuity:* Let $(X_n, Y_n) \sim H_n \in \mathcal{H}$, $n \in \mathbb{N}$, and $(X,Y) \sim H \in \mathcal{H}$ with H_n converging pointwise to H as $n \to \infty$. Then $\lim_{n\to\infty} \kappa(X_n, Y_n) = \kappa(X,Y)$.

Examples of measures of concordance include Spearman's rho, Kendall's tau, Gini's gamma and Blomqvist's beta; see Example 2.2.3 and Remark 2.2.10 for their definitions.

For a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, $d \geq 2$, a bivariate measure of concordance κ can be extended to a matrix of pairwise measures of concordance

$$\kappa(\boldsymbol{X}) = (\ \kappa(X_i, X_j)\ ) \in [-1,1]^{d\times d},$$

as an analog to correlation matrices. *Compatibility* concerns whether a given $d \times d$ matrix R is realizable as a matrix of pairwise κ, that is, whether there exists a d-dimensional continuous random vector $\boldsymbol{X}$ such that $\kappa(\boldsymbol{X}) = R$. Provided that a given matrix R is κ-compatible, *attainability* concerns how to construct $\boldsymbol{X}$ with $\kappa(\boldsymbol{X}) = R$. It is well-known that a matrix $R \in \mathbb{R}^{d\times d}$ is compatible with Pearson's correlation coefficient if it belongs to the so-called *elliptope* $\mathcal{P}_d$, which is the set of positive semi-definite symmetric $[-1,1]$-valued matrices with main diagonal entries equal to one. Any $R \in \mathcal{P}_d$ is attainable since one can construct a d-dimensional continuous random vector $\boldsymbol{X}$ such that $\rho(\boldsymbol{X}) = R$ by taking $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{0}_d, R)$. The main topic of Chapter 2 is to solve the compatibility and the attainability problem for a wide class of measures of concordance.

1.3 Capital allocation problem

In portfolio risk management, *capital allocation* is an essential step to measure the risk of each unit of a portfolio by decomposing the total risk of the whole portfolio. Consider the aggregate loss $S = \sum_{j=1}^{d} X_j$

where $d \geq 2$ is the size of the portfolio, and $X_1, \dots, X_d$ are random variables that represent the losses incurred by exposures $j = 1, 2, \dots, d$ within a fixed time period. Throughout the thesis, a positive value of a loss random variable represents a financial loss, and a negative loss is interpreted as a profit. Let $F_{\boldsymbol{X}}$ be the joint cumulative distribution function (cdf) of $\boldsymbol{X} = (X_1, \dots, X_d)$ with margins $F_1, \dots, F_d$, and let F_S be the cdf of the total loss S. According to Sklar's Theorem, $F_{\boldsymbol{X}}$ can be written by

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d, \tag{1.1}$$

where C is the *copula* of $\boldsymbol{X}$; see Nelsen (2006). A d-dimensional copula is referred to as a d-copula in short. When $F_{\boldsymbol{X}}$ has a probability density function (pdf) $f_{\boldsymbol{X}}$ with marginal densities $f_1, \dots, f_d$, the joint pdf $f_{\boldsymbol{X}}$ can be written by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d), \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{1.2}$$

where c denotes the density of C. Similarly we denote by f_S the pdf of F_S.

Deriving *allocated capitals* is an important task in risk management. A standard procedure of determining allocated capitals involves two steps. The first step is to compute the economic capital $\varrho(S)$ for a risk measure ϱ, which maps a loss random variable to a real number interpreted as capital buffer that is required to cover the loss over a predetermined period of time. One widely used risk measure is *Value-at-Risk (VaR)*. For a random variable $X \sim F$, VaR of X at confidence level $p \in (0, 1)$ is defined by

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : F(x) \geq p\},$$

range Value-at-Risk (RVaR) at confidence levels $0 < p_1 < p_2 \leq 1$ is defined by

$$\mathrm{RVaR}_{p_1,p_2}(X) = \frac{1}{p_2 - p_1} \int_{p_1}^{p_2} \mathrm{VaR}_q(X)\, \mathrm{d}q,$$

and *expected shortfall (ES)* at confidence level $p \in (0, 1)$ is defined by $\mathrm{ES}_p(X) = \mathrm{RVaR}_{p,1}(X)$ provided that $\mathbb{E}[|X|] < \infty$. Note that ES is also known as *C(onditional)VaR, T(ail)VaR, A(verage)VaR* and *C(onditional)T(ail)E(expectation)*. The risk measures VaR, RVaR and ES are law-invariant in the sense that they depend only on the distribution of X. Therefore, we sometimes write $\varrho(F)$ instead of $\varrho(X)$ for these risk measures.

The second step is to allocate the capital $\varrho(S)$ to the d individual exposures. The problem of capital allocation is to determine the vector of allocated capitals $(\mathrm{AC}_1, \dots, \mathrm{AC}_d) \in \mathbb{R}^d$ that satisfies the *full allocation property*

$$\varrho(S) = \sum_{j=1}^{d} \mathrm{AC}_j. \tag{1.3}$$

The *Euler principle*, proposed in Tasche (1995), derives such allocated capitals by utilizing the well-known Euler rule for a function $\boldsymbol{\lambda} \mapsto \varrho(\boldsymbol{\lambda}^\top \boldsymbol{X})$:

$$\varrho(\boldsymbol{\lambda}^\top \boldsymbol{X}) = \sum_{j=1}^{d} \lambda_j \frac{\partial \varrho(\boldsymbol{\lambda}^\top \boldsymbol{X})}{\partial \lambda_j}, \quad \boldsymbol{\lambda} \in \Lambda, \tag{1.4}$$

where $\Lambda \subset \mathbb{R}^d \backslash \{\mathbf{0}\}$ is an open set such that $\mathbf{1}_d \in \Lambda$, and ϱ is a positive homogeneous risk measure, that is, $\varrho(\lambda X) = \lambda \varrho(X)$ holds for $\lambda > 0$. The euler principle is economically justified, for example, in Denault

(2001) and Tasche (1995, 2008), and the resulting allocated capital is also known as the *Aumann-Shapley value* (Aumann and Shapley, 2015) for risk capital allocation problems; see, for example, Boonen et al. (2020), Kalkbrener (2005), and Myers and Read (2001). By taking $\boldsymbol{\lambda} = \mathbf{1}_d$ in Equation (1.4), the vector $(\mathrm{AC}_1^{\varrho}, \dots, \mathrm{AC}_d^{\varrho})$ with

$$\mathrm{AC}_j^{\varrho} = \frac{\partial \varrho(\boldsymbol{\lambda}^\top \boldsymbol{X})}{\partial \lambda_j}\bigg|_{\boldsymbol{\lambda}=\mathbf{1}_d}, \quad j = 1, 2, \dots, d,$$

satisfies the full allocation property (1.3) . In addition, this capital allocation satisfies the so-called *RORAC compatibility*, which means that the profitability of each asset is consistently signaled via the ratio of its return and allocated capital; see Tasche (2008) for details. Since VaR, RVaR and ES are all positive homogeneous, the Euler principle is applicable and the derived allocated capitals, called the *VaR, RVaR* and *ES contributions*, are given by

$$\mathrm{AC}_j^{\mathrm{VaR}_p} = \frac{\partial \mathrm{VaR}_p(\boldsymbol{\lambda}^\top \boldsymbol{X})}{\partial \lambda_j}\bigg|_{\boldsymbol{\lambda}=\mathbf{1}_d} = \mathbb{E}[X_j \mid \{S = \mathrm{VaR}_p(S)\}], \tag{1.5}$$

$$\mathrm{AC}_j^{\mathrm{RVaR}_{p_1,p_2}} = \frac{\partial \mathrm{RVaR}_{p_1,p_2}(\boldsymbol{\lambda}^\top \boldsymbol{X})}{\partial \lambda_j}\bigg|_{\boldsymbol{\lambda}=\mathbf{1}_d} = \mathbb{E}[X_j \mid \{\mathrm{VaR}_{p_1}(S) \le S \le \mathrm{VaR}_{p_2}(S)\}], \tag{1.6}$$

$$\mathrm{AC}_j^{\mathrm{ES}_p} = \frac{\partial \mathrm{ES}_p(\boldsymbol{\lambda}^\top \boldsymbol{X})}{\partial \lambda_j}\bigg|_{\boldsymbol{\lambda}=\mathbf{1}_d} = \mathbb{E}[X_j \mid \{S \ge \mathrm{VaR}_p(S)\}], \tag{1.7}$$

respectively; see Tasche (2001) and Fischer et al. (2018) for derivations.

Computing these risk contributions for a general loss random vector $\boldsymbol{X}$ is known to be challenging since their analytical calculation is rarely possible, and their estimation requires rare-event simulation. Efficient estimation of VaR contributions will be addressed in Chapter 4, and that of a more general class of risk allocations including RVaR and ES contributions as special cases will be considered in Chapter 5. The conditional distribution of $\boldsymbol{X} \mid \{S = \mathrm{VaR}_p(S)\}$ in the formula of VaR contributions (1.5) will be further investigated in Chapter 6 due to its importance in scenario analysis and stress testing of risk allocations.

1.4 A brief introduction to MCMC

In this section we briefly review Markov chain Monte Carlo (MCMC) methods, especially the Metropolis-Hastings (MH) algorithm, for simulating a distribution of interest by constructing a Markov chain whose stationary distribution is the desired one. See, for example, Nummelin (2004) for the general theory of Markov chains. By allowing Markovian-type dependence within the samples, MCMC methods allow one to simulate a wide variety of distributions including the conditional distributions appearing in the risk contributions (1.5), (1.6) and (1.7). MCMC methods will be extensively studied in Chapter 4 and Chapter 5 for efficient estimation of risk contributions.

1.4.1 An overview of the MCMC theory

Let $E \subseteq \mathbb{R}^d$ be a set and $\mathcal{E}$ be a σ-algebra on E. A *Markov chain* is a sequence of E-valued random variables $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots)$ satisfying the Markov property;

$$\mathbb{P}(\boldsymbol{X}^{(n+1)} \in A \mid \boldsymbol{X}^{(k)} = \boldsymbol{x}^{(k)}, k \le n) = \mathbb{P}(\boldsymbol{X}^{(n+1)} \in A \mid \boldsymbol{X}^{(n)} = \boldsymbol{x}^{(n)}),$$

for all $n \ge 1$, $A \in \mathcal{E}$, and $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)} \in E$. A Markov chain is characterized by its *stochastic kernel* $K : E \times \mathcal{E} \to [0, 1]$, given by $\boldsymbol{x} \times A \mapsto K(\boldsymbol{x}, A) = \mathbb{P}(\boldsymbol{X}^{(n+1)} \in A \mid \boldsymbol{X}^{(n)} = \boldsymbol{x})$. If there exists a probability distribution π such that $\pi(A) = \int_E \pi(\mathrm{d}\boldsymbol{x}) K(\boldsymbol{x}, A)$ for any $\boldsymbol{x} \in E$ and $A \in \mathcal{E}$, then π is called the *stationary distribution*. Assuming $K(\boldsymbol{x}, \cdot)$ has a density $k(\boldsymbol{x}, \cdot)$, the *detailed balance condition* (also known as the *reversibility*) with respect to π is given by

$$\pi(\boldsymbol{x}) k(\boldsymbol{x}, \boldsymbol{y}) = \pi(\boldsymbol{y}) k(\boldsymbol{y}, \boldsymbol{x}), \quad \boldsymbol{x}, \boldsymbol{y} \in E, \tag{1.8}$$

and is known as a sufficient condition for the corresponding kernel K to have the stationary distribution π; see Chib and Greenberg (1995).

MCMC methods are widely used for simulating a distribution by generating a Markov chain with the given distribution as stationary distribution π. For some distribution π and a π-measurable vector-valued function $\boldsymbol{h}$ on E, assume that the quantity of interest is

$$\boldsymbol{\pi}(\boldsymbol{h}) = \int_E \boldsymbol{h}(\boldsymbol{x}) \pi(\mathrm{d}\boldsymbol{x}). \tag{1.9}$$

The corresponding MCMC estimator of (1.9) is then given by

$$\hat{\boldsymbol{\pi}}_N(\boldsymbol{h}) = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{h}(\boldsymbol{X}^{(n)}), \tag{1.10}$$

where $(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)})$ is a sample path from time 1 to N (we call it an *N-path*) of a Markov chain whose stationary distribution is π. The distribution π is called the *target distribution*. Since it is determined by the problem at hand, the problem is to find a stochastic kernel K such that it has the stationary distribution π, and sample paths of its Markov chain can easily be generated.

One of the most popular stochastic kernels is the *Metropolis-Hastings (MH)* kernel defined by

$$K(\boldsymbol{x}, \mathrm{d}\boldsymbol{y}) = k(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y} + r(\boldsymbol{x}) \delta_{\boldsymbol{x}}(\mathrm{d}\boldsymbol{y}),$$

where $\delta_{\boldsymbol{x}}$ is the Dirac delta function, $k(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{x}, \boldsymbol{y}) \alpha(\boldsymbol{x}, \boldsymbol{y})$, $q : E \times E \to \mathbb{R}_+$ is a function called a *proposal density* such that $\boldsymbol{x} \mapsto q(\boldsymbol{x}, \boldsymbol{y})$ is measurable for any $\boldsymbol{y} \in E$, and $\boldsymbol{y} \mapsto q(\boldsymbol{x}, \boldsymbol{y})$ is a probability density for any $\boldsymbol{x} \in E$. Furthermore,

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \min\left\{ \frac{\pi(\boldsymbol{y}) q(\boldsymbol{y}, \boldsymbol{x})}{\pi(\boldsymbol{x}) q(\boldsymbol{x}, \boldsymbol{y})}, 1 \right\}, & \text{if } \pi(\boldsymbol{x}) q(\boldsymbol{x}, \boldsymbol{y}) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and $r(\boldsymbol{x}) = 1 - \int_E k(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y}$. It can be shown that the MH kernel has stationary distribution π; see Tierney (1994). Under the three conditions (i)–(iii) where

(i) at least one vector $\boldsymbol{x}^{(0)} \in \mathrm{supp}(\pi)$ is known,

(ii) samples from $q(\boldsymbol{x}, \cdot)$ can be generated for any $\boldsymbol{x} \in E$, and

(iii) the ratio $\pi(\boldsymbol{y})/\pi(\boldsymbol{x})$ can be calculated for any $\boldsymbol{x}, \boldsymbol{y} \in E$,

one can generate an N-path of the desired Markov chain by the *MH algorithm* (Metropolis et al., 1953; Hastings, 1970) given in Algorithm 1.

Algorithm 1 Metropolis-Hastings (MH) algorithm

Require: Random number generator of $q(\boldsymbol{x}, \cdot)$ for $\boldsymbol{x} \in E$, $\boldsymbol{x}^{(0)} \in \mathrm{supp}(\pi)$ and the ratio $\pi(\boldsymbol{y})/\pi(\boldsymbol{x})$ for $\boldsymbol{x}, \boldsymbol{y} \in E$.

Input: Sample size $N \in \mathbb{N}$, proposal density q, and initial value $\boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}$.

Output: Sample path $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}$ of the Markov chain.

for $n = 0, \ldots, N-1$ **do**

1) Generate the candidate $\tilde{\boldsymbol{X}}^{(n)} \sim q(\boldsymbol{X}^{(n)}, \cdot)$.

2) Calculate the *acceptance probability*

$$\alpha_n = \alpha(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n)}) = \min\left[\frac{\pi(\tilde{\boldsymbol{X}}^{(n)}) q(\tilde{\boldsymbol{X}}^{(n)}, \boldsymbol{X}^{(n)})}{\pi(\boldsymbol{X}^{(n)}) q(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n)})}, 1\right]. \tag{1.11}$$

3) Set $\boldsymbol{X}^{(n+1)} = \tilde{\boldsymbol{X}}^{(n)}$ with probability α_n and $\boldsymbol{X}^{(n+1)} = \boldsymbol{X}^{(n)}$ with probability $1 - \alpha_n$.

end for

We call $\alpha_n = \alpha(\boldsymbol{X}^{(n)}, \tilde{\boldsymbol{X}}^{(n)})$ in (1.11) the *acceptance probability* in the nth iteration. The MH estimator (1.10) is then computed based on the N-path $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)})$ generated in Algorithm 1.

Under regularity conditions, the MCMC estimator $\hat{\boldsymbol{\pi}}_N(\boldsymbol{h})$ is *consistent* and *asymptotically normal*. The MCMC estimator is said to be *consistent* if

$$\lim_{N \to \infty} \hat{\boldsymbol{\pi}}_N(\boldsymbol{h}) = \boldsymbol{\pi}(\boldsymbol{h}) \quad \text{a.s.}, \tag{1.12}$$

for any π-integrable function $\boldsymbol{h}$ and any initial state $\boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)} \in \mathrm{supp}(\pi)$. Next, the *central limit theorem (CLT)* holds if

$$\sqrt{N}\{\hat{\boldsymbol{\pi}}_N(\boldsymbol{h}) - \boldsymbol{\pi}(\boldsymbol{h})\} \xrightarrow{d} \mathcal{N}_d(\boldsymbol{0}, \Sigma_{\boldsymbol{h}}) \quad \text{as} \quad N \to \infty, \tag{1.13}$$

where the asymptotic variance matrix is given by

$$\Sigma_{\boldsymbol{h}} = \mathrm{Var}_\pi[\boldsymbol{h}(\boldsymbol{X}^{(1)})] + 2\sum_{k=1}^{\infty} \mathrm{Cov}_\pi[\boldsymbol{h}(\boldsymbol{X}^{(1)}), \boldsymbol{h}(\boldsymbol{X}^{(k+1)})]. \tag{1.14}$$

Since the asymptotic variance (1.14) can rarely be computed in real situations, it needs to be estimated from the sample path $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)})$ generated in Algorithm 1. One popular estimator of $\Sigma_{\boldsymbol{h}}$ is the so-called *batch means estimator*; see Geyer (2011). For an N-path $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)})$, the batch means estimator $\hat{\Sigma}_{\boldsymbol{h},N}$ is defined by

$$\hat{\Sigma}_{\boldsymbol{h},N} = \frac{L_N}{B_N - 1} \sum_{b=1}^{B_N} \{\hat{\boldsymbol{\pi}}_{N,b}(\boldsymbol{h}) - \hat{\boldsymbol{\pi}}_N(\boldsymbol{h})\}\{\hat{\boldsymbol{\pi}}_{N,b}(\boldsymbol{h}) - \hat{\boldsymbol{\pi}}_N(\boldsymbol{h})\}^\top,$$

where L_N and B_N are positive integers satisfying $N = L_N B_N$, and

$$\hat{\boldsymbol{\pi}}_{N,b}(\boldsymbol{h}) = \frac{1}{L_N} \sum_{l=(b-1)L_N}^{bL_N - 1} \boldsymbol{h}(\boldsymbol{X}^{(l)}) \quad \text{for} \quad b = 1, 2, \ldots, B_N.$$

L_N is called the *batch length*, and B_N is the number of batches. Under regularity conditions, the batch means estimator $\hat{\Sigma}_{\boldsymbol{h},N}$ converges to $\Sigma_{\boldsymbol{h}}$ as $N \to \infty$; see Jones et al. (2006) and Vats et al. (2019). By using asymptotic normality of $\hat{\boldsymbol{\pi}}_N(\boldsymbol{h})$ and consistency of $\hat{\Sigma}_{\boldsymbol{h},N}$, one can construct an approximate confidence region of the true quantity $\boldsymbol{\pi}(\boldsymbol{h})$ based on an N-path of the Markov chain.

1.4.2 Choice of the proposal distribution

When implementing the MH algorithm, an appropriate choice of the proposal density q is necessary since it affects the asymptotic variance (1.14). Since Σ_h can rarely be calculated explicitly in real situations, a post-implementation review is usually conducted, that is, the goodness of the selected proposal distribution is evaluated after performing the MH algorithm. In this section, we review two methods for evaluating a selected proposal distribution. We also provide families of proposal distributions for later use.

In practice, there are two prevalent methods to evaluate the performance of the proposal distribution. One is to inspect the *autocorrelation plots* of the marginal sample paths. For an N-path $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)})$, vector-valued measurable function $\boldsymbol{h}(\boldsymbol{X}) = (h_1(\boldsymbol{X}), \ldots, h_d(\boldsymbol{X}))$, and the MH estimator $\hat{\boldsymbol{\pi}}_N(\boldsymbol{h}) = (\hat{\pi}_{N,1}, \ldots, \hat{\pi}_{N,d})$, the sample autocorrelations $\hat{r}_j(k) = \hat{R}_j(k)/\hat{R}_j(0)$ are drawn against the lags $k = 0, 1, 2, \ldots$, where

$$\hat{R}_j(k) = \frac{1}{N-k} \sum_{n=1}^{N-k} \{h_j(\boldsymbol{X}^{(n)}) - \hat{\pi}_{N,j}\}\{h_j(\boldsymbol{X}^{(n+k)}) - \hat{\pi}_{N,j}\}, \quad j = 1, 2, \ldots, d.$$

The asymptotic variance $\Sigma_{\boldsymbol{h}}$ is expected to be small if the autocorrelation plots steadily decline to zero as the lags increase. Another implicative quantity is the *acceptance rate (ACR)*, which is the percentage of times a candidate $\tilde{\boldsymbol{X}}$ is accepted through the whole run. Since an appropriate ACR varies according to the shape and dimension of the target distribution, there is no general standard on the range of ACR; see Rosenthal et al. (2011) and references therein. Meanwhile, altering proposal distribution is generally suggested when an extremely low or high ACR is observed. Figure 1.1 illustrates two typical situations when such extreme ACRs are observed. The first situation, highlighted in red in Figure 1.1, represents the case when the chain tends to be stuck at one point due to, for example, a too high variance of the proposal distribution. The resulting estimator can then have very high variance. The second situation, highlighted in blue, shows the case where the chain moves only around one mode of the target density and does not

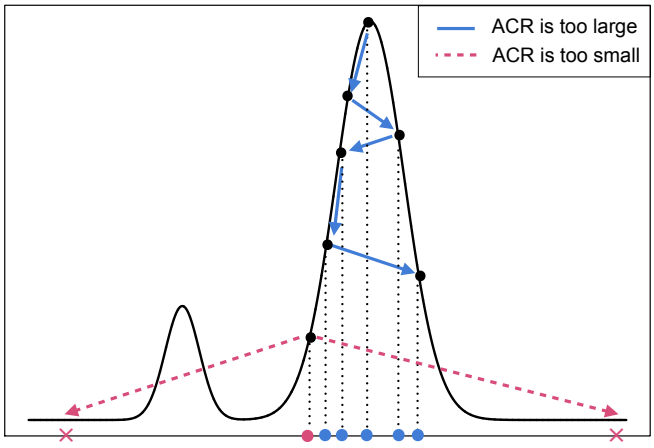

Figure 1.1: Two typical situations when an extremely low or high acceptance rate (ACR) of the Metropolis-Hastings algorithm is observed. In the case highlighted in red, proposed candidates have very low values of the target density and the chain tends to be stuck in one point. In the situation highlighted in blue, the chain only moves around a single mode and does not traverse the entire support of the target distribution.

traverse the entire support of the target distribution. A falsely high ACR may lead to a high bias caused by ignoring other modes.

Typically, proposal distribution q is selected from certain classes of distributions. To find an appropriate q depending on the target distribution, the following classes of proposal distributions are often used due to their simplicity. First, if the proposal function is of the form $q(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y} - \boldsymbol{x})$ for some density f, the candidate $\tilde{\boldsymbol{X}}$ is drawn according to

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{Z}, \quad \text{where} \quad \boldsymbol{Z} \sim f, \tag{1.15}$$

and where $\boldsymbol{X}$ is the current state. This type of q is called the *random walk* proposal distribution. In the case where f is symmetric around the origin, the acceptance probability (1.11) is written simply as $\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left[\frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})}, 1\right]$. Second, when $q(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y})$ for some density f, then the candidate $\boldsymbol{X}$ is updated by

$$\tilde{\boldsymbol{X}} = \boldsymbol{Z}, \quad \text{where} \quad \boldsymbol{Z} \sim f. \tag{1.16}$$

This q is called the *independent* proposal distribution. The random walk and independent proposal distributions often fail to perform well when the target distribution π is heavy-tailed. To overcome this problem, the *mixed preconditioned Crank-Nicolson (MpCN)* proposal distribution is proposed by Kamatani et al. (2018). This proposal distribution updates the candidate according to

$$\tilde{\boldsymbol{X}} = \boldsymbol{\mu} + \rho^{\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu}) + (1-\rho)^{\frac{1}{2}} Z^{-\frac{1}{2}} \cdot \boldsymbol{W}, \tag{1.17}$$

where $\rho \in (0, 1)$, Z follows the gamma distribution with shape parameter $d/2$ and scale parameter $||\Sigma^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})||^2/2$, and $\boldsymbol{W} \sim \mathcal{N}_d(\boldsymbol{0}, \Sigma)$ for some d-vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and $d \times d$ matrix $\Sigma \in \mathcal{M}_+^{d \times d}$. Ideally,

$\boldsymbol{\mu}$ and Σ are set to be $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}]$ and $\Sigma = \mathrm{Cov}[\boldsymbol{X}]$, while in practice, they can be replaced by rough estimates since moments of $\boldsymbol{X}$ are typically unknown. Note that the original MpCN proposed in Kamatani et al. (2018) is the standardized version, that is, $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = \boldsymbol{I}_d$, where $\boldsymbol{I}_d$ is the identity matrix. The acceptance probability (1.11) of the MpCN proposal distribution can be written as

$$\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = \left[\frac{\pi(\tilde{\boldsymbol{X}})}{\pi(\boldsymbol{X})} \left(\frac{||\Sigma^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})||}{||\Sigma^{-\frac{1}{2}}(\tilde{\boldsymbol{X}} - \boldsymbol{\mu})||} \right)^{-d}, 1 \right].$$

One of the key differences between this proposal distribution and the first two simple ones is that in the MpCN, not only the mean but also the variance of the candidate changes with the current state $\boldsymbol{X}$. Since the MpCN proposal distribution admits larger jumps in the tail parts of π, a better acceptance rate can be expected even when π is heavy-tailed.

Chapter 2

Measures of concordance and their compatibility

Measures of concordance have been widely used in insurance and risk management to summarize non-linear dependence among risks modeled by random variables, that Pearson's correlation coefficient cannot capture. However, popular measures of concordance, such as Spearman's rho and Blomqvist's beta, appear as classical correlations of transformed random variables. We characterize a whole class of such concordance measures arising from correlations of transformed random variables, which includes Spearman's rho, Blomqvist's beta and van der Waerden's coefficient as special cases. Compatibility and attainability of square matrices with entries given by such measures are studied, that is, whether a given square matrix of such measures of concordance can be realized for some random vector and how such a random vector can be constructed. Compatibility and attainability of block matrices and hierarchical matrices are also studied due to their practical importance in insurance and risk management. In particular, a subclass of attainable block Spearman's rho matrices is proposed to compensate for the drawback that Spearman's rho matrices are in general not attainable for dimensions larger than four. Another result concerns a novel analytical form of the Cholesky factor of block matrices which allows one, for example, to construct random vectors with given block matrices of van der Waerden's coefficients.

2.1 Introduction

Since the work of Embrechts et al. (2002), copulas have been widely adopted in insurance and risk management to quantify dependence between continuously distributed random variables; see Genest et al. (2009). To summarize the dependence captured by the copula by a single number, measures of concordance are frequently used. For more than two random variables, multivariate measures of concordance exist but are typically not unique extensions of their bivariate counterparts to higher dimensions; see Joe (1990), (Jaworski et al., 2010, Chapter 10) and references therein. Similar to the notion of correlation, matrices of (pairwise) measures of concordance have recently become of interest; see, for example, Embrechts et al.

(2016) (motivated from an application in insurance practice) for the notion of tail dependence. For such matrices of measures of concordance, we study their compatiblity and attainability. *Compatibility* concerns whether a given square matrix can be realized as a matrix of measures of concordance of some random vector, and *attainability* asks how to construct such a random vector. These notions are important in insurance and risk management practice since the entries of matrices of pairwise measures of concordance are often provided as estimates from real data (if available) or from expert opinion based on scenarios (if no data is available or not directly usable to estimate the entries). A primary issue is then to determine whether the given matrix is admissible as a matrix of pairwise measures of concordance and, if so, how an appropriate model may be built on the assumption of admissibility of the given matrix; see Embrechts et al. (2002) and (McNeil et al., 2015, Section 8.4) for a discussion on compatibility and attainability.

Note that compatibility is clear for Pearson's correlation coefficient since a given $[-1, 1]$-valued symmetric matrix P is compatible if and only if it is positive semi-definite and has diagonal entries equal to one. Also, attainability is clear for Pearson's correlation coefficient since any symmetric and positive semi-definite matrix P with ones on the diagonal is attainable by $\boldsymbol{X} = A\boldsymbol{Z}$ where $\boldsymbol{Z}$ is a random vector of independent standard normal distributions and A is the *Cholesky factor* of P, that is, a lower triangular matrix with non-negative diagonal entries and such that $P = AA^\top$.

Although compatibility and attainability of correlation matrices are thus trivial, the limitations of Pearson's correlation coefficient as a dependence measure are well known; see Embrechts et al. (2002). Measures of concordance in the sense of Scarsini (1984) are a remedy for some of the pitfalls of the correlation coefficient and are thus considered more suitable to summarize dependence between risks. Interestingly, such measures can also arise as correlations. Spearman's rho, Blomqvist's beta and van der Waerden's coefficient are prominent examples of measures arising as correlations of transformed variables of ranks.

Block matrices of measures of concordance naturally emerge if the risks of interest are grouped based on business line, industry, country, etc.; see, for example, Huang and Yang (2010). Hierarchical matrices are important special cases of block matrices where a measure of concordance between two variables is determined by an underlying hierarchical tree structure; see Hofert and Scherer (2011) for an application to CDO pricing. Since such matrices are typically high-dimensional, it is practically important to reduce the dimension to solve compatibility and attainability problems in this case.

In this chapter, we answer the following open questions, which naturally arise regarding compatibility and attainability of transformed rank correlation coefficients:

1. Are there more concordance measures which arise as correlations, and if so, how can they be characterized or constructed? (See Section 2.2)
2. What about the compatibility and attainability of matrices of such measures? (See Section 2.3)
3. Can compatibility and attainability be reduced to lower dimensional problems if a matrix has block structure? (See Section 2.4)

2.2 Correlation-based measures of concordance

We start by considering the bivariate case. To this end, let $X_1 \sim F_1$ and $X_2 \sim F_2$ be two continuously distributed random variables with a unique copula C such that $(U_1, U_2) = (F_1(X_1), F_2(X_2)) \sim C$. The measures of concordance of (X_1, X_2) we consider are of the form

$$\kappa_{g_1,g_2}(X_1, X_2) = \rho\big(g_1(F_1(X_1)), g_2(F_2(X_2))\big), \tag{2.1}$$

where $g_1 : [0,1] \to \mathbb{R}$ and $g_2 : [0,1] \to \mathbb{R}$ are measurable functions, and ρ is Pearson's correlation coefficient. Since (2.1) depends only on the copula of (X_1, X_2), we also denote it by $\kappa_{g_1,g_2}(C) = \rho(g_1(U_1), g_2(U_2))$ for $(U_1, U_2) \sim C$. We are interested in conditions on g_1 and g_2 under which (2.1) is a measure of concordance as defined in Definition 1.2.1 The following proposition provides a necessary condition on g_1 and g_2.

Proposition 2.2.1 (Monotonicity of g_1 and g_2)**.** Suppose $g_1 : [0,1] \to \mathbb{R}$ and $g_2 : [0,1] \to \mathbb{R}$ are continuous functions. If κ_{g_1,g_2} defined in (2.1) is a measure of concordance, then g_1 and g_2 must be both increasing or both decreasing, that is,

$$(g_1(u') - g_1(u))(g_2(v') - g_2(v)) \geq 0,$$

for any $0 \leq u < u' \leq 1$ and $0 \leq v < v' \leq 1$.

Proof. For $0 \leq u < u' \leq 1$ and $0 \leq v < v' \leq 1$, there exists a sufficiently large $N \in \mathbb{N}$ and indices $i, i', j, j' \in \{1, \ldots, N\}$ such that

$$\frac{i-1}{N} < u \leq \frac{i}{N}, \quad \frac{i'-1}{N} < u' \leq \frac{i'}{N}, \quad \frac{j-1}{N} < v \leq \frac{j}{N}, \quad \frac{j'-1}{N} < v' \leq \frac{j'}{N}$$

with $(\frac{i-1}{N}, \frac{i}{N}] \cap (\frac{i'-1}{N}, \frac{i'}{N}] = \emptyset$ and $(\frac{j-1}{N}, \frac{j}{N}] \cap (\frac{j'-1}{N}, \frac{j'}{N}] = \emptyset$. Let

$$\delta(x,y) = \begin{cases} 0, & (x,y) \in (\frac{i-1}{N}, \frac{i}{N}] \times (\frac{j-1}{N}, \frac{j}{N}] \cup (\frac{i'-1}{N}, \frac{i'}{N}] \times (\frac{j'-1}{N}, \frac{j'}{N}], \\ 2, & (x,y) \in (\frac{i-1}{N}, \frac{i}{N}] \times (\frac{j'-1}{N}, \frac{j'}{N}] \cup (\frac{i'-1}{N}, \frac{i'}{N}] \times (\frac{j-1}{N}, \frac{j}{N}], \\ 1, & \text{otherwise}, \end{cases}$$

and

$$\tilde{\delta}(x,y) = \begin{cases} 2, & (x,y) \in (\frac{i-1}{N}, \frac{i}{N}] \times (\frac{j-1}{N}, \frac{j}{N}] \cup (\frac{i'-1}{N}, \frac{i'}{N}] \times (\frac{j'-1}{N}, \frac{j'}{N}], \\ 0, & (x,y) \in (\frac{i-1}{N}, \frac{i}{N}] \times (\frac{j'-1}{N}, \frac{j'}{N}] \cup (\frac{i'-1}{N}, \frac{i'}{N}] \times (\frac{j-1}{N}, \frac{j}{N}], \\ 1, & \text{otherwise}, \end{cases}$$

and let Q_N and $\tilde{Q}_N$ be checkerboard copulas having densities δ and $\tilde{\delta}$, respectively; see Carley and Taylor (2002). Then $Q_N \preceq \tilde{Q}_N$ (in concordance order), since for any supermodular function ψ on $(0,1)^2$,

$$\begin{aligned} \int \psi \, \mathrm{d}\tilde{Q}_N - \int \psi \, \mathrm{d}Q_N &= \int \psi \, \mathrm{d}(\tilde{Q}_N - Q_N) \\ &= 2 \int_{(0,1/N)^2} (\psi(i' - 1 + s,\ j' - 1 + t) + \psi(i - 1 + s,\ j - 1 + t) \\ &\quad - \psi(i' - 1 + s,\ j - 1 + t) - \psi(i - 1 + s,\ j' - 1 + t)) \, \mathrm{d}s \, \mathrm{d}t \geq 0, \end{aligned}$$

where the last inequality follows since the integrand is nonnegative for any $(s,t) \in (0,1/N)^2$ by supermodularity of ψ. The inequality $\int \psi \,\mathrm{d}\tilde{Q}_N - \int \psi \,\mathrm{d}Q_N \geq 0$ for any supermodular function ψ implies $Q_N \preceq \tilde{Q}_N$; see Tchen et al. (1980) and Müller and Scarsini (2000). Since κ_{g_1,g_2} is a measure of concordance, coherence of κ_{g_1,g_2} implies that $\kappa_{g_1,g_2}(Q_N) \leq \kappa_{g_1,g_2}(\tilde{Q}_N)$, that is,

$$\begin{aligned} 0 \leq \kappa_{g_1,g_2}(\tilde{Q}_N) - \kappa_{g_1,g_2}(Q_N) &= \int_{(0,1)^2} g_1(U_1) g_2(U_2) \mathrm{d}(\tilde{Q}_N - Q_N) \\ &= 2 \int_{(0,1/N)^2} (g_1(i'-1+s) g_2(j'-1+t) + g_1(i-1+s) g_2(j-1+t) \\ &\quad - g_1(i'-1+s) g_2(j-1+t) - g_1(i-1+s) g_2(j'-1+t)) \,\mathrm{d}s \,\mathrm{d}t; \end{aligned}$$

see Scarsini (1984) for the coherence axiom of a measure of concordance. Since g_1 and g_2 are continuous, apply the intermediate value theorem and let $N \to \infty$ to obtain that

$$g_1(u')g_2(v') + g_1(u)g_2(v) - g_1(u')g_2(v) - g_1(u)g_2(v') = (g_1(u') - g_1(u))(g_2(v') - g_2(v)) \geq 0,$$

which shows that g_1 and g_2 are both increasing or both decreasing. □

By Proposition 2.2.1, g_1 and g_2 must share the same type of monotonicity so that κ_{g_1,g_2} is a measure of concordance. Therefore, it is reasonable to assume that g_1 and g_2 are both increasing functions on $[0,1]$ since, if both are decreasing, then $\kappa_{g_1,g_2} = \kappa_{\tilde{g}_1,\tilde{g}_2}$ for the increasing functions $\tilde{g}_1 = 1 - g_1$ and $\tilde{g}_2 = 1 - g_2$ by invariance of the correlation coefficient under linear transformations. If we relax the assumption of continuity of g_1 and g_2 to left-continuity, then g_1 and g_2 are quantiles of some distributions, say, G_1 and G_2. Recall that for a distribution function $G : \mathbb{R} \to [0,1]$, its *quantile function* is defined by

$$G^{-1}(p) = \inf\{x \in \mathbb{R} : G(x) \geq p\}, \quad p \in (0,1);$$

see Embrechts and Hofert (2013) for properties of G^{-1}. By taking $g_1 = G_1^{-1}$ and $g_2 = G_2^{-1}$, we now define the (G_1, G_2)-*transformed rank correlation coefficient* as follows.

Definition 2.2.2 ((G_1,G_2)-transformed rank correlation coefficient)**.** Let G_1 and G_2 be two distribution functions with quantile functions G_1^{-1} and G_2^{-1}, respectively. For a random vector (X_1, X_2) with continuous margins F_1 and F_2, the (G_1, G_2)-*transformed rank correlation coefficient* is defined by

$$\kappa_{G_1,G_2}(X_1,X_2) = \rho\big(G_1^{-1}(F_1(X_1)), G_2^{-1}(F_2(X_2))\big). \tag{2.2}$$

If $G_1 = G_2 = G$, $\kappa_{G,G}$ is denoted by κ_G and referred to as G-*transformed rank correlation coefficient*.

Example 2.2.3 (Known special cases of κ_{G_1,G_2})**.**

1. If G is the distribution function of the standard uniform distribution $\mathrm{Unif}(0,1)$, we obtain

$$\kappa_G(X_1,X_2) = \rho(F_1(X_1), F_2(X_2))$$

from (2.2). This is known as *Spearman's rho* ρ_S; see Spearman (1904).

2. If G is the distribution function of the symmetric Bernoulli distribution Bern(1/2), that is,

$$G(x) = \begin{cases} 0, & x < 0, \\ 1/2, & 0 \le x < 1, \\ 1, & x \ge 1, \end{cases}$$

then $G^{-1}(p) = \mathbf{1}_{\{1/2<p\le 1\}}$ for $p \in (0,1)$. Therefore, since $U_j = F_j(X_j) \sim \mathrm{Unif}(0,1)$ for $j = 1,2$, κ_{G_1,G_2} in (2.2) is the correlation coefficient of $B_j = G_j^{-1}(F_j(X_j)) \sim \mathrm{Bern}(1/2)$, $j = 1,2$. If C denotes the distribution function of (U_1, U_2) and $G_1 = G_2 = G$, then

$$\begin{aligned} \kappa_G(X_1, X_2) &= \frac{\mathbb{E}(B_1B_2) - \mathbb{E}(B_1)\mathbb{E}(B_2)}{\sqrt{\mathrm{Var}(B_1)\,\mathrm{Var}(B_2)}} = \frac{\mathbb{P}(U_1 > 1/2, U_2 > 1/2) - 1/4}{1/4} \\ &= 4\mathbb{P}(U_1 > 1/2, U_2 > 1/2) - 1 = 4(1 - 1/2 - 1/2 + C(1/2, 1/2)) - 1 \\ &= 4C(1/2, 1/2) - 1 \end{aligned}$$

which equals *Blomqvist's beta* β; see Blomqvist (1950). Note that Blomqvist's beta is also known as *median correlation coefficient*.

3. If G is the distribution function Φ of the standard normal distribution $\mathrm{N}(0,1)$, then

$$\kappa_G(X_1, X_2) = \rho\big(\Phi^{-1}(F_1(X_1)), \Phi^{-1}(F_2(X_2))\big)$$

which equals *van der Waerden's coefficient* ζ; see, for example, Sidak et al. (1999). It is also known as *normal score correlation*.

The first question in the introduction is natural: For which distributions G_1, G_2 does the G_1, G_2-transformed rank correlation κ_{G_1,G_2} lead to a measure of concordance in the sense of Scarsini (1984)? Before answering it, consider the following example in the spirit of Embrechts et al. (2002); another example of this type are the correlation bounds of Bernoulli random variables; see Example 2.2.7. Both examples show that G_1 and G_2 cannot be chosen arbitrarily.

Example 2.2.4 (Log-normal G_1, G_2-functions)**.** For $j = 1,2$, let $\sigma_j > 0$ and G_j be the distribution function of the log-normal distribution $\mathrm{LN}(0, \sigma_j)$. Since κ_{G_1,G_2} is the correlation coefficient of the random vector $(G_1^{-1}(U_1), G_2^{-1}(U_2))$ with $(U_1, U_2) = (F_1(X_1), F_2(X_2))$, its minimal and maximal values are attained when (X_1, X_2) has copula $C = W$ and $C = M$, respectively, where $W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$ is the countermonotone and $M(u_1, u_2) = \min\{u_1, u_2\}$ is the comonotone copula. For different pairs of (σ_1, σ_2), the minimal and maximal (G_1, G_2)-transformed rank correlation coefficients are shown in Figure 2.1 as correlation coefficients of $\mathrm{LN}(0, \sigma_1)$ and $\mathrm{LN}(0, \sigma_2)$. The left-hand side of this figure shows that $\kappa_{G_1,G_2} = -1$ is not attained for any $\sigma_1, \sigma_2 > 0$ and the right-hand side shows that $\kappa_{G_1,G_2} = 1$ is not attained unless $\sigma_1 = \sigma_2$. Consequently, if G_1, G_2 are taken to be log-normal distribution functions, κ_{G_1,G_2} cannot be a measure of concordance since the range axiom of Definition 1.2.1 is violated.

The main result of this section is the following, which provides necessary and sufficient conditions for a transformed rank correlation coefficient to be a measure of concordance in the sense of Scarsini (1984). Recall that two distributions are *of the same type* if one is a location-scale transform of the other.

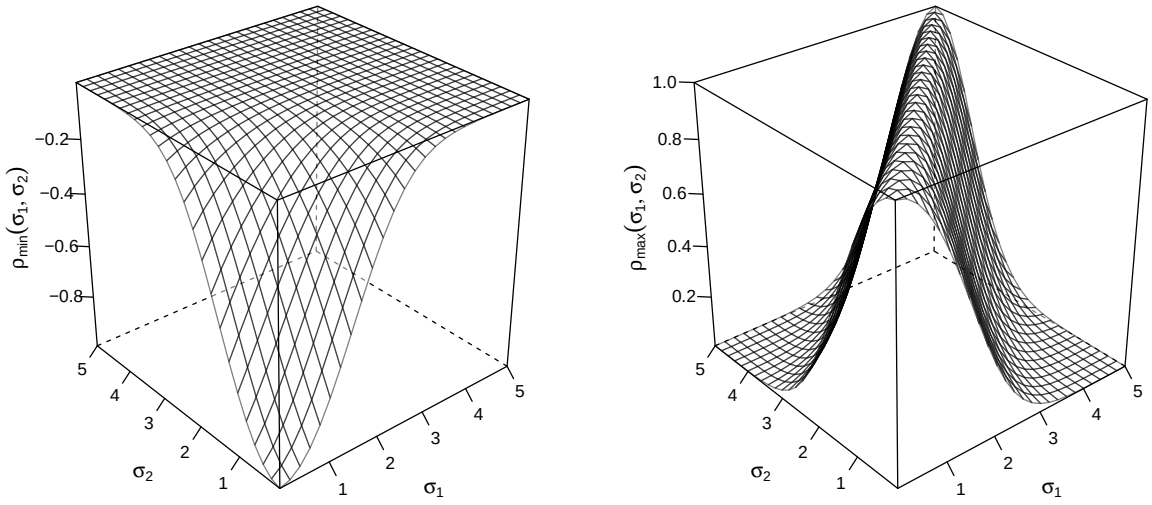

Figure 2.1: Minimal (left) and maximal (right) correlations attained by the (G_1, G_2)-transformed rank correlation coefficient κ_{G_1,G_2} where G_j is the distribution function of $\mathrm{LN}(0, \sigma_j)$, $j = 1, 2$.

Theorem 2.2.5 (Necessary and sufficient conditions for transformed rank correlations to be measures of concordance)**.** Let G_1, G_2 be distribution functions. The (G_1, G_2)-transformed rank correlation coefficient κ_{G_1,G_2} in (2.2) is a measure of concordance if and only if both G_1 and G_2 are of the same type as some non-degenerate symmetric distribution G with finite second moment.

Proof. Let $(X_1, X_2) \sim H$ with copula C and continuous margins F_1, F_2. Then $(U_1, U_2) = (F_1(X_1), F_2(X_2)) \sim C$ so that $(Y_1, Y_2) = (G_1^{-1}(U_1), G_2^{-1}(U_2))$ has copula C and marginal distribution functions G_1, G_2. The transformed rank correlation coefficient $\kappa_{G_1,G_2}(X_1, X_2)$ in (2.2) can then be written as $\kappa_{G_1,G_2}(X_1, X_2) = \rho(Y_1, Y_2)$.

Consider necessity. If either of G_1 and G_2 is degenerate, then $\rho(Y_1, Y_2)$ is not well-defined, which violates the domain axiom of a measure of concordance. Therefore, G_1 and G_2 must be non-degenerate. Next, if either of $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$ is infinite, then $\rho(Y_1, Y_2)$ is not defined, which also violates the domain axiom. Thus, G_1 and G_2 must have finite second moments. For $j = 1, 2$, let $\mu_j = \mathbb{E}(Y_j)$ and $\sigma_j^2 = \mathrm{Var}(Y_j) < \infty$. It is known that $\rho(Y_1, Y_2) = -1$ if and only if $Y_2 \overset{\mathrm{d}}{=} -aY_1 + b$ for some $a, b \in \mathbb{R}$ with $a > 0$ and $\rho(Y_1, Y_2) = 1$ if and only if $\boldsymbol{Y}_2 \overset{\mathrm{d}}{=} cY_1 + d$ for some $c, d \in \mathbb{R}$ with $c > 0$. Note that both distributional equalities must hold simultaneously so that $\kappa_{G_1,G_2}(X_1, X_2) = 1$ when (X_1, X_2) is comonotone and $\kappa_{G_1,G_2}(X_1, X_2) = -1$ when (X_1, X_2) is countermonotone. Since $\sigma_2^2 = a^2\sigma_1^2 = c^2\sigma_1^2$, $a, c > 0$ and $\sigma_1 \neq 0$, we have $a = c$. Furthermore, by taking expectations, $\mu_2 = -c\mu_1 + b$ and $\mu_2 = c\mu_1 + d$, which imply that $\mu_1 = (b-d)/(2c)$ and $\mu_2 = (b+d)/2$. Since $Y_2 - b \overset{\mathrm{d}}{=} -cY_1 \overset{\mathrm{d}}{=} d - Y_2$, adding constant $(b-d)/2$ to both hand sides yield $Y_2 - \mu_2 \overset{\mathrm{d}}{=} \mu_2 - Y_2$. This implies that Y_2 is symmetric about its mean μ_2. Similarly, Y_1 is shown to be symmetric about its mean μ_1. Finally, it follows from $Y_2 \overset{\mathrm{d}}{=} cY_1 + d$ that $G_2(x) = G_1((x-d)/c)$ and thus $G_2^{-1}(u) = d + c\, G_1^{-1}(u)$, which concludes the proof of necessity.

Now consider sufficiency. If G_1 and G_2 are of the same type as some distribution G, then $\kappa_{G_1,G_2}(C) = \kappa_{G,G}(C) = \kappa_G(C)$ for any copula C since correlation coefficient is invariant under positive linear transform;

see Embrechts et al. (2002). Therefore, it suffices to verify the seven axioms of a measure of concordance in Scarsini (1984) for κ_G with G being a non-degenerate symmetric distribution with finite second moment.

1. *Domain:* Since G is non-degenerated with a finite second moment, $\rho(Y_1, Y_2)$ is well-defined for all continuously distributed X_1, X_2.

2. *Symmetry:* To show $\kappa_G(X_1, X_2) = \kappa_G(X_2, X_1)$, it suffices to show

$$\mathbb{E}(G^{-1}(U_1)G^{-1}(U_2)) = \mathbb{E}(G^{-1}(U_2)G^{-1}(U_1))$$

   for any C and $(U_1, U_2) \sim C$, but this is obvious by exchangeability of product.

3. *Coherence:* Let C_1, C_2 be copulas such that $C_1 \preceq C_2$, that is, $C_1(u_1, u_2) \leq C_2(u_1, u_2)$ for all $u_1, u_2 \in [0, 1]$. Then $\kappa_G(C_1) \leq \kappa_G(C_2)$ follows immediately from the Hoeffding's identity; see (McNeil et al., 2015, Lemma 7.27).

4. *Range:* Since $\kappa_G(X_1, X_2) = \rho(Y_1, Y_2)$, we have $-1 \leq \kappa_G(X_1, X_2) \leq 1$. Moreover, since G is symmetric, we have $Y_1 - \mathbb{E}[Y_1] \overset{\mathrm{d}}{=} \mathbb{E}[Y_2] - Y_2$. Together with $Y_1 \overset{\mathrm{d}}{=} Y_2$, the bounds $\kappa_G(X_1, X_2) = -1$ and $\kappa_G(X_1, X_2) = 1$ are attainable when (X_1, X_2) are countermonotone and comonotone, respectively.

5. *Independence:* When X_1, X_2 are independent, so are Y_1, Y_2 and thus $\kappa_G(X_1, X_2) = \rho(Y_1, Y_2) = 0$.

6. *Change of sign:* Let F_{-X_2} be the distribution of $-X_2$. Then it holds that $F_{-X_2}(-x_2) = \mathbb{P}(X_2 > x_2) = 1 - F_2(x_2)$ and thus $F_{-X_2}(-X_2) = 1 - F_2(X_2) = 1 - U_2$. Symmetry of G implies that $G(y) = 1 - G(2\mu_2 - y)$ for $y \in \mathbb{R}$ and thus $G^{-1}(1-p) = 2\mu_2 - G^{-1}(p)$ for $p \in (0, 1)$. Therefore,

$$\begin{aligned}
\kappa(X_1, -X_2) &= \rho\big(G^{-1}(F_{X_1}(X_1)), G^{-1}(F_{-X_2}(-X_2))\big) = \rho(G^{-1}(U_1), G^{-1}(1 - U_2)) \\
&= \rho(G^{-1}(U_1), 2\mu_2 - G^{-1}(U_2)) = \rho(G^{-1}(U_1), -G^{-1}(U_2)) \\
&= -\rho(G_1^{-1}(U_1), G_2^{-1}(U_2)) = -\kappa(X_1, X_2)
\end{aligned}$$

   by invariance and change of sign properties of correlation coefficient.

7. *Continuity:* Let $(X_{n1}, X_{n2}) \sim H_n$, $n \in \mathbb{N}$, and $(X_1, X_2) \sim H$ all have continuous margins with H_n converging pointwise to H as $n \to \infty$. Let C_n denote the copula of H_n, $n \in \mathbb{N}$, and C the one of H. Then $\lim_{n\to\infty} C_n = C$ pointwise. Since $\kappa(X_{n1}, X_{n2})$ and $\kappa(X_1, X_2)$ are correlation coefficients of (Y_{n1}, Y_{n2}) and (Y_1, Y_2) having the same marginal distribution G and copulas C_n and C, respectively, Hoeffding's identity yields that

$$\begin{aligned}
\lim_{n\to\infty} \kappa(X_{n1}, X_{n2}) &= \lim_{n\to\infty} \frac{1}{\sigma_1\sigma_2} \int_{\mathbb{R}^2} (C_n(G(y_1), G(y_2)) - G(y_1)G(y_2)) \,\mathrm{d}\lambda_2(y_1, y_2) \\
&= \frac{1}{\sigma_1\sigma_2} \int_{\mathbb{R}^2} (C(G(y_1), G(y_2)) - G(y_1)G(y_2)) \,\mathrm{d}\lambda_2(y_1, y_2) = \kappa(X_1, X_2), \qquad (2.3)
\end{aligned}$$

   for the Lebesgue measure λ_2 on $\mathbb{R}^2$, where the second equality is justified by the bounded convergence theorem since $C_n(G(y_1), G(y_2)) - G(y_1)G(y_2)$ and $C(G(y_1), G(y_2)) - G(y_1)G(y_2)$ are all uniformly bounded.

□

As seen in the proof of Theorem 2.2.5, if κ_{G_1,G_2} is a measure of concordance, then it must be κ_G for some distribution G which is of the same type as G_1 and G_2. In what follows, we thus focus on G-transformed rank correlation coefficients for which we assume that $G_1 = G_2$.

Remark 2.2.6 (Connection to D_4-invariant measures of concordance)**.** From (2.3) it turns out that (G_1, G_2)-transformed rank correlations κ_{G_1,G_2} form a subclass of D_4-invariant measures of concordance as proposed by Edwards et al. (2005). A measure ν on $(0,1)^2$ is called D_4-*invariant* if it is invariant under transpositions $(x, y) \mapsto (y, x)$ and partial reflections $(x, y) \mapsto (1 - x, y)$. For such measures ν, Edwards et al. (2005) show that the functional

$$C \mapsto \frac{\int_{(0,1)^2}(C - \Pi)\,\mathrm{d}\nu}{\int_{(0,1)^2}(M - \Pi)\,\mathrm{d}\nu} \tag{2.4}$$

is a measure of concordance, where M is the comonotone copula and Π is the independence copula. When G_1 and G_2 are symmetric, the pushforward Lebesgue measure λ_{G_1,G_2} is D_4-invariant and the corresponding measure (2.4) yields our (G_1, G_2)-transformed rank correlation (2.2). Consequently, the sufficiency part of the proof of Theorem 2.2.5 also follows from (Edwards et al., 2005, Theorem 0.6).

According to Theorem 2.2.5, we call a distribution function G *concordance-inducing* if it is non-degenerate, symmetric and has finite second moment. Examples of such distributions include normal, Student t with degrees of freedom $\nu > 2$, continuous and discrete uniform distributions, Laplace and logistic distributions. The following example shows that Bernoulli distributions $\mathrm{Bern}(p)$ are concordance-inducing if and only if they are symmetric, that is, $p = 1/2$.

Example 2.2.7 (Bernoulli G-function)**.** For $j = 1, 2$, let $p_j \in [0, 1]$ and G_j be the distribution of $Y_j \sim \mathrm{Bern}(p_j)$. As discussed in Example 2.2.4, $\kappa_{G_1,G_2}(X_1, X_2) = \rho(Y_1, Y_2)$ and its minimal and maximal values are attained when $C = W$ and $C = M$, respectively. Figure 2.2 illustrates the minimal (left-hand side) and maximal (right-hand side) (G_1, G_2)-transformed rank correlation coefficients as correlations of $\mathrm{Bern}(p_1)$ and $\mathrm{Bern}(p_2)$ for different pairs of (p_1, p_2). The left-hand side of the figure indicates that $\kappa_{G_1,G_2} = -1$ if $p_1 = 1 - p_2$ and this is the only case when Y_1 and $-Y_2$ are of the same type. The right-hand side shows that $\kappa_{G_1,G_2} = 1$ if $p_1 = p_2$, and this is the only case when Y_1 and Y_2 have the same distribution. Since κ_{G_1,G_2} must attain -1 and 1 when $C = W$ and $C = M$, respectively, κ_{G_1,G_2} is a measure of concordance only when $p_1 = p_2 = 1/2$. As a consequence, $\mathrm{Bern}(p)$ is concordance-inducing if and only if $p = 1/2$.

Note that due to the invariance of the correlation coefficient under strictly increasing linear transforms, κ_G is invariant under location-scale transforms of $Y \sim G$. Therefore, if G has bounded support, it may be beneficial to standardize it so that its support is $[0, 1]$. Similarly, if G is supported on $\mathbb{R}$, one can still standardize G to have zero mean and unit variance without changing κ_G. Due to this property, one can see that the quadrant correlation of Mosteller (2006) studied in Raymaekers and Rousseeuw (2019) coincides with Blomqvist's beta.

Uniqueness of G-function up to location-scale transformations follows direclty from (Edwards et al., 2004, Lemma 2.4) or (Edwards et al., 2005, Lemma 0.4).

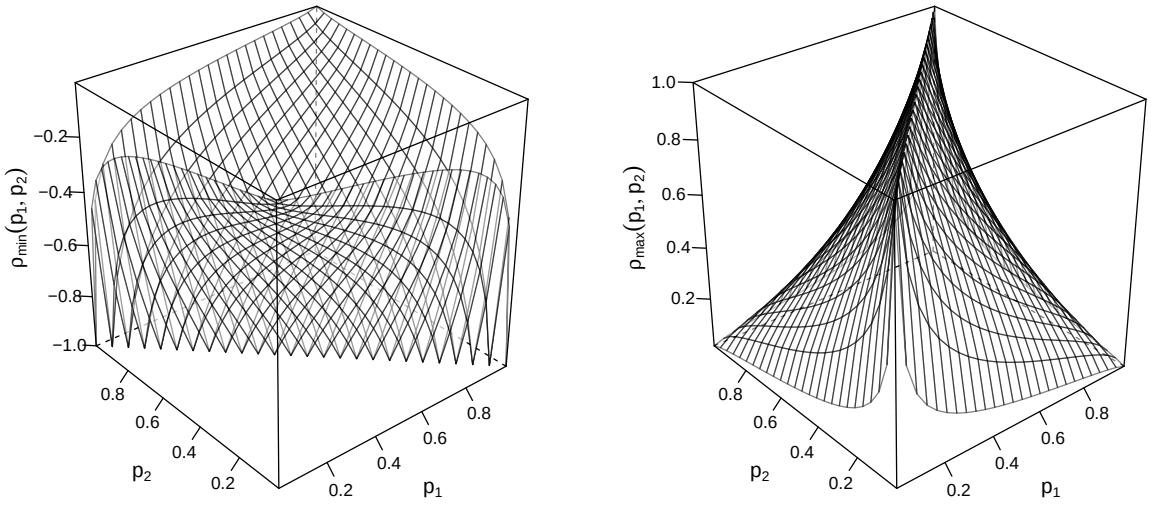

Figure 2.2: Minimal (left) and maximal (right) correlations attained by the (G_1, G_2)-transformed rank correlation coefficient κ_{G_1,G_2} where G_j is the distribution function of $\mathrm{Bern}(p_j)$, $j = 1, 2$.

Proposition 2.2.8 (Uniqueness of G-functions)**.** Let G and G' be two continuous concordance-inducing functions. If $\kappa_G(C) = \kappa_{G'}(C)$ for all 2-copulas, then G and G' are of the same type.

We end this section with a simple linear property of κ_G.

Proposition 2.2.9 (Linearity of κ_G)**.** For $n \in \mathbb{N}$, let $C_1, \dots, C_n$ be 2-copulas and $\alpha_1, \dots, \alpha_n$ be non-negative numbers such that $\alpha_1 + \cdots + \alpha_n = 1$. Then

$$\kappa_G\left(\sum_{i=1}^{n} \alpha_i C_i\right) = \sum_{i=1}^{n} \alpha_i \kappa_G(C_i).$$

Proof. As a mixture, $\sum_{i=1}^{n} \alpha_i C_i$ is a 2-copula from which the equation to prove is an immediate consequence of Hoeffding's identity. □

Remark 2.2.10 (Degree of κ_G)**.** For a general measure of concordance κ, Edwards and Taylor (2009) defined the notion of a *degree* as the maximum degree of the polynomial $t \mapsto \kappa(tC_1 + (1-t)C_2)$, when it is the case, over any two copulas C_1 and C_2. Proposition 2.2.9 shows that κ_G is a measure of concordance of degree one in this sense. Also note that the class of G-transformed rank correlation coefficients is a strict subclass of all measures of concordance of degree one since, for instance, *Gini's coefficient*

$$\gamma(C) = 4 \int_{[0,1]^2} (M(u,v) + W(u,v)) \, \mathrm{d}C(u,v) - 2, \tag{2.5}$$

is of degree one but cannot be represented as (2.2). To see this, note that the G-transformed rank correlation coefficient can be written as

$$k_G(C) = \frac{1}{\sigma^2} \int_{[0,1]^2} G^{-1}(u) G^{-1}(v) \, \mathrm{d}C(u,v) - \left(\frac{\mu}{\sigma}\right)^2, \tag{2.6}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the mean and standard deviation of G, respectively. Expression (2.6) implies that the integrand with respect to the underlying copula C must be of product form $G^{-1}(u)G^{-1}(v)$. Since the integrand in (2.5) cannot be decomposed into such a product form in general, Gini's γ is not a G-transformed rank correlation coefficient. Furthermore, there is no G-function that makes κ_G *Kendall's tau*

$$\tau(X_1, X_2) = 4 \int_{[0,1]^2} C(u, v) \, \mathrm{d}C(u, v) - 1, \tag{2.7}$$

since τ is a measure of concordance of degree two according to Edwards and Taylor (2009).

2.3 Matrices of transformed rank correlation coefficients and their compatibility

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector with continuous margins $F_1, \ldots, F_d$ and copula C. We now consider matrices of (pairwise) G-transformed rank correlation measures, that is, matrices $P \in [-1, 1]^{d \times d}$ with (i, j)th entry given by $\kappa_G(X_i, X_j)$. As in Theorem 2.2.5, G is set to be a distribution function of a non-degenerate, symmetric distribution with finite second moment. We call a given matrix $P \in [-1, 1]^{d \times d}$ *κ_G-compatible* if there exists a d-random vector $\boldsymbol{X}$ such that $P = (\kappa_G(X_i, X_j))$. In this section, we first study this *compatibility problem* for the transformed rank correlation coefficient (2.2) in general and then more specifically for Spearman's rho, Blomqvist's beta and van der Waerden's coefficient. Note that an obvious necessary condition for a given matrix P to be κ_G-compatible is that it is a $[-1, 1]^{d \times d}$ symmetric, positive semi-definite matrix with diagonal elements equal to 1.

2.3.1 A sufficient condition for compatibility of transformed rank correlation coefficients

For a fixed concordance-inducing function G, denote by $\mathcal{K}_G$ the set of all κ_G-compatible matrices. Since $\kappa_G(X_i, X_j) = \rho(Y_i, Y_j)$ with the notation as before, $\mathcal{K}_G$ can be written as

$$\mathcal{K}_G = \{\rho(\boldsymbol{Y}) \mid \boldsymbol{Y} \in \mathcal{F}_d(G, \ldots, G)\},$$

where $\mathcal{F}_d(G, \ldots, G)$ denotes the set of all d-dimensional random vectors with all marginals equal to G. The following corollary follows directly from Proposition 2.2.9.

Corollary 2.3.1 (Convexity of $\mathcal{K}_G$)**.** *$\mathcal{K}_G$ is a convex set for any concordance-inducing function G.*

Let

$$\mathcal{P}_d^{\mathrm{B}}(1/2) = \{\rho(\boldsymbol{B}) : \boldsymbol{B} = (B_1, \ldots, B_d), \ B_j \sim \mathrm{Bern}(1/2), \ j = 1, \ldots, d\}$$

be the set of all correlation matrices of d-dimensional random vectors whose marginals are symmetric Bernoulli distributions. The following proposition provides a sufficient condition for a given matrix to be κ_G-compatible.

Proposition 2.3.2 (A sufficient condition for κ_G-compatibility)**.** For a concordance-inducing function G, it holds that $\mathcal{P}_d^{\mathrm{B}}(1/2) \subseteq \mathcal{K}_G$, that is, a given matrix $P \in [-1,1]^{d\times d}$ is κ_G-compatible if it is a correlation matrix of some random vector with Bern$(1/2)$ margins.

Proof. Fix $P \in \mathcal{P}_d^{\mathrm{B}}(1/2)$. Then there exist $B_1, \ldots, B_d \sim \mathrm{Bern}(1/2)$ such that $\rho(\boldsymbol{B}) = P$ for $\boldsymbol{B} = (B_1, \ldots, B_d)$. For $U \sim \mathrm{Unif}(0,1)$ independent of $\boldsymbol{B}$, define

$$V_j = B_j U + (1 - B_j)(1 - U), \quad j = 1, \ldots, d.$$

Then $V_j \sim \mathrm{Unif}(0,1)$ and thus $Y_j = G^{-1}(V_j) \sim G$, $j = 1, \ldots, d$. Note that $Y_j = G^{-1}(U)$ if $B_j = 1$ and $Y_j = G^{-1}(1-U)$ if $B_j = 0$. Furthermore, since G is concordance-inducing,

$$\rho(G^{-1}(U), G^{-1}(U)) = 1 \quad \text{and} \quad \rho(G^{-1}(U), G^{-1}(1-U)) = -1.$$

Consequently, for all $i, j \in \{1, \ldots, d\}$,

$$\begin{aligned}\rho(Y_i, Y_j) &= \rho(G^{-1}(U), G^{-1}(U))\mathbb{P}(B_i = B_j) + \rho(G^{-1}(U), G^{-1}(1-U))\mathbb{P}(B_i \neq B_j) \\ &= \mathbb{P}(B_i = B_j) - \mathbb{P}(B_i \neq B_j) = 2\mathbb{P}(B_i = B_j) - 1.\end{aligned}$$

Since

$$\begin{aligned}\mathbb{P}(B_i = B_j) &= \mathbb{P}(B_i = 0, B_j = 0) + \mathbb{P}(B_i = 1, B_j = 1) \\ &= \mathbb{P}(1 - B_i = 1, 1 - B_j = 1) + \mathbb{E}(B_i B_j) \\ &= \mathbb{E}((1 - B_i)(1 - B_j)) + \mathbb{E}(B_i B_j) = 2\mathbb{E}(B_i B_j) \\ &= \frac{\rho(B_i, B_j) + 1}{2},\end{aligned}$$

we obtain

$$\rho(Y_i, Y_j) = 2\frac{\rho(B_i, B_j) + 1}{2} - 1 = \rho(B_i, B_j)$$

and thus $P = \rho(\boldsymbol{B}) = \rho(\boldsymbol{Y}) \in \mathcal{K}_G$. □

Note that the construction $Y_j = G^{-1}(B_j U + (1 - B_j)(1 - U))$, $j = 1, \ldots, d$, used in the proof of Proposition 2.3.2 was utilized by Huber and Maric (2015) for the purpose of generating a d-dimensional distribution with given margins G and a correlation matrix P where $P \in \mathcal{P}_d^{\mathrm{B}}(1/2)$.

By Proposition 2.3.2, a given matrix is found to be κ_G-compatible if it belongs to $\mathcal{P}_d^{\mathrm{B}}(1/2)$. The relationship between $\mathcal{K}_G$ and $\mathcal{P}_d^{\mathrm{B}}(1/2)$ depends on the G-function. When G is a symmetric Bernoulli distribution, it holds that $\mathcal{K}_G = \mathcal{P}_d^{\mathrm{B}}(1/2)$, whereas if G is the standard normal distribution function Φ, then $\mathcal{K}_G$ coincides with the set of all correlation matrices $\mathcal{P}_d$, which is strictly larger than $\mathcal{P}_d^{\mathrm{B}}(1/2)$; see Proposition 2.3.4 Part 4 for $\mathcal{K}_\Phi = \mathcal{P}_d$ and Section 2.3.3 for $\mathcal{P}_d^{\mathrm{B}}(1/2) \subset \mathcal{P}_d$. As summarized by the following corollary, $\mathcal{P}_d^{\mathrm{B}}(1/2)$ and $\mathcal{P}_d$ are the smallest and largest set of κ_G compatible matrices for general G.

Corollary 2.3.3 (Upper and lower bounds of $\mathcal{K}_G$)**.** For any concordance-inducing function G, the set of all κ_G-compatible matrices $\mathcal{K}_G$ satisfies $\mathcal{P}_d^{\mathrm{B}}(1/2) \subseteq \mathcal{K}_G \subseteq \mathcal{P}_d$, and both bounds are attainable.

Note that the uniqueness of G attaining the bounds fails and possibly depends on d. For example, when $d \leq 9$, both of $G = \mathrm{Unif}(0,1)$ and Φ attain $\mathcal{K}_G = \mathcal{P}_d$; see Proposition 2.3.4 Part 1, 2 and 4.

We have so far found that the set $\mathcal{P}_d^{\mathrm{B}}(1/2)$ plays important roles for κ_G-compatibility problem. Natural questions regarding $\mathcal{P}_d^{\mathrm{B}}(1/2)$ are how to check a given matrix belongs to $\mathcal{P}_d^{\mathrm{B}}(1/2)$ and how large the set is in comparison to the set of all correlation matrices $\mathcal{P}_d$. These questions will be answered in Section 2.3.3.

2.3.2 Characterizations of specific measures of concordance

In this section, we study the three specific measures of concordance from Example 2.2.3, Spearman's rho, Blomqvist's beta and van der Waerden's coefficient, which are denoted by ρ_{S}, β and ζ, respectively. To this end, let $\mathcal{S}_d$, $\mathcal{B}_d$ and $\mathcal{W}_d$ be the set of $d \times d$-matrices of Spearman's rho, Blomqvist's beta and van der Waerden's coefficients, respectively. As is done in the previous section, denote by $\mathcal{P}_d$ the set of all $d \times d$-correlation matrices, that is, the set of all symmetric, positive semi-definite matrices in $[-1,1]^d$ with diagonal elements one. It is well-known that $\mathcal{P}_d$ is a convex set for any $d \geq 1$. Let $\mathcal{P}_d^{\mathrm{U}}$ and $\mathcal{P}_d^{\mathrm{B}}(p)$, $p \in (0,1)$, be the set of all correlation matrices of d-dimensional random vectors whose marginals are all $\mathrm{Unif}(0,1)$ and all $\mathrm{Bern}(p)$, respectively. By Corollary 2.3.1, $\mathcal{P}_d^{\mathrm{U}}$ and $\mathcal{P}_d^{\mathrm{B}}(p)$ are also convex sets. We can now characterize the sets $\mathcal{S}_d$, $\mathcal{B}_d$ and $\mathcal{W}_d$.

Proposition 2.3.4 (Characterizations of $\mathcal{S}_d$, $\mathcal{B}_d$ and $\mathcal{W}_d$)**.**

1. $\mathcal{P}_d^{\mathrm{U}} = \mathcal{P}_d$ for $d \leq 9$, that is, the set of correlation matrices of random vectors with standard uniform marginals coincides with the set of correlation matrices for $d \leq 9$. For $d \geq 10$, $\mathcal{P}_d^{\mathrm{U}} \subseteq \mathcal{P}_d$.
2. $\mathcal{S}_d = \mathcal{P}_d^{\mathrm{U}}$, that is, the set of Spearman's rho matrices coincides with the set of correlation matrices of random vectors with standard uniform marginals.
3. $\mathcal{B}_d = \mathcal{P}_d^{\mathrm{B}}(1/2)$, that is, the set of Blomqvist's beta matrices coincides with the set of correlation matrices of random vectors with symmetric Bernoulli marginals.
4. $\mathcal{W}_d = \mathcal{P}_d$, that is, the set of van der Waerden's matrices coincides with the set of all correlation matrices.

Proof. Part 1 is from Devroye and Letac (2015), and Part 2 and Part 4 are direct consequences of the definition of Spearman's rho and van der Waerden's coefficient. We thus have left to prove Part 3. Consider "$\subseteq$". Let $(\beta_{ij}) \in \mathcal{B}_d$. Then there exists a d-dimensional random vector $\boldsymbol{X}$ such that $\beta(X_i, X_j) = \beta_{ij}$. By Example 2.2.3 Part 2,

$$\beta_{ij} = \rho(G^{-1}(F_i(X_i)), G^{-1}(F_j(X_j))), \quad i,j = 1,\dots,d,$$

where G is the distribution function of $\mathrm{Bern}(1/2)$. Since $G^{-1}(F_i(X_i)), G^{-1}(F_j(X_j)) \sim \mathrm{Bern}(1/2)$, we obtain that $(\beta_{ij}) \in \mathcal{P}_d^{\mathrm{B}}(1/2)$.

Now consider "$\supseteq$". Let $\boldsymbol{B} = (B_1,\dots,B_d)$ be a d-dimensional symmetric Bernoulli random vector with correlation matrix $\rho(\boldsymbol{B}) = (\rho_{ij})$. Let C be any copula such that

$$\mathbb{P}(B_1 \leq b_1,\dots,B_d \leq b_d) = C(\mathbb{P}(B_1 \leq b_1),\dots,\mathbb{P}(B_d \leq b_d)).$$

Since, for $j = 1, \ldots, d$,

$$\mathbb{P}(B_j \le b_j) = \begin{cases} 0, & \text{if } b_j < 0, \\ 1/2, & \text{if } 0 \le b_j < 1, \\ 1, & \text{if } b_j \ge 1, \end{cases}$$

C is only uniquely determined in $(1/2, \ldots, 1/2)$ inside $[0,1]^d$. Furthermore, for any $(j_1, \ldots, j_d) \in \{0,1\}^d$, the following identity holds:

$$C((1/2)^{j_1}, \ldots, (1/2)^{j_d}) = \mathbb{P}(B_1 \le 1 - j_1, \ldots, B_d \le 1 - j_d).$$

Let $\overline{C}$ be the survival function of C and $\boldsymbol{U} \sim \overline{C}$, so $\mathbf{1} - \boldsymbol{U} \sim C$; in particular, the marginals $F_1, \ldots, F_d$ of $\boldsymbol{U}$ are $\mathrm{Unif}(0,1)$. Let $G(p) = \mathbf{1}_{\{p > 1/2\}}$ be the distribution function of the symmetric Bernoulli distribution. Then

$$\begin{aligned} &\mathbb{P}(G^{-1}(U_1) \le 1 - j_1, \ldots, G^{-1}(U_d) \le 1 - j_d) \\ &= \mathbb{P}(\mathbf{1}_{\{U_1 > 1/2\}} \le 1 - j_1, \ldots, \mathbf{1}_{\{U_d > 1/2\}} \le 1 - j_d) \\ &= \mathbb{P}(1 - U_1 \le (1/2)^{j_1}, \ldots, 1 - U_d \le (1/2)^{j_d}) = C((1/2)^{j_1}, \ldots, (1/2)^{j_d}) \\ &= \mathbb{P}(B_1 \le 1 - j_1, \ldots, B_d \le 1 - j_d), \quad (j_1, \ldots, j_d) \in \{0,1\}^d. \end{aligned}$$

Therefore, we have that $\boldsymbol{B} = (B_1, \ldots, B_d) \overset{\mathrm{d}}{=} (G^{-1}(U_1), \ldots, G^{-1}(U_d))$. Consequently,

$$\beta(U_i, U_j) = \rho(G^{-1}(F_i(U_i)), G^{-1}(F_j(U_j))) = \rho(G^{-1}(U_i), G^{-1}(U_j)) = \rho(B_i, B_j) = \rho_{ij}.$$

Since the random vector $\boldsymbol{U}$ attains (ρ_{ij}) as its Blomqvist's beta matrix, we have $(\rho_{ij}) \in \mathcal{B}_d$. □

Concerning Proposition 2.3.4 Part 1, Devroye and Letac (2015) conjectured that the inclusion relationship among $\mathcal{P}_d^{\mathrm{U}}$ and $\mathcal{P}_d$ is strict for $d \ge 10$. Later Wang et al. (2019) revealed that $\mathcal{P}_d$ is strictly larger than $\mathcal{P}_d^{\mathrm{U}}$ for $d \ge 12$. Although a complete characterization of $\mathcal{P}_d^{\mathrm{U}}$ is still unknown for $d \ge 10$, it is known that $\mathcal{P}_d^{\mathrm{U}}$ and $\mathcal{P}_d$ are not significantly different for any $d \ge 1$ as explained in the following remark.

Remark 2.3.5 ($\mathcal{S}_d$ and $\mathcal{P}_d$)**.** Even for $d \ge 10$, $\mathcal{S}_d$ and $\mathcal{P}_d$ cannot be largely different since a Gauss copula with correlation parameter $P = (\rho_{ij}) \in \mathcal{P}_d$ has Spearman's rho matrix $(\rho_{\mathrm{S},ij})$ with $\rho_{\mathrm{S},ij} = (6/\pi)\arcsin(\rho_{ij}/2)$, or equivalently, $\rho_{ij} = 2\sin(\pi\rho_{\mathrm{S},ij}/6)$. Since $|\rho_{\mathrm{S},ij} - \rho_{ij}| = |\rho_{\mathrm{S},ij} - 2\sin(\pi\rho_{\mathrm{S},ij}/6)| \le 0.0181$, one can find an elementwise close Spearman's rho matrix attained by a Gauss copula for every correlation matrix $P \in \mathcal{P}_d$.

The consequences of Proposition 2.3.4 related to the compatibility problem are as follows. First, Proposition 2.3.4 Part 1 and 2 allow one to check that a given $d \times d$-matrix for $d \le 9$ is ρ_{S}-compatible via checking whether the matrix is a correlation matrix, for example, by trying to compute its Cholesky factor. For $d \ge 10$, a straightforward way to check ρ_{S}-compatibility is not available yet although the sufficient condition in Proposition 2.3.2 is still valid. Second, Proposition 2.3.4 Part 3 states that the set of all Blomqvist's beta matrices are completely characterized by the set of correlation matrices of random vectors with symmetric Bernoulli margins. In Section 2.3.3, we will discuss the problem of checking that a given

matrix belongs to $\mathcal{P}_d^{\mathrm{B}}(1/2)$. Finally, Proposition 2.3.4 Part 4 says that the set of van der Waerden's matrices coincides with the set of all correlation matrices, and thus, checking ζ-compatibility is straightforward. In terms of checking compatibility, this property of van der Waerden's coefficient is an attractive feature that ρ_S and β do not satisfy for any dimension $d \geq 1$. Note that this property is not unique to van der Waerden's coefficient but holds for any elliptical distribution G with finite second moments; see (Joe, 1997, Chapter 4).

2.3.3 Bern(1/2)-compatibility problem

As we have seen in Section 2.2 and 2.3 so far, $\mathcal{P}_d(1/2)$ plays important roles when studying matrix compatibility problems since it coincides with $\mathcal{B}_d$, the set of all Blomqvist's beta matrices, and $\mathcal{P}_d(1/2) \subseteq \mathcal{K}_G$, the set of all κ_G-compatible matrices. If $P \in \mathcal{P}_d(1/2)$, we call P Bern(1/2)-*compatible*. In this section, we address the *membership testing problem* for $\mathcal{P}_d(1/2)$, that is, a test whether a given matrix is Bern(1/2)-compatible or not.

Huber and Maric (2017) presented a characterization of the set $\mathcal{P}_d^{\mathrm{B}}(1/2)$ which can be used for membership testing as we now explain. For $l = 1, \ldots, 2^{d-1}$, let $\boldsymbol{b}(l) = (b_1, \ldots, b_d)$ be the binary expansion of l, that is,

$$\boldsymbol{b}(l) = (b_1, \ldots, b_d) \quad \text{if and only if} \quad l = 1 + \sum_{j=1}^{d} b_j 2^{d-j}.$$

Note that b_1 is equal to 0 for all $l = 1, \ldots, 2^{d-1}$. For each l, let π_l be the d-dimensional distribution which puts equal mass on $\boldsymbol{b}(l) = (b_1, \ldots, b_d)$ and $\mathbf{1} - \boldsymbol{b}(l) = (1 - b_1, \ldots, 1 - b_d)$. One can easily check that the correlation matrix of $\boldsymbol{X} \sim \pi_l$ is given by

$$\rho(X_i, X_j) = 2\mathbf{1}_{\{b_i(l) = b_j(l)\}} - 1, \quad i, j = 1, \ldots, d,$$

where $b_i(l)$ denotes the ith element of $\boldsymbol{b}(l)$. This leads to the following characterization of the set $\mathcal{P}_d^{\mathrm{B}}(1/2)$; see Huber and Maric (2017).

Theorem 2.3.6 (Characterization of $\mathcal{P}_d^{\mathrm{B}}(1/2)$)**.** $\mathcal{P}_d^{\mathrm{B}}(1/2)$ is the convex hull of correlation matrices of the two-point distributions $\pi_1, \ldots, \pi_{2^{d-1}}$, that is,

$$\begin{aligned}\mathcal{P}_d^{\mathrm{B}}(1/2) &= \mathrm{conv}\{\rho(\pi_l) : \ l = 1, \ldots, 2^{d-1}\} \\ &= \left\{ \sum_{l=1}^{2^{d-1}} \alpha_l \rho(\pi_l) : \ \alpha_1, \ldots, \alpha_{2^{d-1}} \geq 0, \ \alpha_1 + \cdots + \alpha_{2^{d-1}} = 1 \right\},\end{aligned}$$

where $\rho(\pi_l)$ is the correlation matrix of π_l.

Remark 2.3.7 (Cut polytope and elliptope)**.** By Theorem 2.3.6, $\mathcal{P}_d^{\mathrm{B}}(1/2)$ coincides with a set known as a *cut polytope*, which is the collection of matrices $\boldsymbol{c}\boldsymbol{c}^\top$ for all $\boldsymbol{c} \in \{-1, 1\}^d$. Moreover, its positive semi-definite relaxation is known to be the *elliptope* $\mathcal{P}_d$; see Laurent and Poljak (1995) and Tropp (2018).

Example 2.3.8 (Cases $d = 2$ and $d = 3$)**.** Write $P = (\rho_{ij}) \in \mathcal{P}_d^{\mathrm{B}}(1/2)$. When $d = 2$, $\rho_{12} = \rho_{21}$ and ρ_{12} can take any value from -1 to 1 since $\rho_{12} = \alpha(+1) + (1-\alpha)(-1) = 2\alpha - 1$ for $\alpha \in [0,1]$. When $d = 3$, the characterization in Theorem 2.3.6 reduces to

$$-1 \le \sum_{1 \le i < j \le 3} \rho_{ij} \le 1 + 2 \min_{1 \le i,j \le 3} \{\rho_{ij}\}. \tag{2.8}$$

In terms of the triple $(\rho_{12}, \rho_{13}, \rho_{23})$ of correlations, (2.8) forms a tetrahedron with vertices $(1,1,1)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. One can check that $\mathcal{P}_d^{\mathrm{B}}(1/2)$ is a strict subset of $\mathcal{P}_d$ for $d \ge 3$. For instance, consider a matrix of the form

$$P(\rho) = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}.$$

Then $P(\rho)$ is a proper correlation matrix if and only if $-1/2 \le \rho \le 1$. On the other hand, the inequality in (2.8) says that $P(\rho) \in \mathcal{P}_d^{\mathrm{B}}(1/2)$ if and only if $-1/3 \le \rho \le 1$. Therefore, if $-1/2 \le \rho < -1/3$, then $P(\rho)$ belongs to $\mathcal{P}_d$ but not to $\mathcal{P}_3^{\mathrm{B}}(1/2)$.

The characterization in Theorem 2.3.6 provides a method to check that a given matrix is Bern(1/2)-compatible.

Proposition 2.3.9 (Checking Bern(1/2)-compatibility)**.** A given matrix $P = (\rho_{ij})$ is Bern(1/2)-compatible if and only if there exist $\alpha_1, \ldots, \alpha_{2^{d-1}} \ge 0$ such that the following $1 + d(d-1)/2$ equations hold

$$\alpha_1 + \cdots + \alpha_{2^{d-1}} = 1, \quad \sum_{l=1}^{2^{d-1}} \alpha_l \mathbf{1}_{\{b_i(l) = b_j(l)\}} = \frac{\rho_{ij} + 1}{2}, \quad 1 \le i < j \le d.$$

Equivalently, the following *phase I linear program* attains zero

$$\min\{z_1 + \cdots + z_{2^{d-1}}\} \quad \text{subject to} \quad \begin{cases} D\boldsymbol{\alpha} + \boldsymbol{z} = \boldsymbol{\lambda}, \\ \boldsymbol{\alpha}, \boldsymbol{z} \ge \mathbf{0}, \end{cases} \tag{2.9}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{2^{d-1}}) \in [0,1]^{2^{d-1}}$, $\boldsymbol{\lambda} = (\lambda_{12}, \lambda_{13}, \lambda_{23}, \ldots, \lambda_{d-1\,d}, 1) \in [0,1]^{1+d(d-1)/2}$ for $\lambda_{ij} = (\rho_{ij} + 1)/2$ and

$$D = \begin{pmatrix} \mathbf{1}_{\{b_1(1)=b_2(1)\}} & \mathbf{1}_{\{b_1(2)=b_2(2)\}} & \cdots & \mathbf{1}_{\{b_1(2^{d-1})=b_2(2^{d-1})\}} \\ \mathbf{1}_{\{b_1(1)=b_3(1)\}} & \mathbf{1}_{\{b_1(2)=b_3(2)\}} & \cdots & \mathbf{1}_{\{b_1(2^{d-1})=b_3(2^{d-1})\}} \\ \mathbf{1}_{\{b_2(1)=b_3(1)\}} & \mathbf{1}_{\{b_2(2)=b_3(2)\}} & \cdots & \mathbf{1}_{\{b_2(2^{d-1})=b_3(2^{d-1})\}} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{1}_{\{b_{d-1}(1)=b_d(1)\}} & \mathbf{1}_{\{b_{d-1}(2)=b_d(2)\}} & \cdots & \mathbf{1}_{\{b_{d-1}(2^{d-1})=b_d(2^{d-1})\}} \\ 1 & 1 & \cdots & 1 \end{pmatrix} \in \{0,1\}^{\left(1+\frac{d(d-1)}{2}\right) \times 2^{d-1}}.$$

Note that the set of constraints in (2.9) is always nonempty since $(\boldsymbol{\alpha}, \boldsymbol{z}) = (\mathbf{0}, \boldsymbol{\lambda})$ is a feasible solution. The phase I linear program can be solved, for example, with the R package lpSolve although it is computationally

Algorithm 2 Simulating random vectors with Bern(1/2) marginals and given correlation matrix P.

Input: Correlation matrix $P \in \mathcal{P}_d^{\mathrm{B}}(1/2)$.

Output: Random vector $\boldsymbol{B} = (B_1, \ldots, B_d)$ with $B_j \sim \mathrm{Bern}(1/2)$, $j = 1, \ldots, d$ and $\rho(\boldsymbol{B}) = P$.

1) For P, solve (2.9) to find $(\alpha_1, \ldots, \alpha_{2^{d-1}})$.
2) Choose the index l with probability α_l, $l \in \{1, \ldots, 2^{d-1}\}$.
3) Set $\boldsymbol{B} = \boldsymbol{b}(l)$ or $\mathbf{1} - \boldsymbol{b}(l)$ with probability $1/2$ each.

demanding for large d. This is to be expected since such problems are known to be NP-complete; see Pitowsky (1991).

Once a (componentwise) non-negative vector $\boldsymbol{\alpha}^*$ such that $D\boldsymbol{\alpha}^* = \boldsymbol{\lambda}$ is obtained, the corresponding symmetric Bernoulli random vector $\boldsymbol{B}$ with correlation matrix $P = (\rho_{ij})$ can be simulated by the following algorithm, which enables us to solve the attainability problem discussed in Section 2.3.4.

Example 2.3.10 (Numerical example for $d = 3$)**.** Consider the two 3×3 matrices

$$P_1 = \begin{pmatrix} 1 & -0.95 & 0.5 \\ -0.95 & 1 & -0.4 \\ 0.5 & -0.4 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & -0.9 & 0.5 \\ -0.9 & 1 & -0.4 \\ 0.5 & -0.4 & 1 \end{pmatrix},$$

both of which can be shown to be positive definite, so correlation matrices. For $d = 3$, the numbers $l = 1, \ldots, 2^{d-1} = 4$ have the binary expansions $\boldsymbol{b}(1) = (0,0,0)$, $\boldsymbol{b}(2) = (0,0,1)$, $\boldsymbol{b}(3) = (0,1,0)$ and $\boldsymbol{b}(4) = (0,1,1)$. The corresponding matrix D is then given by

$$D = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

For P_1, $\lambda_1 = (\lambda_{1,12}, \lambda_{1,13}, \lambda_{1,23}, 1) = (0.025, 0.750, 0.300, 1.000)$. Solving the phase I linear program with the R package lpSolve yields the minimum 0.025 of the objective function $z_1 + z_2 + z_3 + z_4$, which does not attain zero. Therefore, although P_1 is a proper correlation matrix, it is not Bern(1/2)-compatible. For P_2, $\lambda_2 = (0.050, 0.750, 0.300, 1.000)$. By using lpSolve, the objective function is found to achieve zero, and we thus numerically checked that $P_2 \in \mathcal{P}_d^{\mathrm{B}}(1/2)$. These results can also be confirmed with the inequality in (2.8).

One can thus check the compatibility of Blomqvist's beta matrices (or, equivalently, correlation matrices of random vectors with symmetric Bernoulli margins) by solving the phase I linear program (2.9) and by checking whether the objective function attains zero. By the same procedure, the sufficient condition shown in Proposition 2.3.2 can also be checked for general κ_G compatibility.

2.3.4 Attainability of matrices of measures of concordance

We now consider the attainability problem. We call a κ_G-compatible matrix $P \in [-1,1]^{d\times d}$ κ_G-*attainable* if one can construct a random vector $\boldsymbol{X} = (X_1, \dots, X_d)$ such that $\kappa_G(\boldsymbol{X}) = P$. The proof of Proposition 2.3.2 already indicates such a construction principle for a d-dimensional random vector $\boldsymbol{X}$ such that, for a given matrix $P \in \mathcal{P}_d^{\mathrm{B}}(1/2)$, one has $\kappa_G(\boldsymbol{X}) = P$.

Corollary 2.3.11 (κ_G-attainability of $P \in \mathcal{B}_d = \mathcal{P}_d^{\mathrm{B}}(1/2)$)**.** Let $P \in \mathcal{P}_d^{\mathrm{B}}(1/2)$ and the representation $P = \sum_{l=1}^{2^{d-1}} \alpha_l \rho(\pi_l)$ according to Theorem 2.3.6 be given. Then P is κ_G-attainable by $\boldsymbol{X} = (X_1, \dots, X_d)$ defined by

$$X_j = B_j U + (1 - B_j)(1 - U), \quad j = 1, \dots, d, \tag{2.10}$$

where $U \sim \mathrm{Unif}(0,1)$ and $\boldsymbol{B} = (B_1, \dots, B_d)$ is constructed as in Algorithm 2.

Since $\mathcal{B}_d = \mathcal{P}_d^{\mathrm{B}}(1/2)$, that is, the set of Blomqvist's beta matrices coincide with the set of correlation matrices of random vectors with symmetric Bernoulli marginals, all matrices $P \in \mathcal{B}_d$ can be attained by (2.10).

Next, for matrices of pairwise van der Waerden's coefficients ζ, any ζ-compatible matrix is attainable by a multivariate normal distribution.

Corollary 2.3.12 (ζ-attainability of $P \in \mathcal{W}_d = \mathcal{P}_d$)**.** Any matrix $P \in \mathcal{W}_d$ is attainable by the multivariate normal distribution with covariance matrix P.

Finally, for Spearman's rho, ρ_{S}-attainability is not completely solved for dimensions $d \geq 3$. If $P \in \mathcal{P}_d^{\mathrm{B}}(1/2)$, P is ρ_{S}-attainable by Corollary 2.3.11 for $d \geq 3$. If $P \notin \mathcal{P}_d^{\mathrm{B}}(1/2)$, P is known to be ρ_{S}-attainable only when $d = 3$ by the results in Hürlimann (2012), Hürlimann (2014) and Kurowicka and Cooke (2001), where *universal copulas* are studied, that is, explicitly constructed copulas with given correlation matrices. For $d \geq 4$, such a universal copula is still unknown to the best of our knowledge. Accordingly, a general ρ_{S}-compatible matrix P is not known to be attainable when $d \geq 4$.

2.4 Compatibility and attainability for block matrices

In this section, we study the compatibility and attainability of *block matrices* P, that is, matrices containing homogeneous blocks (so blocks of equal entries), possibly with ones on the diagonal. A special case of block matrices are hierarchical matrices, which are introduced in Example 2.4.1. Block matrices naturally appear when clustering algorithms are applied to matrices of measures of concordance or when (rather) sparse, partially exchangeable hierarchical models are designed.

Although all the criteria introduced in Section 2.3 can be directly applied to block correlation matrices, the corresponding computational effort can be large, especially when d is large. The comparably small number of different entries in block or hierarchical matrices is especially attractive for high-dimensional modeling and one expects more efficient ways to check compatibility and attainability for such matrices.

Specifically, compatibility and attainability for Spearman's rho matrices are in demand since, as discussed in Section 2.3, there is no method available to check compatibility for $d \geq 10$, and to check attainability for $d \geq 4$.

2.4.1 Definition and notations

We consider the following symmetric matrix in $[-1,1]^{d\times d}$ with diagonal entries equal to one:

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1S} \\ \vdots & \ddots & \vdots \\ P_{S1} & \cdots & P_{SS} \end{pmatrix}, \quad \text{for } P_{s_1s_2} = \begin{cases} (1-\rho_{ss})I_{d_s} + \rho_{ss}J_{d_s}, & \text{if } s_1 = s_2 = s, \\ \rho_{s_1s_2}J_{d_{s_1}d_{s_2}}, & \text{if } s_1 \neq s_2, \end{cases} \tag{2.11}$$

where I_{d_s} denotes the $d_s \times d_s$ identity matrix, $J_{d_{s_1}d_{s_2}} = \mathbf{1}_{d_{s_1}}\mathbf{1}_{d_{s_2}}^\top \in \mathbb{R}^{d_{s_1}\times d_{s_2}}$ (for $\mathbf{1}_{d_s} = (1,\dots,1) \in \mathbb{R}^{d_s}$) is the $d_{s_1} \times d_{s_2}$ matrix of ones and $J_{d_s} = J_{d_sd_s}$. We call a matrix of Form (2.11) a *block homogeneous matrix*. For notational convenience, let

$$\Gamma_d(a,b) = aI_d + b(J_d - I_d) = (a-b)I_d + bJ_d$$

which is also known as the d-dimensional *compound symmetry matrix*. With this notation, the matrices on the diagonal of P in (2.11) can be written as $P_{ss} = \Gamma_{d_s}(1,\rho_{ss})$.

A matrix of Form (2.11) appears, for example, as a correlation matrix of a random vector with homogeneous correlations within blocks. Let $\boldsymbol{X} = (X_1,\dots,X_d)$ be a d-dimensional random vector which can be divided into S such blocks or groups

$$\boldsymbol{X} = (\boldsymbol{X}_1,\dots,\boldsymbol{X}_S) = (X_{11},\dots,X_{1d_1},\dots,X_{S1},\dots,X_{Sd_S}), \tag{2.12}$$

where d_s is the size of group $s \in \{1,\dots,S\}$. In financial and insurance applications, the groups are often industry sectors, business sectors, regions, etc. If we consider the case where the correlation between two random variables depends only on the groups they belong to, then the resulting correlation matrix of $\boldsymbol{X}$ is block homogeneous of Form (2.11) where $\rho_{s_1s_2}$ represents the correlation coefficient within two (possibly equal) groups s_1 and s_2.

We call a matrix P *block homogeneous* if it is a symmetric $[-1,1]^{d\times d}$ matrix with diagonal entries equal to one, but not necessarily a correlation matrix since positive definiteness of P is not assumed. Note that, for compound symmetry matrices, it is well-known that $\Gamma_d(a,b)$ is positive definite if and only if $-a/(d-1) < b < a$. Therefore, P_{ss}, $s = 1,\dots,S$, is positive definite if and only if $-1/(d_s-1) < \rho_{ss} < 1$.

Example 2.4.1 (Hierarchical matrices)**.** Consider the block homogeneous matrix

$$P = \begin{pmatrix} 1 & 0.4 & 0.4 & 0.4 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.4 & 1 & 0.4 & 0.4 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.4 & 1 & 0.4 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.4 & 1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1 & 0.3 & 0.3 & 0.15 & 0.15 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 1 & 0.3 & 0.15 & 0.15 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 0.3 & 1 & 0.15 & 0.15 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.15 & 0.15 & 0.15 & 1 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.15 & 0.15 & 0.15 & 0.2 & 1 \end{pmatrix} \tag{2.13}$$

with $S = 3$, $(d_1, d_2, d_3) = (4, 3, 2)$, $(\rho_{11}, \rho_{22}, \rho_{33}, \rho_{12}, \rho_{13}, \rho_{23}) = (0.4, 0.3, 0.2, 0.1, 0.1, 0.15)$. This matrix can be described by a tree T_P illustrated in Figure 2.3. For the tree T_P, denote by v_{lm} the mth node (counted

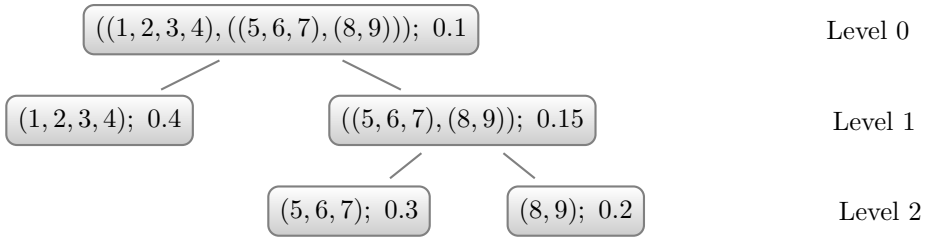

Figure 2.3: Tree representation T_P of the hierarchical correlation matrix P in (2.13).

from the left) at level $l \in \{0, 1, 2\}$. The *leaves* (that is, the terminal nodes) v_{11}, v_{21} and v_{22} represent the groups of variable indices $(1, 2, 3, 4)$, $(5, 6, 7)$ and $(8, 9)$, respectively. Nodes v_{21} and v_{22} are connected by a node v_{12}, and v_{11} and v_{12} are connected by a node v_{01}. To each vertex $v = v_{01}, v_{11}, v_{12}, v_{21}, v_{22}$ (in the set of vertices denoted by $\mathcal{V} = \{v_{01}, v_{11}, v_{12}, v_{21}, v_{22}\}$), a single number $\rho_v = 0.4, 0.3, 0.2, 0.15, 0.1$ is attached, respectively. The vertex v_{01} at the lowest level is called *root*; if two nodes v and v' are connected and v is at lower level than v', then v is a *parent* of v' and v' is a *child* of v. A node v is called *descendant* of another node v' if v' is in the shortest path from v to the root of the tree; note that each single node is regarded as a descendant of itself. Finally, for a pair of two nodes (v, v'), the *lowest common ancestor* is the node at lowest level that has both v and v' as descendants; when $v = v'$, the lowest common ancestor is v itself.

With these notions, the block matrix P is recovered from the tree T_P by defining a matrix with diagonal entries equal to 1 and the (i, j)-entry, for $i \neq j$, equal to the number attached to the descendant of (v_i, v_j) where v_i and v_j are the leaves of groups of variable indices containing i and j, respectively. If a block homogeneous correlation matrix P admits such a tree representation T_P, we call P *hierarchical matrix* and T_P the corresponding *hierarchical tree*. The matrix (2.13) is thus a hierarchical matrix with corresponding tree displayed in Figure 2.3.

2.4.2 Positive (semi-)definiteness

By Corollary 2.3.3, positive (semi-)definiteness is a necessary condition for compatibility of matrices of transformed rank correlation coefficients including Spearman's rho, Blomqvist's beta and van der Waerden's coefficient. In the case of van der Waerden's coefficient, it is even sufficient for compatibility. If a matrix is block homogeneous, it turns out to suffice to check positive semi-definiteness of an $S \times S$ matrix, see Theorem 2.4.3 below. This result can lead to a significant reduction in the computational effort for checking compatibility.

Definition 2.4.2 (Block average map)**.** Let P be a block homogeneous matrix of form (2.11). The *block average map* $P \mapsto \phi(P)$ for $P = (\rho_{ij}) \in \mathbb{R}^{d\times d}$ is defined by

$$\phi(P) = \begin{pmatrix} \tilde{\rho}_{11} & \rho_{12} & \cdots & \rho_{1S} \\ \rho_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{S-1\,S} \\ \rho_{S1} & \cdots & \rho_{S-1\,S} & \tilde{\rho}_{SS} \end{pmatrix} \in \mathbb{R}^{S\times S}, \quad \tilde{\rho}_{ss} = \frac{1 + (d_s - 1)\rho_{ss}}{d_s},\ s = 1, \ldots, S.$$

The block average map ϕ allows one to collapse block matrices (to "ordinary" matrices). If $\boldsymbol{X}$ is a random vector as in (2.12) with $\mathbb{E}(\boldsymbol{X}) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{X}) = P$ where P is as in (2.11), then $\boldsymbol{Y} = (\overline{Y}_1, \ldots, \overline{Y}_S)$ defined by the group averages $\overline{Y}_s = \frac{1}{d_s}\sum_{j=1}^{d_s} X_{sj}$ has covariance matrix $\phi(P)$, that is, $\mathrm{Cov}(\boldsymbol{Y}) = \phi(P)$. Huang and Yang (2010); Roustant and Deville (2017) showed that it suffices to check positive (semi-)definiteness of the matrix $\phi(P) \in \mathbb{R}^{S\times S}$ to obtain positive (semi-)definiteness of $P \in \mathbb{R}^{d\times d}$.

Theorem 2.4.3 (Characterization of positive (semi-)definiteness of block matrices)**.** Let $P \in \mathbb{R}^{d\times d}$ be a block matrix as in (2.11). Then P is positive (semi-)definite if and only if $\phi(P)$ is positive (semi-)definite.

Proof. See Huang and Yang (2010) and Roustant and Deville (2017). □

Example 2.4.4 (Positive definiteness of a hierarchical matrix)**.** Consider P as in (2.13), so $S = 3$, $d = 9$, $(d_1, d_2, d_3) = (4, 3, 2)$ with block average map given by

$$\begin{aligned} \phi(P) &= \begin{pmatrix} (1 + (d_1 - 1)\rho_{11})/d_1 & \rho_{12} & \rho_{13} \\ \rho_{21} & (1 + (d_2 - 1)\rho_{22})/d_2 & \rho_{23} \\ \rho_{31} & \rho_{32} & (1 + (d_3 - 1)\rho_{33})/d_3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1+(4-1)0.4}{4} & 0.1 & 0.1 \\ 0.1 & \frac{1+(3-1)0.3}{3} & 0.15 \\ 0.1 & 0.15 & \frac{1+(2-1)0.2}{2} \end{pmatrix} = \begin{pmatrix} 0.55 & 0.1 & 0.1 \\ 0.1 & 0.5\overline{3} & 0.15 \\ 0.1 & 0.15 & 0.6 \end{pmatrix}. \end{aligned}$$

One can easily check that $\phi(P)$ is positive definite. By Theorem 2.4.3, P is thus positive definite.

2.4.3 Block Cholesky decomposition

The *Cholesky decomposition* of a positive definite (positive semi-definite) matrix $P \in \mathcal{P}_d$ is $P = LL^\top$ for a lower triangular matrix L with positive (non-negative) diagonal elements, which is called the *Cholesky*

factor of P. Such a decomposition of P exists if and only if P is positive (semi-)definite and so can be used to check the latter property computationally.

Cholesky decompositions are of utmost importance in various areas of statistics. In quantitative risk management, they are frequently utilized to construct multivariate elliptical distributions. For example, once the Cholesky factor L of P is computed, the d-dimensional random vector $\boldsymbol{X} = L\boldsymbol{Z}$ for $\boldsymbol{Z} \sim \mathrm{N}_d(0, I_d)$ satisfies $\mathrm{Cov}(\boldsymbol{X}) = LL^\top = P$. This $\boldsymbol{X}$ thus attains a given matrix P of van der Waerden's coefficients; see Corollary 2.3.12. For building hierarchical dependence models after estimating groups of homogeneous models or after applying clustering algorithms (which naturally lead to groups of variables), one often considers block homogeneous correlation matrices or hierarchical matrices (see Example 2.4.1). We will now turn to the question how Cholesky factors of such matrices look like and can be computed more efficiently than in the classical way.

Proposition 2.4.5 (Cholesky factor of block matrices)**.** For a $d \times d$ block homogeneous correlation matrix P of form (2.11), its Cholesky factor L is of the form

$$L = \begin{pmatrix} L_{11} & O & \cdots & O \\ L_{21} & L_{22} & \ddots & O \\ \vdots & \vdots & \ddots & \vdots \\ L_{S1} & L_{S2} & \cdots & L_{SS} \end{pmatrix}$$

where $O = (0)$ represents a block of zeros and, for $s = 1, \ldots, S$, the diagonal matrices are

$$L_{ss} = \begin{pmatrix} \tilde{l}_{ss,1} & 0 & 0 & \cdots & 0 \\ l_{ss,1} & \tilde{l}_{ss,2} & 0 & & \vdots \\ l_{ss,1} & l_{ss,2} & \tilde{l}_{ss,3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ l_{ss,1} & l_{ss,2} & l_{ss,3} & \cdots & \tilde{l}_{ss,d_s} \end{pmatrix} \in \mathbb{R}^{d_s \times d_s}$$

for some $\tilde{l}_{ss,k},\ k = 1, \ldots, S$ and $l_{ss,k},\ k = 1, \ldots, S-1$, and the off-diagonal matrices are

$$L_{s+m,s} = (c_{sm,1}\mathbf{1}_{d_{s+m}}, \ldots, c_{sm,d_s}\mathbf{1}_{d_{s+m}}) \in \mathbb{R}^{d_{s+m} \times d_s}, \quad m = 1, \ldots, S-s$$

for some $(c_{sm,1}, \ldots, c_{sm,d_s})$.

The proof of Proposition 2.4.5 reduces to verify Algorithm 3 at the end of this chapter, which computes the Cholesky factor of a given block homogeneous correlation matrix.

Proof of Algorithm 3 and Proposition 2.4.5. Consider the first iteration $s = 1$. Let

$$L_{11} = \sqrt{P_{11}}, \quad L_{s1} = P_{s1}(L_{11}^\top)^{-1}, \quad s = 2, \ldots, d_S.$$

Since $P_{11} = \Gamma_{d_1}(1, \rho_{11})$ is a compound symmetry matrix, solving the equation $L_{11}L_{11}^\top = P_{11}$ yields that L_{11} is of the form

$$L_{11} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{11,1} & \tilde{l}_{11,2} & 0 & & \vdots \\ l_{11,1} & l_{11,2} & \tilde{l}_{11,3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ l_{11,1} & l_{11,2} & l_{11,3} & \cdots & \tilde{l}_{11,d_1} \end{pmatrix} \in \mathbb{R}^{d_1 \times d_1},$$

where

$$\tilde{l}_{11,j} = \sqrt{1 - \sum_{k=1}^{j-1} l_{11,k}^2}, \quad \text{and} \quad l_{11,j} = \frac{1}{\tilde{l}_{11,j}} \Big(\rho_{11} - \sum_{k=1}^{j-1} l_{11,k}^2 \Big), \quad j = 1, \ldots, d_1.$$

Note that all off-diagonal components in the same column are equal. This set of equations can be solved sequentially for $j = 1, \ldots, d_1$. For $s = 2, \ldots, d_S$, since $P_{s1} = \rho_{s1} J_{d_s d_1}$, $L_{s1} = \rho_{s1} J_{d_s d_1} (L_{11}^\top)^{-1}$ can be written as

$$L_{s1} = (c_{s1,1}\mathbf{1}_{d_s}, \ldots, c_{s1,d_1}\mathbf{1}_{d_s}) \in \mathbb{R}^{d_s \times d_1}, \quad s = 2, \ldots, S,$$

where $(c_{s1,1}, \ldots, c_{s1,d_1})$ can be sequentially determined via

$$c_{s1,j}\tilde{l}_{11,j} + \sum_{k=1}^{j-1} c_{s1,k} l_{11,k} = \rho_{s1}, \quad j = 1, \ldots, d_1.$$

Let $P_{-(1:d_1)}$ be the submatrix of P obtained by removing the first d_1 rows and columns. Let L_{-1} be the Cholesky factor of

$$\overline{P}(1) = P_{-(1:d_1)} - (P_{21}, \ldots, P_{d_s 1})^\top P_{11}^{-1} (P_{21}^\top, \ldots, P_{d_S 1}^\top).$$

Then $LL^\top = P$ for the lower triangle matrix

$$L = \begin{pmatrix} L_{11} & O & \cdots & O \\ L_{21} & & & \\ \vdots & & L_{-1} & \\ L_{S1} & & & \end{pmatrix}.$$

We now show that $\overline{P}(1)$ is a block matrix with diagonal blocks equal to compound symmetric matrices and off-diagonal blocks equal to constant matrices. Since

$$\begin{aligned} &(P_{21}, \ldots, P_{S1})^\top P_{11}^{-1} (P_{21}^\top, \ldots, P_{d_S 1}^\top) \\ &= (\rho_{21} J_{d_2 d_1}, \ldots, \rho_{S1} J_{d_S d_1})^\top P_{11}^{-1} (\rho_{21} J_{d_2 d_1}^\top, \ldots, \rho_{S1} J_{d_S d_1}^\top), \end{aligned}$$

its (i, j)-block for $i, j \in \{1, \ldots, S-1\}$ is given by

$$\rho_{i+1,1}\rho_{j+1,1} J_{d_{i+1}, d_1} P_{11}^{-1} J_{d_{j+1} d_1}^\top = \rho_{i+1,1}\rho_{j+1,1} \mathbf{1}_{d_{i+1}} \mathbf{1}_{d_1}^\top P_{11}^{-1} \mathbf{1}_{d_1} \mathbf{1}_{d_{j+1}}^\top.$$

Since $P_{11} = \Gamma_{d_1}(1, \rho_{11})$, we have that

$$P_{11}\mathbf{1}_{d_1} = (1 + (d_1 - 1)\rho_{11})\mathbf{1}_{d_1}.$$

Moreover,

$$\mathbf{1}_{d_1}^\top P_{11}^{-1} P_{11}\mathbf{1}_{d_1} = \mathbf{1}_{d_1}^\top \mathbf{1}_{d_1} = d_1.$$

Putting these equalities together, we obtain that

$$\mathbf{1}_{d_1}^\top P_{11}^{-1}\mathbf{1}_{d_1} = \frac{d_1}{1 + (d_1 - 1)\rho_{11}}.$$

Therefore, the (i, j)-block of the second term of $\overline{P}_{-1}$ is given by

$$\rho_{i+1,1}\rho_{j+1,1} J_{d_{i+1}d_1} P_{11}^{-1} J_{d_{j+1}d_1}^\top = \frac{d_1\rho_{i+1,1}\rho_{j+1,1}}{1 + (d_1 - 1)\rho_{11}} J_{d_{i+1}d_{j+1}}.$$

Consequently, $\overline{P}(1)$ is a block matrix with (i, i)th block given by

$$\begin{aligned}
&\Gamma_{d_{i+1}}(1, \rho_{i+1,i+1}) + \frac{d_1\rho_{i+1,1}^2}{1 + (d_1 - 1)\rho_{11}} J_{d_{i+1}} \\
&= \Gamma_{d_{i+1}}\left(1 + \frac{d_1\rho_{i+1,1}^2}{1 + (d_1 - 1)\rho_{11}},\ \rho_{i+1,i+1} + \frac{d_1\rho_{i+1,1}^2}{1 + (d_1 - 1)\rho_{11}}\right), \quad i = 1, \ldots, S - 1,
\end{aligned}$$

and with (i, j)th block given by

$$\begin{aligned}
&\rho_{i+1,j+1} J_{d_{i+1},d_{j+1}} + \frac{d_1\rho_{i+1,1}\rho_{j+1,1}}{1 + (d_1 - 1)\rho_{11}} J_{d_{i+1},d_{j+1}} \\
&= \left(\rho_{i+1,j+1} + \frac{d_1\rho_{i+1,1}\rho_{j+1,1}}{1 + (d_1 - 1)\rho_{11}}\right) J_{d_{i+1}d_{j+1}}, \quad i, j \in \{1, \ldots, S - 1\},\ i \neq j.
\end{aligned}$$

Since $\overline{P}(1)$ has the same structure as the initial matrix P, the same procedure can be applied to find a Cholesky factor L_{-1} such that $L_{-1}L_{-1}^\top = \overline{P}(1)$. By iteratively applying this procedure, we obtain the Cholesky factor L of P. □

Algorithm 3 uses only $S(S + 1)/2$ correlation coefficients and the block sizes $d_1, \ldots, d_S$ without the need to consider the full $d \times d$ matrix P, which can lead to significant computational savings especially when d is large and S is small. The following example covers the individual steps of Algorothm 3 with concrete numbers.

Example 2.4.6 (Case of $S = 3$, $(d_1, d_2, d_3) = (4, 3, 2)$)**.** Consider the block homogeneous matrix (2.13). As discussed in Example 2.4.4, the matrix P in (2.13) is positive definite, and thus has a Cholesky factor L.

By applying Algorithm 3, the Cholesky factor L of P is obtained as

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.92 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.26 & 0.88 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.26 & 0.2 & 0.86 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.07 & 0.05 & 0.04 & 0.99 & 0 & 0 & 0 & 0 \\ 0.1 & 0.07 & 0.05 & 0.04 & 0.28 & 0.95 & 0 & 0 & 0 \\ 0.1 & 0.07 & 0.05 & 0.04 & 0.28 & 0.21 & 0.93 & 0 & 0 \\ 0.1 & 0.07 & 0.05 & 0.04 & 0.13 & 0.1 & 0.08 & 0.97 & 0 \\ 0.1 & 0.07 & 0.05 & 0.04 & 0.13 & 0.1 & 0.08 & 0.15 & 0.96 \end{pmatrix}.$$

In the first iteration $s = 1$ of Algorithm 3 with $\overline{P}(1) = P$, the Cholesky factor of the square matrix with entries from the first $d_1 = 4$ columns is computed. By solving (2.16), $P_{11} = \Gamma_{d_1}(1, \rho_{11})$ is decomposed into L_{11} of form (2.15), which is determined by $(\tilde{l}_{11,1}, \tilde{l}_{11,2}, \tilde{l}_{11,3}, \tilde{l}_{11,4}, l_{11,1}, l_{11,2}, l_{11,3}) = (1.00, 0.92, 0.88, 0.88, 0.40, 0.26, 0.20)$. By solving (2.17), L_{21} and L_{31} are determined via $(c_{11,1}, \ldots, c_{11,d_1})$ and $(c_{12,1}, \ldots, c_{12,d_1})$ by $(c_{11,1}, \ldots, c_{11,4}) = (c_{12,1}, \ldots, c_{12,4}) = (0.1, 0.07, 0.05, 0.04)$. For iteration $s = 2$, the submatrix $\overline{P}(2)$ is computed following Step 5) via $(\rho_1^{(2)}, \rho_{1,o}^{(2)}, \rho_2^{(2)}, \rho_{2,o}^{(2)}, \rho_{12}^{(2)}) = (0.98, 0.28, 0.98, 0.18, 0.13)$. By solving (2.16) and (2.17), L_{22} and L_{32} are specified via $(\tilde{l}_{22,1}, \tilde{l}_{22,2}, \tilde{l}_{22,3}, l_{22,1}, l_{22,1}) = (0.99, 0.95, 0.93, 0.28, 0.21)$, and $(c_{21,1}, c_{21,2}) = (0.13, 0.10)$. Finally, the submatrix $\overline{P}(3)$ is given by $\overline{P}(3) = \Gamma_2(0.95, 0.15)$. The Cholesky factor L_{33} is then specified via $(\tilde{l}_{33,1}, \tilde{l}_{33,2}, l_{33,1}) = (0.97, 0.96, 0.15)$ by solving the equations in (2.16).

2.4.4 Attainability for block matrices

In this section, we study compatibility and attainability of measures of concordance for a block homogeneous matrices of form (2.11). We expect that checking compatibility and attainability of a given $d \times d$ block matrix can be reduced to check those of some $S \times S$ matrix for a block size S, which can be much smaller than d.

For van der Waerden's coefficient, we have already seen that Theorem 2.4.3 is available for checking compatibility and that Proposition 2.4.5 is beneficial to attain a given ζ-compatible matrix. For Spearman's rho block matrices, we have the following result.

Proposition 2.4.7 (ρ_{S}-compatible subclass of block matrices)**.** Let P be a $d_1 + \cdots + d_S$ block homogeneous correlation matrix of form (2.11). Let $M = (m_{s_k s_l})$ be a $S \times S$ matrix with $m_{ss} = 1$, $s = 1, \ldots, S$, and

$$m_{s_k s_l} = \frac{d_{s_k} d_{s_l} \rho_{s_k, s_l}}{(1 + (d_{s_k} - 1)\rho_{s_k s_k})(1 + (d_{s_l} - 1)\rho_{s_l s_l})}, \quad s_k, s_l \in \{1, \ldots, S\},\ s_k \neq s_l.$$

If $M \in \mathcal{S}_S$, then P is ρ_{S}-compatible. Moreover, if M is ρ_{S}-attainable, so is P.

Proof. Let $\lambda_s = \tilde{\rho}_{ss} = \frac{1+(d_s-1)\rho_{ss}}{d_s}$. Then positive definiteness of P requires $-1/(d_s - 1) < \rho_{ss} < 1$ and thus it holds that $\lambda_s \in (0, 1)$. Notice that

$$\lambda_s + (1 - \lambda_s)\left(-\frac{1}{d_s - 1}\right) = \rho_{ss}.$$

If $M \in \mathcal{S}_S$, there exists an S-dimensional random vector $\boldsymbol{U} = (U_1, \ldots, U_S)$ with standard uniform margins such that $\rho(\boldsymbol{U}) = M$. For $s \in \{1, \ldots, S\}$, there exists a d_s-dimensional random vector $\boldsymbol{V}_s$ with $\mathrm{Unif}(0,1)$ margins such that its correlation matrix is $\Gamma(1, -1/(d_s - 1))$ for $s \in \{1, \ldots, S\}$; see Murdoch et al. (2001) for a construction. Let $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_S$ be such random vectors independent of each other, and also independent of $\boldsymbol{U}$. For $s \in \{1, \ldots, S\}$, let $B_s \sim \mathrm{Bern}(\lambda_s)$ such that $B_1, \ldots, B_S$ are independent of each other, and independent of $\boldsymbol{U}$ and $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_S$. For $s = 1, \ldots, S$, define a d_s-dimensional random vector

$$\boldsymbol{W}_s = B_s U_s \mathbf{1}_{d_s} + (1 - B_s)\boldsymbol{V}_s. \tag{2.18}$$

One can easily check that $\boldsymbol{W}_s$ has $\mathrm{Unif}(0,1)$ marginals. Moreover, for $s = 1, \ldots, S$,

$$\rho(\boldsymbol{W}_s) = \lambda_s J_{d_s} + (1 - \lambda_s)\Gamma_{d_s}(1, -1/(d_s - 1)) = \Gamma_{d_s}(1, \rho_{ss}) = P_{ss},$$

and for $s_1 \neq s_2$, $i = 1, \ldots, d_{s_1}$, $j = 1, \ldots, d_{s_2}$,

$$\rho(W_{s_1 i}, W_{s_2 j}) = \lambda_{s_1}\lambda_{s_2}\rho(U_{s_1}, U_{s_2}) = \lambda_{s_1}\lambda_{s_2} m_{s_1 s_2} = \rho_{s_1 s_2}.$$

Therefore, $(\boldsymbol{W}_1^\top, \ldots, \boldsymbol{W}_S^\top)$ is a $(d_1 + \cdots + d_S)$-dimensional random vector with correlation matrix P. Since its marginal distributions are all $\mathrm{Unif}(0,1)$, P is ρ_{S}-compatible by Proposition 2.3.4 Part 2. If M is ρ_{S}-attainable by constructing $\boldsymbol{U}$ above, then P is ρ_{S}-attainable via construction (2.18). □

If $S \leq 9$, checking $M \in \mathcal{S}_S$ can be reduced to checking its positive semi-definiteness by Proposition 2.3.4 Part 1 and Part 2. If $S \geq 10$, a sufficient condition is available related to $\mathrm{Bern}(1/2)$-compatibility by Proposition 2.3.2. On attainability of P, M is ρ_{S}-attainable only for the sector size $S = 3$; see the discussion of ρ_{S}-attainability in Section 2.3.4.

Example 2.4.8 (Case with $d = 9$ and $S = 3$)**.** Let P be the block homogeneous correlation matrix defined in (2.13). Since $d \leq 9$, its compatibility can be verified by checking that P is positive semi-definite. In fact, the corresponding matrix M in Proposition 2.4.7 of P is

$$M = \begin{pmatrix} 1 & 0.341 & 0.303 \\ 0.341 & 1 & 0.469 \\ 0.303 & 0.469 & 1 \end{pmatrix}$$

and one can also check that M is positive definite by a simple calculation. Therefore, P is ρ_{S}-compatible by Proposition 2.4.7. Since M is 3-dimensional, P is ρ_{S}-attainable; see the discussion in Section 2.3.4. Therefore, even though P is 9 (> 3)-dimensional, it is ρ_{S}-attainable by construction (2.18).

When a given block homogeneous matrix P is a hierarchical matrix, then the following sufficient condition is available for compatibility and attainability of *any* measure of concordance.

Proposition 2.4.9 (Compatible and attainable hierarchical matrices)**.** For a general measure of concordance κ, a $d \times d$ hierarchical matrix P is κ-compatible and κ-attainable (by a nested or hierarchical Archimedean copula (HAC)) if, for the corresponding hierarchical tree, $0 \leq \rho_v \leq \rho_{v'}$ holds for every pair of nodes (v, v') such that v is a parent of v'.

Proof. Let $\psi_\theta : [0,\infty] \to [0,1]$ be a one-parameter Archimedean generator with $\theta \in \Theta = (\theta_{\min}, \theta_{\max})$, $\theta_{\min} \leq \theta_{\max} \leq \infty$ and let $C_\theta(u_1, u_2) = \psi_\theta(\psi_\theta^{-1}(u_1) + \psi_\theta^{-1}(u_2))$, $u_1, u_2 \in [0,1]$, be the corresponding Archimedean copula family. Suppose $\{\psi_\theta; \theta \in \Theta\}$ satisfies the following conditions:

(1) (Complete monotonicity) $(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\psi_\theta(t) \geq 0$ for any $\theta \in \Theta$ and $k = 0, 1, \ldots$;

(2) (Limiting copulas) $C_{\theta_{\min}} = \lim_{\theta \downarrow \theta_{\min}} C_\theta$ is the independence copula and $C_{\theta_{\max}} = \lim_{\theta \uparrow \theta_{\max}} C_\theta$ is the comonotone copula;

(3) (Positive ordering) if $\theta, \theta' \in \Theta$ such that $\theta \leq \theta'$ then $C_\theta \preceq C_{\theta'}$; and

(4) (Sufficient nesting condition) $\psi_\theta^{-1} \circ \psi_{\theta'}$ is completely monotone for $\theta, \theta' \in \Theta$ if and only if $\theta \leq \theta'$.

Examples of Archimedean copulas satisfying Conditions (1)–(4) are the Clayton and Gumbel copula families with generators given by Laplace transforms of certain gamma and positive stable distributions, respectively; see (Nelsen, 2006, Examples 4.12 and 4.14) and (Hofert, 2010, Tables 2.1 and 2.3). Note that Condition (1) guarantees that the d-dimensional Archimedean copula $C_\theta(u_1, \ldots, u_d) = \psi_\theta(\sum_{j=1}^d \psi_\theta^{-1}(u_j))$ is also a d-copula for any $d \geq 2$; see Kimberling (1974). Together with the continuity and coherence axioms of a measure of concordance, Condition (2) and (3) imply that the map $\kappa(\theta) : \theta \mapsto \kappa(C_\theta)$ is increasing and continuous from Θ to $[0,1]$. Therefore, for every pair of nodes (v, v'), there exist $\theta_v, \theta_{v'} \in \Theta$ such that $\theta_v \leq \theta_{v'}$ and $\kappa(\theta_v) = \rho_v \leq \rho_{v'} = \kappa(\theta_{v'})$. For the hierarchical tree T_P of a given hierarchical matrix P with the corresponding collection of generators $\{\psi_{\theta_v}; v \in \mathcal{V}\}$, Condition (4) thus ensures that there exists a corresponding HAC; see McNeil (2008) and (Joe, 1997, pp. 87) for the sufficient nesting condition and Hofert (2012) and Górecki et al. (2017) for the construction of HACs. By construction, the matrix of pairwise measure of concordance κ is equal to P for this HAC. Thus, P is both κ-compatible and κ-attainable. □

When a hierarchical matrix P satisfies the sufficient condition in Proposition 2.4.9, we call P a *proper hierarchical matrix*. Note that componentwise non-negativity of P is necessary since complete monotonicity (1) of ψ_θ implies that $\Pi \preceq C_\theta$; see (Hofert, 2010, Remark 2.3.2). For sampling from a HAC, see McNeil (2008), Hofert (2011) or Hofert (2012).

Remark 2.4.10 (Positive definiteness of hierarchical matrices)**.** In Proposition 2.4.9, positive definiteness of P was not a necessary assumption. In fact, positive definiteness is impled by the condition $0 \leq \rho_v \leq \rho_{v'}$ for any v and v' such that v is a parent of v' since Proposition 2.4.9 holds for any G-transformed rank correlation coefficient and κ_G-compatible matrices are necessarily positive definite.

Example 2.4.11 (Attainability of hierarchical matrix (2.13) for general κ)**.** By Proposition 2.4.9, the hierarchical matrix P in (2.13) is κ-compatible and κ-attainable for any measure of concordance κ since P is proper as can be easily checked from Figure 2.3. As an example of a model attaining P, let ψ_θ be the generator of Gumbel copula and let C_P be the corresponding HAC given, for each $\boldsymbol{u} \in [0,1]^9$, by

$$C_P(u_1, \ldots, u_9) = C_{v_{01}}\big(C_{v_{11}}(u_1, u_2, u_3, u_4), C_{v_{12}}(C_{v_{21}}(u_5, u_6, u_7), C_{v_{22}}(u_8, u_9))\big),$$

where the Gumbel copula C_v has parameter θ_v such that $\kappa(C_v) = \rho_v$ is attained for every node v. For example, if κ is Blomqvist's beta β, one has $\beta(\theta_v) = \beta(C_v) = 4C_v(1/2, 1/2) - 1 = 2^{2-2^{1/\theta_v}} - 1$, $\theta_v \in [1, \infty)$,

which is continuous and increasing from 0 to $\lim_{\theta_v \to \infty} \beta(\theta_v) = 1$. Therefore, for each $\rho_v = \beta_v$, $v \in \mathcal{V}$, the parameter θ_v is given by $\theta_v = 1/\big(\log_2(2 - \log_2(1 + \beta_v))\big)$.

As another example, when κ is Kendall's tau τ, it is known that $\tau(\theta_v) = \tau(C_{\theta_v}) = (\theta_v - 1)/\theta_v$ for $\theta_v \in [1, \infty)$ and so $\theta_v = 1/(1 - \tau_v)$ where τ_v is the corresponding entry in P in (2.13) or Figure 2.3. Thus, for example, $\tau_{v_{01}} = 0.1$ implies that $\theta_{v_{01}} = 10/9$. The same construction applies to κ being Spearman's rho or van der Waerden's coefficient and the C_v being Clayton copulas, for example. Note that it may sometimes be necessary to find θ_v such that $\kappa(\theta_v) = \kappa_v$ for a given κ_v numerically.

2.5 Conclusion and discussion

We introduced a new class of measures of concordance called transformed rank correlation coefficients, whose members depend on functions G_1 and G_2. Spearman's rho, Blomqvist's beta and van der Waerden's coefficient are obtained as special cases. We provided necessary and sufficient conditions on G_1 and G_2 when transformed rank correlation coefficients are measures of concordance; see Theorem 2.2.5.

For matrices of (pairwise) transformed rank correlation coefficients, a sufficient condition for compatibility and attainability was derived in terms of Bern(1/2)-compatibility; see Proposition 2.3.2 and Corollary 2.3.11 for compatibility and attainability, respectively. We also presented characterizations of the sets of compatible Spearman's rho, Blomqvist's beta and van der Waerden's matrices; see Proposition 2.3.4. This result revealed that, among these measures of concordance, van der Waerden's coefficient may be the most convenient one in terms of checking compatibility and attainability since its compatible set coincides with that of Pearson's linear correlation coefficient.

We then studied compatible and attainable block matrices for which fast methods of checking positive semi-definiteness and of calculating Cholesky factors were derived; see Theorem 2.4.3 and Algorithm 3, respectively. For certain subclasses of block matrices, the problem of checking compatibility and attainability can be reduced to lower dimensions; see Proposition 2.4.7 and Proposition 2.4.9.

Further research is required for compatibility of measures of concordance which cannot be represented as transformed rank correlation coefficients, such as Kendall's tau and Gini's gamma. Another angle to take for future research, which will be also addressed in Chapter 3, is a comparison among different transformed rank correlation coefficients to obtain a clear answer on which measure is the best to be used from a statistical point of view. In terms of block matrices, dimension reduction for checking compatibility of transformed rank correlation coefficients is also an interesting problem for future research.

Algorithm 3 Cholesky decomposition for block matrices

Input: A block correlation matrix P.

Output: The Cholesky decomposition L of P.

1) For $s = 1, \dots, S$, $\overline{P}(1) = P$ and $\overline{P}(s)$, $s \geq 2$ of the form

$$\overline{P}(s) = \begin{pmatrix} P^{(s)}_{1,1} & \cdots & P^{(s)}_{1,S-s+1} \\ \vdots & \ddots & \vdots \\ P^{(s)}_{S-s+1,1} & \cdots & P^{(s)}_{S-s+1,S-s+1} \end{pmatrix}, \tag{2.14}$$

where, for $s_1, s_2 \in \{1, \dots, S-s+1\}$,

$$P^{(s)}_{s_1,s_2} = \begin{cases} \Gamma_{d_{s+t-1}}(\rho^{(s)}_t, \rho^{(s)}_{t,o}), & \text{if } s_1 = s_2 = t \in \{1, \dots, S-s+1\}, \\ \rho^{(s)}_{s_1,s_2} J_{d_{s+s_1-1} d_{s+s_2-1}}, & \text{if } s_1 \neq s_2, \end{cases}$$

for some diagonal entries of diagonal blocks $\rho^{(s)}_t$, off-diagonal entries of diagonal blocks $\rho^{(s)}_{t,o}$, and entries of off-diagonal blocks $\rho^{(s)}_{s_1,s_2}$, do the following.

2-1) Set

$$L_{ss} = \begin{pmatrix} \tilde{l}_{ss,1} & 0 & 0 & \cdots & 0 \\ l_{ss,1} & \tilde{l}_{ss,2} & 0 & & \vdots \\ l_{ss,1} & l_{ss,2} & \tilde{l}_{ss,3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ l_{ss,1} & l_{ss,2} & l_{ss,3} & \cdots & \tilde{l}_{ss,d_s} \end{pmatrix} \in \mathbb{R}^{d_s \times d_s}, \tag{2.15}$$

where

$$\tilde{l}_{ss,j} = \sqrt{\rho^{(s)}_1 - \sum_{k=1}^{j-1} l^2_{ss,k}}, \quad \text{and} \quad l_{ss,j} = \frac{1}{\tilde{l}_{ss,j}} \left(\rho^{(s)}_{1,o} - \sum_{k=1}^{j-1} l^2_{ss,k} \right), \quad j = 1, \dots, d_s. \tag{2.16}$$

2-2) If $s < S$, set, for $m = 1, \dots, S-s$,

$$L_{s+m,s} = (c_{sm,1} \mathbf{1}_{d_{s+m}}, \dots, c_{sm,d_s} \mathbf{1}_{d_{s+m}}) \in \mathbb{R}^{d_{s+m} \times d_s},$$

where $(c_{sm,1}, \dots, c_{sm,d_s})$ can be sequentially determined via

$$c_{sm,j} \tilde{l}_{ss,j} + \sum_{k=1}^{j-1} c_{sm,k} l_{ss,k} = \rho^{(s)}_{m+1,1}, \quad j = 1, \dots, d_s. \tag{2.17}$$

2-3) If $s < S$, set $\overline{P}(s+1)$ to be of form (2.14) with

$$\rho^{(s+1)}_t = \rho^{(s)}_{t+1} + \frac{d_s (\rho^{(s)}_{t+1,1})^2}{\rho^{(s)}_1 + (d_s - 1)\rho^{(s)}_{1o}}, \quad \rho^{(s+1)}_{t,o} = \rho^{(s)}_{t+1,o} + \frac{d_s (\rho^{(s)}_{t+1,1})^2}{\rho^{(s)}_1 + (d_s - 1)\rho^{(s)}_{1o}}, \quad t \in \{1, \dots, S-s\},$$

$$\rho^{(s+1)}_{s_i,s_j} = \rho^{(s)}_{s_i+1,s_j+1} + \frac{d_s \rho^{(s)}_{s_i+1,1} \rho^{(s)}_{s_j+1,1}}{\rho^{(s)}_1 + (d_s - 1)\rho^{(s)}_{1o}}, \quad s_1, s_2 \in \{1, \dots, S-s\}.$$

3) Return the Cholesky factor L whose (i,j)th block is L_{ij} for $i \geq j$ and is O for $i < j$, $i, j = 1, \dots, S$.

Chapter 3

Estimation and comparison of correlation-based measures of concordance

We address the problem of estimating and comparing transformed rank correlation coefficients defined in Chapter 2. We propose a novel framework for comparing transformed rank correlations in terms of the asymptotic variance of their canonical estimators. A general criterion derived from this framework is that concordance-inducing functions with smaller variances of squared random variables are more preferable. In particular, we show that Blomqvist's beta attains the optimal asymptotic variance and Spearman's rho outperforms van der Waerden's coefficient. We also find that the optimal bounds of the asymptotic variance are attained by Kendall's tau.

3.1 Introduction

In Chapter 2, we studied G-transformed rank correlation coefficient defined by

$$\kappa_G(C) = \kappa_G(U, V) = \rho(G^{-1}(U), G^{-1}(V))$$

for a copula C and $(U, V) \sim C$, where G is a distribution function called the concordance-inducing function and G^{-1} is the generalized inverse of G; see Section 2.2 of Chapter 2. Thanks to this representation via Pearson's correlation, this class of measures of concordance has various appealing properties, such as interpretability and ease of studying compatibility and attainability problems discussed in Chapter 2. Another advantage of this class is its ease of estimation since one can estimate κ_G by the sample correlation of pseudo-observations from C transformed by G^{-1} for a given G.

For a given class of transformed rank correlation coefficients, natural questions are which concordance-inducing function is best to use and how to compare different measures of concordance. De Winter et al.

(2016) compared Pearson's linear correlation and Spearman's rho by numerical experiments in terms of bias, variance and robustness to outliers. Various measures of concordance were compared in terms of their power in tests of independence; see, for example, Bhuchongkul (1964), Behnen (1971), Behnen (1972), Luigi Conti and Nikitin (1999), Rödel and Kössler (2004) and Genest and Verret (2005).

We tackle the problem of comparing measures of concordance from the theoretical viewpoint of statistical estimation of κ_G. In our proposed framework, a concordance-inducing function G is more preferable than another one G' if the largest (worst) or smallest (best) asymptotic variance of a canonical estimator $\hat{\kappa}_G$ of κ_G is smaller than that of G' for a certain set of copulas $\mathcal{D}$. Simply put, G is more preferable than G' if $\hat{\kappa}_G$ tends to estimate κ_G more accurately than $\hat{\kappa}_{G'}$ estimates $\kappa_{G'}$ if the underlying copula belongs to $\mathcal{D}$. A general criterion derived from this framework is that concordance-inducing functions with smaller variance $\mathrm{Var}_G(X^2)$ where $X \sim G$ is more preferable. Therefore, heavy-tailed concordance-inducing functions, such as a Student t distribution function, are not recommended in comparison to normal ones. We also find that Spearman's rho, for which G is the uniform distribution, can be outperformed by rank correlations transformed by Beta distributions. Moreover, under certain conditions on $\mathcal{D}$, we prove that Blomqvist's beta attains the optimal worst and best asymptotic variances among all transformed rank correlation coefficients, and Spearman's rho is more preferable than van der Waerden's coefficient. Considering the drawback of Blomqvist's beta that it only depends on the local value $C(1/2, 1/2)$ of a copula C, we also compare transformed rank correlations with Kendall's tau. Based on the representation of Kendall's tau in terms of Pearson's linear correlation coefficient, we find that Kendall's tau also attains the optimal worst and best asymptotic variances if estimators of these measures are compared without being standardized by sample size. Since the correlation-representation of Kendall's tau depends on two independent copies of random vectors following C, Kendall's tau is not optimal any more if the asymptotic variances of these estimators are standardized by sample size. Finally, in a simulation study, we find that the choice of concordance-inducing function G and the strength of dependence of the underlying copula C affect the asymptotic variance of $\hat{\kappa}_G$ more than the kinds of copulas.

This chapter is organized as follows. In Section 3.2 we introduce a framework for comparing G-transformed rank correlations in terms of their asymptotic variances. A canonical estimator of a transformed rank correlation is presented in Section 3.2.1. Section 3.2.2 addresses effects of location-scale transforms of G on the asymptotic variance. The worst and best asymptotic variances among fundamental and Fréchet copulas are provided in Section 3.2.3, and the optimality of Blomqvist's beta is given in Section 3.2.4. Transformed rank correlations and Kendall's tau are compared in Section 3.3. In Section 3.4, a simulation study is conducted to compare asymptotic variances for various parametric copulas and concordance-inducing fuctions. Section 3.5 concludes this work with discussions about directions for future research.

3.2 Estimation of κ_G and their comparison

In this section, we propose a novel framework for comparing G-transformed rank correlations to answer the question which concordance-inducing function is best to be used. In the proposed framework, transformed correlations are compared in terms of the asymptotic variances of their canonical estimators, and one concordance-inducing function G is considered better than another G' if the largest (worst) or

smallest (best) asymptotic variance of an estimator $\hat{\kappa}_G$ of κ_G among a set of copulas $\mathcal{D}$ is smaller than that of $\kappa_{G'}$.

To this end, let $\mathcal{C}_2$ denote the set of all bivariate copulas, that is, all bivariate distributions functions with standard uniform univariate marginal distributions. As defined in Section 1.2, we call $C' \in \mathcal{C}_2$ *more concordant* than $C \in \mathcal{C}_2$, denoted by $C \preceq C'$, if $C(u,v) \leq C'(u,v)$ for all $(u,v) \in [0,1]^2$. The survival function of C is given by $\bar{C}(u,v) = \mathbb{P}(U > u, V > v)$, $(u,v) \in [0,1]$ where $(U,V) \sim C$. For any map $\kappa : \mathcal{C}_2 \to \mathbb{R}$, we identify $\kappa(C)$ with $\kappa(U,V)$ for a random vector $(U,V) \sim C$ defined on a fixed atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The set of all concordance-inducing functions is denoted by $\mathcal{G}$.

3.2.1 Canonical estimator of κ_G

Since κ_G is invariant under location-scale transforms of G, we first consider standardized concordance-inducing functions G with mean zero and variance one. Effects of location-scale transforms of G to estimators of κ_G will be discussed in Section 3.2.2. Assuming that an i.i.d. sample (U_i, V_i), $i = 1, \ldots, n$, $n \in \mathbb{N}$, from C is available, we consider the following canonical estimator of κ_G:

$$\hat{\kappa}_G = \frac{1}{n} \sum_{i=1}^{n} G^{-1}(U_i) G^{-1}(V_i).$$

By the central limit theorem (CLT), $\hat{\kappa}_G$ satisfies the following asymptotic normality:

$$\sqrt{n} \left\{ \hat{\kappa}_G - \kappa_G(C) \right\} \xrightarrow{d} \mathrm{N}(0, \sigma_G^2(C)), \quad \sigma_G^2(C) = \mathrm{Var}(G^{-1}(U) G^{-1}(V)),$$

provided that $\mathrm{Var}(G^{-1}(U)G^{-1}(V)) < \infty$. Writing $X = G^{-1}(U)$ and $Y = G^{-1}(V)$ for $(U,V) \sim C$ and using that $\mathrm{Var}(XY) = \mathbb{E}[(XY)^2] - \mathbb{E}[XY]^2$, $\mathrm{Cov}(X^2, Y^2) = \mathbb{E}[(XY)^2] - \mathbb{E}[X^2]\mathbb{E}[Y^2] = \mathbb{E}[(XY)^2] - 1$ and $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY]$, we have that

$$\sigma_G^2(C) = \mathrm{Var}(XY) = \mathrm{Cov}(X^2, Y^2) + 1 - \mathrm{Cov}(X,Y)^2. \tag{3.1}$$

Since $\mathrm{Cov}(X^2, Y^2) = \mathrm{SD}(X^2)\mathrm{SD}(Y^2)\rho(X^2, Y^2) \leq \mathrm{SD}(X^2)\mathrm{SD}(Y^2)$ and $\mathrm{Cov}(X,Y)^2 = \rho^2(X,Y) \geq 0$, a sufficient condition for $\sigma_G^2(C) < \infty$ is that the fourth moment of G is finite. We consider the sets of optimal concordance-inducing functions and the corresponding optimal bounds in terms of the worst and best asymptotic variances of $\hat{\kappa}_G$, defined by

$$\underline{G}_*(\mathcal{H}, \mathcal{D}) = \operatorname*{arginf}_{G \in \mathcal{H}} \underline{\sigma}_G^2(\mathcal{D}), \quad \underline{\sigma}_*^2(\mathcal{H}, \mathcal{D}) = \inf_{G \in \mathcal{H}} \underline{\sigma}_G^2(\mathcal{D}),$$
$$\overline{G}_*(\mathcal{H}, \mathcal{D}) = \operatorname*{arginf}_{G \in \mathcal{H}} \overline{\sigma}_G^2(\mathcal{D}), \quad \overline{\sigma}_*^2(\mathcal{H}, \mathcal{D}) = \inf_{G \in \mathcal{H}} \overline{\sigma}_G^2(\mathcal{D}),$$

respectively, for $\mathcal{H} \subseteq \mathcal{G}_4$ and $\mathcal{D} \subseteq \mathcal{C}_2$, where

$$\underline{\sigma}_G^2(\mathcal{D}) = \inf_{C \in \mathcal{D}} \sigma_G^2(C), \quad \overline{\sigma}_G^2(\mathcal{D}) = \sup_{C \in \mathcal{D}} \sigma_G^2(C)$$

and

$$\mathcal{G}_4 = \{ G \in \mathcal{G} : \mathbb{E}_G[X] = 0, \ \mathrm{Var}_G(X) = 1 \text{ and } \mathbb{E}_G[X^4] < \infty \}$$

with $\mathbb{E}_G[X]$ and $\mathrm{Var}_G(X)$ being the mean and variance of $X \sim G$, respectively. The sets of attaining G functions $\underline{G}_*(\mathcal{H}, \mathcal{D})$ and $\overline{G}_*(\mathcal{H}, \mathcal{D})$ are defined to be empty if the infima in $\underline{\sigma}^2_G(\mathcal{D})$ and $\overline{\sigma}^2_G(\mathcal{D})$ are not attainable, respectively. Nevertheless, if $\mathcal{H}$ and $\mathcal{D}$ are closed in d_∞, these infima are attainable and thus $\underline{G}_*(\mathcal{H}, \mathcal{D})$ and $\overline{G}_*(\mathcal{H}, \mathcal{D})$ are non-empty; see Section 3.7.3 for details. Calculating the optimal worst and best asymptotic variances $\overline{\sigma}^2_*(\mathcal{H}, \mathcal{D})$ and $\underline{\sigma}^2_*(\mathcal{H}, \mathcal{D})$ is not straightforward since neither $C \mapsto \sigma^2_G(C)$ nor $G \mapsto \sigma^2_G(C)$ have simple linearity; see Section 3.7.2 for details. For simplicity of the discussion, we impose the condition $\mathbb{E}_G[X^4] < \infty$ in $\mathcal{G}_4$ although concordance-inducing functions with $\mathbb{E}_G[X^4] = \infty$ are typically not involved in determining optimal best and worst asymptotic variances. Although an ideal choice of $\mathcal{H}$ is $\mathcal{H} = \mathcal{G}_4$, other choices can also be of interest; for example, $\mathcal{H} = \mathcal{G}_4^{\mathrm{c}}$ where $\mathcal{G}_4^{\mathrm{c}}$ is the set of continuous concordance-inducing functions in $\mathcal{G}_4$, and $\mathcal{H} = \mathcal{G}_4^{\mathrm{b}}$ where $\mathcal{G}_4^{\mathrm{b}}$ is the set of concordance-inducing functions in $\mathcal{G}_4$ with bounded supports. Note that one-sided distributions such that $\mathrm{esssup}(G) = \infty$ and $\mathrm{essinf}(G) < \infty$, or $\mathrm{esssup}(G) < \infty$ and $\mathrm{essinf}(G) = -\infty$, cannot be concordance-inducing since they cannot be radially symmetric. Therefore, the set $\mathcal{G}_4^{\mathrm{b}}$ excludes concordance-inducing functions whose supports are $\mathbb{R}$, and $\mathcal{G}_4 \backslash \mathcal{G}_4^{\mathrm{b}}$ is a set of concordance-inducing functions in $\mathcal{G}_4$ with supports $\mathbb{R}$. The sets of optimal concordance-inducing functions $\overline{G}_*(\mathcal{H}, \mathcal{D})$ and $\underline{G}_*(\mathcal{H}, \mathcal{D})$ are considered as the best choices among the set of concordance-inducing functions $\mathcal{H} \subseteq \mathcal{G}_4$ to accurately estimate κ_G if one believes that $\mathcal{D}$ is the set of underlying copulas which one wants to quantify and compare in terms of their concordance.

Remark 3.2.1 (Reflection invariance of $\sigma^2_G(C)$)**.** Let $\nu_1,\ \nu_2 : \mathcal{C}_2 \to \mathcal{C}_2$ be partial reflections of copulas defined by

$$\nu_1(C)(u, v) = v - C(1 - u, v) \quad \text{and} \quad \nu_2(C)(u, v) = u - C(u, 1 - v), \quad C \in \mathcal{C}_2,$$

respectively, with their composition given by $\nu_1 \circ \nu_2(C)(u, v) = u + v - 1 + C(1 - u, 1 - v)$. For an operator $\varphi : \mathcal{C}_2 \to \mathcal{C}_2$, let $C_\varphi = \varphi(C)$. Then $(1 - U, V) \sim C_{\nu_1}$, $(U, 1 - V) \sim C_{\nu_2}$ and $(1 - U, 1 - V) \sim C_{\nu_1 \circ \nu_2}$ for $(U, V) \sim C$. By radial symmetry of $G \in \mathcal{G}$, we have that $G^{-1}(1 - U) = -G^{-1}(U)$ and $G^{-1}(1 - V) = -G^{-1}(V)$. Therefore, $\sigma^2_G(C)$ is invariant under the reflections ν_1, ν_2, $\nu_1 \circ \nu_2$ in the sense that $\sigma^2_G(C) = \sigma^2_G(C_{\nu_1}) = \sigma^2_G(C_{\nu_2}) = \sigma^2_G(C_{\nu_1 \circ \nu_2})$. This property follows intuitively since $|\kappa_G(C)|$ is also invariant under reflections, and thus one can estimate each of the quantities $\kappa_G(C)$, $\kappa_G(C_{\nu_1})$, $\kappa_G(C_{\nu_2})$ and $\kappa_G(C_{\nu_1 \circ \nu_2})$ from any other.

Remark 3.2.2 (Asymptotic variance of Blomqvist's beta)**.** Schmid and Schmidt (2007) derived an asymptotic variance of Blomqvist's beta. Their asymptotic variance is in general different from ours since we standardize the Bernoulli concordance-inducing function so that it has mean zero and variance one. As they stated, one of the advantages of Blomqvist's beta over other measures of concordance is that Blomqvist's beta admits an explicit form if the copula can be written explicitly. In fact, this advantage can be passed on to a wider class of discrete concordance-inducing functions; see Section 3.6.1 for details.

3.2.2 Optimal location shift of G

Although κ_G is invariant under location-scale transforms of G, the asymptotic variance $\sigma^2_G(C)$ of its canonical estimator $\hat{\kappa}_G$ is not location invariant. To see this, let $G_0 \in \mathcal{G}_4$ be a concordance-inducing function with mean zero and variance one, and let $G_{\mu,\sigma}(x) = G_0(\frac{x-\mu}{\sigma})$ be the corresponding concordance-inducing

function of the same type as G_0 but with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. A canonical estimator of $\kappa_{G_{\mu,\sigma}}$ for known μ and σ is then given by

$$\hat{\kappa}_{G_{\mu,\sigma}} = \frac{1}{n}\sum_{i=1}^{n} \frac{G_{\mu,\sigma}^{-1}(U_i)G_{\mu,\sigma}^{-1}(V_i)}{\sigma^2} - \left(\frac{\mu}{\sigma}\right)^2.$$

By the CLT, $\hat{\kappa}_{G_{\mu,\sigma}}$ is asymptotically normal with asymptotic variance given by

$$\sigma^2_{G_{\mu,\sigma}}(C) = \mathrm{Var}\left(\frac{G_{\mu,\sigma}^{-1}(U)G_{\mu,\sigma}^{-1}(V)}{\sigma^2}\right).$$

Since $G_{\mu,\sigma}^{-1}(U)/\sigma = G_{\mu/\sigma,1}^{-1}(U)$ and $G_{\mu,\sigma}^{-1}(V)/\sigma = G_{\mu/\sigma,1}^{-1}(V)$, one can assume that $\sigma = 1$ without changing the asymptotic variance $\sigma^2_{G_{\mu,\sigma}}(C)$. Therefore, $\sigma^2_G(C)$ for $G \in \mathcal{G}_4$ is invariant under scale transforms of G. On the other hand, $\sigma^2_G(C)$ changes under location transforms of G since shifting G^{-1} by $\mu \in \mathbb{R}$ leads to the asymptotic variance $\mathrm{Var}((X+\mu)(Y+\mu)) = \mathrm{Var}(XY + \mu(X+Y))$ for $X = G^{-1}(U)$ and $Y = G^{-1}(V)$, and it is in general not equal to $\mathrm{Var}(XY)$.

Since the canonical estimator $\hat{\kappa}_{G_{\mu,\sigma}}$ estimates the same quantity κ_{G_0} regardless of the mean μ and variance σ^2 of G, a natural choice of μ under $\sigma = 1$ is such that it minimizes the asymptotic variance $\sigma^2_{G_{\mu,1}}(C)$. For a fixed concordance-inducing function $G_0 \in \mathcal{G}_4$ with mean zero and variance one, denote by $G_\mu(x) = G_0(x-\mu)$ the concordance-inducing function of the same type as G_0 but with mean $\mu \in \mathbb{R}$. For $X = X_0 + \mu \sim G_\mu$ and $Y = Y_0 + \mu \sim G_\mu$ with $X_0 = G_0^{-1}(U)$ and $Y_0 = G_0^{-1}(V)$, the asymptotic variance

$$\begin{aligned}\sigma^2_{G_\mu}(C) &= \mathrm{Var}(XY) = \mathrm{Var}((X_0+\mu)(Y_0+\mu)) = \mathrm{Var}(X_0Y_0 + \mu(X_0+Y_0)) \\ &= \mathrm{Var}(X_0Y_0) + 2\mu\,\mathrm{Cov}(X_0Y_0, X_0+Y_0) + \mu^2\,\mathrm{Var}(X_0+Y_0)\end{aligned}$$

is a quadratic function of $\mu \in \mathbb{R}$ provided that $\mathrm{Var}(X_0+Y_0) > 0$, and thus is minimized when

$$\mu = \mu_* = \mu_*(G_0, C) = -\frac{\mathrm{Cov}(X_0Y_0, X_0+Y_0)}{\mathrm{Var}(X_0+Y_0)}.$$

We call $\mu_*(G_0, C)$ an *optimal shift* of $G_0 \in \mathcal{G}_4$ under $C \in \mathcal{C}_2$. The degenerate case $\mathrm{Var}(X_0+Y_0) = 0$ occurs if and only if $\rho(X_0, Y_0) = -1$, and it is also equivalent to $C = W$; see Embrechts et al. (2002). In this case, $X_0 + Y_0 \overset{\text{a.s.}}{=} 0$ ("a.s." stands for *almost surely*) and thus $\mathrm{Var}(XY) = \mathrm{Var}(X_0Y_0)$, that is, location transforms of G_0 do not change $\sigma^2_{G_0}(C)$. Provided $\mathrm{Var}(X_0+Y_0) > 0$, that is, $C \neq W$, the optimal asymptotic variance is given by

$$\sigma^2_{G_{\mu_*}}(C) = \mathrm{Var}(X_0Y_0) - \frac{\mathrm{Cov}(X_0Y_0, X_0+Y_0)^2}{\mathrm{Var}(X_0+Y_0)}. \tag{3.2}$$

The following proposition states that $\mu_* = 0$ for a certain class of copulas.

Proposition 3.2.3 (Sufficient condition for $\mu_* = 0$)**.** For a copula $C \in \mathcal{C}_2$ and a concordance-inducing function $G_0 \in \mathcal{G}_4$ with mean zero and variance one, $\mu_*(G_0, C) = 0$ holds if C is radially symmetric $C = C_{\nu_1 \circ \nu_2}$, that is, $(U, V) \overset{\text{d}}{=} (1-U, 1-V)$ for $(U, V) \sim C$.

Proof. For $X_0 = G_0^{-1}(U)$ and $Y_0 = G_0^{-1}(V)$ with $(U, V) \sim C$, we have that $\mathbb{E}[X_0+Y_0] = \mathbb{E}[X_0] + \mathbb{E}[Y_0] = 0$, and thus $\mathrm{Cov}(X_0Y_0, X_0+Y_0) = \mathbb{E}[X_0Y_0(X_0+Y_0)] - \mathbb{E}[X_0Y_0]\mathbb{E}[X_0+Y_0] = \mathbb{E}[X_0Y_0(X_0+Y_0)]$. Therefore, it suffices to show that $\mathbb{E}[X_0Y_0(X_0+Y_0)] = 0$ when C is radially symmetric.

For $\varphi \in \{\iota, \nu_1, \nu_2, \nu_1 \circ \nu_2\}$ where $\iota : \mathcal{C}_2 \to \mathcal{C}_2$ is the identity $\iota(C) = C$, denote $(U_\varphi, V_\varphi) \sim C_\varphi$. Since G_0 is radially symmetric, we have that

$$\begin{aligned}(G_0^{-1}(U), G_0^{-1}(V)) &\overset{\mathrm{d}}{=} (-G_0^{-1}(U_{\nu_1}), G_0^{-1}(V_{\nu_1})) \overset{\mathrm{d}}{=} (G_0^{-1}(U_{\nu_2}), -G_0^{-1}(V_{\nu_2})) \\ &\overset{\mathrm{d}}{=} (-G_0^{-1}(U_{\nu_1\circ\nu_2}), -G_0^{-1}(V_{\nu_1\circ\nu_2})).\end{aligned}$$

Moreover, when C is radially symmetric, we have that $C_{\nu_1} = C_{\nu_2}$. Therefore, $(U, V) \overset{\mathrm{d}}{=} (U_{\nu_1\circ\nu_2}, V_{\nu_1\circ\nu_2})$ and $(U_{\nu_1}, V_{\nu_1}) \overset{\mathrm{d}}{=} (U_{\nu_2}, V_{\nu_2})$ for $(U, V) \sim C$. Together with the identity

$$\begin{aligned}1 &= \mathbf{1}_{\{U>1/2, V>1/2\}} + \mathbf{1}_{\{U\le 1/2, V>1/2\}} + \mathbf{1}_{\{U>1/2, V\le 1/2\}} + \mathbf{1}_{\{U\le 1/2, V\le 1/2\}} \\ &= \mathbf{1}_{\{U>1/2, V>1/2\}} + \mathbf{1}_{\{U_{\nu_1}>1/2, V_{\nu_1}>1/2\}} + \mathbf{1}_{\{U_{\nu_2}>1/2, V_{\nu_2}>1/2\}} + \mathbf{1}_{\{U_{\nu_1\circ\nu_2}>1/2, V_{\nu_1\circ\nu_2}>1/2\}},\end{aligned}$$

we have that

$$\begin{aligned}\mathbb{E}[X_0Y_0(X_0+Y_0)] &= \sum_{\varphi\in\{\iota,\nu_1,\nu_2,\nu_1\circ\nu_2\}} \mathbb{E}[\mathbf{1}_{\{U_\varphi>1/2, V_\varphi>1/2\}} G_0^{-1}(U)G_0^{-1}(V)(G_0^{-1}(U)+G_0^{-1}(V))] \\ &= \mathbb{E}[\mathbf{1}_{\{U>1/2,V>1/2\}} G_0^{-1}(U)G_0^{-1}(V)(G_0^{-1}(U)+G_0^{-1}(V))] \\ &\quad - \mathbb{E}[\mathbf{1}_{\{U_{\nu_1\circ\nu_2}>1/2, V_{\nu_1\circ\nu_2}>1/2\}} G_0^{-1}(U_{\nu_1\circ\nu_2})G_0^{-1}(V_{\nu_1\circ\nu_2})(G_0^{-1}(U_{\nu_1\circ\nu_2})+G_0^{-1}(V_{\nu_1\circ\nu_2}))] \\ &\quad + \mathbb{E}[\mathbf{1}_{\{U_{\nu_1}>1/2, V_{\nu_1}>1/2\}} G_0^{-1}(U_{\nu_1})G_0^{-1}(V_{\nu_1})(G_0^{-1}(U_{\nu_1})-G_0^{-1}(V_{\nu_1}))] \\ &\quad - \mathbb{E}[\mathbf{1}_{\{U_{\nu_2}>1/2, V_{\nu_2}>1/2\}} G_0^{-1}(U_{\nu_2})G_0^{-1}(V_{\nu_2})(G_0^{-1}(U_{\nu_2})-G_0^{-1}(V_{\nu_2}))] \\ &= 0,\end{aligned}$$

where the last equality comes from $(U, V) \overset{\mathrm{d}}{=} (U_{\nu_1\circ\nu_2}, V_{\nu_1\circ\nu_2})$ and $(U_{\nu_1}, V_{\nu_1}) \overset{\mathrm{d}}{=} (U_{\nu_2}, V_{\nu_2})$. □

The conditions $C = C_{\nu_1\circ\nu_2}$ and $C_{\nu_1} = C_{\nu_2}$ in Proposition 3.2.3 hold, for example, if C is M, W, Π, a Gaussian copula, t copula or one of their mixtures. For these copulas, location shifts of G_0 do not change the asymptotic variance $\sigma^2_{G_0}(C)$, and thus $\sigma^2_{G_0}(C)$ is invariant under location-scale transforms of G_0. On the other hand, shifting G_0 may improve $\sigma^2_{G_0}(C)$ if C is, for example, a Clayton or Gumbel copula. Nevertheless, we will empirically observe in Section 3.4 that the reduction of the asymptotic variance $\sigma^2_{G_0}(C)$ by the optimal shift μ_* is typically ignorable compared with the first term $\mathrm{Var}(X_0Y_0)$ in (3.2). Based on this observation, in this study we focus on the case $\mu = 0$ and compare asymptotic variances of $\hat{\kappa}_G$ only for standardized concordance-inducing functions G with mean zero and variance one, and comparing the asyptotic variance $\sigma^2_{G_{\mu_*}}(C)$ under the optimal shift is left for future research.

3.2.3 Asymptotic variance for fundamental and Fréchet copulas

In this section we investigate optimal concordance-inducing functions and the corresponding bounds of the best and worst asymptotic variances when $\mathcal{D} \subseteq \mathcal{C}_2$ is a set of fundamental copulas or their mixtures.

Proposition 3.2.4 (Optimal asymptotic variances for sets of fundamental copulas)**.**

1. If $\mathcal{D} = \{\Pi\}$, then $\underline{\sigma}^2_*(\mathcal{H}, \{\Pi\}) = \overline{\sigma}^2_*(\mathcal{H}, \{\Pi\}) = 1$ and $\underline{G}_*(\mathcal{H}, \{\Pi\}) = \overline{G}_*(\mathcal{H}, \{\Pi\}) = \mathcal{H}$ for any $\mathcal{H} \subseteq \mathcal{G}_4$.

2. Suppose $\mathcal{D} = \{M\}$, $\{W\}$ or $\{M, W\}$. Then, for $\mathcal{H} \subseteq \mathcal{G}_4$,

$$\underline{\sigma}^2_*(\mathcal{H}, \mathcal{D}) = \overline{\sigma}^2_*(\mathcal{H}, \mathcal{D}) = \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2),$$
$$\underline{G}_*(\mathcal{H}, \mathcal{D}) = \overline{G}_*(\mathcal{H}, \mathcal{D}) = \operatorname*{arginf}_{G \in \mathcal{H}} \mathrm{Var}_G(X^2).$$

3. If $\mathcal{D} = \{\Pi, M, W\}$, then, for $\mathcal{H} \subseteq \mathcal{G}_4$,

$$\underline{\sigma}^2_*(\mathcal{H}, \{\Pi, M, W\}) = 1 \wedge \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2),$$
$$\overline{\sigma}^2_*(\mathcal{H}, \{\Pi, M, W\}) = 1 \vee \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2).$$

4. Let $\mathcal{H}_{\mathrm{N}}$, $\mathcal{H}_{\mathrm{Unif}}$ and $\mathcal{H}_{\mathrm{Bern}}$ be singletons of normal, uniform and Bernoulli distributions with mean zero and variance one, respectively. Then

$$\underline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \{M, W\}) < \underline{\sigma}^2_*(\mathcal{H}_{\mathrm{Unif}}, \{M, W\}) < \underline{\sigma}^2_*(\mathcal{H}_{\mathrm{N}}, \{M, W\}),$$
$$\overline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \{M, W\}) < \overline{\sigma}^2_*(\mathcal{H}_{\mathrm{Unif}}, \{M, W\}) < \overline{\sigma}^2_*(\mathcal{H}_{\mathrm{N}}, \{M, W\}),$$
$$\underline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \{\Pi, M, W\}) < \underline{\sigma}^2_*(\mathcal{H}_{\mathrm{Unif}}, \{\Pi, M, W\}) < \underline{\sigma}^2_*(\mathcal{H}_{\mathrm{N}}, \{\Pi, M, W\}),$$
$$\overline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \{\Pi, M, W\}) = \overline{\sigma}^2_*(\mathcal{H}_{\mathrm{Unif}}, \{\Pi, M, W\}) < \overline{\sigma}^2_*(\mathcal{H}_{\mathrm{N}}, \{\Pi, M, W\}).$$

Proof. Part 1). (X^2, Y^2) and (X, Y) are both independent random vectors when $(U, V) \sim \Pi$. Therefore, $\mathrm{Cov}(X^2, Y^2) = \mathrm{Cov}(X, Y) = 0$ and thus $\sigma^2_G(\Pi) = 1$ for all $G \in \mathcal{H}$ by (3.1), which gives $\underline{G}_*(\mathcal{H}, \{\Pi\}) = \overline{G}_*(\mathcal{H}, \{\Pi\}) = \mathcal{H}$ and $\underline{\sigma}^2_*(\mathcal{H}, \{\Pi\}) = \overline{\sigma}^2_*(\mathcal{H}, \{\Pi\}) = 1$.
Part 2). When the copula of (X, Y) is M or W, we have that $\mathrm{Cov}(X, Y) = \pm 1$, respectively. Moreover, the copula of (X^2, Y^2) is M since $(X^2, Y^2) \overset{\mathrm{d}}{=} (G^{-1}(U)^2, G^{-1}(U)^2)$ for $U \sim \mathrm{Unif}(0, 1)$ when $C = M$, and $(X^2, Y^2) \overset{\mathrm{d}}{=} (G^{-1}(U)^2, G^{-1}(1-U)^2) = (G^{-1}(U)^2, (-G^{-1}(U))^2) = (G^{-1}(U)^2, G^{-1}(U)^2)$ for $U \sim \mathrm{Unif}(0, 1)$ when $C = W$. Therefore, by (3.1), we have that

$$\sigma^2_G(M) = \overline{\rho}(X^2, Y^2)\,\mathrm{Var}_G(X^2) + 1 - 1^2 = \mathrm{Var}_G(X^2),$$
$$\sigma^2_G(W) = \overline{\rho}(X^2, Y^2)\,\mathrm{Var}_G(X^2) + 1 - (-1)^2 = \mathrm{Var}_G(X^2),$$

where $\overline{\rho}(X^2, Y^2)$ is the maximal correlation coefficient attained by the copula M with the marginal distributions X^2 and Y^2, and $\overline{\rho}(X^2, Y^2) = 1$ since X^2 and Y^2 are of the same type; see Embrechts et al. (2002). Since $\sigma^2_G(M) = \sigma^2_G(W) = \mathrm{Var}_G(X^2)$, we obtain the desired results.
Part 3). The results immediately follow from Part 1 and Part 2.
Part 4). By Part 2, we have that $\sigma^2_*(\mathcal{H}_{\mathrm{N}}, \{M, W\}) = \mathrm{Var}(X^2) = 2$ for $X \sim \mathrm{N}(0, 1)$, $\sigma^2_*(\mathcal{H}_{\mathrm{Unif}}, \{M, W\}) = \mathrm{Var}(12(U - 0.5)^2) = 0.8$ for $U \sim \mathrm{Unif}(0, 1)$ and $\sigma^2_*(\mathcal{H}_{\mathrm{Bern}}, \{M, W\}) = \mathrm{Var}((2B-1)^2) = 0$ for $B \sim \mathrm{Bern}(1/2)$. Together with Part 2 and Part 3 we have the desired inequalities. □

Proposition 3.2.4 Part 1 implies that the choice of the function G does not affect the accuracy of the estimation of κ_G when the underlying copula is the independence copula. Proposition 3.2.4 Part 2 shows that the optimal worst and best asymptotic variances are obtained as the variance of X^2 where $X \sim G \in \mathcal{G}_4$ when the underlying copula is M or W. Proposition 3.2.4 Part 3 gives the optimal worst

and best asymptotic variances when $\mathcal{D}$ is a set of fundamental copulas. Since a small variance of X^2 for $X \sim G$ is preferable in terms of best and worst asymptotic variances when the set of underlying copulas is $\{M, W\}$ or $\{\Pi, M, W\}$, heavy-tailed concordance-inducing functions, such as a Student t distribution with degrees of freedom $4 < \nu < \infty$, are not recommendable choices. Finally, Proposition 3.2.4 Part 4 means that Blomqvist's beta outperforms Spearman's rho and van der Waerden's coefficient, and van der Waerden's coefficient performs worst in terms of the optimal best and worst asymptotic variances when the set of underlying copulas is $\{M, W\}$ or $\{\Pi, M, W\}$.

We now consider a more general class of copulas defined as combinations of the fundamental copulas M, Π and W. A bivariate *Fréchet copula* is defined by

$$C^{\mathrm{F}}_{\boldsymbol{p}} = p_M M + p_\Pi \Pi + p_W W, \quad \boldsymbol{p} = (p_M, p_\Pi, p_W) \in \Delta_3,$$

where $\Delta_3 = \{(p_1, p_2, p_3) \in \mathbb{R}^3 : p_1, p_2, p_3 \geq 0,\ p_1 + p_2 + p_3 = 1\}$ is the standard unit simplex on $\mathbb{R}^3$. Denote by $\mathcal{C}^{\mathrm{F}} = \{C^{\mathrm{F}}_{\boldsymbol{p}} : \boldsymbol{p} \in \Delta_3\}$ the set of all Fréchet copulas. In addition to applications in insurance and finance, Fréchet copulas can be used to approximate bivariate copulas; see Yang et al. (2006). Moreover, for any $G \in \mathcal{G}$, the transformed rank correlation κ_G can take any value in $[-1, 1]$ since, by Proposition 2.2.9,

$$\kappa_G(C^{\mathrm{F}}_{\boldsymbol{p}}) = p_M \kappa_G(M) + p_\Pi \kappa_G(\Pi) + p_W \kappa_G(W) = p_M - p_W \in [-1, 1]. \tag{3.3}$$

The following proposition provides the worst and best asymptotic variances and their attainers when $\mathcal{D} = \mathcal{C}^{\mathrm{F}}$.

Proposition 3.2.5 (Worst and best asymptotic variances for Fréchet copulas)**.** For a concordance-inducing function $G \in \mathcal{G}_4$, the worst and best asymptotic variances on $\mathcal{C}^{\mathrm{F}}$ are given by

$$\overline{\sigma}^2_G(\mathcal{C}^{\mathrm{F}}) = 1 + \mathrm{Var}_G(X^2) \quad \text{and} \quad \underline{\sigma}^2_G(\mathcal{C}^{\mathrm{F}}) = 1 \wedge \mathrm{Var}_G(X^2)$$

with the sets of attaining copulas given by

$$\overline{C}_G(\mathcal{C}^{\mathrm{F}}) = \operatorname*{argsup}_{C \in \mathcal{C}^{\mathrm{F}}} \sigma^2_G(C) = \begin{cases} \left\{\frac{M+W}{2}\right\} & \text{if } \mathrm{Var}_G(X^2) > 0, \\ \left\{p\frac{M+W}{2} + (1-p)\Pi : p \in [0,1]\right\} & \text{if } \mathrm{Var}_G(X^2) = 0, \end{cases}$$

and

$$\underline{C}_G(C^{\mathrm{F}}) = \operatorname*{arginf}_{C \in C^{\mathrm{F}}} \sigma^2_G(C) = \begin{cases} \{M, W\} & \text{if } 0 \leq \mathrm{Var}_G(X^2) < 1, \\ \{M, W, \Pi\} & \text{if } \mathrm{Var}_G(X^2) = 1, \\ \{\Pi\} & \text{if } 1 < \mathrm{Var}_G(X^2), \end{cases}$$

respectively.

Proof. Fix $G \in \mathcal{G}_4$ and $C^{\mathrm{F}}_{\boldsymbol{p}} \in \mathcal{C}^{\mathrm{F}}$ with $\boldsymbol{p} = (p_M, p_\Pi, p_W) \in \Delta_3$. For $X = G^{-1}(U)$ and $Y = G^{-1}(V)$ with $(U, V) \sim C^{\mathrm{F}}_{\boldsymbol{p}}$, we have that $\mathrm{Cov}(X^2, Y^2) = (p_M + p_W)\,\mathrm{Var}_G(X^2)$ and $\mathrm{Cov}(X, Y) = p_M - p_W$. Therefore, by (3.1),

$$\sigma^2_G(C^{\mathrm{F}}_{\boldsymbol{p}}) = (p_M + p_W)v + 1 - (p_M - p_W)^2 =: f(p_M, p_W),$$

where $v = \mathrm{Var}_G(X^2)$. Since the Hessian of f

$$H(p_M, p_W) = \begin{pmatrix} \frac{\partial}{\partial p_M^2} f(p_M, p_W) & \frac{\partial}{\partial p_M p_W} f(p_M, p_W) \\ \frac{\partial}{\partial p_W p_M} f(p_M, p_W) & \frac{\partial}{\partial p_W^2} f(p_M, p_W) \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix},$$

is nonpositive definite, f is a concave function. For $(p_M, p_W) \in \mathbb{R}^2$ such that $0 \le p_M, p_W$ and $p_M + p_W \le 1$, consider the reparametrization $(p, 0) + r(-1, 1) = (p - r, r)$ where $0 \le r \le p \le 1$. Then

$$f(p - r, r) = pv + 1 - (p - r)^2 - r^2 + 2(p - r)r = -4\left(r - \frac{p}{2}\right)^2 + pv + 1,$$

and thus f represents a parabolic cylinder. For a fixed $p \in [0, 1]$, the function $r \mapsto f(p - r, r)$ has a maximum $\overline{f}(p) = pv + 1$ when $r = p/2$, and a minimum $\underline{f}(p) = -p^2 + pv + 1$ when $r = 0$ or $r = p$. Since $v \ge 0$, the maximum of f is given by $v + 1$ with the maximum attained by $p = 1$ when $v > 0$, and by any $p \in [0, 1]$ when $v = 0$. Therefore, we have that $\overline{\sigma}^2_G(\mathcal{C}^{\mathrm{F}}) = v + 1 = \sigma^2_G(C)$ with $C = \frac{M+W}{2}$ when $v > 0$, and with $C = p\left(\frac{M+W}{2}\right) + (1 - p)\Pi$ for any $p \in [0, 1]$ when $v = 0$. For the minimum of f, notice that the function $\underline{f}(p) = -p^2 + pv + 1$, $0 \le p \le 1$, is a concave parabola, and thus the minimum of $\underline{f}(p)$ is attained at $p = 0$ or $p = 1$. With $\underline{f}(0) = 1$ and $\underline{f}(1) = v$, the minimum of f and its attainers are given by $\underline{\sigma}^2_G(\mathcal{C}^{\mathrm{F}}) = 1 \wedge v = \sigma^2_G(C)$ with $C = M$ or W when $0 \le v < 1$, with $C = M$, W or Π when $v = 1$ and with $C = \Pi$ when $v > 1$. □

Note that although $(p_M, p_W) = (1/2, 1/2)$ is the unique point attaining the maximum $v + 1$ of f when $v > 0$, f takes the value v at the points $(p_M, p_W) = (1, 0)$ and $(0, 1)$, and is greater than v on $\{(p_M, p_W) \in [0, 1]^2 : p_M + p_W = 1\}$. Therefore, if $\mathrm{Var}_G(X^2)$ is sufficiently large, the asymptotic variance $\sigma^2_G(C)$ takes large values in $[\mathrm{Var}_G(X^2), \mathrm{Var}_G(X^2) + 1]$ if $C = pM + (1 - p)W$ for $p \in [0, 1]$.

Proposition 3.2.5 immediately leads to the optimal worst and best asymptotic variances on $\mathcal{D} = \mathcal{C}^{\mathrm{F}}$ by the following corollary.

Corollary 3.2.6 (Optimal worst and best asymptotic variances for Fréchet copulas)**.** For $\mathcal{H} \subseteq \mathcal{G}_4$, the optimal worst and best asymptotic variances are given by

$$\overline{\sigma}^2_*(\mathcal{H}, \mathcal{C}^{\mathrm{F}}) = 1 + \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2) \quad \text{and} \quad \underline{\sigma}^2_*(\mathcal{H}, \mathcal{C}^{\mathrm{F}}) = 1 \wedge \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2),$$

with the sets of attaining concorance inducing functions

$$\overline{G}_*(\mathcal{H}, \mathcal{C}^{\mathrm{F}}) = \operatorname*{arginf}_{G \in \mathcal{H}} \mathrm{Var}_G(X^2),$$

$$\underline{G}_*(\mathcal{H}, \mathcal{C}^{\mathrm{F}}) = \begin{cases} \operatorname{arginf}_{G \in \mathcal{H}} \mathrm{Var}_G(X^2), & \text{when } \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2) < 1, \\ \mathcal{H}, & \text{when } \inf_{G \in \mathcal{H}} \mathrm{Var}_G(X^2) \ge 1, \end{cases}$$

respectively.

Compared with the optimal worst and best asymptotic variances from Proposition 3.2.4 Part 3, the lower bound $\underline{\sigma}^2_*(\mathcal{H}, \mathcal{D})$ obtained in Proposition 3.2.6 remains unchanged but the upper bound $\overline{\sigma}^2_*(\mathcal{H}, \mathcal{D})$ increases since the attaining copulas $p\left(\frac{M+W}{2}\right) + (1 - p)\Pi$, $p \in [0, 1]$, are not included in the set $\mathcal{D}$ in Proposition 3.2.4. Similar to the results obtained in Proposition 3.2.4, a small variance of X^2 for $X \sim G$ is preferable in terms of optimal worst and best asymptotic variances.

Remark 3.2.7 (Restrictions of $\mathcal{C}^{\mathrm{F}}$)**.** For a concordance-inducing function $G \in \mathcal{G}_4$, consider the set of Fréchet copulas such that its transformed rank correlation κ_G takes values in $[\underline{k}, \overline{k}]$ for $-1 \le \underline{k} \le \overline{k} \le 1$, that is,

$$\mathcal{C}^{\mathrm{F}}_{\underline{k},\overline{k}}(G) = \{C \in \mathcal{C}^{\mathrm{F}} : \underline{k} \le \kappa_G(C) \le \overline{k}\}.$$

By (3.3), the restriction $\underline{k} \le \kappa_G(C) \le \overline{k}$ reduces to $\underline{k} \le p_M - p_W \le \overline{k}$ and thus $\mathcal{C}^{\mathrm{F}}_{\underline{k},\overline{k}}(G)$ does not depend on the choice of G. Consequently, the maximum and minimum of the asymptotic variance $\sigma^2_G(C)$ on $\mathcal{C}^{\mathrm{F}}_{\underline{k},\overline{k}}(G)$ can be found by calculating $\max f(p_M, p_W)$ and $\min f(p_M, p_W)$ subject to $0 \le p_M, p_W,\ p_M + p_W \le 1$ and $\underline{k} \le p_M - p_W \le \overline{k}$. This maximum and minimum always exist since $(p_M, p_W) \mapsto f(p_M, p_W)$ is bounded, concave and the feasible set is compact in $\mathbb{R}^2$.

3.2.4 Optimality of Blomqvist's beta

In this section, we show that Blomqvist's beta attains the optimal best and worst asymptotic variances under mild conditions on $\mathcal{D} \in \mathcal{C}_2$. The conditions are related to the following properties of copulas.

Definition 3.2.8 (Balancedness of copulas)**.** A copula $C \in \mathcal{C}_2$ is called *balanced* if $p(C) = 1/2$ where $p(C) = C(1/2, 1/2) + \bar{C}(1/2, 1/2)$, *imbalanced* if $p(C) \neq 1/2$, *totally positively imbalanced (TPI)* if $p(C) = 1$ and *totally negatively imbalanced (TNI)* if $p(C) = 0$.

It is straightforward to check that Π is balanced, M is TPI and W is TNI. The following proposition provides the optimal bounds of the worst and best asymptotic variances of Blomqvist's beta.

Proposition 3.2.9 (Asymptotic variance for Blomqvist's beta)**.** For any $\mathcal{D} \subseteq \mathcal{C}_2$, we have that

$$0 \le \underline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \mathcal{D}) \le \overline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \mathcal{D}) = 1.$$

The upper bound $\overline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \mathcal{D}) = 1$ is attained if and only if $\mathcal{D}$ contains a balanced copula, and the lower bound $\underline{\sigma}^2_*(\mathcal{H}_{\mathrm{Bern}}, \mathcal{D}) = 0$ is attained if and only if $\mathcal{D}$ contains a TPI or TNI copula.

Proof. For $\mathcal{H}_{\mathrm{Bern}} = \{G_{\mathrm{Bern}}\}$, we have that $G^{-1}_{\mathrm{Bern}}(u) = \mathbf{2}\mathbf{1}_{\{u>1/2\}} - 1$, $u \in [0,1]$. Therefore, for $(X, Y) = (G^{-1}(U), G^{-1}(V))$ with $(U, V) \sim C$, we have that

$$XY = \begin{cases} 1, & \text{if } \{U \le 1/2,\ V \le 1/2\} \cup \{U > 1/2,\ V > 1/2\}, \\ -1, & \text{if } \{U > 1/2,\ V \le 1/2\} \cup \{U \le 1/2,\ V > 1/2\}. \end{cases}$$

Denoting $p(C) = \mathbb{P}\left(\{U \le 1/2,\ V \le 1/2\} \cup \{U > 1/2,\ V > 1/2\}\right) = C(1/2, 1/2) + \bar{C}(1/2, 1/2)$, the asymptotic variance of $\kappa_{G_{\mathrm{Bern}}}(C)$ is given by $\sigma^2_{G_{\mathrm{Bern}}}(C) = \mathrm{Var}(XY) = 4p(C)(1-p(C))$, which attains its maximum 1 if and only if $p(C) = 1/2$ and attains its minimum 0 if and only if $p(C) = 0$ or 1. Therefore, the desired results follow. □

Remark 3.2.10 (Explicit forms of $\sigma^2_{G_{\mathrm{Bern}}}(C)$)**.** Since $C(1/2,1/2) = \bar{C}(1/2,1/2)$ for any $C \in \mathcal{C}_2$, we have that $p(C) = C(1/2,1/2) + \bar{C}(1/2,1/2) = 2C(1/2,1/2)$. Together with $\kappa_{G_{\mathrm{Bern}}}(C) = 4C(1/2,1/2) - 1$, we have that

$$\sigma^2_{G_{\mathrm{Bern}}}(C) = 4p(C)(1-p(C)) = (1+\kappa_{G_{\mathrm{Bern}}}(C))(1-\kappa_{G_{\mathrm{Bern}}}(C)) = 1 - \kappa^2_{G_{\mathrm{Bern}}}(C).$$

Therefore $\sigma^2_{G_{\mathrm{Bern}}}(C)$ admits an explicit form if $\kappa_{G_{\mathrm{Bern}}}(C)$ does. As an example, $\kappa_{G_{\mathrm{Bern}}}(C) = \frac{2}{\pi}\arcsin(\rho)$ when C is an *elliptical copula* with correlation parameter $\rho \in [-1,1]$. Therefore, $\sigma^2_{G_{\mathrm{Bern}}}(C) = 1 - (\frac{2}{\pi}\arcsin(\rho))^2$, which coincides with the result derived in Schmid and Schmidt (2007, Proposition 9).

With the bounds obtained in Proposition 3.2.9, we can prove the following optimality of Blomqvist's beta.

Corollary 3.2.11 (Optimality of Blomqvist's beta)**.** Consider $\mathcal{D} \subseteq \mathcal{C}_2$ and $\mathcal{H}_{\mathrm{Bern}} \subseteq \mathcal{H}$ for $\mathcal{H} \subseteq \mathcal{G}_4$. If $\Pi \in \mathcal{D}$, then

$$\overline{\sigma}^2_*(\mathcal{H},\mathcal{D}) = 1 \text{ and } \mathcal{H}_{\mathrm{Bern}} \subseteq \overline{G}_*(\mathcal{H},\mathcal{D}). \tag{3.4}$$

If $\mathcal{D}$ includes at least one TPI or TNI copula, then

$$\underline{\sigma}^2_*(\mathcal{H},\mathcal{D}) = 0 \text{ and } \mathcal{H}_{\mathrm{Bern}} \subseteq \underline{G}_*(\mathcal{H},\mathcal{D}). \tag{3.5}$$

Proof. Since $\Pi \in \mathcal{D}$ is balanced, Proposition 3.2.4 Part 1 and Proposition 3.2.9 imply that

$$\sup_{C\in\mathcal{D}} \sigma^2_{G_{\mathrm{Bern}}}(C) = \sigma^2_{G_{\mathrm{Bern}}}(\Pi) = 1 = \sigma^2_G(\Pi) \le \sup_{C\in\mathcal{D}} \sigma^2_G(C)$$

for any $G \in \mathcal{G}_4$. Therefore, the desired results in (3.4) follow. If $\mathcal{D}$ includes at least one TPI or TNI copula denoted by C^*, then Proposition 3.2.9 means that

$$\inf_{C\in\mathcal{D}} \sigma^2_{G_{\mathrm{Bern}}}(C) = \sigma^2_{G_{\mathrm{Bern}}}(C^*) = 0 \le \inf_{C\in\mathcal{D}} \sigma^2_G(C)$$

for any $G \in \mathcal{G}_4$, and thus the results in (3.5) follow. □

Corollary 3.2.11 states that Blomqvist's beta attains the optimal worst asymptotic variance when $\Pi \in \mathcal{D}$, and it attains the optimal best asymptotic variance when some TPI or TNI copula is contained in $\mathcal{D}$. These conditions on $\mathcal{D}$ are mild and satisfied for typical choices of $\mathcal{D}$, such as $\mathcal{C}_2$, $\mathcal{C}_2^{\succeq} = \{C \in \mathcal{C}_2 : C \succeq \Pi\}$ or $\mathcal{C}_2^{\preceq} = \{C \in \mathcal{C}_2 : C \preceq \Pi\}$. Therefore, Blomqvist's beta is typically an optimal choice among transformed rank correlations in terms of both the worst and best asymptotic variances.

In the remainder of this section, we discuss uniqueness of the optimality of Blomqvist's beta, that is, $\overline{G}_*(\mathcal{H},\mathcal{D}) = \mathcal{H}_{\mathrm{Bern}}$ and $\underline{G}_*(\mathcal{H},\mathcal{D}) = \mathcal{H}_{\mathrm{Bern}}$.

We first consider the optimal worst asymptotic variance. For given $\mathcal{H} \subseteq \mathcal{G}_4$ and $\mathcal{D} \subseteq \mathcal{C}_2$, assume that $\mathcal{H}_{\mathrm{Bern}} \subseteq \mathcal{H}$ and that $\mathcal{D}$ contains $C_* = (M+W)/2$. For any $G \in \mathcal{H}$ and $C \in \mathcal{D}$, (3.1) yields

$$\sigma^2_G(C) = \mathrm{Var}_G(X^2)\rho(X^2,Y^2) + 1 - \rho^2(X,Y) \le \mathrm{Var}_G(X^2) + 1$$

with equality if (and only if when $\mathrm{Var}_G(X^2) > 0$) $(X, Y) = (G^{-1}(U), G^{-1}(V))$ with $(U, V) \sim C$ satisfies $\rho(X^2, Y^2) = 1$ and $\rho(X, Y) = 0$. For any $G \in \mathcal{H}$, the copula C_* satisfies these conditions. To see this, let $(X_*, Y_*) = (G^{-1}(U_*), G^{-1}(V_*))$ with $(U_*, V_*) \sim C_*$. By Proposition 3.7.1 in Section 3.7.1, the copula of (X_*^2, Y_*^2) is M and thus $\rho(X_*^2, Y_*^2)$ attains the maximal correlation, which equals 1 since $X_*^2 \overset{\mathrm{d}}{=} Y_*^2$. Therefore, we have that $\rho(X_*^2, Y_*^2) = 1$ and $\rho(X_*, Y_*) = \kappa_G(C_*) = 0$. The worst asymptotic variance is then given by $\overline{\sigma}_G^2(\mathcal{D}) = \mathrm{Var}_G(X^2) + 1$ provided that $C_* \in \mathcal{D}$. Therefore, the optimal worst asymptotic variance $\overline{\sigma}_*^2(\mathcal{D}) = 1$ is attained if and only if $\mathrm{Var}_G(X^2) = 0$, which leads to $\overline{G}_*(\mathcal{H}, \mathcal{D}) = \mathcal{H}_{\mathrm{Bern}}$.

We next consider the optimal best asymptotic variance. For given $\mathcal{H} \subseteq \mathcal{G}_4$ and $\mathcal{D} \subseteq \mathcal{C}_2$, assume that $\mathcal{H}_{\mathrm{Bern}} \subseteq \mathcal{H}$ and that $\mathcal{D}$ contains at least one TPI or TNI copula. Then a given $G \in \mathcal{H}$ satisfies $G \in \underline{G}_*(\mathcal{H}, \mathcal{D})$ if and only if $\underline{\sigma}_G^2(\mathcal{D}) = \underline{\sigma}_{G_{\mathrm{Bern}}}^2(\mathcal{D}) = 0$. Therefore, the following equivalence holds:

$$\begin{aligned} G \in \underline{G}_*(\mathcal{H}, \mathcal{D}) \quad &\Leftrightarrow \quad \text{there exists } C \in \mathcal{D} \text{ s.t. } \sigma_G^2(C) = 0 \\ &\Leftrightarrow \quad G^{-1}(U)G^{-1}(V) \overset{\text{a.s.}}{=} a \text{ for some } a \in \mathbb{R} \text{ and } (U, V) \sim C \in \mathcal{D}. \end{aligned} \tag{3.6}$$

The following proposition provides necessary conditions on $a \in \mathbb{R}$, $G \in \mathcal{H}$ and $C \in \mathcal{D}$ under (3.6).

Proposition 3.2.12 (Necessary conditions on $G \in \underline{G}_*(\mathcal{H}, \mathcal{D})$)**.** Let $G \in \mathcal{H}$ be a given concordance-inducing function such that $G \in \underline{G}_*(\mathcal{H}, \mathcal{D})$. Then $C \in \mathcal{D}$ and $a \in \mathbb{R}$ in (3.6) satisfy the following conditions.

(C1) If $\mathbb{P}(X = 0) > 0$ for $X \sim G$, then $a = 0$ and $\mathbb{P}(X = 0) \geq 1/2$.

(C2) If $\mathbb{P}(X = 0) = 0$, then $a \neq 0$ and the copula C is either TPI or TNI with $0 < a \leq 1$ if C is TPI and $-1 \leq a < 0$ if C is TNI. Moreover, the distribution function $G_+(x) = 2G(x) - 1$, $x > 0$ satisfies $\mathbb{E}_{G_+}[Z] \geq |a|^{1/2}$, $Z \sim G_+$, and

$$G_+(x) = 1 - G_+\left(\frac{|a|}{x}-\right), \quad x > 0. \tag{3.7}$$

In particular it holds that $\mathbb{P}(Z > |a|^{1/2}) = \mathbb{P}(Z < |a|^{1/2})$ for $Z \sim G_+$.

Proof. For $G \in \mathcal{H}$ and $C \in \mathcal{D}$ in (3.6), write $(X, Y) = (G^{-1}(U), G^{-1}(V))$. Under (C1), we have that $\mathbb{P}(XY = 0) > 0$ and thus $a \in \mathbb{R}$ in (3.6) necessarily has to be $a = 0$. If $XY \overset{\text{a.s.}}{=} 0$ holds, then $X \neq 0$ implies that $Y = 0$. Together with $X \overset{\mathrm{d}}{=} Y$, we have that

$$\mathbb{P}(X \neq 0) \leq \mathbb{P}(Y = 0) = \mathbb{P}(X = 0),$$

which leads to the condition $\mathbb{P}(X = 0) > 1/2$.

Next we consider (C2). Since

$$XY \begin{cases} > 0, & \text{if } \{U \leq 1/2,\ V \leq 1/2\} \cup \{U > 1/2,\ V > 1/2\}, \\ < 0, & \text{if } \{U \leq 1/2,\ V > 1/2\} \cup \{U > 1/2,\ V \leq 1/2\}, \end{cases}$$

we have that $\mathbb{P}(XY = 0) = 0$ and thus $a \in \mathbb{R}$ in (3.6) necessarily has to be $a \neq 0$. Since $\mathbb{P}(XY > 0) = p(\{U \leq 1/2,\ V \leq 1/2\} \cup \{U > 1/2,\ V > 1/2\}) = p(C)$ and $\mathbb{P}(XY < 0) = p(\{U \leq 1/2,\ V > 1/2\} \cup \{U >$

$1/2,\ V \le 1/2\}) = 1 - p(C)$, the product XY can never be a constant a.s. if $0 < p(C) < 1$. Therefore, $p(C) = 0$ or 1, and thus C is either TPI or TNI.

Assume that C is TPI. Then $a > 0$ since $\mathbb{P}(XY > 0) = 1$. By the TPI assumption of C, we have that

$$\begin{aligned}
X_+ &= X \mid \{U > 1/2, V > 1/2\} = X \mid \{U > 1/2\} \sim G_+, \\
Y_+ &= Y \mid \{U > 1/2, V > 1/2\} = Y \mid \{V > 1/2\} \sim G_+, \\
X_- &= X \mid \{U \le 1/2, V \le 1/2\} = X \mid \{U \le 1/2\} \sim G_-, \\
Y_- &= X \mid \{U \le 1/2, V \le 1/2\} = Y \mid \{V \le 1/2\} \sim G_-,
\end{aligned}$$

where

$$G_+(x) = \begin{cases} 2G(x) - 1, & \text{if } x > 0, \\ 0, & \text{if } x \le 0, \end{cases} \quad \text{and} \quad G_-(x) = \begin{cases} 1, & \text{if } x > 0, \\ 2G(x), & \text{if } x \le 0. \end{cases}$$

In addition to the equalities $X_+ \overset{\text{d}}{=} Y_+$ and $X_- \overset{\text{d}}{=} Y_-$, we have that $X_+ \overset{\text{d}}{=} -X_-$ and $Y_+ \overset{\text{d}}{=} -Y_-$ since

$$\begin{aligned}
\mathbb{P}(-X_- \le x) &= \mathbb{P}(X_- \ge -x) = 1 - G_-((-x)-) \\
&= \begin{cases} 1 - 1 = 0 & \text{if } x < 0 \\ 1 - 2G((-x)-) = 1 - 2(1 - G(x)) = 2G(x) - 1 & \text{if } x \ge 0 \end{cases} \\
&= G_+(x)
\end{aligned}$$

by radial symmetry of G and the assumption $\mathbb{P}(X = 0) = 0$. Moreover, since $XY \overset{\text{a.s.}}{=} a$, it holds that

$$\begin{aligned}
X_+Y_+ &= XY \mid \{U > 1/2, V > 1/2\} \overset{\text{a.s.}}{=} a, \\
X_-Y_- &= XY \mid \{U \le 1/2, V \le 1/2\} \overset{\text{a.s.}}{=} a.
\end{aligned}$$

Since $X_+Y_+ \overset{\text{a.s.}}{=} a$ and $Y_+ > 0$ a.s., we have that $X_+ \overset{\text{a.s.}}{=} a/Y_+$. Therefore, Jensen's inequality implies that

$$\mathbb{E}[X_+] = \mathbb{E}\left[\frac{a}{Y_+}\right] = a\mathbb{E}\left[\frac{1}{Y_+}\right] \ge \frac{a}{\mathbb{E}[Y_+]} = \frac{a}{\mathbb{E}[X_+]},$$

which yields the mean condition $\mathbb{E}[X_+] \ge \sqrt{a}$.

Since $X_+ \overset{\text{d}}{=} -X_-$ and $\mathrm{Var}(X) = \mathbb{E}[X^2] = 1$, we have that

$$\begin{aligned}
1 = \mathbb{E}[X^2] &= \mathbb{P}\left(U > \frac{1}{2}\right)\mathbb{E}\left[X^2 \;\middle|\; U > \frac{1}{2}\right] + \mathbb{P}\left(U \le \frac{1}{2}\right)\mathbb{E}\left[X^2 \;\middle|\; U \le \frac{1}{2}\right] \\
&= \frac{1}{2}\mathbb{E}[X_+^2] + \frac{1}{2}\mathbb{E}[X_-^2] = \mathbb{E}[X_+^2],
\end{aligned}$$

and thus $\mathbb{E}[X_+^2] = 1$. Using $X_+^2 \overset{\text{a.s.}}{=} (a/Y_+)^2 > 0$ a.s. and Jensen's inequality, we have that

$$1 = \mathbb{E}[X_+^2] = \mathbb{E}\left[\left(\frac{a}{Y_+}\right)^2\right] \ge \frac{a^2}{\mathbb{E}[Y_+^2]} = \frac{a^2}{\mathbb{E}[X_+^2]},$$

which yields $-1 \le a \le 1$. Together with $a > 0$, we have the inequalities $0 < a \le 1$. Moreover, $X_+ \overset{\text{a.s.}}{=} a/Y_+$ implies that, for $x > 0$,

$$G_+(x) = \mathbb{P}(X_+ \le x) = \mathbb{P}\left(\frac{a}{Y_+} \le x\right) = 1 - \mathbb{P}\left(Y_+ < \frac{a}{x}\right) = 1 - G_+\left(\frac{a}{x}-\right),$$

which leads to identity (3.7). The symmetry $\mathbb{P}(Z > a^{1/2}) = \mathbb{P}(Z < a^{1/2})$ for $Z \sim G_+$ is obtained as a special case by taking $x = \sqrt{a} > 0$ in (3.7).

Next assume that C is TNI. Then $a < 0$ since $\mathbb{P}(XY < 0) = 1$. By the TNI assumption, we have that

$$\begin{aligned}
X_+ &= X \mid \{U > 1/2, V \le 1/2\} = X \mid \{U > 1/2\} \sim G_+, \\
Y_+ &= Y \mid \{U \le 1/2, V > 1/2\} = Y \mid \{V > 1/2\} \sim G_+, \\
X_- &= X \mid \{U \le 1/2, V > 1/2\} = X \mid \{U \le 1/2\} \sim G_-, \\
Y_- &= X \mid \{U > 1/2, V \le 1/2\} = Y \mid \{V \le 1/2\} \sim G_-.
\end{aligned}$$

As in the TPI case, it holds that $X_+ \overset{\text{d}}{=} Y_+$, $X_- \overset{\text{d}}{=} Y_-$, $X_+ \overset{\text{d}}{=} -X_+$ and $Y_+ \overset{\text{d}}{=} -Y_+$. Moreover, $XY \overset{\text{a.s.}}{=} a$ implies that

$$\begin{aligned}
X_+Y_- = XY \mid \{U > 1/2, V \le 1/2\} &\overset{\text{a.s.}}{=} a, \\
X_-Y_+ = XY \mid \{U \le 1/2, V > 1/2\} &\overset{\text{a.s.}}{=} a.
\end{aligned}$$

From these equalities, all the necessary conditions derived in the TPI case hold since $X_+(-Y_-) \overset{\text{a.s.}}{=} (-X_-)Y_+ \overset{\text{a.s.}}{=} -a$ with $-Y_-,\ -X_- \sim G_+$ and $-a > 0$. □

By Proposition 3.2.12, not any concordance-inducing function and copula can attain the optimal best asymptotic variance $\sigma_G^2(C) = 0$. The following examples show non-Bernoulli concordance-inducing functions attaining this lower bound.

Example 3.2.13 (Non-Bernoulli concordance-inducing functions in $\underline{G}_*(\mathcal{G}_4, \mathcal{D})$)**.**

1. *The case when* $\mathbb{P}(X = 0) > 0$: Let $X \sim G$ be an equally weighted mixture of 0 and $\mathrm{Unif}(-\sqrt{6}, \sqrt{6})$. Then $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = 1$ and $\mathbb{E}[X^4] < \infty$, and thus $G \in \mathcal{G}_4$. The case falls under (C1) since $\mathbb{P}(X = 0) = 1/2$. Let $M(n, \{J_i\}, \pi, w)$ denote a shuffle-of-M with n being the number of connected components in its support, $\{J_i\} = \{J_1, \dots, J_n\}$ being a finite partition of $[0, 1]$ into n closed subintervals, π being a permutation of $\{1, \dots, n\}$ and $w : \{1, \dots, n\} \to \{-1, 1\}$ being a function indicating whether the strip $J_i \times J_{\pi(i)}$ is flipped ($w(i) = 1$) or not ($w(i) = -1$); see Nelsen (2006, Section 3.2.3). Consider $C_1 = M(4, \cup_{i=1}^4 [(i-1)/4, i/4], \{2, 1, 4, 3\}, \mathbf{1}_4)$, $C_2 = M(4, \cup_{i=1}^4 [(i-1)/4, i/4], \{3, 4, 1, 2\}, \mathbf{1}_4)$ and $C_3 = M(4, \cup_{i=1}^4 [(i-1)/4, i/4], \{2, 4, 1, 3\}, \mathbf{1}_4)$. Then C_1 is TPI, C_2 is TNI and C_3 is neither TPI nor TNI. Moreover, all these shuffle-of-Ms satisfy $\sigma_G^2(C_k) = 0$ for $k = 1, 2, 3$ since $G^{-1}(U)G^{-1}(V) \overset{\text{a.s.}}{=} 0$ with $(U, V) \sim C_k$ for $k = 1, 2, 3$.

2. *The case when* $\mathbb{P}(X = 0) = 0$: Let $X \sim G$ be a discrete uniform distribution on the four points $\{-a/b, -b, b, a/b\}$ where $a = 1/\sqrt{2}$ and $b = \sqrt{1 - \sqrt{2}/2}$ with $b \approx 0.541$ and $a/b \approx 1.307$. Then it is straightforward to check that $G \in \mathcal{G}_4$. Define $(X, Y) = (G^{-1}(U), G^{-1}(V))$ with $(U, V) \sim C_4 =$

$M(4, \cup_{i=1}^{4}[(i-1)/4, i/4], \{2,1,4,3\}, -\mathbf{1}_4)$. Then $(X,Y) = (-a/b, -b)$, $(-b, -a/b)$, $(b, a/b)$ and $(a/b, b)$ are equiprobable, and thus $\sigma_G^2(C_4) = 0$ since $XY \overset{\text{a.s.}}{=} a$. This case belongs to (C2) since C_4 is TPI, $0 < a \leq 1$ and $\mathbb{E}_{G_+}[Z] \approx 0.924 > 0.841 \approx \sqrt{a}$.

To summarize the discussion on the uniqueness of the optimality of Blomqvist's beta, the uniqueness often holds in terms of the worst asymptotic variance whereas uniqueness is typically not fulfilled in terms of the best asymptotic variance. Therefore the Bernoulli concordance-inducing function can be regarded as the most desirable one in terms of the asymptotic variance since it is the unique element in $\overline{G}_*(\mathcal{H}, \mathcal{D}) \cap \underline{G}_*(\mathcal{H}, \mathcal{D})$ under certain assumptions on $\mathcal{D}$.

3.3 Comparison of κ_G and Kendall's tau

In Section 3.2.4, we showed that Blomqvist's beta provides optimal best and worst asymptotic variances under mild conditions on $\mathcal{D} \subseteq \mathcal{C}_2$. However, one of the drawbacks of Blomqvist's beta is that it depends only on the local value $C(1/2, 1/2)$ of the underlying copula C, which attributes to the fact that the corresponding concordance-inducing function G is supported only on two points. In this section, we show that Kendall's tau, a popular measure of concordance, attains the same optimal worst and best asymptotic variances as Blomqvist's beta although Kendall's tau is not a transformed rank correlation coefficient.

Kendall's tau $\tau : \mathcal{C}_2 \to \mathbb{R}$ is defined by

$$\tau(C) = 4 \int_{[0,1]^2} C(u,v) \, \mathrm{d}C(u,v) - 1, \tag{3.8}$$

and is a measure of concordance; see Scarsini (1984). Moreover, it is not a G-transformed rank correlation since τ is not linear with respect to a mixture of copulas; see Remark 2.2.10. Since $\tau(C) = \rho(\mathbf{1}_{\{U \leq \tilde{U}\}}, \mathbf{1}_{\{V \leq \tilde{V}\}})$ where $(U,V) \sim C$ and $(\tilde{U}, \tilde{V}) \sim C$ are independent, Kendall's tau admits the alternative representation

$$\tau(C) = \rho(g(U, \tilde{U}), g(V, \tilde{V})), \quad \text{where} \quad g(l,m) = \begin{cases} 1 & \text{if } l \leq m, \\ -1 & \text{if } l > m, \end{cases} \tag{3.9}$$

by invariance of ρ under location-scale transforms. According to (3.9), we consider the estimator of $\tau(C)$

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} g(U_i, \tilde{U}_i) g(V_i, \tilde{V}_i)$$

where (U_i, V_i) and $(\tilde{U}_i, \tilde{V}_i)$, $i = 1, \ldots, n$, $n \in \mathbb{N}$, are two i.i.d. samples from C. Note that we adopt the estimator $\hat{\tau}$ which is different from the standard estimator defined based on all pairs of samples so that $\hat{\tau}$ is a sum of i.i.d. samples. By the CLT, $\hat{\tau}$ satisfies the following asymptotic normality:

$$\sqrt{n} \{\hat{\tau} - \tau(C)\} \xrightarrow{d} \mathrm{N}(0, \sigma_\tau^2(C)), \quad \sigma_\tau^2(C) = \mathrm{Var}(g(U, \tilde{U}) g(V, \tilde{V})).$$

Similar to the case of G-transformed rank correlations, we consider the following best and worst asymptotic variances among the set of copulas $\mathcal{D} \subseteq C_2$ defined by

$$\underline{\sigma}_\tau^2(\mathcal{D}) = \inf_{C \in \mathcal{D}} \sigma_\tau^2(C) \quad \text{and} \quad \overline{\sigma}_\tau^2(\mathcal{D}) = \sup_{C \in \mathcal{D}} \sigma_\tau^2(C),$$

respectively. The following proposition provides the best and worst asymptotic variances of Kendall's tau.

Proposition 3.3.1 (Best and worst asymptotic variances of Kendall's tau)**.**

1. The asymptotic variance of Kendall's tau satisfies $0 \leq \sigma_\tau^2(C) \leq 1$ for all $C \in \mathcal{C}_2$.
2. For a given $C \in \mathcal{C}_2$, the upper bound $\sigma_\tau^2(C) = 1$ is attained if and only if $\tau(C) = 0$, which holds, for example, when $C = \Pi$ or $C = (M+W)/2$. More generally, $\sigma_\tau^2(C) = 1$ if C satisfies $(U, 1-V) \overset{\mathrm{d}}{=} (U, V)$ or $(1-U, V) \overset{\mathrm{d}}{=} (U, V)$ for $(U, V) \sim C$.
3. For a given $C \in \mathcal{C}_2$, the lower bound $\sigma_\tau^2(C) = 0$ is attained if and only if $C = M$ or W, that is, $\tau(C) = 1$ or -1, respectively.
4. Suppose $\mathcal{H} \subseteq \mathcal{G}_4$ and $\mathcal{D} \subseteq \mathcal{C}_2$ satisfy $\mathcal{H}_{\text{Bern}} \subseteq \mathcal{H}$ and $\Pi \in \mathcal{D}$. Then $\overline{\sigma}_\tau^2(\mathcal{D}) = \overline{\sigma}_*^2(\mathcal{H}, \mathcal{D}) = 1$.
5. Suppose $\mathcal{H} \subseteq \mathcal{G}_4$ and $\mathcal{D} \subseteq \mathcal{C}_2$ satisfy $\mathcal{H}_{\text{Bern}} \subseteq \mathcal{H}$, and $M \in \mathcal{D}$ or $W \in \mathcal{D}$. Then $\underline{\sigma}_\tau^2(\mathcal{D}) = \underline{\sigma}_*^2(\mathcal{H}, \mathcal{D}) = 0$.

Proof. Part 1). Writing $X = g(U, \tilde{U})$ and $Y = g(V, \tilde{V})$, we have that $XY = 1$ when $\{U \leq \tilde{U},\ V \leq \tilde{V}\} \cup \{U > \tilde{U},\ V > \tilde{V}\}$, and $XY = -1$ when $\{U \leq \tilde{U},\ V > \tilde{V}\} \cup \{U > \tilde{U},\ V \leq \tilde{V}\}$. Therefore, $\sigma_\tau^2(C) = \mathrm{Var}(XY) = 4p_\tau(C)(1 - p_\tau(C))$ where

$$\begin{aligned} p_\tau(C) &= \mathbb{P}(\{U \leq \tilde{U},\ V \leq \tilde{V}\} \cup \{U > \tilde{U},\ V > \tilde{V}\}) \\ &= \mathbb{P}(\{U \leq \tilde{U},\ V \leq \tilde{V}\}) + \mathbb{P}(\{U > \tilde{U},\ V > \tilde{V}\}) \\ &= 2\int_{[0,1]^2} C(u, v)\ \mathrm{d}C(u, v) = \frac{\tau(C) + 1}{2}, \end{aligned}$$

with the last equality implied by (3.8). Since $0 \leq p_\tau(C) \leq 1$, we have that $0 \leq \sigma_\tau^2(C) \leq 1$.
Part 2). The upper bound $\sigma_\tau^2(C) = 1$ is attained if and only if $p_\tau(C) = 1/2$, that is, $\tau(C) = 0$. When C satisfies $(U, 1-V) \overset{\mathrm{d}}{=} (U, V)$ or $(1-U, V) \overset{\mathrm{d}}{=} (U, V)$ for $(U, V) \sim C$, then the change of sign axiom of measures of concordance in Definition 1.2.1 implies that $\tau(U, V) = \tau(U, 1-V) = -\tau(U, V)$ or $\tau(U, V) = \tau(1-U, V) = -\tau(U, V)$, either of which yields $\tau(U, V) = 0$. The copulas Π and $(M+W)/2$ are examples of copulas satisfying $(U, 1-V) \overset{\mathrm{d}}{=} (U, V)$ and $(1-U, V) \overset{\mathrm{d}}{=} (U, V)$.
Part 3). The lower bound $\sigma_\tau^2(C) = 0$ is attained if and only if $p_\tau(C) = 1$ or 0, that is, $\tau(C) = 1$ or -1, respectively. By Embrechts et al. (2002, Theorem 3), $\tau(C) = 1$ or -1 if and only of $C = M$ or W, respectively.
Part 4). and Part 5). They are immediate consequences of Part 2, Part 3 and Corollary 3.2.11. □

Remark 3.3.2 (Explicit forms of $\sigma_\tau^2(C)$)**.** As seen in the proof of Proposition 3.3.1 Part 1, $\sigma_\tau^2(C)$ can be written as

$$\sigma_\tau^2(C) = 4p_\tau(C)(1 - p_\tau(C)) = (1 + \tau(C))(1 - \tau(C)) = 1 - \tau^2(C),$$

for any $C \in \mathcal{C}_2$. Therefore, $\sigma_\tau^2(C)$ admits an explicit form if $\tau(C)$ does. For example, $\tau_{G_{\text{Bern}}}(C) = \frac{2}{\pi}\arcsin(\rho)$ when C is an elliptical copula with correlation parameter $\rho \in [-1, 1]$; see Hult and Lindskog (2002). Therefore, $\sigma_\tau^2(C) = 1 - (\frac{2}{\pi}\arcsin(\rho))^2$, which also equals $\sigma_{G_{\text{Bern}}}^2(C)$ as derived in Remark 3.2.10.

Proposition 3.3.1 Parts 1, 2, 3 imply that the asymptotic variance of Kendall's tau has the same upper and lower bounds as those of Blomqvist's beta although different copulas may attain their bounds. By Proposition 3.3.1 Parts 4 and 5, Kendall's tau attains the optimal worst and best asymptotic variances of transformed rank correlations which are also attained by Blomqvist's beta as seen in Proposition 3.2.11. Taking into account the drawback of Blomqvist's beta that it depends only on the local value $C(1/2, 1/2)$ of a copula C, Kendall's tau can be a good alternative of Blomqvist's beta in terms of worst and best asymptotic variances.

Copulas attaining the lower bound $\underline{\sigma}_\tau^2(\mathcal{D}) = 0$ are completely characterized by Proposition 3.3.1 Part 3. Although a given copula $C \in \mathcal{D}$ attains the upper bound $\overline{\sigma}_\tau^2(\mathcal{D}) = 1$ if and only if $\tau(C) = 0$ as seen in the proof of Proposition 3.3.1 Part 2, no characterization of such copulas is known to the best of our knowledge. The following proposition provides a characterization of copulas attaining the upper bound $\overline{\sigma}_\tau^2(\mathcal{D}) = 1$ when $\mathcal{D}$ is the set of Fréchet copulas.

Proposition 3.3.3 (Characterization of copulas attaining $\overline{\sigma}_\tau^2(\mathcal{C}^{\mathrm{F}})$)**.** A Fréchet copula $C = C^{\mathrm{F}}_{(p_M, p_\Pi, p_W)} \in \mathcal{C}^{\mathrm{F}}$ attains the worst asymptotic variance $\overline{\sigma}_\tau^2(\mathcal{C}^{\mathrm{F}}) = 1$ of Kendall's tau if and only if $p_M = p_W \in [0, 1/2]$. Equivalently, C is of the form

$$C = p\frac{M + W}{2} + (1 - p)\Pi, \quad p \in [0, 1].$$

Proof. By Proposition 3.3.1 Part 2, a given copula $C \in \mathcal{C}^{\mathrm{F}}$ attains the upper bound $\overline{\sigma}_\tau^2(\mathcal{C}^{\mathrm{F}}) = 1$ if and only if $\tau(C) = 0$. For a Fréchet copula, we have that $\tau(\mathcal{C}^{\mathrm{F}}_{(p_M, p_\Pi, p_W)}) = (p_M - p_W)(p_M + p_W + 2)/3$; see Nelsen (2006, Example 5.3). Therefore, $\tau(\mathcal{C}^{\mathrm{F}}) = 0$ holds if and only if $p_M = p_W$. □

We now compare Kendall's tau and G-transformed rank correlations when taking the sample size into account. Since Representation (3.9) of Kendall's tau in terms of Pearson's correlation coefficient depends on two independent copies of $(U, V) \sim C$, the estimator $\hat{\tau}$ of $\tau(C)$ requires $2n$ samples from C to construct the estimator with an n-sum. Therefore, if the estimators $\hat{\tau}$ and $\hat{\kappa}_G$ are compared based on their actual variances (instead of their asymptotic variances) $\mathrm{Var}(\hat{\tau}) = \sigma_\tau^2(C)/n$ should be multiplied by 2 to be compared with $\mathrm{Var}(\hat{\kappa}_G) = \sigma_G^2(C)/n$. Based on this discussion, suppose that $\sigma_\tau^2(C)$ in Proposition 3.3.1 is replaced by $\sigma_\tau^{2\star}(C) = 2\sigma_\tau^2(C)$. With this modification, optimality of Kendall's tau in terms of the best asymptotic variance (Proposition 3.3.1 Part 5) remains valid since $\underline{\sigma}_\tau^{2\star}(\mathcal{D}) = 2\underline{\sigma}_\tau^2(\mathcal{D}) = 0 = \underline{\sigma}_*^2(\mathcal{G}_4, \mathcal{D})$. On the other hand, optimality of Kendall's tau in terms of the worst asymptotic variance (Proposition 3.3.1 Part 4) becomes invalid since $\overline{\sigma}_\tau^{2\star}(\mathcal{D}) = 2\overline{\sigma}_\tau^2(\mathcal{D}) = 2 > 1 = \overline{\sigma}_*^2(\mathcal{G}_4, \mathcal{D})$.

Remark 3.3.4 (Alternative estimators of Kendall's tau)**.** One could compare an estimator of the G-transformed rank correlation $\hat{\kappa}_G$ with other estimators of Kendall's tau, such as

$$\hat{\tau}^{\maltese} = \frac{1}{n}\sum_{i=1}^{n} g(U_i, U_{i+1})g(V_i, V_{i+1})$$

where (U_i, V_i), $i = 1, \ldots, n+1$, is an i.i.d. sample from C. Since $Z_n = X_n Y_n \in \{-1, 1\}$ with $X_n = g(U_i, U_{i+1})$ and $Y_n = g(V_i, V_{i+1})$ is a Markov chain with Z_l and Z_m being independent whenever $|l - m| \geq 2$, the Markov chain CLT yields $\sqrt{n}\left\{\hat{\tau}^{\maltese} - \tau(C)\right\} \xrightarrow{d} \mathrm{N}(0, \sigma_\tau^{2\maltese}(C))$ with

$$\sigma_\tau^{2\maltese}(C) = \mathrm{Var}(g(U_1, U_2)g(V_1, V_2)) + \mathrm{Cov}(g(U_1, U_2)g(V_1, V_2), g(U_2, U_3)g(V_2, V_3)), \tag{3.10}$$

where $(U_1, V_1),\ (U_2, V_2),\ (U_3, V_3) \overset{\text{iid}}{\sim} C$. Since $n+1$ samples are required to construct the estimator $\hat{\tau}^{\Phi}$, the modification factor $(n+1)/n$ is asymptotically 1 as $n \to \infty$, and thus we can directly compare $\sigma_\tau^{2\Phi}(C)$ with the asymptotic variance $\sigma_G^2(C)$ of $\hat{\kappa}_G$. One can show that the covariance term in (3.10) equals zero when $C = M$ or W, and thus $\underline{\sigma}_\tau^{2\Phi}(\mathcal{D}) = 0 = \underline{\sigma}_*^2(\mathcal{G}_4, \mathcal{D})$ if $M \in \mathcal{D}$ or $W \in \mathcal{D}$. Therefore, the conclusion that Kendall's tau attains the best asymptotic variance remains valid for the estimator $\hat{\tau}^{\Phi}$. Although the upper bound $\overline{\sigma}_\tau^{2\Phi}(\mathcal{D})$ is not known, $\hat{\tau}$ cannot be more preferable than $\hat{\tau}^{\Phi}$ since

$$\sigma_\tau^{2\Phi}(C) \le \mathrm{Var}(g(U_1, U_2)g(V_1, V_2)) + \mathrm{Var}(g(U_1, U_2)g(V_1, V_2)) \le 1 + 1 = 2 = \overline{\sigma}_\tau^{2\star}(\mathcal{D}).$$

3.4 Simulation study

In this section, we conduct a simulation study to compare the asymptotic variances $\sigma_G^2(C)$ for various copulas $C \in \mathcal{C}_2$ and concordance-inducing functions $G \in \mathcal{G}_4$. For concordance-inducing functions, we consider Bernoulli, uniform and normal distribution functions which correspond to Blomqvist's beta, Spearman's rho and van der Waerden's coefficient, respectively. For comparison, we also consider a Student t distribution function $t(\nu)$ with $\nu = 10$ degrees of freedom and a Beta distribution with shape parameters $(0.5, 0.5)$; note that both are radially symmetric, have finite fourth moments, and thus belong to $\mathcal{G}_4$. The Beta$(0.5, 0.5)$ concordance-inducing function has a different shape from the others since it puts an increasing probability mass as locations farther away from the center. Kendall's tau is also considered for comparison. Besides standardized concordance-inducing functions (mean zero and variance one), we also consider optimally shifted ones as introduced in Section 3.2.2. As underlying copulas, we consider Gaussian C_ρ^{Ga}, Student t $C_{\rho,\nu}^t$ and Clayton copulas C_θ^{Cl} where $\rho \in [-1, 1]$ is a correlation parameter, $\nu > 0$ is the degrees of freedom and $\theta \ge -1$ is a shape parameter. The experiment consists of the following three steps.

1. Set $\rho = -0.99 + 1.98k/49$ for $k = 0, 1, \ldots, 49$, $\nu = 5$ and $\theta = 2\rho/(1-\rho)$ (which yields $\tau(C_\theta^{\mathrm{Cl}}) = \rho$) in $C = C_\rho^{\mathrm{Ga}}$, $C_{\rho,\nu}^t$ and C_θ^{Cl}.
2. For each copula C in Step 1, simulate $(U_1, V_1), \ldots, (U_n, V_n) \overset{\text{iid}}{\sim} C$ with $n = 10^5$.
3. Based on the samples generated in Step 2, estimate $\sigma_G^2(C)$ and $\sigma_\tau^2(C)$ by the sample variances of $G^{-1}(U_i)G^{-1}(V_i)$, $i = 1, \ldots, n$ and of $g(U_i, U_{i+n/2})g(V_i, V_{i+n/2})$, $i = 1, \ldots, n/2$, where G is a standardized, and optimally shifted uniform, Beta$(0.5, 0.5)$, normal, $t(10)$ and symmetric Bernoulli distribution function.

The estimates of $\sigma_G^2(C)$ and $\sigma_\tau^2(C)$ computed in Step 3 are plotted in Figure 3.1. In the remainder of this section, we discuss the observations from these plots.

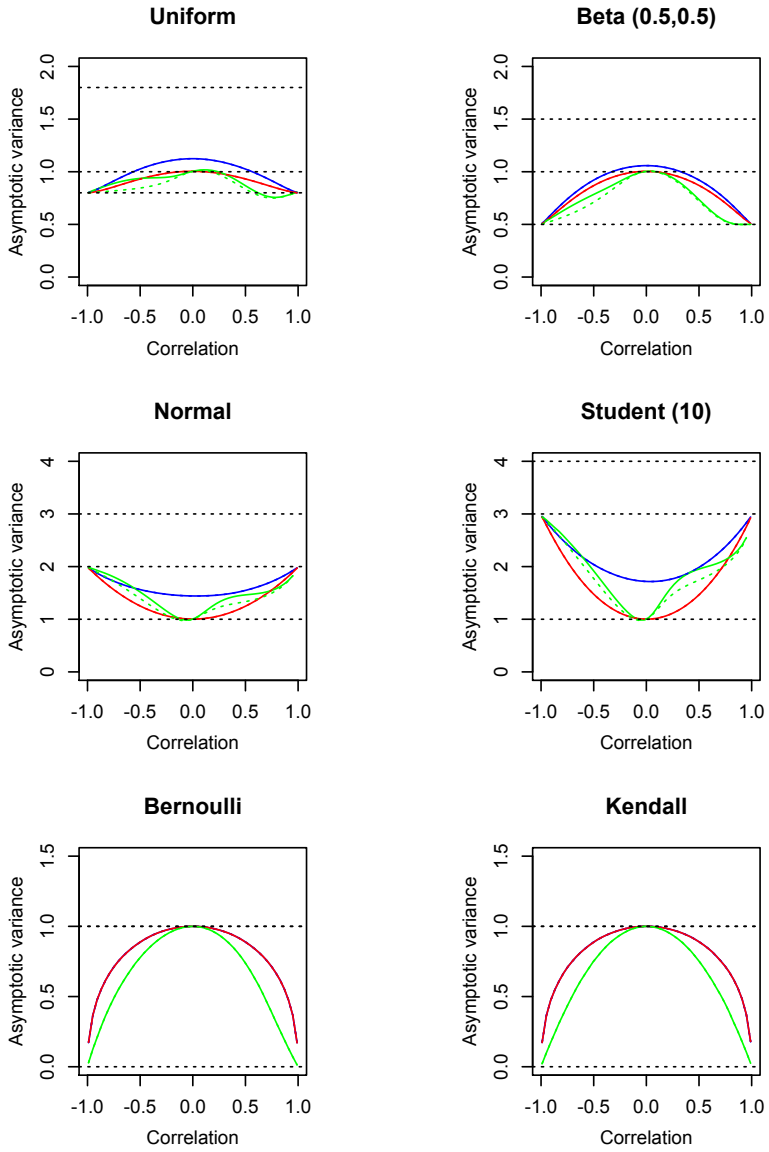

Figure 3.1: Estimates of asymptotic variances $\sigma_G^2(C)$ and $\sigma_\tau^2(C)$ against correlation parameters $\rho \in [-0.99, 0.99]$ of $C = C_\rho^{\mathrm{Ga}}$ (red), $C_{\rho,\nu}^t$ (blue) with $\nu = 5$ and C_θ^{Cl} (green) with $\theta = 2\rho/(1-\rho)$ for G-transformed rank correlation coefficients κ_G (all except bottom-right) and Kendall's tau τ (bottom-right). The concordance-inducing function G is set to be standardized (solid lines) and optimally shifted (dotted lines) uniform, Beta$(0.5, 0.5)$, normal, $t(10)$ and symmetric Bernoulli distribution. The black dotted lines represent $y = 1$, $\mathrm{Var}_G(X^2)$ and $\mathrm{Var}_G(X^2) + 1$ with $\mathrm{Var}_{G_{\mathrm{Bern}}}(X^2) = 0$, $\mathrm{Var}_{G_{\mathrm{U}}}(X^2) = 0.8$, $\mathrm{Var}_{G_{\mathrm{N}}}(X^2) = 2$, $\mathrm{Var}_{G_{t(10)}}(X^2) = 3$, $\mathrm{Var}_{G_{\mathrm{Beta}(0.5,0.5)}}(X^2) = 0.5$ and $\mathrm{Var}_\tau(X^2) = \mathrm{Var}_{G_{\mathrm{Bern}}}(X^2) = 0$.

Symmetry, convexity and concavity. For all copulas C, the curves of $\sigma_G^2(C)$ and $\sigma_\tau^2(C)$ against the correlation parameter ρ were almost symmetric around $\rho = 0$, convex when $\mathrm{Var}_G(X^2) > 1$ (which holds if G is normal or $t(10)$), and concave when $\mathrm{Var}_G(X^2) < 1$ (which holds if G is Bernoulli, uniform and $\mathrm{Beta}(0.5, 0.5)$, and if Kendall's tau is considered). For $C = C_\rho^{\mathrm{Ga}}$ and $C_{\rho,\nu}^t$, the symmetry of the curves is a consequence of the invariance of $\sigma_G^2(C)$ under partial reflections since

$$\sigma_G^2(C_{-\rho}^{\mathrm{Ga}}) = \sigma_G^2(\nu_1(C_\rho^{\mathrm{Ga}})) = \sigma_G^2(C_\rho^{\mathrm{Ga}}) \quad \text{and} \quad \sigma_G^2(C_{-\rho,\nu}^t) = \sigma_G^2(\nu_1(C_{\rho,\nu}^t)) = \sigma_G^2(C_{\rho,\nu}^t)$$

for $\rho \in [-1, 1]$ by Remark 3.2.1. This argument does not apply to Clayton copulas, and thus the curves $\rho \mapsto \sigma_G^2(C_{2\rho/(1-\rho)}^{\mathrm{Cl}})$ and $\rho \mapsto \sigma_\tau^2(C_{2\rho/(1-\rho)}^{\mathrm{Cl}})$ are nearly but not precisely symmetric.

Best and worst asymptotic variances. To see the best and worst asymptotic variances, the bounds $1 \wedge \mathrm{Var}_G(X^2)$, $1 \vee \mathrm{Var}_G(X^2)$ and $1 + \mathrm{Var}_G(X^2)$ derived in Proposition 3.2.4, Corollary 3.2.6 and Proposition 3.3.1 are plotted for each case of κ_G and Kendall's tau with $\mathrm{Var}_\tau(X^2) = \mathrm{Var}_{G_{\mathrm{Bern}}}(X^2)$. For all cases of $C = C_\rho^{\mathrm{Ga}}$, $C_{\rho,\nu}^t$ and C_θ^{Cl}, the best (smallest) $\sigma_G^2(C)$ and $\sigma_\tau^2(C)$ were roughly $1 \wedge \mathrm{Var}_G(X^2)$ and $1 \wedge \mathrm{Var}_\tau(X^2) = 0$ with the lower bound $\sigma_G^2(C) = 1$ attained at $\rho = 0$ when $\mathrm{Var}_G(X^2) > 1$ (normal or $t(10)$), and $\sigma_G^2(C) = \mathrm{Var}_G(X^2)$ attained at $\rho = \pm 1$ when $\mathrm{Var}_G(X^2) < 1$ (Bernoulli, uniform, $\mathrm{Beta}(0.5, 0.5)$ and Kendall). For all cases of copulas, the worst (largest) $\sigma_G^2(C)$ and $\sigma_\tau^2(C)$ were approximately $1 \vee \mathrm{Var}_G(X^2)$ and $1 \vee \mathrm{Var}_\tau(X^2) = 1$ although the curve was slightly above this value with $C = C_{\rho,\nu}^t$ since $C_{0,\nu}^t \neq \Pi$. The upper bound $\sigma_G^2(C) = \mathrm{Var}_G(X^2)$ was attained at $\rho = \pm 1$ when $\mathrm{Var}_G(X^2) > 1$, and $\sigma_G^2(C) = 1$ was attained at $\rho = 0$ when $\mathrm{Var}_G(X^2) < 1$ and in the case of Kendall's tau. Since we only consider specific classes of copulas, the upper bound $1 + \mathrm{Var}_G(X^2)$ derived in Corollary 3.2.6 was not attained except in the cases of Blomqvist's beta and Kendall's tau where $\mathrm{Var}_G(X^2) = 0$ and thus $1 + \mathrm{Var}_G(X^2) = 1 \vee \mathrm{Var}_G(X^2)$.

Choice of G, normal or Student t and uniform or Beta distributions. As seen for the best and worst asymptotic variances, the variance $\mathrm{Var}_G(X^2)$ is an important quantity determining the maximum and minimum of the asymptotic variance $\sigma_G^2(C)$. As theoretically indicated, concordance-inducing functions with smaller $\mathrm{Var}_G(X^2)$ are more preferable in terms of the asymptotic variance of $\hat{\kappa}_G$. Therefore, the normal concordance-inducing function is more preferable than the $t(10)$ since $\mathrm{Var}_{G_{\mathrm{N}}}(X^2) = 2 < 3 = \mathrm{Var}_{G_{t(10)}}(X^2)$. In fact, for all copulas considered, G_{N} had a smaller asymptotic variance than $G_{t(10)}$ even though $t(10)$ is already rather close to $\mathrm{N}(0, 1)$. Interestingly, the $\mathrm{Beta}(0.5, 0.5)$ concordance-inducing function typically had smaller asymptotic variance than the uniform distribution since $\mathrm{Var}_{G_{\mathrm{Beta}(0.5,0.5)}}(X^2) = 0.5 < 0.8 = \mathrm{Var}_{G_{\mathrm{Unif}}}(X^2)$. Therefore, Beta concordance-inducing functions, possibly with different parameters, can be good alternatives to Spearman's rho.

Similarity of Blomqvist's beta and Kendall's tau. The curves of asymptotic variances for Blomqvist's beta and Kendall's tau are seemingly equal for all choices of C. Moreover, the curves for $C = C_\rho^{\mathrm{Ga}}$ and $C_{\rho,\nu}^t$ overlap in the cases of Blomqvist's beta and Kendall's tau. These observations are verified for $C = C_\rho^{\mathrm{Ga}}$ and $C_{\rho,\nu}^t$ since, in these cases, we have that

$$\sigma_\tau^2(C) = \sigma_{G_{\mathrm{Bern}}}^2(C) = 1 - \left(\frac{2}{\pi}\arcsin(\rho)\right)^2$$

as seen in Remark 3.2.10 and 3.3.2. On the other hand, $\sigma^2_\tau(C^{\mathrm{Cl}}_\theta)$ and $\sigma^2_{G_{\mathrm{Bern}}}(C^{\mathrm{Cl}}_\theta)$ are in general different since $\sigma^2_\tau(C^{\mathrm{Cl}}_\theta) = 1 - \tau^2(C^{\mathrm{Cl}}_\theta)$ and $\sigma^2_{G_{\mathrm{Bern}}}(C^{\mathrm{Cl}}_\theta) = 1 - \kappa^2_{G_{\mathrm{Bern}}}(C^{\mathrm{Cl}}_\theta)$ but

$$\tau(C^{\mathrm{Cl}}_\theta) = \frac{\theta}{\theta+2} \quad \text{and} \quad \kappa_{G_{\mathrm{Bern}}}(C^{\mathrm{Cl}}_\theta) = 4(2^{\theta+1}-1)^{-1/\theta} - 1.$$

Strength of dependence and kind of copula. Compared with the choice of concordance-inducing function, the strength of dependence ρ and the kind of C seem to be less influencial on the asymptotic variance $\sigma^2_G(C)$. Furthermore, for any concordance-inducing function, the difference of $\sigma^2_G(C)$ among different copulas $C = C^{\mathrm{Ga}}_\rho$, $C^t_{\rho,\nu}$ and C^{Cl}_θ was typically smaller than difference of $\sigma^2_G(C)$ among different levels of dependence ρ.

Effect of optimal shifts. When $C = C^{\mathrm{Ga}}_\rho$ or $C^t_{\rho,\nu}$, the solid and dotted curves of asymptotic variances overlap (and thus the dotted curves are not visible). When $C = C^{\mathrm{Cl}}_\theta$, the dotted curves do not coincide with, but were close to the solid ones except in the Bernoulli (top left) and Kendall (bottom right) case when two curves seem to overlap. These observations are consistent with Proposition 3.2.3 stating that the asymptotic variance is not reduced by the optimal shift of G when $C = C^{\mathrm{Ga}}_\rho$ or $C^t_{\rho,\nu}$. Even when the copula is C^{Cl}_θ, only a small reduction of the asymptotic variance was observed when optimally shifting G.

In summary, when $\mathrm{Var}_G(X^2) < 1$, the curve of $\sigma^2_G(C)$ is typically symmetric and concave, with the maximum 1 being attained when $\rho = 0$ and the minimum $\mathrm{Var}_G(X^2)$ being attained when $\rho = \pm 1$. When $\mathrm{Var}_G(X^2) > 1$, the curve of $\sigma^2_G(C)$ is typically symmetric and convex with the maximum $\mathrm{Var}_G(X^2)$ attained when $\rho = \pm 1$ and the minimum 1 attained when $\rho = 0$. The curves are not significantly different among different choices of C when the strength of dependence remains the same. Compared with the kind of C, the strength of dependence and the choice of G are more influencial on $\sigma^2_G(C)$. Normal and Beta$(0.5, 0.5)$ concordance-inducing functions are more preferable than $t(10)$ and uniform distributions, respectively. Moreover, Blomqvist's beta and Kendall's tau perform almost the same in terms of their asymptotic variance. Finally, even when C does not satisfy the sufficient conditions of Proposition 3.2.3, the optimal shift of G may not significantly reduce the asymptotic variance $\sigma^2_G(C)$.

3.5 Concluding remark

We addressed the question which measures of concordance to use in terms of best and worst asymptotic variances of their canonical estimators. We proved that Blomqvist's beta attains the optimal best and worst asymptotic variances among all transformed rank correlation coefficients including Spearman's rho and van der Waerden's coefficient. Considering the drawback of Blomqvist's beta that it depends only on the local value $C(1/2, 1/2)$ of a copula C, we also compared transformed rank correlations with the popular measure of concordance Kendall's tau. Based on the representation of Kendall's tau in terms of Pearson's linear correlation coefficient, we found that Kendall's tau also attains the optimal best and worst asymptotic variances if estimators of these measures are compared without being standardized by sample size. Attributing to the fact that the correlation-representation of Kendall's tau depends on two independent copies from the underlying copula, the optimality of Kendall's tau may be violated if the asymptotic variances of these estimators are standardized by sample size. Through a simulation study, we

observed that the curve of the asymptotic variance of a G-transformed rank correlation against the strength of dependence of the underlying copula was typically symmetric and parabolic. Moreover, convexity, maximum and minimum of the asymptotic variance seemed to be determined by $\mathrm{Var}_G(X^2)$. The results of the simulation study supported that concordance-inducing functions G with smaller $\mathrm{Var}_G(X^2)$ are more preferable. Consequently, heavy-tailed concordance-inducing functions, such as Student t distributions with small degrees of freedom, are not recommended in comparison to the normal distribution, and Beta distributions can be good alternatives for uniform distributions (corresponding to Spearman's rho).

Other than Kendall's tau, there are still important measures of concordance which are not included in the class of transformed rank correlations, such as Gini's gamma. Studying a broader framework of comparing these measures of concordance is a part of future research. Further investigation is also required for parametric classes of concordance-inducing functions such as $\mathrm{Beta}(\alpha, \beta)$ with $\alpha = \beta > 0$. Given the limitations of fundamental copulas in practice, another direction of future work is to investgate optimal concordance-inducing functions under more practical choices of sets of underlying copulas, such as the set of parametric copulas or balls of copulas around a given reference copula. A comparison of multivariate measures of concordance, and of matrices of pairwise bivariate measures of concordance are also interesting directions for future research. Finally, it is of interest whether and how the results in this chapter change if all measures of concordance are compared in terms of their asymptotic variance without assuming that marginal distributions are known (and thus only pseudo-samples from copulas are available), and/or if the optimal location shift is applied to the concordance-inducing functions.

3.6 Miscellaneous results

Some miscellaneous results are collected in this section.

3.6.1 A class of discrete concordance-inducing functions

As stated in Schmid and Schmidt (2007), one of the advantages of Blomqvist's beta is that it admits an explicit form whenever the copula is given analytically. This advantage can be extended to a wider class of discrete concordance-inducing functions. For $m \in \mathbb{N}$, $\boldsymbol{z} = (z_1, \ldots, z_m) \in \mathbb{R}_+^m$ and $\boldsymbol{p} = (p_0, p_1, \ldots, p_m) \in \mathbb{R}_+^m$ such that $0 < z_1 < \cdots < z_m$, $p_0 + 2\sum_{i=1}^m p_i = 1$ and $\sum_{i=1}^m p_i z_i^2 = 1/2$, consider a discrete distribution $G_{m,\boldsymbol{z},\boldsymbol{p}}$ supported on $-z_m, \ldots, -z_1, 0, z_1, \ldots, z_m$ with corresponding probabilities $p_m, \ldots, p_1, p_0, p_1, \ldots, p_m$. Then $G_{m,\boldsymbol{z},\boldsymbol{p}}$ is a concordance-inducing function with mean zero and variance one. As a special case, Blomqvist's beta arises when $m = 1$, $z_1 = 1$ and $(p_0, p_1) = (0, 1/2)$. Let $p_+ = p_1 + \cdots + p_m$, $I_{-i} = [p_+ - \sum_{j=1}^i p_j,\ p_+ - \sum_{j=1}^{i-1} p_j]$, $I_0 = [p_+,\ p_+ + p_0]$ and $I_i = [p_+ + p_0 + \sum_{j=1}^{i-1} p_j,\ p_+ + p_0 + \sum_{j=1}^i p_j]$ for $i = 1, \ldots, m$. Then

$$\kappa_{G_{m,\boldsymbol{z},\boldsymbol{p}}}(C) = \mathbb{E}[G_{m,\boldsymbol{z},\boldsymbol{p}}^{-1}(U) G_{m,\boldsymbol{z},\boldsymbol{p}}^{-1}(V)] = \sum_{(i,j) \in \{-m, \ldots, m\}} z_i z_j V_C(I_i \times I_j),$$

and

$$\sigma^2_{G_{m,\boldsymbol{z},\boldsymbol{p}}}(C) = \mathrm{Var}(XY) = \mathbb{E}[(XY)^2] - (\mathbb{E}[XY])^2$$
$$= \sum_{(i,j)\in\{-m,\dots,m\}} z_i^2 z_j^2 V_C(I_i \times I_j) - \left(\sum_{(i,j)\in\{-m,\dots,m\}} z_i z_j V_C(I_i \times I_j) \right)^2,$$

where $z_{-i} = -z_i$ for $i = 1, \dots, m$ and $V_C(A)$, $A \subseteq [0,1]^2$ is a volume of A measured by C. Therefore, $\kappa_{G_{m,\boldsymbol{z},\boldsymbol{p}}}(C)$ and $\sigma^2_{G_{m,\boldsymbol{z},\boldsymbol{p}}}(C)$ admit explicit forms if $V_C(I_i \times I_j)$ can be written explicitly for all $(i,j) \in \{-m,\dots,m\}^2$.

3.7 Properties of $(G, C) \mapsto \sigma^2_G(C)$

In this section, we investigate the map $(G, C) \mapsto \sigma^2_G(C)$ for $C \in \mathcal{C}_2$ and $G \in \mathcal{G}_4$. Since $\sigma^2_G(C)$ admits the representation $\sigma^2_G(C) = \mathrm{Cov}(X^2, Y^2) + 1 - \mathrm{Cov}(X, Y)^2$ for $(X, Y) = (G^{-1}(U), G^{-1}(V))$ and $(U, V) \sim C$, we first study the joint distribution of (X^2, Y^2) in Section 3.7.1. Linearity of $C \mapsto \sigma^2_G(C)$ and continuity of $(G, C) \mapsto \sigma^2_G(C)$ are then investigated in Sections 3.7.2 and 3.7.3, respectively.

3.7.1 Joint distribution of (X^2, Y^2)

We first study the marginal distributions and copula of (X^2, Y^2) for $(X, Y) = (G^{-1}(U), G^{-1}(V))$ and $(U, V) \sim C$. First, by radial symmetry of G, the marginal distribution of X^2 (and that of Y^2) is given by

$$G^{[2]}(x) = \mathbb{P}(X^2 \le x) = \mathbb{P}(-\sqrt{x} \le X \le \sqrt{x}) = G(\sqrt{x}) - G(-\sqrt{x}-)$$
$$= 2G(\sqrt{x}) - 1, \quad x \ge 0.$$

The following proposition describes the copula of (X^2, Y^2) when G is continuous.

Proposition 3.7.1 (Copula of (X^2, Y^2))**.** Let $G \in \mathcal{G}_4^c$ be a continuous concordance-inducing function. For a copula $C \in \mathcal{C}_2$ and $(X, Y) = (G^{-1}(U), G^{-1}(V))$ with $(U, V) \sim C$, the copula of (X^2, Y^2) is given by

$$C^{[2]}(u, v) = \sum_{\varphi\in\{\iota,\nu_1,\nu_2,\nu_1\circ\nu_2\}} \bar{C}_\varphi\left(\frac{1}{2}, \frac{1}{2}\right) C_{\varphi,(1/2,1/2)}\left(\frac{u+1}{2}, \frac{v+1}{2}\right),$$

where $C_\varphi = \varphi(C)$ and $C_{\varphi,(1/2,1/2)}(u, v) = \mathbb{P}\left(U_\varphi \le u,\ V_\varphi \le v \mid U_\varphi > 1/2,\ V_\varphi > 1/2\right)$ for $(U_\varphi, V_\varphi) \sim C_\varphi$.

Proof. By continuity of G, we have that $X > 0$ when $U > 1/2$ and $X \le 0$ when $U \le 1/2$. Therefore,

$$G^{[2]}(X^2) = 2G(\sqrt{X^2}) - 1 = 2G(|X|) - 1$$
$$= \begin{cases} 2G(X) - 1 = 2U - 1, & \text{when } U > 1/2, \\ 2G(-X) - 1 = 2(1 - G(X)) - 1 = 1 - 2U, & \text{when } U \le 1/2. \end{cases}$$

Using this relationship, we have that

$$
\begin{aligned}
C^{[2]}(u,v) &= \mathbb{P}(G^{[2]}(X^2) \le u, G^{[2]}(Y^2) \le v) \\
&= \mathbb{P}\left(U > \frac{1}{2},\ V > \frac{1}{2}\right) \mathbb{P}\left(2U - 1 \le u,\ 2V - 1 \le v \;\middle|\; U > \frac{1}{2},\ V > \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(U \le \frac{1}{2},\ V > \frac{1}{2}\right) \mathbb{P}\left(1 - 2U \le u,\ 2V - 1 \le v \;\middle|\; U \le \frac{1}{2},\ V > \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(U > \frac{1}{2},\ V \le \frac{1}{2}\right) \mathbb{P}\left(2U - 1 \le u,\ 1 - 2V \le v \;\middle|\; U > \frac{1}{2},\ V \le \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(U \le \frac{1}{2},\ V \le \frac{1}{2}\right) \mathbb{P}\left(1 - 2U \le u,\ 1 - 2U \le v \;\middle|\; U \le \frac{1}{2},\ V \le \frac{1}{2}\right) \\
&= \mathbb{P}\left(U > \frac{1}{2},\ V > \frac{1}{2}\right) \mathbb{P}\left(U \le \frac{u+1}{2},\ V \le \frac{v+1}{2} \;\middle|\; U > \frac{1}{2},\ V > \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(1 - U > \frac{1}{2},\ V > \frac{1}{2}\right) \mathbb{P}\left(1 - U \le \frac{u+1}{2},\ V \le \frac{v+1}{2} \;\middle|\; 1 - U > \frac{1}{2},\ V > \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(U > \frac{1}{2},\ 1 - V > \frac{1}{2}\right) \mathbb{P}\left(U \le \frac{u+1}{2},\ 1 - V \le \frac{v+1}{2} \;\middle|\; U > \frac{1}{2},\ 1 - V > \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(1 - U > \frac{1}{2},\ 1 - V > \frac{1}{2}\right) \mathbb{P}\left(1 - U \le \frac{u+1}{2},\ 1 - V \le \frac{v+1}{2} \;\middle|\; 1 - U > \frac{1}{2},\ 1 - V > \frac{1}{2}\right), \\
&= \sum_{\varphi \in \{\iota, \nu_1, \nu_2, \nu_1 \circ \nu_2\}} \bar{C}_\varphi\left(\frac{1}{2}, \frac{1}{2}\right) C_{\varphi,(1/2,1/2)}\left(\frac{u+1}{2}, \frac{v+1}{2}\right).
\end{aligned}
$$

□

As an application of Proposition 3.7.1, let (X, Y) have a fundamental copula. Then the copula of (X^2, Y^2) is given by

$$
C^{[2]} = \begin{cases} \frac{1}{4}\Pi + \frac{1}{4}\Pi + \frac{1}{4}\Pi + \frac{1}{4}\Pi = \Pi, & \text{when } C = \Pi, \\ \frac{1}{2}M + 0 + 0 + \frac{1}{2}M = M, & \text{when } C = M, \\ 0 + \frac{1}{2}M + \frac{1}{2}M + 0 = M, & \text{when } C = W. \end{cases}
$$

Next we study the copula $C^{[2]}$ when C is a convex combination of copulas.

Lemma 3.7.2 (Convex combination of $C^{[2]}$)**.** For a convex combination $\tilde{C}_p = pC + (1-p)C'$ of C and C' where $p \in [0,1]$ and $C, C' \in \mathcal{C}_2$, we have that

$$
\tilde{C}_p^{[2]} = pC^{[2]} + (1-p)C'^{[2]},
$$

provided that the concordance-inducing function $G \in \mathcal{G}_4^c$ is continuous.

Proof. Consider a random vector $(\tilde{U}, \tilde{V}) = B(U,V) + (1-B)(U',V') \sim \tilde{C}_p$ where $(U,V) \sim C$, $(U',V') \sim C'$ and $B \sim \text{Bern}(p)$ are independent of each other. For $(X,Y) = (G^{-1}(U), G^{-1}(V))$, $(X',Y') = (G^{-1}(U'), G^{-1}(V'))$ and $(\tilde{X}, \tilde{Y}) = (G^{-1}(\tilde{U}), G^{-1}(\tilde{V}))$ we have that

$$
\begin{aligned}
(\tilde{X}, \tilde{Y}) &= (G^{-1}(BU + (1-B)U'), G^{-1}(BV + (1-B)V')) \\
&= B(G^{-1}(U), G^{-1}(V)) + (1-B)(G^{-1}(U'), G^{-1}(V')) \\
&= B(X,Y) + (1-B)(X',Y'),
\end{aligned}
$$

and thus $(\tilde{X}^2, \tilde{Y}^2) = B(X^2, Y^2) + (1-B)(X'^2, Y'^2)$. Since $\tilde{X}^2 \overset{\mathrm{d}}{=} \tilde{Y}^2 \sim G^{[2]}$, the copula of $(\tilde{X}^2, \tilde{Y}^2)$ is given by

$$\begin{aligned}
&\tilde{C}_p^{[2]}(u,v) \\
&= \mathbb{P}(G^{[2]}(BX^2 + (1-B)X'^2) \le u,\ G^{[2]}(BY^2 + (1-B)Y'^2) \le v) \\
&= \mathbb{P}(B=1)\mathbb{P}(G^{[2]}(X^2) \le u,\ G^{[2]}(Y^2) \le u) + \mathbb{P}(B=0)\mathbb{P}(G^{[2]}(X'^2) \le u,\ G^{[2]}(Y'^2) \le u) \\
&= pC^{[2]}(u,v) + (1-p)C'^{[2]}(u,v).
\end{aligned}$$

□

3.7.2 Linearity of $C \mapsto \sigma_G^2(C)$

The map $C \mapsto \sigma_G^2(C) = \mathrm{Cov}(X^2, Y^2) + 1 - \mathrm{Cov}(X,Y)^2$ is in general not linear with respect to convex combinations of copulas since $C \mapsto \mathrm{Cov}(X,Y)^2$ is quadratic. On the other hand, the map $C \mapsto \mathrm{Cov}(X^2, Y^2)$ is linear by Lemma 3.7.2, and thus, $C \mapsto \sigma_G^2(C)$ is linear on a restricted subset of $\mathcal{C}_2$ as shown by the following proposition.

Proposition 3.7.3 (Linearity of $C \mapsto \sigma_G^2(C)$)**.** For a continuous concordance-inducing function $G \in \mathcal{G}_4^{\mathrm{c}}$ and a constant $k \in [-1,1]$, the map $C \mapsto \sigma_G^2(C)$ is linear with respect to convex combinations of copulas on $\mathcal{C}_G(k) = \{C \in \mathcal{C}_2 : \kappa_G(C) = k\}$.

Proof. For $\tilde{C}_p = pC + (1-p)C'$, $C, C' \in \mathcal{C}_2$ and $(\tilde{U}, \tilde{V}) \sim \tilde{C}_p$, we have that $(\tilde{X}, \tilde{Y}) = (G^{-1}(\tilde{U}), G^{-1}(\tilde{V})) = (G^{-1}(BU + (1-B)U'), G^{-1}(BV + (1-B)V')) = B(X,Y) + (1-B)(X',Y')$ where $(X,Y) = (G^{-1}(U), G^{-1}(V))$ and $(X',Y') = (G^{-1}(U'), G^{-1}(V'))$ with $(U,V) \sim C$, $(U',V') \sim C'$ and $B \sim \mathrm{Bern}(p)$ being independent. From this representation, we have that $\mathrm{Cov}(\tilde{X}^2, \tilde{Y}^2) = p\,\mathrm{Cov}(X^2, Y^2) + (1-p)\,\mathrm{Cov}(X'^2, Y'^2)$ and $\mathrm{Cov}(\tilde{X}, \tilde{Y}) = p\,\mathrm{Cov}(X,Y) + (1-p)\,\mathrm{Cov}(X',Y')$. If $C \in \mathcal{C}_G(k)$, then $\mathrm{Cov}(X,Y) = \kappa_G(C) = k$ and $\mathrm{Cov}(X',Y') = \kappa_G(C') = k$. Therefore, we have that $\mathrm{Cov}(\tilde{X}, \tilde{Y}) = k$ and thus

$$\begin{aligned}
\sigma_G^2(\tilde{C}_p) &= \mathrm{Var}(\tilde{X}\tilde{Y}) = \mathrm{Cov}(\tilde{X}^2, \tilde{Y}^2) + 1 - \mathrm{Cov}(\tilde{X}, \tilde{Y})^2 \\
&= p\,\mathrm{Cov}(X^2, Y^2) + (1-p)\,\mathrm{Cov}(X'^2, Y'^2) + 1 - (p\,\mathrm{Cov}(X,Y) + (1-p)\,\mathrm{Cov}(X',Y'))^2 \\
&= p(\mathrm{Cov}(X^2, Y^2) + 1 - k^2) + (1-p)(\mathrm{Cov}(X'^2, Y'^2) + 1 - k^2) \\
&= p\,\mathrm{Var}(XY) + (1-p)\,\mathrm{Var}(X'Y') = p\sigma_G^2(C) + (1-p)\sigma_G^2(C'),
\end{aligned}$$

which shows the desired property. □

By Proposition 3.7.3, the supremum and infimum of $\overline{\sigma}_G^2(\mathcal{C}_G(k))$ and $\underline{\sigma}_G^2(\mathcal{C}_G(k))$ are attained by the extremal points of $\mathcal{C}_G(k)$.

3.7.3 Continuity of $(G, C) \mapsto \sigma_G^2(C)$

In this section we study continuity of the map $(G, C) \mapsto \sigma_G^2(C)$, which is important in the discussion of attainability of the suprema and infima in $\overline{\sigma}_G^2(\mathcal{D})$, $\underline{\sigma}_G^2(\mathcal{D})$, $\overline{\sigma}_*^2(\mathcal{H}, \mathcal{D})$ and $\underline{\sigma}_G^2(\mathcal{H}, \mathcal{D})$ for $\mathcal{H} \subseteq \mathcal{G}_4$ and $\mathcal{D} \subseteq \mathcal{C}_2$.

To this end, let the product space $\mathcal{G}_4 \times \mathcal{C}_2$ be metrized by the uniform norm

$$\begin{aligned} d_\infty((G,C),(G',C')) &= d_\infty(G,G') \vee d_\infty(C,C') \\ &= \sup_{x\in\mathbb{R}} |G(x) - G'(x)| \vee \sup_{(u,v)\in[0,1]^2} |C(u,v) - C'(u,v)|. \end{aligned}$$

Then the following propositions show the continuity of the maps $C \mapsto \sigma_G^2(C)$, $G \mapsto \overline{\sigma}_G^2(\mathcal{D})$ and $G \mapsto \underline{\sigma}_G^2(\mathcal{D})$.

Proposition 3.7.4 (Continuity of $C \mapsto \sigma_G^2(C)$)**.** Suppose that $\mathcal{D} \subseteq \mathcal{C}_2$ is a closed subset in $(\mathcal{C}_2, d_\infty)$. Then the map $C \mapsto \sigma_G^2(C)$ is continuous on $\mathcal{D}$ for any $G \in \mathcal{G}_4$. Therefore, the supremum and infimum in $\overline{\sigma}_G^2(\mathcal{D})$ and $\underline{\sigma}_G^2(\mathcal{D})$ are attainable and thus $\overline{C}_G(\mathcal{D}) = \operatorname{argsup}_{C\in\mathcal{D}} \sigma_G^2(C)$ and $\underline{C}_G(\mathcal{D}) = \operatorname{arginf}_{C\in\mathcal{D}} \sigma_G^2(C)$ are non-empty provided that $G \in \mathcal{G}_4$ and that $\mathcal{D} \subseteq \mathcal{C}_2$ is closed.

Proof. For a fixed $G \in \mathcal{G}_4$ and $C_n \in \mathcal{D}$, $n = 1, 2, \ldots$ and $C \in \mathcal{D}$ such that $\lim_{n\to\infty} C_n = C$ pointwise, let $(U_n, V_n) \sim C_n$, $(U, V) \sim C$, $(X_n, Y_n) = (G^{-1}(U_n), G^{-1}(V_n))$, $(X, Y) = (G^{-1}(U), G^{-1}(V))$, $H_n(x, y) = \mathbb{P}(X_n \le x, Y_n \le y)$, $H_n^{[2]}(x, y) = \mathbb{P}(X_n^2 \le x, Y_n^2 \le f)$, $H(x, y) = \mathbb{P}(X \le x, Y \le y)$ and $H^{[2]}(x, y) = \mathbb{P}(X^2 \le x, Y^2 \le f)$ for $(x, y) \in \mathbb{R}^2$. Then the pointwise convergence of $C_n \to C$ implies $H_n \to H$ and $H_n^{[2]} \to H^{[2]}$ pointwise, since $H^{[2]}(x, y) = H(\sqrt{x}, \sqrt{y}) - H(-\sqrt{x}-, \sqrt{y}) - H(\sqrt{x}, -\sqrt{y}-) + H(-\sqrt{x}-, -\sqrt{y}-)$. Therefore, by Hoeffding's identity (McNeil et al., 2015, Lemma 7.27), we have that

$$\begin{aligned} \lim_{n\to\infty} \operatorname{Cov}(X_n^2, Y_n^2) &= \lim_{n\to\infty} \iint_{\mathbb{R}^2} \left(H_n^{[2]}(x, y) - G^{[2]}(x) G^{[2]}(y) \right) \mathrm{d}x\mathrm{d}y \\ &= \iint_{\mathbb{R}^2} \left(H^{[2]}(x, y) - G^{[2]}(x) G^{[2]}(y) \right) \mathrm{d}x\mathrm{d}y \\ &= \operatorname{Cov}(X^2, Y^2). \end{aligned}$$

Interchanging the limit and the integral is justified by Lebesgue's dominated convergence theorem since the integrand $H_n^{[2]}(x, y) - G^{[2]}(x)G^{[2]}(y)$ is bounded above by $M(G^{[2]}(x), G^{[2]}(y)) - G^{[2]}(x)G^{[2]}(y)$, which is integrable as

$$\iint_{\mathbb{R}^2} \left(M(G^{[2]}(x), G^{[2]}(y)) - G^{[2]}(x) G^{[2]}(y) \right) \mathrm{d}x\mathrm{d}y$$

is the maximal covariance between $X^2 \sim G^{[2]}$ and $Y^2 \sim G^{[2]}$ over all copulas, and thus bounded above by $\operatorname{Var}(X^2) < \infty$ provided $G \in \mathcal{G}_4$. Similarly, it holds that $\lim_{n\to\infty} \operatorname{Cov}(X_n, Y_n) = \operatorname{Cov}(X, Y)$. Therefore,

$$\begin{aligned} \lim_{n\to\infty} \sigma_G^2(C_n) &= \lim_{n\to\infty} \left(\operatorname{Cov}(X_n^2, Y_n^2) + 1 - \operatorname{Cov}(X_n, Y_n)^2 \right) \\ &= \operatorname{Cov}(X^2, Y^2) + 1 - \operatorname{Cov}(X, Y)^2 = \sigma_G^2(C), \end{aligned}$$

which completes the proof.

□

Proposition 3.7.5 (Continuity of $G \mapsto \overline{\sigma}_G^2(\mathcal{D})$ and $G \mapsto \underline{\sigma}_G^2(\mathcal{D})$)**.** Let $\mathcal{H} \subseteq \mathcal{G}_4^{\mathrm{b}}$ and $\mathcal{D} \subseteq \mathcal{C}_2$ be closed subsets in $(\mathcal{G}_4^{\mathrm{b}}, d_\infty)$ and $(\mathcal{C}_2, d_\infty)$, respectively. Then the maps $G \mapsto \overline{\sigma}_G^2(\mathcal{D})$ and $G \mapsto \underline{\sigma}_G^2(\mathcal{D})$ are continuous on $\mathcal{H}$. Therefore, the infima in $\overline{\sigma}_*^2(\mathcal{H}, \mathcal{D})$ and $\underline{\sigma}_*^2(\mathcal{H}, \mathcal{D})$ are attainable and thus $\overline{G}_*(\mathcal{H}, \mathcal{D})$ and $\underline{G}_*(\mathcal{D})$ are non-empty provided that $\mathcal{H} \subseteq \mathcal{G}_4^{\mathrm{b}}$ and $\mathcal{D} \subseteq \mathcal{C}_2$ are closed.

Proof. To show the continuity of $G \mapsto \overline{\sigma}^2_G(\mathcal{D})$ and $G \mapsto \underline{\sigma}^2_G(\mathcal{D})$, it suffices to show the continuity of the maps $(G, C) \mapsto \sigma^2_G(C)$ by Berge's maximum theorem (Berge, 1997). Since $|\sigma^2_G(C) - \sigma^2_{G'}(C')| \leq |\sigma^2_G(C) - \sigma^2_{G'}(C)| + |\sigma^2_{G'}(C) - \sigma^2_{G'}(C')|$ for $C, C' \in \mathcal{D}$ and $G, G' \in \mathcal{H}$, it suffices to show the continuity of $C \mapsto \sigma^2_G(C)$ for a fixed $G \in \mathcal{H}$, and of $G \mapsto \sigma^2_G(C)$ for a fixed $C \in \mathcal{D}$. Continuity of $C \mapsto \sigma^2_G(C)$ follows from Proposition 3.7.4. Continuity of $G \mapsto \sigma^2_G(C)$ follows similarly since the pointwise convergence of $G_n \to G$ implies $H_n \to H$ and $H_n^{[2]} \to H^{[2]}$ pointwise; see Nelsen (2006, Lemma 2.1.5). Interchanging the limit and the integrals is verified by the bounded convergence theorem instead since the integrands are bounded, and ranges of the integrals are bounded sets when $G \in \mathcal{G}_4^{\mathrm{b}}$. Therefore, the desired results follow. □

Chapter 4

Estimation of VaR contributions with MCMC

Determining risk contributions for the components of a portfolio based on a given economic capital is an important task in financial risk management. Computing risk contributions according to the widely used Euler allocation principle is in general difficult because of the rare-event simulations involved. In this chapter, we address the problem of estimating risk contributions when the total risk is measured by Value-at-Risk (VaR). Our proposed estimator for such *VaR contributions* is based on the Metropolis-Hasting (MH) algorithm, which is one of the most prevalent Markov chain Monte Carlo (MCMC) methods. We show consistency and asymptotic normality of our MH-based estimator of risk contributions. Numerical experiments based on simulation and real-world data demonstrate that in various risk models, even those having high-dimensional (≈ 500) inhomogeneous margins, our MH estimator has smaller bias and mean squared error compared to existing estimators.

4.1 Introduction

Capital allocation is an important part of a risk analysis, where a given amount of capital is decomposed into a sum of *risk contributions* of each unit's exposure; see, for example, Dev (2004). The *Euler principle* is one of the most well-known rules of risk allocation; see Section 1.3 for details. Calculating risk contributions under the Euler principle poses theoretical and numerical difficulties, especially when the portfolio-wide risk is measured by *Value-at-Risk (VaR)*. Although a simple formula of VaR contributions is derived by Tasche (2001), it can rarely be calculated analytically without a few exceptions; see in Tasche (2004). As is seen in Fan et al. (2012) and Yamai and Yoshiba (2002), the *crude Monte Carlo (MC)* method is the simplest method of computing risk contributions. However, the MC estimator suffers from unignorable bias caused by sample inefficiency and by inevitable numerical modification; see Yamai and Yoshiba (2002) and Section 4.2 of this thesis. To overcome such difficulties, several methods have been proposed in the literature. For instance, Hallerbach (2003) and Tasche and Tibiletti (2004) derived approximations by regarding

VaR contributions as best predictors of individual losses given the total loss. We call this estimator the *generalized regression (GR) estimator*. Glasserman (2005) developed *importance sampling (IS) estimators* with main focus being credit portfolios. The IS method generates samples from the so-called *instrumental distribution* and then adjusts them so that the estimator is consistent. This process of adjustment typically increases the variance of the estimator as a trade-off for enhancing effective sample size. Finding appropriate instrumental distributions to reduce the variance often relies on knowledge of the distribution of the rare events of interest. Finally, Tasche (2009) proposed the *Nadaraya-Watson (NW) estimator*, which is based on the kernel estimation method. Despite its ease of calculation, it still requires importance sampling to achieve an efficient estimation of risk contributions.

In this chapter, we propose a new method for estimating VaR contributions that utilizes the *Markov chain Monte Carlo (MCMC)*, especially the *Metropolis-Hastings (MH) algorithm*; see Section 1.4. Our MH method requires to be able to evaluate the joint loss density. This is often the case when losses are modelled separately by marginal distributions and a copula; see Yoshiba (2013) for various examples. For such loss models, the IS method is not straightforward to apply, since, in general, there is no guidance on the appropriate choice of the instrumental distribution. To the best of our knowledge, no numerically stable estimator of VaR contributions is known uniformly for all risk models. We study the consistency and asymptotic normality of our MH estimator, and provide practical guidelines for the efficient application of the MCMC method to the problem of computing VaR contributions. The proposed method is then carried out for various risk models based on simulations and real-world data. In numerical experiments, we compare the performance of the MH estimator with other existing estimators.

The main difference between our MH and the crude MC methods is that in the former, samples are generated directly from the joint loss distribution given a rare event of interest. In contrast, the MC method generates samples from the unconditional loss distribution, which inevitably wastes a large portion of samples; see Figure 4.1 for the illustration of this difference between MC and MCMC (MH). The right-hand side figure of the MH samples also describes an underlying idea for simulating the conditional distribution of interest, which is to update samples sequentially so that they lie in the rare event region. This feature of the MH method significantly improves sample efficiency without the need for any computational modification as is done in the MC method, which inevitably causes a bias. As a consequence, a small bias and variance can be expected for the MH estimator compared with existing estimators.

This chapter is organized as follows. In Section 4.2 we review existing estimators of VaR contributions. In Section 4.3, we propose the MH estimator which combines the MH method with the estimation of VaR contributions. Consistency and asymptotic normality for the proposed estimator are studied in Section 4.4. Next, in Section 4.5, numerical studies are conducted based on simulation and real-world data. We demonstrate that for various risk models with marginal- and dependence-inhomogeneity and/or high-dimensionality, the MH estimator has smaller bias and mean squared error (MSE) than those of existing estimators. For applying our method to other risk models not presented in this chapter, practical guidelines on the usage of the MH method are also provided. Concluding remarks and discussions are given in Section 4.6. Readers are referred to Section 1.3 and Section 1.4 for notations and preliminaries on the problem of capital allocation and MCMC methods. Since we focus on the problem of estimating VaR contributions $\mathrm{AC}^{\mathrm{VaR}_p} = (\mathrm{AC}_1^{\mathrm{VaR}_p}, \ldots, \mathrm{AC}_d^{\mathrm{VaR}_p})$, we drop the superscript VaR_p and write (1.5) as AC

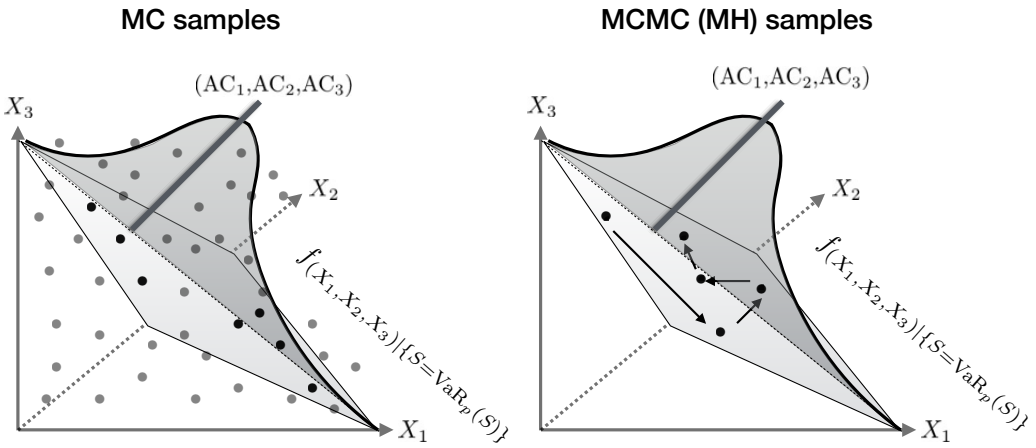

Figure 4.1: The difference between the Monte Carlo (MC, left) and Markov chain Monte Carlo (MCMC, right) methods for estimating the VaR contributions $(\mathrm{AC}_1,\mathrm{AC}_2,\mathrm{AC}_3)= \mathbb{E}[(X_1, X_2, X_3) \mid S = \mathrm{VaR}_p(S)]$, where $X_j, j = 1, 2, 3$ are loss random variables, $S = X_1 + X_2 + X_3$ is the total loss, and $\mathrm{VaR}_p(S)$ is the Value-at-Risk of S with confidence level $p \in (0, 1)$. In the MC method, samples are generated from the unconditional distribution of (X_1, X_2, X_3); only a few samples close enough to the plane $\{(x_1, x_2, x_3) \mid x_1 + x_2 + x_3 = \mathrm{VaR}_p(S)\}$ are used to estimate the allocated capital. On the other hand, the MH method generates samples directly from the conditional joint loss distribution given a rare event of interest, which is denoted by $f_{(X_1,X_2,X_3)|\{S=\mathrm{VaR}_p(S)\}}$.

$= (\mathrm{AC}_1, \ldots, \mathrm{AC}_d)$.

4.2 Existing estimators of VaR contributions

In this section, we review existing estimators of VaR contributions (1.5). Even when the joint density of the portfolio loss vector $f_{\boldsymbol{X}}$ is given explicitly, the analytical computation of $\mathrm{AC}_j = \mathbb{E}[X_j \mid S = \mathrm{VaR}_p(S)]$ is not straightforward since it often requires the joint distribution of (X_j, S), which is in general difficult to derive. A possible case when VaR/ES contributions can be explicitly derived is when $\boldsymbol{X}$ is modelled by a distribution that is *multivariate regularly varying (MRV)*. In this case, $\mathrm{VaR}_p(\boldsymbol{\lambda}^\top \boldsymbol{X})$ and $\mathrm{ES}_p(\boldsymbol{\lambda}^\top \boldsymbol{X})$ asymptotically have explicit formulas as $p \to 1$, and thus VaR and ES contributions can be explicitly derived; see Kley et al. (2016). In spite of their potential appeals, we avert from these formulas since the MRV assumption requires marginal tail-homogeneity. Moreover, these formulas depend on integrals with respect to spectral measures, which is beset with other difficulties.

A possible numerical method to calculate VaR contributions is the *crude MC method*, in which the

pseudo VaR contribution

$$\mathrm{AC}_\delta = \mathbb{E}[\boldsymbol{X} \mid S \in [\mathrm{VaR}_p(S) - \delta, \mathrm{VaR}_p(S) + \delta]\,], \tag{4.1}$$

is computed for a sufficiently small bandwidth $\delta > 0$. Since the probability $\mathbb{P}(S \in [\mathrm{VaR}_p(S)-\delta, \mathrm{VaR}_p(S)+\delta])$ is positive, the right hand side of (4.1) can be written as

$$\mathrm{AC}_\delta = \frac{\mathbb{E}[\boldsymbol{X}\mathbf{1}_{[S\in A_\delta]}]}{\mathbb{P}(S \in A_\delta)}, \quad \text{where} \quad A_\delta = [\mathrm{VaR}_p(S) - \delta, \mathrm{VaR}_p(S) + \delta].$$

This expression allows one to construct the estimator of the pseudo VaR contributions given by

$$\widehat{\mathrm{AC}}^{\mathrm{MC}}_{\delta,N} = \frac{\sum_{n=1}^N \boldsymbol{X}^{(n)}\mathbf{1}_{[S^{(n)}\in A_\delta]}}{\sum_{n=1}^N \mathbf{1}_{[S^{(n)}\in A_\delta]}} = \frac{1}{M_{\delta,N}}\sum_{n=1}^N \boldsymbol{X}^{(n)}\mathbf{1}_{[S^{(n)}\in A_\delta]}, \tag{4.2}$$

where $N > 0$ is the sample size, $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}$ are independent and identically distributed (i.i.d.) samples from $F_{\boldsymbol{X}}$, $S^{(n)} = X_1^{(n)} + \cdots + X_d^{(n)}$ are i.i.d. samples from F_S for $n = 1, \dots, N$, and $M_{\delta,N} = \sum_{n=1}^N \mathbf{1}_{[S^{(n)}\in A_\delta]}$ is the number of samples contained in A_δ. We call (4.2) the *MC estimator*. By setting δ and N sufficiently small and large, respectively, one can expect that the MC estimator approximates the true VaR contributions. Note that this method is available only when δ is positive, since $\mathbb{P}(S \in A_0) = \mathbb{P}(S = \mathrm{VaR}_p(S)) = 0$ by continuity of F_S.

As long as i.i.d. samples from $F_{\boldsymbol{X}}$ can be generated, one can estimate AC_δ by constructing the estimator (4.2). However, this estimator suffers from an inevitable bias caused by changing the condition $\{S = \mathrm{VaR}_p(S)\}$ to $\{S \in A_\delta\}$. The bias of the MC estimator can be decomposed by

$$\widehat{\mathrm{AC}}^{\mathrm{MC}}_{\delta,N} - \mathrm{AC} = b_\delta(N) + b(\delta),$$

where $b_\delta(N) = \widehat{\mathrm{AC}}^{\mathrm{MC}}_{\delta,N} - \mathrm{AC}_\delta$ and $b(\delta) = \mathrm{AC}_\delta - \mathrm{AC}$. The bandwidth δ should be taken as small as possible to reduce $b(\delta)$. However, when δ is quite small, it is difficult to ensure a large enough sample size $M_{\delta,N}$ to keep the first term $b_\delta(N)$ small since $\mathbb{E}[M_{\delta,N}] = N\mathbb{P}(S \in A_\delta)$, and $\mathbb{P}(S \in A_\delta)$ is typically much smaller than $1 - p$.

To overcome this problem, several estimators have been proposed in the literature. One is the *NW kernel estimator* proposed by Tasche (2009), which is defined by

$$\widehat{\mathrm{AC}}^{\mathrm{NW}}_{\phi,h,N} = \frac{\sum_{n=1}^N \boldsymbol{X}^{(n)}\phi\left(\frac{S^{(n)}-\mathrm{VaR}_p(S)}{\Delta}\right)}{\sum_{n=1}^N \phi\left(\frac{S^{(n)}-\mathrm{VaR}_p(S)}{\Delta}\right)}, \tag{4.3}$$

where ϕ is a kernel density and $\Delta > 0$ is the bandwidth. Since this estimator can be interpreted as a smoothing modification of the MC estimator (4.2) by the kernel ϕ, it shares the same bias trade-off explained above. Furthermore, the bias and asymptotic standard deviation of the NW estimator (see, for example, Hansen, 2009) cannot be computed easily because they require an evaluation of the total loss density $f_S(s)$ at $s = \mathrm{VaR}_p(S)$. Next, Hallerbach (2003) and Tasche and Tibiletti (2004) constructed estimators by assuming a regression model among losses of the form

$$\boldsymbol{X} = \boldsymbol{g}_{\boldsymbol{\beta}}(S) + \boldsymbol{\varepsilon},$$

where $\boldsymbol{g}_{\boldsymbol{\beta}}(s) : \mathbb{R} \to \mathbb{R}^d$ is a function parameterized by $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ is an error random vector such that $\mathbb{E}[\boldsymbol{\varepsilon} \mid \{S = \mathrm{VaR}_p(S)\}] = \mathbf{0}$. For an estimator $\hat{\boldsymbol{\beta}}_N$ of $\boldsymbol{\beta}$, we call

$$\widehat{\mathrm{AC}}^{\mathrm{GR}}_{\boldsymbol{g}_{\boldsymbol{\beta}},N} = \boldsymbol{g}_{\hat{\boldsymbol{\beta}}_N}(\mathrm{VaR}_p(S)) \tag{4.4}$$

the *GR estimator*. Although this estimator is intuitive and can easily be computed, it is in general difficult to construct an appropriate model $\boldsymbol{g}_{\boldsymbol{\beta}}$ and estimator $\hat{\boldsymbol{\beta}}_N$ of $\boldsymbol{\beta}$, unless samples from $F_{\boldsymbol{X}|\{S=\mathrm{VaR}_p(S)\}}$ are available.

A notable exception is the case wherein $\boldsymbol{X}$ follows an elliptical distribution. In this case, the following result holds:

$$\mathbb{E}[\boldsymbol{X} \mid \{S = \mathrm{VaR}_p(S)\}] = \mathbb{E}[\boldsymbol{X}] + \frac{\mathrm{Cov}(\boldsymbol{X}, S)}{\mathrm{Var}(S)}(\mathrm{VaR}_p(S) - \mathbb{E}[S]); \tag{4.5}$$

see, for example, Vanduffel and Dhaene (2006) and McNeil et al. (2015, Corollary 8.43). The true VaR contributions are then provided by setting $\boldsymbol{g}_{\boldsymbol{\beta}}(s) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 s$, where

$$\boldsymbol{\beta}_0 = \mathbb{E}[\boldsymbol{X}] - \frac{\mathrm{Cov}(\boldsymbol{X}, S)}{\mathrm{Var}(S)}\mathbb{E}[S] \quad \text{and} \quad \boldsymbol{\beta}_1 = \frac{\mathrm{Cov}(\boldsymbol{X}, S)}{\mathrm{Var}(S)}. \tag{4.6}$$

Since these coefficients are the minimizers of $\mathbb{E}[\boldsymbol{\varepsilon}^2] = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\beta}_0 - \boldsymbol{\beta}_1 S)^2]$, the OLS estimators of (β_0, β_1) are calculated based on the unconditional samples of $\boldsymbol{X}$ and S converges to the true parameters (4.6) as $N \to \infty$.

4.3 The proposed method

As seen in Section 4.2, the essential problem in estimating VaR contributions is that the conditional samples from $F_{\boldsymbol{X}|\{S=\mathrm{VaR}_p(S)\}}$ are unavailable. To solve this problem, we propose a new estimator of VaR contributions that utilizes MCMC method, especially the MH algorithm, to achieve an efficient estimation. Investigation on the consistency and asymptotic normality of our MH-based estimator is provided in the Section 4.4 for certain classes of risk models.

4.3.1 Assumptions and setup

We start by declaring assumptions under which our MH estimator is applicable.

Assumption 4.3.1 (Assumptions to apply MCMC methods)**.** For applying the MH estimator, we assume the following to hold:

(i) an explicit form of the joint loss density $f_{\boldsymbol{X}}$ is given, and thus one can compute the quantity $f_{\boldsymbol{X}}(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathbb{R}^d$;

(ii) a generator of i.i.d. samples from the loss distribution $F_{\boldsymbol{X}}$ is available; and

(iii) neither the explicit form of the total loss density f_S nor the way to compute the quantity $f_S(\mathrm{VaR}_p(S))$ is available.

Note that Assumption 4.3.1 Part (ii) enables us to generate samples from F_S by setting $S^{(n)} = X_1^{(n)} + \cdots + X_d^{(n)}$ where $(X_1^{(n)}, \ldots, X_d^{(n)})$ is the nth sample from $F_{\boldsymbol{X}}$. Assumption 4.3.1 Part (iii) implies that, while f_S can be derived from $f_{\boldsymbol{X}}$ in the form of an integral, the integral is not straightforward to calculate. Such a situation typically occurs when the joint loss density $f_{\boldsymbol{X}}$ is specified through a copula density c and marginal loss densities $f_1, \ldots, f_d$. The resulting joint loss density $f_{\boldsymbol{X}}$ is specified as in formula (1.2).

As is mentioned in Section 4.2, computing VaR contributions involves two steps; the first is to estimate $\mathrm{VaR}_p(S)$ by v, and the second is to estimate VaR contributions $\mathrm{AC} = \mathbb{E}[\boldsymbol{X} \mid \{S = v\}]$. The estimation of $\mathrm{VaR}_p(S)$ in the first step is often conducted with MC simulation. Based on i.i.d. samples $(S^{(1)}, \ldots, S^{(N)})$ from F_S, $\mathrm{VaR}_p(S)$ can be estimated, for example, by $\widehat{\mathrm{VaR}}_p(S) = S^{\lceil Np \rceil}$, where $\lceil Np \rceil$ is the smallest integer greater than or equal to Np, and $S^{\lceil Np \rceil}$ is the $\lceil Np \rceil$th largest sample among the N samples. Since $\widehat{\mathrm{VaR}}_p(S)$ is a deterministic quantity, one can regard it as a constant $v = \widehat{\mathrm{VaR}}_p(S)$.

In the second step, $\mathrm{AC} = \mathbb{E}[\boldsymbol{X} \mid \{S = v\}]$ is estimated. According to the crude MC method, VaR contributions are estimated by (4.2). As explained in Section 4.2, the problem of this two-step procedure is that the estimator of VaR contributions in the second step is typically biased. To address this issue, we develop an MCMC (MH)-based estimator that achieves consistency and high sample efficiency.

4.3.2 The MH estimator of VaR contributions

We propose to estimate VaR contributions by sequentially updating samples so that all samples lie in the set called the v-simplex:

$$\mathcal{S}_v = \{\boldsymbol{x} \in \mathbb{R}^d : x_1 + \cdots + x_d = v\}.$$

The updating rule is established so that the componentwise sum of each sample is preserved and the samples are taken from the distribution $F_{\boldsymbol{X}|\{S=v\}}$. We start to describe the MH-based estimator by reformulating the problem of computing VaR contributions. Since MH requires a density of the target distribution but the d-dimensional density is not well-defined on the degenerated space $\mathcal{S}_v$, we consider the first d' losses $\boldsymbol{X}' = (X_1, \ldots, X_{d'})$ where $d' = d - 1$. Throughout the thesis, the $'$-notation is used to denote quantities related to this non-degenerate distribution in $d-1$ dimensions and should not be confused with matrix transposition for which we will use the $^\top$-symbol. By the full allocation property (1.3), it holds that

$$\mathbb{E}[\boldsymbol{X} \mid \{S = v\}] = (\mathbb{E}[\boldsymbol{X}' \mid \{S = v\}],\ v - \mathbf{1}_{d'}^\top \mathbb{E}[\boldsymbol{X}' \mid \{S = v\}]),$$

where $S = X_1 + \cdots + X_d$. Therefore, computation of VaR contributions $\mathrm{AC} = \mathbb{E}[\boldsymbol{X} \mid \{S = v\}]$ can be reduced to estimating VaR contributions of the d'-subportfolio, denoted by $\mathrm{AC}' = \mathbb{E}[\boldsymbol{X}' \mid \{S = v\}]$. In our method, this quantity AC' is estimated by generating samples directly from $F_{\boldsymbol{X}'|\{S=v\}}$. The conditional joint density of $\boldsymbol{X}'$ given $\{S = v\}$ can be written as

$$f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{x}') = \frac{f_{(\boldsymbol{X}',S)}(\boldsymbol{x}', v)}{f_S(v)} = \frac{f_{\boldsymbol{X}}(\boldsymbol{x}', v - \mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(v)}, \quad \boldsymbol{x}' \in \mathbb{R}^{d'},$$

where the last equation follows from a linear transformation $(\boldsymbol{X}', S) \mapsto \boldsymbol{X}$. At this point, sampling directly from $f_{\boldsymbol{X}'|\{S=v\}}$ is difficult since the total loss density $f_S(v)$ is not easy to evaluate in general.

By taking $E = \mathbb{R}^{d'}$, $h(\boldsymbol{x}) = \boldsymbol{x}$, and $\pi(\boldsymbol{x}) = f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{x})$ in the notation presented in Section 1.4, our problem of estimating VaR contributions can be reduced to estimating $\boldsymbol{\pi}(\boldsymbol{h}) = \mathbb{E}[\boldsymbol{X}' \mid \{S = v\}]$ in (1.9) by MCMC. Even though it is challenging to compute $f_{\boldsymbol{X}'|\{S=v\}}$, we can compute the acceptance probability (1.11) given by

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left[\frac{f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})}{f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y})},\ 1\right] = \min\left[\frac{f_{\boldsymbol{X}}(\boldsymbol{y}, v - \mathbf{1}_{d'}^\top \boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x}, v - \mathbf{1}_{d'}^\top \boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y})},\ 1\right],$$

for any $\boldsymbol{x}, \boldsymbol{y}$ by Assumption 4.3.1 Part (i). Note that the term $f_S(v)$ disappears by considering the ratio $f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{y})/f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{x})$. Therefore, under an appropriate choice of the proposal density q, Algorithm 1 (MH algorithm) allows one to generate an N-path of the Markov chain whose stationary distribution is $\pi(\boldsymbol{x}) = f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{x})$. Based on this sample path, we can then construct the MH estimator $\hat{\boldsymbol{\pi}}_N(\boldsymbol{h})$ defined by (1.10).

The steps to compute the MH estimator of VaR contributions can thus be summarized as follows.

1. Estimate VaR as $v = \widehat{\mathrm{VaR}}_p(S)$ by MC samples.
2. Fix the sample size $N > 0$, proposal distribution q, and initial value $\boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)} \in \mathrm{supp}(\boldsymbol{X}'|\{S = v\})$.
3. Perform Algorithm 1 for the given N, q, and $\boldsymbol{x}^{(0)}$ to generate an N-path $(\boldsymbol{X}'^{(1)}, \ldots, \boldsymbol{X}'^{(N)})$.
4. Set
$$\widehat{\mathrm{AC}}_{q,N}^{\mathrm{MCMC}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{X}^{(n)} \quad \text{where} \quad \boldsymbol{X}^{(n)} = (\boldsymbol{X}'^{(n)}, v - \mathbf{1}_{d'}^\top \boldsymbol{X}'^{(n)}), \tag{4.7}$$
to estimate VaR contributions $\mathrm{AC} = \mathbb{E}[\boldsymbol{X}|\{S = v\}]$.

Note that this procedure can easily be extended to other choices of the function h, that is, one can estimate $\mathbb{E}[h(\boldsymbol{X}) \mid \{S = v\}]$ for general functions h by replacing (4.7) with $\frac{1}{N}\sum_{n=1}^{N} h(\boldsymbol{X}^{(n)})$. Moreover, under regularity conditions, consistency and asymptotic normality of the MH estimator (4.7) hold; see Section 4.4 for more details.

4.4 Consistency and asymptotic normality

In this section, we derive conditions on the copula and marginal distributions under which the corresponding MH estimator of VaR contributions satisfies consistency (1.12) and the CLT (1.13) for some choice of proposal distribution q. This study reveals which proposal distributions are appropriate for a given risk model.

We classify loss distributions $F_{\boldsymbol{X}}$ into two classes; one for which $\mathrm{supp}(\boldsymbol{X}) = \mathbb{R}_+^d = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \geq \boldsymbol{0}\}$ and another for which $\mathrm{supp}(\boldsymbol{X}) = \mathbb{R}^d$. The former corresponds to the case where we model *pure losses*, and the latter to the case of *profits and losses (P&L)*. Our result is mainly about the former case, and we provide some examples for the latter case. It should be emphasized that the former case of pure losses

includes a broad range of loss models. To demonstrate this, let $c_j = \operatorname{essinf}(X_j)$, and set $\tilde{X}_j = X_j - c_j$, $j = 1, \ldots, d$. If $c_j > -\infty$, then $\tilde{X}_j \geq 0$. For $\tilde{S} = \sum_{j=1}^{d} \tilde{X}_j$, the translation invariance of VaR_p implies that

$$\mathrm{VaR}_p(\tilde{S}) = \mathrm{VaR}_p(S) - \sum_{j=1}^{d} c_j.$$

Therefore, the allocated capital of $\tilde{X}_j$ is given by

$$\begin{aligned}\widetilde{\mathrm{AC}}_j &= \mathbb{E}[\tilde{X}_j \mid \{\tilde{S} = \mathrm{VaR}_p(\tilde{S})\}] = \mathbb{E}\left[X_j - c_j \;\middle|\; \left\{S - \sum_{j=1}^{d} c_j = \mathrm{VaR}_p(S) - \sum_{j=1}^{d} c_j\right\}\right] \\ &= \mathbb{E}[X_j \mid \{S = \mathrm{VaR}_p(S)\}] - c_j = \mathrm{AC}_j - c_j.\end{aligned}$$

Consequently, one can estimate $(\mathrm{AC}_1, \ldots, \mathrm{AC}_d)$ by first estimating $(\widetilde{\mathrm{AC}}_1, \ldots, \widetilde{\mathrm{AC}}_d)$ based on the joint distribution of $(\tilde{X}_1, \ldots, \tilde{X}_d)$ such that $\operatorname{supp}(\widetilde{\boldsymbol{X}}) = \mathbb{R}^d_+$ and then subtracting $(c_1, \ldots, c_d)$ from $\widetilde{\mathrm{AC}}$. Therefore, our result for the pure loss case includes the P&L case where the minima of the profits are bounded.

4.4.1 The case of pure losses

When $\operatorname{supp}(\boldsymbol{X}) = \mathbb{R}^d_+$, the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$ is supported on the *v-simplex*

$$\mathcal{S}_v = \{\boldsymbol{x} \in \mathbb{R}^{d'}_+ : x_1 + \cdots + x_{d'} \leq v\}.$$

Because of the compactness of $\mathcal{S}_v$, we can state simple conditions on the marginal loss densities and copula density which lead to consistency and asymptotic normality of the MH estimator.

Theorem 4.4.1. Suppose that the joint distribution $f_{\boldsymbol{X}}$ is supported on $\mathbb{R}^d_+$ and has marginal densities $f_1, \ldots, f_d$ and a copula density c. Then, the consistency and asymptotic normality holds for the MH estimator of VaR contributions if the following conditions $(C1) - (C3)$ hold:

(C1) $\epsilon = \inf_{\boldsymbol{x},\boldsymbol{y} \in \mathcal{S}_v} q(\boldsymbol{x}, \boldsymbol{y}) > 0$,

(C2) $f_j(x)$ is positive and bounded above for any $x \in [0, v]$ for $j = 1, 2, \ldots, d$, and

(C3) $c(\boldsymbol{u})$ is positive and bounded above for any $\boldsymbol{u} \in F_1([0, v]) \times \cdots \times F_d([0, v])$.

Proof. According to Theorem 23 in Roberts and Rosenthal (2004), the CLT holds if the Markov chain is *uniformly ergodic* whenever $\mathbb{E}[||\boldsymbol{X}'||^2 \mid \{S = v\}] < \infty$. Since $X_1, \ldots, X_d \geq 0$, the moment condition is satisfied by the inequality

$$\mathbb{E}[X_i X_j \mid \{S = v\}] \leq \mathbb{E}[(X_1 + \cdots + X_d)^2 \mid \{S = v\}] = v^2 < \infty$$

for any $i, j \in \{1, 2, \ldots, d\}$. Thus, it suffices to show that the Markov chain is uniformly ergodic. According to Theorem 1.3 in Mengersen and Tweedie (1996), the Markov chain is uniformly ergodic if the minorization

condition (Rosenthal, 1995) holds on the whole space $\mathcal{S}_v$; that is, there exists a positive integer n, a positive number $\delta > 0$, and a probability measure ν such that

$$K^n(\boldsymbol{x}, A) > \delta\nu(A), \tag{4.8}$$

for any $\boldsymbol{x} \in \mathcal{S}_v$ and $A \in \mathcal{B}_v = \mathcal{B}(\mathbb{R}^{d'}) \cap \mathcal{S}_v$. Our target distribution can be written as

$$\pi(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_S(v)} = \frac{c(F_1(x_1), \ldots, F_d(x_d))}{f_S(v)} f_1(x_1) \cdots f_d(x_d),$$

where $(x_1, \ldots, x_{d-1}) \in \mathcal{S}_v$ and $x_d = v - \mathbf{1}_d^\top \boldsymbol{x}$. Thus, by conditions (C2), (C3), and since $\mathcal{S}_v \subset [0, v]^{d'}$, we have that

$$l = \inf_{\boldsymbol{x} \in \mathcal{S}_v} \pi(\boldsymbol{x}) > 0, \qquad u = \sup_{\boldsymbol{x} \in \mathcal{S}_v} \pi(\boldsymbol{x}) < \infty. \tag{4.9}$$

Using (4.9) and condition (C1), the minorization condition can be checked as follows. For any $\boldsymbol{x} \in \mathcal{S}_v$, define

$$Q_{\boldsymbol{x}} = \left\{ \boldsymbol{y} \in \mathcal{S}_v : \frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})} \frac{q(\boldsymbol{y}, \boldsymbol{x})}{q(\boldsymbol{x}, \boldsymbol{y})} < 1 \right\}.$$

Then, for any $A \in \mathcal{B}_v$, we have that

$$\begin{aligned}
K(\boldsymbol{x}, A) &= \int_A \{q(\boldsymbol{x}, \boldsymbol{y})\alpha(\boldsymbol{x}, \boldsymbol{y}) + r(\boldsymbol{x})\delta_{\boldsymbol{x}}(\boldsymbol{y})\} \mathrm{d}\boldsymbol{y} \\
&\geq \int_{Q_{\boldsymbol{x}}} q(\boldsymbol{x}, \boldsymbol{y}) \min\left[1, \frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})} \frac{q(\boldsymbol{y}, \boldsymbol{x})}{q(\boldsymbol{x}, \boldsymbol{y})}\right] d\boldsymbol{y} \\
&\quad + \int_{A \setminus Q_{\boldsymbol{x}}} q(\boldsymbol{x}, \boldsymbol{y}) \min\left[1, \frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})} \frac{q(\boldsymbol{y}, \boldsymbol{x})}{q(\boldsymbol{x}, \boldsymbol{y})}\right] d\boldsymbol{y} \\
&= \int_{Q_{\boldsymbol{x}}} \frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})} q(\boldsymbol{y}, \boldsymbol{x}) d\boldsymbol{y} + \int_{A \setminus Q_{\boldsymbol{x}}} q(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} \\
&\geq \frac{\epsilon}{u} \int_{Q_{\boldsymbol{x}}} \pi(\boldsymbol{y}) d\boldsymbol{y} + \epsilon \int_{A \setminus Q_{\boldsymbol{x}}} \frac{\pi(\boldsymbol{y})}{u} d\boldsymbol{y} \\
&= \frac{\epsilon}{u} \pi(A).
\end{aligned}$$

Therefore, the minorization condition holds for $n = 1$, $\delta = \frac{\epsilon}{u} > 0$, and $\nu = \pi$. Consequently, the Markov chain is uniformly ergodic, and thus the CLT holds. Since the minorization condition (4.8) holds, consistency of $\hat{\boldsymbol{\pi}}_N(\boldsymbol{h})$ follows by Theorem 1 in Nummelin (2002). □

An example of a risk model and proposal distribution is given by the following example.

Example 4.4.2. For $j = 1, \ldots, d$, let X_j follow the Pareto distribution with probability density function given by

$$f_j(x_j; \kappa_j, \gamma_j) = \frac{\kappa_j \gamma_j^{\kappa_j}}{(x_j + \gamma_j)^{\kappa_j + 1}}, \quad \kappa_j, \gamma_j > 0 \quad \text{for} \quad x_j > 0. \tag{4.10}$$

Suppose $\boldsymbol{X} = (X_1, \ldots, X_d)$ has a survival Clayton copula with the density given by

$$c(\boldsymbol{u}; \theta) = \frac{\theta^d \Gamma(\frac{1}{\theta} + d)}{\Gamma(\frac{1}{\theta})} \left\{ \prod_{j=1}^d (1 - u_j)^{-\theta - 1} \right\} \left\{ \sum_{j=1}^d (1 - u_j)^{-\theta} - d + 1 \right\}^{-\frac{1}{\theta} - d}, \quad 0 < \theta < \infty. \tag{4.11}$$

Some simple calculations show that the marginal distributions (4.10) satisfies $(C2)$ and the copula (4.11) satisfies $(C3)$ under the mild sufficient condition that $0 < \theta < \log(1-p)/\log(1-\frac{1}{d})$. Therefore, with any choice of proposal distribution q satisfying $(C1)$, the corresponding MH estimator satisfies consistency and asymptotic normality. A possible choice of q is the random walk proposal $q(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y} - \boldsymbol{x})$ with f being the density of a multivariate normal distribution with mean zero. Since $\boldsymbol{y} - \boldsymbol{x} \in [-v, v]^{d'}$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}_v$, $f(\boldsymbol{y} - \boldsymbol{x})$ is always positive.

It is worth noting that the condition $(C3)$ is irrelevant to the copula on the upper tail part $[F_1(v), 1] \times \cdots \times [F_d(v), 1]$. Therefore, $(C3)$ holds even if a copula density explodes at the upper corner, which is often the case with copulas having upper tail dependence. In fact, a more general result holds for survival Archimedean copulas. A d-dimensional *Archimedean copula* with an Archimedean generator ψ is given by

$$C_\psi(\boldsymbol{u}) = \psi\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right), \tag{4.12}$$

where ψ is a continuous and nonincreasing function $\psi : [0, \infty] \to [0, 1]$ satisfying $\psi(0) = 1$, and $\lim_{t\to\infty} \psi(t) = 0$, and is decreasing on $[0, \inf\{t : \psi(t) = 0\}]$. The inverse $\psi^{-1}(u)$ is well-defined on $u \in (0, 1]$ and $\psi^{-1}(0)$ is defined by $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$. Let $\psi^{(j)}$ be the jth derivative of ψ. An Archimedean generator ψ defines a proper d-copula via (4.12) for any $d \geq 1$ if and only if ψ is *completely monotone*, that is, $(-1)^j \psi^{(j)} \geq 0$ on $(0, \infty)$ for all $j = 0, 1, \ldots$; see McNeil et al. (2009). We denote the class of completely monotone generators as Ψ_∞. According to Bernstein's Theorem (see, for example, Feller, 2008), $\psi \in \Psi_\infty$ admits the Laplace–Stieltjes representation $\psi(t) = \mathbb{E}[\mathrm{e}^{-tV}]$ for some positive random variable $V > 0$.

Theorem 4.4.3 (Sufficient condition of $(C3)$ for survival Archimedean copulas)**.** Let $\psi \in \Psi_\infty$ be a completely monotone Archimedean generator. If $\mathbb{E}[V^d] < \infty$ where V is such that $\psi(t) = \mathbb{E}[\mathrm{e}^{-tV}]$, then the survival Archimedean copula $\bar{C}_\psi$ has a density satisfying the condition $(C3)$ in Theorem 4.4.1; moreover, $\bar{C}_\psi$ has a zero lower tail dependence coefficient.

Proof. Denote $\bar{u}_j = F_j(v) < 1$ and $\underline{u}_j = 1 - \bar{u}_j > 0$. The density of the survival Archimedean copula is given by

$$\begin{aligned}
\bar{c}_\psi(\boldsymbol{u}) = c_\psi(\boldsymbol{1} - \boldsymbol{u}) &= \psi^{(d)}\left(\sum_{j=1}^{d} \psi^{-1}(1 - u_j)\right) \prod_{j=1}^{d} \frac{1}{\psi^{(1)}(\psi^{-1}(1 - u_j))} \\
&= (-1)^d \psi^{(d)}(t) \prod_{j=1}^{d} \frac{1}{(-1)\psi^{(1)}(t_j)},
\end{aligned} \tag{4.13}$$

where $t_j = \psi^{-1}(1 - u_j)$ and $t = \sum_{j=1}^{d} t_j$. When $u_j \in [0, F_j(v)]$, we have $0 < \underline{u}_j \leq 1 - u_j \leq 1$ and thus $t_j = \psi^{-1}(1 - u_j) \in [0, \bar{t}_j]$ where $\bar{t}_j = \psi^{-1}(\underline{u}_j) < \infty$. Thus, $0 \leq t = \sum_{j=1}^{d} t_j < \infty$.

Since $\psi \in \Psi_\infty$, it is of the form $\psi(t) = \mathbb{E}[\mathrm{e}^{-tV}]$ for some positive random variable $V > 0$. Therefore, on $0 \leq t < \infty$, we have $0 < (-1)^j \psi^{(j)}(t) < \infty$ for $j = 1$ and $j = d$ since $(-1)^j \psi^{(j)}(t) = \mathbb{E}[V^j \mathrm{e}^{-tV}] > 0$ and $\mathbb{E}[V^j \mathrm{e}^{-tV}] \leq \mathbb{E}[V^j] < \infty$ for $j = 1$ and $j = d$ by assumption. Consequently, the density (4.13) is bounded from below and above.

When $\mathbb{E}[V^d] < \infty$, the corresponding Archimedean copula has an upper tail dependence coefficient

$$\lambda_u(C_\psi) = 2 - 2\lim_{t\to 0}\frac{1-\psi(2t)}{1-\psi(t)} = 2 - 2\lim_{t\to 0}\frac{\psi^{(1)}(2t)}{\psi^{(1)}(t)}$$

where the last equality comes from l'Hôpital's rule. We also have

$$\lim_{t\to 0}\frac{\psi^{(1)}(2t)}{\psi^{(1)}(t)} = \lim_{t\to 0}\frac{(-1)\psi^{(1)}(2t)}{(-1)\psi^{(1)}(t)} = \lim_{t\to 0}\frac{\mathbb{E}[V\mathrm{e}^{-2tV}]}{\mathbb{E}[V\mathrm{e}^{-tV}]} = 1$$

since $\mathbb{E}[V\mathrm{e}^{-2tV}]$ and $\mathbb{E}[V\mathrm{e}^{-tV}]$ converge to $\mathbb{E}[V] < \infty$ as $t \to 0$. Thus, for the survival Archimedean copula, $\lambda_l(\bar{C}_\psi) = \lambda_u(C_\psi) = 0$. □

According to Theorem 4.4.3, the survival Clayton copula satisfies $(C3)$ while the survival Gumbel copula does not because it is known to have a positive lower tail dependence coefficient.

Remark 4.4.4 (Consistency and CLT for copulas with lower tail dependence)**.** Condition $(C3)$ does not hold for elliptical copulas with lower tail dependence, such as a Student t copula with density

$$c^t(\boldsymbol{u};\nu,P) = \frac{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu}{2})}{|P|^{\frac{1}{2}}\Gamma(\frac{\nu+1}{2})^d}\frac{\left(1+\frac{\boldsymbol{x}^\top P^{-1}\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+d}{2}}}{\Pi_{j=1}^d(1+\frac{x_j^2}{\nu})^{-\frac{\nu+1}{2}}}, \quad \nu > 0, \tag{4.14}$$

where $\boldsymbol{x} = (t_\nu^{-1}(u_1),\dots,t_\nu^{-1}(u_d))$ with t_ν being a cumulative distribution function of a univariate Student t distribution with degrees of freedom ν. By carefully checking the proof of Theorem 4.4.1, the consistency and asymptotic normality of $\hat{\boldsymbol{\pi}}_N(\boldsymbol{h})$ still hold under a weaker condition than $(C2)$ and $(C3)$;

$$\frac{q(\boldsymbol{y},\boldsymbol{x})}{\pi(\boldsymbol{x})} = \frac{q(\boldsymbol{y},\boldsymbol{x})f_S(v)}{c(F_1(x_1),\dots,F_d(x_d))f_1(x_1)\cdots f_d(x_d)} \geq L, \quad \boldsymbol{x},\boldsymbol{y}\in\tilde{\mathcal{S}}_v, \tag{4.15}$$

for some positive constant $L > 0$, where $\tilde{\mathcal{S}}_v = \{\boldsymbol{x}\in\mathbb{R}^d : \mathbf{1}_d^\top\boldsymbol{x} = 1\}$. While it is not straightforward to determine, one sufficient condition of (4.15) under $(C1)$ is that π is bounded above on $\tilde{\mathcal{S}}_v$. Another condition is that the proposal density q explodes faster than π. An example of such q can be an independent proposal distribution $q(\boldsymbol{x},\boldsymbol{y}) = f(\boldsymbol{y})$ with f being the density of the Dirichlet distribution $\mathrm{Dir}(\alpha_1,\dots,\alpha_d)$ for $\alpha_1,\dots,\alpha_d < 1$, which explodes to ∞ as $\boldsymbol{x}$ approaches to an axis. Therefore, by choosing such proposal distributions, consistency and CLT can still hold even if a copula density explodes at the lower corner $\boldsymbol{u} = \mathbf{0}$.

4.4.2 The case of profits and losses

In contrast to the case of pure losses, showing asymptotic normality of the MH estimator is challenging for the P&L case. Since the conditional density $f_{\boldsymbol{X}'|\{S=v\}}$ is supported on the unbounded space $\mathbb{R}^{d'}$, a careful study of its tail behavior is necessary. When the original loss random vector $\boldsymbol{X}$ follows an elliptical distribution, the results of Kamatani (2017) can be applicable to justify the CLT of our MH estimator with the MpCN proposal distribution; see Section 1.4. An example of a justification of the CLT for the case where $\boldsymbol{X}$ follows the multivariate Student t distribution is provided below.

Example 4.4.5 (Justification of CLT for multivariate Student t distributions)**.** We demonstrate that the MpCN proposal distribution (1.17) achieves the CLT of VaR contributions when the underlying loss model is a multivariate Student t distribution $t_\nu(\boldsymbol{\mu}, \Sigma)$ with density

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \nu, \Sigma) = \frac{\Gamma(\frac{\nu+d}{2})}{|\pi\nu\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})} \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}}. \tag{4.16}$$

Let $\boldsymbol{X} \sim t_\nu(\boldsymbol{\mu}, \Sigma)$ for $\nu > 2$, $\boldsymbol{\mu} \in \mathbb{R}^d$, and $\Sigma \in \mathcal{M}_+^{d\times d}$ where $\mathcal{M}_+^{d\times d}$ is the set of $d \times d$ positive definite matrices. Throughout the discussion, we set $\boldsymbol{\mu} = \mathbf{0}$ for simplicity. Write

$$\Sigma^{-1} = \begin{pmatrix} A_1 & \boldsymbol{a}_2 \\ \boldsymbol{a}_2^\top & a_3 \end{pmatrix} =: A$$

for $A_1 \in \mathcal{M}^{d'\times d'}(\mathbb{R})$, $\boldsymbol{a}_2 \in \mathbb{R}^{d'}$, and $a_3 \in \mathbb{R}$. Then, it holds that

$$\begin{pmatrix} \boldsymbol{x} \\ v - \mathbf{1}_{d'}^\top \boldsymbol{x} \end{pmatrix}^\top \begin{pmatrix} A_1 & \boldsymbol{a}_2 \\ \boldsymbol{a}_2^\top & a_3 \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ v - \mathbf{1}_{d'}^\top \boldsymbol{x} \end{pmatrix} = (\boldsymbol{x}-\boldsymbol{w})^\top V(\boldsymbol{x}-\boldsymbol{w}) + \eta,$$

where $V = A_1 - \boldsymbol{a}_2\mathbf{1}_{d'}^\top - \mathbf{1}_{d'}\boldsymbol{a}_2^\top + \mathbf{1}_{d'}\mathbf{1}_{d'}^\top \in \mathcal{M}_+^{d'\times d'}$, $\boldsymbol{w} = V^{-1}(vA_3\mathbf{1}_{d'} - v\boldsymbol{a}_2) \in \mathbb{R}^{d'}$, and $\eta = v^2 a_3 - \boldsymbol{w}^\top V \boldsymbol{w} \in \mathbb{R}$. Using this identity, we have that

$$\begin{aligned} f_{\boldsymbol{X}'|\{S=v\}}(\boldsymbol{x}) &\propto f_{\boldsymbol{X}}(\boldsymbol{x}, v - \mathbf{1}_{d'}^\top \boldsymbol{x}) \\ &\propto \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{w})^\top W^{-1}(\boldsymbol{x}-\boldsymbol{w}) + \eta}{\nu}\right)^{-\frac{\nu+d}{2}} \\ &\propto \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{w})^\top W^{-1}(\boldsymbol{x}-\boldsymbol{w})}{\nu+\eta}\right)^{-\frac{\nu+d}{2}}, \end{aligned} \tag{4.17}$$

where $W = V^{-1}$. Provided $\nu + \eta > 0$, $\boldsymbol{X}'|\{S = v\}$ follows a d'-dimensional elliptical distribution with location parameter $\boldsymbol{w}$, scale parameter W, and the density generator $g : \mathbb{R}_+ \to \mathbb{R}_+$ given by

$$g(x) = \left(1 + \frac{x}{\nu+\eta}\right)^{-\frac{\nu+d}{2}}.$$

This type of distribution is called a *Pearson type VII distribution* (Schmidt, 2002).

Consider the MH estimator where the target distribution π is $f_{\boldsymbol{X}'|\{S=v\}}$ and the proposal distribution q is MpCN (1.17). According to Theorem 25 in Roberts and Rosenthal (2004), the CLT holds if the Markov chain is geometrically ergodic and $\mathbb{E}[||\boldsymbol{X}'||^2 \mid \{S = v\}] < \infty$. According to Proposition 5 in Kamatani (2017), the Markov chain with the MpCN proposal distribution is geometrically ergodic if $\mathbb{E}[||\boldsymbol{X}'||^\delta \mid \{S = v\}] < \infty$ for some $\delta > 0$, $\pi(\boldsymbol{x})$ is strictly positive and continuous, and it is symmetrically regularly varying, that is,

$$\lim_{r\to\infty} \frac{\pi(r\boldsymbol{x})}{\pi(r\mathbf{1}_{d'})} = \lambda(\boldsymbol{x}), \tag{4.18}$$

for some function $\lambda : \mathbb{R}^{d'} \to (0, \infty)$ such that $\lambda(\boldsymbol{x}) = 1$ for any $\boldsymbol{x} \in S_W^{d'-1}$, where $S_W^{d'-1} = \{\boldsymbol{x} \in \mathbb{R}^{d'} : ||W^{-\frac{1}{2}}\boldsymbol{x}|| = ||W^{-\frac{1}{2}}\mathbf{1}_{d'}||\}$. We will now show that the moment condition holds and the tail condition (4.18) is also satisfied for $\pi = f_{\boldsymbol{X}'|\{S=v\}}$.

Write $R = ||\boldsymbol{X}'||$. It can be shown that g is regularly varying (see, for example, Resnick, 2013) at ∞ with index $\alpha = -\frac{\nu+d}{2}$; that is,

$$\lim_{r\to\infty} \frac{g(rx)}{g(r)} = x^{-\frac{\nu+d}{2}}, \quad x > 0. \tag{4.19}$$

According to Proposition 3.7 in Schmidt (2002), $f_{R|\{S=v\}}$ is regularly varying with index $-(\nu+1)$. Then, according to Karamata's Theorem (we refer to Resnick, 2013), $F_{R|\{S=v\}}$ is regularly varying with index $-\nu$. Therefore, $\mathbb{E}[R^\delta \mid \{S=v\}] < \infty$ holds for any $\delta < \nu$; see Mikosch (1999). Thus, all the moment conditions above are satisfied as long as $\nu > 2$. In the elliptical case, the tail condition (4.18) is a direct consequence of (4.19). Since $(\boldsymbol{x}-\boldsymbol{w})^\top W^{-1}(\boldsymbol{x}-\boldsymbol{w}) > 0$ for all $\boldsymbol{x} \in \mathbb{R}^{d'}$, it holds that

$$\lim_{r\to\infty} \frac{f_{\boldsymbol{X}'|\{S=v\}}(r\boldsymbol{x})}{f_{\boldsymbol{X}'|\{S=v\}}(r\mathbf{1}_{d'})} = \left(\frac{||W^{-\frac{1}{2}}\boldsymbol{x}||}{||W^{-\frac{1}{2}}\mathbf{1}_{d'}||}\right)^{-(\nu+d)}, \quad \boldsymbol{x} \in \mathbb{R}^{d'}.$$

Thus, by taking

$$\lambda(\boldsymbol{x}) = \left(\frac{||W^{-\frac{1}{2}}\boldsymbol{x}||}{||W^{-\frac{1}{2}}\mathbf{1}_{d'}||}\right)^{-(\nu+d)}$$

in (4.18), $\pi = f_{\boldsymbol{X}'|\{S=v\}}$ is shown to be symmetrically regularly varying. Putting them together, we conclude that the MH estimator with the MpCN proposal distribution satisfies the CLT when the underlying loss vector follows a multivariate Student t distribution with $\nu > 2$ and $\eta > -\nu$. Note that in the numerical experiments in Section 4.5, we set $d = 3$ and $\nu = 4$. Since $\eta + \nu > 0$, the CLT holds true.

4.5 Numerical experiments

In this section, we apply the MH estimator proposed in Section 4.3 to various risk models, and compare its performance with other estimators of VaR contributions. Our simulations and empirical studies based on real-world data show that the MH estimator has smaller bias and lower MSE compared with the other estimators in many scenarios, including high-dimensional ($d \approx 500$) cases. Based on these numerical experiments, we provide practical guidelines on how to choose an appropriate proposal distribution of the MH estimator given a risk model.

4.5.1 Simulation study

Description of the numerical comparison

We consider four risk models that are specified separately by marginal densities and copula density. We adopt heavy-tailed marginal distributions and copulas with upper-tail dependence as they are often applied in risk management. In all risk models, we set the size of the portfolio to $d = 3$. The models are:

(1) The loss random variables X_1, X_2 and X_3 follow homogeneous Pareto distributions (4.10) for $\kappa = 4$ and $\gamma = 3$. The loss random vector (X_1, X_2, X_3) has a d-dimensional survival Clayton copula (4.11) with $\theta = 0.5$.

(2) The losses X_1, X_2 and X_3 have the same marginal distributions as in case (1). Their copula is a Student t copula (4.14) with $\nu = 4$ degrees of freedom, and the dispersion matrix P given by

$$P = \begin{pmatrix} 1 & -0.5 & 0.3 \\ -0.5 & 1 & 0.5 \\ 0.3 & 0.5 & 1 \end{pmatrix}. \tag{4.20}$$

(3) The losses X_1, X_2, and X_3 follow Student t distributions with $\nu = 4$ degrees of freedom, location parameter $\mu = 0$, and scale parameter $\sigma = 1$. A copula of (X_1, X_2, X_3) is a survival Clayton copula (4.11) with $\theta = 0.5$.

(4) The loss random vector (X_1, X_2, X_3) follows a multivariate Student t distribution (4.16) with $\nu = 4$, $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = P$ where P is defined as in (4.20).

Models (1) and (2) consider pure losses, and Models (3) and (4) consider P&L. In all models, marginal distributions have variances of 2 and heavy tails with tail indices 5. Models (1) and (3) possess upper tail dependence with tail coefficients $\lambda^U = 0.025$; see Joe (2014) for formulas for the tail coefficients. Models (2) and (4) have upper, lower, and upper-lower tail dependence with tail coefficients $\lambda^U_{1,2} = \lambda^L_{1,2} = 0.012$, $\lambda^U_{1,3} = \lambda^L_{1,3} = 0.162$, $\lambda^U_{2,3} = \lambda^L_{2,3} = 0.253$, $\lambda^{UL}_{1,2} = \lambda^{LU}_{1,2} = 0.253$, $\lambda^{UL}_{1,3} = \lambda^{LU}_{1,3} = 0.029$, and $\lambda^{UL}_{2,3} = \lambda^{LU}_{2,3} = 0.012$. As inferred by the dispersion matrix (4.20), the first and second losses are negatively dependent, while other pairs of losses are positively dependent.

For each risk model, we compute several estimators of VaR contributions $\mathrm{AC} = \mathbb{E}[\boldsymbol{X} \mid \{S = \mathrm{VaR}_p(S)\}]$ for a confidence level $p = 0.999$ with $\mathrm{VaR}_p(S)$ replaced by its Monte Carlo estimate $v = S^{\lfloor Np \rfloor}$. The estimators we compare are the MC estimator (4.2), NW estimator (4.3) (Tasche, 2009), GR estimator (4.4) (Hallerbach, 2003; Tasche and Tibiletti, 2004), and our MH estimator.

For all estimators, we fix the sample size as $N = 10^6$. Other parameters of the estimators are determined as follows. First, for the MC estimator, we set $\delta > 0$ such that the MC sample size $M_{\delta,N}$ is around 10^3. For a fixed δ, asymptotic normality holds; see, for example, Glasserman (2013). We report the estimate of $\widehat{\mathrm{AC}}^{\mathrm{MC}}_{\delta,N}$ and its approximated standard error for $j = 1, 2, 3$. Second, for the NW estimator, we choose the kernel density ϕ to be the standard normal density. We use the bandwidth $\Delta = 1.06\hat{\sigma}_S N^{-1/5}$ according to Silverman's rule of thumb (Pagan and Ullah, 1999). Although asymptotic normality holds for the NW estimator, its asymptotic variance can hardly be computed because it requires the evaluation of $f_S(v)$. Therefore, we report only the estimate of $\widehat{\mathrm{AC}}^{\mathrm{NW}}_{\phi,h,N}$. Third, for the GR estimator, we choose $g_{\boldsymbol{\beta}}(s) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 s$, and its coefficients are estimated by

$$\hat{\boldsymbol{\beta}}_{N,1} = \frac{\sum_{n=1}^N (\boldsymbol{X}^{(n)} - \bar{\boldsymbol{X}}_N)(S^{(n)} - \bar{S}_N)}{\sum_{n=1}^N (S^{(n)} - \bar{S}_N)^2}, \quad \hat{\boldsymbol{\beta}}_{N,0} = \bar{\boldsymbol{X}}_N - \hat{\boldsymbol{\beta}}_{N,1}\bar{S}_N,$$

where $\bar{\boldsymbol{X}}_N = \frac{1}{N}\sum_{n=1}^N \boldsymbol{X}^{(n)}$ and $\bar{S}_N = \frac{1}{N}\sum_{n=1}^N S^{(n)}$. Under regularity conditions, asymptotic normality holds and thus we report the estimate of $\widehat{\mathrm{AC}}^{\mathrm{GR}}_{g_{\boldsymbol{\beta}},N}$ and its approximated standard error. Finally, for the MH estimator, we choose different proposal distributions depending on risk models (1)–(4). For each risk model, we choose (1) a random walk proposal $q(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y} - \boldsymbol{x})$ with $f \sim \mathcal{N}_d(\mathbf{0}, \hat{\Sigma}_v)$, where $\hat{\Sigma}_v = S^2_{MC}$; (2) an

independent proposal $q(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y})$, where f is the density of the Dirichlet distribution with parameters $(0.323, 0.448, 0.892)$, which are estimated by the maximum likelihood method from pseudo-samples generated by MC; and (3) and (4) the MpCN proposal with $\rho = 0.8$ (a default choice in Kamatani et al., 2018), $\boldsymbol{\mu} = (\widehat{\mathrm{AC}}^{\mathrm{GR}}_{g_{\boldsymbol{\beta}},N}{}', v - \mathbf{1}_{d'}^{\top}\widehat{\mathrm{AC}}^{\mathrm{GR}}_{g_{\boldsymbol{\beta}},N}{}')$, and $\Sigma = S^2_{MC}$ where S^2_{MC} is the estimated variance based on the generated MC samples. In the MH Algorithm, we set the initial values as $\boldsymbol{x}^{(0)} = (v/3, v/3, v/3)$. We estimate the asymptotic variances of the MH estimators by the batch means estimators $\hat{\Sigma}_N$. Following Jones et al. (2006), we choose $L_N = \lfloor N^{\frac{1}{2}} \rfloor = 10^3$ and $B_N = \lfloor N/L_N \rfloor = \lfloor N^{\frac{1}{2}} \rfloor = 10^3$. We report the estimate of $\widehat{\mathrm{AC}}^{\mathrm{MCMC}}_{q,N}$ and its approximated standard error $\hat{\Sigma}^{(j,j)}_N/\sqrt{N}$ for $j = 1, 2, 3$.

Results and discussion

Due to the simplicity of the MC, NW and GR estimators, they were calculated quickly for all risk models. On the other hand, MH estimators took (1) 2.324, (2) 1.425, (3) 3.828, and (4) 3.698 minutes to generate an N-path and to compute the estimators. As mentioned in Section 1.4, the validity of the proposal selection can be inspected by autocorrelation plots and acceptance rates (ACR). Figure 4.2 (v)–(viii) shows the autocorrelation plots of the Markov chains generated by the MH algorithm. The acceptance rate of the MH algorithm in each risk model was (1) 0.566, (2) 0.222, (3) 0.604, and (4) 0.767. In Figure 4.2 (v)–(viii), we can observe that the autocorrelation plots steadily decline below 0.1 by lag h around 100 for all risk models. Together with the observations that the ACRs are moderate, we can state that the choices of the proposal distributions above are appropriate for all risk models.

Before showing the results of the estimation, let us check the shapes of the conditional distributions $F_{\boldsymbol{X}'|\{S=v\}}$ by plotting the N-path generated by the MH algorithm. Figure 4.2 (i)–(iv) shows the contour plots of the generated Markov chains. According to these plots, the features of the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$ in each risk model can be summarized as follows:

(1) *Pareto + survival Clayton:* The contour plot in Figure 4.2 (i) shows that $F_{\boldsymbol{X}'|\{S=v\}}$ has a unique mode. The density steadily decays as it moves away from the mode. In addition, the contour plot seems symmetric with respect to the diagonal line $y = x$.

(2) *Pareto + t copula:* Unlike case (1), $F_{\boldsymbol{X}'|\{S=v\}}$ seems to possess two distinct modes close to the axes. High probabilities are concentrated around the edges of the simplex. Moreover, the contour plot in Figure 4.2 (ii) is asymmetric at the diagonal line $y = x$.

(3) *Student t + survival Clayton:* Although the conditional loss random vector $\boldsymbol{X}' \mid \{S = v\}$ can take negative values, it is supported mostly on the bounded simplex as in case (1). The contour plot in Figure 4.2 (iii) seems unimodal and symmetric around the diagonal. The tails of $F_{\boldsymbol{X}'|\{S=v\}}$ are obviously light.

(4) *Student t + t copula:* In this case, the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$ can be shown to be a *Pearson type VII distribution* (4.17). From the contour plot in Figure 4.2 (iv), we can observe elliptical symmetry and tail-heaviness. Unlike case (3), the loss vector $\boldsymbol{X}' \mid \{S = v\}$ can take large negative values beyond the bounded simplex.

The estimation results are summarized in Table 4.1. For the Models (1)–(4), we report estimates of VaR contributions, their approximated standard errors, biases, and root MSEs (RMSEs) of the estimators MC, NW, GR and MH.

In the first risk model, true VaR contributions are obtained by equally allocating the total VaR since the marginal distributions are homogeneous and the copula is exchangeable. We observed that the MC and NW estimators have relatively large biases in comparison to those of the other estimators. Compared with the MH estimator, the GR estimator still suffers from some inevitable bias although its standard error is quite small. The MC estimator has a relatively large standard error due to sample inefficiency. Overall, the MH estimator outperforms all other estimators in terms of the RMSE.

The second risk model does not allow us to analytically calculate the true VaR contributions. Therefore, the true VaR contributions are computed by Monte Carlo integration, which still works with enough accuracy for the three dimensions considered here. We can observe that existing estimators suffer from biases possibly caused by asymmetry and multi-modality of the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$. In particular, the GR estimator has relatively large bias and RMSE in contrast to the good performance in the first risk model. On the other hand, the MH estimator maintains lower bias and RMSE compared to the other estimators.

In the third risk model, the true VaR contributions are given by equal allocations for the same reason as in case (1). Because of the symmetry and unimodality of the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$, all estimators retain small biases and RMSEs. Together with the results in cases (1) and (2), one can state that the GR estimator performs well as long as $F_{\boldsymbol{X}'|\{S=v\}}$ is symmetric and unimodal. Additionally, the MH estimator reduces bias and RMSE compared with those of MC and NW estimators.

The final risk model provides the true VaR contributions via the formula presented in (4.5). In such an elliptical case, the GR estimator provides quite an accurate estimate. Although the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$ is heavy-tailed as seen in Figure 4.2 (iv), the MH estimator retains high performance in comparison to the MC and NW estimators. The bias of the MH estimator is significantly improved compared with the MC and NW estimators. Moreover, the standard error and RMSE of the MH estimator are lower than those of the MC estimator.

Throughout the numerical study, the MH estimator provided a small bias and RMSE regardless of the shape of the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$. In the case when $F_{\boldsymbol{X}'|\{S=v\}}$ is unimodal and symmetric, the GR estimator also performed well. On the other hand, at least in our numerical experiment, the MC and NW estimators had relatively larger biases and RMSEs compared with the MH and GR estimators.

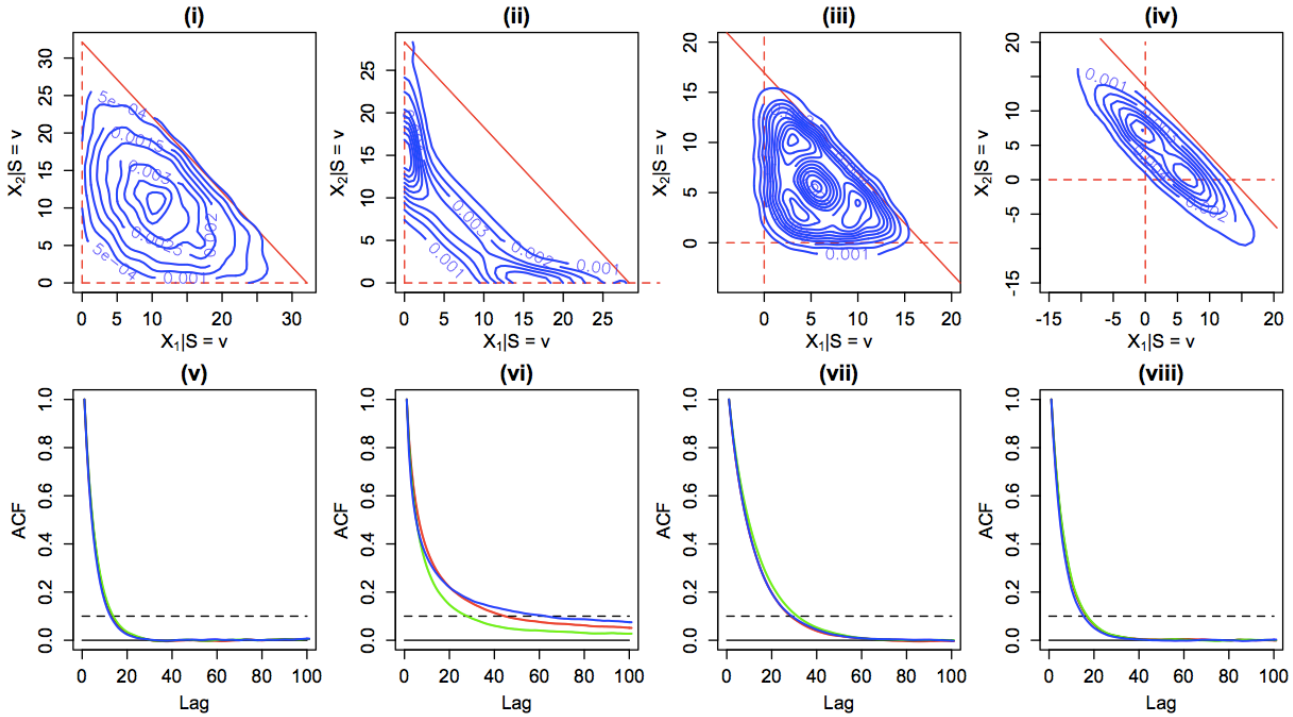

Figure 4.2: Contour plots (i)–(iv) and autocorrelation plots (v)–(viii) of Markov chains generated by the MH algorithm for four different risk models: (i) and (v) Pareto + survival Clayton; (ii) and (vi) Pareto + t copula; (iii) and (vii) Student t + survival Clayton; and (vi) and (viii) Student t + t copula. The red lines represent the edges of the v-simplex, where v is the estimate of $\mathrm{VaR}_p(S)$. The dotted black lines in Plots (vi)–(viii) represent $y = 0.1$. When drawing the contour plots, we used every 100th subsample of the original Markov chains to reduce the dependence among the samples.

Table 4.1: Estimates (biases) and standard errors (root mean squared errors; RMSEs) of the four different estimators of VaR contributions under the four considered risk models†. The best result in each risk model is highlighted in bold font.

	Estimate of AC (Bias):				*Standard error ($\sqrt{MSE}$):*		
Estimator	***MC***	***NW***	***GR***	***MH***	***MC***	***GR***	***MH***
(1) Pareto + survival Clayton: True AC = (10.708, 10.708, 10.708)							
AC_1	10.575	11.744	10.745	**10.708**	0.173	**0.008**	0.019
	(-0.133)	(1.036)	(0.037)	**(0.000)**	(0.218)	(0.038)	**(0.019)**
AC_2	10.138	10.547	10.635	**10.724**	0.169	**0.008**	0.020
	(-0.571)	(-0.161)	(-0.074)	**(0.016)**	(0.595)	(0.074)	**(0.025)**
AC_3	10.389	9.813	10.745	**10.693**	0.178	**0.008**	0.018
	(-0.320)	(-0.896)	(0.037)	**(-0.016)**	(0.366)	(0.038)	**(0.024)**
(2) Pareto + t copula: True AC = (7.198, 8.908, 12.206)							
AC_1	6.835	8.162	7.697	**7.339**	0.238	**0.010**	0.041
	(-0.362)	(0.964)	(0.499)	**(-0.121)**	(0.433)	(0.499)	**(0.132)**
AC_2	8.785	8.355	8.740	**8.765**	0.223	**0.010**	0.028
	(-0.122)	(-0.553)	(-0.167)	**(-0.023)**	(0.255)	(0.168)	**(0.046)**
AC_3	11.913	11.781	11.875	**12.208**	0.134	**0.006**	0.024
	(-0.293)	(-0.426)	(-0.332)	**(0.144)**	(0.322)	(0.332)	**(0.148)**
(3) Student t + survival Clayton: True AC = (5.647, 5.647, 5.647)							
AC_1	5.592	5.693	**5.662**	5.617	0.081	**0.006**	0.018
	(-0.055)	(0.046)	**(0.015)**	(-0.029)	(0.098)	**(0.016)**	(0.034)
AC_2	5.410	5.722	**5.642**	5.665	0.079	**0.006**	0.019
	(-0.236)	(0.076)	**(-0.005)**	(0.018)	(0.249)	**(0.007)**	(0.026)
AC_3	5.473	5.517	**5.636**	5.658	0.082	**0.006**	0.018
	(-0.173)	(-0.130)	**(-0.011)**	**(0.011)**	(0.192)	**(0.012)**	(0.021)
(4) Student t + t copula: True AC = (2.996, 3.745, 6.741)							
AC_1	2.821	3.065	**2.997**	2.940	0.117	**0.007**	0.036
	(-0.176)	(0.069)	**(0.001)**	(-0.056)	(0.211)	**(0.007)**	(0.067)
AC_2	3.772	3.560	**3.742**	3.792	0.109	**0.006**	0.033
	(0.027)	(-0.185)	**(-0.004)**	(0.047)	(0.112)	**(0.007)**	(0.057)
AC_3	6.564	6.852	**6.745**	6.751	0.043	**0.002**	0.011
	(-0.178)	(0.110)	**(0.003)**	(0.010)	(0.183)	**(0.004)**	(0.015)

† The estimate is computed for the Monte Carlo (***MC***), Nadaraya-Watson (***NW***), generalized regression (***GR***), and Metropolis-Hastings (***MH***) estimators. The standard error is computed in all cases except for the NW estimator. The sample size is $N = 10^6$ for all methods.

4.5.2 Empirical study

The numerical study is now extended to a high-dimensional case with real-world data. We used the dataset stockdata in the R-package huge, which consists of stock market data of closing prices from all stocks in the S&P 500 for all the days the market was open in the period of January 1, 2003 to January 1, 2008 (five years). During the time period, there remained $d = 452$ stocks in the S&P 500. The sample size is $T = 1258$. We transformed the data into the log-ratio of prices at time t to the price at time $t-1$, so log-returns.

Most stylized facts on stock returns listed in Chapter 3 of McNeil et al. (2015) are observable in the data. For example, return series are unimodal, leptokurtic, and heavy-tailed with little serial correlation and volatility clusters. Moreover, the d return series are mutually dependent. Taking these observations into account, we adopted a copula-GARCH model with skew-t white noise (ST-GARCH; see, for example, Jondeau and Rockinger, 2006; Huang et al., 2009). In the model, d marginal time series are modelled by GARCH$(1,1)$ and the underlying white noise processes follow skew-t distributions with an inhomogeneous degree of freedom $\nu_j > 0$ and skewness parameter $\gamma_j > 0$; that is, within a fixed time period $\{1,\dots,T\}$ the jth return series $(X_{1,j},\dots,X_{T,j})$ follows

$$X_{t,j} = \mu_j + \sigma_{t,j} Z_{t,j}, \quad \sigma^2_{t,j} = \omega_j + \alpha_j X^2_{t-1,j} + \beta_j \sigma^2_{t-1,j}, \quad Z_{t,j} \overset{\text{iid}}{\sim} \mathrm{ST}(\nu_j, \gamma_j)$$

for $t = 2,\dots,T,\ j = 1,\dots,d$, where $\omega_j > 0, \alpha_j, \beta_j \geq 0$, $\alpha_j + \beta_j < 1$, and $Z_{t,j}$ follows a skew-t distribution $\mathrm{ST}(\nu_j, \gamma_j)$ with density given by

$$f_j(x_j;\nu_j,\gamma_j) = \frac{2}{\gamma + \frac{1}{\gamma}} \left\{ t(x_j,\nu_j)\mathbf{1}_{[x_j \geq 0]} + t(\gamma_j x_j, \nu_j)\mathbf{1}_{[x_j<0]} \right\}, \tag{4.21}$$

where $t(x,\nu)$ is the density of a Student t distribution with $\nu > 0$ degrees of freedom and skewness parameter $\gamma > 0$ with $\gamma = 1$ meaning symmetric; see Fernández and Steel (1998) for more details. The copula among $\boldsymbol{Z}_t = (Z_{t,1},\dots,Z_{t,d})$ is assumed to be a Student t copula with parameters ν and P independent of time t.

We estimated the parameters of the marginal ST-GARCH(1,1) models and the t copula based on the copula approach. First, we fitted the ST-GARCH(1,1) models with the maximum likelihood method to the marginal time series. Then, to obtain pseudo-samples from the copula of $\boldsymbol{Z}$, distributional transforms were applied componentwise to the d-dimensional white noise process extracted from the ST-GARCH models. We finally fit the t copula to them with the method-of-moment using Kendall's tau for the dispersion matrix P and the maximum likelihood method for the degree of freedom ν; see Demarta and McNeil (2005) for more details. The results of the estimation are summarized in Figure 4.3. From (B1) and (B5), we observe that the estimates of means and omegas are almost 0. From (B3), most of the marginal white noise distributions are symmetrical but some are skewed. From (B4), their degrees of freedom range from two to ten, that is, the tail-heaviness of the return series is inhomogeneous over the d assets considered. Finally, (B8) shows that the pairwise correlations among the return series are typically from 0.2 to 0.4, and some have strong positive correlations.

Our goal in this study is to compute the conditional VaR contributions at time $T+1$ given the history $\mathcal{F}_t$. Under the model described above, the marginal distribution of the jth return at time $T+1$ is $X_{T+1,j|\mathcal{F}_T} \sim$

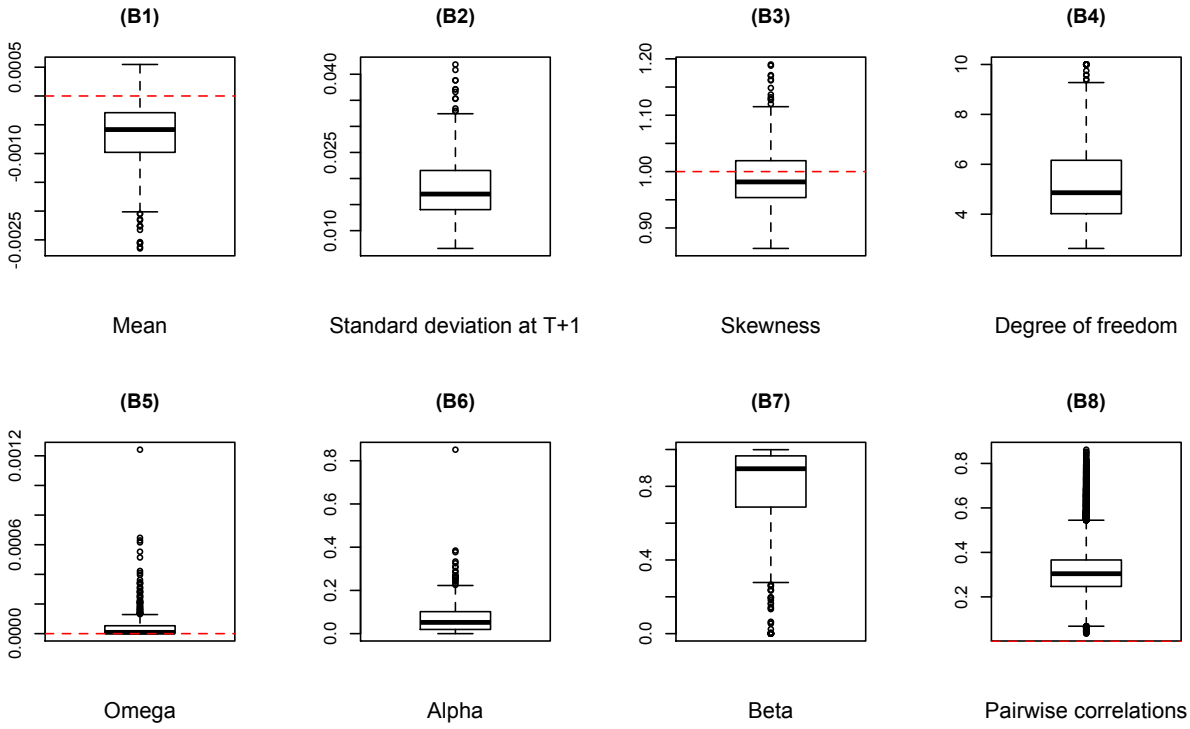

Figure 4.3: Boxplots of d estimates of each parameter (B1) μ_j, (B2) $\sigma_{T+1,j}$, (B3) γ_j, (B4) ν_j, (B5) ω_j, (B6) α_j, (B7) β_j, and (B8) ρ_{j_1,j_2} for $j = 1,\dots,d$ and $j_1, j_2 \in \{1,\dots,d\}$ of the ST-GARCH(1,1) models $X_{t,j} = \mu_j + \sigma_{t,j} Z_{t,j}$, $\sigma_{t,j}^2 = \omega_j + \alpha_j X_{t-1,j}^2 + \beta_j \sigma_{t-1,j}^2$, where $Z_{t,j} \sim \mathrm{ST}(\nu_j, \gamma_j)$, independently and identically for $t = 1,\dots,T+1$, $j = 1,\dots,d$, with a t copula with parameters ν and P. The estimate of the degrees of freedom of the t copula was $\hat{\nu} = 89.039$.

$\mathrm{ST}(\mu_j, \sigma_{t+1,j}^2, \nu_j, \gamma_j)$, that is, a skew t distribution with density $f_j(\frac{x_j - \mu_j}{\sigma_{t+1,j}}; \nu_j, \gamma_j)$ with $f_j(\cdot; \nu_j, \gamma_j)$ defined in (4.21). Their copula is a Student t copula with parameters ν and P. Based on this multivariate model, conditional VaR contributions at time $T+1$ given the history $\mathcal{F}_t$ are estimated by the same procedure as in Section 4.5.1.

We estimated the conditional VaR contributions $(\mathrm{AC}_1^{T+1}, \dots, \mathrm{AC}_d^{T+1})$ with confidence level $p = 0.999$ by using the MC, NW, GR and MH estimators. For MC, $N = 10^5$ samples were generated and the total VaR was estimated as the Npth largest sample among them. The run time of the MC simulation was 2.690 minutes. The MC estimates of VaR contributions were then computed as sample means of the conditional samples whose sums were in the set $A_\delta = [v - \delta, v + \delta]$. The bandwidth was set to be $\delta = 4.8$ so that there were $M_{\delta,N} = 733$ conditional MC samples. Estimates of standard errors were also computed based on these samples. NW, GR and MH estimators were computed analogously to the previous simulation study in Section 4.5.1. For the MH estimator, the MpCN proposal distribution was chosen since the target distribution was expected to be heavy-tailed and elliptical to some extent. The length of the sample path was chosen to be $N = 10^4$, and the run time of the MH algorithm was 5.487 minutes. We inspected

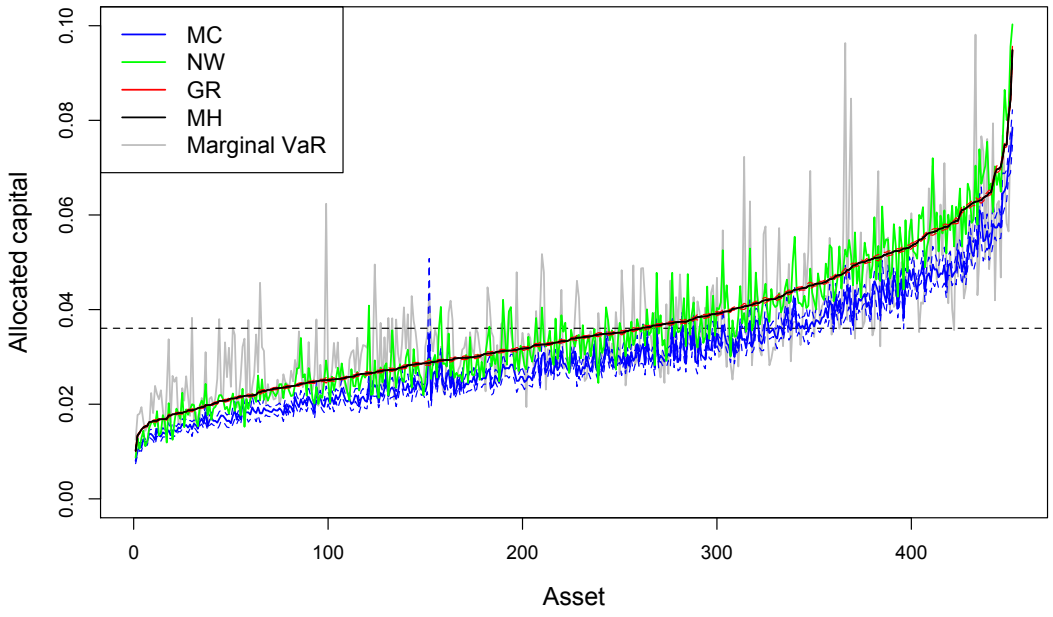

Figure 4.4: Monte Carlo (MC; blue), Nadaraya-Watson (NW; green), Generalized regression (GR; red), and Metropolis-Hastings (MH; black) estimates of conditional VaR contributions at time $T+1$ given $\mathcal{F}_T$ plotted with standardized marginal value-at-risks (gray) and the homogeneously allocated capitals (dotted black line). The colored dotted lines represent the 95% confidence upper or lower bounds of the MC, GR, or MH estimates. The black dashed line is the equal allocation over all the assets.

the autocorrelation plots and ACR to check the validity of the proposal distribution. We observed that all autocorrelations decreased below 0.1 if lags were larger than 40. Together with the ACR 0.983, we concluded that the choice of q was appropriate.

Figure 4.4 shows the MC, NW, GR and MH estimates of the conditional VaR contributions $(\mathrm{AC}_1^{T+1}, \ldots, \mathrm{AC}_d^{T+1})$ of returns at time $T+1$ given the history $\mathcal{F}_T$ plotted with the homogeneously allocated capitals $\mathrm{VaR}_p(S \mid \mathcal{F}_T)/d$ and the standardized marginal VaRs, which is also called the *proportional allocations* in Dhaene et al. (2012), defined by $\mathrm{VaR}_p(X_{T+1,j} \mid \mathcal{F}_T)\Delta_p(\boldsymbol{X}_{T+1} \mid \mathcal{F}_T)$, where $\Delta_p(\boldsymbol{X}_{T+1} \mid \mathcal{F}_T)$ is the so-called *superadditivity ratio* defined by

$$\Delta_p(\boldsymbol{X}_{T+1} \mid \mathcal{F}_T) = \frac{\mathrm{VaR}_p(S \mid \mathcal{F}_T)}{\sum_{j=1}^{d} \mathrm{VaR}_p(X_{T+1,j} \mid \mathcal{F}_T)}.$$

For the MC, GR and MH estimators, the 95% confidence upper and lower bounds are also plotted. On the x-axis, the 452 assets are rearranged in increasing order of the MH estimates.

Compared with the dashed line representing the equal allocation, all the estimated allocated capitals show inhomogeneity among assets. Overall, the estimated VaR contributions are less volatile than the

standardized marginal VaRs, which could be because of the diversification effect. We can also observe that the MH estimates and GR estimates almost coincide for all d assets. The confidence intervals of both estimators are much tighter than those of the MC estimator. NW estimates fluctuate around the line of the MH and GR estimates. On the other hand, the MC estimates deviate from these lines, which indicates that the MC estimator contains inevitable bias. In summary, although the true ACs are unknown, the GR and MH estimators retain stable performance compared with the MC and NW estimators even if the dimension d is large and marginal distributions are inhomogeneous.

4.5.3 Advantages and disadvantages of the MH estimator

We now summarize the advantages and disadvantages of the MH estimator compared with the other estimators. The first advantage is that the MH estimator is consistent whereas this is not always true for the other estimators. As explained in Section 4.2, the MC, NW, and GR estimators have biases which cannot be easily eliminated. In fact, we observed in Table 4.1 that unignorable biases of the MC, NW, and GR estimators sometimes remain even when their standard errors are sufficiently small. In contrast, the MH estimator provides more accurate estimates of VaR contributions as $N \to \infty$ due to its consistency. Since the CLT also holds, confidence intervals for the true VaR contributions are also available. Secondly, the MH estimator has great sample efficiency compared with the MC estimator. While samples are generated from $F_{\boldsymbol{X}}$ and most are discarded in the MC method, no samples are wasted in the MH method since it directly simulates from $F_{\boldsymbol{X}|\{S=v\}}$. Consequently, the MH estimator can achieve low standard errors. Finally, the MH estimator can maintain high performance even when the conditional distribution $F_{\boldsymbol{X}'|\{S=v\}}$ is multimodal or heavy-tailed. As discussed in Section 4.5.1, the performance of the GR estimator highly depends on the shape of $F_{\boldsymbol{X}'|\{S=v\}}$. On the other hand, for the MH estimator, the shape of $F_{\boldsymbol{X}'|\{S=v\}}$ can be directly captured through the proposal distribution q. By choosing an appropriate proposal distribution q according to the shape of $F_{\boldsymbol{X}'|\{S=v\}}$, the MH estimator can attain great performance. This advantage, however, can be seen as a disadvantage from the viewpoint of the simplicity of estimation. In general, estimation with MH requires two steps: the first is to choose a family of proposal distributions, and the second is to determine its parameters. The second step of parameter estimation can be based on the MC samples falling into the set A_δ; these samples are regarded as the *pseudo* samples from $F_{\boldsymbol{X}|\{S=v\}}$. Meanwhile, the first step is not so straightforward. We will discuss this issue in Section 4.5.4 below. Another disadvantage of the MH estimator is that it typically requires a longer run time than other existing estimators. Since MH requires N times simulation from the proposal distribution and evaluation of the acceptance probability (1.11), a careful implementation and proposal selection are necessary to save computational time.

4.5.4 Guidelines for the choice of proposal distribution

A significant drawback of the MH estimator is that the choice of an appropriate proposal distribution q is not as simple as the parameter selections of other existing estimators. An instruction for selecting a proposal distribution is necessary since it highly affects the performance of the MH estimator. In this section, we first investigate the symptoms caused by an inappropriate choice of q. Then, we consider how

to overcome these problems based on the numerical experiments provided above. Practical guidelines for choosing an appropriate proposal distribution are also provided.

An inappropriate choice of q is largely classified into two cases. One is that a proposal distribution q often generates a candidate of which the probability measured by π is quite small. This case occurs, for example, when q does not fully capture the shape of π. In such a case, the Markov chain moves quite slowly and this yields a high asymptotic standard error of the MH estimator. This situation results in quite a low acceptance rate and high autocorrelations. Another case is when q generates only some parts of the whole support of π. This case occurs, for example, when π has distinct local modes and the variance of q is so small that the chain cannot pass between ridges. In such a case, an estimate can be significantly biased, although the acceptance rate and autocorrelation plots are seemingly perfect. This situation results in a distorted plot of MCMC samples whose shapes are completely different from the target distribution π.

How can we detect and avoid such situations? First, as mentioned in Section 1.4.2, it is indispensable to inspect the autocorrelation plots and ACR to prevent the first situation. Additionally, to avoid the second situation, we recommend plotting the generated Markov chain and comparing the samples with the plots of the MC samples whose componentwise sums belong to $A_\delta = [v-\delta, v+\delta]$. Since such MC samples follow the distribution $F_{\boldsymbol{X}|\{S\in A_\delta\}}$, one can detect the distortion of the generated Markov chain by comparing the two scatter plots of $F_{\boldsymbol{X}|\{S=v\}}$ and $F_{\boldsymbol{X}|\{S\in A_\delta\}}$. As an example from our simulation study in Section 4.5.1, Figure 4.5 shows scatter plots of the MC samples whose sums belong to $[v-\delta, v+\delta]$ overlaid with scatter plots of the MH samples. From this figure, we can see that the shapes of the scatter plots of the MH samples bear striking resemblance with those of the MC samples for all risk models. If some part of the support of π is covered by the MC samples but not by the MH samples, the choice of q is questionable.

Finally, through numerical experiments we found that dependence information of the underlying risk model can be helpful for the selection of q. When the copula C of the underlying risk model only has positive dependence for all pairs of loss variables, then the conditional distribution $F_{\boldsymbol{X}|\{S=v\}}$ is likely to be unimodal and light-tailed since positive dependence among $X_1, \dots, X_d$ prevents them from being diversified under the constraint $\{X_1 + \cdots + X_d = v\}$. In Models (1) and (3) in Section 4.5.1, where the copula C has only positive dependence, the contour plots in Figures 4.2 (i) and (iii) show that $F_{\boldsymbol{X}'|\{S=v\}}$ is unimodal and light-tailed. These features facilitate the estimation with MH since simple proposal distributions such as the random walk proposal (1.15) and the independent proposal (1.16) can perform well. Conversely, when C has negative dependence, $F_{\boldsymbol{X}'|\{S=v\}}$ tends to be multimodal or heavy-tailed since negative dependence allows each component of $\boldsymbol{X}$ to take extreme values under $\{X_1 + \cdots + X_d = v\}$. In Models (2) and (4) in Section 4.5.1, where C has negative dependence, Figure 4.2 (ii) indicates that $F_{\boldsymbol{X}'|\{S=v\}}$ is bimodal, and the contour plot in Figure 4.5 (d) shows that $F_{\boldsymbol{X}'|\{S=v\}}$ is heavy-tailed. In such cases, a careful proposal selection is required for achieving an efficient MH estimator. When the losses $X_1, \dots, X_d$ are all nonnegative, then $F_{\boldsymbol{X}'|\{S=v\}}$ is supported on the bounded simplex $\mathcal{S}_v$. Therefore, one can cover the whole support of $F_{\boldsymbol{X}'|\{S=v\}}$ by choosing q as the independent proposal with the distribution defined on the simplex. The uniform distribution on $\mathcal{S}_v$ can be the safest choice. It is also possible to choose other distributions that share the same features of $F_{\boldsymbol{X}'|\{S=v\}}$ observed in the MC samples. For instance, since bimodality is observed in the contour plot in Figure 4.5 (b), we choose q as the independent proposal distribution with f being the Dirichlet distribution on $\mathcal{S}_v$, which can possess two distinct modes around

the edges of the simplex. When $\boldsymbol{X}$ is $\mathbb{R}^d$-valued and negatively dependent, obtaining efficient MCMC algorithm is challenging since the target distribution $F_{\boldsymbol{X}'|\{S=v\}}$ is likely to be multimodal or heavy-tailed. As a special case, when $F_{\boldsymbol{X}}$ is not far away from an elliptical distribution, then $F_{\boldsymbol{X}'|\{S=v\}}$ is likely to be elliptical again. In such a case, even if it is heavy-tailed, the MpCN proposal distribution (1.17) is known to perform well, which is also demonstrated by the simulation study of the risk model (4) in Section 4.5.1 and by the empirical study in Section 4.5.2.

The discussions about choosing an appropriate proposal distribution can be summarized in terms of a flowchart; Figure 4.6. Together with these guidelines, the whole procedure of our MH estimator of VaR contributions presented in this chapter can be summarized as follows.

1. Generate $\boldsymbol{X}_1, \dots, \boldsymbol{X}_M \overset{\text{iid}}{\sim} F_{\boldsymbol{X}}$ by MC.

2. Based on the samples generated in Step 1, estimate VaR by $v = \widehat{\mathrm{VaR}}_p(S)$.

3. For a bandwidth $\delta > 0$, extract subsamples such that $\mathbf{1}_d^\top \boldsymbol{X}_m \in [v-\delta, v+\delta]$ for $m = 1, \dots, M$.

4. Choose a family of proposal distributions according to the guideline in Figure 4.6.

5. Based on the pseudo-samples extracted in Step 3, determine the parameters of the proposal distribution q.

6. For a sample size $N > 0$, proposal density q and the initial value $\boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}$, run Algorithm 1 to generate an N-path $(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)})$ of a Markov chain whose stationary distribution is $f_{\boldsymbol{X}|\{S=v\}}$.

7. To check the validity of the proposal distribution q, compute the acceptance rate, draw the autocorrelation plots, and compare the scatter plots of the MC and MH samples.

8. If the proposal selection is verified in Step 7, compute the MH estimator of VaR contributions (4.7) based on the sample path generated in Step 6. Otherwise, go to Step 4 and choose another proposal distribution.

4.6 Concluding remarks

Computing VaR contributions for a risk model specified by a joint density is in general a difficult task. To this end, we propose the MH estimator of VaR contributions. Its sample efficiency is significantly improved since the MH method generates samples directly from the conditional density given the sum constraint. Moreover, since the MH estimator can capture the features of the risk model more directly than the existing estimators, it can maintain high performance even when the underlying loss distribution is multimodal or heavy-tailed. By the general theory of Markov chains, the MH estimator is consistent and

asymptotically normal. Through simulation and empirical studies based on real-world data, the performance of the MH estimator was compared to those of other existing estimators for various risk models. The numerical results demonstrated that in most risk models, the MH estimator had smaller bias and RMSE compared with the other estimators considered even when the dimension of the portfolio was high, such as $d \approx 500$.

Potential future research includes a theoretical study of the conditional joint distribution of $\boldsymbol{X} \mid \{S = v\}$. Our main interest here is in the influence of the underlying copula of a risk model on the tail behavior and multimodality of the density $f_{\boldsymbol{X}|\{S=v\}}$, which will be partially answered in Chapter 6. We believe that revealing such relationships can provide more promising guidelines for the proposal selection of the MH estimator.

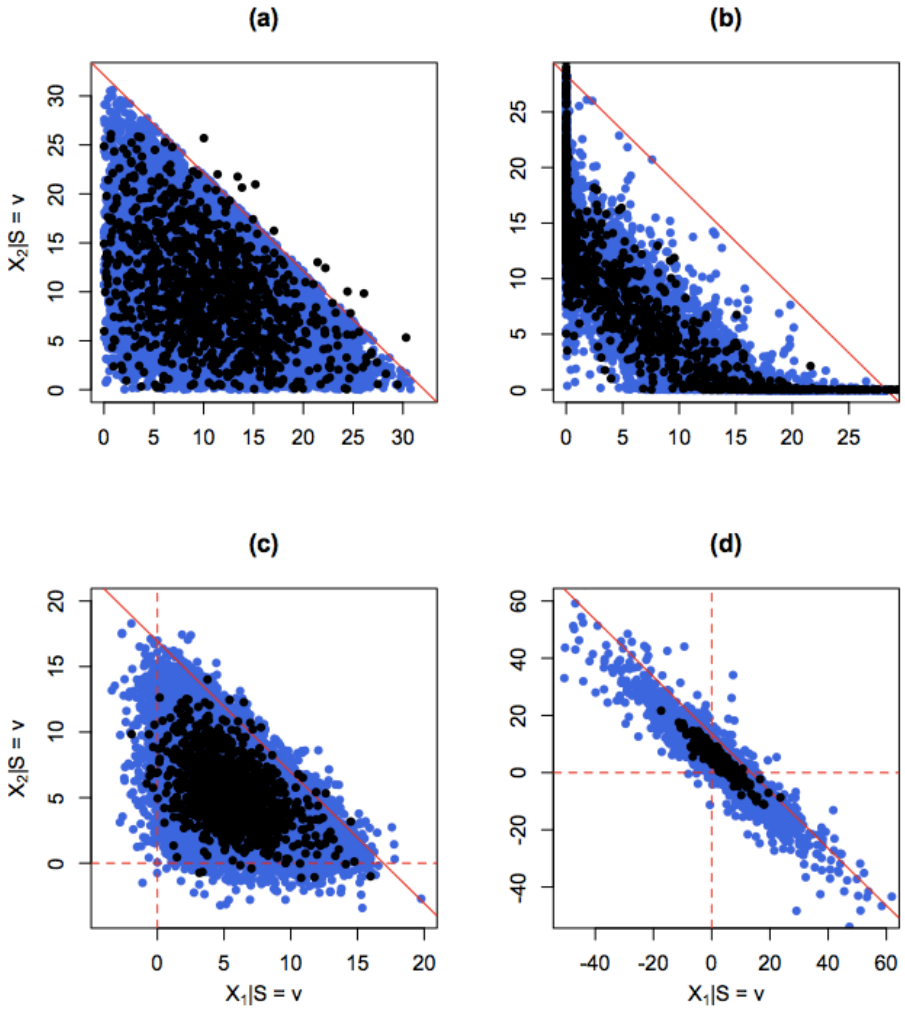

Figure 4.5: Scatter plots of the Monte Carlo (MC; black) and Metropolis-Hastings (MH; blue) samples for different risk models: (a) Pareto + survival Clayton, (b) Pareto + t copula, (c) Student t + survival Clayton, (d) Student t + t copula. The red lines represent the edges of the v-simplex, where v is the estimate of $\mathrm{VaR}_p(S)$. We plot the MC samples generated from $F_{\boldsymbol{X}}$ such that their sums belong to $A_\delta = [v - \delta, v + \delta]$. In the four risk models, the values of δ are (1) 4.8, (2) 3.9, (3) 2.2, and (4) 1.7. When drawing the scatter plots of the MH samples, we only used every 100th sample points among the original sample paths of Markov chains.

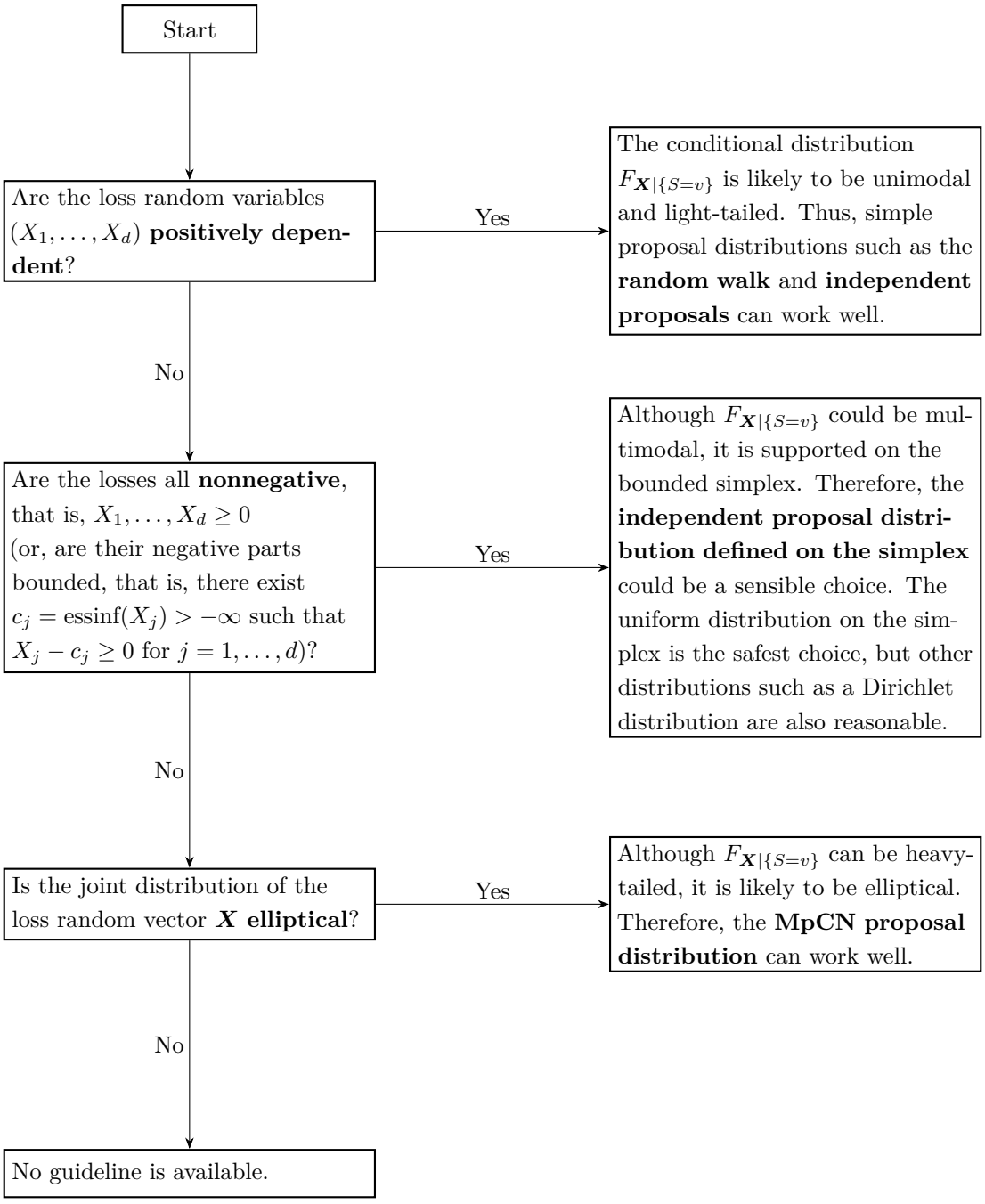

Figure 4.6: Flowchart for choosing the proposal distribution of the Metropolis-Hastings (MH) estimator of value-at-risk contributions.

Chapter 5

Markov Chain Monte Carlo methods for estimating systemic risk allocations

In this chapter, we extend the framework of estimating VaR contributions with MCMC methods presented in Chapter 4 to a more general class of systemic risk measures and risk allocations. We consider a class of allocations whose jth component can be written as some risk measure of the jth conditional marginal loss distribution given the so-called crisis event. By considering a crisis event as an intersection of linear constraints, this class of allocations covers, for example, conditional Value-at-Risk (CoVaR), conditional expected shortfall (CoES), VaR contributions, and range VaR (RVaR) contributions as special cases. For this class of allocations, analytical calculations are rarely available, and numerical computations based on Monte Carlo (MC) methods often provide inefficient estimates due to the rare-event character of the crisis event. We propose an MCMC estimator constructed from a sample path of a Markov chain whose stationary distribution is the conditional distribution given the crisis event. Efficient constructions of Markov chains, such as Hamiltonian Monte Carlo and Gibbs sampler, are suggested and studied depending on the crisis event and the underlying loss distribution. The efficiency of the MCMC estimators is demonstrated in a series of numerical experiments.

5.1 Introduction

In portfolio risk management, *risk allocation* is an essential step to quantify the risk of each unit of a portfolio by decomposing the total risk of the whole portfolio. One of the most prevalent rules to determine risk allocations is the *Euler princple* For the popular risk measures VaR, RVaR, and ES, Euler allocations take the form of conditional expectations of the underlying loss random vector given a certain rare event on the total loss of the portfolio; see Section 1.3 for details. We call this rare event the *crisis event*.

The decomposition of risks is also required in the context of systemic risk measurement. *Systemic risk* is the risk of financial distress of an entire economy as a result of the failure of individual components of the financial system. To quantify such risks, various *systemic risk measures* have been proposed in the literature, such as *conditional VaR (CoVaR)* (Adrian and Brunnermeier (2016)), *conditional expected shortfall (CoES)* (Mainik and Schaanning (2014)) and *marginal expected shortfall (MES)* (Acharya et al. (2017)). These three measures quantify the risk of individuals by taking the VaR, ES and expectation of the individual loss, respectively, under some stressed scenario, that is, given the crisis event. Chen et al. (2013), Hoffmann et al. (2016) and Kromer et al. (2016) proposed an axiomatic characterization of systemic risk measures, where the risk of the aggregated loss in a financial system is first measured and then decomposed into the individual economic entities. Due to the similarity of risk allocations with the derivation of systemic risk measures, we refer to both of them as *systemic risk allocations*. In fact, MES coincides with the Euler allocation of ES, and other Euler allocations can be regarded as special cases of systemic risk measures considered in Gourieroux and Monfort (2013).

Calculating systemic risk allocations given an unconditional joint loss distribution is in general challenging since analytical calculations often require to know the joint distribution of the marginal loss and the aggregated loss. Furthermore, MC estimation suffers from the rare-event character of the crisis event. For computing CoVaR, CoES and MES, Mainik and Schaanning (2014), Bernardi et al. (2017) and Jaworski (2017) derived formulas based on the copula of the marginal and the aggregated loss; Asimit and Li (2018) derived asymptotic formulas based on extreme value theory; and Girardi and Ergün (2013) estimated CoVaR under a multivariate GARCH model. Vernic (2006), Chiragiev and Landsman (2007), Dhaene et al. (2008) and Furman and Landsman (2008) calculated Euler allocations for specific joint distributions. Asimit et al. (2011) derived asymptotic formulas for risk allocations. Furman and Zitikis (2009) and Furman et al. (2018) calculated weighted allocations, which include Euler allocations as special cases, under a Stein-type assumption. Concerning the numerical computation of Euler allocations, Glasserman (2005), Glasserman and Li (2005) and Kalkbrener et al. (2004) considered importance sampling methods, and Siller (2013) proposed the Fourier transform Monte Carlo method, all specifically for credit portfolios. For general copula-based dependence models, analytical calculations of systemic risk allocations are rarely available, and an estimation method is, to the best of our knowledge, only addressed in Targino et al. (2015), where sequential Monte Carlo (SMC) samplers are applied.

We address the problem of estimating systemic risk allocations under general copula-based dependent risks in the case where the copula between the marginal losses and the aggregated loss are not necessarily available. We consider a general class of systemic risk allocations in the form of risk measures of a conditional loss distribution given a crisis event, which includes CoVaR, CoES, MES and Euler allocations as special cases. In our proposed method, the conditional loss distribution, called the *target distribution* π, is simulated by a Markov chain whose stationary distribution is the desired distribution π by sequentially updating the sample path based on the available information from π. While this MCMC method resembles the SMC in Targino et al. (2015), the latter requires a more complicated implementation involving the choice of forward and backward kernels, resampling and move steps, and even MCMC in the move steps. Our suggested approach directly constructs a single sophisticated Markov chain depending on the target distribution of interest. Applications of MCMC to estimating risk allocations have been studied in Chapter 4, specifically for VaR contributions. This chapter explores and demonstrates the applicability of MCMC methods to a

more general class of systemic risk allocations.

Almost all MCMC methods used in practice are of the *Metropolis-Hastings (MH)* type; see Section 1.4. serial correlation, which adversarially affects the efficiency of the As discussed in Chapter 4, an efficient MCMC of MH type is such that the proposal distribution generates a candidate which exhibits low correlation with the current state with sufficiently large acceptance probability. The main difficulty in constructing such an efficient MCMC estimator for systemic risk allocations is that the support of the target distribution π is subject to the constraints determined by the crisis event. For such target distributions, simple MCMC methods such as random walk MH are not efficient since a candidate is immediately rejected if it violates the constraints; see Section 5.3.1 for details.

To tackle this problem, we consider two specific MCMC methods, *Hamiltonian Monte Carlo (HMC)* (Duane et al. (1987)) and the *Gibbs sampler (GS)* (Geman and Geman (1984) and Gelfand and Smith (1990)). In the HMC method, a candidate is generated according to the so-called Hamiltonian dynamics, which leads to a high acceptance probability and low correlation with the current state by accurately simulating the dynamics of sufficiently long length; see Neal et al. (2011) and Betancourt (2017) for an introduction to HMC. Moreover, the HMC candidates always belong to the crisis event by reflecting the dynamics when the chain hits the boundary of the constraints; see Ruján (1997), Pakman and Paninski (2014), Afshar and Domke (2015), Yi and Doshi-Velez (2017) and Chevallier et al. (2018) for this reflection property of the HMC method. An alternative method to handle the constraints is the GS, in which the chain is updated in each component. Since all the components except the updated one remain fixed, a componentwise update is typically subject to weaker constraints. As long as such componentwise updates are feasible, the GS candidates belong to the crisis event, and the acceptance probability is always 1; see Geweke (1991), Gelfand et al. (1992) and Rodriguez-Yam et al. (2004) for the application of the GS to constrained target distributions, and see Gudmundsson and Hult (2014) and Targino et al. (2015) for applications to estimating risk contributions.

Our findings include efficient MCMC estimators of systemic risk allocations achieved via HMC with reflection and GSs. We assume that the unconditional joint loss density is known, possibly through its marginal densities and copula density. Depending on the supports of the marginal loss distributions and the crisis event, different MCMC methods are applicable. We find that if the marginal loss distributions are one-sided, that is, the supports are bounded from the left, then the crisis event is typically a bounded set and HMC shows good performance. On the other hand, if the marginal losses are two-sided, that is, they have both right and left tails, the crisis event is often unbounded and the GSs perform better, provided that random number generators of the conditional copulas are available. Based on the samples generated by the MC method, we propose heuristics to determine the parameters of the HMC and GS methods, for which no manual interaction is required. Since, in the MCMC method, the conditional loss distribution of interest is directly simulated in contrast to MC where rejection is applied based on the unconditional loss distribution, the MCMC method in general outperforms the MC method in terms of the sample size and thus the standard error. This advantage of MCMC becomes more pronounced as the probability of the crisis event becomes smaller. We demonstrate this efficiency of the MCMC estimators of systemic risk allocations by a series of numerical experiments.

This chapter is organized as follows. The general framework of the estimation problem of systemic risk

allocations is introduced in Section 5.2. Our class of systemic risk allocations is proposed in Section 5.2.1 and their estimation via the MC method is presented in Section 5.2.2. Section 5.3 is devoted to MCMC methods for estimating systemic risk allocations. In Section 5.3.1 we formulate our problem of estimating systemic risk allocations in terms of MCMC. HMC and GS for constrained target distributions are then investigated in Sections 5.3.2 and 5.3.3, respectively. In Section 5.4 numerical experiments are conducted including simulation and empirical studies, and a detailed comparison of MC and our introduced MCMC methods. Section 5.5 concludes with practical guidance and limitations of the presented MCMC methods. Readers are referred to Section 1.3 and Section 1.4 for notations and preliminaries concerning the problem of capital allocation and MCMC methods.

5.2 Systemic risk allocations and their estimation

In this section, we define a broad class of systemic risk allocations including Euler allocations, CoVaR and CoES as special cases. Then we describe the MC method to estimate systemic risk allocations.

5.2.1 A class of systemic risk allocations

An *allocation* $A = (A_1, \dots, A_d)$ is a map from a random vector $\boldsymbol{X}$ to $(A_1(\boldsymbol{X}), \dots, A_d(\boldsymbol{X})) \in \mathbb{R}^d$. The sum $\sum_{j=1}^d A_j(\boldsymbol{X})$ can be understood as the capital required to cover the total loss of the portfolio or the economy. The jth component $A_j(\boldsymbol{X})$, $j = 1, \dots, d$ is then the contribution of the jth loss to the total capital $\sum_{j=1}^d A_j(\boldsymbol{X})$. In this chapter, we consider the following class of allocations

$$A^{\varrho_1,\dots,\varrho_d,\mathcal{C}} = (A_1^{\varrho_1,\mathcal{C}}, \dots, A_d^{\varrho_d,\mathcal{C}}), \quad A_j^{\varrho_j,\mathcal{C}}(\boldsymbol{X}) = \varrho_j(X_j \mid \{\boldsymbol{X} \in \mathcal{C}\}),$$

where ϱ_j is a map from a random variable to $\mathbb{R}$ called the jth *marginal risk measure* for $j = 1, \dots, d$, and $\mathcal{C} \subseteq \mathbb{R}^d$ is a set called the *crisis event*. The conditioning set $\{\boldsymbol{X} \in \mathcal{C}\}$ is simply written as $\mathcal{C}$ if there is no confusion. As we now explain, this class of allocations covers well-known allocations as special cases.

We now define various crisis events and marginal risk measures. A typical form of the crisis event is an intersection of a set of linear constraints

$$\mathcal{C} = \bigcap_{m=1}^{M} \left\{ \boldsymbol{h}_m^\top \boldsymbol{x} \geq v_m \right\}, \quad \boldsymbol{h}_m \in \mathbb{R}^d, \quad v_m \in \mathbb{R}, \quad m = 1, \dots, M, \quad M \in \mathbb{N}. \tag{5.1}$$

Several important special cases of the crisis event of Form (5.1) can be provided.

Definition 5.2.1 (VaR, RVaR and ES crisis events)**.** For $S = \sum_{j=1}^d X_j$, the *VaR*, *RVaR* and *ES crisis events* are defined by

$$\begin{aligned}
\mathcal{C}_p^{\mathrm{VaR}} &= \{\boldsymbol{x} \in \mathbb{R}^d \mid \mathbf{1}_d^\top \boldsymbol{x} = \mathrm{VaR}_p(S)\}, \quad p \in (0,1), \\
\mathcal{C}_{p_1,p_2}^{\mathrm{RVaR}} &= \{\boldsymbol{x} \in \mathbb{R}^d \mid \mathrm{VaR}_{p_1}(S) \leq \mathbf{1}_d^\top \boldsymbol{x} \leq \mathrm{VaR}_{p_2}(S)\}, \quad 0 < p_1 < p_2 \leq 1, \\
\mathcal{C}_p^{\mathrm{ES}} &= \{\boldsymbol{x} \in \mathbb{R}^d \mid \mathrm{VaR}_p(S) \leq \mathbf{1}_d^\top \boldsymbol{x}\}, \quad 0 < p < 1, \quad p \in (0,1),
\end{aligned}$$

respectively, where $\mathbf{1}_d$ is the d-dimensional vector of ones.

Definition 5.2.2 (Risk contributions and conditional risk measures)**.** For $j \in \{1, \dots, d\}$, we call $A_j^{\varrho_j, \mathcal{C}}$ of

1. *risk contribution type* if $\varrho_j = \mathbb{E}$;
2. *CoVaR type* if $\varrho_j = \mathrm{VaR}_{p_j}$ for $p_j \in (0,1)$;
3. *CoRVaR type* if $\varrho_j = \mathrm{RVaR}_{p_{j,1}, p_{j,2}}$ for $0 < p_{j,1} < p_{j,2} \leq 1$; and
4. *CoES type* if $\varrho_j = \mathrm{ES}_{p_j}$ for $p_j \in (0,1)$.

The following examples show that $A_j^{\varrho_j, \mathcal{C}}$ coincides with popular allocations for specific choices of marginal risk measure and crisis event.

Example 5.2.3 (Special cases of $A^{\varrho_1, \dots, \varrho_d, \mathcal{C}}$)**.**

(1) *Risk contributions*: If the crisis event is chosen to be $\mathcal{C}_p^{\mathrm{VaR}}$, $\mathcal{C}_{p_1,p_2}^{\mathrm{RVaR}}$ or $\mathcal{C}_p^{\mathrm{ES}}$, the allocations of the risk contribution type $\varrho_j = \mathbb{E}$ reduce to the *VaR, RVaR or ES contributions* defined by

$$\begin{aligned} \mathrm{VaR}_p(\boldsymbol{X}, S) &= \mathbb{E}[\boldsymbol{X} \mid \{S = \mathrm{VaR}_p(S)\}], \\ \mathrm{RVaR}_{p_1,p_2}(\boldsymbol{X}, S) &= \mathbb{E}[\boldsymbol{X} \mid \{\mathrm{VaR}_{p_1}(S) \leq S \leq \mathrm{VaR}_{p_2}(S)\}], \\ \mathrm{ES}_p(\boldsymbol{X}, S) &= \mathbb{E}[\boldsymbol{X} \mid \{S \geq \mathrm{VaR}_p(S)\}], \end{aligned}$$

respectively; see (1.5), (1.6) and (1.7). The ES contributions are also called the marginal expected shortfall (MES) and used as a systemic risk measure; see Acharya et al. (2017).

(2) *Conditional risk measures*: CoVaR and CoES are systemic risk measures defined by

$$\begin{aligned} \mathrm{CoVaR}_{p_1,p_2}^{=}(X_j, S) &= \mathrm{VaR}_{p_2}(X_j \mid \{S = \mathrm{VaR}_{p_1}(S)\}), \\ \mathrm{CoVaR}_{p_1,p_2}(X_j, S) &= \mathrm{VaR}_{p_2}(X_j \mid \{S \geq \mathrm{VaR}_{p_1}(S)\}), \\ \mathrm{CoES}_{p_1,p_2}^{=}(X_j, S) &= \mathrm{ES}_{p_2}(X_j \mid \{S = \mathrm{VaR}_{p_1}(S)\}), \\ \mathrm{CoES}_{p_1,p_2}(X_j, S) &= \mathrm{ES}_{p_2}(X_j \mid \{S \geq \mathrm{VaR}_{p_1}(S)\}), \end{aligned}$$

for $p_1, p_2 \in (0,1)$; see Mainik and Schaanning (2014) and Bernardi et al. (2017). Our CoVaR and CoES type allocations with crisis events $\mathcal{C} = \mathcal{C}^{\mathrm{VaR}_p}$ or $\mathcal{C}^{\mathrm{ES}_p}$ coincide with those defined in the displayed equations.

Remark 5.2.4 (Weighted allocations)**.** For a measurable function $w : \mathbb{R}^d \to \mathbb{R}_+ = [0, \infty)$, Furman and Zitikis (2008) proposed the *weighted allocation* $\varrho_w(\boldsymbol{X})$ with the *weight function* w being defined by $\varrho_w(\boldsymbol{X}) = \mathbb{E}[\boldsymbol{X} w(\boldsymbol{X})]/\mathbb{E}[w(\boldsymbol{X})]$. By taking an indicator function as weight function $w(\boldsymbol{x}) = \mathbf{1}_{\{\boldsymbol{x} \in \mathcal{C}\}}$ and provided that $\mathbb{P}(\boldsymbol{X} \in \mathcal{C}) > 0$, weighted allocation coincides with the risk contribution type systemic allocation $A^{\mathbb{E}, \dots, \mathbb{E}, \mathcal{C}}$.

5.2.2 Monte Carlo estimation of systemic risk allocations

Even if the joint distribution $F_{\boldsymbol{X}}$ of the loss random vector $\boldsymbol{X}$ is known, the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \mathcal{C}$, denoted by $F_{\boldsymbol{X}|\mathcal{C}}$, is typically too complicated to analytically calculate the systemic risk allocations $A^{\varrho_1,\dots,\varrho_d,\mathcal{C}}$. An alternative approach is to numerically estimate them by the MC method as is done in Yamai and Yoshiba (2002) and Fan et al. (2012). To this end, assume that one can generate i.i.d. samples from $F_{\boldsymbol{X}}$. If $\mathbb{P}(\boldsymbol{X} \in \mathcal{C}) > 0$, the MC estimator of $A_j^{\varrho_j,\mathcal{C}}$, $j = 1, \dots, d$ is constructed as follows:

(MC1) *Sample from* $\boldsymbol{X}$: For a sample size $N \in \mathbb{N}$, generate $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)} \overset{\text{ind.}}{\sim} F_{\boldsymbol{X}}$.

(MC2) *Estimate the crisis event*: If the crisis event $\mathcal{C}$ contains unknown quantities, replace them with their estimates based on $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}$. Denote by $\hat{\mathcal{C}}$ the estimated crisis event.

(MC3) *Sample from the conditional distribution of* $\boldsymbol{X}$ *given* $\hat{\mathcal{C}}$: Among $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}$, determine $\tilde{\boldsymbol{X}}^{(n)}$ such that $\tilde{\boldsymbol{X}}^{(n)} \in \hat{\mathcal{C}}$ for all $n = 1, \dots, N$.

(MC4) *Construct the MC estimator*: The MC estimate of $A_j^{\varrho_j,\mathcal{C}}$ is $\varrho_j(\hat{F}_{\tilde{\boldsymbol{X}}})$ where $\hat{F}_{\tilde{\boldsymbol{X}}}$ is the empirical cdf (ecdf) of the $\tilde{\boldsymbol{X}}^{(n)}$'s.

For an example of (MC2), if the crisis event is $\mathcal{C}^{\mathrm{RVaR}}_{p_1,p_2} = \{\boldsymbol{x} \in \mathbb{R}^d \mid \mathrm{VaR}_{p_1}(S) \le \mathbf{1}_d^\top \boldsymbol{x} \le \mathrm{VaR}_{p_2}(S)\}$, then $\mathrm{VaR}_{p_1}(S)$ and $\mathrm{VaR}_{p_2}(S)$ are unknown parameters, and thus they are replaced by $\mathrm{VaR}_{p_1}(\hat{F}_S)$ and $\mathrm{VaR}_{p_2}(\hat{F}_S)$, where $\hat{F}_S$ is the ecdf of the total loss $S^{(n)} = X_1^{(n)} + \cdots + X_d^{(n)}$ for $n = 1, \dots, N$. By the *law of large numbers (LLN)* and the *central limit theorem (CLT)*, the MC estimator of $A^{\varrho_1,\dots,\varrho_d,\mathcal{C}}$ is consistent, and approximate confidence intervals of the true allocations can be constructed based on asymptotic normality; see Glasserman (2005).

As we discussed in Section 4.2 of Chapter 4, MC cannot handle VaR crisis events if S admits a pdf since $\mathbb{P}(\boldsymbol{X} \in \mathcal{C}^{\mathrm{VaR}}_p) = \mathbb{P}(S = \mathrm{VaR}_p(S)) = 0$ and thus no subsample is picked in (MC3) above. A possible remedy (although the resulting estimator suffers from an inevitable bias) is to replace $\mathcal{C}^{\mathrm{VaR}}_p$ with $\mathcal{C}^{\mathrm{RVaR}}_{p-\delta,p+\delta}$ for sufficiently small $\delta > 0$ so that $\mathbb{P}(S \in \mathcal{C}^{\mathrm{RVaR}}_{p-\delta,p+\delta}) = 2\delta > 0$.

The main advantage of MC for estimating systemic risk allocations $A^{\varrho_1,\dots,\varrho_d,\mathcal{C}}$ is that only a random number generator for $F_{\boldsymbol{X}}$ is required for implementing the method. Furthermore, MC is applicable for any choice of the crisis event $\mathcal{C}$ as long as $\mathbb{P}(\boldsymbol{X} \in \mathcal{C}) > 0$. Moreover, the main computational load is simulating $F_{\boldsymbol{X}}$ in (MC1) above, which is typically not demanding. The disadvantage of the MC method is its inefficiency concerning the rare-event characteristics of $\varrho_1, \dots, \varrho_d$ and $\mathcal{C}$. To see this, consider the case where $\mathcal{C} = \mathcal{C}^{\mathrm{RVaR}}_{p_1,p_2}$ and $\varrho_j = \mathrm{RVaR}_{p_1,p_2}$ for $p_1 = 0.95$ and $p_2 = 0.975$. If the MC sample size is $N = 10^5$, there are $N \times (p_2 - p_1) = 2500$ expected subsamples resulting from (MC3). To estimate RVaR_{p_1,p_2} in (MC4) based on this subsample, only $2500 \times (p_2 - p_1) = 62.5$ samples contribute to computing the estimate, which is in general not enough for statistical inference. This effect of sample size reduction is relaxed if ES and/or the ES crisis events are considered, but is more problematic for the VaR crisis event since there is a trade-off concerning reducing bias and MC error when choosing δ; see Section 4.2 in Chapter 4.

5.3 MCMC estimation of systemic risk allocations

To overcome the drawback of the MC method for estimating systemic risk allocations, we introduce MCMC methods which simulate a given distribution by constructing a Markov chain whose stationary distribution is $F_{\boldsymbol{X}|\mathcal{C}}$. In this section, we study how to construct an efficient MCMC estimator for the different choices of crisis events.

5.3.1 MCMC formulation for estimating systemic risk allocations

Numerous choices of proposal densities q are possible to construct an MH kernel. In this section, we consider how to construct an efficient MCMC method for estimating systemic risk allocations $A^{\varrho_1,\dots,\varrho_d,\mathcal{C}}$ depending on the choice of the crisis event $\mathcal{C}$. Our goal is to simulate the conditional distribution $\boldsymbol{X} \mid \mathcal{C}$ directly by constructing a Markov chain whose stationary distribution is

$$\pi(\boldsymbol{x}) = f_{\boldsymbol{X}|\{\boldsymbol{X}\in\mathcal{C}\}}(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\mathbb{P}(\boldsymbol{X}\in\mathcal{C})}\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{C}\}}, \quad \boldsymbol{x}\in E\subseteq\mathbb{R}^d, \tag{5.2}$$

provided $\mathbb{P}(\boldsymbol{X}\in\mathcal{C})>0$. Samples from this distribution can directly be used to estimate systemic risk allocations with crisis event $\mathcal{C}$ and arbitrary marginal risk measures $\varrho_1,\dots,\varrho_d$. Other potential applications are outlined in Remark 5.3.1.

Remark 5.3.1 (Gini shortfall allocation)**.** Samples from the conditional distribution $F_{\boldsymbol{X}|\mathcal{C}_p^{\mathrm{ES}}}$ can be used to estimate, for example, the *tail-Gini coefficient* $\mathrm{TGini}_p(X_j,S) = \frac{4}{1-p}\operatorname{Cov}(X_j, F_S(S) \mid S\geq \mathrm{VaR}_p(S))$ for $p\in(0,1)$, and the *Gini shortfall allocation* (Furman et al. (2017)) $\mathrm{GS}_p(X_j,S) = \mathbb{E}[X_j \mid S\geq\mathrm{VaR}_p(S)] + \lambda\cdot\mathrm{TGini}_p(X_j,S)$, $\lambda\in\mathbb{R}_+$ more efficiently than by applying the MC method. Another application is to estimate risk allocations derived by optimization given a constant economic capital; see Laeven and Goovaerts (2004) and Dhaene et al. (2012).

We now construct a MH algorithm with target distribution (5.2). To this end, we assume that

(A1) the ratio $f_{\boldsymbol{X}}(\boldsymbol{y})/f_{\boldsymbol{X}}(\boldsymbol{x})$ can be evaluated for any $\boldsymbol{x},\boldsymbol{y}\in\mathcal{C}$, and that

(A2) the support of $f_{\boldsymbol{X}}$ is $\mathbb{R}^d$ or $\mathbb{R}^d_+$.

Regarding (A1), the normalization constant of $f_{\boldsymbol{X}}$ and the probability $\mathbb{P}(\boldsymbol{X}\in\mathcal{C})$ are not necessary to be known since they cancel out in the numerator and the denominator of $\pi(\boldsymbol{y})/\pi(\boldsymbol{x})$. In (A2), the loss random vector $\boldsymbol{X}$ refers to the *profit and loss (P&L)* if $\mathrm{supp}(\boldsymbol{X})=\mathbb{R}^d$ and to *pure losses* if $\mathrm{supp}(\boldsymbol{X})=\mathbb{R}^d_+$. Note that the case $\mathrm{supp}(\boldsymbol{X})=[c_1,\infty]\times\cdots\times[c_d,\infty]$, $c_1,\dots,c_d\in\mathbb{R}$, is essentially included in the case of pure losses as long as the marginal risk measures $\varrho_1,\dots,\varrho_d$ are law invariant and translation invariant, and the crisis event is the set of linear constraints of Form (5.1). To see this, define $\tilde{X}_j = X_j - c_j$, $j=1,\dots,d$, $\tilde{\boldsymbol{X}} = (\tilde{X}_1,\dots,\tilde{X}_d)$ and $\boldsymbol{c}=(c_1,\dots,c_d)$. Then $\mathrm{supp}(\tilde{\boldsymbol{X}})=\mathbb{R}^d_+$ and $\boldsymbol{X}\mid\{\boldsymbol{X}\in\mathcal{C}\} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{X}}\mid\{\tilde{\boldsymbol{X}}\in\tilde{\mathcal{C}}\}+\boldsymbol{c}$ where $\tilde{\mathcal{C}}$ is the set of linear constraints with parameters $\tilde{\boldsymbol{h}}_m=\boldsymbol{h}_m$ and $\tilde{v}_m = v_m - \boldsymbol{h}_m^\top\boldsymbol{c}$. By law invariance and translation invariance of $\varrho_1,\dots,\varrho_d$,

$$\varrho_j(X_j\mid\{\boldsymbol{X}\in\mathcal{C}\}) = c_j + \varrho_j(\tilde{X}_j\mid\{\tilde{\boldsymbol{X}}\in\tilde{\mathcal{C}}\}), \quad j=1,\dots,d.$$

Therefore, the problem of estimating $A^{\varrho_1,\dots,\varrho_d,\mathcal{C}}(\boldsymbol{X})$ reduces to that of estimating $A^{\varrho_1,\dots,\varrho_d,\tilde{\mathcal{C}}}(\tilde{\boldsymbol{X}})$ for the shifted loss random vector $\tilde{\boldsymbol{X}}$ (such that $\operatorname{supp}(\boldsymbol{X}) = \mathbb{R}^d_+$) and the modified crisis event of the same form.

For the P&L case, the RVaR and ES crisis events are the set of linear constraints of Form (5.1) with the number of constraints $M = 2$ and 1, respectively. In the case of pure losses, additional d constraints $\boldsymbol{e}_{j,d}^\top \boldsymbol{x} \geq 0$, $j = 1, \dots, d$ are imposed where $\boldsymbol{e}_{j,d}$ is the jth d-dimensional unit vector. Therefore, the RVaR and ES crisis events are of Form (5.1) with $M = d+2$ and $d+1$, respectively. For the VaR crisis event, $\mathbb{P}(\boldsymbol{X} \in \mathcal{C}) = 0$ and thus (5.2) cannot be defined properly. In this case, the allocation $A^{\varrho_1,\dots,\varrho_d,\mathcal{C}^{\mathrm{VaR}}}$ depends on the conditional joint distribution $\boldsymbol{X} \mid \mathcal{C}_p^{\mathrm{VaR}}$ but is completely determined by its first $d' = d-1$ variables $(X_1, \dots, X_{d'}) \mid \mathcal{C}_p^{\mathrm{VaR}}$ since $X_d \mid \mathcal{C}_p^{\mathrm{VaR}} \overset{\mathrm{d}}{=} (\mathrm{VaR}_p(S) - \sum_{j=1}^{d'} X_j) \mid \mathcal{C}_p^{\mathrm{VaR}} \overset{\mathrm{d}}{=} \mathrm{VaR}_p(S) - \sum_{j=1}^{d'} X_j \mid \mathcal{C}_p^{\mathrm{VaR}}$. Estimating systemic risk allocations under the VaR crisis event can thus be achieved by simulating the target distribution

$$\begin{aligned}\pi^{\mathrm{VaR}_p}(\boldsymbol{x}') = f_{\boldsymbol{X}'|\{S=\mathrm{VaR}_p(S)\}}(\boldsymbol{x}) &= \frac{f_{(\boldsymbol{X}',S)}(\boldsymbol{x}', \mathrm{VaR}_p(S))}{f_S(\mathrm{VaR}_p(S))} \\ &= \frac{f_{\boldsymbol{X}}(\boldsymbol{x}', \mathrm{VaR}_p(S) - \mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(\mathrm{VaR}_p(S))} \mathbf{1}_{\{\mathrm{VaR}_p(S) - \mathbf{1}_{d'}^\top \boldsymbol{x}' \in \operatorname{supp}(f_d)\}}, \quad \boldsymbol{x}' \in \mathbb{R}^{d'}, \end{aligned} \tag{5.3}$$

where $\boldsymbol{X}' = (X_1, \dots, X_{d'})$ and the last equation is derived from the linear transformation $(\boldsymbol{X}', S) \mapsto \boldsymbol{X}$ with unit Jacobian. Note that other transformations are also possible; see Betancourt (2012). Under Assumption (A1), the ratio $\pi^{\mathrm{VaR}_p}(\boldsymbol{y})/\pi^{\mathrm{VaR}_p}(\boldsymbol{x})$ can be evaluated and $f_S(\mathrm{VaR}_p(S))$ is not required to be known. In the case of pure losses, the target distribution π^{VaR_p} is subject to d linear constraints $\boldsymbol{e}_{j,d'}^\top \boldsymbol{x}' \geq 0$, $j = 1, \dots, d'$, and $\mathbf{1}_{d'}^\top \boldsymbol{x}' \geq \mathrm{VaR}_p(S)$ where the first d' constraints come from the non-negativity of the losses and the last one is from the indicator in (5.3). Therefore, the crisis event $\mathcal{C}^{\mathrm{VaR}}$ for $(X_1, \dots, X_{d'})$ is of Form (5.1). In the case of P&L, $\operatorname{supp}(f_d) = \mathbb{R}$ and $\mathrm{VaR}_p(S) - \mathbf{1}_{d'}^\top \boldsymbol{x}' \in \operatorname{supp}(f_d)$ holds for any $\boldsymbol{x}' \in \mathbb{R}^{d'}$. Therefore, the target distribution (5.3) is free from any constraints and the problem reduces to constructing an MCMC method with target distribution $\pi(\boldsymbol{x}') \propto f_{\boldsymbol{X}}(\boldsymbol{x}', \mathrm{VaR}_p(S) - \mathbf{1}_{d'}^\top \boldsymbol{x}')$, $\boldsymbol{x}' \in \mathbb{R}^{d'}$. In this chapter the P&L case with VaR crisis event is not investigated further since our focus is the simulation of constrained target distributions; see Chapter 4 for MCMC estimation in the P&L case.

MCMC methods to simulate constrained target distributions require careful design of the proposal density q. A simple MCMC method is *Metropolis-Hastings with rejection* in which the support of the proposal density q may not coincide with that of the target distribution, which is the crisis event $\mathcal{C}$, and a candidate is immediately rejected when it violates the constraints. This construction of MCMC is often inefficient due to a low acceptance probability especially around the bondary of $\mathcal{C}$. An efficient MCMC method in this case can be expected only when the probability mass of π is concentrated near the center of $\mathcal{C}$. In the following sections, we introduce two alternative MCMC methods for the constrained target distributions $F_{\boldsymbol{X}|\mathcal{C}}$ of interest, the HMC method and the GS. Each of them is applicable and can be efficient for different choices of the crisis event and underlying loss distribution functions $F_{\boldsymbol{X}}$.

5.3.2 Estimation with Hamiltonian Monte Carlo

We find that if the HMC method is applicable, it is typically the most preferable method to simulate constrained target distributions because of its efficiency and ease of handling constraints. In Section 5.3.2, we briefly present the HMC method with reflection for constructing a Markov chain supported on the

Algorithm 4 Leapfrog method for Hamiltonian dynamics

Input: Current states $(\boldsymbol{x}(0), \boldsymbol{p}(0))$, stepsize $\epsilon > 0$, gradients ∇U and ∇K.

Output: Updated position $(\boldsymbol{x}(\epsilon), \boldsymbol{p}(\epsilon))$.

1) Set $\boldsymbol{p}\left(\frac{\epsilon}{2}\right) = \boldsymbol{p}(0) - \frac{\epsilon}{2}\nabla U(\boldsymbol{x}(0))$.

2) Set $\boldsymbol{x}(\epsilon) = \boldsymbol{x}(0) + \epsilon \nabla K(\boldsymbol{p}\left(\frac{\epsilon}{2}\right))$.

3) Set $\boldsymbol{p}(\epsilon) = \boldsymbol{p}(\epsilon/2) + \frac{\epsilon}{2}\nabla U(\boldsymbol{x}(\epsilon))$.

constrained space. In Section 5.3.2 we propose a heuristic for determining the parameters of the HMC method based on MC presamples.

Hamiltonian Monte Carlo with reflection

For the possibly unnormalized target density π, consider the *potential energy* $U(\boldsymbol{x})$, *kinetic energy* $K(\boldsymbol{p})$ and the *Hamiltonian* $H(\boldsymbol{x}, \boldsymbol{p})$ defined by

$$U(\boldsymbol{x}) = -\log \pi(\boldsymbol{x}), \quad K(\boldsymbol{p}) = -\log f_K(\boldsymbol{p}) \quad \text{and} \quad H(\boldsymbol{x}, \boldsymbol{p}) = U(\boldsymbol{x}) + K(\boldsymbol{p}),$$

with *position variable* $\boldsymbol{x} \in E$, *momentum variable* $\boldsymbol{p} \in \mathbb{R}^d$ and *kinetic energy density* $f_K(\boldsymbol{p})$ such that $f_K(-\boldsymbol{p}) = f_K(\boldsymbol{p})$. In this chapter, the kinetic energy distribution F_K is set to be the multivariate standard normal with $K(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^\top \boldsymbol{p}$ and $\nabla K(\boldsymbol{p}) = \boldsymbol{p}$; other choices of F_K are discussed in Remark 5.3.3. In the HMC method, a Markov chain augmented on the state space $E \times \mathbb{R}^d$ with the stationary distribution $\pi(\boldsymbol{x}) f_K(\boldsymbol{p})$ is constructed and the desired samples from π are obtained as the first $|E|$-dimensional margins. A process $(\boldsymbol{x}(t), \boldsymbol{p}(t))$, $t \in \mathbb{R}$ on $E \times \mathbb{R}^d$ is said to follow the *Hamiltonian dynamics* if it follows the ordinary differential equation (ODE)

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = \nabla K(\boldsymbol{p}), \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p}(t) = -\nabla U(\boldsymbol{x}). \tag{5.4}$$

Through the Hamiltonian dynamics, the Hamiltonian H is preserved, that is, $\mathrm{d}H(\boldsymbol{x}(t), \boldsymbol{p}(t))/\mathrm{d}t = 0$; moreover, the volume is also preserved in the sense that the map $(\boldsymbol{x}(0), \boldsymbol{p}(0)) \mapsto (\boldsymbol{x}(t), \boldsymbol{p}(t))$ has a unit Jacobian for any $t \in \mathbb{R}$; see Neal et al. (2011). Therefore, the value of the joint target density $\pi \cdot f_K$ remains unchanged by the Hamiltonian dynamics, that is,

$$\pi(\boldsymbol{x}(0)) f_K(\boldsymbol{p}(0)) = \exp(-H(\boldsymbol{x}(0), \boldsymbol{p}(0))) = \exp(-H(\boldsymbol{x}(t), \boldsymbol{p}(t))) = \pi(\boldsymbol{x}(t)) f_K(\boldsymbol{p}(t)), \quad t \geq 0.$$

In practice, the dynamics (5.4) are discretized for simulation by, for example, the so-called *leapfrog method* summarized in Algorithm 4; see Leimkuhler and Reich (2004) for other discretization methods. Note that the evaluation of ∇U does not require the normalization constant of π to be known since $\nabla U = -(\nabla \pi)/\pi$. By repeating the leapfrog method T times with stepsize ϵ, the Hamiltonian dynamics are approximately

Algorithm 5 Hamiltonian Monte Carlo to simulate π

Require: Random number generator of F_K, $\boldsymbol{x}^{(0)} \in \mathrm{supp}(\pi)$, $\pi(\boldsymbol{y})/\pi(\boldsymbol{x})$, $\boldsymbol{x}, \boldsymbol{y} \in E$ and $f_K(\boldsymbol{p}')/f_K(\boldsymbol{p})$, $\boldsymbol{p}, \boldsymbol{p}' \in \mathbb{R}^d$.

Input: Sample size $N \in \mathbb{N}$, kinetic energy density f_K, target density π, gradients of the potential and kinetic energies ∇U and ∇K, stepsize $\epsilon > 0$, integration time $T \in \mathbb{N}$ and initial position $\boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}$.

Output: Sample path $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}$ of the Markov chain.

for $n = 0, \ldots, N-1$ **do**

1) Generate $\boldsymbol{p}^{(n)} \sim F_K$.

2) Set $(\tilde{\boldsymbol{X}}^{(n)}, \tilde{\boldsymbol{p}}^{(n)}) = (\boldsymbol{X}^{(n)}, \boldsymbol{p}^{(n)})$.

3) **for** $t = 1, \ldots, T$,

$$(\tilde{\boldsymbol{X}}^{(n+t/T)}, \tilde{\boldsymbol{p}}^{(n+t/T)}) = \textbf{Leapfrog}(\tilde{\boldsymbol{X}}^{(n+(t-1)/T)}, \tilde{\boldsymbol{p}}^{(n+(t-1)/T)}, \epsilon, \nabla U, \nabla K).$$

end for

4) $\tilde{\boldsymbol{p}}^{(n+1)} = -\boldsymbol{p}^{(n+1)}$.

5) Calculate $p_n = \min\left\{\frac{\pi(\tilde{\boldsymbol{X}}^{(n+1)}) f_K(\tilde{\boldsymbol{p}}^{(n+1)})}{\pi(\boldsymbol{X}^{(n)}) f_K(\boldsymbol{p}^{(n)})}, 1\right\}$.

6) Set $\boldsymbol{X}^{(n+1)} = \mathbf{1}_{\{U \le p_n\}} \tilde{\boldsymbol{X}}^{(n+1)} + \mathbf{1}_{\{U > p_n\}} \boldsymbol{X}^{(n)}$ for $U \sim \mathrm{U}(0,1)$.

end for

simulated with length $T\epsilon$. Due to the discretization error the Hamiltonian is not exactly preserved while it is expected to be almost preserved for ϵ small enough. The discretization error $H(\boldsymbol{x}(T\epsilon), \boldsymbol{p}(T\epsilon)) - H(\boldsymbol{x}(0), \boldsymbol{p}(0))$ is called the *Hamiltonian error*.

All the steps of the HMC method are described in Algorithm 5. In Step 1), the momentum variable is first updated from $\boldsymbol{p}(0)$ to $\boldsymbol{p}$ where $\boldsymbol{p}$ follows the kinetic energy distribution F_K so that the value of the Hamiltonian $H = -\log(\pi \cdot f_K)$ changes. In Step 3), the current state $(\boldsymbol{x}(0), \boldsymbol{p})$ is moved along the level curve of $H(\boldsymbol{x}(0), \boldsymbol{p})$ by simulating the Hamiltonian dynamics. By flipping the momentum in Step 4), the HMC method is shown to be reversible w.r.t. π (c.f. (1.8)) and thus to have the stationary distribution π; see Neal et al. (2011) for details. Furthermore, by the conservation property of the Hamiltonian dynamics, the acceptance probability in Step 5) is expected to be close to 1. Moreover, by taking T sufficiently large, the candidate $\tilde{\boldsymbol{X}}^{(n+1)}$ is expected to be sufficiently decorrelated from the current position $\boldsymbol{X}^{(n)}$.

Consequently, the resulting Markov chain is expected to be efficient.

The remaining challenge for applying the HMC method to our problem of estimating systemic risk allocations is how to handle the constraint $\mathcal{C}$. As we have seen in Sections 5.2.1 and 5.3.1, $\mathcal{C}$ is assumed to be an intersection of linear constraints with parameters $(\boldsymbol{h}_m, v_m)$, $m = 1, \dots, M$, describing hyperplanes. Following the ordinary leapfrog method, a candidate is immediately rejected when the trajectory of the Hamiltonian dynamics penetrates one of these hyperplanes. To avoid it, we modify the leapfrog method according to the reflection technique introduced in Afshar and Domke (2015) and Chevallier et al. (2018). Let $(\boldsymbol{h}, v)$ be the hyperplane which the trajectory of the Hamiltonian dynamics hit at $(\boldsymbol{x}(t), \boldsymbol{p}(t))$. At this time, $(\boldsymbol{x}(t), \boldsymbol{p}(t))$ is immediately replaced by $(\boldsymbol{x}(t), \boldsymbol{p}_r(t))$ where $\boldsymbol{p}_r(t)$ is the *reflected momentum* defined by

$$\boldsymbol{p}_r(t) = \boldsymbol{p}_{\parallel}(t) - \boldsymbol{p}_{\perp}(t),$$

where $\boldsymbol{p}_{\parallel}(t)$ and $\boldsymbol{p}_{\perp}(t)$ are such that $\boldsymbol{p}(t) = \boldsymbol{p}_{\parallel}(t) + \boldsymbol{p}_{\perp}(t)$ and $\boldsymbol{p}_{\parallel}(t)$ and $\boldsymbol{p}_{\perp}(t)$ are parallel and perpendicular to the hyperplane $(\boldsymbol{h}, v)$, respectively. Afshar and Domke (2015) and Chevallier et al. (2018) showed that the map $(\boldsymbol{x}(t), \boldsymbol{p}(t)) \mapsto (\boldsymbol{x}(t), \boldsymbol{p}_r(t))$ preserves the volume and the Hamiltonian, and that this modified HMC method has the stationary distribution π. As long as the initial position $\boldsymbol{x}^{(0)}$ belongs to $\mathcal{C}$, the trajectory of the HMC method never violates the constraint $\mathcal{C}$. The algorithm of this HMC method with reflection is obtained by replacing the **Leapfrog** function call in Step 3) of Algorithm 5 by Algorithm 6 in the end of this chapter. Accordingly, the parameters of the hyperplanes need to be passed as input to Algorithm 5. In Step 3-1) of Algorithm 6 the time t_m at which the trajectory hits the boundary $(\boldsymbol{h}_m, v_m)$ is computed. If $0 < t_m < 1$ for some $m \in \{1, \dots, M\}$, the chain hits the boundary during the dynamics with length ϵ. At the smallest time t_{m^*} among such hitting times, the chain reflects from $(\boldsymbol{x}^*, \boldsymbol{p})$ to $(\boldsymbol{x}_r^*, \boldsymbol{p}_r)$ against the corresponding boundary $(\boldsymbol{h}_{m^*}, v_{m^*})$ as described in Step 3-2-1) of Algorithm 6. The remaining length of the dynamics is $(1 - t_{m^*})\epsilon_{\text{temp}}$ and Step 3) is repeated until the remaining length becomes zero. The trajectory is reflected when it hits a hyperplane and the Markov chain moves within the constrained space with probability one.

Remark 5.3.2 (Roll-back HMC)**.** Yi and Doshi-Velez (2017) proposed *roll-back HMC (RBHMC)*, in which the indicator function $\mathbf{1}_{\{\boldsymbol{x} \in \mathcal{C}\}}$ in the target distribution (5.2) is replaced by a smooth sigmoid function so that the Hamiltonian dynamics naturally move back inwards when the trajectory violates the constraints. HMC with reflection presented in Section 5.3.2 requires to check M boundary conditions at every iteration of the Hamiltonian dynamics. In our problem the number M linearly increases with the dimension d in the case of pure losses, which leads to a linear increase in the computational cost. The RBHMC method avoids such explicit boundary checks, and thus can reduce the computational cost of the HMC method with constrained target distributions. Despite saving computational time, we observed that the RBHMC method requires a careful choice of the stepsize $\epsilon > 0$ and the smoothness parameter of the sigmoid function involved, and we could not find any guidance on how to choose them to guarantee a stable performance.

Choice of parameters for HMC

HMC requires as input two parameters, the *stepsize* ϵ and the *integration time* T. As we now explain, neither of them should be chosen too large nor too small. Since the stepsize ϵ controls the accuracy of

the simulation of the Hamiltonian dynamics, ϵ needs to be small enough to approximately conserve the Hamiltonian; otherwise the acceptance probability can be much smaller than 1. On the other hand, a too small ϵ requires the integration time T to be large for the trajectory to reach far, which is computationally costly. Next, the integration time T needs to be large enough to decorrelate the candidate state with the current state. Meanwhile, the trajectory of the Hamiltonian dynamics may make a U-turn and come back to the starting point if the integration time T is too long; see Neal et al. (2011) for an illustration of this phenomenon.

A notable characteristic of our problem of estimating systemic risk allocations is that the MC sample from the target distribution π is available but its sample size may not be sufficient for statistical inference, and, in the case of the VaR crisis event, the samples only approximately follow the target distribution. We utilize the information of this MC presample to build a heuristic for determining the parameters (ϵ, T); see Algorithm 7 in the end of this chapter. In this heuristic, the initial stepsize is set to be $\epsilon = c_\epsilon d^{-1/4}$ for some constant $c_\epsilon > 0$, say, $c_\epsilon = 1$. This scale was derived in Beskos et al. (2010) and Beskos et al. (2013) under certain assumptions on the target distribution. We determine ϵ through the relationship with the acceptance probability. In Step 2-2-2-1) of Algorithm 7, multiple trajectories are simulated starting from each MC presample with the current stepsize ϵ. In the next Step 2-2-2-2), we monitor the acceptance probability and the distance between the starting and ending points while extending the trajectories. Based on the asymptotic optimal acceptance probability 0.65 (c.f. Gupta et al., 1990; Betancourt et al., 2014) as $d \to \infty$, we set the *target acceptance probability* as

$$\underline{\alpha} = \frac{1 + (d-1) \times 0.65}{d} \in (0.65, 1].$$

The stepsize is gradually decreased in Step 2-1) of Algorithm 7 until the minimum acceptance probability calculated in Step 2-3) exceeds $\underline{\alpha}$. To prevent the trajectory from a U-turn, in Step 2-2-2-3) each trajectory is immediately stopped when the distance begins to decrease. The resulting integration time is set to be the average of these turning points as seen in Step 3). Note that other terminating conditions of extending trajectories are possible; see Hoffman and Gelman (2014) and Betancourt (2016).

At the end of this section, we briefly revisit the choice of the kinetic energy distribution F_K, which is taken to be multivariate standard normal throughout this work. As discussed in Neal et al. (2011), applying the HMC method with target distribution π and kinetic energy distribution $\mathrm{N}(\mathbf{0}, \Sigma^{-1})$ is equivalent to applying HMC with the standardized target distribution $\boldsymbol{x} \to \pi(L\boldsymbol{x})$ and $F_K = \mathrm{N}(\mathbf{0}, I_d)$ where L is the *Cholesky factor* of Σ such that $\Sigma = LL^\top$. By taking Σ to be the covariance matrix of π, the standardized target distribution becomes uncorrelated with unit variances. In our problem, the sample covariance matrix $\hat{\Sigma} = \hat{L}\hat{L}^\top$ calculated based on the MC presample is used alternatively. The new target distribution $\tilde{\pi}(\boldsymbol{y}) = \pi(\hat{L}\boldsymbol{y})|\hat{L}|$ where $|\hat{L}|$ denotes the Jacobian of $\hat{L}$, is almost uncorrelated with unit variances, and thus the standard normal kinetic energy fits well; see Livingstone et al. (2019b). If the crisis event consists of the set of linear constraints $(\boldsymbol{h}_m, v_m)$, $m = 1, \ldots, M$, then the standardized target density is also subject to the set of linear constraints $(\hat{L}^\top \boldsymbol{h}_m, v_m)$, $m = 1, \ldots, M$. Since the ratio $f_{\boldsymbol{X}}(\hat{L}\boldsymbol{y})/f_{\boldsymbol{X}}(\hat{L}\boldsymbol{x})$ can still be evaluated under Assumption (A1), we conclude that the problem remains unchanged after standardization.

Theoretical results of the HMC method with normal kinetic energy are available only when $\mathcal{C}$ is bounded (Cances et al., 2007; Chevallier et al., 2018), or when $\mathcal{C}$ is unbounded and the tail of π is roughly as light as

that of the normal distribution (Livingstone et al., 2019a; Durmus et al., 2017). Boundedness of $\mathcal{C}$ holds for VaR and RVaR crisis events with pure losses; see Section 4.4 in Chapter 4. As is discussed in this chapter, convergence results of MCMC estimators are accessible when the density of the underlying joint loss distribution is bounded from above on $\mathcal{C}$, which is typically the case when the underlying copula does not admit lower tail dependence. For other cases where $\mathcal{C}$ is unbounded or the density explodes on $\mathcal{C}$, no convergence results are available. Potential remedies for the HMC method to deal with heavy-tailed target distributions are discussed in Remark 5.3.3.

Remark 5.3.3 (Riemannian manifold HMC)**.** Livingstone et al. (2019b) indicated that non-normal kinetic energy distributions can potentially deal with heavy-tailed target distributions. In fact, the kinetic energy distribution F_K can even be dependent on the position variable $\boldsymbol{x}$. For example, when $F_K(\cdot|\boldsymbol{x}) = \mathrm{N}(\boldsymbol{0}, G(\boldsymbol{x}))$ for a positive definite matrix $G(\boldsymbol{x}) > 0$ and $\boldsymbol{x} \in E$, the resulting HMC method is known as *Riemannian manifold HMC (RMHMC)* since this case is equivalent to applying HMC on the Riemannian manifold with metric $G(\boldsymbol{x})$; see Girolami and Calderhead (2011). Difficulties in implementing RMHMC are in the choice of metric G and in the simulation of the Hamiltonian dynamics. Due to the complexity of the Hamiltonian dynamics, simple discretization schemes such as the leapfrog method are not applicable, and the trajectory is updated implicitly by solving some system of equations; see Girolami and Calderhead (2011). Various choices of the metric G are studied in Betancourt (2013), Lan et al. (2014) and Livingstone and Girolami (2014) for different purposes. Simulation of RMHMC is studied, for example, in Byrne and Girolami (2013).

5.3.3 Estimation with Gibbs sampler

As discussed in Section 5.3.2, applying HMC methods to heavy-tailed target distributions on unbounded crisis events is not theoretically supported. To deal with this case, we introduce the GS in this section.

Gibbs samplers for estimating systemic risk allocations

The GS is a special case of the MH method in which the proposal density q is completely determined by the target density π via

$$q_{GS}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{i}=(i_1,\dots,i_d)\in\mathcal{I}_d} p_{\boldsymbol{i}} \pi(y_{i_1}|\boldsymbol{x}_{-i_1})\pi(y_{i_2}|y_{i_1}, \boldsymbol{x}_{-(i_1,i_2)}) \cdots \pi(y_{i_d}|\boldsymbol{y}_{-i_d}), \tag{5.5}$$

where $\boldsymbol{x}_{-(j_1,\dots,j_l)}$ is the $(d-l)$-dimensional vector that excludes the components $j_1, \dots, j_l$ from $\boldsymbol{x}$, $\pi(x_j|\boldsymbol{x}_{-j}) = \pi_{j|-j}(x_j|\boldsymbol{x}_{-j})$ is the conditional density of the jth variable of π given all the other components, $\mathcal{I}_d \subseteq \{1, \dots, d\}^d$ is the so-called *index set* and $(p_{\boldsymbol{i}} \in [0,1],\ \boldsymbol{i} \in \mathcal{I}_d)$ is the *index probability distribution* such that $\sum_{\boldsymbol{i}\in\mathcal{I}_d} p_{\boldsymbol{i}} = 1$. For this choice of q, the acceptance probability is always equal to 1; see Johnson (2009). The GS is called *deterministic scan (DSGS)* if $\mathcal{I}_d = \{(1, \dots, d)\}$ and $p_{(1,\dots,d)} = 1$. When the index set is the set of permutations of $(1, \dots, d)$, the GS is called *random permulation (RPGS)*. Finally, the *random scan GS (RSGS)* has the proposal (5.5) with $\mathcal{I}_d = \{1, \dots, d\}^d$ and $p_{(i_1,\dots,i_d)} = p_{i_1} \cdots p_{i_d}$ with probabilities $(p_1, \dots, p_d) \in (0,1)^d$ such that $\sum_{j=1}^d p_j = 1$. These three GSs can be shown to have π as stationary distribution; see Johnson (2009).

Provided that the *full conditional distributions* $\pi_{j|-j}$, $j = 1, \dots, d$ can be simulated, the proposal distribution (5.5) can be simulated by first selecting an index $\boldsymbol{i} \in \mathcal{I}_d$ with probability $p_{\boldsymbol{i}}$ and then replacing the jth component of the current state with a sample from $\pi_{j|-j}$ sequentially for $j = i_1, \dots, i_d$. The main advantage of the GS is that the tails of π are naturally incorporated via full conditional distributions, and thus the MCMC method is expected to be efficient even if π is heavy-tailed. On the other hand, the applicability of the GS is limited to target distributions such that $\pi_{j|-j}$ is available. Moreover, fast simulation methods of $\pi_{j|-j}$, $j = 1, \dots, d$, are required since the computational time linearly increases w.r.t. the dimension d.

In our problem of estimating systemic risk allocations, we find that the GS is applicable when the crisis event is of the form

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^d \text{ or } \mathbb{R}^d_+ \mid v_1 \le \boldsymbol{h}^\top \boldsymbol{x} \le v_2\}, \quad v_1, v_2 \in \mathbb{R} \cup \{\pm\infty\}, \quad \boldsymbol{h} = (h_1, \dots, h_d) \in \mathbb{R}^d \backslash \{\boldsymbol{0}_d\}. \tag{5.6}$$

The RVaR crisis event is obviously a special case of (5.6), and the ES crisis event is included as a limiting case for $v_2 \to \infty$. Furthermore, the full conditional copulas of the underlying joint loss distribution and their inverses are required to be known as we now explain. Consider the target density $\pi = f_{\boldsymbol{X}|\{v_1 \le \boldsymbol{h}^\top \boldsymbol{X} \le v_2\}}$. For its jth full conditional density $\pi_{j|-j}(x_j|\boldsymbol{x}_{-j})$, notice that

$$\{v_1 \le \boldsymbol{h}^\top \boldsymbol{X} \le v_2,\ \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\} = \left\{ \frac{v_1 - \boldsymbol{h}_{-j}^\top \boldsymbol{x}_{-j}}{h_j} \le X_j \le \frac{v_2 - \boldsymbol{h}_{-j}^\top \boldsymbol{x}_{-j}}{h_j},\ \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j} \right\}$$

and thus, for $v_{i,j}(\boldsymbol{x}_{-j}) = (v_i - \boldsymbol{h}_{-j}^\top \boldsymbol{x}_{-j})/h_j$, $i = 1, 2$, we obtain the cdf of $\pi_{j|-j}$ as

$$F_{X_j|\{v_1 \le \boldsymbol{h}^\top \boldsymbol{X} \le v_2,\ \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(x_j) = \frac{F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(x_j) - F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(v_{1,j}(\boldsymbol{x}_{-j}))}{F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(v_{2,j}(\boldsymbol{x}_{-j})) - F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(v_{1,j}(\boldsymbol{x}_{-j}))} \tag{5.7}$$

for $v_{1,j}(\boldsymbol{x}_{-j}) \le x_j \le v_{2,j}(\boldsymbol{x}_{-j})$. Denoting the denominator of (5.7) by $\Delta_j(\boldsymbol{x}_{-j})$, we obtain the quantile function

$$F^{-1}_{X_j|\{v_1 \le \boldsymbol{h}^\top \boldsymbol{X} \le v_2,\ \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(u) = F^{-1}_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}} \left(\Delta_j(\boldsymbol{x}_{-j}) \cdot u + F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(v_{1,j}(\boldsymbol{x}_{-j})) \right).$$

Therefore, if $F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}$ and its quantile function are available, one can simulate the full conditional target densities $\pi_{j|-j}$ with the inversion method; see Devroye (1985). Availability of $F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}$ and its inverse typically depends on the copula of $\boldsymbol{X}$. By Sklar's Theorem (1.1), the jth full conditional distribution of $F_{\boldsymbol{X}}$ can be written as

$$F_{X_j|\{\boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}\}}(x_j) = C_{j|-j}(F_j(x_j) \mid \boldsymbol{F}_{-j}(\boldsymbol{x}_{-j})),$$

where $\boldsymbol{F}_{(j_1,\dots,j_l)}(\boldsymbol{x}_{(j_1,\dots,j_l)}) = (F_{j_1}(x_{j_1}), \dots, F_{j_l}(x_{j_l}))$, $-(j_1, \dots, j_l) = \{1, \dots, d\} \backslash (j_1, \dots, j_l)$ and $C_{j|-j}$ is the jth *full conditional copula* defined by

$$C_{j|-j}(u_j|\boldsymbol{u}_{-j}) = \mathbb{P}(U_j \le u_j \mid \boldsymbol{U}_{-j} = \boldsymbol{u}_{-j}) = \frac{D_{-j} C(\boldsymbol{u})}{D_{-j} C(u_1, \dots, u_{j-1}, 1, u_{j+1}, \dots, u_d)},$$

where D denotes the operator of partial derivatives with respect to the components given as subscripts and $\boldsymbol{U} \sim C$. Assuming the full conditional copula $C_{j|-j}$ and its inverse $C^{-1}_{j|-j}$ are available, one can simulate

$\tilde{X}_j \sim \pi_{j|-j}$ via

$$\begin{aligned} U &\sim \mathrm{U}(0,1), \\ \tilde{U} &= U + (1-U)C_{j|-j}(F_j(v_1(\boldsymbol{x}_{-j}) \mid \boldsymbol{F}_{-j}(\boldsymbol{x}_{-j})), \\ \tilde{X}_j &= F_j^{-1} \circ C_{j|-j}^{-1}(\tilde{U} \mid \boldsymbol{F}_{-j}(\boldsymbol{x}_{-j})). \end{aligned}$$

Examples of copulas for which the full conditional distributions and their inverses are available include normal, Student t, and Clayton copulas; see Cambou et al. (2017). In this case the GS is also applicable to the corresponding survival (π-rotated) copula $\hat{C}$ since

$$\hat{C}_{j|-j}(\boldsymbol{u}) = 1 - C_{j|-j}(1-u_j \,|\, \mathbf{1}_{d'} - \boldsymbol{u}_{-j}), \quad \hat{C}^{-1}_{j|-j}(\boldsymbol{u}) = 1 - C^{-1}_{j|-j}(1-u_j \,|\, \mathbf{1}_{d'} - \boldsymbol{u}_{-j}), \quad j = 1, \dots, d,$$

by the relationship $\tilde{\boldsymbol{U}} = \mathbf{1} - \boldsymbol{U} \sim \hat{C}$ for $\boldsymbol{U} \sim C$. In a similar way, one can also obtain full conditional copulas and their inverses for other rotated copulas; see Hofert et al. (2018, Section 3.4.1) for rotated copulas.

In the end, we remark that even if the full conditional distributions and their inverses are not available, $\pi_{j|-j}$ can be simulated by, for example, the acceptance-rejection method or even the MH algorithm; see Remark 5.3.4.

Remark 5.3.4 (Metropolized Gibbs samplers)**.** Müller (1992) introduced the *Metropolized Gibbs sampler (MGS)* in which the proposal density q in the MH kernel is set to be $q = f_{\boldsymbol{Y}|\{v_1 \le \boldsymbol{h}^\top \boldsymbol{Y} \le v_2\}}$ where $\boldsymbol{Y}$ has the same marginal distributions as $\boldsymbol{X}$ but a different copula C^q for which $C^q_{j|-j}$ and $C^{q,-1}_{j|-j}$ are available so that the GS can be applied to simulate this proposal. This method can be used when the inversion method is not feasible since $C_{j|-j}$ or $C^{-1}_{j|-j}$ are not available. Following the MH algorithm, the candidate is accepted with the acceptance probability (1.11), which can be simply written as

$$\alpha(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \min\left\{ \frac{c(\boldsymbol{F}(\tilde{\boldsymbol{x}}))c^q(\boldsymbol{F}(\boldsymbol{x}))}{c(\boldsymbol{F}(\boldsymbol{x}))c^q(\boldsymbol{F}(\tilde{\boldsymbol{x}}))}, 1 \right\},$$

where c and c^q denote the densities of C and C^q, respectively.

As an example of the MGS, suppose C is the Gumbel copula, for which the full conditional distributions cannot be inverted analytically. One could then choose the survival Clayton copula as the proposal copula C^q above. For this choice of copula, $q_{j|-j}$ is available by the inversion method as discussed in Section 5.3.3. Furthermore, the acceptance probability is expected to be high especially on the upper tail part because the upper threshold copula of C defined as $\mathbb{P}(\boldsymbol{U} > \boldsymbol{v} \mid \boldsymbol{U} > \boldsymbol{u})$, $\boldsymbol{v} \in [\boldsymbol{u}, \mathbf{1}]$, $\boldsymbol{u} \in [0,1]^d$, $\boldsymbol{U} \sim C$ is known to converge to that of a survival Clayton copula when $\lim u_j \to \infty$, $j = 1, \dots, d$; see Juri and Wüthrich (2002), Juri and Wüthrich (2003), Charpentier and Segers (2007) and Larsson and Nešlehová (2011).

Choice of parameters for GS

As discussed in Section 5.3.2, we use information from the MC presamples to determine the parameters of the Gibbs kernel (5.5). Note that standardization of the variables as applied in the HMC method in Section 5.3.2 is not available for the GS since the latter changes the underlying joint losss distribution,

and since the copula after rotating variables is in general not accessible except for in the elliptical case; see Christen et al. (2017). Among the presented variants of GSs, we adopt RSGS since determining d probabilities $(p_1, \dots, p_d)$ is relatively easy whereas RPGS requires $d!$ probabilities to be determined. To this end, we consider the RSGS with the parameters $(p_1, \dots, p_d)$ determined by a heuristic described in Algorithm 8.

The RSGS kernel is simulated in Step 3) and 5) of Algorithm 8 in the end of this chapter. To determine the selection probabilities $p_1, \dots, p_d$, consider a one step update of the RSGS from $\boldsymbol{X}^{(n)}$ to $\boldsymbol{X}^{(n+1)}$ with $\boldsymbol{X}^{(n)} \sim \pi$ and the one step kernel

$$K_{\mathrm{RSGS}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{d} p_j \pi_{j|-j}(y_j | \boldsymbol{x}_{-j}) \mathbf{1}_{\{\boldsymbol{y}_{-j} = \boldsymbol{x}_{-j}\}}.$$

Liu et al. (1995, Lemma 3) implies that

$$\begin{aligned} \mathrm{Cov}(X_j^{(n)}, X_j^{(n+1)}) &= \sum_{i=1}^{d} p_i \mathbb{E}[\mathbb{E}[X_j \mid \boldsymbol{X}_{-i}]] = \sum_{i=1}^{d} p_i \{m_j^{(2)} - \mathbb{E}[\mathrm{Var}(X_j \mid \boldsymbol{X}_{-i})]\} \\ &\propto -\sum_{i=1}^{d} p_i \mathbb{E}[\mathrm{Var}(X_j \mid \boldsymbol{X}_{-i})]), \end{aligned}$$

where $m_j^{(k)}$ is the kth moment of π_j.

For the objective function $\sum_{j=1}^{d} \mathrm{Cov}(X_j^{(n)}, X_j^{(n+1)})$, its minimizer $(p_1^*, \dots, p_d^*)$ under the constraint $\sum_{j=1}^{d} p_j = 1$ satisfies

$$p_j^* \propto \mathbb{E}[\mathrm{Var}(X_j \,|\, \boldsymbol{X}_{-j})]. \tag{5.8}$$

While this optimizer can be computed based on the MC presamples, we observed that its stable estimation is as computationally demanding as estimating the risk allocations themselves. Alternatively, we calculate (5.8) under the assumption that π follows an elliptical distribution. Under this assumption, (5.8) is given by

$$p_j \propto \Sigma_{j,j} - \Sigma_{j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j}$$

where Σ is the covariance matrix of π and Σ_{J_1,J_2}, $J_1, J_2 \subseteq \{1, \dots, d\}$, is the submatrix of Σ with indices in $J_1 \times J_2$. As seen in Step 2) of Algorithm 8, Σ is replaced by its estimate based on the MC presamples.

As is shown in Christen et al. (2017), Gibbs samplers require a large number of iterations to lower the serial correlation when the target distribution has strong dependence. To reduce serial correlations we take every Tth sample in Step 5-2), where $T \in \mathbb{N}$ is called the *thinning interval of times*. Note that we use the same notation T as that of the integration time in HMC since they both represent a repetition time of some single step. Based on the preliminary run with length N_{pre} in Step 3) in Algorithm 8, T is determined as the smallest lag h such that the marginal autocorrelations with lag h are all smaller than the target autocorrelation ρ; see Step 4) in Algorithm 8.

5.4 Numerical experiments

In this section, we demonstrate the performance of the MCMC methods for estimating systemic risk allocations by a series of numerical experiments. We first conduct a simulation study in which true allocations or their partial information are available. Then we perform an empirical study to demonstrate that our MCMC methods are applicable to a more practical setup. Finally, we make more detailed comparisons between the MC and MCMC methods in various setups.

5.4.1 Simulation study

In this simulation study, we compare the estimates and standard errors of the MC and MCMC methods under the low-dimensional risk models described in Section 5.4.1. The results and discussions are summarized in Section 5.4.1.

Model description

We consider the following three-dimensional loss distributions:

(M1) *generalized Pareto distributions (GPDs)* with parameters $(\xi_j, \beta_j) = (0.3, 1)$ and survival Clayton copula with parameter $\theta = 2$ so that Kendall's tau equals $\tau = \theta/(\theta + 2) = 0.5$;

(M2) multivariate Student t distribution with $\nu = 5$ degrees of freedom, location vector $\mathbf{0}$ and dispersion matrix $\Sigma = (\rho_{i,j})$ where $\rho_{j,j} = 1$ and $\rho_{i,j} = |i - j|/d$ for $i, j = 1, \dots, d$, $i \neq j$.

Since the marginals are homogeneous and the copula is exchangeable, the systemic risk allocations under the loss distribution (M1) are all equal provided that the crisis event is invariant under the permutation of the variables. For the loss distribution (M2), by ellipticality of the joint distribution, analytical formulas of risk contribution type systemic risk allocations are available; see McNeil et al. (2015, Corollary 8.43). The parameters of the distributions (M1) and (M2) take into account the stylized facts that the loss distribution is heavy-tailed and extreme losses are positively dependent.

We consider the VaR, RVaR and ES crisis events with confidence levels $p^{\mathrm{VaR}} = 0.99$, $(p_1^{\mathrm{RVaR}}, p_2^{\mathrm{RVaR}}) = (0.975, 0.99)$ and $p^{\mathrm{ES}} = 0.99$, respectively. For each crisis event, the risk contribution, VaR, RVaR and ES type systemic risk allocations are estimated by the MC and MCMC methods, where the parameters of the marginal risk measures VaR, RVaR and ES are set to be $p^{\mathrm{VaR}} = 0.99$, $(p_1^{\mathrm{RVaR}}, p_2^{\mathrm{RVaR}}) = (0.975, 0.99)$ and $p^{\mathrm{ES}} = 0.99$, respectively.

We first conduct the MC simulation for the distributions (M1) and (M2). For the VaR crisis event, the modified event $\mathcal{C}^{\mathrm{mod}} = \{\mathrm{VaR}_{p-\delta}(S) \leq \mathbf{1}_d^\top \boldsymbol{x} \leq \mathrm{VaR}_{p+\delta}(S)\}$ with $\delta = 0.001$ is used to ensure that $\mathbb{P}(\boldsymbol{X} \in \mathcal{C}^{\mathrm{mod}}) > 0$. Based on these MC presamples, the Markov chains are constructed as described in Sections 5.3.2 and 5.3.3. For the MCMC method, (M1) is the case of pure losses and (M2) is the P&L case. Therefore, the HMC method is applied to the distribution (M1) for the VaR and RVaR crisis events, the

GS is applied to (M1) for the ES crisis event and the GS is applied to the distribution (M2) for the RVaR and ES crisis events. The target distribution of (M2) with VaR constraint is free from constraints and was already investigated in Chapter 4; we thus omit this case and consider the five remaining cases.

Note that 99.8% of the MC samples from the unconditional distribution are discarded because of the VaR crisis event and a further 97.5% of them are wasted to estimate the RVaR contributions. Therefore, $1/(0.002 \times 0.025) = 10^5/5 = 20,000$ MC samples are required to obtain one expected MC sample from the conditional distribution. Taking this into account, the sample size of the MC estimator is set to be $N_{\mathrm{MC}} = 10^5$. The sample size of the MCMC estimators is free from such constraints and thus is chosen to be $N_{\mathrm{MCMC}} = 10^4$. Initial values $\boldsymbol{x}_0$ for the MCMC methods are taken as the mean vector calculated from the MC samples. Biases are computed only for the contribution type allocations in the distribution (M2) since the true values are available in this case. For all the five cases, the MC and the MCMC standard errors are computed according to Glasserman (2013, Chapter 1) for MC, and Jones et al. (2006) for MCMC. Asymptotic variances of the MCMC estimators are estimated by the batch means estimator with batch length $L_N = \lceil N^{\frac{1}{2}} \rceil = 100$ and batch size $B_N = \lceil N/L_N \rceil = 100$. The results are summarized in Tables 5.1 and 5.2.

Results and discussion

Since fast random number generators are available for the joint loss distributions (M1) and (M2), the MC estimators are computed almost instantly. On the other hand, the MCMC methods cost around 1.5 minutes for simulating the $N = 10^4$ MCMC samples as reported in Tables 5.1 and 5.2. For the HMC method, the main computational cost consists of calculating gradients $N \times T$ times for the leapfrog method, and calculating the ratio of target densities N times in the acceptance/rejection step, where N is the length of the sample path and T is the integration time. For the GS, simulating an N-sample path requires $N \times T \times d$ random numbers from the full conditional distributions where T here is the thinning interval of times. Therefore, the computational time of the GS linearly increases w.r.t. the dimension d, which can become prohibitive for the GS in high dimensions. To save computational time, MCMC methods in general require careful implementations of calculating the gradients and the ratio of the target densities for HMC, and of simulating the full conditional distributions for the GS.

Next, we inspect the performance of the HMC and GS methods. We observed that autocorrelations of all sample paths steadily decreased below 0.1 if lags are larger than 15. Together with the high ACRs, we conclude that the Markov chains can be considered to be converged. According to the heuristic in Algorithm 7, the stepsize and the integration time for the HMC method are selected to be $(\epsilon, T) = (0.210, 12)$ in Case (I) and $(\epsilon, T) = (0.095, 13)$ in Case (II). As indicated by the small Hamiltonian errors in Figure 5.1, the acceptance rates in both cases are quite close to 1.

For the GS, the thinning interval of times T and the selection probability $\boldsymbol{p}$ are determined as $T = 12$ and $\boldsymbol{p} = (0.221, 0.362, 0.416)$ in Case (III), $T = 10$ and $\boldsymbol{p} = (0.330, 0.348, 0.321)$ in Case (IV) and $T = 4$ and $\boldsymbol{p} = (0.241, 0.503, 0.255)$ in Case (V). For biases of the estimators, observe that in all cases ((I) to (V)), the estimates of the MC method and the MCMC method are close to each other. In Cases (I), (II) and (III), the true allocations are the homogeneous allocations whereas their exact values are not known.

	MC			HMC		
Estimator	$A_1^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_2^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_3^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_1^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_2^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_3^{\varrho,\mathcal{C}}(\boldsymbol{X})$
(I) GPD + survival Clayton with VaR crisis event: $\{S = \mathrm{VaR}_{0.99}(S)\}$						
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}}]$	9.581	9.400	9.829	9.593	9.599	9.619
Standard error	0.126	0.118	0.120	0.007	0.009	0.009
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	12.986	12.919	13.630	13.298	13.204	13.338
Standard error	0.229	0.131	0.086	0.061	0.049	0.060
$\mathrm{VaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	13.592	13.235	13.796	13.742	13.565	13.768
Standard error	0.647	0.333	0.270	0.088	0.070	0.070
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	14.775	13.955	14.568	14.461	14.227	14.427
Standard error	0.660	0.498	0.605	0.192	0.176	0.172
(II) GPD + Survival Clayton with RVaR crisis event: $\{\mathrm{VaR}_{0.975}(S) \leq S \leq \mathrm{VaR}_{0.99}(S)\}$						
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}}]$	7.873	7.780	7.816	7.812	7.802	7.780
Standard error	0.046	0.046	0.046	0.012	0.012	0.011
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	11.790	11.908	11.680	11.686	11.696	11.646
Standard error	0.047	0.057	0.043	0.053	0.055	0.058
$\mathrm{RVaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	12.207	12.382	12.087	12.102	12.053	12.044
Standard error	0.183	0.197	0.182	0.074	0.069	0.069
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	13.079	13.102	13.059	12.859	12.791	12.713
Standard error	0.182	0.173	0.188	0.231	0.218	0.187

Table 5.1: Estimates and standard errors of the MC and HMC estimators of risk contributions, RVaR, VaR and ES type systemic risk allocations under (I) the VaR crisis event and (II) the RVaR crisis event for the loss distribution (M1). The sample size of the MC method is $N_{\mathrm{MC}} = 10^5$ and that of the HMC method is $N_{\mathrm{MCMC}} = 10^4$. The acceptance rate (ACR), stepsize ϵ, integration time T and run time are ACR $= 0.996$, $\epsilon = 0.210$, $T = 12$ and run time $= 1.277$ mins in Case (I), and ACR $= 0.984$, $\epsilon = 0.095$, $T = 13$ and run time $= 1.649$ mins in Case (II).

Estimator	MC			GS		
	$A_1^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_2^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_3^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_1^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_2^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_3^{\varrho,\mathcal{C}}(\boldsymbol{X})$
(III) GPD + survival Clayton with ES crisis event: $\{\mathrm{VaR}_{0.99}(S) \leq S\}$						
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}}]$	15.657	15.806	15.721	15.209	15.175	15.190
Standard error	0.434	0.475	0.395	0.257	0.258	0.261
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	41.626	41.026	45.939	45.506	45.008	45.253
Standard error	1.211	1.065	1.615	1.031	1.133	1.256
$\mathrm{VaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	49.689	48.818	57.488	55.033	54.746	54.783
Standard error	4.901	4.388	4.973	8.079	5.630	3.803
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	104.761	109.835	97.944	71.874	72.588	70.420
Standard error	23.005	27.895	17.908	4.832	4.584	4.313
(IV) Multivariate t with RVaR crisis event: $\{\mathrm{VaR}_{0.975}(S) \leq S \leq \mathrm{VaR}_{0.99}(S)\}$						
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}}]$	2.456	1.934	2.476	2.394	2.060	2.435
Bias	0.019	−0.097	0.038	−0.043	0.029	−0.002
Standard error	0.026	0.036	0.027	0.014	0.023	0.019
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	4.670	4.998	4.893	4.602	5.188	4.748
Standard error	0.037	0.042	0.031	0.032	0.070	0.048
$\mathrm{RVaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	5.217	5.397	5.240	4.878	5.717	5.092
Standard error	0.238	0.157	0.145	0.049	0.174	0.100
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	5.929	5.977	5.946	5.446	6.517	6.063
Standard error	0.204	0.179	0.199	0.156	0.248	0.344
(V) Multivariate t with ES crisis event: $\{S \geq \mathrm{VaR}_{0.99}(S)\}$						
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}}]$	3.758	3.099	3.770	3.735	3.126	3.738
Bias	0.017	−0.018	0.029	−0.005	0.009	−0.003
Standard error	0.055	0.072	0.060	0.031	0.027	0.030
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	8.516	8.489	9.051	8.586	8.317	8.739
Standard error	0.089	0.167	0.161	0.144	0.156	0.158
$\mathrm{VaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	9.256	9.754	10.327	9.454	9.517	9.890
Standard error	0.517	0.680	0.698	0.248	0.293	0.327
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	11.129	12.520	12.946	11.857	12.469	12.375
Standard error	0.595	1.321	0.826	0.785	0.948	0.835

Table 5.2: Estimates and standard errors of the MC and the GS estimators of risk contributions, VaR, RVaR and ES type systemic risk allocations under (III) distribution (M1) and the ES crisis event, (IV) distribution (M2) and the RVaR crisis event, and (V) distribution (M2) and ES crisis event. The sample size of the MC method is $N_{\mathrm{MC}} = 10^5$ and that of the GS is $N_{\mathrm{MCMC}} = 10^4$. The thinning interval of times T, selection probability $\boldsymbol{p}$ and run time are $T = 12$, $\boldsymbol{p} = (0.221, 0.362, 0.416)$ and run time = 107.880 secs in Case (III), $T = 10$, $\boldsymbol{p} = (0.330, 0.348, 0.321)$ and run time = 56.982 secs in Case (IV) and $T = 4$, $\boldsymbol{p} = (0.241, 0.503, 0.255)$ and run time = 22.408 secs in Case (V).

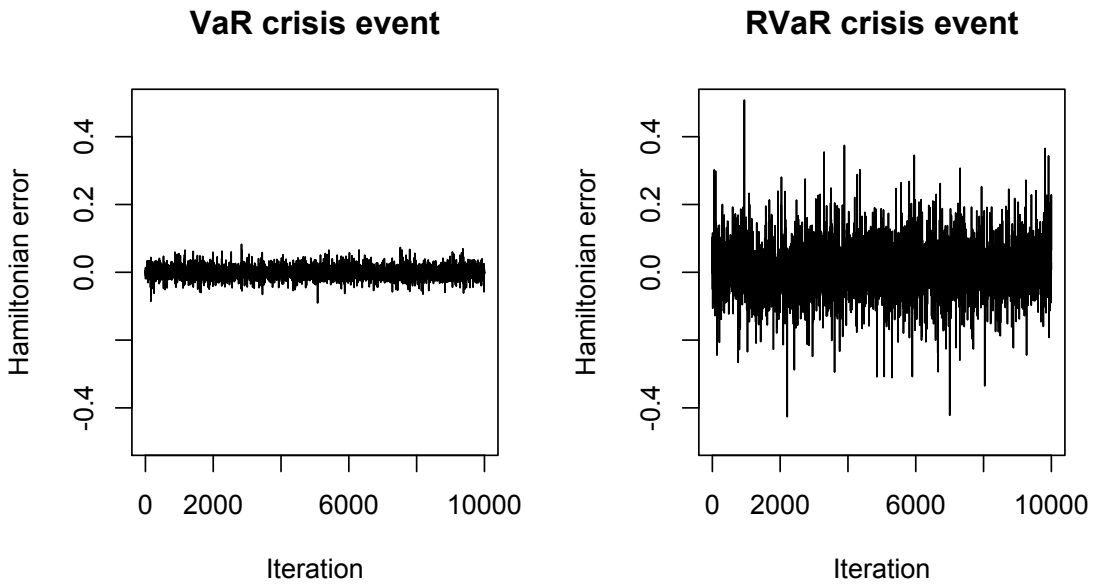

Figure 5.1: Hamiltonian errors of the HMC methods for estimating systemic risk allocations with VaR (left) and RVaR (right) crisis events for the loss distribution (M1). The stepsize and the integration time are set to be $(\epsilon, T) = (0.210, 12)$ in Case (I) and $(\epsilon, T) = (0.095, 13)$ in Case (II).

From the estimates in Tables 5.1 and 5.2, the MCMC estimates are on average more equally allocated compared to those of the MC method especially in Case (III) where heavy-tailedness may lead to quite slow convergence rates of the MC method. Therefore, lower biases of the MCMC estimators are obtained compared to those of the MC estimators. In the case of risk contributions in Case (IV) and (V), exact biases are computed based on ellipticality, and they show that the GS estimator has a smaller bias than the one of the MC estimator.

Although the MC sample size is 10 times larger than that of the MCMC method, the standard error of the latter is in most cases smaller than the MC standard error. This improvement becomes more pronounced as the probability of the crisis event becomes smaller. The largest improvement is observed in Case (I) with VaR crisis event and the smallest one is in Cases (III) and (V) with ES crisis event. MCMC estimates of the risk contribution type allocations have consistently smaller standard errors than the MC ones. For the RVaR, VaR and ES type allocations, the improvement of standard error varies according to the loss models and the crisis event. A notable improvement is observed for ES type allocation in Case (III) although a stable statistical inference is challenging due to the heavy-tailedness of the target distribution.

Overall, the simulation study shows that the MCMC estimators outperform the MC estimators due to the increased effective sample size and their insusceptibility to the probability of the crisis event. The MCMC estimators are especially recommended when the probability of the crisis event is too small for the MC method to simulate sufficiently many samples for a meaningful statistical analysis.

Remark 5.4.1 (Joint loss distributions with negative dependence in the tail)**.** In the above simulation study, we only considered joint loss distributions with positive dependence. Under the existence of positive dependence, the target density $f_{\boldsymbol{X}|\{v_p \leq S \leq v_p\}}$ puts more probability mass around its mean, and the probability decays as the evaluation point moves away from the mean since positive dependence among $X_1, \ldots, X_d$ prevents the components from going in opposite directions (i.e., one component increases and another one decreases) under the sum constraint. This phenomenon leads to the target distributions being more centered and elliptical, which in turn facilitates efficient moves of Markov chains. Although it may not be realistic, joint loss distributions with negative dependence in the tail are also possible. In this case, the target distribution has more variance, heavy tails and is even multimodal since two components can move in opposite directions under the sum constraint. For such cases, constructing efficient MCMC methods becomes more challenging; see Lan et al. (2014) for a remedy for multimodal target distributions with Riemannian manifold HMC.

5.4.2 Empirical Study

In this section, we illustrate our suggested MCMC methods for estimating risk allocations from insurance company indemnity claims. The dataset consists of 1500 liability claims provided by Insurance Services Office. Each claim contains an indemnity payment X_1 and an allocated loss adjustment expense (ALAE) X_2; see Hogg and Klugman (2009) for a description. The joint distribution of losses and expenses is studied, for example in Frees and Valdez (1998) and Klugman and Parsa (1999). Based on Frees and Valdez (1998), we adopt the following parametric model:

(M3) univariate marginals are $X_1 \sim \mathrm{Par}(\lambda_1, \theta_1)$ and $X_2 \sim \mathrm{Par}(\lambda_2, \theta_2)$ with $(\lambda_1, \theta_1) = (14,036, 1.122)$ and $(\lambda_2, \theta_2) = (14,219, 2.118)$, and the copula is the survival Clayton copula with parameter $\theta = 0.512$ (which corresponds to Spearman's rho $\rho_S = 0.310$).

Note that in the loss distribution (M3) the Gumbel copula used in Frees and Valdez (1998) is replaced by the survival Clayton copula since both of them have the same type of tail dependence and the latter possesses more computationally tractable derivatives. The parameter of the survival Clayton copula is determined so that it reaches the same Spearman's rho observed in Frees and Valdez (1998). Figure 5.2 illustrates the data and samples from the distribution (M3). Our goal is to calculate the VaR, RVaR and ES type allocations with VaR, RVaR and ES crisis events for the same confidence levels as in Section 5.4.1. We apply the HMC method to all three crisis events since, due to the infinite and finite variances of X_1 and X_2, respectively, the optimal selection probability of the second variable calculated in Step 2) of Algorithm 8 is quite close to 0, and thus the GS did not perform well. The simulated HMC samples are illustrated in Figure 5.2. The results of estimating the systemic risk allocations are summarized in Table 5.3.

The HMC samples shown in Figure 5.2 indicate that the conditional distributions of interest are successfully simulated from the desired regions. As displayed in Figure 5.3, the Hamiltonian errors of all three HMC methods are sufficiently small, which leads to the high ACRs of 0.997, 0.986 and 0.995 as listed in Table 5.3. We also observed that autocorrelations of all sample paths steadily decreased below 0.1 if lags are larger than 80. Together with the high ACRs, we conclude that the Markov chains can be considered

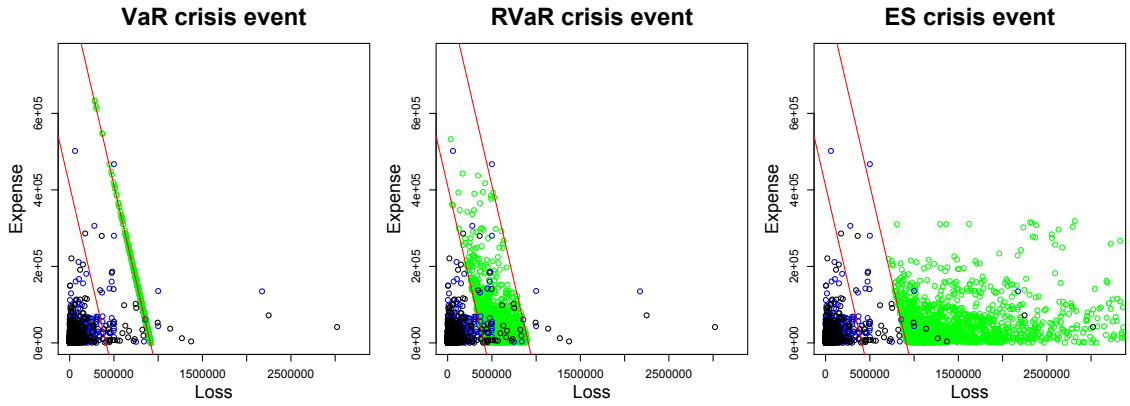

Figure 5.2: Plots of $N = 1500$ MCMC samples (green) with VaR (left), RVaR (center) and ES (right) crisis events. All plots include the data and the MC samples with sample size $N = 1500$ in black and blue dots, respectively. The red lines represent $x_1 + x_2 = \widehat{\mathrm{VaR}}_{p_1}(S)$ and $x_1 + x_2 = \widehat{\mathrm{VaR}}_{p_2}(S)$ where $\widehat{\mathrm{VaR}}_{p_1}(S) = 4.102 \times 10^4$ and $\widehat{\mathrm{VaR}}_{p_2}(S) = 9.117 \times 10^4$ are the MC estimates of $\mathrm{VaR}_{p_1}(S)$ and $\mathrm{VaR}_{p_2}(S)$, respectively, for $p_1 = 0.975$ and $p_2 = 0.99$.

to be converged. Due to the heavy-tailedness of the target distribution in the case of the ES crisis event, the stepsize is very small and the integration time is very large compared to the former two cases of the VaR and RVaR crisis events. As a result, the HMC algorithm in this case has a long run time.

The estimates of the MC and HMC methods are close in all cases except Case (III). In Case (III), the HMC estimates are smaller than the MC ones in almost all cases. Based on the much smaller standard errors of HMC, one could infer that the MC estimates are likely overestimating the allocations due to a small number of extremely large losses, although the corresponding conditional distribution is extremely heavy-tailed and thus no estimation method might be reliable. In terms of the standard error, the estimation of systemic risk allocations by the HMC method is improved in Cases (I) and (III) compared to those of the MC method; the MC standard errors are slightly smaller than those of HMC in Case (II). All results considered, we conclude from this empirical study that the MCMC estimators outperform the MC estimators in terms of standard error. On the other hand, as indicated by the theory of HMC with normal kinetic energy, the HMC method is not recommended for heavy-tailed target distributions due to a long computational time caused by a small stepsize and large integration time determined by Algorithm 8.

5.4.3 Detailed comparison of MCMC with MC

In the previous numerical experiments, we fixed the dimensions of the portfolios and confidence levels of the crisis events. Comparing the MC and MCMC methods after balancing against computational time might be more reasonable although one should keep in mind that run time depends on various external factors, such as the implementation, hardware, workload, programming language or compiler options (and our implementation was not optimized for any of these factors). In this section, we compare the MC and

	MC		HMC	
Estimator	$A_1^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_2^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_1^{\varrho,\mathcal{C}}(\boldsymbol{X})$	$A_2^{\varrho,\mathcal{C}}(\boldsymbol{X})$
(I) VaR crisis event: $\{S = \mathrm{VaR}_{0.99}(S)\}$				
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}}]$	842 465.497	73 553.738	844 819.901	71 199.334
Standard error	7994.573	7254.567	6306.836	6306.836
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	989 245.360	443 181.466	915 098.833	428 249.307
Standard error	307.858	24 105.163	72.568	20 482.914
$\mathrm{VaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	989 765.514	500 663.072	915 534.362	615 801.118
Standard error	4670.966	54 576.957	669.853	96 600.963
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{VaR}})$	990 839.359	590 093.887	915 767.076	761 038.843
Standard error	679.055	75 024.692	47.744	31 211.908
(II) RVaR crisis event: $\{\mathrm{VaR}_{0.975}(S) \leq S \leq \mathrm{VaR}_{0.99}(S)\}$				
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}}]$	528 455.729	60 441.368	527 612.751	60 211.561
Standard error	3978.477	2119.461	4032.475	2995.992
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	846 956.570	349 871.745	854 461.670	370 931.946
Standard error	1866.133	6285.523	2570.997	9766.697
$\mathrm{VaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	865 603.369	413 767.829	871 533.550	437 344.509
Standard error	5995.341	29 105.059	12 780.741	21 142.135
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}})$	882 464.968	504 962.099	885 406.811	529 034.580
Standard error	3061.110	17 346.207	3134.144	23 617.278
(III) ES crisis event: $\{S \geq \mathrm{VaR}_{0.99}(S)\}$				
$\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}}]$	8 663 863.925	137 671.653	2 934 205.458	140 035.782
Standard error	3 265 049.590	10 120.557	165 794.772	14 601.958
$\mathrm{RVaR}_{0.975,0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	35 238 914.131	907 669.462	17 432 351.450	589 309.196
Standard error	2 892 208.689	31 983.660	443 288.649	3471.641
$\mathrm{VaR}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	56 612 082.905	1 131 248.055	20 578 728.307	615 572.940
Standard error	1 353 975.612	119 460.411	1 364 899.752	12 691.776
$\mathrm{ES}_{0.99}(\boldsymbol{X} \mid \mathcal{C}^{\mathrm{ES}})$	503 537 848.192	2 331 984.181	25 393 466.446	649 486.810
Standard error	268 007 317.199	468 491.127	1 138 243.137	7497.200

Table 5.3: Estimates and standard errors of the MC and HMC estimators of RVaR, VaR and ES type systemic risk allocations under the loss distribution (M3) with the (I) VaR crisis event, (II) RVaR crisis event and (III) ES crisis event. The MC sample size is $N_{\mathrm{MC}} = 10^5$ and that of the HMC method is $N_{\mathrm{MCMC}} = 10^4$. The acceptance rate (ACR), stepsize ϵ, integration time T and run time are ACR = 0.997, $\epsilon = 0.015$, $T = 34$ and run time = 2.007 mins in Case (I), ACR = 0.986, $\epsilon = 0.026$, $T = 39$ and run time = 2.689 mins in Case (II), ACR = 0.995, $\epsilon = 5.132 \times 10^{-5}$, $T = 838$ and run time = 44.831 mins in Case (III).

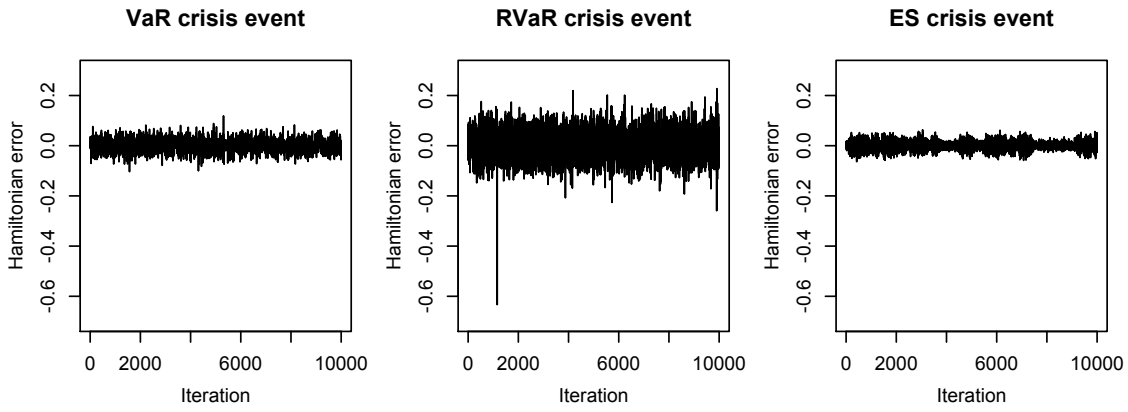

Figure 5.3: Hamiltonian errors of the HMC methods for estimating systemic risk allocations with VaR, RVaR and ES crisis events for the loss distribution (M3). The stepsize and the integration time are chosen as $(\epsilon, T) = (0.015, 34)$, $(\epsilon, T) = (0.026, 39)$ and $(\epsilon, T) = (5.132 \times 10^{-5}, 838)$, respectively.

MCMC methods with different dimensions, confidence levels, and parameters of the HMC methods in terms of bias, standard error and the mean squared error (MSE), adjusted by run time.

In this experiment, we fix the sample size of the MC and MCMC methods as $N_{\mathrm{MC}} = N_{\mathrm{MCMC}} = 10^4$. In addition, we assume $\boldsymbol{X} \sim t_\nu(\mathbf{0}, P)$, i.e., the joint loss follows the multivariate Student t distribution with $\nu = 6$ degrees of freedom, location vector $\mathbf{0}$ and dispersion matrix P, which is the correlation matrix with all off-diagonal entries equal to $1/12$. We let the dimension d of the loss portfolio vary, and consider only risk contribution type systemic risk allocations under VaR, RVaR and ES crisis events as true values of these allocations are available to compare against; see McNeil et al. (2015, Corollary 8.43). If b and σ denote the bias and standard deviation of the MC or MCMC estimator and S the run time, then (under the assumption that run time linearly increasing by sample size) we define the *time-adjusted MSEs* by

$$\mathrm{MSE}_{\mathrm{MC}} = b_{\mathrm{MC}}^2 + \frac{\sigma_{\mathrm{MC}}^2}{\frac{S_{\mathrm{MCMC}}}{S_{\mathrm{MC}}} \times N_{\mathrm{MCMC}}} \quad \text{and} \quad \mathrm{MSE}_{\mathrm{MCMC}} = b_{\mathrm{MCMC}}^2 + \frac{\sigma_{\mathrm{MCMC}}^2}{N_{\mathrm{MCMC}}}.$$

We can then compare the MC and MCMC estimators in terms of bias, standard error and time-adjusted MSE under the following three scenarios:

(A) $\mathrm{VaR}_{0.99}$, $\mathrm{RVaR}_{0.95,0.99}$ and $\mathrm{ES}_{0.99}$ contributions are estimated by the MC, HMC and GS methods for dimensions $d \in \{4, 6, 8, 10\}$. Note that the GS is applied only to RVaR and ES contributions, not to VaR contributions (same in the other scenarios).

(B) For $d = 5$, $\mathrm{VaR}_{p^{\mathrm{VaR}}}$, $\mathrm{RVaR}_{p_1^{\mathrm{RVaR}}, p_2^{\mathrm{RVaR}}}$ and $\mathrm{ES}_{p^{\mathrm{ES}}}$ contributions are estimated by the MC, HMC and

GS methods for confidence levels

$$\begin{aligned} p^{\mathrm{VaR}} &\in \{0.9, 0.99, 0.999, 0.9999\}, \\ (p_1^{\mathrm{RVaR}}, p_2^{\mathrm{RVaR}}) &\in \{(0.9, 0.9999), (0.9, 0.99), (0.99, 0.999), (0.999, 0.9999)\}, \quad \text{and} \\ p^{\mathrm{ES}} &\in \{0.9, 0.99, 0.999, 0.9999\}. \end{aligned}$$

(C) For $d = 5$, $\mathrm{VaR}_{0.9}$, $\mathrm{RVaR}_{0.9,0.99}$ and $\mathrm{ES}_{0.9}$ contributions are estimated by the MC and HMC methods with the parameters $(\epsilon_{\mathrm{opt}}, \epsilon_{\mathrm{opt}})$ (determined by Algorithm 7) and

$$(\epsilon, T) \in \left\{ (10\epsilon_{\mathrm{opt}}, 2T_{\mathrm{opt}}),\ \left(10\epsilon_{\mathrm{opt}}, \frac{T_{\mathrm{opt}}}{2}\right),\ \left(\frac{\epsilon_{\mathrm{opt}}}{10}, 2T_{\mathrm{opt}}\right),\ \left(\frac{\epsilon_{\mathrm{opt}}}{10}, \frac{T_{\mathrm{opt}}}{2}\right) \right\}.$$

In the MC method, the modified VaR contribution $\mathbb{E}[\boldsymbol{X} \mid \mathcal{C}^{\mathrm{RVaR}}_{p-\delta,p+\delta}]$ with $\delta = 0.01$ is computed. Moreover, if the size of the conditional sample for estimating RVaR and ES contributions is less than 100, then the lower confidence level of the crisis event is subtracted by 0.01 so that at least 100 MC presamples are guaranteed. For the sample paths of the MCMC methods, ACR, ACP, and Hamiltonian errors for the HMC methods were inspected and the convergences of the chains were checked as in Section 6.5.2 and 6.5.1.

The results of the comparisons of (A), (B) and (C) are summarized in Figure 5.4, 5.5 and 5.6. In Figure 5.4, the performance of the MC, HMC and GS estimators is roughly similar across dimensions from 4 to 10. For all crisis events, the HMC and GS estimators outperform MC in terms of bias, standard error and time-adjusted MSE. From (A5) and (A8), standard errors of the GS estimators are slightly higher than those of the HMC ones, which result in slightly improved performance of the HMC estimator over the GS in terms of MSE. In Figure 5.5, bias, standard error and MSE of the MC estimator tend to increase as the probability of the conditioning set decreases. This is simply because the size of the conditional samples in the MC method decreases proportional to the probability of the crisis event. On the other hand, the HMC and GS estimators provide a stably better performance than MC since no sample size reduction occurs. As seen in (B4) to (B9) in the cases of $\mathrm{RVaR}_{0.999,0.9999}$ and $\mathrm{ES}_{0.9999}$, however, if the probability of the conditioning event is too small and/or the distribution of the MC presample is too different from the original conditional distribution of interest, then the parameters of the HMC method determined by Algorithm 7 can be entirely different from optimal, which leads to a poor performance of the HMC method as we will see in the next scenario (C). In Figure 5.6, the HMC method with optimally determined parameters from Algorithm 7 is compared to non-optimal parameter choices. First, the optimal HMC estimator outperforms MC in terms of bias, standard error and time-adjusted MSE. On the other hand, from the plots in Figure 5.6 we see that some of the non-optimal HMC estimators are significantly worse than MC. Therefore, a careful choice of the parameters of the HMC method is required to obtain an improved performance of the HMC method in comparison to MC.

5.5 Conclusion, limitations and future work

Efficient calculation of systemic risk allocations is a challenging task, especially when the crisis event has a small probability. To solve this problem for models where a joint loss density is available, we proposed

MCMC estimators where a Markov chain is constructed with the conditional loss distribution given the crisis event as the target distribution. By using HMC and GS, efficient simulation methods from the constrained target distribution are obtained and the resulting MCMC estimator is expected to have a smaller standard error compared to that of the MC estimator. Sample efficiency is significantly improved since the MCMC estimator is computed from samples generated directly from the conditional distribution of interest. Another advantage of the MCMC method is that its performance is less sensitive to the probability of the crisis event, and thus to the confidence levels of the underlying risk measures. We also proposed a heuristic for determining the parameters of the HMC method based on the MC presamples. Numerical experiments demonstrated that our MCMC estimators are more efficient than MC in terms of bias, standard error and time-adjusted MSE. Stability of the MCMC estimation with respect to the probability of the crisis event and efficiency of the optimal parameter choice of the HMC method are also investigated in the experiments.

Based on the results in this chapter, our MCMC estimators can be recommended when the probability of the crisis event is too small for MC to simulate sufficiently many samples for a statistical analysis and/or when unbiased systemic risk allocations under the VaR crisis event are required. The MCMC methods are likely to perform well when the dimension of the portfolio is less than or around 10, losses are bounded from the left and the crisis event is of VaR or RVaR type; otherwise heavy-tailedness and computational time can become challenging. First, a theoretical convergence result of the HMC method is typically not available when the target distribution is unbounded and heavy-tailed, which is the case when the losses are unbounded and/or the crisis event is of ES type; see the case of the ES crisis event in the empirical study in Section 6.5.1. Second, both of the HMC and GS methods suffer from high-dimensional target distributions since the algorithms contain parts of steps where the computational cost linearly increases in the dimension. We observed that, in this case, although the MCMC estimator typically improves bias and standard error compared to MC, the improvement vanishes in terms of time-adjusted MSE due to the long computational time of the MCMC method. Finally, multimodality of joint loss distributions and/or the target distribution is also an undesirable feature since full conditional distributions and their inverses (which are required to implement the GS) are typically unavailable in the former case, and the latter case prevents the HMC method from efficiently exploring the entire support of the target distribution. Potential remedies for heavy-tailed and/or high-dimensional target distributions are the HMC method with a non-normal kinetic energy distribution and roll-back HMC. Further investigation of HMC methods and faster methods for determining the HMC parameters are left for future work.

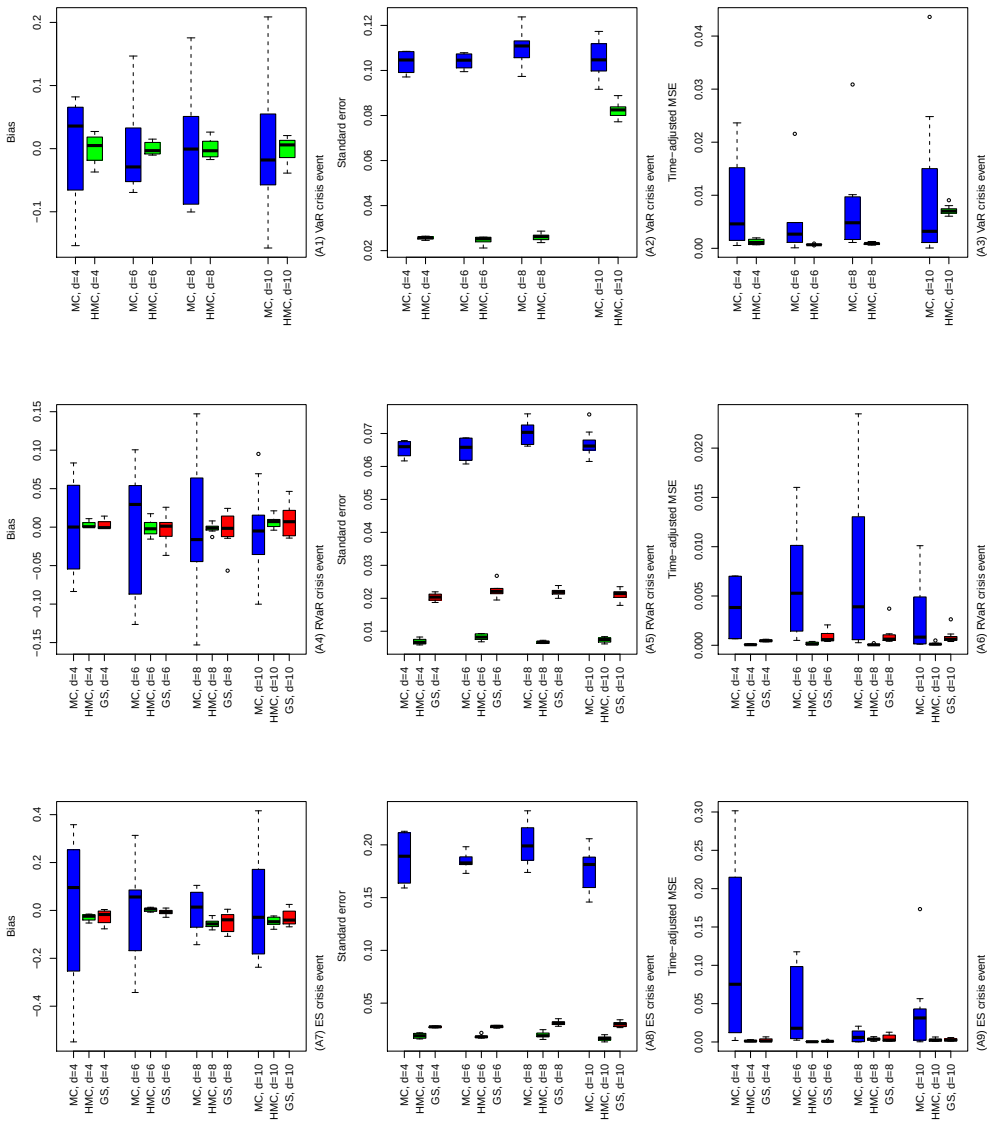

Figure 5.4: Bias (left), standard error (middle) and time-adjusted mean squared error (right) of the MC, HMC and GS estimators of risk contribution type systemic risk allocations under $\mathrm{VaR}_{0.99}$ (top), $\mathrm{RVaR}_{0.95,0.99}$ (middle) and $\mathrm{ES}_{0.99}$ (bottom) crisis events. The underlying loss distribution is $t_\nu(\boldsymbol{\mu}, P)$ where $\nu = 6$, $\boldsymbol{\mu} = \mathbf{0}$ and $P = 1/12 \cdot \mathbf{1}_d\mathbf{1}_d^\top + \mathrm{diag}_d(11/12)$ for portfolio dimensions $d \in \{4, 6, 8, 10\}$. Note that the GS method is applied only to RVaR and ES contributions.

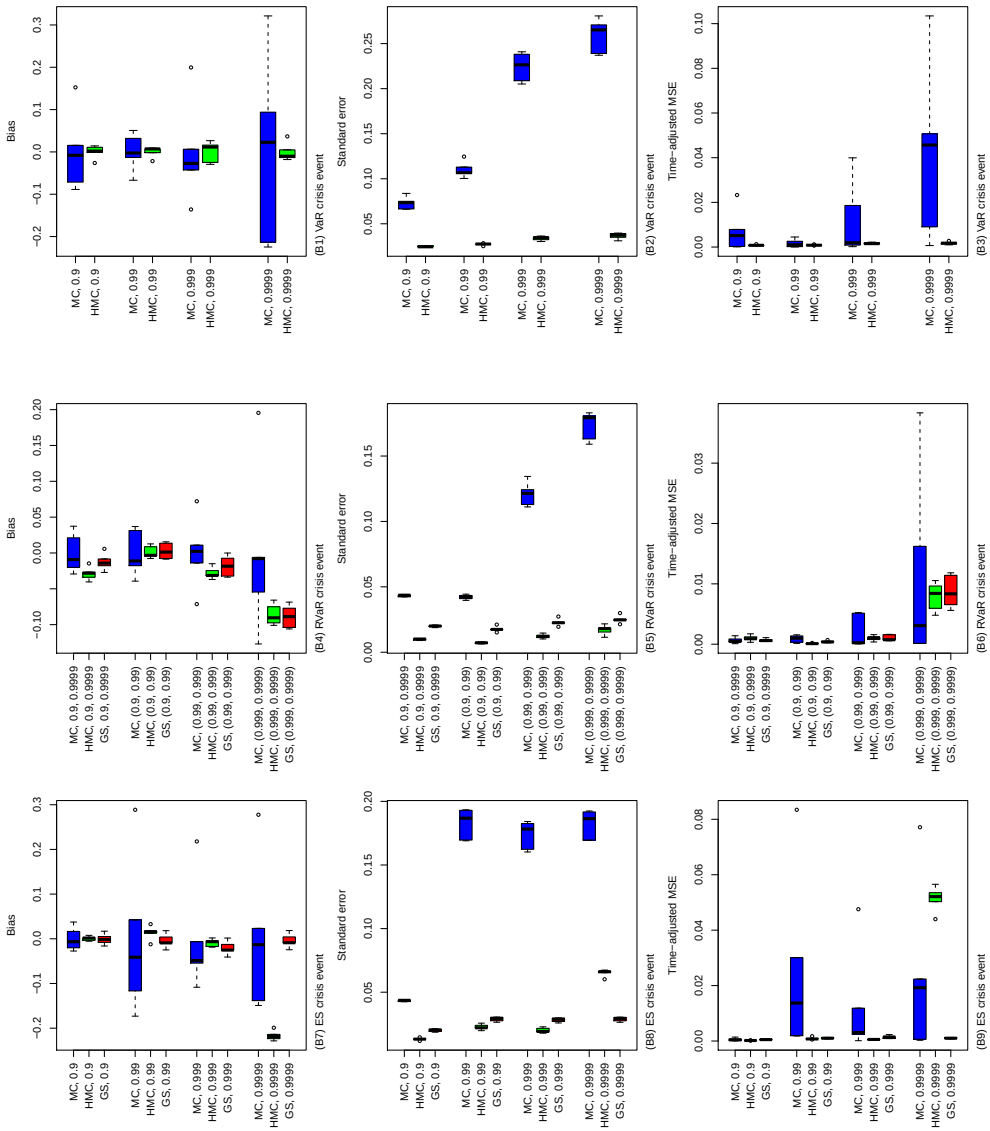

Figure 5.5: Bias (left), standard error (middle) and time-adjusted mean squared error (right) of the MC, HMC and GS estimators of risk contribution type systemic risk allocations with the underlying loss distribution $t_\nu(\boldsymbol{\mu}, P)$ where $\nu = 6$, $\boldsymbol{\mu} = \mathbf{0}$, $P = 1/12 \cdot \mathbf{1}_d\mathbf{1}_d^\top + \mathrm{diag}_d(11/12)$ and $d = 5$. The crisis event is taken differently as $\mathrm{VaR}_{p^{\mathrm{VaR}}}$ (top), $\mathrm{RVaR}_{p_1^{\mathrm{RVaR}},p_2^{\mathrm{RVaR}}}$ (middle) and $\mathrm{ES}_{p^{\mathrm{ES}}}$ (bottom) for confidence levels $p^{\mathrm{VaR}} \in \{0.9, 0.99, 0.999, 0.9999\}$, $(p_1^{\mathrm{RVaR}}, p_2^{\mathrm{RVaR}}) \in \{(0.9, 0.9999), (0.9, 0.99), (0.99, 0.999), (0.999, 0.9999)\}$ and $p^{\mathrm{ES}} \in \{0.9, 0.99, 0.999, 0.9999\}$. Note that the GS method is applied only to RVaR and ES contributions.

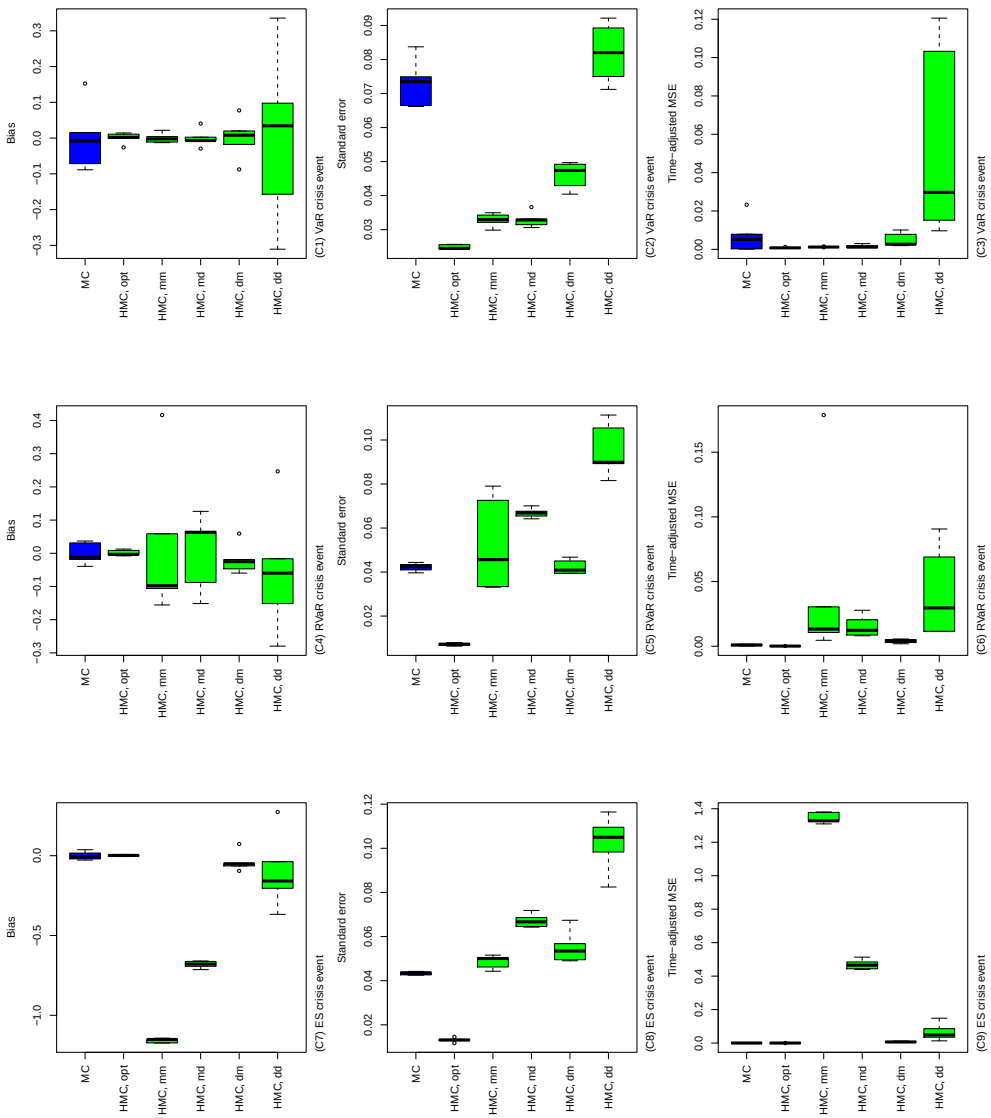

Figure 5.6: Bias (left), standard error (middle) and time-adjusted mean squared error (right) of the MC and HMC estimators of risk contribution type systemic risk allocations under $\mathrm{VaR}_{0.9}$, $\mathrm{RVaR}_{0.9,0.99}$ and $\mathrm{ES}_{0.9}$ crisis events. The underlying loss distribution is $t_\nu(\boldsymbol{\mu}, P)$ where $\nu = 6$, $\boldsymbol{\mu} = \mathbf{0}$, $P = 1/12 \cdot \mathbf{1}_d \mathbf{1}_d^\top + \mathrm{diag}_d(11/12)$ and $d = 5$. The parameters of the HMC method are taken as $(\epsilon_{\mathrm{opt}}, \epsilon_{\mathrm{opt}})$ determined by Algorithm 7 and $(\epsilon, T) \in \{(10\epsilon_{\mathrm{opt}}, 2T_{\mathrm{opt}}),\ (10\epsilon_{\mathrm{opt}}, T_{\mathrm{opt}}/2),\ (\epsilon_{\mathrm{opt}}/10, 2T_{\mathrm{opt}}),\ (\epsilon_{\mathrm{opt}}/10, T_{\mathrm{opt}}/2)\}$. In the labels of the x-axes, each of the five cases $(\epsilon_{\mathrm{opt}}, \epsilon_{\mathrm{opt}})$, $(10\epsilon_{\mathrm{opt}}, 2T_{\mathrm{opt}})$, $(10\epsilon_{\mathrm{opt}}, T_{\mathrm{opt}}/2)$, $(\epsilon_{\mathrm{opt}}/10, 2T_{\mathrm{opt}})$ and $(\epsilon_{\mathrm{opt}}/10, T_{\mathrm{opt}}/2)$ is denoted by HMC.opt, HMC.mm, HMC.md, HMC.dm and HMC.dd, respectively.

Algorithm 6 Leapfrog method with boundary reflection

Input: Current state $(\boldsymbol{x}(0), \boldsymbol{p}(0))$, stepsize $\epsilon > 0$, gradients ∇U and ∇K, and constraints $(\boldsymbol{h}_m, v_m)$, $m = 1, \ldots, M$.

Output: Updated state $(\boldsymbol{x}(\epsilon), \boldsymbol{p}(\epsilon))$.

1) Update $\boldsymbol{p}(\epsilon/2) = \boldsymbol{p}(0) + \epsilon/2 \nabla U(\boldsymbol{x}(0))$.

2) Set $(\boldsymbol{x}, \boldsymbol{p}) = (\boldsymbol{x}(0), \boldsymbol{p}(\epsilon/2))$, $\epsilon_{\text{temp}} = \epsilon$.

3) **while** $\epsilon_{\text{temp}} > 0$

3-1) Compute

$$\begin{aligned} \boldsymbol{x}^* &= \boldsymbol{x} + \epsilon_{\text{temp}} \nabla K(\boldsymbol{p}), \\ t_m &= (v_m - \boldsymbol{h}_m^\top \boldsymbol{x})/(\epsilon \boldsymbol{h}_m^\top \boldsymbol{p}), \quad m = 1, \ldots, M. \end{aligned}$$

3-2) **if** $t_m \in [0, 1]$ for any $m = 1, \ldots, M$,

3-2-1) Set

$$\begin{aligned} m^* &= \text{argmin}\{t_m \mid 0 \le t_m \le 1,\ m = 1, \ldots, M\}, \\ \boldsymbol{x}_r^* &= \boldsymbol{x}^* - 2\frac{\boldsymbol{h}_{m^*}^\top \boldsymbol{x}^* - v_{m^*}}{\boldsymbol{h}_{m^*}^\top \boldsymbol{h}_{m^*}} h_{m^*}, \\ \boldsymbol{p}_r &= \frac{\boldsymbol{x}^* - \boldsymbol{x} - t_{m^*} \epsilon \boldsymbol{p}}{\epsilon(1 - t_{m^*})}. \end{aligned}$$

3-2-2) Set $(\boldsymbol{x}, \boldsymbol{p}) = (\boldsymbol{x}_r^*, \boldsymbol{p}_r)$ and $\epsilon_{\text{temp}} = (1 - t_{m^*})\epsilon_{\text{temp}}$.

else

3-2-3) Set $(\boldsymbol{x}, \boldsymbol{p}) = (\boldsymbol{x}^*, \boldsymbol{p})$ and $\epsilon_{\text{temp}} = 0$.

end if

end while

4) Set $\boldsymbol{x}(\epsilon) = \boldsymbol{x}$ and $\boldsymbol{p}(\epsilon) = \boldsymbol{p} + \frac{\epsilon}{2} \nabla U(\boldsymbol{x})$.

Algorithm 7 Heuristic for determining the stepsize ϵ and integration time T

Input: MC presample $\boldsymbol{X}_1^{(0)}, \ldots, \boldsymbol{X}_{N_0}^{(0)}$, gradients ∇U and ∇K, target acceptance probability $\underline{\alpha}$, initial constant $c_\epsilon > 0$ and the maximum integration time $T_{\max}$ ($c_\epsilon = 1$ and $T_{\max} = 1000$ are set as default values).

Output: Stepsize ϵ and integration time T.

---

1) Set $\alpha_{\min} = 0$ and $\epsilon = c_\epsilon d^{-1/4}$.

2) **while** $\alpha_{\min} < \underline{\alpha}$

2-1) Set $\epsilon = \epsilon/2$.

2-2) **for** $n = 1, \ldots, N_0$

2-2-1) Generate $\boldsymbol{p}_n^{(0)} \sim F_K$.

2-2-2) **for** $t = 1, \ldots, T_{\max}$

2-2-2-1) Set $\boldsymbol{Z}_n^{(t)} = \textbf{Leapfrog}(\boldsymbol{Z}_n^{(t-1)}, \epsilon, \nabla U, \nabla K)$ for $\boldsymbol{Z}_n^{(t-1)} = (\boldsymbol{X}_n^{(t-1)}, \boldsymbol{p}_n^{(t-1)})$.

2-2-2-2) Calculate

$$\alpha_{n,t} = \alpha(\boldsymbol{Z}_n^{(t-1)}, \boldsymbol{Z}_n^{(t)}) \quad \text{and} \quad \Delta_t = ||\boldsymbol{X}_n^{(t)} - \boldsymbol{X}_n^{(0)}|| - ||\boldsymbol{X}_n^{(t-1)} - \boldsymbol{X}_n^{(0)}||.$$

2-2-2-3) **if** $\Delta_t < 0$ and $\Delta_{t-1} > 0$, **break** and set $T_n^* = t - 1$.

end for

end for

2-3) Compute $\alpha_{\min} = \min(\alpha_{n,t} \mid t = 1, 2, \ldots,\ T_n^*,\ n = 1, \ldots, N_0)$

end while

3) Set $T = \lfloor \frac{1}{N_0} \sum_{n=1}^{N_0} T_n^* \rfloor$.

Algorithm 8 Random scan Gibbs sampler (RSGS) with heuristic to determine $(p_1, \ldots, p_d)$

Require: Random number generator of $\pi_{j|-j}$ and $\boldsymbol{x}^{(0)} \in \mathrm{supp}(\pi)$.

Input: MC presample $\tilde{\boldsymbol{X}}_1^{(0)}, \ldots, \tilde{\boldsymbol{X}}_{N_0}^{(0)}$, sample size $N \in \mathbb{N}$, initial state $\boldsymbol{x}^{(0)}$, sample size of the pre-run N_{pre} and the target autocorrelation ρ ($N_{\mathrm{pre}} = 100$ and $\rho = 0.15$ are set as default values).

Output: N sample path $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}$ of the Markov chain.

1) Compute the sample covariance matrix $\hat{\Sigma}$ based on $\tilde{\boldsymbol{X}}_1^{(0)}, \ldots, \tilde{\boldsymbol{X}}_{N_0}^{(0)}$.

2) Set $p_j \propto \hat{\Sigma}_{j,j} - \hat{\Sigma}_{j,-j}\hat{\Sigma}_{-j,-j}^{-1}\hat{\Sigma}_{-j,j}$ and $\boldsymbol{X}^{(0)} = \boldsymbol{X}_{\mathrm{pre}}^{(0)} = \boldsymbol{x}^{(0)}$.

3) **for** $n = 1, \ldots, N_{\mathrm{pre}}$

   3-1) Generate $J = j$ with probability p_j.

   3-2) Update $X_{\mathrm{pre},J}^{(n)} \sim \pi_{J|-J}(\cdot|\boldsymbol{X}_{\mathrm{pre}}^{(n-1)})$ and $\boldsymbol{X}_{\mathrm{pre},-J}^{(n)} = \boldsymbol{X}_{\mathrm{pre},-J}^{(n-1)}$.

   end for

4) Set

$$T = \mathrm{argmin}_{h \in \mathbb{N}_0} \left\{ \text{estimated autocorrelations of } \boldsymbol{X}_{\mathrm{pre}}^{(1)}, \ldots, \boldsymbol{X}_{\mathrm{pre}}^{(N_{\mathrm{pre}})} \text{ with lag } h \ \leq \rho \right\}.$$

5) **for** $n = 1, \ldots, N$, $t = 1, \ldots, T$

   5-1) Generate $J = j$ with probability p_j.

   5-2) Update $X_J^{(n-1+t/T)} \sim \pi_{J|-J}(\cdot|\boldsymbol{X}^{(n-1+(t-1)/T)})$ and $\boldsymbol{X}_{-J}^{(n-1+t/T)} = \boldsymbol{X}_{-J}^{(n-1+(t-1)/T)}$.

   end for

Chapter 6

Modality for scenario analysis and maximum likelihood allocation

We analyze dependence, tail behavior and multimodality of the conditional distribution of a loss random vector given that the aggregate loss equals an exogenously provided capital. This conditional distribution serves as a building block for calculating risk allocations such as the Euler capital allocation of Value-at-Risk. A superlevel set of this conditional distribution can be interpreted as a set of severe and plausible stress scenarios the given capital is supposed to cover. We show that various distributional properties of this conditional distribution are inherited from those of the underlying joint loss distribution. Among these properties, we find that modality of the conditional distribution is an important feature in risk profile related to the number of risky scenarios likely to occur in a stressed situation. Under unimodality, we study a novel risk allocation method called maximum likelihood allocation (MLA), defined as the mode of the conditional distribution given the total capital. Under multimodality, a single vector of allocations can be less sound. To overcome this issue, we investigate the so-called multimodalty adjustment to increasing the soundness of risk allocations. Properties of the conditional distribution, MLA and multimodality adjustment are demonstrated in numerical experiments. In particular, we observe that negative dependence among losses typically leads to multimodality, and thus to multiple risky scenarios and higher multimodality adjustment.

6.1 Introduction

Risk allocation concerns the quantification of the risk of each unit of a portfolio. For a d-dimensional portfolio of risks or losses (typically risk-factor changes) represented by an $\mathbb{R}^d$-valued random vector $\boldsymbol{X} = (X_1, \dots, X_d)$, $d \in \mathbb{N}$, the overall loss $S = X_1 + \cdots + X_d$ is quantified as a total capital $K \in \mathbb{R}$ and typically determined as $K = \varrho(S)$ for a risk measure ρ. The Euler principle, one of the most well-known rules of risk allocation, is applicable when the total capital is determined by a risk measure via $K = \varrho(S)$; see Section 1.3. However, as pointed out by Asimit et al. (2019), the total capital in practice may not always

coincide with the risk measure itself but includes various adjustments such as stress scenarios and liquidity adjustments. In such cases, the capital does not possess the original meaning as a risk measure and the formula under the Euler principle is not available. In addition, there are situations when the total capital is given exogenously as a constant; see Laeven and Goovaerts (2004). For the case when the total capital is regarded as a constant, various allocation methods have been proposed in the literature. One of the main streams found, for example, in Laeven and Goovaerts (2004) and Dhaene et al. (2012), is to derive an allocation as a minimizer of some loss function over a set of allocations $\mathcal{K}_d(K) = \{\boldsymbol{x} \in \mathbb{R}^d : x_1 + \cdots + x_d = K\}$. Another method is to find a confidence level for which the corresponding risk measure coincides with K, and then allocate K by regarding it as measured by a risk measure. For example, if Value-at-Risk (VaR) or Expected Shortfall (ES) are chosen as risk measures, confidence levels $p_{\mathrm{VaR}},\ p_{\mathrm{ES}} \in (0,1)$ are first found such that $K = \mathrm{VaR}_{p_{\mathrm{VaR}}}(S)$ or, respectively, $K = \mathrm{ES}_{p_{\mathrm{ES}}}(S)$ hold. After performing this procedure, the Euler principle becomes applicable to K and the resulting risk allocation of K allocates $\mathbb{E}[X_j \mid \{S = K\}]$ or, respectively, $\mathbb{E}[X_j \mid \{S \geq \mathrm{VaR}_{p_{\mathrm{ES}}}(S)\}]$ to the jth risk X_j; see Section 6.2.1 for details.

Although these methods often provide plausible risk allocations, they sometimes ignore important distributional properties of $\boldsymbol{X}$ related to the soundness of risk allocations and to risky scenarios expected to be covered by the allocated capitals. As we will see in Section 6.2.2, all these allocation methods provide the homogeneous allocation $(K/d, \dots, K/d)$ when $\boldsymbol{X}$ is exchangeable in the sense that $\boldsymbol{X} \overset{\mathrm{d}}{=} (X_{\pi(1)}, \dots, X_{\pi(d)})$ for any permutation $(\pi(1), \dots, \pi(d))$ of $\{1, \dots, d\}$. This homogeneous allocation can be sound when the conditional distribution of $\boldsymbol{X}$ in a stressed situation is unimodal with the mode $(K/d, \dots, K/d)$ since this homogeneous allocation covers the risky scenario most likely to occur in a stressed situation. On the other hand, the same allocation $(K/d, \dots, K/d)$ arises when the conditional distribution in a stressed situation is multimodal and $(K/d, \dots, K/d)$ is supposed to cover multiple risky scenarios *on average*. In this multimodal case, the homogeneous allocation is less sound than in the former unimodal case without identifying the multiple risky scenarios hidden in a single vector of $(K/d, \dots, K/d)$. Consequently, the soundness of risk allocation can depend on the distributional properties of the conditional distribution of $\boldsymbol{X}$ in a stressed situation.

In this chapter, we focus on the conditional distribution of $\boldsymbol{X}$ given $\{S = K\}$. Since $\boldsymbol{X} \mid \{S = K\}$ takes values in $\mathcal{K}_d(K)$, this random vector can be a building block for deriving a risk allocation. For example, the Euler allocation (1.5) arises when $K = \mathrm{VaR}_p(S)$ for some $p \in (0,1)$ and the expectation of $\boldsymbol{X} \mid \{S = K\}$ is considered. In Section 6.2.2 we show that a superlevel set of $\boldsymbol{X} \mid \{S = K\}$ can be regarded as a set of severe and plausible stress scenarios the given capital K is supposed to cover. Based on the motivation provided there, we investigate distributional properties of $\boldsymbol{X} \mid \{S = K\}$ in Section 6.3. We show that dependence, tail behavior and unimodality of $\boldsymbol{X} \mid \{S = K\}$ are typically inherited from those of the underlying unconditional loss $\boldsymbol{X}$, respectively. In addition, we demonstrate by simulation that negative dependence among $\boldsymbol{X}$ typically leads to multimodality of $\boldsymbol{X} \mid \{S = K\}$; see Section 6.5.2. These observations can be useful to detect the hidden risk of multimodality in risk allocation. Furthermore, the properties of $\boldsymbol{X} \mid \{S = K\}$ studied in this chapter are of potential importance in simulation and statistical inference of $\boldsymbol{X} \mid \{S = K\}$ using Markov chain Monte Carlo (MCMC) methods for efficiently simulating the distribution of interest; see Remark 6.2.1 and Section 6.5.3.

We also propose a novel risk allocation method termed *maximum likelihood allocation (MLA)*, which

is defined as the mode of $\boldsymbol{X} \mid \{S = K\}$ under unimodality. Besides the mean (which leads to the Euler allocation of VaR), the mode is also an important summary statistics of $\boldsymbol{X} \mid \{S = K\}$. It can be interpreted as the risky scenario most likely to occur in the stressed situation $\{S = K\}$. By searching for the global mode of $\boldsymbol{X} \mid \{S = K\}$, possibly multiple local modes can be detected. As explained in Section 6.2.2, this procedure of detecting multimodality is beneficial for evaluating the soundness of risk allocations, for discovering hidden multiple scenarios likely to occur in the stressed situation $\{S = K\}$ and for constructing more flexible risk allocations by weighting important scenarios. Definitions and required assumptions on MLA are provided in Section 6.4.1. In Section 6.4.2, we investigate properties of MLA expected to hold for a risk allocation. In Section 6.4.4 we introduce the so-called multimodality adjustment to increase the soundness of risk allocations under multimodality. MLA and multimodality-adjusted allocated capitals are estimated and compared with the Euler allocation in numerical experiments based on real data in Section 6.5.1 and based on simulated data in Section 6.5.2. Concluding remarks are given in Section 6.6.

6.2 Preliminaries

6.2.1 Notation and setup

We consider the case when the capital is an exogenously given constant $K \in \mathbb{R}$. The set of all possible allocations is denoted by

$$\mathcal{K}_d(K) := \{\boldsymbol{x} \in \mathbb{R}^d : x_1 + \cdots + x_d = K\}.$$

If $K = \varrho(S)$ for a risk measure ϱ, the Euler principle determines the jth allocated capital as presented in Section 1.3. Our proposed risk allocation introduced in Section 6.4 is based on the conditional distribution

$$F_{\boldsymbol{X}|\{S=K\}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \le \boldsymbol{x} \mid \{S = K\}), \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{6.1}$$

where $\boldsymbol{X}$ and S are as defined in Section 1.3. The conditional distribution (6.1) is degenerate and its first $d' = d - 1$ components $\boldsymbol{X}' \mid \{S = K\} = (X_1, \dots, X_{d'}) \mid \{S = K\}$ determine the last one via $X_d \mid \{S = K\} = K - (X_1 + \cdots + X_{d'}) \mid \{S = K\}$. Therefore, it suffices to consider the d'-dimensional marginal distribution $F_{\boldsymbol{X}'|\{S=K\}}$. Assuming that $\boldsymbol{X}$ and $(\boldsymbol{X}', S)$ admit densities, $\boldsymbol{X}' \mid \{S = K\}$ also has a density and is given by

$$f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') = \frac{f_{(\boldsymbol{X}',S)}(\boldsymbol{x}', K)}{f_S(K)} = \frac{f_{\boldsymbol{X}}(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(K)}, \quad \boldsymbol{x}' \in \mathbb{R}^{d'}, \tag{6.2}$$

where the last equality follows from an affine transformation $(\boldsymbol{X}', S) \mapsto \boldsymbol{X}$ with unit Jacobian.

6.2.2 A motivating example

The distribution of $\boldsymbol{X} \mid \{S = K\}$ is a primary subject in this chapter. In this section, we provide a motivating example for investigating this distribution from the viewpoint of risk alloation.

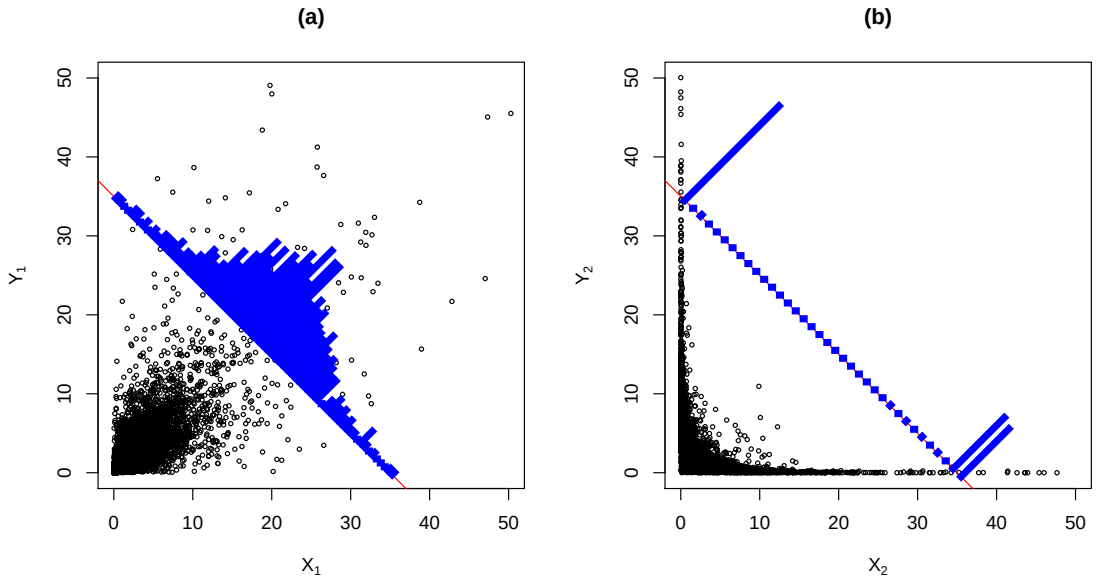

Figure 6.1: Scatter plots (black dots) of (a) (X_1, Y_1) and (b) (X_2, Y_2) such that all of X_1, Y_1, X_2 and Y_2 are identically Pareto distributed with shape parameter 3 and scale parameter 5, and (X_1, Y_1) and (X_2, Y_2) have Student t copulas C^t_{ν,ρ_1} and C^t_{ν,ρ_2}, respectively, where $\nu = 5$ is the degrees of freedom, and $\rho_1 = 0.8$ and $\rho_2 = -0.8$ are the correlation parameters. The red line indicates $x + y = K$ for $K = 35$. Histograms (blue) of the conditional distributions of (a) (X_1, Y_1) and (b) (X_2, Y_2) on the (approximate) set of allocations $\{(x, y) \in \mathbb{R}^2 : K - \delta < x + y < K + \delta\}$, $\delta = 0.5$, are drawn on $\mathcal{K}_d(K) = \{(x, y) \in \mathbb{R}^2 : x + y = K\}$.

To this end, consider two bivariate risks (a) (X_1, Y_1) and (b) (X_2, Y_2) such that all of X_1, Y_1, X_2 and Y_2 are identically Pareto distributed with shape parameter 3 and scale parameter 5, and (X_1, Y_1) and (X_2, Y_2) have Student t copulas C^t_{ν,ρ_1} and C^t_{ν,ρ_2}, respectively, where $\nu = 5$ is the degrees of freedom parameter and $\rho_1 = 0.8$ and $\rho_2 = -0.8$ are the correlation parameters. Suppose that the exogenously given total capital equals $K = 35$. By exchangeability of the risk models (a) and (b), most allocation rules provide the homogeneous allocation $(K/2, K/2) = (17.5, 17.5)$ in both cases (a) and (b). For instance, if K is regarded as VaR or ES at some confidence levels and is allocated according to the Euler principle, then both VaR and ES contributions lead to homogeneous allocations. As we see in Figure 6.1, however, the conditional distributions of (X_1, Y_1) and of (X_2, Y_2) on the set of allocations $\mathcal{K}_d(K)$ differ substantially. Positive dependence among X_1 and Y_1 prevents the two random variables from moving in opposite directions under the constraint $X_1 + Y_1 = K$, which results in unimodality of the conditional distribution on $\mathcal{K}_d(K)$. On the other hand, negative dependence among X_2 and Y_2 allows them to move in opposite directions, which leads to bimodality of the conditional distribution. From the viewpoint of risk management, the homogeneous allocation $(K/2, K/2)$ seems to be a more sound capital allocation in Case (a) because it covers the most likely risky scenario. In Case (b), the two risky scenarios around the corners $(K, 0)$ and $(0, K)$ occur equally likely and the allocation $(K/2, K/2)$ can be understood as an average of these scenarios. However, the likelihood around $(K/2, K/2)$ is quite small and a single vector of the equal allocation $(K/2, K/2)$ obscures

the two distinct risky scenarios. Consequently, the soundness of the allocated capital depends on the modality of the conditional loss distribution, and multiple risky scenarios can be hidden in a single vector resulting from capital allocation.

Inspecting modes of $\boldsymbol{X} \mid \{S = K\}$ can also be regarded as a stress test of risk allocations. Breuer et al. (2009) requires stress scenarios to be severe and plausible. Define the *scenario set with a level of plausibility* $t > 0$ by $L_t(\boldsymbol{X}) = \{\boldsymbol{x} \in \mathbb{R}^d : f_{\boldsymbol{X}}(\boldsymbol{x}) \geq t\}$ where $f_{\boldsymbol{X}}$ is the density function of $\boldsymbol{X}$. Among the scenario set $L_t(\boldsymbol{X})$, the set $L_t(\boldsymbol{X}) \cap \mathcal{K}_d(K)$ can be regarded as a set of most severe scenarios the given total capital K can cover. Using the convention $f_{\boldsymbol{X}|\{S=K\}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\mathbf{1}_d^\top \boldsymbol{x}=K\}}/f_S(K)$, $\boldsymbol{x} \in \mathbb{R}^d$, the set $L_t(\boldsymbol{X}) \cap \mathcal{K}_d(K)$ leads to the superlevel set of $\boldsymbol{X} \mid \{S = K\}$ since

$$\begin{aligned} L_t(\boldsymbol{X}) \cap \mathcal{K}_d(K) &= \{\boldsymbol{x} \in \mathbb{R}^d : f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\mathbf{1}_d^\top \boldsymbol{x}=K\}} \geq t\} \\ &= \{\boldsymbol{x} \in \mathbb{R}^d : f_{\boldsymbol{X}|\{S=K\}}(\boldsymbol{x}) \geq t/f_S(K)\} = L_{t/f_S(K)}(\boldsymbol{X} \mid \{S = K\}). \end{aligned}$$

Throughout the chapter, the superlevel set of $\boldsymbol{X} \mid \{S = K\}$ is treated as a set of *stress scenarios*. In particular, the mode of $\boldsymbol{X} \mid \{S = K\}$ is the most severe and plausible scenario that K can cover since it attains the highest level of plausibility among the stress scenarios. Unimodality of $\boldsymbol{X} \mid \{S = K\}$ (see Definition 6.3.12 for its formal definition) implies that there exists one representative stress scenario the total capital K can cover, and thus the mode is a sound allocation covering the risky scenario most likely to occur. On the other hand, multimodality of $\boldsymbol{X} \mid \{S = K\}$ (again see Definition 6.3.12) means that there are multiple distinct stress scenarios that are severe and plausible, and thus it may not be sufficient to only focus on a single scenario without identifying the other ones.

Remark 6.2.1 (Simulation of $\boldsymbol{X} \mid \{S = K\}$ with MCMC methods)**.** Another motivation for investigating distributional properties of $\boldsymbol{X} \mid \{S = K\}$ is to be able to efficiently simulate this conditional distribution. This is a challenging task since there are no general and tractable sampling methods known for $\boldsymbol{X} \mid \{S = K\}$. Although samples from $\boldsymbol{X}$ satisfying the constraint $\{S = K\}$ can be regarded as samples from $\boldsymbol{X} \mid \{S = K\}$, the probability $\mathbb{P}(S = K)$ is zero, and thus such samples virtually never exist when S admits a density. As done in Chapters 4 and 5, a potential remedy of this problem is to modify the conditioning set $\{S = K\}$ to $\{K - \delta < S < K + \delta\}$ for a small $\delta > 0$ so that $P(K - \delta < S < K + \delta) > 0$. However, this modification distorts the distribution of $\boldsymbol{X} \mid \{S = K\}$ and the resulting estimates of risk allocations are biased. To overcome this issue, we considered MCMC methods for exact simulation from $\boldsymbol{X} \mid \{S = K\}$ in Chapters 4 and 5. Although MCMC methods improve sample efficiency and the resulting estimates are unbiased, their performance highly depends on distributional properties of $\boldsymbol{X} \mid \{S = K\}$, in particular on its modality and heavy-tailedness. From this viewpoint, investigating properties of $\boldsymbol{X} \mid \{S = K\}$ is important for constructing efficient MCMC methods for simulating $\boldsymbol{X} \mid \{S = K\}$.

6.3 Properties of the conditional distribution given a constant sum

In Section 6.2.2, the conditional distribution of $\boldsymbol{X}$ given a constant sum $\{S = K\}$ turned out to play important roles in stress testing of risk allocations. With this motivation, we study the support, dependence,

tail behavior and modality of $\boldsymbol{X} \mid \{S = K\}$ in this section. As introduced in Section 6.2.1, we consider the d'-dimensional random vector $\boldsymbol{X}' \mid \{S = K\}$ for $d' = d - 1$ to avoid the degeneracy of the conditional distribution $\boldsymbol{X} \mid \{S = K\}$.

6.3.1 Support of the conditional distribution

We start with the support of $f_{\boldsymbol{X}'|\{S=K\}}$. By Equation (6.2),

$$\operatorname{supp}(\boldsymbol{X}' \mid \{S = K\}) = \{\boldsymbol{x}' \in \mathbb{R}^{d'} : f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') > 0\} = \{\boldsymbol{x}' \in \mathbb{R}^{d'} : f_{\boldsymbol{X}}(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}') > 0\}.$$

If $X_1, \dots, X_d$ are supported on $\mathbb{R}^d$, we have $\operatorname{supp}(\boldsymbol{X}' \mid \{S = K\}) = \mathbb{R}^{d'}$. Another typical case is when $X_1, \dots, X_d$ are bounded from below, that is, there exists $l_1, \dots, l_d > -\infty$ such that $X_j \geq l_j$ $\mathbb{P}$-a.s. for $j = 1, \dots, d$. In this case, $\operatorname{supp}(\boldsymbol{X}) = (l_1, \infty) \times \cdots \times (l_d, \infty)$ and thus the support of $f_{\boldsymbol{X}'|\{S=K\}}$ is given by

$$\operatorname{supp}(\boldsymbol{X}' \mid \{S = K\}) = \left\{ \boldsymbol{x}' \in \mathbb{R}^{d'} : x_1 > l_1, \dots, x_{d'} > l_{d'}, \sum_{j=1}^{d'} x_j' < K - l_d \right\}. \tag{6.3}$$

If $l_1 = \cdots = l_{d'} = 0$, that is, when $\boldsymbol{X}$ models the nonnegative part of losses, the closure of (6.3) is known as the *K-simplex*. Since the set in (6.3) is bounded, simulation of $\boldsymbol{X}' \mid \{S = K\}$ can be more straightforward than in the former case when $\operatorname{supp}(\boldsymbol{X}' \mid \{S = K\}) = \mathbb{R}^{d'}$. For instance, an independent Metropolis-Hastings (MH) algorithm can be applied by first generating a sample $\boldsymbol{y}'$ uniformly on the set in (6.3) (which is a location-shifted simplex and thus uniform sampling from this set can be achieved by simulating a specific Dirichlet distribution) and then replacing the current state $\boldsymbol{x}'$ with the new state $\boldsymbol{y}'$ with probability $\alpha(\boldsymbol{x}', \boldsymbol{y}') = f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{y}')/f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') = f_{\boldsymbol{X}}(\boldsymbol{y}', K - \mathbf{1}_{d'}^\top \boldsymbol{y}')/f_{\boldsymbol{X}}(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}')$.

6.3.2 The conditional distribution in the elliptical case

Elliptical distributions are important exceptions for which the distribution of $\boldsymbol{X}' \mid \{S = K\}$ can be derived explicitly. For applications of elliptical distributions to risk management, see, for example, Landsman and Valdez (2003), Dhaene et al. (2008) or Chapter 6 of McNeil et al. (2015). The characteristic function of a random vector $\boldsymbol{X}$ is given by $\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(i\boldsymbol{t}^\top \boldsymbol{X})]$, $\boldsymbol{t} \in \mathbb{R}^d$. If a function $\psi(t) : [0, \infty) \to \mathbb{R}$ is such that $\psi(\boldsymbol{t}^\top \boldsymbol{t})$ is a d-dimensional characteristic function, then ψ is called a *characteristic generator*; see Fang (2018) for details. Let Ψ_d denote the class of all characteristic generators. A d-dimensional random vector $\boldsymbol{X}$ is said to have an *elliptical distribution*, denoted by $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$, if its characteristic function can be expressed as

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(i\boldsymbol{t}^\top \boldsymbol{\mu})\ \psi\left(\frac{1}{2}\boldsymbol{t}^\top \Sigma \boldsymbol{t}\right)$$

for a *location vector* $\boldsymbol{\mu} \in \mathbb{R}^d$, *dispersion matrix* $\Sigma \in \mathcal{M}_+^{d\times d}$ and a *characteristic generator* $\psi \in \Psi_d$. When an elliptical distribution $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ admits a density function, it is of the form

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{c_d}{\sqrt{|\Sigma|}} g\left(\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}); d\right), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

for some normalizing constant $c_d > 0$ and a *density generator* $g(\cdot; d)$ satisfying

$$\int_0^\infty t^{d/2-1} g(t; d)\, \mathrm{d}t < \infty;$$

see Fang (2018). We omit the second argument and write $g(\cdot) = g(\cdot\,; d)$ when it can be neglected.

In the following proposition we derive the distribution of $\boldsymbol{X}' \mid \{S = K\}$ provided that $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$.

Proposition 6.3.1 (Ellipticality of $\boldsymbol{X}' \mid \{S = K\}$)**.** Suppose $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$. Then $\boldsymbol{X}' \mid \{S = K\}$ follows an elliptical distribution $\mathcal{E}_{d'}(\boldsymbol{\mu}_K, \Sigma_K, \psi_K)$ for some characteristic generator $\psi_K \in \Psi_{d'}$ and

$$\boldsymbol{\mu}_K = \boldsymbol{\mu}' + \frac{K - \mu_S}{\sigma_S^2}(\Sigma \mathbf{1}_d)' \quad \text{and} \quad \Sigma_K = \Sigma' - \frac{1}{\sigma_S^2}(\Sigma \mathbf{1}_d)'(\Sigma \mathbf{1}_d)'^\top, \tag{6.4}$$

where $\boldsymbol{\mu}'$ and $(\Sigma \mathbf{1}_d)'$ are the first d'-components of $\boldsymbol{\mu}$ and $(\Sigma \mathbf{1}_d)$, respectively, Σ' is the principal submatrix of Σ deleting the dth row and column, $\mu_S = \mathbf{1}_d^\top \boldsymbol{\mu}$ and $\sigma_S^2 = \mathbf{1}_d^\top \Sigma \mathbf{1}_d$. Furthermore, if $\boldsymbol{X}$ admits a density with density generator g, then $\boldsymbol{X}' \mid \{S = K\}$ admits a density with density generator

$$g_K(t) = g(t + \Delta_K) \quad \text{where} \quad \Delta_K = \frac{1}{2}\left(\frac{K - \mu_S}{\sigma_S}\right)^2. \tag{6.5}$$

Proof. Notice that $(\boldsymbol{X}', S) = A\boldsymbol{X} \sim \mathcal{E}_d(A\boldsymbol{\mu}, A\Sigma A^\top, \psi)$ where $A = \begin{pmatrix} \boldsymbol{I}_d & \mathbf{0}_d \\ \mathbf{1}_d^\top & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$. Therefore, the conditional distribution $\boldsymbol{X}' \mid \{S = K\}$ also follows an elliptical distribution with the location parameter $\boldsymbol{\mu}_K$ and the dispersion parameter Σ_K as specified in (6.4). The corresponding characteristic generator ψ_K can be specified through Theorem 2.18 of Fang (2018). If $\boldsymbol{X}$ admits a density with density generator g, then

$$f_{\boldsymbol{X}' \mid \{S=K\}}(\boldsymbol{x}') = \frac{f_{(\boldsymbol{X}', S)}(\boldsymbol{x}', K)}{f_S(K)} \propto g_d\left(\frac{1}{2}(\boldsymbol{x}' - \boldsymbol{\mu}', K - \mu_S)^\top \begin{pmatrix} \Sigma' & (\Sigma \mathbf{1}_d)' \\ (\Sigma \mathbf{1}_d)'^\top & \sigma_S^2 \end{pmatrix}^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}', K - \mu_S)\right).$$

The quadratic term reduces to

$$(\boldsymbol{x}' - \boldsymbol{\mu}', K - \mu_S)^\top \begin{pmatrix} \Sigma' & (\Sigma \mathbf{1}_d)' \\ (\Sigma \mathbf{1}_d)'^\top & \sigma_S^2 \end{pmatrix}^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}', K - \mu_S) = (\boldsymbol{x}' - \boldsymbol{\mu}_K)^\top \Sigma_K^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_K) + \frac{(K - \mu_S)^2}{\sigma_S^2}.$$

Therefore, we have that

$$f_{\boldsymbol{X}' \mid \{S=K\}}(\boldsymbol{x}') \propto g\left(\frac{1}{2}(\boldsymbol{x}' - \boldsymbol{\mu}_K)^\top \Sigma_K^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_K) + \Delta_K\right) = g_K\left(\frac{1}{2}(\boldsymbol{x}' - \boldsymbol{\mu}_K)^\top \Sigma_K^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_K)\right),$$

where $\Delta_K = (K - \mu_S)^2/(2\sigma_S^2)$ and $g_K(t) = g(t + \Delta_K)$ as specified in (6.5). □

As seen in the proof, the characteristic generator ψ_K of $\boldsymbol{X}' \mid \{S = K\}$ is in general different from that of $\boldsymbol{X}$. By Proposition 6.3.1, ellipticality is preserved under conditioning $\{S = K\}$ and thus a change of the shape of the distribution as observed in Figure 6.1 (b) does not occur when $\boldsymbol{X}$ is elliptical. The capital K is typically much larger than the mean of the total loss μ_S in practice. By (6.5), the density generator g_K

is thus typically the tail part of the generator g. Moreover, the location vector $\boldsymbol{\mu}_K$ typically increases in proportion to the sum of covariances $(\Sigma \mathbf{1}_d)'$. As a consequence, more (less) capital is assigned to losses which are positively (negatively) correlated with the other losses. On the other hand, the dispersion matrix Σ_K decreases in proportion to the term $(\Sigma \mathbf{1}_d)'(\Sigma \mathbf{1}_d)'^\top$ and the reduction depends on the variance of the sum.

Example 6.3.2 (Student t distribution)**.** A d-dimensional *Student t distribution* $t_\nu(\boldsymbol{\mu}, \Sigma)$ is an elliptical distribution $\mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ with density generator

$$g(t; d) = \left(1 + \frac{t}{\nu}\right)^{-\frac{d+\nu}{2}}, \quad t \geq 0, \tag{6.6}$$

where $\nu \geq 1$ is the *degrees of freedom* parameter. It is known, for example, from Roth (2012) and Ding (2016) that the conditional distribution of the Student t distribution is again Student t. We can check this closedness property with Proposition 6.3.1. By (6.5), the random variable $\boldsymbol{X}' \mid \{S = K\}$ follows an elliptical distribution $\mathcal{E}_{d'}(\boldsymbol{\mu}_K, \Sigma_K, g_K)$ with density generator (up to a constant) given by

$$g_K(t) = \left(1 + \frac{t}{\nu + \Delta_K}\right)^{-\frac{d+\nu}{2}},$$

for which the corresponding distribution is known as the *Pearson type VII distribution*; see Schmidt (2002). In fact, this distribution reduces to a d'-dimensional Student t distribution since

$$g_K(t) = \left(1 + \frac{t}{\nu + \Delta_K}\right)^{-\frac{d+\nu}{2}} \propto \left(1 + \frac{\nu+1}{\nu + \Delta_K}\frac{t}{\nu+1}\right)^{-\frac{d'+\nu+1}{2}},$$

and the multiplier $(\nu + 1)/(\nu + \Delta_K)$ can be absorbed by redefining the dispersion matrix as $\tilde{\Sigma}_K = (\nu + \Delta_K)\Sigma_K/(\nu + 1)$ for $(\nu + \Delta_K)/(\nu + 1) > 0$. Consequently, $\boldsymbol{X}' \mid \{S = K\}$ has distribution $t_{\nu+1}(\boldsymbol{\mu}_K, \tilde{\Sigma}_K)$. Since the degrees of freedom of $\boldsymbol{X}' \mid \{S = K\}$ increases by 1, $\boldsymbol{X}' \mid \{S = K\}$ has slightly lighter tails than $\boldsymbol{X}$.

6.3.3 Dependence and stochastic order

The dependence structure of $\boldsymbol{X}' \mid \{S = K\}$ is typically described in terms of the dependence among X_j and S for $j = 1, \ldots, d'$. For instance, when $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$, Proposition 6.3.1 yields

$$\begin{aligned}
\operatorname{Cov}[X_i, X_j \mid \{S = K\}] = (\Sigma_K)_{i,j} &= \operatorname{Cov}[X_i, X_j] - \frac{1}{\sigma_S^2}(\Sigma \mathbf{1}_d)_i (\Sigma \mathbf{1}_d)_j \\
&= \operatorname{Cov}[X_i, X_j] - \frac{1}{\sigma_S^2} \operatorname{Cov}[X_i, S] \operatorname{Cov}[X_j, S] \\
&= \sigma_i \sigma_j (\rho_{X_i, X_j} - \rho_{X_i, S}\, \rho_{X_j, S}),
\end{aligned}$$

where $\sigma_j^2 = \operatorname{Var}(X_j)$ and ρ_{X_i, X_j} is the correlation coefficient of (X_i, X_j). In this section we study the dependence, especially the total positivity and its related order of $\boldsymbol{X}' \mid \{S = K\}$ for a general distribution beyond the elliptical case. To this end, define the following concepts.

Definition 6.3.3 (Multivariate total positivity of order 2)**.** Suppose random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ have densities $f_{\boldsymbol{X}}$ and $f_{\boldsymbol{Y}}$, respectively.

1. $\boldsymbol{X}$ is said to be *multivariate totally positively ordered of order 2 (MTP2)* if
$$f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{y}) \leq f_{\boldsymbol{X}}(\boldsymbol{x}\wedge\boldsymbol{y})f_{\boldsymbol{X}}(\boldsymbol{x}\vee\boldsymbol{y}), \quad \text{for all } \boldsymbol{x},\ \boldsymbol{y}\in\mathbb{R}^d.$$
2. $\boldsymbol{X}$ is said to be *multivariate reverse rule of order 2 (MRR2)* if
$$f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{y}) \geq f_{\boldsymbol{X}}(\boldsymbol{x}\wedge\boldsymbol{y})f_{\boldsymbol{X}}(\boldsymbol{x}\vee\boldsymbol{y}), \quad \text{for all } \boldsymbol{x},\ \boldsymbol{y}\in\mathbb{R}^d.$$
3. $\boldsymbol{Y}$ is said to be larger than $\boldsymbol{X}$ in $TP2$*-order*, denoted as $\boldsymbol{X}\leq_{tp}\boldsymbol{Y}$ if
$$f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{Y}}(\boldsymbol{y}) \leq f_{\boldsymbol{X}}(\boldsymbol{x}\wedge\boldsymbol{y})f_{\boldsymbol{Y}}(\boldsymbol{x}\vee\boldsymbol{y}), \quad \text{for all } \boldsymbol{x},\ \boldsymbol{y}\in\mathbb{R}^d.$$

For examples and implied dependence properties of MTP2, MRR2 and TP2 ordered distributions, see Karlin and Rinott (1980a) and Karlin and Rinott (1980b). The following proposition states that the MTP2, MRR2 and TP2 order of $\boldsymbol{X}' \mid \{\mathbf{1}_d^\top\boldsymbol{X}=K\}$ and $\boldsymbol{Y}' \mid \{\mathbf{1}_d^\top\boldsymbol{Y}=K\}$ are inherited from those of $(\boldsymbol{X}', \mathbf{1}_d^\top\boldsymbol{X})$ and $(\boldsymbol{Y}', \mathbf{1}_d^\top\boldsymbol{Y})$.

Proposition 6.3.4 (MTP2, MRR2 and TP2 order of $\boldsymbol{X}' \mid \{S=K\}$)**.** Suppose $(\boldsymbol{X}', S)$ and $(\boldsymbol{Y}', T)$ with $S=\mathbf{1}_d^\top\boldsymbol{X}$ and $T=\mathbf{1}_d^\top\boldsymbol{Y}$ have densities $f_{(\boldsymbol{X}',S)}$ and $f_{(\boldsymbol{Y}',T)}$, respectively.

1. If $(\boldsymbol{X}', S)$ is MTP2 (MRR2) then $\boldsymbol{X}' \mid \{S=K\}$ is MTP2 (MRR2).
2. If $(\boldsymbol{X}', S) \leq_{tp} (\boldsymbol{Y}', T)$ then $\boldsymbol{X}' \mid \{S=K\} \leq_{tp} \boldsymbol{Y}' \mid \{T=K\}$.

Proof. By (6.2) we have, for $\boldsymbol{x}',\ \boldsymbol{y}'\in\mathbb{R}^{d'}$, that
$$\begin{aligned} f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}')f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{y}') &= \frac{f_{(\boldsymbol{X}',S)}(\boldsymbol{x}',K)f_{(\boldsymbol{X}',S\ )}(\boldsymbol{y}',K)}{f_S^2(K)} \\ &\leq \frac{f_{(\boldsymbol{X}',S)}(\boldsymbol{x}'\wedge\boldsymbol{y}',K\wedge K)f_{(\boldsymbol{X}',S)}(\boldsymbol{x}'\vee\boldsymbol{y}',K\vee K)}{f_S^2(K)} \\ &= f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}'\wedge\boldsymbol{y}')f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}'\vee\boldsymbol{y}'), \end{aligned}$$
which proves the first part on MTP2. The MRR2 and TP2 parts are shown in a similar manner. □

The properties of MTP2 (MRR2) and TP2 order have various implications. For example, when $\boldsymbol{X}' \mid \{S=K\}$ is MTP2, then $\boldsymbol{X}' \mid \{S=K\}$ is *positively associated* in the sense that $\mathrm{Cov}[g(X_i), h(X_j) \mid \{S=K\}] \geq 0$ for all increasing functions $g:\mathbb{R}\to\mathbb{R}$ and $h:\mathbb{R}\to\mathbb{R}$. If $\boldsymbol{X}' \mid \{S=K\} \leq_{tp} \boldsymbol{Y}' \mid \{T=K\}$, then $\boldsymbol{X}' \mid \{S=K\} \leq_{st} \boldsymbol{Y} \mid \{T=K\}$, that is, $\mathbb{E}[h(\boldsymbol{X}') \mid \{S=K\}] \leq \mathbb{E}[h(\boldsymbol{Y}') \mid \{T=K\}]$ for all bounded and increasing functions $h:\mathbb{R}^{d'}\to\mathbb{R}$. The readers are referred to Müller and Stoyan (2002) for more implications of the MTP2, MRR2 and TP2 order.

Next, we consider the special but important case when $X_1,\dots,X_d$ are perfectly positively dependent, that is, when $\boldsymbol{X}$ is a *comonotone* random vector $\boldsymbol{X}\overset{\mathrm{d}}{=}(F_1^{-1}(U),\dots,F_d^{-1}(U))$ for some $U\sim\mathrm{Unif}(0,1)$. We treat this special case separately since a comonotone random vector does not admit a density. The following proposition states that $\boldsymbol{X} \mid \{S=K\}$ is degenerate when $\boldsymbol{X}$ is comonotone.

Proposition 6.3.5 ($\boldsymbol{X}' \mid \{S = K\}$ under comonotonicity)**.** Suppose $\boldsymbol{X}$ is a comonotone random vector with continuous and strictly increasing margins $F_1, \dots, F_d$. Then

$$\boldsymbol{X} \mid \{S = K\} = (F_1^{-1}(u^*), \dots, F_d^{-1}(u^*)) \quad \mathbb{P}\text{-a.s.},$$

where $u^* \in [0,1]$ is the unique solution to $\sum_{j=1}^d F_j^{-1}(u) = K$ as an equation of $u \in [0,1]$.

Proof. When $\boldsymbol{X}$ has continuous and strictly increasing margins $F_1, \dots, F_d$, then the quantile functions F_j^{-1}, $j = 1, \dots, d$, are continuous and strictly increasing, and thus the equation $\sum_{j=1}^d F_j^{-1}(u) = K$ has a unique solution u^*. Therefore,

$$\mathbb{P}\Big(\bigcup_{j=1}^d \{X_j \neq F_j^{-1}(u^*)\} \Big| \Big\{ S = K \Big\} \Big) = \mathbb{P}\Big(\bigcup_{j=1}^d \{F_j^{-1}(U) \neq F_j^{-1}(u^*)\} \Big| \Big\{ \sum_{j=1}^d F_j^{-1}(U) = K \Big\} \Big) = 0.$$

□

This result can be understood as an extreme case where positive dependence (comonotonicity) implies unimodality of $\boldsymbol{X} \mid \{S = K\}$ (taking on one point $(F_1^{-1}(u^*), \dots, F_d^{-1}(u^*))$ with probability 1). When $\boldsymbol{X}$ has negative dependence, a wider variety of distributions, possibly multimodal ones, arise as $\boldsymbol{X} \mid \{S = K\}$ compared with the positive dependent case; see the following example for the case that negative dependence of $\boldsymbol{X}$ implies multimodality of $\boldsymbol{X} \mid \{S = K\}$.

Example 6.3.6 ($\boldsymbol{X} \mid \{S = K\}$ under extreme negative dependence)**.** Let $K > 0$ and $X \sim F$ for a continuous distribution function F supported on $[0, \infty)$ such that $X \mid \{X \leq K\}$ is radially symmetric about $K/2$ in the sense that $(X - K/2) \mid \{X \leq K\} \overset{\mathrm{d}}{=} (K/2 - X) \mid \{X \leq K\}$. For $U \sim \mathrm{Unif}(0,1)$ define (X_1, X_2) by

$$\begin{aligned} X_1 &= F^{-1}(U)\mathbf{1}_{\{U \leq F(K)\}} + F^{-1}(U)\mathbf{1}_{\{U > F(K)\}} = F^{-1}(U), \\ X_2 &= (K - F^{-1}(U))\mathbf{1}_{\{U \leq F(K)\}} + F^{-1}(U)\mathbf{1}_{\{U > F(K)\}}. \end{aligned}$$

Then $\mathbb{P}(X_1 \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$ for all $x \geq 0$. Moreover, the conditional radial symmetry of F implies that $\mathbb{P}(K - F^{-1}(U) \leq x,\ U \leq F(K)) = \mathbb{P}(F^{-1}(U) \leq x,\ U \leq F(K))$ and thus that

$$\begin{aligned} \mathbb{P}(X_2 \leq x) &= \mathbb{P}(X_2 \leq x,\ U \leq F(K)) + \mathbb{P}(X_2 \leq x,\ U > F(K)) \\ &= \mathbb{P}(K - F^{-1}(U) \leq x,\ U \leq F(K)) + \mathbb{P}(F^{-1}(U) \leq x,\ U > F(K)) \\ &= \mathbb{P}(F^{-1}(U) \leq x,\ U \leq F(K)) + \mathbb{P}(F^{-1}(U) \leq x,\ U > F(K)) \\ &= \mathbb{P}(F^{-1}(U) \leq x) = F(x), \quad x \geq 0. \end{aligned}$$

Therefore, $X_1 \sim F$ and $X_2 \sim F$. The dependence structure of (X_1, X_2) is a combination of positive and negative dependence. The body part $\{X_1 \leq K\}$ of X_1 and the tail part $\{X_2 > K\}$ of X_2 are mutually exclusive in the sense that $\mathbb{P}(X_1 \leq K,\ X_2 > K) = 0$. Similarly $\mathbb{P}(X_1 > K,\ X_2 \leq K) = 0$. In the tail part, X_1 and X_2 are comonotone in the sense that $(X_1, X_2) = (F^{-1}(U), F^{-1}(U))$ on $\{U > F(K)\}$. In the body part, X_1 and X_2 are countermonotone in the sense that $(X_1, X_2) = (F^{-1}(U), K - F^{-1}(U))$ on $\{U \leq F(K)\}$.

Since $X_1 + X_2 = F^{-1}(U) + K - F^{-1}(U) = K$ on $\{U \leq F(K)\}$ and $X_1 + X_2 = 2F^{-1}(U) > 2K > K$ on $\{U > F(K)\}$, we have that

$$\{X_1 + X_2 = K\} = \{X_1 + X_2 = K,\ U \leq F(K)\} \cup \{X_1 + X_2 = K,\ U > F(K)\} = \{U \leq F(K)\}$$

and thus that

$$(X_1, X_2) \mid \{X_1 + X_2 = K\} = (X_1, X_2) \mid \{U \leq F(K)\} = (F^{-1}(U), K - F^{-1}(U)) \mid \{U \leq F(K)\}.$$

Consequently, $(X_1, X_2) \mid \{S = K\}$ has homogeneous marginal distribution $F_{X|\{X \leq K\}}$ and a countermonotone copula W. Therefore, multimodality of $\boldsymbol{X} \mid \{S = K\}$ appears when, for example, $X \sim F$ has a bimodal distribution on the body part $\{X \leq K\}$.

Remark 6.3.7 (Extension with complete mixability)**.** Example 6.3.6 for constructing (X_1, X_2) based on countermonotonicity can be extended to the multivariate case. Let $K > 0$ and $X \sim F$ for a continuous distribution function F supported on $[0, \infty)$ such that the conditional distribution $F_{X|\{X \leq K\}}$ is *d-completely mixable* with center K for $d \geq 3$, that is, there exists a d-dimensional random vector $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ called the *d-complete mix* such that $Y_j \sim F_{X|\{X \leq K\}}$, $j = 1, \dots, d$, and $Y_1 + \cdots + Y_d = K$ a.s. Such a random vector exists, for example, when $F_{X|\{X \leq K\}}$ admits a decreasing density with $\mathbb{E}[Y_1] = K/d$; see Wang and Wang (2011, Corollary 2.9.). Define $\boldsymbol{X} = (X_1, \dots, X_d)$ by $X_j = Y_j \mathbf{1}_{\{U \leq F(K)\}} + Z_j \mathbf{1}_{\{U > F(K)\}}$ for $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ being the d-complete mix of $F_{X|\{X \leq K\}}$, $U \sim \mathrm{Unif}(0, 1)$, $Z_j \sim F_{X|\{X > K\}}$, $j = 1, \dots, d$ and $\boldsymbol{Y}$, U and $Z_1, \dots, Z_d$ are independent of each other. Then one can check that $X_j \sim F$. Moreover, $\{X_1 + \cdots + X_d = K\} = \{U \leq F(K)\}$ since

$$S = X_1 + \cdots + X_d = K \mathbf{1}_{\{U \leq F(K)\}} + (Z_1 + \cdots + Z_d) \mathbf{1}_{\{U > F(K)\}},$$

and $Z_1 + \cdots + Z_d > dK > K$. Consequently, $\boldsymbol{X} \mid \{X_1 + \cdots + X_d = K\} = \boldsymbol{X} \mid \{U \leq F(K)\} = \boldsymbol{Y}$ a.s. and thus $\boldsymbol{X} \mid \{S = K\}$ is the d-complete mix of $X \mid \{X \leq K\}$. To construct a multimodal $\boldsymbol{X} \mid \{S = K\}$ one can choose $\boldsymbol{Y}$ as an equally weighted mixture of three Dirichlet distributions $\mathrm{Dir}(\alpha, \alpha, \beta)$, $\mathrm{Dir}(\alpha, \beta, \alpha)$ and $\mathrm{Dir}(\beta, \alpha, \alpha)$ for $0 < \alpha < \beta$. This mixture is a 3-complete mix since it has homogeneous marginal distributions and a constant sum. Moreover, $\boldsymbol{Y}$ has three distinct modes when, for example, $\alpha = 2$ and $\beta = 10$, and thus $\boldsymbol{X}' \mid \{S = K\}$ is multimodal.

6.3.4 Tail behavior of the conditional distribution

We now study the tail behavior of $\boldsymbol{X}' \mid \{S = K\}$ through its density. Since boundedness of $\boldsymbol{X}$ from below leads to a bounded support of $\boldsymbol{X}' \mid \{S = K\}$ as shown in Section 6.3.1, we focus on the case when $\boldsymbol{X}$ is supported on $\mathbb{R}^d$. In this case, the support of $\boldsymbol{X}' \mid \{S = K\}$ is $\mathbb{R}^{d'}$ and thus there are $2^{d'}$ orthants to be considered. Hereafter we consider tail behavior only in the first orthant $\{\boldsymbol{x}' \in \mathbb{R}^{d'} : x_1, \dots, x_{d'} > 0\}$ since tails on the other orthants can be discussed similarly. We study the following limiting behaviors of the ratio of densities.

Definition 6.3.8 (Multivariate regular and rapid variation of a density)**.** Let $\boldsymbol{X}$ be a d-dimensional random vector $\boldsymbol{X}$ with a density $f_{\boldsymbol{X}}$.

1. $\boldsymbol{X}$ is called *multivariate regularly varying* with limit function $\lambda : \mathbb{R}^{2d} \to \mathbb{R}_+$ (at ∞ and on the first orthant), denoted by MRV(λ) if

$$\lim_{t\to\infty} \frac{f_{\boldsymbol{X}}(t\boldsymbol{y})}{f_{\boldsymbol{X}}(t\boldsymbol{x})} =: \lambda(\boldsymbol{x},\boldsymbol{y}) > 0 \quad \text{for any} \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d_+, \tag{6.7}$$

   provided the limit function λ exists.

2. $\boldsymbol{X}$ is called *multivariate rapidly varying* (at ∞ and on the first orthant), denoted by MRV(∞) if,

$$\lim_{t\to\infty} \frac{f_{\boldsymbol{X}}(st\boldsymbol{x})}{f_{\boldsymbol{X}}(t\boldsymbol{x})} = \begin{cases} 0, & s > 1, \\ \infty, & 0 < s < 1, \end{cases} \quad \text{for any} \quad s > 0,\ \boldsymbol{x} \in \mathbb{R}^d_+.$$

Note that we adopt the definition of regular variation of densities for its potential application to MCMC methods where the ratio of target densities $f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{y}')/f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}')$ at any two points $\boldsymbol{x}',\ \boldsymbol{y}' \in \mathbb{R}^{d'}$ is of interest; see Section 6.5.3. Taking $\boldsymbol{x} = \mathbf{1}_d$ in (6.7) leads to the standard definition of regular variation introduced, for example, in Resnick (2007). Regular variation is typically described in terms of probability measures or survival functions, and these concepts of variations are connected to regular variation of densities through Resnick (2007, Theorem 6.4.).

The following proposition states that one can find a limit function for $\boldsymbol{X}' \mid \{S = K\}$ based on that of $\boldsymbol{X}$ through the auxiliary random vector $\tilde{\boldsymbol{X}} = (\boldsymbol{X}', K - X_d)$.

Proposition 6.3.9 (Multivariate regular and rapid variation of $\boldsymbol{X}' \mid \{S = K\}$)**.**

1. Assume that $\tilde{\boldsymbol{X}} = (\boldsymbol{X}', K - X_d)$ is MRV($\tilde{\lambda}$). Then $\boldsymbol{X}' \mid \{S = K\}$ is MRV(λ') with limit function

$$\lambda'(\boldsymbol{x}',\boldsymbol{y}') = \tilde{\lambda}((\boldsymbol{x}', \mathbf{1}_{d'}^\top \boldsymbol{x}'), (\boldsymbol{y}', \mathbf{1}_{d'}^\top \boldsymbol{y}')), \quad \boldsymbol{x}',\boldsymbol{y}' \in \mathbb{R}^{d'}_+.$$

2. If $\tilde{\boldsymbol{X}}$ is MRV(∞), then $\boldsymbol{X}' \mid \{S = K\}$ is MRV(∞).

Proof. Let $\tilde{\boldsymbol{X}} = (\boldsymbol{X}', K - X_d)$. Since the density of $\tilde{\boldsymbol{X}}$ is written as $f_{\tilde{\boldsymbol{X}}}(x_1,\dots,x_d) = f_{\boldsymbol{X}}(x_1,\dots,x_{d'}, K - x_d)$, we have, by (6.2), that

$$f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') = \frac{f_{\boldsymbol{X}}(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(K)} = \frac{f_{\tilde{\boldsymbol{X}}}(\boldsymbol{x}', \mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(K)}, \quad \boldsymbol{x}' \in \mathbb{R}^{d'}_+.$$

Therefore, if $\tilde{\boldsymbol{X}}$ has a limit function $\tilde{\lambda}$, then the density of $\boldsymbol{X}' \mid \{S = K\}$ satisfies

$$\lim_{t\to\infty} \frac{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{y}')}{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{x}')} = \lim_{t\to\infty} \frac{f_{\tilde{\boldsymbol{X}}}(t\boldsymbol{y}', t\mathbf{1}_{d'}^\top \boldsymbol{y}')}{f_{\tilde{\boldsymbol{X}}}(t\boldsymbol{x}', t\mathbf{1}_{d'}^\top \boldsymbol{x}')} = \tilde{\lambda}((\boldsymbol{x}', \mathbf{1}_{d'}^\top \boldsymbol{x}'), (\boldsymbol{y}', \mathbf{1}_{d'}^\top \boldsymbol{y}')) =: \lambda'(\boldsymbol{x}',\boldsymbol{y}'),$$

for any $\boldsymbol{x}',\ \boldsymbol{y}' \in \mathbb{R}^{d'}_+$ since $(\boldsymbol{x}', \mathbf{1}_{d'}^\top \boldsymbol{x}'),\ (\boldsymbol{y}', \mathbf{1}_{d'}^\top \boldsymbol{y}') \in \mathbb{R}^d_+$. Similarly, if $\tilde{\boldsymbol{X}}$ is MRV(∞), then

$$\lim_{t\to\infty} \frac{f_{\boldsymbol{X}'|\{S=K\}}(st\boldsymbol{x}')}{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{x}')} = \lim_{t\to\infty} \frac{f_{\tilde{\boldsymbol{X}}}(st\boldsymbol{x}', st\mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_{\tilde{\boldsymbol{X}}}(t\boldsymbol{x}', t\mathbf{1}_{d'}^\top \boldsymbol{x}')} = \begin{cases} 0, & s > 1, \\ \infty, & 0 < s < 1, \end{cases}$$

for any $s > 0$ and $\boldsymbol{x}' \in \mathbb{R}^{d'}_+$. □

The sufficient conditions in Proposition 6.3.9 are more straightforward to check than those in Proposition 6.3.4 since $\tilde{\boldsymbol{X}}$ does not depend on the sum S, and the joint distribution of $\tilde{\boldsymbol{X}}$ can be specified through its marginal distributions and copula. The margins of $\tilde{\boldsymbol{X}}$ are $\tilde{F}_j = F_j$, $j = 1, \dots, d'$, and $\tilde{F}_d(x_d) = \bar{F}_d(K - x_d)$, and the copula $\tilde{C}$ of $\tilde{\boldsymbol{X}}$ is the distribution function of $(U_1, \dots, U_{d'}, 1 - U_d)$ where $\boldsymbol{U} \sim C$ is the copula of $\boldsymbol{X}$. This enables one to find a limit function for $\tilde{\boldsymbol{X}}$; see Li (2013), Li and Wu (2013), Li and Hua (2015) and Joe and Li (2019).

As the following proposition shows, in the elliptical case the limit function is determined by the density generator g.

Proposition 6.3.10 (Multivariate regular and rapid variations for elliptical distribution)**.** Assume $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ admits a density with density generator g continuous on $\mathbb{R}_+$.

1. If g is regularly varying in the sense that

$$\lim_{t\to\infty} \frac{g(tu)}{g(ts)} = \lambda_g(s, u), \quad s, u > 0,$$

then $\boldsymbol{X}' \mid \{S = K\}$ is MRV(λ_K) with

$$\lambda_K(\boldsymbol{x}', \boldsymbol{y}') = \lambda_g(\boldsymbol{x}'^{\top}\Sigma_K^{-1}\boldsymbol{x}',\ \boldsymbol{y}'^{\top}\Sigma_K^{-1}\boldsymbol{y}'), \quad \boldsymbol{x}',\ \boldsymbol{y}' \in \mathbb{R}^{d'}.$$

2. If g is rapidly varying in the sense that

$$\lim_{t\to\infty} \frac{g(st)}{g(t)} = \begin{cases} 0, & s > 1, \\ \infty, & 0 < s < 1, \end{cases}$$

then $\boldsymbol{X}' \mid \{S = K\}$ is MRV(∞).

Proof. Proposition 6.3.1 yields that $\boldsymbol{X}' \mid \{S = K\}$ follows a d'-dimensional elliptical distribution with location vector $\boldsymbol{\mu}_K$, dispersion matrix Σ_K and density generator g_K. If g is regularly varying, then

$$\begin{aligned}
\lim_{t\to\infty} \frac{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{y}')}{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{x}')} &= \lim_{t\to\infty} \frac{g_K\left(\frac{1}{2}(t\boldsymbol{y}' - \boldsymbol{\mu}_K)^{\top}\Sigma_K^{-1}(t\boldsymbol{y}' - \boldsymbol{\mu}_K)\right)}{g_K\left(\frac{1}{2}(t\boldsymbol{x}' - \boldsymbol{\mu}_K)^{\top}\Sigma_K^{-1}(t\boldsymbol{x}' - \boldsymbol{\mu}_K)\right)} \\
&= \lim_{t\to\infty} \frac{g\left(\frac{1}{2}t^2(\boldsymbol{y}' - \boldsymbol{\mu}_K/t)^{\top}\Sigma_K^{-1}(\boldsymbol{y}' - \boldsymbol{\mu}_K/t) + \Delta_K\right)}{g\left(\frac{1}{2}t^2(\boldsymbol{x}' - \boldsymbol{\mu}_K/t)^{\top}\Sigma_K^{-1}(\boldsymbol{x}' - \boldsymbol{\mu}_K/t) + \Delta_K\right)} \\
&= \lim_{t\to\infty} \frac{g(\frac{1}{2}t^2\boldsymbol{y}'^{\top}\Sigma_K^{-1}\boldsymbol{y}')}{g(\frac{1}{2}t^2\boldsymbol{x}'^{\top}\Sigma_K^{-1}\boldsymbol{x}')} = \lambda_g(\boldsymbol{x}'^{\top}\Sigma_K^{-1}\boldsymbol{x}',\ \boldsymbol{y}'^{\top}\Sigma_K^{-1}\boldsymbol{y}') = \lambda_K(\boldsymbol{x}', \boldsymbol{y}'),
\end{aligned}$$

for any $\boldsymbol{x}',\ \boldsymbol{y}' \in \mathbb{R}^{d'}$, where the third equality comes from continuity of g and the fourth equality holds since $\boldsymbol{x}'^{\top}\Sigma_K^{-1}\boldsymbol{x}',\ \boldsymbol{y}'^{\top}\Sigma_K^{-1}\boldsymbol{y}' > 0$. Therefore, $\boldsymbol{X}' \mid \{S = K\}$ is MRV(λ_K). For the rapidly varying case,

$$\lim_{t\to\infty} \frac{f_{\boldsymbol{X}'|\{S=K\}}(st\boldsymbol{x}')}{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{x}')} = \lim_{t\to\infty} \frac{g(\frac{1}{2}t^2s^2\boldsymbol{x}'^{\top}\Sigma_K^{-1}\boldsymbol{x}')}{g(\frac{1}{2}t^2\boldsymbol{x}'^{\top}\Sigma_K^{-1}\boldsymbol{x}')} = \begin{cases} 0, & s > 1, \\ \infty, & 0 < s < 1, \end{cases}$$

for any $s > 0$ and $\boldsymbol{x}',\ \boldsymbol{y}' \in \mathbb{R}^{d'}$ since $s > 1$ if and only if $s^2 > 1$ and $0 < s < 1$ if and only if $0 < s^2 < 1$ for $s > 0$. Therefore, $\boldsymbol{X}' \mid \{S = K\}$ is rapidly varying. □

Example 6.3.11 (Normal and Student t distributions)**.** The multivariate Normal distribution has a rapidly varying density generator $g(t) = \exp(-t)$, and thus its corresponding conditional distribution $\boldsymbol{X}' \mid \{S = K\}$ is also rapidly varying by Proposition 6.3.10 Part 2. Next, suppose $\boldsymbol{X}$ follows a d-dimensional Student t distribution with degrees of freedom $\nu \geq 1$. Its density generator (6.6) is regularly varying with limit function

$$\lim_{t\to\infty} \frac{g(tu)}{g(ts)} = \left(\frac{u}{s}\right)^{-\frac{\nu+d}{2}}, \quad u, s > 0.$$

Consequently, by Proposition 6.3.10 Part 1, $\boldsymbol{X}' \mid \{S = K\}$ is regularly varying with the limit function

$$\lim_{t\to\infty} \frac{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{y}')}{f_{\boldsymbol{X}'|\{S=K\}}(t\boldsymbol{x}')} = \left(\frac{||\Sigma_K^{-\frac{1}{2}}\boldsymbol{y}'||}{||\Sigma_K^{-\frac{1}{2}}\boldsymbol{x}'||}\right)^{-(\nu+d)}, \quad \boldsymbol{x}', \boldsymbol{y}' \in \mathbb{R}_+^{d'},$$

where $||\cdot||$ is an Euclidean norm on $\mathbb{R}^{d'}$.

6.3.5 Unimodality of the conditional distributions

Next we study the modality of $\boldsymbol{X}' \mid \{S = K\}$. Among various definitions of unimodality considered in the literature, we adopt those defined based on the superlevel set

$$L_t(f) = \{\boldsymbol{x} \in \mathbb{R}^d : f(\boldsymbol{x}) \geq t\}, \quad t \in (0,\ \max\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}\}],$$

where f is a density on $\mathbb{R}^d$ which is assumed to be bounded for simplicity so that $\max\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}\}$ exists. By definition, $L_t(f)$ is a decreasing set, that is, $L_{t'}(f) \subseteq L_t(f)$ for $0 < t \leq t'$. We also write $L_t(\boldsymbol{X})$ for $L_t(f)$ if $\boldsymbol{X}$ has density f. A set $A \subseteq \mathbb{R}^d$ is called *star-shaped* about $\boldsymbol{x}_0 \in A$ if, for any $\boldsymbol{y} \in A$, the line segment from $\boldsymbol{x}_0$ to $\boldsymbol{y}$ is in A.

Definition 6.3.12 (Concepts of unimodality)**.** For a bounded density function f on $\mathbb{R}^d$, we call $M(f) = L_{t^*}(f)$ the *mode set* for $t^* = \max\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$. If $L_{t^*}(f) = \{\boldsymbol{m}\}$ then we call $\boldsymbol{m} \in \mathbb{R}^d$ the *mode* of f. Furthermore, f is said to be *weakly unimodal* if $L_t(f)$ is connected, *star unimodal* about the center $\boldsymbol{x}_0 \in \mathbb{R}^d$ if $L_t(f)$ is star-shaped about $\boldsymbol{x}_0$ and *convex unimodal* if $L_t(f)$ is convex, for all $0 < t \leq t^*$. Finally, f is said to be *multimodal* if $L_t(f)$ is not connected for some $0 < t \leq t^*$.

From Definition 6.3.12, convex unimodality implies star unimodality and star unimodality implies weak unimodality. Other notions of unimodality, such as block unimodality, linear unimodality, monotone unimodality, α-unimodality, orthounimodality and Khinchin's unimodality are not introduced in this chapter due to their intractability for our purpose; see Dharmadhikari and Joag-Dev (1988) for a comprehensive discussion on unimodality. Defining notions of unimodality in terms of the shape of the superlevel set $L_t(f)$ fits our purpose in several ways. As mentioned in Section 6.2.2, $L_t(\boldsymbol{X})$ can be understood as a plausible scenario set with $t > 0$ being the level of plausibility. In addition, $L_t(\boldsymbol{X} \mid \{S = K\})$ can be regarded as a set of severe and plausible stress scenarios the total capital K is supposed to cover. From these interpretations, we believe that unimodality should describe tractability of these superlevel sets, such as connectivity and convexity. The superlevel set $L_t(f)$ is also important when f is simulated with MCMC methods since the

ratio of levels of f is a primary quantity of interest for such methods. MCMC methods are required to be specifically designed when $L_t(f)$ is not connected since in this case a Markov chain needs to traverse distinct regions to simulate samples from the entire space.

Note that uniqueness of the maximum of a density f, that is, the mode set of f being a singleton $L_{t^*}(f) = \{\boldsymbol{m}\}$ for $\boldsymbol{m} \in \mathbb{R}^d$, is an important but different concept of unimodality from those in Definition 6.3.12. The notions of unimodality in Definition 6.3.12 concern the overall shape of a density through its superlevel sets whereas uniqueness of the maximum of f is a purely analytical property of the derivative of f. In addition, uniqueness of the maximum is not an appropriate concept of unimodality when the relationship between $\boldsymbol{X}$ and $\boldsymbol{X}' \mid \{S = K\}$ is of interest. In fact, uniqueness of the maximum of $f_{\boldsymbol{X}'|\{S=K\}}$ is equivalent to that of $f_{\boldsymbol{X}}$ on the restricted domain $\mathcal{K}_d(K)$ via (6.2), and thus the uniqueness of the maximum of $f_{\boldsymbol{X}}$ on the entire support $\mathbb{R}^d$ does not provide any information on the shape of $f_{\boldsymbol{X}}$ on $\mathcal{K}_d(K)$ unless the mode of $f_{\boldsymbol{X}}$ on $\mathbb{R}^d$ is in $\mathcal{K}_d(K)$.

The following proposition reveals relationships between unimodality of $\boldsymbol{X}$ and that of $\boldsymbol{X}' \mid \{S = K\}$.

Proposition 6.3.13 (Unimodality of $\boldsymbol{X}' \mid \{S = K\}$)**.**

1. Suppose $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ admits a density with density generator g. If g is decreasing on $\mathbb{R}_+$, then $f_{\boldsymbol{X}'|\{S=K\}}$ is convex unimodal. Furthermore, if the equation $g(t) = \Delta_K$ of $t \in \mathbb{R}_+$ has a unique solution t_K^*, then $f_{\boldsymbol{X}'|\{S=K\}}$ has the mode $\boldsymbol{m} = \boldsymbol{\mu}_K$.
2. If $\boldsymbol{X}$ is convex unimodal, then $\boldsymbol{X}' \mid \{S = K\}$ is convex unimodal.

Proof. 1. By Proposition 6.3.1, $\boldsymbol{X}' \mid \{S = K\}$ follows a d'-dimensional elliptical distribution with location vector $\boldsymbol{\mu}_K$, dispersion matrix Σ_K and density generator g_K. Furthermore, g_K is decreasing if g is. Therefore, for $0 < s \le c_K t_K^* / \sqrt{|\Sigma_K|}$,

$$\begin{aligned} L_s(\boldsymbol{X}' \mid \{S = K\}) &= \left\{ \boldsymbol{x}' \in \mathbb{R}^{d'} : g_K\left(\frac{1}{2} (\boldsymbol{x}' - \boldsymbol{\mu}_K)^\top \Sigma_K^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_K) \right) \ge \frac{s\sqrt{|\Sigma_K|}}{c_K} \right\} \\ &= \left\{ \boldsymbol{x}' \in \mathbb{R}^{d'} : 0 \le (\boldsymbol{x}' - \boldsymbol{\mu}_K)^\top \Sigma_K^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_K) \le 2 \left\{ g^{-1}\left(\frac{s\sqrt{|\Sigma_K|}}{c_K} \right) - \Delta_K \right\} \right\}, \end{aligned}$$

which is a convex set with ellipsoid as surface. Moreover, when $s^* = c_K t_K^* / \sqrt{|\Sigma_K|}$, we have

$$L_{s^*}(\boldsymbol{X}' \mid \{S = K\}) = \left\{ \boldsymbol{x}' \in \mathbb{R}^{d'} : (\boldsymbol{x}' - \boldsymbol{\mu}_K)^\top \Sigma_K^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_K) = 0 \right\} = \{\boldsymbol{\mu}_K\}$$

and thus $\boldsymbol{X}' \mid \{S = K\}$ has a mode $\boldsymbol{\mu}_K$.

2. For $t > 0$ and $\boldsymbol{x}' \in \mathbb{R}^{d'}$, we have the equivalence relation:

$$\boldsymbol{x}' \in L_t(\boldsymbol{X}' \mid \{S = K\}) \quad \text{if and only if} \quad (\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}') \in L_{t f_S(K)}(\boldsymbol{X}) \tag{6.8}$$

since $f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') = f_{\boldsymbol{X}}(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}') / f_S(K)$ and thus

$$L_t(\boldsymbol{X}' \mid \{S = K\}) = \{\boldsymbol{x}' \in \mathbb{R}^{d'} : f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') \ge t\} = \{\boldsymbol{x}' \in \mathbb{R}^{d'} : f_{\boldsymbol{X}}(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}') \ge t f_S(K)\}.$$

Suppose $\boldsymbol{x}'$, $\boldsymbol{y}' \in L_t(\boldsymbol{X}' \mid \{S = K\})$. By (6.8), we have that $(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}')$, $(\boldsymbol{y}', K - \mathbf{1}_{d'}^\top \boldsymbol{y}') \in L_{tf_S(K)}(\boldsymbol{X})$. Since $\boldsymbol{X}$ is convex unimodal, $L_{tf_S(K)}(\boldsymbol{X})$ is a convex set. Therefore, we have, for $\theta \in (0, 1)$, that

$$\begin{aligned}\theta(\boldsymbol{x}', K - \mathbf{1}_{d'}\boldsymbol{x}') + (1-\theta)(\boldsymbol{y}', K - \mathbf{1}_{d'}^\top \boldsymbol{y}') &= (\theta\boldsymbol{x}' + (1-\theta)\boldsymbol{y}',\ \theta(K - \mathbf{1}_{d'}^\top \boldsymbol{x}') + (1-\theta)(K - \mathbf{1}_{d'}^\top \boldsymbol{y}') \\ &= (\theta\boldsymbol{x}' + (1-\theta)\boldsymbol{y}',\ K - \mathbf{1}_{d'}^\top(\theta\boldsymbol{x}' + (1-\theta)\boldsymbol{y}')) \in L_{tf_S(K)}(\boldsymbol{X}),\end{aligned}$$

which implies that $\theta\boldsymbol{x}' + (1-\theta)\boldsymbol{y}' \in L_t(\boldsymbol{X}' \mid S = K)$ by (6.8).

□

Unlike convex unimodality, neither weak unimodality nor star unimodality of $\boldsymbol{X}$ imply any of the unimodality concepts introduced in Definition 6.3.12 for $\boldsymbol{X}' \mid \{S = K\}$. To provide a counterexample, we introduce the following class of distributions.

Definition 6.3.14 (Homothetic density)**.** A d-dimensional random vector $\boldsymbol{X}$ is said to have a *homothetic density*, denoted by $\boldsymbol{X} \sim \mathcal{H}(\boldsymbol{\mu}, D, r)$, with a *location parameter* $\boldsymbol{\mu} \in \mathbb{R}^d$, *shape set* $D \subseteq \mathbb{R}^d$ and a *scaling function* $r : \mathbb{R}_+ \to \mathbb{R}_+$ if $\boldsymbol{X} - \boldsymbol{\mu}$ admits a density f_D satisfying

$$L_t(f_D) = r(t)D = \{s\boldsymbol{x} : 0 \le s \le r(t),\ \boldsymbol{x} \in D\}$$

for some continuous and decreasing function r and a bounded and star-shaped (around $\mathbf{0}$) set $D \in \mathbb{R}^d$ such that

$$\int_0^\infty \mathrm{Leb}_d\left(r(t)D\right) \mathrm{d}t = 1, \tag{6.9}$$

where Leb_d denotes the Lebesgue measure on $\mathbb{R}^d$.

Note that Condition (6.9) is required to ensure that $\int_{\mathbb{R}^d} f_D(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$. To see this, we have

$$\begin{aligned}\int_{\mathbb{R}^d} f_D(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_{\mathbb{R}^d}\int_0^{f_D(\boldsymbol{x})} \mathrm{d}t\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d}\int_0^\infty \mathbf{1}_{\{\boldsymbol{x}\in L_t(f_D)\}}\,\mathrm{d}t\,\mathrm{d}\boldsymbol{x} \\ &= \int_0^\infty \mathrm{Leb}_d\left(L_t(f_D)\right)\mathrm{d}t = \int_0^\infty \mathrm{Leb}_d\left(r(t)D\right)\mathrm{d}t = 1.\end{aligned}$$

Homothetic distributions arise partly from l_p-spherical distributions (Osiewalski, 1993) where the superlevel sets are determined as balls in the l_p-norm, and from a further generalized class of distributions called the v-spherical distributions (Fernandez et al., 1995). Examples of homothetic distributions include skew-normal distributions and rotund-exponential distributions; see Balkema and Nolde (2010). It is straightforward to check that $\boldsymbol{X} \sim \mathcal{H}(\mathbf{0}_d, D, r)$ is star unimodal about $\boldsymbol{x}_0 \in \mathbb{R}^d$ if D is star-shaped about $\boldsymbol{x}_0$, and convex unimodal if D is convex.

Suppose $\boldsymbol{X} \sim \mathcal{H}(\mathbf{0}_d, D, r)$ for a convex set D. Then $\boldsymbol{X}$ is convex unimodal and so is $\boldsymbol{X}' \mid \{S = K\}$ by Proposition 6.3.13. For this homothetic distribution, the superlevel set of $\boldsymbol{X}' \mid \{S = K\}$ embedded in $\mathbb{R}^d$

has the following representation

$$\begin{aligned}
&\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}' \in L_t(\boldsymbol{X}' \mid S = K),\ x_d = K - \mathbf{1}_{d'}^\top \boldsymbol{x}'\} \\
&\quad = \{\boldsymbol{x} \in \mathbb{R}^d : f_{\boldsymbol{X}|\{S=K\}}(\boldsymbol{x}') \geq t,\ x_d = K - \mathbf{1}_{d'}^\top \boldsymbol{x}'\} \\
&\quad = \{\boldsymbol{x} \in \mathbb{R}^d : f(\boldsymbol{x}) \geq t f_S(K)\} \cap \mathcal{K}_d(K) \\
&\quad = r(t f_S(K)) D \cap \mathcal{K}_d(K) = \{s\boldsymbol{x} : \boldsymbol{x} \in D,\ 0 \leq s \leq r(t f_S(K))\} \cap \mathcal{K}_d(K) \\
&\quad = \left\{ \frac{K}{\mathbf{1}_d^\top \boldsymbol{x}} \boldsymbol{x} : \boldsymbol{x} \in D,\ 0 \leq \frac{K}{\mathbf{1}_d^\top \boldsymbol{x}} \leq r(t f_S(K)) \right\} \\
&\quad = \left\{ \frac{K}{\mathbf{1}_d^\top \boldsymbol{x}} \boldsymbol{x} : \boldsymbol{x} \in \bigcup_{k \geq K/r(t f_S(K))} D \cap \mathcal{K}_d(k) \right\},
\end{aligned}$$

that is, the superlevel set $L_t(\boldsymbol{X}' \mid \{S = K\})$ embedded in $\mathbb{R}^d$ is a collection of the projected points of $\boldsymbol{x} \in D$ intersected with the upper half space $\{\boldsymbol{x} \in \mathbb{R}^d : \mathbf{1}_d^\top \boldsymbol{x} \geq K/r(t f_S(K))\}$ onto $\mathcal{K}_d(K)$.

The following example shows that neither weak unimodality nor star unimodality of $\boldsymbol{X}$ imply any of the unimodality concepts introduced in Definition 6.3.12 for $\boldsymbol{X}' \mid \{S = K\}$.

Example 6.3.15. Consider $\boldsymbol{X} \in \mathcal{H}(\mathbf{0}_2, D, r)$ where $D = ([-2,2] \times [-1,1]) \cup ([-1,1] \times [-2,2])$ and $r(t) = \frac{1}{2\sqrt{3}} \exp(-t/2)$. D is star-shaped (and thus connected) around $(0,0)$ and r is a decreasing function. Furthermore, the pair of (D, r) satisfies Condition (6.9) since

$$\int_0^\infty \mathrm{Leb}_2\left(r(t)D\right) \mathrm{d}t = \mathrm{Leb}_2\left(D\right) \int_0^\infty r^2(t) \, \mathrm{d}t = 12 \int_0^\infty \frac{1}{12} \exp(-t) \, \mathrm{d}t = 1.$$

Suppose that the total capital is given by $K = 1/3$. For $t = -2\log(\sqrt{3}/3) \approx 1.098$, we have $r(t) = 1/6$ and thus $L_t(f_D) = D/6 = ([-1/3, 1/3] \times [-1/6, 1/6]) \cup ([-1/6, 1/6] \times [-1/3, 1/3])$. Therefore, $L_t(\boldsymbol{X}' \mid \{S = K\}) = [0, 1/6] \cup [1/3, 1/2]$, which is neither star-shaped nor even connected.

Next we study marginal properties of unimodality. In general, even if $\boldsymbol{X}$ is convex unimodal, it does not imply any unimodality for its marginal distributions; see Balkema and Nolde (2010, Example A.3.) for a counterexample. The following example shows that marginal unimodality also does not imply joint unimodality.

Example 6.3.16 (Marginal unimodality does not imply joint unimodality)**.** Consider the following bivariate density

$$f(u,v) = \frac{9}{4} \mathbf{1}_{\left\{(u,v) \in \bigcup_{i=1}^3 [(i-1)/3, i/3]^2\right\}} + \frac{9}{4} \mathbf{1}_{\{(u,v) \in [1/3, 2/3]^2\}}, \quad (u,v) \in [0,1],$$

which has the convex unimodal marginal densities

$$f_1(u) = f_2(u) = \frac{3}{4} \mathbf{1}_{\{u \in [0,1]\}} + \frac{3}{4} \mathbf{1}_{\{u \in [1/3, 2/3]\}}, \quad u \in [0,1].$$

However, $L_{9/4}(f) = [0, 1/3]^2 \cup [1/3, 2/3]^2 \cup [2/3, 1]^2$ is neither convex nor star-shaped.

Joint unimodality implies marginal unimodality for certain classes of distributions. As is shown in Balkema and Nolde (2010), l_p-spherical distributions form a subclass of homothetic densities for which unimodality is preserved under marginalization. This property also holds for the class of s-concave densities, which is also closed under the operation $\boldsymbol{X} \mapsto \boldsymbol{X}' \mid \{S = K\}$; see Section 6.3.6.

6.3.6 Modality and s-concave densities

As we saw in Section 6.3.5, neither joint unimodality nor marginal unimodality imply the other . However, unimodality is preserved under marginalization for some specific class of densities, so-called s-concave densities. In this section we briefly introduce the connection between unimodality and s-concavity of the conditional distribution given a constant sum.

Definition 6.3.17 (s-concavity)**.** For $s \in \mathbb{R}$, a density f on $\mathbb{R}^d$ is called *s-concave* on a convex set $A \subseteq \mathbb{R}^d$ if

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \geq M_s(f(\boldsymbol{x}), f(\boldsymbol{y}); \theta), \quad \boldsymbol{x}, \boldsymbol{y} \in A, \quad \theta \in (0,1),$$

where M_s is called the *generalized mean* defined, by continuity, as

$$M_s(a,b;\theta) = \begin{cases} \{\theta a^s + (1-\theta)b^s\}^{1/s}, & 0 < s < \infty \text{ or } (-\infty < s < 0 \text{ and } ab \neq 0), \\ 0, & -\infty < s < 0 \text{ and } ab = 0, \\ a^\theta b^{1-\theta}, & s = 0, \\ a \wedge b, & s = -\infty, \\ a \vee b, & s = +\infty, \end{cases}$$

for $s \in \mathbb{R}$, $a,\ b \geq 0$ and $\theta \in (0,1)$.

Definition 6.3.17 of s-concavity is based on densities and can be extended to a measure-based definition for distributions that do not admit a density; see Dharmadhikari and Joag-Dev (1988). For $s = -\infty$, s-concavity is also known as *quasi-concavity* and 0-concavity is also known as *log-concavity*. By definition, for $0 < s < \infty$, f is s-concave if and only if f^s is a concave function. As shown in Dharmadhikari and Joag-Dev (1988), the function $s \mapsto M_s(a,b;\theta)$ is increasing for fixed $(a,b;\theta)$. From this we have that t-concavity of f implies s-concavity for $s < t$. Examples of s-concave densities include the skew-normal distribution (Balkema and Nolde, 2010), Wishart distribution, Dirichlet distribution with certain range of parameters (Dharmadhikari and Joag-Dev, 1988) and the uniform distribution on a convex set in $\mathbb{R}^d$ (Norkin and Roenko, 1991).

Convex unimodality (Definition 6.3.12) is related to s-concavity since a density f is convex unimodal if and only if it is $-\infty$-concave (Dharmadhikari and Joag-Dev, 1988). Therefore, f is convex unimodal if it is s-concave for some $s \in \mathbb{R}$. Furthermore, it is straightforward to show that $\boldsymbol{X}' \mid \{S = K\}$ has an s-concave density if $\boldsymbol{X}$ has. As shown in Dharmadhikari and Joag-Dev (1988) and Saumard and Wellner (2014), s-concavity is preserved under marginalization, convolution and weak-limit for certain ranges of $s \in \mathbb{R}$. Therefore, convex unimodality can also be preserved under these operations if the density $f_{\boldsymbol{X}}$ of $\boldsymbol{X}$ is s-concave.

6.4 Maximum likelihood allocation and multimodality adjustment

In this section we investigate how the modality of $\boldsymbol{X} \mid \{S = K\}$ is incorporated in risk management. Under unimodality, the mode of $\boldsymbol{X} \mid \{S = K\}$ is regarded as the most likely stress scenario covered by the given total capital K. This mode is defined to be a maximum likelihood allocation in Section 6.4.1, and its properties are studied in Section 6.4.2. Under multimodality of $\boldsymbol{X} \mid \{S = K\}$, a single vector of allocations may not be reliable as seen in Section 6.2.2. To overcome this issue, in Section 6.4.4 we consider how to utilize knowledge on multimodality of $\boldsymbol{X} \mid \{S = K\}$ to increase the soundness of risk allocations.

6.4.1 Definition and assumptions

For notational convenience we denote by $\mathcal{U}_d(K)$ the set of all d-dimensional random vectors $\boldsymbol{X}$ such that $\boldsymbol{X}$ and $(\boldsymbol{X}', S)$ admit density functions, and $\boldsymbol{x} \mapsto f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}_d(K)\}}$ has a unique maximum. For $\boldsymbol{X} \in \mathcal{U}_d(K)$, $\boldsymbol{X}' \mid \{S = K\}$ admits a density through (6.2), and $f_{\boldsymbol{X}'|\{S=K\}}$ has a unique maximum attained by the mode of $\boldsymbol{X}' \mid \{S = K\}$. By Proposition 6.3.13, elliptical random vectors with continuous and decreasing density generators form a subclass of $\mathcal{U}_d(K)$. Although some exchangeable random vectors possessing negative dependence, such as Model (b) in Section 6.2.2, may not be included in $\mathcal{U}_d(K)$, we believe that most loss models used in risk management practice are contained in $\mathcal{U}_d(K)$. As explained in Section 6.3.5, uniqueness of the mode of $\boldsymbol{X}' \mid \{S = K\}$ and its unimodality are different concepts, and thus the class $\mathcal{U}_d(K)$ contains multimodal random vectors in the sense that the density $f_{\boldsymbol{X}'|\{S=K\}}$ has multiple local maximizers (we call them the *local modes* of $\boldsymbol{X}' \mid \{S = K\}$). Nevertheless, in this section we primarily consider the MLA to be applied to the unimodal distributions, and we solely focus on the unique global maximizer of $f_{\boldsymbol{X}'|\{S=K\}}$ (not on local ones). As we emphasized in Section 6.2.2, multimodal case should be treat with care, and this case will be further considered in Section 6.4.4. As we will demonstrate in Section 6.5.2, multimodality can be detected by searching for the modes of $f_{\boldsymbol{X}'|\{S=K\}}$.

In the following we define the unique mode of $\boldsymbol{X}' \mid \{S = K\}$ as a risk allocation of K.

Definition 6.4.1 (Maximum likelihood allocation)**.** For $K > 0$ and $\boldsymbol{X} \in \mathcal{U}_d(K)$, the *maximum likelihood allocation (MLA)* on a set $\mathcal{K} \subseteq \mathcal{K}_d(K)$ is defined by

$$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}] = \operatorname{argmax}\{f_{\boldsymbol{X}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{K}\},$$

provided the function $\boldsymbol{x} \mapsto f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}\}}$ has a unique maximum. When $\mathcal{K} = \mathcal{K}_d(K)$, we call it the maximum likelihood allocation.

By (6.2), MLA of K on $\mathcal{K}$ can be equivalently formulated as

$$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}] = \operatorname{argmax}\{f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') : (\boldsymbol{x}', K - \mathbf{1}_{d'}^{\top}\boldsymbol{x}') \in \mathcal{K}\}.$$

By definition, MLA on $\mathcal{K} \subseteq \mathcal{K}_d(K)$ is an allocation of K in the sense that it satisfies the full allocation property $\mathbf{1}_d^{\top}\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}] = K$. We mainly study the case when $\mathcal{K} = \mathcal{K}_d(K)$. However, as we will see in Sections 6.4.2 and 6.5.3, the set $\mathcal{K}$ can be taken so that $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}]$ satisfies some desirable properties for a risk allocation principle.

6.4.2 Properties of MLA

We now investigate properties of MLA as a risk allocation principle; for desirable properties of risk allocation in the case when the capital K is exogenously given as a constant, see Maume-Deschamps et al. (2016). By construction, $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}]$ always satisfies the full allocation property (1.3). The following proposition summarizes other desirable properties of MLA.

Proposition 6.4.2 (Properties of MLA)**.** Suppose $K > 0$ and $\boldsymbol{X} \in \mathcal{U}_d(K)$.

1. *Translation invariance*: $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}+\boldsymbol{c};\ \mathcal{K}_d(K+\mathbf{1}_d^\top \boldsymbol{c})] = \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)] + \boldsymbol{c}$ for $\boldsymbol{c} \in \mathbb{R}^d$.
2. *Positive homogeneity*: $\boldsymbol{K}_{\mathrm{M}}[c\boldsymbol{X};\mathcal{K}_d(cK)] = c\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]$ for $c > 0$.
3. *Symmetry*: For $(i,j) \in \{1,\dots,d\}$, $i \neq j$, let $\tilde{\boldsymbol{X}}$ be a d-dimensional random vector such that $\tilde{X}_j = X_i$, $\tilde{X}_i = X_j$ and $\tilde{X}_k = X_k$, $k \in \{1,\dots,d\}\backslash\{i,j\}$. If $\boldsymbol{X} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{X}}$, then $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]_i = \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]_j$, where $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]_l$ is the lth component of $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]$ for $l = 1,\dots,d$.
4. *Continuity*: Suppose $\boldsymbol{X}_n,\ \boldsymbol{X} \in \mathcal{U}_d(K)$ have densities f_n and f for $n = 1,2,\dots,$ respectively. If f_n is uniformly continuous and bounded for $n = 1,2,\dots,$ and $\boldsymbol{X}_n \to \boldsymbol{X}$ weakly, then $\lim_{n\to\infty} \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}_n;\mathcal{K}_d(K)]$ $= \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]$.

Proof. 1. *Translation invariance*: Let $\tilde{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{c}$, $\tilde{S} = S + \mathbf{1}_d^\top \boldsymbol{c}$ and $\tilde{K} = K + \mathbf{1}_d^\top \boldsymbol{c}$. Since $f_{\boldsymbol{X}+\boldsymbol{c}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{c})$, we have that

$$f_{\tilde{\boldsymbol{X}}'|\{\tilde{S}=\tilde{K}\}}(\tilde{\boldsymbol{x}}') = \frac{f_{(\tilde{\boldsymbol{X}},\tilde{S})}(\tilde{\boldsymbol{x}}',\tilde{K})}{f_{\tilde{S}}(\tilde{K})} = \frac{f_{(\boldsymbol{X}',S)}(\tilde{\boldsymbol{x}}'-\boldsymbol{c}',K)}{f_S(K)} = f_{\boldsymbol{X}'|\{S=K\}}(\tilde{\boldsymbol{x}}'-\boldsymbol{c}').$$

Therefore, uniqueness of the maximizer of $f_{\boldsymbol{X}'|\{S=K\}}$ implies that of $f_{\tilde{\boldsymbol{X}}'|\{\tilde{S}=\tilde{K}\}}$, and these maximizers are related via $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}+\boldsymbol{c};\ \mathcal{K}_d(K+\mathbf{1}_d^\top \boldsymbol{c})] = \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)] + \boldsymbol{c}$.

2. *Positive homogeneity*: Let $\tilde{\boldsymbol{X}} = c\boldsymbol{X}$, $\tilde{S} = cS$ and $\tilde{K} = cK$. Since $f_{c\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}/c)$, we have that

$$f_{\tilde{\boldsymbol{X}}'|\{\tilde{S}=\tilde{K}\}}(\tilde{\boldsymbol{x}}') = \frac{f_{(\tilde{\boldsymbol{X}},\tilde{S})}(\tilde{\boldsymbol{x}}',\tilde{K})}{f_{\tilde{S}}(\tilde{K})} = \frac{f_{(\boldsymbol{X}',S)}(\tilde{\boldsymbol{x}}'/c,K)}{f_S(K)} = f_{\boldsymbol{X}'|\{S=K\}}(\tilde{\boldsymbol{x}}'/c).$$

As seen in the case of translation invariance, this equality implies that $\tilde{\boldsymbol{X}} \in \mathcal{U}_d(\tilde{K})$ and $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(cK)] = c\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X};\mathcal{K}_d(K)]$.

3. *Symmetry*: Without loss of generality, consider $i = 1$ and $j = 2$. Let $\tilde{\boldsymbol{X}} = (X_2, X_1, \boldsymbol{X}_{-(1,2)})$ and $\tilde{S} = \mathbf{1}_d^\top \tilde{\boldsymbol{X}}$, where $\boldsymbol{x}_{-(1,2)}$ is a shorthand for $(x_3,\dots,x_d)$ for $\boldsymbol{x} \in \mathbb{R}^d$. Then $f_{\tilde{\boldsymbol{X}}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\tilde{\boldsymbol{x}})$ for $\boldsymbol{x} = (x_1,x_2,\boldsymbol{x}_{-(1,2)}) \in \mathbb{R}^d$ and $\tilde{\boldsymbol{x}} = (x_2,x_1,\boldsymbol{x}_{-(1,2)}) \in \mathbb{R}^d$. Moreover, when $\boldsymbol{X} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{X}}$, we have $\tilde{\boldsymbol{X}} \in \mathcal{U}_d(K)$ and $f_{\boldsymbol{X}} = f_{\tilde{\boldsymbol{X}}}$. Consequently, we have that

$$f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}') = \frac{f_{\boldsymbol{X}}(\boldsymbol{x}',K-\mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(K)} = \frac{f_{\tilde{\boldsymbol{X}}}(\boldsymbol{x}',K-\mathbf{1}_{d'}^\top \boldsymbol{x}')}{f_S(K)} = \frac{f_{\boldsymbol{X}}(\tilde{\boldsymbol{x}}',K-\mathbf{1}_{d'}^\top \tilde{\boldsymbol{x}}')}{f_S(K)} = f_{\boldsymbol{X}'|\{S=K\}}(\tilde{\boldsymbol{x}}'), \tag{6.10}$$

where the third equation holds since $\mathbf{1}_{d'}^\top \boldsymbol{x}' = \mathbf{1}_{d'}^\top \tilde{\boldsymbol{x}}'$. Now suppose $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_1 \neq \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_2$. Then two distinct vectors $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]$ and $(\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_2,\ \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_1, \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_{-(1,2)})$ attain the maximum of $f_{\boldsymbol{X}'|\{S=K\}}$ by (6.10). Since $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]$ is obtained by the unique maximizer of $f_{\boldsymbol{X}'|\{S=K\}}(\boldsymbol{x}')$, this leads to a contradiction.

4. *Continuity*: When f_n is uniformly continuous and bounded for $n = 1, 2, \dots$, the sequence (f_n) is asymptotically uniformly equicontinuous and bounded in the sense introduced in Sweeting et al. (1986). Together with the assumption that $\boldsymbol{X}_n \to \boldsymbol{X}$ weakly, Theorem 2 of Sweeting et al. (1986) implies that $f_n \to f$ pointwise and uniformly in $\mathbb{R}^d$ for the uniformly continuous density f of $\boldsymbol{X}$. Define $g_n(\boldsymbol{x}') = f_n(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}')$ for $n = 1, 2, \dots$ and $g(\boldsymbol{x}') = f(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}')$, $\boldsymbol{x}' \in \mathbb{R}^{d'}$. By (6.2) and since $\boldsymbol{X}_n,\ \boldsymbol{X} \in \mathcal{U}_d(K)$, the maximizers of g_n and g are uniquely determined. Denote them as $\boldsymbol{x}_n^* = \operatorname*{argmax}_{\boldsymbol{x} \in \mathbb{R}^{d'}} g_n(\boldsymbol{x})$ and $\boldsymbol{x}^* = \operatorname*{argmax}_{\boldsymbol{x} \in \mathbb{R}^{d'}} g(\boldsymbol{x})$. By definition of $\boldsymbol{x}_n^*$, we have that

$$g_n(\boldsymbol{x}_n^*) \geq g_n(\boldsymbol{x}) \quad \text{for any} \quad \boldsymbol{x} \in \mathbb{R}^{d'}.$$

Since g_n converges uniformly to g, it holds that

$$g(\limsup_{n\to\infty} \boldsymbol{x}_n^*) \geq g(\boldsymbol{x}) \quad \text{and} \quad g(\liminf_{n\to\infty} \boldsymbol{x}_n^*) \geq g(\boldsymbol{x}) \quad \text{for any} \quad \boldsymbol{x} \in \mathbb{R}^{d'}.$$

If $\limsup_{n\to\infty} \boldsymbol{x}_n^* > \liminf_{n\to\infty} \boldsymbol{x}_n^*$, then two points attain the maximum of g, which contradicts the uniqueness of the maximizer of g. As a consequence, $\limsup_{n\to\infty} \boldsymbol{x}_n^* = \liminf_{n\to\infty} \boldsymbol{x}_n^* = \lim_{n\to\infty} \boldsymbol{x}_n^* = \boldsymbol{x}^*$ and thus $\lim_{n\to\infty} \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}_n; \mathcal{K}_d(K)] = \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]$.

□

Translation invariance states that a sure loss $\boldsymbol{c} \in \mathbb{R}^d$ requires the same amount of risk allocation and the rest of the total capital is allocated to the random loss $\boldsymbol{X}$. Positive homogeneity means that, for a proportion $c > 0$, $100c\%$ of the loss $\boldsymbol{X}$ requires $100c\%$ of the total capital K and the resulting MLA of $c\boldsymbol{X}$ is $100c\%$ of the allocation derived based on $\boldsymbol{X}$ and K. Symmetry implies that, if exchanging two marginal losses does not change the distribution of the joint loss, then equal amounts of capitals are allocated to them. Finally, continuity ensures that if MLA is calculated based on an estimated model f_n of f, then this estimate of MLA is close to the true MLA as long as f_n correctly estimates f. Note that the assumptions that $\boldsymbol{X}_n$, $n = 1, 2, \dots$, and $\boldsymbol{X}$ belong to $\mathcal{U}_d(K)$ are esssential so that the MLAs of $\boldsymbol{X}_n$, $n = 1, 2, \dots$, and $\boldsymbol{X}$ are well-defined.

Next we cover properties that need to be considered separately.

1. *RORAC compatibility and core compatibility*:

   RORAC compatibility and *core compatibility* are important properties of risk allocations since either of them characterizes Euler allocation; see Tasche (1995) and Denault (2001). However, the definitions of these properties are not meaningful when K is exogenously given as a constant. Moreover, similar constraints as in core compatibility can be additionally imposed on $\mathcal{K}_d(K)$ so that the resulting MLA satisfies desirable core properties; see Section 6.5.3 for details.

2. *Riskless asset*:

   The *riskless asset* condition requires the sure loss $X_j = c_j$ a.s. for $c_j \in \mathbb{R}$ to be covered by the amount of allocated capital c_j. This property needs to be considered separately since in this case $\boldsymbol{X}$ does not admit a density. Suppose that $X_j = c_j \in \mathbb{R}$ a.s. for $j \in I \subseteq \{1, \dots, d\}$ and that $\boldsymbol{X}_{-I} = (X_j, j \in \{1, \dots, d\} \backslash I)$ admits a density $f_{\boldsymbol{X}_{-I}}$. Since

$$(\boldsymbol{X}_I, \boldsymbol{X}_{-I}) \mid \{S = K\} \overset{\mathrm{d}}{=} (\boldsymbol{c}, \boldsymbol{X}_{-I}) \mid \{\mathbf{1}_{|-I|}^\top \boldsymbol{X}_{-I} = K - \mathbf{1}_{|I|}^\top \boldsymbol{c}\} \overset{\mathrm{d}}{=} (\boldsymbol{c}, \boldsymbol{X}_{-I} \mid \{\mathbf{1}_{|-I|}^\top \boldsymbol{X}_{-I} = K - \mathbf{1}_{|I|}^\top \boldsymbol{c}\}), \tag{6.11}$$

   any realization $\boldsymbol{x}$ of $\boldsymbol{X} \mid \{S = K\}$ satisfies $\boldsymbol{x}_I = \boldsymbol{c}$ and the likelihood of $\boldsymbol{x}$ is quantified through the density $f_{\boldsymbol{X}_{-I} \mid \{\mathbf{1}_{|-I|}^\top \boldsymbol{X}_{-I} = K - \mathbf{1}_{|I|}^\top \boldsymbol{c}\}}(\boldsymbol{x}_{-I})$. According to this discussion, a natural extension of the definition of MLA to such a random vector $\boldsymbol{X}$ is

$$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_I = \boldsymbol{c}, \quad \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]_{-I} = \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}_{-I}; \mathcal{K}_{|-I|}(K - \mathbf{1}_{|I|}^\top \boldsymbol{c})], \tag{6.12}$$

   which is compatible with the riskless asset property.

3. *Allocation under comonotonicity*:

   Suppose $\boldsymbol{X}$ is a comonotone random vector with continuous margins $F_1, \dots, F_d$. By Proposition 6.3.5 $\boldsymbol{X} \mid \{S = K\} = (F_1^{-1}(u^*), \dots, F_d^{-1}(u^*))$ a.s., where $u^* \in [0, 1]$ is the unique solution of $\sum_{j=1}^d F_j^{-1}(u) = K$. According to the extended definition (6.12) we have that

$$\boldsymbol{K}_{\mathrm{M}}(\boldsymbol{X}; \mathcal{K}_d(K)) = (F_1^{-1}(u^*), \dots, F_d^{-1}(u^*)).$$

Fallacies in risk allocations

In this section we introduce two properties which intuitively hold but in general do not for the Euler and maximum likelihood allocations. For a d-dimensional random vector $\boldsymbol{X}$ and a real number $K \in \mathbb{R}$, an allocation principle $\boldsymbol{K}$ maps $(\boldsymbol{X}, K)$ to $\boldsymbol{K}(\boldsymbol{X}; K) \in \mathbb{R}^d$ such that $\mathbf{1}_d^\top \boldsymbol{K}(\boldsymbol{X}; K) = K$.

1. *Invariance under independence*:

   For two integers $d, \tilde{d} \geq 2$, consider a d-dimensional random vector $\boldsymbol{X}$ with $S = \mathbf{1}_d^\top \boldsymbol{X}$ and a $\tilde{d}$-dimensional random vector $\tilde{\boldsymbol{X}}$ with $\tilde{S} = \mathbf{1}_{\tilde{d}}^\top \tilde{\boldsymbol{X}}$. For $K, \tilde{K} > 0$, we call a risk allocation *invariant under independence* if

$$\boldsymbol{K}((\boldsymbol{X}, \tilde{\boldsymbol{X}}); K + \tilde{K}) = (\boldsymbol{K}(\boldsymbol{X}; K), \boldsymbol{K}(\tilde{\boldsymbol{X}}; \tilde{K}))$$

   provided that $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ are independent of each other. This property means that risk allocation problems of multiple portfolios independent of each other can be considered separately. Unfortunately, this property does not hold for MLA since

$$\begin{aligned} f_{(\boldsymbol{X}, \tilde{\boldsymbol{X}}) \mid \{S + \tilde{S} = K + \tilde{K}\}}((\boldsymbol{x}, \tilde{\boldsymbol{x}})) &= \frac{f_{(\boldsymbol{X}, \tilde{\boldsymbol{X}})}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \mathbf{1}_{\{\mathbf{1}_d^\top \boldsymbol{x} + \mathbf{1}_{\tilde{d}}^\top \tilde{\boldsymbol{x}} = K + \tilde{K}\}}}{f_{S + \tilde{S}}(K + \tilde{K})} = \frac{f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\tilde{\boldsymbol{X}}}(\tilde{\boldsymbol{x}}) \mathbf{1}_{\{\mathbf{1}_d^\top \boldsymbol{x} + \mathbf{1}_{\tilde{d}}^\top \tilde{\boldsymbol{x}} = K + \tilde{K}\}}}{f_{S + \tilde{S}}(K + \tilde{K})} \\ &\propto f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\tilde{\boldsymbol{X}}}(\tilde{\boldsymbol{x}}) \mathbf{1}_{\{\bigcup_{\{(k, \tilde{k}) \in \mathbb{R}^2, \ k + \tilde{k} = K + \tilde{K}\}} \{\mathbf{1}_d^\top \boldsymbol{x} = k\} \cap \{\mathbf{1}_{\tilde{d}} \tilde{\boldsymbol{x}} = \tilde{k}\}\}} \end{aligned} \tag{6.13}$$

and

$$f_{\boldsymbol{X}|\{S=K\}}(\boldsymbol{x})f_{\tilde{\boldsymbol{X}}|\{\tilde{S}=\tilde{K}\}}(\tilde{\boldsymbol{x}}) \propto f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\mathbf{1}_d^\top \boldsymbol{x}=K\}}f_{\tilde{\boldsymbol{X}}}(\tilde{\boldsymbol{x}})\mathbf{1}_{\{\mathbf{1}_{d'}\tilde{\boldsymbol{x}}=\tilde{K}\}} \tag{6.14}$$

are in general not equal (up to a constant). For example, let $d = d'$ and $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ be two independent and identically distributed standard normal distributions. Then the maximum of (6.13) is attained at $(K + \tilde{K})\mathbf{1}_{2d}/2d$ whereas that of (6.14) is attained at $(K\mathbf{1}_d/d, \tilde{K}\mathbf{1}_d/d)$. The two vectors are not equal unless $K = \tilde{K}$. In this example, the Euler allocations provide the same allocated capitals as MLA, and thus invariance under independence does not hold for Euler allocation either.

2. *Additivity under convolution*:

   Consider two independent d-dimensional random vectors $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ with $S = \mathbf{1}_d^\top \boldsymbol{X}$ and $\tilde{S} = \mathbf{1}_d^\top \tilde{\boldsymbol{X}}$. For $K,\ \tilde{K} > 0$, we call an allocation $\boldsymbol{K}$ *additive under convolution* if

$$\boldsymbol{K}(\boldsymbol{X} + \tilde{\boldsymbol{X}}; K + \tilde{K}) = \boldsymbol{K}(\boldsymbol{X}; K) + \boldsymbol{K}(\tilde{\boldsymbol{X}}; \tilde{K}),$$

   which means that risk allocations of the sum of independent portfolios are calculated as the sum of the allocations of each portfolios. This property neither holds for Euler allocation, nor for MLA. For example, let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and $\tilde{\boldsymbol{X}} \sim \mathrm{N}_d(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$ be two independent normal random vectors for $\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}} \in \mathbb{R}^d$ and $\Sigma, \tilde{\Sigma} \in \mathcal{M}_+^{d\times d}$. By Proposition 6.3.1, Equation (6.4), and Proposition 6.3.13, Part 1, we have that

$$\boldsymbol{K}_{\mathrm{M}}(\boldsymbol{X}; \mathcal{K}_d(K)) = \boldsymbol{\mu}' + \frac{K - \mu_S}{\sigma_S^2}(\Sigma\mathbf{1}_d)' \quad \text{and} \quad \boldsymbol{K}_{\mathrm{M}}(\tilde{\boldsymbol{X}}; \mathcal{K}_d(\tilde{K})) = \tilde{\boldsymbol{\mu}}' + \frac{\tilde{K} - \mu_{\tilde{S}}}{\sigma_{\tilde{S}}^2}(\tilde{\Sigma}\mathbf{1}_d)'.$$

   Similarly, since $\boldsymbol{X} + \tilde{\boldsymbol{X}} \sim \mathrm{N}_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma + \tilde{\Sigma})$, we have that $\sigma_{S+\tilde{S}}^2 = \sigma_S^2 + \sigma_{\tilde{S}}^2$ and that

$$\begin{aligned}\boldsymbol{K}_{\mathrm{M}}(\boldsymbol{X} + \tilde{\boldsymbol{X}}; \mathcal{K}_d(K + \tilde{K})) &= \boldsymbol{\mu}' + \tilde{\boldsymbol{\mu}}' + \frac{K + \tilde{K} - (\mu_S + \mu_{\tilde{S}})}{\sigma_S^2 + \sigma_{\tilde{S}}^2}((\Sigma + \tilde{\Sigma})\mathbf{1}_d)' \\ &= \boldsymbol{\mu}' + \tilde{\boldsymbol{\mu}}' + \left(\frac{\sigma_S^2}{\sigma_S^2 + \sigma_{\tilde{S}}^2}\frac{K - \mu_S}{\sigma_S^2} + \frac{\sigma_{\tilde{S}}^2}{\sigma_S^2 + \sigma_{\tilde{S}}^2}\frac{\tilde{K} - \mu_{\tilde{S}}}{\sigma_{\tilde{S}}^2}\right)((\Sigma + \tilde{\Sigma})\mathbf{1}_d)',\end{aligned}$$

   which is not equal to $\boldsymbol{K}_{\mathrm{M}}(\boldsymbol{X}; \mathcal{K}_d(K)) + \boldsymbol{K}_{\mathrm{M}}(\tilde{\boldsymbol{X}}; \mathcal{K}_d(\tilde{K}))$ unless, for instance, $\Sigma = \tilde{\Sigma}$. Since Euler and maximum likelihood allocations coincide under ellipticality, the same statement holds for Euler allocations.

6.4.3 Discussion on MLA

We now discuss how suitable MLA is as a risk allocation principle and compare Euler and maximum likelihood allocations. Here we define Euler allocation by $\mathbb{E}[\boldsymbol{X} \mid \{S = K\}]$, which are the VaR contributions (1.5) with $K = \mathrm{VaR}_p(S)$ for some confidence level $p \in (0, 1)$. As shown in Proposition 6.4.2, MLA possesses properties naturally required as an allocation such as translation invariance, positive homogeneity and riskless asset. Euler allocation also satisfies these properties since $\mathbb{E}[\boldsymbol{X} + \boldsymbol{c} \mid \{\mathbf{1}_d^\top(\boldsymbol{X} + \boldsymbol{c}) = K + \mathbf{1}_d^\top \boldsymbol{c}\}] = \mathbb{E}[\boldsymbol{X} \mid \{\mathbf{1}_d^\top \boldsymbol{X} = K\}] + \boldsymbol{c}$ for $\boldsymbol{c} \in \mathbb{R}^d$ (translation invariance),

$\mathbb{E}[c\boldsymbol{X} \mid \{\mathbf{1}_d^\top(c\boldsymbol{X}) = cK\}] = c\mathbb{E}[\boldsymbol{X} \mid \{\mathbf{1}_d^\top \boldsymbol{X} = K\}]$ for $c > 0$ (positive homogeneity) and the riskless asset property holds by taking expectation on the both sides of the first equality in (6.11). Note that by Proposition 6.3.1 and Proposition 6.3.13 Part 1, Euler and maximum likelihood allocations coincide when $\boldsymbol{X}$ is elliptically distributed. Therefore, the economic justifications of Euler allocation, such as RORAC compatibility and core-compatibility, also holds for MLA when $\boldsymbol{X}$ is elliptical. Moreover, through the process of estimating MLA, one can detect multimodality of $\boldsymbol{X}' \mid \{S = K\}$ and discover hidden risky scenarios based on which one can evaluate the soundness of risk allocations. On the other hand, the main disadvantage of MLA compared with Euler allocation is that estimating modes becomes more difficult than estimating a mean as the dimension of the portfolio becomes larger. Furthermore, MLA is not well-defined for distributions whose $\operatorname{argmax}\{f_{\boldsymbol{X}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{K}_d(K)\}$ is not a single point. Summarizing these aspects, we believe that MLA and the procedure for searching for (local) modes of $\boldsymbol{X}' \mid \{S = K\}$ are best suited for discovering hidden multiple scenarios likely to occur in the stressed situation $\{S = K\}$, for assessing the soundness of risk allocations in stress testing applications, and eventually for constructing more sound risk allocations based on the multiple scenarios as we will consider in Section 6.4.4.

6.4.4 Multimodality adjustment of risk allocations

Although MLA is investigated in Sections 6.4.1 and 6.4.2, MLA itself does not overcome the issue of multimodality. In this section, we discuss how to utilize local modes of $\boldsymbol{X} \mid \{S = K\}$ discovered in the process of estimating MLAs, and introduce the so-called multimodality adjustment to increase the soundness of capital allocations under multimodality.

To this end, suppose that $M \in \mathbb{N}$ number of scenarios $\boldsymbol{K}_1, \dots, \boldsymbol{K}_M \in \mathcal{K}_d(K)$ are found with corresponding probability weights $w_1, \dots, w_M \in [0, 1]$ such that $\sum_{m=1}^M w_m = 1$. Then multimodality adjustment is defined as follows.

Definition 6.4.3 (Multimodality adjustment of risk allocations)**.** Let $M \in \mathbb{N}$ be the number of scenarios, $\mathcal{X} = \{\boldsymbol{K}_1, \dots, \boldsymbol{K}_M\}$ be the set of scenarios where $\boldsymbol{K}_m \neq \boldsymbol{K}_{m'}$ for any $m,\ m' \in \{1, \dots, M\}$ such that $m \neq m'$, and $\boldsymbol{w} = (w_1, \dots, w_M)$ be the associated probability weights such that $\sum_{m=1}^M w_m = 1$. Then the *multimodality-adjusted allocated capital* is defined by

$$\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda} = \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}} + \sum_{m=1}^M w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}})^+, \tag{6.15}$$

where $\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}} = \sum_{m=1}^M w_m \boldsymbol{K}_m$ is the *baseline allocation*, $\Lambda = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_M) \in \mathbb{R}_+^{d \times M}$ is the matrix of *multimodality loading parameters*, $\boldsymbol{x} \circ \boldsymbol{y} = (x_1 y_1, \dots, x_d y_d)$ for $\boldsymbol{x},\ \boldsymbol{y} \in \mathbb{R}^d$ and $\boldsymbol{x}^+ = (\max(x_1, 0), \dots, \max(x_d, 0))$ for $\boldsymbol{x} \in \mathbb{R}^d$. We call the second term $\sum_{m=1}^M w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}})^+$ of (6.15) the *multimodality adjustment*.

The multimodality-adjusted allocated capital (6.15) consists of the baseline allocation and the additional loading to cover instability due to the existence of multiple scenarios. The probability weight w_m typically represents the likelihood of the scenario $\boldsymbol{K}_m$ to occur, and thus a reasonable choice is $w_m \propto f_{\boldsymbol{X}}(\boldsymbol{K}_m)/f_S(K)$. Experts' assessments on the impact of the loss $\boldsymbol{K}_m$ to the portfolio $\boldsymbol{X}$ can also be incorporated. The baseline allocation is understood as an allocated capital before adjustment of multimodality. Therefore,

$\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}$ in (6.15) can be replaced by Euler allocation if, for example, its economic justification, such as RORAC compatibility and core-compatibility, is regarded as important. To explain the multimodality adjustment, suppose that the scenario $\{\boldsymbol{X} = \boldsymbol{K}_m\}$ occurs with probability w_m. Under this scenario, the portfolio incurs the loss (or profit) $\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}$. When $\boldsymbol{\lambda}_m = \mathbf{1}_d$, the actual amount of loss $(\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}})^+$ contributes to the average $\sum_{m=1}^M w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}})^+$. However, this choice of $\boldsymbol{\lambda}_m$ is too conservative and smaller values of $\boldsymbol{\lambda}_m$ are typically more reasonable since both $\boldsymbol{K}_m$ and $\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}$ sum up to K and thus losses of some units imply profits of others. Therefore, losses of some units can be compensated by the profits of other units, and the multimodality loading parameter $\boldsymbol{\lambda}_m$ can be determined by such risk mitigation or a corresponding insurance contract.

One of the advantages of the multimodality-adjusted allocated capital over MLA is that $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}$ is well-defined even if the global mode of $\boldsymbol{X} \mid \{S = K\}$ is not unique whereas MLA is not well-defined in this case. Next, we will verify that $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}$ measures the risk of multimodality from various viewpoints. First, if $M = 1$, then $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda} = \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}$ and thus the multimodality adjustment is zero. Second, suppose that $M \geq 2$ and $w_m > 0$ for $m = 1, \dots, M$. Then $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda} = \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}$ if and only if $\lambda_{j,m} = 0$ for all $j = 1, \dots, d$ and $m = 1, \dots, M$ such that $K_{m,j} > \bar{K}_{\boldsymbol{w},\mathcal{X},j}$. Therefore, under multimodality, the multimodality adjustment is zero if and only if losses of some units of the portfolio are completely compensated by profits of others. Finally, $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}$ is increasing with respect to the variability of the set of scenarios, which can be understood as a degree of multimodality. To see this, suppose that $\boldsymbol{\lambda}_1 = \cdots = \boldsymbol{\lambda}_M = \boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \in \mathbb{R}^d_+$, and denote by $\boldsymbol{Y}$ the discrete random vector taking points $\boldsymbol{K}_1, \dots, \boldsymbol{K}_M$ with probabilities $w_1, \dots, w_M$. Then the multimodality-adjusted allocated capital (6.15) is written by

$$\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda} = \mathbb{E}[\boldsymbol{Y}] + \boldsymbol{\lambda} \circ \mathbb{E}[(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])^+]. \tag{6.16}$$

Variability of the set of scenarios can then be compared by the so-called convex order of $Y_1, \dots, Y_d$. For two $\mathbb{R}$-valued random variables Y and Y', Y' is said to be larger than Y in the *convex order*, denoted as $Y \leq_{\mathrm{cx}} Y'$, if $\mathbb{E}[\phi(Y)] \leq \mathbb{E}[\phi(Y')]$ for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$ provided the expectations exist; see Shaked and Shanthikumar (2007) for a comprehensive reference. Roughly speaking, convex order compares the variability of random variables and Y' shows more variability than Y if $Y \leq_{\mathrm{cx}} Y'$; for instance, $Y \leq_{\mathrm{cx}} Y'$ implies $\mathbb{E}[Y] = \mathbb{E}[Y']$, $\mathrm{Var}(Y) \leq \mathrm{Var}(Y')$, $\mathrm{ess.inf}(Y') \leq \mathrm{ess.inf}(Y)$, $\mathrm{ess.sup}(Y) \leq \mathrm{ess.sup}(Y')$ and $\mathbb{E}[(Y - a)^+] \leq \mathbb{E}[(Y' - a)^+]$ for all $a \in \mathbb{R}$. Therefore, for two sets of scenarios $\mathcal{X}$ and $\mathcal{X}'$ with associated probabilities $\boldsymbol{w}$ and $\boldsymbol{w}'$, if one shows more variation (multimodal) than the other in the sense that the corresponding discrete random variables satisfy $Y_j \leq_{\mathrm{cx}} Y'_j$ for some $j \in \{1, \dots, d\}$, then $(\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda})_j \leq (\boldsymbol{K}_{\boldsymbol{w}',\mathcal{X}',\Lambda})_j$ holds as desired.

Remark 6.4.4 (Multimodality adjustment for general sets of scenarios)**.** Representation (6.16) bears structural resemblance to Gini shortfall allocations introduced in Furman et al. (2017), and indicates a possible extension of the multimodality adjustment to the case when the set of scenarios is not a discrete set. For instance, by taking $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^d : \mathbf{1}_d^\top \boldsymbol{x} \geq \mathrm{VaR}_p(S)\}$ and $w(\boldsymbol{x}) = f_{\boldsymbol{X} \mid \{S \geq \mathrm{VaR}_p(S)\}}(\boldsymbol{x})$, (6.16) can be interpreted as multimodality-adjusted Euler allocations of Expected shortfall since (6.16) yields

$$\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda} = \mathrm{ES}_p(X_j; S) + \boldsymbol{\lambda} \circ \mathbb{E}[(\boldsymbol{X} - \mathrm{ES}_p(X_j; S)^+ \mid \{S \geq \mathrm{VaR}_p(S)\}],$$

where $\mathrm{ES}_p(X_j; S) = \mathbb{E}[\boldsymbol{X} \mid \{S \geq \mathrm{VaR}_p(S)\}]$ is the Euler allocation of $K = \mathrm{ES}_p(S)$ as derived in (1.7).

Next we study properties of $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}$ considered for MLA in Section 6.4.2. To clarify the relationship between $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}$, the total capital K and the loss distribution of $\boldsymbol{X}$, define $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X};\mathcal{K}_d(K)]$ and $\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X};\mathcal{K}_d(K)]$ to be the multimodality-adjusted allocated capitals (6.15) and their first term $\sum_{m=1}^{M} w_m \boldsymbol{K}_m$, respectively, with $\mathcal{X}$ being the set of local modes $\boldsymbol{K}_1,\dots,\boldsymbol{K}_M$ of $\boldsymbol{x} \mapsto f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}_d(K)\}}$ (assumed to be a discrete set) and with $w_m \propto f_{\boldsymbol{X}}(\boldsymbol{K}_m)$. To this end, we adopt the following definition of local modes.

Definition 6.4.5 (Local modes)**.** For an $\mathbb{R}_+$-valued function f on $\mathbb{R}^d$, $\boldsymbol{x}\in\mathbb{R}^d$ is called a *local mode* of f if there exists $\epsilon>0$ such that

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) \quad \text{for all} \quad \boldsymbol{y}\in\mathcal{N}_\epsilon(\boldsymbol{x}), \tag{6.17}$$

where $\mathcal{N}_\epsilon(\boldsymbol{x}) = \{\boldsymbol{z}\in\mathbb{R}^d : ||\boldsymbol{z}-\boldsymbol{x}|| < \epsilon\}$. If (6.17) holds for any $\epsilon>0$, then $\boldsymbol{x}$ is called a *global mode* of f.

Properties of $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X};\mathcal{K}_d(K)]$ are then summarized as follows.

1. *Translation invariance*: We show that $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}$ is translation invariant in the sense that

$$\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}+\boldsymbol{c};\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c})] = \boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X};\mathcal{K}_d(K)]+\boldsymbol{c} \quad \text{for} \quad \boldsymbol{c}\in\mathbb{R}^d.$$

   To show this, notice that local modes of $\boldsymbol{x}\mapsto f_{\boldsymbol{X}+\boldsymbol{c}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c})\}}$ are given by $\boldsymbol{K}_m+\boldsymbol{c}$, $m=1,\dots,M$, if $\boldsymbol{K}_m$, $m=1,\dots,M$, are the local modes of $\boldsymbol{x}\mapsto f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}_d(K)\}}$. Since $w_m = f_{\boldsymbol{X}}(\boldsymbol{K}_m) = f_{\boldsymbol{X}+\boldsymbol{c}}(\boldsymbol{K}_m+\boldsymbol{c})$, the probability weight assigned to the mth scenario does not change from $(\boldsymbol{X},\mathcal{K}_d(K))$ to $(\boldsymbol{X}+\boldsymbol{c},\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c}))$ for all $m=1,\dots,M$. Therefore, $\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X}+\boldsymbol{c};\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c})] = \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X};\mathcal{K}_d(K)]+\boldsymbol{c}$ and thus

$$\begin{aligned}
\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}+\boldsymbol{c};\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c})] &= \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X}+\boldsymbol{c};\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c})] \\
&\quad + \sum_{m=1}^{M} w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m+\boldsymbol{c}-\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X}+\boldsymbol{c};\mathcal{K}_d(K+\mathbf{1}_d^\top\boldsymbol{c})])^+ \\
&= \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X};\mathcal{K}_d(K)]+\boldsymbol{c}+\sum_{m=1}^{M} w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m-\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X};\mathcal{K}_d(K)])^+ \\
&= \boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X};\mathcal{K}_d(K)]+\boldsymbol{c},
\end{aligned}$$

   which shows translation invariance.

2. *Positive homogeneity*: Multimodality-adjusted allocated capitals are positive homogeneous in the sense that

$$\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[c\boldsymbol{X};\mathcal{K}_d(cK)] = c\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X};\mathcal{K}_d(K)] \quad \text{for} \quad c>0.$$

   This can be checked similarly as translation invariance. The local modes of $\boldsymbol{x}\mapsto f_{c\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}_d(cK)\}}$ are given by $c\boldsymbol{K}_m$, $m=1,\dots,M$, if $\boldsymbol{K}_m$, $m=1,\dots,M$, are the local modes of $\boldsymbol{x}\mapsto f_{\boldsymbol{X}}(\boldsymbol{x})\mathbf{1}_{\{\boldsymbol{x}\in\mathcal{K}_d(K)\}}$, and the probability weight assigned to the mth scenario does not change from $(\boldsymbol{X},\mathcal{K}_d(K))$ to

$(c\boldsymbol{X}, \mathcal{K}_d(cK))$ since $w_m = f_{\boldsymbol{X}}(\boldsymbol{K}_m) = f_{c\boldsymbol{X}}(c\boldsymbol{K}_m)$ for all $m = 1, \dots, M$. Therefore, $\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[c\boldsymbol{X}; \mathcal{K}_d(cK)] = c\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X}; \mathcal{K}_d(K)]$ and thus

$$\begin{aligned}
\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[c\boldsymbol{X}; \mathcal{K}_d(cK)] &= \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[c\boldsymbol{X}; \mathcal{K}_d(cK)] \\
&\quad + \sum_{m=1}^{M} w_m \boldsymbol{\lambda}_m \circ (c\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[c\boldsymbol{X}; \mathcal{K}_d(cK)])^+ \\
&= c\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X}; \mathcal{K}_d(K)] + c \sum_{m=1}^{M} w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}}[\boldsymbol{X}; \mathcal{K}_d(K)])^+ \\
&= c\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)],
\end{aligned}$$

which shows positive homogeneity.

3. *Riskless asset*: Multimodality-adjusted allocated capitals satisfy the riskless asset property in the following sense. Suppose that $X_j = c_j \in \mathbb{R}$ a.s. for $j \in I \subseteq \{1, \dots, d\}$ and that $\boldsymbol{X}_{-I} = (X_j, j \in \{1, \dots, d\} \backslash I)$ admits a density $f_{\boldsymbol{X}_{-I}}$. As we discussed in Section 6.4.2, any realization $\boldsymbol{x}$ of $\boldsymbol{X} \mid \{S = K\}$ satisfies $\boldsymbol{x}_I = \boldsymbol{c}$ where $\boldsymbol{c} = (c_j; j \in I)$, and the likelihood of $\boldsymbol{x}$ is quantified through the density $f_{\boldsymbol{X}_{-I} \mid \{\mathbf{1}_{|-I|}^\top \boldsymbol{X}_{-I} = K - \mathbf{1}_{|I|}^\top \boldsymbol{c}\}}(\boldsymbol{x}_{-I})$. Therefore, reasonable choices of the scenarios $\boldsymbol{K}_1, \dots, \boldsymbol{K}_M \in \mathcal{K}_d(K)$ are such that $(\boldsymbol{K}_m)_I = \boldsymbol{c}$ and $(\boldsymbol{K}_m)_{-I}$ are local modes of $\boldsymbol{X}_{-I} \mid \{\mathbf{1}_{|-I|}^\top \boldsymbol{X}_{-I} = K - \mathbf{1}_{|I|}^\top \boldsymbol{c}\}$. In this case, we have that $(\bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}})_I = \boldsymbol{c}$ and that

$$\left(\sum_{m=1}^{M} w_m \boldsymbol{\lambda}_m \circ (\boldsymbol{K}_m - \bar{\boldsymbol{K}}_{\boldsymbol{w},\mathcal{X}})^+ \right)_I = \sum_{m=1}^{M} w_m (\boldsymbol{\lambda}_m)_I \circ (\boldsymbol{c} - \boldsymbol{c})^+ = \mathbf{0}_{|I|}.$$

Therefore, it holds that $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)]_I = \boldsymbol{c}$ if $\boldsymbol{X}_I = \boldsymbol{c}$ a.s.

4. *Symmetry*: For a reasonable choice of Λ, the multimodality-adjusted allocated capitals satisfy the symmetry property, that is, $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)]_i = \boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)]_j$ for $i, j \in \{1, \dots, d\}$, $i \neq j$, such that $\boldsymbol{X} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{X}}$ where $\tilde{\boldsymbol{X}}$ is a d-dimensional random vector satisfying $\tilde{X}_j = X_i$, $\tilde{X}_i = X_j$ and $\tilde{X}_k = X_k$ for all $k \in \{1, \dots, d\} \backslash \{i, j\}$. For any $\boldsymbol{x} \in \mathbb{R}^d$, denote by $\tilde{\boldsymbol{x}}$ a d-dimensional vector such that $\tilde{x}_j = x_i$, $\tilde{x}_i = x_j$ and $\tilde{x}_k = x_k$ for all $k \in \{1, \dots, d\} \backslash \{i, j\}$. To show the symmetry, suppose that $\boldsymbol{K} \in \mathcal{K}_d(K)$ is a local mode of $\boldsymbol{x} \mapsto f_{\boldsymbol{X}}(\boldsymbol{x}) \mathbf{1}_{\{\boldsymbol{x} \in \mathcal{K}_d(K)\}}$. Then, under $\boldsymbol{X} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{x}}$ is also a local mode of $\boldsymbol{x} \mapsto f_{\boldsymbol{X}}(\boldsymbol{x}) \mathbf{1}_{\{\boldsymbol{x} \in \mathcal{K}_d(K)\}}$ since $f_{\boldsymbol{X}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\tilde{\boldsymbol{y}})$ for any $\boldsymbol{y} \in \mathbb{R}^d$. Hence $\tilde{\boldsymbol{x}}$ satisfies (6.17) for some $\epsilon > 0$. Therefore, any element $\boldsymbol{K}_m = (K_{m,1}, \dots, K_{m,d})$ in $\mathcal{X}$ satisfies either (1) $K_{m,i} = K_{m,j}$ or (2) there exists a unique element $\boldsymbol{K}_{m'} \in \mathcal{X}$ such that $K_{m,i} \neq K_{m,j}$, $K_{m',i} \neq K_{m',j}$, $K_{m',i} = K_{m,j}$, $K_{m,i} = K_{m',j}$ and $K_{m,k} = K_{m',k}$ for all $k \in \{1, \dots, d\} \backslash \{i, j\}$. For such a pair (m, m'), it holds that $w_m = w_{m'}$ since $w_m \propto f_{\boldsymbol{X}}(\boldsymbol{K}_m)$ and $w_{m'} \propto f_{\boldsymbol{X}}(\boldsymbol{K}_{m'}) = f_{\boldsymbol{X}'}(\boldsymbol{K}_m) = f_{\boldsymbol{X}}(\boldsymbol{K}_m)$ as $\boldsymbol{X} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{X}}$. Therefore, if $\boldsymbol{\lambda}_m = \boldsymbol{\lambda}_{m'}$ holds for all pairs of (m, m') in Case (2), it holds that $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)]_i = \boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)]_j$.

The continuity property $\lim_{n \to \infty} \boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}_n; \mathcal{K}_d(K)] = \boldsymbol{K}_{\boldsymbol{w},\mathcal{X},\Lambda}[\boldsymbol{X}; \mathcal{K}_d(K)]$ for a given $\boldsymbol{X}_n$ and $\boldsymbol{X}$ such that $\boldsymbol{X}_n$ converges to $\boldsymbol{X}$ weakly may not be straightforward to verify since a limit of multimodal distributions can be unimodal, and more generally, the number of scenarios may change in n.

We end this section with a remark on the case when multiple measures or models are considered as different scenarios and how to incorporate these scenarios into multimodality-adjusted allocated capitals.

Remark 6.4.6 (Multimodality adjustment for different measures)**.** A single model of a risk may not be sufficient to manage the risk due to changes of an economic situation or due to model uncertainty. For a further risk evaluation, it may be useful to consider multiple measures $\mathbb{Q}_1, \dots, \mathbb{Q}_S$ where $\mathbb{Q}_s$ is a probability measure on $(\Omega, \mathcal{A})$ and $F_{\boldsymbol{X}}^{\mathbb{Q}_s}$ is the distribution function of $\boldsymbol{X}$ under $\mathbb{Q}_s$ for $s = 1, \dots, S$. These multiple measures can be incorporated into the scenario analysis by, for example, considering the (componentwise) maximum of the multimodality-adjusted allocated capitals $\mathcal{K}_{\boldsymbol{w},\mathcal{X},\Lambda}^{\mathbb{Q}_s}(\boldsymbol{X}; \mathcal{K}_d(K))$ calculated based on $F_{\boldsymbol{X}}^{\mathbb{Q}_s}$ for $s = 1, \dots, S$, or considering their mixture with respect to probabilities $q_1, \dots, q_S$ where q_s is associated to the scenario $\mathbb{Q}_s$ determined, for example, proportionally to the sample size available for the distribution $F_{\boldsymbol{X}}^{\mathbb{Q}_s}$.

6.5 Numerical experiments

In this section we conduct an empirical and a simulation study to compute Euler and maximum likelihood allocations, and compare them for various models. Simulation of the conditional distribution given a constant sum is in general challenging. Throughout this section, we adopt the *(crude) Monte Carlo (MC)* method to simulate $\boldsymbol{X}' \mid \{S = K\}$ according to which unconditional samples from $\boldsymbol{X}$ are first generated and those falling in the region $\mathcal{K}_d(K, \delta) = \{\boldsymbol{x} \in \mathbb{R}^d : K - \delta < \mathbf{1}_d^\top \boldsymbol{x} < K + \delta\}$ for a sufficiently small $\delta > 0$ are then extracted. The extracted samples are standardized via $KX_j / \sum_{j=1}^d X_j$ so that their componentwise sum equals K. Finally the standardized samples are used as pseudo-samples from $\boldsymbol{X}' \mid \{S = K\}$. See Section 4.2 for the potential bias caused by this method, and Chapter 4 and Chapter 5 for more sophisticated simulation approaches of $\boldsymbol{X}' \mid \{S = K\}$ based on MCMC methods.

6.5.1 Empirical study

In this section we estimate the proposed MLA nonparametrically for real financial data. We consider daily log-returns of the stock indices FTSE $X_{t,1}$, S&P 500 $X_{t,2}$ and DJI $X_{t,3}$ from January 2, 1990 to March 25, 2004, which contains 3713 days and thus $T = 3712$ log-returns. We consider two portfolios (a) $\boldsymbol{X}_t^{\text{pos}} = (X_{t,1}, X_{t,2}, X_{t,3})$ and (b) $\boldsymbol{X}_t^{\text{neg}} = (X_{t,1}, -X_{t,2}, X_{t,3})$. For each portfolio, we aim at allocating the capital $K = 1$ based on the conditional loss distribution at time $T + 1$ given the history up to and including time T. Taking into account the stylized facts of stock returns listed in Chapter 3 of McNeil et al. (2015) (such as unimodality, heavy-tailedness and volatility clusters), we adopted a copula-GARCH model with marginal skew-t innovations (ST-GARCH; see, for example, Jondeau and Rockinger (2006) and Huang et al. (2009)). We utilize a GARCH$(1, 1)$ model with skew-t innovations with degrees of freedom $\nu_j > 0$ and skewness parameter $\gamma_j > 0$ for the jth marginal time series. That is, within a fixed time period $\{1, \dots, T + 1\}$ the jth return series $(X_{1,j}, \dots, X_{T+1,j})$ follows

$$X_{t,j} = \mu_j + \sigma_{t,j} Z_{t,j}, \quad \sigma_{t,j}^2 = \omega_j + \alpha_j X_{t-1,j}^2 + \beta_j \sigma_{t-1,j}^2, \quad Z_{t,j} \overset{\text{iid}}{\sim} \text{ST}(\nu_j, \gamma_j), \quad j = 1, \dots, d,$$

Table 6.1: Maximum likelihood estimates and estimated standard errors of the ST-GARCH(1,1) model.

	μ_j	ω_j	α_j	β_j	γ_j	ν_j
$X_{t,1}^{\text{pos/neg}}$	0.053	0.006	0.052	0.943	0.969	6.414
SE	0.013	0.002	0.008	0.008	0.021	0.663
$X_{t,2}^{\text{pos}}$	0.050	0.003	0.049	0.950	0.983	6.265
SE	0.013	0.001	0.007	0.007	0.021	0.659
$X_{t,2}^{\text{neg}}$	-0.050	0.003	0.049	0.950	1.018	6.265
SE	0.013	0.001	0.007	0.007	0.022	0.659
$X_{t,3}^{\text{pos/neg}}$	0.031	0.011	0.071	0.920	0.966	10.000
SE	0.014	0.003	0.009	0.010	0.023	1.309

where $\omega_j > 0, \alpha_j, \beta_j \geq 0$, $\alpha_j + \beta_j < 1$, and $Z_{t,j}$ follows a skew-t distribution $\mathrm{ST}(\nu_j, \gamma_j)$ with density given by

$$f_j(x_j; \nu_j, \gamma_j) = \frac{2}{\gamma_j + \frac{1}{\gamma_j}} \left\{ t(x_j, \nu_j) \mathbf{1}_{[x_j \geq 0]} + t(\gamma_j x_j, \nu_j) \mathbf{1}_{[x_j < 0]} \right\}, \tag{6.18}$$

where $t(x, \nu)$ is the density function of a Student t distribution with degrees of freedom $\nu > 0$ and a skewness parameter $\gamma > 0$ with $\gamma = 1$ leading to the standard symmetric case. The copula among the stationary process $\boldsymbol{Z}_t = (Z_{t,1}, \ldots, Z_{t,d})$, denoted as C, is estimated nonparametrically. Under this model, the joint distribution of the returns $\boldsymbol{X}_{T+1|\mathcal{F}_T} = (X_{T+1,1|\mathcal{F}_T}, \ldots, X_{T+1,d|\mathcal{F}_T})$ has marginal distributions $\mathrm{ST}(\mu_j, \sigma_{t+1,j}^2, \nu_j, \gamma_j)$, $j = 1, \ldots, d$, and a copula C, where $\mathrm{ST}(\mu_j, \sigma_{t+1,j}^2, \nu_j, \gamma_j)$ is a skew-t distribution with density $f_j(\frac{x_j - \mu_j}{\sigma_{t+1,j}}; \nu_j, \gamma_j)$ with $f_j(\cdot; \nu_j, \gamma_j)$ defined in (6.18). Parameters of the ST-GARCH(1,1) models are estimated with the maximum likelihood method; the results are summarized in Table 6.1.

For each case of (a) and (b), we take $K = 1$ and estimate the Euler allocation and MLA by a resampling method. After extracting the marginal standardized residuals, we build their pseudo-observations as a pseudo-sample from C. We then generate samples of size $N = 3712$ by resampling with replacement. The samples from C are then marginally transformed by skew-t distributions with parameters specified as in Table 6.1. From these samples of $\boldsymbol{X}_{T+1|\mathcal{F}_T}$, we extract the subsamples falling in the region $\mathcal{K}_d(K, \delta) = \left\{ \boldsymbol{x} \in \mathbb{R}^3 : K - \delta < \sum_{j=1}^3 x_j < K + \delta \right\}$ where $\delta = 0.3$. These samples are then standardized via $K X_{t,j} / \sum_{j=1}^d X_{t,j}$ to add up to K. Scatter plots of the first two components of these data are shown in Figure 6.2.

The 3712 data points lead to 354 and 558 samples from $\boldsymbol{X}_{T+1|\mathcal{F}_T}^{\text{pos}}$ and $\boldsymbol{X}_{T+1|\mathcal{F}_T}^{\text{neg}}$ on $\mathcal{K}_d(K, \delta)$, respectively. Based on these conditional samples, we estimate the Euler allocation $\mathbb{E}[\boldsymbol{X} \mid \{S = K\}]$ and the MLA, that is, the mode of $f_{\boldsymbol{X}|\{S=K\}}$ provided it is unique. The (possibly multiple) modes were estimated by the function kms (*kernel mean shift clustering*, proposed by Fukunaga and Hostetler, 1975) of the R package ks; see Carreira-Perpinán (2015) and Chen et al. (2016) for details and for other methods of estimating modes. For the computational times required to calculate the allocations, computing MLAs took 0.353 seconds in

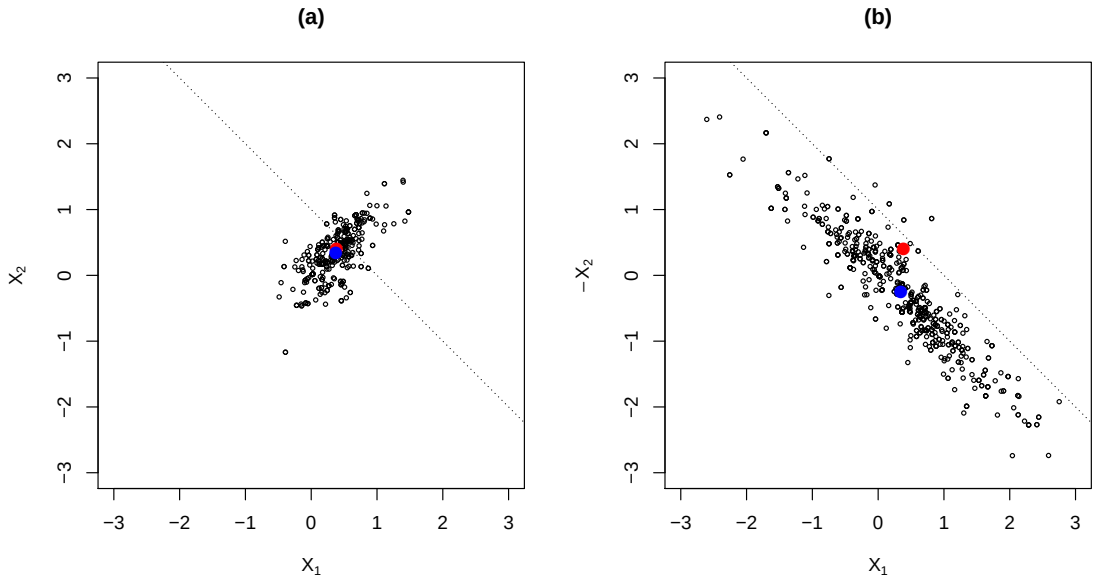

Figure 6.2: Scatter plots (black dots) of the first two components of (a) $\boldsymbol{X}_t^{\text{pos}} = (X_{t,1}, X_{t,2}, X_{t,3})$ and (b) $\boldsymbol{X}_t^{\text{neg}} = (X_{t,1}, -X_{t,2}, X_{t,3})$ for daily log-returns of the stock indices FTSE $X_{t,1}$, S&P 500 $X_{t,2}$ and DJI $X_{t,3}$ falling in the region $\mathcal{K}_d(K, \delta) = \left\{ \boldsymbol{x} \in \mathbb{R}^3 : K - \delta < \sum_{j=1}^3 x_j < K + \delta \right\}$ where $\delta = 0.3$ and $K = 1$. The dotted lines represent the line $x + y = K$. The red dot represents the Euler allocation $\mathbb{E}[\boldsymbol{X}' \mid \{S = K\}]$ and the blue dot represents the maximum likelihood allocation, the mode of $f_{\boldsymbol{X}'|\{S=K\}}$.

Case (a) and 0.431 seconds in Case (b) whereas, in both cases, the Euler allocations were computed almost instantly. As was expected from the ellipticality of the scatter plots in Figure 6.2, the unique mode was discovered in each case. The first two components of the two allocations are pointed out in Figure 6.2.

Next, we estimate the standard errors of the Euler and maximum likelihood allocations using the bootstrap method. We compute the Euler allocation, MLA and their standard errors based on the $B = 100$ number of samples of size $N = 3712$ resampled from the original data with replacement. The results are summarized in Table 6.2.

In Figure 6.2 we can observe that compared with Case (a) the distribution in Case (b) is more spread out and losses take larger absolute values. If the samples are regarded as stressed scenarios, the scenario set in Case (b) contains a wider variety of scenarios than in Case (a) since both positive and negative losses can appear in Case (b) whereas most realizations are positive in Case (a). Nevertheless, as is observed from Table 6.2, in both cases the Euler allocation and the MLA are close to each other also in terms of standard errors. This observation does not conflict with the stylized fact that the joint log-returns nearly follow an elliptical distribution, and thus the mean (Euler allocation) of $\boldsymbol{X} \mid \{S = K\}$ coincides with its mode; see Proposition 6.3.1 and Proposition 6.3.13 Part 1.

Table 6.2: Bootstrap estimates and estimated standard errors of the Euler allocation and MLA of $\boldsymbol{X}^{\mathrm{pos}} = (X_1, X_2, X_3)$ and $\boldsymbol{X}^{\mathrm{neg}} = (X_1, -X_2, X_3)$ for daily log-returns of the stock indices FTSE X_1, S&P 500 X_2 and DJI X_3. The subsample size is $N = 3712$ and the bootstrap sample size is $B = 100$.

	Estimator			Standard error		
	X_1	X_2	X_3	X_1	X_2	X_3
$\mathbb{E}[\boldsymbol{X}^{\mathrm{pos}} \mid \{S = K\}]$	0.378	0.338	0.285	0.019	0.022	0.038
$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}^{\mathrm{pos}}; \mathcal{K}_d(K)]$	0.367	0.365	0.268	0.019	0.024	0.041
$\mathbb{E}[\boldsymbol{X}^{\mathrm{neg}} \mid \{S = K\}]$	0.345	−0.248	0.903	0.037	0.039	0.015
$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}^{\mathrm{neg}}; \mathcal{K}_d(K)]$	0.371	−0.280	0.909	0.040	0.039	0.013

6.5.2 Simulation study

A potential drawback of the nonparametric estimation of Euler and maximum likelihood allocations is that the sample size is often not sufficient for statistical estimation due to the sum constraint. To avoid this issue, one can first fit a parametric model based on the unconditional samples, and then take subsamples of simulated samples from the fitted parametric model to estimate Euler and maximum likelihood allocations. In this section, we consider four models, referred to as (M1), (M2), (M3) and (M4), respectively, with $d = 3$ and having the same marginal distributions $X_1 \sim \mathrm{Par}(2.5, 5)$, $X_2 \sim \mathrm{Par}(2.75, 5)$ and $X_3 \sim \mathrm{Par}(3, 5)$ (where $\mathrm{Par}(\theta, \lambda)$ denotes the Pareto distribution with shape parameter $\theta > 0$ and scale parameter $\lambda > 0$) but different t copulas with degrees of freedom $\nu = 5$ and dispersion matrices

$$P_1 = \begin{pmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0.8 \\ 0.5 & 0.8 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix},$$
$$P_3 = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 1 \end{pmatrix}, \quad P_4 = \begin{pmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & -0.5 \\ 0.5 & -0.5 & 1 \end{pmatrix}, \tag{6.19}$$

respectively. For these parametric models, we first simulate $N = 10^6$ samples from the unconditional distribution and then extract subsamples falling in the region $\mathcal{K}_d(K, \delta)$ with $K = 40$ and $\delta = 1$. These samples from $\boldsymbol{X}' \mid \{S = K\}$ are shown in Figure 6.3. The red point in the figure represents the Euler allocation and the blue points are the (local) modes, which are estimated similarly as in Section 6.5.1.

The computational times required for calculating MLAs were (in seconds) (M1) 18.964, (M2) 11.726, (M3) 15.946, and (M4) 22.762. On the other hand, the Euler allocations, which are simply sample means, were computed almost instantly for all the cases. Compared with the results in Section 6.5.1, we observe that the computational time required to calculate MLA increases more rapidly than the Euler allocation does as the sample size increases.

In Figure 6.3 we can observe that the conditional distribution is more concentrated under positive dependence (Model (M1) and (M2)) and it is more dispersed under negative dependence (Model (M4)).

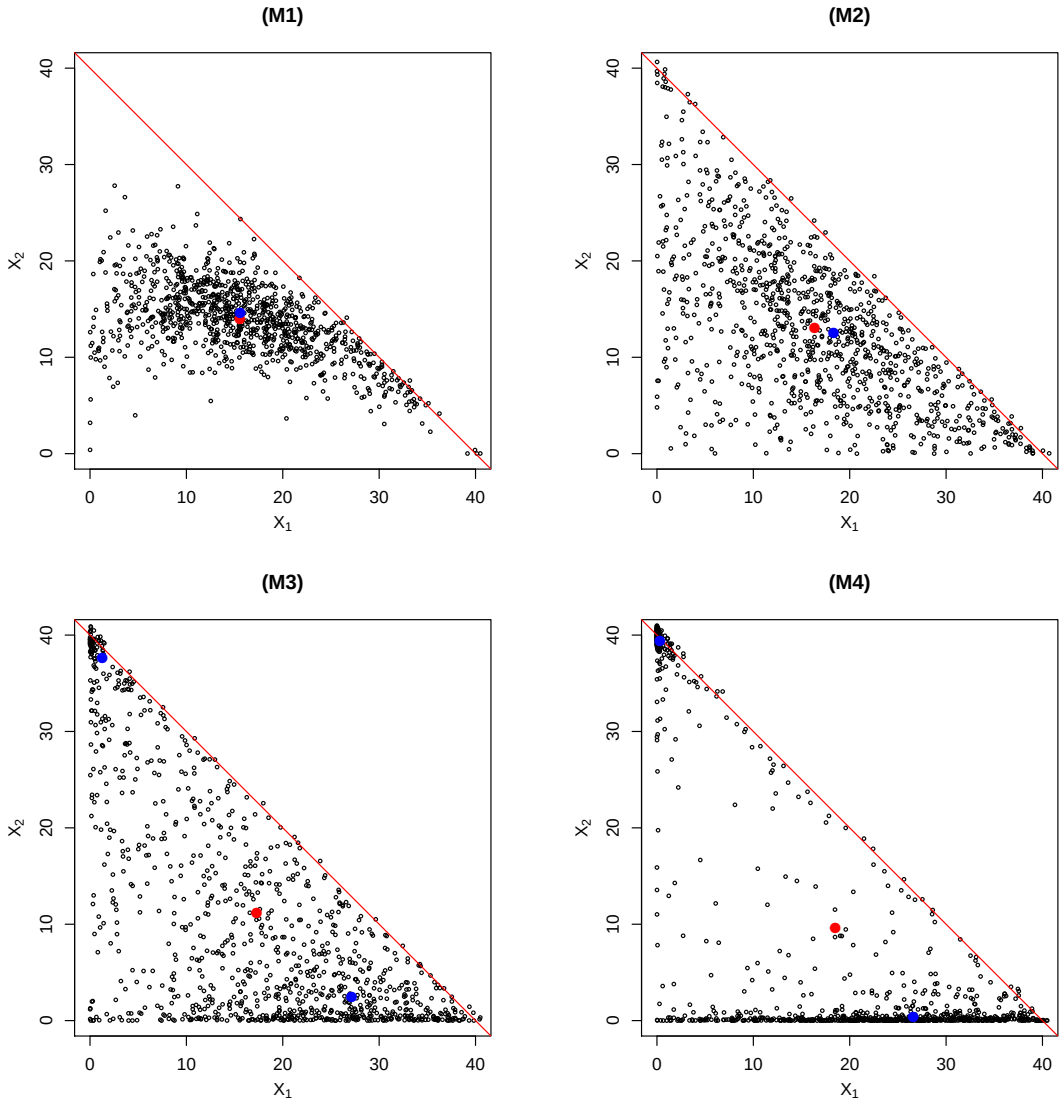

Figure 6.3: Scatter plots (black dots) of the first two components of the four models (M1), (M2), (M3) and (M4) falling in the region $\mathcal{K}_d(K, \delta)$ with $K = 40$ and $\delta = 1$. All the four models have the same marginal distributions $X_1 \sim \text{Par}(2.5, 5)$, $X_2 \sim \text{Par}(2.75, 5)$ and $X_3 \sim \text{Par}(3, 5)$ but different t copulas with parameters provided in (6.19). The red lines represent $x + y = K$. The red dot represents the Euler allocation $\mathbb{E}[\boldsymbol{X}' \mid \{S = K\}]$ and the blue dots represent the (local) modes of $f_{\boldsymbol{X}'|\{S=K\}}$.

Table 6.3: Estimates and estimated standard errors of the Euler allocation and MLA of the four models (M1), (M2), (M3) and (M4) all having the same marginal distributions $X_1 \sim \mathrm{Par}(2.5, 5)$, $X_2 \sim \mathrm{Par}(2.75, 5)$ and $X_3 \sim \mathrm{Par}(3, 5)$ but different t copulas with parameters provided in (6.19). Estimates and estimated standard errors are computed based on 100 replications, each of which utilizing 500 conditional samples falling in the region $\mathcal{K}_d(K, \delta)$ with $K = 40$ and $\delta = 1$.

	Estimator			Standard error		
	X_1	X_2	X_3	X_1	X_2	X_3
(M1) Pareto + t copula: strong positive dependence						
$\mathbb{E}[\boldsymbol{X} \mid \{S = K\}]$	15.549	13.889	10.562	0.336	0.157	0.288
$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]$	15.849	14.434	9.718	0.482	0.213	0.356
(M2) Pareto + t copula: positive dependence						
$\mathbb{E}[\boldsymbol{X} \mid \{S = K\}]$	16.228	13.042	10.562	0.399	0.355	0.288
$\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]$	17.689	12.481	9.830	0.759	0.663	0.475
(M3) Pareto + t copula: independence						
$\mathbb{E}[\boldsymbol{X} \mid \{S = K\}]$	17.479	11.368	10.562	0.517	0.530	0.288
$\boldsymbol{K}_{\mathrm{M},1}[\boldsymbol{X}; \mathcal{K}_d(K)]$	25.678	3.107	11.215	1.185	0.278	1.205
$\boldsymbol{K}_{\mathrm{M},2}[\boldsymbol{X}; \mathcal{K}_d(K)]$	2.639	35.275	2.086	0.973	1.306	0.424
(M4) Pareto + t copula: negative dependence						
$\mathbb{E}[\boldsymbol{X} \mid \{S = K\}]$	19.062	9.272	10.562	0.556	0.614	0.288
$\boldsymbol{K}_{\mathrm{M},1}[\boldsymbol{X}; \mathcal{K}_d(K)]$	28.353	0.684	10.962	2.125	1.646	2.154
$\boldsymbol{K}_{\mathrm{M},2}[\boldsymbol{X}; \mathcal{K}_d(K)]$	0.710	38.385	0.905	1.719	3.537	2.705

Regarding the samples as stress scenarios, the sets in Model (M3) and (M4) are more worrisome than those of Model (M1) and (M2) since the former contain two distinct scenarios, one around the first axis and one around the upper-left corner of the plot region, both of which are likely to occur in the stressed situation $\{S = K\}$. Unimodality of the conditional distribution in Model (M1) and (M2) leads to closer Euler allocation and MLA. For Model (M1) and (M2), the choice of Euler allocation and MLA does not significantly change the resulting allocation. On the other hand, for Model (M3) and (M4), the conditional distributions are multimodal, and thus more careful decision making is required.

To investigate the standard errors of the estimators, we compute the estimates of Euler allocation and (local) modes of $f_{\boldsymbol{X}'|\{S=K\}}$ 100 times for each model. For each repetition, we simulate samples from $\boldsymbol{X}$ so that there are 500 samples in the region $\mathcal{K}_d(K, \delta)$. The estimates and standard errors are computed based on the 100 replications and the results are summarized in Table 6.3. We can again see that for Models (M1) and (M2) the two allocations are close. On the other hand, for Models (M3) and (M4) where the conditional distributions are multimodal, the standard errors of the (local) modes are higher than those of

the Euler allocation.

In the end, we compute the multimodality-adjusted allocated capitals (6.15) in (M3) and (M4). In each case, the parameters are specified as $M = 2$, $\mathcal{X} = \{\boldsymbol{K}_1, \boldsymbol{K}_2\}$ and $\boldsymbol{w} = (w_1, w_2)$ with $w_m \propto f_{\boldsymbol{X}}(\boldsymbol{K}_m)$, where $\boldsymbol{K}_1 = \boldsymbol{K}_{\mathrm{M},1}[\boldsymbol{X}; \mathcal{K}_d(K)] = (26.726, 2.114, 11.158)$ and $\boldsymbol{K}_2 = \boldsymbol{K}_{\mathrm{M},2}[\boldsymbol{X}; \mathcal{K}_d(K)] = (1.505, 37.203, 1.291)$ in (M3), and $\boldsymbol{K}_1 = (28.589, 0.432, 10.978)$ and $\boldsymbol{K}_2 = (0.326, 39.314, 0.358)$ in (M4). In both (M3) and (M4), the first and third units incur losses when $\{\boldsymbol{X} = \boldsymbol{K}_1\}$ occurs, and the second unit incurs a large loss when $\{\boldsymbol{X} = \boldsymbol{K}_2\}$ happens. The probability weights of the scenarios are given by $\boldsymbol{w} = (0.509, 0.490)$ in (M3) and $\boldsymbol{w} = (0.272, 0.727)$ in (M4). The two scenarios $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are almost likely to occur in (M3), and the second scenario $\boldsymbol{K}_2$ is more likely to occur in (M4). Based on $\boldsymbol{w}$ and $\mathcal{X}$, the baseline allocations $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X}}$ are given by $(14.357, 19.323, 6.319)$ in (M3) and $(8.038, 28.705, 3.256)$ in (M4). As seen in Table 6.3, these allocations are not quite close to the Euler allocations since $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X}}$ is calculated based only on the two points $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ in $\mathcal{K}_d(K)$.

We consider two cases when $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X}}$ or Euler allocations are used as baseline allocations. If $\boldsymbol{K}_{\boldsymbol{w},\mathcal{X}}$ are the baseline allocations, the average loss $w_1(\boldsymbol{K}_1 - \boldsymbol{K}_{\boldsymbol{w},\mathcal{X}})^+ + w_2(\boldsymbol{K}_2 - \boldsymbol{K}_{\boldsymbol{w},\mathcal{X}})^+$ in the multimodality adjustment is given by $(6.303, 11.204, 0.000)$ in (M3) and $(5.607, 22.742, 0.000)$ in (M4). If Euler allocation is used as a baseline allocation, the average loss is given by $(4.873, 9.828, 0.000)$ in (M3) and $(2.618, 14.775, 0.428)$ in (M4). In all cases, the average loss incurred in the second unit is larger than those in the first and third units since the second unit incurs a large loss when the second scenario $\{\boldsymbol{X} = \boldsymbol{K}_2\}$ occurs. Moreover, in (M4), the probability that this scenario occurs is higher than that of the first scenario $\{\boldsymbol{X} = \boldsymbol{K}_1\}$. Therefore, the scenario analysis of searching for the modes of $\boldsymbol{X} \mid \{S = K\}$ reveals that higher multimodality adjustment should be applied to $\boldsymbol{X}_2$ and then to $\boldsymbol{X}_1$ to increase the soundness of risk allocations under multimodality.

6.5.3 Simulation of the conditional distribution with MCMC

In Sections 6.2.2, 6.5.1 and 6.5.2, the constraint $\{S = K\}$ was replaced by $\{K - \delta < S < K + \delta\}$ for a small $\delta > 0$ so that $\mathbb{P}(K - \delta < S < K + \delta) > 0$. However, this modification distorts the conditional distribution $\boldsymbol{X} \mid \{S = K\}$ and the resulting estimates of risk allocations suffer from inevitable biases. To overcome this issue, MCMC methods were introduced in Chapters 4 and 5. In this section, we utilize these MCMC methods and compute the Euler allocation and MLA on the restricted set of allocations called the *(atomic) core* defined by

$$\mathcal{K}_d^{\mathrm{C}}(K; r) = \{\boldsymbol{x} \in \mathbb{R}^d : \mathbf{1}_d^\top \boldsymbol{x} = K,\ \boldsymbol{\lambda}^\top \boldsymbol{x} \le r(\boldsymbol{\lambda}),\ \boldsymbol{\lambda} \in \{0,1\}^d\} \subseteq \mathcal{K}_d(K),$$

where $r : \{0,1\}^d \to \mathbb{R}$ is called a *participation profile function* typically determined as $r(\boldsymbol{\lambda}) = \varrho(\boldsymbol{\lambda}^\top \boldsymbol{X})$ for a d-dimensional loss random vector $\boldsymbol{X}$. We call an element of $\mathcal{K}_d^{\mathrm{C}}(K; r)$ a *core allocation*. As explained in Denault (2001), core allocations possess an important property as risk allocations, that is, any subportfolio of $\boldsymbol{X} = (X_1, \dots, X_d)$ of the form $(\lambda_1 X_1, \dots, \lambda_d X_d)$ gains benefit of capital reduction from managing risk as a portfolio $\boldsymbol{X}$. In fact, for a participation profile $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ where $\lambda_j \in \{0,1\}$ represents the presence $(\lambda_j = 1)$ or the absence $(\lambda_j = 0)$ of the jth entity, the total amount of capital required to cover the loss $\boldsymbol{\lambda}^\top \boldsymbol{X}$ is $\boldsymbol{\lambda}^\top \boldsymbol{x}$ for an allocation $\boldsymbol{x} \in \mathcal{K}_d(K)$. The value $r(\lambda) = \varrho(\boldsymbol{\lambda}^\top \boldsymbol{X})$ is interpreted as a stand-alone capital that would have been required if the total loss $\boldsymbol{\lambda}^\top \boldsymbol{X}$ had been managed individually. Therefore, under the

core allocation $\boldsymbol{x} \in \mathcal{K}_d^{\mathrm{C}}(K;r)$, the subportfolio $(\lambda_1 X_1, \dots, \lambda_d X_d)$ gains benefit of capital reduction by $\boldsymbol{\lambda}^\top \boldsymbol{x}$ in comparison to $r(\boldsymbol{\lambda})$.

Given K, r and the joint loss $\boldsymbol{X}$, we are interested in calculating the core-compatible versions of Euler allocation $\mathbb{E}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}]$, MLA $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K;r)]$ and local modes of $f_{\boldsymbol{X}\mid\{\boldsymbol{X}\in\mathcal{K}_d^{\mathrm{C}}(K;r)\}}$ if they exist. However, generating a large number of samples from $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}$ is computationally involved since an unconditional sample $\boldsymbol{X}$ is first filtered by the condition $\boldsymbol{X} \in \{\boldsymbol{x} \in \mathbb{R}^d : K-\delta < \mathbf{1}_d^\top \boldsymbol{x} < K+\delta\} = \mathcal{K}_d(K,\delta)$ for a small $\delta > 0$, and then filtered again by the core condition $\boldsymbol{\lambda}^\top \boldsymbol{X} \le r(\boldsymbol{\lambda})$ for all possible $\boldsymbol{\lambda} \in \{0,1\}^d$. To overcome the issue, we utilize the *Hamiltonian Monte Carlo (HMC) method with reflection* to directly simulate $f_{\boldsymbol{X}'\mid\{\boldsymbol{X}\in\mathcal{K}_d^{\mathrm{C}}(K;r)\}}$. Note that the support of $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}$ is a projection of $\mathcal{K}_d^{\mathrm{C}}(K;r)$ onto $\mathbb{R}^{d'}$, which is an intersection of hyperplanes $\{\boldsymbol{x}' \in \mathbb{R}^{d'} : \boldsymbol{\lambda}^\top(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}') \le r(\boldsymbol{\lambda})\}$ for $\boldsymbol{\lambda} \in \{0,1\}^d$. In the HMC method, a candidate is proposed according to the so-called Hamiltonian dynamics, and the chain reflects at the boundaries $\{\boldsymbol{x}' \in \mathbb{R}^{d'} : \boldsymbol{\lambda}^\top(\boldsymbol{x}', K - \mathbf{1}_{d'}^\top \boldsymbol{x}') = r(\boldsymbol{\lambda})\}$, $\boldsymbol{\lambda} \in \{0,1\}^d$ so that it does not violate the support constraint; see Section 5.3.2.

For a numerical experiment, let $\boldsymbol{X} \sim t_\nu(\mathbf{0}_d, P)$ with $d = 3$, $\nu = 5$ and $P = (\rho_{ij})$ being a correlation matrix with $\rho_{12} = \rho_{23} = 1/3$ and $\rho_{13} = 2/3$. For $p = 0.99$, we set $r(\boldsymbol{\lambda}) = \mathrm{VaR}_p(\boldsymbol{\lambda}^\top \boldsymbol{X})$ for $\boldsymbol{\lambda} \in \{0,1\}^3$ and $K = r(\mathbf{1}_3)$. For $\delta = 0.001$, we first generate $N_{\mathrm{MC}} = 10^6$ samples from $\boldsymbol{X}$ and estimate K and $(r(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \{0,1\}^3)$ from these samples. Then we extract samples of $\boldsymbol{X}$ falling in the region

$$\mathcal{K}_d^{\mathrm{C}}(K,\delta;r) = \mathcal{K}_d(K,\delta) \cap \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{\lambda}^\top \boldsymbol{x} \le r(\boldsymbol{\lambda}),\ \boldsymbol{\lambda} \in \{0,1\}^3 \backslash \{\mathbf{1}_3\}\}.$$

Figure 6.4 (a) shows the first two components of the MC samples from $\boldsymbol{X}$ and the conditional samples falling in $\mathcal{K}_d^{\mathrm{C}}(K,\delta;r)$. Among the $N_{\mathrm{MC}} = 10^6$ samples, 2000 samples were contained in $\mathcal{K}_d(K,\delta)$ and only 189 samples fell in $\mathcal{K}_d^{\mathrm{C}}(K,\delta;r)$. Therefore, this crude simulation method is not efficient since 99.98% of the unconditional samples are discarded.

Instead, we conduct an MCMC simulation to generate $N_{\mathrm{MCMC}} = 10^4$ samples directly from $\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}$. Hyperparameters of the HMC method are estimated based on the 189 MC samples; see Section 5.3.2 in Chapter 5. The resulting stepsize and integration time are $\varepsilon = 0.105$ and $T = 24$, respectively. It took 49.534 seconds to simulate a Markov chain with length $N_{\mathrm{MCMC}} = 10^4$. The resulting acceptance rate was 0.866 and serial correlations were below 0.03 at lag 1. Based on these inspections we conclude that the MCMC method performed correctly. The first 3000 MCMC samples of $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}$ are plotted in Figure 6.4 (b).

By Proposition 6.3.1, $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d(K)\}$ still follows a multivariate Student t distribution, and thus the mode of this conditional distribution is uniquely determined by $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)] = \mathbb{E}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d(K)\}]$ by Part 1 of Proposition 6.3.13. Moreover, when this point is contained in the core $\boldsymbol{K}_d^{\mathrm{C}}(K;r)$, we have $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)] = \boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K;r)]$ since the distributions of $\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d(K)\}$ and $\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}$ share the same mode. We check these observations numerically by calculating the corresponding estimates.

Table 6.4 summarizes the MC and MCMC estimates and standard errors of the Euler and maximum likelihood allocations on $\mathcal{K}_d(K)$ and those on the atomic core $\mathcal{K}_d^{\mathrm{C}}(K;r)$. MC estimates are calculated based on the samples in Figure 6.4 (a) and MCMC estimates are computed based on the samples in Figure 6.4 (b). As expected by theory, the MC estimates of $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d(K)]$ and $\mathbb{E}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d(K)\}]$ were close to each

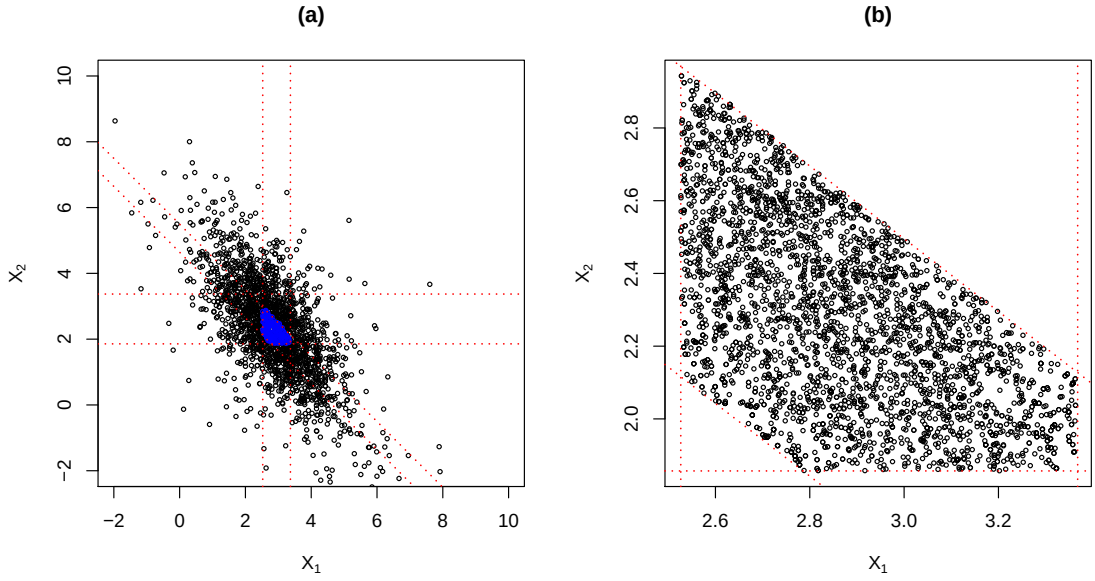

Figure 6.4: Scatter plots of (a) MC samples from $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d(K, \delta)\}$ (black) and $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K, \delta; r)\}$ (blue), and of (b) MCMC samples from $\boldsymbol{X}' \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K; r)\}$ (black) where $\boldsymbol{X} \sim t_\nu(\boldsymbol{0}_d, P)$ with $d = 3$, $\nu = 5$ and $P = (\rho_{i,j})$ being a correlation matrix with $\rho_{1,2} = \rho_{2,3} = 1/3$ and $\rho_{1,3} = 2/3$, $r(\boldsymbol{\lambda}) = \mathrm{VaR}_p(\boldsymbol{\lambda}^\top \boldsymbol{X})$ with $p = 0.99$ for $\boldsymbol{\lambda} \in \{0, 1\}^3$, $K = r(\boldsymbol{1}_3)$ and $\delta = 0.001$. Red lines indicate $\{\boldsymbol{x}' \in \mathbb{R}^2 : \boldsymbol{\lambda}^\top (\boldsymbol{x}', K - \boldsymbol{1}_2^\top \boldsymbol{x}') = r(\boldsymbol{\lambda})\}$ for $\boldsymbol{\lambda} \in \{0, 1\}^3$.

other. We can also observe that the MC and MCMC estimates are close to each other for all the estimators. The standard errors of the MCMC estimator of $\mathbb{E}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K; r)\}]$ were smaller than those of the MC estimator because of sample efficiency. Provided that $\hat{\boldsymbol{K}}_{\mathrm{M}}^{\mathrm{MC}}[\boldsymbol{X}; \mathcal{K}_d(K)]$ belongs to the core $\mathcal{K}_d^{\mathrm{C}}(K; r)$, we expect an estimate of $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K; r)]$ to be close to $\hat{\boldsymbol{K}}_{\mathrm{M}}^{\mathrm{MC}}[\boldsymbol{X}; \mathcal{K}_d(K)]$. Although this was the case for both of the MC and MCMC estimates of $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K; r)]$, the MCMC estimate was slightly closer to $\hat{\boldsymbol{K}}_{\mathrm{M}}^{\mathrm{MC}}[\boldsymbol{X}; \mathcal{K}_d(K)]$ than the MC estimate. Consequently, the MCMC estimator of $\boldsymbol{K}_{\mathrm{M}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K; r)]$ is less biased than the MC estimator.

6.6 Conclusion

Motivated from scenario analysis of risk allocations, we investigated properties of the conditional distribution of $\boldsymbol{X}$ given the constant sum constraint $\{S = K\}$ and introduced the novel risk allocation method called maximum likelihood allocation (MLA). The superlevel set of $\boldsymbol{X} \mid \{S = K\}$ can be regarded as a set of stress (severe and plausible) scenarios, and the modality of $\boldsymbol{X} \mid \{S = K\}$ can be interpreted as a number of distinct risky scenarios, which turned out to be an important feature in risk profile related to the soundness of risk allocations. We then studied properties of $\boldsymbol{X} \mid \{S = K\}$, for example, dependence (Proposition 6.3.4 and Proposition 6.3.5), tail behavior (Proposition 6.3.9 and Proposition 6.3.10) and modality (Proposition 6.3.13), most of which are inherited from those of the unconditional loss $\boldsymbol{X}$. We

Table 6.4: Monte Carlo (superscript "MC") and Markov chain Monte Carlo (superscript "MCMC") estimates and standard errors of the Euler and maximum likelihood allocations on $\mathcal{K}_d(K)$ and those on the atomic core $\mathcal{K}_d^{\mathrm{C}}(K;r)$. The MC sample size of the unconditional sample $\boldsymbol{X}$ is $N_{\mathrm{MC}} = 10^6$ and the sample size of the conditional sample $\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}$ in the MCMC method is $N_{\mathrm{MCMC}} = 10^4$.

	Estimator			Standard error		
	X_1	X_2	X_3	X_1	X_2	X_3
$\hat{\mathbb{E}}^{\mathrm{MC}}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d(K)\}]$	2.865	2.310	2.846	0.026	0.034	0.026
$\hat{\boldsymbol{K}}_{\mathrm{M}}^{\mathrm{MC}}[\boldsymbol{X}; \mathcal{K}_d(K)]$	2.861	2.366	2.793	–	–	–
$\hat{\mathbb{E}}^{\mathrm{MC}}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}]$	2.852	2.267	2.903	0.016	0.019	0.016
$\hat{\boldsymbol{K}}_{\mathrm{M}}^{\mathrm{MC}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K;r)]$	2.838	2.262	2.920	–	–	–
$\hat{\mathbb{E}}^{\mathrm{MCMC}}[\boldsymbol{X} \mid \{\boldsymbol{X} \in \mathcal{K}_d^{\mathrm{C}}(K;r)\}]$	2.876	2.269	2.877	0.002	0.003	0.002
$\hat{\boldsymbol{K}}_{\mathrm{M}}^{\mathrm{MCMC}}[\boldsymbol{X}; \mathcal{K}_d^{\mathrm{C}}(K;r)]$	2.866	2.283	2.871	–	–	–

also investigated how to incorporate the knowledge on the modality of $\boldsymbol{X} \mid \{S = K\}$ for more sound risk management. Under unimodality, we defined MLA as a mode of $\boldsymbol{X} \mid \{S = K\}$, and studied its properties as a risk allocation, such as translation invariance and positive homogeneity (Proposition 6.4.2). Under multimodality, we considered the so-called multimodality adjustment to increase the soundness of risk allocations based on the multiple modes. Euler allocation and MLA were then compared in numerical experiments. Through these experiments, we demonstrated that Euler allocation and MLA lead to close values and $\boldsymbol{X} \mid \{S = K\}$ is typically unimodal when $\boldsymbol{X}$ possesses positive dependence. On the other hand, when the losses are negatively dependent, multimodality is likely to occur and the two allocation principles result in distinct values. For such a case, searching for the modes of $\boldsymbol{X} \mid \{S = K\}$ is beneficial for discovering risky scenarios which cannot be captured by a single vector of risk allocation, and to inspect the soundness of risk allocations. The detected (local) modes can also be useful to increase the soundness of risk allocations by applying the multimodality adjustment.

Although we empirically observed the relationship between multimodality of $\boldsymbol{X} \mid \{S = K\}$ and negative dependence among $\boldsymbol{X}$, this relationship requires further theoretical investigation. Another aspect of future research is to study more distributional properties, such as tail dependence and measures of concordance, of $\boldsymbol{X} \mid \{S = K\}$ especially without assuming the existence of a density. Unlike Euler allocations, estimation of MLAs is not a straightforward problem in general but various methods are known for estimating modes of multivariate distributions and for selections of hyperparameters. For applying the MLA principle in practice, efficient estimation methods of the modes of multivariate loss distributions in high dimensions need to be explored further. An economic justification of the MLA principle is also an interesting direction for future research. In addition, extension of the multimodality adjustment to general sets of scenarios is also an interesting direction of future work since multimodality adjustment proposed in this chapter relies on the assumption that the set of modes is discrete. In the end, efficient simulation approaches of $\boldsymbol{X} \mid \{S = K\}$ may need to rely on MCMC methods as introduced in Section 6.5.3, and further investigation

is required to assess in how far the distributional properties proven in this chapter carry over to MCMC methods since the performance of MCMC methods typically depends on tail-heaviness and modality of the target distribution.

Chapter 7

Conclusion

In this thesis, we addressed various topics in the areas of probability theory, statistics and their applications to quantitative risk management.

In Chapter 2, we introduced a new class of measures of concordance arising from Pearson's linear correlation of transformed random variables. We provided necessary and sufficient conditions on the transformations for which the transformed rank correlation coefficients are measures of concordance. For matrices of pairwise transformed rank correlation coefficients, compatibility and attainability problems were investigated. We derived upper and lower bounds of the compatible sets of transformed rank correlations, and provide an algorithm to simulate compatible matrices when a given matrix is Bern(1/2)-compatible. We then studied dimension reduction of the compatibility and attainability problems for block matrices and hierarchical matrices.

In Chapter 3, we addressed the question which measures of concordance to use in terms of best and worst asymptotic variances of their canonical estimators. We proved that Blomqvist's beta attains the optimal best and worst asymptotic variances among all transformed rank correlation coefficients including Spearman's rho and van der Waerden's coefficient. Moeover, We found that Kendall's tau also attains the optimal best and worst asymptotic variances if estimators of these measures are compared without being standardized by sample size. The results of the simulation study supported that concordance-inducing functions G with smaller $\mathrm{Var}_G(X^2)$, where $X \sim G$, are more preferable. Consequently, heavy-tailed concordance-inducing functions, such as Student t distributions with small degrees of freedom, are not recommended in comparison to the normal distribution, and Beta distributions can be good alternatives for uniform distributions (corresponding to Spearman's rho).

In Chapter 4, we proposed an Metropolis Hastings (MH) estimator of Value-at-Risk (VaR) contributions to achieve efficient computation of VaR contributions especially for a risk model specified by a joint density. Sample efficiency of the MH estimator is significantly improved since the MH method generates samples directly from the conditional density given the sum constraint. By the general theory of Markov chains, the MH estimator is consistent and asymptotically normal. Through simulation and empirical studies, bias and mean squared error (MSE) of the MH estimator were compared with those of other existing estimators.

The numerical experiments demonstrated that in most risk models, the MH estimator had smaller bias and MSE compared with other existing estimators even when the dimension of the portfolio was high, such as $d \approx 500$.

The framework of estimating VaR contributions with Markov Chain Monte Carlo (MCMC) methods was then extended to a more general class of systemic risk allocations in Chapter 5. By using Hamiltonian Monte Carlo (HMC) and Gibbs Sampler (GS), efficient simulation methods of the constrained target distributions are obtained. Sample efficiency is significantly improved since the MCMC estimator is computed from samples generated directly from the conditional distribution of interest. We also proposed a heuristic for determining the parameters of the HMC method based on the Monte Carlo (MC) presamples. Numerical experiments demonstrated that our MCMC estimators are more efficient than MC in terms of bias, standard error and time-adjusted MSE. Stability of the MCMC estimation with respect to the probability of the crisis event and efficiency of the optimal parameter choice of the HMC method were also investigated in the experiments.

In Chapter 6, we investigated properties of the conditional distribution of $\boldsymbol{X}$ given the constant sum constraint $\{S = K\}$ and introduced the novel risk allocation method called maximum likelihood allocation (MLA). The superlevel set of $\boldsymbol{X} \mid \{S = K\}$ can be regarded as a set of stress (severe and plausible) scenarios, and the modality of $\boldsymbol{X} \mid \{S = K\}$ can be interpreted as a number of distinct risky scenarios, which turned out to be an important feature in risk profile related to the soundness of risk allocations. We then studied properties of $\boldsymbol{X} \mid \{S = K\}$, for example, dependence (Proposition 6.3.4 and Proposition 6.3.5), tail behavior (Proposition 6.3.9 and Proposition 6.3.10) and modality (Proposition 6.3.13), most of which are inherited from those of the unconditional loss $\boldsymbol{X}$. We also investigated how to incorporate the knowledge on the modality of $\boldsymbol{X} \mid \{S = K\}$ for more sound risk management. Under unimodality, we defined MLA as a mode of $\boldsymbol{X} \mid \{S = K\}$, and studied its properties as a risk allocation, such as translation invariance and positive homogeneity (Proposition 6.4.2). Under multimodality, we considered the so-called multimodality adjustment to increase the soundness of risk allocations based on the multiple modes. Euler allocation and MLA were then compared in numerical experiments. Through these experiments, we demonstrated that Euler allocation and MLA lead to close values and $\boldsymbol{X} \mid \{S = K\}$ is typically unimodal when $\boldsymbol{X}$ possesses positive dependence. On the other hand, when the losses are negatively dependent, multimodality is likely to occur and the two allocation principles result in distinct values. For such a case, searching for the modes of $\boldsymbol{X} \mid \{S = K\}$ is beneficial for discovering risky scenarios which cannot be captured by a single vector of risk allocation, and to inspect the soundness of risk allocations. The detected (local) modes can also be useful to increase the soundness of risk allocations by applying the multimodality adjustment.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47.

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, 106(7):1705.

Afshar, H. M. and Domke, J. (2015). Reflection, refraction, and hamiltonian monte carlo. In *Advances in Neural Information Processing Systems*, pages 3007–3015. The MIT Press.

Asimit, A. V., Furman, E., Tang, Q., and Vernic, R. (2011). Asymptotics for risk capital allocations based on conditional tail expectation. *Insurance: Mathematics and Economics*, 49(3):310–324.

Asimit, A. V. and Li, J. (2018). Systemic risk: An asymptotic evaluation. *ASTIN Bulletin: The Journal of the IAA*, 48(2):673–698.

Asimit, V., Peng, L., Wang, R., and Yu, A. (2019). An efficient approach to quantile capital allocation and sensitivity analysis. *Mathematical Finance*, 29(4):1131–1156.

Aumann, R. J. and Shapley, L. S. (2015). *Values of non-atomic games*. Princeton University Press, Princeton, New Jersey.

Balkema, G. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432.

Behnen, K. (1971). Asymptotic optimality and are of certain rank-order tests under contiguity. *The Annals of Mathematical Statistics*, pages 325–329.

Behnen, K. (1972). A characterization of certain rank-order tests with bounds for the asymptotic relative efficiency. *The Annals of Mathematical Statistics*, pages 1839–1851.

Berge, C. (1997). *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation.

Bernardi, M., Durante, F., and Jaworski, P. (2017). Covar of families of copulas. *Statistics & Probability Letters*, 120:8–17.

Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M., and Stuart, A. (2010). The acceptance probability of the hybrid monte carlo method in high-dimensional problems. In *AIP Conference Proceedings*, volume 1281, pages 23–26. AIP.

Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M., Stuart, A., et al. (2013). Optimal tuning of the hybrid monte carlo algorithm. *Bernoulli*, 19(5A):1501–1534.

Betancourt, M. (2012). Cruising the simplex: Hamiltonian monte carlo and the dirichlet distribution. In *AIP Conference Proceedings 31st*, volume 1443, pages 157–164. AIP.

Betancourt, M. (2013). A general metric for riemannian manifold hamiltonian monte carlo. In *International Conference on Geometric Science of Information*, pages 327–334. Springer.

Betancourt, M. (2016). Identifying the optimal integration time in hamiltonian monte carlo. *arXiv preprint arXiv:1601.00225*.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*.

Betancourt, M., Byrne, S., and Girolami, M. (2014). Optimizing the integrator step size for hamiltonian monte carlo. *arXiv preprint arXiv:1411.6669*.

Bhuchongkul, S. (1964). A class of nonparametric tests for independence in bivariate populations. *The Annals of Mathematical Statistics*, pages 138–149.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, pages 593–600.

Boonen, T. J., De Waegenaere, A., and Norde, H. (2020). A generalization of the aumann–shapley value for risk capital allocation problems. *European Journal of Operational Research*, 282(1):277–287.

Breuer, T., Jandacka, M., Rheinberger, K., Summer, M., et al. (2009). *How to find plausible, severe, and useful stress scenarios*. Österr. Nationalbank.

Byrne, S. and Girolami, M. (2013). Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845.

Cambou, M., Hofert, M., and Lemieux, C. (2017). Quasi-random numbers for copula models. *Statistics and Computing*, 27(5):1307–1329.

Cances, E., Legoll, F., and Stoltz, G. (2007). Theoretical and numerical comparison of some sampling methods for molecular dynamics. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(2):351–389.

Carley, H. and Taylor, M. D. (2002). A new proof of sklar's theorem. In *Distributions with Given Marginals and Statistical Modelling*, pages 29–34. Springer.

Carreira-Perpinán, M. A. (2015). A review of mean-shift algorithms for clustering. *arXiv preprint arXiv:1503.00687*.

Charpentier, A. and Segers, J. (2007). Lower tail dependence for archimedean copulas: characterizations and pitfalls. *Insurance: Mathematics and Economics*, 40(3):525–532.

Chen, C., Iyengar, G., and Moallemi, C. C. (2013). An axiomatic approach to systemic risk. *Management Science*, 59(6):1373–1388.

Chen, Y.-C., Genovese, C. R., Wasserman, L., et al. (2016). A comprehensive approach to mode clustering. *Electronic Journal of Statistics*, 10(1):210–241.

Chevallier, A., Pion, S., and Cazals, F. (2018). Hamiltonian monte carlo with boundary reflections, and application to polytope volume calculations.

Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

Chiragiev, A. and Landsman, Z. (2007). Multivariate pareto portfolios: Tce-based capital allocation and divided differences. *Scandinavian Actuarial Journal*, 2007(4):261–280.

Christen, J. A., Fox, C., and Santana-Cibrian, M. (2017). Optimal direction gibbs sampler for truncated multivariate normal distributions. *Communications in Statistics-Simulation and Computation*, 46(4):2587–2600.

De Winter, J. C., Gosling, S. D., and Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3):273.

Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International statistical review*, 73(1):111–129.

Denault, M. (2001). Coherent allocation of risk capital. *Journal of Risk*, 4(1):1–34.

Dev, A. (2004). *Economic capital: a practitioner guide*. Risk Books, New York.

Devroye, L. (1985). Non-uniform random variate generation. Springer, New York.

Devroye, L. and Letac, G. (2015). Copulas with prescribed correlation matrix. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, pages 585–601. Springer.

Dhaene, J., Henrard, L., Landsman, Z., Vandendorpe, A., and Vanduffel, S. (2008). Some results on the cte-based capital allocation rule. *Insurance: Mathematics and Economics*, 42(2):855–863.

Dhaene, J., Tsanakas, A., Valdez, E. A., and Vanduffel, S. (2012). Optimal capital allocation principles. *Journal of Risk and Insurance*, 79(1):1–28.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, convexity, and applications*. Elsevier.

Ding, P. (2016). On the conditional distribution of the multivariate t distribution. *The American Statistician*, 70(3):293–295.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2):216–222.

Durmus, A., Moulines, E., and Saksman, E. (2017). On the convergence of hamiltonian monte carlo. *arXiv preprint arXiv:1705.00166*.

Edwards, H. H., Mikusiński, P., and Taylor, M. D. (2004). Measures of concordance determined by d_4-invariant copulas. *International Journal of Mathematics and Mathematical Sciences*, 2004(70):3867–3875.

Edwards, H. H., Mikusiński, P., and Taylor, M. D. (2005). Measures of concordance determined by d_4-invariant measures on $(0,1)^2$. *Proceedings of the American Mathematical Society*, 133(5):1505–1513.

Edwards, H. H. and Taylor, M. D. (2009). Characterizations of degree one bivariate measures of concordance. *Journal of Multivariate Analysis*, 100(8):1777–1791.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432.

Embrechts, P., Hofert, M., and Wang, R. (2016). Bernoulli and tail-dependence compatibility. *The Annals of Applied Probability*, 26(3):1636–1658.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependency in risk management: Properties and pitfalls. In Dempster, M., editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press.

Fan, G., Zeng, Y., and Wong, W. K. (2012). Decomposition of portfolio var and expected shortfall based on multivariate copula simulation. *International Journal of Management Science and Engineering Management*, 7(2):153–160.

Fang, K. W. (2018). *Symmetric multivariate and related distributions*. Chapman and Hall/CRC.

Feller, W. (2008). *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons.

Fernandez, C., Osiewalski, J., and Steel, M. F. (1995). Modeling and inference with v-spherical distributions. *Journal of the American Statistical Association*, 90(432):1331–1340.

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Fischer, M., Moser, T., and Pfeuffer, M. (2018). A discussion on recent risk measures with application to credit risk: Calculating risk contributions and identifying risk concentrations. *Risks*, 6(4):142.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American actuarial journal*, 2(1):1–25.

Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40.

Furman, E., Kuznetsov, A., and Zitikis, R. (2018). Weighted risk capital allocations in the presence of systematic risk. *Insurance: Mathematics and Economics*, 79:75–81.

Furman, E. and Landsman, Z. (2008). Economic capital allocations for non-negative portfolios of dependent risks. *ASTIN Bulletin: The Journal of the IAA*, 38(2):601–619.

Furman, E., Wang, R., and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking & Finance*, 83:70–84.

Furman, E. and Zitikis, R. (2008). Weighted risk capital allocations. *Insurance: Mathematics and Economics*, 43(2):263–269.

Furman, E. and Zitikis, R. (2009). Weighted pricing functionals with applications to insurance: an overview. *North American Actuarial Journal*, 13(4):483–496.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Gelfand, A. E., Smith, A. F., and Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of copulas in finance. *The European Journal of Finance*, 15:609–618.

Genest, C. and Verret, F. (2005). Locally most powerful rank tests of independence for copula models. *Nonparametric Statistics*, 17(5):521–539.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pages 571–578. Fairfax, Virginia: Interface Foundation of North America, Inc.

Geyer, C. (2011). Introduction to markov chain monte carlo. In *Handbook of Markov Chain Monte Carlo*, pages 3–47. Springer, New York.

Girardi, G. and Ergün, A. T. (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, 37(8):3169–3180.

Girolami, M. and Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214.

Glasserman, P. (2005). Measuring marginal risk contributions in credit portfolios. *Journal of Computational Finance*, 9:1–41.

Glasserman, P. (2013). *Monte Carlo methods in financial engineering*. Springer, New York.

Glasserman, P. and Li, J. (2005). Importance sampling for portfolio credit risk. *Management science*, 51(11):1643–1656.

Górecki, J., Hofert, M., and Holeňa, M. (2017). On structure, family and parameter estimation of hierarchical Archimedean copulas. *Journal of Statistical Computation and Simulation*, 87(17):3261–3324.

Gourieroux, C. and Monfort, A. (2013). Allocating systemic risk in a regulatory perspective. *International Journal of Theoretical and Applied Finance*, 16(07):1350041.

Gudmundsson, T. and Hult, H. (2014). Markov chain monte carlo for computing rare-event probabilities for a heavy-tailed random walk. *Journal of Applied Probability*, 51(2):359–376.

Gupta, S., Irbäc, A., Karsch, F., and Petersson, B. (1990). The acceptance probability in the hybrid monte carlo method. *Physics Letters B*, 242(3-4):437–443.

Hallerbach, W. G. (2003). Decomposing portfolio value-at-risk: A general analysis. *Journal of Risk*, 5(2):1–18.

Hansen, B. (2009). Nonparametric regression. Lecture Note.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hofert, M. (2010). *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. Südwestdeutscher Verlag für Hochschulschriften AG & Co. KG. PhD thesis.

Hofert, M. (2011). Efficiently sampling nested Archimedean copulas. *Computational Statistics & Data Analysis*, 55:57–70.

Hofert, M. (2012). A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 82(9):1239–1255.

Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2018). *Elements of Copula Modeling with R*. Springer Use R! Series.

Hofert, M. and Scherer, M. (2011). CDO pricing with nested Archimedean copulas. *Quantitative Finance*, 11(5):775–787.

Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Hoffmann, H., Meyer-Brandis, T., and Svindland, G. (2016). Risk-consistent conditional systemic risk measures. *Stochastic Processes and their Applications*, 126(7):2014–2037.

Hogg, R. V. and Klugman, S. A. (2009). *Loss distributions*, volume 249. John Wiley & Sons.

Huang, J. and Yang, L. (2010). Correlation matrix with block structure and efficient sampling methods. *Journal of Computational Finance*, 14(1):81.

Huang, J.-J., Lee, K.-J., Liang, H., and Lin, W.-F. (2009). Estimating value at risk of portfolio by conditional copula-garch method. *Insurance: Mathematics and economics*, 45(3):315–324.

Huber, M. and Maric, N. (2015). Multivariate distributions with fixed marginals and correlations. *Journal of Applied Probability*, 52(2):602–608.

Huber, M. and Maric, N. (2017). Bernoulli correlations and cut polytopes. *arXiv preprint arXiv:1706.06182*.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied probability*, pages 587–608.

Hürlimann, W. (2012). On trivariate copulas with bivariate linear spearman marginal copulas. *Journal of Mathematics and System Science*, 2(6):368–383.

Hürlimann, W. (2014). A closed-form universal trivariate pair-copula. *Journal of Statistical Distributions and Applications*, 1(1):7.

Jaworski, P. (2017). On conditional value at risk (covar) for tail-dependent copulas. *Dependence Modeling*, 5(1):1–19.

Jaworski, P., Durante, F., Hardle, W. K., and Rychlik, T., editors (2010). *Copula Theory and Its Applications*, volume 198 of *Lecture Notes in Statistics – Proceedings*. Springer.

Joe, H. (1990). Multivariate concordance. *Journal of Multivariate Analysis*, 35(1):12–30.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, Dordrecht.

Joe, H. (2014). *Dependence modeling with copulas*. CRC Press, Florida.

Joe, H. and Li, H. (2019). Tail densities of skew-elliptical distributions. *Journal of Multivariate Analysis*, 171:421–435.

Johnson, A. (2009). Geometric ergodicity of gibbs samplers phd thesis. *University of Minnesota, School of Statistics*.

Jondeau, E. and Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. *Journal of international money and finance*, 25(5):827–853.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for markov chain monte carlo. *Journal of the American Statistical Association*, 101(476):1537–1547.

Juri, A. and Wüthrich, M. V. (2002). Copula convergence theorems for tail events. *Insurance: Mathematics and Economics*, 30(3):405–420.

Juri, A. and Wüthrich, M. V. (2003). Tail dependence from a distributional point of view. *Extremes*, 6(3):213–246.

Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, 15(3):425–437.

Kalkbrener, M., Lotter, H., and Overbeck, L. (2004). Sensible and efficient capital allocation for credit portfolios. *Risk*, 17(1):S19–S24.

Kamatani, K. (2017). Ergodicity of markov chain monte carlo with reversible proposal. *Journal of Applied Probability*, 54(2):638–654.

Kamatani, K. et al. (2018). Efficient strategy for the markov chain monte carlo in high-dimension with heavy-tailed target probability distribution. *Bernoulli*, 24(4B):3711–3750.

Karlin, S. and Rinott, Y. (1980a). Classes of orderings of measures and related correlation inequalities. i. multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498.

Karlin, S. and Rinott, Y. (1980b). Classes of orderings of measures and related correlation inequalities ii. multivariate reverse rule distributions. *Journal of Multivariate Analysis*, 10(4):499–516.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae*, 10:152–164.

Kley, O., Klüppelberg, C., and Reinert, G. (2016). Risk in a large claims insurance market with bipartite graph structure. *Operations Research*, 64(5):1159–1176.

Klugman, S. A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance: mathematics and economics*, 24(1-2):139–148.

Kromer, E., Overbeck, L., and Zilch, K. (2016). Systemic risk measures on general measurable spaces. *Mathematical Methods of Operations Research*, 84(2):323–357.

Kurowicka, D. and Cooke, R. (2001). Conditional, partial and rank correlation for the elliptical copula; dependence modelling in uncertainty analysis. In *Proceedings ESREL*, pages 1795–1802.

Laeven, R. J. and Goovaerts, M. J. (2004). An optimization approach to the dynamic allocation of economic capital. *Insurance: Mathematics and Economics*, 35(2):299–319.

Lan, S., Streets, J., and Shahbaba, B. (2014). Wormhole hamiltonian monte carlo. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Landsman, Z. M. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71.

Larsson, M. and Nešlehová, J. (2011). Extremal behavior of archimedean copulas. *Advances in Applied Probability*, 43(1):195–216.

Laurent, M. and Poljak, S. (1995). On a positive semidefinite relaxation of the cut polytope. *Linear Algebra and its Applications*, 223:439–461.

Leimkuhler, B. and Reich, S. (2004). *Simulating hamiltonian dynamics*, volume 14. Cambridge university press.

Li, H. (2013). Toward a copula theory for multivariate regular variation. In *Copulae in mathematical and quantitative finance*, pages 177–199. Springer.

Li, H. and Hua, L. (2015). Higher order tail densities of copulas and hidden regular variation. *Journal of Multivariate Analysis*, 138:143–155.

Li, H. and Wu, P. (2013). Extremal dependence of copulas: A tail density approach. *Journal of Multivariate Analysis*, 114:99–111.

Liu, J. S., Wong, W. H., and Kong, A. (1995). Covariance structure and convergence rate of the gibbs sampler with various scans. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):157–169.

Livingstone, S., Betancourt, M., Byrne, S., Girolami, M., et al. (2019a). On the geometric ergodicity of hamiltonian monte carlo. *Bernoulli*, 25(4A):3109–3138.

Livingstone, S., Faulkner, M. F., and Roberts, G. O. (2019b). Kinetic energy choice in hamiltonian/hybrid monte carlo. *Biometrika*, 106(2):303–319.

Livingstone, S. and Girolami, M. (2014). Information-geometric markov chain monte carlo methods using diffusions. *Entropy*, 16(6):3074–3102.

Luigi Conti, P. and Nikitin, Y. (1999). Asymptotic efficiency of independence tests based on gini's rank association coefficient, spearman's footrule and their generalizations. *Communications in Statistics-Theory and Methods*, 28(2):453–465.

Mainik, G. and Schaanning, E. (2014). On dependence consistency of covar and some other systemic risk measures. *Statistics & Risk Modeling*, 31(1):49–77.

Maume-Deschamps, V., Rullière, D., and Said, K. (2016). On a capital allocation by minimization of some risk indicators. *European Actuarial Journal*, 6(1):177–196.

McNeil, A. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press, Princeton.

McNeil, A. J., Nešlehová, J., et al. (2009). Multivariate archimedean copulas, d-monotone functions and âĎŞ1-norm symmetric distributions. *The Annals of Statistics*, 37(5B):3059–3097.

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the hastings and metropolis algorithms. *The annals of Statistics*, 24(1):101–121.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Mikosch, T. (1999). Regular variation, subexponentiality and their applications in probability theory. Technical Report 99-013, University of Groningen.

Mosteller, F. (2006). On some useful "inefficient" statistics. In *Selected Papers of Frederick Mosteller*, pages 69–100. Springer.

Müller, A. and Scarsini, M. (2000). Some remarks on the supermodular order. *Journal of Multivariate Analysis*, 73:107–119.

Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks*, volume 389. Wiley New York.

Müller, P. (1992). Alternatives to the gibbs sampling scheme. *Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.5613*.

Murdoch, D. J., Craiu, R. V., and Meng, X.-L. (2001). On the edge: Statistics & computing: Chance and fractals. *Chance*, 14(2):47–52.

Myers, S. C. and Read, J. A. (2001). Capital allocation for insurance companies. *Journal of Risk and Insurance*, pages 545–580.

Neal, R. M. et al. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11).

Nelsen, R. B. (2006). *An introduction to copulas*. Springer, New York.

Norkin, V. and Roenko, N. (1991). α-concave functions and measures and their applications. *Cybernetics and Systems Analysis*, 27(6):860–869.

Nummelin, E. (2002). Mc's for mcmc'ists. *International Statistical Review*, 70(2):215–240.

Nummelin, E. (2004). *General irreducible Markov chains and non-negative operators*. Cambridge University Press, Cambridge.

Osiewalski, J. (1993). Robust bayesian inference in lq-spherical models. *Biometrika*, 80(2):456–460.

Pagan, A. and Ullah, A. (1999). *Nonparametric econometrics*. Cambridge University Press, Cambridge.

Pakman, A. and Paninski, L. (2014). Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542.

Pitowsky, I. (1991). Correlation polytopes: their geometry and complexity. *Mathematical Programming*, 50(1-3):395–414.

Raymaekers, J. and Rousseeuw, P. J. (2019). Fast robust correlation for high-dimensional data. *Technometrics*, pages 1–15.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Resnick, S. I. (2013). *Extreme values, regular variation and point processes*. Springer, New York.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space markov chains and mcmc algorithms. *Probability Surveys*, 1:20–71.

Rödel, E. and Kössler, W. (2004). Linear rank tests for independence in bivariate distributions-power comparisons by simulation. *Computational statistics & data analysis*, 46(4):645–660.

Rodriguez-Yam, G., Davis, R. A., and Scharf, L. L. (2004). Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. *Unpublished manuscript*.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for markov chain monte carlo. *Journal of the American Statistical Association*, 90(430):558–566.

Rosenthal, J. S. et al. (2011). Optimal proposal distributions and adaptive mcmc. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, Florida.

Roth, M. (2012). *On the multivariate t distribution*. Linköping University Electronic Press.

Roustant, O. and Deville, Y. (2017). On the validity of parametric block correlation matrices with constant within and between group correlations. *arXiv preprint arXiv:1705.09793*.

Ruján, P. (1997). Playing billiards in version space. *Neural Computation*, 9(1):99–122.

Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: a review. *Statistics Surveys*, 8:45.

Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8(3):201–218.

Schmid, F. and Schmidt, R. (2007). Nonparametric inference on multivariate versions of blomqvist's beta and related measures of tail dependence. *Metrika*, 66(3):323–354.

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research*, 55(2):301–327.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer Science & Business Media.

Sidak, Z., Sen, P. K., and Hajek, J. (1999). *Theory of rank tests*. Elsevier.

Siller, T. (2013). Measuring marginal risk contributions in credit portfolios. *Quantitative Finance*, 13(12):1915–1923.

Spearman, C. (1904). "general intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292.

Sweeting, T. J. et al. (1986). On a converse to scheffé's theorem. *The Annals of Statistics*, 14(3):1252–1256.

Targino, R. S., Peters, G. W., and Shevchenko, P. V. (2015). Sequential monte carlo samplers for capital allocation under copula-dependent risk models. *Insurance: Mathematics and Economics*, 61:206–226.

Tasche, D. (1995). Risk contributions and performance measurement. Working Paper, Techische Universität München.

Tasche, D. (2001). Conditional expectation as quantile derivative. *arXiv preprint math/0104190*.

Tasche, D. (2004). Capital allocation with creditrisk+. In Gundlach, M. and Lehrbass, F., editors, *CreditRisk+ in the banking industry*, pages 25–44. Springer, New York.

Tasche, D. (2008). Capital allocation to business units and sub-portfolios: the euler principle. In Resti, A., editor, *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, pages 423–453. Risk Books: London.

Tasche, D. (2009). Capital allocation for credit portfolios with kernel estimators. *Quantitative Finance*, 9(5):581–595.

Tasche, D. and Tibiletti, L. (2004). Approximations for the value-at-risk approach to risk-return analysis. *The ICFAI Journal of Financial Risk Management*, 1(4):44–61.

Tchen, A. H. et al. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, 8(4):814–827.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.

Tropp, J. A. (2018). Simplicial faces of the set of correlation matrices. *Discrete & Computational Geometry*, 60(2):512–529.

Vanduffel, S. and Dhaene, J. (2006). Some results on denault's capital allocation rule. *DTEW-AFI_0601*, pages 1–10.

Vats, D., Flegal, J. M., and Jones, G. L. (2019). Multivariate output analysis for markov chain monte carlo. *Biometrika*, 106(2):321–337.

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and economics*, 38(2):413–426.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360.

Wang, B., Wang, R., and Wang, Y. (2019). Compatible matrices of spearman's rank correlation. *Statistics & Probability Letters*, 151:67–72.

Yamai, Y. and Yoshiba, T. (2002). Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization. *Monetary and economic studies*, 20(1):87–121.

Yang, J., Cheng, S., and Zhang, L. (2006). Bivariate copula decomposition in terms of comonotonicity, countermonotonicity and independence. *Insurance: Mathematics and Economics*, 39(2):267–284.

Yi, K. and Doshi-Velez, F. (2017). Roll-back hamiltonian monte carlo. *arXiv preprint arXiv:1709.02855*.

Yoshiba, T. (2013). Risk aggregation by a copula with a stressed condition. Working Paper, Bank of Japan.