Measures for risk, dependence and diversification

by

Liyuan Lin

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

Waterloo, Ontario, Canada, 2024

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:	Paul Glasserman Professor, Graduate School of Business, Columbia University
Supervisors:	Ruodu Wang Professor, Department of Statistics and Actuarial Science, University of Waterloo Alexander Schied Professor, Department of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Mary Hardy Professor Emerita, Department of Statistics and Actuarial Science, University of Waterloo
Internal Member:	Fan Yang Assistant Professor, Department of Statistics and Actuarial Science, University of Waterloo

Internal-External Member: Keisuke Teeple
Assistant Professor, Department of Economics,
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

Two primary tasks in quantitative risk management are measuring risk and managing risk. Risk measures and dependence modeling are important tools for assessing portfolio risk, which have gained much interest in the literature of finance and actuarial science. The assessment of risk further serves to address risk management problems, such as portfolio optimization and risk sharing.

Value-at-Risk (VaR) and Expected Shortfall (ES) are the most widely used risk measures in banking and insurance regulation. The Probability Equivalent Level of VaR-ES (PELVE) is a new risk metric designed to bridge VaR and ES. In Chapter 2, we investigate the theoretical properties of PELVE and address the calibration problem of PELVE, that is, to find a distribution model that yields a given PELVE. Joint mixability, dependence of a random vector with a constant sum, is considered an extreme negative dependence as it represents a perfectly diversified portfolio. Chapter 3 explores the relationship between joint mix and some negative dependence notions in statistics. We further show that the negatively dependent joint mix plays a crucial role in solving the multi-marginal optimal transport problem under the uncertainty in the components of risks.

Diversification is a traditional strategy for mitigating portfolio risk. In Chapter 4, we employ an axiomatic approach to introduce a new diversification measurement called the diversification quotient (DQ). DQ exhibits many attractive properties not shared by existing diversification indices in terms of interpretation for dependence, ability to capture common shocks and tail heaviness, as well as efficiency in portfolio optimization. Chapter 5 provides some technical details and illustrations to support Chapter 4. Moreover, DQ based on VaR and ES have simple formulas for computation. We explore asymptotic behavior of VaR-based DQ and ES-based DQ for large portfolios, the elliptical model, and the multivariate regular varying (MRV) model in Chapter 6, as well as the portfolio optimization problems for the elliptical and MRV models.

Counter-monotonicity, as the converse of comonotonicity, is a natural extreme negative dependence. Chapter 7 conducts a systematic study of pairwise counter-monotonicity. We obtain its stochastic representation, invariance property, interactions with negative as-

sociation, and equivalence to joint mix within the same Fréchet class. We also show that Pareto-optimal allocations for quantile agents exhibit pairwise counter-monotonicity. This finding contrasts sharply with traditional comonotonic allocations for risk-averse agents, inspired further investigation into the appearance of pairwise counter-monotonic allocation in risk-sharing problems. In Chapter 8, we address the risk-sharing problem for agents using distortion riskmetrics, who are not necessarily risk-averse or monotone. Our results indicate that Pareto-optimal allocations for inter-quantile difference agents include pairwise counter-monotonicity. Chapter 9 further explores other decision models in risk-sharing that exhibit pairwise counter-monotonicity in optimal allocations. We introduce a counter-monotonic improvement theorem – a converse result to the widely used comonotonic improvement theorem. Furthermore, we show that pairwise counter-monotonic allocations are Pareto optimal for risk-seeking agents, Bernoulli utility agents, and rank-dependent expected utility agents under certain conditions.

Besides the studies of two extreme negative dependencies, we expand our analysis to dependence modeling through Pearson correlation and copula. In Chapter 10, we characterize all dependence structures for a bivariate random vector that preserve its Pearson correlation coefficient under any common marginal transformations. For multivariate cases, we characterize all invariant correlation matrices and explore the application of invariant correlation in sample duplication. Chapter 11 discusses the selection of copulas when marginals are discontinuous. The checkerboard copula is a desirable choice. We show that the checkboard copula has the largest Shannon entropy and carries the dependence information of the original random vector.

Acknowledgments

I extend my sincerest acknowledge to all those who supported me throughout the development and completion of this thesis. This thesis would not have been possible without their invaluable guidance and assistance.

Foremost, I express my deepest appreciation to my supervisors, Prof. Ruodu Wang and Prof. Alexander Schied. Their patient, motivated, and enthusiastic guidance, along with their continual encouragement, have been fundamental throughout my research journey. I am particularly grateful to Prof. Wang for introducing me to the world of quantitative risk management, which has shaped my research interests. He has provided significant support, both academically and personally, imparting many lessons that will benefit me throughout my life. Additionally, I am thankful to my master's supervisor, Prof. Jingzhen Liu, who first guided me into the academic world and continues to support me beyond my master's program. Working under their mentorship is a precious experience for me.

Acknowledgement is due to the members of my proposal and defense committees: Prof. Paul Glasserman, Prof. Keisuke Teeple, Prof. Mary Hardy and Prof. Fan Yang, for their invaluable time and insightful feedback that have significantly shaped this thesis.

My heartfelt thanks go to my coauthors, Prof. Hirbod Assa, Prof. Xia Han, Prof. Takaaki Koike, Prof. Ruixun Zhang, Dr. Jean-Gabriel Lauzier, and Dr. Chaoyi Zhao for their contributions to our collaborative research and the stimulating discussions that enhanced the quality of this thesis.

Special thanks are due to research fellows from both within and outside the University of Waterloo, including Prof. Fabio Bellini, Prof. Carole Bernard, Prof. Aleš Černý, Prof. An Chen, Prof. Yuyu Chen, Prof. Jan Dhaene, Prof. Paul Embrechts, Prof. Tolulope Fadina, Prof. Zhenzhen Fan, Prof. Sebastian Jaimungal, Prof. Pablo Koch-Medina, Prof. Dimitrios Konstantinides, Prof. Henry Lam, Prof. Hong Li, Prof. Sheldon Lin, Prof. Fangda Liu, Prof. Haiyan Liu, Prof. Peng Liu, Prof. Fabio Maccheroni, Prof. Tiantian Mao, Prof. Massimo Marinacci, Prof. Cosimo Munari, Prof. Silvana Pesenti, Prof. Xiaofei Shi, Prof. Qihe Tang, Prof. Stan Uryasev, Prof. Steven Vanduffel, Prof. Peter Wakker, Prof. Yunran Wei

Prof. Zuoquan Xu, Prof. Kelvin Shuangjian Zhang and Dr. Xiyue Han for their support and insightful comments to my research projects.

Thanks are also extended to Ms. Carla Daniels, Ms. Mary Lou Dufton, Ms. Divya Lala and Mr. Greg Preston and for their administrative and technical support.

I am fortunate to have the backing of friends who provided me with motivation and help during my PhD: Yixuan Fan, Xiaohui Fu, Samuel Gyamerah, Muqiao Huang, Zhenzhen Huang, Zhanyi Jiao, Anqi Jiang, Ying Jie, Wenyuan Li, Hanchen Liu, Yang Liu, Dongxu Ma, Zachary John Van Oosten, Qinghua Ren, Benxuan Shi, Ziyuei Shi Yuan Tian, Peiqi Wang, Qiuqi Wang, Yumin Wang, Zhaowen Wang, Christopher Blier-Wong, Qinyu Wu, Mingren Yin, Jiayue Zhang, Yuanyuan Zhang, Zhenyuan Zhang, Yimiao Zhao.

Lastly, I owe a profound debt of gratitude to my parents, Wenyong Lin and Meichun Zheng, and to my partner, Xintong Li. Their love and support have been my constant source of strength. I dedicate this accomplishment to them.

Dedication

This is dedicated to the one I love.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and preliminaries

1.1 Introduction

Quantifying portfolio risk has gained much interest in the literature of finance and actuarial science. Risk measures, which map a random variable into risk level, are commonly used by financial institutions to gauge and manage exposure. Among various of risk measures, Value-at-Risk (VaR) and Expected Shortfall (ES) are two most widely used risk measures in banking and insurance regulation frameworks such as Basel III/IV, Solvency II and the Swiss Solvency Test. More detailed comparisons between VaR and ES regarding coherence, robustness, and elicitability can be found in Cont et al. (2010), Embrechts et al. (2022), Gneiting (2011) and Kou and Peng (2016).

Compared with VaR, ES is regarded as a more reliable risk measure under extreme scenarios. This has led to suggestions from BCBS (2019) to replace VaR with ES. To bridge between VaR and ES, Li and Wang (2022) proposed the Probability Equivalent Level of VaR-ES (PELVE) to convert the level of VaR to that of ES. Remarkably, PELVE can also measure tail heaviness, reporting higher values for distributions with heavier tails. In Chapter 2, we study the calibration problem of PELVE; that is, to find a distribution model that yields a given PELVE, which may either be obtained from data or from expert opinion. We discuss separately the cases when one-point, two-point and curve constraints are given. For one-point and two-point problems, we construct the distribution model with the Pareto

distribution tail as the calibrated distribution. In the most complicated case of a curve constraint, we reformulate the calibration as an advanced differential equation problem and characterize all potential distributions when PELVE remains constant across all probability levels. We further study some technical properties of PELVE by offering a few new results on monotonicity and convergence. It turns out that the monotonicity of the PELVE is corresponding to the shape of the hazard rate function which is also used as a measurement for tail heaviness.

An essential tool of quantifying multiple risks is dependence modelling. A joint mix, introduced by Wang and Wang (2016), is a random vector with a constant component-wise sum. It is considered an extremely negative dependence since it represents a perfectly diversified portfolio. In Chapter 3, we explore the connection between the joint mix structure and popular notions of negative dependence in statistics, such as negative correlation dependence, negative orthant dependence and negative association (see Block et al. (1982); Lehmann (1966); Alam and Saxena (1981); Joag-Dev and Proschan (1983)). While not all joint mixes are negatively dependent in these senses, we identify some natural classes of joint mixes are. We derive various necessary and sufficient conditions for a joint mix to be negatively dependent, and study the compatibility of these notions. Joint mix is also well known as the solution for the Monge-Kantorovich problem (see Puccetti and Wang (2015)) with convex cost functions. We show that the negatively dependent joint mix serves as the solution for the optimal transport problem under uncertainty of components with quadratic cost function and identical marginal distributions Analysis of this optimal transport problem with heterogeneous marginals reveals a trade-off between negative dependence and the joint mix structure.

Traditionally, diversification has been recognized as an efficient strategy in portfolio risk management. Existing literature on quantifying diversification, such as the diversification ratio (see Choueifaty and Coignard (2008)) and the diversification benefit (see Embrechts et al. (2009)), experiences practical and theoretical deficiencies. In Chapter 4, we establish the first axiomatic theory for diversification indices using six intuitive axioms: non-negativity, location invariance, scale invariance, rationality, normalization, and continuity. The unique class of indices satisfying these axioms, called the diversification quotients (DQs), are defined

based on a parametric family of risk measures. A further axiom of portfolio convexity pins down DQ based on coherent risk measures. DQ has many attractive properties, and it can address several theoretical and practical limitations of existing indices. In particular, for the popular risk measures Value-at-Risk and Expected Shortfall, the corresponding DQ admits simple formulas and it is efficient to optimize in portfolio selection. Moreover, it can properly capture tail heaviness and common shocks, which are neglected by traditional diversification indices. When illustrated with financial data, DQ is intuitive to interpret, and its performance is competitive against other diversification indices. Detailed technical support and auxiliary results for the axiomatic theories are provided in Chapter 5. We further investigate DQ constructed from VaR and ES in Chapter 6. For the popular models of elliptical and multivariate regular varying (MRV) distributions, explicit formulas of DQs based on VaR and ES are available. The portfolio optimization problems for the elliptical and MRV models are also studied. Our results further reveal favourable features of DQ, both theoretically and practically, compared to traditional diversification indices based on a single risk measure.

The other extremely negative dependence notion is called pairwise counter-monotonicity, which is also know as mutually exclusive in actuarial literature (see Dhaene et al. (1999) and Cheung and Lo (2014)). In contrast to the large literature on comonotonicity, there are limit studies of pairwise counter-monotonicity. In Chapter 7, we systematically study pairwise counter-monotonicity. A stochastic representation and an invariance property are established for this dependence structure. We show that pairwise counter-monotonicity implies negative association, and it is equivalent to joint mix dependence if both are possible for the same marginal distributions. Interestingly, we find that Pareto-optimal allocations for risk-sharing among quantile agents are pairwise counter-monotonic.

The risk-sharing scheme can be considered as a function from a random variable to a random vector, which is different from risk allocation problems; see Kalkbrener (2005). A classic result in risk sharing is that the Pareto-optimal allocation for risk averse agents is comonotonic, as derived from the well-known comonotonic improvement theorem introduced in Landsberger and Meilijson (1994). The appearance of pairwise counter-monotonicity in Pareto-optimal allocation for quantile agents highlights the importance of this extremely

negative dependence structure in risk sharing. This result inspired a natural question: what types of decision models will exhibit counter-monotonicity in Pareto-optimal allocations? In Chapter 8, we address the problem of sharing risk among agents whose preferences are modeled by a general class of comonotonic additive and law-based functionals that are not necessarily monotone or convex. These functionals are referred to as distortion risk metrics and include many statistical measures of risk and variability used in portfolio optimization and insurance (see Wang et al. (2020a)). We characterized the set of the Pareto-optimal allocations under both comonotonic or general risk sharing frameworks. Specifically, for three agents using variability measures of the Gini deviation, the mean-median deviation, and the inter-quantile difference, we explicitly solve for Pareto-optimal allocations. An interesting finding is that the Pareto-optimal allocation incorporates a mixture of pairwise counter-monotonicity and conomotonicity for inter-quantile difference agents.

In Chapter 9, we further explore other decision models in risk sharing problem. Counter-monotonic allocations take the form of either "winner-takes-all" lotteries or "loser-loses-all" lotteries, which we respectively refer as jackpot and scapegoat allocations. We first present a converse result of comonotonic improvement called counter-monotonic improvement theorem. It states that one can always improve the allocated risks to each agents in convex order using counter-monotonic allocations for a bounded risk. We investigate the appearance of counter-monotonic allocations in risk-sharing problem for risk-seeking agents, Bernoulli utility agents and rank-dependent expected utility (RDU) agents. The Pareto-optimal allocations, if they exist, must be jackpot allocations when all agents are risk seeking. Scapegoat allocations represent the only fair method to allocate indivisible goods among Bernoulli utility agents. Under certain assumptions, RDU agents prefer jackpot allocations. We provide an application to the mining of cryptocurrencies, showing that, unlike risk-averse miners, RDU miners with limited computing power never join a mining pool. Finally, we characterize the competitive equilibria with risk-seeking agents, providing the first and second fundamental theorems of welfare economics in this context, where all equilibrium allocations are jackpot allocations.

We further explore other concepts of dependence modelling. A simple method to determine the dependence of a random vector is by examining the sign of the correlation coefficient. A useful property of independent samples is that their correlation remains the same after ap-

plying marginal transforms. This invariance property is fundamental in statistical inference (see Bates et al. (2023)), but does not generally hold for dependent samples. In Chapter 10, we are interested in characterizing (X, Y) such that the correlation coefficient remains unchanged after both variables are transformed by the same measurable function. We say that such (X, Y) has an invariant correlation under all transforms. For bivariate cases, we characterize all models of such a random vector via a certain combination of comonotonicity—the strongest form of positive dependence—and independence. In particular, we show that the class of exchangeable copulas with invariant correlation is precisely described by what we call positive Fréchet copulas. In the general multivariate case, we characterize the set of all invariant correlation matrices via the clique partition polytope. Additionally, we propose a positive regression dependent model that accommodates any prescribed invariant correlation matrix. This model turns out to be the joint distribution of samples with duplicate records. In this context, we provide an application of invariant correlation to the statistical inference in the presence of sample duplication. Finally, we show that all our characterization results of invariant correlation, except one special case, remain the same if the common marginal transforms are confined to the set of increasing ones.

A well-known tool used to describe dependence of dependence is the copula. Sklar's theorem states that the copula is unique if and only if the marginals of a random vector are continuous. However, in certain situations such as simulation, stress scenarios, and co-risk measures computation when the marginal distributions exhibit discontinuities, selecting an appropriate copula becomes essential. In Chapter 11, we address the choice of copula when not all marginal distributions of a random vector are continuous. Among all possible copulas, the checkerboard copula, characterized by uniform densities on regions of flexibility, emerges as a natural choice. We demonstrate that the checkerboard copula possesses the largest Shannon entropy, indicating that it carries the least information among all possible copulas for a given random vector. Moreover, the checkerboard copula retains the dependence information of the original random vector.

Chapter 12 concludes this thesis and discusses some open questions that are of potential interest in this field. Each chapter can be read as a separate paper except for Chapters 4–5.

1.2 Preliminaries

Below we give definitions and terminologies that will be used across different chapters. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space, where $\mathcal{F}$ is a σ-field and $\mathbb{P}$ is a probability measure. For $p \in (0, \infty)$, denote by $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables X with $\mathbb{E}[|X|^p] < \infty$ where $\mathbb{E}$ is the expectation under $\mathbb{P}$. Furthermore, $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all essentially bounded random variables, and $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all random variables. Further, denote by $\mathbb{R}_+ = [0, \infty)$ and $\overline{\mathbb{R}} = [-\infty, \infty]$. Terms such as increasing or decreasing functions are in the non-strict sense. Write $X \sim F$ if the random variable X has the distribution function F under $\mathbb{P}$. For $X \in L^0$, ess-sup(X) and ess-inf(X) are the essential supremum and the essential infimum of X, respectively. Two random variables X and Y have the same distribution under $\mathbb{P}$ is denoted by $X \overset{\mathrm{d}}{=} Y$. We treat almost surely (a.s.) equal random variables as identical; this means that all equalities and inequality for random variables hold in the a.s. sense, and we omit "a.s." in all our statements. A random variable X is said to be smaller than a random variable Y in the *convex order*, denoted by $X \leqslant_{\mathrm{cx}} Y$, if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for every convex function $\phi : \mathbb{R} \to \mathbb{R}$ provided that both expectations exist (see Rüschendorf (2013) and Shaked and Shanthikumar (2007)). The order $X \leqslant_{\mathrm{cx}} Y$ means that X is less risky than Y in the sense of Rothschild and Stiglitz (1970). Similarly, X is smaller than Y in the *increasing convex order*, denoted by $X \leqslant_{\mathrm{icx}} Y$ if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for every increasing convex function $\phi : \mathbb{R} \to \mathbb{R}$ provided that both expectations exist. The increasing convex order is referred to as *second-order stochastic dominance (SSD)* in Chapters 4–5.

Let n be a fixed positive integer, and write $[n] = \{1, \dots, n\}$. We always write $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$. A random vector $(X_1, \dots, X_n)$ is *comonotonic* if there exists a random variable Z and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ a.s. for every $i = 1, \dots, n$. A random vector (X, Y) is said to be counter-monotonic if $(X, -Y)$ is comonotonic. An equivalent formulation of comonotonicity is

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0 \quad \text{for } (\mathbb{P} \times \mathbb{P})\text{-almost every } (\omega, \omega') \in \Omega^2.$$

An equivalent formulation of counter-monotonicity is

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \leqslant 0 \quad \text{for } (\mathbb{P} \times \mathbb{P})\text{-almost every } (\omega, \omega') \in \Omega^2.$$

For $n \geqslant 3$, a random vector $\mathbf{X}$ taking values in $\mathbb{R}^n$ is *(pairwise) counter-monotonic* if each pair of its components is counter-monotonic. A *joint mix* (Wang and Wang, 2016) is a random vector $\mathbf{X} = (X_1, \dots, X_n)$ satisfying $X_1 + \cdots + X_n = c$ almost surely (a.s.) for some constant c, called a *center* of the joint mix. A random vector $\mathbf{X} = (X_1, \dots, X_n)$ is *exchangeable* if $\mathbf{X} \overset{\mathrm{d}}{=} \mathbf{X}^\pi$ for all $\pi \in \Pi_n$, where $\mathbf{X}^\pi = (X_{\pi(1)}, \dots, X_{\pi(n)})$.

Let $\mathcal{X}$ be a set of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. A risk measure ρ is a mapping from $\mathcal{X}$ to $\mathbb{R}$. The VaR and ES will be used frequently in this thesis. We use the "small α" through out this thesis: The VaR at level $\alpha \in (0,1)$ (typically very small) is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1-\alpha\}, \quad X \in L^0,$$

and the ES (also called CVaR, TVaR or AVaR) at level $\alpha \in (0,1]$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\,\mathrm{d}\beta, \quad X \in L^1.$$

Moreover, let $\mathrm{VaR}_0(X) = \mathrm{ES}_0(X) = \text{ess-sup}(X)$ and $\mathrm{VaR}_1(X) = \text{ess-inf}(X)$. Note that $\mathrm{ES}_1(X)$ is the mean of X. In Chapter 2, the VaR is defined on L^1 space to define PELVE.

Some standard properties of a risk measure $\rho : \mathcal{X} \to \mathbb{R}$ are collected below.

(i) Constant additivity: $\rho(X + c) = \rho(X) + c$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

(ii) Positive homogeneity: $\rho(\lambda X) = \rho(X)$ for all $\lambda \in (0, \infty)$ and $X \in \mathcal{X}$.

(iii) Subadditivity: $\rho(X + Y) \leqslant \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{X}$.

(iv) Monotonicity: $\rho(X) \leqslant \rho(Y)$ for $X \leqslant Y$.

(v) Comonotonic-additivity: $\rho(X + Y) = \rho(X) + \rho(Y)$ for comonotonic (X, Y).

(vi) Law invariance: $\rho(X) = \rho(Y)$ whenever $X \overset{\mathrm{d}}{=} Y$.

(vii) Continuity: $\rho(X_n) \to \rho(X)$ if $X_n \xrightarrow{L^p} X$, both as $n \to \infty$.

(viii) $\leqslant_{\mathrm{icx}}$-consistency: $\rho(X) \leqslant \rho(Y)$ for all $X, Y \in \mathcal{X}$ whenever $X \leqslant_{\mathrm{icx}} Y$.

(ix) Convex order consistency: $\rho(X) \leqslant \rho(Y)$ whenever $X \leqslant_{\mathrm{cx}} Y$.

Since SSD is equivalent to increasing convex order, $\leqslant_{\text{SSD}}$-consistency, mentioned in Chapter 4–5, is equivalent to $\leqslant_{\text{icx}}$-consistency. In the framework of Artzner et al. (1999), a risk measure is called a monetary risk measure if it satisfies monotonicity and constant additivity; a risk measure is called a coherent risk measure if it satisfies monotonicity, constant additivity, positive homogeneity, and subadditivity. For the case of presentation, we will reintroduce and redefine some notation and important concepts, such as VaR and ES, in each chapter to fit better the specific research problem.

Chapter 2

Calibrating distribution models from PELVE

2.1 Introduction

Value-at-Risk (VaR) and Expected Shortfall (ES, also known as TVaR and CVaR) are the most widely used risk measures for regulation in finance and insurance. The former has gained its popularity due to its simplistic approach toward risk as the risk quantile, and the second one is perceived to be useful as a modification of VaR with more appealing properties, such as tail-sensitivity and subadditivity, as studied in the seminal work of Artzner et al. (1999).

In the Fundamental Review of the Trading Book (FRTB), the Basel Committee on Banking Supervision (BCBS (2019)) proposed to replace VaR at 1% confidence with ES with a 2.5% confidence interval for the internal model-based approach.[1] The main reason, as mentioned in the FRTB, was that ES can better capture tail risk; see Embrechts et al. (2018) for a concrete risk sharing model where tail risk is captured by ES and ignored by VaR. On the other hand, VaR also has advantages that ES does not have, such as elicitability (e.g., Gneiting (2011) and Kou and Peng (2016)) or backtesting tractability (e.g., Acerbi

[1] In this chapter, we use the "small α" convention for VaR and ES. Hence, "VaR at 1% confidence" and "ES at 2.5% confidence" correspond to $\text{VaR}_{99\%}$ and $\text{ES}_{97.5\%}$ in BCBS (2019), respectively.

and Székely (2014)), and the two risk measures admit different axiomatic foundations (see Chambers (2009) and Wang and Zitikis (2021)). We refer to the reviews of Embrechts et al. (2014) and Emmer et al. (2015) for general discussions on VaR and ES, and McNeil et al. (2015) for a standard treatment on risk management including the use of VaR and ES. The technical contrasts of the two risk measures and their co-existence in regulatory practice give rise to great interest from both researchers and practitioners to explore the relationship between them.

To understand the balancing point of VaR and ES during the transition in the FRTB, Li and Wang (2022) proposed the Probability Equivalent Level of VaR-ES (PELVE). The value of PELVE is the multiplier to the tail probability when replacing VaR with ES such that the capital calculation stays unchanged. More precisely, the PELVE of X at level ϵ is the multiplier c such that $\mathrm{ES}_{c\epsilon}(X) = \mathrm{VaR}_{\epsilon}(X)$; such c uniquely exists under mild conditions. For instance, if $\mathrm{VaR}_{1\%}(X) = \mathrm{ES}_{2.5\%}(X)$ for a future portfolio loss X, then PELVE of X at probability level 0.01 is the multiplier 2.5. In this case, replacing $\mathrm{VaR}_{1\%}$ with $\mathrm{ES}_{2.5\%}$ in FRTB does not have much effect on the capital requirement for the bank bearing the loss X. Instead, if $\mathrm{ES}_{2.5\%}(X) > \mathrm{VaR}_{1\%}(X)$, then the bank has a larger capital requirement under the new regulatory risk measure; this is often the case for financial assets and portfolios as shown by the empirical studies in Li and Wang (2022). The PELVE enjoys many convenient properties, and it has been extended in a few ways. In particular, Fiori and Gianin (2022) defined generalized PELVE by replacing VaR and ES with another pair of monotone risk measures $(\rho, \tilde{\rho})$, and Barczy et al. (2022) extended PELVE by replacing ES with a higher-order ES.

For a given distribution model or a data set, its PELVE can be computed or estimated in a straightforward manner. As argued by Li and Wang (2022), the PELVE for a small ϵ may be seen as a summarizing index measuring tail heaviness in a non-limit sense. As such, one may like to generate models for a given PELVE, in a way similar to constructing models for other given statistical information; see e.g., Embrechts et al. (2002, 2016) for constructing multivariate models with a given correlation or tail-dependence matrix. Such statistical information may be obtained either from data or from expert opinion, but there is no a priori guarantee that a corresponding model exists. Since PELVE involves a parameter

$\epsilon \in (0,1)$, its information is represented by a curve. The calibration problem, that is, to find a distribution model for given PELVE values or a given PELVE curve, turns out to be highly non-trivial, and it is the main objective of this chapter.

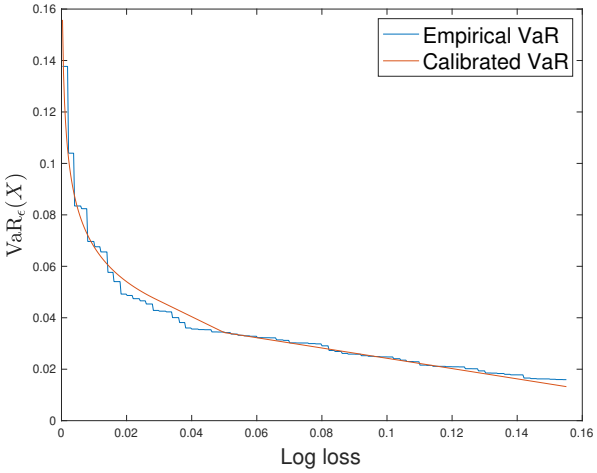
From now on, suppose that we receive some information on the PELVE of a certain random loss from an expert opinion, and we aim to build a distribution model consistent with the supplied information. Since PELVE is location-scale invariant, such a distribution, if it exists, is not unique.

The calibration problem is trivial if we are supplied with only one point on the PELVE curve. As the PELVE curve of the generalized Pareto distribution is a constant when the PELVE is well defined, we can use the generalized Pareto distribution to match the given PELVE value, which has a tail index implied from the expert opinion. The calibration problem becomes more involved if we are supplied with two points on the PELVE curve, because the value of the PELVE at two different probability levels interact with each other. The situation becomes more complicated as the number of points increases, and we further turn to the problem of calibration from a fully specified PELVE curve. Calibrating distribution from the PELVE curve can be reformulated as solving for a function f via the integral equation $\int_0^y f(s)\,\mathrm{d}s = yf(z(y)\,y)$, where the curve z is computed from the PELVE curve. This integral equation can be further converted to an advanced differential equation (see Bellman and Cooke (1963)). For the case that z is a constant curve, we can explicitly obtain all solutions for f. We find other distributions that also have constant PELVE curves other than the simple ones with a Pareto or exponential distribution. As a consequence, a PELVE curve does not characterize a unique location-scale family of distributions; this provides a negative answer to a question posed by Li and Wang (2022, Section 7, Question (iv)). For general function z, we develop a numerical method to compute f.

The calibrated distribution can be used to estimate the value of other risk measures such as VaR and ES at different levels. We illustrate by an empirical example that two points of PELVE give a quite good summary of the tail distribution of risk. Daily log-losses (negative log-returns) of AAPL from Yahoo Finance are collected for the period from January 3, 2012 to December 31, 2021 within total of 2518 observations. We calculate the empirical PELVE at levels 0.01 and 0.05 using the empirical PELVE estimator provided by Li and Wang (2022,

Section 5) with a moving window of 500 trading days. For each pair of two points of PELVE at levels 0.01 and 0.05, we produce a quantile curve from the two empirical PELVE points by our calibration model in Section 2.3.2, which is scaled such that $\mathrm{VaR}_{0.01}$ and $\mathrm{VaR}_{0.05}$ are equal to their empirical values.[2] Figure 2.1 presents the empirical and calibrated quantile curves on December 31, 2021 using 500 trading days prior to that date. The two quantile curves are close to each other, with our calibrated curve being more smooth. We also report the values of $\mathrm{ES}_{0.025}$ of the calibrated distribution, which we call the calibrated $\mathrm{ES}_{0.025}$, and compare it with empirical $\mathrm{ES}_{0.025}$. The left panel of Figure 2.2 shows the curves of empirical and calibrated $\mathrm{ES}_{0.025}$. In the right panel of Figure 2.2, we create a scatter plot using empirical and calibrated $\mathrm{ES}_{0.025}$. Both figures show that the empirical and calibrated $\mathrm{ES}_{0.025}$ curves are quite close.

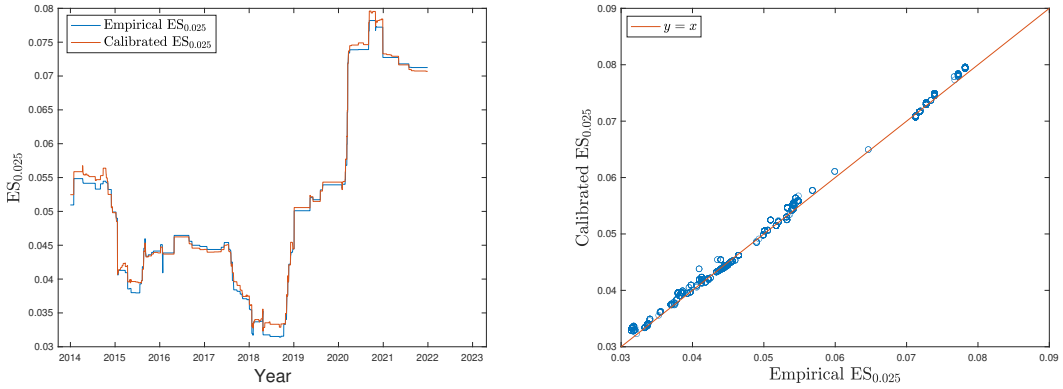
Figure 2.1: Empirical VaR and calibrated VaR



To further enrich the theory of PELVE, we study a few technical properties of PELVE, such as monotonicity and convergence as the probability level goes to 0. A decreasing PELVE indicates a relatively larger impact of ES in risk assessment than VaR moving towards the tail. As we will see, while for the most known parametric distributions the PELVE is decreasing, there exist some examples at some risk levels it is not decreasing. This means that for those examples VaR becomes a stricter risk measure when moving towards the tail. To obtain

[2]Recall that PELVE is location-scale free, and hence we need to pick two free parameters to specify a distribution calibrated from PELVE.

Figure 2.2: Empirical $\mathrm{ES}_{0.05}$ and calibrated $\mathrm{ES}_{0.05}$



conditions for monotonicity, we define the dual PELVE by moving the multiplier c from the ES side to the VaR side. PELVE can be seen as a functional measure of tail heaviness in the sense that a heavier-tailed distribution has a higher PELVE curve (Li and Wang (2022, Theorem 1)). The hazard rate, on the other hand, is another functional measure of tail heaviness. We show that the PELVE is decreasing (increasing) if the inverse of the hazard rate is convex (concave). Monotonicity also leads to conditions for the PELVE to have a limit at the tail, which from the risk management perspective, identifies the ultimate relative positions of ES and VaR in the tail region. From a mathematical perspective, the limit of PELVE at 0 allows us to extend the domain of PELVE to include 0 as a measure of tail heaviness.

The rest of this chapter is organized as follows. Section 2.2 introduces the background and examples of the PELVE. In Section 2.3 we calibrate a distribution from finitely many points in the PELVE curve. Section 2.4 calibrates the distribution from given PELVE curves, where we give a class of explicit solutions for constant PELVE functions and numerical solutions for general PELVE functions. In Section 2.5, we study the monotonicity and convergence of the PELVE. Section 2.6 presents two examples of the model calibration techniques applied to datasets used in insurance. A conclusion is given in Section 2.7. Some technical proofs of results in Sections 2.3, 2.4 and 2.5 are provided in Section 2.8 - 2.10.

2.2 Definitions and background

Let us consider an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the set of the measurable sets and $\mathbb{P}$ is the probability measure. Let L^1 be the set of integrable random variables, i.e., $L^1 = \{X : \mathbb{E}[|X|] < \infty\}$, where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$.

We first define VaR and ES in L^1, the two most popular risk measures. The VaR and at probability level $p \in (0, 1)$ is defined as

$$\mathrm{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1 - p\} = F^{-1}(1 - p), \qquad X \in L^1, \tag{2.1}$$

where F is the distribution of X. The ES at probability level $p \in [0, 1)$ is defined as

$$\mathrm{ES}_p(X) = \frac{1}{p} \int_0^p \mathrm{VaR}_q(X)\, \mathrm{d}q, \qquad X \in L^1.$$

Note that we use the "small α" convention for VaR and ES, which is different from Liu and Wang (2021). Let $\mathrm{VaR}_0(X) = \mathrm{ES}_0(X) = \text{ess-sup}(X)$ and $\mathrm{VaR}_1(X) = \text{ess-inf}(X)$. We have that $\mathrm{ES}_1(X)$ is the mean of X. We will also call $p \mapsto \mathrm{VaR}_p(X)$ the quantile function of X, keeping in mind that in our convention this function is decreasing.[3]

For $\epsilon \in (0, 1)$, the PELVE at level ϵ, proposed by Li and Wang (2022), is defined as

$$\Pi_X(\epsilon) = \inf\left\{c \in [1, 1/\epsilon] : \mathrm{ES}_{c\epsilon}(X) \leqslant \mathrm{VaR}_\epsilon(X)\right\}, \quad X \in L^1,$$

where $\inf(\varnothing) = \infty$. Li and Wang (2022) used $\Pi_\epsilon(X)$ for our $\Pi_X(\epsilon)$, and our choice of notation is due to the fact that the curve $\epsilon \to \Pi_X(\epsilon)$ is the main quantity of interest in this chapter.

The PELVE of X is finite if and only if $\mathrm{VaR}_\epsilon(X) \geqslant \mathbb{E}[X]$. The value of the PELVE is the multiplier c such that $\mathrm{ES}_{c\epsilon}(X) = \mathrm{VaR}_\epsilon(X)$. If $\mathrm{VaR}_p(X)$ is not a constant for $p \in (0, \epsilon]$, then the PELVE is the unique solution for the multiplier. By Theorem 1 in Li and Wang (2022), the PELVE is location-scale invariant. The distribution with a heavy tail will have a higher PELVE value.

If X is a normal distributed random variable and $\epsilon = 1\%$, we have $\Pi_X(\epsilon) \approx 2.5$. It means that $\mathrm{ES}_{2.5\%}(X) \approx \mathrm{VaR}_{1\%}(X)$. That is, the replacement suggested by BCBS is fair for

[3] Throughout the chapter, all terms like "increasing" and "decreasing" are in the non-strict sense.

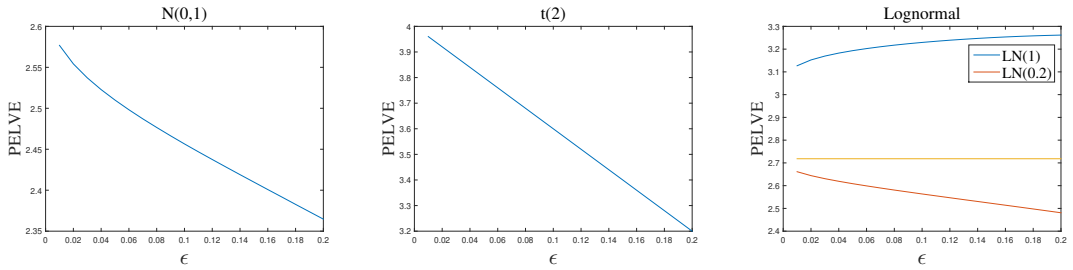
normally distributed risks. In other words, a higher PELVE will result in a higher capital requirement after the replacement.

In this chapter, we are generally interested in the question of which distributions have a specified or partially specified PELVE curve. We first look at a few simple examples.

Example 2.1 (Constant PELVE)**.** We first list some distributions that have constant PELVE curves. From the definition of the PELVE, we know that the PELVE should be larger than 1. As we can see from Table 2.1, the PELVE for the generalized Pareto distribution takes values on $(1, \infty)$. For $X \sim \mathrm{GPD}(\xi)$, we have $1 < \Pi_X(\epsilon) < e$ when $\xi < 0$, $\Pi_X(\epsilon) = e$ when $\xi = 0$ and $\Pi_X(\epsilon) > e$ when $\xi > 0$. Furthermore, if X follows the point-mass distribution δ_c or the Bernoulli distribution, we have $\Pi_X(\epsilon) = 1$.

Example 2.2. Here we present some non-constant PELVE examples. We write $\mathrm{t}(v)$ for the t-distribution with parameter $(0, 1, v)$, and $\mathrm{LN}(\sigma)$ for the log-normal distribution with parameter $(0, \sigma^2)$. As we can see, for normal distribution and t-distribution, the PELVE curve

Figure 2.3: PELVE for normal distribution, t-distribution and lognormal distribution



is decreasing as ϵ increasing. The monotonicity of the PELVE of the lognormal distribution depends on the value of σ. The monotonicity of the PELVE will be further discussed in Section 2.5. For more PELVE examples, see Li and Wang (2022).

2.3 Calibration from finite-point constraints

In this section, we discuss the calibration problem when some points of the PELVE are given. We will focus on the case where one point or two points on the PELVE curve are specified, for which we can explicitly construct a corresponding quantile function.

Table 2.1: Example of constant PELVE

Distribution	Distribution or probability function of X	$\Pi_X(\epsilon)$
δ_c	$\mathbb{P}(X = c) = 1$	$\Pi_X(\epsilon) = 1$ for $\epsilon \in (0, 1)$
$\mathrm{B}(1, p)$	$\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$	$\Pi_X(\epsilon) = 1$ for $\epsilon \in (0, p)$
$\mathrm{U}(0, 1)$	$F(t) = t$ for $t \in (0, 1)$	$\Pi_X(\epsilon) = 2$ for $0 < \epsilon < 1/2$
$\mathrm{EXP}(\lambda)$	$F(t) = 1 - \exp(-\lambda t), \quad \lambda > 0$	$\Pi_X(\epsilon) = e$ for $0 < \epsilon < 1/e$
$\mathrm{GPD}(\xi)$[1]	$F(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-x) & \xi = 0 \end{cases}$	$\Pi_X(\epsilon) = (1 - \xi)^{-\frac{1}{\xi}}$ for $0 < \epsilon < (1 - \xi)^{\frac{1}{\xi}}$

[1] The distribution $\mathrm{GPD}(\xi)$ is called the standard generalized Pareto distribution. As $\mathbb{E}[X] < \infty$ when $\xi < 1$, the PELVE exists only when $\xi < 1$. The support of $\mathrm{GPD}(\xi)$ is $[0, \infty)$ when $\xi > 0$ and $[0, -\frac{1}{\xi}]$ when $\xi < 0$. When $\xi = 0$, the $\mathrm{GPD}(\xi)$ is exactly exponential distribution with $\lambda = 1/\sigma$. There is a three-parameter $\mathrm{GPD}(\mu, \sigma, \xi)$, which is a location-scale transform of standard GPD. Therefore, $\mathrm{GPD}(\mu, \sigma, \xi)$ has the same PELVE as $\mathrm{GPD}(\xi)$.

We first note that the calibrated distribution is not unique. For example, if we are given $\Pi_X(0.01) = 2.5$, we can assume the distribution of X is the Normal distribution or the generalized Pareto distribution with tail parameter ξ satisfying $(1-\xi)^{-1/\xi} = 2.5$ from Table 2.3. Therefore, the distributions obtained in our results are only some possible choices, which we choose to have a generalized Pareto tail, as Pareto tails are standard in risk management applications.

2.3.1 Calibration from a one-point constraint

Based on Table 2.1, we can calibrate the distribution for X from one given PELVE point (ϵ_1, c_1) such that $\Pi_X(\epsilon_1) = c_1$. A simple idea is to take the generalized Pareto distribution when $c_1 > 1$ and δ_c when $c_1 = 1$. We summarize the idea in the following Proposition.

Proposition 2.1. *Let $\epsilon_1 \in (0,1)$ and $c_1 \in [1,\infty)$ such that $c_1\epsilon_1 \leqslant 1$. If $c_1 > 1$, let $\xi \in \mathbb{R}$ such that $(1-\xi)^{-\frac{1}{\xi}} = c_1$. Then, $X \sim GPD(\xi)$ has $\Pi_X(\epsilon_1) = c_1$. If $c_1 = 1$, then $X = k$ for some constant $k \in \mathbb{R}$ has $\Pi_X(\epsilon_1) = c_1$.*

The proof can be directly derived from Table 2.1 and it is omitted. By Proposition 2.1, if we have the value of PELVE at point ϵ_1, we can find a distribution of X which has the same PELVE value at ϵ_1. If we also have the value of VaR at ϵ_1, we can determine the scale parameter (σ) for the GPD distribution or the value of k to match the value of VaR. For Table 2.1, we can see that the calibrated generalized Pareto distribution can also serve as a solution for a more prudent condition $\Pi_X(\epsilon) \geqslant c_1$ when $\epsilon \in (0, \epsilon_1)$.

2.3.2 Calibration from a two-point constraint

The calibration problem would be much more difficult when we are given two points of the PELVE curve. Given two points (ϵ_1, c_1) and (ϵ_2, c_2) such that $\epsilon_1 < \epsilon_2$, we want to find a distribution for $X \in L^1$ such that $\Pi_X(\epsilon_1) = c_1$ and $\Pi_X(\epsilon_2) = c_2$. Nevertheless, the choices of (ϵ_1, c_1) and (ϵ_2, c_2) are not arbitrary. First, we need $1 \leqslant c_1 \leqslant 1/\epsilon_1$ and $1 \leqslant c_2 \leqslant 1/\epsilon_2$ by the definition of the PELVE. Then, we will show that the value of c_2 will be restricted if (ϵ_1, c_1) and ϵ_2 are given.

Lemma 2.1. *For any $X \in L^1$, let $\epsilon_1, \epsilon_2 \in (0,1)$ be such that $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X)$ and $\epsilon_1 < \epsilon_2$. Then, we have $\epsilon_1 \Pi_X(\epsilon_1) \leqslant \epsilon_2 \Pi_X(\epsilon_2)$.*

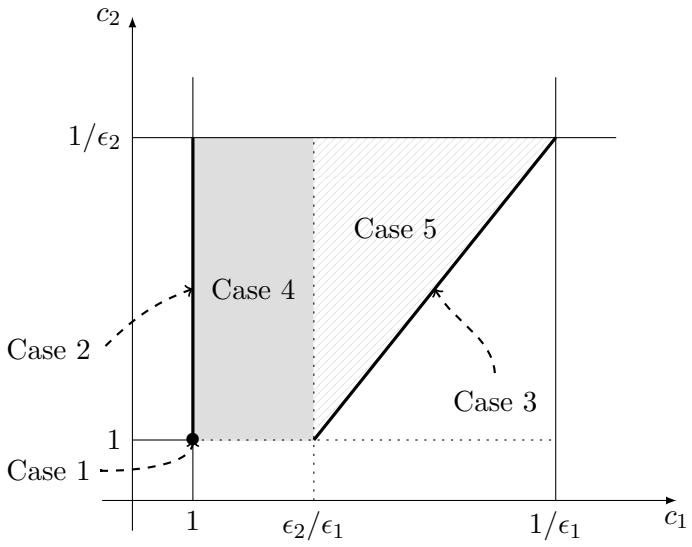
By Lemma 2.1, for given ϵ_1, ϵ_2 and c_1, the value of c_2 is bounded below by both 1 and $c_1\epsilon_1/\epsilon_2$. We also note that if $c_2 = 1$, then $p \mapsto \mathrm{VaR}_p(X)$ is constant on $(0, \epsilon_2)$, which implies $c_1 = 1$. In Section 2.8, Proposition 2.6 shows that the above lower bound is achieved if and only if $\mathrm{VaR}_{\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_2}(X)$.

From the definition of the PELVE and Lemma 2.1, for $\epsilon_1 < \epsilon_2$, the possible choices of (ϵ_1, c_1) and (ϵ_2, c_2) should satisfy $1 \leqslant c_1 \leqslant 1/\epsilon_1$, $1 \leqslant c_2 \leqslant 1/\epsilon_2$ and $c_1\epsilon_1 \leqslant c_2\epsilon_2$. We denote by Δ the admissible set for $(\epsilon_1, c_1, \epsilon_2, c_2)$, that is,

$$\Delta = \{(\epsilon_1, c_1, \epsilon_2, c_2) \in ((0,1) \times [1, \infty))^2 : \epsilon_1 < \epsilon_2,\ c_1\epsilon_1 \leqslant 1,\ c_2\epsilon_2 \leqslant 1,\ c_1\epsilon_1 \leqslant c_2\epsilon_2\}.$$

We illustrate the possible region of (c_1, c_2) with given ϵ_1 and ϵ_2 in Figure 2.4. We divide the region into 5 cases and calibrate the distribution for each case.

Figure 2.4: Admissible region of (c_1, c_2)



The calibration process is to construct a continuous and decreasing quantile function that can satisfy two equivalent conditions between VaR and ES, which are

$$\mathrm{ES}_{c_1\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_1}(X) \quad \text{and} \quad \mathrm{ES}_{c_2\epsilon_2}(X) = \mathrm{VaR}_{\epsilon_2}(X). \tag{2.2}$$

As we can see, only the values of $\mathrm{VaR}_\epsilon(X)$ for $\epsilon \in (0, c_2\epsilon_2]$ matters for the equivalent condition (2.2). Therefore, we focus on constructing $\mathrm{VaR}_\epsilon(X)$ for $\epsilon \in (0, c_2\epsilon_2]$. In addition, we want a continuous calibrated quantile function.

The case $c_1 = 1$ or $c_2 = 1$ is special, which means that $\mathrm{VaR}_\epsilon(X)$ is a constant on the tail part. If $c_1 > 1$, we can set the tail distribution as the generalized Pareto distribution from Table 2.3 such that $\Pi_X(\epsilon_1) = c_1$.

For $\mathbf{z} = (\epsilon_1, c_1, \epsilon_2, c_2) \in \Delta$, we will construct a class of functions, denoted by $G_{\mathbf{z}}$, in five different cases according to Figure 2.4. The function $t \mapsto G_{\mathbf{z}}(t)$ will be our desired quantile function. If $c_1 = 1$, let $\hat{k}, \tilde{k} \in \mathbb{R}$ be any two constants satisfying $\tilde{k} < \hat{k}$. If $c_1 > 1$, let $\xi \in (-\infty, 1)$ be such that $(1-\xi)^{-1/\xi} = c_1$,

$$k(\epsilon) = \begin{cases} \dfrac{1}{\xi}(\epsilon^{-\xi} - 1), & \xi \neq 0, \\ -\log(\epsilon), & \xi = 0, \end{cases}$$

and $k = \int_0^{\epsilon_1} k(\epsilon)\,\mathrm{d}\epsilon$. We first claim that the function $G_{\mathbf{z}}$ can be any arbitrary continuous and decreasing function on $[c_2\epsilon_2, 1)$ since the values of $\mathrm{VaR}_t(X)$ for $t \in [c_2\epsilon_2, 1)$ do not affect its PELVE at ϵ_1 and ϵ_2. The value of $G_{\mathbf{z}}$ on $(0, c_2\epsilon_2]$ is given by

(i) Case 1, $c_2 = 1$ (which implies $c_1 = 1$): $G_{\mathbf{z}}(\epsilon) = \hat{k}$;

(ii) Case 2, $c_1 = 1$ and $1 < c_2 \leqslant 1/\epsilon_2$:

$$G_{\mathbf{z}}(\epsilon) = \begin{cases} \hat{k}, & \epsilon \in (0, \epsilon_1), \\ a_1\epsilon + b_1, & \epsilon \in [\epsilon_1, \epsilon_2), \\ a_2\epsilon + b_2, & \epsilon \in [\epsilon_2, c_2\epsilon_2], \end{cases} \quad \text{where} \quad \begin{cases} a_1 = \dfrac{\tilde{k} - \hat{k}}{\epsilon_2 - \epsilon_1}, \\ b_1 = \hat{k} - a_1\epsilon_1, \\ a_2 = \dfrac{(\tilde{k} - \hat{k})(\epsilon_1 + \epsilon_2)}{(c_2\epsilon_2 - \epsilon_2)^2}, \\ b_2 = \tilde{k} - a_2\epsilon_2; \end{cases}$$

(iii) Case 3, $\epsilon_2/\epsilon_1 < c_1 \leqslant 1/\epsilon_1$ and $c_2 = c_1\epsilon_1/\epsilon_2$:

$$G_{\mathbf{z}}(\epsilon) = \begin{cases} k(\epsilon), & \epsilon \in (0, \epsilon_1), \\ k(\epsilon_1), & \epsilon \in [\epsilon_1, \epsilon_2), \\ a\epsilon + b, & \epsilon \in [\epsilon_2, c_2\epsilon_2], \end{cases} \quad \text{where} \quad \begin{cases} a = \dfrac{2(k(\epsilon_1)\epsilon_1 - k)}{(c_2\epsilon_2 - \epsilon_2)^2}, \\ b = k(\epsilon_1) - a\epsilon_2; \end{cases}$$

(iv) Case 4, $1 < c_1 \leqslant \epsilon_2/\epsilon_1$ and $1 < c_2 \leqslant 1/\epsilon_2$:

$$G_{\mathbf{z}}(\epsilon) = \begin{cases} k(\epsilon), & \epsilon \in (0, c_1\epsilon_1), \\ a_1\epsilon + b_1, & \epsilon \in [c_1\epsilon_1, \epsilon_2), \\ a_2\epsilon + b_2, & \epsilon \in [\epsilon_2, c_2\epsilon_2], \end{cases} \text{ where } \begin{cases} a_1 = -(c_1\epsilon_1)^{-\xi-1}, \\ b_1 = k(c_1\epsilon_1) - a_1 c_1\epsilon_1, \\ a_2 = \dfrac{a_1(\epsilon_2^2 - (c_1\epsilon_1)^2) + 2(k(c_1\epsilon_1) - k(\epsilon_1))c_1\epsilon_1}{(c_2\epsilon_2 - \epsilon_2)^2}, \\ b_2 = a_1\epsilon_2 + b_1 - a_2\epsilon_2; \end{cases}$$

(v) Case 5, $\epsilon_2/\epsilon_1 < c_1 \leqslant 1/\epsilon_1$ and $c_1\epsilon_1/\epsilon_2 < c_2 \leqslant 1/\epsilon_2$:

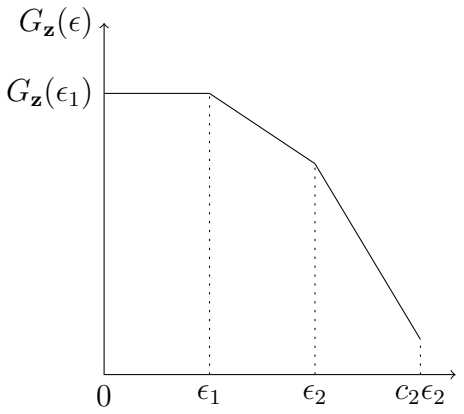
$$G_{\mathbf{z}}(\epsilon) = \begin{cases} k(\epsilon), & \epsilon \in (0, \epsilon_1), \\ a_1\epsilon + b_1, & \epsilon \in [\epsilon_1, \epsilon_2), \\ a_1\epsilon_2 + b_1, & \epsilon \in [\epsilon_2, c_1\epsilon_1), \\ a_2\epsilon + b_2, & \epsilon \in [c_1\epsilon_1, c_2\epsilon_2], \end{cases} \text{ where } \begin{cases} a_1 = \dfrac{k(\epsilon_1)\epsilon_1 - k}{(\epsilon_2 - \epsilon_1)(c_1\epsilon_1 - 1/2(\epsilon_1 + \epsilon_2))}, \\ b_1 = k(\epsilon_1) - a_1\epsilon_1, \\ a_2 = \dfrac{2c_1\epsilon_1(a_1\epsilon_2 + b_1 - k(\epsilon_1))}{(c_1\epsilon_1 - c_2\epsilon_2)^2}, \\ b_2 = a_1\epsilon_2 + b_1 - a_2 c_1\epsilon_1. \end{cases}$$

An illustration of the functions $G_{\mathbf{z}}$ on $[0, c_2\epsilon_2]$ in Case 2 to Case 5 is presented in Figure 2.5, and we omit Case 1 in which $G_{\mathbf{z}}$ is a constant function on $[0, c_2\epsilon_2]$.

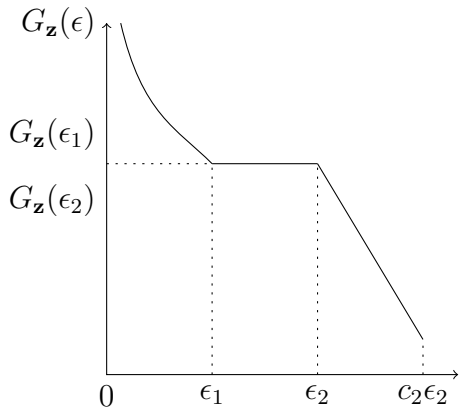
Theorem 2.1. *For $\mathbf{z} = (\epsilon_1, c_1, \epsilon_2, c_2) \in \Delta$, the random variable X with a continuous quantile function given by $t \mapsto \mathrm{VaR}_t(X) = G_{\mathbf{z}}(t)$ satisfies $\Pi_X(\epsilon_1) = c_1$ and $\Pi_X(\epsilon_2) = c_2$.*

Remark 2.1. As we can see from Figure 2.5, some parts of the calibrated quantile function may be flat, corresponding to the existence of atoms in the distribution. This may be considered as undesirable from a modeling perspective, and indeed it is forced by the boundary cases of $(\epsilon_1, c_1, \epsilon_2, c_2) \in \Delta$ in Figure 2.4. The flat parts in Cases 1 to 3 are necessary due to Propositions 2.6. On the other hand, the flat part in Case 5 can be replaced by a strictly decreasing function. For instance, we can replace the flat part with a strictly decreasing linear segment as long as c_2 satisfies the bounds shown in Propositions 2.7 in Section 2.8. Another way is to set $\mathrm{VaR}_\epsilon(X)$ as $k(\epsilon)$ for $\epsilon \in (0, c_1\epsilon_1)$ if $c_2 \leqslant \left(c_1\epsilon_1(\epsilon_1^{-\xi} - (c_1\epsilon_1)^{-\xi})\right) / \left(\epsilon_2^{-\xi} - (c_1\epsilon_1)^{-\xi}\right)$, and this choice is applied in the numerical examples in the Introduction and Section 2.6. The interested reader can see Propositions 2.6 and 2.7 in Section 2.8, where we show that a strictly decreasing quantile function cannot attain the boundary cases $(\epsilon_1, c_1, \epsilon_2, c_2)$, and hence the flat parts are necessary to include and unify these cases.

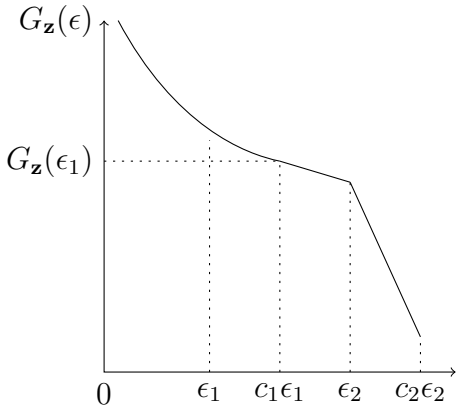
Figure 2.5: An illustration of $G_{\mathbf{z}}$ in cases 2 to 5



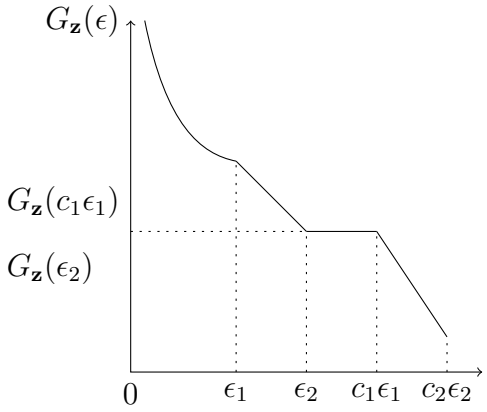
(a) The function $G_{\mathbf{z}}$ in Case 2



(b) The function $G_{\mathbf{z}}$ in Case 3



(c) The function $G_{\mathbf{z}}$ in Case 4



(d) The function $G_{\mathbf{z}}$ in Case 5

We can easily get the distribution of X from $\mathrm{VaR}_\epsilon(X)$. As the PELVE is scale-location invariant, we can scale or move the distribution we get to match more information. For example, if $\mathrm{VaR}_{\epsilon_1}(X)$ and $\mathrm{VaR}_{\epsilon_2}(X)$ are given, we can choose two constants λ and μ such that $\lambda X + \mu$ matches the specified VaR values. In a similar spirit, the calibration problem can be extended to calibrate the distributions from some given ES and VaR values. The two points calibration problem can be regarded as given two ES and VaR values. Calibrating from only ES or VaR would be easy. However, the choices of ES values will also be limited by VaR values if we consider them at the same time, which is the same as the choice of c_1, c_2 as we discussed in this section.

2.3.3 Calibration from an n-point constraint

As we see above, the PELVE calibration problem is quite technical even when only two points on the PELVE curve are given. By extending the constraint to more than two points, the problem will in general become much more complicated. We briefly discuss this problem in this section.

For the n-point constraint problem, we first need to figure out the admissible set for $(\epsilon_i, c_i)_{i=1,\dots,n}$. By Lemma 2.1, the admissible set for the n-point calibration problem is a subset of

$$\{(\epsilon_i, c_i)_{i=1,\dots,n} : 0 < \epsilon_1 < \cdots < \epsilon_n < 1,\ c_1, \dots, c_n \geqslant 1,\ 0 < c_1\epsilon_1 \leqslant \dots \leqslant c_n\epsilon_n \leqslant 1\}.$$

However, it is not clear whether each point in the above set is admissible. There are other constraints for the admissible points such as Proposition 2.7. Once the admissible set is determined, we need to divide the admissible set according to the position of ϵ_i and $c_i\epsilon_i$, $i = 1, \dots, n$. Furthermore, the case $c_i = 1$ and $c_i\epsilon_i = c_j\epsilon_j$ for $i, j = 1, \dots, n$ need special attention as Cases 1, 2 and 3 in the two-point constraint problem. For instance, in the three-point constraint problem, we need to discuss over 10 separate cases.

Below, we only discuss some special cases of $(\epsilon_i, c_i)_{i=1,\dots,n}$. First, if $c_n = 1$, then the problem becomes trivial, as the calibrated quantile functions satisfy $\mathrm{VaR}_t(X) = \hat{k}$ for some $\hat{k} \in \mathbb{R}$ in $[0, c_n\epsilon_n]$.

For the case $c_k\epsilon_k > \epsilon_k \geqslant c_{k-1}\epsilon_{k-1}$ for $k = 3, \dots, n$, we can set the calibrated quantile function in $(0, c_n\epsilon_n]$ recursively. This is because such a configuration of $(\epsilon_i, c_i)_{i=1,\dots,n}$ allows for separation of the constraints, in the sense that we can adjust the values of VaR_t for $t \in [\epsilon_k, c_k\epsilon_k]$ to match PELVE at ϵ_k without disturbing VaR_t for $t \leqslant c_{k-1}\epsilon_{k-1}$. Let $\mathrm{VaR}_t^k(X)$ be the calibrated quantile function from the k-point constraint problem for $k = 2, \dots, n$ where $\mathrm{VaR}_t^2(X)$ follows Theorem 2.1. The calibrated quantile function for the n-point constraint problem is

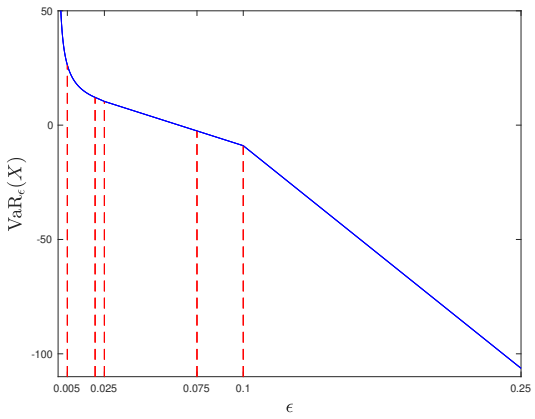
$$\mathrm{VaR}_t^k(X) = \begin{cases} \mathrm{VaR}_t^{k-1}(X), & t \in [0, c_{k-1}\epsilon_{k-1}], \\ a_{k-1}t + b_{k-1}, & t \in (c_{k-1}\epsilon_{k-1}, \epsilon_k], \\ a_k t + b_k, & t \in (\epsilon_k, c_k\epsilon_k], \end{cases} \text{ where } \begin{cases} a_k = \dfrac{a_{k-1}(\epsilon_k^2 + c_{k-1}^2\epsilon_{k-1}^2 - 2c_{k-1}\epsilon_{k-1}^2)}{(c_k\epsilon_k - \epsilon_k)^2}, \\ b_k = a_{k-1}\epsilon_k + b_{k-1} - a_k\epsilon_k. \end{cases}$$

In particular, for $n = 3$, and assuming $c_3\epsilon_3 > \epsilon_3 \geqslant c_2\epsilon_2$, the calibrated function is given by, with $\mathbf{z} = (\epsilon_1, c_1\epsilon_1, \epsilon_2, c_2\epsilon_2) \in \Delta$,

$$\mathrm{VaR}_t(X) = \begin{cases} G_{\mathbf{z}}(t), & t \in [0, c_2\epsilon_2], \\ a_2 t + b_2, & t \in (c_2\epsilon_2, \epsilon_3], \\ a_3 t + b_3, & t \in (\epsilon_3, c_3\epsilon_3], \end{cases} \text{ where } \begin{cases} a_2 = \dfrac{a_1(\epsilon_2^2 - (c_1\epsilon_1)^2) + 2(k(c_1\epsilon_1) - k(\epsilon_1))c_1\epsilon_1}{(c_2\epsilon_2 - \epsilon_2)^2}, \\ b_2 = -(c_1\epsilon_1)^{-\xi-1}(\epsilon_2 - c_1\epsilon_1) - k(c_1\epsilon_1) \\ a_3 = \dfrac{a_2(\epsilon_3^2 + c_2^2\epsilon_2^2 - 2c_2\epsilon_2^2)}{(c_3\epsilon_3 - \epsilon_3)^2}, \\ b_3 = a_2\epsilon_3 + b_2 - a_3\epsilon_3. \end{cases}$$

In Figure 2.6, we show the calibrated quantile function for the case $(\epsilon_1, \epsilon_2, \epsilon_3) = (0.005, 0.025, 0.1)$ and $(c_1, c_2, c_3) = (4, 3, 2.5)$. Note that the condition $c_2\epsilon_2 \leqslant \epsilon_3$ is needed here.

Figure 2.6: Calibrated quantile function when $(\epsilon_1, \epsilon_2, \epsilon_3) = (0.005, 0.025, 0.1)$ and $(c_1, c_2, c_3) = (4, 3, 2.5)$



Although we cannot solve the n-point constraint problem in general, we can instead discuss calibration from a given PELVE curve, which is the problem addressed in the next section.

2.4 Calibration from a curve constraint

By the location-scale invariance properties of the PELVE, we know that the solution cannot be unique. Conversely, it would be interesting to ask whether all solutions can be linearly transformed from a particular solution; that is, for a given function $\epsilon \mapsto \Pi(\epsilon)$, whether

the set $\{X \in \mathcal{X} : \Pi_X = \Pi\}$ is a location-scale class. This question, as well as identifying X satisfying $\Pi_X = \Pi$, is the main objective of this section.

2.4.1 PELVE and dual PELVE

First, we note that calibrated distributions from an entire PELVE curve $\epsilon \mapsto \Pi(\epsilon)$ on $(0, 1)$ would be unnatural, because the existence of the PELVE requires $\mathbb{E}[X] \leqslant \mathrm{VaR}_\epsilon(X)$ which may not hold for ϵ not very small. Thus, the PELVE curve Π_X does not behave well on some parts of $(0, 1)$. To address this issue, we introduce a new notion called the dual PELVE and an integral equation which can help us to calibrate the distribution by differential equations. The dual PELVE is defined by moving the multiplier in PELVE from the ES side to the VaR side.

Definition 2.1. For $X \in L^1$, the dual PELVE function of X at level $\epsilon \in (0, 1]$ is defined as

$$\pi_X(\epsilon) = \inf\left\{d \geqslant 1 : \mathrm{ES}_\epsilon(X) \leqslant \mathrm{VaR}_{\epsilon/d}(X)\right\}, \quad \epsilon \in (0, 1].$$

The existence and uniqueness of $\pi_X(\epsilon)$ can be shown in the same way as the existence and uniqueness of the PELVE. There are advantages and disadvantages of working with both notions; see Li and Wang (2022, Remark 2). In our context, the main advantage of using the dual PELVE is that $\pi_X(\epsilon)$ is finite for all $\epsilon \in (0, 1]$, while $\Pi_X(\epsilon)$ is finite only when $\mathbb{E}[X] \leqslant \mathrm{VaR}_\epsilon(X)$.

Note that for X with a discontinuous quantile function, there may not exist d such that $\mathrm{ES}_\epsilon(X) = \mathrm{VaR}_{\epsilon/d}(X)$. In order to guarantee the above equivalence, we make the following assumption for the quantile function, represented by general function f.

Assumption 2.1. *The function f is strictly decreasing and continuous, and $\int_0^1 |f(s)|\,\mathrm{d}s < \infty$.*

Let $\mathcal{X}$ be the set of $X \in L^1$ with quantile function satisfying Assumption 2.1. The requirement that the quantile function of X is continuous and strictly decreasing is equivalent to that the distribution function is continuous and strictly increasing in $(\text{ess-inf}(X), \text{ess-sup}(X))$; see Embrechts and Hofert (2013). We limit our discussion to random variables $X \in \mathcal{X}$, which include the most common models in risk management.

Proposition 2.2. *For X with quantile function satisfying Assumption 2.1 and $\epsilon \in (0,1)$, we have $\Pi_X(\epsilon/\pi_X(\epsilon)) = \pi_X(\epsilon)$ and $\pi_X(\Pi_X(\epsilon)\epsilon) = \Pi_X(\epsilon)$ if $\mathbb{E}[X] \leqslant \mathrm{VaR}_\epsilon(X)$. Furthermore, $\pi_X(\epsilon)$ is the unique solution $d \geqslant 1$ to the equation*

$$\mathrm{ES}_\epsilon(X) = \mathrm{VaR}_{\epsilon/d}(X).$$

It is straightforward to verify Proposition 2.2. By Proposition 2.2, we can calibrate the distribution functions from dual PELVE instead of PELVE, and the calibrated distributions should satisfy the equation $\mathrm{ES}_\epsilon(X) = \mathrm{VaR}_{\epsilon/d}(X)$.

2.4.2 An integral equation associated with dual PELVE

In order to calibrate distributions from the dual PELVE, we can equivalently focus on quantile functions. Let us consider $X \in \mathcal{X}$ and $f(s) = \mathrm{VaR}_s(X)$. Then, solving $\pi_X(\epsilon)$ is the same as solving z in following equation:

$$\int_0^y f(s)\,\mathrm{d}s = yf(zy) \tag{2.3}$$

for $y = \epsilon$. The solution is $z = 1/\pi_X(y)$. As $f(s) = \mathrm{VaR}_s(X)$, f satisfies Assumption 2.1. Denote by $\mathcal{C}$ the set of all f satisfying Assumption 2.1. For any $f \in \mathcal{C}$, the existence of the solution z is guaranteed by the mean-value theorem and its uniqueness is obvious. For $y \in (0,1]$, let $z_f(y)$ be the solution to (2.3) associated with f. Clearly, $z_f(y) \leqslant 1$ and $y \mapsto yz_f(y)$ is strictly increasing. This is similar to Lemma 2.1 for the two-point case. Obviously, $z_f(y)$ is also location-scale invariant under linear transformation on $f \in \mathcal{C}$. That is, $z_{\lambda f+b} = z_f$ for $\lambda > 0$ and $b \in \mathbb{R}$. Furthermore, z_f is continuous as f is continuous and strictly decreasing. The next proposition is a simple connection between z_f and π_X.

Proposition 2.3. *For any f satisfying Assumption 2.1, $X = f(U)$ for some $U \sim \mathrm{U}(0,1)$ has the dual PELVE $\pi_X(y) = 1/z_f(y)$ for all $y \in (0,1)$ where z_f is solution to (2.3). For X with quantile function satisfying Assumption 2.1, there exists f satisfying Assumption 2.1 such that $X = f(U)$ for some $U \sim \mathrm{U}(0,1)$ and the solution to (2.3) is $z_f(y) = 1/\pi_X(y)$ for all $y \in (0,1)$.*

Proof. For any f satisfying Assumption 2.1, let $F(x) = 1 - f^{-1}(x)$. Hence, F is a continuous and strictly increasing distribution function and $F^{-1}(s) = f(1-s)$ for $s \in (0,1)$. Let $U \sim \mathrm{U}(0,1)$ and $X = F^{-1}(U) = f(1-U)$. Then $X \in \mathcal{X}$ and $X \sim F$. As $F^{-1}(1-s) = f(s)$, we have $\pi_X(y) = 1/z_f(y)$. Take $U' = 1 - U$. We have $X = f(U')$ and $U' \sim \mathrm{U}(0,1)$.

For $X \in \mathcal{X}$, let $f(s) = \mathrm{VaR}_s(X)$. Then, we have $z_f(y) = 1/\pi_X(y)$ for $y \in (0,1]$. Furthermore, we have $F^{-1}(s) = f(1-s)$. Therefore, there exists $U \sim \mathrm{U}(0,1)$ such that $X = f(1-U)$. Let $U' = 1 - U$. Then, we have $X = f(U')$ and $U' \sim \mathrm{U}(0,1)$. □

Proposition 2.3 allows us to study z instead of π for the calibration problem. The integral equation (2.3) can be very helpful in characterizing the distribution from the dual PELVE.

Some examples of π_X and z_f are listed in Table 2.2, which is corresponding to the PELVE presented in Table 2.1.

Table 2.2: Example of π_X and z_f

X	$\pi_X(\epsilon)$	f	z_f
$\mathrm{U}(0,1)$	$\pi_X(\epsilon) = 2$	$f(x) = 1 - x$	$z_f(y) = 1/2$
$\mathrm{Exp}(\lambda)$	$\pi_X(\epsilon) = e$	$f(x) = -\log(x)/\lambda$	$z_f(y) = 1/e$
$\mathrm{GPD}(\xi)$	$\pi_X(\epsilon) = (1-\xi)^{-\frac{1}{\xi}}$	$f(x) = \begin{cases} 1/\xi\left(x^{-\xi} - 1\right), & \xi \neq 0 \\ -\log(x), & \xi = 0 \end{cases}$	$z_f = (1-\xi)^{\frac{1}{\xi}}$

For a given dual PELVE curve π, we find the solution to the integral equation by the following steps.

1. Let $z(y) = \frac{1}{\pi(y)}$ for all $y \in (0,1]$.
2. Find $f \in \mathcal{C}$ that satisfies $\int_0^y f(s)\,\mathrm{d}s = yf\left(z(y)y\right)$ for all $y \in (0,1]$.
3. By Proposition 2.3, $X = f(U)$ for some $U \sim \mathrm{U}(0,1)$ will have the given dual PELVE π.

Therefore, we will focus on characterizing f from a given $z : (0,1] \to (0,1]$ below. Generally, it is hard to characterize f explicitly. We first formulate the problem as an advanced differential equation, which helps us to find solutions.

2.4.3 Advanced differential equations

In this section, we show that the main objective (2.3) can be represented by a differential equation. The use of differential equations in computing risk measures has not been actively developed. The only paper we know is Balbás et al. (2020) which addresses a different problem.

Let us recall the integral equation (2.3) from Section 2.4.2. For a function $f \in \mathcal{C}$, we solve the function $z_f : (0,1) \to \mathbb{R}$ from (2.3). We represent (2.3) by an advanced differential equation using the following steps.

1. Let $\omega_f(y) = y z_f(y)$. It is easy to see that $z_f(y) \leqslant 1$. Hence, ω_f is strictly increasing and continuous on $(0,1]$ and $\omega_f(y) \leqslant y$.

2. Let u_f be the inverse function of ω_f. We have that $u_f : (0, z_f(1)] \mapsto (0,1]$ is a continuous and strictly increasing function and $u_f(w) \geqslant w$.

3. Replacing y with $u_f(w)$ in (2.3), we have $f(w) = \int_0^{u_f(w)} f(w)\,\mathrm{d}s/u_f(w)$.

4. Assume u_f is continuously differentiable. It is clear that f is continuously differentiable on $(0, z_f(1))$. Hence, we can represent (2.3) by the following advanced differential equation
$$f'(w) + \frac{u_f'(w)}{u_f(w)}\left(f(w) - f(u_f(w))\right) = 0.$$

For a given function $z : (0,1] \to \mathbb{R}$, let $u = \omega^{-1}$ such that $\omega(y) = yz(y)$ for $y \in (0,1]$. Then, we solve the function f by the following differential equation
$$f'(w) + \frac{u'(w)}{u(w)} f(w) - \frac{u'(w)}{u(w)} f(u(w)) = 0. \tag{2.4}$$

If $z = 1/\pi_X$ for some $X \in \mathcal{X}$, then u is a strictly increasing and continuous function such that $u(w) \geqslant w$. Furthermore, if z is continuously differentiable, then we can characterize

all $X \in \mathcal{X}$ with $\pi_X = 1/z$ by (2.4). As $u'(w)/u(w) \geqslant 0$ and $u(w) \geqslant w$, (2.4) is a linear advanced differential equation which is well studied in the literature. In Berezansky and Braverman (2011), it is shown that there exists a non-oscillatory solution for (2.4).

2.4.4 The constant PELVE curve

We first solve the case that $z(y) = c$ for all $y \in (0, 1]$ and some constant $c \in (0, 1)$. As we can see from Table 2.2, the power function and logarithm function have constant z_f. If $f(x) = \lambda x^\alpha + b$ for $\alpha > -1$, we can see that $(\alpha + 1)^{-1/\alpha} = c$. In this section, we can characterize all the other solutions which can not be expressed as a linear transformation of the power function. That is, we will see that the set

$$\{f \in \mathcal{C} : z_f(y) = z(y),\ y \in (0, 1]\}$$

is not a location-scale class. Hence, we can answer the question at the beginning of the section; that is, in the case the PELVE is a constant, the set $\{X \in \mathcal{X} : \Pi_X = c\}$ is not a location-scale class.

Theorem 2.2. *For $c \in (0, 1)$, any X with quantile function satisfying Assumption 2.1 and $\pi_X(\epsilon) = 1/c$ for $\epsilon \in (0, 1)$ can be written as $X = f(U)$ for some $U \sim \mathrm{U}(0, 1)$ and f satisfying Assumption 2.1. Furthermore, such f has the form*

$$f(y) = C_1 + C_2 y^\alpha + O\left(y^\zeta\right),$$

where α is the root of $(\alpha + 1)^{-1/\alpha} = c$, $\zeta > \max\{0, \alpha\}$, $C_1, C_2 \in \mathbb{R}$, $C_2\alpha < 0$ and $O(y^\zeta)$ is a function such that $\limsup_{y\to 0} O(y^\zeta)/y^\zeta$ is a constant.

The proof of Theorem 2.2 is provided in Section 2.9. As we can see, Theorem 2.2 characterizes all $X \in \mathcal{X}$ such that $\pi_X(\epsilon) = 1/c$. If $c \in (0, 1/e)$, α is negative. As $\zeta > 0$, we can see that $X = f(U)$ is regularly varying of index α. Hence, one can then consider the Pareto distribution with survival function $S(x) = x^\alpha$ as a representative solution for the tail behavior. An open question is that, in the general case that the PELVE is not necessarily constant, whether all the solutions behave similarly regarding their tail behavior.

Another interesting implication of the theorem and its proof is that one can give a non-trivial solution for z is a constant.

Example 2.3. For $c \in (0,1)$, let (θ, η) be a solution of

$$\begin{cases} c\log c = -\frac{\eta \exp\left(-\frac{\eta}{\tan(\eta)}\right)}{\sin(\eta)}, \\ \theta = -\frac{\eta}{\tan(\eta)}. \end{cases}$$

Then, the function f, given by

$$f(y) = C_1 + C_2 y^{\alpha} + C_3 y^{\zeta} \sin(-\sigma \log(y)), \quad 0 < y < 1, \tag{2.5}$$

satisfies $\int_0^y f(s)\,\mathrm{d}s = yf(cy)$ and Assumption 2.1, where α solves $(\alpha+1)^{-1/\alpha} = c$, $\zeta = \theta/\log c - 1$, $\sigma = -\eta/\log c$, C_2 is a constant such that $C_2\alpha < 0$ and $0 < C_3 < -C_2\alpha/(\zeta + |\sigma|)$.

If we take $C_3 = 0$, we get the simplest power function for $z(x) = c$. If $C_3 \neq 0$, the solution (2.5) is not a linear transformation of the power function solution.

Let us look at the example where $\pi(\epsilon) = 2$ for all $\epsilon \in (0,1]$, which means $z(y) = 1/2$ for $y \in (0,1]$. As we have seen in Table 2.2, $f(y) = 1 - y$ can be a solution that leads to $X \sim \mathrm{U}(0,1)$. Furthermore, according to Example 2.3, we can have another solution

$$f(y) = 1 - y^{\alpha} + Cy^{\zeta} \sin(-\sigma \log(y)),$$

where $\alpha = 1$, $C = 0.05096$, $\zeta = 4.0184$ and $\sigma = -15.4090$. In the left of Figure 2.7, we have depicted the two solutions for f. We can see they are quite different when y goes to 1. In the right of Figure 2.7, we numerically calculate z_f for $f(y) = 1 - y^{\alpha} + Cy^{\zeta}\sin(-\sigma\log(y))$. We can see its numerical value is almost $1/2$ and the discrepancy is due to limited computational accuracy.

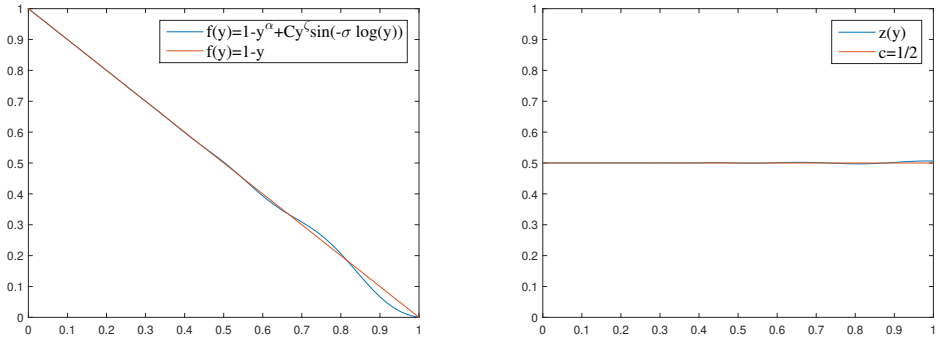
By letting $X = f(U)$, we get $\pi_X(\epsilon) = 2$ for all $\epsilon \in (0,1]$ and such X does not follow the uniform distribution.

2.4.5 A numerical method

In general, it is hard to get an explicit solution to (2.4). Here we present a numerical method to solve (2.4). Let us introduce the following process.

1. Let $a_0 = 1$, $a_1 = a$, ...,$a_n = u^{-1}(a_{n-1})$.

Figure 2.7: Non-unique calibrated functions for $z(y) = 1/2$.



2. For $a \in (0,1)$, let ξ be the solution to $(1-\xi)^{\frac{1}{\xi}} = a$. Let

$$f_0(x) = \begin{cases} \frac{1}{\xi}\left(x^{-\xi} - 1\right), & \xi \neq 0, \\ -\log(x), & \xi = 0, \end{cases} \tag{2.6}$$

on $[a, 1]$.

3. We can solve the following ODE on $[a_2, a_1]$:

$$f_1'(w) + \frac{u'(w)}{u(w)} f_1(w) = \frac{u'(w)}{u(w)} f_0(u(w)), \quad w \in [a_2, a_1].$$

4. Now we can repeat step 3 by induction on $[a_{n+1}, a_n]$ for $n > 1$ by solving

$$f_n'(w) + \frac{u'(w)}{u(w)} f_n(w) = \frac{u'(w)}{u(w)} f_{n-1}(u(w)), \quad w \in [a_{n+1}, a_n].$$

5. In general, the solution for differential equation $\frac{dy}{dx} + P(x)y = Q(x)$ is

$$y = e^{-\int^x P(\lambda)\,\mathrm{d}\lambda}\left[\int^x e^{\int^\lambda P(\varepsilon)\,\mathrm{d}\varepsilon} Q(\lambda)\,\mathrm{d}\lambda + C\right].$$

So, we get the following solution for f_n:

$$f_n(w) = e^{\int_w^{a_n} \frac{u'(\lambda)}{u(\lambda)}\,\mathrm{d}\lambda}\left[f_{n-1}(a_n) - \int_w^{a_n} e^{-\int_\lambda^{a_n} \frac{u'(\varepsilon)}{u(\varepsilon)}\,\mathrm{d}\varepsilon} \frac{u'(\lambda)}{u(\lambda)} f_{n-1}(u(\lambda))\,\mathrm{d}\lambda\right], \; w \in [a_{n+1}, a_n].$$

6. Finally, let $f = f_n$ on $[a_{n+1}, a_n]$.

Note that since we start with a strictly decreasing function, then from equation (2.4) we have

$$f'(w) = \frac{u'(w)}{u(w)}\left(f(u(w)) - f(w)\right) < 0,$$

so f remains strictly decreasing.

The solution produced by the numerical method heavily relies on f_0. The equation (2.4) does not have a unique solution, but the solution from the above process is unique. We set f_0 as (2.6) by assuming z can be extended from $(0, 1]$ to $\mathbb{R}^+$ and set $z(y) = a$ for all $y > 1$. We use this assumption for simplification as we can know that (2.6) satisfies (2.4) for a constant z from Section 2.4.4. This choice of f_0 is the same as the choice of $k(\epsilon)$ in the two-point calibration problem, and this reflects our subjective view of the importance of the Pareto distribution in risk management. Especially, when $z(y) = c$ for some constant c, we have $u(x) = x/c$. Therefore, step 5 gives

$$f_n(w) = \frac{a_n}{w}\left[f_{n-1}(a_n) - \int_w^{a_n} \frac{1}{a_n} f_{n-1}\left(\frac{\lambda}{c}\right) \mathrm{d}\lambda\right].$$

If we set f_0 as (2.6), we can have f_1 also in the form of (2.6). Then, it is obvious that f_n is also in the form of (2.6). Therefore, the numerical method gives the simplest power function or logarithm function when $z(y)$ is a constant on $(0, 1]$ as Table 2.2, which leads to the generalized Pareto distribution for X.

2.4.6 Numerical calibrated quantile function

Now let us explore the method in Section 2.4.5 with simulation. Here we present the results for a few cases. In Figures 2.8 to 2.11, we compare the solution from the numerical method with the standard formula in Table 2.2 in the left panel, and compare $\int_0^y f(s)\,\mathrm{d}s$ with $yf(z(y)y)$ to validate the equation (2.3) in the right panel.

We first try some examples where z is constant as shown in Table 2.2, i.e. $z(x) = 1/2$ (Figure 2.8), $z(x) = 1/e$ (Figure 2.9) and $z(x) = 0.9^{10}$ (Figure 2.10). For Figure 2.8 to 2.10, we can see that the numerical method provides exactly the same function f as Table 2.2.

In Figure 2.11, we check the case $z(x) = \log\left(x/(1 - e^{-x})\right)/x$. The function $f(x) = e^{-x}$ satisfies (2.3). We can see that the solution from the numerical method is close to a function of the form $f(x) = \lambda e^{-x} + b$, which is known to satisfy the integral equation.

Figure 2.8: Calibrated function and validation for $z(x) = 1/2$

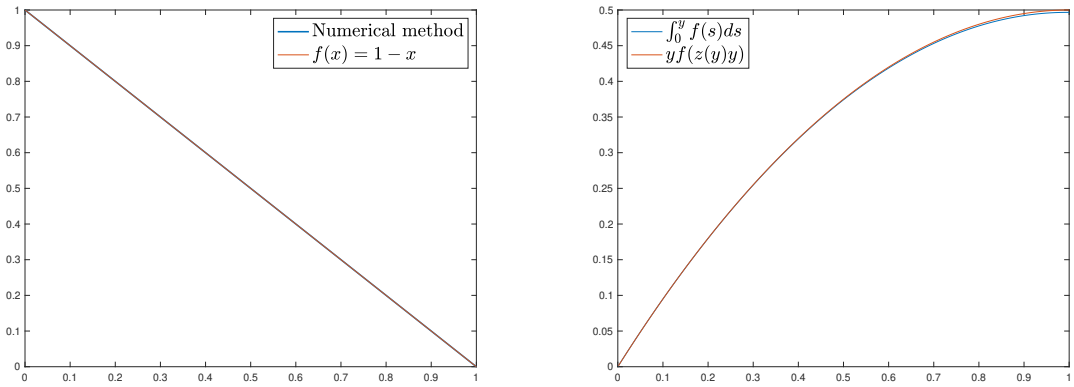
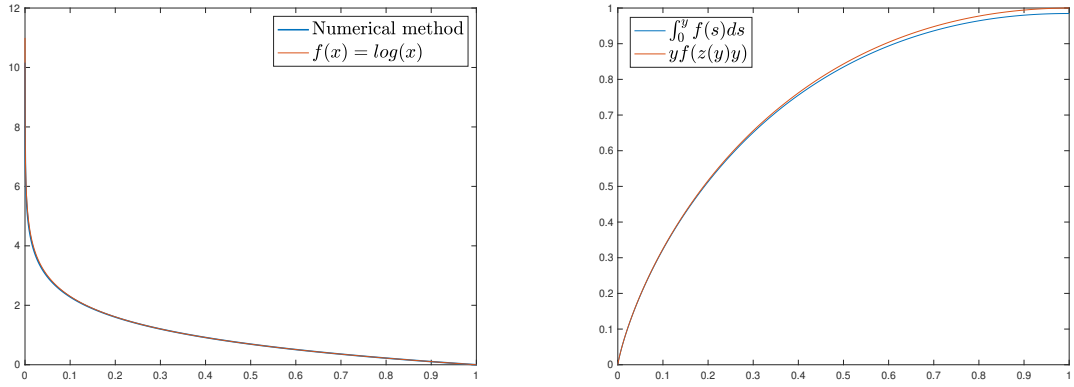


Figure 2.9: Calibrated function and validation for $z(x) = 1/e$



2.5 Technical properties of the PELVE

We now take a turn to study several additional properties of PELVE. In particular, we will obtain results on the monotonicity and convergence of the dual PELVE as well as the PELVE.

2.5.1 Basic properties of dual PELVE

The following proposition that shows the PELVE and dual PELVE share some basic properties such as monotonicity (i), location-scale invariance (ii) and shape relevance (iii)-(iv) below.

Figure 2.10: Calibrated function and validation for $z(x) = 0.9^{10}$

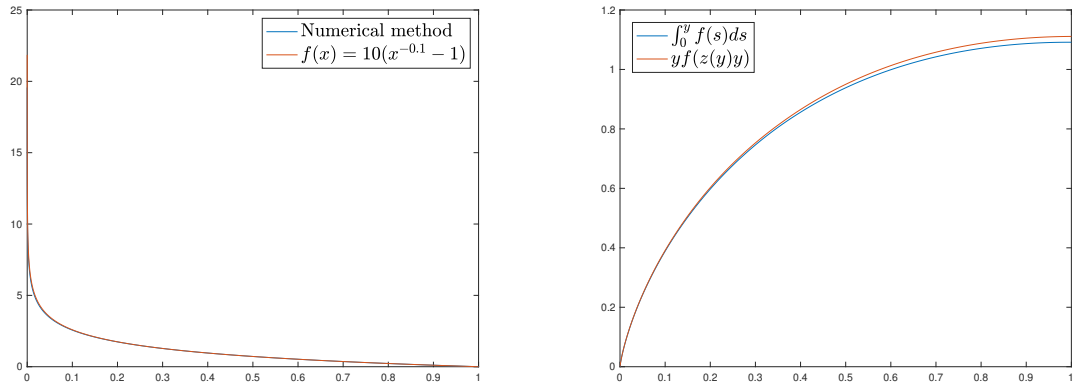
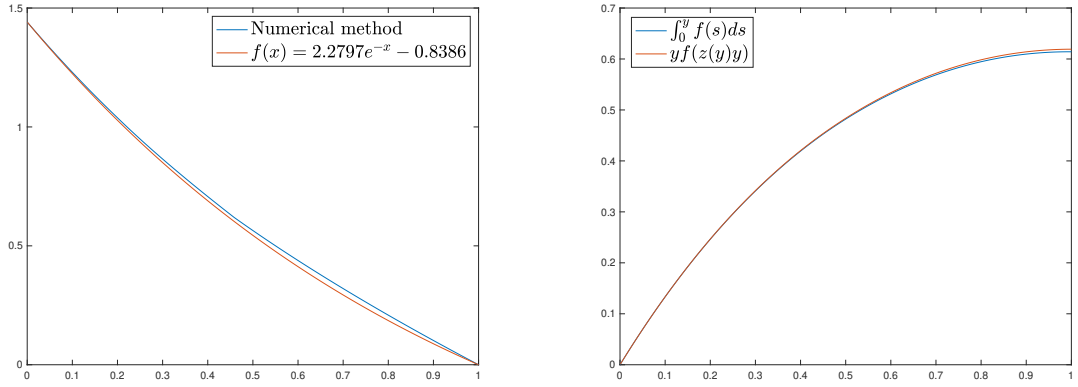


Figure 2.11: Calibrated function and validation for $z(x) = \log\left(x/(1-e^{-x})\right)/x$



Proposition 2.4. *Suppose the quantile function of X satisfies Assumption 2.1 and $\epsilon \in (0, 1]$.*

(i) $\Pi_X(\epsilon)$ *is increasing (decreasing) in* ϵ *if and only if so is* $\pi_X(\epsilon)$.

(ii) For all $\lambda > 0$ *and* $a \in \mathbb{R}$, $\pi_{\lambda X+a}(\epsilon) = \pi_X(\epsilon)$.

(iii) $\pi_{f(X)}(\epsilon) \leqslant \pi_X(\epsilon)$ *for all strictly increasing concave functions:* $f : \mathbb{R} \to \mathbb{R}$ *with* $f(X) \in \mathcal{X}$.

(iv) $\pi_{g(X)}(\epsilon) \geqslant \pi_X(\epsilon)$ *for all strictly increasing convex functions:* $g : \mathbb{R} \to \mathbb{R}$ *with* $g(X) \in \mathcal{X}$.

The statements (ii)-(iv) are parallel to the corresponding statements in Theorem 1 of Li and Wang (2022) on PELVE. The proof of Proposition 2.4 is put in Section 2.10. Proposition

2.4 allows us to study the monotonicity and convergence of the PELVE by analyzing the corresponding properties of the dual PELVE, which is more convenient in many cases. In the following sections, we focus on finding the conditions which make the dual PELVE monotone and convergent at 0. By Proposition 2.4, those conditions can also apply to the PELVE.

2.5.2 Non-monotone and non-convergent examples

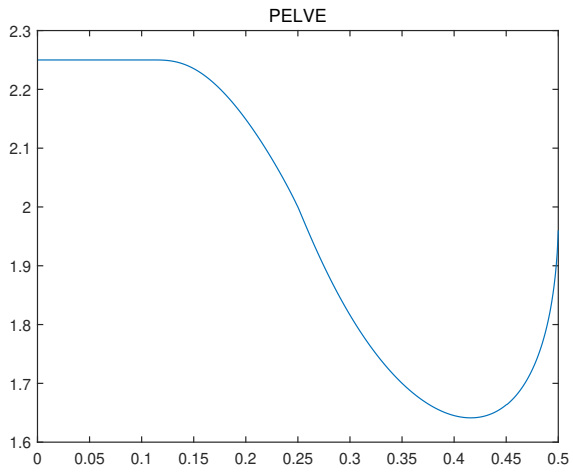
In this section, we study the monotonicity and convergence of dual PELVE. For monotonicity, we have shown that some well-known distributions such as normal distribution, t-distribution and lognormal distribution have monotone PELVE curves in Example 2.3. However, the PELVE is not monotone for all $X \in \mathcal{X}$. Below we provide an example.

Example 2.4 (Non-monotone PELVE)**.** Let us consider the following density function g on $[-2, 2]$,

$$g(x) = \frac{1}{2}\left((x+2)\mathbb{1}_{\{x\in[-2,-1]\}} - x\mathbb{1}_{\{x\in(-1,0]\}} + x\mathbb{1}_{\{x\in(0,1]\}} + (2-x)\mathbb{1}_{\{x\in(1,2]\}}\right).$$

For X with density function g, Figure 2.12 presents the value of $\Pi_X(\epsilon)$ for $\epsilon \in (0, 0.5)$. As one can see, the PELVE is not necessarily decreasing, and so is the dual PELVE.

Figure 2.12: PELVE for X with density g



For the convergence, it is clear that $\pi_X(\epsilon)$ is continuous in $(0,1)$ for $X \in \mathcal{X}$. Therefore, $\lim_{\epsilon\to p} \pi_X(\epsilon)$ exists for all $p \in (0,1)$. However, both $\Pi_X(\epsilon)$ and $\pi_X(\epsilon)$ are not well defined at $\epsilon = 0$. If $\lim_{\epsilon\to 0} \pi_X(\epsilon)$ exists, we can define $\pi_X(0)$ as the limit, and $\Pi_X(0)$ similarly. However, the following example shows that the limit does not exist for some $X \in \mathcal{X}$.

Example 2.5 (No limit at 0)**.** We can construct a random variable $X \in \mathcal{X}$ such that $\lim_{\epsilon\to 0} \pi_X(\epsilon)$ does not exist from the integral equation (2.3) in Section 2.4.2. Equivalently, we will find a continuous and strictly decreasing function $f \in \mathcal{C}$ such that $\lim_{y\to 0} z_f(y)$ does not exist. Let c be the Cantor ternary function on $[0,1]$. Note that $x \mapsto c(x)$ is continuous and increasing on $(0,1)$ and $c(x/3) = c(x)/2$. Let $f(x) = -c(x) - x^{\log 2/\log 3}$. It is clear that $f \in \mathcal{C}$ and $f(x/3) = f(x)/2$. For each $y \in (0,1]$, we have

$$\begin{aligned} yf\left(z_f(y)y\right) &= \int_0^y f(x)\,\mathrm{d}x \\ &= 2\int_0^y f\left(\frac{1}{3}x\right)\,\mathrm{d}x = 6\int_0^{\frac{1}{3}y} f(x)\,\mathrm{d}x = 2yf\left(\frac{1}{3}yz_f\left(\frac{1}{3}y\right)\right) = yf\left(yz_f\left(\frac{1}{3}y\right)\right). \end{aligned}$$

Since f is strictly decreasing, $z_f(y) = z_f(y/3)$ for $y \in (0,1]$. It means that $z_f(y)$ is a constant on $(0,1]$ if $\lim_{y\to 0} z_f(y)$ exists. Now, let us look at two particular points of $z_f(y)$. We can show that $z_f(1) \neq z_f(4/9)$. Let $z = (\log 2/\log 3 + 1)^{-(\log 3/\log 2)}$. Then, we have $1/3 < z \approx 0.46 < 1/2$. For $y = 1$, we have $\int_0^1 c(s)\,\mathrm{d}s = c(z) = 1/2$ and $\int_0^c s^{\log 2/\log 3}\,\mathrm{d}s = z^{\log 2/\log 3}$. Therefore, we get $z_f(1) = z < 1/2$. For $y = 4/9$, we have

$$\begin{aligned} f\left(\frac{4}{9}z_f\left(\frac{4}{9}\right)\right) &= \frac{9}{4}\int_0^{4/9} f(s)\,\mathrm{d}s \\ &= -\frac{9}{4}\left(\frac{1}{\frac{\log 2}{\log 3}+1}\left(\frac{4}{9}\right)^{\frac{\log 2}{\log 3}+1} + \frac{1}{12} + \frac{1}{2}\left(\frac{4}{9} - \frac{1}{3}\right)\right) < -0.68 < f\left(\frac{2}{9}\right) \approx -0.64. \end{aligned}$$

As f is strictly decreasing, we have $(4/9)z_f(4/9) > 2/9$ which implies $z_f(4/9) > 1/2 > z_f(1)$. As a result, $\lim_{y\to 0} z_f(y)$ does not exist. Therefore, we have a continuous and strictly decreasing f such that $\lim_{y\to 0} z_f(y)$ does not exist.

2.5.3 Sufficient condition for monotonicity and convergence

In risk management applications, for a random variable X modeling a random loss, the behavior of its tail is the most important. Let $F^{[p,1]}$ be the upper p-tail distribution of F

(see e.g., Liu and Wang (2021)), namely

$$F^{[p,1]}(x) = \frac{(F(x)-p)_+}{1-p}, \quad x \in \mathbb{R}.$$

We will see that the dual PELVE of $F^{[p,1]}$ is a part of the dual PELVE of F.

Lemma 2.2. *Let F be the distribution function of X with quantile function satisfying Assumption 2.1. For $p \in (0,1)$ and $X' \sim F^{[p,1]}$, it holds*

$$\pi_{X'}(\epsilon) = \pi_X(\epsilon(1-p)).$$

Proof. It is clear that $\mathrm{VaR}_\epsilon(X') = \mathrm{VaR}_{\epsilon(1-p)}(X)$ and $\mathrm{ES}_\epsilon(X') = \mathrm{ES}_{\epsilon(1-p)}(X)$. Therefore,

$$\begin{aligned}\pi_{X'}(\epsilon) &= \inf\{d \geqslant 1 : \mathrm{ES}_\epsilon(X') \leqslant \mathrm{VaR}_{\epsilon/d}(X')\} \\ &= \inf\{d \geqslant 1 : \mathrm{ES}_{\epsilon(1-p)}(X') \leqslant \mathrm{VaR}_{\epsilon(1-p)/d}(X')\} = \pi_X(\epsilon(1-p)).\end{aligned}$$

Thus, we have the desired result. □

The tail distribution can provide a condition to check whether the dual PELVE is decreasing.

Proposition 2.5. *Let F be the distribution function of X with quantile function satisfying Assumption 2.1. If $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is convex (concave) for all $p \in (0,1)$, then π_X and Π_X are decreasing (increasing).*

Proof. For any $p \in (0,1)$, let $X' \sim F^{[p,1]}$. By Lemma 2.2, we have $\pi_{X'}(\epsilon) = \pi_X(\epsilon(1-p))$. Furthermore, we have

$$\big(F^{[p,1]}\big)^{-1}(t) = F^{-1}\left((1-p)t+p\right) = F^{-1}\Big((1-p)F\big(F^{-1}(t)\big)+p\Big), \quad t \in [0,1].$$

Let $U \sim \mathrm{U}(0,1)$, $X = F^{-1}(U)$ and $X' = (F^{[p,1]})^{-1}(U)$.

We assume that $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is a convex function on $(\text{ess-inf}(X), \text{ess-sup}(X))$ first. Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly increasing convex function such that $f(x) = F^{-1}((1-p)F(x)+p)$ for $x \in (\text{ess-inf}(X), \text{ess-sup}(X))$. Then, we have $X' = f(X)$. By Proposition 2.4, we get $\pi_{X'}(\epsilon) \geqslant \pi_X(\epsilon)$. As $\pi_{X'}(\epsilon) = \pi_X(\epsilon(1-p))$, we have $\pi_X(\epsilon(1-p)) \geqslant \pi_X(\epsilon)$ for all $p \in (0,1)$. Thus, π_X is decreasing. By Proposition 2.4, we have Π_X is also decreasing.

On the other hand, if $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is concave, we have $\pi_X(\epsilon(1-p)) \leqslant \pi_X(\epsilon)$ for all $p \in (0,1)$ and π_X is increasing. So is Π_X. □

The condition $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is convex (concave) for all $p \in (0,1)$ is generally hard to check. Intuitively, this condition means that $F^{[p,1]}$ has a less heavy tail compared to F. We can further simplify this condition by using the hazard rate function. For $X \in \mathcal{X}$ with distribution function F and density function f, let $S = 1 - F$ be the survival function and $\eta = f/S$ be the hazard rate function. As F is continuous and strictly increasing, S is continuous and strictly decreasing.

Theorem 2.3. *For X with quantile function satisfying Assumption 2.1, let η be the hazard rate function of X. If $1/\eta$ is second-order differentiable and convex (concave), then π_X and Π_X are decreasing (increasing).*

The proof of Theorem 2.3 is provided in Section 2.10.

Example 2.6. For the normal distribution, we can give a short proof of the convexity of $1/\eta$. Let S be the survival function of the standard normal distribution and f its density. Let $I(x) = 1/\eta(x) = S(x)/f(x) = \exp(x^2/2)\int_{-\infty}^{-x} \exp(-s^2/2)\,\mathrm{d}s$. One can easily see that

$$I'(x) = xI(x) - 1 \tag{2.7}$$

which gives $I''(x) = xI'(x) + I(x)$. This with (2.7) implies that

$$I''(x) = \left(1 + x^2\right) I(x) - x. \tag{2.8}$$

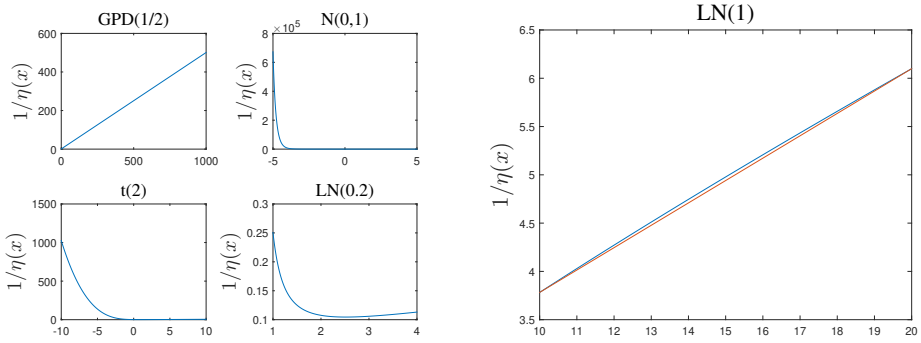
First, consider the negative line i.e., $x < 0$. In this case (2.7) and (2.8) imply $I'(x) = xI(x) - 1 < 0$, and $I''(x) = (1+x^2)I(x) + (-x) > 0$. The implication of the two relations is that I is a convex and decreasing function on negative line. Now we consider the case $x > 0$. In this case, let $i(x) = I'(-x)$. From what we have proved it is clear that i is an increasing function on $x > 0$. On the other hand, we have $I(x) + I(-x) = 1/f(x) = \sqrt{2\pi}\exp(x^2/2)$. This combined with (2.7) gives us

$$I'(x) = x\left(I(x) + I(-x)\right) + i(x) = x\sqrt{2\pi}\exp\left(x^2/2\right) + i(x), x > 0.$$

This means I' is an increasing function on $x > 0$ as it is a summation of two other increasing functions, so I is convex on the positive line as well.

Figure 2.13 presents the curve $1/\eta$ for the generalized Pareto distribution, the Normal distribution, the t-distribution and the Lognormal distribution. For distributions GPD(1/2), N(0, 1) and t(2), we can see that the curves $1/\eta$ are convex, and this coincides with decreasing PELVE shown in Example 2.2. For the Lognormal distribution, the shape of $1/\eta$ depends on σ. As shown in Example 2.2, the PELVE for LN(σ) is visibly decreasing for $\sigma^2 = 0.04$ and increasing for $\sigma = 1$. Corresponding to the above observations, we see that $1/\eta$ is convex for $\sigma^2 = 0.04$ and concave for $\sigma^2 = 1$.

Figure 2.13: $1/\eta$ for GPD(1/2), N(0, 1), t(2), LN(0, 2) and LN(1) in blue curves; in the right panel, the red curve is linear



Corollary 2.1. *If the hazard rate of a random variable X is second-order differentiable and concave, then π_X and Π_X are decreasing.*

Proof. Just note that if η is concave, then $\eta\eta''$ is non-positive. It follows that

$$\left(\frac{1}{\eta}\right)'' = \left(-\frac{\eta'}{\eta^2}\right)' = \frac{2\left(\eta'\right)^2 - \eta\eta''}{\eta^3} \geqslant 0.$$

Thus, $1/\eta$ is convex, and the desired statement follows from Theorem 2.3. □

The corollary above is a result of the fact that the concavity of η implies convexity of $1/\eta$. Therefore, concave η always leads to decreasing PELVE. For example, the Gamma distribution G(α, λ) with density $f(x) = \lambda^\alpha t^{\alpha-1} e^{-\lambda t}/\Gamma(\alpha)$ has concave hazard rate function when $\alpha > 1$. Furthermore, by Theorem 2.3, we can easily find more well-known distributions that have decreasing π_X.

As the tail distribution determines π_X around 0, we can focus on the tail distribution to discuss the convergence of π_X at 0. Note that if the survival distribution function is regularly

varying, then its tail parameter one-to-one corresponds to the limit of Π_X at 0 as shown by Theorem 3 of Li and Wang (2022). Hence, the limit of Π_X, if it exists, can be useful as a measure of tail heaviness, and it is well defined even for distributions that do not have a heavy tail. By the monotone convergence theorem, we have $\lim_{\epsilon\to 0}\pi_X(\epsilon)$ exists if π_X is monotone. The limit may be finite or infinite.

Corollary 2.2. *For X with quantile function satisfying Assumption 2.1, let η be the hazard rate of X. If $1/\eta(x)$ is second-order differentiable and convex (concave) in $(F^{-1}(\delta), \text{ess-sup}(X))$ for some $\delta \in (0,1)$, then $\lim_{\epsilon\to 0}\pi_X(\epsilon)$ exists. In particular, this is true if η is second-order differentiable and concave on $(F^{-1}(\delta), \text{ess-sup}(X))$.*

Proof. Let $X' \sim F^{[\delta,1]}$. Then, the survival function for X' is $S_{X'}(x) = S(x)/(1-p)$ for $x \geqslant F^{-1}(\delta)$. The density function is $f_{X'}(x) = f(x)/(1-p)$ for $x \geqslant F^{-1}(\delta)$. Therefore, the hazard rate function is $\eta_{X'}(x) = f(x)/S(x) = \eta(x)$ for $x \geqslant F^{-1}(\delta)$.

As $1/\eta(x)$ is convex (concave) when $x > F^{-1}(\delta)$, we have $1/\eta_{X'}(x)$ is convex (concave). By Theorem 2.3, we have $\pi_{X'}(\epsilon)$ is decreasing (increasing) on $(0,1)$. As a result, we have $\pi_X(\epsilon)$ is decreasing (increasing) on $(0,\delta)$ and $\lim_{\epsilon\to 0}\pi_X(\epsilon)$ exists.

By Corollary 2.1, if η is concave on $(F^{-1}(\delta), \text{ess-sup}(X))$, we have $1/\eta$ is convex on $(F^{-1}(\delta), \text{ess-sup}(X))$ and $\lim_{\epsilon\to 0}\pi_X(\epsilon)$ also exists. □

Example 2.7. If $\lim_{\epsilon\to 0}\pi_X(\epsilon)$ is a constant, we have $\lim_{\epsilon\to 0}\Pi_\epsilon(X) = \lim_{\epsilon\to 0}\pi_X(\epsilon)$ as $\pi_X(\Pi_X(\epsilon)\epsilon) = \Pi_X(\epsilon)$. We give the numerical values of $\Pi_X(\epsilon)$ at very small probability levels ϵ for normal, t, and log-normal distributions. These distributions do not have a constant PELVE curve, and using Corollary 2.2 we can check that their PELVE have limits. As we can see from Table 2.3, PELVE can still distinguish the heaviness of the tail even when ϵ is very small. The heavier tailed distributions report a higher PELVE value. For the normal distribution and the log-normal distribution with $\sigma = 0.2$, the value of PELVE is close to $e \approx 2.7183$ as $\epsilon \downarrow 0$. From the numerical values, it is unclear whether $\Pi_X(\epsilon) \to e$ for all log-normal distributions, but there is no practical relevance to compute $\Pi_X(\epsilon)$ for $\epsilon < 10^{-11}$ in applications.

Table 2.3: Values of $\Pi_X(\epsilon)$

Distribution	N	LN(1)	LN(0.5)	LN(0.2)	t(2)	t(3)
$\epsilon = 10^{-10}$	2.6884	2.9167	2.7944	2.7290	4.0000	3.3750
$\epsilon = 10^{-11}$	2.6909	2.9077	2.7920	2.7287	4.0000	3.3750

2.6 Applications to datasets used in insurance

In this section, we apply the PEVLE calibration techniques to datasets used in insurance to show how to use the calibrated distribution in estimating risk measure values and simulation.

2.6.1 Dental expenditure data

In this example, we apply the calibration model to the 6494 complete household component's total dental expenditure data from Medical Expenditure Panel Survey for 2020. An earlier version of the same dataset is used by Behan et al. (2010) to study the relationship between worker absenteeism and overweight or obesity. The main purpose of this experiment is to construct tractable models, with continuous and simple quantile functions, which have similar risk measure values as the original dataset, and the same PELVE at certain levels. We present in Figure 2.14 two quantile functions calibrated from $\Pi_X(\epsilon_1)$ and $\Pi_X(\epsilon_2)$, with $(\epsilon_1, \epsilon_2) = (0.01, 0.05)$ and $(\epsilon_1, \epsilon_2) = (0.05, 0.1)$, respectively. The two calibrated quantile functions are scaled up according to the empirical $\text{VaR}_{\epsilon_1}(X)$ and $\text{VaR}_{\epsilon_2}(X)$. By Theorem 2.1, we can calibrate the quantile functions from Case 4 when $(\epsilon_1, \epsilon_2) = (0.01, 0.05)$, and from Case 5 when $(\epsilon_1, \epsilon_2) = (0.05, 0.1)$. As mentioned before, for $(\epsilon_1, \epsilon_2) = (0.05, 0.1)$, we set the calibrated quantile function in $(0, c_1\epsilon_1)$ as the Pareto quantile function. Hence, there is no flat part in the two calibrated quantile functions shown in Figure 2.14. As we can see, both the two calibrated quantile functions fit the empirical quantile functions well. The calibrated quantile function can be regarded as a special parameterized model for tail distribution, which can fit the value of VaR and ES at specified levels. With the parameterized calibrated

model, we can estimate the value of tail risk measures (see Liu and Wang (2021)) such as ES, VaR, and Range-VaR (RVaR), amongst others. In Tables 2.4 and 2.5, we compute the values of ES and RVaR for the calibrated model and compare them with empirical ES and RVaR values, respectively, where the risk measure RVaR is defined as

$$\mathrm{RVaR}_{\alpha,\beta}(X) = \frac{1}{\beta-\alpha}\int_{\beta}^{\alpha} \mathrm{VaR}_{\gamma}(X)\,\mathrm{d}\gamma$$

for $0 \leqslant \alpha < \beta < 1$; see Cont et al. (2010) and Embrechts et al. (2018). As we scale the calibrated quantile function to empirical $\mathrm{VaR}_{\epsilon_1}(X)$ and $\mathrm{VaR}_{\epsilon_2}(X)$, the calibrated ES and empirical ES are identical at levels $\epsilon_1\Pi_X(\epsilon_1)$ and $\epsilon_2\Pi_X(\epsilon_2)$ by the definition of PELVE. For other probability levels, the calibrated ES and RVaR in Tables 2.4 and 2.5 are close to their empirical counterparts. When $(\epsilon_1, \epsilon_2) = (0.01, 0.05)$, it may only be useful to compute calibrated $\mathrm{ES}_p(X)$ for $p < 0.05\Pi_X(0.05) = 0.11591$ because the calibrated quantile function is arbitrary beyond the level 0.11591. If we need to estimate ES or RVaR for a larger probability level, we can choose a higher ϵ_2 as long as $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X)$ is satisfied. For this dataset, the highest ϵ_2 we can use is 0.1983.

Using the methods in Section 2.3, for quantile levels between $(0, \epsilon_1)$, the distribution calibrated from one point (ϵ_1, c_1) is the same as the one calibrated from two points (ϵ_1, c_1) and (ϵ_2, c_2). Hence, the results for ES_p of the one-point calibrated function are also shown in Tables 2.4 and 2.5 in the cells $p \leqslant \epsilon_1$.

Table 2.4: Empirical ES and calibrated ES for the dental expenditure data

p	0.01	0.05	0.1	0.2	0.3
Empirical ES_p	10073.1	5361.7	3624.8	2317.9	1696.7
Calibrated ES_p from $(\epsilon_1, \epsilon_2) = (0.01, 0.05)$	11703.9	5357.7	3759.1	-	-
Calibrated ES_p from $(\epsilon_1, \epsilon_2) = (0.05, 0.1)$	10878.1	5439.6	3711.3	2293.7	1696.4

Figure 2.14: Empirical and calibrated VaR_ϵ for the dental expenditure data

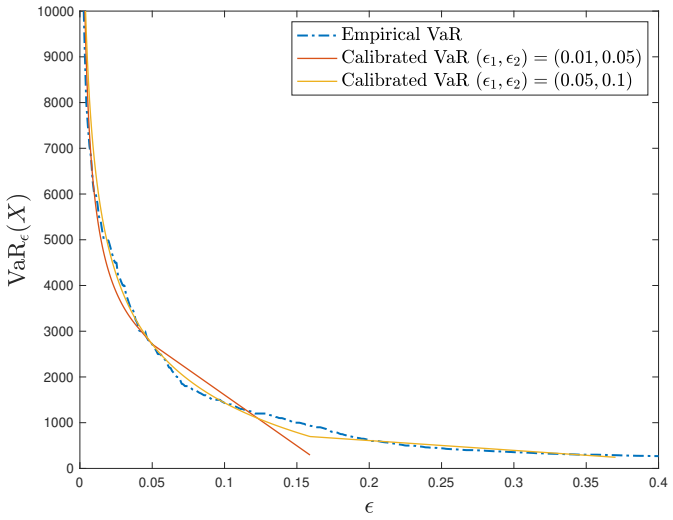


Table 2.5: Empirical RVaR and calibrated RVaR for the dental expenditure data

(α, β)	$(0.01, 0.02)$	$(0.02, 0.05)$	$(0.05, 0.1)$
Empirical $\text{RVaR}_{\alpha,\beta}$	5748.5	3662.4	1887.9
Calibrated $\text{RVaR}_{\alpha,\beta}$ from $(\epsilon_1, \epsilon_2) = (0.01, 0.05)$	5003.7	3360.2	2160.6
Calibrated $\text{RVaR}_{\alpha,\beta}$ from $(\epsilon_1, \epsilon_2) = (0.05, 0.1)$	5634.6	3561.8	1983.1

2.6.2 Hospital costs data

In this example, we apply the calibration process to the Hospital Costs data of Frees (2009) which were originally from the Nationwide Inpatient Sample of the Healthcare Cost and Utilization Project (NIS-HCUP). The data contains 500 hospital costs observations with 244 males and 256 females which can be regarded as the losses of the health insurance policies. Using the calibration model of the two-point constraint problem, we calibrate quantile functions for females and males from PELVE at probability levels $\epsilon_1 = 0.05$ and $\epsilon_2 = 0.1$, which are shown in Figure 2.15. Except for estimating the risk measure, the calibrated distribution is useful in simulation. Assume the insurance company wants to know the top 10% hospital costs; that is $X|X > \text{VaR}_{0.1}(X)$ where X is the hospital costs. There are

only 24 available data for males and 25 available data for females, which would be not enough for making statistically solid decisions. To generate more pseudo-data points, we can simulate data from the calibrated distribution; that is, we simulate data from $F^{[p,1]}$ where F is the calibrated distribution in Figure 2.15. Taking $p = 0.9$, we have $F^{[p,1]}(t) = \mathrm{VaR}_{(1-p)(1-t)}(X)$ with $\mathrm{VaR}_t(X)$ from Figure 2.15. We simulate 1000 data from the calibrated distributions based on PELVE at $\epsilon_1 = 0.05$ and $\epsilon_2 = 0.1$. In Figure 2.16, we present two QQ plots of simulated data against empirical data for females and males respectively. As we can see, the simulated data has a similar distribution as the empirical data. Those simulated pseudo-data points can be used for estimating risk measures or making other decisions. For example, the simulated hospital cost can be used to design health insurance contrasts or set the premium in complex systems, where sometimes methods based on simulated data are more convenient to work with than methods relying on distribution functions. This may be seen as an alternative, smoothed, version of bootstrap; recall that the classic bootstrap sample can only take the values represented in the dataset. Furthermore, we compare the simulated data of hospital costs for females and males in Figure 2.17, which shows that the distribution of the hospital costs for females has a heavier tail than that for males.

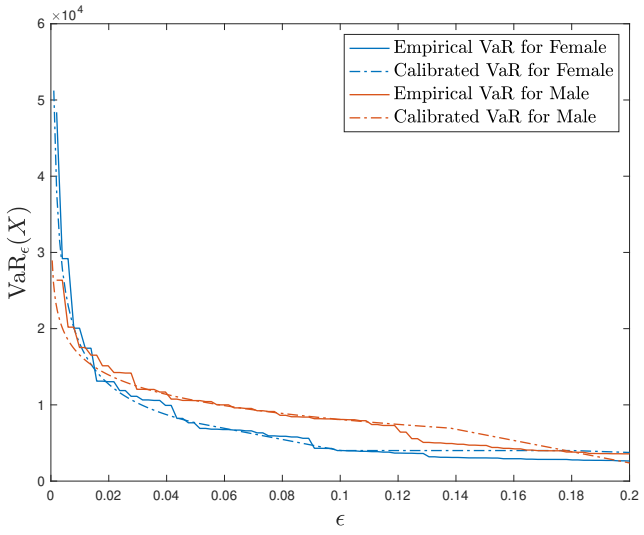
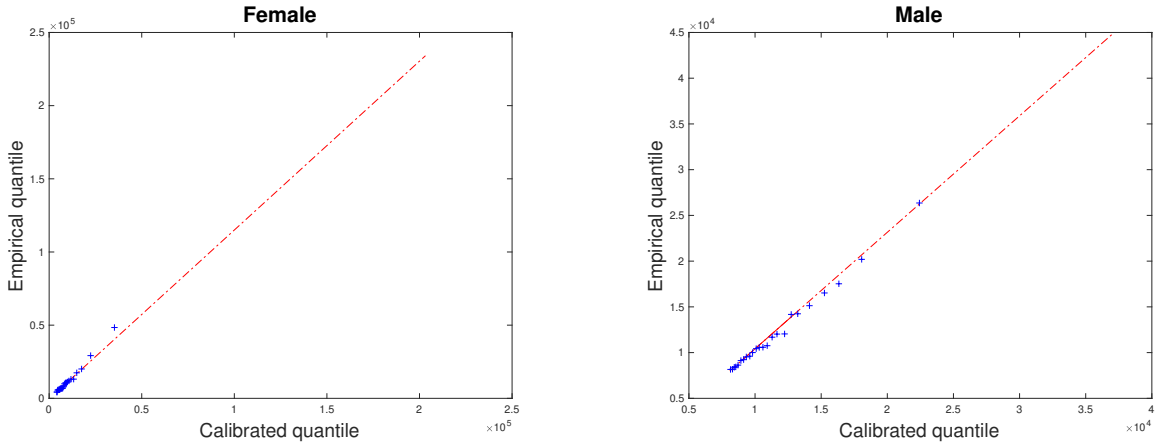


Figure 2.15: Empirical and calibrated VaR_ϵ for the hospital costs data



(a) Hospital costs for female

(b) Hospital costs for male

Figure 2.16: QQ plot: simulated data VS empirical data

2.7 Conclusion

In this chapter, we offer several contributions to the calibration problem and properties of the PELVE. The calibration problem concerns, with some given values from a PELVE curve, how one can build a distribution that has this PELVE. We solve a few settings of calibration based on a one-point constraint, a two-point constraint, or the entire curve constraint. In particular, the calibration for a given PELVE curve involves solving an integral equation $\int_0^y f(s)\,\mathrm{d}s = yf(z(y)\,y)$ for a given function z, and this requires some advanced analysis and a numerical method in differential equations. For the case that z is a constant curve, we can identify all solutions, which are surprisingly complicated. In addition, we see that if π_X is a constant larger than e, which is observed from typical values in financial return data (Li and Wang (2022)), X share the same tail behavior with the corresponding Pareto solution. We also applied our calibration techniques to two datasets used in insurance.

On the technical side, we study whether the PELVE is monotone and whether it converges at 0. We show that the monotonicity of the PELVE is associated with the shape of the hazard rate. If the inverse of the hazard rate is convex (concave), the PELVE is decreasing (increasing). The monotonicity at the tail part of the PELVE leads to conditions to check the convergence of the PELVE at 0. If the inverse of the hazard rate is convex (concave) at the tail of the distribution, the limit of the PELVE at 0 exists.

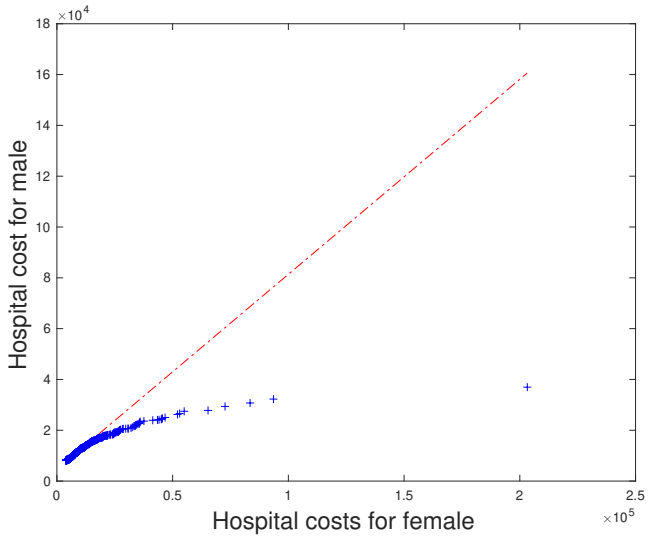


Figure 2.17: QQ plot of simulated data of hospital costs: female VS male

There are several open questions related to PELVE that we still do not fully understand. One particular such question is whether the tail behavior, e.g., tail index, of a distribution is completely determined by its PELVE. We have seen that this holds true in the case of a constant PELVE (see Theorem 2.2), but we do not have a general conclusion. In the case of regularly varying survival functions, Li and Wang (2022, Theorem 3) showed that the limit of PELVE determines its tail parameter, but it is unclear whether this can be generalized to other distributions. Another challenging task is, for a specified curve π on $[0,1]$, to determine whether there exists a model X with $\pi_X = \pi$. The case of n-point constraints for large n may require a new design of verification algorithms. This question concerns the compatibility of given information with statistical models, which has been studied, in other applications of risk management, by Embrechts et al. (2002, 2016) and Krause et al. (2018).

2.8 Omitted proofs in Section 2.3

Proof of Lemma 2.1. As $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X)$ and $\epsilon_1 < \epsilon_2$, $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X) \leqslant \mathrm{VaR}_{\epsilon_1}(X)$. By Proposition 1 in Li and Wang (2022), $\Pi_X(\epsilon_1) < \infty$ and $\Pi_X(\epsilon_2) < \infty$.

For any $\epsilon \in (0,1)$ satisfying $\mathbb{E}[X] \leqslant \mathrm{VaR}_\epsilon(X)$,

$$\begin{aligned}\epsilon \Pi_X(\epsilon) &= \epsilon \inf\{c \in [1, 1/\epsilon] : \mathrm{ES}_{c\epsilon}(X) \leqslant \mathrm{VaR}_\epsilon(X)\} \\ &= \inf\{\epsilon c \in [\epsilon, 1] : \mathrm{ES}_{c\epsilon}(X) \leqslant \mathrm{VaR}_\epsilon(X)\} \\ &= \inf\{k \in [\epsilon, 1] : \mathrm{ES}_k(X) \leqslant \mathrm{VaR}_\epsilon(X)\}.\end{aligned}$$

Let $A(\epsilon) = \{k \in [\epsilon, 1] : \mathrm{ES}_k(X) \leqslant \mathrm{VaR}_\epsilon(X)\}$. For any $k \in A(\epsilon_2)$, we have $1 \geqslant k \geqslant \epsilon_2 > \epsilon_1$ and $\mathrm{ES}_k(X) \leqslant \mathrm{VaR}_{\epsilon_2}(X) \leqslant \mathrm{VaR}_{\epsilon_1}(X)$. Hence, $k \in A(\epsilon_1)$ and this gives $A(\epsilon_2) \subseteq A(\epsilon_1)$. Therefore, $\epsilon_2 \Pi_X(\epsilon_2) = \inf A(\epsilon_2) \geqslant \inf A(\epsilon_1) = \epsilon_1 \Pi_X(\epsilon_1)$. □

Proof of Theorem 2.1. We will check the equivalent condition (2.2) between VaR and ES. Note that if $t \mapsto \mathrm{VaR}_t(X)$ is a constant on $(0, \epsilon)$, then $\Pi_X(\epsilon) = 1$. If $t \mapsto \mathrm{VaR}_t(X)$ is not a constant on $(0, \epsilon)$, then $\Pi_X(\epsilon)$ is the unique solution that satisfies $\mathrm{ES}_{\epsilon \Pi_X(\epsilon)}(X) = \mathrm{VaR}_\epsilon(X)$.

(i) Case 1, $c_2 = 1$. It is clear that $\mathrm{VaR}_t(X)$ is a constant for $t \in (0, c_2\epsilon_2]$ and (2.2) is satisfied. Hence, $\Pi_X(\epsilon_2) = 1$. Moreover, $\mathrm{VaR}_t(X)$ is also a constant for $t \in (0, c_1\epsilon_1]$, which implies $\Pi_X(\epsilon_1) = 1$.

(ii) Case 2, $c_1 = 1$ and $1 < c_2 \leqslant 1/\epsilon_2$. For $t \in (0, \epsilon_1)$, $\mathrm{VaR}_t(X) = G_{\mathbf{z}}(t)$ is a constant for $t \in (0, c_1\epsilon_1)$. Hence, $\Pi_X(\epsilon_1) = 1$. Next, we check whether $\mathrm{ES}_{c_2\epsilon_2}(X) = \mathrm{VaR}_{\epsilon_2}(X)$. The value of $\mathrm{ES}_{c_2\epsilon_2}(X)$ is

$$\begin{aligned}&\mathrm{ES}_{c_2\epsilon_2}(X) \\ &= \frac{1}{c_2\epsilon_2}\left(\int_0^{\epsilon_1} \hat{k}\,\mathrm{d}\epsilon + \int_{\epsilon_1}^{\epsilon_2} (a_1\epsilon + b_1)\,\mathrm{d}\epsilon + \int_{\epsilon_2}^{c_2\epsilon_2} (a_2\epsilon + b_2)\,\mathrm{d}\epsilon\right) \\ &= \frac{1}{c_2\epsilon_2}\left(\epsilon_1\hat{k} + \frac{1}{2}a_1(\epsilon_2^2 - \epsilon_1^2) + b_1(\epsilon_2 - \epsilon_1) + \frac{1}{2}a_2(c_2^2\epsilon_2^2 - \epsilon_2^2) + b_2(c_2\epsilon_2 - \epsilon_2)\right) \\ &= \frac{1}{c_2\epsilon_2}\left(\frac{1}{2}a_1(\epsilon_2 - \epsilon_1)^2 + \hat{k}\epsilon_2 + \frac{1}{2}a_2(c_2\epsilon_2 - \epsilon_2)^2 + \tilde{k}(c_2\epsilon_2 - \epsilon_2)\right) \\ &= \frac{1}{c_2\epsilon_2}\left(\frac{1}{2}(\tilde{k} - \hat{k})(\epsilon_2 - \epsilon_1) + \hat{k}\epsilon_2 + \frac{1}{2}(\tilde{k} - \hat{k})(\epsilon_1 + \epsilon_2) + \tilde{k}(c_2\epsilon_2 - \epsilon_2)\right) = \tilde{k}\end{aligned}$$

The value of $\mathrm{VaR}_{\epsilon_2}(X)$ is $a_2\epsilon_2 + b_2 = \tilde{k}$. Thus, (2.2) is satisfied. As $\mathrm{VaR}_t(X)$ is not a constant for $t \in (0, c_2\epsilon_2)$, we have $\Pi_X(\epsilon_2) = c_2$.

(iii) Case 3, $1 < c_1 \leqslant 1/\epsilon_1$ and $c_2 = \frac{c_1\epsilon_1}{\epsilon_2}$. In this case, we have

$$\mathrm{VaR}_{\epsilon_1}(X) = G_{\mathbf{z}}(\epsilon_1) = k(\epsilon_1) = a\epsilon_2 + b = G_{\mathbf{z}}(\epsilon_2) = \mathrm{VaR}_{\epsilon_2}(X)$$

and $\mathrm{ES}_{c_1\epsilon_1}(X) = \mathrm{ES}_{c_2\epsilon_2}(X)$ as $c_1\epsilon_1 = c_2\epsilon_2$. Thus, we only need to check whether $\mathrm{ES}_{c_2\epsilon_2}(X) = \mathrm{VaR}_{\epsilon_2}(X)$. The value of $\mathrm{ES}_{c_2\epsilon_2}(X)$ is

$$\begin{aligned}
\mathrm{ES}_{c_2\epsilon_2}(X) &= \frac{1}{c_2\epsilon_2}\left(\int_0^{\epsilon_1} k(\epsilon)\,\mathrm{d}\epsilon + \int_{\epsilon_1}^{\epsilon_2} k(\epsilon_1)\,\mathrm{d}\epsilon + \int_{\epsilon_2}^{c_2\epsilon_2} a\epsilon + b\,\mathrm{d}\epsilon\right) \\
&= \frac{1}{c_2\epsilon_2}\left(k + k(\epsilon_1)(\epsilon_2-\epsilon_1) + \frac{1}{2}a(c_2^2\epsilon_2^2 - \epsilon_2^2) + (k(\epsilon_1) - a\epsilon_2)(c_2\epsilon_2 - \epsilon_2)\right) \\
&= \frac{1}{c_2\epsilon_2}\left(k + k(\epsilon_1)(c_2\epsilon_2-\epsilon_1) + \frac{1}{2}a(c_2\epsilon_2 - \epsilon_2)^2\right) \\
&= \frac{1}{c_2\epsilon_2}\left(k + k(\epsilon_1)(c_2\epsilon_2-\epsilon_1) + k(\epsilon_1)\epsilon_1 - k\right) = k(\epsilon_1).
\end{aligned}$$

The value of $\mathrm{VaR}_{\epsilon_2}(X)$ is also $k(\epsilon_1)$. Hence, (2.2) is satisfied and $\Pi_X(\epsilon_1) = c_1$, $\Pi_X(\epsilon_2) = c_2$ because $t \mapsto \mathrm{VaR}_t(X)$ is not a constant on $(0, \epsilon_1)$.

(iv) Case 4, $1 < c_1 \leqslant \epsilon_2/\epsilon_1$ and $1 < c_2 \leqslant 1/\epsilon_2$. The first equivalent condition of (2.2) for $\mathrm{VaR}_{\epsilon_1}(X)$ and $\mathrm{ES}_{c_1\epsilon_1}(X)$ is satisfied because $\mathrm{VaR}_t(X) = k(t)$ is the quantile function for GPD(ξ) with PELVE c_1 and $t \in (0, c_1\epsilon_1)$. Hence, we have $\Pi_X(\epsilon_1) = c_1$. Moreover, $\mathrm{ES}_{c_1\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_1}(X) = k(\epsilon_1)$. We choose $a_1 = k'(c_1\epsilon_1)$ and b_1 such that $a_1c_1\epsilon_1 + b_1 = k(c_1\epsilon_1)$. For the equivalent condition between $\mathrm{ES}_{c_2\epsilon_2}(X)$ and $\mathrm{VaR}_{\epsilon_2}(X)$, we can verify

$$\begin{aligned}
&\mathrm{ES}_{c_2\epsilon_2}(X) = \frac{1}{c_2\epsilon_2}\left(\int_0^{c_1\epsilon_1} k(\epsilon)\,\mathrm{d}\epsilon + \int_{c_1\epsilon_1}^{\epsilon_2} a_1\epsilon + b_1\,\mathrm{d}\epsilon + \int_{\epsilon_2}^{c_2\epsilon_2} (a_2\epsilon + b_2)\,\mathrm{d}\epsilon\right) \\
&= \frac{1}{c_2\epsilon_2}\left(c_1\epsilon_1 k(\epsilon_1) + \frac{1}{2}a_1(\epsilon_2^2 - c_1^2\epsilon_1^2) + b_1(\epsilon_2 - c_1\epsilon_1) + \frac{1}{2}a_2\left(c_2^2\epsilon_2^2 - \epsilon_2^2\right) + b_2\left(c_2\epsilon_2 - \epsilon_2\right)\right) \\
&= \frac{1}{c_2\epsilon_2}\left(c_1\epsilon_1 k(\epsilon_1) + \frac{1}{2}a_1(2c_2\epsilon_2^2 - \epsilon_2^2 - c_1^2\epsilon_1^2) + b_1(c_2\epsilon_2 - c_1\epsilon_1) + \frac{1}{2}a_2\left(c_2\epsilon_2 - \epsilon_2\right)^2\right) \\
&= \frac{1}{c_2\epsilon_2}\left(a_1c_2\epsilon_2^2 + b_1c_2\epsilon_2\right) = a_1\epsilon_2 + b_1 = \mathrm{VaR}_{\epsilon_2}(X).
\end{aligned}$$

Thus, (2.2) is satisfied and we have $\Pi_X(\epsilon_2) = c_2$.

(v) Case 5, $\epsilon_2/\epsilon_1 < c_1 \leqslant 1/\epsilon_1$ and $\frac{c_1\epsilon_1}{\epsilon_2} < c_2 \leqslant 1/\epsilon_2$. The equality between $\mathrm{VaR}_{\epsilon_1}(X)$ and $\mathrm{ES}_{c_1\epsilon_1}(X)$ can be checked by

$$\begin{aligned}
\mathrm{ES}_{c_1\epsilon_1}(X) &= \frac{1}{c_1\epsilon_1}\left(\int_0^{\epsilon_1} k(\epsilon)\,\mathrm{d}\epsilon + \int_{\epsilon_1}^{\epsilon_2} (a_1\epsilon + b_1)\,\mathrm{d}\epsilon + (a_1\epsilon_2 + b_1)(c_1\epsilon_1 - \epsilon_2)\right) \\
&= \frac{1}{c_1\epsilon_1}\left(k + \frac{1}{2}a_1(\epsilon_2^2 - \epsilon_1^2) + (k(\epsilon_1) - a_1\epsilon_1)(\epsilon_2 - \epsilon_1) + (a_1\epsilon_2 + b_1)(c_1\epsilon_1 - \epsilon_2)\right) \\
&= \frac{1}{c_1\epsilon_1}\left(k + a_1(\epsilon_2 - \epsilon_1)(c_1\epsilon_1 - 1/2(\epsilon_2 + \epsilon_1)) + k(\epsilon_1)(c_1\epsilon_1 - \epsilon_1)\right) \\
&= \frac{1}{c_1\epsilon_1}\left(k + k(\epsilon_1)\epsilon_1 - k + k(\epsilon_1)(c_1\epsilon_1 - \epsilon_1)\right) = k(\epsilon_1) = G_{\mathbf{z}}(\epsilon_1) = \mathrm{VaR}_{\epsilon_1}(X).
\end{aligned}$$

The equality between $\mathrm{VaR}_{\epsilon_2}(X)$ and $\mathrm{ES}_{c_2\epsilon_2}(X)$ can be checked by

$$\begin{aligned}
\mathrm{ES}_{c_2\epsilon_2}(X) &= \frac{1}{c_2\epsilon_2}\left(\int_0^{c_1\epsilon_1} k(\epsilon)\,\mathrm{d}\epsilon + \int_{c_1\epsilon_1}^{c_2\epsilon_2} (a_2\epsilon + b_2)\,\mathrm{d}\epsilon\right) \\
&= \frac{1}{c_2\epsilon_2}\left(c_1\epsilon_1 k(\epsilon_1) + \frac{1}{2}a_2(c_2^2\epsilon_2^2 - c_1^2\epsilon_1^2) + b_2(c_2\epsilon_2 - c_1\epsilon_1)\right) \\
&= \frac{1}{c_2\epsilon_2}\left(c_1\epsilon_1 k(\epsilon_1) + \frac{1}{2}a_2(c_2\epsilon_2 - c_1\epsilon_1)^2 + (a_1\epsilon_2 + b_1)(c_2\epsilon_2 - c_1\epsilon_1)\right) \\
&= \frac{1}{c_2\epsilon_2}\left(c_1\epsilon_1 k(\epsilon_1) + c_1\epsilon_1(a_1\epsilon_2 + b_1 - k(\epsilon_1)) + (a_1\epsilon_2 + b_1)(c_2\epsilon_2 - c_1\epsilon_1)\right) \\
&= a_1\epsilon_2 + b_1 = G_{\mathbf{z}}(\epsilon_2) = \mathrm{VaR}_{\epsilon_2}(X)
\end{aligned}$$

Hence, (2.2) is satisfied, and $\Pi_X(\epsilon_1) = c_1$ and $\Pi_X(\epsilon_2) = c_2$.

Therefore, it is checked that X satisfies $\Pi_X(\epsilon_1) = c_1$ and $\Pi_X(\epsilon_2) = c_2$ for all five cases. □

The following propositions address the issue discussed in Remark 2.1 by showing that the boundary cases of $(\epsilon_1, c_1, \epsilon_2, c_2)$ cannot be achieved by strictly decreasing quantile functions, and hence our construction of quantiles with a flat region in Figure 2.5 are needed.

Proposition 2.6. *For any $X \in L^1$, let $\epsilon_1, \epsilon_2 \in (0,1)$ be such that $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X)$ and $\epsilon_1 < \epsilon_2$. Then, $\Pi_X(\epsilon_2) = \max\{1, \Pi_X(\epsilon_1)\epsilon_1/\epsilon_2\}$ if and only if $\mathrm{VaR}_{\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_2}(X)$.*

Proof. Using the same logic as in Lemma 2.1, we have that $\Pi_X(\epsilon_1)$ and $\Pi_X(\epsilon_2)$ are finite.

We first show the "if" statement. Assume $\mathrm{VaR}_{\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_2}(X)$. As $\mathrm{VaR}_{\epsilon}(X)$ is decreasing, we know that $\mathrm{VaR}_{\epsilon}(X)$ is a constant on $[\epsilon_1, \epsilon_2]$.

If $\mathrm{VaR}_{\epsilon}(X) = \mathrm{VaR}_{\epsilon_1}(X)$ for $\epsilon \in (0, \epsilon_1)$, then $\mathrm{VaR}_{\epsilon}(X)$ is a constant on $(0, \epsilon_2]$. Therefore, we can get $\Pi_X(\epsilon_1) = \Pi_X(\epsilon_2) = 1$. Note that $\Pi_X(\epsilon_1)\epsilon_1/\epsilon_2 = \epsilon_1/\epsilon_2 < 1$. Thus, we obtain $\Pi_X(\epsilon_2) = \max\{1, \Pi_X(\epsilon_1)\epsilon_1/\epsilon_2)\}$.

If there exists $\epsilon \in (0, \epsilon_1)$ such that $\mathrm{VaR}_{\epsilon}(X) > \mathrm{VaR}_{\epsilon_1}(X)$, then $\mathrm{ES}_{\epsilon}(X)$ is strictly decreasing on $[\epsilon_1, 1]$. By the equivalent condition between VaR and ES, $\mathrm{VaR}_{\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_2}(X)$ implies $\mathrm{ES}_{\epsilon_1\Pi_X(\epsilon_1)}(X) = \mathrm{ES}_{\epsilon_2\Pi_X(\epsilon_2)}(X)$. Thus, $\epsilon_1\Pi_X(\epsilon_1) = \epsilon_2\Pi_X(\epsilon_2)$. Furthermore, we have

$$\mathrm{VaR}_{\epsilon_1\Pi_X(\epsilon_1)}(X) < \mathrm{ES}_{\epsilon_1\Pi_X(\epsilon_1)}(X) = \mathrm{VaR}_{\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_2}(X).$$

Thus, $\epsilon_1\Pi_X(\epsilon_1) > \epsilon_2$ and we get $\Pi_X(\epsilon_2) = \max\{1, \Pi_X(\epsilon_1)\epsilon_1/\epsilon_2\}$.

Next, we show the "only if" statement. Assume $\Pi_X(\epsilon_2) = \max\{1, \Pi_X(\epsilon_1)\epsilon_1/\epsilon_2\}$.

If $\Pi_X(\epsilon_2) = 1$, then $\mathrm{VaR}_{\epsilon_2}(X) = \mathrm{ES}_{\epsilon_2}(X)$. This implies that $\mathrm{VaR}_{\epsilon}(X)$ is a constant on $(0, \epsilon_2]$, which gives $\mathrm{VaR}_{\epsilon_1}(X) = \mathrm{VaR}_{\epsilon_2}(X)$.

If $\Pi_X(\epsilon_2) = \Pi_X(\epsilon_1)\epsilon_1/\epsilon_2$, then $\epsilon_2\Pi_X(\epsilon_2) = \epsilon_1\Pi_X(\epsilon_1)$. Hence, we have

$$\mathrm{VaR}_{\epsilon_1}(X) = \mathrm{ES}_{\epsilon_1\Pi_X(\epsilon_1)}(X) = \mathrm{ES}_{\epsilon_2\Pi_X(\epsilon_2)}(X) = \mathrm{VaR}_{\epsilon_2}(X).$$

Thus, we complete the proof. □

Proposition 2.7. *For any $X \in L^1$, let $\epsilon_1, \epsilon_2 \in (0,1)$ be such that $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X)$ and $\epsilon_1 < \epsilon_2$. Let $c_1 = \Pi_X(\epsilon_1)$ and $c_2 = \Pi_X(\epsilon_2)$. If $\mathrm{VaR}_{\epsilon_1}(X) > \mathrm{VaR}_{\epsilon_2}(X)$, then*

$$\hat{c} \leqslant c_2 \leqslant \begin{cases} \min\left\{\dfrac{1}{\epsilon_2}, \dfrac{c_1\epsilon_1}{\epsilon_2}\left(\dfrac{\mathrm{VaR}_{\epsilon_1}(X) - \mathrm{VaR}_{c_1\epsilon_1}(X)}{\mathrm{VaR}_{\epsilon_2}(X) - \mathrm{VaR}_{c_1\epsilon_1}(X)}\right)\right\}, & \mathrm{VaR}_{c_1\epsilon_1}(X) < \mathrm{VaR}_{\epsilon_2}(X), \\ \dfrac{1}{\epsilon_2}, & \mathrm{VaR}_{c_1\epsilon_1}(X) \geqslant \mathrm{VaR}_{\epsilon_2}(X), \end{cases}$$

where

$$\hat{c} = \inf\left\{t \in (1, 1/\epsilon_2] : \frac{(t\epsilon_2 - c_1\epsilon_1)\left(\mathrm{VaR}_{\epsilon_2}(X) - \mathrm{VaR}_{t\epsilon_2}(X)\right)}{c_1\epsilon_1\left(\mathrm{VaR}_{\epsilon_1}(X) - \mathrm{VaR}_{\epsilon_2}(X)\right)} \geqslant 1\right\}.$$

Moreover, $\hat{c} \geqslant \max\{1, c_1\epsilon_1/\epsilon_2\}$.

Proof. As $\mathbb{E}[X] \leqslant \mathrm{VaR}_{\epsilon_2}(X)$, we have $c_1 < \infty$ and $c_2 < \infty$. By definition, $c_2 \leqslant 1/\epsilon_2$. From Lemma 2.6, we get $c_2 > \max\{1, c_1\epsilon_1/\epsilon_2\}$. Thus, $c_2 \in (\max\{1, c_1\epsilon_1/\epsilon_2\}, 1/\epsilon_2]$.

Note that $c_1, \epsilon_1, c_2, \epsilon_2$ satisfy the equivalent condition (2.2). We can rewrite (2.2) as

$$\int_0^{c_1\epsilon_1} \mathrm{VaR}_{\epsilon}(X)\,\mathrm{d}\epsilon = c_1\epsilon_1\mathrm{VaR}_{\epsilon_1}(X) \quad \text{and} \quad \int_0^{c_2\epsilon_2} \mathrm{VaR}_{\epsilon}(X)\,\mathrm{d}\epsilon = c_2\epsilon_2\mathrm{VaR}_{\epsilon_2}(X).$$

Therefore, we have

$$\int_{c_1\epsilon_1}^{c_2\epsilon_2} \mathrm{VaR}_{\epsilon}(X)\,\mathrm{d}\epsilon = c_2\epsilon_2\mathrm{VaR}_{\epsilon_2}(X) - c_1\epsilon_1\mathrm{VaR}_{\epsilon_1}(X).$$

Furthermore, by the monotonicity of VaR, we have

$$(c_2\epsilon_2 - c_1\epsilon_1)\mathrm{VaR}_{c_2\epsilon_2}(X) \leqslant \int_{c_1\epsilon_1}^{c_2\epsilon_2} \mathrm{VaR}_{\epsilon}(X)\,\mathrm{d}\epsilon \leqslant (c_2\epsilon_2 - c_1\epsilon_1)\mathrm{VaR}_{c_1\epsilon_1}(X).$$

The two inequality will provide an upper bound and a lower bound for c_2.

An upper bound on c_2. Using $c_2\epsilon_2\mathrm{VaR}_{\epsilon_2}(X)-c_1\epsilon_1\mathrm{VaR}_{\epsilon_1}(X)\leqslant(c_2\epsilon_2-c_1\epsilon_1)\mathrm{VaR}_{c_1\epsilon_1}(X)$, we have

$$c_2\epsilon_2\left(\mathrm{VaR}_{\epsilon_2}(X)-\mathrm{VaR}_{c_1\epsilon_1}(X)\right)\leqslant c_1\epsilon_1\left(\mathrm{VaR}_{\epsilon_1}(X)-\mathrm{VaR}_{c_1\epsilon_1}(X)\right). \tag{2.9}$$

If $\mathrm{VaR}_{c_1\epsilon_1}(X)\geqslant\mathrm{VaR}_{\epsilon_2}(X)$, the left side of (2.9) is less or equal to 0 and the right side of (2.9) is larger or equal to 0 because $\mathrm{VaR}_{\epsilon_1}(X)\geqslant\mathrm{VaR}_{c_1\epsilon_1}(X)$. Therefore, (2.9) is satisfies for any $c_2\in(\max\{1,c_1\epsilon_1/\epsilon_2\},1/\epsilon_2]$. The upper bound for c_2 is unchanged.

On the other hand, if $\mathrm{VaR}_{c_1\epsilon_1}(X)<\mathrm{VaR}_{\epsilon_2}(X)$, we have

$$c_2\leqslant\frac{c_1\epsilon_1}{\epsilon_2}\left(\frac{\mathrm{VaR}_{\epsilon_1}(X)-\mathrm{VaR}_{c_1\epsilon_1}(X)}{\mathrm{VaR}_{\epsilon_2}(X)-\mathrm{VaR}_{c_1\epsilon_1}(X)}\right).$$

Thus, an upper bound for c_2 is $\min\left\{\frac{1}{\epsilon_2},\frac{c_1\epsilon_1}{\epsilon_2}\left(\frac{\mathrm{VaR}_{\epsilon_1}(X)-\mathrm{VaR}_{c_1\epsilon_1}(X)}{\mathrm{VaR}_{\epsilon_2}(X)-\mathrm{VaR}_{c_1\epsilon_1}(X)}\right)\right\}$.

A lower bound on c_2. It holds that

$$(c_2\epsilon_2-c_1\epsilon_1)\mathrm{VaR}_{c_2\epsilon_2}(X)\leqslant c_2\epsilon_2\mathrm{VaR}_{\epsilon_2}(X)-c_1\epsilon_1\mathrm{VaR}_{\epsilon_1}(X).$$

Subtracting $(c_2\epsilon_2-c_1\epsilon_1)\mathrm{VaR}_{\epsilon_2}(X)$ from both sides, we get

$$(c_2\epsilon_2-c_1\epsilon_1)\left(\mathrm{VaR}_{\epsilon_2}(X)-\mathrm{VaR}_{c_2\epsilon_2}(X)\right)\geqslant c_1\epsilon_1\left(\mathrm{VaR}_{\epsilon_1}(X)-\mathrm{VaR}_{\epsilon_2}(X)\right). \tag{2.10}$$

For $t\in(0,1/\epsilon_2)$, let

$$f(t)=(t\epsilon_2-c_1\epsilon_1)\left(\mathrm{VaR}_{\epsilon_2}(X)-\mathrm{VaR}_{t\epsilon_2}(X)\right).$$

As we can see, $f(1)=0$, $f(c_1\epsilon_1/\epsilon_2)=0$ and $f(t)\leqslant 0$ if $t\in[\min\{1,c_1\epsilon_1/\epsilon_2\},\max\{1,c_1\epsilon_1/\epsilon_2\}]$. The f is increasing in the interval $(\max\{1,c_1\epsilon_1/\epsilon_2\},1/\epsilon_2)$, decreasing in $(0,\min\{1,c_1\epsilon_1/\epsilon_2\})$. Hence, by (2.10), the lower bound for c_2 is

$$\hat{c}=\inf\left\{t\in(1,1/\epsilon_2]:\frac{(t\epsilon_2-c_1\epsilon_1)\left(\mathrm{VaR}_{\epsilon_2}(X)-\mathrm{VaR}_{t\epsilon_2}(X)\right)}{c_1\epsilon_1\left(\mathrm{VaR}_{\epsilon_1}(X)-\mathrm{VaR}_{\epsilon_2}(X)\right)}\geqslant 1\right\}.$$

As $c_1\epsilon_1\left(\mathrm{VaR}_{\epsilon_1}(X)-\mathrm{VaR}_{\epsilon_2}(X)\right)>0$, we have $\hat{c}\geqslant\max\{1,c_1\epsilon_1/\epsilon_2\}$. □

2.9 Omitted proofs in Section 2.4.4

Proof of Theorem 2.2. By Proposition 2.3, for any $X\in\mathcal{X}$, we can find $f\in\mathcal{C}$ satisfying (2.3) such that $z_f(y)=1/\pi_X(y)=c$ and $X=f(U)$. As $z(y)=c$ is a continuously differentiable

function, we know that all such f is characterized by the advanced differential equation (2.4). First, we show for any strictly decreasing solution f to (2.4) can be represented as

$$f(y) = C_0 + C_1 y^{\alpha} + O\left(y^{\zeta}\right).$$

Let us start with (2.4). If $z(y) = c$, we need to solve f from

$$f(y) = f(cy) + cyf'(cy), \quad y \in (0, 1].$$

Even though f in the first place is considered on $(0, 1]$, given that $c < 1$, and this final equation, one can expand it to the whole positive line:

$$f(y) = f(cy) + cyf'(cy), \quad y > 0.$$

Next, let $x(t) = e^{-t} f\left(e^{-t}\right)$ for $t \in \mathbb{R}$ and $a = -\log(c) > 0$. This is equivalent to say that $f(y) = x(-\log(y))/y$. This changing variable simply gives the following delayed differential equation:

$$x'(t) = -e^{-a} x(t-a), \quad t \in \mathbb{R}.$$

Since we have assumed that f is strictly decreasing, i.e., $f' < 0$, we have an extra restriction on x. Note that

$$x'(t) = -e^{-t} f\left(e^{-t}\right) - e^{-2t} f'\left(e^{t}\right) = -x(t) - e^{-2t} f'\left(e^{t}\right).$$

Thus, we have $f' < 0 \Leftrightarrow x' + x > 0$. Therefore, we are looking for a solution to the following delay differential equation (DDE):

$$\begin{cases} x'(t) = -e^{-a} x(t-a), \\ x'(t) + x(t) > 0, \end{cases} \quad t \in \mathbb{R}. \tag{2.11}$$

A standard approach of finding the solutions is to assume that they are in the form of a characteristic function $t \mapsto e^{mt}$. Putting this solution inside the equation, we get

$$me^{mt} = -e^{-a} e^{m(t-a)} \quad \Longrightarrow \quad ame^{am} = (-a) e^{(-a)}.$$

This means any solution is given by $x(t) = e^{mt}$ where m solves the characteristic equation

$$l(ma) = l(-a), \tag{2.12}$$

where $l(x) = xe^x$. Let $b = l(-a)$. As $a > 0$, we have $b \in [-1/e, 0)$. This equation has one obvious real solution at $m_1 = -1$. To find m, we need to know about the inverse of l. The inverse of the function l is known as the Lambert W function and plays an essential role in solving delayed and advanced differential equations.

From the Lambert W function, we know that $l(z) = ze^z = b$ has two real solutions when $b \in (-1/e, 0)$ and one real solution when $b = -1/e$. As illustrated by Figure 2.18, if $0 < c < 1/e$, the two real solutions are $z_1 = -a < -1$ and $z_2 = m_2 a > -1$; thus, $-1 < m_2 < 0$. If $0 < c < 1/e$, the two real solutions are $-1 < z_1 = -a < 0$ and $z_2 = m_2 a < -1$; thus, $m_2 < -1$. If $c = 1/e$, there is only one real solution $z_1 = z_2 = -1$; thus $m_2 = m_1 = -1$.

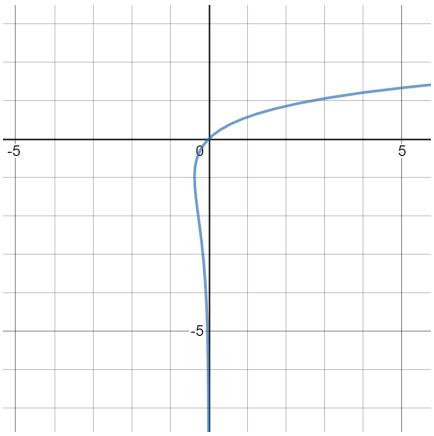


Figure 2.18: Lambert W function.

It is important to note that in general an equation like $l(z) = ze^z = b$ has infinite complex roots. Let $z = \theta + i\eta$, and $b \in [-1/e, 0)$. In that regard, we have

$$\begin{aligned} b = ze^z &= (\theta + i\eta)\, e^{\theta + i\eta} \\ &= (\theta + i\eta)\,(\cos(\eta) + i\sin(\eta)) \\ &= (\theta\cos(\eta) - \eta\sin(\eta)) + i\,(\theta\sin(\eta) + \eta\cos(\eta))\,. \end{aligned}$$

This implies that $\theta\sin(\eta) + \eta\cos(\eta) = 0$, and $b = e^{\theta}(\theta\cos(\eta) - \eta\sin(\eta))$, leading to

$$\eta = 0, b = \theta e^{\theta} \quad \text{or} \quad \theta = -\frac{\eta}{\tan(\eta)}, b = -\frac{\eta \exp\left(-\frac{\eta}{\tan(\eta)}\right)}{\sin(\eta)}.$$

We plot the curves $b = \theta e^{\theta}$ and $\left(-\frac{\eta \exp\left(-\frac{\eta}{\tan(\eta)}\right)}{\sin(\eta)}, -\frac{\eta}{\tan(\eta)}\right)$ to find out the relation between b and the real part of the solution in Figure 2.19. The x-axis is b and the y-axis is θ. The blue curve is associated with $b = \theta e^{\theta}$, which is essentially the principle branch of the Lambert W function. For any b, one can find the real values of the roots by fixing b. For instance, the green dashed line is associated with $b = -0.12$. As one can see, the curves intersect this line in infinite negative values. For $b \in [-1/e, 0)$, we can see that the real roots are greater than the real part of the complex roots. For more explanation of this, see Siewert and Burniston (1973).

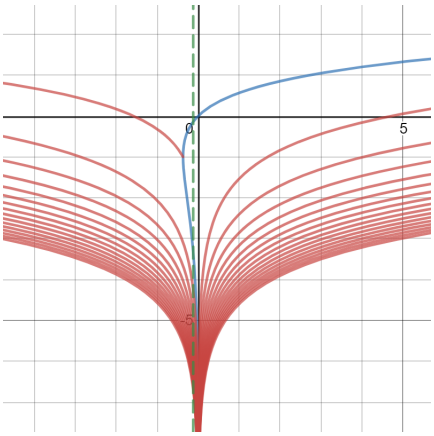


Figure 2.19: The real part of the Lambert W roots.

Now assume that all the complex solutions for $(am)\, e^{am} = (-a)\, e^{(-a)}$ are $m_k = \lambda_k + \sigma_k i$ for $k = 1, 2, 3, \ldots$, where $(\lambda_1, \sigma_1) = (-1, 0)$ and $(\lambda_2, \sigma_2) = (m_2, 0)$. Based on the above discussions, we have

$$\lambda_1 = -1 > \lambda_2 = m_2 > \lambda_3 > \lambda_4 > \ldots \text{ when } c \in (1/e, 1),$$

$$0 > \lambda_2 = m_2 > \lambda_1 = -1 > \lambda_3 > \lambda_4 > \ldots \text{ when } c \in (0, 1/e),$$

and

$$\lambda_1 = \lambda_2 = -1 > \lambda_3 > \lambda_4 > \ldots \text{ when } c = 1/e.$$

Let

$$\mathbb{C} = \left\{ \begin{pmatrix} \lambda & -\sigma \\ \sigma & \lambda \end{pmatrix} = \lambda + \sigma i \mid \lambda, \sigma \in \mathbb{R} \right\}$$

be the set of all complex numbers. Then,

$$x_{\mathbb{C}}(t) = \exp\left\{ \begin{pmatrix} \lambda & -\sigma \\ \sigma & \lambda \end{pmatrix} t \right\}$$

is a complex solution that solves (2.11). It is clear that (2.11) still holds for the linear transform of $x(t)$. Therefore, for any two 2×1 vector A and B,

$$x(t) = A' x_{\mathbb{C}}(t) B = e^{\lambda t} \left(C_1 \cos(\sigma t) + C_2 \sin(\sigma t) \right)$$

is also a solution to (2.11).

In Bellman and Cooke (1963), it is shown that all complex solutions to (2.11) can be represented as follows:

$$x_{\mathbb{C}}(t) = \sum_{k=1}^{\infty} C_k e^{m_k t}.$$

Putting it in the real-valued context, we have that all real-valued solutions are in the following form:

$$x(t) = C_1 e^{-t} + C_2 e^{m_2 t} + \sum_{k=3}^{\infty} e^{\lambda_k t} \left(C_{k,1} \cos(\sigma_k t) + C_{k,2} \sin(\sigma_k t) \right).$$

Now let us check $x' + x > 0$. This means that C_2, $\{C_{k,1}\}_{k \geqslant 3}$ and $\{C_{k,2}\}_{k \geqslant 3}$ for all t must satisfy

$$\begin{aligned} x'(t) + x(t) &= (1 + m_2) C_2 e^{m_2 t} \\ &+ \sum_{k=3}^{\infty} e^{\lambda_k t} \left((\lambda_k C_{k,1} + \sigma_k C_{k,2} + C_{k,1}) \cos(\sigma_k t) + (\lambda_k C_{k,2} - \sigma_k C_{k,1} + C_{k,2}) \sin(\sigma_k t) \right) > 0. \end{aligned}$$

As $\lambda_k < m_2$ for $k \geqslant 3$, we have $\lim_{t \to \infty} \left(x'(t) + x(t) \right) / e^{m_2 t} = (1 + m_2) C_2$. Therefore, we need $C_2 < 0$ if $m_2 < -1$ or $C_2 > 0$ if $m_2 > -1$. That is $C_2(1 + m_2) > 0$. Then the solution can be written as

$$x(t) = C_1 e^{-t} + C_2 e^{m_2 t} + O\left(e^{\lambda_3 t}\right).$$

By a change of variable, we get

$$f(y) = C_1 + C_2 y^{\alpha} + O\left(y^{\zeta}\right),$$

where $\alpha = -(1 + m_2)$ and $\zeta = -(1 + \lambda_3)$. As $\lambda_3 < \min\{-1, m_2\}$ and $C_2(1 + m_2) > 0$, we have $\zeta > \max\{0, \alpha\}$ and $C_2 \alpha < 0$. Also note that since m_2 solves $(am) e^{am} = (-a) e^{(-a)}$, by replacing $m_2 = -1 - \alpha$ and $a = -\log(c)$, we get $(\alpha + 1)^{-1/\alpha} = c$. □

2.10 Omitted proofs in Section 2.5

Proof of Proposition 2.4. Before proving the statements in the proposition, we introduce the following linking function, for $X \in \mathcal{X}$,

$$\Gamma_X(\epsilon) = 1 - F_X(\mathrm{ES}_\epsilon(X)), \qquad \epsilon \in [0,1].$$

As $X \in X$ it is easy to check that Γ_X satisfies Assumption 2.1 for any $X \in \mathcal{X}$. The domain of $\Gamma_X(\epsilon)$ is $[0,1]$ and its range is $[0, 1 - F_X(\mathbb{E}[X])]$.

As $\mathrm{ES}_\epsilon(X) = \mathrm{VaR}_{\Gamma_X(\epsilon)}(X)$, we have

$$\mathrm{VaR}_{\Gamma_X(\epsilon)}(X) = \mathrm{ES}_\epsilon(X) = \mathrm{VaR}_{\epsilon/\pi_X(\epsilon)}(X) \quad \text{for} \quad \epsilon \in (0,1].$$

Hence we have the simple relationship $\Gamma_X(\epsilon) = \epsilon/\pi_X(\epsilon)$. Therefore, $\Pi_X(\epsilon) = \pi_X\left(\Gamma_X^{-1}(\epsilon)\right)$ and $\pi_X(\epsilon) = \Pi_X(\Gamma_X(\epsilon))$. The function Γ_X yields an association between a point on the PELVE on $(0, 1 - F_X(\mathbb{E}[X])]$ and a point on the dual PELVE curve on $(0,1]$ with the same value. Furthermore, we have π_X is continuous on $(0,1)$ as $\pi_X(\epsilon) = \epsilon/\Gamma_X(\epsilon)$ and Γ_X is continuous.

Next, we show the statements (i)-(iv). The equivalence (i) of monotonicity of Π_X and that of $\pi_X(\cdot)$ follows from $\Pi_X(\epsilon) = \pi_X\left(\Gamma_X^{-1}(\epsilon)\right)$, $\pi_X(\epsilon) = \Pi_X(\Gamma_X(\epsilon))$ and that Γ_X is increasing.

For (ii), (iii) and (iv), we first show that Γ_X is location-scale invariant and shape relevant (in the sense of (2.13)). Assume that $f : \mathbb{R} \to \mathbb{R}$ is a strictly increasing concave function such that $f(X) \in \mathcal{X}$. By Jensen's inequality and the dual representation of ES_p, we have

$$\mathrm{ES}_p\left(f(X)\right) \leqslant f\left(\mathrm{ES}_p(X)\right)$$

for all $p \in (0,1)$. This statement can be found in Appendices A.2 in Li and Wang (2022). Therefore,

$$\begin{aligned} \Gamma_{f(X)}(\epsilon) &= 1 - F_{f(X)}\left(\mathrm{ES}_\epsilon\left(f(X)\right)\right) \\ &\geqslant 1 - F_{f(X)}\left(f\left(\mathrm{ES}_\epsilon(X)\right)\right) = 1 - F_X\left(\mathrm{ES}_\epsilon(X)\right) = \Gamma_X(\epsilon). \end{aligned} \tag{2.13}$$

Then, we have $\Gamma_{f(X)}(\epsilon) \geqslant \Gamma_X(\epsilon)$ for all strictly increasing concave functions: $f : \mathbb{R} \to \mathbb{R}$ with $f(X) \in \mathcal{X}$.

For any strictly increasing convex function $g : \mathbb{R} \to \mathbb{R}$ with $g(X) \in \mathcal{X}$, we can take $f(x) = g^{-1}(X)$, which is a strictly increasing concave function. Therefore, we have $\Gamma_{g(X)}(\epsilon) \leqslant \Gamma_X(\epsilon)$ for all strictly increasing convex functions g.

For $\lambda > 0$ and $a \in \mathbb{R}$, we have that $f(x) = \lambda X + a$ is both convex and concave. Therefore, $\Gamma_{\lambda X+a}(\epsilon) = \Gamma_X(\epsilon)$ for all $\epsilon \in [0, 1]$. In conclusion, we have the following results for Γ.

(1) For all $\lambda > 0$ and $a \in \mathbb{R}$, $\Gamma_{\lambda X+a}(\epsilon) = \Gamma_X(\epsilon)$.

(2) $\Gamma_{f(X)}(\epsilon) \geqslant \Gamma_X(\epsilon)$ for all strictly increasing concave functions: $f : \mathbb{R} \to \mathbb{R}$ with $f(X) \in \mathcal{X}$.

(3) $\Gamma_{g(X)}(\epsilon) \leqslant \Gamma_X(\epsilon)$ for all strictly increasing convex functions: $g : \mathbb{R} \to \mathbb{R}$ with $g(X) \in \mathcal{X}$.

Then, we have (ii), (iii) and (iv) from $\pi_X(\epsilon) = \epsilon/\Gamma_X(\epsilon)$. □

Proof of Theorem 2.3. The idea is to prove that if $1/\eta$ is convex (concave), then $x \mapsto F^{-1}((1-p)F(x)+p)$ is convex (concave) for all $p \in (0, 1)$. Then, we can get the desired result by Proposition 2.5. We will use the following steps to show this statement.

Step 1. Let $s(x) = \log(1 - F(x))$ for $x \in (\text{ess-inf}(X), \text{ess-sup}(X))$. Then, s is a continuous and strictly decreasing function and $s(x) < 0$. Let s^{-1} be the inverse function of s. Now, we have

$$F(x) = 1 - e^{s(x)}, \quad x \in (\text{ess-inf}(X), \text{ess-sup}(X))$$

and

$$F^{-1}(t) = s^{-1}(\log(1-t)), \quad t \in (0, 1).$$

Therefore,

$$\begin{aligned} F^{-1}\big((1-p)F(x)+p\big) &= s^{-1}\big(\log(1-(1-p)F(x)-p)\big) \\ &= s^{-1}\Big(\log\big(1-F(x)\big)+\log(1-p)\Big) \\ &= s^{-1}\Big(\log\big(e^{s(x)}\big)+\log(1-p)\Big) \\ &= s^{-1}\Big(s(x)+\log(1-p)\Big). \end{aligned}$$

Let $\theta = \log(1-p)$. It follows that the statement that $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is convex (concave) for all $p \in (0, 1)$ is equivalent to the statement that $x \mapsto s^{-1}(s(x)+\theta)$ is convex (concave) for all $\theta < 0$.

Step 2. Let $g(x) := -s^{-1}(x)$. Then, g is strictly increasing. We will show that if $1/\eta$ is convex (concave), $\log(g'(x))$ is convex (concave).

As $g(x) = -s^{-1}(x) = -S^{-1}(e^x)$, we have

$$g'(x) = \frac{e^x}{f\left(S^{-1}(e^x)\right)}.$$

Let $H(x) := \log(g'(x)) = x - \log\left(f\left(S^{-1}(e^x)\right)\right)$. We have

$$H\left(\log(S(x))\right) = \log S(x) - \log f(x) = -\log \eta(x). \tag{2.14}$$

Then, taking the derivative on both sides of (2.14), we get

$$\begin{aligned} &- H'\big(\log(S(x))\big)\eta(x) = -\frac{\eta'(x)}{\eta(x)} \\ \iff \quad &H'\big(\log(S(x))\big) = \frac{\eta'(x)}{\eta^2(x)} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\eta(x)}\right). \end{aligned}$$

Taking a derivative in both sides again, we get

$$-H''\big(\log(S(x))\big)\eta(x) = -\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{1}{\eta(x)}\right).$$

Then, $1/\eta$ is a convex (concave) function means $H''(x) \geqslant 0$ ($H''(x) \leqslant 0$), which gives that $\log g'(x)$ is convex (concave).

Step 3. For $\theta < 0$, let $G_\theta(x) := s^{-1}(s(x) + \theta)$. We are going to show

$$\lim_{z \to 0} \frac{G_\theta(x+z) - G_\theta(x)}{z} \leqslant \lim_{z \to 0} \frac{G_\theta(x'+z) - G_\theta(x')}{z} \tag{2.15}$$

for all $x < x'$.

We take $z > 0$ first. As s is strictly decreasing, s^{-1} is also strictly decreasing. Then, G_θ is a continuous and strictly increasing function. As $\theta < 0$, we also have $G_\theta(x) > x$. Take arbitrary x, x', y and z such that $x < x'$, $x < y$ and $z > 0$. Let $\theta = s(y) - s(x)$. Then, we have $G_\theta(x) = y$. Define

$$h = G_\theta(x+z) - y, \quad y' = G_\theta(x') \quad \text{and} \quad h' = G_\theta(x'+z) - y'.$$

By the definition of G_θ, we have $s(y+h) = s(x+z) + \theta$, $s(y') = s(x') + \theta$ and $s(y'+h') = s(x'+z) + \theta$. As a result, we have

$$s(y+h) - s(y) = s(x+z) - s(x) \quad \text{and} \quad s(y'+h') - s(y') = s(x'+z) - s(x').$$

By the mean-value theorem, there exists $\zeta \in (y, y+h)$, $\zeta' \in (y', y'+h')$, $\xi \in (x, x+z)$ and $\xi' \in (x', x'+z)$ such that

$$s'(\zeta)h = s'(\xi)z \text{ and } s'(\zeta')h' = s'(\xi')z.$$

Furthermore, x, x', y and y' satisfy $x < x' < y'$ and $x < y < y'$. If z is small enough, then h and h' will also be small enough as G_θ is continuous. Therefore, we have $\xi < \xi' < \zeta'$ and $\xi < \zeta < \zeta'$ when z is small enough.

In Step 2, we have that $\log g'(x)$ is convex when $1/\eta$ is convex. Therefore, we get

$$\log(g'(a)) + \log(g'(b)) \geqslant \log(g'(a')) + \log(g'(b')),$$

for all $a < a' < b$ and $a < b' < b$, which means

$$g'(a)g'(b) \geqslant g'(a')g'(b').$$

As $g(x) = -s^{-1}(x)$, we have

$$\frac{1}{s'(s^{-1}(a))s'(s^{-1}(b))} \geqslant \frac{1}{s'(s^{-1}(a'))s'(s^{-1}(b'))}.$$

As $s^{-1}(x)$ is strictly decreasing, it means that

$$s'(\alpha)s'(\beta) \leqslant s'(\alpha')s'(\beta')$$

for $\alpha > \alpha' > \beta$ and $\alpha > \beta' > \beta$. Therefore, we have $s'(\xi)s'(\zeta') \leqslant s'(\xi')s'(\zeta)$ as $\zeta' > \zeta > \xi$ and $\zeta' > \xi' > \xi$. That is,

$$h = \frac{s'(\xi)z}{s'(\zeta)} \leqslant \frac{s'(\xi')z}{s'(\zeta')} = h'.$$

On the other hand, $h = G_\theta(x+z) - G_\theta(x)$ and $h' = G_\theta(x'+z) - G_\theta(x')$. Therefore,

$$G_\theta(x+z) - G_\theta(x) \leqslant G_\theta(x'+z) - G_\theta(x')$$

when z is small enough. If $z < 0$, we can also get (2.15) by an analogous argument.

Hence, the second-order derivative of G_θ is increasing for each $\theta < 0$, which means that $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is convex for all $p \in (0,1)$ if $1/\eta$ is convex.

An analogous argument yields that $x \mapsto F^{-1}\big((1-p)F(x)+p\big)$ is concave for all $p \in (0,1)$ when $1/\eta$ is concave. □

Chapter 3

Joint mixability and notions of negative dependence

3.1 Introduction

Dependence among multiple sources of randomness has always been an active topic in operations research, statistics, transport theory, economics, and finance; see Denuit et al. (2005), Joe (2014), Rüschendorf (2013), McNeil et al. (2015) and Galichon (2016) for standard textbook treatment in different fields, and the recent work Blanchet et al. (2020) for relevant examples in operations research. In contrast to positive, which received much attention in the literature, considerably fewer studies are found on negative dependence, partially due to its more complicated mathematical nature. For a review and historical account on extremal positive and negative dependence concepts, we refer to Puccetti and Wang (2015).

In the past decade, the notion of joint mixability proposed by Wang et al. (2013), which generalizes complete mixability (Wang and Wang, 2011), has been shown useful for solving many optimization problems involving the dependence of multiple risks. In particular, joint mixability is essential to worst-case bounds on Value-at-Risk and other risk measures under dependence uncertainty (Puccetti and Rüschendorf, 2013; Embrechts et al., 2013; Bernard et al., 2014), as well as bottleneck assignment and scheduling problems (Coffman

and Yannakakis, 1984; Hsu, 1984; Haus, 2015; Bernard et al., 2018).

Joint mixability concerns, for given marginal distributions, the existence of a random vector which has a constant component-wise sum. Such a random vector is called a *joint mix* supported by the given marginal distributions, and it represents a very simple concept of dependence. A joint mix is commonly regarded as a notion of extremal negative dependence; see the review of Puccetti and Wang (2015). The reason why a joint mix represents negative dependence is that it minimizes many objectives which are maximized by comonotonicity. For instance, for fixed marginal distributions of the risks, comonotonicity maximizes the variance, the stop-loss premium, and the Expected Shortfall (ES) of the sum of the risks, whereas a joint mix, if it exists, minimizes these quantities; see e.g., Rüschendorf (2013). As such, joint mixability is seen as the safest dependence structure, as long as risk aggregation is concerned (Embrechts et al., 2014).

Although a joint mix has been treated as a concept of negative dependence, it remains unclear whether it is consistent with classic notions of negative dependence in statistics. Popular notions of negative dependence include negative correlation dependence (NCD), negative orthant dependence (NOD; Block et al., 1982; Lehmann, 1966) and negative association (NA; Alam and Saxena, 1981; Joag-Dev and Proschan, 1983). The connection between joint mixes and these negative dependence concepts is the main object that we address in this chapter. We obtain some necessary and sufficient conditions for a joint mix to be NOD or NA in Section 3.3. Some characterization results are obtained in Section 3.5 for the class of elliptical distributions. In particular, among all elliptical classes, only the Gaussian family supports NOD joint mixes of any dimension.

Since a joint mix may be either negatively dependent or not, a natural question is whether there are special features of negatively dependent joint mixes which are useful in applications. For this question, we consider a multi-marginal optimal transport problem under uncertainty on the set of components. A few optimality results on negatively dependent joint mixes are obtained, and they demonstrate an interesting interplay between joint mixes and negative dependence. In particular, for the special case of quadratic cost, we show that the optimizer has to be an NCD JM in some settings. This is the topic of Section 3.4.

The study of joint mixability was originally motivated by questions in risk manage-

ment and operations research, and it has a strong connection to the theory of multi-marginal optimal transport (Santambrogio, 2015; Pass, 2015) and variance reduction in random sampling (Craiu and Meng, 2001, 2005); see also our Section 3.4. Recently, there is a growing spectrum of applications of joint mixability outside the above fields, including multiple hypothesis testing (Vovk et al., 2022), wireless communications (Besser and Jorswieck, 2020), labor market matching (Boerma et al., 2021), and resource allocation games (Perchet et al., 2022). Results in this chapter connect the two topics of joint mixability and negative dependence, allowing us to bring tools from one area to the other.

This chapter is organized as follows. Section 3.2 introduces the concepts of negative dependence and joint mixability, and summarizes their basic interrelationships. Section 3.3 explores conditions for joint mixes to be negatively dependent. Two results on decompositions of joint mixes into negatively dependent ones are also obtained. Section 3.4 studies a multi-marginal optimal transport problem as an application of negatively dependent joint mixes. Section 3.5 studies joint mixes within the elliptical family, and we obtain a new characterization of the Gaussian family as the only one supporting a negatively dependent elliptical distribution for every dimension. Section 3.6 concludes this chapter with some open questions and potential directions for future research. All the proofs are deferred to Section 3.7.

3.2 Notions of negative dependence

In this section we recall a few classic notions of negative dependence. Throughout, denote by $[n] = \{1, \dots, n\}$ and $\mathbf{1}_n$ the n-vector with all components being 1; the vector $\mathbf{0}_n$ is defined analogously. All inequalities and equalities between (random) vectors are component-wise. For an n-dimensional random vector $\mathbf{X} = (X_1, \dots, X_n)$, denote by $\mathbf{X}^\perp = (X_1^\perp, \dots, X_n^\perp)$ a random vector with independent components such that $X_i \overset{\mathrm{d}}{=} X_i^\perp$, $i \in [n]$, where $\overset{\mathrm{d}}{=}$ stands for equality in distribution. For a set $A \subseteq [n]$, we denote by $\mathbf{X}_A = (X_k)_{k\in A}$. A function $\psi : \mathbb{R}^n \to \mathbb{R}$ is called *supermodular* if $\psi(\mathbf{x}\wedge\mathbf{y})+\psi(\mathbf{x}\vee\mathbf{y}) \geqslant \psi(\mathbf{x})+\psi(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\mathbf{x} \wedge \mathbf{y}$ and $\mathbf{x} \vee \mathbf{y}$ are the component-wise minimum and maximum of $\mathbf{x}$ and $\mathbf{y}$, respectively.

Definition 3.1. Let $\mathbf{X} = (X_1, \dots, X_n)$ be an n-dimensional random vector.

(i) $\mathbf{X}$ is *negative correlation dependent (NCD)* if $\mathrm{cov}(X_i, X_j) \leqslant 0$ for all $i, j \in [n]$ with $i \neq j$.

(ii) $\mathbf{X}$ is *negative upper orthant dependent (NUOD)* if $\mathbb{P}(\mathbf{X} > \mathbf{t}) \leqslant \mathbb{P}(\mathbf{X}^{\perp} > \mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^n$; $\mathbf{X}$ is *negative lower orthant dependent (NLOD)* if $\mathbb{P}(\mathbf{X} \leqslant \mathbf{t}) \leqslant \mathbb{P}(\mathbf{X}^{\perp} \leqslant \mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^n$. If $\mathbf{X}$ is both NLOD and NUOD, then it is *negative orthant dependent (NOD)*.

(iii) $\mathbf{X}$ is *negative supermodular dependent (NSD)* if $\mathbb{E}[\psi(\mathbf{X})] \leqslant \mathbb{E}[\psi(\mathbf{X}^{\perp})]$ for all supermodular functions $\psi : \mathbb{R}^n \to \mathbb{R}$ such that the expectations exist.

(iv) $\mathbf{X}$ is *negatively associated (NA)* if

$$\mathrm{Cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \leqslant 0, \tag{3.1}$$

for any disjoint subsets $A, B \subseteq [n]$ and any real-valued, coordinate-wise increasing functions f and g such that $f(\mathbf{X}_A)$ and $g(\mathbf{X}_B)$ have finite second moments.

(v) $\mathbf{X}$ is *counter-monotonic (CT)* if each pair of its component (X_i, X_j) for $i \neq j$ satisfies $(X_i, X_j) = (f(Z), -g(Z))$ almost surely (a.s.) for some random variable Z and increasing functions f, g.

(vi) $\mathbf{X}$ is a *joint mix* (abbreviated as "$\mathbf{X}$ is JM") if $\sum_{i=1}^n X_i = c$ a.s. for some constant $c \in \mathbb{R}$.

All abbreviations introduced in this section are also used as nouns to represent the corresponding dependence concept. The next definition concerns properties of the marginal distributions that allow for JM random vectors.

Definition 3.2. An n-tuple $(F_1, \dots, F_n)$ of distributions on $\mathbb{R}$ is called *jointly mixable* if there exists a joint mix $\mathbf{X} = (X_1, \dots, X_n)$ such that $X_i \sim F_i$, $i \in [n]$. The constant $c = \sum_{i=1}^n X_i$ is called a *center* of $\mathbf{X}$. In this case, we also say that $(F_1, \dots, F_n)$ *supports a joint mix* $\mathbf{X}$. A distribution F is called *n-completely mixable* if the n-tuple $(F, \dots, F)$ is jointly mixable.

The following implications hold between the above concepts of negative dependence. These implications are either checked directly by definition or shown in the literature, e.g.,

Joag-Dev and Proschan (1983), Christofides and Vaggelatou (2004) and Chapter 7. The case of $n \geqslant 3$ is different from the case $n = 2$.

$$n = 2: \quad \text{JM} \Longrightarrow \text{CT} \Longrightarrow \text{NA} \Longleftrightarrow \text{NSD} \Longleftrightarrow \text{NOD} \Longleftrightarrow \text{NUOD} \Longleftrightarrow \text{NLOD} \Longrightarrow \text{NCD}; \tag{3.2}$$

$$\text{general } n: \quad \text{CT} \Longrightarrow \text{NA} \Longrightarrow \text{NSD} \Longrightarrow \text{NOD} \Longrightarrow \text{NUOD or NLOD} \Longrightarrow \text{NCD}. \tag{3.3}$$

All one-direction implications in (3.2) and (3.3) are strict for $n \geqslant 3$; see Amini et al. (2013) for some examples. In contrast to the case $n = 2$ in (3.2), JM no longer implies any of the properties in (3.3). This can be observed by the following properties of Gaussian random vectors.

Proposition 3.1. *Let $\mathbf{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \Sigma)$ be a Gaussian random vector with mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\Sigma = (\sigma_{ij})_{n\times n}$.*

(a) The followings are equivalent: (i) $\mathbf{X}$ is NA; (ii) $\mathbf{X}$ is NSD; (iii) $\mathbf{X}$ is NUOD; (iv) $\mathbf{X}$ is NLOD; (v) $\mathbf{X}$ is NCD.

(b) The followings are equivalent: (i) $\mathbf{X}$ is JM; (ii) $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$.

(c) For $n = 2$, the followings are equivalent: (i) $\mathbf{X}$ is JM; (ii) $\mathbf{X}$ is CT and $\sigma_{11} = -\sigma_{12}$.

(d) For $n \geqslant 3$, $\mathbf{X}$ is never CT unless at least $n - 2$ components of $\mathbf{X}$ are degenerate.

Proposition 3.1 shows the convenient property of the Gaussian distribution that the concepts of NA, NSD, NOD, NLOD, NUOD and NCD are all equivalent for this class. Parts (a) and (b) immediately tell that, for $n \geqslant 3$, JM does not imply any of these concepts, and none of these concepts implies JM. We will focus mostly on NA, NOD and NCD given their popularity and relative strength in the chains (3.2) and (3.3). For some other notions of negative dependence, see Joe (2014).

3.3 JM and negative dependence

In this section, we explore the relation between JM and negative dependence concepts introduced in Section 3.2 by means of several theoretical results. We first show that a joint

mix is NA under some properties of conditional independence and monotonicity.

Theorem 3.1. *Let* $\mathbf{X}$ *be a joint mix and write* $S_A = \sum_{i\in A} X_i$ *for* $A \subseteq [n]$. *Suppose that*

(a) $\mathbf{X}_A$ *and* $\mathbf{X}_{[n]\setminus A}$ *are independent conditionally on* S_A *for every* $A \subseteq [n]$;

(b) $\mathbb{E}\left[f(\mathbf{X}_A)|S_A\right]$ *is increasing in* S_A *for every increasing function* f *and* $A \subseteq [n]$.

Then $\mathbf{X}$ *is NA.*

Theorem 3.1 can be compared with Theorem 2.6 of Joag-Dev and Proschan (1983), which says that if $\mathbf{X}$ is independent and satisfies (b), then the conditional distribution of $\mathbf{X}$ given $S_{[n]}$ is NA. Since $S_{[n]}$ is a constant for JM and (a) is implied by independence, Theorem 3.1 means that the NA condition in Theorem 2.6 of Joag-Dev and Proschan (1983) holds if the independence assumption is weakened to conditional independence (a), and in addition we assume JM. Note, however, that JM and independence of $\mathbf{X}$ conflict each other unless $\mathbf{X}$ is degenerate.

Most existing examples of NA random vectors are presented by Joag-Dev and Proschan (1983). Although Theorem 3.1 does not directly give new examples of NA random vectors, it can be used to check NA in popular examples.

Example 3.1. We use Theorem 3.1 to check that the uniform distribution on the standard simplex $\Delta_n = \{(x_1, \dots, x_n) \in [0,1]^n : \sum_{i=1}^n x_i = 1\}$ is NA. Let $\mathbf{X}$ follow the uniform distribution over Δ_n which is JM. For every $A \subseteq [n]$, we can check that $(\mathbf{X}_A, \mathbf{X}_{[n]\setminus A})|\{S_A = s\}$ for $s \in (0,1)$ follows a uniform distribution on $(s\Delta_n) \times ((1-s)\Delta_n)$ and condition (a) holds. Condition (b) follows by noting that $\mathbf{X}_A|\{S_A = s\} \overset{\mathrm{d}}{=} s\mathbf{X}_A|\{S_A = 1\}$ for $s \in (0,1)$ and thus $\mathbf{X}_A|\{S_A = s\}$ is stochastically increasing in s.

Example 3.2. The multinomial distribution is known to be NA (Joag-Dev and Proschan, 1983). We show this by virtue of Theorem 3.1. Let $\mathbf{X} \sim \mathrm{MN}_n(k, \mathbf{p})$ follow a multinomial distribution with k trails, n mutually exclusive events and event probabilities $\mathbf{p} = (p_1, \dots, p_n)$. For $s \in \{0, \dots, k\}$ and every $A \subseteq [n]$ with $B = [n] \setminus A$, it holds that $\mathbf{X}_A|\{S_A = s\} \sim \mathrm{MN}_{|A|}(s, \mathbf{p}_A/\sum_{i\in A} p_i)$ and $\mathbf{X}_B|\{S_A = s\} \sim \mathrm{MN}_{|B|}(k-s, \mathbf{p}_B/\sum_{i\in B} p_i)$, where $p_A = (p_i)_{i\in A}$ and $p_B = (p_i)_{i\in B}$. Then conditions (a) and (b) can be checked directly by calculation.

We next focus on exchangeable joint mixes, which exhibit some specific forms of negative dependence. A random vector $\mathbf{X} = (X_1, \dots, X_n)$ (or its distribution) is called *exchangeable* if $\mathbf{X} \overset{\mathrm{d}}{=} \mathbf{X}^\pi$ for all $\pi \in \mathfrak{S}_n$, where $\mathfrak{S}_n$ is the set of all permutations on $[n]$ and $\mathbf{X}^\pi = (X_{\pi(1)}, \dots, X_{\pi(n)})$. First, we note that if $\mathbf{X}$ is CT with identical marginal distributions equal to F, then the distribution of $\mathbf{X}$ is explicitly given by $\mathbb{P}(\mathbf{X} \leqslant \mathbf{x}) = (F(x_1)+\cdots+F(x_n)-n+1)_+$ for $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$; see, for example, Theorem 3.3 of Puccetti and Wang (2015). Clearly, this distribution is exchangeable. Moreover, for any given marginal distribution F which is n-completely mixable, there exists an exchangeable joint mix with marginals F; see Proposition 2.1 of Puccetti et al. (2019). Note that an exchangeable joint mix is NCD because each bivariate correlation coefficient is equal to $-1/(n-1)$. The next proposition states that such an exchangeable joint mix is also negatively dependent in the sense of NSD, NUOD and NULD if so is $\mathbf{X}$.

Proposition 3.2. *If a univariate distribution function F supports an NSD n-joint mix, then F supports an exchangeable NSD n-joint mix. The statement holds true if NSD is replaced by NOD, NUOD or NLOD.*

One may wonder whether Proposition 3.2 holds with NSD replaced by NA. Unfortunately, this question remains open, as our proof for Proposition 3.2 does not extend to NA.

Next, we present a necessary condition for a tuple of distributions to support any negatively dependent joint mixes. This condition is also sufficient when the marginal distributions are Gaussian.

Proposition 3.3. *If a tuple of distributions $(F_1, \dots, F_n)$ with finite variance vector $(\sigma_1^2, \dots, \sigma_n^2)$ supports an NCD joint mix, then*

$$2 \max_{i \in [n]} \sigma_i^2 \leqslant \sum_{i \in [n]} \sigma_i^2. \tag{3.4}$$

Since NCD is weaker than NOD and NA, the necessary condition (3.4) is also necessary for NOD and NA joint mixes.

For a given tuple of distributions $(F_1, \dots, F_n)$ with finite variance vector $(\sigma_1^2, \dots, \sigma_n^2)$, the condition (3.4) is not necessary for the existence of an NCD random vector, since an

independent random vector supported by $(F_1, \dots, F_n)$ always exists and it is NCD. More interestingly, the condition (3.4) is not necessary for the existence of a joint mix either. Indeed, as shown by Wang and Wang (2016, Corollary 2.2), a useful necessary condition for a joint mix supported by $(F_1, \dots, F_n)$ to exist is

$$2 \max_{i \in [n]} \sigma_i \leqslant \sum_{i \in [n]} \sigma_i. \tag{3.5}$$

We note that (3.4) is strictly stronger than (3.5), because for any $j \in [n]$, (3.4) gives

$$\sigma_j^2 \leqslant \sum_{i \in [n] \setminus \{j\}} \sigma_i^2 \implies \sigma_j \leqslant \left(\sum_{i \in [n] \setminus \{j\}} \sigma_i^2 \right)^{1/2} \leqslant \sum_{i \in [n] \setminus \{j\}} \sigma_i,$$

which implies (3.5). It is clear that (3.4) and (3.5) are not equivalent; for example, $(\sigma_1, \sigma_2, \sigma_3) = (2, 2, 3)$ satisfies (3.5) but not (3.4).

By Proposition 2.4 of Wang et al. (2013), if the marginal distributions $F_1, \dots, F_n$ are Gaussian, then the condition (3.5) is necessary and sufficient for a joint mix supported by $(F_1, \dots, F_n)$ to exist. Hence, the condition (3.4), which is strictly stronger than (3.5), is not necessary for a joint mix to exist. On the other hand, (3.4) is generally not sufficient for an NCD joint mix to exist either, since it is well known that a joint mix may not exist even if the marginal distributions are identical. Nevertheless, it turns out that (3.4) is necessary and sufficient for an NCD or NA joint mix to exist for Gaussian marginals.

Theorem 3.2. *A tuple of univariate Gaussian distributions with variance vector $(\sigma_1^2, \dots, \sigma_n^2)$ supports an NCD or NA joint mix if and only if (3.4) holds, that is, $2 \max_{i \in [n]} \sigma_i^2 \leqslant \sum_{i \in [n]} \sigma_i^2$. Moreover, such an NCD or NA joint mix can be chosen as a Gaussian random vector.*

The negative dependence concepts of NA, NSD, NOD, NLOD, NUOD and NCD are all equivalent for multivariate Gaussian distributions, as we see in Proposition 3.1. Hence, (3.4) is also necessary and sufficient for an NSD, NOD, NLOD, or NUOD joint mix to exist for Gaussian marginals. For elliptical distributions (details in Section 3.5), the corresponding statement to Theorem 3.2 holds for NCD but not the other forms of negative dependence; see Proposition 3.7.

Example 3.3. In case $n = 3$, for any marginals with variance vector $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, the covariance of a joint mix $\mathbf{X}$ is uniquely given by

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) & \sigma_3^2 \end{pmatrix};$$

see Xiao and Yao (2020, Corollary 6) for this statement. If $\mathbf{X}$ is Gaussian, it is clear that $\mathbf{X}$ is NA if and only if (3.4) holds. In case $n \geqslant 4$, for Gaussian marginals we can obtain an explicit covariance matrix of an NA joint mix from the proof of Theorem 3.2.

We end this section with two decomposition results of a joint mix into NA joint mixes, one through a random vector decomposition, and one through a mixture decomposition. We first establish a new result showing that any finitely supported discrete joint mix can be decomposed into a linear combination of binary multinomial random vectors. A binary multinomial random vector is a random vector $(X_1, \dots, X_n)$ taking values in $\{0, 1\}^n$ such that $\sum_{i=1}^n X_i = 1$; that is, exactly one of $X_1, \dots, X_n$ takes the value 1. By definition, binary multinomial random vectors are CT (hence NA) and JM.

Theorem 3.3. *Suppose that the random vector* $\mathbf{X}$ *takes values in a finite set. Then,* $\mathbf{X}$ *is JM if and only if it can be represented as a finite linear combination of binary multinomial random vectors.*

Theorem 3.3 generalizes Theorem 2 of Wang (2015) which has a decomposition of a joint mix taking nonnegative integer values as the sum of binary multinomial random vectors. The assumption of finite support in Theorem 3.3 does not seem to be dispensable with the current proof techniques.

Next, using the fact that the distribution of a joint mix can be written as a mixture of discrete uniform (DU) distributions on n points in $\mathbb{R}^n$, we obtain the following decomposition.

Proposition 3.4. *The distribution of any exchangeable joint mix with center* μ *can be written as a mixture of distributions of exchangeable NA joint mixes with center* μ*.*

3.4 A multi-marginal optimal transport problem

To connect JM and negative dependence, a natural question is whether in some applications negatively dependent joint mixes have additional attractive properties that are not shared by other joint mixes. We present an optimal transport problem with uncertainty in this section where a combination of negative dependence and JM naturally appears.

In the multi-marginal optimal transport theory (Santambrogio, 2015; Pass, 2015), a general objective is

$$\text{to minimize } \mathbb{E}[c(X_1,\dots,X_n)] \quad \text{subject to } X_i \sim F_i,\ i\in[n],$$

where $c:\mathbb{R}^n\to\mathbb{R}$ is a cost function and $F_1,\dots,F_n$ are specified marginal distributions. This problem is referred to as the *Monge-Kantorovich problem*. In the context of this chapter, the distributions $F_1,\dots,F_n$ are on $\mathbb{R}$. In all optimization problems we discussed in this section, the constraint is always $X_i\sim F_i$ for each $i\in[n]$ with $F_1,\dots,F_n$ given, and we assume that $F_1,\dots,F_n$ have finite second moments throughout this section.

3.4.1 Optimal transport under uncertainty on the set of components

We will consider a special class of cost functions, leading to the Monge-Kantorovich problem

$$\text{to minimize } \mathbb{E}\left[f\left(\sum_{i=1}^n X_i\right)\right] \quad \text{subject to } X_i\sim F_i,\ i\in[n], \tag{3.6}$$

where $f:\mathbb{R}\to\mathbb{R}$ is a convex function. This special setting is important to JM because, assuming that a joint mix with marginal distributions $F_1,\dots,F_n$ exists, then any joint mix is an optimizer of (3.6) due to Jensen's inequality, and conversely, any optimizer of (3.6) has to be a joint mix if f is strictly convex. As discussed by Puccetti and Wang (2015) and Wang and Wang (2016), one of the main motivations of JM is to solve optimization problems similar to (3.6).

Since joint mixes with given marginal distributions are not unique, we wonder whether a negatively dependent joint mix plays a special role among optimizers to (3.6). This is our

main question to address.

Although each joint mix minimizes (3.6), their distributions can be quite different. For a concrete example, suppose that the marginal distributions are standard Gaussian, and let n be even. With these marginals, $\mathbf{X}^{\mathrm{E}} \sim \mathrm{N}_n(\mathbf{0}_n, P_n^*)$ is an NA joint mix, where P_n^* is a matrix with diagonal entries being 1 and off-diagonal entries being $-1/(n-1)$, and $\mathbf{X}^{\mathrm{A}} = ((-1)^i Z)_{i\in[n]}$, $Z \sim \mathrm{N}_1(0,1)$, is another joint mix which is not NA. Here, "E" stands for "exchangeable" and "A" stands for "alternating". These two joint mixes have the same value $f(0)$ for (3.6). Nevertheless, $\mathbf{X}^{\mathrm{A}}$ may be seen as undesirable in some situations, because some subgroups of its components are comonotonic. Inspired by this, we consider the cost of a subset $K \subseteq [n]$ of risks $f\left(\sum_{i\in K} X_i\right)$. If K is known to the decision maker, then we are back to (3.6) with $(X_i)_{i\in[n]}$ replaced by $(X_i)_{i\in K}$.

In different applications, allowing a flexible choice of K may represent the absence of some risks in a risk aggregation pool, missing particles in a quantum system, an unspecified number of simulation sizes in a sampling program, or uncertainty on the participation of some agents in a risk-sharing game. In each context above, a decision maker may not know K a priori, and hence she may be interested in minimizing a weighted average of the cost, that is

$$C_\mu^f(X_1,\dots,X_n) := \sum_{K\subseteq[n]} \mathbb{E}\left[f\left(\sum_{i\in K} X_i\right)\right]\mu(K), \tag{3.7}$$

where μ is a probability on the sample space $2^{[n]}$, the power set of $[n]$, and $\sum_{i\in K} X_i$ is set to 0 if K is empty; here we slightly abuse the notation by setting $\mu(K) = \mu(\{K\})$, which should not lead to any confusion.

We consider the formulation of uncertainty as in the framework of Gilboa and Schmeidler (1989). With a probability on $2^{[n]}$ uncertain, we consider a set $\mathcal{M}$ of probabilities on $2^{[n]}$, called an uncertainty set. The formulation of (3.7) with uncertainty set $\mathcal{M}$ is

$$\text{to minimize } \sup_{\mu\in\mathcal{M}} C_\mu^f(X_1,\dots,X_n) \quad \text{subject to } X_i \sim F_i,\ i\in[n]. \tag{3.8}$$

The supremum represents a worst-case attitude towards uncertainty, which is axiomatized by Gilboa and Schmeidler (1989) in decision theory. We explain two simple special cases of

(3.8). First, by taking $\mathcal{M}$ as the set of all probabilities on $2^{[n]}$, the objective in (3.8) becomes

$$\sup_{\mu\in\mathcal{M}} C_\mu^f(X_1,\dots,X_n) = \max_{K\subseteq[n]} \mathbb{E}\left[f\left(\sum_{i\in K} X_i\right)\right], \tag{3.9}$$

which represents the situation of having no information on K. Second, by taking $\mathcal{M}$ as the set of all probabilities on $2^{[n]}$ supported by sets K of cardinality $|K| = k \in [n]$, the objective in (3.8) becomes

$$\sup_{\mu\in\mathcal{M}} C_\mu^f(X_1,\dots,X_n) = \max_{K\subseteq[n],\ |K|=k} \mathbb{E}\left[f\left(\sum_{i\in K} X_i\right)\right], \tag{3.10}$$

which represents the situation where one knows how large the subset K is, but not precisely how it is composed.

The problem (3.8) is generally difficult to solve. We will first focus on the homogeneous case where $F = F_1 = \cdots = F_n$, and this will be relaxed in Section 3.4.3. With this interpretation, it is natural to consider uncertainty sets $\mathcal{M}$ that are symmetric. We say that $\mathcal{M}$ is *symmetric* if $\mu \in \mathcal{M}$ implies $\mu_\pi \in \mathcal{M}$ for $\pi \in \mathfrak{S}_n$, where μ_π is a permutation of μ, defined by $\mu_\pi(K) = \mu(\{\pi(i) : i \in K\})$ for $K \subseteq [n]$.

Recall that we are interested in whether a negatively dependent joint mix plays a special role among other joint mixes. The next proposition provides a step in this direction. Also recall that P_n^* is the correlation matrix with off-diagonal entries equal to $-1/(n-1)$.

Proposition 3.5. *Suppose that $\mathcal{M}$ is symmetric, and $\mathbf{X}$ is a joint mix with identical marginals F. Then there exists an exchangeable NCD joint mix $\mathbf{X}^{\mathrm{E}}$ with marginals F and correlation matrix P_n^* such that*

$$\sup_{\mu\in\mathcal{M}} C_\mu^f(\mathbf{X}^{\mathrm{E}}) \leqslant \sup_{\mu\in\mathcal{M}} C_\mu^f(\mathbf{X})$$

for all measurable functions $f : \mathbb{R} \to \mathbb{R}$.

Proposition 3.5 illustrates the intuition that among all joint mixes, the NCD ones with correlation matrix P_n^* are better choices under uncertainty. However, this does not answer whether such NCD joint mixes are optimizers to our main optimal transport problem (3.8). In the next section, we consider the quadratic cost, and show that indeed those NCD joint mixes are solutions to (3.8) for the quadratic cost.

3.4.2 Quadratic cost

We consider the quadratic cost given by $f(x) = x^2$. In this case, we denote by

$$C^2_\mu(X_1, \dots, X_n) := \sum_{K \subseteq [n]} \mathbb{E}\left[\left(\sum_{i \in K} X_i\right)^2\right] \mu(K),$$

and (3.8) becomes, assuming homogeneous marginals,

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \mathbb{E}\left[\left(\sum_{i \in K} X_i\right)^2\right] \mu(K), \quad \text{subject to } X_i \sim F,\ i \in [n]. \tag{3.11}$$

Two other formulations related to the quadratic cost, the repulsive harmonic cost problem and the variance minimization problem, are discussed below in Examples 3.4 and 3.5.

It is clear that, for $\mu \in \mathcal{M}$, its permutation μ_π satisfies $C^2_\mu(\mathbf{X}) = C^2_{\mu_\pi}(\mathbf{X})$. We first show that the exchangeable NCD joint mix is a minimizer to (3.11) if $\mathcal{M}$ is symmetric.

Theorem 3.4. *Suppose that F is n-completely mixable with finite variance and $\mathcal{M}$ is symmetric. Then, each NCD joint mix with marginals F and correlation matrix P^*_n minimizes* (3.11).

If the marginal distribution F is Gaussian, then we can replace NCD by NA, NSD or NOD in Theorem 3.4, since for the Gaussian class NCD is equivalent to these notions (Proposition 3.1).

Theorem 3.4 does not state that all the minimizers to (3.11) are NCD joint mixes. For instance, if $\mathcal{M}$ contains only measures concentrated on K with $|K| = n$, then any joint mix minimizes (3.11); see also Remark 3.1 below for other similar cases. Next, we study the uniqueness of the optimizers for two special choices of $\mathcal{M}$ in (3.9) and (3.10), namely,

$$\text{to minimize } \max_{K \subseteq [n]} \mathbb{E}\left[\left(\sum_{i \in K} X_i\right)^2\right] \quad \text{subject to } X_i \sim F,\ i \in [n], \tag{3.12}$$

and for a fixed $k \in [n]$,

$$\text{to minimize } \max_{K \subseteq [n],\ |K| = k} \mathbb{E}\left[\left(\sum_{i \in K} X_i\right)^2\right] \quad \text{subject to } X_i \sim F,\ i \in [n]. \tag{3.13}$$

In Theorem 3.5 below, we will see that, assuming F has mean zero, negative dependence yields more stable costs in the presence of uncertainty. The exchangeable joint mix with correlation matrix P_n^* minimizes (3.13) for each $k \in [n]$, and this correlation matrix is unique for all minimizers for each $k \in [n] \setminus \{1, n-1, n\}$. As a consequence, all minimizers to (3.12) have the same correlation matrix P_n^* (this holds for $n \geqslant 3$).

Theorem 3.5. *Suppose that $n \geqslant 3$ and the distribution F is n-completely mixable with mean 0 and finite positive variance. A random vector is a minimizer to (3.12) if and only if it is an NCD joint mix with correlation matrix P_n^*. The same conclusion holds true if (3.12) is replaced by (3.13) with any $k \in [n] \setminus \{1, n-1, n\}$.*

Remark 3.1. We briefly comment on the three cases of k excluded from the statement regarding the unique minimizer of (3.13), and it will be clear that uniqueness cannot be expected in these cases. Recall that the marginal distributions of $\mathbf{X}$ are assumed identical.

1. If $k = 1$, then $\mathbb{E}[(\sum_{i \in K} X_i)^2] = \mathbb{E}[X_1^2]$ which does not depend on the dependence structure of $\mathbf{X}$, and hence any coupling minimizes (3.13).
2. If $k = n$, then $K = [n]$ and thus any joint mix minimizes (3.13).
3. If $k = n - 1$, then $\mathbb{E}[(\sum_{i \in K} X_i)^2] = \mathbb{E}[(c - X_1)^2]$ for any joint mix $\mathbf{X}$ with center c. Hence, any joint mix has the same value for (3.13).

Theorem 3.5 implies that for a standard Gaussian F, the exchangeable joint mix $\mathbf{X}^{\mathrm{E}} \sim \mathrm{N}_n(\mathbf{0}_n, P_n^*)$ is a minimizer to both (3.12) and (3.13) for each $k \in [n]$. If $n \geqslant 3$, this minimizer is unique among Gaussian vectors in both cases of (3.12) and (3.13) with $k \in [n] \setminus \{1, n-1, n\}$.

Remark 3.2. As we have seen in Remark 3.1, if $n = 3$, then any joint mix minimizes (3.13) for each $k \in [n]$. The uniqueness statement in Theorem 3.5 implies that the covariance structure of a joint mix is unique for $n = 3$, as we see in Example 3.3.

Below we discuss two specific optimal transport problems related to the quadratic cost.

Example 3.4 (Variance minimization)**.** The quadratic cost minimization problem is equivalent to variance minimization with given marginals. It is clear that

$$C_\mu^2(X_1, \dots, X_n) = \sum_{K \subseteq [n]} \operatorname{var}\left(\sum_{i \in K} X_i\right) \mu(K) + \sum_{K \subseteq [n]} \left(\sum_{i \in K} \mathbb{E}[X_i]\right)^2 \mu(K),$$

and the second term does not depend on the dependence structure of $(X_1, \dots, X_n)$. If $F_1, \dots, F_n$ have zero mean, then the problem (3.11) can be written as

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \operatorname{var}\left(\sum_{i \in K} X_i\right) \mu(K) \quad \text{subject to } X_i \sim F,\ i \in [n].$$

Variance minimization is a classic problem in Monte Carlo simulation (Craiu and Meng, 2001, 2005) and risk management (Rüschendorf, 2013). The above arguments show that the statements in Theorems 3.4 and 3.5 hold true if the objective of quadratic cost $\mathbb{E}[(\sum_{i \in K} X_i)^2]$ is replaced by the variance $\operatorname{var}(\sum_{i \in K} X_i)$.

Example 3.5 (Repulsive harmonic cost)**.** The *repulsive harmonic cost* function is defined by

$$c(x_1, \dots, x_n) = -\sum_{i,j=1}^{n} (x_i - x_j)^2, \qquad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

This cost function originates from the so-called weak interaction regime in Quantum Mechanics; see e.g., Di Marino et al. (2017). Any joint mix minimizes the expected repulsive harmonic cost. To see this, we can rewrite

$$\mathbb{E}[c(X_1, \dots, X_n)] = -2n \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right]. \tag{3.14}$$

Since the first terms on the right-hand side of (3.14) do not depend on the dependence structure of $(X_1, \dots, X_n)$, minimizing $\mathbb{E}[c(X_1, \dots, X_n)]$ is equivalent to minimizing $\mathbb{E}[(\sum_{i=1}^{n} X_i)^2]$, which is clearly minimized if $(X_1, \dots, X_n)$ is a joint mix. Let $c_K(x_1, \dots, x_n) = -\sum_{i,j \in K} (x_i - x_j)^2$, $(x_1, \dots, x_n) \in \mathbb{R}^n$, for $K \subseteq [n]$. The problem (3.11) can be written as

$$\text{to minimize } \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \left(\frac{1}{2}\mathbb{E}[c_K(\mathbf{X})] + |K| \sum_{i \in K} \mathbb{E}[X_i^2]\right) \mu(K) \quad \text{subject to } X_i \sim F,\ i \in [n].$$

The statement in Theorem 3.4 remains true if the objective of quadratic cost $\mathbb{E}[(\sum_{i \in K} X_i)^2]$ is replaced by the cost $\mathbb{E}[c_K(\mathbf{X})]$.

3.4.3 Discussions on heterogeneous marginals

In Theorem 3.5, we assumed that the marginal distributions are identical. This assumption is not dispensable, as the situation for heterogeneous marginals is drastically different

and we do not have general results. In this section, we present a result in the simple case $n = 3$ and provide several examples to discuss some subtle issues and open questions. To illustrate these issues, we focus on the problems (3.12) and (3.13) for $n = 3$. In all examples, we explain with Gaussian marginal distributions, but this assumption can be replaced as long as the covariance matrices in the examples are compatible with the marginals.

Proposition 3.6. *Let $n = 3$. For any tuple of marginal distributions with finite variance vector and zero means, any joint mix, if it exists, minimizes (3.12). If an NCD joint mix exists, then no random vector with any positive bivariate covariance can minimize (3.12).*

Remark 3.3. Proposition 3.6 states that, if a Gaussian triplet supports an NCD joint mix, then it minimizes (3.12), and all Gaussian minimizers must be NCD. It is not clear whether this observation can be extended to $n \geqslant 4$.

Unlike the situation in Theorem 3.5, uniqueness of the covariance matrix does not hold in the setting of Proposition 3.6, as illustrated in the following example.

Example 3.6. Consider Gaussian marginal distributions with variance vector $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (2, 1, 1)$. In this case, (3.4) holds, and an NCD joint mix exists by Theorem 3.2. Both the covariance matrices Σ and Σ' defined by

$$\Sigma = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma' = \begin{pmatrix} 2 & -1/2 & -1 \\ -1/2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

minimize (3.12) subject to the marginal distributions. We can see that Σ corresponds to an NCD joint mix, whereas Σ' corresponds to an NCD random vector, but not a joint mix.

The next example illustrates that, although a joint mix generally minimizes (3.12) in case $n = 3$, NCD may be more relevant than joint mixes for minimizing (3.13) with some $k \neq n$ when the two dependence requirements cannot be simultaneously achieved.

Example 3.7. Consider Gaussian marginal distributions with variance vector $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (4, 1, 1)$. In this case, (3.4) does not hold, and no NCD joint mix exists. Both the covariance

matrices Σ and Σ' defined by

$$\Sigma = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 1 & 1 \\ -2 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma' = \begin{pmatrix} 4 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

minimize (3.12) subject to the marginal distributions. The covariance matrix Σ corresponds to a joint mix, but not NCD. The covariance matrix Σ' corresponds to an NCD random vector, but not a joint mix. Thus, the problem (3.12) admits an NCD minimizing distribution $\mathrm{N}_3(\mathbf{0}_3, \Sigma')$. Moreover, for (3.13) with $k = 2$, the NCD distribution $\mathrm{N}_3(\mathbf{0}_3, \Sigma')$ has a maximum of 3 which is strictly better than the joint mix distribution $\mathrm{N}_3(\mathbf{0}_3, \Sigma)$ with a maximum of 4.

Example 3.7 suggests, informally, that there is a trade-off between a joint mix and NCD when both cannot be attained simultaneously, with a joint mix minimizing (3.13) for $k = n$, and an NCD random vector improving (3.13) from the case of a joint mix for some $1 < k < n$. In fact, (3.13) is not always minimized by NCD random vectors as seen in the following example.

Example 3.8. Consider marginal distributions with variance vector $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (\sigma^2, 1, 1)$ and zero means, where $\sigma > 3$. For any (X_1, X_2, X_3) with the given marginals, we have

$$\mathbb{E}\left[(X_2 + X_3)^2\right] \leqslant 4 < (\sigma - 1)^2 \leqslant \min\left(\mathbb{E}\left[(X_1 + X_2)^2\right], \mathbb{E}\left[(X_1 + X_3)^2\right]\right),$$

and hence

$$\max_{K \subseteq [3],\ |K|=2} \mathbb{E}\left[\left(\sum_{i \in K} X_i\right)^2\right] = \sigma^2 + 1 + 2\sigma \max(\rho_{12}, \rho_{13}), \tag{3.15}$$

where ρ_{ij}, $i, j \in [3]$ is the correlation coefficient of (X_i, X_j). For Gaussian marginals, the minimum of (3.15) is attained if and only if $\rho_{12} = \rho_{13} = -1$. In this case, $\rho_{23} = 1$ is the only possible correlation, and thus the minimizer to (3.15) cannot be NCD.

On the other hand, the next example shows that, if $n = 3$, there always exists an NCD minimizer to (3.12) for Gaussian marginals.

Example 3.9. Let (X_1, X_2, X_3) follow a multivariate Gaussian distribution with equicorrelation matrix P_3^*; i.e., all pairwise correlation coefficients are $-1/2$. The variances σ_1^2, σ_2^2

and σ_3^2 are assumed to satisfy $\sigma_1 \leqslant \sigma_2 \leqslant \sigma_3$ without the loss of generality and the means for prescribed marginal distributions are assumed to be zero. We can easily verify that each of $\mathbb{E}\left[(X_1+X_2+X_3)^2\right]$ and $\mathbb{E}\left[(X_i+X_j)^2\right]$, $i,j \in [3]$, is smaller than or equal to σ_3^2. Hence, (X_1, X_2, X_3) attains the lower bound

$$\max_{K\subseteq[3]} \mathbb{E}\left[\left(\sum_{i\in K} X_i\right)^2\right] = \sigma_3^2 = \max_{i\in[3]} \sigma_i^2,$$

and thus it minimizes (3.12).

Remark 3.4. For $n \geqslant 4$, it is not clear whether there always exists an NCD minimizer to (3.12) under a general heterogeneous marginal constraint.

3.5 Elliptical distributions

Elliptical distributions form a tractable class of joint mixes for arbitrary dimensions. In this section, we investigate negative dependence properties of such elliptical joint mixes.

An n-dimensional *elliptical distribution* is a family of multivariate distributions defined through the characteristic function

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[\exp\left(\mathrm{i}\mathbf{t}^\top \mathbf{X}\right)\right] = \exp\left(\mathrm{i}\mathbf{t}^\top \boldsymbol{\mu}\right)\psi(\mathbf{t}^\top \Sigma \mathbf{t}), \qquad \mathbf{t} \in \mathbb{R}^n, \tag{3.16}$$

for some location parameter $\boldsymbol{\mu} \in \mathbb{R}^n$, $n \times n$ positive semi-definite symmetric matrix $\Sigma \in \mathbb{R}^{n\times n}$ and the so-called *characteristic generator* $\psi : \mathbb{R}_+ \to \mathbb{R}$, where $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geqslant 0\}$. See Section 6 of McNeil et al. (2015) for more properties. We denote an elliptical distribution by $\mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ and refer to $\boldsymbol{\mu}$ as the *location vector* and Σ the *dispersion matrix*. We say that an elliptical distribution $\mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ is non-degenerate if all its marginals are non-degenerate (i.e., not a point-mass). Equivalently, the diagonal entries of Σ are positive. As presented in Proposition 6.27 of McNeil et al. (2015), a random vector $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ with $\mathrm{rank}(\Sigma) = k$ admits the stochastic representation $\mathbf{X} = \boldsymbol{\mu} + RA\mathbf{S}$, where $\mathbf{S}$ is the uniform distribution on the unit sphere on $\mathbb{R}^k$, the radial random variable $R \geqslant 0$ is independent of $\mathbf{S}$, and $A \in \mathbb{R}^{n\times k}$ is such that $AA^\top = \Sigma$. With this representation, we have that $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{X}) = \mathbb{E}[R^2]\,\Sigma/k$ provided $\mathbb{E}[R^2] < \infty$.

We first present a simple lemma on elliptical joint mixes which will be useful for later discussions.

Lemma 3.1. *An n-dimensional elliptically distributed random vector $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ is a joint mix if and only if $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$ or $\psi = 1$ on $\mathbb{R}_+$.*

Negative dependence of such an elliptical joint mix is the main topic of this section. We next provide a characterization for NCD-JM as an extension to Theorem 3.2.

Proposition 3.7. *Suppose that ψ is the characteristic generator of an n-dimensional elliptical distribution. A tuple of univariate distributions $(\mathrm{E}_1(\mu_i, \sigma_i^2, \psi), i \in [n])$ supports an NCD joint mix if and only if (3.4) holds, that is, $2\max_{i\in[n]} \sigma_i^2 \leqslant \sum_{i\in[n]} \sigma_i^2$. Moreover, such an NCD joint mix can be chosen to follow an elliptical distribution.*

As we showed in Proposition 3.1, for Gaussian random vectors, NA, NSD and NOD are all equivalent to NCD; that is, the bivariate correlations are non-positive. Recall that P_n^* is an $n \times n$ matrix whose diagonal entries are 1 and off-diagonal entries are $-1/(n-1)$. Together with Lemma 3.1, the matrix P_n^* is the only choice of Σ with diagonal entries being 1 such that $\mathbf{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \Sigma)$ is an exchangeable NA (and thus NSD and NOD) joint mix.

One may hope that non-Gaussian elliptical distributions can represent NOD, NSD and NA joint mixes for $n \geqslant 3$. The following result states that Gaussian family is characterized as the only elliptical family which admits such a negatively dependent n-joint mix for all n. For a characteristic generator ψ, denote by $\mathcal{E}(\psi)$ the class of all non-degenerate random vectors following an elliptical distribution with characteristic generator ψ. In what follows, a class $\mathcal{E}(\psi)$ is a *Gaussian variance mixture family* if there exists a nonnegative random variable W such that each member $\mathbf{X}$ admits the stochastic representation $\mathbf{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + \sqrt{W} A \mathbf{Z}$, where $\boldsymbol{\mu} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times k}$, and $\mathbf{Z}$ is a k-dimensional standard Gaussian independent of W.

Theorem 3.6. *Let ψ be a characteristic generator.*

(i) The class $\mathcal{E}(\psi)$ contains an NCD n-joint mix for all $n \geqslant 2$ if and only if the class $\mathcal{E}(\psi)$ is a Gaussian variance mixture family.

(ii) The class $\mathcal{E}(\psi)$ contains an NOD, NSD, or NA n-joint mix for all $n \geqslant 2$ if and only if the class $\mathcal{E}(\psi)$ is Gaussian.

Theorem 3.6 shows a clear contrast between NCD and other concepts of negative dependence. As seen in the proof of Theorem 3.6, NCD does not restrict the class of elliptical distributions since ψ generates an n-dimensional elliptical distribution for every $n \in \mathbb{N}$ if and only if the corresponding elliptical class is a Gaussian variance mixture family (Fang et al., 1990, Section 2.6). Note that a class of multivariate t distributions (with a common degree of freedom) is an example of a Gaussian variance mixture family. On the other hand, NOD, NSD and NA characterize Gaussian. This result stems from the fact that multivariate Gaussian distribution is the only one among elliptical distributions such that independence is equivalent to uncorrelatedness.

3.6 Conclusion

This chapter has focused on the relationship between JM and classic notions of negative dependence such as NOD and NA. Various connections between these concepts are obtained, and some conditions for a joint mix to be negatively dependent are derived. In particular, an exchangeable negatively dependent joint mix solves a multi-marginal optimal transport problem for quadratic cost under uncertainty on the participation of agents.

Negative dependence is always studied with many technical challenges. Although our main questions are addressed or partially addressed in this chapter, they give rise to many questions that remain open. We list a few of them that we find particularly interesting.

1. Under what conditions, possibly stronger than exchangeability and NOD, is a joint mix NA? An example of NOD joint mix that is not NA can be found in Section 3.2 of Malinovsky and Rinott (2023) in the context of knockout tournaments with a nonrandom draw.

2. Under what general conditions, other than Gaussian, do we know a tuple of distributions supports an NA joint mix? For a fixed $n \geqslant 3$, this question is not clear even within the elliptical class.

3. It is unclear whether the decomposition result in Theorem 3.3 can be generalized to joint mixes that take infinitely many values or that are continuously distributed.

4. Assuming homogeneous marginals, does an exchangeable joint mix solve problem (3.8) for a general convex cost function f? In Theorem 3.4, we showed that this holds true for quadratic cost. We also know that a joint mix is an optimizer for the general convex cost problem without uncertainty. These observations seem to hint at the possible optimality of some exchangeable joint mix for general convex cost under uncertainty, but we do not have a proof.

5. Do negatively dependent joint mixes play an important role in optimization problems other than the ones considered in Section 3.4? It is also unclear how results in Section 3.4 can be extended to heterogeneous marginal distributions with dimension higher than 3. Two unsolved questions have already been mentioned in Remarks 3.3 and 3.4.

These questions yield new challenges to dependence theory and require future research.

3.7 Proofs of all results

Some notions of negative dependence introduced in Section 3.2 are related to stochastic orders. We first introduce some concepts of stochastic order.

For two n-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{X}$ is said to be less than $\mathbf{Y}$ in *lower concordance order* (denoted by $\mathbf{X} \leqslant_{\mathrm{cL}} \mathbf{Y}$) if $\mathbb{P}(\mathbf{X} \leqslant \mathbf{t}) \leqslant \mathbb{P}(\mathbf{Y} \leqslant \mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^n$, *upper concordance order* (denoted by $\mathbf{X} \leqslant_{\mathrm{cU}} \mathbf{Y}$) if $\mathbb{P}(\mathbf{X} > \mathbf{t}) \leqslant \mathbb{P}(\mathbf{Y} > \mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^n$, *concordance order* (denoted by $\mathbf{X} \leqslant_{\mathrm{c}} \mathbf{Y}$) if $\mathbf{X} \leqslant_{\mathrm{cL}} \mathbf{Y}$ and $\mathbf{X} \leqslant_{\mathrm{cU}} \mathbf{Y}$, and in *supermodular order* (denoted by $\mathbf{X} \leqslant_{\mathrm{sm}} \mathbf{Y}$) if $\mathbb{E}[\psi(\mathbf{X})] \leqslant \mathbb{E}[\psi(\mathbf{Y})]$ for all supermodular functions $\psi : \mathbb{R}^n \to \mathbb{R}$ such that the expectations exist. Using these notations of stochastic order, the notions of negative dependence NLOD, NUOD, NOD and NSD for an n-dimensional random vector $\mathbf{X} = (X_1, \dots, X_n)$ are denoted by $\mathbf{X} \leqslant_{\mathrm{cL}} \mathbf{X}^{\perp}$, $\mathbf{X} \leqslant_{\mathrm{cU}} \mathbf{X}^{\perp}$, $\mathbf{X} \leqslant_{\mathrm{c}} \mathbf{X}^{\perp}$ and $\mathbf{X} \leqslant_{\mathrm{sm}} \mathbf{X}^{\perp}$, respectively, where we recall that $\mathbf{X}^{\perp} = (X_1^{\perp}, \dots, X_n^{\perp})$ is a random vector with independent components such that $X_i \overset{\mathrm{d}}{=} X_i^{\perp}$, $i \in [n]$.

Proof of Proposition 3.1. In part (a), the implication from (v) to (i) is shown by Joag-Dev and Proschan (1983). The other implications follow from (3.3). Parts (b) and (c) can be easily checked by definition. Finally, part (d) follows from the fact that a CT random vector for $n \geqslant 3$ cannot have continuous marginal distributions (Dall'Aglio, 1972; Puccetti and Wang, 2015). □

Proof of Theorem 3.1. Note that to show NA, it suffices to show (3.1) for A, B that form a partition of $[n]$, as we can choose increasing functions in (3.1) that only depend on a subset of A and B. Let f and g be two increasing functions on $\mathbb{R}^d$ and $\mathbb{R}^{n-d}$, respectively, where d is the cardinality of A. Note that

$$\operatorname{cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) = \mathbb{E}[\operatorname{cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)|S_A)] + \operatorname{cov}(\mathbb{E}[f(\mathbf{X}_A)|S_A], \mathbb{E}[g(\mathbf{X}_B)|S_A]);$$

see (1.1) of Joag-Dev and Proschan (1983). Using conditional independence (a), we get

$$\operatorname{cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) = \operatorname{cov}(\mathbb{E}[f(\mathbf{X}_A)|S_A], \mathbb{E}[g(\mathbf{X}_B)|S_A]).$$

Since $S_A + S_B$ is a constant, condition (b) implies that $\mathbb{E}[f(\mathbf{X}_A)|S_A]$ is an increasing function of S_A, and $\mathbb{E}[g(\mathbf{X}_B)|S_A] = \mathbb{E}[g(\mathbf{X}_B)|S_B]$ is a decreasing function of S_A. This shows that their covariance is non-positive. Therefore, $\mathbf{X}$ is NA. □

Proof of Proposition 3.2. Denote by $\mathbf{X} = (X_1, \dots, X_n)$ the NSD n-joint mix with joint distribution $F_{\mathbf{X}}$. Let $\mathbf{X}^\Pi = (X_{\Pi(1)}, \dots, X_{\Pi(n)})$ be an exchangeable joint mix, where Π follows a uniform distribution on $\mathfrak{S}_n$ and is independent of $\mathbf{X}$. Obviously $\mathbf{X}^\Pi$ is a joint mix, and has the same marginal distributions as $\mathbf{X}$. Let $\bar{F} = \frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} F_{\mathbf{X}^\pi}$ be the distribution function of $\mathbf{X}^\Pi$. Then $\bar{F}$ is exchangeable. Moreover, $\bar{F}$ is NSD since

$$\mathbb{E}[\psi(\mathbf{X}^\Pi)] = \frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} \mathbb{E}[\psi(\mathbf{X}^\Pi)] \leqslant \frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} \mathbb{E}[\psi(\mathbf{X}^\perp)] = \mathbb{E}[\psi(\mathbf{X}^\perp)]$$

for every supermodular function ψ such that the expectations above exist. Other cases of NOD, NUOD and NLOD are shown analogously. □

Proof of Proposition 3.3. Without loss of generality, assume σ_n^2 is the maximum of $\{\sigma_1^2, \dots, \sigma_n^2\}$. Note that NCD implies that the bivariate correlations are non-positive. If $(X_1, \dots, X_n)$ is

an NCD joint mix where $X_i \sim F_i$, $i \in [n]$, then

$$\sigma_n^2 = \mathrm{var}(X_n) = \mathrm{var}(X_1 + \cdots + X_{n-1}) \leqslant \sum_{i=1}^{n-1} \mathrm{var}(X_i) = \sum_{i=1}^{n-1} \sigma_i^2,$$

which yields (3.4) by adding σ_n^2 to both sides. □

Proof of Theorem 3.2. The necessity follows from Proposition 3.3, and below we show sufficiency. Suppose that (3.4) holds. Without loss of generality, we can assume $\sigma_n \geqslant \sigma_{n-1} \geqslant \cdots \geqslant \sigma_1$. It suffices to consider $n \geqslant 3$ and $\sigma_{n-1} > 0$, and otherwise the problem is trivial. Moreover, the location parameters of the Gaussian distributions are not relevant, and they are assumed to be 0.

Let λ be a constant such that

$$\lambda^2 \sum_{i=1}^{n-1} \sigma_i^2 + (1-\lambda^2)\sigma_{n-1}^2 = \sigma_n^2. \tag{3.17}$$

By (3.4), we have $\sum_{i=1}^{n-1} \sigma_i^2 \geqslant \sigma_n^2 \geqslant \sigma_{n-1}^2$, and this ensures that we can take $\lambda \in [0,1]$.

Let P_n^* be a matrix with diagonal entries being 1 and off-diagonal entries being $-1/(n-1)$, and let $P_n^\perp$ be a matrix with diagonal entries being 1 and off-diagonal entries being 0. Take $\mathbf{Y} = (Y_1, \dots, Y_{n-1}) \sim \mathrm{N}_{n-1}(\mathbf{0}_{n-1}, P_{n-1}^\perp)$ and

$$\mathbf{Z}^{(m)} = (Z_m^{(m)}, \dots, Z_n^{(m)}) \sim \mathrm{N}_{n-m+1}(\mathbf{0}_{n-m+1}, P_{n-m+1}^*), \qquad m = 1, \dots, n-1,$$

such that $\mathbf{Y}, \mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(n-1)}$ are independent. Note that $\mathbf{Z}^{(n-1)} = (Z_{n-1}^{(n-1)}, Z_n^{(n-1)}) \sim \mathrm{N}_2(\mathbf{0}_2, P_2^*)$ is 2-dimensional, and each $\mathbf{Z}^{(m)}$ is a joint mix.

For notational simplicity, let the function d be given by $d(a,b) = (a^2 - b^2)^{1/2}$ for $a \geqslant b \geqslant 0$. Note that $a^2 = d(a,b)^2 + b^2$. Moreover, for $k = 1, \dots, n-1$, let

$$\alpha_k = d(\sigma_k, \sigma_{k-1}) = \left(\sigma_k^2 - \sigma_{k-1}^2\right)^{1/2},$$

with $\sigma_0 = 0$, and thus $\alpha_1 = \sigma_1$. For $k = 1, \dots, n-1$, let

$$X_k = \lambda \sigma_k Y_k + d(1,\lambda) \sum_{j=1}^{k} \alpha_j Z_k^{(j)}.$$

Moreover, let

$$X_n = -\lambda Y^* + d(1,\lambda)\sum_{j=1}^{n-1}\alpha_j Z_n^{(j)}, \quad \text{where } Y^* = \sum_{k=1}^{n-1}\sigma_k Y_k.$$

For $k = 1, \ldots, n-1$, using independence among $Z_k^{(1)}, \ldots, Z_k^{(k)}$, we get

$$\text{var}\left(\sum_{j=1}^{k}\alpha_j Z_k^{(j)}\right) = \sum_{i=1}^{k}\alpha_j^2 = \sigma_1^2 + d(\sigma_2,\sigma_1)^2 + \cdots + d(\sigma_k,\sigma_{k-1})^2 = \sigma_k^2.$$

Hence, $\sum_{j=1}^{k}\alpha_j Z_k^{(j)} \sim \text{N}_1(0,\sigma_k^2)$, and again using independence of Y_k and $\sum_{j=1}^{k}\alpha_j Z_k^{(j)}$, we get $X_k \sim \text{N}_1(0,\sigma_k^2)$. By (3.17), we have

$$\text{var}(X_n) = \text{var}\left(\lambda\sum_{k=1}^{n-1}\sigma_k Y_k\right) + \text{var}\left(d(1,\lambda)\sum_{j=1}^{n-1}\alpha_j Z_n^{(j)}\right) = \lambda^2\sum_{i=1}^{n-1}\sigma_i^2 + (1-\lambda^2)\sigma_{n-1}^2 = \sigma_n^2.$$

Hence, $X_n \sim \text{N}_1(0,\sigma_n^2)$.

Next, we show that $(X_1, \ldots, X_n)$ is a joint mix. We can directly compute

$$\begin{aligned}\sum_{k=1}^{n}X_k &= \sum_{i=k}^{n-1}\lambda\sigma_k Y_k + d(1,\lambda)\sum_{k=1}^{n-1}\sum_{j=1}^{k}\alpha_j Z_k^{(j)} - \lambda\sum_{k=1}^{n-1}\sigma_k Y_k + d(1,\lambda)\sum_{j=1}^{n-1}\alpha_j Z_n^{(j)}\\ &= d(1,\lambda)\sum_{j=1}^{n-1}\sum_{k=j}^{n}\alpha_j Z_k^{(j)} = 0,\end{aligned}$$

where the last equality follows from the fact that $\mathbf{Z}^{(j)}$ is a joint mix for each $j = 1, \ldots, n-1$.

We check that $(X_1, \ldots, X_n)$ is NA. This follows from the fact that $(X_1, \ldots, X_n)$ is the weighted sum of several independent NA random vectors $(\sigma_1 Y_1, \ldots, \sigma_{n-1}Y_{n-1}, -Y^*)$ and $(\mathbf{0}_{m-1}, \mathbf{Z}^{(m)})$ for $m = 1, \ldots, n-1$. Alternatively, one can check that all non-zero terms in $\text{cov}(X_i, X_j)$ are negative for $i \neq j$ as follows. For X_k, X_l with $k, l \leqslant n-1$ and $k < l$, we have

$$\begin{aligned}\text{cov}(X_k, X_l) &= \text{cov}\left(\lambda\sigma_k Y_k + d(1,\lambda)\sum_{j=1}^{k}\alpha_j Z_k^{(j)}, \lambda\sigma_l Y_l + d(1,\lambda)\sum_{i=1}^{l}\alpha_i Z_l^{(i)}\right)\\ &= \lambda^2\sigma_k\sigma_l\text{cov}(Y_k, Y_l) + \lambda\sigma_k d(1,\lambda)\sum_{i=1}^{l}\alpha_i\text{cov}(Z_l^{(i)}, Y_k)\\ &\quad + \lambda\sigma_l d(1,\lambda)\sum_{j=1}^{k}\alpha_j\text{cov}(Z_k^{(j)}, Y_l) + d^2(1,\lambda)\sum_{j=1}^{k}\sum_{i=1}^{l}\alpha_j\alpha_i\text{cov}(Z_k^{(j)}, Z_l^{(i)})\\ &= -d^2(1,\lambda)\sum_{j=1}^{k}\frac{\alpha_j^2}{n-j} \leqslant 0.\end{aligned}$$

For X_k, X_n for all $k \leqslant n-1$, we have

$$\begin{aligned}
\operatorname{cov}(X_k, X_n) &= \operatorname{cov}\left(\lambda\sigma_k Y_k + d(1,\lambda)\sum_{j=1}^{k}\alpha_j Z_k^{(j)}, -\lambda\sum_{i=1}^{n-1}\sigma_i Y_i + d(1,\lambda)\sum_{i=1}^{n-1}\alpha_i Z_n^{(i)}\right) \\
&= -\lambda^2\sigma_k\sum_{i=1}^{n-1}\sigma_i\operatorname{cov}(Y_k, Y_i) + \lambda\sigma_k d(1,\lambda)\sum_{i=1}^{n-1}\alpha_i\operatorname{cov}(Z_n^{(i)}, Y_k) \\
&\quad - \lambda d(1,\lambda)\sum_{j=1}^{k}\sum_{i=1}^{n-1}\sigma_i\alpha_j\operatorname{cov}(Z_k^{(j)}, Y_i) + d^2(1,\lambda)\sum_{j=1}^{k}\sum_{i=1}^{n-1}\alpha_j\alpha_i\operatorname{cov}(Z_k^{(j)}, Z_n^{(i)}) \\
&= -\lambda^2\sigma_k^2 - d^2(1,\lambda)\sum_{j=1}^{k}\frac{\alpha_j^2}{n-j} \leqslant 0.
\end{aligned}$$

Finally, the joint mix can be chosen as multivariate Gaussian by the construction of $(X_1, \ldots, X_n)$ as the sum of Gaussian vectors. □

Proof of Theorem 3.3. The "if" statement is straightforward, and we will check the "only if" statement. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a joint mix and denote by $c = \sum_{i=1}^{n} X_i \in \mathbb{R}$. First, suppose that each component of $\mathbf{X}$ is positive. Denote by $V \subset \mathbb{R}$ the set of all possible values taken by random variables of the form $\sum_{i=1}^{j} X_i$ for $j = 0, \ldots, n$, with the convention that $\sum_{j=1}^{0} X_j = 0$. Clearly, V is finite. The elements of V are denoted by $v_0, v_1, \ldots, v_K$ such that $v_0 < v_1 < \cdots < v_K$. Our assumptions imply that $v_0 = 0$, $v_1 > 0$ and $v_K = c$ because $\sum_{i=1}^{n} X_i = c$. For $k \in [K]$ and $i \in [n]$, let

$$Y_{k,i} = \mathbb{1}_{\{\sum_{j=1}^{i} X_j \geqslant v_k\}} - \mathbb{1}_{\{\sum_{j=1}^{i-1} X_j \geqslant v_k\}}$$

and let $\mathbf{Y}_k = (Y_{k,1}, \ldots, Y_{k,n})$. Since each X_j is positive, the value of $Y_{k,i}$ is either 0 or 1, and

$$\sum_{i=1}^{n} Y_{k,i} = \mathbb{1}_{\{\sum_{j=1}^{n} X_j \geqslant v_k\}} - \mathbb{1}_{\{0 \geqslant v_k\}} = \mathbb{1}_{\{c \geqslant v_k\}} = 1.$$

Therefore, $\mathbf{Y}_k$ follows a binary multinomial distribution for each $k \in [K]$. Let $\mathbf{X}_k = (v_k - v_{k-1})\mathbf{Y}_k$ for $k \in [K]$ with $v_0 = 0$. Note that for $i \in [n]$,

$$\sum_{k=1}^{K} X_{k,i} = \sum_{k=1}^{K}(v_k - v_{k-1})\left(\mathbb{1}_{\{\sum_{j=1}^{i} X_j \geqslant v_k\}} - \mathbb{1}_{\{\sum_{j=1}^{i-1} X_j \geqslant v_k\}}\right) = \sum_{j=1}^{i} X_j - \sum_{j=1}^{i-1} X_j = X_i,$$

where we used the identity $\sum_{k=1}^{K}(v_k - v_{k-1})\mathbb{1}_{\{x \geqslant v_k\}} = x$ for $x \in V$. Therefore, $\sum_{k=1}^{K} \mathbf{X}_k = \mathbf{X}$, showing that $\mathbf{X}$ can be represented as a finite linear combination of binary multinomial random vectors.

If some components of $\mathbf{X}$ are not positive, we can take $m \in \mathbb{R}$ such that $X_i > m$ for each $i \in [n]$. Applying the above result, we know that $(X_1 - m, \dots, X_n - m)$ can be decomposed as the sum of JM joint mixes. Note that $\mathbf{X} = (X_1 - m, \dots, X_n - m) + (m, \dots, m)$, and $(m, \dots, m)$ is m times the sum of n binary multinomial random vectors $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$. Hence, $\mathbf{X}$ admits a finite linear combination of binary multinomial random vectors. □

Proof of Proposition 3.4. Let G be the joint distribution of an exchangeable joint mix. Let us write

$$\widetilde{G}(A) = \int_{\mathbb{R}^n} \delta_{\mathbf{a}}(A) \, \mathrm{d}G(\mathbf{a}), \qquad A \in \mathcal{B}(\mathbb{R}^n),$$

where $\delta_{\mathbf{a}}$ is the point-mass at $\mathbf{a}$. By exchangeability, we have $G(A^\pi) = G(A)$ for $\pi \in \mathfrak{S}_n$ and $A \in \mathcal{B}(\mathbb{R}^n)$, where A^π is π applied to elements of A. Therefore,

$$G(A) = \int_{\mathbb{R}^n} \delta_{\mathbf{a}^\pi}(A) \, \mathrm{d}G(\mathbf{a}).$$

Taking an average of the above formula over $\mathfrak{S}_n$, we have

$$G(A) = \int_{\mathbb{R}^n} U_{\mathbf{a}}(A) \, \mathrm{d}G(\mathbf{a}).$$

It is known that each $U_{\mathbf{a}}$ is NA (Joag-Dev and Proschan, 1983, Theorem 2.11). Moreover, the center of the joint mix distributed as $U_{\mathbf{a}}$ is μ since G is supported on $\{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 + \cdots + x_n = \mu\}$. □

Proof of Proposition 3.5. As $\mathcal{M}$ is symmetric, we have $\sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}) = \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\pi)$ for all $\pi \in \mathfrak{S}_n$. Let Π be uniformly distributed on $\mathfrak{S}_n$ and independent of $\mathbf{X}$. Plugging $\mathbf{X}^\Pi$ in the objective (3.7), we have

$$\begin{aligned}
\sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\Pi) &= \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \mathbb{E}\left[f\left(\sum_{i \in K} X_i^\Pi \right) \right] \mu(K) \\
&= \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \mathbb{E}\left[f\left(\sum_{i \in K} X_i^\pi \right) \right] \mu(K) \\
&\leqslant \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \sup_{\mu \in \mathcal{M}} \sum_{K \subseteq [n]} \mathbb{E}\left[f\left(\sum_{i \in K} X_i^\pi \right) \right] \mu(K) \\
&= \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}^\pi) = \sup_{\mu \in \mathcal{M}} C_\mu^f(\mathbf{X}).
\end{aligned}$$

Hence, $\sup_{\mu\in\mathcal{M}} C_\mu^f(\mathbf{X}^\Pi) \leqslant \sup_{\mu\in\mathcal{M}} C_\mu^f(\mathbf{X})$. Furthermore, as $\mathbf{X}$ is a joint mix, we have that $\mathbf{X}^\Pi$ is an exchangeable NCD joint mix with marginals F and correlation matrix P_n^*. □

Proof of Theorem 3.4. Without loss of generality, we assume that the variance of F is 1. Using the same argument in the proof of Proposition 3.5, for any $\mathbf{X}$ with identical marginals F, we have $\sup_{\mu\in\mathcal{M}} C_\mu^2(\mathbf{X}^\Pi) \leqslant \sup_{\mu\in\mathcal{M}} C_\mu^2(\mathbf{X})$, where Π is uniformly distributed on $\mathfrak{S}_n$. Let $\mathbf{X}_\rho$ be a random vector with identical marginals F and a correlation matrix whose off-diagonal entries are all ρ. Since correlation matrices are positive semi-definite, we have $\rho \in [-1/(n-1), 1]$, with $\rho = -1/(n-1)$ attainable since F is n-completely mixable. The value of $\sup_{\mu\in\mathcal{M}} C_\mu^2(\mathbf{X})$ only depends on the correlation matrix. Therefore, it suffices to find an optimizer of the form $\mathbf{X}_\rho$ for some $\rho \in [-1/(n-1), 1]$. Note that

$$\begin{aligned}\sup_{\mu\in\mathcal{M}} C_\mu^2(\mathbf{X}_\rho) &= \sup_{\mu\in\mathcal{M}} \sum_{K\subseteq[n]} \left(\operatorname{var}\left(\sum_{i\in K} X_i\right) + \left(\mathbb{E}\left[\sum_{i\in K} X_i\right]\right)^2 \right) \mu(K) \\ &= \sup_{\mu\in\mathcal{M}} \sum_{k=1}^{n} \sum_{K\subseteq[n],|K|=k} \left(k + (k^2-k)\rho + k\mathbb{E}[X_1]\right) \mu(K).\end{aligned}$$

It is clear that $\sup_{\mu\in\mathcal{M}} C_\mu^2(\mathbf{X}_\rho)$ increases in ρ. Therefore, the minimum is achieved at $\rho^* = -1/(n-1)$, which implies that $\mathbf{X}_{\rho^*}$ is an NCD joint mix with correlation matrix P_n^*. As the value of (3.11) only depends on the correlation matrix, we have the desired result. □

Proof of Theorem 3.5. The "if" part is shown by Theorem 3.4 by choosing $\mathcal{M}$ as both (3.12) and (3.13) are special cases of (3.11). Next, we show the "only if" part by showing that the correlation matrix of the minimizer to (3.12) or (3.13) with any $k \in [n] \setminus \{1, n-1, n\}$ is P_n^*.

Without loss of generality, we assume that the variance of F is 1. As the mean of F is zero, we have $\mathbb{E}[(\sum_{i\in K} X_i)^2] = \operatorname{var}(\sum_{i\in K} X_i)$. By plugging an NCD joint mix $\mathbf{X}^{\mathrm{E}}$ with correlation matrix P_n^* into (3.12), the optimal value for (3.12) is

$$\max_{K\subseteq[n]} \mathbb{E}\left[\left(\sum_{i\in K} X_i\right)^2\right] = \max_{K\subseteq[n]} \operatorname{var}\left(\sum_{i\in K} X_i^{\mathrm{E}}\right) = \max_{k\in[n]} \frac{k(n-k)}{n-1} = \frac{k^*(n-k^*)}{n-1}, \tag{3.18}$$

where $k^* = \lfloor n/2 \rfloor$.

First, we consider the case $n = 3$. In this case, $[n] \setminus \{1, n-1, n\}$ is empty, and we only need to show that P_n^* is the unique correlation matrix of the minimizer to (3.12). Suppose

that $\mathbf{X}$ with covariance matrix Σ is a minimizer to (3.12). By (3.18), optimal value for (3.12) is 1. Hence,

$$\operatorname{var}\left(\sum_{i\in K} X_i\right) \leqslant 1 \qquad \text{for each } K \text{ with } |K| = 2,$$

and this implies

$$\sigma_{ij} \leqslant -1/2, \qquad \text{for } i \neq j. \tag{3.19}$$

Since Σ is positive semi-definite, we have $\sum_{i,j\in[3]} \sigma_{ij} \geqslant 0$, which leads to $3+2\sigma_{12}+2\sigma_{13}+2\sigma_{23} \geqslant 0$, implying $\sigma_{12}+\sigma_{13}+\sigma_{23} \geqslant -3/2$. Together with (3.19), we get $\sigma_{12} = \sigma_{13} = \sigma_{23} = -1/2$, and hence $\Sigma = P_n^*$.

Next, we consider the case $n \geqslant 4$. We first show that the correlation matrix of the minimizer to (3.13) is unique for $1 < k < n-1$. Fix $k \in [n] \setminus \{1, n-1, n\}$. Suppose that $\mathbf{X}$ with covariance matrix Σ is a minimizer to (3.13). Our goal is to show $\Sigma = P_n^*$.

Let K_ℓ, $\ell = 1, \ldots, n_k$, be all subsets of $[n]$ with cardinality k, where $n_k = \binom{n}{k}$. Then we have

$$\frac{1}{n_k}\sum_{\ell=1}^{n_k} \operatorname{var}\left(\sum_{i\in K_\ell} X_i\right) \geqslant k - \frac{k(k-1)}{n-1} = \frac{k(n-k)}{n-1}. \tag{3.20}$$

As $\mathbf{X}$ is a minimizer, for each K with $|K| = k$, we have

$$\operatorname{var}\left(\sum_{i\in K} X_i\right) = \mathbb{E}\left[\left(\sum_{i\in K} X_i\right)^2\right] \leqslant \max_{K\subseteq[n],\ |K|=k} \mathbb{E}\left[\left(\sum_{i\in K} X_i^{\mathrm{E}}\right)^2\right] = \frac{k(n-k)}{n-1}. \tag{3.21}$$

Combining (3.20) and (3.21), we have

$$\operatorname{var}\left(\sum_{i\in K} X_i\right) = \frac{k(n-k)}{n-1} \qquad \text{for each } K \text{ with } |K| = k.$$

Take $k = 2$. For any $i, j \in [n]$ such that $i \neq j$, the above equation implies

$$\operatorname{var}(X_i + X_j) = \operatorname{var}(X_i) + \operatorname{var}(X_j) + 2\operatorname{cov}(X_i, X_j) = \frac{2(n-2)}{n-1}.$$

As a result, we have $\operatorname{cov}(X_i, X_j) = -1/(n-1)$ for all $i, j \in [n]$ such that $i \neq j$. Hence, we conclude that $\Sigma = P_n^*$.

Finally, note that k^* in (3.18) satisfies $1 < k^* < n-1$ for $n \geqslant 4$. We have justified that the correlation matrix for optimizers to (3.13) with $k = k^*$ is unique. Therefore, by using (3.18), the correlation matrix for optimizers to (3.12) is also unique.

The above arguments show that, for $n \geqslant 3$, if $\mathbf{X}$ is a minimizer to (3.12), then the correlation matrix of $\mathbf{X}$ is P_n^*, which implies that $\mathbf{X}$ is an NCD joint mix with correlation matrix P_n^*. The same conclusion holds true if (3.12) is replaced by (3.13) with any $k \in [n] \setminus \{1, n-1, n\}$. □

Proof of Proposition 3.6. As the given marginals have zero means, for any $(Y_1, \dots, Y_n)$ with variance vector $(\sigma_1^2, \dots, \sigma_n^2)$,

$$\max_{K \subseteq [n]} \mathbb{E}\left[\left(\sum_{i \in K} Y_i\right)^2\right] = \max_{K \subseteq [n]} \operatorname{var}\left(\sum_{i \in K} Y_i\right) \geqslant \max_{i \in [n]} \operatorname{var}(Y_i) = \max_{i \in [n]} \sigma_i^2.$$

In case $n = 3$, a joint mix $\mathbf{X}$ with variance vector $(\sigma_1^2, \dots, \sigma_n^2)$ satisfies

$$\max_{K \subseteq [3],\ |K|=1} \operatorname{var}\left(\sum_{i \in K} X_i\right) = \max_{K \subseteq [3],\ |K|=2} \operatorname{var}\left(\sum_{i \in K} X_i\right) = \max_{i \in [3]} \operatorname{var}(X_i) = \max_{i \in [3]} \sigma_i^2,$$

and $\operatorname{var}(X_1 + X_2 + X_3) = 0$. Hence, the joint mix minimizes (3.12).

To show that no positive covariance is allowed, suppose that (X_1, X_2, X_3) is a minimizer to (3.12) and $\operatorname{cov}(X_i, X_j) > 0$ for some $i \neq j$. We have

$$\max_{K \subseteq [n]} \mathbb{E}\left[\left(\sum_{i \in K} X_i\right)^2\right] \geqslant \mathbb{E}\left[(X_i + X_j)^2\right] = \operatorname{var}(X_i + X_j) > \sigma_i^2 + \sigma_j^2 \geqslant \max\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right), \quad (3.22)$$

where the last inequality follows from the necessary condition (3.4) of the existence of an NCD joint mix. Since we have seen that the optimal value of (3.12) is $\max_{i \in [3]} \sigma_i^2$, (3.22) implies that (X_1, X_2, X_3) does not minimize (3.12). □

Proof of Lemma 3.1. One of the key properties of elliptical distributions is that they are closed under linear transformations, which is clear from (3.16). Hence, for $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$, the random variable $\sum_{i=1}^n X_i$ follows $\mathrm{E}_n(\mathbf{1}_n^\top \boldsymbol{\mu}, \mathbf{1}_n^\top \Sigma \mathbf{1}_n, \psi)$, which is degenerate if and only if $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$. Hence, $\mathbf{X}$ is a joint mix if and only if $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$ or $\psi = 1$ on $\mathbb{R}_+$. □

Proof of Proposition 3.7. Necessity follows from Proposition 3.3. To show sufficiency, let $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$, where Σ is the dispersion matrix of the multivariate Gaussian distribution constructed in the proof of Theorem 3.2. Then $X_i \sim \mathrm{E}_1(\mu_i, \sigma_i, \psi)$, $i \in [n]$. Moreover, it is checked in the proof of Theorem 3.2 that $\mathbf{1}_n^\top \Sigma \mathbf{1}_n = 0$ and $\sigma_{ij} \leqslant 0$ for $i, j \in [n]$ such that $i \neq j$. Therefore, $\mathbf{X}$ is the desired NCD joint mix. □

Proof of Theorem 3.6. The statement (i) immediately follows from the facts that ψ generates an n-dimensional elliptical distribution for every $n \in \mathbb{N}$ if and only if the corresponding elliptical class is a Gaussian variance mixture family (Fang et al., 1990, Section 2.6), and that $\mathrm{E}_n(\mathbf{0}_n, P_n^*, \psi)$ is an NCD joint mix, where P_n^* is an $n \times n$ matrix whose diagonal entries are 1 and off-diagonal entries are $-1/(n-1)$.

To show (ii), we need two lemmas.

Lemma 3.2 (Corollary 4 of Yin (2021))**.** *Let* $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ *and* $\mathbf{Y} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma', \psi)$ *be two elliptical distributions such that* $\Sigma = (\sigma_{ij})$ *and* $\Sigma' = (\sigma'_{ij})$ *satisfy* $\sigma_{ii} = \sigma'_{ii}$ *for all* $i \in [n]$*. Then* $\mathbf{X} \leqslant_{\mathrm{cU}} \mathbf{Y}$ *if and only if* $\sigma_{ij} \leqslant \sigma'_{ij}$ *for all* $i \neq j$*.*

Let $P_n^\perp$ be the identity matrix, which is the correlation matrix of an independent random vector. Although Lemma 3.2 implies that $\mathrm{E}_n(\boldsymbol{\mu}, P_n^*, \psi) \leqslant_{\mathrm{c}} \mathrm{E}_n(\boldsymbol{\mu}, P_n^\perp, \psi)$ for general elliptical distributions, $\mathrm{E}_n(\boldsymbol{\mu}, P_n^*, \psi)$ is not necessarily NOD in general since $\mathrm{E}_n(\mathbf{0}_n, P_n^\perp, \psi)$ does not have independent components. In fact, an elliptical distribution $\mathrm{E}_n(\mathbf{0}_n, P_n^\perp, \psi)$ is not NOD unless it is Gaussian.

Lemma 3.3. *The elliptical distribution* $\mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ *where* Σ *is diagonal is not NOD unless it is Gaussian.*

Proof. Assume that $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ is NOD. Since NOD is location invariant, it suffices to show the case when $\boldsymbol{\mu} = \mathbf{0}_n$. When $\mathbf{X}$ is NOD, then so is (X_1, X_2), that is,

$$\mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2) \leqslant \mathbb{P}(X_1 \leqslant x_1)\mathbb{P}(X_2 \leqslant x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$

Since (X_1, X_2) and $(-X_1, X_2)$ are identically distributed, we have

$$\mathbb{P}(X_1 \geqslant x_1, X_2 \leqslant x_2) \leqslant \mathbb{P}(X_1 \geqslant x_1)\mathbb{P}(X_2 \leqslant x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2,$$

and similarly, by symmetry,

$$\mathbb{P}(X_1 \geqslant x_1, X_2 \geqslant x_2) \leqslant \mathbb{P}(X_1 \geqslant x_1)\mathbb{P}(X_2 \geqslant x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2,$$

$$\mathbb{P}(X_1 \leqslant x_1, X_2 \geqslant x_2) \leqslant \mathbb{P}(X_1 \leqslant x_1)\mathbb{P}(X_2 \geqslant x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$

Adding the above four inequalities together, we get $1 \leqslant 1$. Hence, each of them is an equality. However, (X_1, X_2) follows a bivariate elliptical distribution with generator ψ, and thus X_1

and X_2 are not independent unless it is Gaussian; see Theorem 4.11 of Fang et al. (1990). Therefore, $\mathbf{X}$ cannot be NOD unless it is Gaussian. □

Now we are ready to prove Theorem 3.6. The "if" statement follows from Proposition 3.1. It remains to show the "only if" statement. Let ψ be a characteristic generator different from that of the Gaussian distribution. For $n \geqslant 2$, let $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \psi)$ be an NOD joint mix where Σ has positive diagonal entries. We start by observing from Lemma 3.2 that if $\sigma_{ij} > 0$ for $i \neq j$, then the bivariate projection (X_i, X_j) of $\mathbf{X}$ satisfies $(X_i, X_j) \geqslant_{\mathrm{c}} (X_i', X_j')$ where $(X_i', X_j') \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma_{ij}', \psi)$ with

$$\Sigma_{ij}' = \begin{pmatrix} \sigma_{ii} & 0 \\ 0 & \sigma_{jj} \end{pmatrix}.$$

Using Lemma 3.3, we know that (X_i', X_j') is not NOD, that is, there exists $(x_i, x_j) \in \mathbb{R}^2$ such that

$$\mathbb{P}(X_i' \leqslant x_i, X_j' \leqslant x_j) > \mathbb{P}(X_i' \leqslant x_i)\mathbb{P}(X_j' \leqslant x_j); \tag{3.23}$$

note that it suffices to consider the inequality needed for NLOD (not NUOD) by symmetry of the elliptical distribution and location invariance of NOD. Therefore, we have that

$$\mathbb{P}(X_i \leqslant x_i, X_j \leqslant x_j) > \mathbb{P}(X_i' \leqslant x_i, X_j' \leqslant x_j). \tag{3.24}$$

The two inequalities (3.23) and (3.24) imply that

$$\mathbb{P}(X_i \leqslant x_i, X_j \leqslant x_j) > \mathbb{P}(X_i \leqslant x_i)\mathbb{P}(X_j \leqslant x_j),$$

that is, (X_i, X_j) is not NOD. This leads to a contradiction.

Next, we assume $\sigma_{ij} \leqslant 0$ for all $i \neq j$. Since $\mathbf{a}^\top \Sigma \mathbf{a} \geqslant 0$ for all $\mathbf{a} \in \mathbb{R}^n$ and Σ has positive diagonal entries, we can take $\mathbf{a} = (1/\sqrt{\sigma_{11}}, \dots, 1/\sqrt{\sigma_{nn}})$, and this yields

$$\sum_{i,j=1}^{n} \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = n + \sum_{i \neq j} \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \geqslant 0.$$

Hence, there exist i, j with $i \neq j$ such that

$$\rho_{ij} := \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \geqslant -\frac{1}{n-1}.$$

Since NOD is location-scale invariant, the NOD of (X_i, X_j) implies that $\mathrm{E}_2(\mathbf{0}_2, P_{ij}, \psi)$ is NOD, where

$$P_{ij} = \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix}.$$

Taking a limit as $n \to \infty$, and noting that NOD is closed under weak convergence (Muller and Stoyan, 2002), we conclude that $\mathrm{E}_2(\mathbf{0}_2, P_2^{\perp}, \psi)$ is also NOD, which contradicts Lemma 3.3 if $\mathcal{E}(\psi)$ is not Gaussian.

Finally, by Proposition 3.1, this characterization result follows if NOD is replaced by NSD or NA. □

Chapter 4

Diversification quotients: Quantifying diversification via risk measures

4.1 Introduction

Portfolio diversification refers to investment strategies that spread out among many assets, usually with the hope to reduce the volatility or risk of the resulting portfolio. A mathematical formalization of diversification in a portfolio selection context was made by Markowitz (1952), and some early literature on diversification includes Sharpe (1964), Samuelson (1967), Levy and Sarnat (1970) and Fama and Miller (1972), amongst others.

Although diversification is conceptually simple, the question of how to measure diversification *quantitatively* is never well settled. An intuitive, but non-quantitative, approach is to simply count the number of distinct stocks or industries of substantial weight in the portfolio; see e.g., Green and Hollifield (1992), Denis et al. (2002) and DeMiguel et al. (2009) in different contexts. This approach is heuristic as it does not involve statistical or stochastic modeling. The second approach is to compute a quantitative index of the portfolio model, based on e.g., the volatility, variance, an expected utility, or a risk measure; this idea is certainly along the direction of Markowitz (1952). In addition, one may empirically address diversification by combining both approaches; see e.g., Tu and Zhou (2011) for the performance of different diversified portfolio strategies, D'Acunto et al. (2019) in the context of

robo-advising, and Berger and Eeckhoudt (2021) from the perspective of risk aversion and ambiguity aversion. Green and Hollifield (1992) studied conditions under which the two approaches are roughly in-line with each other.

In this chapter, we take the second approach by assigning a quantifier, called a *diversification index*, to each modeled portfolio. Carrying the idea of Markowitz (1952), we start our journey with a simple index, the diversification ratio (DR) based on the standard deviation (SD). For a random vector $\mathbf{X} = (X_1, \dots, X_n)$ representing future random losses and profits of individual components in a portfolio in one period,[1] DR based on SD is defined as

$$\mathrm{DR}^{\mathrm{SD}}(\mathbf{X}) = \frac{\mathrm{SD}\left(\sum_{i=1}^{n} X_i\right)}{\sum_{i=1}^{n} \mathrm{SD}(X_i)}; \tag{4.1}$$

see Choueifaty and Coignard (2008). One can also replace SD by variance. Intuitively, with a smaller value indicating a stronger diversification, the index $\mathrm{DR}^{\mathrm{SD}}$ quantifies the improvement of the portfolio SD over the sum of SD of its components, and it has several convenient properties. Nevertheless, it is well-known that SD is a coarse, non-monotone and symmetric measurement of risk, making it unsuitable for many risk management applications, especially in the presence of heavy-tailed and skewed loss distributions; see Embrechts et al. (2002) for thorough discussions.

Risk measures, in particular the Value-at-Risk (VaR) and the Expected Shortfall (ES), are more flexible quantitative tools, widely used in both financial institutions' internal risk management and banking and insurance regulatory frameworks, such as Basel III/IV (BCBS (2019)) and Solvency II (EIOPA (2011)). ES has many nice theoretical properties and satisfies the four axioms of coherence (Artzner et al. (1999)), whereas VaR is not subadditive in general, but it enjoys other practically useful properties; see Embrechts et al. (2014, 2018), Emmer et al. (2015) and the references therein for more discussions on the issues of VaR versus ES.

Some indices of diversification based on various risk measures have been proposed in the literature. For a given risk measure ϕ, an example of a diversification index is DR in (4.1) with SD replaced by ϕ; see Tasche (2007). For a review of diversification indices, see

[1] We focus on the one-period losses to establish the theory. This is consistent with the vast majority of literature on risk measures and decision models.

Koumou (2020). We find several demerits of DR built on a general risk measure ϕ such as VaR or ES in Section 4.2. A natural question is whether we can design a suitable index based on risk measures to quantify the magnitude of diversification, which avoids the deficiencies of DR. Answering this and related questions is the main purpose of this chapter.

We take an axiomatic approach to find our desirable diversification indices. Axiomatic approaches for risk and decision indices have been prolific in economic and statistical decision theories; see e.g., the recent discussions of Gilboa et al. (2019) and the monographs Gilboa (2009) and Wakker (2010). Closely related to diversification indices, risk measures (Artzner et al. (1999); Frittelli and Rosazza Gianin (2002); Föllmer and Schied (2016)) and acceptability indices (Cherny and Madan (2009)) also admit sound axiomatic foundation; the particular cases of VaR and ES are studied by Chambers (2009) and Wang and Zitikis (2021).

In Section 4.3, as our main contributions, we establish the first axiomatic foundation of diversification indices.[2] This axiomatic theory leads to the class of diversification quotients (DQs), the main object of this chapter, which have an interpretation parallel to DR. Six simple axioms—non-negativity, location invariance, scale invariance, rationality, normalization, and continuity—are introduced and justified for their desirability in quantifying diversification. Their interpretations are self-evident and they describe the basic requirements for a diversification index. In Theorem 4.1, these six axioms characterize DQ based on monetary and positive homogeneous risk measures. A seventh axiom of portfolio convexity, planting an intuitive ordering over portfolio weights in the index, further pins down DQ based on coherent risk measures in Theorem 4.2. Further, Proposition 4.1 gives conditions for which such DQ has the range of a standard interval. Portfolio convexity means that, with a given list of assets, combining a portfolio with a better-diversified one does not lead to worse diversification than the original portfolio, reflecting a fundamental principle in economics (Mas-Colell et al. (1995)). The financial interpretation of DQ is that it quantifies the improvement of a risk-level parameter (such as the parameter in VaR or ES) caused by pooling assets, and

[2] A different list of desirable axioms for diversification indices is studied by Koumou and Dionne (2022). Their framework is mathematically different from ours as their diversification indices are mappings of portfolio weights, instead of mappings of portfolio random vectors. They did not provide axiomatic characterization results.

this is discussed in Section 4.3.4.

A detailed analysis of the properties of DQ based on general risk measures is discussed in Section 4.4, which reveals that DQ has many appealing features, both theoretically and practically. In addition to standard operational properties (Proposition 4.2), DQ has intuitive behaviour for several benchmark portfolio scenarios (Theorem 4.3). Moreover, DQ allows for consistency with stochastic dominance (Proposition 4.3) and a fair comparison across portfolio dimensions (Proposition 4.4). We proceed to focus on VaR and ES in Section 4.5. It turns out that DQs based on VaR and ES have convenient alternative formulations (Theorem 4.4) and a natural range of $[0, n]$ and $[0, 1]$, respectively (Proposition 4.5). Further, they report intuitive comparisons between normal and t-models and it has the nice feature that it can capture heavy tails and common shocks.

In Section 4.6, efficient algorithms for DQs based on VaR and ES in portfolio optimization based on empirical observations are obtained (Proposition 4.6). Our new diversification index is applied to financial data in Section 4.7, where several empirical observations highlight the advantages of DQ. We conclude this chapter in Section 4.8 by discussing a number of implications and promising future directions for DQ. Some additional results, proofs, and some omitted numerical results are relegated to the E-Companion.

Notation. Throughout this chapter, $(\Omega, \mathcal{F}, \mathbb{P})$ is an atomless probability space, on which almost surely equal random variables are treated as identical. A risk measure ϕ is a mapping from $\mathcal{X}$ to $\mathbb{R}$, where $\mathcal{X}$ is a convex cone of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ representing losses faced by a financial institution or an investor (i.e., a sign flip from Artzner et al. (1999)), and $\mathcal{X}$ is assumed to include all constants (i.e., degenerate random variables). For $p \in (0, \infty)$, denote by $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables X with $\mathbb{E}[|X|^p] < \infty$ where $\mathbb{E}$ is the expectation under $\mathbb{P}$. Furthermore, $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all (essentially) bounded random variables, and $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all random variables. Write $X \sim F$ if the random variable X has the distribution function F under $\mathbb{P}$, and $X \overset{\mathrm{d}}{=} Y$ if two random variables X and Y have the same distribution. Further, denote by $\mathbb{R}_+ = [0, \infty)$ and $\overline{\mathbb{R}} = [-\infty, \infty]$. Terms such as increasing or decreasing functions are in the non-strict sense. For $X \in L^0$, ess-sup(X) and ess-inf(X) are the essential supremum and the essential infimum of X, respectively. Let n be a fixed positive integer representing the number of assets in a

portfolio, and write $[n] = \{1, \dots, n\}$. It does not hurt to think about $n \geqslant 2$ although our results hold also (trivially) for $n = 1$. The vector $\mathbf{0}$ represents the n-vector of zeros, and we always write $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_n)$.

4.2 Preliminaries and motivation

The main object of this chapter, a *diversification index* D is a mapping from $\mathcal{X}^n$ to $\overline{\mathbb{R}}$, which is used to quantify the magnitude of diversification of a risk vector $\mathbf{X} \in \mathcal{X}^n$ representing portfolio losses. Our convention is that a smaller value of $D(\mathbf{X})$ represents a stronger diversification in a sense specified by the design of D.

As the evaluation of diversification is closely related to that of risk, diversification indices in the literature are often defined through risk measures. An example of a diversification index is the diversification ratio (DR) mentioned in the Introduction based on measures of variability such as the standard deviation (SD) and variance (var):

$$\mathrm{DR}^{\mathrm{SD}}(\mathbf{X}) = \frac{\mathrm{SD}\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \mathrm{SD}(X_i)} \quad \text{and} \quad \mathrm{DR}^{\mathrm{var}}(\mathbf{X}) = \frac{\mathrm{var}\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \mathrm{var}(X_i)},$$

with the convention $0/0 = 0$. We refer to Rockafellar et al. (2006), Furman et al. (2017) and Bellini et al. (2022) for general measures of variability. DRs based on SD and var satisfy the three simple properties below, which can be easily checked.

[+] Non-negativity: $D(\mathbf{X}) \geqslant 0$ for all $\mathbf{X} \in \mathcal{X}^n$.

[LI] Location invariance: $D(\mathbf{X} + \mathbf{c}) = D(\mathbf{X})$ for all $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ and all $\mathbf{X} \in \mathcal{X}^n$.

[SI] Scale invariance: $D(\lambda \mathbf{X}) = D(\mathbf{X})$ for all $\lambda > 0$ and all $\mathbf{X} \in \mathcal{X}^n$.

The first property, [+], simply means that diversification is measured in non-negative values, where 0 typically represents a fully diversified or hedged portfolio (in some sense). The property [LI] means that injecting constant losses or gains to components of a portfolio, or changing the initial price of assets in the portfolio,[3] does not affect its diversification index. The property [SI] means that rescaling a portfolio does not affect its diversification

[3] Recall that X_i represents the loss from asset i. Suppose that two agents purchased the same portfolio of

index. The latter two properties are arguably natural, although they are not satisfied by some diversification indices used in the literature (see (4.2) below). A diversification index satisfying both [LI] and [SI] is called location-scale invariant.

Next, we define the two popular risk measures in banking and insurance practice. The VaR at level $\alpha \in [0,1)$ is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1-\alpha\}, \quad X \in L^0,$$

and the ES (also called CVaR, TVaR or AVaR) at level $\alpha \in (0,1)$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\,\mathrm{d}\beta, \quad X \in L^1,$$

and $\mathrm{ES}_0(X) = \text{ess-sup}(X) = \mathrm{VaR}_0(X)$, which may be ∞. The probability level α above is typically very small, e.g., 0.01 or 0.025 in BCBS (2019); note that we use the "small α" convention. Artzner et al. (1999) introduced *coherent* risk measures $\phi : \mathcal{X} \to \mathbb{R}$ as those satisfying the following four properties.

[M] Monotonicity: $\phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ with $X \leqslant Y$.[4]

[CA] Constant additivity: $\phi(X+c) = \phi(X) + c$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

[PH] Positive homogeneity: $\phi(\lambda X) = \lambda\phi(X)$ for all $\lambda \in (0,\infty)$ and $X \in \mathcal{X}$.

[SA] Subadditivity: $\phi(X+Y) \leqslant \phi(X) + \phi(Y)$ for all $X, Y \in \mathcal{X}$.

ES satisfies all four properties above, whereas VaR does not satisfy [SA]. We say that a risk measure is *monetary* if it satisfies [CA] and [M], and it is *MCP* if it satisfies [M], [CA] and [PH]. For discussions and interpretations of these properties, we refer to Föllmer and Schied (2016).

Some diversification indices are defined via risk measures, such as DR (e.g., Bürgi et al. (2008), Mainik and Rüschendorf (2010) and Embrechts et al. (2015)) and the diversification

assets but at different prices of each asset. Denote by $\mathbf{X}$ the portfolio loss vector of agent 1. The portfolio loss vector of agent 2 is $\mathbf{X} + \mathbf{c}$, where $\mathbf{c}$ is the vector of differences between their purchase prices. The two agents should have the same level of diversification regardless of their purchase prices, as they hold the same portfolio.

[4]The inequality $X \leqslant Y$ between two random variables X and Y is pointwise.

benefit (DB, e.g., Embrechts et al. (2009) and McNeil et al. (2015)). For a risk measure ϕ, DR and DB based on ϕ are defined as[5]

$$\mathrm{DR}^{\phi}(\mathbf{X}) = \frac{\phi\left(\sum_{i=1}^{n} X_i\right)}{\sum_{i=1}^{n} \phi(X_i)} \quad \text{and} \quad \mathrm{DB}^{\phi}(\mathbf{X}) = \sum_{i=1}^{n} \phi(X_i) - \phi\left(\sum_{i=1}^{n} X_i\right). \tag{4.2}$$

In contrast to DR, a larger value of DB represents a stronger diversification, but this convention can be easily modified by flipping the sign to consider $-\mathrm{DB}^{\phi}$. By definition, DR is the ratio of the pooled risk to the sum of the individual risks, and thus a measurement of how substantially pooling reduces risk; similarly, DB measures the difference instead of the ratio.

DR has a number of deficiencies. First, the value of DR^{ϕ} is not necessarily non-negative, violating [+]. Since the risk measure ϕ may take negative values,[6] it would be difficult to interpret the case where either the numerator or denominator in DR is negative, and this makes optimization of DR troublesome. An example is a portfolio of credit default losses, where VaR of individual losses is often 0 or negative but VaR of the portfolio loss is positive; see McNeil et al. (2015, Example 2.25). Second, for common risk measures, DR violates [LI], meaning that adding a risk-free asset changes the value of DR. Third, DR is not necessarily quasi-convex in portfolio weights; this point is more subtle and will be explained later. In addition to the above drawbacks, we also find that DR has wrong incentives for some simple models; for instance, it suggests that an iid portfolio of t-distributed risks is less diversified than a portfolio with a common shock and the same marginals; see Section 4.5.2 for details. Similarly to DR, the index DB satisfies [LI] for ϕ satisfying [CA], but it does not satisfy [SI] for common risk measures, and it may take both positive and negative values.

In financial applications, the risk measures VaR and ES are specified in regulatory documents such as BCBS (2019) and EIOPA (2011), and therefore it is beneficial to stick to VaR or ES as the risk measure when assessing diversification. Both $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ satisfy [SI], but they do not satisfy [+] or [LI].[7] It remains unclear how one can define a

[5] If the denominator in the definition of $\mathrm{DR}^{\phi}(\mathbf{X})$ is 0, then we use the convention $0/0 = 0$ and $1/0 = \infty$.

[6] A negative value of a risk measure has a concrete meaning as the amount of capital to be withdrawn from a portfolio position while keeping it acceptable; see Artzner et al. (1999).

[7] An impossibility result (Proposition 5.1) is presented in Section 5.3, which suggests that it is not possible to construct non-trivial diversification indices like DR and DB satisfying [+], [LI] and [SI].

diversification index based on VaR or ES satisfying these properties. In the remainder of this chapter, we will introduce and study a new index of diversification to bridge this gap.

4.3 Diversification indices: An axiomatic theory

In this section, we fix $\mathcal{X} = L^\infty$ as the standard choice in the literature of axiomatic theory of risk measures. In addition to [+], [LI] and [SI] introduced in Section 4.2, we propose four new axioms. The first six axioms together characterize a new class of diversification indices, that is, diversification quotients (DQ) based on MCP risk measures. With the seventh axiom of portfolio convexity, we further pin down the class of DQ based on coherent risk measures.

4.3.1 Axioms of rationality, normalization, and continuity

We first present three axioms, which depend on a risk measure ϕ. These three axioms are standard and weak in the sense that they do not impose a specific functional structure on D other than some forms of monotonicity, normalization, and continuity.

For a risk measure ϕ, we say that two vectors $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ are *ϕ-marginally equivalent* if $\phi(X_i) = \phi(Y_i)$ for each $i \in [n]$, and we denote this by $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$. In other words, if an agent evaluates risks using the risk measure ϕ, then she would view the individual components of $\mathbf{X}$ and those of $\mathbf{Y}$ as equally risky. Similarly, denote by $\mathbf{X} \overset{\mathrm{m}}{\succeq} \mathbf{Y}$ if $\phi(X_i) \leqslant \phi(Y_i)$ for each $i \in [n]$, and by $\mathbf{X} \overset{\mathrm{m}}{\succ} \mathbf{Y}$ if $\phi(X_i) < \phi(Y_i)$ for each $i \in [n]$. The other three desirable axioms are presented below, and they are built on a given risk measure ϕ, such as VaR or ES, typically specified exogenously by financial regulation.

$[\mathrm{R}]_\phi$ Rationality: $D(\mathbf{X}) \leqslant D(\mathbf{Y})$ for $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ satisfying $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$ and $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$.

To interpret the axiom $[\mathrm{R}]_\phi$, consider two portfolios $\mathbf{X}$ and $\mathbf{Y}$ satisfying $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$. If further $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$ holds, then the total loss from portfolio $\mathbf{X}$ is always less or equal to that from portfolio $\mathbf{Y}$, making the portfolio $\mathbf{X}$ safer than $\mathbf{Y}$. Since the individual components

in $\mathbf{X}$ and those in $\mathbf{Y}$ are equally risky, the fact that $\mathbf{X}$ is safer in aggregation is a result of the different diversification effects in $\mathbf{X}$ and $\mathbf{Y}$, leading to the inequality $D(\mathbf{X}) \leqslant D(\mathbf{Y})$. This axiom is called rationality because a rational agent always prefers to have smaller losses.

Next, we formulate our idea about normalizing representative values of the diversification index. First, we assign the zero portfolio $\mathbf{0}$ the value $D(\mathbf{0}) = 0$, as it carries no risk in every sense.[8] A natural benchmark of a non-diversified portfolio is one in which all components are the same. Such a portfolio $\mathbf{X}^{\mathrm{du}} = (X, \dots, X)$ will be called a *duplicate* portfolio, and we may, ideally, wish to assign the value $D(\mathbf{X}^{\mathrm{du}}) = 1$. However, since the zero portfolio $\mathbf{0}$ is also duplicate but $D(\mathbf{0}) = 0$, we will require the weaker assumption $D(\mathbf{X}^{\mathrm{du}}) \leqslant 1$ for duplicate portfolios.[9] Lastly, we should understand for what portfolios $D(\mathbf{X}) \geqslant 1$ needs to occur. We say that a portfolio $\mathbf{X}^{\mathrm{wd}} = (X_1, \dots, X_n)$ is *worse than duplicate*, if there exists a duplicate portfolio $\mathbf{X}^{\mathrm{du}} = (X, \dots, X)$ such that $\mathbf{X}^{\mathrm{wd}} \overset{\mathrm{m}}{\succ} \mathbf{X}^{\mathrm{du}}$ and $\sum_{i=1}^{n} X_i \geqslant nX$. Intuitively, this means that each component of $\mathbf{X}^{\mathrm{wd}}$ is strictly less risky than X, but putting them together always incurs a larger loss than nX; in this case, diversification creates nothing but a penalty to the risk manager, and we assign $D(\mathbf{X}^{\mathrm{wd}}) \geqslant 1$.[10] Existence of worse-than-duplicate portfolios is characterized in Section 5.4.1. Putting all of the considerations above, we propose the following normalization axiom.

$[\mathrm{N}]_\phi$ Normalization: $D(\mathbf{0}) = 0$, $D(\mathbf{X}) \leqslant 1$ if $\mathbf{X}$ is duplicate, and $D(\mathbf{X}) \geqslant 1$ if $\mathbf{X}$ is worse than duplicate.

Finally, we propose a continuity axiom which is mainly for technical convenience.

$[\mathrm{C}]_\phi$ Continuity: For $\{\mathbf{Y}^k\}_{k\in\mathbb{N}} \subseteq \mathcal{X}^n$ and $\mathbf{X} \in \mathcal{X}^n$ satisfying $\mathbf{Y}^k \overset{\mathrm{m}}{\simeq} \mathbf{X}$ for each k, if $(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i^k)_+ \xrightarrow{L^\infty} 0$ as $k \to \infty$, then $(D(\mathbf{X}) - D(\mathbf{Y}^k))_+ \to 0$.

The axiom $[\mathrm{C}]_\phi$ is a special form of semi-continuity. To interpret it, consider portfolios $\mathbf{X}$ and $\mathbf{Y}$ that are marginally equivalent. If the sum of components of $\mathbf{X}$ is not much worse than

[8] Indeed, the value of $D(\mathbf{0})$ may be rather arbitrary; this is the case for DR where 0/0 occurs.

[9] Theorem 4.3 gives some mild conditions that yield $D(\mathbf{X}^{\mathrm{du}}) = 1$ for the class D characterized in this section.

[10] Such situations may be regarded as diversification disasters; see Ibragimov et al. (2011).

that of $\mathbf{Y}$ in L^∞, then the axiom says that the diversification of $\mathbf{X}$ is not much worse than the diversification of $\mathbf{Y}$. This property allows for a special form of stability or robustness[11] with respect to statistical errors when estimating the distributions of portfolio losses.

One can check that the axioms $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ and $[\mathrm{C}]_\phi$ are satisfied by $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ if we only consider positive portfolio vectors. The axioms are not satisfied by $\mathrm{DR}^{\mathrm{SD}}$ because SD is not monotone and hence the inequalities $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$ and $\sum_{i=1}^n X_i \geqslant nX$ used in $[\mathrm{R}]_\phi$ and $[\mathrm{N}]_\phi$ are not relevant for SD.

4.3.2 Portfolio convexity

The next axiom, different from the three above, imposes a natural form of convexity on the diversification index. Portfolio diversification is intrinsically connected to convexity of ordering relations. Quoting Mas-Colell et al. (1995, p. 44), "Convexity can also be viewed as the formal expression of a basic inclination of economic agents for diversification." For this purpose, we propose an axiom of portfolio convexity in this section.

Let a random vector $\mathbf{X} \in \mathcal{X}^n$ represent losses from n assets and a vector $\mathbf{w} = (w_1, \dots, w_n) \in \Delta_n$ of portfolio weights, where Δ_n is the standard n-simplex, given by

$$\Delta_n = \{\mathbf{x} \in [0,1]^n : x_1 + \cdots + x_n = 1\}.$$

The total loss of the portfolio is $\mathbf{w}^\top \mathbf{X}$. We write $\mathbf{w} \odot \mathbf{X} = (w_1 X_1, \dots, w_n X_n)$, which is the portfolio loss vector with the weight $\mathbf{w}$. The *portfolio convexity* axiom is formulated below.

[PC] Portfolio convexity: The set $\{\mathbf{w} \in \Delta_n : D(\mathbf{w} \odot \mathbf{X}) \leqslant d)\}$ is convex for each $\mathbf{X} \in \mathcal{X}^n$ and $d \in \overline{\mathbb{R}}$.

Intuitively, portfolio convexity means that, for a given vector $\mathbf{X}$ of assets, combining a portfolio strategy with a better-diversified one on the same set of assets does not result in a portfolio that is less diversified than the original portfolio. As convexity is the decision-theoretic counterpart of diversification, [PC] is desirable for diversification indices.

[11] In the literature of statistical robustness, often a different metric than the L^∞ metric is used; see Huber and Ronchetti (2009) for a general treatment. Our choice of formulating continuity via the L^∞ metric is standard in the axiomatic theory of risk mappings on L^∞.

Remark 4.1. Axiom [PC] is equivalent to *quasi-convexity* of $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$ for each $\mathbf{X} \in \mathcal{X}^n$; that is, $D((\lambda \mathbf{w} + (1-\lambda)\mathbf{w}') \odot \mathbf{X}) \leqslant D(\mathbf{w} \odot \mathbf{X}) \vee D(\mathbf{w}' \odot \mathbf{X})$ for all $\lambda \in [0,1]$, $\mathbf{w}, \mathbf{w}' \in \Delta_n$ and $\mathbf{X} \in \mathcal{X}^n$.

Remark 4.2. Convexity or quasi-convexity of $\mathbf{X} \mapsto D(\mathbf{X})$ is not natural or desirable. For instance, combining two diversified portfolios (X, Y) and (Y, X) may result in a duplicate portfolio; see Example 5.1 in Section 5.4.2. Convexity of $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$, which is stronger than [PC], is not desirable either; see Example 5.2 in Section 5.4.2.

The four axioms introduced above, together with the three in Section 4.2, lead to a class of diversification indices, which we define next.

4.3.3 Characterization results

We first formally introduce the diversification index DQ relying on a parametric class of risk measures, which will be characterized in two results below.

Definition 4.1. Let $\rho = (\rho_\alpha)_{\alpha \in I}$ be a class of risk measures indexed by $\alpha \in I = (0, \overline{\alpha})$ with $\overline{\alpha} \in (0, \infty]$ such that ρ_α is decreasing in α. For $\mathbf{X} \in \mathcal{X}^n$, the *diversification quotient* based on the class ρ at level $\alpha \in I$ is defined by

$$\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \frac{\alpha^*}{\alpha}, \quad \text{where } \alpha^* = \inf\left\{ \beta \in I : \rho_\beta\left(\sum_{i=1}^{n} X_i\right) \leqslant \sum_{i=1}^{n} \rho_\alpha(X_i) \right\}, \tag{4.3}$$

with the convention $\inf(\varnothing) = \overline{\alpha}$.

We first characterize DQ based on MCP risk measures by six axioms without [PC].

Theorem 4.1. *A diversification index $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ satisfies* [+], [LI], [SI], $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ *and* $[\mathrm{C}]_\phi$ *for some MCP risk measure ϕ if and only if D is* $\mathrm{DQ}^{\rho}_{\alpha}$ *for some α and decreasing class ρ of MCP risk measures. Moreover, in both directions of the above equivalence, it can be required that $\rho_\alpha = \phi$.*

Theorem 4.1 gives the first axiomatic characterization of diversification indices, to the best of our knowledge. The proof techniques to show the important "only if" statement of

Theorem 4.1 are based on a sophisticated analysis of an auxiliary mapping

$$R : \mathcal{X} \to [0, \infty], \ R(X) = \inf \left\{ D(\mathbf{X}) : X \leqslant \sum_{i=1}^{n} X_i, \ \mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{0} \right\},$$

and this is explained in Section 5.2.

Next, we incorporate portfolio convexity into our axiomatic framework. For this purpose, it is natural to build the diversification indices based on risk measures with convexity. When formulated on monetary risk measures, convexity represents the idea that diversification reduces the risk; see Föllmer and Schied (2016). For risk measures that are not constant additive, Cerreia-Vioglio et al. (2011) argued that quasi-convexity is more suitable than convexity to reflect the consideration of diversification; moreover, convexity and quasi-convexity are equivalent if [CA] holds. A risk measure is *linear* if it satisfies $\phi(aX+bY) = a\phi(X)+b\phi(Y)$ for all $X, Y \in \mathcal{X}$ and $a, b \in \mathbb{R}$. Since linear risk measures correspond to expectations (under monotonicity), which do not reflect diversification, we will focus on non-linear ones. The next theorem characterizes DQ based on coherent risk measures.

Theorem 4.2. *Suppose $n \geqslant 4$ and ϕ is a non-linear coherent risk measure. A diversification index $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ satisfies* [+], [LI], [SI], $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$, $[\mathrm{C}]_\phi$ *and* [PC] *if and only if D is DQ^ρ_α for some α and decreasing class ρ of coherent risk measures with $\rho_\alpha = \phi$.*

The conditions $n \geqslant 4$ and non-linearity of ϕ are essential to the proof of Theorem 4.2. They are harmless for financial applications since typical portfolios have more than a few components, and common risk measures are not linear.

Although portfolio convexity is crucial for diversification indices, making Theorem 4.2 a central result, we present Theorem 4.1 separately for the following reasons. First, Theorem 4.1 reveals the fundamental properties needed to pin down the form of DQ and this helps to clarify the role of [PC]. Second, the proof of Theorem 4.2 is technically built on Theorem 4.1. Third, the class of DQ characterized by Theorem 4.1 allows for DQ based on VaR, which is popular in financial regulation.

In the next proposition, we show that for sub-linear risk measures, DQ satisfies [PC] (thus, the "if" direction of Theorem 4.2 does not need [M] and [CA] for ρ), and its range is $[0, 1]$ under mild conditions, avoiding non-degeneracy. A risk measure is sub-linear if it

Table 4.1: Summary of axioms satisfied by diversification indices DR^{ϕ}, DB^{ϕ} and $\mathrm{DQ}^{\rho}_{\alpha}$ (with $\phi = \rho_\alpha$), where $\mathcal{X}_+$ is the set of non-negative elements in $\mathcal{X}$ and $\alpha \in (0,1)$

Index	Domain	[+]	[LI]	[SI]	$[\mathrm{R}]_\phi$	$[\mathrm{N}]_\phi$	$[\mathrm{C}]_\phi$	[PC]
$\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$	$\mathcal{X}^n$	×	×	√	×	×	×	×
$\mathrm{DR}^{\mathrm{VaR}_\alpha}$	$\mathcal{X}_+^n$	√	×	√	√	√	√	×
$\mathrm{DR}^{\mathrm{ES}_\alpha}$	$\mathcal{X}_+^n$	√	×	√	√	√	√	√
$\mathrm{DR}^{\mathrm{SD}}$	$\mathcal{X}^n$	√	√	√	×	×	×	√
$\mathrm{DR}^{\mathrm{var}}$	$\mathcal{X}^n$	√	√	√	×	×	×	×
$-\mathrm{DB}^{\mathrm{VaR}_\alpha}$	$\mathcal{X}^n$	×	√	×	√	×	√	×
$-\mathrm{DB}^{\mathrm{ES}_\alpha}$	$\mathcal{X}^n$	×	√	×	√	×	√	√
$\mathrm{DQ}^{\mathrm{VaR}}_\alpha$	$\mathcal{X}^n$	√	√	√	√	√	√	×
$\mathrm{DQ}^{\mathrm{ES}}_\alpha$	$\mathcal{X}^n$	√	√	√	√	√	√	√

satisfies subadditivity and positive homogeneity (equivalently, convexity and positive homogeneity).

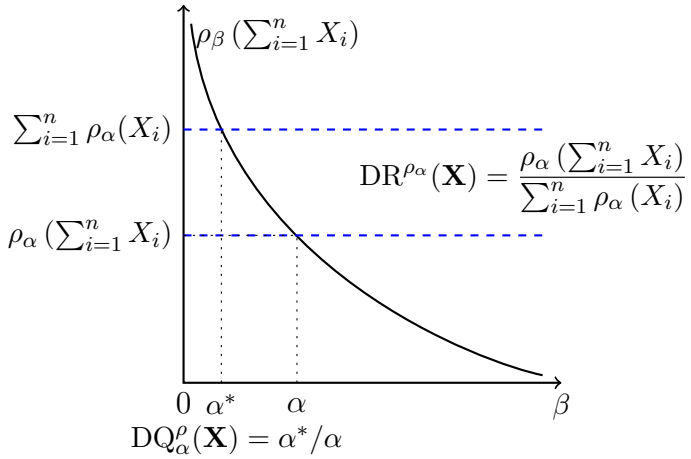
Proposition 4.1. *Let $\rho = (\rho_\beta)_{\beta\in I}$ be a decreasing class of sub-linear risk measures and $\alpha \in I$. Then $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies* [PC]. *If $n \geqslant 3$, ρ_α is non-linear and there exists $X \in \mathcal{X}$ such that $\beta \mapsto \rho_\beta(X)$ is strictly decreasing, then $\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0,1]$.*

Given a sub-linear risk measure ρ_α, the conditions in Proposition 4.1 for $\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0,1]$ are mild and satisfied by e.g., DQ based on the family of ES. In contrast to DQ, DR based on sub-linear risk measures may not satisfy [PC] since the denominator in (4.2) may be negative. For a clear comparison, we summarize in Table 4.1 the axioms satisfied by the diversification indices that appear in this chapter.

4.3.4 Interpretation of DQ

DQ based on MCP or coherent risk measures have been characterized axiomatically, but we have not interpreted the meaning of DQ in (4.3). For an interpretation, consider a decreasing class of risk measures $(\rho_\beta)_{\beta\in I}$. The values of risk measures typically represent the

Figure 4.1: Conceptual symmetry between DQ and DR



capital requirement of a risky asset, and hence β is interpreted as a parameter of risk level (as in VaR_β or ES_β), that is, a smaller β means a larger capital requirement for the same risk. Notice from (4.3) that, under mild conditions, α^* is uniquely determined by

$$\rho_{\alpha^*}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \rho_\alpha(X_i).$$

Therefore, α^* is the parameter of risk level achieved by pooling, assuming that the portfolio maintains the same total capital requirement assessed by ρ_α when there is no pooling, that is, $\sum_{i=1}^n \rho_\alpha(X_i)$. As $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = \alpha^*/\alpha$, DQ is the ratio of the risk-level parameters before and after pooling. To summarize,

the index DQ quantifies the improvement of the risk-level parameter caused by pooling assets.

In the most relevant case $\rho_\alpha\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \rho_\alpha(X_i)$, we present in Figure 4.1 the conceptual symmetry between DQ, which measures the improvement by pooling in the horizontal direction, and DR, which measures an improvement in the vertical direction. In particular, in the case of VaR, DQ measures the probability improvement, whereas DR measures the quantile improvement; see Theorem 4.4 and (4.7) below.

Remark 4.3. The idea of improvement of risk level is closely related to acceptability indices, proposed by Cherny and Madan (2009). More precisely, an acceptability index for a loss

$X \in \mathcal{X}$ is defined by $\mathrm{AI}^{\rho}(X) = \sup\{\gamma \in \mathbb{R}_+ : \rho_{1/\gamma}(X) \leqslant 0\}$ for a decreasing class of coherent risk measures $(\rho_\gamma)_{\gamma\in\mathbb{R}_+}$, which has visible similarity to α^* in (4.3); see Kováčová et al. (2020) for optimization of acceptability indices. If ρ is a class of risk measures satisfying [CA], then

$$\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \frac{1}{\alpha}\left(\mathrm{AI}^{\rho}\left(\sum_{i=1}^{n}(X_i - \rho_\alpha(X_i))\right)\right)^{-1}.$$

Dhaene et al. (2012) studied several methods for capital allocation, among which the quantile allocation principle computes a capital allocation $(C_1, \dots, C_n)$ such that $\sum_{i=1}^n C_i = \mathrm{VaR}_\alpha(\sum_{i=1}^n X_i)$ and $C_i = \mathrm{VaR}_{c\alpha}(X_i)$ for some $c > 0$. The constant c appearing as a nuisance parameter in the above rule has a visible mathematical similarity to $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$. Mafusalov and Uryasev (2018) studied the so-called buffered probability of exceedance, which is the inverse of the ES curve $\beta \mapsto \mathrm{ES}_\beta(X)$ at a specific point $x \in \mathbb{R}$; note that α^* in (4.3) is obtained by inverting the ES curve $\beta \mapsto \mathrm{ES}_\beta(\sum_{i=1}^n X_i)$ at $\sum_{i=1}^n \mathrm{ES}_\alpha(X_i)$.

We close the section with discussions on the construction of DQ. First, DQ can be constructed from any monotonic parametric family of risk measures. All commonly used risk measures belong to a monotonic family, as this includes VaR, ES, expectiles (e.g., Bellini et al. (2014)), mean-variance (e.g., Markowitz (1952) and Maccheroni et al. (2009)), and entropic risk measures (e.g., Föllmer and Schied (2016)); some choices do not guarantee all axioms to hold. Our results imply that using ES or expectiles guarantees all axioms and non-degeneracy for DQ. In addition, there are ways to construct DQ from a single risk measure; see Section 5.4.3. DQ can also be axiomatized using preferences instead of risk measures; see Section 5.4.4.

DQ can be used as a normative tool for measuring diversification. In this context, the choice of the parametric family of risk measures is up to the user, and DQ serves as a versatile tool that accommodates various risk attitudes. The choice of risk measures (e.g., VaR, ES) and the determination of the confidence level (α) should be aligned with the risk tolerance, objectives, and regulatory requirements of the decision maker. For instance, conservative investors, prioritizing capital preservation, may gravitate towards the family of ES at a high level α, which reflects an assessment of downside risk, whereas those with aggressive risk preferences may opt for VaR or ES at a lower level α. Most generally, we would recommend the use of DQ based on ES, which has a natural and strong connection to financial regulation

and tail risk management, and the parameter α allows for flexibility in the assessment of tail risk.

4.4 Properties of DQ

In this section, we study the properties of DQ defined in Definition 4.1. For the greatest generality, we do not impose any properties of risk measures in the decreasing family $\rho = (\rho_\alpha)_{\alpha\in I}$, i.e., the family ρ is not limited to MCP or coherent risk measures, so that our results can be applied to more flexible contexts in which some of the seven axioms are relaxed. In this section, $\mathcal{X}$ is not restricted to L^∞.

4.4.1 Basic properties

We first make a few immediate observations by the definition of DQ. From (4.3), we can see that computing DQ^ρ_α is to invert the decreasing function $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$ at $\sum_{i=1}^n \rho_\alpha(X_i)$. For the cases of VaR and ES, $I = (0,1)$, $\alpha^* \in [0,1]$, and DQ has simple formulas; see Theorem 4.4 in Section 4.5. For a fixed value of $\sum_{i=1}^n \rho_\alpha(X_i)$, DQ is larger if the curve $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$ is larger, and DQ is smaller if the curve $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$ is smaller. This is consistent with our intuition that a diversification index is large if there is little or no diversification, thus a large value of the portfolio risk, and a diversification index is small if there is strong diversification.

In Theorem 4.1, we have seen that DQ satisfies [SI] and [LI] if ρ is a class of MCP risk measures. These properties of DQ can be obtained based on a more general version of properties [CA] and [PH] of risk measures, allowing us to include SD and the variance. The results are summarized in Proposition 4.2, which are straightforward to check by definition.

$[\mathrm{CA}]_m$ Constant additivity with $m \in \mathbb{R}$: $\phi(X + c) = \phi(X) + mc$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

$[\mathrm{PH}]_\gamma$ Positive homogeneity with $\gamma \in \mathbb{R}$: $\phi(\lambda X) = \lambda^\gamma \phi(X)$ for all $\lambda \in (0,\infty)$ and $X \in \mathcal{X}$.

Proposition 4.2. *Let $\rho = (\rho_\alpha)_{\alpha\in I}$ be a class of risk measures decreasing in α. For each $\alpha \in I$,*

(i) if ρ_β *satisfies* $[\mathrm{PH}]_\gamma$ *with the same* γ *across* $\beta \in I$*, then* DQ^ρ_α *satisfies* [SI].

(ii) if ρ_β *satisfies* $[\mathrm{CA}]_m$ *with the same* m *across* $\beta \in I$*, then* DQ^ρ_α *satisfies* [LI].

(iii) if ρ_α *satisfies* [SA]*, then* DQ^ρ_α *takes value in* $[0, 1]$.

It is clear that [CA] is $[\mathrm{CA}]_m$ with $m = 1$ and [PH] is $[\mathrm{PH}]_\gamma$ with $\gamma = 1$. More properties of DQs on the important families of VaR and ES will be discussed in Section 4.5. In particular, we will see that the ranges of $\mathrm{DQ}^{\mathrm{VaR}}_\alpha$ and $\mathrm{DQ}^{\mathrm{ES}}_\alpha$ are $[0, n]$ and $[0, 1]$, respectively.

Example 4.1 (Liquidity and temporal consistency)**.** In risk management practice, liquidity and time-horizon of potential losses need to be taken into account; see BCBS (2019, p.89). If liquidity risk is of concern, one may use a risk measure with $[\mathrm{PH}]_\gamma$ with $\gamma > 1$ to penalize large exposures of losses. For such risk measures, DQ^ρ_α remains scale invariant, as shown by Proposition 4.2. On the other hand, if the risk associated to the loss $\mathbf{X}(t)$ at different time spots $t > 0$ is scalable by a function $f > 0$ (usually of the order $f(t) = \sqrt{t}$ in standard models such as the Black-Scholes), then DQ is consistent across different horizons in the sense that $\mathrm{DQ}^\rho_\alpha(\mathbf{X}(t)) = \mathrm{DQ}^\rho_\alpha(\mathbf{X}(s))$ for two time spots $s, t > 0$, given that $\rho_\beta(X_i(t)) = f(t)\rho_\beta(X_i(1))$ for $i \in [n]$, $t > 0$ and $\beta \in I$.

Next, we explain that the values taken by DQ are consistent with our usual perceptions of portfolio diversification. For a given risk measure ϕ and a portfolio risk vector $\mathbf{X}$, we consider the following three situations which yield intuitive values of DQ.

(i) There is no insolvency risk with pooled individual capital, i.e., $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n \phi(X_i)$ a.s.;

(ii) diversification benefit exists, i.e., $\phi\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \phi(X_i)$;

(iii) the portfolio relies on a single asset, i.e., $\mathbf{X} = (\lambda_1 X, \dots, \lambda_n X)$ for some $X \in \mathcal{X}$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}_+$. A duplicate portfolio relies on a single asset.

The above three situations receive special attention because they intuitively correspond to very strong diversification, some diversification, and no diversification, respectively. Naturally, we would expect DQ to be very small for (i), DQ to be smaller than 1 for (ii), and

DQ to be 1 for (iii). It turns out that the above intuitions all check out under very weak conditions that are satisfied by commonly used classes of risk measures.

Before presenting this result, we fix some technical terms. For a class ρ of risk measures ρ_α decreasing in α, we say that ρ is *non-flat from the left* at (α, X) if $\rho_\beta(X) > \rho_\alpha(X)$ for all $\beta \in (0, \alpha)$, and ρ is *left continuous* at (α, X) if $\alpha \mapsto \rho_\alpha(X)$ is left continuous. A random vector $(X_1, \dots, X_n)$ is *comonotonic* if there exists a random variable Z and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ a.s. for every $i \in [n]$. A risk measure is *comonotonic-additive* if $\phi(X + Y) = \phi(X) + \phi(Y)$ for comonotonic (X, Y). Each of ES and VaR satisfies comonotonic-additivity, as well as any distortion risk measures (Yaari (1987), Kusuoka (2001)) and signed Choquet integrals (Wang et al. (2020b)). We denote by $\rho_0 = \lim_{\alpha \downarrow 0} \rho_\alpha$. Note that $\rho_0 = \text{ess-sup}$ for common classes ρ such as VaR, ES, expectiles, and entropic risk measures.

Theorem 4.3. *For given $\mathbf{X} \in \mathcal{X}^n$ and $\alpha \in I$, if ρ is left continuous and non-flat from the left at $(\alpha, \sum_{i=1}^n X_i)$, the following hold.*

(i) Suppose $\rho_0 \leqslant \text{ess-sup}$. If for ρ_α there is no insolvency risk with pooled individual capital, then $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 0$. The converse holds true if $\rho_0 = \text{ess-sup}$.

(ii) Diversification benefit exists if and only if $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) < 1$.

(iii) If ρ_α satisfies [PH] *and $\mathbf{X}$ relies on a single asset, then $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 1$.*

(iv) If ρ_α is comonotonic-additive and $\mathbf{X}$ is comonotonic, then $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 1$.

In (i), we see that if there is no insolvency risk with pooled individual capital, then $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 0$. In typical models, such as some elliptical models in Section 4.5.2, $\sum_{i=1}^n X_i$ is unbounded from above unless it is a constant. Hence, for such models and ρ satisfying $\rho_0 = \text{ess-sup}$, $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 0$ if and only if $\sum_{i=1}^n X_i$ is a constant, thus full hedging is achieved. This is also consistent with our intuition of full hedging as the strongest form of diversification. The existence of diversification benefit is the main idea behind coherent risk measures of Artzner et al. (1999). By (ii), DQ and DR agree on whether diversification benefit exists under mild conditions, and this intuition is consistent with Artzner et al. (1999).

Remark 4.4. We require ρ to be left continuous and non-flat from the left to make the inequality in (ii) holds strictly. This requirement excludes, in particular, trivial cases like $\mathbf{X} = \mathbf{c} \in \mathbb{R}^n$ which gives $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 0$ by definition. In case the conditions fail to hold, $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) < 1$ may not guarantee $\rho_\alpha\left(\sum_{i=1}^n X_i\right) < \sum_{i=1}^n \rho_\alpha(X_i)$, but it implies the non-strict inequality $\rho_\alpha\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i)$, and thus the portfolio risk is not worse than the sum of the individual risks.

4.4.2 Stochastic dominance and dependence

In this section, we discuss the consistency of DQ with respect to stochastic dominance, as well as the best and worst cases for DQ among all dependence structures with given marginal distributions of the risk vector.

For a diversification index, monotonicity with respect to stochastic dominance yields consistency with common decision-making criteria such as the expected utility model and the rank-dependent utility model. A random variable X (representing random loss) is dominated by a random variable Y in second-order stochastic dominance (SSD) if $\mathbb{E}[f(X)] \leqslant \mathbb{E}[f(Y)]$ for all decreasing concave functions $f : \mathbb{R} \to \mathbb{R}$ provided that the expectations exist, and we denote this by $X \leqslant_{\mathrm{SSD}} Y$.[12] A risk measure ϕ is *SSD-consistent* if $\phi(X) \geqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ whenever $X \leqslant_{\mathrm{SSD}} Y$. SSD consistency is known as strong risk aversion in the classic decision theory literature (Rothschild and Stiglitz (1970)). SSD-consistent monetary risk measures, which include all law-invariant convex risk measures such as ES, admit an ES-based characterization (Mao and Wang (2020)).

Proposition 4.3. *Assume that $\rho = (\rho_\alpha)_{\alpha \in I}$ is a decreasing class of SSD-consistent risk measures. For $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ and $\alpha \in I$, if $\sum_{i=1}^n \rho_\alpha(X_i) \leqslant \sum_{i=1}^n \rho_\alpha(Y_i)$ and $\sum_{i=1}^n X_i \leqslant_{\mathrm{SSD}} \sum_{i=1}^n Y_i$, then $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) \geqslant \mathrm{DQ}_\alpha^\rho(\mathbf{Y})$.*

Proposition 4.3 follows from the simple observation that

$$\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i)\right\} \subseteq \left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n Y_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(Y_i)\right\},$$

[12] If X and Y represent gains instead of losses, then SSD is typically defined via increasing concave functions.

and we omit the proof.

Assume ρ is a class of SSD-consistent risk measures (e.g., law-invariant convex risk measures). Proposition 4.3 implies that if the sum of marginal risks is the same for $\mathbf{X}$ and $\mathbf{Y}$ (this holds in particular if $\mathbf{X}$ and $\mathbf{Y}$ have the same marginal distributions), then DQ is decreasing in SSD of the total risk. The dependence structures which maximize or minimize DQ for $\mathbf{X}$ with specified marginal distributions are discussed in Section 5.5.1. For instance, a comonotonic portfolio has the largest DQ (thus the smallest diversification) among all portfolios with the same marginal distributions; this observation is related to Proposition 4.2 (iii) and Theorem 4.3 (iv).

4.4.3 Consistency across dimensions

All properties in the previous sections are discussed under the assumption that the dimension $n \in \mathbb{N}$ is fixed. Letting n vary allows for a comparison of diversification between portfolios with different dimensions. In this section, we slightly generalize our framework by considering a diversification index D as a mapping on $\bigcup_{n\in\mathbb{N}} \mathcal{X}^n$; note that the input vector $\mathbf{X}$ of DQ and DR can naturally have any dimension n. We present two more useful properties of DQ in this setting. For $\mathbf{X} \in \mathcal{X}^n$ and $c \in \mathbb{R}$, $(\mathbf{X}, c)$ is the $(n+1)$-dimensional random vector obtained by pasting $\mathbf{X}$ and c, and $(\mathbf{X}, \mathbf{X})$ is the $(2n)$-dimensional random vector obtained by pasting $\mathbf{X}$ and $\mathbf{X}$.

[RI] Riskless invariance: $D(\mathbf{X}, c) = D(\mathbf{X})$ for all $n \in \mathbb{N}$, $\mathbf{X} \in \mathcal{X}^n$ and $c \in \mathbb{R}$.

[RC] Replication consistency: $D(\mathbf{X}, \mathbf{X}) = D(\mathbf{X})$ for all $n \in \mathbb{N}$ and $\mathbf{X} \in \mathcal{X}^n$.

Riskless invariance means that adding a risk-free asset into the portfolio $\mathbf{X}$ does not affect its diversification. For instance, the Sharpe ratio of the portfolio does not change under such an operation. Replication consistency means that replicating the same portfolio composition does not affect D. Both properties are arguably desirable in most applications due to their natural interpretations.

Proposition 4.4. *Let $\rho = (\rho_\alpha)_{\alpha\in I}$ be a class of risk measures decreasing in α. For $\alpha \in I$,*

(i) If ρ_β *satisfies* $[\mathrm{CA}]_{\mathrm{m}}$ *with* $m \in \mathbb{R}$ *for* $\beta \in I$ *and* $\rho_\alpha(0) = 0$ *then* DQ^ρ_α *satisfies* [RI].

(ii) If ρ_β *satisfies* [PH] *for* $\beta \in I$*, then* DQ^ρ_α *satisfies* [RC].

We further show in Proposition 5.7 that if [RI] is assumed, then the only option for DR is to use a non-negative ϕ (which is a subclass of DQ). Thus, if [RI] is considered as desirable, then DQ becomes useful compared to DR as it offers more choices, and in particular, it works for any classes ρ of monetary risk measures with $\rho_\alpha(0) = 0$ including VaR and ES. Both DQ and DR satisfy [RC] and [RI] for MCP risk measures.

Example 4.2. Let ϕ be a risk measure satisfying [CA], such as ES_α or VaR_α. Suppose that $\phi(\sum_{i=1}^n X_i) = 1$ and $\sum_{i=1}^n \phi(X_i) = 2$, and thus $\mathrm{DR}^\phi(\mathbf{X}) = 1/2$. If a non-random payoff of $c > 0$ is added to the portfolio, then $\mathrm{DR}^\phi(\mathbf{X}, -c) = (1-c)/(2-c)$, which turns to 0 as $c \uparrow 1$, and it becomes negative as soon as $c > 1$. Hence, DR^ϕ is improved or made negative by including constant payoffs (either as a new asset or added to an existing asset). This creates problematic incentives in optimization. On the other hand, DQ does not suffer from this problem due to [LI] and [RI].

4.5 DQ based on the classes of VaR and ES

Since VaR and ES are the two most common classes of risk measures in practice, we focus on the theoretical properties of $\mathrm{DQ}^{\mathrm{VaR}}_\alpha$ and $\mathrm{DQ}^{\mathrm{ES}}_\alpha$ in this section. We fix the parameter range $I = (0, 1)$, and we choose $\mathcal{X}^n$ to be $(L^0)^n$ when we discuss $\mathrm{DQ}^{\mathrm{VaR}}_\alpha$ and $(L^1)^n$ when we discuss $\mathrm{DQ}^{\mathrm{ES}}_\alpha$, but all results hold true if we fix $\mathcal{X} = L^1$.

4.5.1 General properties

We first provide alternative formulations of $\mathrm{DQ}^{\mathrm{VaR}}_\alpha$ and $\mathrm{DQ}^{\mathrm{ES}}_\alpha$. The formulations offer clear interpretations and simple ways to compute the values of DQs. The formula (4.6) below can be derived from the optimization formulation for the buffered probability of exceedance in Proposition 2.2 of Mafusalov and Uryasev (2018).

Theorem 4.4. *For a given* $\alpha \in (0,1)$, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ *and* $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ *have the alternative formulas*

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \frac{1}{\alpha}\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right), \quad \mathbf{X} \in \mathcal{X}^n, \tag{4.4}$$

and

$$\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \frac{1}{\alpha}\mathbb{P}\left(Y > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)\right), \quad \mathbf{X} \in \mathcal{X}^n, \tag{4.5}$$

where $Y = \mathrm{ES}_U(\sum_{i=1}^n X_i)$ *and* $U \sim \mathrm{U}[0,1]$. *Furthermore, if* $\mathbb{P}(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)) > 0$, *then*

$$\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \frac{1}{\alpha}\min_{r\in(0,\infty)} \mathbb{E}\left[\left(r\sum_{i=1}^n (X_i - \mathrm{ES}_\alpha(X_i)) + 1\right)_+\right], \tag{4.6}$$

and otherwise $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 0$.

As a first observation from Theorem 4.4, it is straightforward to compute $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ on real or simulated data by applying (4.4) and (4.5) to the empirical distribution of the data.

Theorem 4.4 also gives $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ a clear economic interpretation as the improvement of insolvency probability when risks are pooled, making the discussion in Section 4.3.4 more concrete. Suppose that $X_1, \dots, X_n$ are continuously distributed and they represent losses from n assets. The total pooled capital is $s_\alpha = \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)$, which is determined by the marginals of $\mathbf{X}$ but not the dependence structure. An agent investing only in asset X_i with capital computed by VaR_α has an insolvency probability $\alpha = \mathbb{P}(X_i > \mathrm{VaR}_\alpha(X_i))$. On the other hand, by Theorem 4.4, α^* is the probability that the pooled loss $\sum_{i=1}^n X_i$ exceeds the pooled capital s_α. The improvement from α to α^*, computed by α^*/α, is precisely $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$. From here, it is also clear that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) < 1$ is equivalent to $\mathbb{P}\left(\sum_{i=1}^n X_i > s_\alpha\right) < \alpha$.

To compare $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ with $\mathrm{DR}^{\mathrm{VaR}_\alpha}$, recall that the two diversification indices can be rewritten as

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \frac{\mathbb{P}\left(\sum_{i=1}^n X_i > s_\alpha\right)}{\alpha} \quad \text{and} \quad \mathrm{DR}^{\mathrm{VaR}_\alpha}(\mathbf{X}) = \frac{\mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right)}{s_\alpha}. \tag{4.7}$$

From (4.7), we can see a clear symmetry between DQ, which measures the probability improvement, and DR, which measures the quantile improvement. DQ and DR based on ES have a similar comparison.

The range of DQ based on VaR is different from that based on ES, which is $[0,1]$ by Proposition 4.1. We summarize them below.

Proposition 4.5. *For $\alpha \in (0,1)$ and $n \geqslant 2$, $\{\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0, \min\{n, 1/\alpha\}]$ and $\{\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0,1]$.*

Both $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ take values on a bounded interval. In contrast, the diversification ratio $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ is unbounded, and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ is bounded above by 1 only when the ES of the total risk is non-negative.

Remark 4.5. It is a coincidence that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ for $\alpha < 1/n$ and $\mathrm{DR}^{\mathrm{var}}$ both have a maximum value n. The latter maximum value is attained by a risk vector $(X/n, \dots, X/n)$ for any $X \in L^2$.

4.5.2 Capturing heavy tails and common shocks

In this section, we analyze three simple normal and t-models to illustrate some features of DQ regarding heavy tails and common shocks in the portfolio models. Here, we only present some key observations. A detailed study of DQs based on VaR and ES for elliptical distributions and multivariate regularly varying models, including explicit formulas to compute DQ for these models, can be found in Chapter 6.

Let $\mathbf{Z} = (Z_1, \dots, Z_n)$ be an n-dimensional standard normal random vector, and let ξ^2 have an inverse gamma distribution independent of $\mathbf{Z}$. Denote by $\mathrm{it}_n(\nu)$ the joint distribution with n independent t-marginals $\mathrm{t}(\nu, 0, 1)$, where the parameter ν represents the degrees of freedom; see McNeil et al. (2015) for t-models. The model $\mathbf{Y} = (Y_1, \dots, Y_n) \sim \mathrm{it}_n(\nu)$ can be stochastically represented by

$$Y_i = \xi_i Z_i, \quad \text{for } i \in [n], \tag{4.8}$$

where $\xi_1, \dots, \xi_n$ are iid following the same distribution as ξ, and independent of $\mathbf{Z}$. In contrast, a joint t-distributed random vector $\mathbf{Y}' = (Y_1', \dots, Y_n') \sim \mathrm{t}(\nu, \mathbf{0}, I_n)$ has a stochastic representation $\mathbf{Y}' = \xi \mathbf{Z}$, that is,

$$Y_i' = \xi Z_i, \quad \text{for } i \in [n]. \tag{4.9}$$

Table 4.2: DQs/DRs based on VaR, ES, SD and var, where $\alpha = 0.05$, $n = 10$ and $\nu = 3$; numbers in bold indicate the most diversified among $\mathbf{Z}, \mathbf{Y}, \mathbf{Y}'$ according to the index D

D	$\mathrm{DQ}_\alpha^{\mathrm{VaR}}$	$\mathrm{DQ}_\alpha^{\mathrm{ES}}$	$\mathrm{DR}^{\mathrm{VaR}_\alpha}$	$\mathrm{DR}^{\mathrm{ES}_\alpha}$	$\mathrm{DR}^{\mathrm{SD}}$	$\mathrm{DR}^{\mathrm{var}}$
$\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, I_n)$	$\mathbf{2.0 \times 10^{-6}}$	$\mathbf{1.9 \times 10^{-9}}$	**0.3162**	0.3162	**0.3162**	**1**
$\mathbf{Y} \sim \mathrm{it}_n(3)$	0.0235	0.0124	0.3569	**0.2903**	**0.3162**	**1**
$\mathbf{Y}' \sim \mathrm{t}(3, \mathbf{0}, I_n)$	0.0502	0.0340	**0.3162**	0.3162	**0.3162**	**1**
$D(\mathbf{Z}) < D(\mathbf{Y})$	Yes	Yes	Yes	No	No	No
$D(\mathbf{Y}) < D(\mathbf{Y}')$	Yes	Yes	No	Yes	No	No

In other words, $\mathbf{Y}'$ is a standard normal random vector multiplied by a heavy-tailed common shock ξ. All three models $\mathbf{Z}, \mathbf{Y}, \mathbf{Y}'$ have the same correlation matrix, the identity matrix I_n.

Because of the common shock ξ in (4.9), large losses from components of $\mathbf{Y}'$ are more likely to occur simultaneously, compared to $\mathbf{Y}$ in (4.8), which does not have a common shock. Indeed, $\mathbf{Y}'$ is tail dependent (Example 7.39 of McNeil et al. (2015)) whereas $\mathbf{Y}$ is tail independent. As such, at least intuitively (if not rigorously), diversification for portfolio $\mathbf{Y}'$ should be considered as weaker than $\mathbf{Y}$, although both models are uncorrelated and have the same marginals.[13] By the central limit theorem, for $\nu > 2$, the component-wise average of $\mathbf{Y}$ (scaled by its variance) is asymptotically normal as n increases, whereas the component-wise average of $\mathbf{Y}'$ is always t-distributed. Hence, one may intuitively expect the order $D(\mathbf{Z}) < D(\mathbf{Y}) < D(\mathbf{Y}')$ to hold.

In Tables 4.2 and 4.3, we present DQ and DR for a few different models based on $\mathrm{N}(\mathbf{0}, I_n)$, $\mathrm{t}(\nu, \mathbf{0}, I_n)$, and $\mathrm{it}_n(\nu)$. We choose $n = 10$ and $\nu = 3$ or 4,[14] and thus we have five models in total. As we see from Tables 4.2 and 4.3, DQs based on both VaR and ES report a lower value for $\mathrm{it}_n(\nu)$ and a larger value for $\mathrm{t}(\nu, \mathbf{0}, I_n)$, meaning that diversification is weaker for the common shock t-model (4.9) than the iid t-model (4.8). For the iid normal model, the

[13] On a related note, as discussed by Embrechts et al. (2002), correlation is not a good measure of diversification in the presence of heavy-tailed and skewed distributions.

[14] Most financial asset log-loss data have a tail-index between $[3, 5]$, which corresponds to $\nu \in [3, 5]$; see e.g., Jansen and De Vries (1991).

Table 4.3: DQs/DRs based on VaR, ES, SD and var, where $\alpha = 0.05$, $n = 10$ and $\nu = 4$; numbers in bold indicate the most diversified among $\mathbf{Z}, \mathbf{Y}, \mathbf{Y}'$ according to the index D

D	$\mathrm{DQ}_\alpha^{\mathrm{VaR}}$	$\mathrm{DQ}_\alpha^{\mathrm{ES}}$	$\mathrm{DR}^{\mathrm{VaR}_\alpha}$	$\mathrm{DR}^{\mathrm{ES}_\alpha}$	$\mathrm{DR}^{\mathrm{SD}}$	$\mathrm{DR}^{\mathrm{var}}$
$\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, I_n)$	$\mathbf{2.0 \times 10^{-6}}$	$\mathbf{1.9 \times 10^{-9}}$	**0.3162**	0.3162	**0.3162**	**1**
$\mathbf{Y} \sim \mathrm{it}_n(4)$	0.0050	0.0017	0.3415	**0.2828**	**0.3162**	**1**
$\mathbf{Y}' \sim \mathrm{t}(4, \mathbf{0}, I_n)$	0.0252	0.0138	**0.3162**	0.3162	**0.3162**	**1**
$D(\mathbf{Z}) < D(\mathbf{Y})$	Yes	Yes	Yes	No	No	No
$D(\mathbf{Y}) < D(\mathbf{Y}')$	Yes	Yes	No	Yes	No	No

diversification is the strongest according to DQ. In contrast, DR sometimes reports that the iid t-model has a larger diversification than the common shock t-model, which is counter-intuitive. In the setting of both Tables 4.2 and 4.3, a risk manager governed by $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ would prefer the iid portfolio over the common shock portfolio, but the preference is flipped if the risk manager uses $\mathrm{DR}^{\mathrm{VaR}_\alpha}$. A more detailed analysis on this phenomenon for varying $\alpha \in (0, 0.1]$ is presented in Figure 5.1 in Section 5.6, and consistent results are observed.

4.6 Portfolio selection with DQ

Next, we focus on the optimal diversification problem

$$\min_{\mathbf{w} \in \Delta_n} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X}) \quad \text{and} \quad \min_{\mathbf{w} \in \Delta_n} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w} \odot \mathbf{X}); \tag{4.10}$$

recall that a smaller value of DQ means better diversification.[15] Recall from Table 4.1 that the first optimization is not quasi-convex and the second one is quasi-convex (Proposition 4.1). We do not say that optimizing a diversification index has a decision-theoretic benefit; here we simply illustrate the advantage of DQ in computation and optimization. Whether optimizing diversification is desirable for individual or institutional investors is an open-ended question

[15] A possible alternative formulation to (4.10) is to use DQ as a constraint instead of an objective in the optimization. This is mathematically similar to a risk measure constraint (e.g., Basak and Shapiro (2001), Rockafellar and Uryasev (2002) and Mafusalov and Uryasev (2018)), but with a different interpretation, as DQ is not designed to measure or control risk.

which goes beyond this chapter; we refer to Van Nieuwerburgh and Veldkamp (2010), Boyle et al. (2012) and Choi et al. (2017) for relevant discussions.

For the portfolio weight $\mathbf{w}$, DQ based on VaR at level $\alpha \in (0,1)$ is given by

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X}) = \frac{1}{\alpha} \inf \left\{ \beta \in (0,1) : \mathrm{VaR}_\beta \left(\sum_{i=1}^n w_i X_i \right) \leqslant \sum_{i=1}^n w_i \mathrm{VaR}_\alpha(X_i) \right\},$$

and DQ based on ES is similar. In what follows, we fix $\alpha \in (0,1)$ and $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$, where $\mathcal{X}$ is L^0 for VaR and L^1 for ES, as in Section 4.5. Write $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^n$ and $\mathbf{x}_\alpha^\rho = (\rho_\alpha(X_1), \dots, \rho_\alpha(X_n))$ for a given risk measure ρ.

Proposition 4.6. *Fix $\alpha \in (0,1)$ and $\mathbf{X} \in \mathcal{X}^n$. The optimization of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X})$ in (4.10) can be solved by*

$$\min_{\mathbf{w} \in \Delta_n} \mathbb{P}\left(\mathbf{w}^\top \left(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{VaR}} \right) > 0 \right). \tag{4.11}$$

Assuming $\mathbb{P}(X_i > \mathrm{ES}_\alpha(X_i)) > 0$ for each $i \in [n]$, the optimization of $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w} \odot \mathbf{X})$ in (4.10) can be solved by the convex program

$$\min_{\mathbf{v} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}} \mathbb{E}\left[\left(\mathbf{v}^\top \left(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}} \right) + 1 \right)_+ \right], \tag{4.12}$$

and the optimal $\mathbf{w}^*$ is given by $\mathbf{v}/\|\mathbf{v}\|_1$.

Proposition 4.6 offers efficient algorithms to optimize $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ in real-data applications. The values of $\mathbf{x}_\alpha^{\mathrm{VaR}}$ and $\mathbf{x}_\alpha^{\mathrm{ES}}$ can be computed by many existing estimators of the individual losses (see e.g., McNeil et al. (2015)). In particular, a simple way to estimate these risk measures is to use an empirical estimator. More specifically, if we have data $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$ sampled from $\mathbf{X}$ satisfying some ergodicity condition (being iid would be sufficient), then the empirical version of the problem (4.11) is

$$\text{minimize} \quad \sum_{j=1}^N \mathbb{1}_{\left\{ \mathbf{w}^\top \left(\mathbf{X}^{(j)} - \widehat{\mathbf{x}}_\alpha^{\mathrm{VaR}} \right) > 0 \right\}} \quad \text{over } \mathbf{w} \in \Delta_n, \tag{4.13}$$

where $\widehat{\mathbf{x}}_\alpha^{\mathrm{VaR}}$ is the empirical estimator of $\mathbf{x}_\alpha^{\mathrm{VaR}}$ based on sample $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$; see McNeil et al. (2015). Write $\mathbf{y}^{(j)} = \mathbf{X}^{(j)} - \widehat{\mathbf{x}}_\alpha^{\mathrm{VaR}}$ and $z_j = \mathbb{1}_{\{\mathbf{w}^\top \mathbf{y}^{(j)} > 0\}}$ for $j \in [n]$. Problem (4.13) involves a chance constraint (see e.g., Luedtke (2014) and Liu et al. (2016)). By using the big-M

method (see e.g., Shen et al. (2010)) via choosing a sufficient large M (e.g., it is sufficient if M is larger than the components of $\mathbf{y}^{(j)}$ for all j), (4.13) can be converted into the following linear integer program:

$$\begin{aligned} &\text{minimize} && \textstyle\sum_{j=1}^{N} z_j \\ &\text{subject to} && \mathbf{w}^\top \mathbf{y}^{(j)} - M z_j \leqslant 0, \quad \textstyle\sum_{i=1}^{n} w_i = 1, \\ & && z_j \in \{0,1\}, \quad w_i \geqslant 0 \quad \text{for all } j \in [N] \text{ and } i \in [n]. \end{aligned} \tag{4.14}$$

Similarly, the optimization problem (4.12) for $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ can be solved the empirical version of the problem (4.12), which is a convex program:

$$\text{minimize} \quad \sum_{j=1}^{N} \max\left\{\mathbf{v}^\top \left(\mathbf{X}^{(j)} - \widehat{\mathbf{x}}_\alpha^{\mathrm{ES}}\right) + 1, 0\right\} \quad \text{over } \mathbf{v} \in \mathbb{R}_+, \tag{4.15}$$

where $\widehat{\mathbf{x}}_\alpha^{\mathrm{ES}}$ is the empirical estimator of $\mathbf{x}_\alpha^{\mathrm{ES}}$ based on sample $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(N)}$. Both problems (4.14) and (4.15) can be efficiently solved by modern optimization programs, such as CVX programming (see e.g., Matmoura and Penev (2013)).

Additional linear constraints, such as those on budget or expected return, can be easily included in (4.11)-(4.15), and the corresponding optimization problems can be solved similarly.

Tie-breaking needs to be addressed when working with (4.13) since its objective function takes integer values. In dynamic portfolio selection, it is desirable to avoid adjusting positions too drastically or frequently. Therefore, in the real-data analysis in Section 4.7.3, among tied optimizers, we pick the closest one (in L^1-norm $\|\cdot\|_1$ on $\mathbb{R}^n$) to a given benchmark $\mathbf{w}_0$, the portfolio weight of the previous trading period. With this tie-breaking rule, we solve

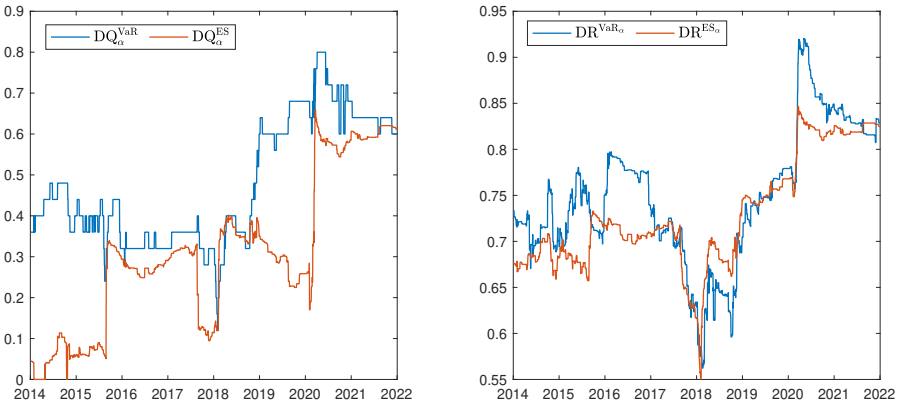
$$\text{minimize} \quad \|\mathbf{w} - \mathbf{w}_0\|_1 \quad \text{over } \mathbf{w} \in \Delta_n \quad \text{subject to} \quad \sum_{j=1}^{N} \mathbb{1}_{\{\mathbf{w}^\top \mathbf{y}^{(j)} > 0\}} \leqslant m^*, \tag{4.16}$$

where m^* is the optimum of (4.13). A tie-breaking for (4.15) may need to be addressed similarly since (4.15) is not strictly convex.

4.7 Numerical illustrations

To illustrate the performance of DQ, we collect historical asset prices from Yahoo Finance and conduct three sets of numerical experiments based on the data. We use the period

Figure 4.2: DQs and DRs based on VaR and ES with $\alpha = 0.05$



from January 3, 2012, to December 31, 2021, with a total of 2518 observations of daily losses and 500 trading days for the initial training. In Section 4.7.1, we first compare DQs and DRs based on VaR and ES. In Section 4.7.2, we calculate the values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ under different selections of stocks. Finally, we construct portfolios by minimizing $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ and $\mathrm{DR}^{\mathrm{SD}}$ and by the mean-variance criterion in Section 4.7.3.

4.7.1 Comparing DQ and DR

We first identify the largest stock in each of the S&P 500 sectors ranked by market cap in 2012. Among these stocks, we select the 5 largest stocks[16] to build our portfolio. We compute $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ on each day using the empirical distribution in a rolling window of 500 days, where we set $\alpha = 0.05$.

Figure 4.2 shows that the values of DQ and DR are between 0 and 1. This corresponds to the observation in Theorem 4.3 that $\mathrm{DQ}_\alpha^\rho < 1$ is equivalent to $\mathrm{DR}^{\rho_\alpha} < 1$. DQ has a similar temporal pattern to DR in the above period of time, with a large jump when COVID-19 exploded, which is more visible for DQ than for DR. We remind the reader that DQ and DR are not meant to be compared on the same scale, and hence the fact that DQ has a larger range than DR should be taken lightly. We also note that the values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ are in discrete grids. This is because the empirical distribution function takes value in multiples of

[16]XOM from ENR, AAPL from IT, BRK/B from FINL, WMT from CONS, and GE from INDU.

$1/N$ there N is the sample size (500 in this experiment) and hence $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ takes the values $k/(N\alpha)$ for an integer k; see (4.4). If a smooth curve is preferred, then one can employ a smoothed VaR through linear interpolation. This is a standard technique for handling VaR; see McNeil et al. (2015, Section 9.2.6) and Li and Wang (2022, Remark 8 and Section B).

4.7.2 DQ for different portfolios

In this section, we fix $\alpha = 0.05$ and calculate the values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ under different portfolio compositions of stocks. We consider portfolios with the following stock compositions:

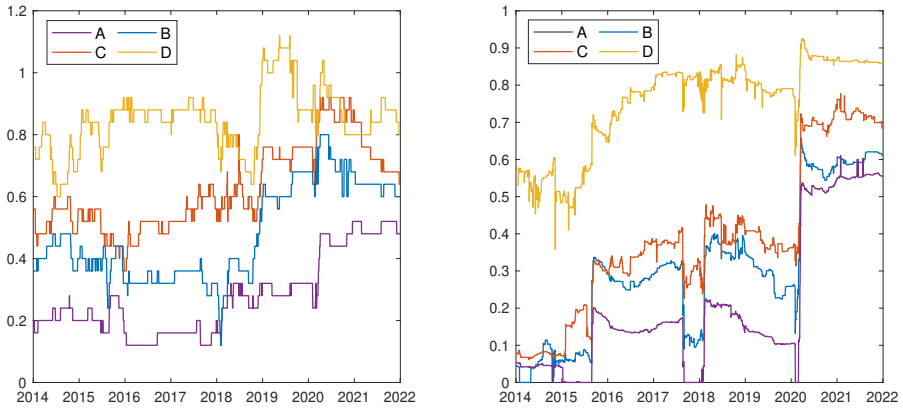
(A) the two largest stocks from each of the 10 different sectors of S&P 500;

(B) the largest stock from each of 5 different sectors of S&P 500 (as in Section 4.7.1);

(C) the 5 largest stocks, AAPL, MSFT, IBM, GOOGL and ORCL, from the Information Technology (IT) sector;

(D) the 5 largest stocks, BRK/B, WFC, JPM, C and BAC, from the Financials (FINL) sector.

Figure 4.3: DQs based on VaR (left) and ES (right) with $\alpha = 0.05$



We make a few observations from Figure 4.3. Both $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ provide similar

comparative results. The order (A)$\leqslant$(B)$\leqslant$(C)$\leqslant$(D) is consistent with our intuition.[17] First, portfolio (A) of 20 stocks has the strongest diversification effect among the four compositions. Second, portfolio (B) across 5 sectors has stronger diversification than (C) and (D) within one sector. Third, portfolio (C) of 5 stocks within the IT sector has a stronger diversification than portfolio (D) of 5 stocks within the FINL sector, consistent with the fact that the stocks in the IT sector are less correlated. Moreover, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ for the FINL sector is larger than 1 during some period of time, which means that there is no diversification benefit if risk is evaluated by VaR. All DQ curves based on ES show a large up-ward jump around the COVID-19 outbreak; such a jump also exists for curves based on VaR but it is less pronounced.

4.7.3 Optimal diversified portfolios

In this section, we fix $\alpha = 0.1$ and build portfolios via $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, $\mathrm{DR}^{\mathrm{SD}}$, and the mean-variance criterion in the Markowitz (1952) model.[18] The optimal portfolio problems using $\mathrm{DR}^{\mathrm{SD}}$ and the Markowitz model are well studied in literature; see e.g. Choueifaty and Coignard (2008). We compare these portfolio wealth with the equal weighted (EW) portfolio and the simple buy-and-hold (BH) portfolio. For an analysis on the EW strategy, see DeMiguel et al. (2009).

We apply the algorithms in Proposition 4.6 to optimize $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, which are extremely fast. A tie-breaking is addressed for each objective as in (4.16). Minimization of $\mathrm{DR}^{\mathrm{SD}}$ and the Markowitz model can be solved by existing algorithms. The initial wealth is set to 1, and the risk-free rate is $r = 2.84\%$, which is the 10-year yield of the US treasury bill in Jan 2014. The target annual expected return for the Markowitz portfolio is set to 10%. We optimize the portfolio weights in each month with a rolling window of 500 days. That is, in each month, roughly 21 trading days, starting from January 2, 2014, we use the preceding 500 trading days to compute the optimal portfolio weights using the method described above.

[17] The observations here are consistent with those from applying $\mathrm{DR}^{\mathrm{SD}}$ (which is also a DQ) in the same setting; see Section 5.8.

[18] One may try other portfolio criteria other than mean-variance. For instance, Levy and Levy (2004) found that portfolio strategies based on prospect theory perform similarly to the mean-variance strategies.

The portfolio is rebalanced every month. We choose the 4 largest stocks from each of the 10 different sectors of S&P 500 ranked by market cap in 2012 as the portfolio compositions (40 stocks in total). The portfolio performance is reported in Figure 4.4, and the cumulative distribution of the sorted portfolio weights, averaged over each month, is shown in Figure 4.5. Summary statistics, including the annualized return (AR), the annualized volatility (AV), the Sharpe ratio (SR), and the average trading proportion (ATP), are reported in Table 4.4.[19]

From these results, we can see that the portfolio optimization strategies based on minimizing DQ perform quite well, similarly to those based on $\mathrm{DR}^{\mathrm{SD}}$, and better than the Markowitz strategy. Moreover, ATP and portfolio weight distribution are similar across the strategies based on the three diversification indices and the Markowitz strategy. In contrast, the EW and BH strategies have more uniform portfolio weight distributions and smaller ATP, as anticipated. We remark that it is not our intention to analyze which diversification strategy generates the highest return, which is a challenging question that needs a separate study; also, we do not suggest diversification should or should not be optimized in practice. The empirical results here are presented to illustrate how our proposed diversification indices work in the context of portfolio selection. More empirical results with some other datasets and portfolio strategies are given in Section 5.8, and the results show similar patterns.

4.8 Concluding remarks

In this chapter, we put forward six axioms to jointly characterize a new class of indices of diversification, and a seventh axiom to specialize this class. The new diversification index DQ has favourable features both theoretically and practically, and it is contrasted with its competitors, in particular DR. At a high level, because of the conceptual symmetry in Figure

[19] ATP is an approximation of trading costs, and it is computed as the average of $\sum_{t=1}^{T} |w_i^t - w_i^{t-}|$ over $i = 1, \dots, n$, where $T = 96$ is the total number of months, w_i^t is the portfolio weight of asset i at the beginning of month t, and w_i^{t-} is the portfolio weight of asset i at the end of month $t-1$, with w_i^{1-} set to w_i^1. Note that for BH, ATP is 0 because there is no trading, whereas for EW, ATP is positive, as rebalancing occurs at the end of each month.

Figure 4.4: Wealth processes for portfolios, 40 stocks, Jan 2014 - Dec 2021

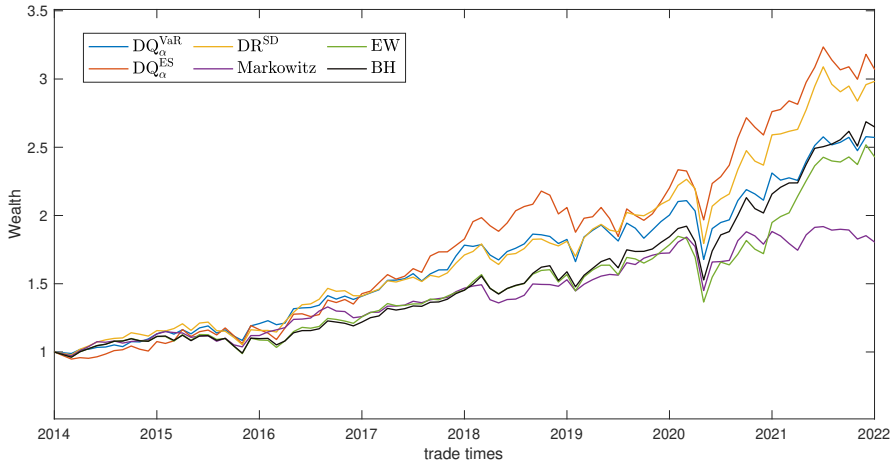
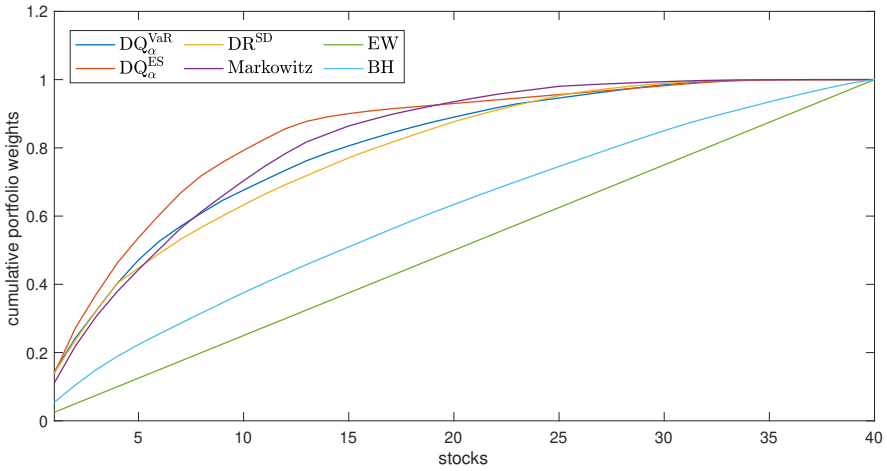


Figure 4.5: Cumulative portfolio weights, 40 stocks, Jan 2014 - Dec 2021



4.1 (see also (4.7)), we expect both DQ and DR to have advantages and disadvantages in different applications, and none should fully dominate the other. Nevertheless, we find many attractive features of DQ through the results in this chapter, which suggest that DQ may be a better choice in many situations.

We summarize these features below. Some of these features are shared by DR, but many are not. (i) DQ defined on a class of MCP risk measures can be uniquely characterized by six intuitive axioms (Theorem 4.1). DQ defined on a class of coherent risk measures can be uniquely characterized by further adding the axiom of portfolio convexity (Theorem 4.2).

Table 4.4: Annualized return (AR), annualized volatility (AV), Sharpe ratio (SR), and average trading proportion (ATP) for different portfolio strategies from Jan 2014 to Dec 2021

%	$\mathrm{DQ}_\alpha^{\mathrm{VaR}}$	$\mathrm{DQ}_\alpha^{\mathrm{ES}}$	$\mathrm{DR}^{\mathrm{SD}}$	Markowitz	EW	BH
AR	12.56	14.59	14.36	7.93	11.91	12.88
AV	14.64	15.74	14.99	12.98	15.92	14.34
SR	66.40	74.66	76.85	39.22	56.95	70.02
ATP	19.29	14.75	15.61	18.79	4.43	0

These two results lay an axiomatic foundation for using DQ as a diversification index. (ii) DQ further satisfies many properties for common risk measures (Propositions 4.1-4.4). These properties are not shared by the corresponding DR. (iii) DQ is intuitive and interpretable with respect to dependence and common perceptions of diversification (Theorem 4.3). (iv) DQ can be applied to a wide range of risk measures, such as the regulatory risk measures VaR and ES, as well as expectiles. In cases of VaR and ES, DQ has simple formulas and convenient properties (Theorem 4.4 and Proposition 4.5). (v) Portfolio optimization of DQs based on VaR and ES can be computed very efficiently (Proposition 4.6). (vi) DQ can be easily applied to real data and it produces results that are consistent with our usual perception of diversification (Section 4.7).

Among the class of DQ, for most applications, we generally recommend the use of DQ based on ES for the following reasons: (a) it satisfies all seven axioms of intuitive appeal; (b) it has a simple optimization formula that is very convenient in portfolio optimization; (c) it is closely connected to financial regulation as ES is the standard risk measure of Basel IV; (d) it has a flexible parameter α that allows for reflecting the sensitivity to the tail risk of the decision maker; (e) it is conceptually easy to interpret as the (usually unique) level β of the ES family such that $\mathrm{ES}_\beta(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)$.

We also mention a few interesting questions on DQ, which call for thorough future study. (i) DQ is defined through a class of risk measures. It would be interesting to formulate DQ using expected utility or behavioral decision models, to analyze the decision-theoretic implications of DQ. For instance, DQ based on entropic risk measures can be equivalently

formulated using exponential utility functions. Alternatively, one may also build DQ directly from acceptability indices (see Remark 4.3). (ii) To compute DQ, one needs to invert the decreasing function $\beta \mapsto \rho_\beta(\sum_{i=1}^n X_i)$. In the case of VaR and ES, the formula for this inversion is simple (Theorem 4.4). For more complicated classes of risk measures, this computation may be complicated and requires detailed analysis. (iii) For general distributions and risk measures other than VaR and ES, finding analytical formulas or efficient algorithms for optimal diversification using either DQ or DR is a challenging task. (iv) Further analysis of DQ without scale-invariance, such as those built on star-shaped risk measures (Castagnoli et al. (2022)), may further generalize the domain of application.

Chapter 5

Diversification quotients: Technical details and illustrations

5.1 Outline

We organize this chapter as follows. The proofs of the main results in Chapter 4, Theorems 4.1–4.4, are presented in Section 5.2. Additional results, discussions, and proofs of propositions are presented in Section 5.3 (for Section 4.2), 5.4 (for Section 4.3), 5.5 (for Section 4.4), 5.6 (for Section 4.5), and 5.7 (for Section 4.6). Finally, in Section 5.8, we present other examples for the optimal portfolio problem that complement the empirical studies in Section 4.7.3.

5.2 Proofs of Theorems 4.1–4.4

Proof of Theorem 4.1. For $\mathbf{X} \in \mathcal{X}^n$ and a risk measure $\phi : \mathcal{X} \to \mathbb{R}$, denote by $S(\mathbf{X}) = \sum_{i=1}^n X_i$ and $\mathbf{X}_\phi = (X_1 - \phi(X_1), \dots, X_n - \phi(X_n))$.

We first verify the "if" statement. Using the definition of DQ and properties of MCP risk measures, it is straightforward to verify [+], [LI], [SI]. Below we check the other three axioms.

To show $[\mathrm{R}]_\phi$, for $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ such that $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$ and $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n Y_i$, we have $\sum_{i=1}^n \rho_\alpha(X_i) = \sum_{i=1}^n \rho_\alpha(Y_i)$ and $\rho_\beta(\sum_{i=1}^n X_i) \leqslant \rho_\beta(\sum_{i=1}^n Y_i)$ for all $\beta \in I$. Hence,

$$\begin{aligned}\mathrm{DQ}_\alpha^\rho(\mathbf{X}) &= \frac{1}{\alpha} \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i)\right\} \\ &\leqslant \frac{1}{\alpha} \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n Y_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(Y_i)\right\} = \mathrm{DQ}_\alpha^\rho(\mathbf{Y}).\end{aligned}$$

To show $[\mathrm{N}]_\phi$, it is straightforward that $\mathrm{DQ}_\alpha^\rho(\mathbf{0}) = 0$. Let $\mathbf{X} = (X, \dots, X)$ for any $X \in \mathcal{X}$. We have

$$\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = \frac{1}{\alpha} \inf\{\beta \in I : \rho_\beta(nX) \leqslant n\rho_\alpha(X)\} \leqslant \frac{\alpha}{\alpha} = 1.$$

If $\mathbf{Y} \overset{\mathrm{m}}{\succ} (X, \dots, X)$ and $\sum_{i=1}^n Y_i \geqslant nX$, then $\sum_{i=1}^n \rho_\alpha(Y_i) < n\rho_\alpha(X) \leqslant \rho_\alpha(\sum_{i=1}^n Y_i)$. Hence, $\mathrm{DQ}_\alpha^\rho(\mathbf{Y}) \geqslant 1$.

To show $[\mathrm{C}]_\phi$, for $\mathbf{X} \in \mathcal{X}^n$, we have $a_{\mathbf{X}}^* = \inf\{\beta \in I : \rho_\beta(S(\mathbf{X}_{\rho_\alpha})) \leqslant 0\}$. If $a_{\mathbf{X}}^* = 0$, it is clear that $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 0$ and $[\mathrm{C}]_\phi$ holds as $\mathrm{DQ}_\alpha^\rho(\mathbf{Y}^k) \geqslant 0$ for any $\mathbf{Y}^k \in \mathcal{X}^n$. Now, we assume $a_{\mathbf{X}}^* > 0$. For any $0 \leqslant \beta < a_{\mathbf{X}}^*$, we have $\rho_\beta(S(\mathbf{X}_{\rho_\alpha})) > 0$. Since $\mathbf{Y}^k \overset{\mathrm{m}}{\simeq} \mathbf{X}$ for each k and $(S(\mathbf{X}) - S(\mathbf{Y}^k))_+ \xrightarrow{L^\infty} 0$ as $k \to \infty$, for any $\epsilon > 0$, there exists K such that $S(\mathbf{X}_{\rho_\alpha}) - S(\mathbf{Y}^k_{\rho_\alpha}) \leqslant \epsilon$ for all $k \geqslant K$. For any $0 < \delta < a_{\mathbf{X}}^*$, let $0 < \epsilon < \rho_{a_{\mathbf{X}}^* - \delta}(S(\mathbf{X}_{\rho_\alpha}))$. It is clear that $0 < \epsilon < \rho_\beta(S(\mathbf{X}_{\rho_\alpha}))$ for all $0 < \beta < a_{\mathbf{X}}^* - \delta$. Hence, for all $0 < \beta < a_{\mathbf{X}}^* - \delta$, there exists K such that $0 < \rho_\beta(S(\mathbf{X}_{\rho_\alpha})) - \epsilon \leqslant \rho_\beta(S(\mathbf{Y}^k_{\rho_\alpha}))$ for all $k \geqslant K$, which implies $a_{\mathbf{Y}^k}^* \geqslant a_{\mathbf{X}}^* - \delta$. Therefore, $(\mathrm{DQ}_\alpha^\rho(\mathbf{X}) - \mathrm{DQ}_\alpha^\rho(\mathbf{Y}^k))_+ \to 0$.

Next, we show the "only if" statement. Assume that $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ satisfies [+], [LI], [SI], $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ and $[\mathrm{C}]_\phi$. Note that $\mathbf{X}_\phi \overset{\mathrm{m}}{\simeq} \mathbf{Y}_\phi$ for all $\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$ since $\phi(X - \phi(X)) = 0$ for any $X \in \mathcal{X}$. Hence, by using $[\mathrm{R}]_\phi$, we know that $S(\mathbf{X}_\phi) = S(\mathbf{Y}_\phi)$ implies $D(\mathbf{X}_\phi) = D(\mathbf{Y}_\phi)$. Further, by [LI], we have $D(\mathbf{X}) = D(\mathbf{Y})$ if $S(\mathbf{X}_\phi) = S(\mathbf{Y}_\phi)$. This means that $D(\mathbf{X})$ is determined by $S(\mathbf{X}_\phi)$. Define the mapping

$$R : \mathcal{X} \to [0, \infty],\ R(X) = \inf\{D(\mathbf{X}) : X \leqslant S(\mathbf{X}_\phi),\ \mathbf{X} \in \mathcal{X}^n\}, \tag{5.1}$$

with the convention $\inf \varnothing = \infty$. Next, we will verify several properties of R.

(a) $R(S(\mathbf{X}_\phi)) = D(\mathbf{X})$ for $\mathbf{X} \in \mathcal{X}^n$. The inequality $R(S(\mathbf{X}_\phi)) \leqslant D(\mathbf{X})$ follows directly from (5.1). To see the opposite direction of the inequality, suppose $R(S(\mathbf{X}_\phi)) < D(\mathbf{X})$.

By (5.1), there exists $\mathbf{Y} \in \mathcal{X}^n$ such that $D(\mathbf{Y}) < D(\mathbf{X})$ and $S(\mathbf{X}_\phi) \leqslant S(\mathbf{Y}_\phi)$. This contradicts $[\mathrm{R}]_\phi$ of D.

(b) $R(\lambda X) = R(X)$ for all $\lambda > 0$ and $X \in \mathcal{X}$. This follows directly from (5.1), [SI] of D and positive homogeneity of ϕ which gives $(\lambda \mathbf{X})_\phi = \lambda \mathbf{X}_\phi$.

(c) $R(X) \leqslant R(Y)$ for all $X, Y \in \mathcal{X}$ with $X \leqslant Y$. This follows directly from (5.1).

(d) $R(0) = 0$. This follows directly from (5.1) and $D(\mathbf{0}) = 0$ in $[\mathrm{N}]_\phi$.

(e) $\lim_{c \downarrow 0} R(S(\mathbf{X}_\phi) - c) = R(S(\mathbf{X}_\phi))$ for $\mathbf{X} \in \mathcal{X}^n$. Let $X = S(\mathbf{X}_\phi)$. By (c), we have $\lim_{c \downarrow 0} R(X - c) \leqslant R(X)$. Assume $\lim_{c \downarrow 0} R(X - c) < R(X)$; that is, there exists $\delta > 0$ such that $R(X - c) < R(X) - \delta$ for all $c > 0$. Let $c_k = 1/k$ for $k \in \mathbb{N}$. By (5.1), there exists a sequence $\{\mathbf{Y}^k\}_{k \in \mathbb{N}}$ such that $X - c_k \leqslant S(\mathbf{Y}^k_\phi)$ and $D(\mathbf{Y}^k_\phi) < D(\mathbf{X}_\phi) - \delta$. For $\{\mathbf{Y}^k_\phi\}_{k \in \mathbb{N}}$, we have $0 \leqslant (S(\mathbf{X}_\phi) - S(\mathbf{Y}^k_\phi))_+ \leqslant c_k$, which implies $(S(\mathbf{X}_\phi) - S(\mathbf{Y}^k_\phi))_+ \xrightarrow{L^\infty} 0$ as $k \to \infty$. By $[\mathrm{C}]_\phi$, we have $(D(\mathbf{X}_\phi) - D(\mathbf{Y}^k_\phi))_+ \to 0$; that is, for any $\delta > 0$, there exists $K \in \mathbb{N}$ such that $D(\mathbf{X}_\phi) - \delta \leqslant D(\mathbf{Y}^k_\phi)$ for all $k > K$, which is a contradiction. Therefore, we have $\lim_{c \downarrow 0} R(S(\mathbf{X}_\phi) - c) = R(S(\mathbf{X}_\phi))$.

Let $I = (0, \infty)$. For each $\beta \in (0, \infty)$, let $\mathcal{A}_\beta = \{X \in \mathcal{X} : R(X) \leqslant \beta\}$. Since R is monotone, $\mathcal{A}_\beta$ is a decreasing set; i.e., $X \in \mathcal{A}_\beta$ implies $Y \in \mathcal{A}_\beta$ for all $Y \leqslant X$. Moreover, $\mathcal{A}_\beta$ is conic; i.e., $X \in \mathcal{A}_\beta$ implies $\lambda X \in \mathcal{A}_\beta$ for all $\lambda > 0$. Moreover, we have $\mathcal{A}_\beta \subseteq \mathcal{A}_\gamma$ for $\beta \leqslant \gamma$, and $\mathcal{A}_\beta \neq \varnothing$ since $0 \in \mathcal{A}_0$.

Let $\rho_\beta(X) = \inf\{m \in \mathbb{R} : X - m \in \mathcal{A}_\beta\}$ for $\beta \in I$. Since ρ_β is defined via a conic acceptance set, $(\rho_\beta)_{\beta \in I}$ is a class of MCP risk measures; see Föllmer and Schied (2016). It is also clear that ρ_β is decreasing in β. Note that $X \in \mathcal{A}_\beta$ implies $\rho_\beta(X) \leqslant 0$. Hence,

$$R(X) = \inf\{\beta \in I : R(X) \leqslant \beta\} = \inf\{\beta \in I : X \in \mathcal{A}_\beta\} \geqslant \inf\{\beta \in I : \rho_\beta(X) \leqslant 0\}.$$

For $X \in \{S(\mathbf{X}_\phi) : \mathbf{X} \in \mathcal{X}^n\}$, using (e), we have $R(X - m) \leqslant \beta$ for all $m > 0$ implies $R(X) \leqslant \beta$. Then we have

$$\inf\{\beta \in I : \rho_\beta(X) \leqslant 0\} = \inf\{\beta \in I : R(X - m) \leqslant \beta \text{ for all } m > 0\} \geqslant \inf\{\beta \in I : R(X) \leqslant \beta\}.$$

Therefore, $\inf\{\beta \in I : \rho_\beta(S(\mathbf{X}_\phi)) \leqslant 0\} = \inf\{\beta \in I : R(S(\mathbf{X}_\phi)) \leqslant \beta\} = R(S(\mathbf{X}_\phi))$ for all $\mathbf{X} \in \mathcal{X}^n$. Using (a), we get, for all $\mathbf{X} \in \mathcal{X}^n$,

$$D(\mathbf{X}) = R(S(\mathbf{X}_\phi)) = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \phi(X_i)\right\}.$$

Take a duplicate portfolio $\mathbf{X} = (X, \dots, X)$. Together with [PH] of ρ_β, $D(\mathbf{X}) \leqslant 1$ implies

$$D(X, \dots, X) = \inf\{\beta \in I : \rho_\beta(X) \leqslant \phi(X)\} \leqslant 1,$$

which is equivalent to $\rho_\beta(X) \leqslant \phi(X)$ for $\beta > 1$. For $m < \phi(X)$, take any $\mathbf{Y} = (Y_1, \dots, Y_n)$ satisfying $S(\mathbf{Y}_\phi) \geqslant X - m$; such $\mathbf{Y}$ may not exist. Let $\mathbf{Z} = (Y_1 - \phi(Y_1) + m/n, \dots, Y_n - \phi(Y_n) + m/n)$, yielding $\sum_{i=1}^n Z_i = \sum_{i=1}^n Y_i - \sum_{i=1}^n \phi(Y_i) + m \geqslant X$ and $(Z_1 \dots, Z_n) \overset{\mathrm{m}}{\succ} (X/n, \dots, X/n)$. Hence, $\mathbf{Z}$ is worse than duplicate. By [LI] and $[\mathrm{N}]_\phi$, we have $D(\mathbf{Y}) = D(\mathbf{Z}) \geqslant 1$. Since $D(\mathbf{Y}) \geqslant 1$ for all such $\mathbf{Y}$, by (5.1), we have $R(X - m) \geqslant 1$. The above observation implies $\rho_{1-\epsilon}(X) \geqslant \phi(X)$ for any $\epsilon > 0$ since $\rho_{1-\epsilon}(X) = \inf\{m \in \mathbb{R} : R(X - m) \leqslant 1 - \epsilon\}$. Therefore, we get $\rho_{1-\epsilon} \geqslant \phi \geqslant \rho_{1+\epsilon}$ for all $\epsilon > 0$. Let $\tilde{\rho}_1 = \phi$, and $\tilde{\rho}_\beta = \rho_\beta$ for $\beta \neq 1$. The class $\tilde{\rho} = (\tilde{\rho}_\beta)_{\beta \in I}$ of MCP risk measures is decreasing in β by the above argument. Moreover, for any $X \in \mathcal{X}$, since the two decreasing curves $\beta \mapsto \rho_\beta(X)$ and $\beta \mapsto \tilde{\rho}_\beta(X)$ differ at only one point, their (left) inverses coincide, and we have, for all $\mathbf{X} \in \mathcal{X}^n$,

$$\inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \phi(X_i)\right\} = \inf\left\{\beta \in I : \tilde{\rho}_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \tilde{\rho}_1(X_i)\right\},$$

which implies $D = \mathrm{DQ}_1^{\tilde{\rho}}$ on $\mathcal{X}^n$. A reparametrization via $\hat{\rho}_\beta = \tilde{\rho}_{\beta/\alpha}$ leads to $D = \mathrm{DQ}_\alpha^{\hat{\rho}}$ and $\hat{\rho}_\alpha = \phi$. □

The next two remarks are useful in the proof of Theorem 4.2.

Remark 5.1. In the proof of Theorem 4.1, the constructed class of risk measures $(\rho_\beta)_{\beta \in I}$ exhibits right continuity in β. This is established based on the condition $\bigcap_{\beta > \beta^*} \mathcal{A}_\beta = \mathcal{A}_{\beta^*}$.

Remark 5.2. For a non-linear coherent risk measure ϕ, there exists $Y \in \mathcal{X}$ such that $\phi(Y) + \phi(-Y) > 0$. Suppose otherwise. Since ϕ is coherent risk measure, we have $\phi(Y) + \phi(-Y) \geqslant 0$, and this implies $\phi(Y) + \phi(-Y) = 0$ for all $Y \in \mathcal{X}$. We obtain $\phi(Y) \leqslant \phi(X + Y) + \phi(-X) = \phi(X + Y) - \phi(X)$ and $\phi(X + Y) \leqslant \phi(X) + \phi(Y)$ for any $X, Y \in \mathcal{X}$. This implies $\phi(X + Y) = \phi(X) + \phi(Y)$ for any $X, Y \in \mathcal{X}$, contradicting the non-linearity of ϕ.

Proof of Theorem 4.2. For the"if" statement, since a coherent risk measure is also MCP, it follows that DQ_α^ρ satisfies [+], [LI], [SI], $[\mathrm{R}]_\phi$ and $[\mathrm{N}]_\phi$, $[\mathrm{C}]_\phi$ by Theorem 4.1. Next, we show that DQ_α^ρ satisfies [PC].

For any $\mathbf{X} \in \mathcal{X}^n$, let $r_\beta^{\mathbf{X}} : \Delta_n \to \mathbb{R}$ be given by

$$r_\beta^{\mathbf{X}}(\mathbf{w}) = \rho_\beta\left(\sum_{i=1}^n w_i X_i\right) - \sum_{i=1}^n \rho_\alpha\left(w_i X_i\right)$$

for $\beta \in I$. From [PH] of ρ_α, we have $r_\beta^{\mathbf{X}}(\mathbf{w}) = \rho_\beta\left(\sum_{i=1}^n w_i X_i\right) - \mathbf{w}^\top \mathbf{x}_\alpha^\rho$. Convexity of ρ_β implies convexity of $\mathbf{w} \mapsto r_\beta^{\mathbf{X}}(\mathbf{w})$. Hence, for the portfolio weight $\lambda\mathbf{w} + (1-\lambda)\mathbf{v} \in \Delta_n$, DQ based on ρ at level $\alpha \in (0,1)$ is given by

$$\begin{aligned}
\mathrm{DQ}_\alpha^\rho((\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \odot \mathbf{X}) &= \frac{1}{\alpha} \inf\left\{\beta \in I : r_\beta^{\mathbf{X}}(\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \leqslant 0\right\} \\
&\leqslant \frac{1}{\alpha} \inf\left\{\beta \in I : \lambda r_\beta^{\mathbf{X}}(\mathbf{w}) + (1-\lambda) r_\beta^{\mathbf{X}}(\mathbf{v}) \leqslant 0\right\} \\
&\leqslant \frac{1}{\alpha} \max\left\{\inf\{\beta \in I : r_\beta^{\mathbf{X}}(\mathbf{w}) \leqslant 0\}, \inf\{\beta \in I : r_\beta^{\mathbf{X}}(\mathbf{v}) \leqslant 0\}\right\} \\
&= \max\left\{\mathrm{DQ}_\alpha^\rho(\mathbf{w} \odot \mathbf{X}), \mathrm{DQ}_\alpha^\rho(\mathbf{v} \odot \mathbf{X})\right\},
\end{aligned}$$

which gives us quasi-convexity of $\mathbf{w} \mapsto \mathrm{DQ}_\alpha^\rho(\mathbf{w} \odot \mathbf{X})$. By Remark 4.1, we have that DQ_α^ρ satisfies [PC].

For the "only if" statement, we have constructed a class of MCP risk measures $\rho = (\rho_\beta)_{\beta \in (0,\infty)}$ with $\rho_\alpha = \phi$ in the proof of Theorem 4.1. We will further show that ρ is a class of convex risk measures using [PC].

For any $\lambda \in [0,1]$, $\mathbf{w}, \mathbf{v} \in \Delta_n$, $\mathbf{X} \in \mathcal{X}^n$ and $\beta \in (0,\infty)$, if $r_\beta^{\mathbf{X}}(\mathbf{w}) \leqslant 0$ and $r_\beta^{\mathbf{X}}(\mathbf{v}) \leqslant 0$, then we have $\mathrm{DQ}_\alpha^\rho(\mathbf{w} \odot \mathbf{X}) \leqslant \beta$ and $\mathrm{DQ}_\alpha^\rho(\mathbf{v} \odot \mathbf{X}) \leqslant \beta$. By [PC], we have $\mathrm{DQ}_\alpha^\rho((\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \odot \mathbf{X}) \leqslant \beta$. As discussed in Remark 5.1, ρ_β is right-continuous for any $X \in \mathcal{X}$. Hence, we have $r_\beta^{\mathbf{X}}(\lambda\mathbf{w} + (1-\lambda)\mathbf{v}) \leqslant 0$. That is, the set $\{\mathbf{w} \in \Delta_n : r_\beta^{\mathbf{X}}(\mathbf{w}) \leqslant 0\}$ is convex for any $\mathbf{X} \in \mathcal{X}^n$ and $\beta \in I$.

Let $\mathrm{Conv}\{X - \rho_\alpha(X) : X \in \mathcal{X}\}$ be the convex hull of $\{X - \rho_\alpha(X) : X \in \mathcal{X}\}$. Next, we show that $\mathrm{Conv}\{X - \rho_\alpha(X) : X \in \mathcal{X}\} = \{X \in \mathcal{X} : \rho_\alpha(X) \leqslant 0\}$. For any $Z \in \mathrm{Conv}\{X - \rho_\alpha(X) : X \in \mathcal{X}\}$, there exist $(\lambda_1, \dots, \lambda_n) \in \Delta_n$ and $X_1, \dots, X_n \in \mathcal{X}$ such that $Z = \sum_{i=1}^n \lambda_i (X_i - \rho_\alpha(X_i))$. Since ρ_α is convex, we have

$$\rho_\alpha(Z) = \rho_\alpha\left(\sum_{i=1}^n \lambda_i (X_i - \rho_\alpha(X_i))\right) \leqslant \sum_{i=1}^n \lambda_i \rho_\alpha(X_i - \rho_\alpha(X_i)) = 0.$$

Hence, $\mathrm{Conv}\{X-\rho_\alpha(X): X\in\mathcal{X}\}\subseteq\{X\in\mathcal{X}:\rho_\alpha(X)\leqslant 0\}$.

On the other hand, since ρ_α is a non-linear coherent risk measure, as noted in Remark 5.2, there exists Y such that $\rho_\alpha(Y)+\rho_\alpha(-Y)>0$. Let $Y'=1-Y$. For any $Z\in\mathcal{X}$ with $\rho_\alpha(Z)\leqslant 0$, we can find $\theta>0$ such that

$$Z=(1-2\lambda)(X-\rho_\alpha(X))+\lambda(\theta Y-\rho_\alpha(\theta Y))+\lambda(\theta Y'-\rho_\alpha(\theta Y'))$$

with $\lambda=-\rho_\alpha(Z)/(\theta\rho_\alpha(Y)+\theta\rho_\alpha(-Y))$ and $X=1/(1-2\lambda)Z$. It is clear that $\lambda\in[0,1/2]$ holds if θ is sufficiently large. Hence, $Z\in\mathrm{Conv}\{X-\rho_\alpha(X):X\in\mathcal{X}\}$. This implies $\mathrm{Conv}\{X-\rho_\alpha(X):X\in\mathcal{X}\}\supseteq\{X\in\mathcal{X}:\rho_\alpha(X)\leqslant 0\}$.

Therefore, for any $X,Y\in\mathcal{X}$ with $\rho_\alpha(X)\leqslant 0$ and $\rho_\alpha(Y)\leqslant 0$, we can find $\mathbf{w},\mathbf{v}\in\Delta^n$ and $\mathbf{X}\in\mathcal{X}^n$ with $n\geqslant 4$ such that $X=\mathbf{w}^\top(\mathbf{X}-\mathbf{x}^\rho_\alpha)$ and $Y=\mathbf{v}^\top(\mathbf{X}-\mathbf{x}^\rho_\alpha)$. Since $r_\beta(\mathbf{w})=\rho_\beta\left(\mathbf{w}^\top(\mathbf{X}-\mathbf{x}^\rho_\alpha)\right)$, we have $\rho_\beta(X)=r^{\mathbf{X}}_\beta(\mathbf{w})$, $\rho_\beta(Y)=r^{\mathbf{X}}_\beta(\mathbf{v})$ and $\rho_\beta(\lambda X+(1-\lambda)Y)=r^{\mathbf{X}}_\beta(\lambda\mathbf{w}+(1-\lambda)\mathbf{v})$. If $\rho_\beta(X)\leqslant 0$ and $\rho_\beta(Y)\leqslant 0$, we have $r^{\mathbf{X}}_\beta(\lambda\mathbf{w}+(1-\lambda)\mathbf{v})\leqslant 0$ for any $\lambda\in[0,1]$; that is $\rho_\beta(\lambda X+(1-\lambda)Y)\leqslant 0$. Hence, ρ_β is quasi-convex. Since ρ_β is MCP, we further have ρ_β is coherent. □

Proof of Theorem 4.3. (i) As $\sum_{i=1}^n X_i\leqslant\sum_{i=i}^n\rho_\alpha(X_i)$ a.s. and $\rho_0\leqslant$ ess-sup, it is clear that $\alpha^*=0$ in (4.3), which implies $\mathrm{DQ}^\rho_\alpha(\mathbf{X})=0$. Conversely, if $\mathrm{DQ}^\rho_\alpha(\mathbf{X})=0$, then $\alpha^*=0$. By definition of ρ_0 and DQ^ρ_α, this implies $\rho_0(\sum_{i=1}^n X_i)\leqslant\sum_{i=1}^n\rho_\alpha(X_i)$, and hence $\sum_{i=1}^n X_i\leqslant\sum_{i=1}^n\rho_\alpha(X_i)$ a.s.

(ii) We first show the "only if" statement. As ρ is left continuous and non-flat from the left at $(\alpha,\sum_{i=1}^n X_i)$ and $\sum_{i=1}^n\rho_\alpha(X_i)-\rho_\alpha\left(\sum_{i=1}^n X_i\right)>0$, there exists $\delta>0$ such that

$$\rho_\beta\left(\sum_{i=1}^n X_i\right)-\rho_\alpha\left(\sum_{i=1}^n X_i\right)<\sum_{i=1}^n\rho_\alpha(X_i)-\rho_\alpha\left(\sum_{i=1}^n X_i\right)$$

for all $\beta\in(\alpha-\delta,\alpha)$. Hence, we have $\alpha^*\leqslant\alpha-\delta<\alpha$, which leads to $\mathrm{DQ}^\rho_\alpha(\mathbf{X})<1$.

Next, we show the "if" statement. As $\mathrm{DQ}^\rho_\alpha(\mathbf{X})<1$, we have $\alpha>\alpha^*$. By (4.3), there exists $\beta\in(\alpha^*,\alpha)$ such that

$$\sum_{i=1}^n\rho_\alpha(X_i)\geqslant\rho_\beta\left(\sum_{i=1}^n X_i\right).$$

Because ρ is non-flat from the left at $(\alpha, \sum_{i=1}^n X_i)$, we have

$$\sum_{i=1}^n \rho_\alpha(X_i) \geqslant \rho_\beta\left(\sum_{i=1}^n X_i\right) > \rho_\alpha\left(\sum_{i=1}^n X_i\right).$$

(iii) If ρ_α satisfies [PH], for $\mathbf{X} = (\lambda_1 X, \dots, \lambda_n X)$ where $\lambda_1, \dots, \lambda_n \geqslant 0$, we have

$$\alpha^* = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n \lambda_i X\right) \leqslant \sum_{i=1}^n \lambda_i \rho_\alpha(X)\right\}.$$

It is clear that $\rho_\alpha\left(\sum_{i=1}^n \lambda_i X\right) = \left(\sum_{i=1}^n \lambda_i\right)\rho_\alpha(X)$. Together with the non-flat condition and $\rho_\beta\left(\sum_{i=1}^n \lambda_i X\right) > \sum_{i=1}^n \lambda_i \rho_\alpha(X)$ for all $\beta < \alpha$, we have $\alpha^* = \alpha$, and thus $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 1$.

(iv) If ρ_α is comonotonic-additive and $\mathbf{X}$ is comonotonic, then

$$\alpha^* = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \rho_\alpha(X_i) = \rho_\alpha\left(\sum_{i=1}^n X_i\right)\right\},$$

which, together with the non-flat condition, implies that $\alpha^* = \alpha$, and thus $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 1$. □

Proof of Theorem 4.4. We first show (4.4). For any $X \in \mathcal{X}$, $t \in \mathbb{R}$ and $\alpha \in (0,1)$, by Lemma 1 of Guan et al. (2022), $\mathbb{P}(X > t) \leqslant \alpha$ if and only if $\mathrm{VaR}_\alpha(X) \leqslant t$. Hence,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) &= \inf\left\{\beta \in (0,1) : \mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) \leqslant \beta\right\} \\ &= \inf\left\{\beta \in (0,1) : \mathrm{VaR}_\beta\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right\},\end{aligned}$$

and (4.4) follows. The formula (4.5) for $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ follows from a similar argument to (4.4) by noting that Y is a random variable with $\mathrm{VaR}_\alpha(Y) = \mathrm{ES}_\alpha(\sum_{i=1}^n X_i)$.

Next, we show the last statement of the theorem. If $\mathbb{P}(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)) = 0$, then $\mathrm{DQ}_\alpha^{\mathrm{ES}}(X) = 0$ by Theorem 4.3 (i).

Below, we assume $\mathbb{P}(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)) > 0$. The formula (4.6) is very similar to Proposition 2.2 of Mafusalov and Uryasev (2018), where we additionally show that the minimizer to (4.6) is not 0. Here we present a self-contained proof based on the well-known formula of ES (Rockafellar and Uryasev (2002)),

$$\mathrm{ES}_\beta(X) = \min_{t \in \mathbb{R}}\left\{t + \frac{1}{\beta}\mathbb{E}\left[(X-t)_+\right]\right\}, \quad \text{for } X \in \mathcal{X} \text{ and } \beta \in (0,1).$$

Using this formula, we obtain, by writing $X_i' = X_i - \mathrm{ES}_\alpha(X_i)$ for $i \in [n]$,

$$\begin{aligned}
\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) &= \frac{1}{\alpha} \inf\left\{ \beta \in (0,1) : \mathrm{ES}_\beta\left(\sum_{i=1}^n X_i\right) - \sum_{i=1}^n \mathrm{ES}_\alpha(X_i) \leqslant 0 \right\} \\
&= \frac{1}{\alpha} \inf\left\{ \beta \in (0,1) : \min_{t\in\mathbb{R}} \left\{ t + \frac{1}{\beta}\mathbb{E}\left[\left(\sum_{i=1}^n X_i' - t \right)_+ \right] \right\} \leqslant 0 \right\} \\
&= \frac{1}{\alpha} \inf\left\{ \beta \in (0,1) : \frac{1}{\beta}\mathbb{E}\left[\left(\sum_{i=1}^n X_i' - t \right)_+ \right] \leqslant -t \text{ for some } t \in \mathbb{R} \right\} \\
&= \frac{1}{\alpha} \inf\left\{ \beta \in (0,1) : \mathbb{E}\left[\left(r\sum_{i=1}^n X_i' + 1 \right)_+ \right] \leqslant \beta \text{ for some } r \in (0,\infty) \right\} \\
&= \frac{1}{\alpha} \inf_{r\in(0,\infty)} \mathbb{E}\left[\left(r\sum_{i=1}^n X_i' + 1 \right)_+ \right].
\end{aligned}$$

Let $f : [0,\infty) \to [0,\infty),\ r \mapsto \mathbb{E}[(r\sum_{i=1}^n X_i' + 1)_+]$. It is clear that $f(0) = 1$. Moreover,

$$f(r) \geqslant r\mathbb{E}\left[(X_i')_+\right] \to \infty \quad \text{as } r \to \infty.$$

By Theorem 4.1 (iii), we have $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) \leqslant 1$, and hence $\inf_{r\in(0,\infty)} f(r) \leqslant \alpha < 1$. The continuity of f yields $\inf_{r\in(0,\infty)} f(r) = \min_{r\in(0,\infty)} f(r)$, and thus (4.6) holds. □

5.3 Additional results for Section 4.2

In this section, we present an impossibility result showing a conflicting nature of the three natural properties [+], [LI] and [SI] for some diversification indices defined via risk measures. As mentioned in Section 4.2, the most commonly used diversification indices depend on $\mathbf{X}$ through its values assessed by some risk measure ϕ. That is, given a risk measures ϕ and a portfolio $\mathbf{X}$, the diversification index can be written as

$$D(\mathbf{X}) = R\left(\phi\left(\sum_{i=1}^n X_i\right), \phi(X_i), \dots, \phi(X_n)\right) \text{ for some function } R : \mathbb{R}^{n+1} \to \overline{\mathbb{R}}. \tag{5.2}$$

We will say that D is *ϕ-determined* if (5.2) holds. Often, one may further choose R so that $D(\mathbf{X})$ decreases in $\phi(\sum_{i=1}^n X_i)$ and increases in $\phi(X_i)$ for each $i \in [n]$, for a proper interpretation of measuring diversification.

We show that a diversification index based on an MCP risk measure, such as VaR or ES satisfying all three properties [+], [LI] and [SI] can take at most 3 different values. In

this case, we will say that the diversification index D is *degenerate*. In fact, this result can be extended to more general properties $[\mathrm{PH}]_\gamma$ and $[\mathrm{CA}]_m$ with $\gamma \in \mathbb{R}$ and $m \in \mathbb{R}$ of the risk measure ϕ, with definitions given at the beginning of Section 4.4.

Proposition 5.1. *Fix $n \geqslant 1$. Suppose that a risk measure ϕ satisfies $[\mathrm{PH}]_\gamma$ and $[\mathrm{CA}]_\mathrm{m}$ with $\gamma \in \mathbb{R}$ and $m \neq 0$. A diversification index D is ϕ-determined and satisfies [+], [LI] and [SI] if and only if for all $\mathbf{X} \in \mathcal{X}^n$,*

$$D(\mathbf{X}) = C_1 \mathbb{1}_{\{d<0\}} + C_2 \mathbb{1}_{\{d=0\}} + C_3 \mathbb{1}_{\{d>0\}}, \tag{5.3}$$

where $d = \mathrm{DB}^\phi(\mathbf{X}) = \sum_{i=1}^n \phi(X_i) - \phi\left(\sum_{i=1}^n X_i\right)$ for some $C_1, C_2, C_3 \in \mathbb{R}_+ \cup \{\infty\}$.

We first present a lemma to prepare for the proof of Proposition 5.1.

Lemma 5.1. *A function $R : \mathbb{R}^{n+1} \to \overline{\mathbb{R}}$ satisfies, for all $x_0 \in \mathbb{R}$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ and $\lambda > 0$, (i) $R\left(x_0 + \sum_{i=1}^n c_i, \mathbf{x} + \mathbf{c}\right) = R(x_0, \mathbf{x})$ and (ii) $R(\lambda x_0, \lambda \mathbf{x}) = R(x_0, \mathbf{x})$, if and only if there exist $C_1, C_2, C_3 \in \overline{\mathbb{R}}$ such that*

$$R(x_0, \mathbf{x}) = C_1 \mathbb{1}_{\{r<0\}} + C_2 \mathbb{1}_{\{r=0\}} + C_3 \mathbb{1}_{\{r>0\}}, \tag{5.4}$$

where $r = \sum_{i=1}^n x_i - x_0$, for all $x_0 \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.

Proof. First, we show that R in (5.4) satisfies (i) and (ii). Assume $r < 0$. For any $\mathbf{c} \in \mathbb{R}^n$ and $\lambda > 0$, it is clear that $x_0 + \sum_{i=1}^n c_i < \sum_{i=1}^n (x_i + c_i)$ and $\lambda x_0 < \sum_{i=1}^n \lambda x_i$. Therefore, (i) and (ii) are satisfied. The cases of $r = 0$ and $r > 0$ follow by the same argument.

Next, we verify the "only if" part. Given $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ satisfying $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, let $\mathbf{c} = \mathbf{y} - \mathbf{x}$. For any $x_0 \in \mathbb{R}$, we have $\sum_{i=1}^n c_i = \sum_{i=1}^n (y_i - x_i) = 0$. Therefore,

$$R(x_0, \mathbf{x}) = R\left(x_0 + \sum_{i=1}^n c_i, \mathbf{x} + \mathbf{c}\right) = R(x_0, \mathbf{y}).$$

Thus, the value of $R(x_0, \mathbf{x})$ only depends on x_0 and $\sum_{i=1}^n x_i$. Let $\tilde{R} : \mathbb{R}^2 \to \overline{\mathbb{R}}$ be a function such that $\tilde{R}\left(x_0, \sum_{i=1}^n x_i\right) = R(x_0, \mathbf{x})$. From the properties of R, $\tilde{R}$ satisfies $\tilde{R}(a + c, b + c) = \tilde{R}(a, b)$ for any $c \in \mathbb{R}$, and $\tilde{R}(\lambda a, \lambda b) = R(a, b)$ for any $\lambda > 0$. Hence, we have

$$\tilde{R}(a, b) = \tilde{R}(a - b, 0) = \tilde{R}(1, 0) \quad \text{for} \quad a > b,$$

$$\tilde{R}(a,b) = \tilde{R}(0,b-a) = \tilde{R}(0,1) \quad \text{for} \quad a < b,$$

and

$$\tilde{R}(a,b) = \tilde{R}(a-a,b-a) = \tilde{R}(0,0) \quad \text{for} \quad a = b.$$

Let $C_1 = \tilde{R}(1,0)$, $C_2 = \tilde{R}(0,0)$ and $C_3 = \tilde{R}(0,1)$. We have $R(x_0,\mathbf{x}) = \tilde{R}(x_0, \sum_{i=1}^n x_i)$, which has the form in (5.4). □

Proof of Proposition 5.1. Let us first prove sufficiency. By definition, D satisfies [+] and D is ϕ-determined. Next, we prove D satisfies [LI] and [SI]. Similarly to Lemma 5.1, we only prove the case $d < 0$. It is straightforward that

$$\phi\left(\sum_{i=1}^n \lambda X_i\right) = \lambda^\gamma \phi\left(\sum_{i=1}^n X_i\right) < \lambda^\gamma \sum_{i=1}^n \phi(X_i) = \sum_{i=1}^n \phi(\lambda X_i),$$

and

$$\phi\left(\sum_{i=1}^n (X_i + c_i)\right) = \phi\left(\sum_{i=1}^n X_i\right) + m\sum_{i=1}^n c_i < \sum_{i=1}^n (\phi(X_i) + mc_i) = \sum_{i=1}^n \phi(X_i + c_i).$$

Thus, we have $D(\lambda\mathbf{X}) = C_1$ and $D(\mathbf{X}+\mathbf{c}) = C_1$, which completes the proof of sufficiency.

Next, we show the necessity. Define the set

$$\mathcal{A} = \left\{ \left(\phi\left(\sum_{i=1}^n X_i\right), \phi(X_1), \dots, \phi(X_n) \right) : (X_1,\dots,X_n) \in \mathcal{X}^n \right\}.$$

Note that ϕ satisfies $[\mathrm{PH}]_\gamma$ with $\gamma \neq 0$ since $[\mathrm{CA}]_m$ for $m \neq 0$ implies $\rho(2) \neq \rho(1)$, which in turn implies $\gamma \neq 0$. We always write $\mathbf{x} = (x_1,\dots,x_n)$ and $\mathbf{c} = (c_1,\dots,c_n)$. Consider the two operations $(x_0,\mathbf{x}) \mapsto (x_0 + \sum_{i=1}^n c_i, \mathbf{x}+\mathbf{c})$ for some $\mathbf{c} \in \mathbb{R}^n$ and $(x_0,\mathbf{x}) \mapsto (\lambda x_0, \lambda\mathbf{x})$ for some $\lambda > 0$. Let $r(x_0,\mathbf{x}) = \sum_{i=1}^n x_i - x_0$. By using $[\mathrm{CA}]_m$ and $[\mathrm{PH}]_\gamma$ of ϕ, we have that (see also the proof of Lemma 5.1) the regions $\mathcal{A}_+ := \{(x_0,\mathbf{x}) : r(x_0,\mathbf{x}) > 0\}$, $\mathcal{A}_0 := \{(x_0,\mathbf{x}) : r(x_0,\mathbf{x}) = 0\}$ and $\mathcal{A}_- := \{(x_0,\mathbf{x}) : r(x_0,\mathbf{x}) < 0\}$ are closed under the above two operations, and each of them is connected via the above two operations. Therefore, $\mathcal{A}$ is the union of some of $\mathcal{A}_+$, $\mathcal{A}_0$ and $\mathcal{A}_-$.

We define a function $R : \mathbb{R}^{n+1} \to \overline{\mathbb{R}}$. For $(x_0,\mathbf{x}) \in \mathcal{A}$, let $R(x_0,\mathbf{x}) = D(X_1,\dots,X_n)$, where $(X_1,\dots,X_n)$ is any random vector with $x_0 = \phi(\sum_{i=1}^n X_i)$ and $\mathbf{x} = (\phi(X_1),\dots,\phi(X_n))$. The choice of $(X_1,\dots,X_n)$ is irrelevant since D is ϕ-determined. For $(x_0,\mathbf{x}) \in \mathbb{R}^{n+1} \setminus \mathcal{A}$, let $R(x_0,\mathbf{x}) = 0$. We will verify that R satisfies conditions (i) and (ii) in Lemma 5.1.

For $(x_0, \mathbf{x}) \in \mathcal{A}$, there exists $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ such that $x_0 = \phi(\sum_{i=1}^n X_i)$ and $\mathbf{x} = (\phi(X_1), \ldots, \phi(X_n))$. For any $\mathbf{c} \in \mathbb{R}^n$, using $[\mathrm{CA}]_m$ with $m \neq 0$ of ϕ and [LI] of D, we obtain

$$\begin{aligned} R\left(x_0 + \sum_{i=1}^n c_i, \mathbf{x} + \mathbf{c}\right) &= R\left(\phi\left(\sum_{i=1}^n X_i\right) + \sum_{i=1}^n c_i, \phi(X_1) + c_1, \ldots, \phi(X_n) + c_n\right) \\ &= R\left(\phi\left(\sum_{i=1}^n \left(X_i + \frac{c_i}{m}\right)\right), \phi\left(X_1 + \frac{c_1}{m}\right), \ldots, \phi\left(X_n + \frac{c_n}{m}\right)\right) \\ &= D\left(\mathbf{X} + \mathbf{c}/m\right) = D(\mathbf{X}) = R\left(x_0, \mathbf{x}\right). \end{aligned}$$

Using $[\mathrm{PH}]_\gamma$ with $\gamma \neq 0$ of ϕ and [SI] of D, for any $\lambda > 0$, we obtain

$$\begin{aligned} R(\lambda x_0, \lambda \mathbf{x}) &= R\left(\lambda\phi\left(\sum_{i=1}^n X_i\right), \lambda\phi(X_1), \ldots, \lambda\phi(X_n)\right) \\ &= R\left(\phi\left(\sum_{i=1}^n \lambda^{1/\gamma} X_i\right), \phi(\lambda^{1/\gamma} X_1), \ldots, \phi(\lambda^{1/\gamma} X_n)\right) \\ &= D(\lambda^{1/\gamma}\mathbf{X}) = D(\mathbf{X}) = R\left(x_0, \mathbf{x}\right). \end{aligned}$$

Hence, R satisfies (i) and (ii) in Lemma 5.1 on $\mathcal{A}$. By definition, R satisfies (i) and (ii) also on $\mathbb{R}^{n+1} \setminus \mathcal{A}$. Since $\mathcal{A}$ and $\mathbb{R}^{n+1} \setminus \mathcal{A}$ are both closed under the two operations, we know that R satisfies (i) and (ii) on $\mathbb{R}^{n+1}$.

Using Lemma 5.1, we have R has the representation (5.4), which gives

$$D(\mathbf{X}) = C_1 \mathbb{1}_{\{d<0\}} + C_2 \mathbb{1}_{\{d=0\}} + C_3 \mathbb{1}_{\{d>0\}}$$

with $d = \sum_{i=1}^n \phi(X_i) - \phi\left(\sum_{i=1}^n X_i\right)$ and $C_1, C_2, C_3 \in \overline{\mathbb{R}}$ for all $\mathbf{X} \in X^n$. As D satisfying [+], we have $C_1, C_2, C_3 \in \mathbb{R}_+ \cup \{\infty\}$. □

5.4 Additional results and proofs for Section 4.3

5.4.1 Existence of worse-than-duplicate portfolios

We discuss the existence of worse-than-duplicate portfolios. First, note that if a vector $\mathbf{X}^{\mathrm{wd}} = (X_1^{\mathrm{wd}}, \ldots, X_n^{\mathrm{wd}})$ is worse than a duplicate portfolio $\mathbf{X}^{\mathrm{du}} = (X, \ldots, X)$ under a given

MCP risk measure ϕ, then we have

$$\phi\left(\sum_{i=1}^{n} X_i^{\mathrm{wd}}\right) \geqslant \phi(nX) = n\phi(X) > \sum_{i=1}^{n} \phi(X_i^{\mathrm{wd}})$$

and thus ϕ violates subadditivity with $\mathbf{X}^{\mathrm{wd}}$. Therefore, a necessary condition for the existence of a vector that is worse than a duplicate under a MCP risk measure ϕ is that ϕ violates subadditivity.

We further provide a necessary and sufficient condition for the existence or non-existence of duplicate portfolios.

Lemma 5.2. *For a monotone risk measure ϕ, there exists a worse-than-duplicate portfolio if and only if there exist $X, X_1, \dots, X_n \in \mathcal{X}$ with $X_1 + \cdots + X_n = nX$ such that $\phi(X) > \phi(X_i)$ for $i \in [n]$.*

Proof. This follows directly by monotonicity. □

A risk measure $\phi : \mathcal{X} \to \mathbb{R}$ is *scale-continuous* if the mapping $\lambda \mapsto \phi(\lambda X)$ on $(0, 1)$ is continuous for every X. This condition is very weak; for instance it is weaker than continuity on any L^p-space $\mathcal{X}$.

Proposition 5.2. *For a monotone risk measure ϕ scale-continuous on $\mathcal{X} = L^\infty$, there exists no worse-than-duplicate portfolio if and only if ϕ is quasi-convex.*

Proof. If ϕ is quasi-convex, then for any $X_1, \dots, X_n$,

$$\phi\left(\frac{X_1}{n} + \cdots + \frac{X_n}{n}\right) \leqslant \max\{\phi(X_1), \dots, \phi(X_n)\}. \tag{5.5}$$

By Lemma 5.2, there exists no worse-than-duplicate portfolio. Conversely, if exists no worse-than-duplicate portfolio, then (5.5) holds for all $X_1, \dots, X_n$. It suffices to verify that this implies quasi-convexity of ϕ. That is, we need to show that for $\lambda \in (0, 1)$ and X_1, X_2,

$$\phi(\lambda X_1 + (1 - \lambda)X_2) \leqslant \max\{\phi(X_1), \phi(X_2)\}. \tag{5.6}$$

First, suppose that $\lambda = p/n^q$ where $p, q \in \mathbb{N}$. Repeatedly applying (5.5) q times, we get, for all $Y_1, \dots, Y_m$ where $m = n^q$,

$$\phi\left(\sum_{i=1}^{m} \frac{Y_i}{m}\right) \leqslant \max\{\phi(Y_1), \dots, \phi(Y_m)\}. \tag{5.7}$$

Letting $Y_i = X_1$ for $i \leqslant p$ and $Y_i = X_2$ for $i > p$ in (5.7), we get (5.6). Next, consider a general $\lambda \in (0,1)$. Let $X_1' = \lambda X_1/t$ and $X_2' = (1-\lambda)X_2/(1-t)$, where $t = p/n^q \in (0,1)$ for some $p, q \in \mathbb{N}$. Using (5.7), we get

$$\phi(\lambda X_1 + (1-\lambda)X_2) = \phi(tX_1' + (1-t)X_2') \leqslant \max\{\phi(X_1'), \phi(X_2')\}. \tag{5.8}$$

Sending $t \to \lambda$ and using continuity we obtain the desired result. □

5.4.2 Examples and proofs related to portfolio convexity

In the first example, we show that convexity or quasi-convexity of $\mathbf{X} \mapsto D(\mathbf{X})$ should not hold for a diversification index D.

Example 5.1 (Quasi-convexity on $\mathcal{X}^n$ is not desirable)**.** Let $(X, Y) \in \mathcal{X}^2$ represent any diversified portfolio (e.g., with iid normal components), and assume that $Z := (X+Y)/2$ is not a constant. Since the portfolio (Z, Z) relies only on one asset and has no diversification benefit, for a good diversification index D we naturally want $D(Z, Z)$ to be larger than both $D(X, Y)$ and $D(Y, X)$; recall that $D(Z, Z) = 1$ in the setting of Theorem 4.3 (iii). This argument shows that it is unnatural to require D to be convex or quasi-convex on $\mathcal{X}^2$; the case of $\mathcal{X}^n$ is similar. Indeed, if a real-valued D satisfies [SI] and convexity on $\mathcal{X}^n$, then it is a constant; this is shown in the proposition below.

Proposition 5.3. *A mapping $D : \mathcal{X}^n \to \mathbb{R}$ satisfies* [SI] *and convexity if and only if $D(\mathbf{X}) = c$ for all $\mathbf{X} \in \mathcal{X}$ and some constant $c \in \mathbb{R}$.*

Proof. If D is a constant for all $\mathbf{X} \in \mathcal{X}^n$, it is clear that D satisfies [SI] and convexity. Next we will show the "only if" part. Let $d_0 = D(\mathbf{0}) \in \mathbb{R}$.

(i) If $d_0 \geqslant D(\mathbf{X})$ for all $\mathbf{X} \in \mathcal{X}^n$ and there exists $\mathbf{X}_0$ such that $D(\mathbf{X}_0) < d_0$, then

$$D\left(\frac{1}{2}\mathbf{X}_0 + \frac{1}{2}(-\mathbf{X}_0)\right) = D(\mathbf{0}) > \frac{1}{2}D(\mathbf{X}_0) + \frac{1}{2}D(-\mathbf{X}_0),$$

which contradicts the convexity of D.

(ii) If there exists $\mathbf{X}_0$ such that $d_0 < D(\mathbf{X}_0)$, then, by [SI] of D,

$$D\left(\frac{1}{2}\mathbf{0} + \frac{1}{2}\mathbf{X}_0\right) = D(\mathbf{X}_0) > \frac{1}{2}D(\mathbf{0}) + \frac{1}{2}D(\mathbf{X}_0),$$

which contradicts the convexity of D.

By (i) and (ii), we can conclude that D only takes the value d_0. □

From the proof of Proposition 5.3, we see that the conflict between convexity and [SI] holds for real-valued mappings on any closed convex cone, not necessarily on $\mathcal{X}^n$.

In the second example, we see that convexity of $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$ is not desirable either for a good diversification index.

Example 5.2 (Convexity in $\mathbf{w}$ is not desirable)**.** Let Z be standard normal and $\epsilon > 0$ be a small constant. Consider a portfolio vector $\mathbf{X} = ((1-\epsilon)Z, -\epsilon Z)$. Let $\mathbf{w} = (1, 0)$ and $\mathbf{v} = (\epsilon, 1-\epsilon)$. Note that $\mathbf{w} \odot \mathbf{X} = (1-\epsilon)(Z, 0)$ and $\mathbf{v} \odot \mathbf{X} = (\epsilon - \epsilon^2)(Z, -Z)$. The portfolio $\mathbf{w} \odot \mathbf{X}$ is not diversified since it has only one non-zero component, and the portfolio $\mathbf{v} \odot \mathbf{X}$ is perfectly hedged since the sum of its components is 0. Hence, for a good diversification index D, it should hold that $D(\mathbf{w} \odot \mathbf{X}) = 1$ and $D(\mathbf{v} \odot \mathbf{X}) = 0$; Theorem 4.3 confirms this. On the other hand, the portfolio

$$\left(\frac{1}{2}\mathbf{w} + \frac{1}{2}\mathbf{v}\right) \odot \mathbf{X} = \frac{1}{2}\left((1-\epsilon^2)Z, -(\epsilon - \epsilon^2)Z\right)$$

is not well diversified since its second component is very small compared to its first component. Intuitively, for $\epsilon \approx 0$, we expect $D((\mathbf{w}/2 + \mathbf{v}/2) \odot \mathbf{X}) \approx 1 > D(\mathbf{w} \odot \mathbf{X})/2 + D(\mathbf{v} \odot \mathbf{X})/2$. This shows that $\mathbf{w} \mapsto D(\mathbf{w} \odot \mathbf{X})$ is not convex. One can verify that this is indeed true if D is DQ or DR based on commonly used risk measures such as SD, VaR ($\alpha < 1/2$) and ES.

Proof of Proposition 4.1. Since the proof of Theorem 4.2 solely relies on convexity and positive homogeneity of ρ_α to show [PC], it is clear that DQ^ρ_α satisfies [PC].

The subadditivity of ρ_α implies that DQ^ρ_α takes value in $[0, 1]$. Consequently, $\{\mathrm{DQ}^\rho_\alpha(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} \subseteq [0, 1]$. We only need to show $[0, 1] \subseteq \{\mathrm{DQ}^\rho_\alpha(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\}$.

Since ρ_α is non-linear and sub-linear, there exists Y such that $\rho_\alpha(Y)+\rho_\alpha(-Y)>0$ following the argument of Remark 5.2. Consider $\mathbf{X}^\theta=(X,\theta Y,-\theta Y,0,\dots,0)$ with $\theta\geqslant 0$. We have

$$\mathrm{DQ}^\rho_\alpha(\mathbf{X}^\theta)=\frac{1}{\alpha}\inf\{\beta\in I:\rho_\beta(X)\leqslant\rho_\alpha(X)+\theta\rho_\alpha(Y)+\theta\rho_\alpha(-Y)\}.$$

It is clear that $\mathrm{DQ}^\rho_\alpha(\mathbf{X}^0)=1$, and there exists $\tilde\theta$ such that $\mathrm{DQ}^\rho_\alpha(\mathbf{X}^{\tilde\theta})=0$ with $\rho_\alpha(X)+\tilde\theta\rho_\alpha(Y)+\tilde\theta\rho_\alpha(-Y)>\rho_0(X)$. Since $\beta\mapsto\rho_\beta(X)$ is strictly decreasing, its generalized inverse is continuous and we can conclude $\{\mathrm{DQ}^\rho_\alpha(\mathbf{X}^\beta):\beta\in[0,\tilde\theta]\}=[0,1]$. Hence, $[0,1]\subseteq\{\mathrm{DQ}^\rho_\alpha(\mathbf{X}):\mathbf{X}\in\mathcal{X}^n\}$. □

5.4.3 Constructing DQ from a single risk measure

In this section, we discuss how to construct DQ from only a single risk measure ϕ. For commonly used risk measures like VaR and ES, a natural family ρ with $\rho_\alpha=\phi$ exists. If in some applications one needs to use a different ϕ which does not belong to an existing family, we will need to construct a family of risk measures for ϕ.

First, suppose that ϕ is MCP. A simple approach is to take $\rho_\alpha=(1-\alpha)\text{ess-sup}+\alpha\phi$ for $\alpha\in(0,1)$. Clearly, $\rho_1=\phi$. As ϕ is MCP, we have $\phi(X)\leqslant\phi(\text{ess-sup}(X))=\text{ess-sup}(X)$ for all $X\in L^\infty$. Hence, ρ is a decreasing class of MCP risk measures. Therefore, DQ^ρ_1 satisfies the six axioms in Theorem 4.1. Moreover, by checking the definition, this DQ has an explicit formula

$$\mathrm{DQ}^\rho_1(\mathbf{X})=\left(\frac{\text{ess-sup}\left(\sum_{i=1}^n X_i\right)-\sum_{i=1}^n\phi(X_i)}{\text{ess-sup}\left(\sum_{i=1}^n X_i\right)-\phi\left(\sum_{i=1}^n X_i\right)}\right)_+.$$

If $\sum_{i=1}^n X_i\leqslant\sum_{i=1}^n\phi(X_i)$, we have $\text{ess-sup}(\sum_{i=1}^n X_i)\leqslant\sum_{i=1}^n\phi(X_i)$ and $\mathrm{DQ}^\rho_1(\mathbf{X})=0$; this is also reflected by Theorem 4.3 when $\text{ess-sup}(\sum_{i=1}^n X_i)>\phi(\sum_{i=1}^n X_i)$.

For any arbitrary risk measure ϕ, we can always define the decreasing family $\{\phi_+/\alpha:\alpha\in I\}$ for constructing DQ; here ϕ_+ is the positive part of ϕ. This approach leads to DQs that are also DRs.

Proposition 5.4. *For a given $\phi:\mathcal{X}\to\mathbb{R}_+$, let $\rho=(\phi/\alpha)_{\alpha\in(0,\infty)}$. For $\alpha\in(0,\infty)$, we have $\mathrm{DQ}^\rho_\alpha=\mathrm{DR}^\phi$. The same holds if $\rho=(b\mathbb{E}+c\phi/\alpha)_{\alpha\in(0,\infty)}$ for some $b\in\mathbb{R}$ and $c>0$ and $\mathcal{X}=L^1$.*

Proof. First, we compute α^* by the definition of $\mathrm{DQ}^{\rho}_{\alpha}$. For any $\mathbf{X} \in (L^1)^n$,

$$\begin{aligned}\alpha^* &= \inf\left\{\beta \in (0,\infty) : b\mathbb{E}\left[\sum_{i=1}^n X_i\right] + \frac{c}{\beta}\phi\left(\sum_{i=1}^n X_i\right) \leqslant b\sum_{i=1}^n \mathbb{E}\left[X_i\right] + \sum_{i=1}^n \frac{c}{\alpha}\phi(X_i)\right\} \\ &= \inf\left\{\beta \in (0,\infty) : \frac{\phi\left(\sum_{i=1}^n X_i\right)}{\beta} \leqslant \frac{\sum_{i=1}^n \phi(X_i)}{\alpha}\right\}.\end{aligned}$$

If $\phi(\sum_{i=1}^n X_i) = 0$ and $\sum_{i=1}^n \phi(X_i) = 0$, then $\alpha^* = 0$. If $\phi(\sum_{i=1}^n X_i) > 0$ and $\sum_{i=1}^n \phi(X_i) = 0$, then $\alpha^* = \infty$ because the set on which the infimum is taken is empty. If $\phi(\sum_{i=1}^n X_i) > 0$ and $\sum_{i=1}^n \phi(X_i) > 0$, then $\alpha^* = \alpha\phi(\sum_{i=1}^n X_i)/\sum_{i=1}^n \phi(X_i)$. Hence, $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \mathrm{DR}^{\phi}(\mathbf{X})$ holds for all $\mathbf{X} \in (L^1)^n$. By the same argument, for $\rho = (\phi/\alpha)_{\alpha\in(0,\infty)}$, we get $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \mathrm{DR}^{\phi}(\mathbf{X})$ for all $\mathbf{X} \in \mathcal{X}^n$. □

As a result of Proposition 5.4, DQ built on the family ρ of the mean-SD functions given by $\rho_\alpha(X) = \mathbb{E}[X] + \mathrm{SD}(X)/\alpha$ is precisely $\mathrm{DR}^{\mathrm{SD}}$.

5.4.4 Axiomatization of DQ using preferences

The axioms $[\mathrm{R}]_\phi$, $[\mathrm{N}]_\phi$ and $[\mathrm{C}]_\phi$ are formulated based on an exogenously specified risk measure ϕ, usually by financial regulation. This choice can also be endogenized in the context of internal decision making. In this section, we provide an axiomatization of DQ as in Theorem 4.1 without specifying a risk measure ϕ. We first define the preference of a decision maker over risks. A preference relation $\succeq$ is defined by a non-trivial total preorder[1] on $\mathcal{X}$. As usual, $\succ$ and $\simeq$ correspond to the antisymmetric and equivalence relations, respectively. On the preference $\succeq$ of risk, the relation $X \succeq Y$ means the agent prefers X to Y for any $X, Y \in \mathcal{X}$. We will use the following axioms.

[A1] $X \leqslant Y \implies X \succeq Y$.

[A2] $X \succeq Y \implies X + c \succeq Y + c$ for any $c \in \mathbb{R}$.

[1] A preorder is a binary relation on $\mathcal{X}$, which is reflexive and transitive. A binary relation $\succeq$ is reflexive if $X \succeq X$ for all $X \in \mathcal{X}$, and transitive if $X \succeq Y$ and $Y \succeq Z$ imply $X \succeq Z$. A non-trivial total preorder is a preorder that in addition is complete, that is, $X \succeq Y$ or $Y \succeq X$ for all $X, Y \in \mathcal{X}$, and there exist at least two alternatives X, Y such that X is preferred over Y strictly.

[A3] $X \succeq Y \implies \lambda X \succeq \lambda Y$ for any $\lambda > 0$.

[A4] For any $X \in \mathcal{X}$, there exists $c \in \mathbb{R}$ such that $X \simeq c$.

The four axioms are rather standard and we only briefly explain them. The axiom [A1] means that the agent always prefers a smaller loss. The axioms [A2] and [A3] mean that if the agent prefers one random loss over another, then this is preserved under any strictly increasing linear transformations. The axiom [A4] implies that any random losses can be equally favourable as a constant loss which is commonly referred to as a certainty equivalence.

A numerical representation of a preference $\succeq$ is a mapping $\phi : \mathcal{X} \to \mathbb{R}$, such that $X \succeq Y \iff \phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$. In other words, $\succeq$ is the preference of an agent favouring less risk evaluated via ϕ. There is a simple relationship between preferences satisfying [A1]-[A4] and MCP risk measures.

Lemma 5.3. *A preference satisfies* [A1]–[A4] *if and only if it can be represented by an MCP risk measure* ϕ.

Proof. The "if" statement is straightforward to check, and we will show the "only if" statement. The preference $\succeq$ can be represented by a risk measure ϕ through $X \succeq Y \iff \phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ since $\succeq$ is separable by [A1] and [A4]; see Debreu (1954) and Drapeau and Kupper (2013). If $\phi(0) = \phi(1)$, then by using [A1]–[A3], the preference $\succeq$ is trivial, contradicting our assumption on $\succeq$. Hence, using [A1], $\phi(0) < \phi(1)$, we can further let $\phi(0) = 0$ and $\phi(1) = 1$. It is then straightforward to verify that ϕ is MCP from [A1]-[A3]. □

Similarly to Section 4.3, but with the preference $\succeq$ replacing the risk measure ϕ, we denote by $\mathbf{X} \overset{\mathrm{m}}{\simeq} \mathbf{Y}$ if $X_i \simeq Y_i$ for each $i \in [n]$, by $\mathbf{X} \overset{\mathrm{m}}{\succeq} \mathbf{Y}$ if $X_i \succeq Y_i$ for each $i \in [n]$, and by $\mathbf{X} \overset{\mathrm{m}}{\succ} \mathbf{Y}$ if $X_i \succ Y_i$ for each $i \in [n]$. With this new formulation and everything else unchanged, the axioms of rationality, normalization and continuity are now denoted by $[\mathrm{R}]_{\succeq}$, $[\mathrm{N}]_{\succeq}$ and $[\mathrm{C}]_{\succeq}$.

Proposition 5.5. *A diversification index* $D : \mathcal{X}^n \to \overline{\mathbb{R}}$ *satisfies* [+], [LI], [SI], $[\mathrm{R}]_{\succeq}$, $[\mathrm{N}]_{\succeq}$ *and* $[\mathrm{C}]_{\succeq}$ *for some preference* $\succeq$ *satisfying* [A1]–[A4] *if and only if* D *is* $\mathrm{DQ}^{\rho}_{\alpha}$ *for some decreasing families* ρ *of MCP risk measures. Moreover, in both directions of the above equivalence, it can be required that* ρ_α *represents* $\succeq$.

Proof. The proof follows from Theorem 4.1 by noting that Lemma 5.3 allows us to convert between a preference $\succeq$ satisfying [A1]-[A4] and an MCP risk measure ϕ. □

Theorem 4.2 also admits a formulation via preferences similar to Proposition 5.5.

5.4.5 Uniqueness of the risk measure family representing a DQ

Proposition 5.6 below shows that the choice of the risk measure family is unique up to strictly increasing transformation of the parameter if the ordering structure on portfolio diversification is specified by a given ordering relation $\succeq$ on $\mathcal{X}^n$ that can be numerically represented by a DQ.

Proposition 5.6. *Let $n \geqslant 3$, $I = [0,1]$, $\alpha \in I$ and ϕ is a positively homogeneous risk measure with $\phi(Y)+\phi(-Y) > 0$ for some $Y \in \mathcal{X}$. Suppose that a weak order $\succeq$ is numerically represented by both $\mathrm{DQ}^{\rho}_{\alpha}$ and $\mathrm{DQ}^{\tau}_{\alpha}$ such that $\{\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = \{\mathrm{DQ}^{\tau}_{\alpha}(\mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\} = [0,1]$, where $\rho = (\rho_\beta)_{\beta\in I}$ and $\tau = (\tau_\beta)_{\beta\in I}$ are continuous decreasing families of risk measures satisfying $\phi = \rho_\alpha = \tau_\alpha$. Then, there exists a strictly increasing $f : (0,\alpha) \to (0,\alpha)$ such that $\tau_\beta = \rho_{f(\beta)}$ for all $\beta \in (0,\alpha)$.*

Proof. Since $\succeq$ is represented by both $\mathrm{DQ}^{\rho}_{\alpha}$ and $\mathrm{DQ}^{\tau}_{\alpha}$, there exists a strictly increasing function $g : [0,1] \to [0,1]$ such that $\mathrm{DQ}^{\rho}_{\alpha} = g(\mathrm{DQ}^{\tau}_{\alpha})$. Let $f(\beta) = g(\beta/\alpha)$ for $\beta \in (0,\alpha)$.

Assume that there exists $\beta^* \in (0,\alpha)$ such that $\tau_{\beta^*}(X) > \rho_{f(\beta^*)}(X)$. By positive homogeneity of ϕ and $\phi(Y) + \phi(-Y) > 0$, there exists $\epsilon > 0$ such that $\tau_{\beta^*}(X) > \phi(X) + \phi(\epsilon Y) + \phi(-\epsilon Y) > \rho_{f(\beta^*)}(X)$. Let $\mathbf{X} = (X, \epsilon Y, -\epsilon Y, 0, \ldots, 0)$. Since $\beta \mapsto \rho_\beta(X)$ and $\beta \mapsto \tau_\beta(X)$ are continuous, we have

$$g(\mathrm{DQ}^{\tau}_{\alpha}(\mathbf{X})) = f\left(\inf\left\{\beta \in I : \tau_\beta(X) \leqslant \phi(X) + \phi(\epsilon Y) + \phi(-\epsilon Y)\right\}\right) > f(\beta^*)$$

and

$$\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \inf\left\{\beta \in I : \rho_\beta(X) \leqslant \phi(X) + \phi(\epsilon Y) + \phi(-\epsilon Y)\right\} \leqslant f\left(\beta^*\right),$$

which contradicts $\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = g(\mathrm{DQ}^{\tau}_{\alpha}(\mathbf{X}))$. □

5.5 Additional results and proofs for Section 4.4

In this section, we present additional results, proofs, and discussions supplementing Sections 4.4.2 and 4.4.3.

5.5.1 Worst-case and best-case dependence for DQ (Section 4.4.2)

We assume that two random vectors $\mathbf{X}$ and $\mathbf{Y}$ have the same marginal distributions, and we study the effect of the dependence structure. We will assume that a tuple of distributions $\mathbf{F} = (F_1, \dots, F_n)$ is given and each component has a finite mean. Let

$$\mathcal{Y}_{\mathbf{F}} = \{(X_1, \dots, X_n) : X_i \sim F_i \text{ for each } i = 1, \dots, n\}.$$

For $\mathbf{X}, \mathbf{Y} \in \mathcal{Y}_{\mathbf{F}}$, we say that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in sum-convex order, denoted by $\mathbf{X} \leqslant_{\mathrm{scx}} \mathbf{Y}$, if $\sum_{i=1}^n X_i \geqslant_{\mathrm{SSD}} \sum_{i=1}^n Y_i$; see Corbett and Rajaram (2006). We refer to Shaked and Shanthikumar (2007) for a general treatment of multivariate stochastic orders. With arbitrary dependence structures, the best-case value and worst-case value of DQ_α^ρ are given by

$$\inf_{\mathbf{X} \in \mathcal{Y}_{\mathbf{F}}} \mathrm{DQ}_\alpha^\rho(\mathbf{X}) \quad \text{and} \quad \sup_{\mathbf{X} \in \mathcal{Y}_{\mathbf{F}}} \mathrm{DQ}_\alpha^\rho(\mathbf{X}).$$

For some mapping on $\mathcal{X}^n$, finding the best-case and worst-case values and structures over $\mathcal{Y}_{\mathbf{F}}$ is known as a problem of risk aggregation under dependence uncertainty; see Bernard et al. (2014) and Embrechts et al. (2015).

If $\rho = (\rho_\alpha)_{\alpha \in I}$ is a class of SSD-consistent risk measures such as ES, then, by Proposition 4.3, DQ_α^ρ is consistent with the sum-convex order on $\mathcal{Y}_{\mathbf{F}}$. This leads to the following observations on the corresponding dependence structures.

(i) It is well-known (e.g., Rüschendorf (2013)) that the $\leqslant_{\mathrm{scx}}$-largest element of $\mathcal{Y}_{\mathbf{F}}$ is comonotonic, and thus a comonotonic random vector has the largest DQ_α^ρ in this case. Note that such ρ does not include VaR. Indeed, as we have seen from Proposition 4.5, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 1$ for comonotonic $\mathbf{X}$ under mild conditions, which is not equal to its largest value n.

(ii) In case $n = 2$, the $\leqslant_{\mathrm{scx}}$-smallest element of $\mathcal{Y}_{\mathbf{F}}$ is counter-comonotonic, and thus a comonotonic random vector has the smallest $\mathrm{DQ}^{\rho}_{\alpha}$.

(iii) For $n \geqslant 3$, the $\leqslant_{\mathrm{scx}}$-smallest elements of $\mathcal{Y}_{\mathbf{F}}$ are generally hard to obtain. If each pair (X_i, X_j) is counter-monotonic for $i \neq j$, then $\mathbf{X}$ is a $\leqslant_{\mathrm{scx}}$-smallest element of $\mathcal{Y}_{\mathbf{F}}$. Pairwise counter-monotonicity puts very strong restrictions on the marginal distributions. For instance, it rules out all continuous marginal distributions; see Puccetti and Wang (2015).

(iv) If a joint mix, i.e., a random vector with a constant component-wise sum, exists in $\mathcal{Y}_{\mathbf{F}}$, then any joint mix is a $\leqslant_{\mathrm{scx}}$-smallest element of $\mathcal{Y}_{\mathbf{F}}$ by Jensen's inequality. See Puccetti and Wang (2015) and Wang and Wang (2016) for results on the existence of joint mixes. In case a joint mix does not exist, the $\leqslant_{\mathrm{scx}}$-smallest elements are obtained by Bernard et al. (2014) and Jakobsons et al. (2016) under some conditions on the marginal distributions such as monotonic densities.

In optimization problems over dependence structures (see e.g., Rüschendorf (2013) and Embrechts et al. (2015)), the above observations yield guidelines on where to look for the optimizing structures.

5.5.2 Proofs and related discussions on RI and RC (Section 4.4.3)

Here we present the proof of Proposition 4.4 and an additional result (Proposition 5.7) on the properties RI and RC.

Proof of Proposition 4.4. (i) For any $n \in \mathbb{N}$, $\mathbf{X} \in (L^p)^n$ and $c \in \mathbb{R}$, by $[\mathrm{CA}]_m$ of $(\rho_\alpha)_{\alpha\in I}$,

$$\begin{aligned}\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}, c) &= \frac{1}{\alpha}\inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^{n} X_i + c\right) \leqslant \sum_{i=1}^{n}\rho_\alpha(X_i) + \rho_\alpha(c)\right\}\\ &= \frac{1}{\alpha}\inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^{n} X_i\right) + mc \leqslant \sum_{i=1}^{n}\rho_\alpha(X_i) + mc\right\} = \mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}),\end{aligned}$$

and hence $\mathrm{DQ}^{\rho}_{\alpha}$ satisfies [RI].

(ii) For any $n \in \mathbb{N}$ and $\mathbf{X} \in (L^p)^n$, by [PH] of $(\rho_\alpha)_{\alpha \in I}$,

$$\mathrm{DQ}^\rho_\alpha(\mathbf{X}, \mathbf{X}) = \frac{1}{\alpha} \inf \left\{ \beta \in I : \rho_\beta \left(2 \sum_{i=1}^n X_i \right) \leqslant 2 \sum_{i=1}^n \rho_\alpha(X_i) \right\} = \mathrm{DQ}^\rho_\alpha(\mathbf{X}),$$

and hence DQ^ρ_α satisfies [RC]. □

Proposition 5.7. *Let $\phi : L^p \to \mathbb{R}$ be a continuous and law-invariant risk measure.*

(i) Suppose that DR^ϕ is not degenerate for some input dimension. Then DR^ϕ satisfies [RI] *and* [+] *if and only if ϕ satisfies* $[\mathrm{CA}]_0$, [±] *and $\phi(0) = 0$.*

(ii) If ϕ satisfies [PH]*, then DR^ϕ satisfies* [RC].

Proof. (i) We first show the "if" part. If ϕ satisfies $[\mathrm{CA}]_0$ and $\phi(0) = 0$, then $\phi(c) = \phi(0) = 0$ for all $c \in \mathbb{R}$. For any $n \in \mathbb{N}$, $\mathbf{X} \in (L^p)^n$ and $c \in \mathbb{R}$,

$$\mathrm{DR}^\phi(\mathbf{X}, c) = \frac{\phi\left(\sum_{i=1}^n X_i + c\right)}{\sum_{i=1}^n \phi(X_i) + \phi(c)} = \frac{\phi\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \phi(X_i)} = \mathrm{DR}^\phi(\mathbf{X}).$$

Thus, DR^ϕ satisfies [RI].

For the "only if" part, we first assume $\phi(0) \neq 0$. Since DR^ϕ satisfies [RI], for all $n \in \mathbb{N}$, $c \in \mathbb{R}$ and $\mathbf{X} = \mathbf{0} \in \mathbb{R}^n$, we have

$$\mathrm{DR}^\phi(\mathbf{X}, c) = \frac{\phi(c)}{n\phi(0) + \phi(c)} = \mathrm{DR}^\phi(\mathbf{X}) = \frac{\phi(0)}{n\phi(0)} = \frac{1}{n}.$$

The above equality means that $\phi(c) = n\phi(0)/(n-1)$ holds for any $n \in \mathbb{N}$ and $c \in \mathbb{R}$, and thus we have $\phi(0) = 0$, which violates the assumption $\phi(0) \neq 0$. Hence, $\phi(0) = 0$.

If there exists $c_1 \in \mathbb{R}$ such that $\phi(c_1) \neq 0$, then by [RI] and $\phi(0) = 0$, we have

$$\mathrm{DR}^\phi(c_1, 0, 0, \dots, 0, c) = \frac{\phi(c_1 + c)}{\phi(c_1) + \phi(c)} = \mathrm{DR}^\phi(c_1, 0, 0, \dots, 0) = \frac{\phi(c_1)}{\phi(c_1)} = 1,$$

and thus $\phi(c_1 + c) = \phi(c_1) + \phi(c)$ as long as $\phi(c_1)$ or $\phi(c)$ is not zero. If both of $\phi(c_1)$ and $\phi(c)$ are 0, then $\phi(c_1 + c) = 0$. To sum up, ϕ is additive on $\mathbb{R}$. Since ϕ is also continuous on $\mathbb{R}$, we know that ϕ is linear, that is, $\phi(c) = \beta c$ for some $\beta \in \mathbb{R}$.

Suppose that there exists X such that $\phi(X) \neq 0$; otherwise there is nothing to show. Using [RI] and $\phi(0) = 0$, we have, for $c \in \mathbb{R}$,

$$\mathrm{DR}^\phi(X, 0, 0, \dots, 0, c) = \frac{\phi(X + c)}{\phi(X) + \phi(c)} = \mathrm{DR}^\phi(X, 0, \dots, 0) = 1,$$

which implies $\phi(X+c) = \phi(X) + \phi(c) = \phi(X) + \beta c$.

Using the fact that DR^ϕ is not degenerate for some dimension n, there exists $\mathbf{X} = (X_1, \dots, X_n)$ such that $\mathrm{DR}^\phi(\mathbf{X}) \in \mathbb{R}\backslash\{0,1\}$. Note that $\phi(\sum_{i=1}^n X_i) \neq 0$ and $\sum_{i=1}^n \phi(X_i) \neq 0$. Hence,

$$\mathrm{DR}^\phi(\mathbf{X}, 1) = \frac{\phi(\sum_{i=1}^n X_i + 1)}{\sum_{i=1}^n \phi(X_i) + \phi(1)} = \frac{\phi(\sum_{i=1}^n X_i) + \beta}{\sum_{i=1}^n \phi(X_i) + \beta} = \mathrm{DR}^\phi(\mathbf{X}) = \frac{\phi(\sum_{i=1}^n X_i)}{\sum_{i=1}^n \phi(X_i)}.$$

This implies $\beta = 0$, $\phi(c) = 0$ for all $c \in \mathbb{R}$ and $\phi(X+c) = \phi(X)$ for all $X \in L^p$ such that $\phi(X) \neq 0$. For any $X \in L^p$ such that $\phi(X) = 0$ and $c \in \mathbb{R}$, we have

$$\mathrm{DR}^\phi(X, 0, \dots, 0, c) = \frac{\phi(X+c)}{\phi(X) + \phi(c)} = \frac{\phi(X+c)}{\phi(X)} = \frac{\phi(X)}{\phi(X)} = \mathrm{DR}^\phi(X, 0, \dots, 0),$$

which implies $\phi(X+c) = 0 = \phi(X)$. Therefore, ϕ satisfies $[\mathrm{CA}]_0$.

Finally, we show ϕ is either non-negative or non-positive by considering the following three cases.

(a) Assume that there exists $X \in L^p$ such that $\phi(X) + \phi(-X) > 0$. If there exists $Y \in L^p$ such that $\phi(Y) < 0$, then by continuity of ϕ and $\phi(0) = 0$, there exists $m > 0$ such that $0 < -\phi(mY) < \phi(X) + \phi(-X)$. We have

$$\mathrm{DR}^\phi(mY, X, -X, 0, \dots, 0) = \frac{\phi(mY)}{\phi(mY) + (\phi(X) + \phi(-X))} < 0,$$

which contradicts the fact that DR^ϕ is non-negative. Hence, $\phi(Y) \geqslant 0$ for all $Y \in L^\infty$.

(b) By the same argument, if there exists $X \in L^p$ such that $\phi(X) + \phi(-X) < 0$, then $\phi(Y) \leqslant 0$ for all $Y \in L^\infty$.

(c) Assume $\phi(X) + \phi(-X) = 0$ for all $X \in L^\infty$. Suppose that there exists $Y \in L^\infty$ such that $\phi(Y) < 0$. Using Lemma 1 of Wang and Wu (2020) again, there exist $Z, Z' \in L^\infty$ satisfying $Z \overset{\mathrm{d}}{=} Z'$ and $Z - Z' \overset{\mathrm{d}}{=} Y - \mathbb{E}[Y]$. For $\mathbf{Z} = (Z, -Z', 0, \dots, 0)$, using the law invariance of ϕ, we have

$$\mathrm{DR}^\phi(\mathbf{Z}) = \frac{\phi(Z - Z')}{\phi(Z) + \phi(-Z')} = \frac{\phi(Y - \mathbb{E}[Y])}{\phi(Z) + \phi(-Z')} = \frac{\phi(Y)}{\phi(Z) + \phi(-Z)} = \frac{\phi(Y)}{0} = -\infty,$$

which contradicts $\mathrm{DR}^\phi(\mathbf{Z}) \geqslant 0$. Hence, $\phi(X) \geqslant 0$ for all $X \in L^\infty$. Together with $\phi(X) + \phi(-X) = 0$, we get $\phi(X) = 0$. To extend this to L^p, we simply

use continuity. For $X \in L^p$, let $Y_M = (X \wedge M) \vee (-M)$. Hence, $Y_M \in L^\infty$ and $Y_M \xrightarrow{L^p} X$ as $M \to \infty$. As a result, we have $\phi(X) = \lim_{M\to\infty} \phi(Y_M) = 0$.

In conclusion, we have $\phi(Y) \geqslant 0$ or $\phi(Y) \leqslant 0$ for all $X \in L^p$. Case (c) is not possible because it contradicts that DR^ϕ is not degenerate. Cases (a) and (b) are possible, corresponding to, for instance, (a) $\phi = \mathrm{SD}$; (b) $\phi = -\mathrm{SD}$.

(ii) If ϕ satisfies [PH], then for any $n \in \mathbb{N}$ and $\mathbf{X} \in (L^p)^n$,

$$\mathrm{DR}^\phi(\mathbf{X}, \mathbf{X}) = \frac{\phi\left(2\sum_{i=1}^n X_i\right)}{2\sum_{i=1}^n \phi(X_i)} = \frac{\phi\left(\sum_{i=1}^n X_i\right)}{\sum_{i=1}^n \phi(X_i)} = \mathrm{DR}^\phi(\mathbf{X}).$$

Hence, DR^ϕ satisfies [RC]. □

In Proposition 5.7, we show that if [RI] is assumed, then the only option for DR is to use a non-negative ϕ (we can use $-\phi$ if ϕ is non-positive) such as var or SD. By Proposition 5.4, all such DRs belong to the class of DQs.

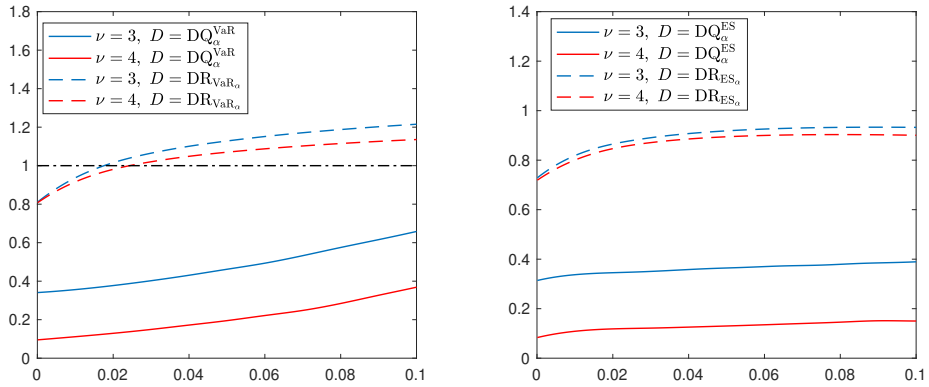
5.6 Additional results and proofs for Section 4.5

In this section, we present the proof for Proposition 4.5 and an additional numerical result to complement those in Section 4.5.2.

Proof of Proposition 4.5. This statement on ES follows from Proposition 4.1; for the one on VaR, see Theorem 1 (i) of Chapter 6. □

We look at the models $\mathbf{Y}'$ and $\mathbf{Y}$ in the setting of Tables 4.2 and 4.3. In Figure 5.1, we observe that the values of $D(\mathbf{Y}')/D(\mathbf{Y})$ for $D = \mathrm{DQ}_\alpha^{\mathrm{VaR}}$ or $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ are always smaller than 1 for $\alpha \in (0, 0.1]$, while the values of $D(\mathbf{Y}')/D(\mathbf{Y})$ for $D = \mathrm{DR}^{\mathrm{VaR}_\alpha}$ are only smaller than 1 when α is relatively small. We always observe that, if the desired relation $D(\mathbf{Y}')/D(\mathbf{Y}) < 1$ holds for $D = \mathrm{DR}^{\mathrm{VaR}_\alpha}$ or $\mathrm{DR}^{\mathrm{ES}_\alpha}$ then it holds for $D = \mathrm{DQ}_\alpha^{\mathrm{VaR}}$ or $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, but the converse does not hold. This means that if the iid model is preferred to the common shock model by DR, then it is also preferred by DQ, but in many situations, it is only preferred by DQ not by DR. Similarly to Tables 4.2 and 4.3, the iid normal model shows a stronger diversification according to DQ, and this is not the case for DR.

Figure 5.1: $D(\mathbf{Y}')/D(\mathbf{Y})$ based on VaR and ES for $\alpha \in (0, 0.1]$ with fixed $n = 10$



5.7 Proofs for Section 4.6

Proof of Proposition 4.6. For the case of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$, (4.4) in Theorem 4.4 gives that to minimize $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ is equivalent to minimize

$$\mathbb{P}\left(\mathbf{w}^\top \mathbf{X} > \mathbf{w}^\top \mathbf{x}_\alpha^{\mathrm{VaR}}\right) = \mathbb{P}\left(\mathbf{w}^\top \left(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{VaR}}\right) > 0\right) \qquad \text{over } \mathbf{w} \in \Delta_n.$$

Next, we discuss the case of $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})$. Let $f(\mathbf{v}) = \mathbb{E}[(\mathbf{v}^\top(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1)_+]$ for $\mathbf{v} \in \mathbb{R}_+^n$. It is clear that f is convex. Furthermore, for any $i \in [n]$, we have, for almost every $\mathbf{v} \in \mathbb{R}_+^n$,

$$\begin{aligned}
\frac{\partial f}{\partial v_i}(\mathbf{v}) &= \mathbb{E}\left[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\{\mathbf{v}^\top(\mathbf{X}-\mathbf{x}_\alpha^{\mathrm{ES}})+1>0\}}\right] \\
&= \mathbb{E}\left[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\left\{\{\mathbf{v}^\top(\mathbf{X}-\mathbf{x}_\alpha^{\mathrm{ES}})+1>0\}\cap\{X_i-\mathrm{ES}_\alpha(X_i)>0\}\right\}}\right] \\
&\quad + \mathbb{E}\left[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\left\{\{\mathbf{v}^\top(\mathbf{X}-\mathbf{x}_\alpha^{\mathrm{ES}})+1>0\}\cap\{X_i-\mathrm{ES}_\alpha(X_i)<0\}\right\}}\right].
\end{aligned}$$

The set $\{(\mathbf{v}^\top\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1 > 0\} \cap \{X_i - \mathrm{ES}_\alpha(X_i) > 0\}$ increases in v_i and the set $\{(\mathbf{v}^\top\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1 > 0\} \cap \{X_i - \mathrm{ES}_\alpha(X_i) < 0\}$ decreases in v_i. Hence, $v_i \mapsto \partial f/\partial v_i(\mathbf{v})$ is increasing. Furthermore, $\partial f/\partial v_i(\mathbf{v}) \to \mathbb{E}[(X_i - \mathrm{ES}_\alpha(X_i))\mathbb{1}_{\{X_i-\mathrm{ES}_\alpha(X_i)>0\}}] > 0$ as $v_i \to \infty$. Also, $\partial f/\partial v_i(\mathbf{v}) \to \mathbb{E}[X_i - \mathrm{ES}_\alpha(X_i)] < 0$ as $\mathbf{v} \downarrow \mathbf{0}$ component-wise. Hence, there exists a minimizer $\mathbf{v}^*$ of the problem $\min_{\mathbf{v}\in\mathbb{R}_+^n\setminus\{\mathbf{0}\}} \mathbb{E}[(\mathbf{v}^\top(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1)_+]$.

Let $A = \{\mathbf{v} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\} : \mathbb{P}(\mathbf{v}(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) > 0) > 0\}$ and $B = \{\mathbf{v} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\} : \mathbb{P}(\mathbf{v}(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) > 0) = 0\}$. If B is empty, it is clear that $\min_{\mathbf{w}\in\Delta_n} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w} \odot \mathbf{X}) = \min_{\mathbf{v}\in\mathbb{R}_+^n\setminus\{\mathbf{0}\}} \mathbb{E}[(\mathbf{v}^\top(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1)_+]$ by Theorem 4.4.

If B is not empty, assume $\mathbf{v}^* \in A$. For any $\mathbf{v}_A \in A$, $\mathbf{v}_B \in B$ and $k > 0$, we have

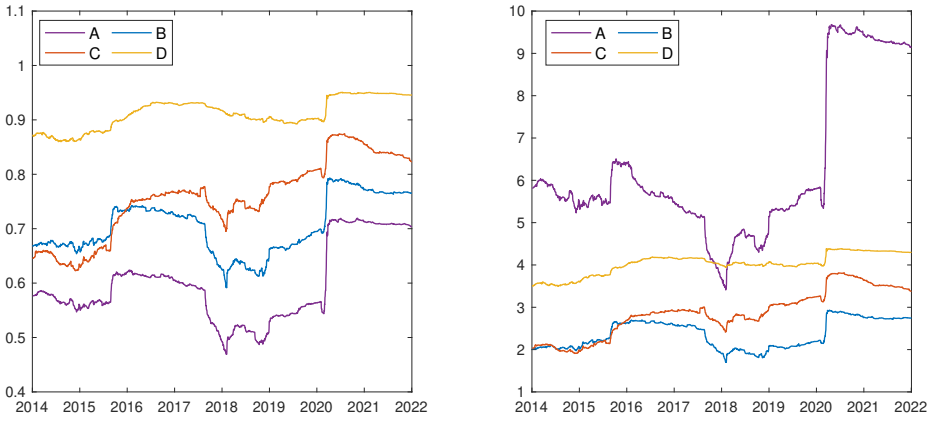
$$\mathbb{E}\left[\left((\mathbf{v}_A + k\mathbf{v}_B)^\top(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1\right)_+\right] \leqslant \mathbb{E}\left[\left(\mathbf{v}_A^\top(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) + 1\right)_+\right].$$

This implies $f(\mathbf{v}^* + k\mathbf{v}_B) = f(\mathbf{v}^*)$ for all $k > 0$, which contradicts $\partial f/\partial v_i(\mathbf{v}) > 0$ as $v_i \to \infty$. Hence, we have $\mathbf{v}^* \in B$. For $\mathbf{w}^* = \mathbf{v}^*/\|\mathbf{v}^*\|$, we have $\mathbb{P}((\mathbf{w}^*)^\top(\mathbf{X} - \mathbf{x}_\alpha^{\mathrm{ES}}) > 0) = 0$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w}^* \odot \mathbf{X}) = 0$ by Theorem 4.4, which means that $\mathbf{w}^*$ is the minimizer of the problem $\min_{\mathbf{w}\in\Delta_n} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w} \odot \mathbf{X})$. □

5.8 Additional empirical results for Section 4.7

In this section, we present some omitted empirical results to complement those in Sections 4.7.2 and 4.7.3. In Section 4.7.2, the values of DQs based on VaR and ES are reported under different portfolio compositions of stocks during the period from 2014 to 2022. Using the same stock compositions in (A)-(D), we calculate the values of DRs based on SD and var (recall that they are also DQs), to see how they perform. The results are reported in Figure 5.2.

Figure 5.2: DRs based on SD (left) and var (right)

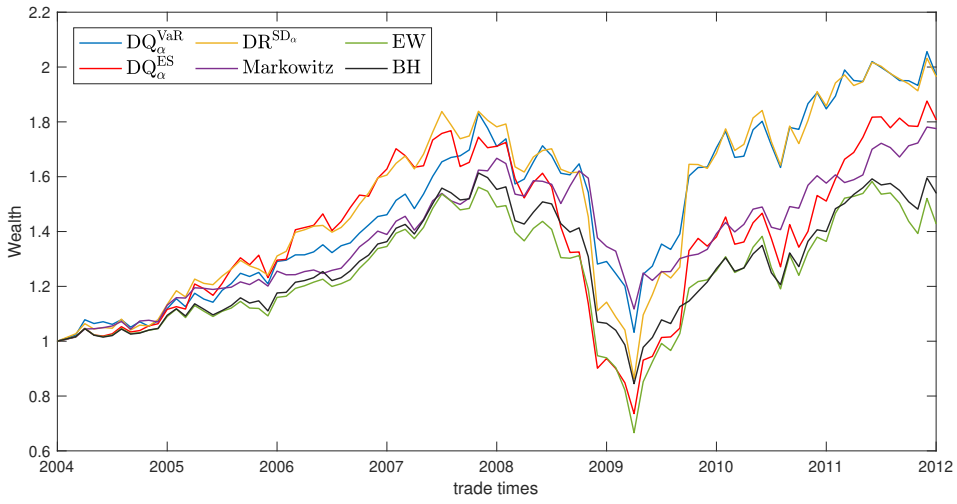


We can see that the same intuitive order (A)⩽(B)⩽(C)⩽(D) as in Figure 4.3 in Section 4.7.2 holds for $\mathrm{DR}^{\mathrm{SD}}$, showing some consistency between DQs based on VaR and ES and $\mathrm{DR}^{\mathrm{SD}}$. The values of $\mathrm{DR}^{\mathrm{SD}}$ are between 0 and 1. On the other hand, the values of $\mathrm{DR}^{\mathrm{var}}$ are

all larger than 1, and portfolio (A) of 20 stocks has the weakest diversification effect according to $\mathrm{DR}^{\mathrm{var}}$ among the four compositions. This is not in line with our intuition, but is to be expected since variance has a different scaling effect than SD, and more correlated stocks lead to a larger value of $\mathrm{DR}^{\mathrm{var}}$ in general. For example, $\mathrm{DR}^{\mathrm{var}}$ equals 1 even for an iid normal model of arbitrarily large dimension (which is often considered as quite well-diversified), and $\mathrm{DR}^{\mathrm{var}}$ equals n if the portfolio has one single asset. These observations show that $\mathrm{DR}^{\mathrm{var}}$ is difficult to interpret if it is used to measure diversification across dimensions.

In Section 4.7.3, we used the period from January 3, 2012, to December 31, 2021, to build up the portfolios. Next, we consider two different datasets from Section 4.7.3, first using the period 2002-2011 and second using 20 instead of 40 stocks, to see how the results vary.

Figure 5.3: Wealth processes for portfolios, 40 stocks, Jan 2004 - Dec 2011



For the first experiment, we choose the four largest stocks from each of the 10 different sectors of S&P 500 ranked by market cap in 2002 as the portfolio compositions and use the period from January 3, 2002, to December 31, 2011, to build up the portfolio. The risk-free rate $r = 4.38\%$, and the target annual expected return for the Markowitz portfolio is set to 5% due to infeasibility of setting 10%. The results are reported in Figures 5.3, 5.4 and Table 5.1.

For the second experiment, we choose the top two stocks from each sector to build the

Figure 5.4: Cumulative portfolio weights, 40 stocks, Jan 2004 - Dec 2011

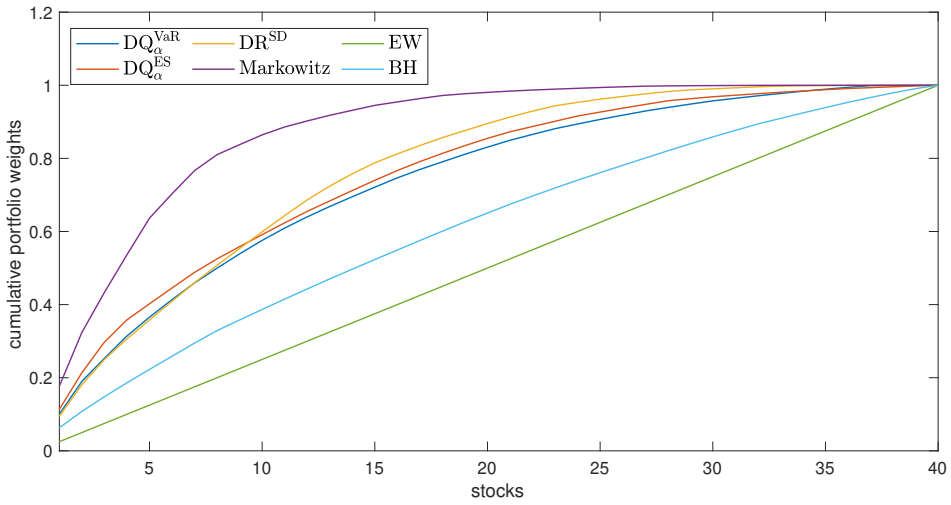


Table 5.1: Annualized return (AR), annualized volatility (AV), Sharpe ratio (SR) and average trading proportion (ATP) for different portfolio strategies from Jan 2004 to Dec 2011

%	DQ_α^{VaR}	DQ_α^{ES}	DR^{SD}	Markowitz	EW	BH
AR	9.46	8.13	9.10	7.98	5.30	6.23
AV	16.65	21.45	20.92	11.98	20.15	15.53
SR	30.48	17.47	22.58	30.06	4.57	11.94
ATP	37.23	28.59	20.24	24.56	5.04	0

Table 5.2: Annualized return (AR), annualized volatility (AV), Sharpe ratio (SR) and average trading proportion (ATP) for different portfolio strategies from Jan 2014 to Dec 2021

%	DQ_α^{VaR}	DQ_α^{ES}	DR^{VaR_α}	DR^{ES_α}	DR^{SD}	Markowitz	EW	BH
AR	13.54	14.79	12.77	13.85	14.37	8.59	12.74	14.22
AV	13.43	15.90	14.41	14.53	14.29	12.74	14.68	13.96
SR	79.69	75.17	68.89	75.79	80.67	45.14	67.40	81.54
ATP	16.07	19.24	64.77	57.56	11.81	15.19	4.45	0

Figure 5.5: Wealth processes for portfolios, 20 stocks, Jan 2014 - Dec 2021

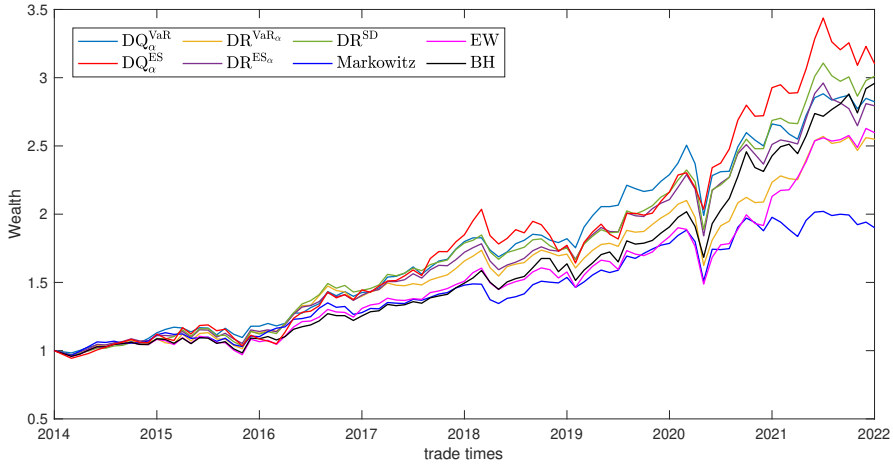
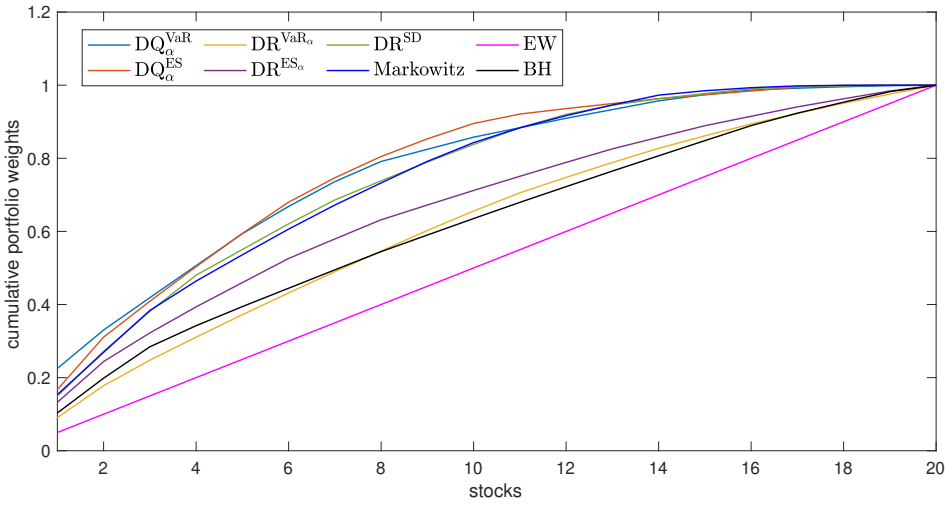


Figure 5.6: Cumulative portfolio weights, 20 stocks, Jan 2014 - Dec 2021



portfolios, and all other parameters are the same as in Section 4.7.3. The results, including two other portfolios built by $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$, are reported in Figures 5.5 and 5.6 and Table 5.2. Since we do not find an efficient algorithm for computing $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$, we use the preceding 500 trading days to compute the optimal portfolio weights using the random sampling method, which is relatively slow and not very stable. (If the previous month has an optimal weight $\mathbf{w}_{t-1}^*$, then 10^5 new weights are sampled from $\lambda\mathbf{w}_{t-1}^*+(1-\lambda)\Delta_n$, where λ is chosen as 0.9. Tie-breaking is done by picking the one that is closest to $\mathbf{w}_{t-1}^*$. We set $\mathbf{w}_0^* = (1/n, \ldots, 1/n)$.) The results show similar observations to those in Section 4.7.3. The

additional observation from Table 5.2 is that DR strategies have much larger ATP than the others, but this may be partially caused by our random sampling algorithms to optimize DR.

Chapter 6

Diversification quotients based on VaR and ES

6.1 Introduction

In order to mitigate risks in portfolios of financial investment quantitatively, a common approach is to compute a quantitative index of the portfolio model, based on e.g., the volatility, variance, an expected utility or a risk measure, following the seminal idea of Markowitz (1952) on portfolio diversification. In the literature, one of the most prominent examples of the diversification index based on a general risk measure is defined by Tasche (2007) which is referred as diversification ratio (DR). Choueifaty and Coignard (2008) investigated the theoretical and empirical properties of DR in portfolio construction and compared the behavior of the resulting portfolio to common, simple strategies. See Embrechts et al. (2015) and Koumou and Dionne (2022) for theories of DR and other diversification indices. Bürgi et al. (2008) defined a closely related notion of DR which is called the diversification gain and explored various methods of modeling dependence and their influence on diversification gain.

Different from the traditional diversification indices such as DR in the above literature, Chapter 4 proposed six axioms – non-negativity, location invariance, scale invariance, rationality, normalization and continuity – which jointly characterize a new diversification index,

called the diversification quotient (DQ), whose definition is based on a class of risk measures decreasing in an index α. All commonly used risk measures belong to a monotonic parametric family, and this includes VaR, ES, expectiles, mean-variance, and entropic risk measures. They argued that DQ has many appealing features both theoretically and practically, while these properties, in particular the six axioms above, are not shared by DR based on VaR, ES, or any other commonly used risk measure. Moreover, portfolio optimization of DQs based on VaR and ES can be computed very efficiently, and thus DQ can be easily applied to real data.

Most properties of DQ are studied by Chapter 4 for a general class of risk measures. In this chapter, we focus on specific risk measures, in particular, the Value-at-Risk (VaR) and the Expected Shortfall (ES). Even though VaR has been criticized because of its lack of subadditivity and ES requires the loss to have a finite mean, VaR and ES are still the two most common classes of risk measures in practice, widely employed in global banking and insurance regulatory frameworks; see Basel III/IV (BCBS (2019)) and Solvency II (EIOPA (2011)). More theoretical properties and discussions of VaR and ES can be found in, e.g., Artzner et al. (1999), Embrechts et al. (2014, 2018), Emmer et al. (2015) and the references therein. We pay particular attention to two popular models in finance and insurance, namely, elliptical and multivariate regular variation (MRV) distributions. Elliptical distributions, including normal and t-distributions as special cases, are the most standard tools for quantitative risk management (McNeil et al. (2015)). They have been studied for DR with convenient properties; see Cui et al. (2022) and the references therein. The MRV model is widely used in Extreme Value Theory for investigating the portfolio diversification; see, e.g., Mainik and Rüschendorf (2010), Mainik and Embrechts (2013) and Bignozzi et al. (2016).

This chapter is an extension of Chapter 4 in which an axiomatic framework of diversification indices is proposed and general properties of DQ are studied. As a new concept of diversification index, studying properties such as explicit formulas and limiting behavior of DQ under specific risk measures and special risk models will help us to better understand and use DQ in risk management applications. In addition, the advantages of DQ and the connection between DQ and DR are clearer under the elliptical and MRV models, revealing many attractive features of choosing DQ instead of DR to quantify diversification risk,

especially for tail heaviness and common shocks.

The chapter is organized as follows. In Section 6.2, the definition of DQ and some preliminaries on risk measures are collected. In Section 6.3, we study general properties for DQs based on VaR and ES. Since DQs based on VaR and ES have natural ranges of $[0, n]$ and $[0, 1]$, respectively, some special dependence structures of the portfolio that correspond to the special values of 0, 1, and n are constructed with clear interpretation for values in between (Theorem 6.1). In Section 6.4, we focus on DQ for large portfolios. By the Law of Large Numbers, we show that DQs based VaR and ES for a portfolio with independent components tend to 0 as the number of assets in the portfolio increases to infinity (Theorem 6.2). The limits for DQs based on VaR and ES for portfolios with exchangeable components do not necessarily tend to 0. We show that the upper bound for the limit decreases in the bivariate correlation coefficient. (Proposition 6.1). In Section 6.5, DQ is applied to elliptical models; explicit formulas and the limiting behavior of DQs based on VaR and ES are available (Proposition 6.2 and Theorem 6.3). Moreover, we present several numerical results for the two most important elliptical distributions used in finance and insurance, namely the multivariate normal distribution and the multivariate t-distribution, and show that DQ can properly capture tail heaviness. As a popular tool for modeling heavy-tailed phenomena, MRV models for DQ are studied in Section 6.6. Furthermore, we generalize the results to the optimal portfolio selection problem in Section 6.7. Under elliptical models, the optimization problem can boil down to a well-studied problem (see e.g., Choueifaty and Coignard (2008)) and a limiting result in MRV models is also derived (Theorem 6.4 and Proposition 6.5). We conclude the chapter in Section 6.8.

6.2 Diversification quotients

Throughout this chapter, $(\Omega, \mathcal{F}, \mathbb{P})$ is an atomless probability. The atomless assumption in our context is very weak and it is widely used in statistics and risk management; see Delbaen (2002) and Section A.3 of Föllmer and Schied (2016) for details of atomless probability spaces. Almost surely equal random variables are treated as identical. A risk measure ϕ is a mapping from $\mathcal{X}$ to $\mathbb{R}$, where $\mathcal{X}$ is a convex cone of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ repre-

senting losses faced by a financial institution or an investor, and $\mathcal{X}$ is assumed to include all constants (i.e., degenerate random variables). For $p \in (0,\infty)$, denote by $L^p = L^p(\Omega,\mathcal{F},\mathbb{P})$ the set of all random variables X with $\mathbb{E}[|X|^p] < \infty$ where $\mathbb{E}$ is the expectation under $\mathbb{P}$. Furthermore, $L^\infty = L^\infty(\Omega,\mathcal{F},\mathbb{P})$ is the space of all essentially bounded random variables, and $L^0 = L^0(\Omega,\mathcal{F},\mathbb{P})$ is the space of all random variables. Write $X \sim F$ if the random variable X has the distribution function F under $\mathbb{P}$, and $X \overset{\mathrm{d}}{=} Y$ if two random variables X and Y have the same distribution. We always write $\mathbf{X} = (X_1,\dots,X_n)$ and $\mathbf{0}$ for the n-vector of zeros. Further, denote by $[n] = \{1,\dots,n\}$, $\mathbb{R}_+ = [0,\infty)$ and $\overline{\mathbb{R}} = [-\infty,\infty]$. Terms such as increasing or decreasing functions are in the non-strict sense. For $X \in \mathcal{X}$, ess-sup(X) and ess-inf(X) are the essential supremum and the essential infimum of X, respectively.

A *diversification index* D is a mapping from $\mathcal{X}^n$ to $\overline{\mathbb{R}}$, which is used to quantify the magnitude of diversification of a risk vector $\mathbf{X} \in \mathcal{X}^n$ representing portfolio losses. Our convention is that a smaller value of $D(\mathbf{X})$ represents a stronger diversification. Measuring diversification is closely related to risk measures. Some standard properties of a risk measure $\phi : \mathcal{X} \to \mathbb{R}$ are collected below.

[M] Monotonicity: $\phi(X) \leqslant \phi(Y)$ for all $X, Y \in \mathcal{X}$ with $X \leqslant Y$.

[CA] Constant additivity: $\phi(X + c) = \phi(X) + c$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

[PH] Positive homogeneity: $\phi(\lambda X) = \lambda\phi(X)$ for all $\lambda \in (0,\infty)$ and $X \in \mathcal{X}$.

[SA] Subadditivity: $\phi(X + Y) \leqslant \phi(X) + \phi(Y)$ for all $X, Y \in \mathcal{X}$.

The two popular classes of risk measures in banking and insurance practice are VaR and ES. The VaR at level $\alpha \in [0,1)$ is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1-\alpha\}, \quad X \in L^0,$$

and the ES (also called CVaR, TVaR or AVaR) at level $\alpha \in (0,1)$ is defined as

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\,\mathrm{d}\beta, \quad X \in L^1,$$

and $\mathrm{ES}_0(X) = \text{ess-sup}(X) = \mathrm{VaR}_0(X)$ which may be ∞. The probability level α above is typically very small, e.g., 0.01 or 0.025 in BCBS (2019); note that we use the "small α"

convention as in Chapter 4. Both VaR and ES satisfy the properties [M], [CA] and [PH], while ES also satisfies the property [SA].

To measure diversification quantitatively, we recall a new index, called diversification quotient (DQ), as follows.

Definition 6.1. Let $\rho = (\rho_\alpha)_{\alpha\in I}$ be a class of risk measures indexed by $\alpha \in I = (0, \bar{\alpha})$ with $\bar{\alpha} \in (0, \infty]$ such that ρ_α is decreasing in α. For $\mathbf{X} \in \mathcal{X}^n$, the diversification quotient based on the class ρ at level $\alpha \in I$ is defined by

$$\mathrm{DQ}^{\rho}_{\alpha}(\mathbf{X}) = \frac{\alpha^*}{\alpha}, \qquad \text{where } \alpha^* = \inf\left\{\beta \in I : \rho_\beta\left(\sum_{i=1}^{n} X_i\right) \leqslant \sum_{i=1}^{n} \rho_\alpha\left(X_i\right)\right\}$$

with the convention $\inf(\varnothing) = \bar{\alpha}$.

Remark 6.1. The value of $\mathrm{DQ}^{\rho}_{\alpha}$ depends on how the class $\rho = (\rho_\alpha)_{\alpha\in I}$ is parametrized. For instance, one could, hypothetically, use a different parametrization $\mathrm{VaR}'_\alpha = \mathrm{VaR}_{\alpha^2}$ for the class VaR, although there is no real reason to do so. The value of $\mathrm{DQ}^{\mathrm{VaR}'}_{\alpha}$ is generally different from $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha^2}$, but they generate the same order; that is, $\mathrm{DQ}^{\mathrm{VaR}'}_{\alpha}(\mathbf{X}) \leqslant \mathrm{DQ}^{\mathrm{VaR}'}_{\alpha}(\mathbf{Y})$ if and only if $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha^2}(\mathbf{X}) \leqslant \mathrm{DQ}^{\mathrm{VaR}}_{\alpha^2}(\mathbf{Y})$, which can be checked by definition. Therefore, different parametrizations do not affect the application of DQ in portfolio optimization.

Theorem 4.1 in Chapter 4 characterized a subclass of DQ via six axioms: non-negativity, location invariance, scale invariance, rationality, normalization and continuity; such DQs are defined on the class of risk measures satisfying [M], [CA] and [PH]. DQ is defined based on a monotonic parametric class of risk measures. All commonly used risk measures belong to a monotonic parametric family; for instance, this includes VaR, ES, expectiles, mean-variance, and entropic risk measures; see Föllmer and Schied (2016) for a general treatment of risk measures.

In finance and insurance, the risk measures VaR and ES play prominent roles, as they are specified in regulatory documents such as BCBS (2019) and EIOPA (2011). We will focus on VaR or ES as the risk measures assessing diversification by DQ in this chapter. In particular, both VaR and ES satisfy the properties [M], [CA] and [PH], and hence $\mathrm{DQ}^{\mathrm{VaR}}_{\alpha}$ and $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}$ satisfy the six above axioms.

Another popular diversification index is the diversification ratio (see e.g., Tasche (2007) and Embrechts et al. (2015)), defined as

$$\mathrm{DR}^{\phi}(\mathbf{X}) = \frac{\phi\left(\sum_{i=1}^{n} X_i\right)}{\sum_{i=1}^{n} \phi(X_i)}, \tag{6.1}$$

where ϕ is a suitably chosen risk measure, such as VaR_α, ES_α, variance (var), or standard deviation (SD). Although DR generally does not satisfy some of the six axioms, we will compare DQ and DR in several parts of the chapter.

6.3 DQ based on VaR and ES

In this section, we will focus on the theoretical properties of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$. For VaR and ES, the interval in Definition 6.1 has a natural range of $I = (0, 1)$. Similarly to Chapter 4, we let $\mathcal{X}^n$ be $(L^0)^n$ when we discuss $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $(L^1)^n$ when we discuss $\mathrm{DQ}_\alpha^{\mathrm{ES}}$. To compute $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, we first define the superquantile transform (Liu et al. (2021, Example 4)). The term "superquantile" is an alternative name for ES; see Rockafellar et al. (2014).

Definition 6.2. The *superquantile transform* of a distribution F with finite mean is a distribution $\widetilde{F}$ with quantile function $p \mapsto \mathrm{ES}_{1-p}(X)$ for $p \in (0, 1)$, where $X \sim F$.

The following alternative formulas for DQs based on VaR and ES will be useful later. They are shown in Theorem 4.4 of Chapter 4 For a given $\alpha \in (0, 1)$, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ can be computed by

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \frac{1 - F\left(\sum_{i=1}^{n} \mathrm{VaR}_\alpha(X_i)\right)}{\alpha} \quad \text{and} \quad \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \frac{1 - \widetilde{F}\left(\sum_{i=1}^{n} \mathrm{ES}_\alpha(X_i)\right)}{\alpha}, \tag{6.2}$$

where F is the distribution of $\sum_{i=1}^{n} X_i$ and $\widetilde{F}$ is the superquantile transform of F.

Remark 6.2. Let $S = \sum_{i=1}^{n} X_i$. If S has a continuous and strictly monotone quantile function, then (6.2) can be rewritten as

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \frac{1}{\alpha}\mathbb{P}\left(S > \sum_{i=1}^{n} \mathrm{VaR}_\alpha(X_i)\right), \quad \mathbf{X} \in \mathcal{X}^n,$$

and

$$\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \frac{1}{\alpha}\mathbb{Q}\left(S > \sum_{i=1}^{n} \mathrm{ES}_\alpha(X_i)\right), \quad \mathbf{X} \in \mathcal{X}^n,$$

for some probability measure $\mathbb{Q}$. To give a formula for $\mathbb{Q}$, let F be the distribution of S, and $\alpha_0 = 1 - F(\mathbb{E}[S])$. There exists an increasing and continuous function $g : (0,1) \to [0,1]$ such that $\mathrm{ES}_{g(\alpha)}(S) = \mathrm{VaR}_\alpha(S)$ for all $\alpha \in (0, \alpha_0)$ and $g(\alpha) = 1$ for $\alpha \in [\alpha_0, 1)$. We can express $\mathbb{Q}$ by $\mathrm{d}\mathbb{Q}/\,\mathrm{d}\mathbb{P} = g'(1 - F(S))$.

Remark 6.3. DQ based on ES admits another convenient formula in Theorem 4.4. If we have $\mathbb{P}(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)) > 0$, then

$$\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \frac{1}{\alpha} \min_{r \in (0,\infty)} \mathbb{E}\left[\left(r \sum_{i=1}^n (X_i - \mathrm{ES}_\alpha(X_i)) + 1 \right)_+ \right], \tag{6.3}$$

and otherwise $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 0$. The main advantage of this formula of $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ is computation and optimization. In particular, this formula allows us to write the portfolio optimization problem of $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ as a convex program; this is shown in Proposition 4.6 of Chapter 4.

Next, we see that if $\alpha \in (0, 1/n)$, there are three special values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, which are 0, 1 and n, corresponding to different representative dependence structures. The last value of n is based on a useful inequality

$$\mathrm{VaR}_{n\alpha}\left(\sum_{i=1}^n X_i \right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i) \tag{6.4}$$

from Corollary 1 of Embrechts et al. (2018), and its sharpness is stated in Corollary 2 therein. For $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, there are two special numbers, 0 and 1, because ES is a class of subadditive risk measures. As a natural question, we wonder for what types of dependence structures these special values are attained. Next, we address this question.

We first present the concept of risk concentration in Wang and Zitikis (2021) which will be useful to understand the dependence structures corresponding to special values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$.

Definition 6.3 (Tail event and α-concentrated)**.** Let X be a random variable and $\alpha \in (0,1)$.

(i) A tail event of X is an event $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$ such that $X(\omega) \geqslant X(\omega')$ holds for a.s. all $\omega \in A$ and $\omega' \in A^c$, where A^c stands for the complement of A.

(ii) A random vector $(X_1, \dots, X_n)$ is α-concentrated if its component share a common tail event of probability α.[1]

Theorem 4 of Wang and Zitikis (2021) gives that a random vector $(X_1, \dots, X_n)$ is α-concentrated for all $\alpha \in (0,1)$ if and only if it is comonotonic, and hence the dependence notion of α-concentration is weaker than comonotonicity. A random vector $(X_1, \dots, X_n)$ is *comonotonic* if there exists a random variable Z and increasing functions $f_1, \dots, f_n$ on $\mathbb{R}$ such that $X_i = f_i(Z)$ a.s. for every $i \in [n]$.

We first address the case that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n$, which involves the dependence concepts of both risk concentration and mutual exclusivity (see Dhaene et al. (1999)). Thus, to arrive at the maximum value of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n$, one requires a dependence structure that is a combination of positive and negative dependence. This phenomenon is common in problems in VaR aggregation; see Puccetti and Wang (2015) for extremal dependence concepts. For this purpose, we propose the *α-concentration-exclusion* (α-CE) model for $\alpha \in (0, 1/n)$, which is a random vector $\mathbf{X} \in \mathcal{X}^n$ satisfying four conditions:

(i) $\mathbb{P}\left(X_i > \mathrm{VaR}_\alpha(X_i)\right) = \alpha$;

(ii) $\mathbb{P}(X_i \geqslant \mathrm{VaR}_\alpha(X_i)) \geqslant n\alpha$;

(iii) $\{X_i > \mathrm{VaR}_\alpha(X_i)\}$, $i \in [n]$, are mutually exclusive;

(iv) $(X_1, \dots, X_n)$ are $(n\alpha)$-concentrated.

For a class ρ of risk measures ρ_α decreasing in α, we say that ρ is *non-flat from the left* at (α, X) if $\rho_\beta(X) > \rho_\alpha(X)$ for all $\beta \in (0, \alpha)$, and ρ is *left continuous* at (α, X) if $\alpha \mapsto \rho_\alpha(X)$ is left continuous.

Remark 6.4. For any given $X \in L^0$, if VaR is non-flat from the left at $(n\alpha, X)$, then there exists α-CE random vector $\mathbf{X} \in \mathcal{X}^n$ such that $\sum_{i=1}^n X_i = X$. For instance, let $A = \{X > \mathrm{VaR}_{n\alpha}(X)\}$. As VaR is non-flat from the left at $(n\alpha, X)$, we have $\mathbb{P}(A) = n\alpha$.

[1] Wang and Zitikis (2021) used the "large α" convention, and hence our α-concentration corresponds to their $(1-\alpha)$-concentration.

Let $(A_1, \dots, A_n)$ be a partition of A with $\mathbb{P}(A_i) = \alpha$ for $i \in [n]$. Also, let $X_i = (X - m)\mathbf{1}_{A_i}$ for $i \in [n-1]$ and $X_n = (X - m)\mathbf{1}_{\{A_n \cup A^c\}} + m$ where $m = \mathrm{VaR}_{n\alpha}(X)$ is a constant. It follows that $\sum_{i=1}^n X_i = X$, and it is clear that $\mathbf{X} = (X_1, \dots, X_n)$ is an α-CE model; such a construction is essentially the one in Embrechts et al. (2018, Theorem 2). More generally, we give a sufficient condition for $\mathbf{X}$ to satisfy the α-CE model. A random vector (X, Y) is said to be counter-monotonic if $(X, -Y)$ is comonotonic. If each pair (X_i, X_j) is counter-monotonic for $i \neq j$, and for each $i \in [n]$, $\mathbb{P}(X_i > \mathrm{VaR}_\alpha(X_i)) = \alpha$ and $\mathrm{VaR}_\alpha(X_i) = \text{ess-inf}(X_i)$, then $\mathbf{X}$ follows an α-CE model. For recent results on pairwise counter-monotonicity, see Chapter 7.

In the next result, we summarize several dependence structures that correspond to special values 0, 1 and n of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and the special values 0 and 1 of $\mathrm{DQ}_\alpha^{\mathrm{ES}}$.

Theorem 6.1. *For $\alpha \in (0,1)$ and $n \geqslant 2$, the following hold:*

(i) $\left\{\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \mid \mathbf{X} \in \mathcal{X}^n\right\} = [0, \min\{n, 1/\alpha\}]$ *and* $\left\{\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) \mid \mathbf{X} \in \mathcal{X}^n\right\} = [0,1]$.

(ii) For ρ being VaR *or* ES, $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 0$ *if and only if* $\sum_{i=1}^n X_i \leqslant \sum_{i=1}^n \rho_\alpha(X_i)$ *a.s. In case* $\sum_{i=1}^n X_i$ *is a constant,* $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 0$ *if* $\alpha < 1/n$ *and* $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 0$.

(iii) For ρ being VaR *or* ES, *if $\mathbf{X}$ is α-concentrated, then* $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) \leqslant 1$. *If, in addition, ρ is continuous and non-flat from the left at* $(\alpha, \sum_{i=1}^n X_i)$, *then* $\mathrm{DQ}_\alpha^\rho(\mathbf{X}) = 1$.

(iv) If $\alpha < 1/n$ *and $\mathbf{X}$ has an α-CE model, then* $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n$ *and* $\mathrm{DQ}_{n\alpha}^{\mathrm{ES}}(\mathbf{X}) = 1$.

Proof. (i) We first prove the case of VaR. By Corollary 1 of Embrechts et al. (2018), we have

$$\mathrm{VaR}_{n\alpha}\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i),$$

which implies $\alpha^* \leqslant n\alpha$, and hence $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \leqslant n$. By definition, $\alpha^* \in [0,1]$, and hence $0 \leqslant \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \leqslant 1/\alpha$. To summarize, $\left\{\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \mid \mathbf{X} \in \mathcal{X}^n\right\} \subseteq [0, \min\{n, 1/\alpha\}]$.

Next, we show that every point in the interval $[0, \min\{n, 1/\alpha\}]$ is attainable by $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$. Take any $\mathbf{X} \in \mathcal{X}^n$ and let $a = \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$. Since $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ satisfies [LI], we can replace each component X_i of $\mathbf{X}$ with $X_i - \mathrm{VaR}_\alpha(X_i)$ for $i \in [n]$. Hence, it is safe to assume that VaR_α of each component of $\mathbf{X}$ is 0. Let $\mathbf{Z} = \mathbf{X}\mathbb{1}_A$ where $A \in \mathcal{F}$ is independent of $\mathbf{X}$ and

$\mathbb{P}(A) = p \in (0,1)$. Since the mapping $F \mapsto \mathrm{VaR}_\alpha(X)$ where $X \sim F$ has convex level sets (e.g., Gneiting (2011)), VaR_α of each component of $\mathbf{Z}$ is 0. By (6.2), we have

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{Z}) = \frac{1}{\alpha}\mathbb{P}\left(\sum_{i=1}^n Z_i > 0\right) = \frac{p}{\alpha}\mathbb{P}\left(\sum_{i=1}^n X_i > 0\right) = p\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}).$$

Since $p \in (0,1)$ is arbitrary, any point in $[0,a]$ belongs to the range of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$. To complete the proof, it suffices to construct $\mathbf{X}$ such that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \min\{n, 1/\alpha\}$.

In case $\alpha \geqslant 1/n$, let $\mathbf{X}$ follow an n-dimensional multinomial distribution with parameters $(1/n, \dots, 1/n)$. It is clear that $\sum_{i=1}^n X_i = 1$. Since $\alpha \geqslant 1/n$, then $\mathrm{VaR}_\alpha(X_i) = 0$. In this case, by (6.2), $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 1/\alpha$. In case $\alpha < 1/n$, we can find $\mathbf{X}$ satisfying $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n$, which is constructed in part (iv) of the proof below.

Next, we prove the case of ES. Since ES satisfies [SA], the range of $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ is contained in $[0,1]$. Take any $t \in [0,2]$, and let each of X_1 and X_2 follow a uniform distribution on $[-1,1]$ such that $X_1 + X_2$ is uniformly distributed on $[-t,t]$. The existence of such (X_1, X_2) is shown by Theorem 3.1 of Wang and Wang (2016). Let $X_i = 0$ for $i = 3, \dots, n$. We can easily compute $\mathrm{ES}_\alpha(X_1) = \mathrm{ES}_\alpha(X_2) = 1 - \alpha$ and $\mathrm{ES}_\beta(X_1 + X_2) = t(1-\beta)$. Hence,

$$\mathrm{DQ}_\alpha^{\mathrm{ES}}(X_1, \dots, X_n) = \frac{1}{\alpha}\inf\{\beta \in (0,1) : t(1-\beta) \leqslant 2 - 2\alpha\} = \frac{1}{\alpha}\left(1 - \frac{2-2\alpha}{t}\right)_+.$$

For letting t vary in $[0,2]$, we get that every point in $[0,1]$ is attained by $\mathrm{DQ}_\alpha^{\mathrm{ES}}$.

(ii) The first part follows directly from Theorem 4.3 (i) of Chapter 4. In particular, if $\sum_{i=1}^n X_i$ is a constant, we have $\mathrm{VaR}_0\left(\sum_{i=1}^n X_i\right) = \mathrm{VaR}_{n\alpha}\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)$ for $\alpha < 1/n$, and $\mathrm{ES}_0\left(\sum_{i=1}^n X_i\right) = \mathrm{ES}_\alpha\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)$. Thus, we have $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 0$ if $\alpha < 1/n$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 0$.

(iii) By Theorem 6 in Wang and Zitikis (2021), if $\mathbf{X}$ is α-concentrated, we have

$$\mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha\left(X_i\right),$$

which implies $\alpha^* \leqslant \alpha$ and then $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \leqslant 1$. Further, as VaR is continuous and non-flat from the left at $(\alpha, \sum_{i=1}^n X_i)$, by Theorem 6 in Wang and Zitikis (2021), the inequality above is an equality. Thus, we have $\alpha^* = \alpha$, which leads to $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 1$. Moreover, from Theorem 5 of Wang and Zitikis (2021), we know that $\mathrm{ES}_\alpha\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{ES}_\alpha\left(X_i\right)$ if

$(X_1, \dots, X_n)$ is α-concentrated. Combining with the fact that $\mathrm{ES}_\alpha(\sum_{i=1}^n X_i)$ is non-flat from left at $(\alpha, \mathbf{X})$, we have $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 1$.

(iv) As $X_1, \dots, X_n$ are $(n\alpha)$-concentrated, there exists an event B such that B is a tail event for all X_i and $\mathbb{P}(B) = n\alpha$. Let $B_i = \{X_i > \mathrm{VaR}_\alpha(X_i)\}$. By Lemma A.3 of Wang and Zitikis (2021), we have $\{X_i > \mathrm{VaR}_{n\alpha}(X_i)\} \subseteq B$. As $\mathrm{VaR}_\alpha(X_i) \geqslant \mathrm{VaR}_{n\alpha}(X_i)$, it gives $B_i \subseteq B$ for all $i \in [n]$. From $\mathbb{P}(X_i \geqslant \mathrm{VaR}_\alpha(X_i)) \geqslant n\alpha$, we know that $X_i(\omega) \geqslant \mathrm{VaR}_\alpha(X_i)$ for all $\omega \in B$. Further, as $B_1, \dots, B_n$ are mutually exclusive, we have $X_i(\omega) > \mathrm{VaR}_\alpha(X_i)$ and $X_j(\omega) = \mathrm{VaR}_\alpha(X_j)$ for all $\omega \in B_i$ and $j \neq i$. Hence, for all $\omega \in \bigcup_{i=1}^n B$, we have $\sum_{i=1}^n X_i(\omega) > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)$ while $\sum_{i=1}^n X_i(\omega) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)$ for $\omega \in (\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n A_i^c$. Therefore, if $\alpha < 1/n$,

$$\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) = \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n \mathbb{P}(B_i) = n\alpha.$$

By (6.2), we have $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n$.

For the case of ES, as $X_1, \dots, X_n$ are $(n\alpha)$-concentrated, by Theorem 5 of Wang and Zitikis (2021), we have $\mathrm{ES}_{n\alpha}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{ES}_{n\alpha}\left(X_i\right)$. Together with the fact that $\beta \mapsto \mathrm{ES}_\beta\left(\sum_{i=1}^n X_i\right)$ is strictly decreasing at $\beta = n\alpha$, we get that $\mathrm{DQ}_{n\alpha}^{\mathrm{ES}}(\mathbf{X}) = 1$. □

Note that comonotonicity is stronger than α-concentration, and hence it is a sufficient condition for (iii) in Theorem 6.1 replacing α-concentration.

In summary, both $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ take values on a bounded interval. In contrast, the diversification ratio $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ is unbounded, and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ is bounded above by 1 only when the ES of the total risk is non-negative. The continuous ranges of DQs also give more information on diversification. Moreover, similarly to the continuity axiom of preferences (e.g., Föllmer and Schied (2016)), a bounded interval can provide mathematical convenience for applications. The values of DQs are simple to interpret. To be specific, for $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, its value is 0 if there is a very good hedge in the sense of Theorem 6.1 (ii); its value is 1 if there is strong positive dependence such as comonotonicity, and its value is n if there is strong negative dependence conditional on the tail event. For $\mathrm{DQ}_\alpha^{\mathrm{ES}}$, its value is 0 if there is a very good hedge in the sense of Theorem 6.1 (ii) and its value is 1 if there is strong positive dependence such as comonotonicity or α-concentration.

6.4 Diversification for large portfolios

In this section, we will focus on the asymptotic behavior of DQ for large portfolios. First, since the independent portfolio is widely recognized as an effectively diversified portfolio, we anticipate that DQ for this type of portfolio would be close to zero as n tends to ∞.

Theorem 6.2. *Let $X_1, X_2, \dots$ be a sequence of uncorrelated random variables in L^2. Assume $\sup_{i\in\mathbb{N}} \operatorname{var}(X_i) < \infty$ and $\inf_{i\in\mathbb{N}}\{\rho_\alpha(X_i) - \mathbb{E}[X_i]\} > 0$. For $\alpha \in (0,1)$ and ρ being* VaR *or* ES*,*

$$\lim_{n\to\infty} \mathrm{DQ}^{\rho}_{\alpha}(X_1, \dots, X_n) = 0. \tag{6.5}$$

Proof. Let $\mathbf{X}_n = (X_1, \dots, X_n)$ and $S_n = \sum_{i=1}^n X_i$. As $\mathrm{DQ}^{\rho}_{\alpha}$ is location invariant, we can assume that $\mathbb{E}[X_i] = 0$ for $i = 1, 2, \dots$. Hence, by the L^2-Law of Large Numbers in the form of Durrett (2019, Theorem 2.2.3), we have $S_n/n \xrightarrow{L^2} 0$. (In fact, L^1 convergence is sufficient to prove our result.)

We first prove the case of VaR. Note that $S_n/n \xrightarrow{L^2} 0$ implies $\lim_{n\to\infty} \mathbb{P}(S_n/n > x) = 0$ for all $x > 0$. Let $\epsilon = \inf_{i\in\mathbb{N}}\{\rho_\alpha(X_i) - \mathbb{E}[X_i]\}$. As $\mathrm{VaR}_\alpha(X_i) > \epsilon$, $i = 1, 2, \dots$, we have

$$\mathbb{P}\left(S_n > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) \leqslant \mathbb{P}\left(S_n/n > \epsilon\right) \to 0.$$

Thus, $\lim_{n\to\infty} \mathbb{P}(S_n > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)) = 0$. By (6.2), we have

$$\lim_{n\to\infty} \mathrm{DQ}^{\mathrm{VaR}}(\mathbf{X}_n) = \lim_{n\to\infty} \frac{1}{\alpha}\mathbb{P}\left(S_n > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right) = 0.$$

Next, we prove the case of ES. As ES is a convex distortion risk measure, ES is L^1-continuous (see Rüschendorf (2013, Corollary 7.10)). Further, since $\mathrm{ES}_\beta(0) = 0$, we have $\mathrm{ES}_\beta(S_n/n) \to 0$ as $n \to \infty$ for all $\beta \in (0,1)$. As a result, for every $\beta \in (0,1)$, there exists N_β such that $\mathrm{ES}_\beta(S_n/n) < \epsilon$ for all $n > N_\beta$. Therefore, we have

$$\alpha^* = \inf\left\{\beta \in (0,1) : \mathrm{ES}_\beta(S_n) \leqslant \sum_{i=1}^n \mathrm{ES}_\alpha(X_i)\right\} \leqslant \inf\left\{\beta \in (0,1) : \mathrm{ES}_\beta(S_n/n) \leqslant \epsilon\right)\} \to 0$$

as $n \to \infty$. Hence, we have $\mathrm{DQ}^{\mathrm{ES}}_{\alpha}(\mathbf{X}_n) = \alpha^*/\alpha \to 0$ as $n \to \infty$. □

Note that Theorem 6.2 does not imply that all independent portfolios are good hedges, because (6.5) holds under some assumptions. In case the components of the portfolio have very heavy tails, DQ based on VaR can be close to n even if the individual losses are iid, as we will see in Theorem 6.3 below.

Remark 6.5. In the special case that $X_1, X_2, \dots$ are iid, Theorem 6.2 implies that, if $\rho_\alpha(X_1) > \mathbb{E}[X_1]$, we have

$$\lim_{n\to\infty} \mathrm{DQ}_\alpha^\rho(X_1, \dots, X_n) = 0$$

for ρ being VaR or ES.

Next, we focus on portfolios with exchangeable components, which may represent a homogeneous subgroup of assets from a large asset pool. An infinite sequence of random variables $X_1, X_2, \dots$ is said to be exchangeable if $(X_1, \dots, X_n) \overset{\mathrm{d}}{=} (X_{\pi(1)}, \dots, X_{\pi(n)})$ for all $n \geqslant 2$ and $\pi \in \mathfrak{S}_n$, where $\mathfrak{S}_n$ is the set of permutations of $[n]$. Exchangeability is closely related to iid sequence of random variables due to de Finetti's theorem, which says that any infinite exchangeable sequence is conditionally iid. However, for the exchangeable portfolio, the value of DQ does not necessarily converge to 0 as n goes to infinity. By the Birkhoff–Khinchin theorem (see Aleksandr and Khinchin (1949)), if $\mathbb{E}[|X_1|] < \infty$, we have $\sum_{i=1}^n X_i/n \to \mathbb{E}[X_1|\mathcal{G}]$ a.s. for some sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. By (6.2), we get

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(X_1, \dots, X_n) \to \frac{1 - F\left(\mathrm{VaR}_\alpha(X_1)\right)}{\alpha} \quad \text{as} \quad n \to \infty,$$

and

$$\mathrm{DQ}_\alpha^{\mathrm{ES}}(X_1, \dots, X_n) \to \frac{1 - \widetilde{F}\left(\mathrm{ES}_\alpha(X_1)\right)}{\alpha} \quad \text{as} \quad n \to \infty,$$

where F is the distribution of $\mathbb{E}[X_1|\mathcal{G}]$ and $\widetilde{F}$ is the superquantile transform of F.

The above formulas depend on $\mathcal{G}$ which may not be explicit. In the next proposition, we derive an upper bound on the limit.

Proposition 6.1. *Let X_1, $X_2, \dots$ be a sequence of exchangeable random variables in L^2. Denote by $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{var}(X_1)$ and $r = \mathrm{corr}(X_1, X_2)$. For $\alpha \in (0, 1)$ and ρ being* VaR *or* ES*, if $\rho_\alpha(X_1) > \mu$, then*

$$\lim_{n\to\infty} \mathrm{DQ}_\alpha^\rho(X_1, \dots, X_n) \leqslant \frac{1}{\alpha} \frac{r\sigma^2}{r\sigma^2 + (\rho_\alpha(X_1) - \mu)^2}. \tag{6.6}$$

Proof. Let $S_n = \sum_{i=1}^n X_i$. As $(X_1, \dots, X_n)$ is exchangeable, we have $\mathbb{E}[S_n] = n\mu$ and $\mathrm{var}(S_n) = (n + n(n-1)r)\sigma^2$. The mean and variance of S_n imply the bound

$$\rho_\beta(S_n) \leqslant n\mu + \sigma\sqrt{n + n(n-1)r}\sqrt{\frac{1-\beta}{\beta}}$$

for all $\beta \in (0,1)$; see Table 1 of Li et al. (2018). As a result, we have

$$\begin{aligned}\mathrm{DQ}_\alpha^\rho(X_1, \dots, X_n) &\leqslant \frac{1}{\alpha}\inf\left\{\beta \in (0,1) : n\mu + \sigma\sqrt{n + n(n-1)r}\sqrt{\frac{1-\beta}{\beta}} \leqslant n\rho_\alpha(X_1)\right\} \\ &= \frac{1}{\alpha}\frac{\frac{1+(n-1)r}{n}\sigma^2}{\frac{1+(n-1)r}{n}\sigma^2 + (\rho_\alpha(X_1) - \mu)^2}.\end{aligned}$$

Sending $n \to \infty$, we get the desired result. □

The upper bound (6.6) on $\lim_{n\to\infty} \mathrm{DQ}_\alpha^\rho(X_1, \dots, X_n)$ in Proposition 6.1 decreases as the correlation r between assets decreases. Intuitively, this means that less positive dependence leads to greater diversification. In particular, if $r \downarrow 0$, then $\lim_{n\to\infty} \mathrm{DQ}_\alpha^\rho(X_1, \dots, X_n) \to 0$. The upper bound (6.6) holds true also without exchangeability, as long as the average of the bivariate correlations of assets converges to r and all assets are identically distributed.

6.5 Elliptical models

The most commonly used classes of multivariate distributions are the elliptical models which include the multivariate normal and t-distributions as special cases. For a general treatment of elliptical models in risk management, see McNeil et al. (2015). In this section, we study DQs based on VaR and ES for elliptical models.

6.5.1 Explicit formulas for DQ

A random vector $\mathbf{X}$ is *elliptically distributed* if its characteristic function can be written as

$$\psi(\mathbf{t}) = \mathbb{E}\left[\exp\left(\mathtt{i}\mathbf{t}^\top \mathbf{X}\right)\right] = \exp\left(\mathtt{i}\mathbf{t}^\top \boldsymbol{\mu}\right)\tau\left(\mathbf{t}^\top \Sigma \mathbf{t}\right),$$

for some $\boldsymbol{\mu} \in \mathbb{R}^n$, positive semi-definite matrix $\Sigma \in \mathbb{R}^{n\times n}$, and $\tau : \mathbb{R}_+ \to \mathbb{R}$ called the characteristic generator. We denote this distribution by $\mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$. We will assume that Σ is not a matrix of zeros. Each marginal distribution of an elliptical distribution is a one-dimensional elliptical distribution with the same characteristic generator. The most common examples of elliptical distributions are normal and t-distributions. An n-dimensional t-distribution $\mathrm{t}(\nu, \boldsymbol{\mu}, \Sigma)$ with $\nu > 0$ has density function f given by (if $|\Sigma| > 0$)

$$f(\mathbf{x}) = \frac{\Gamma\left((\nu+n)/2\right)}{\Gamma(\nu/2)\nu^{n/2}\pi^{n/2}\left|\Sigma\right|^{1/2}} \left(1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^{-(\nu+n)/2},$$

where Γ is the gamma function and $|\Sigma|$ is the determinant of the dispersion matrix Σ.

We remind the reader that for elliptical models, VaR and ES behave very similarly. For instance, VaR_α is subadditive for $\alpha \in (0, 1/2)$ in this setting; see McNeil et al. (2015, Theorem 8.28). Moreover, for $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$ and $\mathbf{a} \in \mathbb{R}^n$, both $\mathrm{VaR}_\alpha(\mathbf{a}^\top \mathbf{X})$ and $\mathrm{ES}_\alpha(\mathbf{a}^\top \mathbf{X})$ have the form $y\sqrt{\mathbf{a}^\top \Sigma \mathbf{a}} + \mathbf{a}^\top \boldsymbol{\mu}$ for some constant y being $y_\alpha^{\mathrm{VaR}} := \mathrm{VaR}_\alpha(Y)$ or $y_\alpha^{\mathrm{ES}} := \mathrm{ES}_\alpha(Y)$ where $Y \sim \mathrm{E}_1(0, 1, \tau)$. As a consequence, the behaviour of DQ based on VaR is similar to that based on ES, except for the case of infinite mean.

For a positive semi-definite matrix Σ, we write $\Sigma = (\sigma_{ij})_{n\times n}$, $\sigma_i^2 = \sigma_{ii}$, and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$, and define the constant

$$k_\Sigma = \frac{\sum_{i=1}^n \left(\mathbf{e}_i^\top \Sigma \mathbf{e}_i\right)^{1/2}}{\left(\mathbf{1}^\top \Sigma \mathbf{1}\right)^{1/2}} = \frac{\sum_{i=1}^n \sigma_i}{\left(\sum_{i,j}^n \sigma_{ij}\right)^{1/2}} \in [1, \infty), \tag{6.7}$$

where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ and $\mathbf{e}_1, \dots, \mathbf{e}_n$ are the column vectors of the $n \times n$ identity matrix I_n. Moreover, $k_\Sigma = 1$ if and only if $\Sigma = \boldsymbol{\sigma}\boldsymbol{\sigma}^\top$, which means that $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$ is comonotonic.

Explicit formulas and the limiting behavior of DQs based on VaR and ES for elliptical models are given by the following few results.

Proposition 6.2. *Suppose that* $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$. *We have, for* $\alpha \in (0, 1)$,

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \frac{1 - F(k_\Sigma \mathrm{VaR}_\alpha(Y))}{\alpha} \quad \textit{and} \quad \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \frac{1 - \widetilde{F}(k_\Sigma \mathrm{ES}_\alpha(Y))}{\alpha},$$

where $Y \sim \mathrm{E}_1(0, 1, \tau)$ *with distribution function* F, *and* $\widetilde{F}$ *is the superquantile transform of* F *in* (6.2). *Moreover,*

(i) $\alpha \mapsto \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ *takes value in* $[0,1]$ *on* $(0,1/2]$ *and it takes value in* $[1,2]$ *on* $(1/2,1)$*;*

(ii) $k_\Sigma \mapsto \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ *is decreasing for* $\alpha \in (0,1/2]$ *and increasing for* $\alpha \in (1/2,1)$*;*

(iii) $k_\Sigma \mapsto \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})$ *is decreasing for* $\alpha \in (0,1)$*.*

Proof. We first consider the case of VaR. Since $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$, the linear structure of ellipitical distributions gives $\sum_{i=1}^n X_i \sim \mathrm{E}_1(\mathbf{I}^\top \boldsymbol{\mu}, \mathbf{I}^\top \Sigma \mathbf{I}, \tau)$. That is, $\sum_{i=1}^n X_i \overset{\mathrm{d}}{=} \sum_{i=1}^n \mu_i + \|\mathbf{1}^\top A\|_2 Y$, where A is the Cholesky decomposition of Σ. Also, we have $\mathrm{VaR}_\alpha(X_i) = \mu_i + \|\mathbf{e}_i^\top A\|_2 \mathrm{VaR}_\alpha(Y)$. By (6.2),

$$\begin{aligned}\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) &= \frac{1}{\alpha}\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i}^n \mu_i + \|\mathbf{e}_i^\top A\|_2 \mathrm{VaR}_\alpha(Y)\right) \\ &= \frac{1}{\alpha}\mathbb{P}\left(\sum_{i=1}^n \mu_i + \|\mathbf{1}^\top A\|_2 Y > \sum_{i}^n \mu_i + \|\mathbf{e}_i^\top A\|_2 \mathrm{VaR}_\alpha(Y)\right) = \frac{1 - F(k_\Sigma \mathrm{VaR}_\alpha(Y))}{\alpha}.\end{aligned}$$

By replacing VaR with ES and $\sum_{i=1}^n X_i$ with $\mathrm{ES}_U(\sum_{i=1}^n X_i)$, we can get the first formula of $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})$.

(i) For $\alpha \in (0,1/2]$, we have $\mathrm{VaR}_\alpha(Y) \leqslant k_\Sigma \mathrm{VaR}_\alpha(Y)$ and $1-\alpha \leqslant F(k_\Sigma \mathrm{VaR}_\alpha(Y)) \leqslant 1$. Hence, $0 \leqslant \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \leqslant 1$.

For $\alpha \in (1/2,1)$, $\mathrm{VaR}_\alpha(Y) \geqslant k_\Sigma \mathrm{VaR}_\alpha(Y)$ and $\alpha \leqslant 1 - F(k_\Sigma \mathrm{VaR}_\alpha(Y)) \leqslant 1$. Hence, $1 \leqslant \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \leqslant 1/\alpha \leqslant 2$.

(ii) If $\alpha \in (0,1/2]$, then $\mathrm{VaR}_\alpha(Y) \geqslant 0$, and thus $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ decreases in k_Σ. If $\alpha \in (1/2,1)$, then $\mathrm{VaR}_\alpha(Y) \leqslant 0$, and thus $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ increases in k_Σ.

(iii) For $\alpha \in (0,1)$, $\mathrm{ES}_\alpha(Y) \geqslant 0$. Hence, $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})$ increases in k_Σ. □

In the discussions below, we will assume $\alpha \in (0,1/2)$, which is the most common setting in risk management. In Proposition 6.2, we see that, for $\alpha \in (0,1/2)$, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \in [0,1]$. This is in contrast to Theorem 6.1, where the range of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ is $[0,n]$ instead of $[0,1]$, when we do not restrict to elliptical models. This phenomenon should not be surprising, because, as we mentioned before, VaR_α for $\alpha \in (0,1/2)$ is similar to ES_α for elliptical models, and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ has range $[0,1]$.

In case $Y \sim \mathrm{E}_1(0,1,\tau)$ has a positive density on $\mathbb{R}$, we can see from Proposition 6.2 that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 1$ if and only if $k_\Sigma = 1$ (i.e., $\mathbf{X}$ is comonotonic) or $\mathrm{VaR}_\alpha(Y) = 0$ (i.e., $\alpha = 1/2$). Similarly, $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = 1$ if and only if $k_\Sigma = 1$.

In case the elliptical distribution is asymptotically uncorrelated, we will see that, as $n \to \infty$, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \to 0$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) \to 0$. This is consistent with our intuition that, if the individual risks are asymptotically uncorrelated, then full diversification can be achieved asymptotically, thus the diversification index goes to 0. The value $\mathrm{AC}_\Sigma = \sum_{i,j}^n \sigma_{ij}/(\sum_{i=1}^n \sigma_i)^2 = 1/k_\Sigma^2$ will be called the average correlation (AC) of Σ.

Proposition 6.3. *Suppose that* $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$.

(i) Let $Y \sim \mathrm{E}_1(0,1,\tau)$ *and* f *be the density function of* Y. *We have*

$$\lim_{\alpha \downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \lim_{x \to \infty} k_\Sigma \frac{f(k_\Sigma x)}{f(x)} \quad \text{if } \mathrm{VaR}_0(Y) = \infty \text{ and the limit exists}, \tag{6.8}$$

and $\lim_{\alpha \downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 0$ *if* $\mathrm{VaR}_0(Y) < \infty$.

(ii) If $\lim_{n\to\infty} \mathrm{AC}_\Sigma = 0$, *then*

$$\lim_{n\to\infty} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \lim_{n\to\infty} \mathrm{DQ}_\beta^{\mathrm{ES}}(\mathbf{X}) = 0$$

for $\alpha \in (0, 1/2)$ *and* $\beta \in (0,1)$.

Proof. (i) If $\mathrm{VaR}_0(Y) < \infty$, then $\mathrm{VaR}_0(Y) \leqslant k_\Sigma \mathrm{VaR}_0(Y)$ as $k_\Sigma \geqslant 1$. Hence, $\mathrm{DQ}_0^{\mathrm{VaR}}(\mathbf{X}) = 0$. If $\mathrm{VaR}_0(Y) = \infty$, then $\mathrm{VaR}_0(Y) > k_\Sigma \mathrm{VaR}_\alpha(Y)$ for $\alpha > 0$. Therefore,

$$\lim_{\alpha\to 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \lim_{\alpha \to 0} \frac{1 - F\left(k_\Sigma \mathrm{VaR}_\alpha(Y)\right))}{\alpha} = \lim_{\alpha \to 0} k_\Sigma \frac{f\left(k_\Sigma \mathrm{VaR}_\alpha(Y)\right))}{f(\mathrm{VaR}_\alpha(Y)))} = \lim_{x\to\infty} k_\Sigma \frac{f\left(k_\Sigma x\right)}{f(x)},$$

and we get the desired result.

(ii) We only show the proof of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ as the result for $\mathrm{DQ}_\beta^{\mathrm{ES}}$ can be obtained along the same analogy. By Proposition 6.2, it is clear that $\mathrm{AC}_\Sigma \to \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ is increasing for $\alpha \in (0, 1/2)$ and $\mathrm{AC}_\Sigma \to \mathrm{DQ}_\beta^{\mathrm{ES}}(\mathbf{X})$ is increasing for $\alpha \in (0,1)$. Moreover, if AC_Σ goes to 0 as $n \to \infty$, we have $\lim_{n\to\infty} k_\Sigma = \infty$. Thus, we have $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \to 0$ as $n \to \infty$ by Proposition 6.2. □

Explicit formulas of (6.8) for normal and t-distributions are provided in Section 6.5.2.

Remark 6.6. In general, we do not have a limiting result for $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ in the form of Proposition 6.3 (i). If $\mathbf{X} \sim \mathrm{t}(\nu, \boldsymbol{\mu}, \Sigma)$ for $\nu > 1$, then $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ has the same limit as $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ in (6.8) as $\alpha \downarrow 0$ because $\mathrm{VaR}_\alpha(Y)/\mathrm{ES}_\alpha(Y)$ has a constant limit $(\nu-1)/\nu$ for a t-distributed Y by the Karamata theorem; see Theorem A.7 of McNeil et al. (2015).

From the results above, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})$ depend on both τ and α. In sharp contrast, DR of a centered elliptical distribution is always $1/k_\Sigma$, which ignores the shape of the distribution. More precisely, for $\mathbf{X} \sim \mathrm{E}_n(\mathbf{0}, \Sigma, \tau)$ and $\alpha \in (0, 1/2)$, we have

$$\mathrm{DR}^{\mathrm{VaR}_\alpha}(\mathbf{X}) = \frac{\mathrm{VaR}_\alpha(\sum_{i=1}^n X_i)}{\sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)} = \frac{\left(\sum_{i,j}^n \sigma_{ij}\right)^{1/2} \mathrm{VaR}_\alpha(Y)}{\sum_{i=1}^n \sigma_i \mathrm{VaR}_\alpha(Y)} = \frac{1}{k_\Sigma}, \tag{6.9}$$

and similarly, $\mathrm{DR}^{\mathrm{ES}_\alpha}(\mathbf{X}) = 1/k_\Sigma$. Note that in this case, $\mathrm{DR}^{\mathrm{VaR}_\alpha}$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}$ do not depend on τ, α or whether the risk measure is VaR or ES. Indeed, DR based on var or SD also has the same value $1/k_\Sigma$.

For $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$ with $\boldsymbol{\mu} \neq \mathbf{0}$, $\mathrm{DR}^{\mathrm{VaR}_\alpha}(\mathbf{X})$ and $\mathrm{DR}^{\mathrm{ES}_\alpha}(\mathbf{X})$ depend also on $\boldsymbol{\mu}$, which is arguably undesirable as it conflicts location invariance. Nevertheless, $\lim_{\alpha \downarrow 0} \mathrm{DR}^{\mathrm{VaR}_\alpha}(\mathbf{X}) = 1/k_\Sigma$ if $\mathrm{VaR}_0(Y) = \infty$ (i.e., the value taken by Y is unbounded from above), and this limit does not depend on $\boldsymbol{\mu}$. On the other hand, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$ has a limit in (6.8) which depends on both k_Σ and τ. The above observations suggest that DQ is more comprehensive than DR by utilizing the information on the shape of the distribution.

A similar result to Proposition 6.3 (ii) holds for DR of centered elliptical distributions. More precisely, If $\alpha \in (0, 1/2)$, $\boldsymbol{\mu} = \mathbf{0}$, and $\lim_{n\to\infty} \mathrm{AC}_\Sigma = 0$, then we have $\lim_{n\to\infty} \mathrm{DR}^{\mathrm{VaR}_\alpha}(\mathbf{X}) = 0$ by (6.9), and similarly, $\lim_{n\to\infty} \mathrm{DR}^{\mathrm{ES}_\alpha}(\mathbf{X}) = 0$. These limits do not hold if $\boldsymbol{\mu} \neq \mathbf{0}$.

6.5.2 Normal and t-distributions

Next, we take a close look at the two most important elliptical distributions used in finance and insurance, namely the multivariate normal distribution and the multivariate t-distribution. The explicit formulas for DQ for these distributions are available through the explicit formulas of VaR and ES; see Examples 2.14 and 2.15 of McNeil et al. (2015).

Chapter 4 proposed three simple models where the components of portfolio vectors follow the iid normal model, iid t-model and the common shock t-model, respectively, and showed that the diversification is the strongest according to DQ for the iid normal model and the iid t-model has a smaller DQ than the common shock t-model. In contrast, DR reports a similar value for all three models; see their Section 5.2 for details. Therefore, DQ has the nice feature that it can capture heavy tails and common shocks.

We present some formulas and numerical results for correlated normal and t-models. We focus our discussions mainly on $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ as the case of $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ is similar. We first compute the limit of DQ as $\alpha \downarrow 0$ according to (6.8). By direct calculation,

$$\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \mathbb{1}_{\{k_\Sigma=1\}} \quad \text{if } \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma); \tag{6.10}$$

$$\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = k_\Sigma^{-\nu} \quad \text{if } \mathbf{X} \sim \mathrm{t}(\nu, \boldsymbol{\mu}, \Sigma). \tag{6.11}$$

The above two values properly reflect the fact that the normal distribution is tail independent unless $k_\Sigma = 1$ (i.e., comonotonic), whereas the t-distribution is tail dependent; see Examples 7.38 and 7.39 of McNeil et al. (2015). DQ is able to capture this phenomenon well, by providing, for α close to 0, $\mathrm{DQ}_\alpha^{\mathrm{VaR}} \approx 0$ (strong diversification) for normal distribution and $\mathrm{DQ}_\alpha^{\mathrm{VaR}} \approx k_\Sigma^{-\nu}$ (moderate diversification for common choices of Σ and ν; see Figure 6.3) for a t-distribution. On the other hand, DR of centered normal and t-distributions is always $1/k_\Sigma$, which fails to distinguish the tail of the t-distribution from that of the normal distribution (see (6.9)).

For numerical illustrations, we consider two specific dispersion matrices, parameterized by $r \in [0, 1]$ and $n \in \mathbb{N}$,

$$\Sigma_1 = (\sigma_{ij})_{n\times n}, \quad \text{where } \sigma_{ii} = 1 \text{ and } \sigma_{ij} = r \text{ for } i \neq j, \text{ and}$$

$$\Sigma_2 = (\sigma_{ij})_{n\times n}, \quad \text{where } \sigma_{ii} = 1 \text{ and } \sigma_{ij} = r^{|j-i|} \text{ for } i \neq j.$$

Note that Σ_1 represents an equicorrelated model and Σ_2 represents an autoregressive model AR(1). For $r = 0$, $r = 1$ or $n = 2$, these two dispersion matrices are identical. We take four models $\mathbf{X}_i \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma_i)$ and $\mathbf{Y}_i \sim \mathrm{t}(\nu, \boldsymbol{\mu}, \Sigma_i)$, $i = 1, 2$, and we will let r, ν, α, n vary. Note that the location $\boldsymbol{\mu}$ does not matter in computing DQ, and we can simply take $\boldsymbol{\mu} = \mathbf{0}$. The default parameters are set as $r = 0.3$, $n = 4$, $\nu = 3$ and $\alpha = 0.05$ if not explained otherwise.

Figure 6.1: DQ and DR based on VaR for $\nu \in (0, 10]$ and ES for $\nu \in (1, 10]$ with fixed $\alpha = 0.05$, $r = 0.3$ and $n = 4$

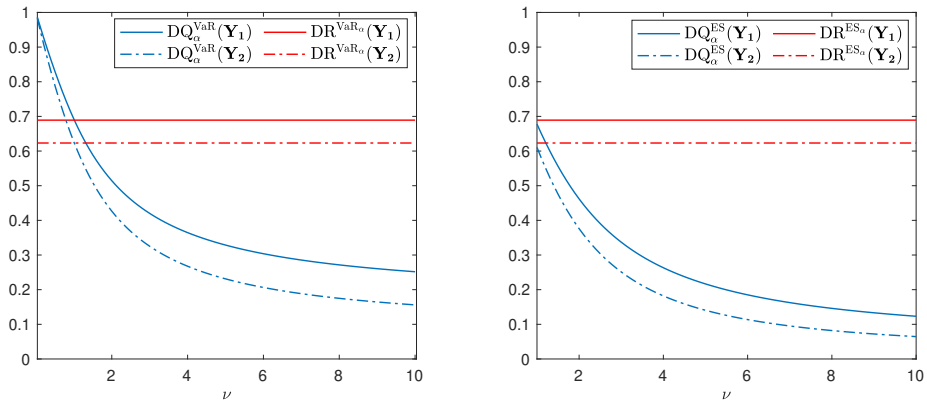
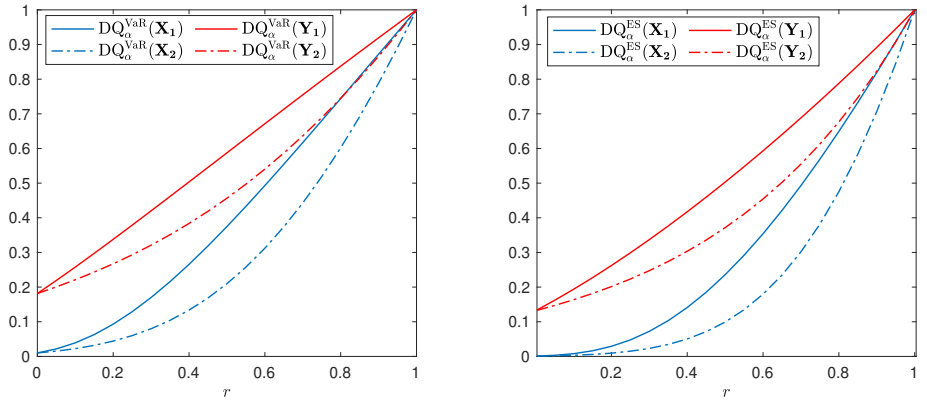


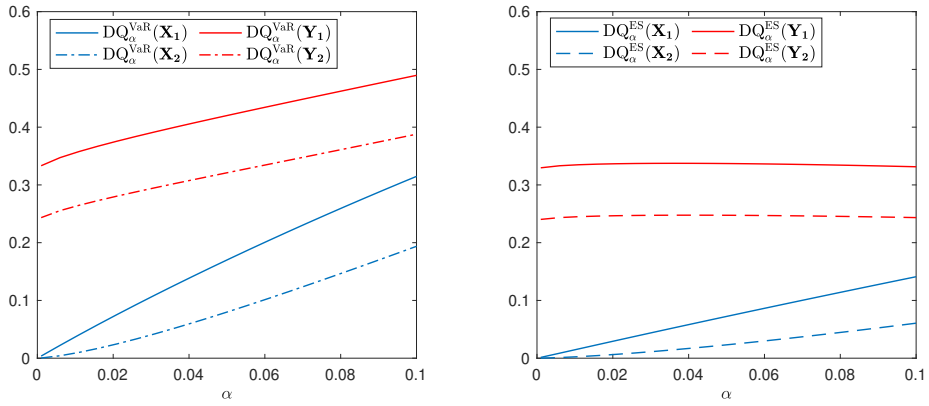
Figure 6.2: DQ based on VaR and ES for $r \in [0, 1]$ with fixed $\alpha = 0.05$, $\nu = 3$, and $n = 4$



DQ for the t-models as the parameter of degrees of freedom ν varies

Figure 6.1 presents the values of DQ for the t-models with varying ν, where $\nu \in (0, 10]$ for VaR and $\nu \in (1, 10]$ for ES. We observe a monotonic relation that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ are decreasing in ν. In particular, if ν is close to 0, we see that $\mathrm{DQ}_\alpha^{\mathrm{VaR}} \approx 1$ which means there is almost no diversification effect for such super heavy-tailed models. On the other hand, DR completely ignores ν and always reports the same value. Note that the values of DQ and DR are not directly comparable as they are not on the same scale.

Figure 6.3: DQ based on VaR and ES for $\alpha \in (0, 0.1)$ with fixed $\nu = 3$, $r = 0.3$ and $n = 4$



DQ for elliptical models as the correlation parameter r varies

In Figure 6.2, we report how DQ changes over $r \in [0, 1]$ in the four models. Intuitively, for r close to 1 which corresponds to comonotonicity, DQ is close to 1 in all models since there is no or very weak diversification in this case. More interestingly, for r close to 0, there is very strong diversification for the normal models, meaning $\mathrm{DQ}_\alpha^{\mathrm{VaR}} \approx 0$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}} \approx 0$, whereas for the t-models, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ are clearly away from 0. Note that the components of a t-distribution are tail dependent even for zero or negative correlation (see Example 7.39 of McNeil et al. (2015)). Hence, DQ is able to capture dependence created by the common factor in the t-model, in addition to its correlation structure.

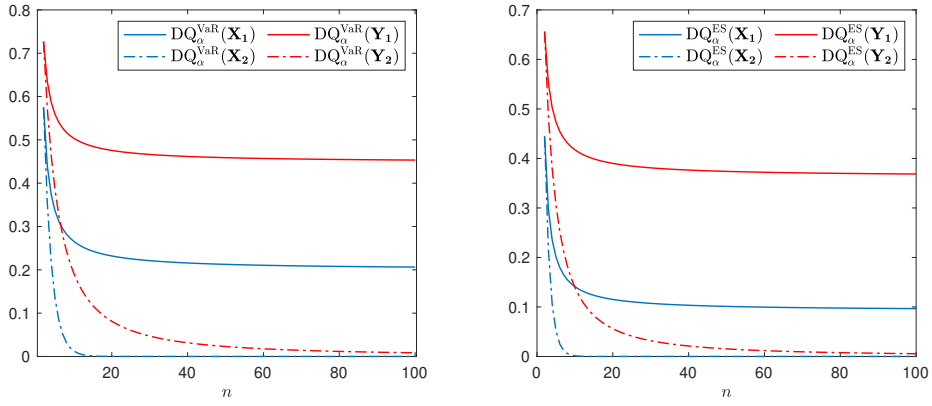
DQ for varying α and its limit

In Figure 6.3, we report $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ for $\alpha \in (0, 1)$ in the four models with correlation matrices specified in Section 6.5.2. We can see from Figure 6.3 that DQ can be non-monotonic with respect to α (see the curves of $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ for $\mathbf{X}_i \sim \mathrm{t}(\nu, \boldsymbol{\mu}, \Sigma_i)$). In addition, we can compute $k_{\Sigma_1} = 1.4510$ and $k_{\Sigma_2} = 1.6046$. Hence, it can be anticipated from Proposition 6.2 that, since DQ is decreasing in k_Σ, models with Σ_1 has larger DQ than the corresponding models with Σ_2. Moreover, as $\alpha \downarrow 0$, we can see that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ converges to its corresponding limits in (6.10) and (6.11); also note that $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ has the same limits as $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ for t-distributions as discussed in Remark 6.6.

DQ for elliptical models as the dimension n varies

Figure 6.4 is related to Section 6.5.2 and reports how DQ changes over $n \in [2, 100]$ in the four models. We choose $r = 0.5$ in this experiment for better visibility. As we can see, DQ decreases to 0 for models with the AR(1) dispersion Σ_2, and DQ converges to a non-zero constant for models with the equicorrelated dispersion Σ_1. This is consistent with Proposition 6.3 (ii) because $\mathrm{AC}_{\Sigma_1} \to r$ and $\mathrm{AC}_{\Sigma_2} \to 0$ as $n \to \infty$.

Figure 6.4: DQs based on VaR and ES for $n \in [2, 100]$ with fixed $\alpha = 0.05$, $r = 0.5$ and $\nu = 3$



Cross-comparison between DQ based on VaR and ES

One may be tempted to compare values of DQ based on VaR to those based on ES. Although we see from Figure 6.3 that the curve $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ often dominates the curve $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ for the same model, such a comparison is not meaningful, since VaR and ES are not meant to be compared at the same level α. For a fair comparison, one needs to associate a VaR level α to an ES level $c\alpha$ where $c \geqslant 1$ is PELVE of Li and Wang (2022) defined via $\mathrm{ES}_{c\alpha}(X) = \mathrm{VaR}_\alpha(X)$ for X being normally or t-distributed; note that the location and scale of X do not matter. The values of c, $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_{c\alpha}^{\mathrm{ES}}$ for $\alpha = 0.01$ are summarized in Table 6.1. As we observe from Table 6.1, the values of DQs based on VaR and ES are quite close when the probability level is calibrated via PELVE. This is consistent with the afore-mentioned fact that VaR behaves similarly to ES in the setting of elliptical models.

Table 6.1: Values of DQs based on VaR at level $\alpha = 0.01$ and ES at level $c\alpha$, where $n = 4$ and $r = 0.3$

	c	$c\alpha$	$\mathrm{DQ}_{\alpha}^{\mathrm{VaR}}$	$\mathrm{DQ}_{c\alpha}^{\mathrm{ES}}$
$\mathbf{X}_1 \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma_1)$	2.58	0.0258	0.0369	0.0377
$\mathbf{X}_2 \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma_2)$	2.58	0.0258	0.0024	0.0025
$\mathbf{Y}_1 \sim \mathrm{t}(3, \boldsymbol{\mu}, \Sigma_1)$	3.31	0.0331	0.3558	0.3373
$\mathbf{Y}_2 \sim \mathrm{t}(3, \boldsymbol{\mu}, \Sigma_2)$	3.31	0.0331	0.2094	0.1961

6.6 Multivariate regularly varying models

Heavy-tailed distributions are known to exhibit complicated and even controversial phenomena in finance (see e.g., Ibragimov et al. (2011)), and they are typically modelled via multivariate regularly varying (MRV) models, important objects in Extreme Value Theory. Such models are particularly relevant for tail risk measures such as VaR and ES at high levels (McNeil et al. (2015)). In particular, MRV models have been applied to DR based on VaR (e.g., Mainik and Rüschendorf (2010) and Mainik and Embrechts (2013)). Since $\mathrm{VaR}_\alpha(X)/\mathrm{ES}_\alpha(X) \to (\gamma - 1)/\gamma$ as $\alpha \downarrow 0$ for $X \in \mathrm{RV}_\gamma$ with finite mean (see e.g., McNeil et al. (2015, p.154)), we only present the case of VaR.

Definition 6.4. A random vector $\mathbf{X} \in \mathcal{X}^n$ has an MRV model with some $\gamma > 0$ if there exists a Borel probability measure Ψ on the unit sphere $\mathbb{S}^n := \{\mathbf{s} \in \mathbb{R}^n : \|\mathbf{s}\| = 1\}$ such that for any $t > 0$ and any Borel set $S \subseteq \mathbb{S}^n$ with $\Psi(\partial S) = 0$,

$$\lim_{x\to\infty} \frac{\mathbb{P}(\|\mathbf{X}\| > tx,\ \mathbf{X}/\|\mathbf{X}\| \in S)}{\mathbb{P}(\|\mathbf{X}\| > x)} = t^{-\gamma}\Psi(S),$$

where $\|\cdot\|$ is the L_1-norm (one could use any other norm equivalent to the L_1-norm). We call γ the tail index of $\mathbf{X}$ and Ψ the spectral measure of $\mathbf{X}$. This is written as $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$.

The univariate regular variation with tail index γ is defined as

$$\text{for all } t > 0,\ \lim_{x\to\infty} \frac{1 - F_X(tx)}{1 - F_X(x)} = t^{-\gamma},$$

where F is the distribution function of X. We write $X \in \mathrm{RV}_\gamma$ for this property. As a consequence of $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$, $\|\mathbf{X}\|$ satisfies univariate regular variation with the same tail index γ.

Regular variation is one of the basic notions for describing heavy-tailed distributions and dependence in the tails. In what follows, we limit our discussion to $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ under the non-degeneracy condition:

$$\Psi\left(\{\mathbf{s} \in \mathbb{S}^n : \mathbf{s} \in (0,\infty)^n\}\right) > 0.$$

Note that if $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ satisfies non-degeneracy condition, we have $\mathbf{w}^\top \mathbf{X} \in \mathrm{RV}_\gamma$ (See Mainik and Embrechts (2013)).

Let $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ be a random vector with identical marginals. If $X_1, \dots, X_n$ have a finite mean, then VaR is asymptotically subadditive in the following sense (see e.g., Embrechts et al. (2009))

$$\mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i) \quad \text{for } \alpha \text{ close enough to } 0,$$

but the inequality is reversed if $X_1, \dots, X_n$ do not have a finite mean. Next, in contrast to Proposition 6.2 and Remark 6.5, we will show that DQ based on VaR can be arbitrarily close to n even if the individual losses are iid.

Theorem 6.3. *Suppose that $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ and $\mathbf{X}$ has positive joint density on the support of $\mathbf{X}$. Then,*

$$\lim_{\alpha \downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \eta_{\mathbf{1}} \left(\sum_{i=1}^n \eta_{\mathbf{e}_i}^{1/\gamma}\right)^{-\gamma}, \tag{6.12}$$

where $\eta_{\mathbf{x}} = \int_{\mathbb{S}^n} \left(\mathbf{x}^\top \mathbf{s}\right)_+^\gamma \Psi(\mathrm{d}\mathbf{s})$ for $\mathbf{x} \in \mathbb{R}^n$. Moreover, if $X_1, \dots, X_n$ are iid random variables, then $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \to n^{1-\gamma}$ as $\alpha \downarrow 0$.

Proof. A more general result of (6.12) and its proof are shown in Proposition 6.5, where the asymptotic behavior of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ for weighted portfolios is investigated. Since DQ is scale-invariant, by taking $\mathbf{w} = (1/n, \dots, 1/n)$ in Proposition 6.5, it gives

$$\lim_{\alpha \downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \lim_{\alpha \downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(1/nX_1, \dots, 1/nX_n) = \frac{\eta_{\mathbf{w}}}{\left(\sum_{i=1}^n w_i \eta_{\mathbf{e}_i}^{1/\gamma}\right)^\gamma},$$

where $\eta_{\mathbf{w}} = n^{-\gamma} \int_{\mathbb{S}^n} \left(\mathbf{1}^\top s\right)_+^\gamma \Psi(\,\mathrm{d}\mathbf{s}) = n^{-\gamma}\eta_{\mathbf{1}}$. As a result, we have

$$\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \eta_{\mathbf{1}} \left(\sum_{i=1}^n \eta_{\mathbf{e}_i}^{1/\gamma} \right)^{-\gamma}.$$

If $X_1, \dots, X_n$ are iid non-negative random variables, by Example 3.1 of Embrechts et al. (2009), we have

$$\eta_{\mathbf{1}}^{1/\gamma} = \lim_{\alpha\downarrow 0} \frac{\mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right)}{\mathrm{VaR}_\alpha(X_1)} = n^{1/\gamma},$$

which implies that $\eta_{\mathbf{1}} = n$. Moreover,

$$(\eta_{\mathbf{e}_i})^{1/\gamma} = \lim_{\alpha\downarrow 0} \frac{\mathrm{VaR}_\alpha(X_i)}{\mathrm{VaR}_\alpha(X_1)} = 1.$$

Hence, $\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n^{1-\gamma}$. Further, if $\gamma \downarrow 0$, then $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \to n$. □

The α-CE model in Theorem 6.1 with $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = n$ is complicated and involves both positive and negative dependence. Theorem 6.3 suggests that $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \approx n$ can be obtained for some very heavy-tailed iid model with γ close to 0. Therefore, the upper bound n on $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ is relevant when analyzing very heavy-tailed risks such as catastrophe losses; we refer to Embrechts et al. (1997) for a general treatment of heavy-tailed risks in insurance and finance.

Remark 6.7. Suppose that $X_1, \dots, X_n$ are iid random variables with $X_1 \in \mathrm{RV}_\gamma$ having positive density over its support. We have $\mathbf{X} = (X_1, \dots, X_n) \in \mathrm{MRV}_\gamma(\Psi)$ by Kulik and Soulier (2020, Example 2.1.4), and thus $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) \to n^{1-\gamma}$ as $\alpha \downarrow 0$.

Remark 6.8. We note that the intersection between elliptical distributions and MRV distributions is non-empty. For $\mathbf{X} \sim \mathrm{E}_n(\mu, \Sigma, \tau)$, we have

$$\mathbf{X} \overset{\mathrm{d}}{=} \mu + RAU,$$

where $A \in \mathbb{R}^{n\times n}$ satisfying $AA^\top = \Sigma$, U is uniformly distributed on the Euclidean sphere $\mathbb{S}_2^d$ and R is a non-negative random variable that is independent of U. Theorem 4.3 of Hult and Lindskog (2002) showed that $\mathbf{X}$ has an MRV model if and only if $R \in \mathrm{RV}_\gamma$ for some $\gamma > 0$. Assume that the elliptically distributed $\mathbf{X}$ is in $\mathrm{MRV}_\gamma(\Psi)$ with $\gamma > 0$. As a result, we have

$Y \sim E_1(0,1,\tau) \in \mathrm{RV}_\gamma$. Let f be the density of Y. Following Proposition 6.3 (i) and the fact that $\mathrm{VaR}_\alpha(Y)/\mathrm{ES}_\alpha(Y) \to (\gamma-1)/\gamma$ as $\alpha \downarrow 0$ for $Y \in \mathrm{RV}_\gamma$ with finite mean, we have

$$\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = \lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \lim_{x\to\infty} k_\Sigma \frac{f(k_\Sigma x)}{f(x)} = k_\Sigma^{-\gamma}.$$

If $\mathbf{X}$ follows an elliptical distribution in the MRV class, then $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})$ has the same limit as $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})$. For example, if $\mathbf{X} \sim t(\nu,\mu,\Sigma)$, we have $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ with $\gamma = \nu$ as we have shown in (6.11) that $\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = \lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X}) = k_\Sigma^{-\nu}$.

To end this section, we show that if there exists an asset with a strictly heavier tail than the other assets in the portfolio, then DQ based on VaR tends to 1 as $\alpha \downarrow 0$.

Proposition 6.4. *Suppose $X_i \in \mathrm{RV}_{\gamma_i}$ for $i \in [n]$ such that $\gamma_1 < \min_{i=2,\dots,n} \gamma_i$. If $X_1,\dots,X_n$ have positive densities on their support, then $\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) = 1$.*

Proof. Since $\gamma_1 < \min_{i=2,\dots,n} \gamma_i$, X_1 has a heavier tail than $X_2,\dots,X_n$. As a result, we have $\sum_{i=1}^n X_i \in \mathrm{RV}_{\gamma_1}$ regardless of the dependence between all random variables (See Kulik and Soulier (2020, Lemma 1.3.2)), that is,

$$\lim_{x\to\infty} \frac{\mathbb{P}\left(\sum_{i=1}^n X_i > x\right)}{\mathbb{P}(X_1 > x)} = 1.$$

Moreover, X_1 having a heavier tail than $X_2,\dots,X_n$ also implies that $\lim_{\alpha\downarrow 0} \mathrm{VaR}_\alpha(X_i)/\mathrm{VaR}_\alpha(X_1) = 0$ for all $i = 2,\dots,n$, and thus $\lim_{\alpha\downarrow 0} \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)/\mathrm{VaR}_\alpha(X_1) = 1$. Therefore, we have

$$\begin{aligned}
\lim_{\alpha\downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X}) &= \lim_{\alpha\downarrow 0} \frac{\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right)}{\alpha} \\
&= \lim_{\alpha\downarrow 0} \frac{\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right)}{\mathbb{P}(X_1 > \mathrm{VaR}_\alpha(X_1))} \\
&= \lim_{\alpha\downarrow 0} \frac{\mathbb{P}\left(\sum_{i=1}^n X_i > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)\right)}{\mathbb{P}(X_1 > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i))} \frac{\mathbb{P}(X_1 > \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i))}{\mathbb{P}(X_1 > \mathrm{VaR}_\alpha(X_1))} \\
&= \lim_{\alpha\downarrow 0} \left(\frac{\sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)}{\mathrm{VaR}_\alpha(X_1)}\right)^{-\gamma_1} = 1.
\end{aligned}$$

Thus, we get the desired result. □

Proposition 6.4 illustrates the intuitive fact that, if the tail of one asset is strictly heavier than the others, then the portfolio has no diversification in the tail region, i.e., as $\alpha \downarrow 0$.

6.7 Optimization for the elliptical models and MRV models

We analyze portfolio diversification for a random vector $\mathbf{X} \in \mathcal{X}^n$ representing losses from n assets and a vector $\mathbf{w} = (w_1, \dots, w_n) \in \Delta_n$ of portfolio weights, where

$$\Delta_n := \{\mathbf{x} \in [0,1]^n : x_1 + \cdots + x_n = 1\}.$$

The total loss of the portfolio is $\mathbf{w}^\top \mathbf{X}$. We write $\mathbf{w} \odot \mathbf{X} = (w_1 X_1, \dots, w_n X_n)$ which is the portfolio loss vector with the weight $\mathbf{w}$. For a portfolio selection problem, we need to treat $\mathrm{DQ}_\alpha^\rho(\mathbf{w} \odot \mathbf{X})$ as a function of the portfolio weight $\mathbf{w}$.

Chapter 4 studied the following optimization diversification problem

$$\min_{\mathbf{w} \in \Delta_n} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X}) \quad \text{and} \quad \min_{\mathbf{w} \in \Delta_n} \mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w} \odot \mathbf{X}); \tag{6.13}$$

for general $\mathbf{X}$. Moreover, efficient algorithms are obtained to optimize $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ in real-data applications; see their Sections 6.2 and 7. In this section, we focus on the portfolio optimization problems for elliptical and MRV models.

For the elliptical models, the optimization of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$, $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ boils down to maximizing $k_{\mathbf{w}\Sigma\mathbf{w}^\top}$ in (6.7) since DQ of $\mathbf{w} \odot \mathbf{X}$ is decreasing in $k_{\mathbf{w}\Sigma\mathbf{w}^\top}$. We assume that Σ is invertible, and write $\Sigma = (\sigma_{ij})_{n\times n}$, with diagonal entries $\sigma_{ii} = \sigma_i^2$, $i \in [n]$, and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$. Note that

$$k_{\mathbf{w}\Sigma\mathbf{w}^\top} = \frac{\mathbf{w}^\top \boldsymbol{\sigma}}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}},$$

and we immediately give the optimizer of (6.13) for the elliptical models.

Theorem 6.4. *Suppose that* $\mathbf{X} \sim \mathrm{E}_n(\boldsymbol{\mu}, \Sigma, \tau)$, Σ *is invertible and* $\alpha \in (0, 1/2)$, *then the vector*

$$\mathbf{w}^* = \arg\max_{\mathbf{w} \in \Delta_n} \frac{\mathbf{w}^\top \boldsymbol{\sigma}}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}} \tag{6.14}$$

minimizes (6.13), *that is,*

$$\min_{\mathbf{w} \in \Delta_n} \mathrm{DQ}_\alpha^\rho(\mathbf{w} \odot \mathbf{X}) = \mathrm{DQ}_\alpha^\rho(\mathbf{w}^* \odot \mathbf{X}) \tag{6.15}$$

for ρ *being* VaR *or* ES.

The optimization problem (6.14) is well studied in the literature, and the existence and uniqueness of the solution can be verified if Σ is invertible, see, e.g. Choueifaty and Coignard (2008). Note that the optimizer for problem (6.15) does not depend on the tail probability level α. It is straightforward to see that

$$\arg\min_{\mathbf{w}\in\Delta_n} \mathrm{DR}^{\rho_\alpha}(\mathbf{w}\odot\mathbf{X}) = \arg\max_{\mathbf{w}\in\Delta_n} \frac{\mathbf{w}^\top\boldsymbol{\mu}+\mathbf{w}^\top\boldsymbol{\sigma}\rho_\alpha(Y)}{\mathbf{w}^\top\boldsymbol{\mu}+\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}}\rho_\alpha(Y)}$$

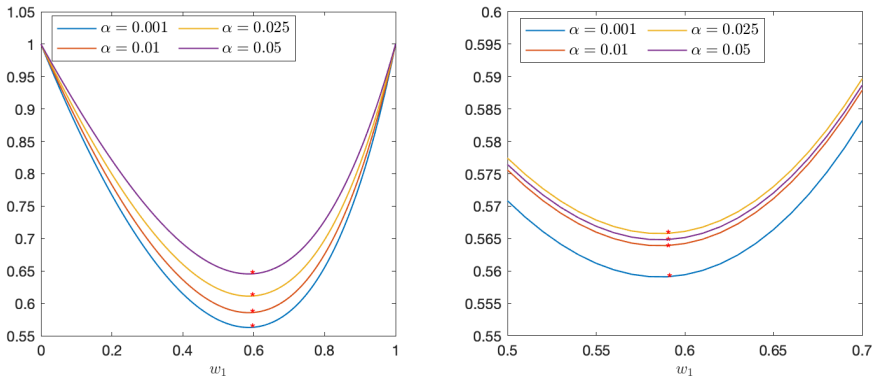
for ρ being VaR or ES and $Y\sim \mathrm{E}_1(0,1,\tau)$. This optimizer is the same as that of (6.15) if $\boldsymbol{\mu}=\mathbf{0}$. This shows that for centered elliptical models, optimizing DQ and optimizing DR are equivalent problems, both of which are further equivalent to optimizing DR based on SD (assuming it exists). This is intuitive as for a fixed τ, centered elliptical distributions are parameterized by their dispersion matrices.

Example 6.1. Assume that $\mathbf{X}\sim \mathrm{t}(\nu,\boldsymbol{\mu},\Sigma)$ where $\nu=3$ and the dispersion matrix is given by

$$\Sigma=\begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}.$$

Clearly, DQ does not depend on $\boldsymbol{\mu}$. We show the curves of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w}\odot\mathbf{X})$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w}\odot\mathbf{X})$ against the weight w_1 with various values of $\alpha=0.001,0.01,0.025,0.05$. It can be anticipated from (6.14) that although DQ depends on α, the optimizer does not. By solving (6.14), we get $w_1^*=0.5860$ and $w_2^*=0.4140$, which corresponds to the observations in Figure 6.5. Recall that $\mathrm{DQ}_\alpha^{\mathrm{ES}}$ is quite flat when α varies in Figure 6.3, and hence curves of $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w}\odot\mathbf{X})$ look similar for different α.

Figure 6.5: Values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w}\odot\mathbf{X})$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{w}\odot\mathbf{X})$ for $w_1\in[0,1]$



Next, we turn to the MRV model. The following result gives the limit of DQ of the portfolio $\mathbf{w} \odot \mathbf{X}$ where $\mathbf{X}$ follows an MRV model. Due to the same reason stated in Section 6.6, we only present the case of VaR. In the proofs below, for any positive functions f and g, we write $f(x) \simeq g(x)$ as $x \to x_0$ to represent $\lim_{x \to x_0} f(x)/g(x) = 1$.

Proposition 6.5. *Suppose that* $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ *and* $\mathbf{X}$ *has positive joint density on the support of* $\mathbf{X}$. *Then, for* $\mathbf{w} \in \Delta_n$,

$$\lim_{\alpha \downarrow 0} \mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X}) = f(\mathbf{w}),$$

where $f(\mathbf{w}) = \eta_{\mathbf{w}} / \left(\sum_{i=1}^n w_i \eta_{\mathbf{e}_i}^{1/\gamma} \right)^\gamma$ *and* $\eta_{\mathbf{x}} = \int_{\mathbb{S}^n} \left(\mathbf{x}^\top \mathbf{s} \right)_+^\gamma \Psi(\mathrm{d}\mathbf{s})$ *for* $\mathbf{x} \in \mathbb{R}^n$.

Proof. If $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ with $\gamma \in (0, 1)$, we have (Lemma 2.2 of Mainik and Embrechts (2013))

$$\lim_{\alpha \downarrow 0} \frac{\mathrm{VaR}_\alpha \left(\sum_{i=1}^n w_i X_i \right)}{\mathrm{VaR}_\alpha \left(\|\mathbf{X}\|_1 \right)} = \eta_{\mathbf{w}}^{1/\gamma},$$

and

$$\lim_{\alpha \downarrow 0} \sum_{i=1}^n \frac{w_i \mathrm{VaR}_\alpha \left(X_i \right)}{\mathrm{VaR}_\alpha \left(\|\mathbf{X}\|_1 \right)} = \sum_{i=1}^n w_i \eta_{\mathbf{e}_i}^{1/\gamma},$$

where $\|\mathbf{X}\|_1 = \sum_{i=1}^n |X_i|$. As $\mathbf{X}$ has positive joint density, VaR_α is continuous for $\sum_{i=1}^n w_i X_i$. Then we have $\mathrm{VaR}_{\alpha^*}(\sum_{i=1}^n w_i X_i) = \sum_{i=1}^n w_i \mathrm{VaR}_\alpha(X_i)$. Thus, it follows that

$$\frac{\mathrm{VaR}_\alpha \left(\sum_{i=1}^n w_i X_i \right)}{\mathrm{VaR}_{\alpha^*}(\sum_{i=1}^n w_i X_i)} \to \frac{\eta_{\mathbf{w}}^{1/\gamma}}{\sum_{i=1}^n w_i \eta_{\mathbf{e}_i}^{1/\gamma}} \qquad \text{as } \alpha \downarrow 0.$$

Since $\sum_{i=1}^n w_i X_i \in \mathrm{RV}_\gamma$, for $c > 0$,

$$\frac{\mathrm{VaR}_\alpha \left(\sum_{i=1}^n w_i X_i \right)}{\mathrm{VaR}_{c\alpha} \left(\sum_{i=1}^n w_i X_i \right)} \simeq \left(\frac{1}{c} \right)^{-1/\gamma} \qquad \text{as } \alpha \downarrow 0.$$

Let $c = \alpha^*/\alpha$, we have

$$\left(\frac{\alpha}{\alpha^*} \right)^{-1/\gamma} \to \frac{\eta_{\mathbf{w}}^{1/\gamma}}{\sum_{i=1}^n w_i \eta_{\mathbf{e}_i}^{1/\gamma}}.$$

Hence,

$$\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X}) = \frac{\alpha^*}{\alpha} \to \frac{\eta_{\mathbf{w}}}{\left(\sum_{i=1}^n w_i \eta_{\mathbf{e}_i}^{1/\gamma} \right)^\gamma}.$$

The desired result is obtained. □

Proposition 6.5 allows us to approximately optimize $\mathrm{DQ}_\alpha^{\mathrm{VaR}}$ by minimizing $f(\mathbf{w})$. For $\mathbf{X} \in \mathrm{MRV}_\gamma(\Psi)$ with $\gamma > 1$, by assuming $\Psi\left(\left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = 0\right\}\right) = 0$ for any $\mathbf{a} \in \mathbb{R}^n$, which means that all components are relevant for the extremes of $\mathbf{X}$, the existence and uniqueness of $\mathbf{w}^* = \arg\min_{\mathbf{w}\in\Delta_n} f(\mathbf{w})$ are guaranteed. In fact, the existence of $\mathbf{w}^*$ is due to the continuity of $f(\mathbf{w})$ and the compactness of Δ_n. To show uniqueness, we can rewrite the above minimization problem as

$$\min_{\mathbf{w}\in\Delta_n} \eta_{\mathbf{w}} \quad \text{s.t.} \quad \sum_{i=1}^{d} w_i \eta_{\mathbf{e}_i}^{1/\gamma} = 1.$$

Note that the set of constraints is compact and $\eta_{\mathbf{w}}$ is strictly convex, and hence $\mathbf{w}^*$ is unique.

Example 6.2. Assume that Y_1 and Y_2 are iid following a standard t-distribution with $\nu > 1$ degrees of freedom. A random vector $\mathbf{X} = (X_1, X_2)$ is defined as

$$\mathbf{X} = A\mathbf{Y} \quad \text{with} \quad A = \begin{pmatrix} 1 & 0 \\ r & \sqrt{1-r^2} \end{pmatrix}.$$

The random vectors $\mathbf{X}$ and $\mathbf{Y}$ are not elliptically distributed. Using the results in Mainik and Embrechts (2013), we have

$$\frac{\eta_{\mathbf{w}}}{\eta_{\mathbf{1}_1}} = (w_1 + w_2 r)^\nu + \left(w_2\sqrt{1-r^2}\right)^\nu,$$

and

$$\frac{\eta_{\mathbf{w}}}{\eta_{\mathbf{1}_2}} = \frac{(w_1 + w_2 r)^\nu + \left(w_2\sqrt{1-r^2}\right)^\nu}{r^\nu + \sqrt{1-r^2}^\nu}.$$

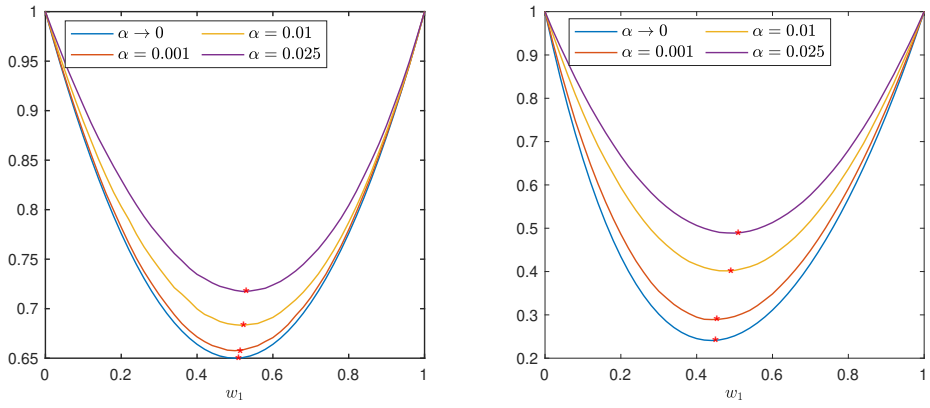
Hence,

$$f(\mathbf{w}) = \left(w_1 \left((w_1 + w_2 r)^\nu + \left(w_2\sqrt{1-r^2}\right)^\nu \right)^{-\frac{1}{\nu}} + w_2 \left(\frac{(w_1 + w_2 r)^\nu + \left(w_2\sqrt{1-r^2}\right)^\nu}{r^\nu + \sqrt{1-r^2}^\nu} \right)^{-\frac{1}{\nu}} \right)^{-\nu}.$$

Take $r = 0.3$. We show the curves of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X})$ against w_1 for $\alpha = 0.001, 0.01, 0.025$ and $\nu = 2, 4$. Also, we use $f(\mathbf{w})$ to approximate $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w} \odot \mathbf{X})$ as α tends to 0. From Figure 6.6, we can see that the optimizer w_1^* is converging to the one that maximizes $f(\mathbf{w})$ as α tends to 0.

Remark 6.9. Some negative dependence concepts yield small values of DQ. The joint mix dependence usually leads to a zero DQ as we see in Theorem 6.1 (ii). The negative dependence

Figure 6.6: Values of $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{w}\odot\mathbf{X})$ with $\nu=2$ (left) and $\nu=4$ (right)



concept of Lee and Ahn (2014), weaker than joint mix dependence, does not necessarily lead to a small value of DQ. For instance, the portfolio vector $\mathbf{X}=(X,-\epsilon X)$ is counter-monotonic for $\epsilon>0$, but its DQ can be close to 1 for small ϵ. In particular, we have $\mathrm{DQ}_\alpha^{\mathrm{VaR}}(\mathbf{X})\approx 0.9333$ and $\mathrm{DQ}_\alpha^{\mathrm{ES}}(\mathbf{X})\approx 0.9044$ for $\alpha=0.05$ and $\epsilon=0.01$ when X follows a standard normal distribution.

6.8 Conclusion

The DQs based on VaR and ES are investigated in this chapter, following the theory of DQ in Chapter 4. In particular, for elliptical and MRV models, these DQs have simple forms. Comparisons between DQ and DR illustrate some attractive features of DQ. These results enhance the theory and applications of DQ.

We summarize some features below. (i) In cases of VaR and ES, DQs have simple formulas, in a way comparable to DRs. (ii) DQs based on VaR and ES take values in bounded intervals and have natural ranges of $[0,n]$ and $[0,1]$, respectively. The special values 0, 1 and n which correspond to special dependence structures can be constructed. (iii) DQs based on VaR and ES for elliptical distributions and MRV models have convenient expressions and it can capture heavy tails in an intuitive way. (iv) Portfolio optimization for elliptical models boils down to a well-studied problem in the literature. For centered elliptical models, optimizing DQ and optimizing DR are equivalent problems.

We discuss some future directions for the research of DQ. As a potential alternative to ES, expectiles (Bellini et al. (2014)) have received increasing attention in the recent literature; indeed, they are the only elicitable coherent risk measures (Ziegel (2016)). It would be interesting to formulate DQ based on expectiles and investigate its properties that are different from DQ based on ES or VaR. As another interesting class of risk measures, the optimized certainty equivalents (Ben-Tal and Teboulle (2007)) are introduced from decision-theoretic criterion based on utility functions. It would be useful to construct DQ based on utility functions or optimized certainty equivalents and analyze the decision-theoretic implications of DQ.

Chapter 7

Pairwise counter-monotonicity

7.1 Introduction

Dependence modelling is a crucial part of modern quantitative studies in economics, finance, and insurance (McNeil et al. (2015)). Comonotonicity and counter-monotonicity are known as the strongest forms of positive and negative dependence, respectively. In quantitative risk management, assuming knowledge of the marginal distributions, comonotonicity corresponds to the most dangerous dependence structure (Denneberg (1994) and Dhaene et al. (2002, 2006)) for the aggregate risk, whereas counter-monotonicity corresponds to the safest. In dimensions higher than 2, by counter-monotonicity we mean pairwise counter-monotonicity (Dall'Aglio (1972)), which has been studied under the name of mutual exclusivity in the actuarial literature (Dhaene et al. (1999) and Cheung and Lo (2014)).[1]

Despite the obvious similarity in their definitions, comonotonicity and counter-monotonicity are asymmetric in several major senses. For instance, comonotonicity admits a stochastic representation (see Lemma 7.1 below), but such a representation is not known for pairwise counter-monotonicity. Moreover, for any given tuple of marginal distributions, a comonotonic random vector with these marginal distributions always exists, but a pairwise counter-

[1]Mutual exclusivity is defined using joint exceedance probability (see Section 7.2). The two definitions are shown to be equivalent first by Dall'Aglio (1972, Lemma 2) and in a more precise form by Cheung and Lo (2014, Theorem 4.1).

monotonic one may not exist unless quite restrictive conditions on the marginal distributions are satisfied, as first studied by Dall'Aglio (1972). In particular, a pairwise counter-monotonic random vector cannot have continuous marginal distributions. Comonotonicity has many important roles in economics, finance and actuarial science, and as such it has received great attention in the literature, as in axiomatization of preferences (Yaari (1987); Schmeidler (1989)), risk measures (Kusuoka (2001)) and premium principles (Wang et al. (1997)), risk sharing (Landsberger and Meilijson (1994); Jouini et al. (2008)), insurance design (Huberman et al. (1983); Carlier and Dana (2003)), risk aggregation (Embrechts et al. (2015)), and optimal transport (Rüschendorf (2013)).

In sharp contrast to the rich literature on comonotonicity, research on pairwise counter-monotonicity is quite limited. As a dependence concept, pairwise counter-monotonicity has been studied by Dall'Aglio (1972), Hu and Wu (1999), Dhaene et al. (1999) and Cheung and Lo (2014), but the list of relevant studies do not grow much longer. In contrast to the relatively limited studies on pairwise counter-monotonicity, this dependence structure appears naturally in many economic contexts, such as lottery tickets, Bitcoin mining, gambling, and mutual aid platforms, whenever payment events are mutually exclusive. In particular, the interest in studying pairwise counter-monotonicity has grown in the recent risk sharing literature. A pairwise counter-monotonic structure is the essential building block of any optimal allocation for agents using Value-at-Risk (VaR, which are quantiles) and quantile-related risk measures; such problems are studied by Embrechts et al. (2018) and generalized by Weber (2018), Embrechts et al. (2020), Liu et al. (2022) and Xia et al. (2023). Moreover, counter-monotonicity, when possible, serves as the best-case dependence structure in risk aggregation for some common risk measures, and, in some contexts, it also serves as the worst-case dependence structure for VaR (see Example 7.1 in Section 7.2).

This chapter is dedicated to a systematic study of pairwise counter-monotonicity. As comonotonicity and counter-monotonicity are classic and prominent concepts in mathematics and its applications with a long history, at least since the seminal work of Hardy et al. (1934), one may guess that there is not much more to discover about them. To our pleasant surprise, we offer, through the development of this chapter, many new results on counter-monotonicity, some of which are motivated by recent developments in risk management.

We obtain a new stochastic representation for pairwise counter-monotonic random vectors using their component-wise sum in Theorem 7.1, which will be useful for many other results in this chapter. The second result, Theorem 7.2, establishes that counter-monotonicity is preserved under increasing transforms on disjoint sets of components of a random vector, which is an invariance property proposed by Joag-Dev and Proschan (1983) satisfied by negative association (Alam and Saxena (1981)). Using this invariance property, we obtain in Theorem 7.3 that counter-monotonicity implies negative association. The notion of negative association is stronger than many other forms of negative dependence, such as negative orthant dependence (Block et al. (1982)) and negative supermodular dependence (Hu (2000)). In particular, Theorem 7.3 surpasses a result of Dhaene et al. (1999) showing that counter-monotonicity implies negative supermodular dependence.

Another negative dependence concept is joint mix dependence (Wang and Wang (2011, 2016)), which can be used to optimize many quantities in risk aggregation; see Wang et al. (2013) and Rüschendorf (2013). To connect counter-monotonicity and joint mix dependence, we fully characterize all Fréchet classes (Joe (1997)) which are compatible with both dependence concepts in Theorem 7.4; it turns out that the two notions, when both exist in the same Fréchet class, are equivalent. Finally, we show in Theorem 7.5 that in the context of risk sharing for quantile agents (Embrechts et al. (2018)), under some mild conditions on the total loss, there always exists a pairwise counter-monotonic Pareto-optimal allocation, and any pairwise counter-monotonic allocation is Pareto optimal for some agents. As a consequence, pairwise counter-monotonic random vectors are natural for agents that are not risk averse. This is in stark contrast to comonotonic allocations, which appear prominently for risk-averse agents (in the sense of Rothschild and Stiglitz (1970)) as a consequence of comonotonic improvements introduced by Landsberger and Meilijson (1994).

7.2 Preliminaries

We first define comonotonicity and counter-monotonicity for bivariate random variables. Fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The probability space does not need to be atomless in Sections 7.2-7.4. We treat almost surely (a.s.) equal random variables as identical; this

means that all equalities and inequality for random variables hold in the a.s. sense, and we omit "a.s." in all our statements. Terms like "increasing" are in the non-strict sense. Let n be a positive integer and $[n] = \{1, \dots, n\}$. Throughout, we consider $n \geqslant 2$.

A bivariate random vector (X, Y) is *comonotonic* if there exist increasing functions f, g and a random variable Z such that $(X, Y) = (f(Z), g(Z))$. A bivariate random vector (X, Y) is *counter-monotonic* if $(X, -Y)$ is comonotonic. An equivalent formulation of comonotonicity is

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0 \text{ for } (\mathbb{P} \times \mathbb{P})\text{-almost every } (\omega, \omega') \in \Omega^2.$$

An equivalent formulation of counter-monotonicity is

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \leqslant 0 \text{ for } (\mathbb{P} \times \mathbb{P})\text{-almost every } (\omega, \omega') \in \Omega^2.$$

Next, we define these concepts in dimensions higher than 2. For $n \geqslant 3$, a random vector $\mathbf{X}$ taking values in $\mathbb{R}^n$ is *(pairwise) comonotonic* if each pair of its components is comonotonic, and it is *(pairwise) counter-monotonic* if each pair of its components is counter-monotonic.[2] We will say "pairwise counter-monotonicity" to emphasize the case $n \geqslant 3$ and simply say "counter-monotonicity" when we also include dimension 2. We always omit "pairwise" for comonotonicity, for which the distinction between dimensions $n = 2$ and $n \geqslant 3$ is unnecessary.

There are many equivalent ways of formulating comonotonicity and counter-monotonicity; see Puccetti and Wang (2015, Section 3.2) for a review. For instance, they can be formulated using joint distributions. A comonotonic random vector and a counter-monotonic random vector have, respectively, the largest and the smallest joint distribution functions among all random vectors with the same marginals. With given marginals, the largest (resp. smallest) joint distribution function is known as the Fréchet-Hoeffding upper (resp. lower) bound.

A stochastic representation of comonotonicity, which follows from Denneberg (1994, Proposition 4.5), is presented in the next lemma.

Lemma 7.1 (Denneberg (1994)). *Let $(X_1, \dots, X_n)$ be a random vector and denote by $S = \sum_{i=1}^n X_i$. The following are equivalent.*

[2] We also say that random variables $X_1, \dots, X_n$ are comonotonic (counter-monotonic), which means that the random vector $(X_1, \dots, X_n)$ is comonotonic (counter-monotonic).

(i) $(X_1, \dots, X_n)$ *is comonotonic.*

(ii) *There exist increasing functions* $f_1, \dots, f_n$ *and a random variable* Z *such that* $X_i = f_i(Z)$ *for all* $i \in [n]$.

(iii) *There exist continuously increasing functions* $f_1, \dots, f_n$ *such that* $X_i = f_i(S)$ *for all* $i \in [n]$.

Lemma 7.1 implies that a comonotonic vector can be represented by increasing functions of the sum S. Such a representation result does not exist for pairwise counter-monotonicity, since the sum S cannot determine the components $(X_1, \dots, X_n)$ in the presence of negative dependence.

Although quite different from comonotonicity, pairwise counter-monotonicity also has a special structure, presented below in Lemma 7.2, which is a restatement of Lemma 2 and Theorem 3 of Dall'Aglio (1972). This result will be useful in a few places in the chapter. The current form of this lemma can be found in Theorem 4.1 of Cheung and Lo (2014) and Proposition 3.2 of Puccetti and Wang (2015). Denote by ess-inf$X$ and ess-sup$X$ the essential infimum and essential supremum of a random variable X, respectively.

Lemma 7.2 (Dall'Aglio (1972))**.** *If at least three of* $X_1, \dots, X_n$ *are non-degenerate, pairwise counter-monotonicity of* $(X_1, \dots, X_n)$ *means that one of the following two cases holds true:*

$$\mathbb{P}(X_i > \text{ess-inf}X_i,\ X_j > \text{ess-inf}X_j) = 0 \text{ for all } i \neq j; \tag{7.1}$$

$$\mathbb{P}(X_i < \text{ess-sup}X_i,\ X_j < \text{ess-sup}X_j) = 0 \text{ for all } i \neq j. \tag{7.2}$$

A necessary condition for (7.1) *is* $\sum_{i=1}^{n} \mathbb{P}(X_i > \text{ess-inf}X_i) \leqslant 1$, *and a necessary condition for* (7.2) *is* $\sum_{i=1}^{n} \mathbb{P}(X_i < \text{ess-sup}X_i) \leqslant 1$.

In the actuarial literature, mutual exclusivity of $(X_1, \dots, X_n)$ is defined as either (7.1) or (7.2); see Cheung and Lo (2014).

Pairwise counter-monotonicity imposes strong constraints on the marginal distributions. For instance, the necessary condition in case of (7.1) is equivalent to $\sum_{i=1}^{n} \mathbb{P}(X_i = \text{ess-inf}X_i) \geqslant n - 1$, and it implies, in particular, that $X_1, \dots, X_n$ are bounded from below.

Moreover, given $n \geqslant 3$ non-degenerate marginal distributions, a pairwise counter-monotonic random vector exists if and only if one of the two necessary conditions on the marginal distributions holds (Theorem 3 of Dall'Aglio (1972)).

Example 7.1. We illustrate the special role of counter-monotonicity in risk aggregation with a simple model. Let $F_1, \dots, F_n$ be Bernoulli distributions with mean $\epsilon \in (0, 1/n)$. These distributions may represent losses from credit default events in a pre-specified period, which usually occur with a small probability. In risk aggregation problems (e.g., Embrechts et al. (2013, 2015)), we are interested in the minimum (best-case) value or maximum (worst-case) value of

$$\rho\left(\sum_{i=1}^{n} X_i\right) \quad \text{with the marginal condition } X_i \sim F_i,\ i \in [n], \tag{7.3}$$

where ρ is a risk measure, and $\sum_{i=1}^n X_i$ represents the total loss from a portfolio of defaultable bonds, with the probability of default ϵ estimated from the credit rating of these bonds, assumed to be equal for simplicity. We consider two choices of ρ, which lead to opposite conclusions.

(a) Let ρ be a risk measure that is consistent with convex order. Such risk measures are characterized by Mao and Wang (2020), and they include all law-invariant coherent, as well as convex, risk measures, such as the Expected Shortfall (Föllmer and Schied (2016)). The *minimum* value of (7.3) is obtained by a counter-monotonic random vector $(X_1, \dots, X_n)$. This result holds for other marginal distributions as long as a counter-monotonic random vector with these marginal distributions exists; see e.g., Lemma 3.6 of Cheung and Lo (2014).

(b) Let $\rho : X \mapsto \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1 - \alpha\}$, which is the risk measure VaR_α in Section 7.6. Further, assume that $\alpha/\epsilon \in (n/2, n)$. The *maximum* value of (7.3) is obtained by a counter-monotonic random vector $(X_1, \dots, X_n)$, as explained below. First, since $\sum_{i=1}^n X_i$ only takes integer values, so does $\rho(\sum_{i=1}^n X_i)$. If $(X_1, \dots, X_n)$ is counter-monotonic, then $\sum_{i=1}^n X_i$ follows a Bernoulli distribution with mean $n\epsilon > \alpha$, and hence $\rho(\sum_{i=1}^n X_i) = 1$. Moreover, for any $X_1, \dots, X_n$ with the specified marginal distributions, if $\rho(\sum_{i=1}^n X_i) \geqslant 2$ then $\mathbb{E}[\sum_{i=1}^n X_i] \geqslant 2\alpha > n\epsilon$, a contradiction, thus showing $\rho(\sum_{i=1}^n X_i) \leqslant 1$.

The interpretation of the above two cases is that, for credit default losses, using a coherent risk measure and using VaR may lead to opposite conclusions on which dependence structure is safe or dangerous, and both cases highlight the important role of counter-monotonicity.

7.3 Stochastic representation

We provide in this section a stochastic representation of pairwise counter-monotonicity. To explain the result, let Π_n be the set of all n-compositions of Ω, that is,

$$\Pi_n = \left\{ (A_1, \dots, A_n) \in \mathcal{A}^n : \bigcup_{i \in [n]} A_i = \Omega \text{ and } A_1, \dots, A_n \text{ are disjoint} \right\}.$$

In other words, a composition of Ω is a partition of Ω with order. Denote by $\mathcal{X}_{\pm}$ the set of all nonnegative random variables and nonpositive random variables.

Theorem 7.1. *Let $(X_1, \dots, X_n)$ be a random vector and denote by $S = \sum_{i=1}^n X_i$. Suppose that at least three of $X_1, \dots, X_n$ are non-degenerate. The following are equivalent.*

(i) $(X_1, \dots, X_n)$ is pairwise counter-monotonic.

(ii) There exist $m_1, \dots, m_n \in \mathbb{R}$, $(A_1, \dots, A_n) \in \Pi_n$ and $Z \in \mathcal{X}_{\pm}$ such that

$$X_i = Z \mathbb{1}_{A_i} + m_i \quad \text{for all } i \in [n]. \tag{7.4}$$

(iii) There exists $(A_1, \dots, A_n) \in \Pi_n$ such that

$$X_i = (S - m) \mathbb{1}_{A_i} + m_i \quad \text{for all } i \in [n], \tag{7.5}$$

where either $m_i = \text{ess-inf} X_i$ for $i \in [n]$ or $m_i = \text{ess-sup} X_i$ for $i \in [n]$, and $m = \sum_{i=1}^n m_i$.

Proof. The implication (iii)⇒(ii) is straightforward. To see (ii)⇒(i), take $i, j \in [n]$ with $i \neq j$, and we check a few cases of $\omega, \omega' \in \Omega$. If $\omega, \omega' \notin A_i$, then $X_i(\omega) = X_i(\omega') = m_i$, and hence

$$(X_i(\omega) - X_i(\omega'))(X_j(\omega) - X_j(\omega')) = 0. \tag{7.6}$$

Similarly, (7.6) holds if $\omega, \omega' \notin A_j$. If $(\omega, \omega') \in A_i \times A_j$ or $(\omega, \omega') \in A_j \times A_i$, then

$$(X_i(\omega) - X_i(\omega'))(X_j(\omega) - X_j(\omega')) = -Z(\omega)Z(\omega') \leqslant 0.$$

This shows that (X_i, X_j) is counter-monotonic, and hence, $(X_1, \dots, X_n)$ is pairwise counter-monotonic.

Next, we show the implication (i)⇒(iii). By Lemma 7.2, it suffices to consider (7.1) and (7.2). Suppose that (7.1) holds. Let $B_i = \{X_i > \text{ess-inf} X_i\}$ and $m_i = \text{ess-inf} X_i$ for $i \in [n]$. Clearly $B_1, \dots, B_n$ are (a.s.) disjoint events, and $S \geqslant \sum_{i=1}^n m_i = m$. Using (7.1), if event B_i occurs, then $X_j = m_j$ for $j \neq i$, and $S = X_i + \sum_{j=1}^n m_j - m_i$. Moreover, if B_i does not occur, then $X_i = m_i$. Therefore, we have

$$X_i = (S - m + m_i)\mathbb{1}_{B_i} + m_i \mathbb{1}_{B_i^c} = (S - m)\mathbb{1}_{B_i} + m_i, \quad \text{for } i \in [n]. \tag{7.7}$$

Let $B = \{S = m\}$ and it is clear that $(B, B_1, \dots, B_n)$ is a composition of Ω. Let $A_1 = B_1 \cup B$, and $A_2 = B_2, \dots, A_n = B_n$. Since $S - m = 0$ on B, (7.7) yields (7.5). If (7.2) holds instead of (7.1), then we can analogously show (7.5) with $m_i = \text{ess-sup} X_i$ for $i \in [n]$. □

Theorem 7.1 shows that pairwise counter-monotonicity can be represented by the sum S and a composition $(A_1, \dots, A_n)$. In contrast, comonotonicity can be represented by the sum S and increasing continuous functions $f_1, \dots, f_n$, as in Lemma 7.1. This representation result will be instrumental in proving the other results of this chapter. Another direct consequence of Theorem 7.1 is that if at least three components of a pairwise counter-monotonic random vector are non-degenerate, then either the components are all bounded from below or they are all bounded from above; this can also be seen from Lemma 7.2.

Example 7.2. A simple pairwise counter-monotonic random vector in the form of (7.4) and (7.5), which will be referred to repeatedly in the following sections, is given by

$$X_i = \mathbb{1}_{A_i} \text{ for } i \in [n] \text{ where } (A_1, \dots, A_n) \in \Pi_n. \tag{7.8}$$

Such $(X_1, \dots, X_n)$ may represent the outcome of n lottery tickets, exactly one of which randomly wins a reward of 1, or the reward to Bitcoin miners computing the next block in the Bitcoin blockchain; see Leshno and Strack (2020).

Remark 7.1. In parts (ii) and (iii) of Theorem 7.1, we can replace $(A_1, \dots, A_n) \in \Pi_n$ by $A_1, \dots, A_n$ being disjoint events, and the equivalence relations in the theorem remain true.

In the case at most two components of $(X_1, \dots, X_n)$ are non-degenerate, the stochastic representation of counter-monotonicity is quite different from Theorem 7.1. When $n = 2$, (X_1, X_2) is counter-monotonic if and only if there exist increasing functions f_1, f_2 such that

$$X_1 = f_1(X_1 - X_2) \quad \text{and} \quad X_2 = f_2(X_2 - X_1);$$

this statement follows by applying Lemma 7.1 to the comonotonic random vector $(X_1, -X_2)$. Note that the difference $X_1 - X_2$ replaces the summation $S = X_1 + X_2$ in Lemma 7.1. The sum of two counter-monotonic random variables represents the loss from a hedged portfolio and it has been studied by Cheung et al. (2014) and Chaoubi et al. (2020).

7.4 Invariance property and negative association

Negative association appears in various natural probabilistic and statistical contexts, such as permutation distributions, sampling without replacement, negatively correlated Gaussian distributions and tournament scores; see Joag-Dev and Proschan (1983) and the more recent paper Chi et al. (2022) for many examples.

A random vector $\mathbf{X} = (X_1, \dots, X_n)$ is said to be *negatively associated* if for any disjoint subsets $I, J \subseteq [n]$, and any real-valued, coordinate-wise increasing functions f, g, we have

$$\mathrm{Cov}(f(\mathbf{X}_I), g(\mathbf{X}_J)) \leqslant 0, \tag{7.9}$$

where $\mathbf{X}_I = (X_k)_{k \in I}$ and $\mathbf{X}_J = (X_k)_{k \in J}$, provided that $f(\mathbf{X}_I)$ and $g(\mathbf{X}_J)$ have finite second moments. Negative association is stronger than many other notions of negative dependence, such as negative supermodular dependence (shown by Christofides and Vaggelatou (2004)) and negative orthant dependence (shown by Joag-Dev and Proschan (1983)).

Remark 7.2. Negative association is invariant under increasing marginal transforms. Therefore, if $f(\mathbf{X}_I)$ and $g(\mathbf{X}_J)$ are continuously distributed, then NA implies that (7.9) holds with the covariance operator replaced by Spearman's rank correlation coefficient or another similar concordance measure; see McNeil et al. (2015, Chapter 7).

We first present a self-consistency property of both comonotonicity and counter-monotonicity in the spirit of Property P6 of Joag-Dev and Proschan (1983) for negative association. To the best of our knowledge, this self-consistency property is not found in the literature even for the case of comonotonicity, although its proof is straightforward.

Theorem 7.2. *The following statements hold.*

(i) Increasing functions of subsets of a set of comonotonic random variables are comonotonic.

(ii) Increasing functions of disjoint subsets of a set of counter-monotonic random variables are counter-monotonic.

Proof. (i) Let $\mathbf{X} = (X_1, \dots, X_n)$ be a comonotonic random vector. By Lemma 7.1, there exist increasing functions $f_1, \dots, f_n$ and a random variable Z such that $X_i = f_i(Z)$ for all $i \in [n]$. For $I_1, \dots, I_m \subseteq [n]$ and increasing functions $g_j : \mathbb{R}^{|I_j|} \to \mathbb{R}$, $j \in [m]$, let $Y_j = g_j(\mathbf{X}_{I_j})$, $j \in [m]$, where $|\cdot|$ is the cardinality of a set. That is, $Y_j = g_j \circ f_{I_j}(Z)$ where $f_{I_j} = (f_i)_{i \in I_j}$. As the composition of increasing functions, $g_i \circ f_{I_j}$ is increasing on $\mathbb{R}$. Thus, $(Y_1, \dots, Y_m)$ is a comonotonic vector.

(ii) Let $\mathbf{X} = (X_1, \dots, X_n)$ be a pairwise counter-monotonic random vector. If at most two of $X_1, \dots, X_n$ are non-degenerate, the desired statement holds trivially. Next, we assume that at least three of $X_1, \dots, X_n$ are non-degenerate. For disjoint subsets $I_1, \dots, I_m$ of $[n]$ and increasing functions $g_j : \mathbb{R}^{|I_j|} \to \mathbb{R}$, $j \in [m]$, let $Y_j = g_j(\mathbf{X}_{I_j})$, $j \in [m]$. By Theorem 7.1, there exist $\mathbf{m} = (m_1, \dots, m_n) \in \mathbb{R}^n$, $(A_1, \dots, A_n) \in \Pi_n$ and $Z \in \mathcal{X}_{\pm}$ such that $X_i = Z \mathbb{1}_{A_i} + m_i$ for all $i \in [n]$. Without loss of generality, assume $Z \geqslant 0$. For $i \in [n]$ and $j \in [m]$, if A_i occurs, then $X_i = Z + m_i$ and $X_k = m_k$ for $k \neq i$, which means $Y_j = g_j(\mathbf{X}_{I_j}) \geqslant g_j(\mathbf{m}_{I_j})$. If A_i does not occur, then $Y_j = g_j(\mathbf{m}_{I_j})$. Let $Z_j = \sum_{i \in I_j} \left(g_j(\mathbf{X}_{I_j}) - g_j(\mathbf{m}_{I_j}) \right) \mathbb{1}_{A_i} \geqslant 0$. It follows that

$$\begin{aligned} Y_j &= \sum_{i \in I_j} g_j(\mathbf{X}_{I_j}) \mathbb{1}_{A_i} + g_j(\mathbf{m}_{I_j}) \left(1 - \sum_{i \in I_j} \mathbb{1}_{A_i} \right) \\ &= Z_j \mathbb{1}_{\bigcup_{i \in I_j} A_i} + g_j(\mathbf{m}_{I_j}) = \left(\sum_{k=1}^{m} Z_k \right) \mathbb{1}_{\bigcup_{i \in I_j} A_i} + g_j(\mathbf{m}_{I_j}). \end{aligned}$$

By using Theorem 7.1 and the fact that $\sum_{k=1}^{m} Z_k \geqslant 0$, we conclude that $(Y_1, \dots, Y_m)$ is pairwise counter-monotonic. □

Remark 7.3. For Theorem 7.2 (i), an equivalent statement is that increasing functions of comonotonic random variables are comonotonic. This is because one can choose the subsets as $[n]$ and take functions on $\mathbb{R}^n$ which are constant in some dimensions. We use the current presentation of statement (i) to show a contrast to statement (ii).

What we will use from Theorem 7.2 is the second statement, which leads to the next result in this section; that is, counter-monotonicity implies negative association. Since negative association is stronger than negative supermodular dependence, this result surpasses Theorem 12 of Dhaene et al. (1999), which states that counter-monotonicity is stronger than negative supermodular dependence.

Theorem 7.3. *Counter-monotonicity implies negative association.*

Proof. Let $\mathbf{X}$ be an n-dimensional counter-monotonic random vector. Take disjoint subsets $I, J \subseteq [n]$ and coordinate-wise increasing functions $f : \mathbb{R}^{|I|} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|} \to \mathbb{R}$, where $|\cdot|$ is the cardinality of a set. By Theorem 7.2 (ii), $f(\mathbf{X}_I)$ and $g(\mathbf{X}_J)$ are counter-monotonic. The Fréchet-Hoeffding inequality (see e.g., Corollary 3.28 of Rüschendorf (2013)) yields $\mathbb{E}[f(\mathbf{X}_I)g(\mathbf{X}_J)] \leqslant \mathbb{E}[f(\mathbf{X}_I)]\mathbb{E}[g(\mathbf{X}_J)]$ provided that the expectations exist. Hence, $\mathbf{X}$ is negatively associated. □

Joag-Dev and Proschan (1983, Theorem 2.11) already noted that the lottery-type random vector (7.8) in Example 7.2 is negatively associated.

The result in Theorem 7.3 has a straightforward interpretation, as counter-monotonicity is the extreme form of negative dependence, which intuitively should imply other notions of negative dependence, among which negative association is considered a strong notion; see Amini et al. (2013) for a comparison of several notions of negative dependence.

Counter-monotonicity is also stronger than several other notions of negative dependence which are not implied by negative association. These notions include conditional decreasing in sequence and negative dependence in sequence (see Joag-Dev and Proschan (1983, Remark

2.16)) and negative dependence through stochastic ordering (see Block et al. (1985)). These implications can be checked directly with Theorem 7.2, thus highlighting its usefulness.

Remark 7.4. A random vector $\mathbf{X}$ is positively associated if $\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) \geqslant 0$ for all real-valued, coordinate-wise increasing functions f, g (Esary et al. (1967)). Comonotonicity implies positive association because $(f(\mathbf{X}), g(\mathbf{X}))$ is comonotonic by Theorem 7.2, and the covariance of a comonotonic pair of random variables is non-negative due to the Fréchet-Hoeffding inequality.

7.5 Joint mix dependence and Fréchet classes

Another type of extremal negative dependence structure is the notion of joint mixes. In this section, we study the connection between counter-monotonicity and joint mix dependence.

From now on, assume that the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is atomless. A random vector $(X_1, \dots, X_n)$ is a *joint mix* if $\sum_{i=1}^n X_i$ is a constant c, and in this case we say that *joint mix dependence* holds for $(X_1, \dots, X_n)$. The constant c is called the center of $(X_1, \dots, X_n)$, and it is obvious that $c = \sum_{i=1}^n \mathbb{E}[X_i]$ if the expectations of $X_1, \dots, X_n$ are finite. Joint mix dependence is regarded as a concept of extremal negative dependence due to its opposite role to comonotonicity in risk aggregation problems; see Puccetti and Wang (2015) and Wang and Wang (2016).

The lottery-type random vector in Example 7.2 satisfies both counter-monotonicity and joint mix dependence. In the case $n = 2$, joint mix dependence is strictly stronger than counter-monotonicity. This result cannot be extended to $n \geqslant 3$. For example, $(X, X, -2X)$ is a joint mix that is not counter-monotonic. A weaker notion than joint mix dependence is proposed by Lee and Ahn (2014), which does not imply, and is not implied by, counter-monotonicity in dimension $n \geqslant 3$.

Joint mix dependence and counter-monotonicity share some similarities. First, for a random vector $(X_1, \dots, X_n)$ with its sum $S = X_1 + \cdots + X_n$, if either pairwise counter-monotonicity or joint mix dependence holds, then X_i and $S - X_i$ are counter-monotonic

for each $i \in [n]$. The case of pairwise counter-monotonicity is verified by Theorem 7.2, and the case of joint mix dependence is verified by definition. Second, both dependence notions impose strong conditions on the marginal distributions. The condition for pairwise counter-monotonicity is given in Lemma 7.2, and that for joint mix dependence is much more sophisticated; see Wang and Wang (2016) for some sufficient conditions as well as necessary ones. This is in contrast to concepts such as comonotonicity, independence, and negative association, for which the existence of the corresponding random vectors is always guaranteed for any given marginal distribution. Both joint mix dependence and counter-monotonicity are used in the tail region to obtain lower bounds for risk aggregation with given marginal distributions, as studied by Bernard et al. (2014) and Cheung et al. (2017), respectively.

The next result characterizes marginal distributions that are compatible with both counter-monotonicity and joint mix dependence. For this, we need some notation and terminology. In what follows, we will use distribution functions to represent distributions. For an n-tuple $(F_1, \dots, F_n)$ of distributions on $\mathbb{R}$, a *Fréchet class* (see Joe (1997, Chapter 3)) is defined as

$$\mathcal{F}_n(F_1, \dots, F_n) = \{\text{distribution of } (X_1, \dots, X_n) : X_i \sim F_i,\ i \in [n]\}.$$

We say that a Fréchet class $\mathcal{F}_n(F_1, \dots, F_n)$ supports counter-monotonicity (resp. joint mix dependence) if there exists a counter-monotonic random vector (resp. a joint mix) whose distribution is in this class. Let δ_x be the distribution function of a point-mass at $x \in \mathbb{R}$, and denote by Θ_n the standard n-simplex, that is, $\Theta_n = \{(p_1, \dots, p_n) \in [0,1]^n : \sum_{i=1}^n p_i = 1\}$. Two distributions F and G are symmetric if $F(x) = 1 - G(c - x)$, $x \in \mathbb{R}$ for some $c \in \mathbb{R}$. In other words, if X has distribution F, then $c - X$ has distribution G.

It turns out that all Fréchet classes $\mathcal{F}_n(F_1, \dots, F_n)$ which support both counter-monotonicity and joint mix dependence can be characterized explicitly. If at least three of $F_1, \dots, F_n$ are non-degenerate, then $F_1, \dots, F_n$ are two-point distributions given by

$$F_i = p_i \delta_{a+m_i} + (1 - p_i)\delta_{m_i} \text{ for } i \in [n], \text{ where } a, m_1, \dots, m_n \in \mathbb{R} \text{ and } (p_1, \dots, p_n) \in \Theta_n. \tag{7.10}$$

If at most two of $F_1, \dots, F_n$ are non-degenerate, then

$$F_i \text{ and } F_j \text{ are symmetric for some } i, j \in [n], \text{ and } F_k \text{ is degenerate for all } k \in [n] \setminus \{i, j\}. \tag{7.11}$$

Theorem 7.4. *A Fréchet class supports both counter-monotonicity and joint mix dependence if and only if one of* (7.10) *and* (7.11) *holds. In case both are supported, counter-monotonicity and joint mix dependence are equivalent for this Fréchet class.*

Proof. We first prove the equivalence statement in the last part of the theorem. Suppose that the Fréchet class $\mathcal{F}_n(F_1, \dots, F_n)$ supports both counter-monotonicity and joint mix dependence. Puccetti and Wang (2015, Theorem 3.8) shows that if a Fréchet class supports a counter-monotonic random vector, then a random vector is counter-monotonic if and only if it is Σ-counter-monotonic, and moreover, a joint mix is always Σ-counter-monotonic. Using these two facts, a joint mix is counter-monotonic for this Fréchet class. For the conserve statement, note that in $\mathcal{F}_n(F_1, \dots, F_n)$ there exists a unique distribution function

$$F(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} F_i(x_i) - d + 1 \right)_+, \qquad (x_1, \dots, x_n) \in \mathbb{R}^n$$

of a counter-monotonic random vector (Theorem 3.3 of Puccetti and Wang (2015)). Since a joint mix with marginal distributions $F_1, \dots, F_n$ is counter-monotonic, its distribution must coincide with F. This shows that F is the distribution of a joint mix.

Next, we prove the first part of the theorem. For the "if" statement, assume that a Fréchet class $\mathcal{F}_n(F_1, \dots, F_n)$ supports both counter-monotonicity and joint mix dependence. By the above argument, $\mathcal{F}_n(F_1, \dots, F_n)$ supports a pairwise counter-monotonic joint mix $(X_1, \dots, X_n)$. First, consider the case that at least three of $F_1, \dots, F_n$ are non-degenerate. Using (7.5),

$$X_i = (c - m)\mathbb{1}_{A_i} + m_i, \quad \text{for} \quad i \in [n],$$

where $(A_1, \dots, A_n) \in \Pi_n$, c is the center of $(X_1, \dots, X_n)$, either $m_i = \operatorname{ess\text{-}inf}(X_i)$ for all $i \in [n]$ or $m_i = \operatorname{ess\text{-}sup}(X_i)$ for all $i \in [n]$, and $m = \sum_{i=1}^{n} m_i$. It is clear that F_i has the form (7.10) by setting $a = c - m$. If at most two of $F_1, \dots, F_n$ are degenerate, say F_i and F_j, then a joint mix $(X_1, \dots, X_n)$ with marginal distributions $F_1, \dots, F_n$ satisfies $X_i = c - X_j$ for some $c \in \mathbb{R}$, and X_k is a constant for each $k \in [n] \setminus \{i, j\}$. This implies (7.11).

Finally, we verify the converse statement. If $(F_1, \dots, F_n)$ has the form (7.10), then take $X_i = a\mathbb{1}_{A_i} + m_i$ with $(A_1, \dots, A_n) \in \Pi_n$ satisfying $\mathbb{P}(A_i) = p_i$ for $i \in [n]$, and we have $(X_1, \dots, X_n)$ is counter-monotonic by Theorem 7.1 and $\sum_{i=1}^n X_i = a + \sum_{i=1}^n m_i$. If $(F_1, \dots, F_n)$ has the form (7.11), then by taking X_i with distribution F_i, $X_j = c - X_i$ with distribution F_j and $c \in \mathbb{R}$, and X_k with distribution F_k for each $k \in [n] \setminus \{i, j\}$, we can directly verify that $(X_1, \dots, X_n)$ is a counter-monotonic joint mix. □

From the proof of Theorem 7.4 (ii), if at least three components of a pairwise counter-monotonic joint mix $\mathbf{X} = (X_1, \dots, X_n)$ are non-degenerate, then it has the form

$$X_i = a\mathbb{1}_{A_i} + m_i, \quad \text{for} \quad i \in [n]$$

where $(A_1, \dots, A_n) \in \Pi_n$, $a \in \mathbb{R}$ and $\mathbf{m} = (m_1, \dots, m_n) \in \mathbb{R}^n$. If $a \neq 0$, then the random vector $(\mathbf{X} - \mathbf{m})/a$ has a categorical distribution with n categories and probability vector $(\mathbb{P}(A_1), \dots, \mathbb{P}(A_n))$.

Remark 7.5. Theorem 7.4 characterizes a Fréchet class that supports both counter-monotonicity and joint mix dependence. Fréchet classes that support (non-degenerate) pairwise counter-monotonicity are fully described by the conditions in Lemma 7.2. Whether a given Fréchet class supports joint mix dependence is a very challenging problem, with existing result summarized in Puccetti and Wang (2015) and Wang and Wang (2016). In risk aggregation problems, the notion of joint mix dependence is more relevant, because a joint mix usually "approximately exists" for large dimensions, which leads to the main idea behind the Rearrangement Algorithm; see Embrechts et al. (2013, 2014), Bernard and Vanduffel (2015) and Bernard et al. (2017). In contrast, counter-monotonicity is more relevant for risk sharing problems, which we discuss in the next section.

7.6 Optimal allocations in risk sharing for quantile agents

We now formally establish the link between counter-monotonicity and Pareto-optimal allocations in risk sharing problems for quantile agents.

We first describe the basic setting. A quantile agent assesses risk by its quantile, also known as the risk measure Value-at-Risk (VaR) in risk management. Following the convention of Embrechts et al. (2018), the VaR at level $\alpha \in (0,1)$ is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1-\alpha\}, \quad X \in \mathcal{X},$$

where $\mathcal{X}$ is the set of all random variables in the probability space. Moreover, write $\mathrm{VaR}_\alpha = -\infty$ on $\mathcal{X}$ for $\alpha \geqslant 1$, although our agents use VaR_α for $\alpha \in (0,1)$. It is important to highlight that quantile agents with level $\alpha \in (0,1)$ are not risk averse (Rothschild and Stiglitz (1970)).

We consider the risk sharing problem for $n \geqslant 3$ quantile agents with levels $\alpha_1, \dots, \alpha \in (0,1)$. For a given $S \in \mathcal{X}$, the set of *allocations* of S is

$$\mathbb{A}_n(S) = \left\{ (X_1, \dots, X_n) \in \mathcal{X}^n : \sum_{i=1}^n X_i = S \right\}.$$

An allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(S)$ is *Pareto optimal* if for any $(Y_1, \dots, Y_n) \in \mathbb{A}_n(S)$ satisfying $\mathrm{VaR}_{\alpha_i}(Y_i) \leqslant \mathrm{VaR}_{\alpha_i}(X_i)$ for all $i \in [n]$, we have $\mathrm{VaR}_{\alpha_i}(Y_i) = \mathrm{VaR}_{\alpha_i}(X_i)$ for all $i \in [n]$. Pareto optimality of $(X_1, \dots, X_n) \in \mathbb{A}_n(S)$ is equivalent to

$$\sum_{i=1}^n \mathrm{VaR}_{\alpha_i}(X_i) = \inf\left\{ \sum_{i=1}^n \mathrm{VaR}_{\alpha_i}(Y_i) : (Y_1, \dots, Y_n) \in \mathbb{A}_n(S) \right\} = \mathrm{VaR}_{\sum_{i=1}^n \alpha_i}(S), \tag{7.12}$$

where the first equality is Embrechts et al. (2018, Proposition 1) and the second equality is Embrechts et al. (2018, Corollary 2). Using (7.12), we obtain that the existence of a Pareto-optimal allocation is equivalent to $\sum_{i=1}^n \alpha_i < 1$; this is also given by Theorem 3.6 of Wang and Wu (2020). For this reason, we say that the n quantile agents are *compatible* if $\sum_{i=1}^n \alpha_i < 1$ holds, meaning that a Pareto-optimal allocation exists for some S, and equivalently, for every S.

The following theorem shows that, under some conditions of the total risk S to share, the risk sharing problem for any quantile agents admits a pairwise counter-monotonic Pareto-optimal allocation, and every pairwise counter-monotonic allocation is Pareto optimal for some agents. Moreover, comonotonic allocations are never Pareto optimal. Recall that by Lemma 7.2, a pairwise counter-monotonic random vector $(X_1, \dots, X_n)$ satisfies either (7.1) or (7.2).

Theorem 7.5. *For $S \in \mathcal{X}$, the following hold.*

(i) If S is bounded from below, then for any compatible quantile agents there exists a pairwise counter-monotonic allocation of S which is Pareto optimal.

(ii) If $\mathbb{P}(S = \text{ess-inf}S) > 0$, then every type-(7.1) pairwise counter-monotonic allocation of S is Pareto optimal for some quantile agents.

(iii) If S is continuously distributed, then a comonotonic allocation of S is never Pareto optimal for any quantile agents.

Proof. (i) Let $\alpha_1, \dots, \alpha_n \in (0,1)$ be the VaR levels of the quantile agents. Compatibility of the agents means $\sum_{i=1}^n \alpha_i < 1$. In this case, a Pareto-optimal allocation $(X_1, \dots, X_n)$ of S is given by Theorem 2 of Embrechts et al. (2018), with the form

$$X_i = (X - m)\mathbb{1}_{A_i},\ i \in [n-1] \quad \text{and} \quad X_n = (X - m)\mathbb{1}_{A_n} + m$$

for some $(A_1, \dots, A_n) \in \Pi_n$. By setting $m = \text{ess-inf}S$, $(X_1, \dots, X_n)$ is pairwise counter-monotonic by Theorem 7.1.

(ii) Note that shifting $X_1, \dots, X_n$ by arbitrary constants, and adjusting S correspondingly, does not affect its Pareto optimality due to (7.12). Moreover, (7.1) guarantees that at most one of $X_1, \dots, X_n$ is not bounded from below, and further $\mathbb{P}(S = \text{ess-inf}S) > 0$ guarantees that this can only happen if all $X_1, \dots, X_n$ are bounded from below. Therefore, we can, without loss of generality, assume $\text{ess-inf}X_i = 0$ for each $i \in [n]$.

Let $B = \{S = \text{ess-inf}S\}$ and $A = \bigcup_{i=1}^n \{X_i > 0\}$. First, if $\mathbb{P}(B \cap A) = 0$, then we let $\alpha_i = \mathbb{P}(X_i > 0) + \mathbb{P}(B)/(2n) > 0$ for $i \in [n]$. Note that

$$\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \mathbb{P}(X_i > 0) + \frac{1}{2}\mathbb{P}(B) = \mathbb{P}(A) + \frac{1}{2}\mathbb{P}(B) < \mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B) \leqslant 1.$$

It is clear that $\text{VaR}_{\alpha_i}(X_i) = 0$ for each $i \in [n]$, leading to $\sum_{i=1}^n \text{VaR}_{\alpha_i}(X_i) = 0 \leqslant \text{ess-inf}S \leqslant \text{VaR}_{\sum_{i=1}^n \alpha_i}(S)$. Note that

$$\sum_{i=1}^n \text{VaR}_{\alpha_i}(X_i) \leqslant \text{VaR}_{\sum_{i=1}^n \alpha_i}(S) \implies (X_1, \dots, X_n) \text{ is Pareto optimal.} \tag{7.13}$$

This is because Corollary 1 of Embrechts et al. (2018) gives $\sum_{i=1}^{n} \mathrm{VaR}_{\alpha_i}(X_i) \geqslant \mathrm{VaR}_{\sum_{i=1}^{n} \alpha_i}(S)$, and this leads to $\sum_{i=1}^{n} \mathrm{VaR}_{\alpha_i}(X_i) = \mathrm{VaR}_{\sum_{i=1}^{n} \alpha_i}(S)$ in (7.12), which gives Pareto optimality of $(X_1, \dots, X_n)$ as we see in part (i).

Next, assume $\mathbb{P}(B \cap A) > 0$. Then there exists $j \in [n]$ such that $\mathbb{P}(B \cap \{X_j > 0\}) > 0$. Let $\epsilon = \mathbb{P}(B \cap \{X_j > 0\})/(2n)$. Take $\alpha_i = \mathbb{P}(X_i > 0) + \epsilon > 0$ for $i \in [n] \setminus \{j\}$ and $\alpha_j = \mathbb{P}(\{X_j > 0\} \setminus B) + \epsilon$. By Lemma 7.2,

$$1 \geqslant \sum_{i=1}^{n} \mathbb{P}(X_i > 0) = \sum_{i=1}^{n} (\alpha_i - \epsilon) + \mathbb{P}(B \cap \{X_j > 0\}) = \sum_{i=1}^{n} \alpha_i + n\epsilon,$$

and hence $\sum_{i=1}^{n} \alpha_i < 1$. By definition of $\alpha_1, \dots, \alpha_n$, we have $\mathrm{VaR}_{\alpha_i}(X_i) = 0$ for $i \in [n] \setminus \{j\}$. Moreover, note that $X_j = S$ on $\{X_j > 0\}$ and

$$\mathbb{P}(\{X_j = \text{ess-inf} S\} \cap \{X_j > 0\}) = \mathbb{P}(B \cap \{X_j > 0\}) = 2n\epsilon,$$

which implies $\mathbb{P}(X_j > \text{ess-inf} S) = \mathbb{P}(X_j > 0) - 2n\epsilon < \alpha_j$. Therefore, $\mathrm{VaR}_{\alpha_j}(X_j) \leqslant \text{ess-inf} S$, leading to $\sum_{i=1}^{n} \mathrm{VaR}_{\alpha_i}(X_i) \leqslant \text{ess-inf} S \leqslant \mathrm{VaR}_{\sum_{i=1}^{n} \alpha_i}(S)$. Hence, we obtain Pareto optimality of $(X_1, \dots, X_n)$ via (7.13).

(iii) For a comonotonic allocation $(X_1, \dots, X_n)$ of S, using decreasing monotonicity of $\alpha \mapsto \mathrm{VaR}_\alpha$ and comonotonic additivity of VaR_α, we have

$$\sum_{i=1}^{n} \mathrm{VaR}_{\alpha_i}(X_i) \geqslant \sum_{i=1}^{n} \mathrm{VaR}_{\beta}(X_i) = \mathrm{VaR}_{\beta}(S),$$

where we write $\beta = \max\{\alpha_1, \dots, \alpha_n\}$. As S is continuously distributed, $\mathrm{VaR}_\alpha(S)$ is strictly decreasing in α. Noting that $\beta < \sum_{i=1}^{n} \alpha_i$, we have $\mathrm{VaR}_{\beta}(S) > \mathrm{VaR}_{\sum_{i=1}^{n} \alpha_i}(S)$. Therefore, the comonotonic allocation $(X_1, \dots, X_n)$ is not Pareto optimal by (7.12). □

Theorem 7.5 states that allocations with a pairwise counter-monotonic structure solve the problem of sharing risk among quantile agents. For instance, the lottery-type allocation in Example 7.2 is Pareto optimal for some quantile agents. Further, Theorem 7.5 (iii) states that comonotonic allocations can never be Pareto optimal for quantile agents if the total risk is continuously distributed. As mentioned, this is in stark contrast with the risk sharing problem with risk-averse agents, for which comonotonic allocations are always optimal. The latter result, due to the comonotonic improvements of Landsberger and Meilijson (1994), is

well-known; see also Jouini et al. (2008) and Rüschendorf (2013). Moreover, when all agents are strictly risk averse, only comonotonic allocations are Pareto optimal (see Proposition 8.5 of Chapter 8 for the case when preferences are modelled by strictly concave distortion functions).

As a symmetric statement to Theorem 7.5, if a random vector $(X_1, \dots, X_n)$ is pairwise counter-monotonic of type (7.2), then it is the maximizer of a risk sharing problem for some quantile agents.

Theorem 7.5 (i) assumes that S is bounded from below. This is needed because any type-(7.1) pairwise counter-monotonic allocation is bounded from below. Theorem 7.5 (ii) assumes $\mathbb{P}(S = \text{ess-inf}S) > 0$. In case $\mathbb{P}(S > \text{ess-inf}S) = 0$, a pairwise counter-monotonic allocation of type (7.1) may not be Pareto optimal for any quantile agents with levels in $(0, 1)$. A counter-example is provided in Example 7.3 below. Theorem 7.5 (iii) assumes that S is continuously distributed. This condition is also needed for the result to hold. For instance, if $S = 1$, then the allocation $(1/n, \dots, 1/n)$ is Pareto optimal for any compatible quantile agents, violating the impossibility statement on Pareto optimality.

Example 7.3. Suppose that S is uniformly distributed on $[0, 1]$, and $X_i = S\mathbb{1}_{A_i}$ with $(A_1, \dots, A_n) \in \Pi_n$ independent of S with $\mathbb{P}(A_i) > 0$ for each $i \in [n]$. We will see that the pairwise counter-monotonic allocation $(X_1, \dots, X_n)$ is not Pareto optimal for any quantile agents with levels $\alpha_1, \dots, \alpha_n \in (0, 1)$. If $\sum_{i=1}^n \alpha_i \geqslant 1$, there does not exist any Pareto-optimal allocation. If $\sum_{i=1}^n \alpha_i < 1$, then

$$\sum_{i=1}^n \mathrm{VaR}_{\alpha_i}(X_i) = \sum_{i=1}^n \left(1 - \frac{\alpha_i}{\mathbb{P}(A_i)}\right)_+ = \sum_{i=1}^n \left(\frac{\mathbb{P}(A_i) - \alpha_i}{\mathbb{P}(A_i)}\right)_+$$

and

$$\mathrm{VaR}_{\sum_{i=1}^n \alpha_i}(S) = 1 - \sum_{i=1}^n \alpha_i = \sum_{i=1}^n (\mathbb{P}(A_i) - \alpha_i) \leqslant \sum_{i=1}^n \left(\frac{\mathbb{P}(A_i) - \alpha_i}{\mathbb{P}(A_i)}\right)_+ = \sum_{i=1}^n \mathrm{VaR}_{\alpha_i}(X_i).$$

Using the condition (7.12), if $(X_1, \dots, X_n)$ is Pareto optimal, then the inequality above is an equality; this implies $\alpha_i = \mathbb{P}(A_i)$ for each $i \in [n]$. However, this further implies $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \mathbb{P}(A_i) = 1$ conflicting $\sum_{i=1}^n \alpha_i < 1$.

The next example illustrates that for the same S in Example 7.3 and compatible quantile

agents, a pairwise counter-monotonic Pareto-optimal allocation exists as implied by Theorem 7.5 (i).

Example 7.4. Let S be uniformly distributed on $[0,1]$ and $\alpha_1,\dots,\alpha_n \in (0,1)$ with $\sum_{i=1}^n \alpha_i < 1$. Take $(A_1,\dots,A_n) \in \Pi_n$ such that $\bigcup_{i=1}^{n-1} A_i = \{S \geqslant 1-\sum_{i=1}^{n-1}\alpha_i\}$ and $\mathbb{P}(A_i) = \alpha_i$ for $i \in [n-1]$. Let $X_i = S\mathbb{1}_{A_i}$ for $i \in [n]$. We can verify that $\mathrm{VaR}_{\alpha_i}(X_i) = 0$ for $i \in [n-1]$ and

$$\mathrm{VaR}_{\alpha_n}(X_n) = \mathrm{VaR}_{\alpha_n}\left(S\mathbb{1}_{\{S<1-\sum_{i=1}^{n-1}\alpha_i\}}\right) = 1-\sum_{i=1}^{n}\alpha_i = \mathrm{VaR}_{\sum_{i=1}^n \alpha_i}(S).$$

This shows that $(X_1,\dots,X_n)$ is Pareto optimal. It is also pairwise counter-monotonic by Theorem 7.1. Note that although the allocation $(X_1,\dots,X_n)$ here has the same form $(S\mathbb{1}_{A_1},\dots,S\mathbb{1}_{A_n})$ as the one in Example 7.3, the specification of $(A_1,\dots,A_n)$ is different in the two examples, leading to opposite conclusions on optimality.

Remark 7.6. One may notice that the condition on S in Theorem 7.5 part (ii) and that in part (iii), although both quite weak, are actually conflicting. This is not a coincidence, because comonotonicity and counter-monotonicity have a non-empty intersection: A random vector is both comonotonic and counter-monotonic if and only if it has at most one non-degenerate component. Therefore, we cannot have both conclusions in parts (ii) and (iii) for the same S.

Remark 7.7. As shown by Embrechts et al. (2018), the same pairwise counter-monotonic allocation which is Pareto optimal for quantile agents is also optimal for the more general Range Value-at-Risk (RVaR) agents. Therefore, the conclusion in Theorem 7.5 also applies to the RVaR agents. Another appearance of pairwise counter-monotonicity in optimal allocations is obtained by Chapter 8, where it is shown that for agents using inter-quantile differences, a Pareto-optimal allocation is the sum of two pairwise counter-monotonic random vectors. All discussions above assume homogeneous beliefs; that is, all agents use the same probability measure $\mathbb{P}$. In the setting of heterogeneous beliefs, Embrechts et al. (2020) showed that for Expected Shortfall agents, a Pareto-optimal allocation above certain constant level also has a pairwise counter-monotonic structure; see their Proposition 3. Generally, agents using the dual utility model of Yaari (1987), including the quantile-based models above, have quite different features in risk sharing and other optimization problems compared to those with

expected utility agents. For the optimal payoff of Yaari agents in portfolio choice, see Boudt et al. (2022).

Example 7.5. We illustrate that counter-monotonicity may also be the structure of an optimal allocation outside the dual utility of Yaari (1987). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be an atomless probability space, $S = 1$ and $\alpha > 0$. Consider the problem

$$\text{to maximize} \quad \sum_{i=1}^{n} \mathbb{E}\left[\alpha \mathbb{1}_{\{X_i \geqslant 1\}}\right] \quad \text{subject to } (X_1, \dots, X_n) \in \mathbb{A}_n(S) \text{ and } X_i \geqslant 0, \text{ for } i \in [n].$$

It is straightforward to verify that the set of maximizers is

$$\mathbb{A}^* = \left\{(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}) \in \mathbb{A}_n(S) : (A_1, \dots, A_n) \in \Pi_n\right\},$$

which contains only counter-monotonic allocations. This problem can be interpreted as the problem of sharing $S = 1$ among n expected utility maximizers with common utility function $u(x) = \alpha \mathbb{1}_{\{x \geqslant 1\}}$ for $\alpha > 0$. The optimization problem is thus a social planner's problem, and the set $\mathbb{A}^*$ contains all Pareto-optimal allocations for this problem. The allocations satisfying $\mathbb{P}(A_i) = \mathbb{P}(A_j)$ for every $i \neq j$ are of particular interest, as they are common in auction theory as the *random tie-breaking rule*. The variable S can be understood as an indivisible good that was auctioned, and the parameter α as the net utility of a series of n agents with quasi-linear utilities $v(X, t) = \theta X - t$ having bid the same amount $0 \leqslant t < \theta$. It is straightforward to see that these allocations are the only *fair* allocations, in the sense that all agents have the same expected utility. In other words, a fair lottery (which is counter-monotonic) is the only fair way to distribute the indivisible good among people who value it equally.

7.7 Conclusion

We provide a series of technical results on the representation (Theorem 7.1) and invariance property (Theorem 7.2) of pairwise counter-monotonicity, as well as their connection to negative association (Theorem 7.3), joint mix dependence (Theorem 7.4), and optimal allocations for quantile agents (Theorem 7.5). This chapter is motivated by the recently increasing attention in counter-monotonicity and negative dependence, and it fills the gap between the relatively scarce studies on pairwise counter-monotonicity in the literature and

the wide appearance of this dependence structure in modern applications, in particular, in risk sharing problems with agents that are not using expected utilities.

In general, studies of negative dependence and positive dependence are highly asymmetric in nature, with negative dependence being more challenging to study in various applications of risk management and statistics. In addition to the negative dependence concepts we considered in this chapter, some other notions have been studied in the recent literature, and the interested reader is referred to Amini et al. (2013), Lee and Ahn (2014), Lee et al. (2017) and Chi et al. (2022), as well as the monographs of Joe (1997, 2014).

Chapter 8

Risk sharing, measuring variability, and distortion riskmetrics

8.1 Introduction

Anne, Bob and Carole are sharing a random financial loss. After negotiating their respective expected returns, each of them prefers to minimize a statistical measure of variability of their allocated risk. While agreeing on the distribution of the total loss, and that the variance is a poor metric of riskiness, each of them has their own favourite tool for measuring risks. Anne, as an economics student, likes the Gini deviation (GD) because of its intuitive appearance as an economic index. Bob, as a computer science student, prefers the mean-median deviation (MMD) because it minimizes the mean absolute error. Finally, Carole, as a statistics student, finds that an inter-quantile difference (IQD) is the most representative of her preference, as she does not worry about extreme events for this particular risk. How should Anne, Bob and Carole optimally share risks among themselves?

The reader familiar with risk sharing problems may immediately realize two notable features of such a problem. First, the preferences are not monotone, different from standard decision models in the literature. Second, and most crucially, Carole's preference is neither convex nor consistent with second-order stochastic dominance. This alludes to the possibility of non-comonotonic Pareto-optimal allocations, in contrast to the comonotonic ones, which

are well studied in the literature (e.g., Landsberger and Meilijson, 1994; Jouini et al., 2008; Carlier et al., 2012; Rüschendorf, 2013).

The GD, the MMD and the IQD are measures of distributional variability. Variability is used to characterize the concept of risk as in the classic work of Markowitz (1952) and Rothschild and Stiglitz (1970). For this reason, we also call them *riskmetrics*, which include also *risk measures* in the literature, often associated with monotonicity (e.g., Föllmer and Schied, 2016). As the most popular measure of variability, the variance is known to be a coarse metric; Embrechts et al. (2002) discussed various flaws of using variance and correlation in financial risk management. Anne's decision criterion has been proposed in Shalit and Yitzhaki (1984), which considers an optimal portfolio problem *à la* Markowitz (1952), but with the variance replaced by the GD.[1] Formally, the authors analyze the investor's problem $\min_X \mathrm{GD}(X)$ subject to $\mathbb{E}[X] \geqslant R$, for a given rate $R \geqslant 0$ of return proportional to the investor's risk aversion. As with mean-variance preferences (e.g., Markowitz, 2014; Maccheroni et al., 2013), the decision criterion can thus be viewed as the problem of maximizing $\mathbb{E}[X] - \eta \mathrm{GD}(X)$, for $\eta \geqslant 0$ being the Lagrange multiplier of the problem. While the decision criterion $\mathbb{E}[X] - \eta \mathrm{GD}(X)$ seems natural, it is not monotone unless η is less than or equal to one, in which case the investor's preference belongs to those of Yaari (1987). The other measures MMD and IQD also have sound foundations and long history in statistics and its applications (Yule, 1911, Chapter 6). Slightly different from MMD, Konno and Yamazak (1991) studied portfolio optimization using the mean-absolute deviation from the mean. Risk sharing problems with convex risk measures are well studied (e.g., Barrieu and El Karoui, 2005, Jouini et al., 2008 and Filipović and Svindland, 2008), but the classes of riskmetrics mentioned above do not belong to convex risk measures in general.

In this chapter, we address the problem of sharing risk among agents that uses *distortion riskmetrics* as their preferences. Distortion riskmetrics are evaluation functionals that are characterized by comonotonic additivity and law invariance (Wang et al., 2020a). This rich family includes many measures of risk and variability, and in particular, the mean, the GD, the MMD, the IQD, and their linear combinations. Distortion riskmetrics are closely related to Choquet integrals and rank-dependent utilities widely used in decision theory (e.g., Yaari,

[1] The authors use the term *Gini's mean difference*.

1987; Schmeidler, 1989; Carlier and Dana, 2003); for a comprehensive treatment of distortions in decisions and economics, see Wakker (2010). The combination of the mean and GD or that of the mean and MMD, as well as other distortion riskmetrics, are used as premium principles in the insurance literature; see Denneberg (1990). Several variability measures within the class of distortion riskmetrics are studied by Grechuk et al. (2009), Furman et al. (2017) and Bellini et al. (2022).

While we analyze the general problem of sharing risk amongst distortion riskmetrics agents, non-monotone and non-cash-additive evaluation functionals receive greater attention. This is for a few reasons. First, the special case of sharing risk with cash-additive and law-invariant functional is well studied, and more so when the functionals are monotone, but the general case is less understood. Second, the formalism we introduce allows us to generalize the example above and consider individuals that analyze their risks with different variability measures. This is critical because we aim to understand how the act of measuring risk differently gives rise to incentives to trade it. Third, technically, relaxing monotonicity and convexity allows us to deal with maximization and minimization problems of risk in a unified framework.

The following simple example, by considering the GD and MMD agent only, illustrates the structure of a Pareto-optimal allocation as an insurance contract.

Example 8.1. Consider the problem of sharing a random loss X between Anne (A) and Bob (B) only. Recall that Bob evaluates its allocation X_B using the mean-median deviation $\mathrm{MMD}(X_B)$. Similarly, Anne's allocation is X_A which she evaluates with the Gini deviation $\mathrm{GD}(X_A)$. We will show (in Section 8.6) that any Pareto-optimal allocation takes the form

$$X_A = X \wedge \ell - X \wedge d, \quad X_B = X - X_A,$$

where $\ell \geqslant d$ will be specified later. We can interpret this as a situation where X is Bob's potential loss and Anne provides some degree of insurance for Bob. The contract (transfer function) is thus simply the random variable X_A. Notice that (i) when $\ell \geqslant X \geqslant d = 0$ there is full insurance, (ii) when $\ell = d$ there is no insurance and (iii) for other choices of $\ell > d$, the contract is a simple deductible d with an upper limit ℓ. Further, we show that each Pareto-optimal allocation minimizes $\lambda \mathrm{MMD}(X_B) + (1-\lambda)\mathrm{GD}(X_A)$ for some $\lambda \in [0,1]$.

The previous example is interesting because it confirms the intuition that the act of measuring risk differently leads to incentives to trade it. Yet, the "shape" of the solution above is not surprising, as both the Gini deviation and mean-median deviation are convex order consistent functionals, and so exhibit risk aversion of Rothschild and Stiglitz (1970). Just as in the increasing distortion case, risk-minimizing (utility-maximizing) Pareto-optimal allocations are comonotonic when the distortion riskmetrics' distortion function is concave (convex), because concavity of the distortion function is equivalent to convex order consistency.

The situation for IQD agents like Carole is more sophisticated. The distortion function of IQD is discontinuous, non-concave, non-monotone, and takes value zero on both tails of the distribution. The preference induced by IQD does not correspond to decision criterion typically considered in the literature, whereas the preferences induced by quantiles, called quantile maximization, have been axiomatized by Rostek (2010). IQD is a standard measure of dispersion used in statistics such as in box plots, and its most popular special case in statistics is the inter-quartile difference, which measures the difference between the 25% and 75% quantiles of data.

The general theory of risk sharing between agents using distortion riskmetrics is laid out in Section 8.3. A convenient feature of distortion riskmetrics is that they are convex order consistent if and only if the distortion function is concave (Wang et al., 2020a, Theorem 3). This enables the characterization of Pareto-optimal allocations for such agents using the comonotonic improvement, a notion introduced in Landsberger and Meilijson (1994) to characterize the optimal sharing of risk among risk-averse expected utility maximizers; see also Ludkovski and Rüschendorf (2008) and Rüschendorf (2013). Non-concave distortion functions lead to substantial challenges and to non-comonotonic optimal allocations, with limited recent results obtained by Embrechts et al. (2018) and Weber (2018) for some increasing distortion riskmetrics.

We study optimality within the subset of comonotonic allocations, which we refer to as the comonotonic risk sharing problem, for general distortion riskmetrics which are not necessarily convex in Section 8.4. We show that the utility possibility frontier of distortion riskmetrics is always a convex set when restricted to the subset of comonotonic allocations.

By the Hanh-Banach Theorem, we can always find comonotonic Pareto-optimal allocations by optimizing a linear combination of the agents' welfare. This simple but valuable result "essentially comes for free" by the comonotonic additivity and positive homogeneity of distortion riskmetrics. In particular, it does not require the convexity of the evaluation functionals. Moreover, this comonotonic setting allows us to easily incorporate heterogeneous beliefs as in the setting of Embrechts et al. (2020), which we study in Section 8.11 for the interested reader.

With IQD agents, the set of optimal allocations can dramatically differ when defined on the whole set of allocations or the subset of comonotonic ones, as shown by results in Sections 8.3.2 and 8.4.2. We show the surprising result that Pareto-optimal allocations are precisely those which solve a sum optimality problem, which is not true for other variability measures such as GD or MMD. Closed-form Pareto-optimal allocations are obtained, which can be decomposed as the sum of two pairwise counter-monotonic allocations. This observation complements the optimal allocations for quantile agents obtained by Embrechts et al. (2018) which are pairwise counter-monotonic.

Combining results obtained in Sections 8.3 and 8.4, the general problem of sharing risks between IQD agents (like Carole) and agents with concave and symmetric distortion functions (like Anne and Bob) mentioned in the beginning of the chapter is solved in Section 8.5 and further illustrated in Section 8.6. We obtain a sum-optimal allocation which features a combination of comonotonicity and pairwise counter-monotonicity. These two structures are, respectively, regarded as extremal positive and negative dependence concepts; see Puccetti and Wang (2015) for an overview of these dependence concepts and Chapter 7 for a stochastic representation of pairwise counter-monotonicity. More specifically, there exists an event on which the obtained Pareto-optimal allocation is comonotonic, and two events on which the sum-optimal allocation is pairwise counter-monotonic. To the best of our knowledge, this is the first article to obtain such a type of sum-optimal or Pareto-optimal allocation. Moreover, none of our results relies on continuity of the distortion functions. We conclude the chapter in Section 8.7 with a few remarks, and all proofs are put in Sections 8.3 - 8.12.

8.2 Preliminaries

8.2.1 Distortion riskmetrics

For a measurable space $(\Omega, \mathcal{F})$ and a finite set function $\nu : \mathcal{F} \to \mathbb{R}$ with $\nu(\varnothing) = 0$, the signed *Choquet integral* of a random variable $X : \Omega \to \mathbb{R}$ is the integral

$$\int X \, \mathrm{d}\nu = \int_0^\infty \nu(X > x) \, \mathrm{d}x + \int_{-\infty}^0 (\nu(X > x) - \nu(\Omega)) \, \mathrm{d}x, \tag{8.1}$$

provided these integrals are finite. Let n be a positive integer and let $[n] = \{1, \dots, n\}$. The random variables $X_1, \dots, X_n$ are *comonotonic* if there exists a collection of increasing functions $f_i : \mathbb{R} \to \mathbb{R}$, $i \in [n]$, and a random variable Z such that $X_i = f_i(Z)$ for all $i \in [n]$. Two random variables X_1, X_2 are *counter-monotonic* if $X_1, -X_2$ are comonotonic. The random variables $X_1, \dots, X_n$ are *pairwise counter-monotonic* if X_i, X_j are counter-monotonic for each pair of distinct i, j (see Puccetti and Wang (2015) and Chapter 7). Terms like "increasing" or "decreasing" are in the non-strict sense.

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is an atomless probability space where almost surely equal random variables are treated as identical. Let $\mathcal{X}$ be a set of random variables on this space. For simplicity, we assume throughout that $\mathcal{X} = L^\infty$, the set of essentially bounded random variables, and we will inform the reader when a result can be extended to larger spaces. A *distortion riskmetric* ρ_h is the mapping from $\mathcal{X}$ to $\mathbb{R}$,

$$\rho_h(X) = \int X \, \mathrm{d}\,(h \circ \mathbb{P}) = \int_0^\infty h(\mathbb{P}(X > x)) \, \mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}(X > x)) - h(1)) \, \mathrm{d}x, \tag{8.2}$$

where h is in the set $\mathcal{H}^{\mathrm{BV}}$ of all possibly non-monotone *distortion functions*, i.e.,

$$\mathcal{H}^{\mathrm{BV}} = \{h : [0,1] \to \mathbb{R} \mid h \text{ is of bounded variation and } h(0) = 0\}.$$

We now recall some properties of distortion riskmetrics that we use throughout. Any distortion riskmetric ρ_h always satisfies the following four properties as a function $\rho : \mathcal{X} \to \mathbb{R}$.

1. *Law invariance:* $\rho(X) = \rho(Y)$ for $X \overset{\mathrm{d}}{=} Y$.

2. *Positive homogeneity:* $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda > 0$ and $X \in \mathcal{X}$.

3. *Comonotonic additivity:* $\rho(X+Y) = \rho(X)+\rho(Y)$ whenever X and Y are comonotonic.

4. *Translation invariance:* $\rho(X + c) = \rho(X) + c\rho(1)$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

As a special case of translation invariance with $\rho(1) = 1$, ρ is *cash additive* if $\rho(X + c) = \rho(X)+c$ for $x \in \mathbb{R}$ and $X \in \mathcal{X}$. For a distortion riskmetric ρ_h, cash additivity means $h(1) = 1$. We also say *location invariance* for $h(1) = 0$ and *reverse cash additivity* for $h(1) = -1$. We note that although we use the general term "cash additivity" as in the literature of risk measures, the values of random variables may be interpreted as non-monetary, such as carbon dioxide emissions, as long as they can be transferred between agents in an additive fashion.

A distortion riskmetric ρ_h may also satisfy the following properties depending on h. A random variable X is said to be smaller than a random variable Y in the *convex order*, denoted by $X \leqslant_{\mathrm{cx}} Y$, if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for every convex function $\phi : \mathbb{R} \to \mathbb{R}$, provided that both expectations exist.

5. *Increasing monotonicity*: $\rho(X) \leqslant \rho(Y)$ whenever $X \leqslant Y$.

6. *Convex order consistency*: $\rho(X) \leqslant \rho(Y)$ whenever $X \leqslant_{\mathrm{cx}} Y$.

7. *Subadditivity*: $\rho(X + Y) \leqslant \rho(X) + \rho(Y)$ for every $X, Y \in \mathcal{X}$.

We also say that ρ is *monotone* if either ρ or $-\rho$ is increasing. Increasing and cash-additive functionals are called *monetary risk measure* (Föllmer and Schied, 2016) or *niveloids* (Cerreia-Vioglio et al., 2014). For a distortion riskmetric ρ_h, increasing monotonicity means that h is increasing, and either subadditivity or convex order consistency is equivalent to the concavity of h by Theorem 3 of Wang et al. (2020a).

Distortion riskmetrics are precisely all law-invariant and comonotonic-additive mappings satisfying a form of continuity; see Wang et al. (2020b) on L^∞ and Wang et al. (2020a) on general spaces. The subset of increasing normalized distortion functions is denoted by $\mathcal{H}^{\mathrm{DT}}$, that is,

$$\mathcal{H}^{\mathrm{DT}} = \{h : [0, 1] \to \mathbb{R} \mid h \text{ is increasing, } h(0) = 0 \text{ and } h(1) = 1\}.$$

If $h \in \mathcal{H}^{\mathrm{DT}}$, then ρ_h is called a *dual utility* of Yaari (1987). Recall that a Yaari agent is strongly risk averse when the distortion function h is concave (Yaari, 1987). Hence, we slightly abuse nomenclature and simply say that a distortion riskmetric agent is risk averse when its distortion function is concave, regardless of whether it is increasing or not. This is consistent with the concept of increasing in risk introduced by Rothschild and Stiglitz (1970).

Any distortion riskmetric admits a quantile representation (Lemma 1 of Wang et al. (2020a)). For a concise presentation of results, we define quantiles by counting losses *from large to small*.[2] Formally, we define the left quantile of a random variable $X \in \mathcal{X}$ as $Q_t^-(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1-t\}$, and the right quantile as $Q_t^+(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) > 1-t\}$, where $\inf \varnothing = \infty$, ess-sup $= Q_0^-$ and ess-inf $= Q_1^+$ by convention.

Lemma 8.1. *For $h \in \mathcal{H}^{\mathrm{BV}}$ and $X \in \mathcal{X}$ such that $\rho_h(X)$ is well-defined (it may take values $\pm\infty$),*

(i) if h is right-continuous, then $\int X \,\mathrm{d}(h \circ \mathbb{P}) = \int_0^1 Q_t^+(X)\,\mathrm{d}h(t)$;

(ii) if h is left-continuous, then $\int X \,\mathrm{d}(h \circ \mathbb{P}) = \int_0^1 Q_t^-(X)\,\mathrm{d}h(t)$;

(iii) if $Q_t^-(X)$ is continuous on $(0,1)$, then $\int X \,\mathrm{d}(h\circ\mathbb{P}) = \int_0^1 Q_t^-(X)\,\mathrm{d}h(t) = \int_0^1 Q_t^+(X)\,\mathrm{d}h(t)$.

There are two main advantages of working with non-monotone distortion functions. First, as monotonicity is not assumed, results on maxima and minima are symmetric; we only need to analyze one of them. Second, distortion riskmetrics include many more functionals in risk management, such as variability measures, which never have a monotone distortion function. We will make extensive use of three variability measures which appeared in the introduction. They are well defined on spaces larger than L^∞, although we state our main results on $\mathcal{X} = L^\infty$.

The first measure of variability we use extensively is the Gini deviation (GD)

$$\mathrm{GD}(X) = \frac{1}{2}\mathbb{E}[|X^* - X^{**}|] = \int X \,\mathrm{d}(h_{\mathrm{GD}} \circ \mathbb{P})$$

[2]It will be clear from Theorem 8.2 that this nontraditional choice of notation significantly simplifies the presentation of several results; this is also the case in Embrechts et al. (2018).

for $X \in L^1$, $h_{\mathrm{GD}}(t) = t - t^2$, $t \in [0,1]$ and X^*, X^{**} independent copies of X. Its distortion function is depicted in Figure 8.1 (a). As our second measure of variability, the mean-median deviation (MMD) is defined by

$$\mathrm{MMD}(X) = \min_{x \in \mathbb{R}} \mathbb{E}[|X - x|] = \mathbb{E}[|X - Q^-_{1/2}(X)|] = \int X \, \mathrm{d}(h_{\mathrm{MMD}} \circ \mathbb{P})$$

for $X \in L^1$ and $h_{\mathrm{MMD}}(t) = t \wedge (1-t)$, $t \in [0,1]$; see Figure 8.1 (b). The mean-median deviation is sometimes called the mean (or average) absolute deviation from the median and is well known for its statistical robustness. Both the mean-median deviation and the Gini deviation have concave distortions and thus are convex order consistent. Lastly, the inter-quantile difference (IQD) is defined by

$$\mathrm{IQD}_\alpha(X) = Q^-_\alpha(X) - Q^+_{1-\alpha}(X) = \int X \, \mathrm{d}(h_{\mathrm{IQD}} \circ \mathbb{P})$$

for $X \in L^0$ and $h_{\mathrm{IQD}}(t) = \mathbb{1}_{\{\alpha < t < 1-\alpha\}}$, $t \in [0,1]$ and $\alpha \in [0, 1/2)$. See Figure 8.1 (c) for its distortion function. We further set $\mathrm{IQD}_\alpha = 0$ for $\alpha \in [1/2, \infty)$, but this is only for the purpose of unifying the presentation of some results. Our formulation of IQD is slightly different from the definition used by Bellini et al. (2022) where IQD_α is defined as $Q^+_\alpha - Q^-_{1-\alpha}$, but this difference is minor. For $X \in \mathcal{X}$ and $\alpha \in [0, 1/2)$, a convenient formula (see Theorem 1 of Bellini et al. (2022)) is

$$\mathrm{IQD}_\alpha(X) = Q^-_\alpha(X) + Q^-_\alpha(-X), \tag{8.3}$$

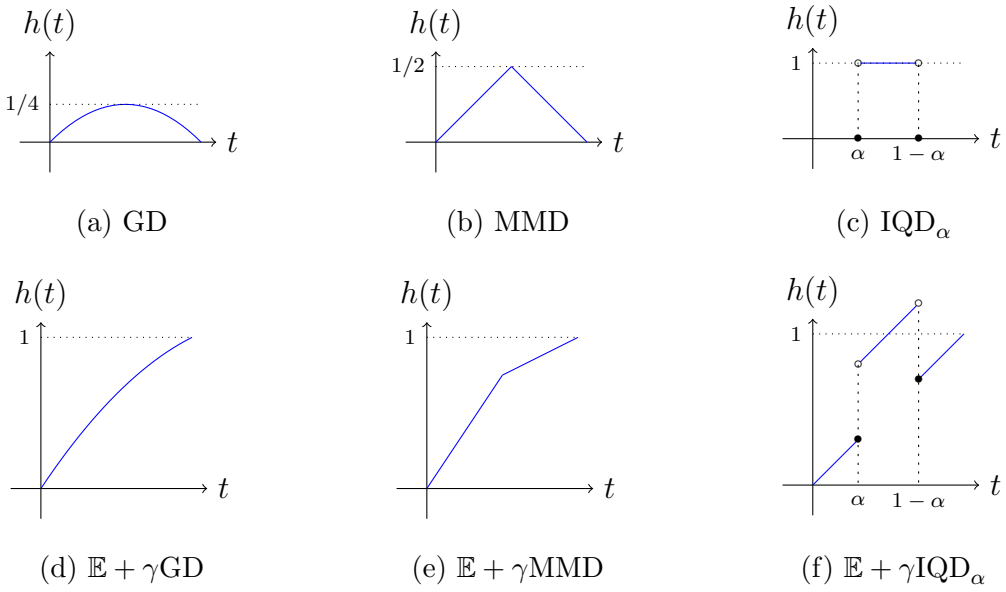
and this is due to $Q^+_{1-\alpha}(X) = -Q^-_\alpha(-X)$.

Consider now a preference functional $\mathcal{I}$ of the form

$$\mathcal{I}(X) = \theta \mathbb{E}(X) + \gamma D(X)$$

for $\theta \geqslant 0$, $\gamma \in \mathbb{R}$ and $D(X)$ a variability measure. The version of $\mathcal{I}$ with $\theta = 1$ and $\gamma < 0$ is widely used in modern portfolio theory (as an objective to maximize). There, the random variable X denotes the gains, the parameter γ indicates the degree of risk aversion and $D(X)$ is a variability measure chosen to replace the variance. This yields the "Mean-D" preferences nomenclature common in the literature. The version of $\mathcal{I}$ with X being a loss, $\theta \geqslant 1$ and $\gamma \geqslant 0$ is common in the insurance/reinsurance literature, where it is called a distortion-deviation

Figure 8.1: Distortion functions for GD, MMD, IQD and $\mathbb{E} + \gamma D$, where $\gamma = 1/2$



(a) GD (b) MMD (c) IQD_α

(d) $\mathbb{E} + \gamma\text{GD}$ (e) $\mathbb{E} + \gamma\text{MMD}$ (f) $\mathbb{E} + \gamma\text{IQD}_\alpha$

premium principle. For instance, Denneberg (1990) suggests the premium principle $\theta = 1$ and $D(X) = \text{MMD}(X)$. The functional $\mathcal{I}$ is a distortion riskmetric as long as D is one, and so we adopt the convention of denoting such functional by ρ_h and interpreting X as losses. Panels (d)-(f) of Figure 8.1 illustrate the distortion functions of $\mathbb{E} + \gamma D$.

8.2.2 Risk sharing problems

There are n agents sharing a total loss $X \in \mathcal{X}$. Suppose that agent $i \in [n]$ has a preference modelled by a distortion riskmetric ρ_{h_i} with smaller values preferred. Given $X \in \mathcal{X}$, we define the set of *allocations* of X as

$$\mathbb{A}_n(X) = \left\{ (X_1, \dots, X_n) \in \mathcal{X}^n : \sum_{i=1}^n X_i = X \right\}. \tag{8.4}$$

The *inf-convolution* $\square_{i=1}^n \rho_{h_i}$ of n distortion riskmetrics $\rho_{h_1}, \dots, \rho_{h_n}$ is defined as

$$\square_{i=1}^n \rho_{h_i}(X) := \inf \left\{ \sum_{i=1}^n \rho_{h_i}(X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n(X) \right\}, \quad X \in \mathcal{X}.$$

That is, the inf-convolution of n distortion riskmetrics is the infimum over aggregate welfare for all possible allocations. For a general treatment of inf-convolution in risk sharing problems, see Rüschendorf (2013).

Let $\rho_{h_1}, \dots, \rho_{h_n}$ be distortion riskmetrics and $X \in \mathcal{X}$. The allocation $(X_1, \dots, X_n)$ is *sum optimal in* $\mathbb{A}_n(X)$ if $\square_{i=1}^n \rho_{h_i}(X) = \sum_{i=1}^n \rho_{h_i}(X_i)$. An allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is *Pareto optimal in* $\mathbb{A}_n(X)$ if for any $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ satisfying $\rho_{h_i}(Y_i) \leqslant \rho_{h_i}(X_i)$ for all $i \in [n]$, we have $\rho_{h_i}(Y_i) = \rho_{h_i}(X_i)$ for all $i \in [n]$.

Part of our analysis is conducted for the constrained problem where the allocations are confined to the set of comonotonic allocations, that is,

$$\mathbb{A}_n^+(X) = \{(X_1, \dots, X_n) \in \mathbb{A}_n(X) : \ X_1, \dots, X_n \text{ are comonotonic}\},$$

We first make a useful observation about $\mathbb{A}_n^+(X)$ below, which is a generalization of Denneberg (1994, Proposition 4.5) in the case of $n = 2$.

Proposition 8.1. *The random variables $X_1, \dots, X_n$ are comonotonic if and only if there exist increasing functions $f_i : \mathbb{R} \to \mathbb{R}$ such that $X_i = f_i(\sum_{i=1}^n X_i)$, $i \in [n]$ and $\sum_{i=1}^n f_i(x) = x$ for $x \in \mathbb{R}$.*

Proof. We first show that the statement holds for $n = 3$. Assume that X_1, X_2 and X_3 are comonotonic. We have that $X_1 + X_2$ and X_3 are comonotonic by the definition of comonotonicity. By Denneberg (1994, Proposition 4.5), there exist increasing functions u and v such that

$$X_1 + X_2 = u\,(X_1 + X_2 + X_3)\,, \quad X_3 = v\,(X_1 + X_2 + X_3)$$

and $u(x) + v(x) = x$ for all $x \in \mathbb{R}$. Since X_1 and X_2 are comonotonic, there exist increasing functions g_1 and g_2 such that $X_1 = g_1(X_1 + X_2)$, $X_2 = g_2(X_1 + X_2)$ and $g_1(x) + g_2(x) = x$ for $x \in \mathbb{R}$. Let $f_1 = g_1 \circ u$, $f_2 = g_2 \circ u$ and $f_3 = v$. It is clear that $X_1 = f_1(X_1 + X_2 + X_3)$, $X_2 = f_2(X_1 + X_2 + X_3)$, $X_3 = f_3(X_1 + X_2 + X_3)$ and $f_1(x) + f_2(x) + f_3(x) = x$ for all $x \in \mathbb{R}$. By repeating the above argument, we can show the "only if" part for $n \geqslant 3$. The "if" part follows directly from the definition of comonotonicity. □

Proposition 8.1 implies that if $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$, then we can set $X = Z$ in the definition of comonotonicity while guaranteeing that for every $\omega \in \Omega$ it is $\sum_{i=1}^n X_i(\omega) = X(\omega)$.

The *comonotonic inf-convolution* $\boxplus_{i=1}^{n} \rho_{h_i}$ of distortion riskmetrics $\rho_{h_1}, \dots, \rho_{h_n}$ is defined as

$$\boxplus_{i=1}^{n} \rho_{h_i}(X) := \inf \left\{ \sum_{i=1}^{n} \rho_{h_i}(X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n^+(X) \right\}.$$

Let $\rho_{h_1}, \dots, \rho_{h_n}$ be distortion riskmetrics and $X \in \mathcal{X}$. An allocation $(X_1, \dots, X_n)$ is *sum optimal in* $\mathbb{A}_n^+(X)$ when $\boxplus_{i=1}^{n} \rho_{h_i}(X) = \sum_{i=1}^{n} \rho_{h_i}(X_i)$. An allocation $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ is *Pareto optimal in* $\mathbb{A}_n^+(X)$ if for any $(Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X)$ satisfying $\rho_{h_i}(Y_i) \leqslant \rho_{h_i}(X_i)$ for all $i \in [n]$, we have $\rho_{h_i}(Y_i) = \rho_{h_i}(X_i)$ for all $i \in [n]$.

The set of comonotonic allocations $\mathbb{A}_n^+(X)$ is a strict subset of the set of all possible allocations $\mathbb{A}_n(X)$. Hence, the sequel refers to the problem of sharing risk in $\mathbb{A}_n(X)$ and $\mathbb{A}_n^+(X)$ as *unconstrained* and *comonotonic* risk sharing, respectively.

8.3 Unconstrained risk sharing

This section tackles the unconstrained risk sharing problem. It is divided into two subsections. The first aims at providing general results and the second subsection characterizes the unconstrained risk sharing problem with IQD agents. There, we show that sum-optimal allocations involve pairwise counter-monotonicity, an extreme form of negative dependence between the agents' risk.

8.3.1 Pareto optimality, sum optimality, and comonotonic improvement

In all results, we will always assume that agents have preferences modelled by $\rho_{h_1}, \dots, \rho_{h_n}$ where $h_1, \dots, h_n \in \mathcal{H}^{\mathrm{BV}}$, with one exception which will be specified clearly. The value of $h(1)$ is important for a distortion riskmetric ρ_h because, by translation invariance, it pins down the value attributed to a sure gain or loss.

Proposition 8.2. *Let $X \in \mathcal{X}$. Then*

(i) If a Pareto-optimal allocation in either $\mathbb{A}_n^+(X)$ *or* $\mathbb{A}_n(X)$ *exists then* $h_i(1)$, $i \in [n]$, *are all* 0, *all positive, or all negative;*

(ii) If $\boxplus_{i=1}^n \rho_{h_i}(X) > -\infty$, *then* $h_1(1) = \cdots = h_n(1)$.

The proof of Proposition 8.2 highlights the role of translation invariance. Notice that since distortion riskmetrics are positively homogeneous, the value of $h(1)$ can be interpreted as a constant marginal utility of money. For (i), we thus assume by contradiction that $(X_1, \dots, X_n)$ is Pareto optimal but that $h_i(1)$, $i \in [n]$, are not all zero or all of the same sign. We can organize a (cash) transfer $c_1, \dots, c_n$ between agents such that $\sum_{i=1}^n c_i = 0$ and the allocation $(X_1 + c_1, \dots, X_n + c_n)$ strictly improves upon $(X_1, \dots, X_n)$, an absurd. This condition implies that, in order for the risk sharing problem to be meaningful, all agents must agree on whether they like or dislike an increase of their allocation. In the former case, $X_1, \dots, X_n$ may represent a good like monetary gains, and in the latter case, they may represent bad, like carbon dioxide emissions. For (ii), when the value of $h(1)$ differs between agents, a similar type of transfer strictly reduces the sum of welfare $\sum_{i=1}^n \rho_{h_i}$, and so the value attained by the inf-convolution $\boxplus_{i=1}^n \rho_{h_i}$ is arbitrarily small.

For $h \in \mathcal{H}^{\mathrm{BV}}$, we write $\tilde{h} = h/|h(1)|$ if $h(1) \neq 0$ and $\tilde{h} = h$ if $h(1) = 0$. If $h(1) \neq 0$, then $\tilde{h}(1) = \pm 1$. Note that replacing h_i with its normalized version $\tilde{h}_i$ does not change the preference of agent i. Hence, we sometimes consider in our proofs distortion riskmetrics that are either all cash additive or all reverse cash additive. While this normalization does change the value attained by the inf-convolution, it is without loss of generality for characterizing Pareto optimality.

We now state our first theorem, a generalization of Proposition 1 of Embrechts et al. (2018) stated for monetary risk measures.

Theorem 8.1. *Suppose that* $h_i(1) \neq 0$ *for some* $i \in [n]$. *An allocation* $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ *is Pareto optimal in* $\mathbb{A}_n(X)$ *if and only if* $\sum_{i=1}^n \rho_{\tilde{h}_i}(X_i) = \square_{i=1}^n \rho_{\tilde{h}_i}(X)$.

Theorem 8.1 states that Pareto optimality and sum optimality can be translated into each other via normalization whenever the distortion riskmetrics are not location invariant. The picture for location-invariant distortion riskmetrics is, however, drastically different,

as we only have one direction. The next statement considers this setting. Its proof is straightforward and thus omitted.

Proposition 8.3. *Suppose that $h_i(1) = 0$ for all $i \in [n]$. For an allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$, it holds that (i)⇒(ii):*

(i) $\sum_{i=1}^n \lambda_i \rho_{h_i}(X_i) = \square_{i=1}^n (\lambda_i \rho_{h_i})(X)$ *for some* $(\lambda_1, \dots, \lambda_n) \in (0, \infty)^n$;

(ii) $(X_1, \dots, X_n)$ *is Pareto optimal in* $\mathbb{A}_n(X)$.

One might be interested in the converse statement of Proposition 8.3, asking whether the Pareto optimality of $(X_1, \dots, X_n)$ implies the existence of a set of $(\lambda_1, \dots, \lambda_n) \in [0, \infty)^n \setminus \{\mathbf{0}\}$ such that $\sum_{i=1}^n \lambda_i \rho_{h_i}(X_i) = \square_{i=1}^n (\lambda_i \rho_{h_i})(X)$. We see in this chapter that this claim holds in three cases: when agents have $h_i(1) > 0$ or $h_i(1) < 0$ (Theorem 8.1); when all agents are IQD (Theorem 8.2); when they have concave distortion functions (a combination of Propositions 8.4 and 8.7). However, we do not know whether this holds true for general distortion functions with $h_1(1) = \cdots = h_n(1) = 0$; see also the discussion after Proposition 8.7.

In view of Proposition 8.3, we say that an allocation $(X_1, \dots, X_n)$ of X is $\boldsymbol{\lambda}$-optimal if

$$\square_{i=1}^n \rho_{\lambda_i h_i}(X) = \sum_{i=1}^n \rho_{\lambda_i h_i}(X_i). \tag{8.5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$. Clearly, $\boldsymbol{\lambda}$-optimality is equivalent to sum optimality when the preferences are specified as $(\lambda_1 \rho_{h_1}, \dots, \lambda_n \rho_{h_n})$, and conversely, sum optimality is $(1, \dots, 1)$-optimality. Therefore, we encounter no additional technical complications when solving either of them.

The following result follows from the well-known result of comonotonic improvement of Landsberger and Meilijson (1994)[3] and the fact that distortion riskmetrics are convex order consistent when the distortion functions h_i are concave (Theorem 3 of Wang et al. (2020a)). A *comonotonic improvement* of $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is a random vector $(Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X)$ such that $Y_i \leqslant_{\mathrm{cx}} X_i$ for all $i \in [n]$. Such a comonotonic improvement always exists for any $(X_1, \dots, X_n)$.

[3]See Rüschendorf (2013) for this result on general spaces.

Proposition 8.4. *Suppose that $h_1, \dots, h_n$ are concave. It holds that $\square_{i=1}^n \rho_{h_i} = \boxplus_{i=1}^n \rho_{h_i}$. Moreover, for any $X \in \mathcal{X}$, if there exists a Pareto-optimal allocation in $\mathbb{A}_n(X)$, then there exists a comonotonic Pareto-optimal allocation in $\mathbb{A}_n(X)$.*

Next, we prove that if $h_1, \dots, h_n$ are strictly concave, then the set of optimal allocations in $\mathbb{A}_n(X)$ is exactly that of those in $\mathbb{A}_n^+(X)$. This is because comonotonic improvements lead to a strict increase in welfare when the probability distortions h_i are strictly concave. We state this result formally in Corollary 8.1 as a consequence of the following ancillary lemma:

Lemma 8.2. *For two random variables $X, Y \in \mathcal{X}$, the following are equivalent:*

(i) $X \overset{\mathrm{d}}{=} Y$;

(ii) $\rho_h(X) = \rho_h(Y)$ *for all concave* $h \in \mathcal{H}^{\mathrm{BV}}$;

(iii) $\rho_h(X) \leqslant \rho_h(Y)$ *for all concave* $h \in \mathcal{H}^{\mathrm{BV}}$, *in which the equality holds for a strictly concave* h;

(iv) $X \leqslant_{\mathrm{cx}} Y$ *and* $\rho_h(X) = \rho_h(Y)$ *for a strictly concave* $h \in \mathcal{H}^{\mathrm{BV}}$.

Corollary 8.1. *If $X \leqslant_{\mathrm{cx}} Y$ and $X \overset{\mathrm{d}}{\neq} Y$, then $\rho_h(X) < \rho_h(Y)$ for any strictly concave h.*

Remark 8.1. The equivalence in Lemma 8.2 holds true for any random variables X, Y with finite mean, by requiring that $\rho_h(X)$ and $\rho_h(Y)$ are finite for the strictly concave function h in (iii) and (iv). This follows by noting that we did not use the boundedness of X and Y in the proof.

Proposition 8.5. *Suppose that $h_1, \dots, h_n$ are strictly concave and $X \in \mathcal{X}$.*

(i) Every Pareto-optimal allocation in $\mathbb{A}_n(X)$ is comonotonic.

(ii) If for each $i \in [n]$, $h_i = a_i h_1$ for some $a_i > 0$ then an allocation is Pareto optimal in $\mathbb{A}_n(X)$ if and only if it is comonotonic.

As mentioned previously, Proposition 8.4 and 8.5 are generalizations of well-known results in the literature. Both can easily be extended to L^p for $p \geqslant 1$ instead of $\mathcal{X} = L^\infty$ provided that every ρ_{h_i} is finite when defined on L^p.[4]

[4]Conditions for the finiteness of ρ_h on L^p are provided in Proposition 1 of Wang et al. (2020a).

8.3.2 IQD agents and negatively dependent optimal allocations

We characterize the sum-optimal allocations on general spaces when agents evaluate their risk with the IQD measure of variability. We start with the problem of sharing risk among IQD agents only. In this setting, agent $i \in [n]$ has IQD_{α_i} as their preference where $\alpha_i \in [0, 1/2)$.

For a random variable X on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define tail events as in Wang and Zitikis (2021). For $\beta \in [0,1]$, we say that an event $A \in \mathcal{F}$ is a *right (resp. left) β-tail event* of X if $\mathbb{P}(A) = \beta$ and $X(\omega) \geqslant X(\omega')$ (resp. $X(\omega) \leqslant X(\omega')$) holds for a.s. all $\omega \in A$ and $\omega' \in A^c$, where A^c stands for the complement of A.

Theorem 8.2. *For $X \in \mathcal{X}$ and the IQD risk sharing problem in $\mathbb{A}_n(X)$ with $\alpha_1, \dots, \alpha_n \in [0, 1/2)$, let $\alpha = \sum_{i=1}^n \alpha_i$.*

(i) An allocation of X is Pareto optimal if and only if it is sum optimal.

(ii) For $\lambda_1, \dots, \lambda_n \geqslant 0$ and $\lambda = \bigwedge_{i=1}^n \lambda_i$,

$$\square_{i=1}^{n} (\lambda_i \mathrm{IQD}_{\alpha_i}) = \left(\bigwedge_{i=1}^{n} \lambda_i \right) \mathrm{IQD}_{\sum_{i=1}^n \alpha_i} = \lambda \mathrm{IQD}_{\alpha}. \tag{8.6}$$

In particular, $\square_{i=1}^n \mathrm{IQD}_{\alpha_i} = \mathrm{IQD}_{\alpha}$.

(iii) A class of Pareto-optimal allocations of $X \in \mathcal{X}$ for IQD agents is given by

$$X_i = (X - c)\mathbb{1}_{A_i \cup B_i} + a_i (X - c)\left(1 - \mathbb{1}_{A \cup B}\right) + c_i, \qquad i \in [n], \tag{8.7}$$

where, by setting $\beta = \alpha \wedge (1/2)$,

(a) $\{A_i\}_{i=1}^n$ and $\{B_i\}_{i=1}^n$ are partitions of a right β-tail event A and a left β-tail event B of X with A, B disjoint, respectively, satisfying $\mathbb{P}(A_i) = \mathbb{P}(B_i) = \alpha_i \beta / \alpha$, $i \in [n]$;

(b) $a_i \geqslant 0$ for $i \in [n]$ and $\sum_{i=1}^n a_i = 1$;

(c) $c \in [Q_{1/2}^-(X), Q_{1/2}^+(X)]$ and $\sum_{i=1}^n c_i = c$.

Remark 8.2. The allocation (8.7) satisfies $\sum_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}(X_i) = \square_{i=1}^n (\lambda_i \mathrm{IQD}_{\alpha_i})(X)$ by setting $a_i = 0$ for $i \in [n]$ such that $\lambda_i > \lambda$.

The surprising ingredient of Theorem 8.2, part (i) is that, for IQD agents, sum optimality is indeed equivalent to Pareto optimality, which is the case for cash-additive distortion riskmetrics (Theorem 8.1). However, for general agents with $h_1(1) = \cdots = h_n(1) = 0$, Pareto optimality is not necessarily equivalent to sum optimality, because different choices of $(\lambda_1, \dots, \lambda_n)$ in Proposition 8.3 lead to different Pareto-optimal allocations, which are not necessarily sum optimal (see Proposition 8.3 as well as Section 8.6). As a consequence of this result, Pareto-optimal allocations for IQD agents are precisely those for agents using the mean-risk preferences with risk measured by IQD,

$$\rho_{h_i}(X_i) = \mathbb{E}[X_i] + \mathrm{IQD}_{\alpha_i}(X_i), \quad i \in [n],$$

because both solve the same sum optimality problem by noting that $\sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X]$ for any allocation $(X_1, \dots, X_n)$ of X.

In part (ii) of Theorem 8.2, we see that the inf-convolution of several IQD is an IQD. Related to this observation, Embrechts et al. (2018) showed that the inf-convolution of several quantiles is again a quantile.

Figure 8.2 illustrates an example of the Pareto-optimal allocation (8.7) in Theorem 8.2, part (iii). The dependence structure of this allocation warrants some further discussion. Without loss of generality, assume $c_1 = \cdots = c_n = 0$ (this implies that a median of X is $c = 0$), and assume that X is continuously distributed. Note that (a.s.) $X > 0$ if event A occurs and $X < 0$ if event B occurs.

First, suppose $\alpha \geqslant 1/2$ so that $\mathbb{P}(A \cup B) = 1$. In this case, we have $X_i = X\mathbb{1}_{A_i \cup B_i}$ for $i \in [n]$. The random vector $(X\mathbb{1}_{A_i}, X\mathbb{1}_{A_j})$ for $i \neq j$ is counter-monotonic because $A_i \cap A_j = \varnothing$ and $X > 0$ on A. This implies $(X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n})$ is *pairwise counter-monotonic*. From the above analysis, we can see that conditional on A, $(X_1, \dots, X_n)$ is pairwise counter-monotonic, and so is it conditional on B; that is $(X_1, \dots, X_n)$ is a mixture of two pairwise counter-monotonic vectors. Moreover, $(X_1, \dots, X_n)$ is also the sum of the two pairwise counter-monotonic vectors $(X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n})$ and $(X\mathbb{1}_{B_1}, \dots, X\mathbb{1}_{B_n})$. We can check that $(X_i(\omega) - X_j(\omega))(X_i(\omega') - X_j(\omega')) < 0$ for $\omega \in A_i$ and $\omega' \in A_j$, and $(X_i(\omega) - X_j(\omega))(X_i(\omega') - X_j(\omega')) > 0$ for $\omega \in A_i$ and $\omega' \in B_j$. Therefore, the allocation $(X_1, \dots, X_n)$ is not comonotonic, yet it is not pairwise counter-monotonic either. This is illustrated by the "vertical slicing" in

Figure 8.2: A Pareto-optimal allocation in (8.7), where the shaded area represents the allocation to agent 1 minus c_1, that is, $X_1 - c_1 = (X-c)\mathbb{1}_{A_i \cup B_i} + a_i(X-c)\mathbb{1}_{(A\cup B^c)}$

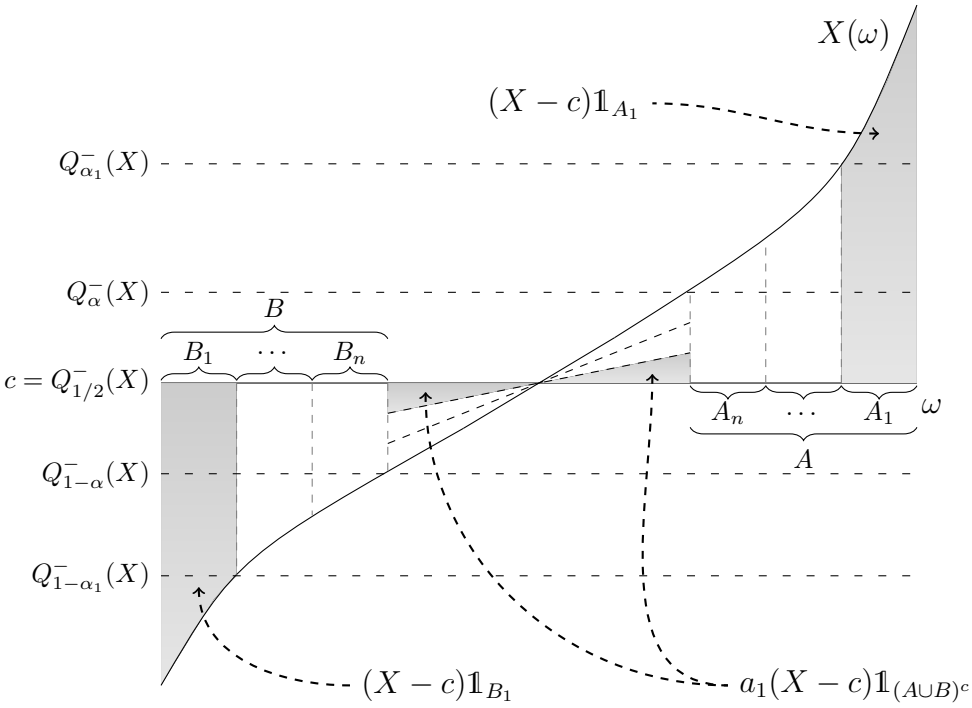


Figure 8.2, where on A and B pairwise counter-monotonicity holds.

As discussed above, we can describe $(X_1, \dots, X_n)$ as either the sum or the mixture of two pairwise counter-monotonic vectors. Pairwise counter-monotonicity is a form of extreme negative dependence that extends the concept of counter-monotonicity to the case of $n \geqslant 3$ agents; see Puccetti and Wang (2015) and Chapter 7 for more details. This observation is in contrast to the optimal allocations for quantile agents in Theorem 1 of Embrechts et al. (2018), which are indeed pairwise counter-monotonic.

If $0 < \alpha < 1/2$, then the term $a_i X \mathbb{1}_{(A\cup B)^c}$ appears in the allocation of every agent. Note that conditional on $(A \cup B)^c$, $(X_1, \dots, X_n)$ becomes comonotonic. This is illustrated by "proportional slicing" in the middle part of Figure 8.2. This local comonotonicity will become crucial in Section 8.5, where we study the risk sharing problem among several IQD agents and risk-averse agents.

As hinted by Propositions 8.4 and 8.5, solving Pareto-optimal allocations for risk-averse agents requires us to study comonotonic risk sharing, which is the topic of the next section.

8.4 Comonotonic risk sharing

We now turn to the important case of comonotonic risk sharing. As before, we first provide theoretical results and then proceed to analyze further the special case of sharing risks with IQD agents.

8.4.1 Pareto optimality, sum optimality, and explicit allocations

The next result is similar to Theorem 8.1, but for comonotonic risk sharing. We omit its proof because it does not provide new insights.

Proposition 8.6. *Suppose that $h_i(1) \neq 0$ for some $i \in [n]$. Then, $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ is Pareto optimal in $\mathbb{A}_n^+(X)$ if and only if $\sum_{i=1}^n \rho_{\tilde{h}_i}(X_i) = \boxplus_{i=1}^n \rho_{\tilde{h}_i}(X)$.*

We now show that $\boldsymbol{\lambda}$-optimality in $\mathbb{A}_n^+(X)$ pins down Pareto optimality. This result is stated in a stronger form than Proposition 8.3 for the corresponding notions of optimality in $\mathbb{A}_n(X)$.

Proposition 8.7. *Suppose that $h_i(1) = 0$ for all $i \in [n]$. For an allocation $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$, it holds that (i)⇒(ii)⇒(iii):*

(i) $\sum_{i=1}^n \lambda_i \rho_{h_i}(X_i) = \boxplus_{i=1}^n (\lambda_i \rho_{h_i})(X)$ *for some* $(\lambda_1, \dots, \lambda_n) \in (0, \infty)^n$*;*

(ii) $(X_1, \dots, X_n)$ *is Pareto optimal in* $\mathbb{A}_n^+(X)$*;*

(iii) $\sum_{i=1}^n \lambda_i \rho_{h_i}(X_i) = \boxplus_{i=1}^n (\lambda_i \rho_{h_i})(X)$ *for some* $(\lambda_1, \dots, \lambda_n) \in [0, \infty)^n \setminus \{\mathbf{0}\}$.

Comonotonicity plays an important role in the proof of Proposition 8.7. The comonotonic additivity of distortion riskmetrics guarantees that the utility possibility frontier S is a convex set when restricted to $\mathbb{A}_n^+(X)$. This needs not be true on $\mathbb{A}_n(X)$. In this case, we cannot use the Hanh-Banach Theorem to obtain the existence of the Pareto weights $(\lambda_1, \dots, \lambda_n)$, which is the reason why we did not state a "converse statement" in Proposition 8.3. Propositions 8.4 and 8.7 together yield that if all agents have concave distortion functions, then any Pareto-optimal allocation in $\mathbb{A}_n(X)$, which yields the same welfare for

all agents as a Pareto-optimal allocation in $\mathbb{A}_n^+(X)$, must satisfy (iii). If their distortion functions are strictly concave, then by Proposition 8.5, every Pareto-optimal allocation can be found through an inf-convolution.

We now aim to characterize further the set of Pareto-optimal allocations in $\mathbb{A}_n^+(X)$. The following result generalizes Proposition 5 of Embrechts et al. (2018) for dual utilities.

Theorem 8.3. *Suppose that* $h_1(1) = \cdots = h_n(1)$. *Then*

$$\boxplus_{i=1}^{n} \rho_{h_i} = \rho_{h_\wedge},$$

where $h_\wedge(t) = \min\{h_1(t), \dots, h_n(t)\}$, *and* $\rho_{h_\wedge}$ *is finite on* $\mathcal{X}$. *Moreover, a sum-optimal allocation* $(X_1, \dots, X_n)$ *in* $\mathbb{A}_n^+(X)$ *is given by* $X_i = f_i(X)$, $i = 1, \dots, n$, *where*

$$f_i(x) = \int_0^x g_i(t)\,\mathrm{d}t, \quad \text{and} \quad g_i(x) = \frac{1}{|M_x|}\mathbb{1}_{\{i \in M_x\}}, \qquad x \in \mathbb{R}, \tag{8.8}$$

and where $M_x = \{j \in [n] : h_j(\mathbb{P}(X > x)) = h_\wedge(\mathbb{P}(X > x))\}$. *The sum-optimal allocation is unique up to constant shifts almost surely if and only if* $|M_x| = 1$ *for* μ_X*-almost every* x, *where* μ_X *is the distribution measure of* X.

A key step in the proof of Theorem 8.3 is the following lemma, which gives a convenient alternative formula for $\rho_h(f(X))$. The lemma generalizes Lemma 2.1 of Cheung and Lo (2017) for dual utilities in the context of optimal reinsurance design.

Lemma 8.3. *For any* $h \in \mathcal{H}^{\mathrm{BV}}$, *random variable* X *bounded from below, and increasing Lipschitz function* f *with right-derivative* g, *we have*

$$\rho_h(f(X)) = \int_0^\infty g(x)h(\mathbb{P}(X > x))\,\mathrm{d}x + \int_{-\infty}^0 g(x)(h(\mathbb{P}(X > x) - h(1))\,\mathrm{d}x. \tag{8.9}$$

The results in Theorem 8.3 can be extended to domains like $\{X \in L^p : X_- \in L^\infty\}$ for $p \geqslant 0$ as long as $\rho_{h_1}, \dots, \rho_{h_n}$ are finite on this domain. This is because Lemma 8.3 only requires boundedness from below. The next example illustrates the uniqueness statement in Theorem 8.3, which gives not only unique sum-optimal allocations in $\mathbb{A}_n^+(X)$, but also unique Pareto-optimal ones, up to constant shifts.

Example 8.2. Suppose that $\rho_{h_1} = \beta_1\mathbb{E}+\gamma_1\mathrm{GD}$, $\rho_{h_2} = \beta_2\mathbb{E}+\gamma_2\mathrm{MMD}$ and $\rho_{h_3} = \beta_3\mathbb{E}+\gamma_3\mathrm{IQD}_\alpha$ for some $\beta_i, \gamma_i > 0$, $i = 1, 2, 3$, and $\alpha \in [0, 1/2)$. For any continuously distributed $X \in \mathcal{X}$, the Pareto-optimal allocation in $\mathbb{A}_3^+(X)$ is unique up to constant shifts. To see this, by Proposition 8.6, any Pareto-optimal allocation (X_1, X_2, X_3) in $\mathbb{A}_3^+(X)$ satisfies $\sum_{i=1}^3 \rho_{\tilde{h}_i}(X_i) = \boxplus_{i=1}^3 \rho_{\tilde{h}_i}(X)$. Noting that for each $1 \leqslant i < j \leqslant 3$, $\tilde{h}_i(t) = \tilde{h}_j(t)$ for at most two points $t \in (0, 1)$, by Theorem 8.3, the allocation (X_1, X_2, X_3) is unique up to constant shifts.

By replacing h_i with $\lambda_i h_i$ for some $\lambda_i \geqslant 0$, we obtain the following corollary, which helps to identify $\boldsymbol{\lambda}$-optimal allocations in conjunction with Theorem 8.3.

Corollary 8.2. *Let $\boldsymbol{\lambda} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$ be a vector and $\boxplus_{i=1}^n \rho_{\lambda_i h_i}$ be finite. Then $\boxplus_{i=1}^n \rho_{\lambda_i h_i} = \rho_{h_{\boldsymbol{\lambda}}}$, where $h_{\boldsymbol{\lambda}}(t) = \min\{\lambda_1 h_1(t), \dots, \lambda_n h_n(t)\}$ for $t \in [0, 1]$.*

By Proposition 8.2, the inf-convolution $\boxplus_{i=1}^n \rho_{\lambda_i h_i}$ being finite implies that $\lambda_i h_i(1)$ are equal for all $i \in [n]$. Corollary 8.2 is thus only useful for the case of location-invariant distortion riskmetrics ($h_i(1) = 0$, $i \in [n]$), as otherwise we simply normalize $\lambda_i = 1$, $i \in [n]$. Theorem 8.3's characterization of $\boldsymbol{\lambda}$-optimality in $\mathbb{A}_n^+(X)$ extends to location-invariant distortion riskmetrics by setting $M_x = \{i \in [n] : \lambda_i h_i(\mathbb{P}(X > x)) = h_{\boldsymbol{\lambda}}(\mathbb{P}(X > x))\}$ in (8.8).

For cash-additive and reverse cash-additive distortion riskmetrics, Proposition 8.6 and Theorem 8.3 together yield a full characterization of Pareto-optimal allocations in $\mathbb{A}_n^+$. It remains to characterize those for location-invariant distortion riskmetrics. The next proposition gives an answer for a large class of such distortion riskmetrics.

Proposition 8.8. *Suppose $h_i(1) = 0$ and $h_i(t) > 0$ for all $i \in [n]$ and all $t \in (0, 1)$. Then the allocation $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ is Pareto optimal if and only if there exists $K \subseteq [n]$ and a vector $\boldsymbol{\lambda} \in (0, \infty)^{\#K}$ such that $(X_i)_{i\in K}$ solves $\boxplus_{i\in K} \rho_{\lambda_i h_i}(X)$, and X_i, $i \notin K$ are constants.*

The assumption that $h_i(t) > 0$ for all $i \in [n]$ and all $t \in (0, 1)$ is critical for the characterization of Proposition 8.8. This condition has a natural interpretation, as it is equivalent to $\rho_{h_i}(X) > 0$ for all non-degenerate X and it is satisfied by many variability measures; it is part of the definition of deviation measures of Rockafellar et al. (2006). But this assumption rules out IQD, which we study in the next section.

8.4.2 IQD agents and positively dependent optimal allocations

We start with the comonotonic risk sharing problem among IQD agents. The following proposition gives the corresponding statements, parallel to Theorem 8.2, on Pareto optimality and inf-convolution in this setting. The sum-optimal allocations are given by Theorem 8.3.

Proposition 8.9. *Consider $X \in \mathcal{X}$ and the IQD risk sharing problem in $\mathbb{A}_n^+(X)$ with $\alpha_1, \dots, \alpha_n \in [0, 1/2)$.*

(i) An allocation of X is Pareto optimal if and only if it is sum optimal.

(ii) For $\lambda_1, \dots, \lambda_n \geqslant 0$,

$$\boxplus_{i=1}^{n}(\lambda_i \mathrm{IQD}_{\alpha_i}) = \left(\bigwedge_{i=1}^{n} \lambda_i\right) \mathrm{IQD}_{\bigvee_{i=1}^{n} \alpha_i}.$$

In particular, $\boxplus_{i=1}^{n} \mathrm{IQD}_{\alpha_i} = \mathrm{IQD}_{\bigvee_{i=1}^{n} \alpha_i}$.

Comparing Theorem 8.2 with Proposition 8.9, we note that for $\alpha_1, \dots, \alpha_n \in (0, 1/2)$, we have $\sum_{i=1}^{n} \alpha_i > \bigvee_{i=1}^{n} \alpha_i$, which implies that

$$\boxplus_{i=1}^{n}(\lambda_i \mathrm{IQD}_{\alpha_i})(X) - \square_{i=1}^{n}(\lambda_i \mathrm{IQD}_{\alpha_i})(X) > 0 \tag{8.10}$$

for any continuously distributed X. This further implies that the Pareto-optimal allocations in $\mathbb{A}_n(X)$ are disjoint from those in $\mathbb{A}_n^+(X)$. The difference in (8.10) can be interpreted as the welfare gain of allowing agents to share risks in non-comonotonic arrangements.

8.5 Several IQD and risk-averse agents

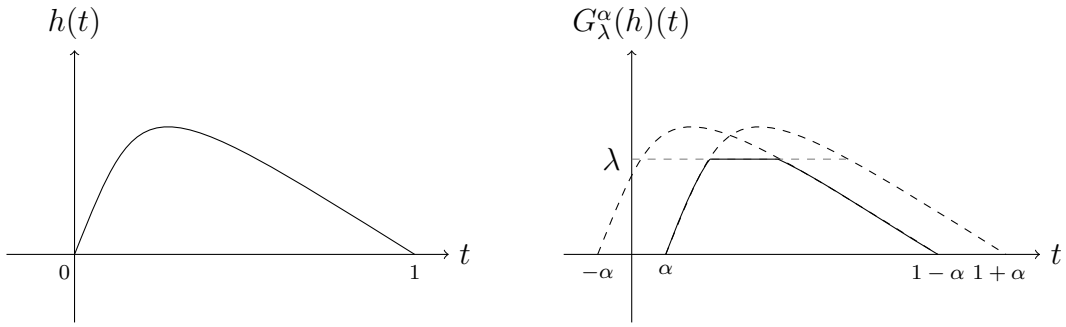
Combining results established in Sections 8.3 and 8.4, we are now able to tackle the unconstrained risk sharing problem for IQD and risk-averse agents. We consider agents from the following two sets: the IQD agents, modelled by distortion functions in

$$\mathcal{H}^{\mathrm{IQD}} = \{t \mapsto \mathbb{1}_{\{\alpha < t < 1-\alpha\}} : \alpha \in [0, 1/2)\}$$

and the risk-averse agents, modelled by distortion functions in

$$\mathcal{H}^{\mathrm{C}} = \{h \in \mathcal{H}^{\mathrm{BV}} |\ h(1) = 0,\ h \text{ is concave}\}.$$

Figure 8.3: An illustration of the transform G_λ^α



(a) h (b) $G_\lambda^\alpha(h)$

That is, $\mathcal{H}^{\mathrm{IQD}}$ is the set of all distortion functions for IQD variability measures and $\mathcal{H}^{\mathrm{C}}$ is the set of location-invariant concave distortion functions $h \in \mathcal{H}^{\mathrm{BV}}$. Notice that each $h \in \mathcal{H}^{\mathrm{C}}$ is increasing in $[0, s]$ and decreasing $[s, 1]$ for some $s \in (0, 1)$. Define the mapping $G_\lambda^\alpha : \mathcal{H}^{\mathrm{C}} \to \mathcal{H}^{\mathrm{BV}}$ for $\alpha \in [0, 1/2)$ and $\lambda \geqslant 0$ as

$$G_\lambda^\alpha(h)(t) = (h(t-\alpha) \wedge h(t+\alpha) \wedge \lambda)\, \mathbb{1}_{\{\alpha < t < 1-\alpha\}} \quad \text{for } t \in [0,1].$$

The mapping G_λ^α transforms a concave distortion function to another distortion function with value 0 on $[0, \alpha] \cup [1-\alpha, 1]$. See Figure 8.3 for an illustration of this transform. For $\alpha \geqslant 1/2$, we define $G_\lambda^\alpha(h) = 0$.

We will see in the next proposition that the function G_λ^α plays an important role because of the inf-convolution of $\lambda\mathrm{IQD}_\alpha$ and ρ_h for $h \in \mathcal{H}^{\mathrm{C}}$ satisfies

$$(\lambda\mathrm{IQD}_\alpha) \square \rho_h = \rho_{G_\lambda^\alpha(h)}.$$

This formula is a special case of (8.11) in Theorem 8.4 below.

Theorem 8.4. *Let $C \subseteq [n]$ and $I = [n] \setminus C$. Suppose that $h_i \in \mathcal{H}^{\mathrm{C}}$ for $i \in C$ and $h_i \in \mathcal{H}^{\mathrm{IQD}}$ for $i \in I$ with IQD parameter α_i. Denote by $\alpha = \sum_{i \in I} \alpha_i$.*

(i) For $\lambda_1, \dots, \lambda_n \geqslant 0$, denoting by $\lambda = \bigwedge_{i \in I} \lambda_i$ and $h = \bigwedge_{i \in S} (\lambda_i h_i)$, we have

$$\square_{i=1}^{n} (\lambda_i \rho_{h_i}) = \rho_{G_\lambda^\alpha(h)}. \tag{8.11}$$

(ii) A Pareto-optimal allocation is given by

$$X_i = (X - c)\mathbb{1}_{A_i \cup B_i} + Y_i + c_i, \tag{8.12}$$

where, by denoting by $\beta = \alpha \wedge (1/2)$,

(a) $\{A_i\}_{i=1}^n$ *and* $\{B_i\}_{i=1}^n$ *are partitions of a right* β*-tail event* A *and a left* β*-tail event* B *of* X *with* A, B *disjoint, respectively, satisfying* $\mathbb{P}(A_i) = \mathbb{P}(B_i) = \alpha_i\beta/\alpha$ *for* $i \in I$ *and* $A_i = B_i = \varnothing$ *for* $i \in C$*;*

(b) $(Y_1, \dots, Y_n)$ *is a Pareto-optimal allocation of* $(X - c)\mathbb{1}_{(A\cup B)^c}$ *for preferences with distortion functions* $h_1', \dots, h_n'$ *where* $h_i' = h_i$ *if* $i \in C$ *and* $h_i(t) = \mathbb{1}_{\{t\in(0,1)\}}$ *for* $i \in I$*.*

(c) $c \in [Q_{1/2}^-(X), Q_{1/2}^+(X)]$ *and* $\sum_{i=1}^n c_i = c$*.*

The type of allocation characterized in Theorem 8.4 has some special features. In contrast to the allocation with several IQD agents only in Theorem 8.2, the risk in case of the event $(A \cup B)^c$ are optimally shared among *risk-averse* agents with distortion functions $h_1', \dots, h_n'$, which are all concave. To solve for the allocation $(Y_1, \dots, Y_n)$ in (b) of Theorem 8.4, we can conveniently convert the problem into a comonotonic allocation problem as guaranteed by Proposition 8.4, and this allocation is fully solved by Theorem 8.3, Corollary 8.2, and Proposition 8.8, thus yielding an explicit solution to the problem in this section.

Remark 8.3. Let $C \subseteq [n]$ and $I = [n] \setminus C$. Suppose that $h_i \in \mathcal{H}^{\mathrm{C}}$ for $i \in C$ and $h_i \in \mathcal{H}^{\mathrm{IQD}}$ for $i \in I$ with IQD parameter α_i. For any $\lambda_1, \dots, \lambda_n \geqslant 0$ it is $\boxplus_{i=1}^n(\lambda_i\rho_{h_i}) = \rho_{h_{\boldsymbol{\lambda}}}$, where $h_{\boldsymbol{\lambda}} = \bigwedge_{i\in[n]} \lambda_i h_i$. The distortion function $h_{\boldsymbol{\lambda}}$ takes value 0 on $[0, \bigvee_{i\in I} \alpha_i] \cup [\bigvee_{i\in I} \alpha_i, 1]$; on the other hand, the distortion function $G_\lambda^\alpha(h)$ from Theorem 8.4 takes value 0 on $[0, \sum_{i\in I} \alpha_i] \cup [\sum_{i\in I} \alpha_i, 1]$.

8.6 GD, MMD and IQD agents

We now provide examples of the results obtained in Section 8.3 and 8.4. Some calculation details are put in Section 8.12. The following two subsections come back on the risk

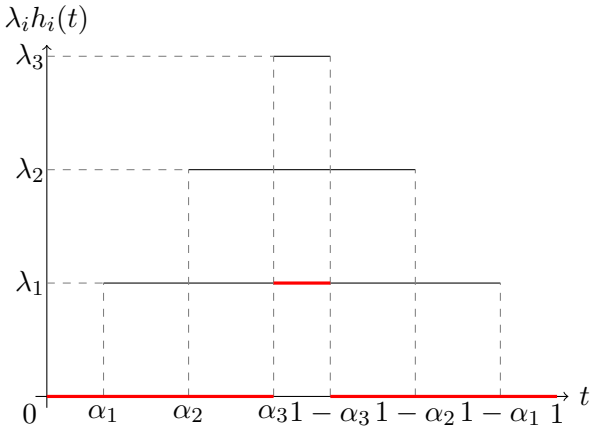
sharing problem with several IQD agents and explains further the allocations found in Section 8.3.2. The last two subsections analyze the risk sharing problem when agents consider the Gini and mean-median deviations as the relevant statistical measures of risk.

8.6.1 Several IQD agents

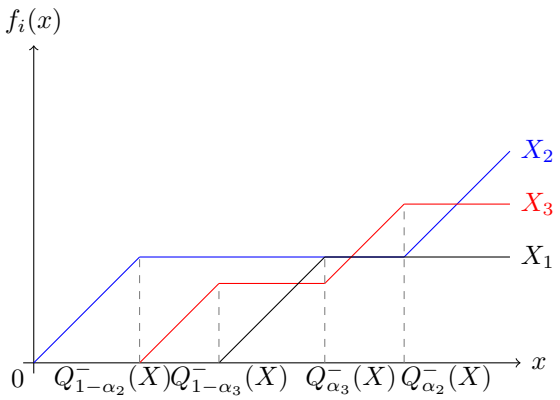
The difference between the two sum-optimal allocations found in Theorem 8.2 and Proposition 8.9 is important.

In contrast, Figure 8.4 illustrates some comonotonic allocations that are $\boldsymbol{\lambda}$-optimal (and also Pareto optimal and sum optimal; see Proposition 8.9) when restricted to the subset $\mathbb{A}_n^+(X)$. The solution for $\boxplus_{i=1}^n(\lambda_i \mathrm{IQD}_{\alpha_i})$ is not unique as $|M_x|$ can be larger than 1. The figure depicts a particular case when simultaneously $\alpha_1 < \alpha_2 < \alpha_3$ and $\lambda_1 < \lambda_2 < \lambda_3$. The left panel shows the distortion function of each agent multiplied by the corresponding λ, and the lower envelope $h_{\boldsymbol{\lambda}}(t)$. Figure 8.4b presents a sum-optimal allocation where all three agents take non-zero risks. Comonotonic sum-optimal allocations are not unique, because the allocation where agent 3 takes all risks in the α_3-tails and agent 1 takes the rest is also sum optimal. As discussed before, comonotonic sum-optimal allocations are generally not sum optimal in $\mathbb{A}_n(X)$.

Figure 8.4: Distortion functions and the sum-optimal allocation for $\boxplus_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}$



(a) Distortion functions for $\lambda_i \mathrm{IQD}_{\alpha_i}$, $i = 1, 2, 3$



(b) Allocation for $\boxplus_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}$

8.6.2 The GD, MMD, and IQD problem

We now turn to the allocations characterized by Theorem 8.4. Consider the problem of sharing risk between Anne, Bob and Carole, i.e., the case when there is only one GD agent, one MMD and IQD agent. Let $\alpha < 1/2$ and $\lambda_1, \lambda_2, \lambda_3 > 0$ and consider the inf-convolution

$$\inf_{(X_1,X_2,X_3)\in\mathbb{A}(X)} \{\lambda_1 \mathrm{GD}(X_1) + \lambda_2 \mathrm{MMD}(X_2) + \lambda_3 \mathrm{IQD}_\alpha(X_3)\}.$$

Without loss of generality we assume $Q^-_{1/2}(X) = 0$ for the convenience of presentation, so that c in Theorem 8.4 is taken as 0.

Let A be a right α-tail event and $B \subsetneq A^c$ be a left α-tail event of X, where A and B are disjoint sets. All the α-tail risks must go to the IQD agent. That is, every sum-optimal allocation requires that the IQD agent takes the whole risk on $A \cup B$.

It remains to share risk "in the middle", that is, on the event $(A \cup B)^c$. We note by $Y = X\mathbb{1}_{(A\cup B)^c}$, which has an optimal allocation $(Y_1, \dots, Y_n)$ in Theorem 8.4 which is comonotonic on $(A \cup B)^c$. This is done in the same fashion as we do later for comonotonic risk sharing, with the caveat that the IQD agent might take on some risk depending on the weights λ_1, λ_2 and λ_3. Define $c_1 = 1/2 - \sqrt{1/4 - \lambda_3/\lambda_1} + \alpha$, $c_2 = \lambda_3/\lambda_2 + \alpha$ and $c_3 = 1 - \lambda_2/\lambda_1 + \alpha$. If $c_1 \in (\alpha, 1/2)$, then $\lambda_1 h_{\mathrm{GD}}(t)$ and $\lambda_3 h_{\mathrm{IQD}}(t)$ cross twice on $(0, 1)$, once at $c_1 - \alpha$ and then once again at $1 - c_1 + \alpha$. If $c_2 \in (\alpha, 1/2)$, then $\lambda_2 h_{\mathrm{MMD}}(t)$ and $\lambda_3 h_{\mathrm{IQD}}(t)$ cross twice on $(0, 1)$, once at $c_2 - \alpha$ and then once again at $1 - c_2 + \alpha$. Similarly, if $c_3 \in (\alpha, 1/2)$ then $\lambda_1 h_{\mathrm{GD}}(t)$ and $\lambda_2 h_{\mathrm{MMD}}(t)$ cross at $c_3 - \alpha$ and $1 - c_3 + \alpha$. Note that $c_2 > \alpha$ and $\alpha \leqslant c_1 \leqslant 1/2 + \alpha$ whenever $1/4 \geqslant \lambda_3/\lambda_1$.

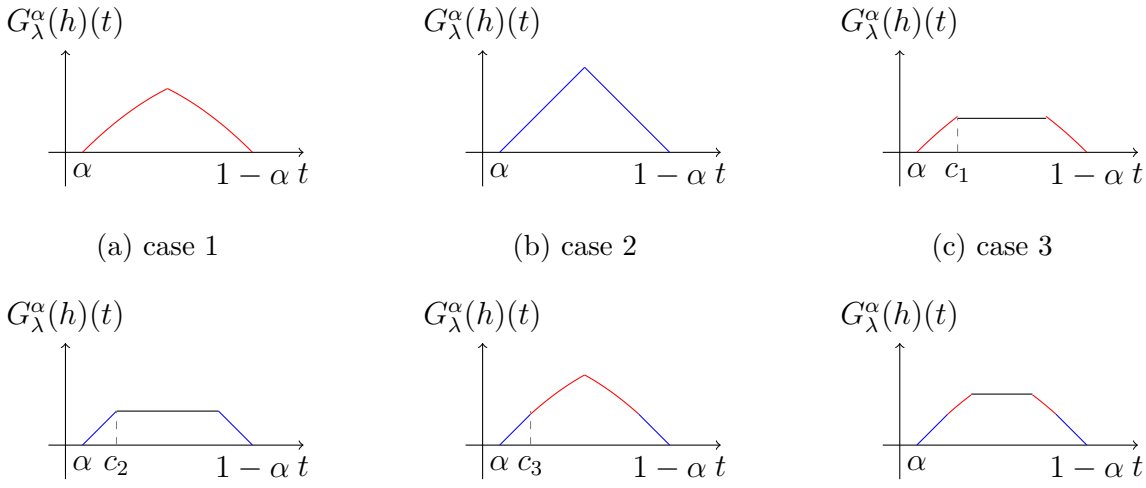
We have six cases to handle; the details can be found in Section 8.12. Figure 8.5 plots the function $G^\alpha_\lambda(h)$ for $h = \min\{\lambda_1 h_{\mathrm{GD}}, \lambda_2 h_{\mathrm{MMD}}\}$. The red, blue and black colour denote the risk that goes to the GD agent, the MMD agent and the IQD agent, respectively.

We present the Pareto-optimal allocations (X_1, X_2, X_3) in the six cases below. These allocations are generally not comonotonic, but they are comonotonic on the event $(A \cup B)^c$. Recall that Y stands for $X\mathbb{1}_{(A\cup B)^c}$.

Case 1, $c_1 \geqslant 1/2$ and $c_3 \leqslant \alpha$: $X_1 = Y$, $X_2 = 0$ and $X_3 = X\mathbb{1}_{A\cup B}$.

Case 2, $c_2 \geqslant 1/2$ and $c_3 \geqslant 1/2$: $X_1 = 0$, $X_2 = Y$ and $X_3 = X\mathbb{1}_{A\cup B}$.

Figure 8.5: The function $G^{\alpha}_{\lambda}(h)$



(a) case 1 (b) case 2 (c) case 3

(d) case 4 (e) case 5 (f) case 6

Case 3, $c_3 \leqslant \alpha < c_1 < 1/2$: $X_1 = X - X_3$, $X_2 = 0$ and $X_3 = X\mathbb{1}_{A\cup B} + Y \wedge Q^-_{c_1}(X) - Y \wedge Q^-_{1-c_1}(X)$.

Case 4, either $\alpha < c_1 < c_3 < 1/2$ or $\alpha < c_2 < 1/2 < c_3$: $X_1 = 0$, $X_2 = X - X_3$ and

$$X_3 = X\mathbb{1}_{A\cup B} + Y \wedge Q^-_{c_2}(X) - Y \wedge Q^-_{1-c_2}(X).$$

Case 5, $\alpha < c_3 < 1/2 < c_1$: $X_2 = X - X_1 - X_3$, $X_3 = X\mathbb{1}_{A\cup B}$ and $X_1 = Y \wedge Q^-_{c_3}(X) - Y \wedge Q^-_{1-c_3}(X)$.

Case 6, $\alpha < c_3 \leqslant c_1 < 1/2$: $X_1 = Y \wedge Q^-_{c_3}(X) - Y \wedge Q^-_{c_1}(X) + Y \wedge Q^-_{1-c_1}(X) - Y \wedge Q^-_{1-c_3}(X)$,

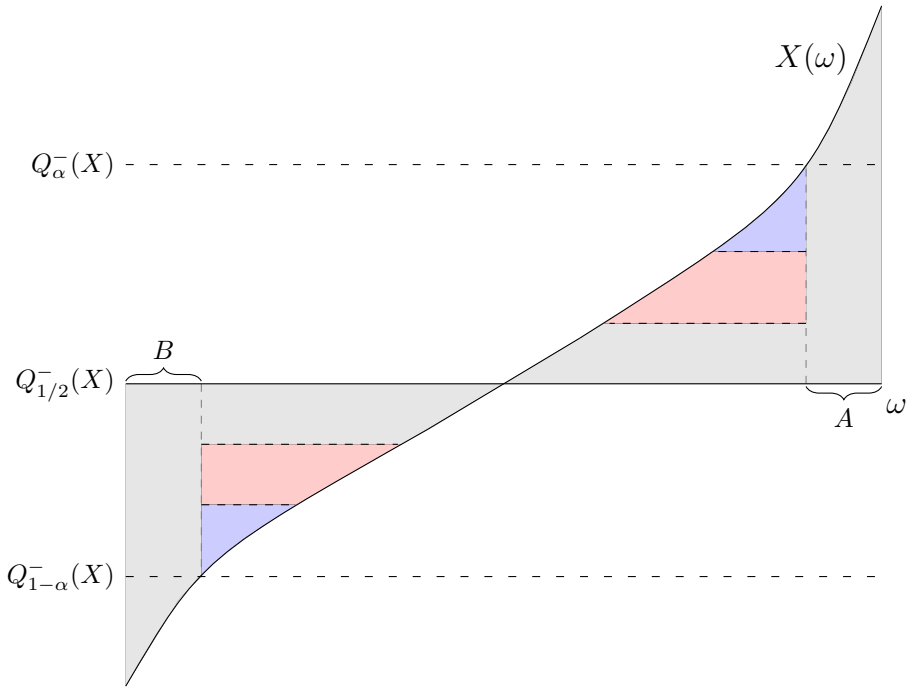
$$X_2 = X - X_1 - X_3 \quad \text{and} \quad X_3 = X\mathbb{1}_{A\cup B} + Y \wedge Q^-_{c_1}(X) - Y \wedge Q^-_{1-c_1}(X).$$

The allocation in case 6 is showing a particularly rich feature, and we depict it in Figure 8.6.

8.6.3 Insurance between two GD and MMD agents

We next solve the insurance example (Example 8.1) presented in the introduction. Consider two individuals, Anne and Bob, who evaluate their risk with GD and MMD, respectively.

Figure 8.6: A Pareto-optimal allocation for Anne, Bob and Carole, where the red, blue and gray areas represent the allocations to Anne (GD), Bob (MMD) and Carole (IQD) respectively, up to constant shifts



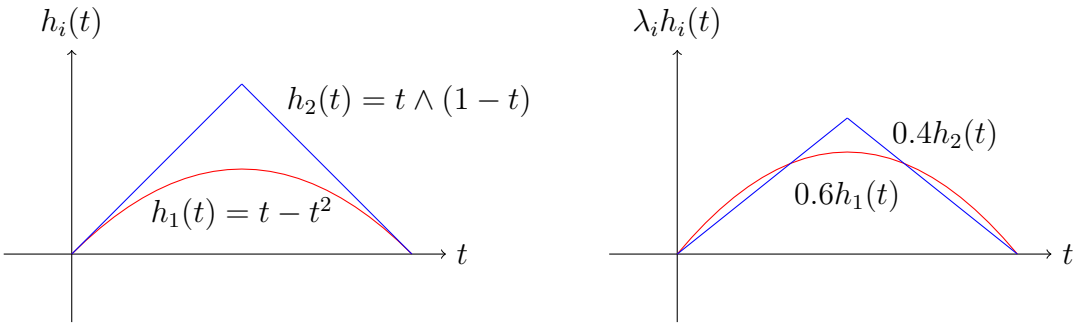
That is, set $h_1 = h_{\mathrm{GD}}$ and $h_2 = h_{\mathrm{MMD}}$. (Or, they could use $\mathbb{E} + \lambda_1 \mathrm{GD}$ and $\mathbb{E} + \lambda_2 \mathrm{MMD}$, which would not change our analysis.) This setting is simpler than the three-agent problem in Section 8.6.2, and it offers a clearer visualization of the Pareo-optimal allocation.

Both h_1 and h_2 are strictly concave, and, by Proposition 8.5, any Pareto-optimal allocation in $\mathbb{A}_n(X)$ is comonotonic. By Proposition 8.7, each Pareto-optimal allocation can be found by solving the inf-convolution $\boxplus_{i=1}^2(\lambda_i \rho_{h_i})$ for some Pareto weights $(\lambda_1, \lambda_2) \in [0,\infty)^2 \setminus \{\mathbf{0}\}$. Consider the normalized ones $\lambda_1 = \lambda \in [0,1]$ and $\lambda_2 = 1 - \lambda$. Figure 8.7 depicts the functions $h_i(t)$ and $\lambda_i h_i(t)$.

By positive homogeneity it is $\lambda \rho_{h_1}(X_1) = \rho_{\lambda h_1}(X_1)$ for Anne, and similarly for Bob. By Corollary 8.2, we have $\boxplus_{i=1}^2 \rho_{\lambda_i h_i} = \rho_{h_{\boldsymbol{\lambda}}}$, where $h_{\boldsymbol{\lambda}}(t) = \min\{\lambda h_1(t), (1-\lambda) h_2(t)\}$. That is, the sum-optimal allocation gives all the marginal t-quantile risk to the individual with the lowest $\lambda_i h_i(t)$.

The condition of Theorem 8.3 is satisfied, and so the (λ_1, λ_2)-optimal allocation is unique

Figure 8.7: Distortion functions of GD and MMD agents



(a) h_1 is GD, h_2 is MMD

(b) λh_2 and $(1-\lambda)h_1$ when $\lambda = 0.4$

up to constant shifts. Any Pareto-optimal allocation takes the form

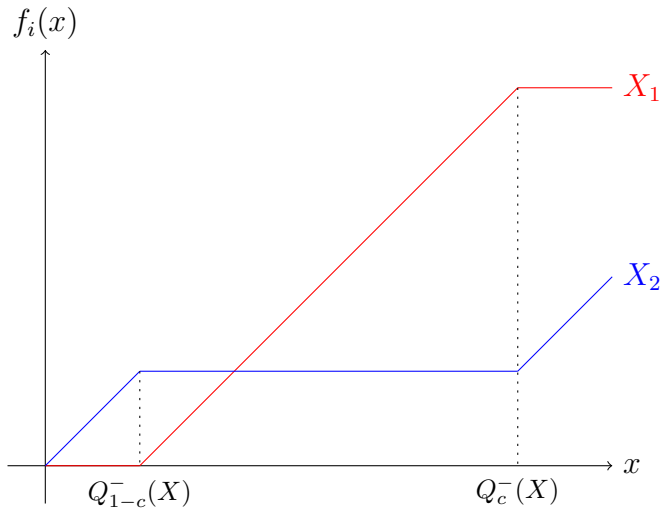
$$X_1 = X \wedge Q_c^-(X) - X \wedge Q_{1-c}^-(X) + k \quad \text{and} \quad X_2 = X - X_1,$$

where $c \in [0, 1/2]$ and $k \in \mathbb{R}$ is a constant. We can interpret this as a situation where the first individual insures the potential losses X of the second one. The contract (transfer function) is the random variable X_1, while its price is k, the latter which needs to be negotiated between the two agents. Next, we argue that the mapping $\lambda \mapsto c$ is surjective.

(i) If $\lambda < 1/2$ we have $\lambda h_1 < (1-\lambda)h_2$ everywhere and so $c = 0$. That is, the GD agent bears all the risk and provides full insurance. (ii) Similarly, if $\lambda > 2/3$ we have $\lambda h_1 > (1-\lambda)h_2$ everywhere and so $c = 1/2$. It is the MMD agent that bears all the risk and no insurance is provided. Finally, (iii) if $1/2 < \lambda < 2/3$ then $\lambda h_1 > (1-\lambda)h_2$ on both $(0, (2\lambda-1)/\lambda)$ and $((1-\lambda)/\lambda, 1)$ and $\lambda h_1 < (1-\lambda)h_2$ on $((2\lambda-1)/\lambda, (1-\lambda)/\lambda)$. Hence, $c = (2\lambda-1)/\lambda$ and the contract is a simple deductible $Q_{1-c}^-(X)$ with an upper limit $Q_c^-(X)$. This type of allocation is depicted in Figure 8.8.

The constant k can take any value because by location invariance, for any $k \in \mathbb{R}$, we have $\rho_{h_i}(X_i + k) = \rho_{h_i}(X_i) + h_i(1)k = \rho_{h_i}(X_i)$ and the price of the insurance does not affect Pareto optimality. This observation remains true if agents use $\mathbb{E} + \lambda_i \rho_{h_i}$ instead of ρ_{h_i}.

Figure 8.8: A Pareto-optimal allocation for the MMD and GD pair



8.6.4 Risk sharing with several mixed GD-MMD agents

We conclude with the problem of sharing risk among many agents $i \in [n]$ evaluating their risks with the variability measure

$$\rho_{h_i}(X_i) = \int X_i \, \mathrm{d}\left((a_i h_{\mathrm{GD}} + (1-a_i)h_{\mathrm{MMD}}) \circ \mathbb{P}\right) = a_i \mathrm{GD}(X_i) + (1-a_i)\mathrm{MMD}(X_i),$$

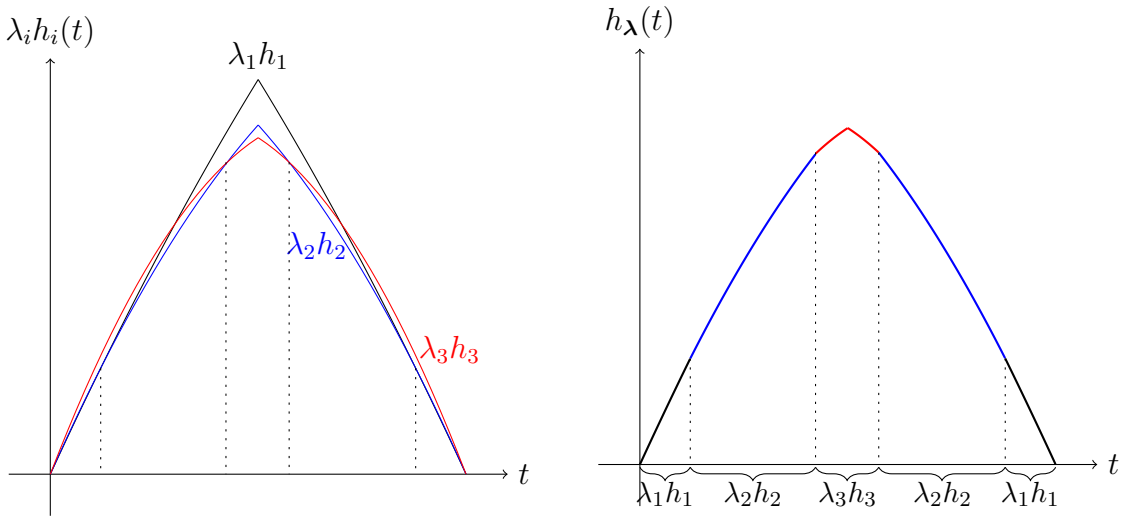
$a_i \in [0,1]$. It is easily verified that for every $i \in [n]$ the distortion function $h_i = a_i h_{\mathrm{GD}} + (1-a_i)h_{\mathrm{MMD}}$ is strictly concave and satisfies $h_i(1) = 0$. We can therefore invoke Theorem 8.3, Corollary 8.2 and Proposition 8.8 to characterize the set of Pareto-optimal allocations. Consider the usual normalization of the Pareto weights $\sum_{i=1}^n \lambda_i = 1$ with $\lambda_i > 0$ and notice that

$$\boxplus_{i=1}^n \rho_{\lambda_i h_i} = \rho_{h_{\boldsymbol{\lambda}}},$$

where $h_{\boldsymbol{\lambda}}(t) = \min\{\lambda_1 h_1(t), \dots, \lambda_n h_n(t)\}$.

Deriving every agent's allocation (contract) in a closed-form solution is a bit more cumbersome. Yet, Theorem 8.3 and Corollary 8.2 still fully pin down the shape of the optimal allocation, and we can visualize it easily. Consider the case when $0 < \lambda_1 a_1 < \lambda_2 a_2 < \cdots < \lambda_n a_n$ and set $M_x = \{i \in [n] : \lambda_i h_i(\mathbb{P}(X > x)) = h_{\boldsymbol{\lambda}}(\mathbb{P}(X > x))\}$ as before. We have that $|M_x| = 1$ μ_X-almost surely, so the sum-optimal allocation is unique up to constant shifts for any $\boldsymbol{\lambda}$. Figure 8.9 shows an example with three agents.

Figure 8.9: Distortion functions for mixed GD-MMD agents, where $a_1 = 1/4,\ a_2 = 1/2,\ a_3 = 3/4$ and $\boldsymbol{\lambda} = (0.31, 0.32, 0.37)$



(a) Distortion function $\lambda_i h_i$

(b) Lower envelope: $h_{\boldsymbol{\lambda}}(t) = \min_{i\in[3]} \lambda_i h_i(t)$

As we obtained in the previous application, $h_{\boldsymbol{\lambda}}$ induces a partition of the state space on which only one agent takes the full marginal risk. That is, the Pareto-optimal allocation's shape is similar to the payoff obtained with a collection of straight deductibles insurance contracts with upper limits. For instance, the part of the risk that goes to agent 2 is

$$X_2 = X \wedge b - X \wedge a + X \wedge d - X \wedge c$$

for $0 < a < b < c < d < \infty$ implicitly defined through the lower envelope $h_{\boldsymbol{\lambda}}(t)$.

8.7 Conclusion

We summarize the chapter with a few remarks on the results that we obtained. While we focused on the case $\mathcal{X} = L^\infty$, all the results of this article generalize to larger spaces provided all the decision criteria ρ_{h_i} and the inf-convolution $\square_{i=1}^n \rho_{h_i}(X)$ are finite on the larger space. We emphasized when key results can be readily generalized, but this finiteness property must often be verified case-by-case. For example, the results on risk sharing with IQD agents can be extended to L^0 because the IQD is bounded from below. This property does not generalize to other functionals.

The unconstrained risk sharing problem for non-concave distortion functions typically leads to non-comonotonic sum-optimal allocations without explicit forms, and they can be difficult to analyze. Although we obtained several results on necessary or sufficient conditions for Pareto and sum optimality (Theorem 8.1 and Propositions 8.2-8.5), a full characterization of the Pareto-optimal or sum-optimal allocations for arbitrary distortion riskmetrics is beyond the current techniques.

The case of IQD agents is, nevertheless, special, although they do not have concave distortion functions. For this setting, we can fully characterize all Pareto-optimal allocations via sum-optimal ones, and the inf-convolution for such distortion riskmetrics admit concise formulas (Theorem 8.2 and Proposition 8.9):

$$\operatorname*{\square}_{i=1}^{n}(\lambda_i \mathrm{IQD}_{\alpha_i}) = \left(\bigwedge_{i=1}^{n}\lambda_i\right)\mathrm{IQD}_{\sum_{i=1}^{n}\alpha_i} \quad \text{and} \quad \operatorname*{\boxplus}_{i=1}^{n}(\lambda_i \mathrm{IQD}_{\alpha_i}) = \left(\bigwedge_{i=1}^{n}\lambda_i\right)\mathrm{IQD}_{\bigvee_{i=1}^{n}\alpha_i},$$

and their particular instances

$$\operatorname*{\square}_{i=1}^{n}\mathrm{IQD}_{\alpha_i} = \mathrm{IQD}_{\sum_{i=1}^{n}\alpha_i} \quad \text{and} \quad \operatorname*{\boxplus}_{i=1}^{n}\mathrm{IQD}_{\alpha_i} = \mathrm{IQD}_{\bigvee_{i=1}^{n}\alpha_i}.$$

These formulas may be compared with the quantile inf-convolutions formulas obtained by Embrechts et al. (2018) and Liu et al. (2022)

$$\operatorname*{\square}_{i=1}^{n}Q^{-}_{\alpha_i} = Q^{-}_{\sum_{i=1}^{n}\alpha_i}, \quad \operatorname*{\square}_{i=1}^{n}Q^{+}_{\alpha_i} = Q^{+}_{\sum_{i=1}^{n}\alpha_i}, \text{and} \quad \operatorname*{\boxplus}_{i=1}^{n}Q^{-}_{\alpha_i} = Q^{-}_{\bigvee_{i=1}^{n}\alpha_i} \quad \text{and} \quad \operatorname*{\boxplus}_{i=1}^{n}Q^{+}_{\alpha_i} = Q^{+}_{\bigvee_{i=1}^{n}\alpha_i}.$$

In the context of risk sharing, these results show that a representative agent (using the inf-convolution as its reference) of several IQD agents is again an IQD agent, and similarly, the representative agent of several quantile agents is again a quantile agent.

When the distortion functions are concave, or, when we constrain ourselves to the set of comonotonic allocations, the risk sharing problem becomes much more tractable, and we obtain explicit allocations which are Pareto optimal or sum optimal (Theorem 8.3). This builds on the comonotonic improvement *à la* Landsberger and Meilijson (1994), when the distortion riskmetrics are convex order consistent. A high-level summary is that all results that were established for increasing distortion riskmetrics, in particular, Yaari (1987)'s dual utilities, can be extended in parallel to non-increasing ones without extra efforts (these results are summarized in Propositions 8.6-8.8). This opens up various application areas where risks are traditionally studied with only increasing distortion riskmetrics.

Combining the results for IQD agents and for risk-averse agents, we are able to solve risk sharing problems among these agents, whose Pareto-optimal allocations are found explicitly (Theorem 8.4). Various examples of risk sharing among these agents are presented in Section 8.6.

It remains unclear to us whether our analysis can be generalized to distortion riskmetrics other than IQD, which are not convex (i.e., with non-concave distortion functions), and how large the class of such tractable risk functionals is. As far as we are aware, the unconstrained risk sharing problems for non-convex risk measures and variability measures have very limited explicit results (e.g., Embrechts et al. (2018), Weber (2018) and Liu et al. (2022)), and further investigation is needed for a better understanding of the challenges and their solutions.

8.8 Proofs of results in Section 8.3

Proof of Proposition 8.2. (i) Let $(X_1, \dots, X_n)$ be a Pareto-optimal allocation in $\mathbb{A}_n(X)$. We will show, without loss of generality, that any of the three following hypotheses leads to a contradiction of the Pareto optimality of $(X_1, \dots, X_n)$: (1) if simultaneously $h_1(1) = 0$ and $h_2(1) > 0$; (2) if simultaneously $h_1(1) < 0$ and $h_2(1) > 0$ and (3) if simultaneously $h_1(1) = 0$ and $h_2(1) < 0$.

Consider the allocation $(X_1 + c, X_2 - c, X_3, \dots, X_n)$. Clearly, the allocation belongs to $\mathbb{A}_n(X)$. Recall that by translation invariance it is $\rho_{h_1}(X_1 + c) = \rho_{h_1}(X_1) + ch_1(1)$ and $\rho_{h_2}(X_2 - c) = \rho_{h_2}(X_2) - ch_2(1)$.

Suppose (1) first so that $h_1(1) = 0$ and $h_2(1) > 0$. Setting $c > 0$ we have that $\rho_{h_1}(X_1 + c) = \rho h_1(X_1)$ and $\rho_{h_2}(X_2 - c) < \rho_{h_2}(X_2)$ contradicting the Pareto optimality of $(X_1, \dots, X_n)$. For (2), we have $\rho_{h_1}(X_1 + c) < \rho h_1(X_1)$ and $\rho_{h_2}(X_2 - c) < \rho_{h_2}(X_2)$ as $h_1(1) < 0$ and $h_2(1) > 0$. For (3), the case when $h_1(1) = 0$ and $h_2(1) < 0$, we can choose $c < 0$, which leads to a similar contradiction of the Pareto optimality of $(X_1, \dots, X_n)$.

The case when $(X_1, \dots, X_n)$ is Pareto optimal in $\mathbb{A}_n^+(X)$ is identical, and we conclude that $h_i(1)$ are either all zero, all positive, or all negative.

(ii) We show that if there exist $i, j \in [n]$ such that $h_i(1) \neq h_j(1)$, then $\boxplus_{i=1}^n \rho_{h_i}(X) = -\infty$

for any $X \in \mathcal{X}$. Without loss of generality, let $h_1(1) < h_2(1)$ and consider a $c > 0$. Given $X \in \mathcal{X}$, for any allocation $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ we have that

$$\rho_{h_1}(X_1 + c) + \rho_{h_2}(X_2 - c) = \rho_{h_1}(X_1) + \rho_{h_2}(X_2) + c(h_1(1) - h_2(1)).$$

Consider now the allocation $(X_1 + c, X_2 - c, X_3, \dots, X_n)$. Taking the limit $c \to \infty$ we have $\sum_{i=1}^n \rho_{h_i}(X_i) = -\infty$ and so $\boxplus_{i=1}^n \rho_{h_i}(X) = -\infty$. □

Proof of Theorem 8.1. For the "if" part, since every $\rho_{\tilde{h}_i}$, $i \in [n]$, are finite and $\sum_{i=1}^n \rho_{\tilde{h}_i}(X_i) = \square_{i=1}^n \rho_{\tilde{h}_i}(X)$ it holds that $\square_{i=1}^n \rho_{\tilde{h}_i}(X)$ is finite. It is thus clear that, by definition, the allocation $(X_1, \dots, X_n)$ is Pareto optimal for agents using $\rho_{\tilde{h}_1}, \dots, \rho_{\tilde{h}_n}$ as their preferences. We now show that $(X_1, \dots, X_n)$ is also Pareto optimal for agents using $\rho_{h_1}, \dots, \rho_{h_n}$ as their preferences. Since $\tilde{h}_i = h_i/|h_i(1)|$ for $i \in [n]$, we have $\rho_{h_i}(X) = |h_i(1)|\rho_{\tilde{h}_i}(X)$ for all $X \in \mathcal{X}$ and $i \in [n]$. Suppose $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ is such that $\rho_{h_i}(Y_i) \leqslant \rho_{h_i}(X_i)$ for all $i \in [n]$; we obtain that $\rho_{\tilde{h}_i}(Y_i) \leqslant \rho_{\tilde{h}_i}(X_i)$ for all $i \in [n]$. As $(X_1, \dots, X_n)$ is Pareto optimal for agents using $\rho_{\tilde{h}_1}, \dots, \rho_{\tilde{h}_n}$ as their preferences, it holds that $\rho_{\tilde{h}_i}(Y_i) = \rho_{\tilde{h}_i}(X_i)$. Hence, $\rho_{h_i}(Y_i) = \rho_{h_i}(X_i)$ for all $i \in [n]$, and $(X_1, \dots, X_n)$ is Pareto optimal for agents using $\rho_{h_1}, \dots, \rho_{h_n}$ as their preferences.

Next, we show the "only if" part. Let $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ be a Pareto-optimal allocation in $\mathbb{A}_n(X)$. By Proposition 8.2, we have $h_i(1)$, $i \in [n]$, are either all positive or all negative; that is $\tilde{h}_i(1)$, $i \in [n]$, are all 1 or -1. We first consider the case where $\tilde{h}_i(1) = 1$ for $i \in [n]$. Assume by contradiction that $\sum_{i=1}^n \rho_{\tilde{h}_i}(X_i) > \square_{i=1}^n \rho_{\tilde{h}_i}(X)$. There exists an allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ such that $\sum_{i=1}^n \rho_{\tilde{h}_i}(Y_i) < \sum_{i=1}^n \rho_{\tilde{h}_i}(X_i)$. Set $c_i = \rho_{\tilde{h}_i}(X_i) - \rho_{\tilde{h}_i}(Y_i)$, $i = 1, \dots, n$ and notice that $c = \sum_{i=1}^n c_i > 0$. Hence,

$$(Y_1 + c_1 - c/n, \dots, Y_n + c_n - c/n) \in \mathbb{A}_n(X)$$

and by translation invariance for every $i \in [n]$ it is

$$\rho_{\tilde{h}_i}(Y_i + c_i - c/n) = \rho_{\tilde{h}_i}(Y_i + c_i) - c/n < \rho_{\tilde{h}_i}(Y_i + c_i) = \rho_{\tilde{h}_i}(X_i),$$

contradicting the Pareto optimality of $(X_1, \dots, X_n)$. The case $\tilde{h}_i(1) = -1$, $i \in [n]$, is analogous. □

Proof of Lemma 8.2. The implications (i)⇒(ii)⇒(iii)⇒(iv) are all straightforward, where (iii)⇒(iv) follows from the fact that $X \leqslant_{\mathrm{cx}} Y$ is equivalent to $\rho_h(X) \leqslant \rho_h(Y)$ holding for all concave $h \in \mathcal{H}^{\mathrm{BV}}$ by Theorem 2 of Wang et al. (2020b).

We next show (iv)⇒(i). Suppose for the purpose of contradiction that h is strictly concave, $X \leqslant_{\mathrm{cx}} Y$, $\rho_h(X) = \rho_h(Y)$, and $X \overset{\mathrm{d}}{\neq} Y$. For $t \in (0,1)$ and $\epsilon > 0$ with $[t-\epsilon, t+\epsilon] \subseteq (0,1)$, let $Y_{t,\epsilon}$ be a random variable such that $Q^-_s(Y_{t,\epsilon}) = (2\epsilon)^{-1} \int_{t-\epsilon}^{t+\epsilon} Q^-_r(Y)\,\mathrm{d}r$ for $s \in [t-\epsilon, t+\epsilon]$, and $Q^-_s(Y_{t,\epsilon}) = Q^-_s(Y)$ otherwise. By construction, $Y_{t,\epsilon} \leqslant_{\mathrm{cx}} Y$.

We claim that there exist $t \in (0,1)$ and $\epsilon > 0$ such that $X \leqslant_{\mathrm{cx}} Y_{t,\epsilon} \overset{\mathrm{d}}{\neq} Y$. To see this, consider the function $\mu_Z : [0,1] \to \mathbb{R}$, $t \mapsto \int_0^t Q^-_s(Z)\,\mathrm{d}s$ for $Z \in \mathcal{X}$. Note that $X \leqslant_{\mathrm{cx}} Y$ if and only if $\mu_X \geqslant \mu_Y$ and $\mu_X(1) = \mu_Y(1)$; see e.g., Theorem 3.A.5 of Shaked and Shanthikumar (2007). Note that μ_X and μ_Y are continuous convex functions. Since $X \overset{\mathrm{d}}{\neq} Y$, we have $\mu_X(t) > \mu_Y(t)$ for some $t \in (0,1)$. Because μ_X is concave and $\mu_X(1) = \mu_Y(1)$, we can and will choose t such that μ_Y is not locally linear at t; this gives $Y_{t,\epsilon} \overset{\mathrm{d}}{\neq} Y$. Since μ_Y and μ_X are continuous and $\mu_X(t) > \mu_Y(t)$, there exists $\epsilon > 0$ small enough such that

$$\inf_{s\in[t-\epsilon,t+\epsilon]} \mu_X(s) > \sup_{s\in[t-\epsilon,t+\epsilon]} \mu_Y(s) + 4\epsilon M,$$

where $M = \sup_{s\in(t-\epsilon,t+\epsilon)} |Q^-_s(Y)|$. Using the above inequality and

$$|\mu_{Y_{t,\epsilon}} - \mu_Y| \leqslant \int_{t-\epsilon}^{t+\epsilon} |Q^-_s(Y_{t,\epsilon}) - Q^-_s(Y)|\,\mathrm{d}s \leqslant 4\epsilon M,$$

we get $\mu_X(s) > \mu_Y(s) + 4\epsilon M \geqslant \mu_{Y_{t,\epsilon}}(s)$ for $s \in [t-\epsilon, t+\epsilon]$. Moreover, $\mu_{Y_{t,\epsilon}}(s) = \mu_Y(s) \leqslant \mu_X(s)$ for $s \in [0,1] \setminus [t-\epsilon, t+\epsilon]$. Therefore, we get $X \leqslant_{\mathrm{cx}} Y_{t,\epsilon}$.

Note that $X \leqslant_{\mathrm{cx}} Y_{t,\epsilon} \leqslant_{\mathrm{cx}} Y$ implies $\rho_h(X) \leqslant \rho_h(Y_{t,\epsilon}) \leqslant \rho_h(Y)$, and further $\rho_h(X) = \rho_h(Y_{t,\epsilon}) = \rho_h(Y)$ since $\rho_h(X) = \rho_h(Y)$. Since h is concave, it is continuous on $[t-\epsilon, t+\epsilon] \subseteq (0,1)$. Using Lemma 3 of Wang et al. (2020b), we get

$$\rho_h(Y) - \rho_h(Y_{t,\epsilon}) = \int_{t-\epsilon}^{t+\epsilon} \left(Q^-_s(Y) - Q^-_s(Y_{t,\epsilon})\right) \mathrm{d}h(s) = \int_{t-\epsilon}^{t+\epsilon} \left(Q^-_s(Y) - Q^-_s(Y_{t,\epsilon})\right) h'(s)\,\mathrm{d}s,$$

where h' represents the right derivative of h. Since $Q^-_s(Y)$ is not a constant for $s \in [t-\epsilon, t+\epsilon]$, and h is strictly concave, by the Fréchet-Hoeffding inequality, we have

$$\int_{t-\epsilon}^{t+\epsilon} \left(Q^-_s(Y) - Q^-_s(Y_{t,\epsilon})\right) h'(s)\,\mathrm{d}s > \frac{1}{2\epsilon} \int_{t-\epsilon}^{t+\epsilon} \left(Q^-_s(Y) - Q^-_s(Y_{t,\epsilon})\right) \mathrm{d}s \int_{t-\epsilon}^{t+\epsilon} h'(s)\,\mathrm{d}s = 0.$$

Hence, $\rho_h(Y) - \rho_h(Y_{t,\epsilon}) > 0$, a contradiction to $\rho_h(Y_{t,\epsilon}) = \rho_h(Y)$. Therefore, (iv)⇒(i) holds. □

Proof of Proposition 8.5. (i) follows from Corollary 8.1 observing that comonotonic improvements strictly improve welfare. For (ii), the "only if" part is directly shown by (i). We only show the "if" part. As the normalization of h_i, $i \in [n]$, will not change the preferences, we only consider the case when $a_i = a_j = a$ for all $i, j \in [n]$. Let $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$. By comonotonic additivity and positive homogeneity it is $\sum_{i=1}^n \rho_{a_i h_1}(X_i) = a\rho_{h_1}(X)$. Let $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$. By subadditivity we have $\sum_{i=1}^n \rho_{a_i h_1}(Y_i) \geqslant a\rho_{h_1}\left(\sum_{i=1}^n Y_i\right) = a\rho_{h_1}(X)$. Hence, a comonotonic allocation $(X_1, \dots, X_n)$ always solves $\square_{i=1}^n \rho_{a_i h}(X)$, and thus it is Pareto optimal. □

Proof of Theorem 8.2. We first prove part (ii) and then use it to prove part (i). Let us first verify $\square_{i=1}^n(\lambda_i \mathrm{IQD}_{\alpha_i}) \geqslant \lambda \mathrm{IQD}_\alpha$. Using (8.3) and the fact that an IQD is non-negative, if $\alpha < 1/2$, then for $X \in \mathcal{X}$,

$$\begin{aligned}
\operatorname*{\square}_{i=1}^n (\lambda_i \mathrm{IQD}_{\alpha_i}) &\geqslant \lambda \operatorname*{\square}_{i=1}^n \mathrm{IQD}_{\alpha_i}(X) \\
&= \lambda \inf\left\{ \sum_{i=1}^n Q^-_{\alpha_i}(X_i) + \sum_{i=1}^n Q^-_{\alpha_i}(-X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n(X) \right\} \\
&\geqslant \lambda \operatorname*{\square}_{i=1}^n Q^-_{\alpha_i}(X) + \lambda \operatorname*{\square}_{i=1}^n Q^-_{\alpha_i}(-X) \\
&= \lambda Q^-_{\sum_{i=1}^n \alpha_i}(X) + \lambda Q^-_{\sum_{i=1}^n \alpha_i}(-X) = \lambda \mathrm{IQD}_\alpha(X),
\end{aligned}$$

where the second-last equality is due to Corollary 2 of Embrechts et al. (2018). If $\alpha \geqslant 1/2$, then $\square_{i=1}^n(\lambda_i \mathrm{IQD}_{\alpha_i}) \geqslant 0 = \lambda \mathrm{IQD}_\alpha$ holds automatically.

Next, we verify $\square_{i=1}^n(\lambda_i \mathrm{IQD}_{\alpha_i}) \leqslant \lambda \mathrm{IQD}_\alpha$ by showing that the construction of the allocation $(X_1, \dots, X_n)$ of $X \in \mathcal{X}$ in (8.7) satisfies $\sum_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}(X_i) = \lambda \mathrm{IQD}_\alpha(X)$. This will prove part (ii) as well as Remark 8.2. First, it is straightforward to verify $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$. Since IQD is location invariant, we can, without loss of generality, assume $c = c_1 = \cdots = c_n = 0$; i.e., 0 is a median of X. Note that this leads to the simplified form

$$X_i = X \mathbb{1}_{A_i \cup B_i} + a_i X \left(1 - \mathbb{1}_{A \cup B}\right), \qquad i \in [n].$$

If $\alpha \geqslant 1/2$, then it suffices to verify that $\mathrm{IQD}_{\alpha_i}(X_i) = 0$, which follows because $\mathbb{P}(X_i > 0) \leqslant \mathbb{P}(A_i) \leqslant \alpha_i$ and symmetrically, $\mathbb{P}(X_i < 0) \leqslant \mathbb{P}(B_i) \leqslant \alpha_i$.

Next, assume $\alpha < 1/2$. We have $\mathbb{P}(\{X > Q_\alpha^-(X)\} \cap A^c) = 0$ by Lemma A.3 of Wang and Zitikis (2021). For $i \in [n]$, we can compute

$$\mathbb{P}(X_i > a_i Q_\alpha^-(X)) \leqslant \mathbb{P}(A_i) + \mathbb{P}(\{X_i > a_i Q_\alpha^-(X)\} \setminus A_i) \leqslant \alpha_i + \mathbb{P}(\{X > Q_\alpha^-(X)\} \cap A^c) \leqslant \alpha_i.$$

This implies $Q_{\alpha_i}^-(X_i) \leqslant a_i Q_\alpha^-(X)$. Using a symmetric argument, we get $Q_{1-\alpha_i}^+(X_i) \geqslant a_i Q_{1-\alpha}^+(X)$. It follows that

$$\mathrm{IQD}_{\alpha_i}(X_i) = a_i Q_\alpha^-(X) - a_i Q_{1-\alpha}^+(X) \leqslant a_i \mathrm{IQD}_\alpha(X).$$

Therefore, $\sum_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}(X_i) \leqslant \sum_{i=1}^n \lambda_i a_i \mathrm{IQD}_\alpha(X)$. Taking $a_i = 0$ for all $i \in [n]$ with $\lambda_i > \lambda$ gives the desired inequality $\sum_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}(X_i) \leqslant \lambda \mathrm{IQD}_\alpha(X)$.

Putting the above arguments together, we prove (ii), that is, $\square_{i=1}^n \lambda_i \mathrm{IQD}_{\alpha_i}(X) = \lambda \mathrm{IQD}_\alpha(X)$. In particular,

$$\mathrm{IQD}_{\alpha_i}(X_i) = a_i \mathrm{IQD}_\alpha(X) \quad \text{and} \quad \sum_{i=1}^n \mathrm{IQD}_{\alpha_i}(X_i) = \mathrm{IQD}_\alpha(X) = \operatorname*{\square}_{i=1}^n \mathrm{IQD}_{\alpha_i}(X), \tag{8.13}$$

and thus $(X_1, \dots, X_n)$ is sum optimal.

Next, we show part (i). The "if" statement follows from Proposition 8.3, and we will show the "only if" statement. Take any Pareto-optimal allocation $(Y_1, \dots, Y_n)$ of X. Write $x = \mathrm{IQD}_\alpha(X)$, $y_i = \mathrm{IQD}_{\alpha_i}(Y_i)$ for $i \in [n]$, and $y = \sum_{i=1}^n y_i$. It suffices to show $y = x$. If $y = 0$, there is nothing to show; next we assume $y > 0$. For the allocation $(X_1, \dots, X_n)$ in (8.7), we have $\mathrm{IQD}_{\alpha_i}(X_i) = a_i \mathrm{IQD}_\alpha(X) = a_i x$ by (8.13). Let $a_i = y_i/y$ for $i \in [n]$, which sums up to 1. If $x < y$, then $\mathrm{IQD}_{\alpha_i}(X_i) = x y_i / y \leqslant y_i = \mathrm{IQD}_{\alpha_i}(Y_i)$ for $i \in [n]$, and strict inequality holds as soon as $y_i > 0$, conflicting Pareto optimality of $(Y_1, \dots, Y_n)$. Hence, we obtain $x = y$.

Finally, part (iii) on Pareto optimality of $(X_1, \dots, X_n)$ follows by combining (i) and (8.13). □

8.9 Proofs of results in Section 8.4

Proof of Proposition 8.7. (i) ⇒ (ii) is analogous to Theorem 8.1.

(ii) $\Rightarrow$ (iii) Let $S = \{(\rho_{h_1}(X_1), \dots, \rho_{h_n}(X_n) : (X_1, \dots, X_n) \in \mathbb{A}_n^+(X)\}$ be the utility possibility frontier of the set of comonotonic allocations. We claim that S is a convex set. First, notice that $\mathbb{A}_n^+(X)$ is a convex set, as for any two allocations $\mathbf{X} = (X_1, \dots, X_n), \mathbf{Y} = (Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X)$ and every $\xi \in [0,1]$ we have $\xi\mathbf{X} + (1-\xi)\mathbf{Y} \in \mathbb{A}_n^+(X)$. Set $\mathbf{x} = (\rho_{h_1}(X_1), \dots, \rho_{h_n}(X_n)) \in S$ and $\mathbf{y} = (\rho_{h_1}(Y_1), \dots, \rho_{h_n}(Y_n)) \in S$ two vectors of utility achieved by allocation $\mathbf{X}$ and $\mathbf{Y}$. By comonotonic additivity and positive homogeneity for every $\xi \in [0,1]$ and for every $i \in [n]$ it is

$$\begin{aligned}\rho_{h_i}(\xi X_i + (1-\xi)Y_i) &= \rho_{h_i}(\xi X_i) + \rho_{h_i}((1-\xi)Y_i) \\ &= \xi\rho_{h_i}(X_i) + (1-\xi)\rho_{h_i}(Y_i) \\ &= \xi x_i + (1-\xi)y_i\end{aligned}$$

and $\xi\mathbf{x} + (1-\xi)\mathbf{y} \in S$. Notice now that the utility vector $(\rho_{h_1}(X_1), \dots, \rho_{h_n}(X_n))$ of a Pareto-optimal allocation always belongs to the boundary of S.

Let $V = \{(v_1, \dots, v_n) : v_i \leqslant \rho_{h_i}(X_i) \text{ for } i \in [n]\}$ where $(X_1, \dots, X_n)$ is Pareto optimal. It is clear that V is a non-empty convex set. Next, we clarify that $V \cap S = \{\mathbf{x}\}$. Assume $\mathbf{v} = (v_1, \dots, v_n) \in V \cap S$. As $\mathbf{v} \in S$, there exists an allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X)$ such that $\rho_{h_i}(Y_i) = v_i$ for all $i \in [n]$. Furthermore, as $\mathbf{v} \in V$, we have $\rho_{h_i}(Y_i) = v_i \leqslant \rho_{h_i}(X_i)$ for all $i \in [n]$. As $(X_1, \dots, X_n)$ is a Pareto-optimal allocation, we get $v_i = \rho_{h_i}(Y_i) = \rho_{h_i}(X_i) = x_i$ for all $i \in [n]$. Hence, $\mathbf{v} = \mathbf{x}$ and $V \cap S = \{\mathbf{x}\}$.

Therefore, by the Separating Hyperplane Theorem, there exists $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \setminus \mathbf{0}$ such that $\sum_{i=1}^n \lambda_i \rho_{h_i}(X_i) = \inf_{\mathbf{x} \in S} \sum_{i=1}^n \lambda_i x_i = \inf_{\mathbf{X} \in \mathbb{A}_n^+(X)} \sum_{i=1}^n \lambda_i \rho_{h_i}(X_i)$ and $\sum_{i=1}^n \lambda v_i \leqslant \sum_{i=1}^n \lambda_i \rho_{h_i}(X_i)$ for any $(v_1, \dots, v_n) \in V$.

We are left to show that $\lambda_i \geqslant 0$ for every $i \in [n]$. Let $\mathbf{v} = \mathbf{x} - (1, 0, \dots, 0)$. We have $\mathbf{v} \in V$. Hence, we have $\lambda_1 \geqslant 0$ as $\sum_{i=1}^n \lambda v_i \leqslant \sum_{i=1}^n \lambda_i \rho_{h_i}(X_i)$. Similarly, we obtain $\lambda_i \geqslant 0$ for all $i \in [n]$. □

Proof of Theorem 8.3. We first show that $\boxplus_{i=1}^n \rho_{h_i} = \rho_{h_\wedge}$. Let $h_\wedge(t) = \min\{h_1(t), \dots, h_n(t)\}$. For every $i \in [n]$, we have $h_i(1) = h_\wedge(1)$ and $h_i \leqslant h_\wedge$ on $[0,1]$; hence, it is $\rho_{h_\wedge}(X) \leqslant \rho_{h_i}(X)$ for every $X \in \mathcal{X}$. By comonotonic additivity of $\rho_{h_\wedge}$, for every $(X_1, \dots, X_n)$ in $\mathbb{A}_n^+(X)$ we

have

$$\sum_{i=1}^{n} \rho_{h_i}(X_i) \geqslant \sum_{i=1}^{n} \rho_{h_\wedge}(X_i) = \rho_{h_\wedge}\left(\sum_{i=1}^{n} X_i\right) = \rho_{h_\wedge}(X)$$

and thus $\boxplus_{i=1}^{n} \rho_{h_i} \geqslant \rho_{h_\wedge}$. Conversely, notice that for every $i \in [n]$ the function f_i in (8.8) is Lipschitz continuous and non-decreasing because g_i is non-negative and bounded. Using Lemma 8.3, we get

$$\rho_{h_i}(f_i(X)) = \int_0^\infty g_i(s) h_i(\mathbb{P}(X > s))\,\mathrm{d}s + \int_{-\infty}^0 g_i(s)(h_i(\mathbb{P}(X > s) - h_i(1))\,\mathrm{d}s. \tag{8.14}$$

It follows that

$$\begin{aligned}
\sum_{i=1}^{n} \rho_{h_i}(f_i(X)) &= \sum_{i=1}^{n} \int_0^\infty g_i(s) h_i(\mathbb{P}(X > s))\,\mathrm{d}s + \int_{-\infty}^0 g_i(s)(h_i(\mathbb{P}(X > s) - h_i(1))\,\mathrm{d}s \\
&= \int_0^\infty \sum_{i=1}^{n} g_i(s) h_i(\mathbb{P}(X > s))\,\mathrm{d}s + \int_{-\infty}^0 \sum_{i=1}^{n} g_i(s)(h_i(\mathbb{P}(X > s) - h_i(1))\,\mathrm{d}s \\
&= \int_0^\infty h_\wedge(\mathbb{P}(X > s))\,\mathrm{d}s + \int_{-\infty}^0 (h_\wedge(\mathbb{P}(X > s) - h_\wedge(1))\,\mathrm{d}s \\
&= \rho_{h_\wedge}(X) \geqslant \boxplus_{i=1}^{n} \rho_{h_i}(X).
\end{aligned}$$

Hence, $\boxplus_{i=1}^{n} \rho_{h_i} = \rho_{h_\wedge}$.

Next, we show that the solution is unique up to constant shifts almost surely if and only if $|M_x| = 1$ for μ_X-almost every x, where μ_X is the distribution measure of X.

Since the above argument of $\sum_{i=1}^{n} \rho_{h_i}(f_i(X)) = \boxplus_{i=1}^{n} \rho_{h_i}(X)$ only requires $\sum_{i \in M_x} g_i(x) = 1$ for almost every x, any allocation $(f_1(X), \dots, f_n(X))$ in (8.8) with g_i replaced by

$$g_i(x) = \mathbb{1}_{\{i = \min M_x\}} \text{ or } g_i(x) = \mathbb{1}_{\{i = \max M_x\}}, \qquad x \in \mathbb{R},$$

also satisfies sum optimality. Therefore, if $|M_x| = 1$ does not hold almost surely, there are multiple optimal allocations that are not constant shifts from each other.

Conversely, we show that if $|M_x| = 1$ for μ_X-almost every x then every sum-optimal allocation is almost surely equal to the one in (8.8).

For any increasing and Lipschitz function k with right-derivative w, we have, by Lemma 8.3,

$$\rho_h(k(X)) - \rho_g(k(X)) = \int_{-\infty}^\infty w(s)(h(\mathbb{P}(X > s)) - g(\mathbb{P}(X > s)))\,\mathrm{d}s.$$

This means $\rho_h(k(X)) = \rho_g(k(X))$ with $h \geqslant g$ if and only if $k'(s) = 0$ almost surely for s such that $h(\mathbb{P}(X > s)) > g(\mathbb{P}(X > s))$. Note that if $(k_1(X), \dots, k_n(X)) \in \mathbb{A}_n^+(X)$ is sum optimal, then

$$\sum_{i=1}^{n} \rho_{h_i}(k_i(X)) = \rho_{h_\wedge}(X) = \sum_{i=1}^{m} \rho_{h_\wedge}(k_i(X)).$$

This implies that $w_i(x) = 0$ as soon as $h_i(\mathbb{P}(X > x)) > h_\wedge(\mathbb{P}(X > x))$, where w_i is the right-derivative of k_i. Moreover, $w_i(x) = 1$ if $h_i(\mathbb{P}(X > x)) = h_\wedge(\mathbb{P}(X > x))$ since $\sum_{j=1}^{n} w_j(x) = 1$ for almost every x. and thus w_i is uniquely determined μ_X-a.s., implying that k_i is unique μ_X-a.s. up to a constant shift. □

Proof of Lemma 8.3. Without loss of generality we assume $X \geqslant 0$ and $f(X) \geqslant 0$. Denote by $\nu = h \circ \mathbb{P}$. We have

$$\begin{aligned}
\rho_h(f(X)) - \int_0^\infty g(x)h(\mathbb{P}(X > x))\,\mathrm{d}x &= \int_0^\infty \nu(f(X) > y)\,\mathrm{d}y - \int_0^\infty g(x)\nu(X > x)\,\mathrm{d}x \\
&= \int_0^\infty g(x)\nu(f(X) > f(x))\,\mathrm{d}x - \int_0^\infty g(x)\nu(X > x)\,\mathrm{d}x \\
&= \int_0^\infty g(x)(\nu(f(X) > f(x)) - \nu(X > x))\,\mathrm{d}x.
\end{aligned}$$

Note that $\mathbb{P}(f(X) > f(x)) \leqslant \mathbb{P}(X > x)$ for all x. If $\mathbb{P}(f(X) > f(x)) < \mathbb{P}(X > x)$, then there exists $z > x$ such that $f(z) = f(x)$. This implies that $g(x) = 0$ for any point x with $\nu(f(X) > f(x)) - \nu(X > x) \neq 0$. Therefore,

$$\rho_h(f(X)) - \int_0^\infty g(x)h(\mathbb{P}(X > x))\,\mathrm{d}x = 0.$$

The case of general X bounded from below can be obtained by constant shifts on both X and f. □

Proof of Proposition 8.8. It is clear that since $h_i(1) = 0$ and $h_i(t) > 0$ for all $i \in [n]$ and all $t \in (0, 1)$, we have that $\rho_{h_i}(X) \geqslant 0$ for all $i \in [n]$, with equality only if X is a constant. We first show the "if" statement. Suppose, by contradiction, that $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ is not Pareto optimal but that it solves $\boxplus_{i \in K} \rho_{\lambda_i h_i}(X - \sum_{i \notin K} X_i)$ for $K = \{i \in [n] : X_i \notin \mathbb{R}\}$ and $\boldsymbol{\lambda} \in (0, \infty)^{\#K}$. Our contradiction hypothesis implies that there exists a $(Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X)$ such that simultaneously $\rho_{h_i}(Y_i) \leqslant \rho_{h_i}(X_i)$ for every $i \in [n]$ and $\rho_{h_j}(Y_j) < \rho_{h_j}(X_j)$ for some $j \in [n]$. Notice that if $i \notin K$ it is

$$0 \leqslant \rho_{h_i}(Y_i) \leqslant \rho_{h_i}(X_i) = 0$$

and so it must be the case that $\rho_{h_i}(Y_i) < \rho_{h_i}(X_i)$ for some $i \in K$, a contradiction with the hypothesis that $(X_i)_{i\in K}$ solves $\boxplus_{i\in K} \rho_{\lambda_i h_i}(X - \sum_{i\notin K} c_i) = \boxplus_{i\in K} \rho_{\lambda_i h_i}(X)$, where the equality follows because of location invariance of $\boxplus_{i\in K} \rho_{\lambda_i h_i}$.

Conversely, let $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ be Pareto optimal and define $K = \{i \in [n] : X_i \notin \mathbb{R}\}$; this gives that $\sum_{i\notin K} X_i$ is a constant. Recall that $\rho_{h_i}(X_i) = 0$ for every $i \notin K$, and $\rho_{h_i}(X_i) > 0$ for every $i \in K$. It is clear that $(X_i)_{i\in K}$ is a Pareto-optimal allocation of $X - \sum_{i\notin K} X_i$ for the collection $(\rho_{h_i})_{i\in K}$. By Proposition 8.7, there exists a $\boldsymbol{\lambda} \in [0,\infty)^{\#K} \setminus \{\mathbf{0}\}$ such that $\sum_{i\in K} \lambda_i \rho_{h_i}(X_i) = \boxplus_{i\in K}(\lambda_i \rho_{h_i})(X - \sum_{j\notin K} X_j) = \boxplus_{i\in K} \rho_{\lambda_i h_i}(X)$. As $\rho_{h_i}(X_i) > 0$ for $i \in K$, we have $\boxplus_{i\in K}(\lambda_i \rho_{h_i})(X) > 0$. It must be the case that $\lambda_i > 0$ for all $i \in K$, as otherwise, we have $\boxplus_{i\in K}(\lambda_i \rho_{h_i})(X) = 0$, a contradiction. □

Proof of Proposition 8.9. Part (ii) follows directly from Corollary 8.2, so it remains to show part (i). Let $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ be Pareto optimal. Then there exists $(\lambda_1, \dots, \lambda_n) \in [0,\infty)^n$, with $\lambda = \bigwedge_{i=1}^n \lambda_i > 0$, such that

$$\sum_{i=1}^n (\lambda_i \mathrm{IQD}_{\alpha_i})(X_i) = \lambda \mathrm{IQD}_{\bigvee_{i=1}^n \alpha_i}(X).$$

Using the fact that an IQD is non-negative and part (ii), we get

$$\lambda \boxplus_{i=1}^n \mathrm{IQD}_{\alpha_i}(X) \leqslant \sum_{i=1}^n (\lambda \mathrm{IQD}_{\alpha_i})(X_i) \leqslant \sum_{i=1}^n (\lambda_i \mathrm{IQD}_{\alpha_i})(X_i) = \lambda \mathrm{IQD}_{\bigvee_{i=1}^n \alpha_i}(X) = \lambda \boxplus_{i=1}^n \mathrm{IQD}_{\alpha_i}(X),$$

and so $(X_1, \dots, X_n)$ is sum optimal. □

8.10 Proofs of results in Section 8.5

We first present a lemma that we will use in the proof of Theorem 8.4.

Lemma 8.4. *For $\alpha \in [0, 1/2)$, $\lambda > 0$ and $h \in \mathcal{H}^{\mathrm{C}}$ it is*

$$(\lambda \mathrm{IQD}_{\alpha}) \square \rho_h = \rho_{G_\lambda^\alpha(h)}. \tag{8.15}$$

Proof of Lemma 8.4. We first verify that $\lambda \mathrm{IQD}_\alpha(X_1) + \rho_h(X_2) \geqslant \rho_{G_\lambda^\alpha(h)}(X)$ for any $(X_1, X_2) \in \mathbb{A}_2(X)$. As both IQD_α and ρ_h are location invariant, we can, without loss of generality, assume the allocation (X_1, X_2) satisfies $Q^-_{1/2}(X_1) = 0$. Let A be a right α-tail event of X_1 and

$B \subseteq A^c$ be a left α-tail event of X_1. Hence, $\mathbb{P}(A) = \mathbb{P}(B) = \alpha$ and $X_1(\omega_B) \leqslant X_1(\omega) \leqslant X_1(\omega_A)$ for a.s. $\omega_A \in A$, $\omega_B \in B$ and $\omega \in (A \cup B)^c$. Let $X_1^* = X_1 \mathbb{1}_{\{(A\cup B)^c\}}$ and $h^* = h \wedge \lambda$. Recall that $\mathrm{IQD}_0 = Q_0^- - Q_1^+$ is the range functional. It is straightforward to verify that $\mathrm{IQD}_\alpha(X_1) = \mathrm{IQD}_0(X_1^*)$ and that h^* is concave. Further, notice that $\lambda \mathrm{IQD}_0 \geqslant \rho_{h^*}$, $\rho_h \geqslant \rho_{h^*}$ and ρ_{h^*} is subadditive. Therefore,

$$\lambda \mathrm{IQD}_\alpha(X_1) + \rho_h(X_2) = \lambda \mathrm{IQD}_0(X_1^*) + \rho_h(X_2) \geqslant \rho_{h^*}(X_1^*) + \rho_{h^*}(X_2) \geqslant \rho_{h^*}\left(X_1^* + X_2\right).$$

As $Q_{1/2}^-(X_1) = 0$, we have, in the a.s. sense, $X_1 \geqslant 0$ on A and $X_1 \leqslant 0$ on B; that is, $X_1^* + X_2 = X$ on $(A \cup B)^c$, $X_1^* + X_2 \geqslant X$ on B, and $X_1^* + X_2 \leqslant X$ on A. For any $x \in \mathbb{R}$, we have

$$\begin{aligned}\mathbb{P}\left(X_1^* + X_2 > x\right) &\geqslant \mathbb{P}\left(X > x,\, (A \cup B)^c\right) + \mathbb{P}\left(X > x,\, B\right) \\ &\geqslant \mathbb{P}\left(X > x\right) - \mathbb{P}\left(A\right) = \mathbb{P}\left(X > x\right) - \alpha,\end{aligned}$$

and similarly, $\mathbb{P}\left(X_1^* + X_2 \leqslant x\right) \geqslant \mathbb{P}\left(X \leqslant x\right) - \alpha$. Therefore,

$$\mathbb{P}\left(X > x\right) - \alpha \leqslant \mathbb{P}\left(X_1^* + X_2 > x\right) \leqslant \mathbb{P}\left(X > x\right) + \alpha.$$

Let $s \in \mathbb{R}$ be such that $x \mapsto h^*(\mathbb{P}(X_1^* + X_2 > x))$ is increasing on $(-\infty, s]$ and decreasing on $[s, \infty)$. Such s exists since h^* is first increasing and then decreasing. By treating $h^*(t) = 0$ if t is outside $[0, 1]$, we have

$$\begin{aligned}\rho_{h^*}(X_1^* + X_2) &= \int_{-\infty}^{s} h^*(\mathbb{P}(X_1^* + X_2 > x))\,\mathrm{d}x + \int_{s}^{\infty} h^*(\mathbb{P}(X_1^* + X_2 > x))\,\mathrm{d}x \\ &\geqslant \int_{-\infty}^{s} h^*(\mathbb{P}(X > x) - \alpha)\,\mathrm{d}x + \int_{s}^{\infty} h^*(\mathbb{P}(X > x) + \alpha)\,\mathrm{d}x \\ &\geqslant \int_{-\infty}^{\infty} \min\left\{h^*\left(\mathbb{P}(X > x) + \alpha\right), h^*(\mathbb{P}(X > x) - \alpha)\right\}\,\mathrm{d}x = \rho_{G_\lambda^\alpha(h)}(X).\end{aligned}$$

Therefore, we have $\lambda \mathrm{IQD}_\alpha(X_1) + \rho_h(X_2) \geqslant \rho_{G_\lambda^\alpha(h)}(X)$.

Next, we give an allocation $(X_1, X_2) \in \mathbb{A}_2(X)$ that attains the lower bound $\rho_{G_\lambda^\alpha(h)}(X)$. Define the function $f(s) = h^*(\mathbb{P}(X > x) + \alpha) - h^*(\mathbb{P}(X > x) - \alpha)$ where $h^*(t) = 0$ if t is outside $[0, 1]$. Since h^* is concave, the function $s \mapsto f(s)$ is increasing on the set of s with $\mathbb{P}(X > s) \in [\alpha, 1 - \alpha]$. Moreover, $f(s) \leqslant 0$ for $s \leqslant Q_{1-\alpha}^-(X)$ and $f(s) \geqslant 0$ for $s \geqslant Q_\alpha^+(X)$. Hence, there exists $s^* \in [Q_{1-\alpha}^-(X), Q_\alpha^+(X)]$ such that $f(s) \geqslant 0$ for $s < s^*$ and $f(s) \leqslant 0$ for $s > s^*$.

Let A be a right α-tail event of X and $B \subseteq A^c$ be a left α-tail event of X. Write $T = A \cup B$. Let $(Y_1, Y_2) \in \mathbb{A}_2^+(X\mathbb{1}_{T^c} + s^*\mathbb{1}_T)$ be a $(\lambda, 1)$-optimal allocation for $(\mathrm{IQD}_\alpha, \rho_h)$. Define $X_1 = (X - s^*)\mathbb{1}_T + Y_1$ and $X_2 = Y_2$; clearly $(X_1, X_2) \in \mathbb{A}_2(X)$. By Theorem 8.3, we have

$$\lambda \mathrm{IQD}_\alpha(X_1) + \rho_h(X_2) = \lambda \mathrm{IQD}_0(Y_1) + \rho_h(Y_2) = \rho_{h^*}(X\mathbb{1}_{T^c} + s^*\mathbb{1}_T).$$

Note that

$$\begin{aligned}
\rho_{h^*}(X\mathbb{1}_{T^c} + s^*\mathbb{1}_T) &= \int_{Q^-_{1-\alpha}(X)}^{Q^+_\alpha(X)} h^*(\mathbb{P}(X\mathbb{1}_{T^c} + s^*\mathbb{1}_T > x))\,\mathrm{d}x \\
&= \int_{Q^-_{1-\alpha}(X)}^{s^*} h^*(\mathbb{P}(X > x,\ T^c) + 2\alpha)\,\mathrm{d}x + \int_{s^*}^{Q^+_\alpha(X)} h^*(\mathbb{P}(X > x,\ T^c))\,\mathrm{d}x \\
&= \int_{Q^-_{1-\alpha}(X)}^{s^*} h^*(\mathbb{P}(X > x) + \alpha)\,\mathrm{d}x + \int_{s^*}^{Q^+_\alpha(X)} h^*(\mathbb{P}(X > x) - \alpha)\,\mathrm{d}x \\
&= \int_{Q^-_{1-\alpha}(X)}^{Q^+_\alpha(X)} \min\{h^*(\mathbb{P}(X > x) + \alpha), h^*(\mathbb{P}(X > x) - \alpha)\}\,\mathrm{d}x = \rho_{G^\alpha_\lambda(h)}(X),
\end{aligned}$$

where the second-last equality is due to the definition of s^*. Therefore, the lower bound $\rho_{G^\alpha_\lambda(h)}(X)$ can be attained. Thus, $(\lambda\mathrm{IQD}_\alpha)\square\rho_h = \rho_{G^\alpha_\lambda(h)}(X)$. □

Proof of Theorem 8.4. As the cases $I = [n]$ and $S = [n]$ follow from Theorems 8.2 and 8.3 respectively, we assume that the sets I and S are non-empty.

(i) The equality $\square_{i=1}^n(\lambda_i\rho_{h_i}) = \rho_{G^\alpha_\lambda(h)}$ follows from Lemma 8.4, Theorems 8.2 and 8.3, and the fact that the inf-convolution is associative (Lemma 2 of Liu et al. (2020)), which together yield

$$\square_{i=1}^n(\lambda_i\rho_{h_i}) = \left(\square_{i\in I}(\lambda_i\rho_{h_i})\right) \square \left(\square_{i\in S}(\lambda_i\rho_{h_i})\right) = (\lambda\mathrm{IQD}_\alpha)\square\rho_h = \rho_{G^\alpha_\lambda(h)}.$$

(ii) Without loss of generality, we assume $c = c_1 = \cdots = c_n = 0$ and let $Y = X\mathbb{1}_{(A\cup B)^c}$. If $\alpha \geqslant 1/2$, it is straightforward to check that $(X_1, \dots, X_n)$ is Pareto optimal as $\rho_{h_i}(X_i) = 0$ for $i \in [n]$. Now, we assume $\alpha < 1/2$.

We first show that $\rho_{h_i}(X_i) \leqslant \rho_{h_i'}(Y_i)$ for all $i \in [n]$. Note that $\rho_{h_i}(X_i) = \rho_{h_i}(Y_i) = \rho_{h_i'}(Y_i)$ for all $i \in C$. We are left to show $\mathrm{IQD}_{\alpha_i}(X_i) \leqslant \mathrm{IQD}_0(Y_i)$ for all $i \in I$. As $X(\omega) \leqslant 0$ a.s. for $\omega \in B_i$,

$$\mathbb{P}(X_i \leqslant Q_0^-(Y_i)) = \mathbb{P}(X\mathbb{1}_{A_i\cup B_i} + Y_i \leqslant Q_0^-(Y_i)) \geqslant \mathbb{P}(B_i) + \mathbb{P}((A_i\cup B_i)^c) = \alpha_i + 1 - 2\alpha_i = 1 - \alpha_i.$$

That is, $Q^-_{\alpha_i}(X_i) \leqslant Q^-_0(Y_i)$. Similarly, $Q^+_{1-\alpha_i}(X_i) \geqslant Q^+_1(Y_i)$. Hence, $\rho_{h_i}(X) = \mathrm{IQD}_{\alpha_i}(X_i) \leqslant \mathrm{IQD}_0(Y_i) = \rho_{h'_i}(Y_i)$ for all $i \in I$.

Let $(Y'_1, \dots, Y'_n)$ be a comonotonic improvement of $(Y_1, \dots, Y_n)$. The definition of comonotonic improvement and Pareto optimality of $(Y_1, \dots, Y_n)$ imply that $\rho_{h'_i}(Y_i) = \rho_{h'_i}(Y'_i)$ for all $i \in [n]$. First, if there exists some $i \in C$ such that $h_i(t) = 0$ on $[0, 1]$, then Pareto optimality of $(Y'_1, \dots, Y'_n)$ implies that $\rho_{h'_i}(Y'_i) = 0$ for each $i \in [n]$. This in turn implies that $\rho_{h_i}(X_i) = 0$ for each $i \in [n]$, and hence $(X_1, \dots, X_n)$ is Pareto optimal. Below, we assume for each $i \in C$, $h_i(t) > 0$ for some $t \in (0, 1)$, which gives that $h_i(t) > 0$ for all $t \in (0, 1)$ due to concavity.

As $h'_i(1) = 0$ and $h'_i(t) > 0$ for all $i \in [n]$ and $t \in (0, 1)$, by Proposition 8.8, Pareto optimality of $(Y'_1, \dots, Y'_n)$ implies that there exist $K \subseteq [n]$ and a vector $\boldsymbol{\lambda} \in (0, \infty)^{\#K}$ such that $(Y'_i)_{i \in K}$ solves $\boxplus_{i \in K} \rho_{\lambda_i h'_i}(Y)$, and Y'_i, $i \notin K$ are constants. Denote by $h^* = \bigwedge_{i \in C \cap K}(\lambda_i h_i)$ and $\lambda^* = \bigwedge_{i \in I \cap K} \lambda_i > 0$; here, we set $\inf \varnothing = \infty$. Putting together several observations above, we get

$$\sum_{i \in K} \lambda_i \rho_{h_i}(X_i) \leqslant \sum_{i \in K} \rho_{\lambda_i h'_i}(Y_i) = \sum_{i \in K} \rho_{\lambda_i h'_i}(Y'_i) = \boxplus_{i \in K} \rho_{\lambda_i h'_i}(Y) = \rho_{h^* \wedge \lambda^*}(Y), \tag{8.16}$$

where the first inequality holds because $\rho_{h_i}(X_i) \leqslant \rho_{h'_i}(Y_i)$ for all $i \in [n]$, the first equality holds because $\rho_{h'_i}(Y_i) = \rho_{h'_i}(Y'_i)$ for all $i \in [n]$, the second equality is due to $\boldsymbol{\lambda}$-optimality of $(Y'_i)_{i \in K}$ whose component-wise sum is Y plus a constant, and the last equality is due to Theorem 8.3. Furthermore, for $i \notin K$, we have $0 \leqslant \rho_{h_i}(X_i) \leqslant \rho_{h'_i}(c_i) = 0$; that is $\rho_{h_i}(X_i) = 0$. Note that

$$\rho_{h^* \wedge \lambda^*}(Y) = \rho_{h^* \wedge \lambda^*}(X \mathbb{1}_{(A \cup B)^c}) = \rho_{G^\alpha_{\lambda^*}(h^*)}(X). \tag{8.17}$$

Take $\beta \geqslant \lambda^*$. If $i \in C \setminus K$, then $X_i = Y'_i$ is a constant. Write $Z = \sum_{i \in I \cup K} X_i$. Using (8.16) and (8.17), we get

$$\sum_{i \in K} \lambda_i \rho_{h_i}(X_i) + \sum_{i \in I \setminus K} \beta \rho_{h_i}(X_i) \leqslant \rho_{G^\alpha_{\lambda^*}(h^*)}(X) = \rho_{G^\alpha_{\lambda^*}(h^*)}\left(X - \sum_{i \in C \setminus K} X_i\right) = \rho_{G^\alpha_{\lambda^*}(h^*)}(Z). \tag{8.18}$$

Using part (i), we have

$$\left(\mathop{\square}_{i \in K} (\lambda_i \rho_{h_i})\right) \square \left(\mathop{\square}_{i \in I \setminus K} (\beta \rho_{h_i})\right) = \rho_{G^\alpha_{\lambda^*}(h^*)}.$$

Therefore, (8.18) implies that $(X_i)_{i\in I\cup K} \in \mathbb{A}_n(Z)$ minimizes $\sum_{i\in K} \lambda_i \rho_{h_i}(X_i) + \sum_{i\in I\setminus K} \beta \rho_{h_i}(X_i)$. Since also $\rho(X_i) = 0$ for $i \notin K$, we conclude that $(X_1, \dots, X_n)$ is Pareto optimal. □

8.11 Heterogeneous beliefs in comonotonic risk sharing

We considered throughout an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This assumption entails that every individual $i \in [n]$ agrees on the fundamentals of the risk to be shared. We explain in this section that all our results on comonotonic risk sharing can be extended to incorporate heterogeneous beliefs with almost no extra effort; this is not true for the unconstrained setting of risk sharing in Section 8.3. Our characterization of comonotonic risk sharing extends the main results of Liu (2020), which focus on dual utilities. See also Embrechts et al. (2020), Boonen and Ghossoub (2020) and Liebrich (2021) for risk sharing with risk measures and heterogeneous beliefs.

Let $(\Omega, \mathcal{F})$ be a measurable space that allows for atomless probability measures and denote by $\mathbb{P}_i$ the atomless probability measure that agent $i \in [n]$ considers. That is, every individual $i \in [n]$ believes the probability space $(\Omega, \mathcal{F}, \mathbb{P}_i)$ is the true one. Let $\mathcal{P}$ be the set of atomless probability measures on the measurable space $(\Omega, \mathcal{F})$ and let $\ll$ denote absolute continuity. As before, every individual evaluates their risk with the distortion riskmetric

$$\rho_{h_i}^{\mathbb{P}_i}(X) = \int X \, \mathrm{d}\left(h_i \circ \mathbb{P}_i\right).$$

For a probability measure $\mathbb{P}$, we define the corresponding left quantile as $Q_t^{\mathbb{P}}(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant 1 - t\}$.

The next lemma is instrumental in proving this section's main result:

Lemma 8.5. *Let $\mathbb{P}_0, \mathbb{P} \in \mathcal{P}$ be such that $\mathbb{P}_0 \ll \mathbb{P}$, let $h \in \mathcal{H}^{\mathrm{BV}}$ and let $X \in \mathcal{X}$ admits a density under $\mathbb{P}$. The function $g(t) = h(\mathbb{P}_0(X > Q_t^{\mathbb{P}}(X)))$, $t \in [0,1]$, satisfies $\rho_h^{\mathbb{P}_0}(f(X)) = \rho_g^{\mathbb{P}}(f(X))$ for any increasing functions $f : \mathbb{R} \to \mathbb{R}$.*

Proof of Lemma 8.5. Let $g(t) = h(\mathbb{P}_0(X > Q_t^{\mathbb{P}}(X)))$ for $t \in [0,1]$, where $Q_t^{\mathbb{P}}(X)$ is the left quantile under the measure $\mathbb{P}$. We first show that $g(\mathbb{P}(X > x)) = h(\mathbb{P}_0(X > x))$ for all $x \in \mathbb{R}$. It is clear that $g(\mathbb{P}(X > x)) = h(\mathbb{P}_0(X > Q^{\mathbb{P}}_{\mathbb{P}(X\leqslant x)}(X)))$. By the definition

of $Q_t^{\mathbb{P}}$, we have $Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X) \leqslant x$. For $x \in \mathbb{R}$, if $Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X) = x$, then it is clear that $g(\mathbb{P}(X > x)) = h(\mathbb{P}_0(X > x))$. If $Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X) < x$, we have $\mathbb{P}(Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X) < X \leqslant x) = 0$. As $\mathbb{P}_0 \ll \mathbb{P}$, we have $\mathbb{P}_0(Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X) < X \leqslant x) = 0$. Hence,

$$h(\mathbb{P}_0(X > Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X))) = h\left(\mathbb{P}_0(x \geqslant X > Q^{\mathbb{P}}_{\mathbb{P}(X>x)}(X)) + \mathbb{P}_0(X > x)\right) = h(\mathbb{P}_0(X > x)).$$

Taking $t \uparrow 1$, we obtain $g(1) = h(1)$.

Next, let show that $\rho_h^{\mathbb{P}_0}(f(X)) = \rho_g^{\mathbb{P}}(f(X))$ for any increase function $f : \mathbb{R} \to \mathbb{R}$. Denote by $f^{-1}(x) = \inf\{y : f(y) > x\}$ the inverse function of f. As $\mathbb{P}(X = x) = 0$ and $\mathbb{P}_0 \ll \mathbb{P}$, we have $\mathbb{P}(X = x) = \mathbb{P}_0(X = x)$ for all $x \in \mathbb{R}$. Hence,

$$\begin{aligned}
\rho_h^{\mathbb{P}_0}(f(X)) &= \int_0^\infty h(\mathbb{P}_0(f(X) > x))\,\mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}_0(f(X) > x)) - h(1))\,\mathrm{d}x \\
&= \int_0^\infty h\left(\mathbb{P}_0(X > f^{-1}(x)) + \mathbb{P}_0(X = f^{-1}(x))\mathbb{1}_{\{f(f^{-1}(x))>x\}}\right)\,\mathrm{d}x \\
&\quad + \int_{-\infty}^0 \left(h\left(\mathbb{P}_0(X > f^{-1}(x)) + \mathbb{P}_0(X = f^{-1}(x))\mathbb{1}_{\{f(f^{-1}(x))>x\}}\right) - h(1)\right)\,\mathrm{d}x \\
&= \int_0^\infty h\left(\mathbb{P}_0(X > f^{-1}(x))\right)\,\mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}_0(X > f^{-1}(x))) - h(1))\,\mathrm{d}x \\
&= \int_0^\infty g(\mathbb{P}(X > f^{-1}(x)))\,\mathrm{d}x + \int_{-\infty}^0 (g(\mathbb{P}(X > f^{-1}(x))) - h(1))\,\mathrm{d}x = \rho_g^{\mathbb{P}}(f(X)),
\end{aligned}$$

as desired. □

Lemma 8.5 states that if a belief $\mathbb{P}_0$ is absolutely continuous with respect to a probability measure $\mathbb{P}$ and if a random variable X is continuous under $\mathbb{P}$, then we can always find a distortion function g such that the two distortion riskmetrics $\rho_h^{\mathbb{P}_0}$ and $\rho_g^{\mathbb{P}}$ are exactly the same for every random variable $Y = f(X)$ comonotonic with X.

Our last result states that when every belief is sufficiently "well-behaved", then the comonotonic risk sharing problem with heterogeneous beliefs is equivalent to a comonotonic risk sharing problem with homogeneous belief $\mathbb{P}$.

Proposition 8.10. *Let $\mathbb{P}_1, \dots, \mathbb{P}_n \in \mathcal{P}$, $h_1, \dots, h_n \in \mathcal{H}^{\mathrm{BV}}$ be given and let $X \in \mathcal{X}$ admit a density under all $\mathbb{P}_1, \dots, \mathbb{P}_n$. There exist a probability measure $\mathbb{P} \in \mathcal{P}$ and a collection of distortion functions $g_1, \dots, g_n \in \mathcal{H}^{\mathrm{BV}}$ such that the allocation $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ is Pareto optimal for $(\rho_{h_1}^{\mathbb{P}_1}, \dots, \rho_{h_n}^{\mathbb{P}_n})$ if and only if it is Pareto optimal for $(\rho_{g_1}^{\mathbb{P}}, \dots, \rho_{g_n}^{\mathbb{P}})$.*

The proof of Proposition 8.10 is straightforward. The essential step is to notice that we can find a probability measure $\mathbb{P}$ such that X admits a density under $\mathbb{P}$ and for which $\mathbb{P}_i \ll \mathbb{P}$, $i \in [n]$, and then invoke Lemma 8.5. The proof simply takes $\mathbb{P}$ as an average of the beliefs $\mathbb{P}_i$, although other such $\mathbb{P}$ would have worked.

Proof of Proposition 8.10. Let $\mathbb{P} = 1/n \sum_{i=1}^n \mathbb{P}_i$ and $g_i(t) = h_i(\mathbb{P}_i(X > Q_t^{\mathbb{P}}(X)))$ for $t \in [0,1]$. It is clear that X also has a density function under $\mathbb{P}$ and $\rho_{h_i}^{\mathbb{P}_i}(f(X)) = \rho_{g_i}^{\mathbb{P}}(f(X))$ for increasing functions f and $i \in [n]$ by Lemma 8.5. Hence, $(\rho_{h_1}^{\mathbb{P}_1}, \dots, \rho_{h_n}^{\mathbb{P}_n})$ and $(\rho_{g_1}^{\mathbb{P}}, \dots, \rho_{g_n}^{\mathbb{P}})$ have the same class of Pareto-optimal allocations. □

8.12 Omitted details in Section 8.6

We present the functions $G_\lambda^\alpha(h)$ for Cases 1 to 6 in Section 8.6.2 which yield the allocations that we present in that section.

Case 1: When $c_1 \geqslant 1/2$ and $c_3 \geqslant 1/2$ it is

$$G_\lambda^\alpha(h)(t) = \lambda_2 \left((t-\alpha) \wedge (1-t-\alpha)\right) \mathbb{1}_{\{\alpha<t<1-\alpha\}}.$$

Case 2: When $c_2 \geqslant 1/2$ and $c_3 \leqslant \alpha$ it is

$$G_\lambda^\alpha(h)(t) = \lambda_3 \left([(t-\alpha)(1+\alpha-t)] \wedge [(t+\alpha)(1-\alpha-t)]\right) \mathbb{1}_{\{\alpha<t<1-\alpha\}}.$$

Case 3: When either $\alpha < c_2 < c_3 < 1/2$ or $\alpha < c_1 < 1/2 < c_3$ it is

$$G_\lambda^\alpha(h)(t) = \left(\lambda_2[(t-\alpha) \wedge (1-t-\alpha)] \wedge \lambda_1\right) \mathbb{1}_{\{\alpha<t<1-\alpha\}}.$$

Case 4: When $c_3 \leqslant \alpha < c_2 < 1/2$ it is

$$G_\lambda^\alpha(h)(t) = \left(\lambda_3[(t-\alpha)(1+\alpha-t)] \wedge [(t+\alpha)(1-\alpha-t)] \wedge \lambda_1\right) \mathbb{1}_{\{\alpha<t<1-\alpha\}}.$$

Case 5: When $\alpha < c_3 < 1/2 < c_2$, it is

$$G_\lambda^\alpha(h)(t) = \begin{cases} 0, & t \in [0,\alpha] \cup [1-\alpha, 1], \\ \lambda_2(t-\alpha), & t \in (\alpha, c_3), \\ \lambda_3(t-\alpha)(1-t+\alpha), & t \in [c_3, 1/2), \\ \lambda_3(t+\alpha)(1-t-\alpha), & t \in [1/2, 1-c_3), \\ \lambda_2(1-\alpha-t), & t \in [1-c_3, 1-\alpha). \end{cases}$$

Case 6: When $\alpha < c_3 \leqslant c_2 < 1/2$ it is

$$G_\lambda^\alpha(h)(t) = \begin{cases} 0, & t \in [0,\alpha] \cup [1-\alpha, 1], \\ \lambda_2(t-\alpha), & t \in (\alpha, c_3), \\ \lambda_3(t-\alpha)(1-t+\alpha), & t \in [c_3, c_2), \\ \lambda_1 & t \in [c_2, 1-c_2), \\ \lambda_3(t+\alpha)(1-t-\alpha), & t \in [1-c_2, 1-c_3), \\ \lambda_2(1-t-\alpha), & t \in [1-c_3, 1-\alpha). \end{cases}$$

Chapter 9

Negatively dependent optimal risk sharing

9.1 Introduction

The problem of sharing risk and its mathematical underpinnings are pivotal in understanding the economic behaviours of agents. When agents are risk-averse expected utility maximizers, the risk sharing problem behaves similarly to the general equilibrium of an exchange economy with aggregate risks (Arrow and Debreu, 1954; Arrow, 1964; Radner, 1968). An important observation from this literature is that, under strict risk aversion, Pareto-optimal allocations are comonotonic, i.e., they are increasing functions of the total wealth. This can be interpreted as agents being "on the same boat" when losses or gains occur.

As comonotonicity is an extreme form of positive dependence, one might wonder if a converse statement exists for risk-seeking agents, i.e., the agents in an optimal allocation being "in opposite boats" when losses or gains occur. Unfortunately, it is well known that the most extreme form of negative dependence is generally not tractable with three or more random variables, and thus, the question is technically very challenging and not well understood.

This article addresses this gap by investigating negative dependence in risk sharing, and in particular, the extreme form of negative dependence, which we refer to as counter-

monotonicity. To establish comonotonic optimal allocations in the classic literature, a central mathematical tool is the *comonotonic improvement theorem* of Landsberger and Meilijson (1994), which states that for any random vector, there exists a comonotonic random vector whose components are *less risky* than those of the given random vector, in the sense of Rothschild and Stiglitz (1970). This establishes the important intuition that risk-averse agents always prefer comonotonic allocations.

Parallel to this classic finding, our main result, Theorem 9.1, referred to as the *counter-monotonic improvement theorem*, states that for any random vector bounded from below (or above), there exists a counter-monotonic random vector whose components are *riskier* than those of the given random vector. The counter-monotonic improvement theorem uses the stochastic representation of counter-monotonicity recently obtained by Chapter 7. In Proposition 9.2, we provide a simplification of this stochastic representation that makes transparent that any counter-monotonic allocation resembles extreme forms of gambling as either "winner-takes-all" or "loser-loses-all" (drawing straws) lotteries. We respectively define the normalized version of "winner-takes-all" and "loser-loses-all" allocations as *jackpot* and *scapegoat* allocations.

To appreciate the optimality of the jackpot and the scapegoat allocations, we need to depart from the standard utility theory of risk-averse agents. An immediate consequence of our main theorem is that for the problem of sharing risk among strictly risk-seeking agents, all Pareto-optimal allocations are jackpot allocations. However, this set may be empty in some situation, and this can be resolved by imposing constraints on the allocations or restricting the effective domain of their Bernoulli utility function.

To understand the role of scapegoat allocations, we then analyze the problem of sharing risk among agents that have the same discontinuous utility function. We first consider the problem of sharing risk among Dirac utility agents, defined as expected utility maximizers for which the Bernoulli utility function is an indicator function. A key property of this problem is that allocations which give a constant endowment to all agents but one are always Pareto optimal. The choice of whose allocation varies can be random, as if all agents were "drawing straws". The optimal allocations must thus be (payoff equivalent to) scapegoat allocations when the endowment is probabilistically too small. In this case, Pareto-optimal allocations

cannot be simultaneously comonotonic and fair, where we define fairness as all agents having the same expected utility.[1] We show that a similar result holds for agents with piecewise linear Bernoulli utility function with one jump, demonstrating that this result does not rely on the satiation of the underlying preference relation.

We proceed to consider agents modelled by rank-dependent expected utility (RDU) of Quiggin (1993), with a particular focus on agents with inverted S-shaped probability distortions as in the cumulative prospect theory of Tversky and Kahneman (1992). RDU agents with inverted S-shaped distortion can exhibit a combination of risk-averse and risk-seeking behaviours. Assuming that all agents are modelled by the same RDU, we find conditions where these agents prefer fair jackpot allocations to any other fair allocations. We show that if the number of agents is large, then only jackpot allocations can be both fair and Pareto optimal; as a consequence, comonotonic and fair allocations cannot be Pareto optimal.

We conclude with a simplified game-theoretical model of cryptocurrency mining where agents can choose to form a mining pool. Leshno and Strack (2020) already observed that risk-averse agents have an incentive to form mining pools because it allows them to reduce the variability of their payoff. Clearly, the payoff of joining the pool is a mean-preserving contraction of the payoff for mining alone when at least another agent joins the pool. Joining the pool is thus a weakly dominant strategy for risk-averse agents, with strict dominance if at least one other agent joins. However, RDU agents can behave the opposite depending on the size of their computing power. RDU agents with large computing power behave as risk-averse agents. But if their computing power is small so that their probability of mining the coin is also small, then mining alone can be a weakly dominant strategy, with strict dominance whenever at least one agent joins the pool. These novel results suggest that a richer model of pool formation is required to better understand the interaction between crypto-miners.

It is natural to ask if and how the counter-monotone improvement theorem can be used in other contexts than risk-sharing. We investigate competitive equilibria with risk-seeking agents and obtain all results of typical interest in general equilibrium, including the first and second fundamental theorem of welfare economics. We also show that all jackpot

[1] Fair allocations maximize the Rawlsian social welfare function in this particular setup.

allocations are Pareto optimal, hereby extending the results of Section 9.4. The second welfare theorem thus implies that all jackpot allocations are competitive equilibrium for some initial endowments. The analysis of competitive equilibria is highly technical and is thus relegated in Section 9.9. We emphasize that we do not know if a similar analysis can be performed for agents with other types of risk-seeking decision criteria, as the construction of the equilibrium pricing measure is quite delicate and tailored to the specific setting we study.

Next, we review the literature. Section 9.2 contains all preliminaries, including a formal statement of the classic result of comonotonic improvement. Section 9.3 states and proves our main result, the counter-monotonic improvement theorem. This is also where we review the stochastic representation of counter-monotonicity and define jackpot and scapegoat allocations. Sections 9.4, 9.5 and 9.6 consider respectively the risk sharing problem with risk-seeking agents, agents with a discontinuous Bernoulli utility function and RDU agents. Section 9.7 analyzes the choice of joining a crypto-currency mining pool, and the conclusion discusses avenues for further research. Section 9.9 analyzes competitive equilibria with risk-seeking agents, and Section 9.10 contains the proofs.

9.1.1 Literature review

The technique of comonotonic improvement was initially introduced in Landsberger and Meilijson (1994), and subsequently extended in Dana and Meilijson (2003), Ludkovski and Rüschendorf (2008) and Carlier et al. (2012). We refer to Rüschendorf (2013) for an up-to-date formal treatment and Section 9.2 for more details. For the formal treatment of counter-monotonicity, we refer to Puccetti and Wang (2015) for a general overview and to Chapter 7 for the stochastic representation of counter-monotonicity. In the actuarial literature, counter-monotonicity in dimension greater than two is also called mutual exclusivity; see Dhaene et al. (1999) and Cheung and Lo (2014). See also our discussion in Section 9.3.

While the economics and finance literature does not always employ the terms comonotonicity and comonotonic allocations, these concepts have long been a subject of interest in these fields. Of direct relevance is the case of economies with a constant aggregate endowment, where all allocations that are constant across states are comonotonic. In this context,

comonotonic allocations are sometimes called "no-betting" or "risk-free allocations." See, for instance, in the literature on risk sharing under heterogeneous beliefs and ambiguity: Billot et al. (2000), Rigotti et al. (2008), and Strzalecki and Werner (2011). More recently, Beissner et al. (2023) analyzes no-betting allocations on probability spaces with two RDU agents. Chateauneuf et al. (2000) analyzes comonotonic allocations with aggregate risk when all agents are ambiguity-averse Choquet expected utility maximizers.

In contrast to comonotonicity, the concept of counter-monotonicity received much less attention in the economics and finance literature related to risk sharing. A notable exception is quantile-based risk sharing. A key property of counter-monotonic allocations and, more generally, of negatively dependent allocations is their optimality in this setting. See Embrechts et al. (2018) and Weber (2018) for quantile-based risk sharing problems on probability spaces with Pareto-optimal counter-monotonic allocations and Embrechts et al. (2020) for the case of heterogeneous beliefs. Chapter 8 contains a risk sharing problem where the optimal allocations entail both positive and negative dependence. Specifically, the authors show that when sharing risk with agents that consider the inter-quantile difference as their measure of variability, any Pareto-optimal allocation entails counter-monotonicity on the tails of the distribution of the aggregate risk.

We refer to Quiggin (1993) for RDU agents, although we are mostly interested in the inverted S-shaped probability distortion functions considered in the cumulative prospect theory of Tversky and Kahneman (1992) (and Kahneman and Tversky (1979)). We show that RDU agents using the Kahnmenan-Tversky inverted S-shaped distortion function can behave as risk-seeking agents provided the risk is to be shared among a large number of individuals.

Our crypto-currency example is inspired by the axiomatic characterization of Leshno and Strack (2020), which obtains an impossibility result for risk-averse miners. In a nutshell, mining pools can provide a reward scheme that is a mean-preserving contraction of the payoff of individual mining. Risk-averse agents always prefer to mine in a pool. On the opposite, we find conditions for which RDU agents behave as risk-seeking agents and prefer to mine alone despite having a concave Bernoulli utility function.

9.2 Preliminaries: risk sharing and comonotonic improvements

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and denote by $\mathcal{X}$ a corresponding L^p space, where almost surely equal objects are treated as equal. While Sections 9.2 and 9.3 consider $\mathcal{X} = L^1$ for generality, Sections 9.4, 9.5 and 9.6 consider the more standard setting $\mathcal{X} = L^\infty$, the set of all bounded random variables. Let n be a positive integer and write $[n] := \{1, \dots, n\}$. We are mainly interested in the situation where $n \geqslant 3$ agents share a random outcome $X \in \mathcal{X}$.

Definition 9.1. An *allocation* of $X \in \mathcal{X}$ is an element of the set

$$\mathbb{A}_n(X) := \left\{ (X_1, \dots, X_n) \in \mathcal{X}^n : \sum_{i=1}^n X_i = X \right\}.$$

A foundational idea of risk sharing is that if all agents are strictly risk averse and know the probability measure $\mathbb{P}$, then all Pareto-optimal allocations are comonotonic (see e.g., Rüschendorf (2013)). At a formal level, this is typically proved using the technique of *comonotonic improvements*, as introduced in Landsberger and Meilijson (1994) for a finite state space. The result was subsequently extended to L^∞ (Dana and Meilijson, 2003) and L^1 (Ludkovski and Rüschendorf, 2008). We provide some background to understand the scope of the comonotonic improvement technique.

The two random variables X, Y are said to be *comonotonic* if

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0 \quad \text{for } (\mathbb{P} \times \mathbb{P})\text{-almost every } (\omega, \omega') \in \Omega^2,$$

and the collection of random variables $X_1, \dots, X_n$ is comonotonic if all its component are pairwise comonotonic. Alternatively, the random variables $X_1, \dots, X_n$ are comonotonic if there exists a collection of increasing functions $f_i : \mathbb{R} \to \mathbb{R}$, $i \in [n]$, and a random variable Z such that $X_i = f_i(Z)$ for all $i \in [n]$ (recall that equalities are in the $\mathbb{P}$-almost sure sense). The latter definition comes from the stochastic representation of comonotonicity given by Denneberg's Lemma (see Denneberg, 1994, Proposition 4.5), and if $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$, then one can set $Z = X$ in the preceding definition.

A random variable X is said to be smaller than a random variable Y in the *convex order*, denoted by $X \leqslant_{\text{cx}} Y$, if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for every convex function $\phi : \mathbb{R} \to \mathbb{R}$ provided

that both expectations exist (see Rüschendorf (2013) and Shaked and Shanthikumar (2007)). The order $X \leqslant_{\text{cx}} Y$ means that X is less risky than Y in the sense of Rothschild and Stiglitz (1970). Notice that if $X \leqslant_{\text{cx}} Y$, then $\mathbb{E}[X] = \mathbb{E}[Y]$, meaning that the convex order compares random variable with the same mean. Similarly, X is smaller than Y in the increasing convex order, denoted by $X \leqslant_{\text{icx}} Y$, if $\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)]$ for every increasing convex function $\phi : \mathbb{R} \to \mathbb{R}$ provided the expectations exist. The proof of the next proposition is in Rüschendorf (2013, Theorem 10.50).

Proposition 9.1 (Comonotonic improvements)**.** *Let $X_1, \dots, X_n \in L^1$ and $X = \sum_{i=1}^n X_i$. Then there exists a $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ such that (i) $(Y_1, \dots, Y_n)$ is comonotonic and (ii) for every $i \in [n]$ it is $Y_i \leqslant_{\text{cx}} X_i$.*

Now, assume that X represents a monetary payoff so that greater values are preferred, and let $\rho_i : \mathcal{X} \to \mathbb{R}$ denote the decision criterion used by agent $i \in [n]$. An allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is *Pareto-optimal* in $\mathbb{A}_n(X)$ if for any $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ satisfying $\rho_i(Y_i) \geqslant \rho_i(X_i)$, all $i \in [n]$, we have $\rho_i(Y_i) = \rho_i(X_i)$, all $i \in [n]$. Let $(\lambda_1, \dots, \lambda_n)$ be a vector of positive numbers, usually called a Negishi weight vector. We say that an allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is *sum-optimal* in $\mathbb{A}_n(X)$ with respect to $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ if $(X_1, \dots, X_n)$ maximizes $\sum_{i=1}^n \lambda_i \rho_i(X_i)$ subject to the constraint $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$. We use the term sum optimality for the case $\lambda_i = 1$, all $i \in [n]$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let all agents have homogeneous beliefs, i.e., everyone agrees on the probability measure $\mathbb{P}$. The concept of strict risk aversion translates to a strict preference for random variables that are lower in the convex order. That is, if $X_i <_{\text{cx}} Y_i$ (meaning $X_i \leqslant_{\text{cx}} Y_i$ but $Y_i \not\leqslant_{\text{cx}} X_i$) then $\rho_i(X_i) > \rho_i(Y_i)$. We can derive from Proposition 9.1 that the set of Pareto-optimal allocations contains only comonotonic allocations when all agents are strictly risk averse.

9.3 Counter-monotonicity and counter-monotonic improvement

Comonotonicity is an extreme type of positive dependence. This article contends with the opposite situation: negatively dependent optimal allocations, which are much less studied in the literature. We first define this dependence concept. In all statements, X represents a random wealth, so greater values are preferred; negative values of X are allowed.

Throughout this section, we consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Two random variables X, Y are *counter-monotonic* if the two random variables $X, -Y$ are comonotonic. An allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is pairwise counter-monotonic if for every $i \neq j$ the random variables X_i, X_j are counter-monotonic. Pairwise counter-monotonicity is the generalization of counter-monotonicity for the case $n \geqslant 3$, but the concept is not always well-defined for dimensions $n \geqslant 3$. We use the simpler term counter-monotonicity throughout.

The next lemma, due to Dall'Aglio (1972), gives necessary conditions for a random vector $(X_1, \dots, X_n)$ to be counter-monotonic.

Lemma 9.1 (Dall'Aglio (1972))**.** *If at least three of $X_1, \dots, X_n$ are non-degenerate, counter-monotonicity of $(X_1, \dots, X_n)$ means that one of the following two cases holds true:*

$$\mathbb{P}(X_i > \text{ess-inf} X_i,\ X_j > \text{ess-inf} X_j) = 0 \text{ for all } i \neq j; \tag{9.1}$$

$$\mathbb{P}(X_i < \text{ess-sup} X_i,\ X_j < \text{ess-sup} X_j) = 0 \text{ for all } i \neq j. \tag{9.2}$$

A necessary condition for (9.1) is $\sum_{i=1}^{n} \mathbb{P}(X_i > \text{ess-inf} X_i) \leqslant 1$, and a necessary condition for (9.2) is $\sum_{i=1}^{n} \mathbb{P}(X_i < \text{ess-sup} X_i) \leqslant 1$.

Let Π_n be the set of all n-compositions of Ω, that is,

$$\Pi_n = \left\{ (A_1, \dots, A_n) \in \mathcal{F}^n : \bigcup_{i \in [n]} A_i = \Omega \text{ and } A_1, \dots, A_n \text{ are disjoint} \right\}.$$

In other words, a composition of Ω is a partition of Ω with order. The next proposition simplifies the stochastic representation of counter-monotonicity given in Theorem 7.1 of Chapter 7.

Proposition 9.2. *For $X \in \mathcal{X}$, suppose that at least three of $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ are non-degenerate. Then, $(X_1, \dots, X_n)$ is counter-monotonic if and only if there exist constants $m_1, \dots, m_n$ and $(A_1, \dots, A_n) \in \Pi_n$ such that*

$$X_i = (X - m)\mathbb{1}_{A_i} + m_i \quad \text{for all } i \in [n] \text{ with } m = \sum_{i=1}^{n} m_i \leqslant \text{ess-inf} X, \tag{9.3}$$

or

$$X_i = (X - m)\mathbb{1}_{A_i} + m_i \quad \text{for all } i \in [n] \text{ with } m = \sum_{i=1}^{n} m_i \geqslant \text{ess-sup} X. \tag{9.4}$$

Remark 9.1. As in Denneberg's Lemma, the underlying probability measure $\mathbb{P}$ is not used in the stochastic representation of counter-monotonicity of Chapter 7, and so the allocations characterized in Proposition 9.2 are also well defined on measurable spaces without specified probability, as long as the null sets are specified.

The allocation $(X, 0, \dots, 0)$ is counter-monotonic by taking $A = \Omega$ and $m = m_1 = \text{ess-inf} X$, and it is trivial to verify that it is also comonotonic. Notice now that the allocations defined in equation (9.3) and (9.4) echo the allocations in Lemma 9.1. In (9.3), for every $\omega \in \Omega$, at most one agent receives more than their essential infimum. Conversely, in (9.4), at most one agent receives less than their essential supremum. This is the "winner-takes-all" and "loser-loses-all" structure of counter-monotonic allocations.

The most curious case of both (9.3) and (9.4) when $m_1 = \cdots = m_n = 0$, given by

$$X_i = X\mathbb{1}_{A_i} \text{ for all } i \in [n], \text{ where } (A_1, \dots, A_n) \in \Pi_n, \tag{9.5}$$

will draw our special attention. Note that $m = 0$ implies that either $X \geqslant 0$ or $X \leqslant 0$ holds, resulting in two different cases.

Definition 9.2. An allocation $(X_1, \dots, X_n)$ is a *jackpot allocation* if (9.5) holds for some $X \geqslant 0$, and it is a *scapegoat allocation* if (9.5) holds for some $X \leqslant 0$.

A comparison of a jackpot allocation and a comonotonic allocation is illustrated in Figure 9.1. Although sharing the formula (9.5), a jackpot allocation and a scapegoat allocation have very different meanings. In a jackpot allocation, the total wealth X is nonnegative (e.g.,

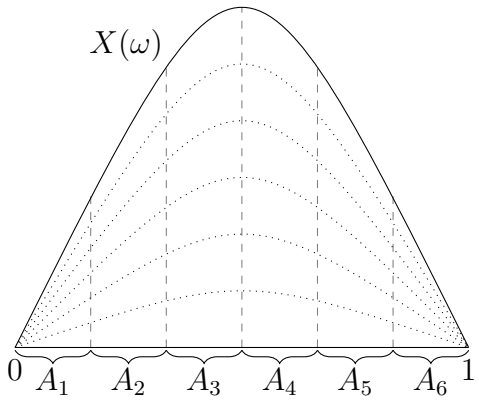


Figure 9.1: An illustration of positively and negatively dependent allocations, where a comonotonic allocation is $X_i = X/n$ for $i \in [n]$ (the area between two dotted curves) and a jackpot allocation is $X_i = X\mathbb{1}_{A_i}$ for $i \in [n]$ with $\Omega = [0, 1]$ (the area between two dashed lines).

a prize), and for each realization of the world ω, only one agent "wins", i.e., receives all positive payoff, and all other agents receive nothing. In a scapegoat allocation, the total wealth is nonpositive (e.g., a loss), and only one agent "loses", i.e., suffers the loss. Both types of allocations are often observed in daily life. For instance, the simple lottery ticket (only one winner) is a jackpot allocation, and the "designated driving" is a scapegoat allocation.

The next result shows a special role of the jackpot and the scapegoat allocations among all counter-monotonic allocations. Using Lemma 9.1, an allocation $(X_1, \dots, X_n)$ is a jackpot allocation if and only if

$$X_i \geqslant 0 \text{ and } \mathbb{P}(X_i \wedge X_j > 0) = 0 \text{ for all } i \neq j, \tag{9.6}$$

where $a \wedge b$ means $\min\{a, b\}$. Therefore, being a jackpot allocation is a property of the joint distribution of $(X_1, \dots, X_n)$. The probabilistic mixture of two random vectors with joint distributions F and G is another random vector with joint distribution $\lambda F + (1 - \lambda)G$ for some $\lambda \in [0, 1]$. The next result yields that jackpot allocations are closed under probabilistic mixtures. The same holds for scapegoat allocations by symmetry.

Proposition 9.3. *A probabilistic mixture of two jackpot allocations is again a jackpot allocation.*

For two general counter-monotonic allocations other than jackpot and scapegoat allocations, their mixture is not necessarily counter-monotononic, even if they both belong to the same type (9.3) or (9.4).

Next is our main result.

Theorem 9.1. *Let $X_1, \dots, X_n \in L^1$ be nonnegative and $X = \sum_{i=1}^n X_i$. Assume that there exists a uniform random variable U independent of X. Then, there exists $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ such that (i) $(Y_1, \dots, Y_n)$ is counter-monotonic; (ii) $Y_i \geqslant_{\mathrm{cx}} X_i$ for $i \in [n]$; (iii) $Y_1, \dots, Y_n$ are nonnegative. Moreover, $(Y_1, \dots, Y_n)$ can be chosen as a jackpot allocation.*

Remark 9.2. The boundedness from below of $X_1, \dots, X_n$ is necessary to obtain the existence of jackpot allocations. A similar statement can be made for scapegoat allocations, which then requires the boundedness of $X_1, \dots, X_n$ from above instead (e.g., $X_i \leqslant 0$ for all $i \in [n]$). The proof follows from observing that in this case, $-X_1, \dots, -X_n$ satisfies the assumptions of Theorem 9.1 and is thus omitted.

Theorem 9.1 gives a converse to the comonotonic improvements for bounded random variables. We obtain that jackpot allocations will always be preferred by risk-seeking agents.

Before moving on, we emphasize that the technical assumption that there exists a uniform random variable U independent of X is not completely innocuous. Intuitively, we can interpret it as assuming that any allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ can be "implemented" with randomization devices like flipping coins or spinning roulette wheels. At a technical level, this assumption guarantees that the inf-(sup-)convolution of law-invariant functionals is law-invariant Liu et al. (see 2020).

9.4 Risk-seeking agents in expected utility theory

From now on, let us focus on $\mathcal{X} = L^\infty$. Counter-monotonic allocations on probability spaces are not necessarily interesting when we restrict our attention to the most popular

preferences, which are modelled by concave Bernoulli utility functions. To see why, let $u_i : \mathbb{R} \to \mathbb{R}$ be a twice-differentiable Bernoulli utility function, assume that every individual $i \in [n]$ shares the same risk attitude and consider the expected utility criterion $\rho_i(X_i) = \mathbb{E}[u(X_i)]$. We can always trivially find counter-monotonic Pareto-optimal allocations when all individuals are risk neutral. The reader can convince themselves by simply setting u_i as the identity and observing that any allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is Pareto optimal.[2] Thus, anything goes with risk-neutrality, and there is little to say about counter-monotonicity in this context.

As mentioned, a foundational result in risk-sharing is the comonotonicity of Pareto optimal allocations when individuals are strictly risk averse. This result implies that there cannot be counter-monotonic Pareto-optimal allocations (besides the trivial counter-monotonic allocations) when the utility functions are strictly concave. It thus seems natural to consider strictly risk-seeking individuals. However, the next informal argument shows that Pareto-optimal allocations do not exist in the general case.

Suppose that $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is Pareto optimal and consider two strictly risk-seeking individuals $i \neq j$. We can construct another feasible allocation $(X_1', \dots, X_n') \in \mathbb{A}_n(X)$ by finding a non-trivial partition $A \cup B$ of Ω on which we create an arbitrary transfer of wealth between i and j. Say, if $\omega \in A$, then i gives one billion dollars to j and vice-versa if $\omega \in B$. The strict convexity of u_i and u_j implies, by Jensen's inequality, that both individuals are strictly better off, contradicting the Pareto optimality of $(X_1, \dots, X_n)$.

We know of two ways to make the problem sensible. The first is to impose lower bounds on the allocation, so $X_i \geqslant a$ for $a \in \mathbb{R}$ and $i \in [n]$. The case $a = 0$ is of particular interest because it can be interpreted as a no-short selling/borrowing constraint. The second approach is to restrict the set of allocations by restricting the effective domain of u; we emphasize that this is a common strategy in the empirical literature, where concave power utilities $u(x) = x^\alpha$, $0 < \alpha \leqslant 1$, are used extensively. In what follows, all utility functions are mappings from $\mathbb{R}$ to $\mathbb{R} \cup \{-\infty\}$ and not constantly $-\infty$.

Theorem 9.2. *Let $X \geqslant 0$ in L^∞ and such that there is a uniform independent of it and let*

[2] We implicitly consider a.s. bounded allocations.

u_i be increasing and strictly convex on $[0,\infty)$ and taking value $-\infty$ for all $x<0$ for each i. Then all Pareto-optimal allocations are jackpot allocations.

It is straightforward to see that we obtain a similar result for risk-seeking agents if we restrict the set of feasible allocations to allocations satisfying $X_i \geqslant 0$, all $i \in [n]$.

Theorem 9.2 obtains that all Pareto-optimal allocations must be counter-monotonic allocations with $m_i = 0$ for all $i \in [n]$, i.e., jackpot allocation. It is natural to wonder whether the converse is true, that is, whether all counter-monotonic allocations of X satisfying $m_i = 0$, all $i \in [n]$, are Pareto optimal. Proposition 9.12 in Section 9.9 shows that the answer is "Yes" in the general case. However, the construction is technical, and the next proposition shows the simplified case $X = x > 0$.

Proposition 9.4. *Let $x > 0$ be given.*

(i) If $u_1, \dots, u_n$ are strictly increasing and concave functions, then all comonotonic allocations of x are Pareto optimal.

(ii) If $u_1, \dots, u_n$ are increasing functions with $u_i(x) > u_i(0)$, all $i \in [n]$ that are convex on $[0,\infty)$ and taking value $-\infty$ for all $x < 0$, then all jackpot allocations are Pareto optimal.

While all the allocations found above are Pareto optimal, not all are equal from a welfare point of view. When all agents use the same decision criterion and have the same Bernoulli utility, we say that an allocation is fair if all agents achieve the same (Ex-ante) welfare.

Definition 9.3. Let all agents have the same decision criterion ρ and the same Bernoulli utility function u. Then an allocation $(X_1, \dots, X_n)$ is *fair* if $\rho_i(X_i) = \rho_i(X_j)$ for all $i \neq j$.

We emphasize that this notion of fairness is cardinal, and so we only define it for the case where every agent is identical in order to avoid general interpersonal comparisons of welfare.

Definition 9.4. An allocation $(X_1, \dots, X_n)$ is *distributionally fair* if $X_i \overset{\mathrm{d}}{=} X_j$ for all $i \neq j$.

For a counter-monotonic allocation to be distributionally fair, one must thus require that the underlying $(A_1, \ldots, A_n) \in \Pi_n$ be such that $\mathbb{P}(A_i) = \mathbb{P}(A_j) = 1/n$, all $i \neq j$. Clearly, a counter-monotonic allocation is fair if it is distributionally fair, and all distributionally fair allocations are fair when all agents share the same decision criterion and Bernoulli utility. Equipped with this distinction, we close this section with an example that highlights the role of counter-monotonic allocations as a different way to take convex combinations.

Example 9.1. Set $x = 1$ and consider the problem

$$\text{to maximize} \quad \sum_{i=1}^{n} \mathbb{E}[u(X_i)] \quad \text{subject to } (X_1, \ldots, X_n) \in \mathbb{A}_n(1) \ \text{ and } X_1, \ldots, X_n \geqslant 0.$$

Set $a_i \in \mathbb{R}^n$ as $a_1 = (1, 0, \ldots, 0)$, $a_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $a_n = (0, \ldots, 1)$, and denote by δ the Dirac delta function. The collection $(a_i, \delta_{a_i})_{i\in[n]}$ denotes all the allocations giving the whole $x = 1$ to one agent with certainty and corresponds to all the extreme points of the utility possibility set. With risk-averse agents, one must have comonotonic allocation, and we are taking "convex combinations along the a_is". In this case, the only fair comonotonic allocation is the pair $\big((1/n, \ldots, 1/n), \delta_{(1/n,\ldots,1/n)}\big)$ that gives everyone $x_i = 1/n$ with certainty.

Counter-monotonic allocations are like "taking convex combination along the δ_{a_i}s". In this case, all fair allocations must be like the allocation $(a_i, \delta_{a_i}/n)_{i\in[n]}$, i.e. they must give $x_i = 1$ to one agent with probability $1/n$. With strictly risk-seeking agents, only this type of "convex combination" preserves Pareto optimality. In the general case, the non-atomicity of the probability space combined with the assumption that there exists a uniform independent of X precisely guarantees that we can find these "convex combinations along the probabilities".

9.5 Discontinuous Bernoulli utilities

The previous section obtained the optimality of jackpot allocations when all agents are risk-seeking (provided that Pareto-optimal allocations exist). These allocations have a direct interpretation as a "winner-takes-all" lottery where a (random) prize is (potentially non-randomly) given to only one winner. We now turn our attention to scapegoat allocations

and present situations where they are optimal. In order to simplify the treatment, consider the following:

Assumption 9.1. *The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, the set of random variables is $\mathcal{X} = L^\infty$, and $X \in \mathcal{X}$ is such that there exists a uniform U independent of X.*

The sequel always assumes Assumption 9.1.

9.5.1 Pareto optimal allocations with Dirac utility

For every $i \in [n]$ set the decision criterion $\rho_i : \mathcal{X} \to \mathbb{R}$ as $\rho_i(X_i) = \mathbb{E}\left[\alpha \mathbb{1}_{\{X_i \geqslant 1\}}\right]$; we will refer to these agents as Dirac agents. In Chapter 7, we considered the special case of the risk-sharing problem with Dirac agents where $X = 1$ and $X_i \geqslant 0$. We interpreted the variable $X = 1$ as an indivisible good that was auctioned and the utility function as the net utility of n agents with the quasi-linear utilities $v(X, t) = \theta X - t$ having bid the same amount $\theta - t = \alpha$.

It is straightforward to see that the set

$$\{(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}) \in \mathbb{A}_n(X) : (A_1, \dots, A_n) \in \Pi_n\}$$

consists of Pareto-optimal jackpot allocations. We thus interpreted the allocations satisfying $\mathbb{P}(A_i) = \mathbb{P}(A_j)$ for every $i \neq j$ as the random tie-breaking rule. Those are distributionally fair, and thus, they also are fair allocations because all agents have the same expected utility. We observed that a fair lottery (which is counter-monotonic) is the only fair way to distribute the indivisible good among people who value it equally.

This section analyzes further the problem of sharing risk among Dirac agents. We assume $\alpha = 1$ for simplicity and without loss of generality. We first establish that for every $X \in \mathcal{X}$ there exists a counter-monotonic allocation which is Pareto optimal. Let us first set the Negishi weights to one so that we search for the allocations $(X_1, \dots, X_n)$ that solve

$$\text{to maximize} \quad \sum_{i=1}^{n} \rho_i(X_i) \quad \text{subject to } (X_1, \dots, X_n) \in \mathbb{A}_n(X).$$

Proposition 9.5. *The allocation $X_1 = X_2 = \cdots = X_{n-1} = 1$ and $X_n = X - (n-1)$ is Pareto optimal.*

The trivial counter-monotonic allocation in Proposition 9.5 has the characteristic that agent n above potentially gives everything to its peers. Notice now that

$$\Upsilon(X) := \left\{ \left(\mathbb{1}_{A_1}(X-(n-1)) + \mathbb{1}_{A_1^c}, \ldots, \mathbb{1}_{A_n}(X-(n-1)\right) + \mathbb{1}_{A_n^c}) \in \mathbb{A}_n(X) : (A_1, \ldots, A_n) \in \Pi_n \right\}$$

consist exclusively of sum-optimal allocations. We can interpret this set as the set of all allocations that "randomizes who gets or loses everything", much like drawing straws but with potentially unfair probabilities. When $\mathbb{P}(X \leqslant n) = 1$, the set $\Upsilon(X)$ boils down to

$$\mathcal{T}(X) := \{(\mathbb{1}_{A_1}(X-n)+1, \ldots, \mathbb{1}_{A_n}(X-n)+1) \in \mathbb{A}_n(X) : (A_1, \ldots, A_n) \in \Pi_n\},$$

the set of scapegoats allocations of X that are shifted so that it satisfies $m_i = 1$, all $i \in [n]$.

Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$, $\lambda_i > 0$, be a vector of Negishi weights and $f : \mathcal{X}^n \times \mathbb{R}^n_{++} \to \mathbb{R}$ be

$$f_{\boldsymbol{\lambda}}(\mathbf{X}) = \mathbb{E}\left[\sum_{i=1}^{n} \lambda_i \mathbb{1}_{\{X_i \geqslant 1\}}\right]$$

for $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{A}_n(X)$. By convention, we consider $\lambda_1, \ldots, \lambda_n$ in decreasing order.

Proposition 9.6. *If* $\mathbf{X}^*$ *maximizes* $f_{\boldsymbol{\lambda}}(\mathbf{X})$ *then* $\mathbf{X}^*$ *also maximizes* $f_{\mathbf{1}}(\mathbf{X})$.

Proposition 9.6 informs us that with Dirac agents, it suffices to characterize the set of sum-optimal allocations in order to understand the whole set of allocations that are sum-optimal for some Negishi weights $\boldsymbol{\lambda}$. This, of course, means that $\Upsilon(X)$ contains all allocations of interest.

While all the allocations in $\Upsilon(X)$ maximize the sum of expected utilities, not all are equal from a welfare point of view. Recall that a Rawlsian social welfare function considers only the utility of the worst-off agent and that maximizing a Rawlsian social welfare function involves focusing on allocations where $\rho_i = \rho_j$, all $i \neq j$.[3] This motivates our focus on fair allocations, where in our case we have that an allocation $(X_1, \ldots, X_n) \in \mathbb{A}_n(X)$ is fair if for every $i \neq j$, $i, j \in [n]$, it is $\rho_i(X_i) = \rho_i(X_j)$.

[3] In our context, a social welfare function $\mathcal{W}$ is a mapping $(\rho_i)_{i\in[n]} = \boldsymbol{\rho} \mapsto \mathcal{W}(\boldsymbol{\rho}) \in \mathbb{R}$ that ranks allocation, so higher value are better. The Rawlsian social welfare function is $\mathcal{W}(\boldsymbol{\rho}) = \min_{i\in[n]}\{\rho_i\}$. Thus if two allocations $\mathbf{X}, \mathbf{Y}$ are such that $\sum_{i=1}^n \rho_i(X_i) = \sum_{i=1}^n \rho_i(Y_i)$, but $\boldsymbol{X}$ is fair and $\boldsymbol{Y}$ is not, then a Rawlsian social welfare function ranks $\boldsymbol{X}$ strictly higher than $\boldsymbol{Y}$.

Since we assumed that there exists a uniform U independent of X, we can always find a partition $(A_1, \ldots, A_n) \in \Pi_n$ independent of X such that $\mathbb{P}(A_i) = 1/n$ for all $i \in [n]$. Thus, the set $\Upsilon(X)$ always contains distributionally fair allocations that give the same expected utility to all agents, and $\Upsilon(X)$ always contains a fair allocation.

Proposition 9.7. *If $\mathbb{P}(X < n) > 0$ then there exists no allocation which is simultaneously comonotonic, fair and sum-optimal.*

To summarize, we obtained that the Pareto optimality of an allocation $(X_1, \ldots, X_n)$ requires that for every $\omega \in \Omega$, there is at most one agent $i \in [n]$ for which $X_i(\omega) < 1$. If $\mathbb{P}(X < n) = 0$, one can always find comonotonic allocations (satisfying $X_i \geqslant 1$, all $i \in [n]$) that are Pareto optimal. In particular, fair ones exist. But when $\mathbb{P}(X < n) > 0$, it is no longer possible for a fair comonotonic allocation to have at most one agent i for which $X_i(\omega) < 1$.

The idea of a scapegoat allocation and of drawing straws to randomize the scapegoat is quite alien to standard welfare analysis, and one might be interested in imposing a lower bound on the allocation. Once again, the constraint $X_i \geqslant 0$, all $i \in [n]$, is particularly interesting because it can be interpreted as a borrowing constraint. Clearly, imposing constraints on the allocation can impact the aggregate welfare when $\mathbb{P}(X < n) > 0$. While we do not fully characterize the impacts of such constraints, we observe that they sometimes imply that jackpot allocations are Pareto optimal, as in the auction example above.

Corollary 9.1. *Let $X \geqslant 0$ and consider the constraints $X_i \geqslant 0$, all $i \in [n]$. If $\mathbb{P}(X < 2) = 1$ then all Pareto-optimal allocations are payoff equivalent to a jackpot allocation: if $(X_1, \ldots, X_n)$ is constrained Pareto optimal then there exists a feasible counter-monotonic allocation $(Y_1, \ldots, Y_n)$ such that $\rho_i(X_i) = \rho(Y_i)$ for all $i \in [n]$.*

Next, we extend our analysis to piecewise linear Bernoulli utility with one jump.

9.5.2 Piecewise linear Bernoulli utility with one jump

Assume now that piecewise linear Bernoulli utility functions $u_i(X_i) = a_i X_i + b_i \mathbb{1}_{\{X_i \geqslant 1\}}$, so that $\rho_i(X_i) = \mathbb{E}[u_i(X_i)] = a_i \mathbb{E}[X_i] + b_i \mathbb{P}(X_i \geqslant 1)$. For simplicity, we assume that $a_i = a$

and $b_i = 1$, for all $i \in [n]$. These agents are a combination of risk-neutral agents and Dirac agents, and it is easy to verify that ρ_i is monotone.

Consider once again the set

$$\Upsilon(X) = \left\{ \left(\mathbb{1}_{A_1}(X - (n-1)) + \mathbb{1}_{A_1^c}, \dots, \mathbb{1}_{A_n}(X - (n-1)) + \mathbb{1}_{A_n^c} \right) \in \mathbb{A}_n(X) : (A_1, \dots, A_n) \in \Pi_n \right\}.$$

As before, when $\mathbb{P}(X \leqslant n) = 1$, the set $\Upsilon(X)$ boils down to

$$\mathcal{T}(X) = \{ (\mathbb{1}_{A_1}(X - n) + 1, \dots, \mathbb{1}_{A_n}(X - n) + 1) \in \mathbb{A}_n(X) : (A_1, \dots, A_n) \in \Pi_n \},$$

the set of Pareto-optimal scapegoat allocations satisfying $m_i = 1$, all $i \in [n]$. Once again by Assumption 9.1 we can find a partition $(A_1, \dots, A_n) \in \Pi_n$ independent of X such that $\mathbb{P}(A_i) = 1/n$ for all $i \in [n]$, so $\Upsilon(X)$ contains a fair allocation.

Suppose first that $\mathbb{P}(X < n) = 0$. It is easy to verify that the comonotonic allocation $X_i = X/n$, $i \in [n]$, is Pareto optimal and fair. In other words, when $\mathbb{P}(X < n) = 0$, the problem goes in a similar fashion to the problem of sharing risk among Dirac agents. Our interest thus lies again in the case where $\mathbb{P}(X < n) > 0$. Let $(X_1, \dots, X_n) \in \Upsilon(X)$ be fair, and for simplicity chose it so that $(A_1, \dots, A_n) \in \Pi_n$ is independent of X. Computing the expected utility for all $i \in [n]$ we have

$$\begin{aligned}
\mathbb{E}[u_i(X_i)] &= a\mathbb{E}[X_i] + \mathbb{P}(X_i \geqslant 1) \\
&= a\mathbb{E}[\mathbb{1}_{A_i^C} + \mathbb{1}_{A_i}(X - (n-1))] + \mathbb{P}(A_i^C) + \mathbb{P}(A_i)\mathbb{P}(X \geqslant n) \\
&= a\left(\mathbb{P}(A_i^C) + \mathbb{P}(A_i)\mathbb{E}[X] - \mathbb{P}(A_i)(n-1)\right) + \mathbb{P}(A_i^C) + \mathbb{P}(A_i)\mathbb{P}(X \geqslant n) \\
&= \frac{a\mathbb{E}[X] + (n-1) + \mathbb{P}(X \geqslant n)}{n}
\end{aligned}$$

since $\mathbb{P}(A_i^C) = (n-1)/n = \mathbb{P}(A_i)(n-1)$. Summing over $i \in [n]$ we obtain

$$\sum_{i=1}^{n} \mathbb{E}[u_i(X_i)] = a\mathbb{E}[X] + (n-1) + \mathbb{P}(X \geqslant n).$$

Proposition 9.8. *Let $\mathbb{P}(X < n) > 0$. Then a comonotonic and fair allocation $(X_1, \dots, X_n)$ cannot be Pareto optimal.*

To summarize, the key property of the jump in the Bernoulli utility function is that it creates an incentive to concentrate losses on at most one agent. This property is not driven

by the satiation of preferences, as was implicitly suggested by the risk-sharing problem with Dirac agents, but rather by the sharp gains in utility at the discontinuity threshold. Thus, when there is a positive probability of not having enough to share, the corresponding optimal allocation cannot be simultaneously comonotonic and fair if it is to be Pareto optimal, as the latter requires concentrating the losses.

9.6 Rank-dependent utility agent

We now analyze the problem of sharing risk among RDU agents. A function $h : [0,1] \to [0,1]$ is called a probability distortion if it is non-decreasing and satisfies $h(0) = 0$ and $h(1) = 1$. An agent is RDU if its decision criterion is $\rho_h(X) = \int u(X)\,\mathrm{d}h \circ \mathbb{P}$, where $u : \mathbb{R} \to \mathbb{R}$ is a Bernoulli utility function, h is a probability distortion and where the integral is in the sense of Choquet. When the Bernoulli utility function is linear, we obtain Yaari (1987)'s dual utility with decision criterion $\rho_h(X) = \int X\,\mathrm{d}h \circ \mathbb{P}$. Yaari agents are risk seeking (risk averse) if the probability distortion function h is concave (convex). We remind the reader that when $0 < \gamma < 1$ the Kahneman-Tversky (KT) distortion function $h_{\mathrm{KT}}(t) = \frac{t^\gamma}{(t^\gamma + (1-t)^\gamma)^{1/\gamma}}$ is inverted S-shaped, i.e. concave-convex.

Assumption 9.2. *The utility Bernoulli function $u : \mathbb{R} \to \mathbb{R}$ is increasing and differentiable on $[0,\infty)$, weakly concave, satisfies $u(0) = 0$ and such that $u(x) = -\infty$ for all $x < 0$. The probability distortion function $h : [0,1] \to [0,1]$ is concave-convex.*

Notice that the exponential Bernoulli utility $u(x) = x^\alpha$ for $0 < \alpha \leqslant 1$ satisfies Assumption 9.2. We emphasize that RDU agents satisfying Assumption 9.2 with u linear on $[0,\infty)$ are not Yaari agents. This distinction will come back in our later discussion.

Denoting by $F_X^{-1}(1-t)$ the quantile function of X, the quantile representation of $\rho_h(X)$ is

$$\rho_h(X) = \int u(X)\,\mathrm{d}h \circ \mathbb{P} = \int_0^\infty h\left(\mathbb{P}\left(u\left(X\right) > t\right)\right)\,\mathrm{d}t = \int_0^1 u\left(F_X^{-1}(1-t)\right)\,\mathrm{d}h(t).$$

We denote by $\overline{h} : [0,1] \to [0,1]$ the concave envelope of h, which we define as the smallest concave function such that $h(t) \leqslant \overline{h}(t)$ for all $t \in [0,1]$. We have that $\overline{h} = h^{**}$ for

h^{**} the biconjugate of h; clearly $\overline{h}$ is a probability distortion. Since $\overline{h}$ dominates h pointwise for any $X \in \mathcal{X}$ we have

$$\rho_{\overline{h}}(X) = \int_0^1 u\left(F_X^{-1}(1-t)\right) \, \mathrm{d}\overline{h}(t) \geqslant \int_0^1 u\left(F_X^{-1}(1-t)\right) \, \mathrm{d}h(t) = \rho_h(X)$$

and $\rho_{\overline{h}}$ gives an upper-bound to the value attained by ρ_h. By construction, if the Bernoulli utility is linear, then the decision criterion $\rho_{\overline{h}}$ behaves as a risk-seeking agent, similarly to Section 9.4. Recalling the KT distortion h_{KT} is always concave-convex when $0 < \gamma < 1$, with strict concavity on a subset, we obtain that its concave envelope $\overline{h}_{\mathrm{KT}}$ is strictly concave on a subset.

Our goal now is to find conditions guaranteeing the Pareto-optimality of fair jackpot allocations. To do so we introduce a strengthening of the notion of distributional fairness:

Definition 9.5. A random vector $(X_1, \ldots, X_n) \in \mathcal{X}^n$ is exchangeable if $(X_1, \ldots, X_n) \overset{\mathrm{d}}{=} \left(X_{\pi(1)}, \ldots, X_{\pi(n)}\right)$ for all permutation $\pi \in \mathfrak{S}_n$ where $\mathfrak{S}_n$ is the set of all permutations on $[n]$.

Clearly if $(X_1, \ldots, X_n) \in \mathcal{X}^n$ is exchangeable then $X_i \overset{\mathrm{d}}{=} X_j$ for all $i \neq j$, so $(X_1, \ldots, X_n)$ is distributionally fair. Suppose that $X_i \geqslant 0$ for all $i \in [n]$ and that there is a uniform distribution independent of $(X_1, \ldots, X_n)$. Then $(X_1, \ldots, X_n)$ has an exchangeable counter-monotonic improvement $(Y_1, \ldots, Y_n)$ such that for all $i \in [n]$, $Y_i \geqslant 0$ and $\mathbb{P}(Y_i > 0) = 1/n$. That is:

Corollary 9.2. *In the setting of Theorem 9.1, we further assume that $(X_1, \ldots, X_n)$ is exchangeable. Then, there exists an exchangeable jackpot allocation $(Y_1, \ldots, Y_n) \in \mathbb{A}_n(X)$ such that $Y_i \geqslant_{\mathrm{cx}} X_i$ and $\mathbb{P}(Y_i > 0) \leqslant 1/n$ for $i \in [n]$.*

Having established that the counter-monotonic improvement of an exchangeable allocation can be exchangeable, we now consider the risk-sharing problem with agents that have a linear Bernoulli utility function on $[0, \infty)$.

Theorem 9.3. *Assume Assumptions 9.1 and 9.2. Let all agents be RDU agents with the same utility function $u(x) = ax$ for some $a > 0$ and probability distortion h such that $h = \overline{h}$ for $t \in [0, 1/n]$. Let $(X_1, \ldots, X_n) \in \mathbb{A}_n(X)$ be a non-negative exchangeable allocation. Then there exists an exchangeable jackpot allocation $(Y_1, \ldots, Y_n) \in \mathbb{A}_n(X)$ that Pareto improves upon $(X_1, \ldots, X_n)$.*

Theorem 9.4. *Assume Assumptions 9.1 and 9.2. Let all agents be RDU agents with the same utility function $u(x) = ax$ for some $a > 0$ and probability distortion h such that $h = \bar{h}$ for $t \in [0, 1/n]$. For $X > 0$, there exists an exchangeable jackpot Pareto-optimal allocation.*

Proof. We say an allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ satisfying the following equation is a sum optimal allocation for $\rho_{\bar{h}}$:

$$\sum_{i=1}^{n} \rho_{\bar{h}}(X_i) = \max_{(X_1,\dots,X_n)\in\mathbb{A}_n(X)} \sum_{i=1}^{n} \rho_{\bar{h}}(X_i).$$

We first show that there exists a sum optimal allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ for $\rho_{\bar{h}}$. As $u(x) = -\infty$ for $x < 0$, it is clear that $\rho_{\bar{h}}(X_i) \leqslant \rho_{\bar{h}}(X)$ for all $i \in [n]$. Hence, $\sum_{i=1}^{n} \rho_{\bar{h}}(X_i) \leqslant n\rho_{\bar{h}}(X) < \infty$ for all $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$. Let $u* = \sup\{\sum_{i=1}^{n} \rho_{\bar{h}}(X_i); (X_1, \dots, X_n) \in \mathbb{A}_n(X)\}$. For any $\epsilon > 0$, we can find $\mathbf{X}^{\epsilon} = (X_1^{\epsilon}, \dots, X_n^{\epsilon})$ such that $\sum_{i=1}^{n} \rho_{\bar{h}}(X_i^{\epsilon}) \geqslant u^* - \epsilon$. Now, take $\epsilon = 1/m$ for $m \in \mathbb{N}$, we can get a sequence of random vectors $\{\mathbf{X}^m = (X_1^m, \dots, X_n^m)\}_{m\in\mathbb{N}}$ such that $\sum_{i=1}^{n} \rho_{\bar{h}}(X_i^m) \geqslant u^* - 1/m$ for all $m \in \mathbb{N}$. By Helly's Selection Theorem, we can find a subsequence $\{\mathbf{X}^{m_k}\}_{k\in\mathbb{N}}$ such that $\{\mathbf{X}^{m_k}\}_{k\in\mathbb{N}}$ converges in distribution to $\mathbf{X} = (X_1, \dots, X_n)$. For any $i \in [n]$, we have $\{X_i^{m_k}\}_{k\in\mathbb{N}}$ converges in distribution to X_i and ess-sup$(X_i^{m_k}) \leqslant$ ess-sup(X) for all $k \in \mathbb{N}$. Hence, $\{X, X_i^{m_1}, X_i^{m_2}, \dots\}$ is h-uniformly integrable and $\rho_{\bar{h}}(X_i^{m_k}) \to \rho_{\bar{h}}(X_i)$ as $k \to \infty$ for all $i \in [n]$ by Wang et al. (2020b, Theorem 6). Furthermore, as $\sum_{i=1}^{n} X_i^{m_k} = X$ for all $k \geqslant 1$, we have $\sum_{i=1}^{n} X_i \overset{\mathrm{d}}{=} X$. Hence, we can always find $(X_1^*, \dots, X_n^*)$ with the same distribution as $(X_1, \dots, X_n)$ and $\sum_{i=1}^{n} X_i^* = X$ such that

$$u^* \geqslant \sum_{i=1}^{n} \rho_{\bar{h}}(X_i^*) = \lim_{k\to\infty} \sum_{i=1}^{n} \rho_{\bar{h}}(X_i^{m_k}) \geqslant \lim_{k\to\infty} \left(u^* - \frac{1}{m_k}\right) = u^*.$$

Therefore, $(X_1^*, \dots, X_n^*)$ is a sum optimal allocation for $\rho_{\bar{h}}$.

Let $\mathbf{X}^{\Pi} = (X_{\pi(1)}^*, \dots, X_{\pi(n)}^*)$ where $(\pi(1), \dots, \pi(n))$ is a random vector uniformly distributed in $\mathfrak{S}_n$. Hence, it is clear that $\mathbf{X}^{\Pi}$ is an exchangeable allocation in $\mathbb{A}_n(X)$ and the marginals of $\mathbf{X}^{\Pi}$ are

$$F_{X_i^{\Pi}}(x) = \frac{1}{n} \sum_{i=1}^{n} F_i(x) \quad \text{for all } i \in [n] \text{ and } x \in \mathbb{R}.$$

That is, the distribution of X_i^{Π} is a mixture of the distributions of $X_1^*, \dots, X_n^*$. Since $\bar{h}$ is concave, we have $\rho_{\bar{h}}$ concave on mixtures by Wang et al. (2020b, Theorem 3). Hence,

$\rho_{\bar{h}}(X_i^{\Pi}) \geqslant (1/n)\sum_{i=1}^n \rho_{\bar{h}}(X_i^*)$ for all $i \in [n]$. As a result, we have

$$\sum_{i=1}^n \rho_{\bar{h}}(X_i^{\Pi}) = n\rho_{\bar{h}}(X_1^{\Pi}) \geqslant \sum_{i=1}^n \rho_{\bar{h}}(X_i^*) = \max_{(X_1,\dots,X_n)\in\mathbb{A}_n(X)} \sum_{i=1}^n \rho_{\bar{h}}(X_i).$$

Hence, $\mathbf{X}^{\Pi}$ is also a sum optimal allocation for $\rho_{\bar{h}}$.

Let $\mathbf{Y} = (Y_1, \dots, Y_n)$ be an exchangeable jackpot Pareto improvement allocation of $\mathbf{X}^{\Pi}$ in Theorem 9.3. By the proof of Theorem 9.3, we have

$$\sum_{i=1}^n \rho_h(Y_i) = \sum_{i=1}^n \rho_{\bar{h}}(Y_i) \geqslant \sum_{i=1}^n \rho_{\bar{h}}(X_i^{\Pi}) = \max_{(X_1,\dots,X_n)\in\mathbb{A}_n(X)} \sum_{i=1}^n \rho_{\bar{h}}(X_i) \geqslant \max_{(X_1,\dots,X_n)\in\mathbb{A}_n(X)} \sum_{i=1}^n \rho_h(X_i).$$

where the last inequality comes from $\bar{h} \geqslant h$. As a result, we have $\mathbf{Y}$ satisfies

$$\sum_{i=1}^n \rho_h(Y_i) = \max_{(X_1,\dots,X_n)\in\mathbb{A}_n(X)} \sum_{i=1}^n \rho_h(X_i).$$

That is, $\mathbf{Y}$ is a sum optimal allocation for ρ_h. Hence, it is clear that $\mathbf{Y}$ is a Pareto optimal allocation. □

A direct implication of Theorem 9.3 is that if an agent with a linear Bernoulli utility uses a distortion function h that is strictly concave on the segment $[0, 1/n]$, then the same agent precisely behaves as the risk-seeking agent of Section 9.4 when comparing the two allocations.

When is n large and all agents have a linear Bernoulli utility, all fair Pareto-optimal allocations are counter-comonotonic, and the exchangeable jackpot allocation strictly Pareto dominates the exchangeable comonotonic allocation. Fair Pareto-optimal allocations thus cannot be comonotonic. However, we can have non-trivial unfair comonotonic allocations that are Pareto optimal:

Example 9.2. Assume all RDU agents have a linear Bernoulli utility function and the same distortion function h. Let $X \geqslant 0$ and consider the following allocation: $X_1 = X_2 = X/2$ and for all $i \neq 1, 2$, $X_i = 0$. This allocation is non-trivially comonotonic, and one can build a jackpot allocation where $Y_1 = X\mathbb{1}_{A_1}$ and $Y_2 = X\mathbb{1}_{A_2}$ with $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1/2$. Yet, the jackpot allocation need not improve upon the original allocation. This happens when the condition $h = \overline{h}$ for $t \in [0, 1/2]$ is not satisfied. The economic intuition is that the

Pareto optimality of comonotonic allocations can happen when some agents have a large enough share of the aggregate endowment so that gambling with others does not create a high enough reward.

The symmetry of behaviour between RDU agents and the risk-seeking agents of Section 9.4 goes beyond the cases considered above. Yaari agents are strictly risk seeking when the distortion function h is strictly concave, and we can reproduce the argument of Section 9.4 to show that Pareto-optimal allocations do not exist with strictly risk-seeking Yaari agents. Pareto-optimal allocations also do not exist for Yaari agents with concave-convex distortions when the number of agents n is large. This is seen by concavifying the distortion function h and observing that if $\overline{h}$ is concave and n, our Yaari agents behave as risk-seeking agents. Once again, the assumption that $u(x) = -\infty$ for $x < 0$ in Assumption 9.2 is instrumental in guaranteeing that Pareto-optimal allocations exist; it might be replaced by constraining the set of feasible allocations to non-negative allocations.

The observation that the exchangeable jackpot allocation sometimes leads to strict Pareto improvements hints at the possibility that Theorem 9.3 might be extended to some cases where the Bernoulli utility is strictly concave on $[0, \infty)$. While the general case is still an open question, the next section's cryptocurrency example shows that counter-monotonic payoff can indeed be preferred by RDU agents with concave Bernoulli utility when we restrict their choice set.

9.7 Cryptocurrency mining: to pool or not to pool

Let us consider n miners who need to decide whether they mine by themselves or join a mining pool. For all $i \in [n]$, the actions set is $\mathcal{A}_i = \mathcal{A} = \{H, P\}$, where H denotes mining by themselves (from "Home"), and P denotes joining the pool. We consider two types of miners. The first type of miner is behavioural, which we define as an RDU agent with concave-convex probability distortion functions. Let $y \in \mathbb{R}$ and let $\phi_y : [0, 1] \to \mathbb{R}$ be $\phi_y(x) = h(x)u(y/x)$. Behavioural miners satisfy the following assumption:

Assumption 9.3. *The utility Bernoulli function $u : \mathbb{R} \to \mathbb{R}$ is increasing and differentiable on $[0, \infty)$, weakly concave, satisfies $u(0) = 0$ and such that $u(x) = -\infty$ for all $x < 0$. The probability distortion function $h : [0, 1] \to [0, 1]$ is concave-convex. Further, there exists a unique a $p_0 \in (0, 1)$ such that for every $y \geqslant 0$, $\phi_y(x)$ decreases on $(0, p_0)$ and $\phi_y(p_0) \geqslant \phi_y(z)$ for all $z > p_0$.*

The second type of miner is the strictly risk-averse miner. We still denote by $\rho(X) = \int u(X)\, \mathrm{d}h \circ \mathbb{P}$ their decision criterion to unify notation, but do not specify the shape of u or h and simply assume that ρ is strictly risk-averse.[4]

Let $k \in \mathbb{N}_+$ denote the number of behavioural miners so that there are $n - k$ miners risk-averse miners. Each miner $i \in [n]$ has a computational power c_i, and the probability of mining the next coin is proportional to the total computing power as in Leshno and Strack (2020). Normalizing $\sum_{i=1}^n c_i = 1$ we can take an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and define the event

$$A_i := \{\omega \in \Omega : \text{ agent } i \text{ mines the coin}\}$$

so that $\mathbb{P}(A_i) = c_i$, all $i \in [n]$, and $(A_1, \ldots, A_n) \in \Pi_n$. Let $v > 0$ denote the given value of the coin. For all $i \in [n]$ we normalize the monetary payoff of mining from home as $v\mathbb{1}_{A_i}$ and set $u_i(0) = 0$ so that the expected payoff of a home miner i is

$$\rho_i(H) = h_i(c_i)u_i(v) + (1 - h(c_i))u_i(0) = h_i(c_i)u_i(v).$$

Let Po denote the set of agents that join the pool, i.e. $\mathrm{Po} = \{i \in [n] : a_i = P\}$. We assume that the pool uses the conditional mean risk-sharing rule so that the monetary payoff of agent $i \in \mathrm{Po}$ conditional on the pool mining the coin is $\frac{vc_i}{\sum_{j\in\mathrm{Po}} c_j}$. Note that $\bigcup_{j\in\mathrm{Po}} A_j$ is the event that the pool mines the coin, and the unconditional monetary payoff of agent $i \in \mathrm{Po}$ is

$$\frac{vc_i}{\sum_{j\in\mathrm{Po}} c_j}\mathbb{1}_{\bigcup_{j\in\mathrm{Po}} A_j}.$$

When Po is given the expected utility of miner $i \in \mathrm{Po}$ is thus

$$\rho_i(P) = h\left(\sum_{j\in\mathrm{Po}} c_j\right) u\left(\frac{vc_i}{\sum_{j\in\mathrm{Po}} c_j}\right).$$

[4]The criterion $\rho(X) = \int u(X)\, \mathrm{d}h \circ \mathbb{P}$ can be risk-averse in the following three cases: (1) u concave and h the identity; (2) u linear and h convex and (3) u concave can h convex.

Since we focus on pure-strategy Nash equilibria, we slightly abuse notation and denote by a_i the strategy profile of agent $i \in [n]$ playing action a_i with probability one.[5] Let $a_{-i} = (a_j)_{j \neq i} \in \mathcal{A}^{n-1}$ and let $\rho_i(a_i, a_{-i})$ be the expected utility of action a_i given the action profile a_{-i}. We let $\sigma_i : \mathcal{A}^{n-1} \rightrightarrows \mathcal{A}$ be the best-reply correspondence of agent i so that $\sigma_i(a_{-i}) = \arg\max_{a_i \in \mathcal{A}} \rho_i(a_i, a_{-i})$. A pure-strategy Nash equilibrium of the crypto mining game is a profile of actions $(a_1^*, \ldots, a_n^*) \in \mathcal{A}^n$ such that for every $i \in [n]$, $a_i^* \in \sigma_i\left(a_{-i}^*\right)$.

Notice now that

$$\mathbb{E}\left[\frac{vc_i}{\sum_{j\in \mathrm{Po}} c_j}\mathbb{1}_{\bigcup_{j\in \mathrm{Po}} A_j}\right] = \frac{vc_i}{\sum_{j\in \mathrm{Po}} c_j}\mathbb{P}\left(\bigcup_{j\in \mathrm{Po}} A_j\right) = vc_i = \mathbb{E}\left[v\mathbb{1}_{A_i}\right]$$

and $v\mathbb{1}_{A_i}$ is a mean-preserving spread of $\frac{vc_i}{\sum_{j\in \mathrm{Po}} c_j}\mathbb{1}_{\bigcup_{j\in \mathrm{Po}} A_j}$ since both are Bernoulli random variables.

Clearly if $a_{-i} = (H, \ldots, H)$ we have

$$\frac{vc_i}{\sum_{j\in \mathrm{Po}} c_j}\mathbb{1}_{\bigcup_{j\in \mathrm{Po}} A_j} = v\mathbb{1}_{A_i}$$

and $\rho_i(H, a_{-i}) = \rho(P, a_{-i})$. The following lemma is thus trivial:

Lemma 9.2. *The action profile $\mathbf{a}^* = (H, \ldots, H) \in \mathcal{A}^n$ constitutes a Nash equilibrium.*

The lemma means that it can happen that the pool never forms in equilibrium.

Proposition 9.9. *Let agent i be strictly risk averse. Then*

$$\sigma_i(a_{-i}) = \begin{cases} \{H, P\} & \text{if } a_j = H \text{ for all } j \neq i \\ P & \text{otherwise} \end{cases}$$

and P is a weakly dominant strategy for agent i.

Of course, this implies that H is a weakly dominated strategy for risk-averse agents and the Nash equilibrium $(H, \ldots, H) \in \mathcal{A}^n$ does not survive the iterated elimination of weakly dominated strategies. Further, if at least one agent chooses P, then all strictly risk-averse agents choose P. The equilibria where all risk-averse agents choose P are thus of greater interest.

[5] That is, we denote by a_i the strategy profile δ_{a_i}, where δ is again the Dirac delta function.

Proposition 9.10. *Let agent i be a RDU agent with decision criterion $\rho_h(X)$ satisfying Assumption 9.3. Then*

(i) If $c_i \leqslant p_0$ and $\phi_y(x)$ strictly decreases on $(0, p_0)$ we have

$$\sigma_i(a_{-i}) = \begin{cases} \{H, P\} & \text{if } a_j = H \text{ for all } j \neq i \\ H & \text{otherwise.} \end{cases}$$

(ii) If there is a $p^* \in [0,1]$ for which $h(t)/t$ is strictly increasing for $t > p^*$ and if $c_i \geqslant p^*$, then

$$\sigma_i(a_{-i}) = \begin{cases} \{H, P\} & \text{if } a_j = H \text{ for all } j \neq i \\ P & \text{otherwise.} \end{cases}$$

We immediately obtain the following corollary:

Corollary 9.3. *Let all behavioural agents $j \in [k]$ be such that $\phi_y(x)$ strictly decreases on $(0, p_0)$ and such that $c_j \in (0, p_0)$. Then the action profile $(a_i^*)_{i\in[n]}$ is a pure-strategy Nash Equilibrium, where*

$$a_i^* = \begin{cases} H & \text{if } j \in k \\ P & \text{otherwise.} \end{cases}$$

Moreover, $(a_i^*)_{i\in[n]}$ is the unique pure-strategy Nash Equilibrium left after performing the iterated deletion of weakly dominated strategies.

The takeaway of Proposition 9.10 and its corollary is that behavioural agents can have somewhat "bang-bang" strategies. On the one hand, they can behave as risk-seeking agents and mine from home. This happens when they have proportionally small computing power so that c_i is in $(0, p_0)$ a subset of the concave part of h. This result both complements and contrasts the results of the previous section, as RDU agents with strictly concave Bernoulli utility functions can behave as risk-seeking agents when their choice set is restricted.

On the other hand, behavioural agents can strictly prefer to join the pool when their computing power is large in proportion to the total computer power and their probability

of winning is high. This happens for two reasons. First, when $h(t)/t$ strictly increases for $t \in [p^*, 1]$, we have that $c_i \in [p^*, 1]$ corresponds to a set of probability where the agent is risk-averse. While the effect of risk aversion is clear, a second, less obvious reason comes from our assumption that the pool uses the conditional mean risk-sharing rule. This assumption implies that the high contribution of player i to the pool's computing power translates into a large share of the value of the coin. In other words, joining the pool lets player i "hedge" some of its risk. This property would not be obvious if the pool had a different risk-sharing rule. For instance, imposing an upper bound on the share that each individual miner can impact the equilibria.

The optimal strategy of behavioural agents is unclear when $c_i \in (p_0, p^*)$. This is because the best-reply of agent i can now vary as a function of the other agents' action. A complete equilibrium analysis is out of the scope of this article, but we believe it would be interesting to analyze the optimal pool formation as a function of both the computing power of the agents, the risk-sharing rule and the ability to divide its computing power among different pools.

9.8 Conclusion

Our main result, the counter-monotonic improvement theorem, lays the foundation for analyzing risk sharing with counter-monotonic allocations, the most extreme forms of negatively dependent allocation. This theorem allowed us to shed light on Pareto-optimal allocations when the risk is to be shared among risk-seeking agents, agents with a discontinuous Bernoulli utility function and RDU agents with inverted S-shaped probability distortion functions.

However, these characterizations of counter-monotonic Pareto-optimal allocations beg for more questions than it answers. Can competitive equilibria be counter-monotonic? If yes, under which conditions? What happens if we lift the assumption of the underlying risk and probability space that is well-understood by everyone?

The first two questions are natural extensions of our analysis of counter-monotonic risk sharing. While we analyze the competitive equilibria with risk-seeking agents in Section

9.9, it is unclear to us if and how these questions can be answered in the general case. The key issue is finding a price vector, as this is usually done using fixed-point theorems relying on some continuity property of the excess demand correspondence. Unfortunately, we do not currently know how the excess demand correspondences behave in general, as the Pareto optimality of counter-monotonic allocations requires "bang-bang" behaviour of the underlying preferences of agents.

We have many reasons to believe that the counter-monotonicity of Pareto-optimal allocations is possible when sharing risk under heterogeneous beliefs or ambiguity. Our stochastic representation of counter-monotonicity holds on general measurable spaces, and a superficial look at the no-betting allocations literature suggests that imposing strong assumptions on the beliefs held by agents might do the trick. In the case of ambiguity-averse agents, Billot et al. (2000) suggests that the emptiness of the intersection of the core of the agents' capacity is likely to be a necessary condition for the Pareto optimality of counter-monotonic allocations. This condition is unlikely to be sufficient, as this assumption does not rule out the comonotonicity of some agents' allocation. We hope that further investigations will shed light on this issue, as a characterization of extreme betting behaviour with ambiguity-averse agents would be highly counter-intuitive.

9.9 Competitive equilibria with risk-seeking agents

Fix an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let L^∞_+ be the set of nonnegative random variables in L^∞ which are not constantly 0.

Assumption 9.4. *All agents are expected utility agents with a common utility function u, which is convex on $[0, \infty)$. The total wealth in the economy is $X \in L^\infty_+$, and the vector of initial endowment, denoted by $(\xi_1, \dots, \xi_n) \in \mathbb{A}_n(X)$, has nonnegative components.*

We always make the above assumption.

9.9.1 Explicit construction of the equilibria

A pricing measure is a probability measure Q with $Q(X > 0) = 1$. Consider the individual optimization problem for agent $i \in [n]$:

$$\text{maximize } \mathbb{E}[u(X_i)] \quad \text{subject to } \mathbb{E}^Q[X_i] \leqslant \mathbb{E}^Q[\xi_i];\ 0 \leqslant X_i \leqslant X. \tag{9.7}$$

The tuple $(X_1, \dots, X_n, Q)$ is a *competitive equilibrium* if (a) individual optimality: X_i solves (9.7) for each $i \in [n]$; and (b) market clearance: $\sum_{i=1}^n X_i = X$. In this case, $(X_1, \dots, X_n)$ is an *equilibrium allocation*, and Q is an *equilibrium pricing measure*.

Proposition 9.11. *Let Q be given by*

$$\frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}} = \frac{u(X)}{X}\frac{1}{\mathbb{E}[u(X)/X]} \quad \textit{with the convention } 0/0 = 0, \tag{9.8}$$

and let

$$\begin{aligned} &(X_1, \dots, X_n) = (X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n}) \\ &\textit{for some } (A_1, \dots, A_n) \in \Pi_n \textit{ such that } \mathbb{E}^Q[X\mathbb{1}_{A_i}] = \mathbb{E}^Q[\xi_i] \textit{ for } i \in [n]. \end{aligned} \tag{9.9}$$

Then $(X_1, \dots, X_n, Q)$ is a competitive equilibrium.

Proof. Denote by $x_i = \mathbb{E}^Q[\xi_i]$ and $z = \mathbb{E}[u(X)/X] \geqslant 0$. It follows that

$$\mathbb{E}^Q[X_i] = \mathbb{E}^Q[X\mathbb{1}_{A_i}] = x_i,$$

and hence the budget constraint is satisfied for each $i \in [n]$. Moreover,

$$\mathbb{E}[u(X_i)] = \mathbb{E}[u(X)\mathbb{1}_{A_i}] = \mathbb{E}\left[X\frac{u(X)}{X}\mathbb{1}_{A_i}\right] = z\mathbb{E}^Q[X\mathbb{1}_{A_i}] = zx_i.$$

For any Y_i satisfying $0 \leqslant Y_i \leqslant X$ and the budget constraint $\mathbb{E}^Q[Y_i] \leqslant x_i$, using the fact that $x \mapsto u(x)/x$ is increasing, we have

$$\mathbb{E}[u(Y_i)] = \mathbb{E}\left[X_i\frac{u(Y_i)}{Y_i}\right] \leqslant \mathbb{E}\left[Y_i\frac{u(X)}{X}\right] = z\mathbb{E}^Q[Y_i] \leqslant zx_i = \mathbb{E}[u(X_i)].$$

Therefore, $(X_1, \dots, X_n, Q)$ satisfies individual optimality. Market clearance also holds, because $\sum_{i=1}^n \mathbb{1}_{A_i} = 1$. □

Remark 9.3. The proof of Proposition 9.11 only requires $x \mapsto u(x)/x$ to be an increasing function. Under our assumption $u(0) = 0$, this is a weaker condition than convexity of u.

9.9.2 Uniqueness of the equilibrium

Let $\mathcal{L}$ be the set of random variables Y in L^∞_+ such that there exists a standard uniform random variable independent of Y.

Theorem 9.5. *Suppose that u is strictly convex on $[0,\infty)$ and $X \in \mathcal{L}$.*

(i) All equilibrium allocations $(X_1,\dots,X_n)$ have the form (9.9).

(ii) If at least two of $\xi_1,\dots,\xi_n$ are not 0, then the equilibrium pricing measures is uniquely given by (9.8).

Proof. (i) Let $(X_1,\dots,X_n,Q)$ be a competitive equilibrium. By the counter-monotonic improvement theorem, there exists an allocation $(Y_1,\dots,Y_n) \in \mathbb{A}_n(X)$ such that $Y_i \geqslant_{\mathrm{cx}} X_i$ and $Y_i \geqslant 0$ for each $i \in [n]$. If $\mathbb{E}^Q[Y_i] < \mathbb{E}^Q[\xi_i]$ for some $i \in [n]$, then there exists $b > 0$ such that $\mathbb{E}^Q[Y_i + b(X - Y_i)] \leqslant \mathbb{E}^Q[\xi_i]$. Hence, $Y_i + b(X - Y_i)$ satisfies the budget constraint for agent i. Moreover, by strict convexity of u (implying strict increasing monotonicity), we have

$$\mathbb{E}[u(Y_i + b(X - Y_i))] > \mathbb{E}[u(Y_i)] \geqslant \mathbb{E}[u(X_i)],$$

and hence X_i is not optimal for agent i, a contradiction. Therefore, we conclude that $\mathbb{E}^Q[Y_i] \geqslant \mathbb{E}^Q[\xi_i]$ for all $i \in [n]$. Since $\sum_{i=1}^n \mathbb{E}^Q[Y_i] = \mathbb{E}^Q[X] = \sum_{i=1}^n \mathbb{E}^Q[X_i]$, we further obtain $\mathbb{E}^Q[Y_i] = \mathbb{E}^Q[\xi_i]$ for all $i \in [n]$. Hence, $(Y_1,\dots,Y_n)$ satisfies the budget constraint. For each $i \in [n]$, individual optimality gives $\mathbb{E}[u(X_i)] \geqslant \mathbb{E}[u(Y_i)]$ and convex order gives $\mathbb{E}[u(X_i)] \leqslant \mathbb{E}[u(Y_i)]$, together leading to $\mathbb{E}[u(X_i)] = \mathbb{E}[u(Y_i)]$, and thus $Y_i \overset{\mathrm{d}}{=} X_i$. The rest of the proof follows from the same argument as Theorem 9.2, justifying that $(X_1,\dots,X_n)$ is a jackpot allocation, thus with the form $(X\mathbb{1}_{A_1},\dots,X\mathbb{1}_{A_n})$ in (9.9).

(ii) Suppose that $(X_1,\dots,X_n,Q)$ is a competitive equilibrium. Using (i), we can write $(X_1,\dots,X_n) = (X\mathbb{1}_{A_1},\dots,X\mathbb{1}_{A_n})$ for some $(A_1,\dots,A_n) \in \Pi_n$ in (9.9). Let P be the conditional probability measure of $\mathbb{P}$ on $\{X > 0\}$, and let $p = \mathbb{P}(X > 0)$. Let $\eta = \mathrm{d}Q/\mathrm{d}P$ and define a probability measure R by

$$\frac{\mathrm{d}R}{\mathrm{d}Q} = \frac{X}{c}, \text{ where } c = \mathbb{E}^Q[X],$$

and Note that for any $A \in \mathcal{F}$, we have

$$\mathbb{E}[u(X\mathbb{1}_A)] = \frac{1}{p}\mathbb{E}^P[u(X\mathbb{1}_A)] = \frac{1}{p}\mathbb{E}^R\left[\frac{\mathrm{d}P}{\mathrm{d}Q}\frac{\mathrm{d}Q}{\mathrm{d}R}u(X)\mathbb{1}_A\right] = \frac{1}{p}\mathbb{E}^R\left[\frac{cu(X)}{\eta X}\mathbb{1}_A\right].$$

Denote by $Z = cu(X)/(\eta X)$. Individual optimality of $(X_1, \dots, X_n)$ implies that for any $i \in [n]$ and any $A \in \mathcal{F}$ satisfying $\mathbb{E}^Q[X\mathbb{1}_A] \leqslant \mathbb{E}^Q[X\mathbb{1}_{A_i}]$, we have

$$\mathbb{E}^R\left[Z\mathbb{1}_A\right] = p\mathbb{E}[u(X\mathbb{1}_A)] \leqslant p\mathbb{E}[u(X\mathbb{1}_{A_i})] = \mathbb{E}^R\left[Z\mathbb{1}_{A_i}\right].$$

Note that $\mathbb{E}^Q[X\mathbb{1}_A] \leqslant \mathbb{E}^Q[X\mathbb{1}_{A_i}]$ is equivalent to $R(A) \leqslant R(\mathbb{1}_{A_i})$. Take $A \in \mathcal{F}$ such that $R(A) = R(\mathbb{1}_{A_i})$ and Z and $\mathbb{1}_A$ are comonotonic. Suppose that Z is not a constant. The Fréchet-Hoeffding inequality gives

$$\operatorname{cov}(Z, \mathbb{1}_{A_i}) \geqslant \operatorname{cov}(Z, \mathbb{1}_A) \geqslant 0,$$

and $\operatorname{cov}(Z, \mathbb{1}_A) > 0$ if $R(A) \in (0, 1)$. Since at least two of $\xi_1, \dots, \xi_n$ are not 0, by (9.9), at least two of $A_1, \dots, A_n$ have positive probability under R. Therefore, $\operatorname{cov}(Z, \mathbb{1}_{A_i}) > 0$ for at least one i. However, $\sum_{i=1}^n \operatorname{cov}(Z, \mathbb{1}_{A_i}) = \operatorname{cov}(Z, 1) = 0$, a contradiction. Hence, Z is a constant. Therefore, η is equal to a constant times $u(X)/X$, showing that Q has the form (9.8). □

In case only one of $\xi_1, \dots, \xi_n$ is not 0, say ξ_i, the equilibrium allocation is $X_i = X$ and $X_j = 0$ for $j \in [n] \setminus \{i\}$, and the equilibrium pricing measure is arbitrary.

The equilibrium pricing density $\mathrm{d}Q/\,\mathrm{d}\mathbb{P}$ is increasing in X, and which is more expensive for states with larger X. This is in sharp contrast to the case of risk-averse expected utility agents, where the pricing density is cheaper for states with larger X.

9.9.3 Existence of the equilibrium

The next lemma shows that the competitive equilibrium in Proposition 9.11 always exists even without assuming the existence of a uniform random variable independent of X.

Lemma 9.3. *For any probability Q, there exists $(A_1, \dots, A_n) \in \Pi_n$ satisfying (9.9).*

Proof. Let U be a uniform transform of X (i.e., $F_X^{-1}(U) = X$ a.s. and U is uniformly distributed on $[0,1]$), and let $A_1 = \{0 \leqslant U \leqslant a_1\}$ where a_1 satisfies

$$\int X \mathbb{1}_{\{0 \leqslant U \leqslant a_1\}} \, \mathrm{d}Q = \mathbb{E}^Q[\xi_1].$$

Since $a \mapsto \int X \mathbb{1}_{\{0 \leqslant U \leqslant a\}} \, \mathrm{d}Q$ is continuous on $[0,1]$ and takes value in $[0, \mathbb{E}^Q[X]]$, such a_1 exists. Next, let $A_2 = \{a_1 \leqslant U \leqslant a_2\}$ where a_2 satisfies

$$\int X \mathbb{1}_{\{a_1 \leqslant U \leqslant a_2\}} \, \mathrm{d}Q = \mathbb{E}^Q[\xi_2].$$

Since $a \mapsto \int X \mathbb{1}_{\{a_1 \leqslant U \leqslant a\}} \, \mathrm{d}Q$ is continuous on $[a_1, 1]$ and takes value in $[0, \mathbb{E}^Q[X] - \mathbb{E}^Q[\xi_1]]$ (note that $\mathbb{E}^Q[X] - \mathbb{E}^Q[\xi_1] \geqslant \mathbb{E}^Q[\xi_2]$), such a_2 exists. Repeating this procedure yields the desirable $(A_1, \dots, A_n)$. □

Let Q be given in (9.8). Suppose $X \in \mathcal{L}$. In this case, the composition in Lemma 9.3 is much simpler: we can take A_i as an event independent of X with probability $\mathbb{E}^Q[\xi_i]/\mathbb{E}^Q[X]$ for each $i \in [n]$. Then, $(X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n}, Q)$ is a competitive equilibrium. It has a simple interpretation: the probability of winning the jackpot reward for agent i is $\mathbb{E}^Q[\xi_i]$, which is proportional to $\mathbb{E}[\xi_i u(X)/X]$. In particular, if $\xi_i = X/n$ for each $i \in [n]$, then we obtain a fair (exchangeable) jackpot allocation.

9.9.4 Welfare theorems

We now establish the first welfare theorem.

Proposition 9.12. *Every equilibrium allocation of $X \in \mathcal{L}$ is Pareto optimal.*

Proof. By Theorem 9.5, any equilibrium allocation $(X_1, \dots, X_n)$ has the form $(X_1, \dots, X_n) = (X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n})$ for some $(A_1, \dots, A_n) \in \Pi_n$. Note that

$$\sum_{i=1}^n \mathbb{E}[u(X_i)] = \sum_{i=1}^n \mathbb{E}[u(X)\mathbb{1}_{A_i}] = \mathbb{E}[u(X)].$$

Convexity of u implies $u(x+y) \geqslant u(x)+u(y)$ for all $x, y \geqslant 0$. For any allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$,

$$\sum_{i=1}^n \mathbb{E}[u(Y_i)] = \mathbb{E}\left[\sum_{i=1}^n u(Y_i)\right] \leqslant \mathbb{E}\left[u\left(\sum_{i=1}^n Y_i\right)\right] = \mathbb{E}[u(X)].$$

Therefore, $(X_1, \dots, X_n)$ is sum-optimal, and hence Pareto optimal. □

Proposition 9.12 also shows that all jackpot allocations are Pareto optimal in this setting. Next we establish the second welfare theorem.

Proposition 9.13. *Suppose that u is strictly convex on $[0,\infty)$. Every Pareto-optimal allocation of $X \in \mathcal{L}$ is an equilibrium allocation for some initial endowments.*

Proof. Suppose that $(X_1,\dots,X_n)$ is a Pareto-optimal allocation of X. By Theorem 9.2, every Pareto-optimal allocation is a jackpot allocation; that is, it admits a representation $(X_1,\dots,X_n) = (X\mathbb{1}_{A_1},\dots,X\mathbb{1}_{A_n})$ for some $(A_1,\dots,A_n) \in \Pi_n$. Let Q be given by (9.8). Further, let

$$a_i = \frac{\mathbb{E}^Q[X\mathbb{1}_{A_i}]}{\mathbb{E}^Q[X]}$$

and $\xi_i = a_i X$ for each $i \in [n]$. It follows that

$$\mathbb{E}^Q[X_i] = \mathbb{E}^Q[X\mathbb{1}_{A_i}] = a_i\mathbb{E}^Q[X] = \mathbb{E}^Q[\xi_i].$$

Therefore, (9.9) is satisfied. Using Proposition 9.11, we get that $(X_1,\dots,X_n,Q)$ is a competitive equilibrium. □

To summarize all results, we obtain the following theorem.

Theorem 9.6. *Suppose that u is strictly convex on $[0,\infty)$. For an allocation of $X \in \mathcal{L}$, the following are equivalent.*

(i) It is Pareto optimal;

(ii) it is an equilibrium allocation for some initial endowments;

(iii) it is a jackpot allocation.

9.10 Omitted proofs

9.10.1 Proofs of Section 9.3

Proof of Proposition 9.2. The "if" part follows from the fact that $\sum_{i=1}^n X_i = X$ and Theorem 7.1 of Chapter 7. We will show the "only if" part.

Assume $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is counter-monotonic. By Theorem 7.1 of Chapter 7.1, there exists $(A_1, \dots, A_n) \in \Pi_n$ such that

$$X_i = (X - m)\mathbb{1}_{A_i} + m_i \quad \text{for all } i \in [n],$$

where either $m_i = \text{ess-inf} X_i$ for $i \in [n]$ or $m_i = \text{ess-sup} X_i$ for $i \in [n]$, and $m = \sum_{i=1}^n m_i$. If $m_i = \text{ess-inf} X_i$ for all $i \in [n]$, we have $m = \sum_{i=1}^n \text{ess-inf}(X_i) \leqslant \text{ess-inf}(\sum_{i=1}^n X_i) \leqslant \text{ess-inf} X$. If $m_i = \text{ess-sup} X_i$ for all $i \in [n]$, we have $m = \sum_{i=1}^n \text{ess-sup}(X_i) \geqslant \text{ess-sup}(\sum_{i=1}^n X_i) \geqslant \text{ess-sup} X$. □

Proof of Proposition 9.3. Let $(Z_1, \dots, Z_n)$ be a probabilistic mixture of $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$, which are two jackpot allocations. It follows that $Z_i \geqslant 0$ for all $i \in [n]$ and $\mathbb{P}(Z_i \wedge Z_j > 0) = \lambda \mathbb{P}(X_i \wedge X_j > 0) + (1 - \lambda)\mathbb{P}(Y_i \wedge Y_j > 0) = 0$ for $i \neq j$. Therefore, using (9.6) we know that $(Z_1, \dots, Z_n)$ is a jackpot allocation. □

Proof of Theorem 9.1. The case $X = 0$ is trivial and will be excluded below. Let U be a standard uniform random variable independent of X. First we argue that we can assume that U is independent of $X_1, \dots, X_n$. Otherwise, we can find two iid standard uniform random variables U and V independent of X, and $(\hat{X}_1, \dots, \hat{X}_n)$ measurable to $\sigma(X, V)$ such that $(\hat{X}_1, \dots, \hat{X}_n, X)$ is identically distributed to $(X_1, \dots, X_n, X)$. Clearly, U is independent of $(\hat{X}_1, \dots, \hat{X}_n, X)$, and all desired statements follow if we could prove them for $(\hat{X}_1, \dots, \hat{X}_n)$.

Write

$$Z_i = \frac{\sum_{j=1}^i X_j}{X} \mathbb{1}_{\{X>0\}} \text{ for } i \in [n] \text{ and } Z_0 = 0.$$

Define the event $A_i = \{Z_{i-1} \leqslant U < Z_i\}$ for $i \in [n]$. Clearly, $A_1, \dots, A_n$ are disjoint and $\mathbb{P}(\bigcup_{i \in [n]} A_i) = 1$. Let $Y_i = X\mathbb{1}_{A_i}$ for $i \in [n]$. It is clear that $(Y_1, \dots, Y_n)$ is counter-monotonic. For $i \in [n]$,

$$\begin{aligned} \mathbb{E}\left[Y_i \mid X_1, \dots, X_n\right] &= \mathbb{E}\left[X\mathbb{1}_{\{Z_{i-1} \leqslant U < Z_i\}} \mid X_1, \dots, X_n\right] \\ &= \mathbb{E}\left[X(Z_i - Z_{i-1}) \mid X_1, \dots, X_n\right] = X\frac{X_i}{X}\mathbb{1}_{\{X>0\}} = X_i, \end{aligned}$$

where in the last equality we used the fact that $X_i = 0$ if $X = 0$. Therefore, X_i is a conditional expectation of Y_i, yielding the desired order $X_i \leqslant_{\text{cx}} Y_i$ via Jensen's inequality. □

9.10.2 Proofs of Section 9.4

Proof of Theorem 9.2. Suppose $(X_1, \ldots, X_n) \in \mathbb{A}_n(X)$ is a Pareto-optimal allocation. By Theorem 9.1, there is a jackpot allocation $(Y_1, \ldots, Y_n) \in \mathbb{A}_n(X)$ such that for all $i \in [n]$ and $Y_i \geqslant_{\mathrm{cx}} X_i$. As u_i is strictly convex, we have $\mathbb{E}[u(Y_i)] = \mathbb{E}[u(X_i)]$ by Pareto optimality of $(X_1, \ldots, X_n)$. By Shaked and Shanthikumar (2007, Theorem 3.A.43), we obtain that $Y_i =_{\mathrm{st}} X_i$ in the usual stochastic order, and thus $Y_i \overset{\mathrm{d}}{=} X_i$.

We now want to show that $(X_1, \ldots, X_n)$ is counter-monotonic. Let var and cov denote respectively the variance and covariance. For any given $1 \leqslant i < j \leqslant n$, we have that Y_i and Y_j are counter-monotonic, $X_i \overset{\mathrm{d}}{=} Y_i$ and $X_j \overset{\mathrm{d}}{=} Y_j$. Therefore, $\mathrm{cov}(X_i, X_j) \geqslant \mathrm{cov}(Y_i, Y_j)$. Furthermore, by the fact that $\sum_{i=1}^n X_i = \sum_{i=1}^n Y_i = X$, we have

$$\mathrm{var}(X) = \mathrm{var}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \sum_{j=1}^n \mathrm{cov}(Y_i, Y_j) \leqslant \sum_{i=1}^n \sum_{i=1}^n \mathrm{cov}(X_i, X_j) = \mathrm{var}\left(\sum_{i=1}^n X_i\right) = \mathrm{var}\left(X\right).$$

Hence, $\mathrm{cov}(X_i, X_j) = \mathrm{cov}(Y_i, Y_j)$ for $i, j \in [n]$.

By the Hoeffding's identity, we have for all $i \neq j$ and

$$\begin{aligned} &\iint \left(\mathbb{P}\left(X_i \leqslant t, X_j \leqslant s\right) - \mathbb{P}\left(Y_i \leqslant t\right)\mathbb{P}\left(Y_j \leqslant s\right)\right) \mathrm{d}t\, \mathrm{d}s \\ &= \iint \left(\mathbb{P}\left(Y_i \leqslant t, Y_j \leqslant s\right) - \mathbb{P}\left(X_i \leqslant t\right)\mathbb{P}\left(X_j \leqslant s\right)\right) \mathrm{d}t\, \mathrm{d}s. \end{aligned}$$

Given that (Y_i, Y_j) and (X_i, X_j) have the same marginals and (Y_i, Y_j) is counter-monotonic, we have $\mathbb{P}\left(X_i \leqslant t, X_j \leqslant s\right) \geqslant \mathbb{P}\left(Y_i \leqslant t, Y_j \leqslant s\right)$. Therefore, we have $\mathbb{P}\left(X_i \leqslant t, X_j \leqslant s\right) = \mathbb{P}\left(Y_i \leqslant t, Y_j \leqslant s\right)$ for almost every $t, s \in \mathbb{R}$. Thus, for every $i \neq j$, it is $\mathbb{P}\left(X_i > 0, X_j > 0\right) = \mathbb{P}\left(Y_i > 0, Y_j > 0\right) = 0$ and $(X_1, \ldots, X_n)$ is counter-monotonic, and further it is a jackpot allocation by (9.6), as desired. □

Proof of Proposition 9.4. (i) It is clear that any comonotonic allocation of x is the set of $(x_1, \ldots, x_n) \in \mathbb{R}^n$ such that $\sum_{i=1}^x X_i = x$. We first assume that the allocation $(x_1, \ldots, x_n) \in \mathbb{R}^n$ is not Pareto optimal; that is, there exists $(Y_1, \ldots, Y_n) \in \mathbb{A}_n(x)$ such that $\mathbb{E}[u_i(Y_i)] \geqslant \mathbb{E}[u(x_i)] = u(x_i)$ for all $i \in [n]$, with strict inequality for some $i \in [n]$. Let $y_i = \mathbb{E}[Y_i]$ for $i \in [n]$. We have that $(y_1, \ldots, y_n) \in \mathbb{A}_n(X)$ because

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \mathbb{E}[Y_i] = \mathbb{E}\left[\sum_{i=1}^n Y_i\right] = \mathbb{E}[x] = x.$$

Since u_i is concave we obtain that $u(y_i) \geqslant \mathbb{E}[u(Y_i)] \geqslant u(x_i)$ for all $i \in [n]$, with $u(y_i) > u(x_i)$ for some $i \in [n]$. Furthermore, as u_i is strictly increasing, we have $y_i \geqslant x_i$ for all $i \in [n]$ and $y_i > x_i$ for some $i \in [n]$. Therefore, $\sum_{i=1}^n y_i > \sum_{i=1}^n x_i = x$, a contradiction. Hence, the allocation $(x_1, \dots, x_n) \in \mathbb{R}^n$ is Pareto optimal.

(ii) Let $(X_1, \dots, X_n) \in \mathbb{A}_n(x)$ be a counter-monotonic allocation of x satisfying $m_i = 0$ for all $i \in [n]$. By Proposition 9.2 we have

$$(X_1, \dots, X_n) = (x\mathbb{1}_{A_1}, \dots, x\mathbb{1}_{A_n}) \text{ for some } (A_1, \dots, A_n) \in \Pi_n.$$

Let $p_i = \mathbb{P}(X_i)$ for $i \in [n]$. We have $\sum_{i=1}^n p_i = 1$.

Assume that $(X_1, \dots, X_n) \in \mathbb{A}_n(x)$ is not Pareto optimal. There is a $(Y_1, \dots, Y_n) \in \mathbb{A}_n(x)$ such that $\mathbb{E}[u_i(Y_i)] \geqslant \mathbb{E}[u_i(X_i)] = p_i u_i(x) + (1 - p_i)u(0)$ for all $i \in [n]$, with strict inequalities for some $i \in [n]$. It is clear that $Y_i \geqslant 0$ for all $i \in [n]$. By Theorem 9.1, we can always find a nonnegative counter-monotonic allocation $(Y_1', \dots, Y_n') \in \mathbb{A}_n(x)$ such that $Y_i' \geqslant_{\mathrm{cx}} Y_i$. As $\sum_{i=1}^n Y_i' = x$, we have

$$Y_i' = (x - m)\mathbb{1}_{B_i} + m_i \text{ for some } m_1, \dots m_n \geqslant 0, m = \sum_{i=1}^n m \leqslant x, \text{ and } (B_1, \dots, B_n) \in \Pi_n.$$

Furthermore, it is clear that the allocation $\left(\hat{Y}_1, \dots, \hat{Y}_n\right) = (x\mathbb{1}_{B_1}, \dots, x\mathbb{1}_{B_n}) \in \mathbb{A}_n(x)$ satisfies $\mathbb{E}\left[u_i\left(\hat{Y}_i\right)\right] \geqslant \mathbb{E}\left[u_i\left(Y_i'\right)\right] \geqslant \mathbb{E}[u_i(X_i)]$ for all $i \in [n]$ and $\mathbb{E}\left[u_i\left(\hat{Y}_i\right)\right] > \mathbb{E}[u_i(X_i)]$ for some $i \in [n]$. Let $q_i = \mathbb{P}(B_i)$ for $i \in [n]$ so we have $\sum_{i=1}^n q_i = 1$. On the other hand, we also have

$$q_i u_i(x) + (1 - q_i)u_i(0) = \mathbb{E}\left[u_i\left(\hat{Y}_i\right)\right] \geqslant \mathbb{E}[u(X_i)] = p_i u_i(x) + (1 - p_i)u_i(0)$$

for all $i \in [n]$ and strictly inequalities hold for some $i \in [n]$. That is, we have $q_i \geqslant p_i$ for all $i \in [n]$ and $q_i > p_i$ for some $i \in [n]$. As a result, $\sum_{i=1}^n q_i > \sum_{i=1}^n p_i = 1$, a contradiction. Hence, $(X_1, \dots, X_n)$ is Pareto optimal. □

9.10.3 Proofs of Section 9.5

Proof of Proposition 9.5. The allocation $(X_1, \dots, X_n)$ is feasible because $\sum_{i=1}^n X_i = n - 1 + (X - (n - 1)) = X$. Notice that since for all $i \in [n - 1]$ it is $\mathbb{P}(X_i = 1) = 1$ and since

$\mathbb{P}(X_n \geqslant 1) = \mathbb{P}(X \geqslant n)$ it holds that

$$\sum_{i=1}^{n} \rho_i(X_i) = \sum_{i=1}^{n} \mathbb{P}(X_i \geqslant 1) = n - 1 + \mathbb{P}(X \geqslant n).$$

Consider an alternative allocation $(Y_1, Y_2, \ldots, Y_n)$ that satisfies the feasibility condition $\sum_{i=1}^{n} Y_i = X$. It is

$$\sum_{i=1}^{n} \rho_i(Y_i) = \sum_{i=1}^{n} \mathbb{P}(Y_i \geqslant 1) = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \geqslant 1\}}\right] \leqslant n - 1 + \mathbb{P}(X \geqslant n)$$

and $(Y_1, \ldots, Y_n)$ cannot strictly improve upon $(X_1, \ldots, X_n)$. □

Proof of Proposition 9.6. Suppose by contraposition that $\mathbf{X}^*$ does not maximize $f_{\mathbf{1}}(\mathbf{X})$. Recalling that $\max_{\mathbf{X}} f_{\mathbf{1}}(\mathbf{X}) = n - \mathbb{P}(X < n)$ we have that

$$n - \mathbb{P}(X < n) > \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{X_i^* \geqslant 1\}}\right] = \sum_{i=1}^{n} p_i$$

for $p_i = \mathbb{P}(X_i \geqslant 1) \in [0, 1]$. Rearranging the previous inequality yields $\sum_{i=1}^{n}(1 - p_i) > \mathbb{P}(X < n)$. Let $\lambda_1, \ldots, \lambda_n$ be in decreasing order and notice that

$$\sum_{i=1}^{n} \lambda_i(1 - p_i) \geqslant \sum_{i=1}^{n} \lambda_n(1 - p_i) > \lambda_n \mathbb{P}(X < n).$$

Rearranging the previous inequality we have

$$\max_{\mathbf{X}} f_{\boldsymbol{\lambda}}(\mathbf{X}) \geqslant \sum_{i=1}^{n} \lambda_i - \lambda_n \mathbb{P}(X < n) > \sum_{i=1}^{n} \lambda_i p_i = \mathbb{E}\left[\sum_{i=1}^{n} \lambda_i \mathbb{1}_{\{X_i^* \geqslant 1\}}\right] = f_{\boldsymbol{\lambda}}(\mathbf{X}^*),$$

and hence $\mathbf{X}^*$ does not maximize $f_{\boldsymbol{\lambda}}(\mathbf{X})$. □

Proof of Proposition 9.7. Suppose by contradiction that $(X_1, \ldots, X_n)$ is a comonotonic, fair and optimal allocation.

Fairness implies that $\mathbb{P}(X_1 \geqslant 1) = \cdots = \mathbb{P}(X_n \geqslant 1)$. We claim that comonotonicity of $X_1, \ldots, X_n$ implies that $\{X_i \geqslant 1\} = \{X_j \geqslant 1\}$ for any $i, j = 1, \ldots, n$; otherwise, there exist $\omega \in \{X_i \geqslant 1\} \setminus \{X_j \geqslant 1\}$ and $\omega' \in \{X_j \geqslant 1\} \setminus \{X_i \geqslant 1\}$ such that $(X_i(\omega) - X_j(\omega))(X_i(\omega') - X_j(\omega')) < 0$. Therefore, $\mathbb{P}(X_i \geqslant 1) = \mathbb{P}(X \geqslant n)$ for all $i = 1, \ldots, n$. This implies

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{X_i \geqslant 1\}}\right] = n(1 - \mathbb{P}(X < n)) < n - \mathbb{P}(X < n)$$

which shows that $(X_1, \ldots, X_n)$ is not optimal. □

Proof of Corollary 9.1. Assume that $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is a constrained Pareto-optimal allocation. Let $A_i = \{X_i \geqslant 1\}$. As $X_i \geqslant 0$ and $\mathbb{P}(X < 2) = 1$, we have $\mathbb{P}(A_i \cap A_j) = 0$ for all $i, j \in [n]$ such that $i \neq j$; otherwise, we will have $\mathbb{P}(X \geqslant 2) > 0$. Let $B_i = \left(\bigcap_{j\neq i}^n A_j^c\right) \cap A_i$. It is clear that $B_1, \dots, B_n$ are disjoint. Furthermore,

$$\mathbb{P}(A_i) = \mathbb{P}(B_i) + \mathbb{P}\left(A_i \cap \left(\bigcup_{j\neq i}^n A_j\right)\right) = \mathbb{P}(B_i) + \mathbb{P}\left(\bigcup_{j\neq i}^n A_i \cap A_j\right) = \mathbb{P}(B_i).$$

Let $X = \sum_{i=1}^n X_i$. Take the allocation

$$Y_i = X\mathbb{1}_{B_i} \text{ for } i \in [n-1] \text{ and } Y_n = X\mathbb{1}_{\Omega\setminus\bigcup_{i=1}^{n-1} B_i}.$$

It is clear that $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$. Note that $B_i \subseteq A_i \subseteq \{X \geqslant 1\}$ for $i \in [n]$. For $i \in [n-1]$,

$$\mathbb{P}(Y_i \geqslant 1) = \mathbb{P}(X\mathbb{1}_{B_i} \geqslant 1) = \mathbb{P}\left(\{X \geqslant 1\} \cap B_i\right) = \mathbb{P}(B_i) = \mathbb{P}(X_i \geqslant 1).$$

For $i = n$, as $B_n \subseteq \Omega \setminus \bigcup_{i=1}^{n-1} B_i$, we have

$$\mathbb{P}(Y_n \geqslant 1) = \mathbb{P}\left(\{X \geqslant 1\} \cap \left(\Omega \setminus \bigcup_{i=1}^{n-1} B_i\right)\right) \geqslant \mathbb{P}(B_n) = \mathbb{P}(X_n \geqslant 1).$$

Hence, $\rho_i(Y_i) \geqslant \rho_i(X_i)$ for all $i \in [n]$. As $(X_1, \dots, X_n)$ is Pareto optimal, we have $\rho_i(Y_i) = \rho_i(X_i)$ for all $i \in [n]$ □

Proof of Proposition 9.8. Fairness implies that $a\mathbb{E}[X_1]+\mathbb{P}(X_1 \geqslant 1) = \cdots = a\mathbb{E}[X_n]+\mathbb{P}(X_n \geqslant 1)$. We claim that comonotonicity of $X_1, \dots, X_n$ implies that $\{X_i \geqslant 1\} = \{X_j \geqslant 1\}$ for any $i, j = 1, \dots, n$; otherwise, there exist $\omega \in \{X_i \geqslant 1\} \setminus \{X_j \geqslant 1\}$ and $\omega' \in \{X_j \geqslant 1\} \setminus \{X_i \geqslant 1\}$ such that $(X_i(\omega) - X_j(\omega))(X_i(\omega') - X_j(\omega')) < 0$. Therefore, $\mathbb{P}(X_i \geqslant 1) = \mathbb{P}(X \geqslant n)$ for all $i = 1, \dots, n$. Summing we have

$$\sum_{i=1}^n (a\mathbb{E}[X_i] + \mathbb{P}(X_i \geqslant 1)) = a\mathbb{E}[X] + \sum_{i=1}^n \mathbb{P}(X_i \geqslant 1) = a\mathbb{E}[X] + n\mathbb{P}(X \geqslant n).$$

Since $\mathbb{P}(X \geqslant n) < 1$, the allocation $(Y_1, \dots, Y_n) = (\mathbb{1}_{A_i^C} + \mathbb{1}_{A_i}(X - (n-1)))_{i\in[n]}$ Pareto dominates $(X_1, \dots, X_n)$, where $(A_1, \dots, A_n) \in \Pi_n$ is such that $\mathbb{P}(A_i) = 1/n$ for all $i \in [n]$. □

9.10.4 Proofs of Section 9.6

Proof of Corollary 9.2. Following the proof of Theorem 9.1, we have

$$\mathbb{P}(A_i) = \mathbb{E}[\mathbb{E}[\mathbb{1}_{Z_{i-1} \leqslant U < Z_i} | X_1, \dots, X_n]] = \mathbb{E}\left[\frac{X_i}{X} \mathbb{1}_{\{X>0\}}\right].$$

Since $(X_1, \dots, X_n)$ is exchangeable, we have $\mathbb{P}(A_i) = \mathbb{P}(A_j)$ for all $i \neq j$. Hence, $(Y_1, \dots, Y_n) = (X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n})$ is an exchangeable jackpot allocation with $Y_i \geqslant_{\mathrm{cx}} X_i$ for all $i \in [n]$. As $A_1, \dots, A_n$ are disjoint and $(Y_1, \dots, Y_n)$ is exchangeable, we have

$$n\mathbb{P}(Y_i > 0) = \sum_{j=1}^{n} \mathbb{P}(Y_j > 0) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1$$

for all $i \in [n]$. Hence, we have (iv). □

Proof of Theorem 9.3. By Corollary 9.2, there exists an exchangeable jackpot allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$ such that $Y_i \geqslant_{\mathrm{cx}} X_i$ and $\mathbb{P}(Y_i > 0) \leqslant 1/n$ for all $i \in [n]$.

As $\bar{h} \geqslant h$, we have $\rho_h(X) \leqslant \rho_{\bar{h}}(X)$. Since $\mathbb{P}(Y_i > 0) \leqslant 1/n$, we have $\mathbb{P}(u(Y_i) > 0) = \mathbb{P}(Y_i > 0) \leqslant 1/n$. Given that $h = \bar{h}$ for $t \in [0, 1/n]$, we have

$$\rho_h(Y_i) = \int_0^\infty h\left(\mathbb{P}\left(u(Y_i) > t\right)\right) \mathrm{d}t = \int_0^\infty \bar{h}\left(\mathbb{P}\left(u(Y_i) > t\right)\right) \mathrm{d}t = \rho_{\bar{h}}(Y_i).$$

As $Y_i \geqslant_{\mathrm{cx}} X_i$, we have $u(Y_i) = aY_i \geqslant_{\mathrm{cx}} aX_i = u(X_i)$. By the fact that $\bar{h}$ is concave, we get $\rho_{\bar{h}}(Y_i) \geqslant \rho_{\bar{h}}(X_i)$. In conclusion, we have $\rho_h(X_i) \leqslant \rho_{\bar{h}}(X_i) \leqslant \rho_{\bar{h}}(Y_i) = \rho_h(Y_i)$ for all $i \in [n]$. □

9.10.5 Proofs of Section 9.7

Proof of Proposition 9.9. Suppose there is a $j \neq i$ such that $a_j = P$. Then, it is clear that $v\mathbb{1}_{A_i} >_{\mathrm{cx}} \frac{vc_i}{\sum_{j \in \mathrm{Po}} c_j} \mathbb{1}_{\bigcup_{j \in \mathrm{Po}} A_j}$ and by strict risk-aversion we have $\rho_i(P, a_{-i}) > \rho_i(H, a_{-i})$, as desired. □

Proof of Proposition 9.10. (i) If $c_i \leqslant p_0$ then by Assumption 9.3 we have

$$\rho_i(H, a_{-i}) = u_i(v) h_i(c_i) = u_i\left(\frac{vc_i}{c_i}\right) h_i(c_i) \geqslant u_i\left(\frac{vc_i}{\sum_{j \in \mathrm{Po}} c_j}\right) h_i\left(\sum_{j \in \mathrm{Po}} c_j\right) = \rho_i(P, a_{-i}).$$

If there is at least one coplayer $j \neq i$ such that $a_j = P$ then the strict inequality obtains whenever ϕ_y strictly decreases on $(0, p_0)$.

(ii) By concavity of u we have $au(x/a) \geqslant bu(x/b)$ for $a > b$. Suppose there is a $p^* \in [0,1]$ for which $h(t)/t$ is strictly increasing for $t > p^*$. If $c_i \geqslant p^*$ and there is at least one agent in the pool, we obtain

$$\rho_i(H, a_{-i}) = c_i u_i \left(\frac{vc_i}{c_i} \right) \frac{h_i(c_i)}{c_i} < \left(\sum_{j \in \mathrm{Po}} c_j \right) u_i \left(\frac{vc_i}{\sum_{j \in \mathrm{Po}} c_j} \right) \frac{h_i \left(\sum_{j \in \mathrm{Po}} c_j \right)}{\sum_{j \in \mathrm{Po}} c_j} = \rho_i(P, a_{-i}).$$

Hence, $\sigma_i(a_{-i}) = P$. □

Chapter 10

Invariant correlation under marginal transforms

10.1 Introduction

The (Pearson) correlation coefficient is one of the most popular measures of quantifying the strength of statistical dependence. It is also called the linear correlation coefficient, reflecting on the fact that it measures linear dependence among random variables. As such, the correlation coefficient is preserved under common linear transforms. When non-linear transforms are applied, the correlation coefficient typically changes except some specific dependence structures such as independence. In fact, this invariance is a useful and notable property of independent samples, and plays a fundamental role in statistics and probability theory.

To define this invariance property more formally, a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with $\mathrm{Var}(X_i) \in (0, \infty)$, $i \in [d] = \{1, \ldots, d\}$, is said to have an invariant correlation matrix $R = (r_{ij})_{d \times d}$ if

$$\mathrm{Corr}(X_i, X_j) = \mathrm{Corr}(g(X_i), g(X_j)) = r_{ij} \tag{10.1}$$

for all measurable functions $g : \mathbb{R} \to \mathbb{R}$ such that $\mathrm{Corr}(g(X_i), g(X_j))$ is well defined for all $i, j \in [d]$. Although this property appears strong, it is satisfied by some specific models

existing in the literature. One example is the dataset with duplicate records, where some individuals appear multiple times. Such a dataset may arise when the data are collected from multiple sources and curated by multiple groups. This phenomenon of *sample duplication* also occurs in bootstrap samples where observations are resampled with replacement. Another example is the *conformal p-value*, a useful statistical tool in machine learning and non-parametric inference (Vovk et al., 2005). The recent work Bates et al. (2023) showed that (10.1) holds for a set of null conformal p-values to discuss a broad class of combination tests.

Motivated by these specific models, this chapter addresses the following questions. (i) How large is the class of models with this invariance property? (ii) Whether and how is this property connected to other dependence concepts? (iii) How can we leverage this property in statistical applications? (iv) What if the invariance property is confined to a smaller class of transforms relevant for applications, such as the class of monotone transforms?

Since (10.1) is essentially a bivariate property, we first focus on the case when $d = 2$. It is straightforward to check that a pair of random variables (X, Y) has an invariant correlation $r \in [-1, 1]$ if X and Y have an identical marginal F and their joint distribution H is given by

$$H(x, y) = r \min(F(x), F(y)) + (1 - r)F(x)F(y), \quad x, y \in \mathbb{R}. \tag{10.2}$$

We say that this model or its distribution is *r-Fréchet* due to its connection to the Fréchet copulas; see Denuit et al. (2005); Yang et al. (2006) for applications in risk modeling. If $r \in [0, 1]$, this model is a probabilistic mixture of comonotonic and independent cases, where two random variables X and Y are said to be *comonotonic* if $X = f(Z)$ and $Y = h(Z)$ for some random variable Z and two increasing functions $f, h : \mathbb{R} \to \mathbb{R}$. All terms like "increasing" and "decreasing" are in the non-strict sense in this chapter.

Since invariant correlation is a strong property, one may wonder whether models of the form (10.2) are exhaustive. It turns out that this is not the case, but not far away from the truth. As one of our main contributions, we give a complete characterization of all bivariate random vectors having an invariant correlation (10.1) using the asymmetric analogs of independence and the r-Fréchet model, which we call *quasi-independence* (Definition 10.2)

Table 10.1: Summary of characterization results for bivariate distributions with invariant correlation, where QI stands for quasi-independence, QF stands for the quasi-Fréchet model, IN stands for independence, IS stands for models with identical marginal supports, and $\varnothing$ means no such model.

<table>
<tr><th colspan="2">marginals F, G</th><th>$r = 0$</th><th>$r \neq 0$</th></tr>
<tr><td colspan="2">$F = G$</td><td>QI (Theorem 10.1/10.3)</td><td>QF (Theorem 10.3)</td></tr>
<tr><td rowspan="3">$F \neq G$</td><td>both are bi-atomic</td><td rowspan="2">IN (Proposition 10.4)</td><td>IS (Proposition 10.5)</td></tr>
<tr><td>precisely one is bi-atomic</td><td rowspan="2">$\varnothing$ (Theorem 10.2)</td></tr>
<tr><td>both are not bi-atomic</td><td>QI (Theorem 10.1)</td></tr>
</table>

and the *quasi-r-Fréchet model* (Definition 10.3), respectively. Our characterization results are summarized in Table 10.1. We prove that zero invariant correlation is equivalent to quasi-independence if neither of the marginals is bi-atomic (i.e., supported on two-points), and to independence otherwise. We also find that, for non-identical marginals, non-zero invariant correlation is not possible except when both marginals are bi-atomic. For the case of identical marginals, it turns out that a model has a non-zero invariant correlation r if and only if it is quasi-r-Fréchet. In particular, when X and Y are exchangeable (i.e., $(X, Y) \overset{\mathrm{d}}{=} (Y, X)$, where $\overset{\mathrm{d}}{=}$ stands for equality in distribution), then the r-Fréchet model (10.2) characterizes the model (X, Y) with invariant correlation r.

Such a complete characterization cannot be expected for a general multivariate case since (10.1) is a requirement on bivariate margins. For a general $d \geqslant 2$, we prove that the set of all invariant correlation matrices attained by random vectors with continuous marginals coincides with the *clique partition polytope* (Grötschel and Wakabayashi, 1990). Based on this characterization, we provide a numerical procedure to check whether a given matrix is admissible as an invariant correlation matrix. For $k \in \mathbb{N}$ and iid observations $Y_1, \ldots, Y_k$ from

a continuous and strictly increasing distribution F, we then consider the model:

$$\mathbf{X} = \Gamma\mathbf{Y} = (\mathbf{Z}_1^\top\mathbf{Y}, \ldots, \mathbf{Z}_d^\top\mathbf{Y}), \quad \text{where } \Gamma = (\mathbf{Z}_1, \ldots, \mathbf{Z}_d)^\top \text{ and } \mathbf{Y} = (Y_1, \ldots, Y_k)^\top, \quad (10.3)$$

with $\mathbf{Z}_i$, $i \in [d]$, being $\{0,1\}^k$-valued Bernoulli random vector such that $\mathbf{Z}_i^\top\mathbf{Z}_i = 1$ and Γ being independent of $\mathbf{Y}$. We show that the model (10.3) has an identical marginal F and an invariant correlation matrix; moreover, for $k \geqslant d$, this model accommodates any admissible invariant correlation matrix. In addition, we show that this model has *positive regression dependence* (Lehmann, 1966), an important dependence concept for controlling p-values in the context of multiple testing (Benjamini and Yekutieli, 2001). By associating the model (10.3) with a dataset containing duplicate records, we then apply our results to the problem of *sample duplication*, which is known to have adverse effects in statistical inference and machine learning (Allamanis, 2019; Zhao et al., 2021). We exemplify how to leverage the invariant correlation property of such a dataset for more efficient statistical inference.

In many applications, only increasing transforms of the random variables are relevant since non-increasing transforms do not preserve the joint distribution of the ranks of the variables. In view of this, we also study a variant of invariant correlation where we require (10.1) to hold only for increasing transforms instead of all transforms. By definition, this requirement is weaker than the original invariant correlation property (10.1). A natural question is whether this formulation allows for more models than those characterized by (10.1). It turns out that invariant correlation confined to increasing transforms does not accommodate more models except for the case when both random variables are bi-atomic.

This chapter is organized as follows. In Section 10.2, we begin with introducing notation, definitions and basic properties of invariant correlation. Section 10.3 is then devoted to a full characterization of invariant correlation in the bivariate setting. Invariant correlation matrix is studied in Section 10.4, where a model admitting any prescribed invariant correlation matrix is proposed. This section also explores positive regression dependence of this model and an application to the problem of sample duplication. Section 10.5 provides the characterization of invariant correlation under increasing transforms. Discussion and directions for future research are given in Section 10.6. Proofs and auxiliary results are deferred to Sections 10.7 and 10.8.

10.2 Definitions and basic properties

In this section, we provide definitions and basic properties of invariant correlation and related models. We fix an atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$ throughout this chapter. Denote by $\mathcal{L}^2$ the set of all non-degenerate real-valued random variables with finite variance. For $d \in \mathbb{N}$, denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-algebra on $\mathbb{R}^d$. We also write $\mathbf{0}_d = (0, \dots, 0)^\top$ and $\mathbf{1}_d = (1, \dots, 1)^\top$ for the d-dimensional vector of zeros and ones, respectively. For a random vector $\mathbf{X}$, denote by $F_{\mathbf{X}}$ its distribution function.

For $X, Y \in \mathcal{L}^2$, the *Pearson correlation coefficient* of (X, Y) is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}.$$

We are interested in the following invariance of correlation under certain transforms of (X, Y). In this context, we call a function $g : \mathbb{R} \to \mathbb{R}$ *admissible* (for (X, Y), omitted if clear) if it is measurable and $\mathrm{Corr}(g(X), g(Y))$ is well defined, i.e., $g(X), g(Y) \in \mathcal{L}^2$. Strictly monotone functions with bounded derivatives are always admissible.

Definition 10.1. Let $r \in [-1, 1]$. A bivariate random vector (X, Y) is said to have an *invariant correlation* r if

$$\mathrm{Corr}(X, Y) = \mathrm{Corr}(g(X), g(Y)) = r \quad \text{for all admissible functions } g. \tag{10.4}$$

The set of all bivariate random vectors with invariant correlation r is denoted by IC_r. Moreover, let $\mathrm{IC} = \bigcup_{r \in [-1,1]} \mathrm{IC}_r$, which is the set of all bivariate random vectors having any invariant correlation.

The concept of invariance correlation can be naturally formulated dimension larger than 2. For $d \geqslant 2$, a d-dimensional random vector $\mathbf{X}$ is said to have an *invariant correlation matrix* $R = (r_{ij})_{d \times d}$ if $(X_i, X_j) \in \mathrm{IC}_{r_{ij}}$ for every pair $(i, j) \in [d]^2$. The set of all such random vectors is denoted by IC_R. For a distribution function H on $\mathbb{R}^d$, we write $\mathbf{X} \sim H$ if $\mathbf{X}$ is distributed according to H, and $H \sim \mathrm{IC}_R$ means $\mathbf{X} \in \mathrm{IC}_R$ for some $\mathbf{X} \sim H$. Denote by $\mathrm{IC}^d = \bigcup_{R \in \mathcal{P}_d} \mathrm{IC}_R$ the collection of all d-dimensional random vectors with invariant correlation matrix, where $\mathcal{P}_d$ is the set of all $d \times d$ correlation matrices. A measurable function g is admissible for $\mathbf{X}$ if $g(X_i) \in \mathcal{L}^2$ for each $i \in [d]$. Note that $\mathrm{IC}^2 = \mathrm{IC}$ is a special case. Although the problem

of invariance correlation is more general in dimension $d \geqslant 2$, its theoretical challenge lies mostly in the case $d = 2$, because the correlation coefficient is naturally a bivariate concept.

Let us first look at a few basic examples of invariant correlation.

Example 10.1. If a d-dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)$ is independent, then so is $(g(X_1), \dots, g(X_d))$ for every measurable g. Hence, we have $\mathbf{X} \in \mathrm{IC}_{I_d}$, where $I_d = \mathrm{diag}(\mathbf{1}_d)$. In dimension $d = 2$, we have $(X_1, X_2) \in \mathrm{IC}_0$.

Example 10.2. Suppose that A is a random event with $\mathbb{P}(A) = r \in \mathcal{I}$, and let $\mathbf{X} = (X, \dots, X)$ and $\mathbf{X}^\perp = (X_1, \dots, X_d)$, where $X, X_1, \dots, X_d$ are iid random variables independent of A. Let

$$\mathbf{Y} = (Y_1, \dots, Y_d) = \mathbf{X}\mathbb{1}_A + \mathbf{X}^\perp \mathbb{1}_{A^c}. \tag{10.5}$$

It is easy to verify that, for $i, j \in [d]$, $i \neq j$, and for every admissible g,

$$\mathrm{Corr}(g(Y_i), g(Y_j)) = \frac{\mathrm{cov}(g(Y_i), g(Y_j))}{\mathrm{var}(g(X))} = \frac{\mathrm{var}(g(X))\mathbb{P}(A)}{\mathrm{var}(g(X))} = \mathbb{P}(A) = r.$$

Hence, we have $\mathbf{Y} \in \mathrm{IC}^d$ where all off-diagonal elements in the invariant correlation matrix equal r.

The following simple properties follow directly from the definition of invariant correlation, and will be used repeatedly in the subsequent analyses.

Proposition 10.1. *If $\mathbf{X} = (X_1, \dots, X_d) \in \mathrm{IC}_R$ for $R \in \mathcal{P}_d$ and h is an admissible function for $\mathbf{X}$, then $(h(X_1), \dots, h(X_d)) \in \mathrm{IC}_R$.*

For a d-dimensional random vector $\mathbf{X}$, its copula C is a distribution function on $[0,1]^d$ with standard uniform marginals such that $F_{\mathbf{X}}(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$ for $(x_1, \dots, x_n) \in \mathbb{R}^n$, where $F_1, \dots, F_n$ are the marginals of $\mathbf{X}$; see Joe (2014) for background on copulas. A copula of $\mathbf{X}$ is unique on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_d)$, where $\mathrm{Ran}(F)$ is the range of a distribution funtion F. In particular, if $\mathbf{X}$ has continuous marginals, then its copula is unique. It is known that every d-dimensional copula C satisfies $W_d(\mathbf{u}) \leqslant C(\mathbf{u}) \leqslant M_d(\mathbf{u})$ for all $\mathbf{u} = (u_1 \dots, u_d) \in \mathcal{I}^d$, where $W_d(\mathbf{u}) = \max(u_1 + \cdots + u_d + 1 - d, 0)$ and $M_d(\mathbf{u}) = \min(u_1, \dots, u_d)$ are called *Fréchet bounds*. Note that M_d is a d-dimensional copula for every

$d \geqslant 2$, and W_d is a copula only when $d = 2$. These bounds, together with the *independence copula* $\Pi_d(\mathbf{u}) = \prod_{i=1}^d u_i$, $\mathbf{u} \in \mathcal{I}^d$, play fundamental roles in dependence analysis.

For $d = 2$, we write $M = M_2$, $\Pi = \Pi_2$ and $W = W_2$ in short. A (bivariate) *Fréchet copula* is a 2-dimensional copula defined as

$$C^{\mathrm{F}}_{r,s}(u, v) = rM(u, v) + sW(u, v) + (1 - r - s)\Pi(u, v), \quad (u, v) \in \mathcal{I}^2, \tag{10.6}$$

which is parametrized by $r, s \in \mathcal{I}$ such that $r + s \leqslant 1$. If $s = 0$ in (10.6), a Fréchet copula is called *positive* and we denote it by C^{F}_r. In fact, when $d = 2$ and the identical distribution of X, X_1, X_2 in Example 10.2 is continuous, then $\mathbf{X}$ and $\mathbf{X}^\perp$ has the copula M and Π, respectively. Moreover, the copula of $\mathbf{Y}$ is the positive Fréchet copula C^{F}_r. It is easy to see that a bivariate random vector (Y_1, Y_2) has a positive Fréchet copula C^{F}_r if and only if the stochastic representation (10.5) with $d = 2$ holds almost surely for some event A such that $\mathbb{P}(A) = r$.

Let $\mathbf{X}$ be a random vector with continuous and identical marginals. A useful consequence of Proposition 10.1 is that it suffices to consider the copula of $\mathbf{X}$ to analyze invariant correlation of $\mathbf{X}$.

Corollary 10.1. *Suppose that $\mathbf{X} \sim H$ has identical continuous and strictly increasing marginals, a correlation matrix $R \in \mathcal{P}_d$ and a copula C. Then $H \sim \mathrm{IC}_R$ if and only if $C \sim \mathrm{IC}_R$.*

We next introduce some special structures on bivariate distributions, which turn out to be equivalent to invariant correlation in Sections 10.3 and 10.5. We first present a concept of dependence which is close to independence.

Definition 10.2. We say that a random vector $(X, Y) \sim H$ with marginals F and G is *quasi-independent* if

$$H(x, y) + H(y, x) = F(x)G(y) + F(y)G(x) \quad \text{for all } x, y \in \mathbb{R}. \tag{10.7}$$

By a standard probabilistic argument, (10.7) is equivalent to

$$\mathbb{P}(X \in A, Y \in B) + \mathbb{P}(X \in B, Y \in A) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) + \mathbb{P}(X \in B)\mathbb{P}(Y \in A) \tag{10.8}$$

for all $A, B \in \mathcal{B}(\mathbb{R})$; see Lemma 10.1 in Section 10.8.1. Clearly, independence implies quasi-independence, but the opposite implication is not true. Quasi-independence plays an important role in characterizing the dependence structure for $(X, Y) \in \mathrm{IC}_0$.

We then consider an extension of the positive Fréchet copula so that the marginals are not standard uniform and the dependence structure is not exchangeable. Note, however, that the marginal distributions are assumed to be identical in the following concept.

Definition 10.3. Let (X, Y) be a random vector with identical marginal F and joint distribution H. We say that (X, Y) is *quasi-r-Fréchet* for some $r \in [-1, 1]$ if

$$\frac{H(x, y) + H(y, x)}{2} = r \min(F(x), F(y)) + (1 - r)F(x)F(y), \quad \text{for all } x, y \in \mathbb{R}. \tag{10.9}$$

The need to introduce quasi-independence and quasi-Fréchet model stems from the permutation invariance of the correlation coefficient $\mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$, that is, correlation coefficient does not capture directional (non-exchangeable) dependence. Note that we require identical marginals of (X, Y) in the quasi-Fréchet model, but no such requirement is imposed on quasi-independence. When (X, Y) has an identical marginal, the quasi-Fréchet model with $r = 0$ reduces to quasi-independence.

Denote by μ, μ_+ and $\mu_\perp$ the probability measures induced by the cumulative distribution functions H, H_+ and $H_\perp$, respectively, where $H_+(x, y) = \min(F(x), F(y))$ and $H_\perp(x, y) = F(x)F(y)$, $(x, y) \in \mathbb{R}$. By the standard measure-theoretic argument (see Lemma 10.2 in Section 10.8.1), we have that the condition (10.9) is equivalent to

$$\frac{\mu(A \times B) + \mu(B \times A)}{2} = r\mu_+(A \times B) + (1 - r)\mu_\perp(A \times B), \quad \text{for all } A, B \in \mathcal{B}(\mathbb{R}). \tag{10.10}$$

The term "quasi-Fréchet" reflects its connection to the Fréchet copula in (10.6), although here r may be negative. When (X, Y) is exchangeable, the quasi-r-Fréchet model reduces to (10.2), which again yields the positive Fréchet copula C_r^{F} when the identical marginal is the standard uniform.

10.3 Bivariate invariant correlation

In this section we focus on the bivariate case $d = 2$. We will characterize dependence structures of bivariate random vectors (X, Y) with invariant correlation. The main results of this section are summarized in Table 10.1. For a random variable X, the support of the distribution of X is given by $\mathrm{Supp}(X) = \{x \in \mathbb{R} : \mathbb{P}(x - \epsilon < X \leqslant x + \epsilon) > 0 \text{ for all } \epsilon > 0\}$.

10.3.1 Zero invariant correlation

We first characterize the joint distribution in the case (X, Y) has an invariant correlation 0. As we will see later, this case is fundamentally different from the case $r \neq 0$.

In Example 10.1, we have seen that $(X, Y) \in \mathrm{IC}_0$ if (X, Y) is independent. A natural first question is whether independence is the only possibility for IC_0. In the next result, we show that zero invariant correlation is equivalent to quasi-independence.

Theorem 10.1. *For $X, Y \in \mathcal{L}^2$, $(X, Y) \in \mathrm{IC}_0$ if and only if (X, Y) is quasi-independent.*

For a random vector (X_1, X_2), we call (X_{π_1}, X_{π_2}) the *random rearrangement* of (X_1, X_2), where (π_1, π_2) is a random vector uniformly distributed on $\{(1, 2), (2, 1)\}$ independent of (X_1, X_2). Although quasi-independence does not imply independence, the following proposition shows that quasi-independence of (X_1, X_2) implies independence of its random rearrangement if X_1 and X_2 have the same distribution.

Proposition 10.2. *Suppose that $X_1, X_2 \in \mathcal{L}^2$ have the same distribution. Then $(X_1, X_2) \in \mathrm{IC}_0$ if and only if its random rearrangement is independent.*

In particular, if (X_1, X_2) is exchangeable, that is, $(X_2, X_1) \overset{\mathrm{d}}{=} (X_1, X_2)$, then zero invariant correlation is equivalent to independence.

Let $|A|$ be the cardinality of a set $A \subseteq \mathbb{R}$; $|A| = \infty$ if A is infinite. A random variable X is *n-atomic* if $|\mathrm{Supp}(X)| = n$; that is, its distribution is supported on n distinct points. We use the term "bi-atomic" in case $n = 2$ and "tri-atomic" in case $n = 3$. These terms are applied to both random variables and their distributions. Theorem 10.1 identifies the joint distribution of $(X, Y) \in \mathrm{IC}_0$ when X and Y are atomic random variables.

Proposition 10.3. *Suppose X is m-atomic and Y is n-atomic with $2 \leqslant m \leqslant n$. Let $P = (p_{ij})_{n\times n}$ with $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$ for $i \in [m]$, $j \in [n]$ and $p_{ij} = 0$ for $m < i \leqslant n$, $j \in [n]$, $\mathbf{p} = P\mathbf{1}_n$ and $\mathbf{q} = P^\top \mathbf{1}_n$. Then the followings are equivalent.*

(i) $(X, Y) \in \mathrm{IC}_0$.

(ii) $P + P^\top = \mathbf{p}\mathbf{q}^\top + \mathbf{q}\mathbf{p}^\top$.

(iii) $P = \mathbf{p}\mathbf{q}^\top + S$, *where* $S = (s_{ij})_{n\times n}$ *satisfies: (a)* $p_i q_j + s_{ij} \geqslant 0$ *for each* (i, j)*; (b) each of its row sums and column sums is* 0*; and (c)* $s_{ij} = -s_{ji}$ *for each* (i, j)*, i.e., S is anti-symmetric.*

For the special case $m = n = 3$ and $x_i = y_i$ for $i \in [3]$, we have an explicit representation for the probability matrix P in Example 10.3.

Example 10.3. Assume that X and Y are tri-atomic random variables with identical supports. Then, $(X, Y) \in \mathrm{IC}_0$ if and only if the probability matrix is given by $P = \mathbf{p}\mathbf{q}^\top + S$, where $S = (s_{ij})_{3\times 3}$ is an anti-symmetric matrix such that $s_{ii} = 0$ for $i \in [3]$ and $s_{12} = s_{23} = s_{31} = \epsilon \in [-\min(p_1q_2, p_2q_3, p_3q_1), \min(p_1q_3, p_2q_1, p_3q_2)]$.

By taking $p_1 \downarrow 0$, we have $\epsilon \to 0$ in Example 10.3. Hence, it is expected that zero invariant correlation is characterized by independence if one of the marginals is bi-atomic. This is formally stated in the next proposition.

Proposition 10.4. *Let $X, Y \in \mathcal{L}^2$ where X is bi-atomic. If (X, Y) is quasi-independent, then (X, Y) is independent. In particular, $(X, Y) \in \mathrm{IC}_0$ if and only if (X, Y) is independent.*

At this point, we have fully characterized zero invariant correlation and its relation to independence. Next, we consider general invariant correlation r for $r \in [-1, 1]$.

10.3.2 Invariant correlation with non-identical marginals

The main message of this section is that, for X and Y with different marginals, having a non-zero invariant correlation is impossible, except for the very special case of bi-atomic distributions.

First, suppose that X and Y are bi-atomic random variables. The following result shows that $(X, Y) \in \mathrm{IC}$ if X and Y have the same support. In this case, the invariant correlation can be any number in $[-1, 1]$. In particular, we have that $\mathrm{Corr}(X, Y) = 1$ implies comonotonicity, $\mathrm{Corr}(X, Y) = 0$ implies independence, and $\mathrm{Corr}(X, Y) = -1$ implies counter-monotonicity (i.e., X and $-Y$ are comonotonic). However, if X and Y have different supports, then independence is the only possible case of invariant correlation.

Proposition 10.5. *Let X and Y be bi-atomic random variables. Then $(X, Y) \in \mathrm{IC}$ if and only if either (i) X and Y have the same support, or (ii) (X, Y) is independent.*

Note that case (ii) in Proposition 10.5 is equivalent to zero invariant correlation. Therefore, for a pair of bi-atomic random variables, IC_r for $r \neq 0$ is characterized as the set of all models with identical marginal supports.

Next, we obtain the following observations by carefully investigating tri-atomic models; see Lemma 10.4 in Section 10.8.3 for details.

(i) For bi-atomic X and tri-atomic Y such that $\mathrm{Supp}(X) \subsetneq \mathrm{Supp}(Y)$, we have $(X, Y) \in \mathrm{IC}$ if and only if $(X, Y) \in \mathrm{IC}_0$.

(ii) Let X and Y be tri-atomic random variables with the same support. If $(X, Y) \in \mathrm{IC}$, then $F_X = F_Y$ or $\mathrm{Corr}(X, Y) = 0$.

If X and Y are general random variables with different distributions, then we can take h in Proposition 10.1 to transform X and Y into bi-atomic or tri-atomic random variables, so that we can apply Proposition 10.5 and the above observations (i) and (ii). This argument leads to the main result of this section.

Theorem 10.2. *Suppose that $X, Y \in \mathcal{L}^2$ have different distributions and $|\mathrm{Supp}(Y)| > 2$. Then $(X, Y) \in \mathrm{IC}$ if and only $(X, Y) \in \mathrm{IC}_0$ (which is characterized in Theorem 10.1).*

Together with Proposition 10.5, Theorem 10.2 characterizes all $(X, Y) \in \mathrm{IC}$ with different marginals. In particular, the only possibility for having different marginals in $(X, Y) \in \mathrm{IC}_r$, $r \neq 0$, is that X and Y are both bi-atomic.

10.3.3 Invariant correlation with identical marginals

In this section we explore the remaining case of $(X,Y) \in \mathrm{IC}$ where $F_X = F_Y$. We start with n-atomic random variables X and Y. Assume that X and Y take values in $\mathcal{X} = \{x_1, \dots, x_n\}$, $n \in \mathbb{N}$, with $x_1 < \cdots < x_n$. Define the $n \times n$ probability matrix $P = (p_{ij})_{n\times n}$ of (X,Y) as $p_{ij} = \mathbb{P}(X = x_i, Y = x_j)$. Let $\mathbf{p} = P\mathbf{1}_n$ be the marginal probability of X and Y. Let $D = \mathrm{diag}(\mathbf{p})$. For $\mathbf{x} = (x_1, \dots, x_n)^\top$, we have $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbf{x}^\top \mathbf{p}$, $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = \mathbf{x}^\top D\mathbf{x}$, $\mathbb{E}[XY] = \mathbf{x}^\top P\mathbf{x}$, and thus the Pearson correlation coefficient of (X,Y) is given by

$$\mathrm{Corr}(X,Y) = \frac{\mathbf{x}^\top P\mathbf{x} - \mathbf{x}^\top \mathbf{p}\mathbf{p}^\top \mathbf{x}}{\sqrt{\mathbf{x}^\top D\mathbf{x} - \mathbf{x}^\top \mathbf{p}\mathbf{p}^\top \mathbf{x}}\sqrt{\mathbf{x}^\top D\mathbf{y} - \mathbf{x}^\top \mathbf{p}\mathbf{p}^\top \mathbf{x}}} = \frac{\mathbf{x}^\top (P - \mathbf{p}\mathbf{p}^\top)\mathbf{x}}{\mathbf{x}^\top (D - \mathbf{p}\mathbf{p}^\top)\mathbf{x}}.$$

Proposition 10.6. *Let X and Y be identically distributed n-atomic random variables. Then $(X,Y) \in \mathrm{IC}_r$ for some $r \in [-1,1]$ if and only if (X,Y) is quasi-r-Fréchet, that is, the probability matrix P satisfies*

$$\frac{P + P^\top}{2} = rD + (1-r)\mathbf{p}\mathbf{p}^\top. \tag{10.11}$$

Furthermore, the invariant correlation r satisfies

$$\underline{r}_{\mathbf{p}} \leqslant r \leqslant 1, \quad \textit{where} \quad \underline{r}_{\mathbf{p}} = \max\left(\max_{j\in[n]} \left(-\frac{p_j}{1-p_j} \right), \max_{i,j\in[n], i\neq j} \left(1 - \frac{1}{p_i p_j} \right) \right). \tag{10.12}$$

If (X,Y) in Proposition 10.6 is exchangeable, then $P^\top = P$ in (10.11), which yields $P = rD + (1-r)\mathbf{p}\mathbf{p}^\top$. That is, P is r-Fréchet as defined in (10.2).

Remark 10.1. We show that $-1/(n-1) \leqslant \underline{r}_{\mathbf{p}} < 0$ for any $\mathbf{p}$. The upper bound for $\underline{r}_{\mathbf{p}}$ is trivial since all terms in $\underline{r}_{\mathbf{p}}$ are negative. Since

$$\max_{j\in[n]} \left(-\frac{p_j}{1-p_j} \right) = -\left(\frac{1}{\min(p_1, \dots, p_n)} - 1 \right)^{-1}, \tag{10.13}$$

we have $\underline{r}_{\mathbf{p}} \uparrow 0$ if $\min(p_1, \dots, p_n) \downarrow 0$. To show the lower bound for $\underline{r}_{\mathbf{p}}$, notice that the minimum of (10.13) is $-1/(n-1)$ when $\min(p_1, \dots, p_n) = 1/n$, and this is attained if and only if $p_i = 1/n$, $i \in [n]$. Consequently, we have the lower bound $-1/(n-1)$. Note that, for $\mathbf{p} \neq (1/n)\mathbf{1}_n$, (10.13) is strictly greater than $-1/(n-1)$ and thus $\underline{r}_{\mathbf{p}} > -1/(n-1)$.

We next present the main result of this section, which covers the case when X and Y have identical marginals.

Theorem 10.3. *Let $X, Y \in \mathcal{L}^2$ be identically distributed. Then $(X,Y) \in \mathrm{IC}_r$ if and only if (X,Y) is quasi-r-Fréchet.*

As seen in Proposition 10.6 and Remark 10.1, the invariant correlation r can be negative when the support of F is finite. On the other hand, we will see that $r \geqslant 0$ if $X \sim F$ is continuous or the support of X is countable but infinite.

Next, we consider the case when the support of X and Y is discrete, that is, $\mathcal{X} = \{x_i : i \in \mathcal{I}\}$, $\mathcal{I} \subseteq \mathbb{N}$. Write $\mathbb{P}(X = x_i, Y = x_j) = p_{ij} > 0$, $i, j \in \mathcal{I}$, and $\mathbb{P}(X = x_i) = p_i > 0$, $i \in \mathcal{I}$. In this special case, Theorem 10.3 can be stated in the following form, which is more general than Proposition 10.6 since $|\mathcal{X}| = \infty$ is allowed. For identically distributed discrete random variables X and Y, we have $(X,Y) \in \mathrm{IC}_r$ for some $r \in [-1,1]$ if and only if

$$\frac{p_{ij} + p_{ji}}{2} = r p_i \mathbb{1}_{\{i=j\}} + (1-r) p_i \, p_j \quad \text{for all } i, j \in \mathcal{I}, \tag{10.14}$$

with $\underline{r}_{\mathbf{p}} \leqslant r \leqslant 1$, where

$$\underline{r}_{\mathbf{p}} = \max\left\{ \sup_{j \in \mathcal{I}} \left(-\frac{p_j}{1-p_j} \right), \sup_{i,j \in \mathcal{I}, i \neq j} \left(1 - \frac{1}{p_i p_j} \right) \right\}.$$

In particular, if $|\mathcal{X}| = \infty$, we have $\inf_{i \in \mathcal{I}} p_i = 0$, and hence $\underline{r}_{\mathbf{p}} = 0$ and $0 \leqslant r \leqslant 1$.

Next, we consider the case when $F_X = F_Y = F$ for a continuous and strictly increasing distribution F. Let $(U,V) = (F(X), F(Y)) \sim C$ be the unique copula of (X,Y). By using Corollary 10.1, Theorem 10.3 can be stated as follows.

Corollary 10.2. *Suppose that X and Y have identical continuous and strictly increasing marginals. Then the following are equivalent:*

(i) $(X,Y) \in \mathrm{IC}_r$;

(ii) *the copula C of (X,Y) satisfies*

$$\frac{C(u,v) + C(v,u)}{2} = r M(u,v) + (1-r)\Pi(u,v) \quad \textit{for all } (u,v) \in \mathcal{I}; \tag{10.15}$$

(iii) $r \in [0,1]$ *and the positive Fréchet copula C_r^{F} is the copula of the random rearrangement of (X,Y).*

Unlike the case of atomic random variables, negative invariant correlation is not possible when the support of X and Y is uncountable. By Corollary 10.2, the positive Fréchet copula C_r^{F} is characterized as an exchangeable copula with invariant correlation r. Non-exchangeable copulas satisfying (10.15) are constructed in the next example.

Example 10.4. For $r \in \mathcal{I}$, let P be a 3×3 stochastic matrix such that $(P + P^\top)/2 = (1/9)\mathbf{1}_3\mathbf{1}_3^\top$. Let C_P be the checkerboard copula (see Section 4.1.1 of Durante and Sempi (2016)) associated with P with the interior of each grid filled by the independence copula. Let $C = rM + (1-r)C_P$. Then it is straightforward to check that the copula C satisfies (10.15). On the other hand, C may not be a Fréchet copula since C_P is not exchangeable when P is not symmetric.

10.4 Invariant correlation matrices

In this section, we study invariant correlation matrices as the multivariate extension of invariant correlation in dimension 2.

10.4.1 Characterization of invariant correlation matrices

Recall that a d-dimensional random vector $\mathbf{X}$ has an *invariant correlation matrix* $R = (r_{ij})_{d\times d}$ if every pair (i, j) of its components has an invariant correlation r_{ij} for $i, j \in [d]$. We confine ourselves to the case when $\mathbf{X}$ is continuous and its marginals are identical. Denote by Θ_d the set of all possible $d \times d$ invariant correlation matrices of continuous random vectors with identical marginals. Our goal in this section is to understand the set Θ_d and its corresponding models.

For continuous marginals, Corollary 10.2 immediately leads to the following result. For a d-copula C and $i, j \in [d]$, $i \neq j$, its (i, j)th marginal is denoted by C_{ij}.

Proposition 10.7. *For a continuous d-dimensional random vector $\mathbf{X}$ with identical marginals, $\mathbf{X} \in \mathrm{IC}^d$ holds if and only if for every $i, j \in [d]$ with $i \neq j$, there exists $r_{ij} \in [0, 1]$ such that*

the copula C_{ij} of (X_i, X_j) satisfies

$$\frac{C_{ij} + C_{ji}}{2} = r_{ij} M + (1 - r_{ij})\Pi.$$

Let $k \in \mathbb{N}$. For $i \in [d]$, let $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ik})^\top$ be a $\{0,1\}^k$-valued Bernoulli random vector with $\mathbf{Z}_i^\top \mathbf{Z}_i = 1$. In other words, $\mathbf{Z}_i$ has a categorical distribution with $\sum_{j=1}^k Z_{ij} = 1$. Write the matrix $\Gamma = (Z_{ij})_{d\times k}$, and call it a $d \times k$ *categorical random matrix*. Let $\mathbf{U}$ be an independent uniform random vector on $[0,1]^k$ independent of Γ. Consider the following model

$$\mathbf{X} = \Gamma \mathbf{U} = (\mathbf{Z}_1^\top \mathbf{U}, \dots, \mathbf{Z}_d^\top \mathbf{U}). \tag{10.16}$$

We first show that $\mathbf{X}$ has an invariant correlation matrix.

Proposition 10.8. *For $\mathbf{X}$ in (10.16), the following statements hold:*

(i) $\mathbf{X}$ has standard uniform marginals;

(ii) $\mathbf{X}$ has an invariant correlation matrix;

(iii) The correlation matrix of $\mathbf{X}$ (equal to its tail-dependence matrix) is given by $\mathbb{E}[\Gamma\Gamma^\top]$.

Example 10.5. Take $k = d+1$ and let $Z_{ij} = 0$ for all $j \in [d] \setminus \{i\}$. This implies $Z_{ik} = 1 - Z_{ii}$ for $i \in [d]$. For this model, we have $X_i = Z_{ii}U_i + Z_{ik}U_{d+1} = (1 - Z_{ik})U_i + Z_{ik}U_{d+1}$, $i \in [d]$, and the correlation coefficient between X_i and X_j is given by $\mathbb{E}[Z_{ik}Z_{jk}]$.

Remark 10.2. Let $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$ be iid observations from a continuous and strictly increasing distribution F, and let Γ be as in (10.16) independent of $\mathbf{Y}$. As an immediate consequence from Proposition 10.8, the model

$$\Gamma \mathbf{Y} = (\mathbf{Z}_1^\top \mathbf{Y}, \dots, \mathbf{Z}_d^\top \mathbf{Y}) \tag{10.17}$$

has an identical marginal distribution F and an invariant correlation matrix $\mathbb{E}[\Gamma\Gamma^\top]$.

Denote by $\mathcal{Z}_{d,k}$ the set of all matrices of the form $\mathbb{E}[\Gamma\Gamma^\top]$ for some $d \times k$ categorical random matrix Γ. Proposition 10.8 indicates that $\mathcal{Z}_{d,k}$ is a subset of Θ_d. Lemma 10.6 in Section 10.8.5 shows that $\mathcal{Z}_{d,k}$ is the convex hull of the collection of all clique partition

points (Grötschel and Wakabayashi, 1990; Fiebig et al., 2017), which we will explain below. Let $\mathbf{A} = (A_1, \ldots, A_k)$ be a partition of $[d]$, where some of A_s, $s \in [k]$, may be empty. The *clique partition point* $\Sigma^{\mathbf{A}} = (\Sigma^{\mathbf{A}}_{ij})_{d\times d}$ for the partition $\mathbf{A} = (A_1, \ldots, A_k)$ of $[d]$ is defined as

$$\Sigma^{\mathbf{A}}_{ij} = \sum_{s=1}^{k} \mathbb{1}_{\{i,j\in A_s\}}, \quad i, j \in [d]. \tag{10.18}$$

In other words, $\Sigma^{\mathbf{A}}$ can be decomposed into disjoint submatrices of all entries 1 and the other entries are 0. We can verify that $\Sigma^{\mathbf{A}} \in \mathcal{Z}_{d,k}$, as it is the correlation matrix of the following model of the form (10.16):

$$X^{\mathbf{A}}_i = \sum_{s=1}^{k} \mathbb{1}_{\{i\in A_s\}} U_s, \quad i \in [d]. \tag{10.19}$$

Denote by $\mathcal{S}_{d,k}$ the collection of all possible clique partition points (10.18) for all possible k partitions of $[d]$. In what follows we characterize Θ_d in terms of $\mathcal{Z}_{d,k}$ and $\mathcal{S}_{d,k}$.

Theorem 10.4. *For $d \geqslant 2$, $\Theta_d = \mathcal{Z}_{d,d} = \bigcup_{k\in\mathbb{N}} \mathcal{Z}_{d,k} = \mathrm{Conv}(\mathcal{S}_{d,d})$.*

The set $\mathrm{Conv}(\mathcal{S}_{d,d})$ is called the *clique partition polytope* in dimension d, which is known to be relevant to matrix compatibility of other dependence measures; see Section 10.7.1 for details. Theorem 10.4 implies that all invariant correlation matrices are realized by the model (10.16). Nevertheless, not all exchangeable random vectors with an invariant correlation matrix can be represented by (10.16). For instance, any d-dimensional random vector with pairwise independent components has an invariant correlation matrix I_d but is not necessarily modelled by (10.16). Generally, it is not possible to completely identify a multivariate model from bivariate properties. A class of copulas whose bivariate marginals are Fréchet copulas is studied by Yang et al. (2009).

We next consider the *membership testing problem* of Θ_d, which aims at determining whether a given correlation matrix $R = (r_{ij})_{d\times d}$ belongs to Θ_d; that is, R is realized as an invariant correlation matrix of some d-dimensional random vector. Note that the number of vertices of the clique partition polytope equals the number of different partitions of $[d]$, and this number is known as the *Bell number* N_d, which can be computed explicitly (Grötschel and Wakabayashi, 1990). The number N_d grows rapidly in d. For example, $N_3 = 5$, $N_4 = 15$,

$N_5 = 52$ and $N_{10} > 10^5$. By Theorem 10.4, we have $R = (r_{ij})_{d\times d} \in \Theta_d$ if and only if there exists $\alpha_1, \ldots, \alpha_{N_d} \geqslant 0$ such that $\sum_{k=1}^{N_d} \alpha_k = 1$ and $\sum_{k=1}^{N_d} \alpha_k \Sigma^{(k)} = R$, where $\mathcal{S}_{d,d} = \left\{\Sigma^{(k)} : k \in [N_d]\right\}$, $\Sigma^{(k)} = (\sigma_{ij}^{(k)})_{d\times d}$, is a collection of all vertices of the clique correlation polytope. These linear constraints on $\boldsymbol{\alpha}$ can be summarized into $V_d\boldsymbol{\alpha} = \tilde{\mathbf{r}}$, where $\tilde{\mathbf{r}} \in [0,1]^{\tilde{d}}$ and $V_d \in \{0,1\}^{\tilde{d}\times N_d}$ are defined in (10.21) and (10.22) below with $\tilde{d} = 1 + d(d-1)/2$. By introducing an auxiliary variable $\mathbf{z} \in \mathbb{R}^{\tilde{d}}$, a given correlation matrix $R = (r_{ij})_{d\times d} \in [0,1]^{d\times d}$ is in Θ_d if and only if the following linear program attains zero:

$$\min_{\mathbf{z}\in\mathbb{R}^{\tilde{d}}, \boldsymbol{\alpha}\in\mathbb{R}^{N_d}} \mathbf{1}_{\tilde{d}}^\top \mathbf{z} \quad \text{subject to} \begin{cases} V_d\boldsymbol{\alpha} + \mathbf{z} = \tilde{\mathbf{r}}, \\ \boldsymbol{\alpha} \geqslant \mathbf{0}_{N_d} \text{ and } \mathbf{z} \geqslant \mathbf{0}_{\tilde{d}}, \end{cases} \tag{10.20}$$

where

$$\tilde{\mathbf{r}} = (r_{12}, r_{13}, r_{23}, \ldots, r_{d-1\,d}, 1) \tag{10.21}$$

and

$$V_d = \begin{pmatrix} \sigma_{12}^{(1)} & \sigma_{12}^{(2)} & \cdots & \sigma_{12}^{(N_d)} \\ \sigma_{13}^{(1)} & \sigma_{13}^{(2)} & \cdots & \sigma_{13}^{(N_d)} \\ \sigma_{23}^{(1)} & \sigma_{23}^{(2)} & \cdots & \sigma_{23}^{(N_d)} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{(d-1)\,d}^{(1)} & \sigma_{(d-1)\,d}^{(2)} & \cdots & \sigma_{(d-1)\,d}^{(N_d)} \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{10.22}$$

Note that any correlation matrix with at least one negative entry is immediately excluded from Θ_d. For $R = (r_{ij})_{d\times d} \in [0,1]^{d\times d}$, the set of constraints in (10.20) is always nonempty since it contains the pair $(\boldsymbol{\alpha}, \mathbf{z}) = (\mathbf{0}_{N_d}, \tilde{\mathbf{r}})$. The above linear program is solved, for example, with the R package lpSolve although it can be computationally demanding for large d.

Remark 10.3. Suppose that the linear program (10.20) attains 0 at $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ for a given correlation matrix $R = (r_{ij})_{d\times d}$. Then one can simulate a d-dimensional random vector with invariant correlation R as a mixture of the models of the form (10.19) with $\boldsymbol{\alpha}^*$ being the vector of mixture weights to $\{\Sigma^{(k)} : k \in [N_d]\}$.

10.4.2 Relationship to positive regression dependence

Positive regression dependence (Lehmann, 1966) is a concept of dependence known to

be useful to control the false discovery rate (FDR) through the procedure of Benjamini and Hochberg (1995) in the context of multiple testing (Benjamini and Yekutieli, 2001). In this section, we explore the relationship between this dependence property and invariant correlation.

We start with the definition. A set $A \subseteq \mathbb{R}^d$ is said to be *increasing* if $\mathbf{x} \in A$ implies $\mathbf{y} \in A$ for all $\mathbf{y} \geqslant \mathbf{x}$.

Definition 10.4. A d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is said to have *positive regression dependence on the subset* $\mathcal{N} \subseteq [d]$ *(PRDS)* if for any index $i \in \mathcal{N}$ and increasing set $A \subseteq \mathbb{R}^d$, the function $x \mapsto \mathbb{P}(\mathbf{X} \in A \mid X_i = x)$ is increasing. The case of $\mathcal{N} = [d]$ is simply called *positive regression dependence (PRD).*

Let us first consider $d = 2$. A direct consequence from Lemma 10.2 in Section 10.8.1 is that every exchangeable random vector (X_1, X_2) with invariant correlation $r \geqslant 0$ is PRD since (10.10) implies that, for every increasing set A and $x \in \mathbb{R}$, $\mathbb{P}\left((X_1, X_2) \in A \mid X_1 = x\right) = r\mathbb{1}_{\{(x,x)\in A\}} + (1-r)\mathbb{P}\left((x, X_2) \in A\right)$, which is increasing in x. Note, however, that without exchangeability PRD is not implied by invariant correlation; see Example 10.8 in Section 10.7.3.

For $d \geqslant 3$, there also exists an exchangeable model which has an invariant correlation matrix with non-negative entries but does not have PRD; this is reported in Example 10.9 in Section 10.7.3. Therefore, a dependence structure admitting an invariant correlation matrix does not imply PRD even if exchangeability is additionally assumed. This is not surprising as invariant correlation is essentially a pairwise property, and pairwise dependence does not determine overall dependence. Nevertheless, the next proposition shows that the model (10.16) has PRD, and hence also PRDS for any subset $\mathcal{N} \subseteq [d]$. Together with Theorem 10.4, this result indicates that every invariant correlation matrix admits a model with PRD.

Proposition 10.9. *The random vector* $\mathbf{X} = \Gamma\mathbf{U}$ *in* (10.16) *has PRD.*

Since both the properties of PRD and invariance correlation are preserved under increasing transforms, we immediately obtain that the model

$$\mathbf{X} = \Gamma\mathbf{V}, \quad \text{where } \mathbf{V} = (V_1, \ldots, V_d)^\top \text{ and } V_i = g(U_i) \text{ for } i \in [d], \tag{10.23}$$

has PRD and an invariant correlation matrix for any increasing and admissible g with $g(0) = 0$. For example, by choosing $g(x) = \lceil nx \rceil / n$, we obtain the discrete version of (10.16) where $V_1, \ldots, V_d$ are independent and uniformly distributed on $[n]/n$.

10.4.3 Application to the problem of sample duplication

Sample duplication is the phenomenon that some sample is repeated several times in the dataset. It occurs typically when the data is collected from multiple sources and curated by multiple groups. The presence of duplicate records has adverse effects when, for example, training or testing machine learning algorithms (Allamanis, 2019; Zhao et al., 2021). A simple solution to this problem is to de-duplicate the data by deleting such multiple records, which is also referred to as *record linkage*, *reference matching* and *copy detection*; see Heidari et al. (2020) and the references therein. However, de-duplication is not always feasible due to the time complexity where it commonly requires to measure similarity of samples for each pair of records. To solve this issue, various alternative approaches have been considered by leveraging some structures and specific assumptions on the frequency of sample duplication. Yet, it is a non-trivial task to verify these specific structures and assumptions Xie et al. (2013); Heidari et al. (2020). In this section, we demonstrate the benefit of modeling the correlation matrix of the duplication structure for stable statistical inference in the presence of sample duplication.

We focus on the approach of using the original dataset with duplicate records, and propose a versatile method to reduce adverse effects of sample duplication in the general context of optimization problem. We associate the dataset $\mathbf{X} = (X_1, \ldots, X_d)$, possibly containing duplicate records, with the form (10.17). For a continuous and strictly increasing function F, suppose that we are interested in estimating $\mathbb{E}[\mathrm{S}(Y; \boldsymbol{\theta})] = \int \mathrm{S}(y; \boldsymbol{\theta}) F(\mathrm{d}y)$ based on the observations from F, where $\boldsymbol{\theta} \in \mathbb{R}^q$ is a vector of parameters and $\mathrm{S} : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ is a loss function.

We define several random elements. For the random set of indices $I = \{j \in [k] : \text{ there exists } i \in [d] \text{ such that } Z_{ij} = 1\}$, the size of iid samples with no duplicate record is given by $N = |I|$. In addition, let $W_j = \sum_{i=1}^{d} Z_{ij}$ be the number of times the jth sample in

$\mathbf{Y}$ appears in the dataset $\mathbf{X}$. By its definition, it holds that $\sum_{j\in[k]} W_j = d$.

The empirical average score based on the dataset $\mathbf{X}$ can be decomposed as follows (Allamanis, 2019; Zhao et al., 2021):

$$\frac{1}{d}\sum_{i\in[d]} \mathrm{S}(X_i;\boldsymbol{\theta}) = \frac{1}{d}\sum_{j\in[k]} W_j\,\mathrm{S}(Y_j;\boldsymbol{\theta}) = \frac{N}{d}\frac{1}{N}\sum_{j\in I}\mathrm{S}(Y_j;\boldsymbol{\theta}) + \frac{d-N}{d}\frac{1}{d-N}\sum_{j\in I}(W_j-1)\,\mathrm{S}(Y_j;\boldsymbol{\theta}), \tag{10.24}$$

where the terms $(d-N)/d$ and $(1/(d-N))\sum_{j\in I}(W_j-1)\,\mathrm{S}(Y_j;\boldsymbol{\theta})$ are called *duplication factor* and *duplication bias*, respectively. Moreover, $(1/N)\sum_{j\in I}\mathrm{S}(Y_j;\boldsymbol{\theta})$ is the term estimating the score with iid observations.

Despite these terms used in the literature, any convex combination of the form:

$$\sum_{i\in[d]} \lambda_i\,\mathrm{S}(X_i;\boldsymbol{\theta}) \quad \text{where } \boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_d) \in \mathbb{R}^d \text{ are such that } \sum_{i\in[d]}\lambda_i = 1, \tag{10.25}$$

which includes (10.24) with $\boldsymbol{\lambda} = \mathbf{1}_d/d$, is an unbiased estimator of $\mathbb{E}[\mathrm{S}(Y;\boldsymbol{\theta})]$ by Remark 10.2. However, an adverse effect of sample duplication lies in the variance of the estimator since the variance of (10.24) is inflated as follows:

$$\mathrm{Var}\left(\frac{1}{d}\sum_{i\in[d]}\mathrm{S}(X_i;\boldsymbol{\theta})\right) = \left(1+\frac{2}{d}\sum_{i,j\in[d]\,i<j}\mathrm{Corr}(X_i,X_j)\right)\mathrm{Var}\left(\frac{1}{d}\sum_{i\in[d]}\mathrm{S}(Y_i;\boldsymbol{\theta})\right).$$

Let $\Sigma = \mathrm{Cov}(\mathbf{X})$ and $P = \mathrm{Corr}(\mathbf{X})$. It is well-known that the minimum variance unbiased estimator among those of the form (10.25) is uniquely attained when

$$\boldsymbol{\lambda}^* = \frac{\mathbf{1}_d^\top\Sigma^{-1}}{\mathbf{1}_d^\top\Sigma^{-1}\mathbf{1}_d} = \frac{\mathbf{1}_d^\top P^{-1}}{\mathbf{1}_d^\top P^{-1}\mathbf{1}_d},$$

with the minimum variance given by $(\mathbf{1}_d^\top\Sigma^{-1}\mathbf{1}_d)^{-1}$. This optimally weighted estimator is available when we specify the correlation structure of the dataset, for example, by assuming or inferring the underlying duplication structure. Since the correlation matrix of $\mathbf{X}$ is invariant under any transform as in Remark 10.2, the same optimal weight $\boldsymbol{\lambda}^*$ can be used for any scoring function S evaluated at any point $\boldsymbol{\theta}$.

Example 10.6 (A small simulation study)**.** Suppose that $\mathbf{X}$ follows the model (10.17), where $d = k = 1000$ and F is the standard normal distribution. For $k^* \in [k]$, let $\mathbf{Z}_1,\ldots,\mathbf{Z}_d$

independently follow Multi$(1, \mathbf{p}_i)$ with $\mathbf{p}_i = (\mathbf{1}_{k^*}, \mathbf{0}_{k-k^*})/k^*$ for $i \in [d/2]$ and $\mathbf{p}_i = \mathbf{e}_i$ for $i \in [d]\backslash[d/2]$. In this model, the first group $[d/2]$ contains duplicate records and the second group $[d]\backslash[d/2]$ consists of iid observations. With this duplication structure, $\mathrm{Corr}(X_i, X_j)$ is $1/k^*$ if $i, j \in [d/2]$ and is 0 otherwise. Let $\mathrm{S}(x;\theta) = (x-\theta)^2$, $x, \theta \in \mathbb{R}$, be the squared loss. We then estimate $\mathbb{E}[\mathrm{S}(X;\theta)]$, $X \sim F$, evaluated at $\theta = 0$ by (10.25) with $\boldsymbol{\lambda} = \mathbf{1}_d/d$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, respectively. To check the stability of these estimators, we generate $n = 1000$ independent replications of these estimates. For $k^* = 10, 50, 100$, we find that the sample variances of the estimates with the optimal weight $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ are reduced by $92\%, 71\%, 51\%$, respectively, compared to those with $\boldsymbol{\lambda} = \mathbf{1}_d/d$.

When the correlation matrix is estimated or given externally, it may be necessary to check its eligibility, that is, the existence of a model of duplication structure admitting the specified correlation matrix. Such a numerical procedure is provided in Section 10.4.1 although it turned out to be computationally demanding if no specific structure of correlation matrix is assumed.

10.5 Invariant correlation under increasing transforms

In many applications, only increasing transforms of the random variables are relevant, as non-increasing transforms do not preserve the copula among random variables. In view of this, we study a variant of invariant correlation where we require (10.4) to hold only for increasing transforms, instead of all transforms. By definition, this requirement is weaker than invariant correlation defined in Section 10.2. The interesting question is then whether this formulation allows for more models than those characterized in Section 10.3.

Definition 10.5. Let $r \in [-1, 1]$. A bivariate random vector (X, Y) is said to have an *invariant correlation r under increasing transforms* if (10.4) holds for all admissible increasing functions g. The set of all such random vectors is denoted by $\mathrm{IC}_r^{\uparrow}$.

Denote by $\mathrm{IC}^{\uparrow} = \bigcup_{r\in[-1,1]} \mathrm{IC}_r^{\uparrow}$. Note that $\mathrm{IC}_r \subseteq \mathrm{IC}_r^{\uparrow}$ for every $r \in [-1, 1]$, and $\mathrm{IC} \subseteq \mathrm{IC}^{\uparrow}$ by their definitions.

Table 10.2: Relationship between IC and $\mathrm{IC}^{\uparrow}$, where $\mathrm{IC} \subseteq \mathrm{IC}^{\uparrow}$ always holds. The abbreviations mean sets corresponding to those in Table 10.1. Note that QI is equivalent to IN if at least one of the marginals is bi-atomic (Proposition 10.4). We slightly abuse the notation so that equality is understood between corresponding subsets.

<table>
<tr><th colspan="2">Marginals F, G</th><th>$r = 0$</th><th>$r \neq 0$</th></tr>
<tr><td colspan="2">$F = G$</td><td rowspan="3">$\mathrm{IC}_0 = \mathrm{IC}_0^{\uparrow} = \mathrm{QI}$ (Theorem 10.5)</td><td>$\mathrm{IC}_r = \mathrm{IC}_r^{\uparrow} = \mathrm{QF}$ (Theorem 10.5)</td></tr>
<tr><td rowspan="2">$F \neq G$</td><td>both bi-atomic</td><td>$\mathrm{IC}_r \neq \mathrm{IC}_r^{\uparrow}$ (Example 10.7)</td></tr>
<tr><td>not both bi-atomic</td><td>$\mathrm{IC}_r = \mathrm{IC}_r^{\uparrow} = \varnothing$ (Theorem 10.5)</td></tr>
</table>

We summarize in Table 10.2 the results on the relationship between IC and $\mathrm{IC}^{\uparrow}$. Our main message is that invariant correlation confined to increasing transforms does not accommodate more models except for the case when both random variables are bi-atomic. Therefore, in most cases, we can safely treat the two formulations as equivalent.

The results of Proposition 10.1 and Corollary 10.1 can be easily extended to $\mathrm{IC}_r^{\uparrow}$. We collect these useful properties in the following corollary. The proofs are analogous to those of Proposition 10.1 and Corollary 10.1, and thus omitted.

Corollary 10.3. *The following statements hold.*

(i) Let $(X, Y) \in \mathrm{IC}_r^{\uparrow}$ for some $r \in [-1, 1]$, and let h be an admissible increasing function for (X, Y). Then, $(h(X), h(Y)) \in \mathrm{IC}_r^{\uparrow}$.

(ii) Suppose that $(X, Y) \sim H$ has identical continuous and strictly increasing marginal distributions, a correlation coefficient $r \in [-1, 1]$, and a copula C. Then $H \sim \mathrm{IC}_r^{\uparrow}$ if and only if $C \sim \mathrm{IC}_r^{\uparrow}$.

We first show in the following example that the property of invariant correlation under increasing transforms is not always equivalent to that under all transforms.

Example 10.7. Let X and Y be non-identically distributed bi-atomic random variables. In contrast to the result in Proposition 10.5, (X, Y) always admits an invariant correlation under increasing transforms. This is because the increasingness of g implies that $g(X)$ and $g(Y)$ are increasing linear functions of X and Y, respectively. Hence, $(X, Y) \in \mathrm{IC}^{\uparrow}$ regardless of the dependence structure of (X, Y). As a result, we have $\mathrm{IC} \subsetneq \mathrm{IC}^{\uparrow}$.

Except for the bi-atomic distributions, we show in the next theorem that IC_r and $\mathrm{IC}_r^{\uparrow}$ are equivalent.

Theorem 10.5. *Let $X, Y \in \mathcal{L}^2$ and $r = \mathrm{Corr}(X, Y)$. Assume that one of the following conditions holds: (i) $r = 0$; (ii) X and Y have identical distributions; (iii) X and Y have different distributions and $|\mathrm{Supp}(Y)| > 2$. Then $(X, Y) \in \mathrm{IC}_r$ if and only if $(X, Y) \in \mathrm{IC}_r^{\uparrow}$. In particular, if (iii) holds, then $r = 0$.*

Remark 10.4. By Item (ii) in Corollary 10.3, the result in Corollary 10.2 remains valid with IC_r in the statements replaced by $\mathrm{IC}_r^{\uparrow}$.

Under the condition (iii), Theorem 10.5 further implies that, for $r \neq 0$, there is no model (X, Y) in $\mathrm{IC}_r^{\uparrow}$ or IC_r.

10.6 Concluding remarks

Our main results on a full characterization of models with invariant correlation can be briefly summarized below. Except for the very special case of bi-atomic distributions, invariant correlation is characterized by quasi-independence (Theorem 10.1) and quasi-Fréchet models (Theorem 10.3), with non-identical marginal distributions excluded unless the correlation is zero (Theorem 10.2). The same holds true when transforms are confined to be increasing (Theorem 10.5). We also identify the set of all compatible invariant correlation matrices (Theorem 10.4).

Several aspects and generalizations of invariant correlation require further research. For instance, it would be interesting to understand whether a higher dimensional correlation measure, instead of the matrix of bivariate correlations, can be used to naturally formulate

an invariant correlation property, and whether it has interesting implications similar to our results in the bivariate case. Since PRD and PRDS are dependence concepts that are not determined by their bivariate marginals, a higher dimensional notion of invariant correlation may be more naturally connected to PRD and PRDS. Although a connection of invariant correlation to multiple testing and FDR control has been briefly discussed in Section 10.7.2, its applications and relevance for statistics are not yet clear. A further question concerns whether restricting the marginal transforms to a smaller but practically relevant class would characterize different and potentially useful models. These questions require further studies.

10.7 Auxiliary results

10.7.1 Invariant correlation matrices and other dependence matrices

In this section, we summarize the connection of Θ_d, the set of all possible $d \times d$ invariant correlation matrices of continuous random vectors with identical marginals, to the collection of tail-dependence matrices and other dependence matrices. Let (X_1, X_2) be a bivariate random vector with continuous marginal distributions F_1 and F_2. The *(lower) tail-dependence coefficient* of (X_1, X_2) is defined by $\lambda = \lim_{u \downarrow 0} \mathbb{P}\left(F_1(X_1) \leqslant u, F_2(X_2) \leqslant u\right)/u$ given that the limit exists. For a function $g : \mathbb{R} \to \mathbb{R}$, the *g-transformed rank correlation* of (X_1, X_2) is defined by $\kappa_g(X_1, X_2) = \mathrm{Corr}(g(F_1(X_1)), g(F_2(X_2)))$, provided that it is well defined. The *tail-dependence matrix* of a d-dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)$ with continuous marginal distributions has the tail-dependence coefficient of (X_i, X_j) in its (i, j)th entry for $i, j \in [d]$. Analogously, the κ_g-matrix of $\mathbf{X}$ is a $d \times d$ matrix whose (i, j)th entry is $\kappa_g(X_i, X_j)$ for $i, j \in [d]$. The set of all possible $d \times d$ tail-dependence matrices is called the *tail-dependence compatibility set* and is denoted by $\mathcal{T}_d$; see Embrechts et al. (2016); Fiebig et al. (2017); Krause et al. (2018); Shyamalkumar and Tao (2020) for recent studies. Similarly, the d-dimensional *compatibility set* of κ_g, denoted by $\mathcal{K}_d(g)$, is the collection of all possible κ_g-matrices in dimension d.

The next proposition shows that, for $\mathbf{X}$ with identical continuous marginals and an

invariant correlation matrix R, the model $\mathbf{X}$ has the same tail-dependence matrix and κ_g-matrix R.

Proposition 10.10. *Let* $\mathbf{X}$ *be a* d*-dimensional continuous random vector with identical marginals. If* $\mathbf{X} \in \mathrm{IC}^d$*, then its correlation matrix and its tail-dependence matrix coincide, as well as its* κ_g*-matrices for any* g *admissible for the standard uniform distribution.*

Proposition 10.10 implies that $\Theta_d \subset \mathcal{T}_d$ and that $\Theta_d \subset \mathcal{K}_d(g)$ for wide varieties of g. For instance, Θ_d is a subset of the compatibility sets of Blomqvist's beta and Spearman's rho, studied by Hofert and Koike (2019) and Wang et al. (2019), respectively. Since the compatibility set of Blomqvist's beta is equal to that of *Kendall's tau* (McNeil et al., 2022) and is smaller than that of *Gini's gamma* (Koike and Hofert, 2023), our result implies that Θ_d is contained in all of these compatibility sets. The connection of the clique partition polytope to $\mathcal{T}_d$ is given in Proposition 22 of Fiebig et al. (2017). This result, together with Theorem 10.4 and Proposition 10.10, leads to the following relationship between Θ_d and $\mathcal{T}_d$: $\Theta_d = \mathcal{T}_d$ for $d \leqslant 4$, and $\Theta_d \subsetneq \mathcal{T}_d$ for $d \geqslant 5$.

10.7.2 Conformal p-values

Conform p-value, studied by Bates et al. (2023) in the context of outlier detection, is an interesting example of invariant correlation. Let $S_1, \dots, S_n$ be scores of the null training sample (computed from some score function on the data), where n is a fixed positive integer. Scores of the test sample are denoted by S_{n+i}, $i \in [d]$, and the corresponding conformal p-value is given by

$$P_i = \frac{1}{n+1} + \frac{1}{n+1} \sum_{k=1}^{n} \mathbb{1}_{\{S_k \leqslant S_{n+i}\}}. \tag{10.26}$$

Assume that $\{S_i : i \in [n+d]\}$ is a sequence of independent random variables and $S_k \sim F_0$, $k \in [n]$. The above p-values are testing the sequence of null hypotheses $\mathrm{H}_{0,i} : S_{n+i} \sim F_0$, $i \in [d]$.

Proposition 10.11. *Let* $\mathcal{N} \subseteq [d]$ *be any set of null indices, that is,* $\mathrm{H}_{0,i}$ is true for $i \in \mathcal{N}$.

Then,

$$\mathbb{P}\left(P_i = \frac{j_i}{n+1},\ i \in \mathcal{N}\right) = \frac{N_1! \cdots N_{n+1}!}{(n+m)(n+m-1)\cdots(n+1)}, \quad j_i \in [n+1],$$

where $m = |\mathcal{N}|$ and $N_j = |\{i \in \mathcal{N} : j_i = j\}|$, $j \in [n+1]$. In particular, P_i is uniformly distributed on $[n+1]/(n+1)$ for every $i \in \mathcal{N}$. Moreover, $(P_i)_{i\in\mathcal{N}}$ has the invariant correlation matrix $(r_{ij})_{m\times m}$ such that $r_{ij} = 1/(n+2)$ for every $i, j \in [m]$, $i \neq j$.

We can take $\mathcal{N}$ to be the set of all null indices in Proposition 10.11. The fact that $(P_i)_{i\in\mathcal{N}}$ has an invariant correlation matrix has been shown in Lemma 2.1 of Bates et al. (2023), which also contains the joint distribution of (P_i, P_j) for $i, j \in \mathcal{N}$. Proposition 10.11 further gives the joint distribution of $(P_i)_{i\in\mathcal{N}}$. On the other hand, the full vector of p-values, $(P_i)_{i\in[d]}$ does not have an invariant correlation matrix in general. Note that the marginal distributions of $(P_i)_{i\in[d]}$ are generally different for null and non-null components, and there is some positive correlation between them by design in (10.26). Hence, our results in Section 10.3 explain that invariant correlation for $(P_i)_{i\in[d]}$ is not possible. Nevertheless, the vector of conformal p-values has PRDS on $\mathcal{N}$ as shown by Theorem 2.4 of Bates et al. (2023).

For $m = 2$, the vector of null conformal p-values follows the model (10.23) with $k = n+2$, $g(x) = \lfloor (n+1)x \rfloor/(n+1)$ and $\mathbf{Z}_1, \mathbf{Z}_2$ being iid multinomials with the number of trials 1 and uniform event probabilities on $[k]$. It is left open whether the vector of null conformal p-values can be written of the form (10.23) for $m \geqslant 3$.

10.7.3 Invariant correlaton and concepts of dependence

In Section 10.4.2, we have explored the relationship between invariant correlation and positive regression dependence (PRD). In this section, we show by examples that models with invariant correlation matrix do not always have PRD.

We start from the bivariate case $d = 2$. Following Lehmann (1966), a random vector (X, Y) is called *positive quadrant dependent (PQD)* (also called *positive orthant dependent*), if $\mathbb{P}(X \leqslant x, Y \leqslant y) \geqslant \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y)$ for every $(x, y) \in \mathbb{R}^2$. *Negative quadrant dependence (NQD)* is analogously defined by $\mathbb{P}(X \leqslant x, Y \leqslant y) \leqslant \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y)$ for every $(x, y) \in \mathbb{R}^2$. It is shown in Lemma 4 of Lehmann (1966) that PRD implies PQD. In the

next example, we construct a random vector that has zero invariant correlation but is neither PQD nor NQD. This example particularly indicates that non-negative invariant correlation does not lead to PRD.

Example 10.8. Let (X, Y) be the model in Example 10.3 with the identical marginal support $[3]$, $p_i = q_i = 1/3$ for $i \in [3]$ and $\epsilon = 1/9$. Then $\mathbb{P}(X \leqslant 2, Y \leqslant 1) = 1/9$, $\mathbb{P}(X \leqslant 1, Y \leqslant 2) = 1/3$ and $\mathbb{P}(X \leqslant 1)\mathbb{P}(Y \leqslant 2) = \mathbb{P}(X \leqslant 2)\mathbb{P}(Y \leqslant 1) = 2/9$. Hence (X, Y) is neither PQD nor NQD.

We have seen in Section 10.4.2 that, for the bivariate case, invariant correlation implies PRD under the additional assumption of exchangeability. The next example shows that this is not the case for $d = 3$, that is, there exists an exchangeable model which admits a non-negative invariant correlation matrix but does not have PRD.

Example 10.9. Let $d = 3$ and consider a member of the Farlie-Gumbel-Morgenstein family of copulas:

$$C_\theta(u_1, u_2, u_3) = u_1u_2u_3 + \theta u_1u_2u_3(1 - u_1)(1 - u_2)(1 - u_3), \quad \theta \in [-1, 1].$$

Note that C_θ is an exchangeable copula for any $\theta \in [-1, 1]$. Moreover, all bivariate marginals of C_θ are Π, and thus it has an invariant correlation matrix I_3.

We will show that the model $\mathbf{U} \sim C_\theta$ does not satisfy the following weaker version of PRDS than that in Definition 10.4:

$$s \mapsto \mathbb{P}(\mathbf{U} \in A \mid U_i \leqslant s) \text{ is increasing in } s \text{ for any index } i \in \mathcal{N} \text{ and increasing set } A \subseteq \mathbb{R}^d.$$

To this end, let $i = 1$, $s \in [0, 1]$, $\mathbf{t} = (t_1, t_2, t_3) \in [0, 1]^3$, $t_1 = 0$, and $A = (t_1, \infty) \times (t_2, \infty) \times (t_3, \infty)$. Then A is an increasing set and

$$\begin{aligned}\mathbb{P}(\mathbf{U} \in A \mid U_1 \leqslant s) &= \mathbb{P}(\mathbf{U} > \mathbf{t} \mid U_1 \leqslant s) \\ &= \frac{1}{s}\mathbb{P}(U_1 \leqslant s\,, t_2 < U_2,\, t_3 < U_3) = \frac{1}{s}\left\{(1 - t_2)(1 - t_3) - \overline{C}(s, t_2, t_3)\right\},\end{aligned}$$

where $\overline{C}(\mathbf{u}) = \mathbb{P}(\mathbf{U} > \mathbf{u})$ is the joint survival function of C. By calculation, we have that $\mathbb{P}(\mathbf{U} \in A \mid U_1 \leqslant s) = 1 - t_2 - t_3 + C(s, t_2, t_3)/s$. Therefore, $(\partial/\partial s)\mathbb{P}(\mathbf{U} \in A \mid U_1 \leqslant s) = \{s\,\partial_1 C(s, t_2, t_3) - C(s, t_2, t_3)\}/(s^2)$. When $t_2 = t_3 = 0.5$ and $\theta = 1$, this derivative is a constant $-1/16$ by a simple calculation. Therefore, $\mathbb{P}(\mathbf{U} \in A \mid U_1 \leqslant s)$ is a decreasing function in $s \in [0, 1]$.

10.8 Proofs of all results

10.8.1 Proofs in Section 10.2

Proof of Proposition 10.1. Let $(X, Y) \in \mathrm{IC}_r$ and g be an admissible function for the pair $(h(X), h(Y))$. The function $g \circ h$ is admissible for (X, Y). Since (X, Y) has an invariant correlation r, we have $\mathrm{Corr}(g \circ h(X), g \circ h(Y)) = \mathrm{Corr}(X, Y) = r$. Since an admissible function g is arbitrary, we have that $(h(X), h(Y)) \in \mathrm{IC}_r$. Applying this observation to all pairs of components of $(X_1, \ldots, X_d)$ establishes the desired result. □

Proof of Corollary 10.1. Suppose that $\mathbf{X} \in \mathrm{IC}^d$ with continuous and strictly increasing marginals F. By taking $h = F$ in Proposition 10.1, we have that $(F(X_1), \ldots, F(X_n)) \in \mathrm{IC}^d$; hence $C \sim \mathrm{IC}^d$. Suppose next that $C \sim \mathrm{IC}^d$. There exists a uniform random vector $\mathbf{U} = (U_1, \ldots, U_n) \in \mathrm{IC}^d$. Let $\mathbf{X} = (F^{-1}(U_1), \ldots, F^{-1}(U_n))$. By taking $h = F^{-1}$ in Proposition 10.1, we have that $\mathbf{X} \in \mathrm{IC}^d$. □

The next lemma justifies that the formulations of quasi-independence via (10.7) and (10.8) are equivalent.

Lemma 10.1. *For a random vector $(X, Y) \sim H$ with marginals F and G, (X, Y) is quasi-independent if and only if (10.8) holds for all $A, B \in \mathcal{B}(\mathbb{R})$.*

Proof of Lemma 10.1. It is clear that (10.8) implies quasi-independence. We will show the "only if" part. Assume that (X, Y) is quasi-independent. For $B \in \mathcal{B}(\mathbb{R})$, define $\mathcal{D}_B(X, Y) = \{A \in \mathcal{B}(\mathbb{R}) : (10.8) \text{ holds for } A, B\}$. In what follows, we check that $\mathcal{D}_B(X, Y)$ is a λ-system for any $B \in \mathcal{B}(\mathbb{R})$.

1. It is clear that $\mathbb{R} \in \mathcal{D}_B(X, Y)$.

2. Let $A_1, A_2 \in \mathcal{D}_B(X,Y)$ such that $A_1 \subseteq A_2$. For $A_2 \setminus A_1$, we have

$$\begin{aligned}
&\mathbb{P}(X \in A_2 \setminus A_1, Y \in B) + \mathbb{P}(X \in B, Y \in A_2 \setminus A_1,) \\
&= \mathbb{P}(X \in A_2, Y \in B) - \mathbb{P}(X \in A_1, Y \in B) + \mathbb{P}(X \in B, Y \in A_2) - \mathbb{P}(X \in B, Y \in A_1) \\
&= \mathbb{P}(X \in A_2)\mathbb{P}(Y \in B) - \mathbb{P}(X \in A_1)\mathbb{P}(Y \in B) \\
&\quad + \mathbb{P}(X \in B)\mathbb{P}(Y \in A_2) - \mathbb{P}(X \in B)\mathbb{P}(Y \in A_1) \\
&= \mathbb{P}(X \in A_2 \setminus A_1)\mathbb{P}(Y \in B) + \mathbb{P}(X \in B)\mathbb{P}(Y \in A_2 \setminus A_1).
\end{aligned}$$

Hence, $A_2 \setminus A_1 \in \mathcal{D}_B(X,Y)$.

3. Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ be an increasing sequence of sets in $\mathcal{D}_B(X,Y)$. For $\bigcup_{i=1}^{\infty} A_i$, we have

$$\begin{aligned}
&\mathbb{P}\left(A \in \bigcup_{i=1}^{\infty} A_i, Y \in B\right) + \mathbb{P}\left(A \in B, Y \in \bigcup_{i=1}^{\infty} A_i\right) \\
&= \lim_{n\to\infty} \mathbb{P}(A \in A_n, Y \in B) + \lim_{n\to\infty} \mathbb{P}(A \in B, Y \in A_n) \\
&= \lim_{n\to\infty} \mathbb{P}(A \in A_n)\mathbb{P}(Y \in B) + \lim_{n\to\infty} \mathbb{P}(A \in B)\mathbb{P}(Y \in A_n) \\
&= \mathbb{P}\left(A \in \bigcup_{i=1}^{\infty} A_i\right)\mathbb{P}(Y \in B) + \mathbb{P}(A \in B)\mathbb{P}\left(Y \in \bigcup_{i=1}^{\infty} A_i\right).
\end{aligned}$$

Hence, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}_B(X,Y)$.

Let $B = (-\infty, y]$ for some $y \in \mathbb{R}$. Then quasi-independence of (X,Y) implies that $L := \{(-\infty, x] : x \in \mathbb{R}\} \subseteq \mathcal{D}_B(X,Y)$. It is straightforward to check that L is a π-system. Therefore, Sierpiński–Dynkin's π-λ theorem yields $\sigma(L) \subseteq \mathcal{D}_B(X,Y)$, where $\sigma(L)$ is the smallest σ-algebra containing L. As $\sigma(L) = \mathcal{B}(\mathbb{R})$, we have $\mathcal{D}_B(X,Y) = \mathcal{B}(\mathbb{R})$ for any $B = (-\infty, y]$ with $y \in \mathbb{R}$.

Next, let $\mathcal{D}(X,Y) = \{B \in \mathcal{B}(\mathbb{R}) : \mathcal{D}_B(X,Y) = \mathcal{B}(\mathbb{R})\}$. It is clear that $\mathcal{D}(X,Y)$ is a λ-system and $L \subseteq \mathcal{D}(X,Y)$. Hence, we have $\mathcal{D}(X,Y) = \mathcal{B}(\mathbb{R})$ by the same argument as above. Therefore, we obtain (10.8) for all $A, B \in \mathcal{B}(\mathbb{R})$. □

The next lemma justifies the formulations of the quasi-Fréchet model via (10.9) and (10.10). We omit the proof since it is analogous to that of Lemma 10.1.

Lemma 10.2. *For a random vector $(X,Y) \sim H$ with the marginals F, (X,Y) is quasi-r-Fréchet if and only if (10.10) holds.*

10.8.2 Proofs in Section 10.3.1

Proof of Theorem 10.1. We first show the "only if" part. As $(X, Y) \in \mathrm{IC}_0$, we have $\mathrm{cov}(g(X), g(Y)) = 0$ for any admissible g. For any $A \in \mathcal{B}(\mathbb{R})$, taking $g(x) = \mathbb{1}_A(x)$, we have $\mathrm{cov}(g(X), g(Y)) = \mathrm{cov}(\mathbb{1}_A(X), \mathbb{1}_A(Y)) = 0$. For any $A, B \in \mathcal{B}(\mathbb{R})$, let $g(x) = \mathbb{1}_A(x) + \mathbb{1}_B(x)$, which leads to

$$\begin{aligned}
\mathrm{cov}(g(X), g(Y)) &= \mathrm{cov}\left(\mathbb{1}_A(X) + \mathbb{1}_B(X), \mathbb{1}_A(Y) + \mathbb{1}_B(Y)\right) \\
&= \mathrm{cov}\left(\mathbb{1}_A(X), \mathbb{1}_B(Y)\right) + \mathrm{cov}\left(\mathbb{1}_B(X), \mathbb{1}_A(Y)\right) \\
&= \mathbb{P}\left(X \in A, Y \in B\right) + \mathbb{P}\left(X \in B, Y \in A\right) \\
&\quad - \mathbb{P}\left(X \in A\right)\mathbb{P}\left(Y \in B\right) - \mathbb{P}\left(X \in B\right)\mathbb{P}\left(Y \in A\right).
\end{aligned}$$

As $\mathrm{cov}(g(X), g(Y)) = 0$, we have that (X, Y) satisfies (10.8) for all $A, B \in \mathcal{B}(\mathbb{R})$. Hence, (X, Y) is quasi-independent by Lemma 10.1.

Next, we show that quasi-independence implies $\mathrm{cov}(g(X), g(Y)) = 0$ for any admissible g. As quasi-independence implies (10.8), by taking $A = B$, we have $\mathbb{P}(X \in A, Y \in A) = \mathbb{P}(X \in A)\mathbb{P}(Y \in A)$ for any $A \in \mathcal{B}(\mathbb{R})$. Hence, $\mathrm{cov}(g(X), g(Y)) = 0$ for any indicator functions $g = \mathbb{1}_A$.

First, let $g(x) = \sum_{i=1}^n c_i \mathbb{1}_{A_i}(x)$ be an admissible simple function where $A_1, \ldots, A_n \subseteq \mathbb{R}$ are disjoint measurable sets, and $c_1, \ldots c_n \in \mathbb{R}$ are real numbers. For the simple function g, we have

$$\begin{aligned}
\mathrm{cov}(g(X), g(Y)) &= \sum_{i=1}^n \sum_{j=1}^n c_i c_j \mathrm{cov}\left(\mathbb{1}_{A_i}(X), \mathbb{1}_{A_j}(Y)\right) \\
&= \sum_{1 \leqslant i < j \leqslant n} \left(c_i c_j \mathrm{cov}\left(\mathbb{1}_{A_i}(X), \mathbb{1}_{A_j}(Y)\right) + c_j c_i \mathrm{cov}\left(\mathbb{1}_{A_j}(X), \mathbb{1}_{A_i}(Y)\right)\right) \\
&= \sum_{1 \leqslant i < j \leqslant n} c_i c_j \Big(\mathbb{P}\left(X \in A_i, Y \in A_j\right) - \mathbb{P}\left(X \in A_i\right)\mathbb{P}\left(Y \in A_j\right) \\
&\quad + \mathbb{P}\left(X \in A_j, Y \in A_i\right) - \mathbb{P}\left(X \in A_j\right)\mathbb{P}\left(Y \in A_i\right) \Big) = 0.
\end{aligned}$$

Therefore, we have $\mathrm{cov}(g(X), g(Y)) = 0$ for all admissible simple functions g.

Second, let g be an admissible non-negative function. Admissibility of g implies that $\mathbb{E}[g(X)] < \infty$, $\mathbb{E}[g(Y)] < \infty$ and $\mathbb{E}[g(X)g(Y)] < \infty$. For a non-negative admissible function

g, there exists a sequence of non-negative simple functions $\{g_n\}_{n\geqslant 1}$ such that $g_n \uparrow g$ pointwise. Thus, we can get

$$\begin{aligned}
\operatorname{cov}\left(g(X), g(Y)\right) &= \operatorname{cov}\left(\lim_{n\to\infty} g_n(X), \lim_{n\to\infty} g_n(Y)\right) \\
&= \mathbb{E}\left[\lim_{n\to\infty} g_n(X) \lim_{n\to\infty} g_n(Y)\right] - \mathbb{E}\left[\lim_{n\to\infty} g_n(X)\right]\mathbb{E}\left[\lim_{n\to\infty} g_n(Y)\right] \\
&= \mathbb{E}\left[\lim_{n\to\infty} g_n(X) g_n(Y)\right] - \mathbb{E}\left[\lim_{n\to\infty} g_n(X)\right]\mathbb{E}\left[\lim_{n\to\infty} g_n(Y)\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[g_n(X) g_n(Y)\right] - \lim_{n\to\infty} \mathbb{E}\left[g_n(X)\right]\mathbb{E}\left[g_n(Y)\right] \\
&= \lim_{n\to\infty} \left\{\mathbb{E}\left[g_n(X) g_n(Y)\right] - \mathbb{E}\left[g_n(X)\right]\mathbb{E}\left[g_n(Y)\right]\right\} \\
&= \lim_{n\to\infty} \operatorname{cov}(g_n(X), g_n(Y)) = 0.
\end{aligned}$$

Therefore, we have $\operatorname{cov}(g(X), g(Y)) = 0$ for all admissible non-negative function g.

Finally, let g be an admissible function. Define $g_+ = \max(g, 0)$ and $g_- = -\min(g, 0)$. It is clear that g_+ and g_- are non-negative admissible functions and $g = g_+ - g_-$. Let $G = g_+ + g_-$. We have that G is also a non-negative admissible function. Hence, we get $\operatorname{cov}(g_+(X), g_+(Y)) = 0$, $\operatorname{cov}(g_-(X), g_-(Y)) = 0$ and $\operatorname{cov}(G(X), G(Y)) = 0$, which imply that $\operatorname{cov}(g_+(X), g_-(Y)) + \operatorname{cov}(g_-(X), g_+(Y)) = 0$. As a result, we have $\operatorname{cov}\left(g(X), g(Y)\right) = \operatorname{cov}(g_+(X) - g_-(X), g_+(Y) - g_-(Y)) = 0$. Therefore, we can conclude $(X, Y) \in \mathrm{IC}_0$. □

Proof of Proposition 10.2. For all $A, B \in \mathcal{B}(\mathbb{R})$, we have

$$\mathbb{P}(X_{\pi_1} \in A, X_{\pi_2} \in B) = \frac{1}{2}\mathbb{P}(X_1 \in A, X_2 \in B) + \frac{1}{2}\mathbb{P}(X_1 \in B, X_2 \in A).$$

Since X_1 and X_2 have the same distribution, independence of (X_{π_1}, X_{π_2}) is equivalent to

$$\begin{aligned}
&\mathbb{P}(X_{\pi_1} \in A, X_{\pi_2} \in B) = \mathbb{P}(X_{\pi_1} \in A)\mathbb{P}(X_{\pi_2} \in B) \\
&= \mathbb{P}(X_1 \in A)\mathbb{P}(X_2 \in B) = \mathbb{P}(X_1 \in B)\mathbb{P}(X_2 \in A).
\end{aligned}$$

Hence, independence of (X_{π_1}, X_{π_2}) is equivalent to

$$\mathbb{P}(X_1 \in A, X_2 \in B) + \mathbb{P}(X_1 \in B, X_2 \in A) = \mathbb{P}(X_1 \in A)\mathbb{P}(X_2 \in B) + \mathbb{P}(X_1 \in B)\mathbb{P}(X_2 \in A),$$

which is also equivalent to $(X_1, X_2) \in \mathrm{IC}_0$ by Theorem 10.1. □

Proof of Proposition 10.3. We can get "(i) $\Leftrightarrow$ (ii)" from Theorem 10.1. We only show "(ii) $\Leftrightarrow$ (iii)".

Suppose that $P + P^\top = \mathbf{p}\mathbf{q}^\top + \mathbf{q}\mathbf{p}^\top$. Let $S = P^\top - \mathbf{q}\mathbf{p}^\top$. We first verify that S satisfies (a), (b) and (c).

1. As $s_{ij} + p_i q_j = p_{ij} \geqslant 0$, we have that S satisfies (a).

2. The jth row sum of S is $\sum_{i=1}^n p_{ij} - \sum_{i=1}^n q_j p_i = q_j - q_j = 0$. The ith column sum of S is $\sum_{j=1}^n p_{ij} - \sum_{j=1}^n p_i q_j = p_i - p_i = 0$. Hence, S satisfies (b).

3. We have $S + S^\top = P^\top + P - \mathbf{q}\mathbf{p}^\top - \mathbf{p}\mathbf{q}^\top = O$ where O is an $n \times n$ matrix with all elements equal to 0. Hence, S satisfies (c).

Next, let $P = \mathbf{p}\mathbf{q}^\top + S$ with S satisfies (a), (b) and (c). We verify that P is a valid probability matrix and satisfies (ii). First, it is clear that $p_{ij} \geqslant 0$ as $p_{ij} = p_i q_j + s_{ij} \geqslant 0$. Second, the marginals is $\mathbf{q}$ and $\mathbf{p}$ as $\sum_{j=1}^n p_{ij} = \sum_{j=1}^n (p_i q_j + s_{ij}) = p_i$ and $\sum_{i=1}^n p_{ij} = \sum_{i=1}^n (p_i q_j + s_{ij}) = q_j$. Finally, since $S + S^\top = O$, we have $P + P^\top = \mathbf{p}\mathbf{q}^\top + S + \mathbf{q}\mathbf{p}^\top + S^\top = \mathbf{p}\mathbf{q}^\top + \mathbf{q}\mathbf{p}^\top$. Therefore, we can conclude "(ii) $\Leftrightarrow$ (iii)". □

Proof of Proposition 10.4. It is clear that independent (X, Y) satisfies invariant correlation with $\text{Corr}(X, Y) = 0$. By Theorem 10.1, we only need to show that quasi-independence implies independence under the condition of Proposition 10.4. Without loss of generality, assume that the support of X is $\{1, 2\}$. To show that X and Y are independent, it is enough to show that $\mathbb{P}(X = 1, Y \in A) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A)$ for all $A \in \mathbb{R}$. We show this statement in each of the following 4 cases.

Case 1: Assume $\{1, 2\} \subseteq A$. We have $\mathbb{P}(X \in A) = 1$ and $\mathbb{P}(X \in A, Y = 1) = \mathbb{P}(Y = 1)$. Hence, $\mathbb{P}(X \in A, Y = 1) = \mathbb{P}(X \in A)\mathbb{P}(Y = 1)$. In this case, quasi-independence means $\mathbb{P}(X = 1, Y \in A) + \mathbb{P}(X \in A, Y = 1) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A) + \mathbb{P}(X \in A)\mathbb{P}(Y = 1)$. Thus, we can get $\mathbb{P}(X = 1, Y \in A) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A)$.

Case 2: Assume $1 \notin A$ and $2 \notin A$. We have $\mathbb{P}(X \in A) = 0$ and $\mathbb{P}(X \in A, Y = 1) = 0$. Hence, $\mathbb{P}(X \in A, Y = 1) = \mathbb{P}(X \in A)\mathbb{P}(Y = 1)$. Therefore, we have $\mathbb{P}(X = 1, Y \in A) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A)$ using the same argument as in Case 1.

Case 3: Assume $1 \in A$, $2 \notin A$. First, using Case 2, we have $\mathbb{P}(X = 1, Y \in A \setminus \{1\}) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A \setminus \{1\})$. Furthermore, quasi-independence implies $\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1)$. Therefore,

$$\mathbb{P}(X = 1, Y \in A) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 1, Y \in A \setminus \{1\}) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A).$$

Case 4: Assume $1 \notin A$, $2 \in A$. By the same argument as in Case 3, we have $\mathbb{P}(X = 2, Y \in A) = \mathbb{P}(X = 2)\mathbb{P}(Y \in A)$. Hence,

$$\begin{aligned}\mathbb{P}(X = 1, Y \in A) &= \mathbb{P}(Y \in A) - \mathbb{P}(X = 2, Y \in A) \\ &= \mathbb{P}(Y \in A) - \mathbb{P}(X = 2)\mathbb{P}(Y \in A) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A).\end{aligned}$$

In conclusion, we have $\mathbb{P}(X = 1, Y \in A) = \mathbb{P}(X = 1)\mathbb{P}(Y \in A)$ for all $A \subseteq \mathbb{R}$. As $\mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 1$, we can also obtain $\mathbb{P}(X = 2, Y \in A) = \mathbb{P}(X = 2)\mathbb{P}(Y \in A)$ for all $A \subseteq \mathbb{R}$. Hence, X and Y are independent. □

10.8.3 Proofs in Section 10.3.2

The proof of Proposition 10.5 is built on the following lemma.

Lemma 10.3. *If X and Y are bi-atomic random variables with the same support, then $(X, Y) \in$ IC.*

Proof of Lemma 10.3. As the support of X and Y has only two points, $g(X)$ and $g(Y)$ are linear transformations of X and Y. Therefore, we have $\mathrm{Corr}(g(X), g(Y)) = \mathrm{Corr}(aX+b, aY+b) = \mathrm{Corr}(X, Y)$ for any admissible g. Hence, $(X, Y) \in$ IC regardless of the dependence structure of (X, Y). □

Proof of Proposition 10.5. The "if" part of (ii) is clear from Example 10.1 and Lemma 10.3. We show the "only if" part below.

Let $\{x_1, x_2\}$ and $\{y_1, y_2\}$ be the support of X and Y, respectively, with $x_1 < x_2$ and $y_1 < y_2$. Assume X and Y have different supports. We can easily find a function g such that $g(x_1) < g(x_2)$ and $g(y_1) > g(y_2)$. Then $\mathrm{Corr}(g(X), g(Y)) = -\mathrm{Corr}(X, Y)$ since $g(X)$

is an increasing linear transformation of X and $g(Y)$ is a decreasing linear transformation of Y. Therefore, $(X,Y) \in \mathrm{IC}$ implies $\mathrm{Corr}(X,Y) = 0$. For $r \neq 0$, $(X,Y) \in \mathrm{IC}_r$ implies that X and Y have the same support. If $(X,Y) \in \mathrm{IC}_0$, by Proposition 10.4, X and Y are independent. □

Before showing Theorem 10.2 for the general case, we first analyze the case of the tri-atomic distribution in the following lemma.

Lemma 10.4. *Suppose* $\mathrm{Supp}(X) \subseteq \mathrm{Supp}(Y) = \{x_1, x_2, x_3\}$. *If* $(X,Y) \in \mathrm{IC}$, *then* $F_X = F_Y$ *or* $\mathrm{Corr}(X,Y) = 0$.

Proof of Lemma 10.4. Since a linear transform of g does not change $\mathrm{Corr}(g(X), g(Y))$, it suffices to consider the case $g(x_1) = a$, $g(x_2) = 0$, and $g(x_3) = 1$. Let $p_i = \mathbb{P}(X = x_i)$, $q_j = \mathbb{P}(Y = x_j)$, $p_{ij} = \mathbb{P}(X = x_i, Y = x_j)$ and $s_{ij} = p_{ij} - p_i q_j$ for $i,j \in [3]$. As $\mathrm{Supp}(X) \subseteq \mathrm{Supp}(Y) = \{x_1, x_2, x_3\}$, we can assume that $0 < p_i < 1$ for $i = 1, 3$ and $0 < q_i < 1$ for $j \in [3]$. Write the matrices $P = (p_{ij})_{3\times 3}$ and $S = (s_{ij})_{3\times 3}$, and they are connected by $P = \mathbf{p}\mathbf{q}^\top + S$, where $\mathbf{p} = (p_1, p_2, p_3)^\top$ and $\mathbf{q} = (q_1, q_2, q_3)^\top$. Note that each of the row sums and column sums of S is 0, that is, $\sum_{k=1}^3 s_{kj} = 0 = \sum_{k=1}^3 s_{ik}, \quad i,j \in [3]$.

Define the function $f : \mathbb{R} \to \mathbb{R}$ as

$$\begin{aligned} f(a) :&= \mathrm{Corr}(g(X), g(Y)) = \frac{a^2 p_{11} + a(p_{13} + p_{31}) + p_{33} - (ap_1 + p_3)(aq_1 + q_3)}{\sqrt{a^2 p_1 + p_3 - (ap_1 + p_3)^2}\sqrt{a^2 q_1 + q_3 - (aq_1 + q_3)^2}} \\ &= \frac{a^2 s_{11} + a(s_{13} + s_{31}) + s_{33}}{\sqrt{a^2(p_1 - p_1^2) - 2ap_1p_3 + (p_3 - p_3^2)}\sqrt{a^2(q_1 - q_1^2) - 2aq_1q_3 + (q_3 - q_3^2)}}. \end{aligned}$$

If $(X,Y) \in \mathrm{IC}_r$, the equation $f(a) = r$ holds for all $a \in \mathbb{R}$. Hence, by matching the coefficients in front of a^k, $k \in [4]$, in the equation $f(a) = r$, we obtain the follow equations:

$$s_{11}^2 = r^2(p_1 - p_1^2)(q_1 - q_1^2), \tag{10.27a}$$

$$s_{11}(s_{13} + s_{31}) = -r^2\left(q_1q_3(p_1 - p_1^2) + p_1p_3(q_1 - q_1^2)\right), \tag{10.27b}$$

$$2s_{11}s_{33} + (s_{13} + s_{31})^2 = r^2\left((p_1 - p_1^2)(q_3 - q_3^2) + (p_3 - p_3^2)(q_1 - q_1^2) + 4p_1q_1p_3q_3\right), \tag{10.27c}$$

$$s_{33}(s_{13} + s_{31}) = -r^2\left(p_1p_3(q_3 - q_3^2) + q_1q_3(p_3 - p_3^2)\right), \tag{10.27d}$$

$$s_{33}^2 = r^2(p_3 - p_3^2)(q_3 - q_3^2). \tag{10.27e}$$

Next, we will show that if $r \neq 0$, we have $\mathbf{p} = \mathbf{q}$.

As p_1, p_3, q_1q_3 take values in $(0,1)$ and $r \neq 0$, we have that s_{11}, s_{33} and $s_{13} + s_{31}$ are non-zero. Hence, (10.27a), (10.27b), (10.27d) and (10.27e) yield

$$\frac{s_{11}^2}{s_{33}^2} = \frac{\left(q_1q_3(p_1 - p_1^2) + p_1p_3(q_1 - q_1^2)\right)^2}{\left(p_1p_3(q_3 - q_3^2) + q_1q_3(p_3 - p_3^2)\right)^2} = \frac{(p_1 - p_1^2)(q_1 - q_1^2)}{(p_3 - p_3^2)(q_3 - q_3^2)}.$$

Simplifying the equation, we get

$$\left\{q_1^2q_3^2(p_1 - p_1^2)(p_3 - p_3^2) - p_1^2p_3^2(q_1 - q_1^2)(q_3 - q_3^2)\right\}\left\{(p_1 - p_1^2)(q_3 - q_3^2) - (p_3 - p_3^2)(q_1 - q_1^2)\right\} = 0.$$

This equation is further simplified to

$$(q_1q_3p_2 - p_1p_3q_2)\,(p_2q_3 - q_2p_3) = 0, \tag{10.28}$$

by using $p_1 + p_2 + p_3 = 1$ and $q_1 + q_2 + q_3 = 1$. We can observe that if $p_2 = 0$, then (10.28) cannot hold. Hence, we consider the case when $0 < p_2 < 1$ below.

In this case, we can switch the roles of indices $1, 2, 3$ as p_i and q_i take value in $(0,1)$ for all $i \in [3]$. Hence, we have $f(a) = r$ for all $a \in \mathbb{R}$ if and only if

$$(q_1q_3p_2 - p_1p_3q_2)\,(p_2q_3 - q_2p_3) = 0, \tag{10.29}$$

$$(q_2q_1p_3 - p_2p_1q_3)\,(p_3q_1 - q_3p_1) = 0, \tag{10.30}$$

$$(q_3q_2p_1 - p_3p_2q_1)\,(p_1q_2 - q_1p_2) = 0. \tag{10.31}$$

Now we consider all possible cases when (10.29), (10.30) and (10.31) hold simultaneously.

First, assume that all the first terms in (10.29), (10.30) and (10.31) equal 0; that is $q_1q_3p_2 = p_1p_3q_2$, $q_2q_1p_3 = p_2p_1q_3$ and $q_3q_2p_1 = p_3p_2q_1$. Thus, we have $q_1q_2q_3 = p_1p_2p_3$. As a result, we get $p_1^2 = q_1^2$, $p_2^2 = q_2^2$ and $p_3^2 = q_3^2$, which lead to $p_1 = q_1$, $p_2 = q_2$ and $p_3 = q_3$.

Second, assume that all the second terms in (10.29), (10.30) and (10.31) equal 0; that is $p_2q_3 = q_2p_3$, $p_3q_1 = q_3p_1$ and $p_1q_2 = q_1p_3$. Thus, we have $q_1/p_1 = q_2/p_2 = q_3/p_3$, which leads to $p_1 = q_1$, $p_2 = q_2$ and $p_3 = q_3$. We observe that one of the three equations $p_2q_3 = q_2p_3$, $p_3q_1 = q_3p_1$ and $p_1q_2 = q_1p_3$ is redundant. Hence, if two of the second terms in (10.29), (10.30) and (10.31) equal 0, the third one also equals 0.

In the final step, we consider the case that two of the first terms in (10.29), (10.30) and (10.31) equal 0. Without loss of generality, we assume $q_1q_3p_2 = p_1p_3q_2$, $q_2q_1p_3 = p_2p_1q_3$ and $p_1q_2 = q_1p_2$. By $q_1q_3p_2 = p_1p_3q_2$ and $p_1q_2 = q_1p_2$, we get $p_3 = q_3$. Combining with $q_2q_1p_3 = p_2p_1q_3$, we have $p_1^2 = q_1^2$ and $p_2^2 = q_2^2$, which gives $p_1 = q_1$ and $p_2 = q_2$.

In conclusion, if $(X, Y) \in \mathrm{IC}_r$ with $r \neq 0$, then $\mathbf{p} = \mathbf{q}$, or equivalently, $F_X = F_Y$. □

Proof of Theorem 10.2. The "if" part is clear. For the "only if" part, we will show that if $(X, Y) \in \mathrm{IC}$, then $\mathrm{Corr}(X, Y) = 0$. We first discuss the case X and Y have different supports. Let $A = \mathrm{Supp}(X)$ and $B = \mathrm{Supp}(Y)$.

(a) Suppose $A \cap B \neq A$ and $A \cap B \neq B$. In this case, we can find a function g such that $g(X)$ and $g(Y)$ are bi-atomic random variables with different supports. By Proposition 10.1 and Proposition 10.5, we have that $(g(X), g(Y)) \in \mathrm{IC}$ and $g(X)$ and $g(Y)$ are independent, which gives $\mathrm{Corr}(g(X), g(Y)) = 0$. As $(X, Y) \in \mathrm{IC}$, we have $\mathrm{Corr}(X, Y) = 0$.

(b) Suppose $A \subsetneq B$. We can find a function such that $g(X)$ is a bi-atomic random variable and $g(Y)$ is a tri-atomic random variable. By Proposition 10.1 and Lemma 10.4, we have that $(g(X), g(Y)) \in \mathrm{IC}$ and $\mathrm{Corr}(g(X), g(Y)) = 0$. Hence, $\mathrm{Corr}(X, Y) = 0$. If $B \subsetneq A$, we can also have $\mathrm{Corr}(X, Y) = 0$.

Next, we discuss the case X and Y have the same support. Let $S = \mathrm{Supp}(X) = \mathrm{Supp}(Y)$. Because X and Y have different distributions, there exists a set $A \subsetneq S$ such that $\mathbb{P}(X \in A) \neq \mathbb{P}(Y \in A)$. Furthermore, we can also find $B \subsetneq S$ such that $B \cap A = \emptyset$ and $A \cup B \subsetneq S$ as S contains more than two points. Thus, the sets A, B and $S/(A \cup B)$ are three non-empty and exclusive sets. Let $g(x) = \mathbb{1}_A + 2\mathbb{1}_B + 3\mathbb{1}_{S/(A\cup B)}$. It is clear that $g(X)$ and $g(Y)$ are tri-atomic random variables with the same support [3] but have different distributions. Therefore, by Lemma 10.4, we have $\mathrm{Corr}(g(X), g(Y)) = 0$, and thus $\mathrm{Corr}(X, Y) = 0$. □

10.8.4 Proofs in Section 10.3.3

Proof of Proposition 10.6. We first show the necessity. If $(X, Y) \in \mathrm{IC}_r$, then

$$\mathrm{Corr}(g(X), g(Y)) = \frac{\mathbf{z}^\top (P - \mathbf{p}\mathbf{p}^\top)\mathbf{z}}{\mathbf{z}^\top (D - \mathbf{p}\mathbf{p}^\top)\mathbf{z}} = r, \quad \mathbf{z} = g(\mathbf{x}) = (g(x_1), \ldots g(x_n))^\top, \tag{10.32}$$

holds for all admissible function g, that is, for all real vectors $\mathbf{z} \in \breve{\mathbb{R}}^n$ with $\breve{\mathbb{R}}^n = \mathbb{R}^n \backslash \{c\mathbf{1}_n : c \in \mathbb{R}\}$. Equation (10.32) is equivalent to

$$\mathbf{z}^\top (P - rD - (1-r)\mathbf{p}\mathbf{p}^\top)\mathbf{z} = 0 \quad \text{for all } \mathbf{z} \in \breve{\mathbb{R}}^n. \tag{10.33}$$

This implies that $(P + P^\top - 2rD - 2(1-r)\mathbf{p}\mathbf{p}^\top)\mathbf{z} = \mathbf{0}_n$ for all $\mathbf{z} \in \breve{\mathbb{R}}^n$. By taking $\mathbf{z} = \mathbf{e}_i$ for $i \in [n]$, we have that $P + P^\top - 2rD - 2(1-r)\mathbf{p}\mathbf{p}^\top = O$, and thus we obtain (10.11).

Next, we show the sufficiency. Suppose that the probability matrix P satisfies (10.11). Then we have that

$$P - rD - (1-r)\mathbf{p}\mathbf{p}^\top + P^\top - rD - (1-r)\mathbf{p}\mathbf{p}^\top = O.$$

Therefore, for every $\mathbf{z} \in \mathbb{R}^n$, we have

$$\begin{aligned}\mathbf{z}^\top (P - rD - (1-r)\mathbf{p}\mathbf{p}^\top)\mathbf{z} &= \mathbf{z}^\top (P - rD - (1-r)\mathbf{p}\mathbf{p}^\top)^\top \mathbf{z} = \mathbf{z}^\top (P^\top - rD - (1-r)\mathbf{p}\mathbf{p}^\top)\mathbf{z} \\ &= -\mathbf{z}^\top (P - rD - (1-r)\mathbf{p}\mathbf{p}^\top)\mathbf{z},\end{aligned}$$

and thus $\mathbf{z}^\top (P - rD - (1-r)\mathbf{p}\mathbf{p}^\top)\mathbf{z} = 0$. Therefore, we have $(X, Y) \in \mathrm{IC}_r$.

In order for $rD + (1-r)\mathbf{p}\mathbf{p}^\top$ to be a probability matrix, $r \in [-1, 1]$ has to satisfy $0 \leqslant rp_j + (1-r)p_j^2 \leqslant 1$ for all $j \in [n]$ and $0 \leqslant (1-r)p_i p_j \leqslant 1$ for all $i, j \in [n]$, $i \neq j$, Equivalently, $-p_j/(1-p_j) \leqslant r \leqslant 1 + 1/p_j$ for all $j \in [n]$, and $1 - 1/(p_i p_j) \leqslant r$ for all $i, j \in [n]$, $i \neq j$. Therefore, r satisfies (10.12). □

Proof of Theorem 10.3. We first show the necessity. It suffices to show (10.9) for every $x, y \in \mathrm{Supp}(X)$ since the cumulative distribution functions change only at such points.

The case that X and Y are bi-atomic is verified in Proposition 10.6. In what follows, we suppose $X, Y \in \mathcal{L}^2$ are not bi-atomic. Let x and y be any two distinct points in $\mathrm{Supp}(X)$. In view of Proposition 10.6, it suffices to consider the case when there exist three distinct points $z_1, z_2, z_3 \in \mathrm{Supp}(X)$, $z_1 < z_2 < z_3$, with two of them equal x and y. Let $z_0 = -\infty$ and $z_4 = \infty$. Define

$$h(t) = \sum_{i=0}^{3} \mathbb{1}_{\{t > z_i\}}, \quad t \in \mathbb{R}. \tag{10.34}$$

Since h is admissible, Proposition 10.1 implies that $(h(X), h(Y)) \in \mathrm{IC}_r$. Note that $h(X)$ and $h(Y)$ are identical random variables taking the value i with probability $p_i = \mathbb{P}(z_{i-1} < X \leqslant z_i)$ for $i \in [4]$, respectively. Since $z_1, z_2, z_3 \in \mathrm{Supp}(X)$, we have $p_1, p_2, p_3 > 0$. We exclude the index $i = 4$ in the following proof when $p_4 = 0$. The joint probability matrix of $(h(X), h(Y))$, denoted by $P = (p_{ij})_{4\times 4}$, is then given by

$$\begin{aligned} p_{ij} = \mathbb{P}\left(h(X) = i, h(Y) = j\right) &= \mathbb{P}(z_{i-1} < X \leqslant z_i,\ z_{j-1} < Y \leqslant z_j) \\ &= H(z_i, z_j) - H(z_{i-1}, z_j) - H(z_i, z_{j-1}) + H(z_{i-1}, z_{j-1}) =: p_{ij}(H) \end{aligned}$$

for $i, j \in [4]$. Let $H^\top(x, y) = H(y, x)$ be the distribution function of (Y, X). By Proposition 10.6, we have that

$$\frac{p_{ij}(H) + p_{ij}(H^\top)}{2} = r p_i \mathbb{1}_{\{i=j\}} + (1-r) p_i p_j \quad \text{for all } i, j \in [4]. \tag{10.35}$$

Since $p_{11}(H) = p_{11}(H^\top) = H(z_1, z_1)$ and $p_1 = F(z_1)$, we have $H(z_1, z_1) = r \min(F(z_1), F(z_1)) + (1-r)F(z_1)F(z_1)$. Next, by taking $(i, j) = (1, 2)$ in (10.35), we have

$$\begin{aligned} \frac{H(z_1, z_2) + H(z_2, z_1)}{2} &= H(z_1, z_1) + \frac{p_{12}(H) + p_{12}(H^\top)}{2} \\ &= r \min(F(z_1), F(z_1)) + (1-r)F(z_1)F(z_1) + (1-r)F(z_1)\{F(z_2) - F(z_1)\} \\ &= r \min(F(z_1), F(z_2)) + (1-r)F(z_1)F(z_2). \end{aligned}$$

By repeating this calculation for all pairs of indices $(i, j) \in [4]^2$, we eventually obtain

$$\frac{H(z_i, z_j) + H(z_i, z_j)}{2} = r \min(F(z_i), F(z_j)) + (1-r)F(z_i)F(z_j),$$

for all $(i, j) \in [4]^2$. Since the set $\{z_1, z_2, z_3\}$ includes x and y, we have (10.9) for (x, x), (x, y), (y, x) and (y, y). Since x and y, $x \neq y$ are taken arbitrary in $\mathrm{Supp}(X)$, we obtain the desired identity (10.9).

Next, we show the sufficiency. We show $(X, Y) \in \mathrm{IC}_r$ when $(X, Y) \sim H$ satisfies (10.9).

Let g be an admissible function. Denote by H_g and F_g the joint and marginal distributions of $(g(X), g(Y))$, respectively. By taking $A = \{t \in \mathbb{R} : g(t) \leqslant x\}$ and $B = \{t \in \mathbb{R} : g(t) \leqslant y\}$ in (10.10), we have that

$$\frac{H_g(x, y) + H_g(y, x)}{2} = r \min(F_g(x), F_g(y)) + (1-r)F_g(x)F_g(y), \quad (x, y) \in \mathbb{R}. \tag{10.36}$$

Note that $(x, y) \mapsto H_g(y, x)$ is the distribution function of $(g(Y), g(X))$ and the correlation coefficient is invariant under permutation, that is, $\operatorname{cov}(g(X), g(Y)) = \operatorname{cov}(g(Y), g(X))$. Together with the Fréchet-Hoeffding identity (Lemma 2 of Lehmann, 1966), (10.36) implies that

$$\begin{aligned}
&\operatorname{cov}(g(X), g(Y)) + \operatorname{cov}(g(X), g(Y)) \\
&= 2 \int_{\mathbb{R}} \int_{\mathbb{R}} \{r \min(F_g(x), F_g(y)) + (1-r) F_g(x) F_g(y) - F_g(x) F_g(y)\} \, \mathrm{d}x \, \mathrm{d}y \\
&= 2r \int_{\mathbb{R}} \int_{\mathbb{R}} \{\min(F_g(x), F_g(y)) - F_g(x) F_g(y)\} \, \mathrm{d}x \, \mathrm{d}y \\
&= 2r \operatorname{var}(g(X)) < \infty.
\end{aligned}$$

Therefore, we have $\operatorname{Corr}(g(X), g(Y)) = r$ and we conclude $(X, Y) \in \mathrm{IC}_r^{\uparrow}$. □

Proof of Corollary 10.2. It suffices to show that the range of the admissible invariant correlation is $0 \leqslant r \leqslant 1$. Suppose (10.15) holds. Since the LHS of (10.15) is a copula, so is the RHS. Therefore, we have that $V_{rM+(1-r)\Pi}([\underline{a}, \overline{a}] \times [\underline{b}, \overline{b}]) \geqslant 0$ for all $\underline{a}, \overline{a}, \underline{b}, \overline{b} \in \mathcal{I}$ such that $\underline{a} \leqslant \overline{a}$ and $\underline{b} \leqslant \overline{b}$, where V_C is a C-volume of a copula C. Consider the case when $\underline{a} \leqslant \underline{b} \leqslant \overline{b} \leqslant \overline{a}$. Then, for $l := \overline{a} - \underline{a}$, we have that

$$V_{rM+(1-r)\Pi}([\underline{a}, \overline{a}] \times [\underline{b}, \overline{b}]) = r(\overline{b} - \underline{b}) + (1-r)(\overline{a} - \underline{a})(\overline{b} - \underline{b}) \geqslant 0,$$

that is, $r + (1-r)l \geqslant 0$. Equivalently, we have $r \geqslant -l/(1-l)$. By letting $l \downarrow 0$, we obtain $r \geqslant 0$. □

10.8.5 Proofs in Section 10.4.1

Proof of Proposition 10.8. (i) For $i \in [d]$ and $t \in [0, 1]$, we have

$$\mathbb{P}(X_i \leqslant t) = \sum_{j=1}^{k} \mathbb{P}(Z_{ij} = 1) \mathbb{P}(U_j \leqslant t) = t \sum_{j=1}^{k} \mathbb{E}[Z_{ij}] = t.$$

Hence, X_i has the standard uniform distribution.

(ii) We can check that the copula of (X_1, X_2) is given by

$$\begin{aligned}
\mathbb{P}(X_1 \leqslant u_1, X_2 \leqslant u_2) &= \sum_{j=1}^{k} \mathbb{P}(Z_{1j} Z_{2j} = 1) \min(u_1, u_2) + \left(1 - \sum_{j=1}^{k} \mathbb{P}(Z_{1j} Z_{2j} = 1)\right) u_1 u_2 \\
&= \mathbb{E}[\mathbf{Z}_1^\top \mathbf{Z}_2] \min(u_1, u_2) + \left(1 - \mathbb{E}[\mathbf{Z}_1^\top \mathbf{Z}_2]\right) u_1 u_2,
\end{aligned}$$

and this argument also applies to the other pairs. Therefore, each pair of components of $\mathbf{X}$ has a positive Fréchet copula, and hence it has an invariant correlation. By Proposition 10.7, we know that $\mathbf{X}$ has an invariant correlation matrix.

(iii) From (ii), we know that the correlation between X_1 and X_2 is $\mathbb{E}[\mathbf{Z}_1^\top \mathbf{Z}_2]$, and similarly for the other pairs. Therefore, the correlation matrix of $\mathbf{X}$ is $\mathbb{E}[\Gamma\Gamma^\top]$. □

For the proof of Theorem 10.4, we first present the following two lemmas.

Lemma 10.5. *For $k \in \mathbb{N}$, the following statements hold:*

(i) $\mathcal{Z}_{d,k}$ is increasing in k;

(ii) $\mathcal{Z}_{d,k}$ is convex;

(iii) $\mathcal{Z}_{d,k} \subseteq \Theta_d$. In particular, any matrix in $\mathcal{Z}_{d,k}$ is a correlation matrix.

Proof of Lemma 10.5. (i) It suffices to note $\mathcal{Z}_{d,k-1} \subseteq \mathcal{Z}_{d,k}$ for $k \geqslant 2$, which can be checked by taking $Z_{ik} = 0$ for all $i \in [d]$ in (10.16).

(ii) This can be checked by the fact that for any event A independent of two $d \times k$ categorical random matrices Γ and Γ', the matrix $\mathbb{1}_A\Gamma + (1 - \mathbb{1}_A)\Gamma'$ is a $d \times k$ categorical random matrix.

(iii) This follows from Proposition 10.8. □

Lemma 10.6. $\mathcal{Z}_{d,k} = \mathrm{Conv}(\mathcal{S}_{d,k})$.

Proof of Lemma 10.6. We have that $\mathcal{Z}_{d,k} \supseteq \mathrm{Conv}(\mathcal{S}_{d,k})$ since $\mathcal{Z}_{d,k} \supseteq \mathcal{S}_{d,k}$ and the set $\mathcal{Z}_{d,k}$ is convex. Therefore, it suffices to prove that $\mathcal{Z}_{d,k} \subseteq \mathrm{Conv}(\mathcal{S}_{d,k})$.

Let Γ be a $d \times k$ categorical random matrix such that $\mathbb{E}[\Gamma\Gamma^\top] \in \mathcal{Z}_{d,k}$. Let $\gamma = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d)^\top \in \{0,1\}^{d\times k}$ be any realization of the random matrix Γ. Define $A(\gamma) = (A_1(\gamma), \ldots, A_k(\gamma))$, where $A_s(\gamma) = \{i \in [d] : \gamma_{is} = 1\}$, $s \in [k]$. Then $A(\gamma)$ is a partition

of $[d]$ into k subsets since $\sum_{s=1}^{k}\gamma_{is}=1$ and thus every $i\in[d]$ belongs to only one of the partitioned subsets. Moreover, we have $\gamma^\top\gamma\in\mathcal{S}_{d,k}$ since

$$(\gamma^\top\gamma)_{ij}=\boldsymbol{\gamma}_i^\top\boldsymbol{\gamma}_j=\sum_{s=1}^{k}\gamma_{is}\gamma_{js}=\sum_{s=1}^{k}\mathbb{1}_{\{\gamma_{is}=1\,\gamma_{js}=1\}}=\sum_{s=1}^{k}\mathbb{1}_{\{i,j\in A_s(\gamma)\}}.$$

Therefore, we have that $\mathbb{E}[\Gamma\Gamma^\top]=\sum_\gamma\gamma\gamma^\top\mathbb{P}(\Gamma=\gamma)\in\mathrm{Conv}(\mathcal{S}_{d,k})$. As a result, we have $\mathcal{Z}_{d,k}=\mathrm{Conv}(\mathcal{S}_{d,k})$. □

Proof of Theorem 10.4. The last two equalities follow directly from Part (i) of Lemma 10.5, Lemma 10.6 and the fact that $\mathcal{S}_{d,k}=\mathcal{S}_{d,d}$ for $k\geqslant d$. Moreover, Proposition 10.8 implies that $\Theta_d\supseteq\mathcal{Z}_{d,d}$. Therefore, it suffices to prove $\Theta_d\subseteq\mathrm{Conv}(\mathcal{S}_{d,d})$.

Let $\mathbf{Y}=(Y_1,\ldots,Y_d)^\top$ be a continuous random vector with standard uniform marginals and invariant correlation $R=(r_{ij})_{d\times d}$. Let $L=\{(u,v)\in[0,1]:u<v\}$. Then its volume with respect to M and Π are $V_M(L)=0$ and $V_\Pi(L)=1/2$, respectively. By Proposition 10.7, we have that

$$\mathbb{P}(Y_i\neq Y_j)=\mathbb{P}((Y_i,Y_j)\in L)+\mathbb{P}((Y_j,Y_i)\in L)=2r_{ij}V_M(L)+2(1-r_{ij})V_\Pi(L)=1-r_{ij},$$

and hence $\mathbb{P}(Y_i=Y_j)=r_{ij}$ for all $i,j\in[d]$.

For $\mathbf{y}\in[0,1]^d$, let $A(\mathbf{y})=(A_1(\mathbf{y}),\ldots,A_d(\mathbf{y}))$ be any partition of $[d]$ with d sets such that two indices i and j are in the same partitioned subset if and only if $y_i=y_j$. In addition, define the matrix $\Sigma(\mathbf{y})=(\Sigma_{ij}(\mathbf{y}))_{d\times d}$ by $\Sigma_{ij}(\mathbf{y})=\sum_{s=1}^{d}\mathbb{1}_{\{i,j\in A_s(\mathbf{y})\}}$. Note that, although $\mathbf{y}$ takes a value on the uncountable set $[0,1]^d$, there is only a finite number of different partitions of $[d]$, and thus so is $\Sigma(\mathbf{y})$. Let $\Sigma^{(1)},\ldots,\Sigma^{(N)}$, $N\in\mathbb{N}$, be all possible such matrices. For the given random vector $\mathbf{Y}$, let $\alpha_n=\mathbb{P}(\Sigma(\mathbf{Y})=\Sigma^{(n)})$, $n\in[N]$. Then

$$r_{ij}=\mathbb{P}(Y_i=Y_j)=\mathbb{E}[\mathbb{1}_{\{Y_i=Y_j\}}]=\mathbb{E}\left[\sum_{s=1}^{d}\mathbb{1}_{\{i,j\in A_s(\mathbf{Y})\}}\right]=\mathbb{E}[\Sigma_{ij}(\mathbf{Y})].$$

Since $\Sigma^{(n)}\in\mathcal{S}_{d,d}$ for every $n\in[N]$ and $\sum_{n=1}^{N}\alpha_n=1$, we conclude that $R=\mathbb{E}[\Sigma(\mathbf{Y})]=\sum_{n=1}^{N}\alpha_n\Sigma^{(n)}\in\mathrm{Conv}(\mathcal{S}_{d,d})$. Therefore, we obtain the desired result. □

10.8.6 Proofs in Section 10.4.2

Proof of Proposition 10.9. Fix $i \in [d]$ and an increasing set $A \subseteq \mathbb{R}^d$. For every $s \in [k]$, the joint law of (Z_{is}, X_i) is identical to that of (Z_{is}, U_s), where Z_{is} is the (i,s)th element of Γ. Indeed, for any $A \subseteq [0,1]$, we have that

$$\mathbb{P}(Z_{is} = 1,\ X_i \in A) = \mathbb{P}\left(Z_{is} = 1,\ \sum_{l=1}^{k} Z_{il} U_l \in A\right) = \mathbb{P}(Z_{is} = 1,\ U_s \in A).$$

This implies the independence between Z_{is} and X_i since

$$\mathbb{P}(Z_{is} = 1,\ X_i \in A) = \mathbb{P}(Z_{is} = 1,\ U_s \in A) = \mathbb{P}(Z_{is} = 1)\mathbb{P}(U_s \in A) = \mathbb{P}(Z_{is} = 1)\mathbb{P}(X_i \in A).$$

Therefore, we have that, for $x \in [0,1]$,

$$\begin{aligned}\mathbb{P}(\mathbf{X} \in A \mid X_i = x) &= \sum_{s=1}^{k} \mathbb{P}(\mathbf{X} \in A \mid X_i = x, Z_{is} = 1)\mathbb{P}(Z_{is} = 1 \mid X_i = x) \\ &= \sum_{s=1}^{k} \mathbb{P}(\mathbf{X} \in A \mid U_s = x, Z_{is} = 1)\mathbb{P}(Z_{is} = 1),\end{aligned}$$

and hence it suffices to prove that $x \mapsto \mathbb{P}(\mathbf{X} \in A \mid U_s = x, Z_{is} = 1)$ is increasing for every $s \in [k]$. Let γ_{-s} be a $d \times (k-1)$ matrix obtained from deleting the sth column from γ, and $\mathbf{u}_{-s}$ be a $(k-1)$-dimensional vector obtained from deleting the sth element from $\mathbf{u}$. Since U_s is independent of Γ and $\mathbf{U}_{-s}$, we have that

$$\begin{aligned}\mathbb{P}(\mathbf{X} \in A \mid U_s = x, Z_{is} = 1) &= \mathbb{P}(\Gamma_{-s}\mathbf{U}_{-s} + x(Z_{is}, \dots, Z_{ds})^\top \in A \mid U_s = x, Z_{is} = 1) \\ &= \mathbb{P}(\Gamma_{-s}\mathbf{U}_{-s} + x(Z_{is}, \dots, Z_{ds})^\top \in A \mid Z_{is} = 1).\end{aligned}$$

The probability in the last expression is increasing in x since $(Z_{is}, \dots, Z_{ds}) \geqslant \mathbf{0}_d$ and A is an increasing set. □

10.8.7 Proofs in Section 10.5

We first present some lemmas for the proof of Theorem 10.5.

Lemma 10.7. *Suppose that X is bi-atomic and Y is tri-atomic. If $(X,Y) \in \mathrm{IC}_r^{\uparrow}$ for some $r \in [-1,1]$, then $r = 0$.*

Proof of Lemma 10.7. Assume $\mathrm{Supp}(X) = \{x_1, x_2\}$ for some $x_1 < x_2$ and $\mathrm{Supp}(Y) = \{y_1, y_2, y_3\}$ for some $y_1 < y_2 < y_3$. Let $P = (p_{ij})_{2\times 3}$ be the probability matrix, $\mathbf{p} = (p_1, p_2)^\top$ and $\mathbf{q} = (q_1, q_2, q_3)^\top$ be the marginal distributions of X and Y, respectively, with $p_i = \mathbb{P}(X = x_i)$ and $q_j = \mathbb{P}(Y = y_j)$ for $i \in [2]$ and $j \in [3]$ and $S = (s_{ij})_{2\times 3} = P - \mathbf{p}^\top \mathbf{q}$. We have $p_i > 0$ for $i = 1, 2$ and $q_j > 0$ for $j \in [3]$.

Since a linear transform of g does not change $\mathrm{Corr}(g(X), g(Y))$, we can fix $g(x_1) = 0$ and $g(x_2) = 1$. Assume $g(y_1) = z_1$, $g(y_2) = z_2$ and $g(y_3) = z_3$ with $z_1 \leqslant z_2 \leqslant z_3$. As $(X, Y) \in \mathrm{IC}_r^\uparrow$, we have

$$\mathrm{Corr}(g(X), g(Y)) = \frac{z_1 s_{21} + z_2 s_{22} + z_3 s_{23}}{\sqrt{p_2 - p_2^2}\sqrt{z_1^2 q_1 + z_2^2 q_2 + z_3^2 q_3 - (z_1 q_1 + z_2 q_2 + z_3 q_3)^2}} = r.$$

By matching the coefficients in front of z_1^2, z_2^2, z_3^2, $z_1 z_2$, $z_1 z_3$ and $z_2 z_3$ terms, we get the system

$$s_{21}^2 = r^2 (p_2 - p_2^2)(q_1 - q_1^2), \tag{10.37a}$$

$$s_{22}^2 = r^2 (p_2 - p_2^2)(q_2 - q_2^2), \tag{10.37b}$$

$$s_{32}^2 = r^2 (p_2 - p_2^2)(q_3 - q_3^2), \tag{10.37c}$$

$$s_{21} s_{22} = r^2 (p_2 - p_2^2) q_1 q_2, \tag{10.37d}$$

$$s_{21} s_{23} = r^2 (p_2 - p_2^2) q_1 q_3, \tag{10.37e}$$

$$s_{22} s_{23} = r^2 (p_2 - p_2^2) q_2 q_3. \tag{10.37f}$$

As g can be any increasing function, at most one of z_1, z_2, z_3 can be fixed as 0. If $z_3 = 0$, we have (10.37a), (10.37b) and (10.37d) hold simultaneously. Thus, $r \neq 0$ implies $q_3 = 0$. If $z_2 = 0$, we have (10.37a), (10.37c) and (10.37e) hold simultaneously. Thus, $r \neq 0$ implies $q_2 = 0$. If $z_1 = 0$, we have (10.37b), (10.37c) and (10.37f) hold simultaneously. Thus, $r \neq 0$ implies $q_1 = 0$. If none of z_1, z_2, z_3 is 0, then we have $q_1 = q_2 = q_3 = 0$. As $q_1, q_2, q_3 > 0$, we have $r = 0$. □

Lemma 10.8. *Suppose both X and Y are tri-atomic random variables with the same support. If $(X, Y) \in \mathrm{IC}_r^\uparrow$ for some $r \in [-1, 1]$, then $r = 0$ or $F_X = F_Y$.*

Proof of Lemma 10.8. As $p_i = \mathbb{P}(X = x_i) > 0$ and $q_i = \mathbb{P}(Y = y_j) > 0$ for all $i, j \in [3]$, we can use the similar argument in the proof of Lemma 10.4 to show that if $(X, Y) \in \mathrm{IC}_r^\uparrow$ with $r \neq 0$, then $F_X = F_Y$. □

Lemma 10.9. *Suppose X is m-atomic and Y is n-atomic with $m \geqslant 2$, $n > 2$ and* $\mathrm{Supp}(X) \neq \mathrm{Supp}(Y)$. *If $(X, Y) \in \mathrm{IC}_r^{\uparrow}$ for some $r \in [-1, 1]$, then $r = 0$.*

Proof of Lemma 10.9. It suffices to show that, for such X and Y, there exists an increasing function g such that $g(X)$ is bi-atomic and $g(Y)$ is tri-atomic, or $g(X)$ is tri-atomic and $g(Y)$ is bi-atomic. Then, we can use Lemma 10.7 to get $r = 0$. The existence of g can be checked directly by exhausting all possibilities of the supports of X and Y. For instance, if there exists $y_0 \in \mathrm{Supp}(Y)$ but not in $\mathrm{Supp}(X)$, and each of the events $\{X < y_0\}$, $\{X > y_0\}$, $\{Y < y_0\}$ and $\{Y > y_0\}$ has non-zero probability, then the function $g : y \mapsto \mathbb{1}_{\{y > y_0\}} + \mathbb{1}_{\{y \geqslant y_0\}}$ is sufficient. We omit the details of all other cases here. □

Proof of Theorem 10.5. It is clear that $\mathrm{IC}_r \subseteq \mathrm{IC}_r^{\uparrow}$ for all $r \in [-1, 1]$, which implies the "only if" part. We will show the "if" part below.

(i) If $(X, Y) \in \mathrm{IC}_0^{\uparrow}$, then $\mathrm{cov}(g(X), g(Y)) = 0$ for all admissible decreasing functions g. By taking $g(x) = \mathbb{1}_{\{x \leqslant a\}}$, we have $\mathrm{cov}(\mathbb{1}_{\{X \leqslant a\}}, \mathbb{1}_{\{Y \leqslant a\}}) = 0$ for every $a \in \mathbb{R}$. For every fixed $a, b \in \mathbb{R}$, let $g(x) = \mathbb{1}_{\{x \leqslant a\}} + \mathbb{1}_{\{x \leqslant b\}}$. As $\mathrm{cov}(\mathbb{1}_{\{X \leqslant a\}}, \mathbb{1}_{\{Y \leqslant a\}}) = 0$ and $\mathrm{cov}(\mathbb{1}_{\{X \leqslant b\}}, \mathbb{1}_{\{Y \leqslant b\}}) = 0$, we have from $\mathrm{cov}(g(X), g(Y)) = 0$ that $\mathrm{cov}\left(\mathbb{1}_{\{X \leqslant a\}}, \mathbb{1}_{\{Y \leqslant b\}}\right) + \mathrm{cov}\left(\mathbb{1}_{\{X \leqslant b\}}, \mathbb{1}_{\{Y \leqslant a\}}\right) = 0$. Hence (X, Y) is quasi-independent, and thus $(X, Y) \in \mathrm{IC}_0$ by Theorem 10.1.

(ii) Assume X and Y have identical distributions. By Theorem 10.3, it only remains to show that (10.9) holds when $(X, Y) \in \mathrm{IC}_r^{\uparrow}$.

Notice that the necessity part of the proof of Theorem 10.3 directly applies with a few modifications since, for example, the function (10.34) is increasing and admissible, and thus Item (i) in Corollary 10.3 is used instead of Proposition 10.1. By carefully checking the proof of Theorem 10.3, we only need to show that Proposition 10.6 holds for the case of $\mathrm{IC}_r^{\uparrow}$. Using $\mathrm{IC}_r^{\uparrow} \subseteq \mathrm{IC}_r$, we know that n-atomic identically distributed random variables X and Y satisfy $(X, Y) \in \mathrm{IC}_r^{\uparrow}$ if r satisfies (10.12) and the probability matrix P satisfies (10.11). We show the remaining "only if" part of this statement.

To this end, let X and Y be n-atomic identically distributed random variables taking values in $\mathcal{X} = \{x_1, \dots, x_n\}$, $n \in \mathbb{N}$, with $x_1 < \cdots < x_n$. Write $P = (p_{ij})_{n\times n}$, $p_{ij} = \mathbb{P}(X = x_i, Y = x_j)$, with $\mathbf{p} = P\mathbf{1}_n$ and $D = \operatorname{diag}(\mathbf{p})$.

Suppose that $(X, Y) \in \mathrm{IC}_r^{\uparrow}$ and let $A := P + P^\top - 2rD - 2(1-r)\mathbf{p}\mathbf{p}^\top$. Since (10.32) holds for all admissible increasing functions g, (10.33) holds for all $\mathbf{z} \in \breve{\mathbb{R}}^n$ such that $z_1 < \cdots < z_n$. Therefore, for such $\mathbf{z}$'s, we have $A\mathbf{z} = \mathbf{0}_n$. For $\mathbf{z} \in \breve{\mathbb{R}}^n$ such that $z_1 < \cdots < z_n$, there exists a sufficiently small $\delta > 0$ such that $\mathbf{z} + \epsilon\mathbf{e}_j \in \breve{\mathbb{R}}^n$ is still increasing for all $\epsilon < \delta$ and $j \in [n]$. Since $A\mathbf{z} = A(\mathbf{z}+\epsilon\mathbf{e}_j) = \mathbf{0}_n$, we obtain $A\mathbf{e}_j = \mathbf{0}_n$ for all $j \in [n]$, and thus $A = O$. Therefore, we have (10.11). The range condition (10.12) of r necessarily holds for $rD + (1-r)\mathbf{p}\mathbf{p}^\top$ to be a probability matrix.

(iii) For $X, Y \in \mathcal{L}^2$ with different distributions and $|\mathrm{Supp}(Y)| > 2$, we will show that $(X, Y) \in \mathrm{IC}_0^{\uparrow}$ implies $r = 0$.

As X and Y have different distributions, there exists $(a, b]$ such that $\mathbb{P}(X \in (a, b]) \neq \mathbb{P}(Y \in (a, b])$. Hence, we can find an increasing function g such that $g(X)$ is m-atomic and $g(Y)$ is n-atomic for some $m \geqslant 2$ and $n > 2$ where $g(X)$ and $g(Y)$ have different distributions. Hence, without loss of generality, we can assume X is m-atomic and Y is n-atomic for some $m \geqslant 2$ and $n > 2$ with different distributions.

If $\mathrm{Supp}(X) \neq \mathrm{Supp}(Y)$, we have $r = 0$ by Lemma 10.9. Assume $\mathrm{Supp}(X) = \mathrm{Supp}(Y)$. As X, Y have different distributions, we can always find $a \in \mathrm{Supp}(X)$ such that $\mathbb{P}(X > a) \neq \mathbb{P}(Y > a)$. For such a, let $g(x) = \mathbb{1}_{\{x>a\}} + \mathbb{1}_{\{x>b\}} + \mathbb{1}_{\{x>c\}}$ for some $b, c \in \mathrm{Supp}(X)$ with $a < b < c$. Thus, $g(X)$ and $g(Y)$ are tri-atomic random variables with the same support but different distribution. By Lemma 10.8, we have $r = 0$.

Therefore, we have $(X, Y) \in \mathrm{IC}_0^{\uparrow}$. Using Item (i) above, we have $(X, Y) \in \mathrm{IC}_0$. □

10.8.8 Proofs in Section 10.7

Proof of Proposition 10.10. Since the statement on κ_g-matrices is obvious, we will show that $\mathbf{X}$ has the tail-dependence matrix R. Without loss of generality, we can assume that the identical marginal of $\mathbf{X}$ is the standard uniform distribution. Fix $i, j \in [d]$. Since $\lambda_{ij} = \lambda_{ji}$,

Proposition 10.7 leads to

$$\lambda_{ij} = \frac{\lambda_{ij} + \lambda_{ji}}{2} = \lim_{u \downarrow 0} \frac{C_{ij}(u,u) + C_{ji}(u,u)}{2u} = \lim_{u \downarrow 0} \frac{r_{ij} M(u,u) + (1 - r_{ij}) \Pi(u,u)}{u} = r_{ij}. \quad \square$$

Proof of Proposition 10.11. Since $S_{n+1}, \ldots, S_{n+d}$ are independent conditional on the null training sample $\mathcal{D} = \{S_i : i \in [n]\}$, we have

$$\begin{aligned}
\mathbb{P}\left(P_i = \frac{j_i}{n+1}, i \in \mathcal{N}\right) &= \mathbb{E}_{\mathcal{D}}\left[\mathbb{P}\left(P_i = \frac{j_i}{n+1}, i \in \mathcal{N} \,\middle|\, \mathcal{D}\right)\right] \\
&= \mathbb{E}_{\mathcal{D}}\left[\prod_{i \in \mathcal{N}} \mathbb{P}\left(P_i = \frac{j_i}{n+1} \,\middle|\, \mathcal{D}\right)\right] = \mathbb{E}_{\mathcal{D}}\left[\prod_{i \in \mathcal{N}} \left(S_{(j_i)} - S_{(j_i - 1)}\right)\right]
\end{aligned}$$

for $j_i \in [n+1]$, where $0 = S_{(0)} < S_{(1)} < \cdots < S_{(n)} < S_{(n+1)} = 1$ are the order statistics of the null training sample $(S_1 \ldots, S_n)$. Since the conformal p-values are independent of the distribution of scores, we can assume without loss of generality that $S_1, \ldots, S_n$ are independently distributed of the standard uniform distribution. Let $T_j = S_{(j)} - S_{(j-1)}$, $j \in [n+1]$. Then $(T_1, \ldots, T_{n+1})$ follows the Dirichlet distribution with the parameter vector $\mathbf{1}_{n+1} \in \mathbb{R}^{n+1}$. Therefore,

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}\left[\prod_{i \in \mathcal{N}} \left(S_{(j_i)} - S_{(j_i - 1)}\right)\right] &= \mathbb{E}_{\mathcal{D}}\left[T_1^{N_1} \cdots T_{n+1}^{N_{n+1}}\right] \\
&= \frac{B(1 + N_1, \ldots, 1 + N_{n+1})}{B(\mathbf{1}_{n+1})} = \frac{N_1! \cdots N_{n+1}!}{(n+m)(n+m-1)\cdots(n+1)},
\end{aligned}$$

where B is the $(n+1)$-dimensional Beta function.

In particular, when $m = 1$, we have $\mathbb{P}(P_i = j/(n+1)) = 1/(n+1)$, $j \in [n+1]$. When $m = 2$, we have

$$\mathbb{P}\left(P_i = \frac{j}{n+1}, P_{i'} = \frac{j'}{n+1}\right) = \frac{1}{n+2} \cdot \frac{1}{n+1} \mathbb{1}_{\{j = j'\}} + \left(1 - \frac{1}{n+2}\right) \cdot \frac{1}{n+1} \cdot \frac{1}{n+1},$$

which is the exchangeable case of the model (10.14) and thus has the invariant correlation $1/(n+2)$. $\square$

Chapter 11

The checkerboard copula and dependence concepts

11.1 Introduction

The copula theory has been actively studied over the past few decades with many applications in statistics, finance, engineering, and the natural sciences; for an introduction, see the monographs of Nelsen (2006) and Joe (2014).

It is well known through Sklar's theorem (Nelsen (2006, Theorem 2.10.9)) that the copula of a random vector is unique if and only if it has continuous marginal distributions. Genest and Nešlehová (2007) discussed difficulties in identifying copulas for discrete distributions. The purpose of this chapter is to understand whether it is possible to identify a canonical copula for a random vector in some sense if it does not have continuous marginal distributions.

To answer this question, we seek inspiration from three applications. Let $\mathbf{X} = (X_1, \dots, X_d)$ be a d-dimensional random vector with $d \geqslant 2$, which may have non-unique copulas. Denote by $\mathcal{C}_{\mathbf{X}}$ the set of all copulas of $\mathbf{X}$. For a random variable X, its probability integral transform U is a uniform random variable on $[0, 1]$ satisfying $F^{-1}(U) = X$ almost surely (a.s.), where F is the distribution function of X and F^{-1} is the quantile function of X. Let $(U_1, \dots, U_d)$ be any vector of probability integral transforms of $X_1, \dots, X_d$ with a joint distribution C;

certainly, C is a copula of $\mathbf{X}$. All random variables live in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1. **Simulating from the copula of X.** One of the most popular applications of copulas in finance is to model default correlation, as famously done by Li (2000); see McNeil et al. (2015) for discussions. In such applications, one needs to simulate from the copula of $\mathbf{X}$, where $\mathbf{X}$ may have non-continuous marginal distributions (e.g., losses from default events). Assume that we can simulate $\mathbf{X}$, and we also have knowledge of all marginal distributions of $\mathbf{X}$. How can we find a reasonable copula $C \in \mathcal{C}_{\mathbf{X}}$ to simulate from, that is determined only by $\mathbf{X}$ but not by any particular modeling choices (such as the Gaussian copula)?

2. **Stressing the distribution of X.** In sensitivity analysis and risk management, it is often necessary to stress, or distort, the distribution of $\mathbf{X}$ to obtain post-stress distributions. In the stressing mechanisms studied by Millossovich et al. (2021), one needs to find a stressed probability measure Q_1 by using $\mathrm{d}Q_1/\,\mathrm{d}\mathbb{P} = g(U_1)$ for a non-negative increasing function g with $\int_0^1 g(u)\,\mathrm{d}u = 1$, such as $g(u) = 2u$. The simple interpretation of Q_1 is to gradually increase the weight of realizations $\omega \in \Omega$ at which X_1 is large. Similarly, one can simultaneously stress all components of $\mathbf{X}$ by considering a measure Q such that $\mathrm{d}Q/\,\mathrm{d}\mathbb{P} = (1/d)\sum_{i=1}^d g_i(U_i)$ or $\mathrm{d}Q/\,\mathrm{d}\mathbb{P} = c\prod_{i=1}^d g_i(U_i)$ with a normalizing constant $c > 0$ ($c = 1$ if $U_1, \dots, U_d$ are independent), where g_i are non-negative increasing functions with $\int_0^1 g_i(u)\,\mathrm{d}u = 1$. If we are only interested in the post-stress distribution $\hat{F}_1^{Q_1}$ of X_1 under Q_1, the choice of the copula $C \in \mathcal{C}_{\mathbf{X}}$ is irrelevant. However, the choice of the copula $C \in \mathcal{C}_{\mathbf{X}}$ matters for the distribution $\hat{F}_i^{Q}$ of X_i under Q, as well as for the distribution $\hat{F}_i^{Q_1}$ of X_i under Q_1.

3. **Computing a co-risk measure.** Co-risk measures (e.g., Adrian and Brunnermeier (2016)) are calculated for the conditional distribution of a random variable X_2 given some event related to X_1. A classic example is the Marginal Expected Shortfall at level $p \in (0,1)$, which is defined as, assuming that X_1 is continuously distributed,

$$\rho(X_2|X_1) := \mathbb{E}[X_2|X_1 > F_1^{-1}(p)] = \mathbb{E}[X_2|U_1 > p].$$

Generally, ρ is the mean of X_2 given a (not necessarily unique) p-tail event of X_1 in the sense of Wang and Zitikis (2021). This risk measure ρ does not depend on the choice of $C \in \mathcal{C}_{\mathbf{X}}$ if X_1 is continuously distributed (p-tail event is unique a.s.); however, it may depend on $C \in \mathcal{C}_{\mathbf{X}}$ if X_1 has some points of mass. Other co-risk measures, such as CoVaR (Adrian and Brunnermeier, 2016), also face the same issue.

All of the above contexts point to the question of choosing a good copula $C \in \mathcal{C}_{\mathbf{X}}$, which we address in this chapter. We first offer a new characterization of all copulas of a given random vector in Section 11.2 in Theorem 11.1. In Section 11.3, we give some intuitive and heuristic arguments for the questions above, leading to the proposal of using the checkerboard copula, that is, the unique copula of $\mathbf{X}$ that is as uniform as possible in regions where the copulas of $\mathbf{X}$ are not uniquely determined, formally defined in Definition 11.1. Although the arguments in Section 11.3 are heuristic, the use of the checkerboard copula indeed has a theoretical justification, which we present in Section 11.4. The checkerboard copula has the maximum Shannon entropy among all possible copulas of $\mathbf{X}$, as shown in Theorem 11.2. In Section 11.5, we show in Theorem 11.3 that the checkerboard copula preserves various dependence concepts that are satisfied by $\mathbf{X}$. This result is intuitive, but the proof requires serious technical analysis. We discuss two applications of our results in diversification penalty and induced order statistics in Section 11.6. Section 11.7 concludes this chapter.

11.2 Copulas for a discrete random vector

Let $d \geqslant 2$ be an integer and $[d] = \{1, \dots, d\}$. All inequalities are interpreted componentwise when applied to vectors. All random variables live in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbf{X} = (X_1, \dots, X_d)$ be a d-dimensional random vector, $F_1, \dots, F_d$ be the marginal distributions of $\mathbf{X}$, and $\mathrm{Ran}(F_i)$ be the range of F_i for $i \in [d]$. By Sklar's theorem, the copula of $\mathbf{X}$ is uniquely determined on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_d)$ but undetermined in other regions. Therefore, when the marginal distribution F_i is not continuous for some $i \in [d]$, the copula of $\mathbf{X}$ may not be unique. In this section, we give a concrete representation for any copulas of $\mathbf{X}$.

We start with the observation that, if a random variable X is continuously distributed, the random variable

$$U_X := F_X(X)$$

will be uniformly distributed over $[0,1]$, where F_X is the cumulative distribution function of X. More generally, regardless of whether X is continuously distributed, we can define its probability integral transform

$$U_X := F_X(X-) + V_X(F_X(X) - F_X(X-)), \tag{11.1}$$

where $F_X(x-) = \lim_{y \uparrow x} F_X(x) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$ and $V_X \sim \mathrm{U}[0,1]$ is independent of X, assumed to exist.[1] The probability integral transform U_X satisfies $U_X \sim \mathrm{U}[0,1]$ and $F_X^{-1}(U_X) = X$ a.s. (see e.g., Rüschendorf (2013, Proposition 1.3)). Therefore, the probability integral transform (11.1) converts any random variable X to a $\mathrm{U}[0,1]$ distributed random variable U_X using V_X.

We extend this idea to the case of a random vector $\mathbf{X}$. Let $\mathbf{V} = (V_1, \dots, V_d)$ be a random vector with $\mathrm{U}[0,1]$ marginals such that V_i is independent of X_i for each $i \in [d]$. Denote the set of such $\mathbf{V}$ by $\mathcal{V}_{\mathbf{X}}$. Similar to (11.1), let us define the probability integral transform for $\mathbf{X} = (X_1, \dots, X_d)$:

$$U_i := F_i(X_i-) + V_i(F_i(X_i) - F_i(X_i-)), \quad i \in [d]. \tag{11.2}$$

It immediately follows that $U_i \sim \mathrm{U}[0,1]$ and $F_i^{-1}(U_i) = X_i$ a.s.. Therefore, $\mathbf{U} = (U_1, \dots, U_d)$ is a random vector with uniform marginals. This technique of constructing random vectors with uniform marginals has been used in the literature; see e.g., Moore and Spruill (1975) and Nešlehová (2007).

Let $C_{\mathbf{X}}^{\mathbf{V}}$ be the copula of $\mathbf{U}$. Because $F_i^{-1}(U_i) = X_i$ a.s. for each $i \in [d]$, we have

$$C_{\mathbf{X}}^{\mathbf{V}}(F_1(x_1), \dots, F_d(x_d)) = \mathbb{P}(U_1 \leqslant F_1(x_1), \dots, U_d \leqslant F_d(x_d)) = \mathbb{P}(X_1 \leqslant x_1, \dots, X_d \leqslant x_d)$$

for any $(x_1, \dots, x_d) \in \mathbb{R}^d$. Hence, $C_{\mathbf{X}}^{\mathbf{V}}$ is a copula of $\mathbf{X}$.

[1] This assumption is safe as we are interested in distributional properties, and we can extend the probability space to include such independent V_X, if necessary.

According to (11.2), the copula $C_{\mathbf{X}}^{\mathbf{V}}$ is determined by the joint distribution of $(\mathbf{X}, \mathbf{V})$. In particular, the copula $C_{\mathbf{X}}^{\mathbf{V}}$ does not depend on the choice of V_i for i such that X_i is continuously distributed because, for these i, U_i in (11.2) is a.s. equal to $F_i(X_i)$. While for i such that X_i is discrete, V_i does have an impact on the copula $C_{\mathbf{X}}^{\mathbf{V}}$.

In general, the choice of $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$ for constructing the copula $C_{\mathbf{X}}^{\mathbf{V}}$ may not be unique. This is because $\mathcal{V}_{\mathbf{X}}$ allows two types of dependence that might be present in the construction of $\mathbf{V}$: First, the components of $\mathbf{V}$ may be mutually dependent. Second, V_i may depend on X_j for $i \neq j$. Naturally, a different choice of $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$ often leads to a different copula $C_{\mathbf{X}}^{\mathbf{V}}$; see the following example.

Example 11.1. Assume that $d = 2$, X_1 is a constant, and X_2 is continuously distributed. It is well known that any copula is a copula of $\mathbf{X}$ in this case. For instance, by choosing V_1 to be independent of X_2, $C_{\mathbf{X}}^{\mathbf{V}}$ is the independence copula, and by choosing $V_1 = F_2(X_2)$, $C_{\mathbf{X}}^{\mathbf{V}}$ is the comonotonic copula.

The following result says that all copulas of X can be realized by some $C_{\mathbf{X}}^{\mathbf{V}}$. Hence, (11.2) gives a stochastic representation for any copula of $\mathbf{X}$. The representation is quite intuitive, but we did not find it in the literature, so we provide a self-contained proof.

Theorem 11.1. *Assume that there exists a continuously distributed random variable independent of a random vector* $\mathbf{X}$. *A copula* C *is a copula of* $\mathbf{X}$ *if and only if* $C = C_{\mathbf{X}}^{\mathbf{V}}$ *for some* $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$.

Proof. We have seen that $C_{\mathbf{X}}^{\mathbf{V}}$ is a copula of $\mathbf{X}$. It suffices to show the "only if" statement. Let C be a copula of $\mathbf{X}$, take $\mathbf{U}' = (U_1', \dots, U_d') \sim C$, and write $\mathbf{X}' = \left(F_1^{-1}(U_1'), \dots, F_d^{-1}(U_d')\right)$. Because C is a copula of $\mathbf{X}$, for $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, we have

$$\begin{aligned}\mathbb{P}(\mathbf{X} \leqslant \mathbf{x}) = C\left(F_1(x_1), \dots, F_d(x_d)\right) &= \mathbb{P}\left(U_1' \leqslant F_1(x_1), \dots, U_d' \leqslant F_d(x_d)\right) \\ &= \mathbb{P}\left(F_1^{-1}(U_1') \leqslant x_1, \dots, F_d^{-1}(U_d') \leqslant x_d\right) = \mathbb{P}(\mathbf{X}' \leqslant \mathbf{x}).\end{aligned}$$

Hence, $\mathbf{X} \overset{\mathrm{d}}{=} \mathbf{X}'$. Take $\mathbf{U}^* = (U_1^*, \dots, U_d^*)$ such that $(\mathbf{X}, \mathbf{U}^*) \overset{\mathrm{d}}{=} (\mathbf{X}', \mathbf{U}')$, and we then have $\mathbf{X} = \left(F_1^{-1}(U_1^*), \dots, F_d^{-1}(U_d^*)\right)$ a.s.. Furthermore, take $V' \sim \mathrm{U}[0,1]$ which is independent of $(\mathbf{X}, \mathbf{U}^*)$. The existence of $\mathbf{U}^*$ and V' is guaranteed by the assumption of the existence of a

continuously distributed random variable independent of $\mathbf{X}$. For $i \in [d]$, let $\mathbf{V} = (V_1, \dots, V_d)$ be given by

$$V_i = \frac{U_i^* - F_i(X_i-)}{F_i(X_i) - F_i(X_i-)} \mathbb{1}_{\{F_i(X_i) > F_i(X_i-)\}} + V' \mathbb{1}_{\{F_i(X_i) = F_i(X_i-)\}}.$$

Fix $i \in [d]$ below. Let D_i be the set of discontinuity points of F_i. Note that for $x \in D_i$, we have

$$\mathbb{P}\left(U_i^* \in [F_i(x-), F_i(x)] \middle| X_i = x\right) = 1 \quad \text{and} \quad \mathbb{P}\left(U_i^* \in [F_i(x-), F_i(x)] \middle| X_i \neq x\right) = 0.$$

Because U_i^* is uniformly distributed over $[0,1]$, U_i^* is uniform on $[F_i(x-), F_i(x)]$ conditional on $X_i = x \in D_i$. Thus,

$$\mathbb{P}(U_i^* \leqslant u | X_i = x) = \frac{u - F_i(x-)}{F_i(x) - F_i(x-)}, \qquad u \in [F_i^{-1}(x-), F_i^{-1}(x)].$$

Therefore, for $u \in [0,1]$,

$$\begin{aligned}
\mathbb{P}(V_i \leqslant u | X_i) &= \mathbb{P}\left(\frac{U_i^* - F_i(X_i-)}{F_i(X_i) - F_i(X_i-)} \leqslant u \middle| X_i \right) \mathbb{1}_{\{X_i \in D_i\}} + \mathbb{P}(V' \leqslant u) \mathbb{1}_{\{X_i \notin D_i\}} \\
&= \mathbb{P}\left(U_i^* \leqslant u(F_i(X_i) - F_i(X_i-)) + F_i(X_i-) \middle| X_i\right) \mathbb{1}_{\{X_i \in D_i\}} + u \mathbb{1}_{\{X_i \notin D_i\}} \\
&= u \mathbb{1}_{\{X_i \in D_i\}} + u \mathbb{1}_{\{X_i \notin D_i\}} = u.
\end{aligned}$$

Hence, V_i follows $\mathrm{U}[0,1]$ and is independent of X_i. Note that, by the construction, U_i^*, V_i, and X_i satisfy $U_i^* = F_i(X_i-) + V_i(F_i(X_i) - F_i(X_i-))$ a.s., and hence $\mathbf{U}^* \sim C_{\mathbf{X}}^{\mathbf{V}}$. This shows $C = C_{\mathbf{X}}^{\mathbf{V}}$. □

Theorem 11.1 implies $\mathcal{C}_{\mathbf{X}} = \{C_{\mathbf{X}}^{\mathbf{V}} : \mathbf{V} \in \mathcal{V}_{\mathbf{X}}\}$. Note that $\mathcal{C}_{\mathbf{X}}$ is a singleton if and only if all marginal distributions of $\mathbf{X}$, $F_1, \dots, F_d$, are continuous functions.

11.3 Motivating arguments for the checkerboard copula

Theorem 11.1 gives the entire class of copulas for $\mathbf{X}$. We now consider which $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$ can answer the three motivating questions in Section 11.1, which all point to the same unique choice of $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$.

1. **Simulating from the copula of X.** A natural approach to simulating from the copula of $\mathbf{X}$ with some atoms in the marginal distributions is by first simulating a pair of $(\mathbf{X}, \mathbf{V})$, and then applying the probability integral transform using (11.2). Theorem 11.1 shows that all copulas of $\mathbf{X}$ can be simulated this way. For this purpose, the simplest and most natural choice of $\mathbf{V}$ is $\mathbf{V} \sim \mathrm{U}\left([0,1]^d\right)$ which is independent of $\mathbf{X}$. In fact, we could not think of an argument against the use of this particular $\mathbf{V}$ in the context of simulation.

2. **Stressing the distribution of X.** To understand how the choice of $\mathbf{V}$ affects the stressed distribution of X_2, we look at the simple example in Example 11.1 with $g(u) = 2u$. Choosing V_1 independent of X_2 would lead to $\hat{F}_2^{Q_1} = F_2$, whereas choosing $V_1 = F_2(X_2)$ would lead to $\hat{F}_2^{Q_1} = (F_2)^2$. Because we are interested in the effect of stressing X_1 on X_2, and X_1 is a constant in this example, it is natural to choose a V_1 that affects the distribution of X_2 minimally, which is achieved when V_1 is independent of X_2. Translating this argument into the general d-dimensional setting suggests choosing $\mathbf{V} \sim \mathrm{U}\left([0,1]^d\right)$ independent of $\mathbf{X}$.

3. **Computing a co-risk measure.** To understand how the choice of $\mathbf{V}$ affects the value of the co-risk measure, we again look at Example 11.1. We have $\rho(X_2|X_1) = \mathbb{E}[X_2]$ if V_1 is independent of X_2, and $\rho(X_2|X_1) = \mathrm{ES}_p(X_2)$ if $V_1 = F_2(X_2)$, where $\mathrm{ES}_p(X_2) = \mathbb{E}[X_2|U_2 > p]$ is the Expected Shortfall of X_2 at level p. The interpretation of ρ as the mean of X_2 on a tail event of X_1 suggests that it is natural to choose V_1 independent of X_2, because X_1 is a constant and its tail event should not affect X_2.

In all the considerations above, $\mathbf{V} \sim \mathrm{U}\left([0,1]^d\right)$ independent of $\mathbf{X}$ appears to be a good choice. Let us denote this by $\mathbf{V}_{\mathbf{X}}^{\perp}$ and the corresponding copula by $C_{\mathbf{X}}^{\perp}$, where $\perp$ reflects that independence is used twice to construct $\mathbf{V}$ (within components of $\mathbf{V}$ and between $\mathbf{V}$ and $\mathbf{X}$). From the three motivating examples above, the choice of the particular copula $C_{\mathbf{X}}^{\perp}$ is natural and has several unique features. This choice has been known as the checkerboard copula.

Definition 11.1. The copula $C_{\mathbf{X}}^{\perp}$ is called the *checkerboard copula* of $\mathbf{X}$.

The copula $C_{\mathbf{X}}^{\perp}$ is also called the multilinear extension copula of $\mathbf{X}$; see Genest et al.

(2017) for its properties, its empirical process, and a history. One notable property is that $X_1, \ldots, X_d$ are independent if and only if $C^{\perp}_{\mathbf{X}}$ is the independence copula.

The rest of this chapter focuses on the properties and applications of the checkerboard copula.

11.4 Entropy maximization

Given the natural choice of $C^{\perp}_{\mathbf{X}}$ in the applications in Section 11.3, it should have some unique properties within the class $\mathcal{C}_{\mathbf{X}}$. The applications seem to suggest that $C^{\perp}_{\mathbf{X}}$ relies less on external information compared to other choices of $\mathbf{V}$. Such consideration is typically studied via entropy. Indeed, as argued by Jaynes (1957), the maximum-entropy distribution should be the only unbiased choice given available information. If a copula C has a density function c, then its Shannon (differential) entropy is defined as

$$H(C) = -\int_{[0,1]^d} c(\mathbf{u}) \log c(\mathbf{u}) \, \mathrm{d}\mathbf{u}.$$

One problem with the above formulation is that a copula C often does not have a density. We set $H(C) = -\infty$ if C does not have a density, which is intuitive and can be seen as a limiting case. However, even the checkerboard copula $C^{\perp}_{\mathbf{X}}$ may not have a density if the distribution of $\mathbf{X}$ has some singular continuous part. This issue may be solved by considering other measures of information, but for now, let us stick to the Shannon entropy, which is the most popular notion in information theory. We would like to compare $H(C^{\perp}_{\mathbf{X}})$ with $H(C)$ for $C \in \mathcal{C}_{\mathbf{X}}$, or equivalently, $H(C^{\mathbf{V}}_{\mathbf{X}})$ for other choices of $\mathbf{V} \in \mathcal{V}_{\mathbf{X}}$. The main result of this section is to show that $H(C^{\perp}_{\mathbf{X}})$ has the largest entropy among all other choices.

Theorem 11.2. *For $C \in \mathcal{C}_{\mathbf{X}}$, we have $H(C^{\perp}_{\mathbf{X}}) \geqslant H(C)$.*

The proof of Theorem 11.2 essentially boils down to showing the following lemma, which states that the density of the checkerboard copula can be expressed as the conditional expectation for the density of other possible copulas in $\mathcal{C}_{\mathbf{X}}$. From this lemma and Jensen's inequality, Theorem 11.2 follows.

Lemma 11.1. *For $C \in \mathcal{C}_{\mathbf{X}}$, if the density c of C exists, then the density $c^{\perp}$ of $C^{\perp}_{\hat{\mathbf{X}}}$ exists. Moreover, we have $c^{\perp}(\mathbf{U}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$, where $\mathbf{U} = (U_1, \ldots, U_d) \sim \mathrm{U}\left([0,1]^d\right)$, $\hat{\mathbf{X}} = \left(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)\right)$, and $F_1, \ldots, F_d$ are the marginals of $\mathbf{X}$.*

Proof. Since $\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$ is $\sigma(\hat{\mathbf{X}})$-measurable, there exists a function $f : \mathbb{R}^d \to [0,1]$ such that $f(\hat{\mathbf{X}}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$ (in the almost sure sense). Let $c^{\perp}$ be a function $[0,1]^d \to [0,1]$ defined as $c^{\perp}(\mathbf{u}) = f\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right)$ for any $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$. We claim that $c^{\perp}$ is the density of $C^{\perp}_{\hat{\mathbf{X}}}$. This claim implies that $c^{\perp}$ exists and $c^{\perp}(\mathbf{U}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$.

To prove this claim, let $\mathbf{U}^c \sim C$, $\mathbf{U}^{\perp} \sim C^{\perp}_{\hat{\mathbf{X}}}$, and $R = \prod_{i=1}^d \mathrm{Ran}(F_i)$. We first show that $\int_A c^{\perp}(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbb{P}(\mathbf{U}^{\perp} \in A)$ for the following two types of the set A.

(i) Let $A = \prod_{i=1}^d [0, a_i]$ with $\mathbf{a} = (a_1, \ldots, a_d) \in R$. We have $\mathbb{1}_{\{\mathbf{U} \leqslant \mathbf{a}\}} = \mathbb{1}_{\left\{\hat{\mathbf{X}} \leqslant \left(F_1^{-1}(a_1), \ldots, F_d^{-1}(a_d)\right)\right\}}$. Therefore,

$$\begin{aligned}
\int_A c^{\perp}(\mathbf{u})\,\mathrm{d}\mathbf{u} &= \int_{\prod_{i=1}^d [0,a_i]} f\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right)\,\mathrm{d}u_1 \cdots \mathrm{d}u_d \\
&= \mathbb{E}\left[f(\hat{\mathbf{X}})\mathbb{1}_{\{\mathbf{U} \leqslant \mathbf{a}\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]\mathbb{1}_{\left\{\hat{\mathbf{X}} \leqslant \left(F_1^{-1}(a_1), \ldots, F_d^{-1}(a_d)\right)\right\}}\right] \\
&= \mathbb{E}\left[c(\mathbf{U})\mathbb{1}_{\{\mathbf{U} \leqslant \mathbf{a}\}}\right] = \mathbb{P}(\mathbf{U}^c \leqslant \mathbf{a}) = \mathbb{P}(\mathbf{U}^{\perp} \leqslant \mathbf{a}),
\end{aligned}$$

where the last equality holds because

$$\mathbb{P}\left(\mathbf{U}^c \leqslant \mathbf{a}\right) = \mathbb{P}\left(\mathbf{X} \leqslant \left(F_1^{-1}(a_1), \ldots, F_d^{-1}(a_d)\right)\right) = \mathbb{P}(\mathbf{U}^{\perp} \leqslant \mathbf{a}).$$

This further implies that $\int_A c^{\perp}(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbb{P}(\mathbf{U}^{\perp} \in A)$ for any $A = \prod_{i=1}^d A_i$ such that $A_i \in \{[0, a_i] : a_i \in \mathrm{Ran}(F_i)\} \cup \{(F_i(x_i-), F_i(x_i)] : x_i \text{ is a discontinuity point of } F_i\}$ for $i \in [d]$.

(ii) Let $A = \prod_{i=1}^k [0, a_i] \times \prod_{j=k+1}^d (s_j, t_j]$ with $k \in \{0, 1, \ldots, d\}$ such that $a_i \in \mathrm{Ran}(F_i)$ for $i \in [k]$ and $(s_j, t_j] \cap \mathrm{Ran}(F_j) = \varnothing$ for $j \in [d] \setminus [k]$. For $j \in [d] \setminus [k]$, denote by $x_j = F_j^{-1}(s_j)$, and thus $(s_j, t_j] \subseteq (F_j(x_j-), F_j(x_j))$. By the definition of $c^{\perp}$, for fixed $u_i \in [0, a_i]$ and $i \in [k]$, $c^{\perp}(u_1, \ldots, u_k, v_{k+1}, \ldots, v_d)$ is a constant for all $(v_{k+1}, \ldots, v_d) \in$

$\prod_{j=k+1}^{d}(F_j(x_j-), F_j(x_j))$. Therefore, let $B = \prod_{i=1}^{k}[0, a_i] \times \prod_{j=k+1}^{d}(F_j(x_j-), F_j(x_j))$, we have

$$\int_A c^\perp(\mathbf{u})\,\mathrm{d}\mathbf{u} = \left(\prod_{j=k+1}^{d} \frac{t_j - s_j}{F_j(x_j) - F_j(x_j-)}\right) \int_B c^\perp(\mathbf{u})\,\mathrm{d}\mathbf{u}.$$

Let $\mathbf{V} = (V_1, \ldots, V_d) \sim \mathrm{U}\left([0,1]^d\right)$ be independent of $\mathbf{X}$, and for $j \in [d] \setminus [k]$, denote by $s'_j = (s_j - F_j(x_j-))/(F_j(x_j) - F_j(x_j-))$ and $t'_j = (t_j - F_j(x_j-))/(F_j(x_j) - F_j(x_j-))$. Hence,

$$\prod_{j=k+1}^{d} \frac{t_j - s_j}{F_j(x_j) - F_j(x_j-)} = \mathbb{P}\left(V_j \in (s'_j, t'_j] \text{ for all } j \in [d] \setminus [k]\right).$$

In addition, by (i), we can get

$$\int_B c^\perp(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \in B) = \mathbb{P}\left(X_i \leqslant F^{-1}(a_i),\ X_j = x_j \text{ for all } i \in [k],\ j \in [d] \setminus [k]\right).$$

Therefore,

$$\begin{aligned}
\int_A c^\perp(\mathbf{u})\,\mathrm{d}\mathbf{u} &= \mathbb{P}\left(\bigcap_{j \in [d]\setminus[k]} \{V_j \in (s'_j, t'_j]\}\right) \mathbb{P}\left(\bigcap_{i\in[k], j\in[d]\setminus[k]} \{X_i \leqslant F^{-1}(a_i),\ X_j = x_j\}\right) \\
&= \mathbb{P}\left(\bigcap_{i\in[k], j\in[d]\setminus[k]} \left\{X_i \leqslant F^{-1}(a_i),\ X_j = x_j,\ V_i \in [0,1],\ V_j \in (s'_j, t'_j]\right\}\right) \\
&= \mathbb{P}\left(\mathbf{U}^\perp \in \prod_{i=1}^{k}[0, a_i] \times \prod_{j=k+1}^{d}(s_j, t_j]\right) = \mathbb{P}(\mathbf{U}^\perp \in A).
\end{aligned}$$

By the same argument, we have $\int_A c^\perp(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \in A)$ for any $A = \prod_{i=1}^{d} A_i$ such that $A_i \in \{[0, a_i] : a_i \in \mathrm{Ran}(F_i)\} \cup \{(s_i, t_i] : (s_i, t_i] \cap \mathrm{Ran}(F_i) = \varnothing\}$ for $i \in [d]$.

For any $\mathbf{a} = (a_1, \ldots, a_d) \in [0,1]^d$, the region $\prod_{i=1}^{d}[0, a_i]$ can always be represented by an at most countable disjoint union of regions studied in (i) and (ii). Hence, we can obtain

$$\int_{\prod_{i=1}^{d}[0,a_i]} c^\perp(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbb{P}(\mathbf{U}^\perp \leqslant \mathbf{a}).$$

This proves our claim that $c^\perp$ is the density of $C^\perp_{\mathbf{X}}$. □

Proof of Theorem 11.2. If $H(C) = -\infty$, there is nothing to show. Hence, it suffices to consider the case that C has a density, which we denote by c. By Lemma 11.1, we have $c^\perp(\mathbf{U}) = \mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}]$ where $c^\perp$ is the density of $C^\perp_{\mathbf{X}}$, $\mathbf{U} = (U_1, \ldots, U_d) \sim \mathrm{U}\left([0,1]^d\right)$, and

$\hat{\mathbf{X}} = \big(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)\big)$ with $F_1, \ldots, F_d$ as the marginals of $\mathbf{X}$. Define a function $g(x) = x \log x$ for $x \in (0, \infty)$. It is clear that g is convex. By the fact that $\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}] = c^{\perp}(\mathbf{U})$ and Jensen's inequality, we have

$$H(C^{\perp}_{\mathbf{X}}) = -\mathbb{E}[g(c^{\perp}(\mathbf{U}))] = -\mathbb{E}[g(\mathbb{E}[c(\mathbf{U})|\hat{\mathbf{X}}])] \geqslant -\mathbb{E}[\mathbb{E}[g(c(\mathbf{U}))|\hat{\mathbf{X}}]] = -\mathbb{E}[g(c(\mathbf{U}))] = H(C).$$

Thus, $H(C^{\perp}_{\mathbf{X}}) \geqslant H(C)$ for all $C \in \mathcal{C}_{\mathbf{X}}$. □

11.5 Checkerboard copula and dependence concepts

In this section, we study how the checkerboard copula preserves dependence concepts. This question is motivated by a problem raised in the context of diversification in Chen et al. (2024b), which we describe in Section 11.6.1.

11.5.1 Dependence concepts

We first define several notions of positive dependence, introduced and studied by Lehmann (1966), Esary et al. (1967), and Benjamini and Yekutieli (2001), and the corresponding notions of negative dependence, introduced and studied by Lehmann (1966), Alam and Saxena (1981), Block et al. (1982, 1985), Joag-Dev and Proschan (1983), and Chen et al. (2024a).

In what follows, for $i \in [d]$ and an d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$, write $\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$, and for $A, B \subseteq [d]$, write $\mathbf{X}_A = (X_k)_{k \in A}$ and $\mathbf{X}_B = (X_k)_{k \in B}$. A set $S \subseteq \mathbb{R}^d$ is *decreasing* if $\mathbf{x} \in S$ implies $\mathbf{y} \in S$ for all $\mathbf{y} \leqslant \mathbf{x}$.

Definition 11.2. A random vector $\mathbf{X}$ is

(i) (a) *positively associated* (PA) if for every pair of subsets A, B of $[d]$ and any functions f and g both increasing or decreasing coordinatewise, provided the covariance below exists,

$$\operatorname{cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \geqslant 0;$$

(b) *negatively associated* (NA) if for every pair of disjoint subsets A, B of $[d]$ and any functions f and g both increasing or decreasing coordinatewise, provided the covariance below exists,

$$\operatorname{cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \leqslant 0;$$

(ii) (a) *positively regression dependent* (PRD) if for every $i \in [d]$, the random variable $\mathbb{E}[g(\mathbf{X}_{-i})|X_i]$ is an increasing function of X_i for any coordinatewise increasing function g such that the conditional expectation exists;

(b) *negatively regression dependent* (NRD) if for every $i \in [d]$, the random variable $\mathbb{E}[g(\mathbf{X}_{-i})|X_i]$ is a decreasing function of X_i for any coordinatewise increasing function g such that the conditional expectation exists;

(iii) (a) *weakly positively associated* (WPA) if for any $i \in [d]$, decreasing set $S \subseteq \mathbb{R}^{d-1}$, and $x \in \mathbb{R}$ with $\mathbb{P}(X_i \leqslant x) > 0$,

$$\mathbb{P}(\mathbf{X}_{-i} \in S \mid X_i \leqslant x) \geqslant \mathbb{P}(\mathbf{X}_{-i} \in S);$$

(b) *weakly negatively associated* (WNA) if for any $i \in [d]$, decreasing set $S \subseteq \mathbb{R}^{d-1}$, and $x \in \mathbb{R}$ with $\mathbb{P}(X_i \leqslant x) > 0$,

$$\mathbb{P}(\mathbf{X}_{-i} \in S \mid X_i \leqslant x) \leqslant \mathbb{P}(\mathbf{X}_{-i} \in S);$$

(iv) (a) *positively orthant dependent* (POD) if for all $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\mathbb{P}(\mathbf{X} \leqslant \mathbf{x}) \geqslant \prod_{i=1}^d \mathbb{P}(X_i \leqslant x_i)$ and $\mathbb{P}(\mathbf{X} > \mathbf{x}) \geqslant \prod_{i=1}^d \mathbb{P}(X_i > x_i)$;

(b) *negatively orthant dependent* (NOD) if for all $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\mathbb{P}(\mathbf{X} \leqslant \mathbf{x}) \leqslant \prod_{i=1}^d \mathbb{P}(X_i \leqslant x_i)$ and $\mathbb{P}(\mathbf{X} > \mathbf{x}) \leqslant \prod_{i=1}^d \mathbb{P}(X_i > x_i)$.

Moreover, we say that a distribution or a copula is PA, PRD, WPA, POD, NA, NRD, WNA, or NOD if the corresponding random vector is.

Note that the definition of PA does not require A and B to be disjoint, whereas the definition of NA requires this.

The relationship between the above notions is summarized below (see e.g., Chen et al. (2024a)).

$$\text{PA} \Longrightarrow \text{WPA}; \quad \text{PRD} \Longrightarrow \text{WPA}; \quad \text{WPA} \Longrightarrow \text{POD};$$

$$\text{NA} \Longrightarrow \text{WNA}; \quad \text{NRD} \Longrightarrow \text{WNA}; \quad \text{WNA} \Longrightarrow \text{NOD}.$$

Within the class of multivariate normal distributions, the four concepts of positive dependence are equivalent, and each is equivalent to having nonnegative bivariate correlation coefficients; similarly, the four concepts of negative dependence are equivalent, and each is equivalent to having nonpositive bivariate correlation coefficients.

In the sequel, we use $\mathfrak{D}$ to represent one of the following: PA, PRD, WPA, POD, NA, NRD, WNA, or NOD. Our question is whether these properties are properties purely based on copulas. It turns out that the checkerboard copula can help answer this question.

11.5.2 The checkerboard copula preserves dependence

We first present a self-consistency property of those negative dependence concepts in the spirit of Joag-Dev and Proschan (1983, Property P_6) for NA.

Lemma 11.2. *If $f_1, \ldots, f_d$ are increasing functions and $\mathbf{X}$ satisfies $\mathfrak{D}$, $(f_1(X_1), \ldots, f_d(X_d))$ also satisfies $\mathfrak{D}$.*

Proof. We only show the result for the concepts of negative dependence, as the case of positive dependence is similar.

The self-consistency properties of NA and NOD are shown in Joag-Dev and Proschan (1983, Property P_6) and Lehmann (1966, Lemma 1), respectively. We will show the properties for NRD and WNA. Let $\mathbf{Y} = (f_1(X_1), \ldots, f_d(X_d))$.

1. Assume $\mathbf{X}$ is NRD. Fix $i \in [d]$. Let g be a coordinatewise increasing function and $g' = g \circ (f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_d)$. As a result, we have g' is a coordinatewise increasing function and $g(\mathbf{Y}_{-i}) = g'(\mathbf{X}_{-i})$. For any $y \in \mathbb{R}$, let $A_y = \{x : f_i(x) = y\}$. We have $\{Y_i = y\} = \{X_i \in A_y\}$. Therefore, $\mathbb{E}[g(\mathbf{Y}_{-i})|Y_i = y] = \mathbb{E}[g'(\mathbf{X}_{-i})|X_i \in A_y]$. Assume

$y_1 < y_2$. For any $x_1 \in A_{y_1}$ and $x_2 \in A_{y_2}$, we have $x_1 \leqslant x_2$; hence, $\mathbb{E}[g'(\mathbf{X}_{-i})|X_i = x_1] \geqslant \mathbb{E}[g'(\mathbf{X}_{-i})|X_i = x_2]$. Thus,

$$\begin{aligned}\mathbb{E}[g'(\mathbf{X}_{-i})|X_i \in A_{y_1}] &= \mathbb{E}[\mathbb{E}[g'(\mathbf{X}_{-i})|X_i]|X_i \in A_{y_1}] \\ &\geqslant \mathbb{E}[\mathbb{E}[g'(\mathbf{X}_{-i})|X_i]|X_i \in A_{y_2}] = \mathbb{E}[g'(\mathbf{X}_{-i})|X_i \in A_{y_2}],\end{aligned}$$

which implies that $\mathbb{E}[g(\mathbf{Y}_{-i})|Y_i = y_1] \geqslant \mathbb{E}[g(\mathbf{Y}_{-i})|Y_i = y_2]$; hence $\mathbf{Y}$ is NRD.

2. Assume $\mathbf{X}$ is WNA. For $i \in [d]$, let $S \subseteq \mathbb{R}^{d-1}$ be a decreasing set, and

$$S_i^f = \{(x_1, \dots x_{i-1}, x_{i+1}, \dots, x_d) : (f_1(x_1), \dots, f_{i-1}(x_{i-1}), f_{i+1}(x_{i+1}), \dots, f_d(x_d)) \in S\}.$$

It is clear that $\{\mathbf{Y}_{-i} \in S\} = \{\mathbf{X}_{-i} \in S_i^f\}$. For any $\mathbf{x}_1 \leqslant \mathbf{x}_2$ and $\mathbf{x}_2 \in S_i^f$, we have $f_k(x_{1,k}) \leqslant f_k(x_{2,k})$ for all $k \in [d] \setminus \{i\}$. Furthermore, because S is decreasing, we have $\mathbf{x}_1 \in S_i^f$, which implies S_i^f is a decreasing set. For any $y \in \mathbb{R}$ with $\mathbb{P}(Y_i \leqslant y) > 0$, let $x = \sup\{t \in \mathbb{R} : f_i(t) \leqslant y\}$. If $f_i(x) \leqslant y$, we have $\{Y_i \leqslant y\} = \{X_i \leqslant x\}$ and $\mathbb{P}(X_i \leqslant x) > 0$. Therefore,

$$\mathbb{P}(\mathbf{Y}_{-i} \in S|Y_i \leqslant y) = \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f|X_i \leqslant x\right) \leqslant \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f\right) = \mathbb{P}(\mathbf{Y}_{-i} \in S),$$

which implies that $\mathbf{Y}$ is WNA. If $f_i(x) > y$, we have $\{Y_i \leqslant y\} = \{X_i < x\}$ and $\mathbb{P}(X_i < x) > 0$. Therefore,

$$\begin{aligned}\mathbb{P}(\mathbf{Y}_{-i} \in S, Y_i \leqslant y) &= \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f, X_i < x\right) \\ &= \lim_{t \uparrow x} \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f, X_i \leqslant t\right) \\ &\leqslant \lim_{t \uparrow x} \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f\right) \mathbb{P}(X_i \leqslant t) \\ &= \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f\right) \lim_{t \uparrow x} \mathbb{P}(X_i \leqslant t) \\ &= \mathbb{P}\left(\mathbf{X}_{-i} \in S_i^f\right) \mathbb{P}(X_i < x) = \mathbb{P}(\mathbf{Y}_{-i} \in S)\mathbb{P}(Y_i \leqslant y),\end{aligned}$$

which implies that $\mathbb{P}(\mathbf{Y}_{-i} \in S|Y_i \leqslant y) \leqslant \mathbb{P}(\mathbf{Y}_{-i} \in S)$ and $\mathbf{Y}$ is WNA. □

The following theorem demonstrates that the checkerboard copula of $\mathbf{X}$ preserves the dependence information of $\mathbf{X}$.

Theorem 11.3. *A random vector* $\mathbf{X}$ *satisfies* $\mathfrak{D}$ *if and only if it has a copula that satisfies* $\mathfrak{D}$*. Moreover, the copula can be chosen as the checkerboard copula* $C^{\perp}_{\mathbf{X}}$*.*

Proof. The "if" part follows from Lemma 11.2 because, for $\mathbf{U} = (U_1, \ldots, U_d)$ following the copula of $\mathbf{X}$ that satisfies $\mathfrak{D}$, we have $(X_1, \ldots, X_d) = \left(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)\right)$ and F_i^{-1} is increasing for all $i \in [d]$.

Now we show the "only if" part. Let $\mathbf{U} = (U_1, \ldots, U_d)$ be the random vector given by (11.2) with $\mathbf{V} = (V_1, \ldots, V_d) \sim \mathrm{U}\left([0,1]^d\right)$ independent of $\mathbf{X}$. Hence, we have $\mathbf{U} \sim C^{\perp}_{\mathbf{X}}$ and $C^{\perp}_{\mathbf{X}}$ is a copula of $\mathbf{X}$. Note that, for any $i \in [d]$, given V_i, we have that U_i is an increasing function of X_i. Hence, by Lemma 11.2, $\mathbf{X}$ satisfies $\mathfrak{D}$ implies that $\mathbf{U}|\mathbf{V}$ also satisfies $\mathfrak{D}$.

Assume $\mathbf{X}$ is NA. For any given pair of disjoint subsets A, B of $[d]$ and any given functions f and g both increasing or decreasing coordinatewise, we have

$$\begin{aligned}\operatorname{cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)) &= \mathbb{E}[\operatorname{cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)|\mathbf{V})] + \operatorname{cov}\left(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}]\right) \\ &\leqslant 0 + \operatorname{cov}\left(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}_A], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}_B]\right) = 0,\end{aligned}$$

where the inequality follows from $\mathbf{U}|\mathbf{V}$ is NA, and the last equality follows from the independence between $\mathbf{V}_A$ and $\mathbf{V}_B$. Hence, $\mathbf{U}$ is NA.

Assume $\mathbf{X}$ is NRD. For any fixed i and k, by (11.2), there exist x and v such that $\{U_i = k\} = \{X_i = x, V_i = v\}$. Then, for any coordinatewise increasing function g, by the independence between V_i and $(X_i, \mathbf{U}_{-i})$, we have

$$\mathbb{E}[g(\mathbf{U}_{-i})|U_i = k] = \mathbb{E}[g(\mathbf{U}_{-i})|X_i = x, V_i = v] = \mathbb{E}[g(\mathbf{U}_{-i})|X_i = x].$$

Because $\mathbf{U}_{-i}$ is a function of $\mathbf{X}_{-i}$ and $\mathbf{V}_{-i}$, we can let h be the function such that $g(\mathbf{U}_{-i}) = h(\mathbf{X}_{-i}, \mathbf{V}_{-i})$. Then, due to the independence between $\mathbf{V}_{-i}$ and $\mathbf{X}$,

$$\mathbb{E}[g(\mathbf{U}_{-i})|X_i = x] = \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{V}_{-i})|X_i = x] = \int_{[0,1]^{d-1}} \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x] \mathrm{d}\mathbf{v}_{-i},$$

where $\mathbf{v}_{-i} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_d)$. Therefore, for any $k_1 \leqslant k_2$, there exist x_1 and x_2 such that

$$\begin{aligned}\mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_1] &= \int_{[0,1]^{d-1}} \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_1] \mathrm{d}\mathbf{v}_{-i}, \\ \mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_2] &= \int_{[0,1]^{d-1}} \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_2] \mathrm{d}\mathbf{v}_{-i}.\end{aligned}$$

In addition, by (11.2), we must have $x_1 \leqslant x_2$. Note that given $\mathbf{v}_{-i}$, $h(\mathbf{X}_{-i}, \mathbf{v}_{-i})$ is a coordinatewise increasing function of $\mathbf{X}_{-i}$. Hence, we have $\mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_1] \geqslant \mathbb{E}[h(\mathbf{X}_{-i}, \mathbf{v}_{-i})|X_i = x_2]$ for any $\mathbf{v}_{-i}$. Therefore, $\mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_1] \geqslant \mathbb{E}[g(\mathbf{U}_{-i})|U_i = k_2]$ and $\mathbf{U}$ is NRD.

Assume $\mathbf{X}$ is WNA. For any $i \in [d]$, decreasing set $S \subseteq \mathbb{R}^{d-1}$, and $x \in \mathbb{R}$ with $\mathbb{P}(U_i \leqslant x) > 0$,

$$\begin{aligned}
\mathbb{P}(\mathbf{U}_{-i} \in S, U_i \leqslant x) &= \mathbb{E}[\mathbb{P}(\mathbf{U}_{-i} \in S, U_i \leqslant x \mid \mathbf{V})] \\
&\leqslant \mathbb{E}[\mathbb{P}(\mathbf{U}_{-i} \in S|\mathbf{V}_{-i})\mathbb{P}(U_i \leqslant x \mid V_i)] \\
&= \mathbb{E}[\mathbb{P}(\mathbf{U}_{-i} \in S|\mathbf{V}_{-i})]\mathbb{E}[\mathbb{P}(U_i \leqslant x \mid V_i)] \\
&= \mathbb{P}(\mathbf{U}_{-i} \in S)\mathbb{P}(U_i \leqslant x).
\end{aligned}$$

Hence, $\mathbf{U}$ is WNA.

Assume $\mathbf{X}$ is NOD. For any $t_1, \ldots, t_d \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{P}(U_1 \leqslant t_1, \ldots, U_d \leqslant t_d) &= \mathbb{E}[\mathbb{P}(U_1 \leqslant t_1, \ldots, U_d \leqslant t_d|V_1, \ldots, V_d)] \\
&\leqslant \mathbb{E}[\mathbb{P}(U_1 \leqslant t_1|V_1) \cdots \mathbb{P}(U_d \leqslant t_d|V_d)] \\
&= \mathbb{E}[\mathbb{P}(U_1 \leqslant t_1|V_1)] \cdots \mathbb{E}[\mathbb{P}(U_d \leqslant t_d|V_d)] \\
&= \mathbb{P}(U_1 \leqslant t_1) \cdots \mathbb{P}(U_d \leqslant t_d).
\end{aligned}$$

Similarly, we can show

$$\mathbb{P}(U_1 > t_1, \ldots, U_d > t_d) \leqslant \mathbb{P}(U_1 > t_1) \cdots \mathbb{P}(U_d > t_d).$$

Hence, $\mathbf{U}$ is NOD.

In conclusion, if $\mathbf{X}$ satisfies $\mathfrak{D}$, then $\mathbf{U}$ satisfies $\mathfrak{D}$, where $\mathfrak{D}$ is one of the four concepts of negative dependence.

To show the case of positive dependence, we follow a similar route. We take the same $\mathbf{U}$ as above. Assume $\mathbf{X}$ is PA. Because $U_i|V_i$ is an increasing function of X_i, by Lemma 11.2, $\mathbf{U}|\mathbf{V}$ is also PA. Thus, for any given pair of subsets A, B of $[d]$ and any given functions f and g both coordinatewise increasing or decreasing, we have

$$\begin{aligned}
\operatorname{cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)) &= \mathbb{E}[\operatorname{cov}(f(\mathbf{U}_A), g(\mathbf{U}_B)|\mathbf{V})] + \operatorname{cov}\left(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}]\right) \\
&\geqslant \operatorname{cov}\left(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}_A], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}_B]\right).
\end{aligned}$$

Moreover, given $\mathbf{X}$, U_i is an increasing function of V_i. Hence, $\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}_A]$ and $\mathbb{E}[f(\mathbf{U}_B)|\mathbf{V}_B]$ are coordinatewise increasing (or decreasing) with respect to $\mathbf{V}_A$ and $\mathbf{V}_B$, respectively, if f and g are both coordinatewise increasing (or decreasing). Because $\mathbf{V}$ is PA, we have

$$\operatorname{cov}\left(\mathbb{E}[f(\mathbf{U}_A)|\mathbf{V}_A], \mathbb{E}[g(\mathbf{U}_B)|\mathbf{V}_B]\right) \geqslant 0,$$

implying that $\mathbf{U}$ is PA. The proofs for other positive dependence concepts are similar. □

11.6 Two applications

We provide two applications in this section to highlight the usefulness of Theorem 11.3.

11.6.1 An application on diversification penalty

For random variables X and Y, let $X \geqslant_{\mathrm{st}} Y$ represent $\mathbb{P}(X > x) \geqslant \mathbb{P}(Y > x)$ for all $x \in \mathbb{R}$; this is called the stochastic order. Chen et al. (2024a,b) studied the problem of diversification penalty; that is, whether

$$X \leqslant_{\mathrm{st}} \sum_{i=1}^{d} \theta_i X_i \text{ for all } (\theta_1, \dots, \theta_d) \in \Delta_d, \text{ where } X, X_1, \dots, X_d \text{ are identically distributed,} \tag{11.3}$$

holds under certain marginal distributions and dependence structures. Here, Δ_d is the standard simplex defined by $\Delta_d = \{(\theta_1, \dots, \theta_d) \in [0,1]^d : \theta_1 + \cdots + \theta_d = 1\}$. When X is interpreted as a loss, (11.3) intuitively means that the non-diversified portfolio X is less dangerous than the diversified portfolio $\sum_{i=1}^{d} \theta_i X_i$. This seems counter-intuitive at first glance, but it indeed happens in the model of Chen et al. (2024a), where X has infinite mean.

Define the set, for some dependence concept $\mathfrak{D}$ in Section 11.5.1,

$$\mathcal{F}_{\mathfrak{D}} = \{\text{distribution of } X : (11.3) \text{ holds for all } (X_1, \dots, X_d) \text{ that satisfy } \mathfrak{D}\}.$$

Chen et al. (2024a) showed that the Pareto(1) distribution belongs to $\mathcal{F}_{\mathrm{WNA}}$, and hence also to $\mathcal{F}_{\mathrm{NA}}$, $\mathcal{F}_{\mathrm{NRD}}$, and $\mathcal{F}_{\mathrm{IN}}$, where IN stands for independence. Moreover, Chen et al. (2024b, Proposition 1) showed that $\mathcal{F}_{\mathfrak{D}}$ for $\mathfrak{D}$ being WNA, NA, or IN is closed under strictly

increasing convex transforms on the random variables. Our next result, which relies on our Theorem 11.3, addresses non-strictly increasing f and other notions of dependence, thus generalizing the above result.

Proposition 11.1. *Each of $\mathcal{F}_{\mathfrak{D}}$ is closed under increasing convex transforms on the random variable.*

Proof. Below we first show that each of $\mathcal{F}_{\mathfrak{D}}$ is closed under strictly increasing convex transforms on the random variable, that is, if the distribution of X is in $\mathcal{F}_{\mathfrak{D}}$, so is the distribution of $f(X)$ for a strictly increasing convex f. Assume that $F \in \mathcal{F}_{\mathfrak{D}}$, X follows F, and $Y = f(X)$, where f is strictly increasing and convex. Because f is strictly increasing, if $(Y_1, \ldots, Y_d)$ satisfies $\mathfrak{D}$, so does $(X_1, \ldots, X_d)$, where $X_i = f^{-1}(Y_i)$ for $i \in [d]$, by Lemma 11.2. Because each of $X, X_1, \ldots, X_d$ has a distribution $F \in \mathcal{F}_{\mathfrak{D}}$, we have $X \leqslant_{\mathrm{st}} \sum_{i=1}^{d} \theta_i X_i$, and this gives, using the convexity of f,

$$Y = f(X) \leqslant_{\mathrm{st}} f\left(\sum_{i=1}^{d} \theta_i X_i\right) \leqslant \sum_{i=1}^{d} \theta_i f(X_i) = \sum_{i=1}^{d} \theta_i Y_i. \tag{11.4}$$

To address the case that f is not strictly increasing, Theorem 11.3 allows us to find the above $(X_1, \ldots, X_d)$ that satisfies $\mathfrak{D}$ and such that $Y_i = f(X_i)$ for $i \in [d]$. In particular, using Theorem 11.3, we can construct $(U_1, \ldots, U_d)$ that follows the checkerboard copula of $(Y_1, \ldots, Y_d)$ and satisfies $\mathfrak{D}$, such that

$$(Y_1, \ldots, Y_d) = (f \circ g(U_1), \ldots, f \circ g(U_d)),$$

where g is the quantile function of X and $f \circ g$ is the quantile function of Y. Setting $(X_1, \ldots, X_d) = (g(U_1), \ldots, g(U_d))$, we get that $(X_1, \ldots, X_d)$ satisfies $\mathfrak{D}$, and this leads to (11.4). □

11.6.2 An application on induced order statistics

Here we demonstrate another application of Theorem 11.3 in characterizing the distribution of induced order statistics. Consider N independent and identically distributed bivariate random vectors

$$\begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix}, \begin{pmatrix} \xi_2 \\ \eta_2 \end{pmatrix}, \ldots, \begin{pmatrix} \xi_N \\ \eta_N \end{pmatrix}.$$

Note that, for $i \neq j$, (ξ_i, η_i) and (ξ_j, η_j) are independent and identically distributed, but ξ_i and η_i may be correlated and have different marginal distributions. We rank these bivariate vectors according to their first components, ξ_i:

$$\begin{pmatrix} \xi_{1:N} \\ \eta_{[1:N]} \end{pmatrix}, \begin{pmatrix} \xi_{2:N} \\ \eta_{[2:N]} \end{pmatrix}, \dots, \begin{pmatrix} \xi_{N:N} \\ \eta_{[N:N]} \end{pmatrix}, \tag{11.5}$$

where $\xi_{1:N} \leqslant \xi_{2:N} \leqslant \cdots \leqslant \xi_{N:N}$ are the order statistics of $\xi_1, \xi_2, \dots, \xi_N$. The notation $\eta_{[i:N]}$ represents the i-th *induced order statistic* (Bhattacharya, 1974), where the order is induced by another variable ξ_i. The induced order statistics $\eta_{[1:N]}, \dots, \eta_{[N:N]}$ are also referred to as *concomitants* of the order statistics $\xi_{1:N}, \dots, \xi_{N:N}$ (David, 1973).

In the context of constructing impact portfolios, Lo et al. (2024) investigated the joint distribution of $(\eta_{[1:N]}, \dots, \eta_{[N:N]})$. In particular, they proved a representation theorem for the joint distribution of $(\eta_{[1:N]}, \dots, \eta_{[N:N]})$ using the copula of (ξ_i, η_i). Furthermore, they demonstrated that if ξ_i is not continuously distributed, the representation theorem holds if and only if the copula of (ξ_i, η_i) is chosen as the (bivariate) checkerboard copula in this chapter. This reveals a potential application of the checkerboard copula in portfolio construction.

Lo et al. (2024) also showed that the rank of the odd-order moments of induced order statistics relies on the copula of (ξ_i, η_i). Assume that C is a copula of (ξ_i, η_i). Lo et al. (2024, Theorem EC.5) proved that, for any $k = 0, 1, \dots$, if C is PRD, we have

$$\mathbb{E}\left(\eta_{[1:N]}^{2k+1}\right) \leqslant \mathbb{E}\left(\eta_{[2:N]}^{2k+1}\right) \leqslant \cdots \leqslant \mathbb{E}\left(\eta_{[N:N]}^{2k+1}\right), \tag{11.6}$$

and if C is NRD, we have

$$\mathbb{E}\left(\eta_{[1:N]}^{2k+1}\right) \geqslant \mathbb{E}\left(\eta_{[2:N]}^{2k+1}\right) \geqslant \cdots \geqslant \mathbb{E}\left(\eta_{[N:N]}^{2k+1}\right). \tag{11.7}$$

In particular, the copula C can be chosen as the checkerboard copula. Therefore, using Theorem 11.3, we directly obtain the following result.

Proposition 11.2. *For any $k = 0, 1, \dots$, (11.6) holds if (ξ_i, η_i) is PRD, and (11.7) holds if (ξ_i, η_i) is NRD.*

The difference between Proposition 11.2 and Lo et al. (2024, Theorem EC.5) is that the latter imposes the dependence assumption (PRD or NRD) on the copula of (ξ_i, η_i), while the

former imposes a more natural assumption on the random vector (ξ_i, η_i) directly, which is only possible due to our Theorem 11.3.

11.7 Conclusion

We discussed the choice of copula when the marginal distributions are not necessarily continuous. Among all the choices of copulas for a given random vector, the checkerboard copula is the most convenient and natural selection in applications such as simulating from the copula, stressing the distribution, and computing a co-risk measure. It is shown that the checkerboard copula is the most unbiased choice in the sense that it has the largest Shannon entropy among all possible copulas for a given random vector. Moreover, the checkerboard copula can preserve the dependence information of the underlying random vector. This preservation property is applied to identify suitable distributions in the context of diversification penalty studied by Chen et al. (2024a,b) and to determine the ranks of the moments of induced order statistics in the context of impact portfolios studied by Lo et al. (2024).

Chapter 12

Summary and potential future work

12.1 Summary

My thesis explores various topics in quantitative risk management, including risk measures, dependence modelling, diversification measures, and risk sharing. In this section, we conclude with the key results for all chapters of this thesis.

Chapter 2 primarily addresses the calibration problem and properties of the PELVE. We developed the calibration distribution model for a one-point constraint (Proposition 2.1), a two-point constraint (Theorem 2.1), and the curve constraint (Section 2.4.3). Additional theoretical results of PELVE include studies on the monotonicity (Theorem 2.3) and convergence (Corollary 2.2) properties.

Chapter 3 focuses on the relationship between JM and classic notions of negative dependence. Theorem 3.1 gives the condition for a joint mix to be negative association. Theorem 3.2 provides the necessary conditions for the existence of negatively dependent joint mixes. Theorem 3.4 solves a multi-marginal optimal transport problem for quadratic costs under uncertainty on the participation of agents by using the negatively dependent joint mix.

Chapter 4 has formulated two axiomatic characterizations for diversification quotient through Theorem 4.1 and Theorem 4.2. Propositions 4.1-4.4 examine the fundamental properties of DQ for common risk measures, some of which are not shared by DR. DQ is intuitive and interpretable with respect to dependence and common perceptions of diversification

(Theorem 4.3). The alternative formulas for $\mathrm{DQ}^{\mathrm{VaR}}$ and $\mathrm{DQ}^{\mathrm{ES}}$ are provided in Theorem 4.4, which is convenient in the portfolio optimization of DQs (Proposition 4.6). The proof of main results and additional results supporting the discussion in Chapter 4 are included in Chapter 5. More properties of $\mathrm{DQ}^{\mathrm{VaR}}$ and $\mathrm{DQ}^{\mathrm{ES}}$ are studies in Chapter 6. Theorem 6.1 shows the range of DQs based on VaR and ES and the corresponding dependence structures for special values. Using the law of large numbers, we show that the value of DQs based on VaR and ES for large portfolios converges to 0 (Theorem 6.2), which aligns with the intuition that the diversification degree increases with the number of assets. More asymptotic behaviors of DQs based on VaR and ES for elliptical distributions and MRV models are presented in Proposition 6.3 and Theorem 6.3. Theorem 6.4 and Proposition 6.5 solved the portfolio optimization for elliptical models and MRV models.

Chapter 7 provides a series of technical results on pairwise counter-monotonicity, including the representation (Theorem 7.1), invariance property (Theorem 7.2), connection to negative association (Theorem 7.3), joint mix dependence (Theorem 7.4), and optimal allocations for quantile agents (Theorem 7.5). Chapter 8 further studies the risk-sharing problem for agents using distortion riskmetrics. We obtained several results on the necessary or sufficient conditions for Pareto and sum optimality (Theorem 8.1 and Propositions 8.2-8.5) in comonotonic and general allocation sets. For IQD agents, we fully characterize all Pareto-optimal allocations via the inf-convolution (Theorem 8.2 and Proposition 8.9). The Pareto-optimal allocations for IQD agents exhibit pairwise comonotonicity structures. For general distortion riskmetrics, we obtained explicit allocations that are Pareto optimal or sum optimal in comonotonic allocation sets (Theorem 8.3). Chapter 9 further investigates the appearance of pairwise counter-monotonicity in the risk-sharing problem. The main result, the counter-monotonic improvement theorem (Theorem 9.1), lays the foundation for analyzing risk sharing with counter-monotonic allocations. This theorem has allowed us to shed light on Pareto-optimal allocations when the risk is to be shared among risk-seeking agents (Theorem 9.2), agents with a discontinuous Bernoulli utility function (Proposition 9.5), and RDU agents with inverted S-shaped probability distortion functions (Theorem 9.3).

Chapter 10 and Chapter 11 explore dependence modelling through correlation and copulas. Characterization results for invariant correlation with general and increasing transfor-

mations are summarized in Table 10.1 and Table 10.2, respectively. Quasi-independence and quasi-Fréchet models play important roles in these studies. We also identify the set of all compatible invariant correlation matrices in Theorem 10.4. Theorem 11.1 provides a representation for all copulas of a given random vector, with checkerboard copulas as a special case. We further show that the checkerboard copula has the largest Shannon Entropy (Theorem 11.2) and captures the dependence information from original random vectors (Theorem 11.3).

12.2 Potential future work

In this section, I briefly discuss some open questions related to PELVE, diversification indices, negative dependence modeling and risk-sharing problems, that will be explored in the future. .

12.2.1 On PELVE

A.1 We have discussed the valid input of PELVE in the one-point and two-point constrain problems in Section 2.3, but the valid input of PELVE functions in the curve constrain problem is still unknown. Determining whether a given function π on $[0,1]$ can be a PELVE function for some $X \in \mathcal{X}$ remains a challenging task.

A.2 We have provided only one solution for the calibrated distribution model for the three cases we studied in Chapter 2. A natural extension is to characterize all the distribution functions that satisfy a given PELVE.

A.3 Since the PELVE value is related to the tail heaviness of a distribution, an interesting question is whether the tail behavior, e.g., the maximal finite moment index, of a non-regularly varying distribution is completely determined by its PELVE.

12.2.2 On negative dependence modelling and risk sharing

B.1 We already know from Chapters 3 and 7 that joint mixability and pairwise counter-monotonicity provide the safest aggregation of risks. It remains to explore what kind of dependence structure yields the safest aggregation risk when a Fréchet class does not support joint mix dependence or pairwise counter-monotonicity. Several research directions are worth further investigation, such as d-countermonotonicity in Lee and Ahn (2014) and Σ-countermonotonicity in Puccetti and Wang (2015). I will further explore the relationship between these forms of negative dependence.

B.2 The optimal transport problem under uncertainty is not fully solved in Chapter 3. The solution for the general setting of cost functions, uncertainty sets, and heterogeneous marginal distributions remains unknown.

B.3 It remains unclear to us whether our analysis in Chapter 8 can be generalized to distortion risk metrics other than IQD, which are not convex (i.e., those with non-concave distortion functions). As far as we are aware, the unconstrained risk-sharing problems for non-convex risk measures and variability measures have very limited explicit results (e.g., Embrechts et al. (2018), Weber (2018), and Liu et al. (2022)), further investigation is needed for a better understanding of the challenges and their solutions.

B.4 Chapter 9 discusses some decision models that exhibit pairwise counter-monotonicity in Pareto-optimal allocations. A more ambitious goal is to characterize all decision models that utilize extremely negatively dependent allocations under some reasonable conditions.

B.5 Inspired by the treatment effect, the upper and lower bounds on $\mathrm{VaR}_p(X+Y)$ under the order constraint $X \leqslant Y$ have been studied by Chen et al. (2022). If we use X and Y to represent the outcomes before and after a treatment or intervention, such as in drug and policy experiments, the value of $Y - X$ represents the effect of the treatment. Hence, bounds on $\mathrm{VaR}_p(Y-X)$ under the order constraint $X \leqslant Y$ are also of interest A useful dependence structure is the maximal coupling mentioned in Thorisson (2000), which can solve the lower bound on $\mathrm{VaR}_p(Y-X)$ in the most trivial

case where $\mathrm{VaR}_p(Y - X) = 0$.

12.2.3 On diversification indices

C.1 Although we have shown that VaR-based DQ and ES-based DQ have simple formulations, as discussed in Chapter 4, the formulation of DQs based on other risk measures, such as expectiles and entropic risk measures, remains unclear. Our current results in an on-going project show that DQ based on expectiles is closely related to the Ω-ratio. For a given $\alpha \in [0, 1/2]$, the DQ based on expectiles has the alternative formula:

$$\mathrm{DQ}_\alpha^{\mathrm{ex}}(\mathbf{X}) = \frac{1}{\alpha}\left\{\mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathrm{ex}_\alpha(X_i))\right)_+\right] \Big/ \mathbb{E}\left[\left|\sum_{i=1}^n (X_i - \mathrm{ex}_\alpha(X_i))\right|\right]\right\}, \quad \mathbf{X} \in \mathcal{X}^n.$$

With this alternative formula, we can further explore other theoretical properties and practical applications of expectiles-based DQ, including but not limited to portfolio optimization and asymptotic behavior under MRV and elliptical models.

C.2 Currently, we use the estimators of DQs based on VaR and ES, as shown in Chapter 4, for the experiments in Section 4.7. However, their unbiasedness, efficiency, and consistency have not yet been shown. It is important to explore these statistical properties of the VaR-based and ES-based DQ estimators for their future usage.

C.3 We can extend the axiomatic framework of diversification indices in the following ways.

(i) We established the axiomatic framework for DQ based on monotone, constant additive, and positive homogeneous risk measures in Theorem 4.1, as well as for coherent risk measures in Theorem 4.2. It remains to extend the axiomatic framework to DQ based on other risk measures, such as convex risk measures.

(ii) Since DQs based on VaR and ES offer substantial theoretical and practical benefits, an interesting question is whether we can further character $\mathrm{DQ}^{\mathrm{VaR}}$ and $\mathrm{DQ}^{\mathrm{ES}}$ with more reasonable axioms.

(iii) The existing studies on the axiomatic framework for diversification indices are limited. It remains to fill this gap in the literature with characterization results

for other diversification indices, such as the diversification ratio and diversification benefit.

Bibliography

Acerbi, C. and Székely, B. (2014). Backtesting expected shortfall. *Risk Magazine*, 17 Nov 2014.

Adrian, T. and Brunnermeier, M. K. (2016). CoVaR. *American Economic Review*, **106**(7), 1705–1741.

Alam, K. and Saxena, K. M. L. (1981). Positive dependence in multivariate distributions. *Communications in Statistics-Theory and Methods*, **10**(12), 1183–1196.

Aleksandr, I. and Khinchin, A. (1949). *Mathematical Foundations of Statistical Mechanics*. Courier Corporation.

Allamanis, M. (2019). The adverse effects of code duplication in machine learning models of code. In *Proceedings of the 2019 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, 143–153.

Amini, M., Nili Sani, H. R. and Bozorgnia, A. (2013). Aspects of negative dependence structures. *Communications in Statistics-Theory and Methods*, **42**(5), 907–917.

Arrow, K. J. (1964). The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies*, **31**(2), 91–96.

Arrow, K. J., and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, **22**(3), 265–290.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Balbás, A., Balbás, B. and Balbás, R. (2017). Differential equations connecting VaR and CVaR. *Journal of Computational and Applied Mathematics*, **326**, 247–267.

Barczy, M., Nedényi, F. K. and Sütő, L. (2022). Probability equivalent level of Value at Risk and higher-order Expected Shortfalls. *Insurance: Mathematics and Economics*, **108**, 107–128.

Barrieu, P. and El Karoui, N. (2005). Inf-convolution of risk measures and optimal risk transfer.

Finance and Stochastics, **9**, 269–298.

Basak, S. and Shapiro, A. (2001). Value-at-Risk based risk management: Optimal policies and asset prices. *Review of Financial Studies*, **14**(2), 371–405.

Bates, S., Candès, E., Lei, L., Romano, Y. and Sesia, M. (2023). Testing for outliers with conformal p-values. *Annals of Statistics*, **51**(1), 149–178.

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. BIS online publication No. bcbs457.

Behan, D. F., Cox, S. H., Lin, Y., Pai, J., Pedersen, H. W. and Yi, M. (2010). Obesity and its relation to mortality and morbidity costs. *Society of Actuaries Research Projects – Life Insurance*, December 2010.

Beissner, P., Boonen, T., and Ghossoub, M. (2023). (No-) Betting Pareto Optima Under Rank-Dependent Utility. *Mathematics of Operations Research*. https://doi.org/10.1287/moor.2022.0317.

Bellini, F., Fadina, T., Wang, R., and Wei, Y. (2022). Parametric measures of variability induced by risk measures. *Insurance: Mathematics and Economics*, **106**, 270–284.

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **54**(1), 41–48.

Bellman, R. and Cooke, K. L. (1963). *Differential-difference Equations*. The RAND Corporation.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**(1), 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**(4), 1165–1188.

Ben-Tal, A. and Teboulle, M. (2007). An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, **17**(3), 449–476.

Berezansky, L. and Braverman, E. (2011). On nonoscillation of advanced differential equations with several terms. *Abstract and Applied Analysis*, **2011**, 637142.

Berger, L. and Eeckhoudt, L. (2021). Risk, ambiguity, and the value of diversification. *Management Science*, **67**(3), 1639–1647.

Bernard, C., Bondarenko, O. and Vanduffel, S. (2018). Rearrangement algorithm and maximum entropy. *Annals of Operations Research*, **261**(1), 107–134.

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

Bernard, C., Rüschendorf, L. and Vanduffel, S. (2017). VaR bounds with variance constraint. *Journal of Risk and Insurance*, **84**(3), 923–959.

Bernard, C. and Vanduffel, S. (2015). A new approach to assessing model risk in high dimensions. *Journal of Banking and Finance*, **58**, 166–178.

Besser, K. and Jorswieck, E. A. (2020). Reliability bounds for dependent fading wireless channels. *IEEE Transactions on Wireless Communications*, **59**(9), 5833–5845.

Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *The Annals of Statistics*, **2**(5), 1034–1039.

Billot, A., Chateauneuf, A., Gilboa, I., and Tallon, J.M. (2000). Sharing beliefs: between agreeing and disagreeing. *Econometrica*, **68**(3), 685–694.

Bignozzi, V., Mao, T., Wang, B. and Wang, R. (2016). Diversification limit of quantiles under dependence uncertainty. *Extremes*, **19**(2), 143–170.

Blanchet, J., Lam, H., Liu, Y. and Wang, R. (2020). Convolution bounds on quantile aggregation. *arXiv:* 2007.09320

Block, H. W., Savits, T. H. and Shaked, M. (1982). Some concepts of negative dependence. *Annals of Probability*, **10**(3), 765–772.

Block, H. W., Savits, T. H., and Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics and probability letters*, **3**(2), 81–86.

Boerma, J., Tsyvinski, A. and Zimin, A. P. (2021). Sorting with team formation. *arXiv*: 2109.02730.

Boonen, T. J. and Ghossoub, M. (2020). Bilateral risk sharing with heterogeneous beliefs and exposure constraints. *ASTIN Bulletin*, **50**(1), 293–323.

Boudt, K., Dragun, K. and Vanduffel, S. (2022). The optimal payoff for a Yaari investor. *Quantitative Finance*, **22**(10), 1839–1852.

Boyle, P., Garlappi, L., Uppal, R. and Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management Science*, **58**(2), 253–272.

Bürgi, R., Dacorogna, M. M. and Iles, R. (2008). Risk aggregation, dependence structure and

diversification benefit. *SSRN*: 1468526.

Carlier, G. and Dana, R. A. (2003). Pareto efficient insurance contracts when the insurer's cost function is discontinuous. *Economic Theory*, **21**(4), 871–893.

Carlier, G., Dana, R.-A. and Galichon, A. (2012). Pareto efficiency for the concave order and multivariate comonotonicity. *Journal of Economic Theory*, **147**, 207–229.

Castagnoli, E., Cattelan, G., Maccheroni, F., Tebaldi, C. and Wang, R. (2022). Star-shaped risk measures. *Operations Research*, **70**(5), 2637–2654.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011). Risk measures: Rationality and diversification. *Mathematical Finance*, **21**(4), 743–774.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Rustichini, A. (2014). Niveloids and their extensions: Risk measures on small domains. *Journal of Mathematical Analysis and Applications*, **413(1)**, 343-360.

Chambers, C. P. (2009). An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance*, **19**(2), 335–342.

Chaoubi, I., Cossette, H., Gadoury, S.-P. and Marceau, E. (2020). On sums of two counter-monotonic risks. *Insurance: Mathematics and Economics*, **92**, 47–60.

Chateauneuf, A., Dana, R. A., and Tallon, J.M. (2000). Optimal risk-sharing rules and equilibria with Choquet-expected-utility. *Journal of Mathematical Economics*, **34**(2), 191–214.

Chi, Z., Ramdas, A. and Wang, R. (2022). Multiple testing under negative dependence. *arXiv*: 2212.09706.

Cherny, A. and Madan, D. (2009). New measures for performance evaluation. *The Review of Financial Studies*, **22**(7), 2571–2606.

Chen, Y., Embrechts, P. and Wang, R. (2024b). Risk exchange under infinite-mean Pareto models. *arXiv*: 2403.20171.

Chen, Y., Embrechts, P. and Wang, R. (2024a). An unexpected stochastic dominance: Pareto distributions, dependence, and diversification. *Operations Research*. https://doi.org/10.1287/opre.2022.0505.

Chen, Y., Lin, L., and Wang, R. (2022). Risk aggregation under dependence uncertainty and an order constraint. *Insurance: Mathematics and Economics*, **102**, 169–187.

Cheung, K. C., Denuit, M. and Dhaene, J. (2017). Tail mutual exclusivity and Tail-VaR lower

bounds. *Scandinavian Actuarial Journal*, **2017**(1), 88–104.

Cheung, K. C., Dhaene, J., Lo, A. and Tang, Q. (2014). Reducing risk by merging counter-monotonic risks. *Insurance: Mathematics and Economics*, **54**, 58–65.

Cheung, K. C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance: Mathematics and Economics*, **55**, 180–190.

Cheung, K. C. and Lo, A. (2017). Characterizations of optimal reinsurance treaties: a cost-benefit approach. *Scandinavian Actuarial Journal*, **2017**(1), 1–28.

Choi, N., Fedenia, M., Skiba, H. and Sokolyk, T. (2017). Portfolio concentration and performance of institutional investors worldwide. *Journal of Financial Economics*, **123**(1), 189–208.

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *The Journal of Portfolio Management*, **35**(1), 40–51.

Christofides, T. C. and Vaggelatou, E. (2004). A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, **88**(1), 138–151.

Coffman, E. G. and Yannakakis, M. (1984). Permuting elements within columns of a matrix in order to minimize maximum row sum. *Mathematics of Operations Research*, **9**(3), 384–390.

Corbett, C. J. and Rajaram, K. (2006). A generalization of the inventory pooling effect to non-normal dependent demand. *Manufacturing and Service Operations Management*, **8**(4), 351–358.

Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.

Craiu, R. V. and Meng, X. L. (2001). On the edge: Statistics & computing: Chance and fractals. *Chance*, **14**(2), 47–52.

Craiu, R. V. and Meng, X. L. (2005). Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo. *Annals of Statistics*, **33**(2), 661–697.

Cui, H., Tan, K. S., Yang, F. and Zhou, C. (2022) Asymptotic analysis of portfolio diversification. *Insurance: Mathematics and Economics*, **106**, 302–325.

D'Acunto, F., Prabhala, N. and Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, **32**(5), 1983–2020.

Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica*, **9**, 131–150.

Dana, R-A. and Meilijson, I. (2003). Modelling agents' preferences in complete markets by second

order stochastic dominance. *Peprint*, Université Paris Dauphine.

David, H. A. (1973). Concomitants of order statistics. *Bulletin of the International Statistical Institute*, **45**(1), 295–300.

Debreu, G. (1954). Representation of a preference ordering by a numerical function. *Decision Processes*, **3**, 159–165.

Delbaen, F. (2002). Coherent risk measures on general probability spaces. *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, 1–37. Eds: K. Sandmann, P. J. Schönbucher.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, **22**(5), 1915–1953.

Denis, D. J., Denis, D. K. and Yost, K. (2002). Global diversification, industrial diversification, and firm value. *The Journal of Finance*, **57**(5), 1951–1979.

Denneberg, D. (1990). Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, **20**, 181–190.

Denneberg, D. (1994). *Non-additive Measure and Integral.* Springer Science & Business Media.

Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). *Actuarial Theory for Dependent Risks.* Wiley: Chichester, UK.

Dhaene, J. and Denuit, M. (1999). The safest dependence structure among risks. *Insurance: Mathematics and Economics*, **25**(1), 11–21.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vynche, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**(1), 3–33.

Dhaene, J., Tsanakas, A., Valdez, E. A. and Vanduffel, S. (2012). Optimal capital allocation principles. *Journal of Risk and Insurance*, **79**(1), 1–28.

Dhaene, J., Vanduffel, S., Goovaerts, M.J., Kaas, R., Tang, Q. and Vynche, D. (2006). Risk measures and comonotonicity: A review. *Stochastic Models*, **22**, 573–606.

Di Marino, S., Gerolin, A. and Nenna, L. (2017). Optimal transportation theory with repulsive costs. In: *Topological Optimization and Optimal Transport* (Eds: Bergounioux et al.), 204–256.

Drapeau, S. and Kupper, M. (2013). Risk preferences and their robust representation. *Mathematics of Operations Research*, **38**(1), 28–62.

Durante, F. and Sempi, C. (2016) *Principles of Copula Theory*. Boca Raton: CRC press.

Durrett, R. (2019). *Probability: Theory and Examples*. Vol. 49. Cambridge University Press.

EIOPA (2011). Equivalence assessment of the Swiss supervisory system in relation to articles 172, 227 and 260 of the Solvency II Directive, EIOPA-BoS-11-028.

Embrechts, P., Furrer, H. and Kaufmann, R. (2009). Different kinds of risk. In *Handbook of financial time series* (pp. 729–751). Springer, Berlin, Heidelberg.

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. *Annals of Applied Probability*, **26**(3), 1636–1658.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

Embrechts, P., Lambrigger, D. and Wüthrich, M. (2009). Multivariate extremes and the aggregation of dependent risks: Examples and counter-examples. *Extremes*, **12**(2), 107–127.

Embrechts, P., Liu, H., Mao, T. and Wang, R. (2020). Quantile-based risk sharing with heterogeneous beliefs. *Mathematical Programming Series B*, **181**(2), 319–347.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (Ed. Dempster, M. A. H.), 176–223, Cambridge University Press.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.

Embrechts, P., Schied, A. and Wang, R. (2022). Robustness in the optimization of risk measures. *Operations Research*, **70**(1), 95–110.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Emmer, S., Kratz, M. and Tasche, D. (2015). What is the best risk measure in practice? A

comparison of standard measures. *Journal of Risk*, **18**(2), 31–60.

Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics*, **38**(5), 1466–1474.

Fama, E.F. and Miller, M.H. (1972). *The Theory of Finance.* Holt, Rinehart, and Winston, New York.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions.* London: Chapman & Hall.

Fiebig, U. R., Strokorb, K. and Schlather, M. (2017). The realization problem for tail correlation functions. *Extremes*, **20**, 121–168.

Filipovic, D and Svindland, G. (2008). Optimal capital and risk allocations for law- and cash-invariant convex functions. *Finance and Stochastics*, **12**(3), 423–439.

Fiori, A. M. and Gianin, E. R. (2022). Generalized PELVE and applications to risk measures. *European Actuarial Journal*, **13**(1), 307–339.

Föllmer, H. and Knispel, T. (2011). Entropic risk measures: Coherence vs. convexity, model ambiguity, and robust large deviations. *Stochastics and Dynamics*, **11**(2/3), 333–351.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* Fourth Edition. Walter de Gruyter, Berlin.

Frees, E. W. (2009). *Regression Modeling with Actuarial and Financial Applications.* Cambridge University Press.

Frittelli, M. and Rosazza Gianin, E. (2002). Putting order in risk measures. *Journal of Banking and Finance*, **26**,1473–1486.

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocation and heavy-tailed risks. *Journal of Banking and Finance*, **83**, 70–84.

Galichon, A. (2016). *Optimal Transport Methods in Economics.* Princeton University Press.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, **37**(2), 475–515.

Genest, C., Nešlehová, J. G. and Rémillard, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *Journal of Multivariate Analysis*, **159**, 82–110.

Gilboa, I. (2009). *Theory of Decision under Uncertainty.* Cambridge University Press, Cambridge.

Gilboa, I., Postlewaite, A. and Samuelson, L. (2019). What are axiomatizations good for? *Theory*

and Decision, **86**(3–4), 339–359.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**, 141–153.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494), 746–762.

Grechuk, B., Molyboha, A. and Zabarankin, M. (2009). Maximum entropy principle with general deviation measures. *Mathematics of Operations Research*, **34**(2), 445–467.

Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, **47**(5), 1785–1809.

Grötschel, M. and Wakabayashi, Y. (1990). Facets of the clique partitioning polytope. *Mathematical Programming*, **47**, 367–387.

Guan, Y., Jiao, Z. and Wang, R. (2022). A reverse Expected Shortfall optimization formula. *arXiv*: 2203.02599.

Hardy, G. H., Littlewood, J. E. and Pólya, G (1934). *Inequalities*. Cambridge University Press.

Haus, U. (2015). Bounding stochastic dependence, complete mixability of matrices, and multidimensional bottleneck assignment problems. *Operations Research Letters*, **43**(1), 74–79.

Heidari, A., Kushagra, S. and Ilyas, I. F. (2020). On sampling from data with duplicate records. *arXiv preprint arXiv:2008.10549*.

Hofert, M. and Koike, T. (2019). Compatibility and attainability of matrices of correlation-based measures of concordance. *ASTIN Bulletin*, **49**(3), 885–918.

Hsu, W. L. (1984). Approximation algorithms for the assembly line crew scheduling problem. *Mathematics of Operations Research*, **9**(3), 376–383.

Hu, T. (2000). Negatively superadditive dependence of random variables with applications. *Chinese Journal of Applied Probability and Statistics*, **16**(2), 133–144.

Hu, T. and Wu, Z. (1999). On dependence of risks and stop-loss premiums. *Insurance: Mathematics and Economics*, **24**(3), 323–332.

Huber, P. J. and Ronchetti E. M. (2009). *Robust Statistics.* Second ed., Wiley Series in Probability and Statistics. Wiley, New Jersey. First ed.: Huber, P. (1980).

Huberman, G., Mayers, D. and Smith Jr, C. W. (1983). Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, **14**(2), 415–426.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, **34**(3), 587–608.

Ibragimov, R., Jaffee, D. and Walden, J. (2011). Diversification disasters. *Journal of Financial Economics*, **99**(2), 333–348.

Jakobsons, E., Han, X. and Wang, R. (2016). General convex order on risk aggregation. *Scandinavian Actuarial Journal*, **2016**(8), 713-740.

Jansen, D. W. and De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, **73**(1), 18–24.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, **106**(4), 620–628.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Annals of Statistics*, **11**, 286–295.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H. (2014). *Dependence Modeling with Copulas*. London: Chapman & Hall.

Jouini, E., Schachermayer, W. and Touzi, N. (2008). Optimal risk sharing for law invariant monetary utility functions. *Mathematical Finance*, **18**(2), 269–292.

Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47**(2), 263–292.

Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, **15**(3), 425–437.

Krause, D., Scherer, M., Schwinn, J. and Werner, R. (2018). Membership testing for Bernoulli and tail-dependence matrices. *Journal of Multivariate Analysis*, **168**, 240–260.

Koike, T. and Hofert, M. (2023). Matrix compatibility and correlation mixture representation of generalized Gini's gamma. *Canadian Journal of Statistics*, **51**(4), 1111–1125.

Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, **37**(5), 519–531.

Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5), 1056–1072.

Koumou, G. B. (2020). Diversification and portfolio theory: A review. *Financial Markets and Portfolio Management*, **34**(3), 267–312.

Koumou, G. B. and Dionne, G. (2022). Coherent diversification measures in portfolio theory: An axiomatic foundation. *Risks*, **10**(11), 205.

Kováčová, G., Rudloff, B., Cialenco, I. (2020). Acceptability maximization. *arXiv*: 2012.11972.

Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series.* Berlin: Springer.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.

Landsberger, M. and Meilijson, I. (1994). Co-monotone allocations, Bickel-Lehmann dispersion and the Arrow-Pratt measure of risk aversion. *Annals of Operations Research*, **52**(2), 97–106.

Lee, W. and Ahn, J. Y. (2014). On the multidimensional extension of countermonotonicity and its applications. *Insurance: Mathematics and Economics*, **56**, 68–79.

Lee, W., Cheung, K. C. and Ahn, J. Y. (2017). Multivariate countermonotonicity and the minimal copulas. *Journal of Computational and Applied Mathematics*, **317**, 589–602.

Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(5), 1137–1153.

Leshno, J. D. and Strack, P. (2020). Bitcoin: An axiomatic approach and an impossibility theorem. *American Economic Review: Insights*, **2**(3), 269–286.

Levy, H. and Levy, M. (2004). Prospect theory and mean-variance analysis. *The Review of Financial Studies*, **17**(4), 1015–1041.

Levy, H. and Sarnat, M. (1970). International diversification of investment portfolios. *The American Economic Review*, **60**(4), 668–675.

Li, D. X. (2000). On default correlation: A copula function approach. *Journal of Fixed Income*, **9**(4): 43–54.

Li, H. and Wang, R. (2022). PELVE: Probability equivalent level of VaR and ES. *Journal of Econometrics*, https://doi.org/10.1016/j.jeconom.2021.12.012.

Li, L., Shao, H., Wang, R. and Yang, J. (2018). Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics*, **9**(1), 190–218.

Liebrich, F. (2021). Risk sharing under heterogeneous beliefs without convexity. *arXiv:* 2108.05791

Liu, H. (2020). Weighted comonotonic risk sharing under heterogeneous beliefs. *ASTIN Bulletin*, **50**(2), 647–673.

Liu, F., Mao, T., Wang, R. and Wei, L. (2022). Inf-convolution, optimal allocations, and model

uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519.

Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.

Liu, P. and Schied, A. and Wang, R. (2021). Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research*, **46**(4), 1490–1512.

Liu, P., Wang, R. and Wei, L. (2020). Is the inf-convolution of law-invariant preferences law-invariant? *Insurance: Mathematics and Economics*, **91**, 144–154.

Liu, X., Küçükyavuz, S. and Luedtke, J. (2016). Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, **157**(1), 219–243.

Lo, A. W., Wu, L., Zhang, R. and Zhao, C. (2024). Optimal impact portfolios with general dependence and marginals. *Operations Research*. https://doi.org/10.1287/opre.2023.0400.

Ludkovski, M. and Rüschendorf, L. (2008). On comonotonicity of Pareto optimal risk sharing. *Statistics and Probability Letters*, **78**(10), 1181–1188.

Luedtke, J. (2014). A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming*, **146**(1), 219–244.

Maccheroni, F., Marinacci, M. and Ruffino, D. (2013). Alpha as ambiguity: Robust mean-variance portfolio analysis. *Econometrica*, **81**(3), 1075–1113.

Maccheroni, F., Marinacci, M., Rustichini, A. and Taboga, M. (2009). Portfolio selection with monotone mean-variance preferences. *Mathematical Finance*, **19**(3), 487–521.

Mainik, G. and Embrechts, P. (2013). Diversification in heavy-tailed portfolios: Properties and pitfalls. *Annals of Actuarial Science*, **7**(1), 26–45.

Mainik, G. and Rüschendorf, L (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, **14**(4), 593–623.

Malinovsky, Y. and Rinott, Y. (2023). On tournaments and negative dependence. *Journal of Applied Probability*, **60**(3), 945–954.

Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, **7**(1), 77–91.

Markowitz, H. (2014). Mean–variance approximations to expected utility. *European Journal of Operational Research*, **234**(2), 346–355.

Mas-Colell, A., Whinston, M.D. and Green, J.R. (1995). *Microeconomic Theory*. Oxford University Press.

Matmoura, Y. and Penev, S. (2013). Multistage optimization of option portfolio using higher order coherent risk measures. *European Journal of Operational Research*, **227**(1), 190–198.

Mafusalov A. and Uryasev S. (2018). Buffered probability of exceedance: Mathematical properties and optimization. *SIAM Journal on Optimization*, **28**(2), 1077–1103.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.

McNeil, A. J., Nešlehová, J. G. and Smith, A. D. (2022). On attainability of Kendall's tau matrices and concordance signatures. *Journal of Multivariate Analysis*, **191**, 105033.

Millossovich, P., Tsanakas, A. and Wang, R. (2021). A theory of multivariate stress testing. *SSRN*: 3966204.

Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics*, **3**(3), 599–616.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley.

Nelsen, R. (2006). *An Introduction to Copulas*. Springer, New York, Second Edition.

Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, **98**(3), 544–567.

Pass, B. (2015). Multi-marginal optimal transport: Theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, **49**(6), 1771–1790.

Perchet, V., Rigollet, P. and Gouic, T. L. (2022). An algorithmic solution to the Blotto game using multi-marginal couplings. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, 208–209.

Puccetti, G., Rigo, P., Wang, B. and Wang, R. (2019). Centers of probability measures without the mean. *Journal of Theoretical Probability*, **32**(3), 1482–1501.

Puccetti, G. and Rüschendorf, L. (2013). Sharp bounds for sums of dependent risks. *Journal of Applied Probability,* **50**(1), 42–53.

Puccetti, G. and Wang R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.

Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model*. Springer Science

& Business Media.

Radner, R. (1968). Competitive equilibrium under uncertainty. *Econometrica*, **36**(1), 31–58.

Rigotti, L., Shannon, C., and Strzalecki, T. (2008). Subjective beliefs and ex ante trade. *Econometrica*, **76**(5), 1167–1190.

Rockafellar, R. T., Royset, J. O. and Miranda, S. I. (2014). Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research*, **234**(1), 140–154.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.

Rockafellar, R. T., Uryasev, S. and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, **10**(1), 51–74.

Rostek, M. (2010). Quantile maximization in decision theory. *Review of Economic Studies*, **77**, 339–371.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.

Samuelson, P. A. (1967). General proof that diversification pays. *Journal of Financial and Quantitative Analysis*, **2**(1), 1–13.

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians*. Springer, New York.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics.

Shalit, H. and Yitzhaki, S. (1984). Mean-Gini, portfolio theory, and the pricing of risky assets. *The Journal of Finance*, **39**(5), 1449–1468.

Sharpe, (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19**(3), 425–442.

Shen, S., Smith, J. C. and Ahmed, S. (2010). Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science*, **56**(10), 1794–1814.

Shyamalkumar, N. D. and Tao, S. (2020). On tail dependence matrices: The realization problem

for parametric families. *Extremes*, **23**(2), 245–285.

Siewert, C. E. and Burniston, E. E. (1973). Exact analytical solutions of $ze^z = a$. *Journal of Mathematical Analysis and Applications*, **43**(3), 626–632.

Strzalecki, T., and Werner, J. (2011). Efficient allocations under ambiguity. *Journal of Economic Theory*, **146**(3), 1173–1194.

Tasche, D. (2007). Capital allocation to business units and sub-portfolios: The Euler principle. *arXiv*: 0708.2542.

Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, **99**(1), 204–215.

Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.

Van Nieuwerburgh, S. and Veldkamp, L. (2010). Information acquisition and under-diversification. *The Review of Economic Studies*, **77**(2), 779–805.

Vovk, V., Gammerman, A. and Shafer, G. (2005). *Algorithmic Learning in a Random World*. New York: Springer.

Vovk, V., Wang, B. and Wang, R. (2022). Admissible ways of merging p-values under arbitrary dependence. *Annals of Statistics*, **50**(1), 351–375.

Wakker, P.P. (2010). *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press, Cambridge.

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. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.

Wang, B., Wang, R. and Wang, Y. (2019). Compatible matrices of Spearman's rank correlation. *Statistics and Probability Letters*, **151**, 67–72.

Wang, Q., Wang, R. and Wei, Y. (2020a). Distortion riskmetrics on general spaces. *ASTIN Bulletin*, **50**(4), 827–851.

Wang, R. (2015). Current open questions in complete mixability. *Probability Surveys*, **12**, 13–32.

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-

at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

Wang, R., Wei, Y. and Willmot, G. E. (2020b). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.

Wang, R. and Wu, Q. (2020). Dependence and risk attitudes: An equivalence. *SSRN*: 3707709.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**, 1413–1429.

Wang, S., Young, V. R. and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**(2), 173–183.

Weber, S. (2018). Solvency II, or how to sweep the downside risk under the carpet. *Insurance: Mathematics and Economics*, **82**, 191–200.

Xia, Z., Zou, Z. and Hu, T. (2023). Inf-convolution and optimal allocations for mixed-VaRs. *Insurance: Mathematics and Economics*, **108**, 156–164.

Xiao, Y. and Yao, J. (2020). A note on joint mix random vectors. *Communications in Statistics-Theory and Methods*, **49**(12), 3063–3072.

Xie, F., Condict, M. and Shete, S. (2013). Estimating duplication by content-based sampling. In *2013 USENIX Annual Technical Conference (USENIX ATC 13)*, 181–186.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

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.

Yang, J., Qi, Y. and Wang, R. (2009). A class of multivariate copulas with bivariate Fréchet marginal copulas. *Insurance: Mathematics and Economics*, **45**(1), 139–147.

Yin, C. (2021). Stochastic orderings of multivariate elliptical distributions. *Journal of Applied Probability*, **58**(2), 551-568.

Yule, G. U. (1911). *An Introduction to the Theory of Statistics*. Charles Griffin and Company.

Zhao, Y., Li, L., Wang, H., Cai, H., Bissyandé, T. F., Klein, J. and Grundy, J. (2021). On the impact of sample duplication in machine-learning-based android malware detection. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, **30**(3), 1–38.

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, **26**(4), 901–918.