# Ordering and Inequalities for Mixtures on Risk Aggregation

Yuyu Chen* Peng Liu† Yang Liu‡ Ruodu Wang§

June 22, 2021

**Abstract**

Aggregation sets, which represent model uncertainty due to unknown dependence, are an important object in the study of robust risk aggregation. In this paper, we investigate ordering relations between two aggregation sets for which the sets of marginals are related by two simple operations: distribution mixtures and quantile mixtures. Intuitively, these operations "homogenize" marginal distributions by making them similar. As a general conclusion from our results, more "homogeneous" marginals lead to a larger aggregation set, and thus more severe model uncertainty, although the situation for quantile mixtures is much more complicated than that for distribution mixtures. We proceed to study inequalities on the worst-case values of risk measures in risk aggregation, which represent conservative calculation of regulatory capital. Among other results, we obtain an order relation on VaR under quantile mixture for marginal distributions with monotone densities. Numerical results are presented to visualize the theoretical results and further inspire some conjectures. Finally, we provide applications on portfolio diversification under dependence uncertainty and merging p-values in multiple hypothesis testing, and discuss the connection of our results to joint mixability.

**Keywords:** aggregation set; distribution mixture; quantile mixture; risk measure; joint mixability

*Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: y937chen@uwaterloo.ca

†Department of Mathematical Sciences, University of Essex, UK. Email: peng.liu@essex.ac.uk

‡Corresponding author. Department of Mathematical Sciences, Tsinghua University, China. Email: yang-liu16@mails.tsinghua.edu.cn

§Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: wang@uwaterloo.ca

# 1 Introduction

Robust risk aggregation has been studied extensively with applications in banking and insurance. A typical problem in this area is to compute the worst-case values of some risk measures for an aggregate loss with unknown dependence structure. Two popular regulatory risk measures used in industry are Value-at-Risk (VaR) and the Expected Shortfall (ES); see McNeil et al. (2015) and the references therein. The worst-case value of ES in risk aggregation is explicit since ES is a coherent risk measure (Artzner et al. (1999)), whereas the worst-case value of VaR in risk aggregation generally does not admit analytical formulas, which is a known challenging problem (see e.g., Embrechts et al. (2013, 2015)). See Cai et al. (2018) on robust risk aggregation for general risk measures, and Eckstein et al. (2020) on computation of robust risk aggregation using neural networks.

The above robust risk aggregation problem involves taking the supremum of a risk measure over an *aggregation set*. Fix an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{M}$ be the set of cdfs[1] on $\mathbb{R}$. For $F \in \mathcal{M}$, $X \sim F$ means that the cdf of a random variable $X$ is $F$. Moreover, let $\mathcal{M}_1$ denote the set of cdfs on $\mathbb{R}$ with finite mean. For $\mathbf{F} = (F_1, \dots, F_n) \in \mathcal{M}^n$, the aggregation set (Bernard et al. (2014)) is defined as

$$\mathcal{D}_n(\mathbf{F}) = \{\text{cdf of } X_1 + \cdots + X_n : X_i \sim F_i,\ i = 1, \dots, n\}. \tag{1}$$

The obvious interpretation is that $\mathcal{D}_n(\mathbf{F})$ fully describes model uncertainty associated with known marginal distributions $F_1, \dots, F_n$ but unknown dependence structure. The separate modeling of marginals and dependence is a standard practice in quantitative risk modeling, often involving copula techniques; see e.g., McNeil et al. (2015). An analytical characterization of $\mathcal{D}_n(\mathbf{F})$ for a given $\mathbf{F}$ is very difficult and challenging. The only available analytical results are in Mao et al. (2019) for standard uniform marginals.

The main objective of this paper is to compare model uncertainty of risk aggregation for $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$ which represent two possible models of marginals. The strongest form of comparison is set inclusion between two aggregation sets $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\mathbf{G})$. It turns out that such a strong relation may be achievable if $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$ are related by the simple operations of distribution mixtures and quantile mixtures. Distribution mixture produces a tuple whose components are convex combinations of the given distributions and quantile mixture yields a tuple whose components are given by convex combinations of the given quantiles. Both types of operations are common in statistics and risk management, as they correspond to simple operations on the parameters in statistical models or on portfolio construction; see Section 7 for an example. Moreover, if $\mathbf{G}$ is obtained from $\mathbf{F}$ via a distribution or quantile mixture, then the mean (assumed to be finite) of any element of $\mathcal{D}_n(\mathbf{G})$ is the same as that of any element of $\mathcal{D}_n(\mathbf{F})$, making the comparison fair. To the best of our knowledge, this paper is the first systematic study on the order relation between $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\mathbf{G})$ for different $\mathbf{F}$ and $\mathbf{G}$, thus comparing model uncertainty at the level of all possible distributions.

In some cases, a strong comparison via set inclusion is not possible, but we can compare values of a

[1] In this paper, we treat probability measures on $\mathcal{B}(\mathbb{R})$ and cdfs on $\mathbb{R}$ as equivalent objects.

chosen risk measure. For a law-invariant risk measure[2] $\rho : \mathcal{M} \to \mathbb{R}$, we denote by $\overline{\rho}(\mathbf{F})$ the worst-case value of $\rho$ in risk aggregation for $\mathbf{F} \in \mathcal{M}^n$, that is,

$$\overline{\rho}(\mathbf{F}) = \sup\{\rho(F) : F \in \mathcal{D}_n(\mathbf{F})\}.$$

We shall compare $\overline{\rho}(\mathbf{F})$ with $\overline{\rho}(\mathbf{G})$, thus the worst-case values of a risk measure under model uncertainty, which usually represent conservative calculation of regulatory risk capital (e.g., Embrechts et al. (2013)). Certainly, $D_n(\mathbf{F}) \subset D_n(\mathbf{G})$ implies $\overline{\rho}(\mathbf{F}) \leqslant \overline{\rho}(\mathbf{G})$ for all risk measures $\rho$, implying that the first comparison is stronger than the second one.[3]

Our study brings insights to two relevant problems in risk management. First, suppose that $\mathbf{F}$ and $\mathbf{G}$ are two possible statistical models for the marginal distributions in a risk aggregation setting. Our results allow for a comparison of model uncertainty associated with the two models, regardless of the choice of risk measures. Although a completely unknown dependence structure is sometimes unrealistic, it is commonly agreed that the dependence structure in a risk model is difficult to accurately specify (e.g., Embrechts et al. (2013) and Bernard et al. (2017)). Hence, a comparison of the magnitude of model uncertainty is an important practical issue. On the other hand, the general conclusions remain valid even if the marginal distributions are not completely specific (see the discussion in Section 9 on the presence of marginal uncertainty), and thus the assumption of known marginal distributions in our study is not harmful.

Second, our results provide an analytical way to establish inequalities on the worst-case risk measures in the form $\overline{\rho}(\mathbf{F}) \leqslant \overline{\rho}(\mathbf{G})$. Sometimes the worst-case risk measure is difficult to calculate for $\mathbf{F}$, but it may be easier to calculate for $\mathbf{G}$. For instance, formulas on worst-case VaR are available for some homogeneous marginal distributions in Wang et al. (2013) and Puccetti and Rüschendorf (2013), but explicit results on heterogeneous marginal distributions are limited (see Blanchet et al. (2020) for a recent treatment). Therefore, we can use the analytical formula $\overline{\rho}(\mathbf{G})$, if available, as an upper bound on $\overline{\rho}(\mathbf{F})$, and this leads to interesting applications in other fields; see Section 7 for applications on portfolio diversification and multiple hypothesis testing and Section 8 for a connection to joint mixability.

Our theoretical contributions are briefly summarized below. In Sections 2 and 3, we analyze general relations on distribution and quantile mixtures. The general message of our results is that the more "homogeneous" the distribution tuple is, the larger its corresponding aggregation set $\mathcal{D}_n$ is. In particular, the set inclusion is established for any tuples connected by distribution mixtures in Theorem 2.2; that is, $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$ if $\mathbf{G}$ is a distribution mixture of $\mathbf{F}$. The problem for quantile mixtures is much more challenging. The set inclusion is established for uniform marginals in Proposition 3.4. For other families of distributions, such a general relationship does not hold, as discussed with some examples.

In Section 4, we obtain inequalities between the worst-case values of some risk measure $\rho$ in risk aggrega-

[2]We conveniently treat law-invariant risk measures as mappings on $\mathcal{M}$, although it is conventional to treat them as mappings on a space of random variables. The two settings are equivalent for law-invariant risk measures.

[3]In this paper, the set inclusion "$\subset$" is non-strict; the strict set inclusion is "$\subsetneq$". Similarly, the terms "increasing" and "decreasing" are in the non-strict sense.

tion with marginals related by distribution or quantile mixtures. Although quantile mixtures do not satisfy the relationship $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$ in general, we can prove an order property between $\overline{\rho}(\mathbf{G})$ and $\overline{\rho}(\mathbf{F})$ for commonly used risk measures. Most remarkably, in Theorem 4.4, we show that under a monotone density assumption, VaR satisfies this order property for a quantile mixture. Section 5 is dedicated to the most interesting special case of Pareto risk aggregation, with a special focus on the case of infinite mean.

Numerical results are presented in Section 6 to illustrate the obtained results. In Section 7, we provide two applications: portfolio diversification under dependence uncertainty and merging p-values in multiple hypothesis testing. Some further technical discussions on distribution and quantile mixtures are put in Section 8. Section 9 concludes the paper by presenting several open mathematical challenges related to quantile mixtures. Some proofs and further properties of Pareto risk aggregation are put in the Appendix.

# 2 Distribution mixtures

In this section we put our focus on one of the two operations: distribution mixture. The main objective is to establish some ordering relationships on the set $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\mathbf{G})$ where $\mathbf{G}$ is a distribution mixture of $\mathbf{F}$. For greater generality, we investigate a more general $f$-aggregation set $\mathcal{D}_f(\mathbf{F})$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a measurable and symmetric function.[4] Similarly to (1), for $\mathbf{F} = (F_1, \dots, F_n) \in \mathcal{M}^n$, the $f$-aggregation set is defined as

$$\mathcal{D}_f(\mathbf{F}) = \{\text{cdf of } f(X_1, \dots, X_n) : X_i \sim F_i,\ i = 1, \dots, n\}.$$

It is clear that $\mathcal{D}_n$, defined in (1), becomes a specific case of $D_f$ if $f$ is a sum function ($f(x_1, \dots, x_n) = \sum_{j=1}^n x_j$). We first present some properties of the $f$-aggregation set.

**Lemma 2.1.** *For an $n$-symmetric function $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$, $\lambda \in [0, 1]$ and an $n$-permutation $\pi$, the following hold.*

*(i)* $\mathcal{D}_f(\mathbf{F}) = \mathcal{D}_f(\pi(\mathbf{F}))$.

*(ii)* $\lambda\mathcal{D}_f(\mathbf{F}) + (1 - \lambda)\mathcal{D}_f(\mathbf{G}) \subset \mathcal{D}_f(\lambda\mathbf{F} + (1 - \lambda)\mathbf{G})$. *In particular,*

  *(a)* $\lambda\mathcal{D}_f(\mathbf{F}) + (1 - \lambda)\mathcal{D}_f(\mathbf{F}) = \mathcal{D}_f(\mathbf{F})$.

  *(b)* $\mathcal{D}_f(\mathbf{F}) \cap \mathcal{D}_f(\mathbf{G}) \subset \mathcal{D}_f(\lambda\mathbf{F} + (1 - \lambda)\mathbf{G})$.

*Proof.* (i) holds because of the symmetry of $f$. To prove (ii), for any $H \in \lambda\mathcal{D}_f(\mathbf{F}) + (1 - \lambda)\mathcal{D}_f(\mathbf{G})$, there exist $X_1 \sim F_1, \dots, X_n \sim F_n$, $Y_1 \sim G_1, \dots, Y_n \sim G_n$ and an event $A \in \mathcal{F}$ independent of $X_1, \dots, X_n, Y_1, \dots, Y_n$ such that $\mathbb{P}(A) = \lambda$ and $(f(X_1, \dots, X_n)\mathbb{1}_A + f(Y_1, \dots, Y_n)\mathbb{1}_{A^c}) \sim H$. We notice that

$$f(X_1, \dots, X_n)\mathbb{1}_A + f(Y_1, \dots, Y_n)\mathbb{1}_{A^c} = f\left(X_1\mathbb{1}_A + Y_1\mathbb{1}_{A^c}, \dots, X_n\mathbb{1}_A + Y_n\mathbb{1}_{A^c}\right),$$

[4] A function $f$ is symmetric if $f(\mathbf{x}) = f(\pi(\mathbf{x}))$ for any $\mathbf{x} \in \mathbb{R}^n$ and $n$-permutation $\pi$.

and $(X_i\mathbb{1}_A + Y_i\mathbb{1}_{A^c}) \sim \lambda F_i + (1-\lambda)G_i$ for any $i = 1, \ldots, n$. Thus we have $H \in \mathcal{D}_f(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$. This completes the proof of (ii). $\square$

We briefly fix some notation and convention. Let $\Delta_n$ be the standard simplex given by $\Delta_n = \{(\lambda_1, \ldots, \lambda_n) \in [0,1]^n : \sum_{i=1}^n \lambda_i = 1\}$. Recall that a doubly stochastic matrix is a square matrix of nonnegative real numbers, each of whose rows and columns sums to 1 (i.e. each row or column is in $\Delta_n$). Denote by $\mathcal{Q}_n$ the set of $n \times n$ doubly stochastic matrices. All vectors should be treated as column vectors. For $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \Delta_n$ and $\mathbf{F} = (F_1, \ldots, F_n) \in \mathcal{M}^n$, their dot product is $\boldsymbol{\lambda} \cdot \mathbf{F} = \sum_{i=1}^n \lambda_i F_i \in \mathcal{M}$. For a matrix $\Lambda = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)^\top \in \mathcal{Q}_n$ and $\mathbf{F} \in \mathcal{M}^n$, their product is $\Lambda\mathbf{F} = (\boldsymbol{\lambda}_1 \cdot \mathbf{F}, \ldots, \boldsymbol{\lambda}_n \cdot \mathbf{F}) \in \mathcal{M}^n$.

The vector $\Lambda\mathbf{F}$ is a distribution mixture of $\mathbf{F}$, and we will call it the $\Lambda$-*mixture* of $\mathbf{F}$ to emphasize the reliance on $\Lambda$. Indeed, $\Lambda\mathbf{F}$ can be seen as a vector of weighted averages of $\mathbf{F}$. In particular, by choosing $\Lambda = (\frac{1}{n})_{n\times n}$ (here $(x)_{n\times n}$ means an $n \times n$ matrix with identical number $x \in \mathbb{R}$), we get the vector $(F, \ldots, F)$ where $F$ is the average of components of $\mathbf{F}$. Note that if $\mathbf{F} \in \mathcal{M}_1^n$, then the mean of any element of $\mathcal{D}_n(\mathbf{F})$ is the same as that of $\mathcal{D}_n(\Lambda\mathbf{F})$.

The first result below suggests that the set of aggregation for a tuple of distributions is smaller than that for the weighted averages. The proof is elementary, but the result allows us to observe the important phenomenon that *more homogeneous marginals lead to a larger aggregation set.*

**Theorem 2.2.** *For an $n$-symmetric function $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{F} \in \mathcal{M}^n$ and $\Lambda \in \mathcal{Q}_n$, $\mathcal{D}_f(\mathbf{F}) \subset \mathcal{D}_f(\Lambda\mathbf{F})$. In particular, $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\Lambda\mathbf{F})$.*

*Proof.* Let $\Pi_1, \ldots, \Pi_{n!}$ be all different $n$-permutation matrices, i.e. $\Pi_k\mathbf{F}$ is a permutation of $\mathbf{F}$. By Birkhoff's Theorem (Theorem 2.A.2 of Marshall et al. (2011)), the set $\mathcal{Q}_n$ of doubly stochastic matrices is the convex hull of permutation matrices, that is, for any $\Lambda \in \mathcal{Q}_n$, there exists $(\lambda_1, \ldots, \lambda_{n!}) \in \Delta_{n!}$, such that

$$\Lambda = \sum_{k=1}^{n!} \lambda_k \Pi_k.$$

Note that $\mathcal{D}_f(\mathbf{F}) = \mathcal{D}_f(\Pi_k\mathbf{F})$ for $k = 1, \ldots, n!$ by Lemma 2.1(i). Further, by Lemma 2.1(ii-b), we have,

$$\mathcal{D}_f(\mathbf{F}) = \bigcap_{k=1}^{n!} \mathcal{D}_f(\Pi_k\mathbf{F}) \subset \mathcal{D}_f\left(\sum_{k=1}^{n!} \lambda_k \Pi_k(\mathbf{F})\right) = \mathcal{D}_f(\Lambda\mathbf{F}).$$

This completes the theorem. $\square$

As the sum aggregation is the most common in financial applications, we will mainly discuss $\mathcal{D}_n$ instead of $\mathcal{D}_f$ in the following context, while keeping in mind that most results on $\mathcal{D}_n$ can be extended naturally to $\mathcal{D}_f$.

**Corollary 2.3.** *For $\mathbf{F} = (F_1, \ldots, F_n) \in \mathcal{M}^n$ and $\Lambda \in \mathcal{Q}_n$, $\mathcal{D}_n(\Lambda\mathbf{F}) \subset \mathcal{D}_n(F, \ldots, F)$ where $F = \frac{1}{n}\sum_{i=1}^n F_i$.*

By taking $\Lambda$ as the identity in Corollary 2.3, we obtain the set inclusion $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(F, \dots, F)$, which was given in Theorem 3.5 of Bernard et al. (2014) to find the bounds on VaR for heterogeneous marginal distributions.

The doubly stochastic matrices are closely related to majorization order. For $\boldsymbol{\lambda}, \boldsymbol{\gamma} \in \mathbb{R}^n$, we say that $\boldsymbol{\lambda}$ dominates $\boldsymbol{\gamma}$ in *majorization order*, denoted by $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$, if $\sum_{i=1}^n \phi(\gamma_i) \leqslant \sum_{i=1}^n \phi(\lambda_i)$ for all continuous convex functions $\phi$. There are several equivalent conditions for this order; see Section 1.A.3 of Marshall et al. (2011). One equivalent condition that is relevant to Theorem 2.2 is that $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$ if and only if there exists $\Lambda \in \mathcal{Q}_n$ such that $\boldsymbol{\gamma} = \Lambda \boldsymbol{\lambda}$. We can similarly define majorization order between $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$, denoted by $\mathbf{G} \prec \mathbf{F}$, if $\mathbf{G} = \Lambda \mathbf{F}$ for some $\Lambda \in \mathcal{Q}_n$. Then, we have the following corollary.

**Corollary 2.4.** *For $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$, if $\mathbf{G} \prec \mathbf{F}$, then $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$.*

**Example 2.5** (Bernoulli distributions)**.** We apply Theorem 2.2 to Bernoulli distributions. Let $B_p$ be a Bernoulli cdf with (mean) parameter $p \in [0, 1]$. Note that a mixture of Bernoulli distributions is still Bernoulli, and more precisely, for $\mathbf{p} = (p_1, \dots, p_n) \in [0, 1]^n$ and $\mathbf{q} = (q_1, \dots, q_n) = \Lambda \mathbf{p}$, we have $\Lambda(B_{p_1}, \dots, B_{p_n}) = (B_{q_1}, \dots, B_{q_n})$. Therefore, by Theorem 2.2, for any $\mathbf{p}, \mathbf{q} \in [0, 1]^n$ with $\mathbf{q} \prec \mathbf{p}$, we have $\mathcal{D}_n(B_{p_1}, \dots, B_{p_n}) \subset \mathcal{D}_n(B_{q_1}, \dots, B_{q_n})$. This result will be used later to discuss joint mixability (see Section 8) of Bernoulli distributions. For instance, we can set $\mathbf{p} = (0.2, 0.8)$,

$$\Lambda = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \text{ and } \mathbf{q} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix} (0.2, 0.8) = (0.65, 0.35).$$

Note that $\Lambda(B_{0.2}, B_{0.8}) = (B_{0.65}, B_{0.35})$. Hence $\mathcal{D}_n(B_{0.2}, B_{0.8}) \subset \mathcal{D}_n(B_{0.65}, B_{0.35})$.

Next, we discuss how $\Lambda$-mixtures affect the lower sets with respect to convex order. A distribution $F \in \mathcal{M}_1$ is called smaller than a distribution $G \in \mathcal{M}_1$ in *convex order*, denoted by $F \prec_{\mathrm{cx}} G$, if

$$\int \phi \, \mathrm{d}F \leqslant \int \phi \, \mathrm{d}G \quad \text{for all convex} \quad \phi : \mathbb{R} \to \mathbb{R}, \tag{2}$$

provided that both integrals exist (finite or infinite); see Müller and Stoyan (2002) and Shaked and Shanthikumar (2007) for an overview on convex order and the related notion of second-order stochastic dominance. For a given distribution $F \in \mathcal{M}_1$, denote by $\mathcal{C}(F)$ the set of all distributions in $\mathcal{M}_1$ dominated by $F$ in convex order, that is,

$$\mathcal{C}(F) = \{G \in \mathcal{M}_1 : G \prec_{\mathrm{cx}} F\}.$$

For any distributions $F$ and $G$, we denote by $F \oplus G$ the distribution with quantile function $F^{-1} + G^{-1}$.[5] Moreover, define

$$\mathcal{C}(F_1, \dots, F_n) = \mathcal{C}(F_1 \oplus \dots \oplus F_n).$$

[5] In other words, $F \oplus G$ is the distribution of the sum of two comonotonic random variables with respective distributions $F$ and $G$. Two random variables $X$ and $Y$ are said to be *comonotonic*, if there exists a random variable $U$ and two increasing functions $f, g$ such that $X = f(U)$ and $Y = g(U)$ almost surely. Such $U$ can be chosen as a standard uniform random variable ($U \sim \mathrm{U}[0, 1]$), and $f$ and $g$ can be chosen as the inverse distribution functions of $X$ and $Y$, respectively.

The following lemmas give a simple link between the sets $\mathcal{D}_n$ and $\mathcal{C}$; see e.g., Lemma 1 of Mao et al. (2019).

**Lemma 2.6.** *For* $\mathbf{F} \in \mathcal{M}_1^n$, $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{C}(\mathbf{F})$.

Similarly to the set $\mathcal{D}_n(\mathbf{F})$ in Theorem 2.2, $\mathcal{C}(\mathbf{F})$ also satisfies an order with respect to $\Lambda$-mixture.

**Theorem 2.7.** *For* $\mathbf{F} \in \mathcal{M}_1^n$ *and* $\Lambda \in \mathcal{Q}_n$, *we have* $\mathcal{C}(\mathbf{F}) \subset \mathcal{C}(\Lambda\mathbf{F})$.

*Proof.* Note that $F_1 \oplus \cdots \oplus F_n \in \mathcal{D}_n(\mathbf{F})$ since $F_1 \oplus \cdots \oplus F_n$ corresponds to the sum of comonotonic random variables with respective distributions $F_1, \dots, F_n$. Using Theorem 2.2 and Lemma 2.6, we have $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\Lambda\mathbf{F}) \subset \mathcal{C}(\Lambda\mathbf{F})$. This implies $F_1 \oplus \cdots \oplus F_n \in \mathcal{C}(\Lambda\mathbf{F})$. By definition, $\mathcal{C}(\mathbf{F}) \subset \mathcal{C}(\Lambda\mathbf{F})$. $\square$

# 3 Quantile mixtures

In Section 2, we have seen a set inclusion between $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\mathbf{G})$ where $\mathbf{G}$ is a distribution mixture of $\mathbf{F}$. The general message from Theorem 2.2 is that distribution mixtures enlarge the aggregation sets. As distribution mixture corresponds to the arithmetic average of distribution functions, it would then be of interest to see whether a "harmonic average" of $F_1, \dots, F_n$ would give similar properties. By saying "harmonic average" of $F_1, \dots, F_n$, we mean the distribution $F$ with $F^{-1} = \frac{1}{n}\sum_{i=1}^n F_i^{-1}$, i.e., the average of quantiles. We shall call this type of average as *quantile mixture*.

In many statistical applications, marginal distributions of a multi-dimensional object are modelled in the same location-scale family (such as Gaussian, elliptical, or uniform family). The quantile mixture of such distributions is still in the same family, whereas the distribution mixture is typically no longer in the family. Moreover, a quantile mixture also corresponds to the combination of comonotonic random variables (such as combining an asset price with a call option on it), and hence finds its natural position in finance. As such, it is rather important and practical to consider quantile mixtures.

*Remark* 3.1. The two types of mixtures are both basic operations on distributions and often lead to qualitatively very different mathematical results. As a famous example in decision theory, the axiom of linearity on distribution mixtures leads to the classic von Neumann-Morgenstern expected utility theory, whereas the axiom of linearity on quantile mixtures leads to the dual utility theory of Yaari (1987).

For a matrix $\Lambda$ of non-negative elements (not necessarily in $\mathcal{Q}_n$) and $\mathbf{F} \in \mathcal{M}^n$, let $\Lambda \otimes \mathbf{F}$ be a vector of distributions $\mathbf{G}$ such that componentwise, $\mathbf{G}^{-1}$ is equal to $\Lambda\mathbf{F}^{-1}$. If $\Lambda \in \mathcal{Q}_n$, we call $\mathbf{G} = \Lambda \otimes \mathbf{F}$ the $\Lambda$-*quantile mixture* of $\mathbf{F}$. If $\mathbf{F} \in \mathcal{M}_1^n$, then the mean of any element of $\mathcal{D}_n(\mathbf{F})$ is the same as that of $\mathcal{D}_n(\Lambda \otimes \mathbf{F})$, similarly to the case of distribution mixture. This suggests that one may compare $\mathcal{D}_n(\mathbf{F})$ with $\mathcal{D}_n(\Lambda \otimes \mathbf{F})$, just like what we did in Section 2 for distribution mixture.

The first natural candidates for us to look at are $\mathcal{D}_n(F_1, \dots, F_n)$ and $\mathcal{D}_n(F, \dots, F)$ where $F^{-1} = \frac{1}{n}\sum_{i=1}^n F_i^{-1}$, thus the quantile version of Corollary 2.3. Unfortunately, the sets $\mathcal{D}_n(F_1, \dots, F_n)$ and $\mathcal{D}_n(F, \dots, F)$ are not necessarily comparable, as seen from the following example.

**Example 3.2.** Take $F_1$ as a binary uniform distribution (with probability 1/2 at each point) on $\{0,1\}$ and $F_2$ as a binary uniform distribution on $\{0,3\}$. Clearly, $F$ is a binary uniform distribution on $\{0,2\}$. $\mathcal{D}_2(F_1,F_2)$ contains distributions supported on $\{0,1,3,4\}$ and $\mathcal{D}_2(F,F)$ contains distributions supported on $\{0,2,4\}$. Therefore, these two sets do not have a relation of set inclusion.

On the other hand, as a trivial example, if $F_2,\dots,F_n$ are point masses (without loss of generality, we assume that they are point masses at 0), then $F$ satisfies $F^{-1}=F_1^{-1}/n$. In this case, $\mathcal{D}_n(F_1,\dots,F_n)=\{F_1\}\subset\mathcal{D}_n(F,\dots,F)$ holds trivially. Therefore, we can expect that the inclusion $\mathcal{D}_n(F_1,\dots,F_n)\subset\mathcal{D}_n(F,\dots,F)$ may hold under some special settings.

Below, we note that both $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\Lambda\otimes\mathbf{F})$ have the same convex-order maximal element. This is in sharp contrast to the case of mixtures in Theorem 2.7. Proposition 3.3 can be verified directly by definition.

**Proposition 3.3.** *For $\mathbf{F}\in\mathcal{M}_1^n$ and $\Lambda\in\mathcal{Q}_n$, we have $\mathcal{C}(\mathbf{F})=\mathcal{C}(\Lambda\otimes\mathbf{F})$.*

As we see from Example 3.2, $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\Lambda\otimes\mathbf{F})$ are not necessarily comparable. In Mao et al. (2019), a non-trivial result is established for the aggregation of standard uniform distributions, which leads to an interesting observation along this direction.

**Proposition 3.4.** *Suppose that $F_1,\dots,F_n$ are uniform distributions, $n\geqslant 3$, and $\Lambda=(\frac{1}{n})_{n\times n}$. Then $\mathcal{D}_n(\mathbf{F})\subset\mathcal{D}_n(\Lambda\otimes\mathbf{F})$.*

*Proof.* Note that the components of $\Lambda\otimes\mathbf{F}$ are uniform distributions with equal length. By Theorem 5 of Mao et al. (2019), we have $\mathcal{D}_n(\Lambda\otimes\mathbf{F})=\mathcal{C}_n(\Lambda\otimes\mathbf{F})$. Using Proposition 3.3, we have $\mathcal{C}_n(\mathbf{F})=\mathcal{C}_n(\Lambda\otimes\mathbf{F})$. Lemma 2.6 further yields $\mathcal{D}_n(\mathbf{F})\subset\mathcal{C}_n(\mathbf{F})$. Putting the above results together, we obtain $\mathcal{D}_n(\mathbf{F})\subset\mathcal{D}_n(\Lambda\otimes\mathbf{F})$. □

It is unclear whether $\mathcal{D}_n(\mathbf{F})\subset\mathcal{D}_n(\Lambda\otimes\mathbf{F})$ under some other conditions, similarly to Proposition 3.4. Note that the set inclusion $\mathcal{D}_n(\mathbf{F})\subset\mathcal{D}_n(\Lambda\otimes\mathbf{F})$ would help us to obtain semi-explicit formulas for bounds on risk measures (such as VaR), since by choosing $\Lambda=(\frac{1}{n})_{n\times n}$, the marginal distributions of $\Lambda\otimes\mathbf{F}$ are the same, and formulas for VaR bounds in e.g., Wang et al. (2013) and Bernard et al. (2014) are applicable; see Section 4.

There are several sharp contrasts regarding distribution and quantile mixtures. In addition to the contrast on order relations that we see from Theorem 2.2 and Example 3.2, the two notions also treat location shifts on the marginal distributions very differently. This point will be explained in Section 8.1.

## 4 Bounds on the worst-case values of risk measures

This section is dedicated to exploring the inequalities between the worst-cases value of risk measures in risk aggregation with different marginal distribution tuples. Our main results in Sections 2 and 3 will help to find the inequalities in Proposition 4.3.

### 4.1 Risk measures

We pay a particular attention to the popular regulatory risk measure VaR, which is a quantile functional. For $F \in \mathcal{M}$, for $p \in (0,1)$, define the risk measure $\mathrm{VaR}_p : \mathcal{M} \to \mathbb{R}$ as

$$\mathrm{VaR}_p(F) = F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geqslant p\}.$$

Another popular regulatory risk measure is $\mathrm{ES}_p : \mathcal{M}_1 \to \mathbb{R}$ for $p \in (0,1)$, given by

$$\mathrm{ES}_p(F) = \frac{1}{1-p}\int_p^1 F^{-1}(u)\mathrm{d}u.$$

Given marginals $\mathbf{F}$, the worst-case value of VaR in risk aggregation with unknown dependence structure is then defined as

$$\overline{\mathrm{VaR}}_p(\mathbf{F}) = \sup\{\mathrm{VaR}_p(G) : G \in \mathcal{D}_n(\mathbf{F})\}.$$

In other words, $\overline{\mathrm{VaR}}_p(\mathbf{F})$ is the largest value of $\mathrm{VaR}_p$ of the aggregate risk $X_1 + \cdots + X_n$ over all possible dependence structures among $X_i \sim F_i,\ i = 1, \ldots, n$. Similarly, the worst-case value of ES in risk aggregation is defined as $\overline{\mathrm{ES}}_p(\mathbf{F}) = \sup\{\mathrm{ES}_p(G) : G \in \mathcal{D}_n(\mathbf{F})\}$.

The worst-case value of ES in risk aggregation is easy to calculate since ES is consistent with convex order. On the other hand, worst-case value of VaR in risk aggregation generally does not admit any analytical formula, which is a challenging problem; results under some specific cases are given in Wang et al. (2013), Puccetti and Rüschendorf (2013) and Bernard et al. (2014). To obtain approximations for $\overline{\mathrm{VaR}}_p(\mathbf{F})$, one may use the asymptotic equivalence between VaR and ES in Embrechts et al. (2015) and then directly apply ES bounds, or use a numerical algorithm such as the rearrangement algorithm of Puccetti and Rüschendorf (2012) and Embrechts et al. (2013).

We will discuss a general relationship on risk measures for different aggregation sets. A *risk measure* is a functional $\rho : \mathcal{M}_\rho \to \mathbb{R}$, where $\mathcal{M}_\rho \subset \mathcal{M}$ is the set of distributions of some financial losses. For instance, if $\rho$ is the mean, then $\mathcal{M}_\rho$ is naturally chosen as the set of distributions with finite mean. We denote by $\overline{\rho}(\mathbf{F})$ the worst-case value of $\rho$ in risk aggregation for $\mathbf{F} \in \mathcal{M}^n$, that is, assuming $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{M}_\rho$,

$$\overline{\rho}(\mathbf{F}) = \sup\{\rho(G) : G \in \mathcal{D}_n(\mathbf{F})\}.$$

### 4.2 Inequalities implied by stochastic dominance

Quite obviously, one can compare the worst-case values of some risk measures for two tuples of distributions satisfying some stochastic dominance, which we briefly discuss here.

A distribution $F \in \mathcal{M}$ is smaller than a distribution $G$ in *stochastic order* (also first-order stochastic dominance), denoted by $F \prec_{\mathrm{st}} G$, if $F \geqslant G$. For $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$, we say that $\mathbf{F}$ is smaller than $\mathbf{G}$ in stochastic order, denoted by $\mathbf{F} \prec_{\text{st}} \mathbf{G}$, if $F_i \prec_{\text{st}} G_i$, $i = 1, \dots, n$. Analogously, for $\mathbf{F}, \mathbf{G} \in \mathcal{M}_1^n$, we say that $\mathbf{F}$ is smaller than $\mathbf{G}$ in convex order, denoted by $\mathbf{F} \prec_{\text{cx}} \mathbf{G}$, if $F_i \prec_{\text{cx}} G_i$, $i = 1, \dots, n$.

We define two relevant common properties of risk measures. A risk measure $\rho$ is *monotone* if $\rho(F) \leqslant \rho(G)$ whenever $F \prec_{\text{st}} G$; it is *consistent with convex order* if $\rho(F) \leqslant \rho(G)$ whenever $F \prec_{\text{cx}} G$. Almost all risk measures used in practice are monotone; ES is consistent with convex order whereas VaR is not. Monetary risk measures (see Föllmer and Schied (2016)) that are consistent with convex order are characterized by Mao and Wang (2020) and they admit an ES-based representation. In particular, all lower semi-continuous convex risk measures, including ES and expectiles (e.g., Ziegel (2016) and Delbaen et al. (2016)), are consistent with convex order; we refer to Föllmer and Schied (2016) for an overview on risk measures.

Now we state in Proposition 4.1 that one can compare the worst-case values of some risk measures for $\mathbf{F}$ and $\mathbf{G}$ if $\mathbf{F}$ is smaller than $\mathbf{G}$ in stochastic order or convex order.

**Proposition 4.1.** *Let $\rho$ be a risk measure and $\mathbf{F}, \mathbf{G} \in \mathcal{M}^n$ with $\mathcal{D}_n(\mathbf{F}), \mathcal{D}_n(\mathbf{G}) \subset \mathcal{M}_\rho$.*

*(i) If $\rho$ is monotone and $\mathbf{F} \prec_{\text{st}} \mathbf{G}$, then $\overline{\rho}(\mathbf{F}) \leqslant \overline{\rho}(\mathbf{G})$.*

*(ii) If $\rho$ is consistent with convex order and $\mathbf{F} \prec_{\text{cx}} \mathbf{G}$ with $\mathbf{F}, \mathbf{G} \in \mathcal{M}_1^n$, then $\overline{\rho}(\mathbf{F}) \leqslant \overline{\rho}(\mathbf{G})$.*

*Proof.* (i) is straightforward to verify. We next focus on (ii). Since $F_1 \oplus \cdots \oplus F_n$ is the largest distribution in $D_n(\mathbf{F})$ with respect to convex order and $\rho$ is consistent with convex order, we have $\overline{\rho}(\mathbf{F}) = \rho(F_1 \oplus \cdots \oplus F_n)$. Similarly, $\overline{\rho}(\mathbf{G}) = \rho(G_1 \oplus \cdots \oplus G_n)$. Note that $\mathbf{F} \prec_{\text{cx}} \mathbf{G}$ means $F_i \prec_{\text{cx}} G_i$, $i = 1, \dots, n$. For all $p \in (0,1)$, using comonotonic-additivity of $\mathrm{ES}_p$, we have

$$\mathrm{ES}_p(F_1 \oplus \cdots \oplus F_n) = \sum_{i=1}^n \mathrm{ES}_p(F_i) \leqslant \sum_{i=1}^n \mathrm{ES}_p(G_i) = \mathrm{ES}_p(G_1 \oplus \cdots \oplus G_n),$$

which gives $F_1 \oplus \cdots \oplus F_n \prec_{\text{cx}} G_1 \oplus \cdots \oplus G_n$ (see e.g., Theorem 3.A.5 of Shaked and Shanthikumar (2007)). $\square$

In the following result, we will show that the distribution tuples and their $\Lambda$-mixture or $\Lambda$-quantile mixture typically do not satisfy stochastic order or convex order, unless the mixture operation is essentially identical ($\Lambda \mathbf{F} = \mathbf{F}$ or $\Lambda \otimes \mathbf{F} = \mathbf{F}$). The proof of Proposition 4.2 is put in Appendix A.2.

**Proposition 4.2.** *Suppose $\Lambda \in \mathcal{Q}_n$. The statements within each of (i)-(iv) are equivalent.*

*(i) For $\mathbf{F} \in \mathcal{M}^n$, (a) $\Lambda \mathbf{F} \prec_{\text{st}} \mathbf{F}$; (b) $\mathbf{F} \prec_{\text{st}} \Lambda \mathbf{F}$; (c) $\Lambda \mathbf{F} = \mathbf{F}$.*

*(ii) For $\mathbf{F} \in \mathcal{M}^n$, (a) $\Lambda \otimes \mathbf{F} \prec_{\text{st}} \mathbf{F}$; (b) $\mathbf{F} \prec_{\text{st}} \Lambda \otimes \mathbf{F}$; (c) $\Lambda \otimes \mathbf{F} = \mathbf{F}$.*

*(iii) For $\mathbf{F} \in \mathcal{M}_1^n$, (a) $\Lambda \otimes \mathbf{F} \prec_{\text{cx}} \mathbf{F}$; (b) $\mathbf{F} \prec_{\text{cx}} \Lambda \otimes \mathbf{F}$; (c) $\Lambda \otimes \mathbf{F} = \mathbf{F}$.*

*(iv) For $\mathbf{F} \in \mathcal{M}_1^n$, (a) $\Lambda \mathbf{F} \prec_{\text{cx}} \mathbf{F}$; (b) $\Lambda \mathbf{F} = \mathbf{F}$.*

An implication of Proposition 4.2 is that the result on stochastic order in Proposition 4.1 cannot be applied to compare the worst-case values of risk measures for $\mathbf{F}$ and $\Lambda \mathbf{F}$ or $\mathbf{F}$ and $\Lambda \otimes \mathbf{F}$. Nevertheless, this comparison can be conducted by applying our findings in Sections 2 and 3 and some other techniques. This will be the task in the next subsection.

## 4.3 Inequalities generated by distribution/quantile mixtures

In the following, we will obtain inequalities between the worst-case values of risk measures for $\mathbf{F}$ and $\Lambda\mathbf{F}$ or $\mathbf{F}$ and $\Lambda \otimes \mathbf{F}$. First, we apply Theorem 2.2 and Proposition 3.3 and immediately obtain the following result.

**Proposition 4.3.** *Let $\rho$ be a risk measure and $\Lambda \in \mathcal{Q}_n$.*

*(i) For $\mathbf{F} \in \mathcal{M}^n$ with $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{M}_\rho$ and $\mathcal{D}_n(\Lambda\mathbf{F}) \subset \mathcal{M}_\rho$, we have $\overline{\rho}(\mathbf{F}) \leqslant \overline{\rho}(\Lambda\mathbf{F})$;*

*(ii) For $\mathbf{F} \in \mathcal{M}_1^n$ with $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{M}_\rho$ and $\mathcal{D}_n(\Lambda \otimes \mathbf{F}) \subset \mathcal{M}_\rho$, if $\rho$ is consistent with convex order, then $\overline{\rho}(\mathbf{F}) = \overline{\rho}(\Lambda \otimes \mathbf{F}) = \rho(F_1 \oplus \cdots \oplus F_n)$.*

Note that in Proposition 4.3, the inequality for distribution mixture is valid for all risk measures whereas the equality for quantile mixture is constrained to risk measures consistent with convex order. As $\mathrm{ES}_p$ is a special case of risk measures consistent with convex order, we immediately get $\overline{\mathrm{ES}}_p(\mathbf{F}) \leqslant \overline{\mathrm{ES}}_p(\Lambda\mathbf{F})$ and $\overline{\mathrm{ES}}_p(\mathbf{F}) = \overline{\mathrm{ES}}_p(\Lambda \otimes \mathbf{F})$. Since VaR is not consistent with convex order, (ii) of Proposition 4.3 cannot be applied to VaR. Nevertheless, using a recent result on $\overline{\mathrm{VaR}}$ in Blanchet et al. (2020), we obtain an inequality between $\overline{\mathrm{VaR}}$ for some special marginals and $\overline{\mathrm{VaR}}$ of their corresponding quantile mixture. Denote by $\mathcal{M}_D$ (respectively, $\mathcal{M}_I$) the set of distributions with decreasing (respectively, increasing) densities on their support. Moreover, let $\mathcal{M}_D^n = (\mathcal{M}_D)^n$ and $\mathcal{M}_I^n = (\mathcal{M}_I)^n$.

**Theorem 4.4.** *For $p \in (0,1)$, $\Lambda \in \mathcal{Q}_n$, and $\mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$, we have*

$$\overline{\mathrm{VaR}}_p(\mathbf{F}) \leqslant \overline{\mathrm{VaR}}_p(\Lambda \otimes \mathbf{F}).$$

*Proof.* We start with some preliminaries. Define the upper VaR at level $p$ for a cdf $F$ as

$$\mathrm{VaR}_p^*(F) = \inf\{x \in \mathbb{R} : F(x) > p\}, \quad p \in (0,1).$$

The worst-case value of the upper VaR in risk aggregation is $\overline{\mathrm{VaR}}_p^*(\mathbf{F}) = \sup\{\mathrm{VaR}_p^*(G) : G \in \mathcal{D}_n(\mathbf{F})\}$. For $\mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$ and $p \in (0,1)$, Lemma 4.5 of Bernard et al. (2014) gives

$$\overline{\mathrm{VaR}}_p^*(\mathbf{F}) = \overline{\mathrm{VaR}}_p(\mathbf{F}).$$

Using Lemma A.1 in Appendix A.1 (paraphrased from Theorem 2 of Blanchet et al. (2020)), we have

$$\overline{\mathrm{VaR}}_p(\mathbf{F}) = \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^n \frac{1}{(1-p)(1-\beta)} \int_{p+(1-p)(\beta-\beta_i)}^{1-(1-p)\beta_i} \mathrm{VaR}_u(F_i)\mathrm{d}u, \tag{3}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$, $\beta = \sum_{i=1}^n \beta_i$ and $\mathbb{B}_n = \{\boldsymbol{\beta} \in [0,1)^n : \beta < 1\}$. Note that

$$\Lambda \otimes \mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n \;\text{ if }\; \mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n.$$

Consequently, for $p \in (0,1)$,

$$
\begin{aligned}
\overline{\mathrm{VaR}}_p(\Lambda \otimes \mathbf{F}) &= \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^{n} \frac{1}{(1-p)(1-\beta)} \int_{p+(1-p)(\beta-\beta_i)}^{1-(1-p)\beta_i} \left( \sum_{j=1}^{n} \Lambda_{i,j} \mathrm{VaR}_u(F_j) \right) \mathrm{d}u \\
&= \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \Lambda_{i,j} M_{i,j}(\boldsymbol{\beta}),
\end{aligned}
$$

where the function $M : \mathbb{B}_n \to \mathbb{R}^{n\times n}$, mapping an $n$-dimensional vector to an $n \times n$ matrix, is given by

$$
M_{i,j}(\boldsymbol{\beta}) = \frac{1}{(1-p)(1-\beta)} \int_{p+(1-p)(\beta-\beta_i)}^{1-(1-p)\beta_i} \mathrm{VaR}_u(F_j)\mathrm{d}u, \quad i,j = 1,\dots,n.
$$

We can rewrite (3) as

$$
\overline{\mathrm{VaR}}_p(\mathbf{F}) = \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^{n} M_{i,i}(\boldsymbol{\beta}).
$$

Let $\Pi_1, \dots, \Pi_{n!}$ be all different $n$-permutation matrices, i.e., $\Pi_k \boldsymbol{\beta}$ is a permutation of $\boldsymbol{\beta}$ for each $k$. By Birkhoff's Theorem (Theorem 2.A.2 of Marshall et al. (2011)), for $\Lambda \in \mathcal{Q}_n$, there exists $(\lambda_1, \dots, \lambda_{n!}) \in \Delta_{n!}$ such that $\Lambda = \sum_{k=1}^{n!} \lambda_k \Pi_k$. Hence, by writing $\Pi_k \boldsymbol{\beta} = (\beta_1^k, \dots, \beta_n^k)$ for each $k$, we have

$$
\begin{aligned}
\sum_{i=1}^{n} \sum_{j=1}^{n} \Lambda_{i,j} M_{i,j}(\boldsymbol{\beta}) &= \frac{1}{(1-p)(1-\beta)} \sum_{i=1}^{n} \int_{p+(1-p)(\beta-\beta_i)}^{1-(1-p)\beta_i} \left( \sum_{j=1}^{n} \Lambda_{i,j} \mathrm{VaR}_u(F_j) \right) \mathrm{d}u \\
&= \frac{1}{(1-p)(1-\beta)} \sum_{i=1}^{n} \sum_{k=1}^{n!} \lambda_k \int_{p+(1-p)(\beta-\beta_i^k)}^{1-(1-p)\beta_i^k} \mathrm{VaR}_u(F_i)\mathrm{d}u \\
&= \sum_{k=1}^{n!} \lambda_k \sum_{i=1}^{n} \frac{1}{(1-p)(1-\beta)} \int_{p+(1-p)(\beta-\beta_i^k)}^{1-(1-p)\beta_i^k} \mathrm{VaR}_u(F_i)\mathrm{d}u \\
&= \sum_{k=1}^{n!} \lambda_k \sum_{i=1}^{n} M_{i,i}(\Pi_k \boldsymbol{\beta}).
\end{aligned}
$$

Using the above facts, we finally obtain

$$
\begin{aligned}
\overline{\mathrm{VaR}}_p(\mathbf{F}) = \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^{n} M_{i,i}(\boldsymbol{\beta}) &= \sum_{k=1}^{n!} \lambda_k \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^{n} M_{i,i}(\Pi_k \boldsymbol{\beta}) \\
&\leqslant \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{k=1}^{n!} \lambda_k \sum_{i=1}^{n} M_{i,i}(\Pi_k \boldsymbol{\beta}) = \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \Lambda_{i,j} M_{i,j}(\boldsymbol{\beta}) = \overline{\mathrm{VaR}}_p(\Lambda \otimes \mathbf{F}).
\end{aligned}
$$

This completes the proof of the theorem. □

The restriction of marginals to distributions with monotone densities in Theorem 4.4 is because of applying Lemma A.1. This assumption is common in the literature of VaR bounds (e.g., Wang et al. (2013)). We may expect Theorem 4.4 to hold for more general classes of $\mathbf{F}$; this is supported by the numerical results in Figure 4. Moreover, for $\Lambda \in \mathcal{Q}_n$ and and $\mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$, we may expect $\overline{\rho}(\mathbf{F}) \leqslant \overline{\rho}(\Lambda \otimes \mathbf{F})$ for other risk measures $\rho$ than VaR (Theorem 4.4) and those consistent with convex order (Proposition 4.3). Unfortunately, we are unable to prove the above statements in general. Some related open questions are listed in Section 9.

*Remark* 4.5. The condition $\mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$ in Theorem 4.4 can be relaxed to that the $p$-tail distributions of $F_1, \dots, F_n$ are all in $\mathcal{M}_D$ or all in $\mathcal{M}_I$.[6] This should be clear since only the $p$-tail distributions are involved in the proof of Theorem 4.4. This condition often holds if $p$ is close to 1, and it allows for Theorem 4.4 to be applied to many common distributions in risk management.

Next, we study location-scale distribution families. Let $T_x(F)$ be a shift of $F \in \mathcal{M}$ by adding a constant $x \in \mathbb{R}$ to its location, that is, $T_x(F)$ is the distribution of $X + x$ for $X \sim F$. For $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\mathbf{F} = (F_1, \dots, F_n) \in \mathcal{M}^n$, we use the notation $\mathbf{T_x}(\mathbf{F}) = (T_{x_1}(F_1), \dots, T_{x_n}(F_n))$. Moreover, for $\lambda \geqslant 0$, we denote by $F^\lambda$ the distribution of $\lambda X$ for $X \sim F$ and write $\mathbf{F}^{\boldsymbol{\lambda}} = (F^{\lambda_1}, \dots, F^{\lambda_n})$.

**Corollary 4.6.** *For $p \in (0,1)$, $F \in \mathcal{M}_D \cup \mathcal{M}_I$ , $\boldsymbol{\lambda}, \boldsymbol{\gamma} \in \mathbb{R}_+^n$, and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, if $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$ and $\sum_{i=1}^n x_i \leqslant \sum_{i=1}^n y_i$, then*

$$\overline{\mathrm{VaR}}_p(\mathbf{T_x}(\mathbf{F}^{\boldsymbol{\lambda}})) \leqslant \overline{\mathrm{VaR}}_p(\mathbf{T_y}(\mathbf{F}^{\boldsymbol{\gamma}})). \tag{4}$$

*Proof.* By Section 1.A.3 of Marshall et al. (2011), $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$ if and only if there exists $\Lambda \in \mathcal{Q}_n$ such that $\boldsymbol{\gamma} = \Lambda\boldsymbol{\lambda}$. This implies $\mathbf{F}^{\boldsymbol{\gamma}} = \Lambda \otimes \mathbf{F}^{\boldsymbol{\lambda}}$. By Theorem 4.4, it follows that $\overline{\mathrm{VaR}}_p(\mathbf{F}^{\boldsymbol{\lambda}}) \leqslant \overline{\mathrm{VaR}}_p(\mathbf{F}^{\boldsymbol{\gamma}})$. Moreover, observe that

$$\overline{\mathrm{VaR}}_p(\mathbf{T_x}(\mathbf{F}^{\boldsymbol{\lambda}})) = \overline{\mathrm{VaR}}_p(\mathbf{F}^{\boldsymbol{\lambda}}) + \sum_{i=1}^n x_i \quad \text{and} \quad \overline{\mathrm{VaR}}_p(\mathbf{T_y}(\mathbf{F}^{\boldsymbol{\gamma}})) = \overline{\mathrm{VaR}}_p(\mathbf{F}^{\boldsymbol{\gamma}}) + \sum_{i=1}^n y_i.$$

By the fact that $\sum_{i=1}^n x_i \leqslant \sum_{i=1}^n y_i$, we prove (4). □

# 5 Bounds on risk measures for Pareto risk aggregation

In this section we study the worst-case risk measure for a portfolio of Pareto risks, and the risk measure is not necessarily consistent with convex order. Throughout this section, we assume that $\rho$ is a monotone risk measure, such as VaR.

One particular situation of interest for risk aggregation with non-convex risk measures is when the risks in the portfolio do not have a finite mean. Note that for a portfolio without finite mean, any non-constant risk measure that is consistent with convex order (including convex risk measures) will have an infinite value. Therefore, one has to use a non-convex risk measure such as VaR to assess risks in this situation.

Arguably, the most important class of heavy-tailed risk distributions is the class of Pareto distributions due to their regularly varying tails and their prominent appearance in extreme value theory; see e.g., Embrechts et al. (1997). A common parameterization of Pareto distributions is given by, for $\theta, \alpha > 0$,

$$P_{\alpha,\theta}(x) = 1 - \left(\frac{\theta}{x}\right)^\alpha, \quad x \geqslant \theta.$$

[6] The $p$-tail distribution of $F$ is the distribution of $F^{-1}(U)$ where $U$ is uniform on $[p, 1]$; see e.g., Rockafellar and Uryasev (2002).

Note that if $X \sim P_{\alpha,1}$, then $\theta X \sim P_{\alpha,\theta}$, and thus $\theta$ is a scale parameter. Moreover, the mean of $P_{\alpha,\theta}$ is infinite if and only if $\alpha \in (0,1]$. Limited by the current techniques, we confine ourselves to portfolios of risks with a fixed $\alpha$ and possibly different $\theta$.

For $\alpha > 0$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in (0,\infty)^n$, let $\mathbf{P}_{\alpha,\boldsymbol{\theta}} = (P_{\alpha,\theta_1}, \dots, P_{\alpha,\theta_n})$. We are interested in the worst-case value $\overline{\rho}(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$. We first note some simple properties of the above quantity, which are straightforward to check (a simple proof is put in Appendix A.3).

**Proposition 5.1.** *Let $\rho$ be a monotone risk measure on $\mathcal{M}$. For $\alpha > 0$ and $\boldsymbol{\theta} \in (0,\infty)^n$,*

*(i)* $\Lambda \otimes \mathbf{P}_{\alpha,\boldsymbol{\theta}} = \mathbf{P}_{\alpha,\Lambda\boldsymbol{\theta}}$ *for all* $\Lambda \in (0,\infty)^{n\times n}$*;*

*(ii)* $\overline{\rho}(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ *is decreasing in* $\alpha$*;*

*(iii)* $\overline{\rho}(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ *is increasing in each component of* $\boldsymbol{\theta}$*.*

The next result contains an ordering relationship on the aggregation of Pareto risks. In particular, we show that for $\alpha \in (0,1]$, which means the mean of the distribution is infinite, the quantile mixture leads to an even larger worst-case value of risk aggregation than the distribution mixture (this statement is generally not true for $\alpha > 1$; see the figures in Section 6). This result is not implied by any comparisons obtained in the previous sections, and it seems to be rather specialized for Pareto distributions, as seen from the proof. It is unclear at the moment whether the result can be generalized to other types of distributions without a finite mean.

**Theorem 5.2.** *Let $\rho$ be a monotone risk measure on $\mathcal{M}$. For $\alpha \in (0,1]$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in (0,\infty)^n$, and $\Lambda \in \mathcal{Q}_n$, we have $\overline{\rho}(\mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \overline{\rho}(\Lambda \mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \overline{\rho}(\mathbf{P}_{\alpha,\Lambda\boldsymbol{\theta}})$.*

*Proof.* The first inequality follows directly from Theorem 2.2. Next we focus on the second inequality. Recall that $\Lambda = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n)^\top \in \mathcal{Q}_n$ and let $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \dots, \lambda_{j,n})$ for $j = 1, \dots, n$. For any fixed $j \in \{1, \dots, n\}$, denote the cdf of $(\Lambda \mathbf{P}_{\alpha,\boldsymbol{\theta}})_j$ by $F_j$, then

$$F_j(x) = \sum_{i=1}^n \lambda_{j,i} \left(1 - \left(\frac{\theta_i}{x}\right)^\alpha\right)_+, \quad x \in \mathbb{R}.$$

For some fixed $x > 0$ and $\alpha \in (0,1]$, define $g(t) := 1 - (t/x)^\alpha$, $t \geqslant 0$. Note that $g$ is a convex function on $[0,\infty)$. Hence

$$F_j(x) \geqslant \sum_{i=1}^n \lambda_{j,i} \left(1 - \left(\frac{\theta_i}{x}\right)^\alpha\right) \geqslant 1 - \left(\frac{\sum_{i=1}^n \lambda_{j,i}\theta_i}{x}\right)^\alpha.$$

This implies

$$F_j(x) \geqslant G_j(x),\ x \geqslant 0,$$

where $G_j = (\mathbf{P}_{\alpha,\Lambda\boldsymbol{\theta}})_j$. As $F_j \leqslant_{\mathrm{st}} G_j$ for $j = 1, \dots, n$ and $\rho$ is monotone, by Proposition 4.1(i), we have the second inequality. $\square$

Next, we combine the results of Theorems 4.4-5.2 and Propositions 4.3-5.1 with a special focus on $\mathrm{VaR}_p$, $p \in (0,1)$. The proof is straightforward and omitted.

**Proposition 5.3.** *For $p \in (0,1)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in (0,\infty)^n$, and $\Lambda \in \mathcal{Q}_n$,*

*(i) If $\alpha \in (0,\infty)$, $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \overline{\mathrm{VaR}}_p(\Lambda \mathbf{P}_{\alpha,\boldsymbol{\theta}})$;*

*(ii) If $\alpha \in (0,\infty)$, $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\Lambda\boldsymbol{\theta}})$;*

*(iii) If $\alpha \in (0,1]$, $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \overline{\mathrm{VaR}}_p(\Lambda \mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\Lambda\boldsymbol{\theta}})$.*

Proposition 5.3 is useful for the application in Section 7.2 on multiple hypothesis testing, where $P^r$ follows a Pareto distribution for a p-value $P$ and $r < 0$. Some further properties of $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ are put in Appendix A.4.

# 6 Numerical illustration

Define a $3 \times 3$ doubly stochastic matrix by

$$\Lambda = 0.8 \times I_3 + 0.2 \times \left(\frac{1}{3}\right)_{3\times 3}, \tag{5}$$

where $I_3$ is the $3 \times 3$ identity matrix. In this section, we consider a sequence of doubly stochastic matrices $\{\Lambda^k\}_{k\in\mathbb{N}}$ to numerically illustrate the ordering relationships and inequalities obtained throughout the paper. Note that $\Lambda^k$ is more "homogeneous" as $k$ grows larger, and $\Lambda^k \to (\frac{1}{3})_{3\times 3}$ as $k \to \infty$. The general messages obtained from the numerical examples are listed as follows.

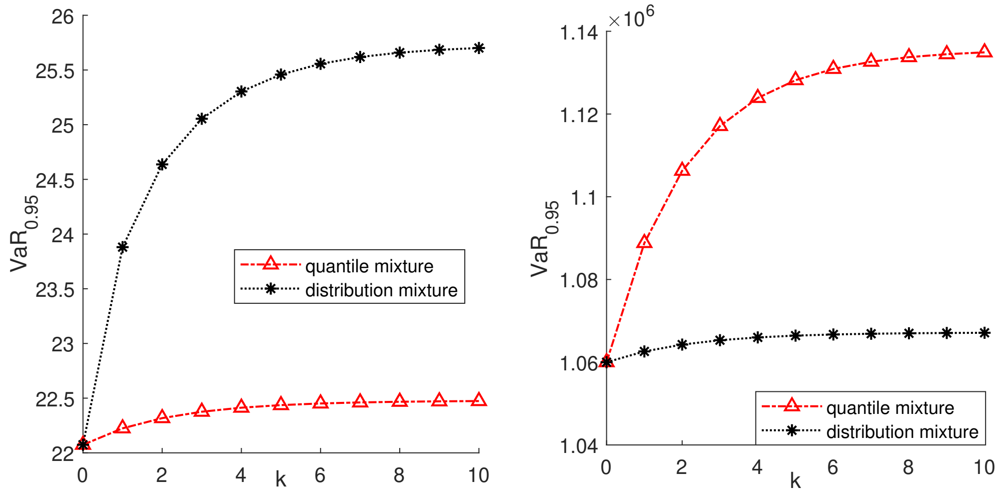
1. For general marginals, the value of $\overline{\mathrm{VaR}}$ becomes larger after making a distribution mixture (Proposition 4.3(i)); this is shown in all figures.

2. For marginals with monotone densities, with a quantile mixture, the value of $\overline{\mathrm{VaR}}$ becomes larger (Theorem 4.4); see Figures 1-3. Numerical examples in Figure 4 indicate that Theorem 4.4 may also hold for marginals with non-monotone densities. Nevertheless, the order does not hold for arbitrary marginals. A counterexample, involving discrete marginals, is provided in Figure 5.

3. For Pareto distributions with infinite mean, the value of $\overline{\mathrm{VaR}}$ of the quantile mixture is larger than that of the distribution mixture (Proposition 5.3(iii)); see Figure 1(b). This relationship does not hold for Pareto distributions with finite mean; see Figure 1(a).

## 6.1 Illustration of theoretical results

In this subsection, we discuss marginals with monotone densities ($\mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$). We have $\Lambda^k \otimes \mathbf{F}, \Lambda^k \mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$. According to Lemma A.1 in Appendix A.1, we obtain a formula (Equation (3)) for $\mathbf{F}$, $\Lambda^k \otimes \mathbf{F}$ and $\Lambda^k \mathbf{F}$ and numerically compute the exact values for $\overline{\mathrm{VaR}}$.




(a) Pareto distribution with finite mean ($\alpha = 3$) (b) Pareto distribution with infinite mean ($\alpha = 1/3$)

Figure 1: Quantile mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \otimes \mathbf{P}_{\alpha,\boldsymbol{\theta}}) = \overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\Lambda^k\boldsymbol{\theta}})$; Distribution mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \mathbf{P}_{\alpha,\boldsymbol{\theta}})$. Setting: $p = 0.95$; $\boldsymbol{\theta} = (1, 2, 3)$, $X_i \sim \mathrm{Pareto}(\alpha, \theta_i)$, $i = 1, 2, 3$; $\Lambda$ is defined by (5); $k = 0, 1, \ldots, 10$.

In Figure 1, we consider Pareto distributions with finite mean ($\alpha = 3$) and infinite mean ($\alpha = 1/3$), respectively. The ordering relationships in Proposition 5.3(i)-(ii) for Pareto distributions with the same $\alpha$ are visualized as the curves in Figure 1 are all increasing in $k$. In Figure 1(b), it turns out that for the case with infinite mean the quantile mixture gives larger value of $\overline{\mathrm{VaR}}$ than that given by the distribution mixture. This coincides with the conclusion in Proposition 5.3(iii). Interestingly, we observe from Figure 1(a) that the value of $\overline{\mathrm{VaR}}$ given by distribution mixture is larger than the one with quantile mixture, which is contrary to the case with infinite mean (Figure 1(b)). It is an open question whether this conclusion is true for general doubly stochastic matrices $\Lambda$ and all $\alpha > 1$.

We next focus on Pareto distributions with different $\alpha$ in Figure 2. First observe that the curves of quantile mixture and distribution mixture in Figure 2 are both increasing in $k$, which is consistent with Theorem 4.4 and Proposition 4.3(i). Comparing the two curves, it is shown that value for the distribution mixture in this case is smaller than the one for quantile mixture.

Heterogeneous distribution families with decreasing densities are considered in Figure 3. As we can see, the curves are both increasing in Figure 3, which coincides with the statements in Theorem 4.4 and Proposition 4.3(i). We can also observe that the value for distribution mixture is smaller than the corresponding one for quantile mixture in Figure 3, which is the same as it has been shown in Figure 2.

### 6.2 Conjectures for general distributions

Explicit expressions for $\overline{\mathrm{VaR}}_p(\mathbf{F})$ are unavailable for general marginal distributions. Fortunately, we can approximate the value of $\overline{\mathrm{VaR}}_p(\mathbf{F})$ using the rearrangement algorithm (RA) of Embrechts et al. (2013) and get an upper bound on $\overline{\mathrm{VaR}}_p(\mathbf{F})$ using (12) in Lemma A.1.

For distributions with non-monotone densities including Gamma and Weibull, the curves of both distri-

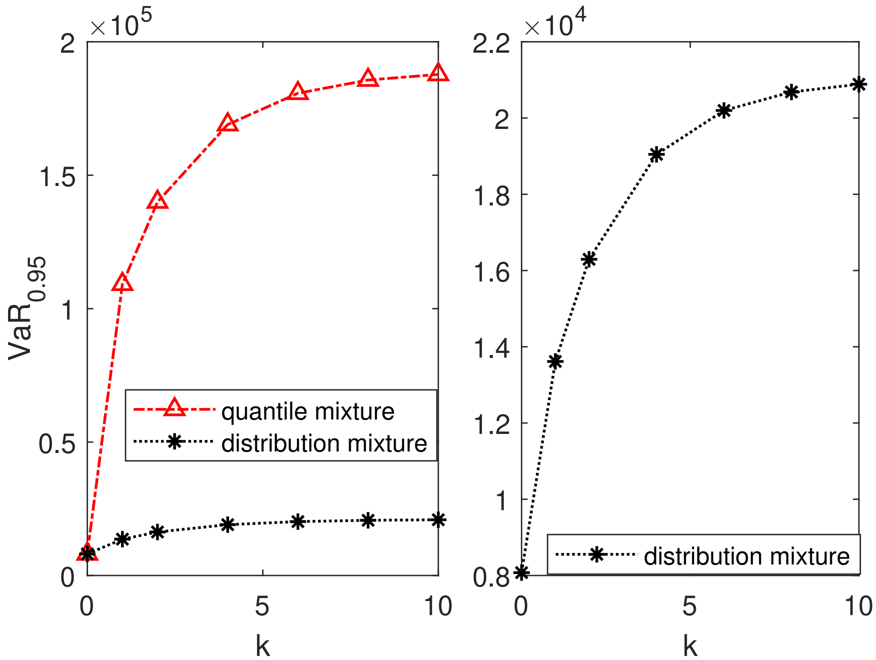



Figure 2: Quantile mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \otimes \mathbf{F})$; Distribution mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \mathbf{F})$. Setting: $p = 0.95$; $\boldsymbol{\alpha} = (1/3, 4, 5)$, $\boldsymbol{\theta} = (1, 2, 3)$, $X_i \sim \mathrm{Pareto}(\alpha_i, \theta_i)$, $i = 1, 2, 3$; $\Lambda$ is defined by (5); $k = 0, 1, 2, 4, 6, 8, 10$. The right panel zooms in on the range of the distribution mixture.

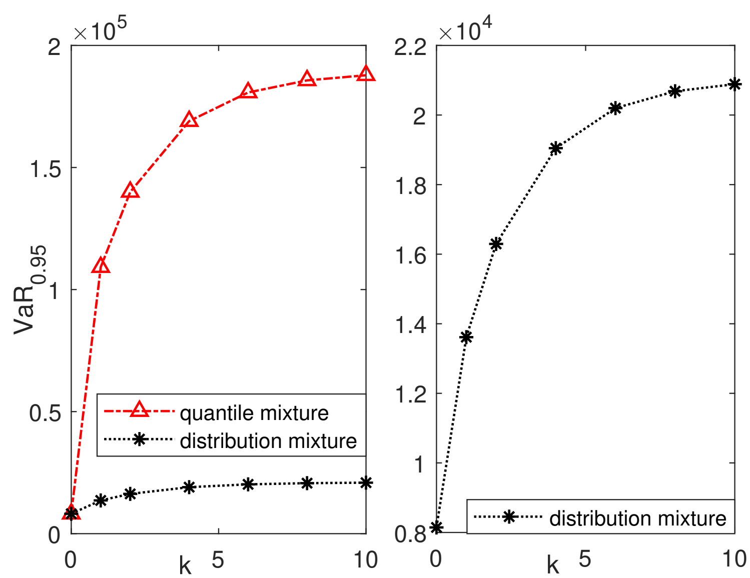
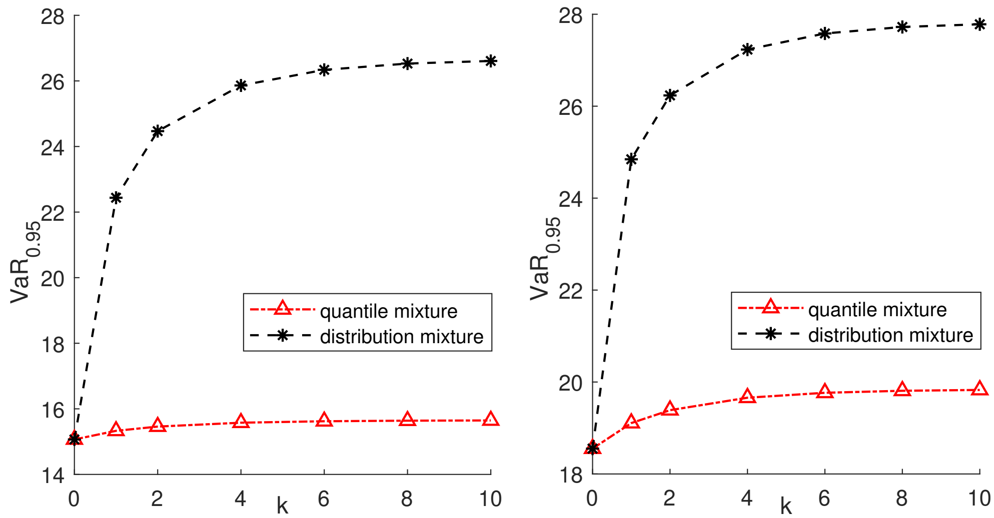



Figure 3: Quantile mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \otimes \mathbf{F})$; Distribution mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \mathbf{F})$. Setting: $p = 0.95$; $X_1 \sim \mathrm{Pareto}(1/3, 1)$, $X_2 \sim \Gamma(1, 2)$, $X_3 \sim \mathrm{Weibull}(1, 1/2)$; $\Lambda$ is defined by (5); $k = 0, 1, 2, 4, 6, 8, 10$. The right panel zooms in on the range of the distribution mixture.

bution and quantile mixtures in Figure 4 are increasing in $k$. The result on distribution mixture is consistent with Proposition 4.3(i), and the result on quantile mixture seems to suggest that the conclusion in Theorem 4.4 may be valid for more general distributions with non-monotone densities. This conjectured extension of Theorem 4.4 would hold if (3) holds for more general distributions, which is a difficult question.




(a) $X_3 \sim \text{Pareto}(3,1)$ (b) $X_3 \sim \text{LogNormal}(0,1)$

Figure 4: Quantile mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \otimes \mathbf{F})$; Distribution mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \mathbf{F})$. Setting: $p = 0.95$; $X_1 \sim \Gamma(5,1)$, $X_2 \sim \text{Weibull}(1,5)$, left panel: $X_3 \sim \text{Pareto}(3,1)$, right panel: $X_3 \sim \text{LogNormal}(0,1)$; $\Lambda$ is defined by (5); $k = 0,1,2,4,6,8,10$.

The above observation is no longer true for discrete distributions. We observe in Figure 5 that the curve of the quantile mixture is not increasing at some points (in this example, we have chosen a small $p = 0.01$ for illustration). This shows that the claim in Theorem 4.4 cannot be extended to arbitrary, in particular discrete, distributions.

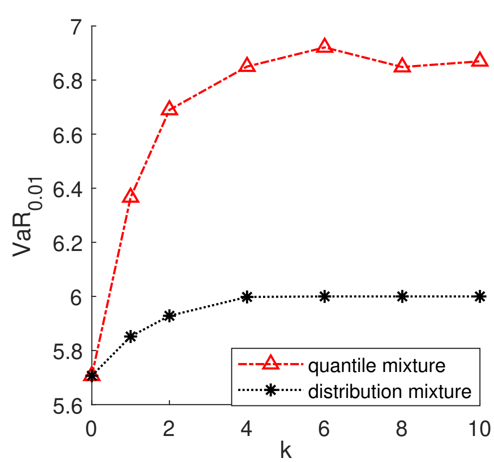



Figure 5: Quantile mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \otimes \mathbf{F})$; Distribution mixture: $\overline{\mathrm{VaR}}_p(\Lambda^k \mathbf{F})$. Setting: $p = 0.01$; $X_1 \sim \text{Binomial}(10, 0.1)$, $X_2 \sim \Gamma(5,1)$, $X_3 \sim \text{Weibull}(1,5)$; $\Lambda$ is defined by (5); $k = 0,1,2,4,6,8,10$.

# 7 Applications

## 7.1 Portfolio diversification with dependence uncertainty

We discuss applications of our results to portfolio diversification in the presence of dependence uncertainty. In this section, we treat risk measures as functionals on the space of random variables, that is, for a random variable $X$ and a risk measure $\rho$, we write $\rho(X) = \rho(F)$ if $X \sim F$.

For tractability, we consider a simple setting where the vector of losses $(X_1, \ldots, X_n)$ has identical marginal distributions $F$. A classic portfolio selection problem is to choose a portfolio position $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \Delta_n$ to minimize

$$R_\rho(\boldsymbol{\lambda}) := \rho\left(\sum_{i=1}^n \lambda_i X_i\right). \tag{6}$$

Alternatively, one may consider an objective which involves both risk and return, such as maximizing the quantity $\mathbb{E}[-\sum_{i=1}^n \lambda_i X_i] - \alpha\rho(\sum_{i=1}^n \lambda_i X_i)$ for some $\alpha > 0$ (e.g., $\alpha$ may arise as a Lagrangian multiplier); in our setting, this problem is equivalent to (6) since $\mathbb{E}[\sum_{i=1}^n \lambda_i X_i]$ is constant over $\boldsymbol{\lambda} \in \Delta_n$. Intuitively, for two portfolio positions $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$, we can say that $\boldsymbol{\gamma}$ is more diversified than $\boldsymbol{\lambda}$ if $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$, since in this case $\boldsymbol{\gamma}$ can be obtained from averaging components of $\boldsymbol{\lambda}$, i.e., $\boldsymbol{\gamma} = \Lambda\boldsymbol{\lambda}$ for some $\Lambda \in \mathcal{Q}_n$ (see Section 2). Due to diversification effect, one may expect, under the assumption that the marginal distributions of $(X_1, \ldots, X_n)$ are identical,

$$R_\rho(\boldsymbol{\gamma}) \leqslant R_\rho(\boldsymbol{\lambda}) \quad \text{if } \boldsymbol{\gamma} \text{ is more diversified than } \boldsymbol{\lambda}. \tag{7}$$

Note that $(\frac{1}{n}, \ldots, \frac{1}{n}) \prec \boldsymbol{\lambda} \prec (1, 0, \ldots, 0)$ for any portfolio position $\boldsymbol{\lambda}$, meaning that the most diversified portfolio is the equally weighted one, and the least diversified portfolio is concentrating on a single source of risk.

To compute the value of $R_\rho(\boldsymbol{\lambda})$ in (6) requires a full specification of the joint distribution of $(X_1, \ldots, X_n)$. In the presence of dependence uncertainty, we may take a worst-case approach by minimizing

$$\overline{R}_\rho(\boldsymbol{\lambda}) := \sup\left\{\rho\left(\sum_{i=1}^n \lambda_i Y_i\right) : Y_1, \ldots, Y_n \sim F\right\}. \tag{8}$$

Under the setting of optimizing (8), our intuition is that diversification should not yield any benefit, since the portfolio may not have any diversification effect due to unknown dependence; see Wang and Zitikis (2021) for discussions on the absence of diversification effect within the Fundamental Review of the Trading Book from the Basel Committee on Banking Supervision (BCBS (2019)). Hence, one may expect, as the marginal distributions are identical, that

$$\overline{R}_\rho(\boldsymbol{\gamma}) = \overline{R}_\rho(\boldsymbol{\lambda}) \quad \text{even if } \boldsymbol{\gamma} \text{ is more diversified than } \boldsymbol{\lambda} \text{ (in fact, for all } \boldsymbol{\gamma} \text{ and } \boldsymbol{\lambda}). \tag{9}$$

A similar observation is made in Proposition 1 of Pflug and Pohl (2018), which says that for a subadditive, comonotonic-additive and positively homogeneous risk measure, diversification under dependence uncertainty does not decrease the aggregate risk. These assumptions on the risk measure are not necessary for our result below.

The next proposition, based on Theorem 4.4 and Proposition 4.3, shows that, under some extra conditions, the two intuitive equations (7) and (9) hold for risk measures consistent with convex order. For VaR, one arrives at a statement in the reverse direction: the more diversified portfolio has a larger risk under dependence uncertainty.

**Proposition 7.1.** *Suppose that $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$, $(X_1, \dots, X_n)$ has identical marginal distributions $F$ with finite mean, and $\rho$ is a risk measure.*

*(i) If $\rho$ is consistent with convex order and $(X_1, \dots, X_n)$ is exchangeable,*[7] *then $R_\rho(\boldsymbol{\gamma}) \leqslant R_\rho(\boldsymbol{\lambda})$.*

*(ii) If $\rho$ is consistent with convex order, then $\overline{R}_\rho(\boldsymbol{\gamma}) = \overline{R}_\rho(\boldsymbol{\lambda})$.*

*(iii) If $\rho = \mathrm{VaR}_p$ for some $p \in (0,1)$ and $F \in \mathcal{M}_D \cup \mathcal{M}_I$, then $\overline{R}_\rho(\boldsymbol{\gamma}) \geqslant \overline{R}_\rho(\boldsymbol{\lambda})$.*

*Moreover, in (i) and (iii), the inequalities are generally not equalities.*

*Proof.* Write $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$. Take $X \sim F$, and let $\mathbf{F}$ and $\mathbf{G}$ be the tuples of marginal distributions of $(\gamma_1 X, \dots, \gamma_n X)$ and $(\lambda_1 X, \dots, \lambda_n X)$, respectively. Using $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$, there exists $\Lambda \in \mathcal{Q}_n$ such that $\mathbf{F} = \Lambda \otimes \mathbf{G}$. Since $X_1, \dots, X_n \sim F$, $\mathbf{F}$ and $\mathbf{G}$ are also tuples of marginal distributions of $(\gamma_1 X_1, \dots, \gamma_n X_n)$ and $(\lambda_1 X_1, \dots, \lambda_n X_n)$, respectively. Hence, we have

$$\overline{R}_\rho(\boldsymbol{\gamma}) = \overline{\rho}(\mathbf{F}) = \overline{\rho}(\Lambda \otimes \mathbf{G}) \quad \text{and} \quad \overline{R}_\rho(\boldsymbol{\lambda}) = \overline{\rho}(\mathbf{G}). \tag{10}$$

(i) As $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$ and $(X_1, \dots, X_n)$ is exchangeable, by Theorem 3.A.35 of Shaked and Shanthikumar (2007), we have $\sum_{i=1}^n \gamma_i X_i \prec_{\mathrm{cx}} \sum_{i=1}^n \lambda_i X_i$. Hence, $R_\rho(\boldsymbol{\gamma}) \leqslant R_\rho(\boldsymbol{\lambda})$. The inequality is strict when, for instance, $\rho = \mathrm{ES}_{0.5}$, $X_1, \dots, X_n$ are iid normal, $\boldsymbol{\gamma} = (\frac{1}{n}, \dots, \frac{1}{n})$, and $\boldsymbol{\lambda} = (1, 0, \dots, 0)$.

(ii) This follows directly from Proposition 4.3(ii) and (10).

(iii) The inequality $\overline{R}_\rho(\boldsymbol{\gamma}) \geqslant \overline{R}_\rho(\boldsymbol{\lambda})$ follows directly from Theorem 4.4 and (10). The inequality is strict in, for instance, the situation of Figure 1(a), where $F$ is a Pareto distribution with $\alpha = 3$. ☐

We make a few observations from Proposition 7.1. For identical marginal distributions in $\mathcal{M}_D$ or $\mathcal{M}_I$, under dependence uncertainty, VaR yields a bigger risk if the portfolio is more diversified. This may be seen as another disadvantage of VaR, which is well known to be problematic regarding diversification. In contrast, any risk measure consistent with convex order, such as ES, would simply ignore diversification effect in this setting (where diversification benefit is unjustifiable). Moreover, without dependence uncertainty, for an exchangeable vector of losses, a risk measure consistent with convex order rewards diversification, and there is no such general relationship for VaR. For the inequality in Proposition 7.1 (iii), it suffices to require the $p$-tail distribution of $F$ to be in $\mathcal{M}_D \cup \mathcal{M}_I$; see Remark 4.5.

[7] A random vector $\mathbf{X}$ is exchangeable if $\mathbf{X}$ is identically distributed as $\pi(\mathbf{X})$ for any permutation $\pi$.

### 7.2 Merging p-values in hypothesis testing

In this subsection, we apply our results to p-merging methods following the setup of Vovk and Wang (2020). A random variable $P$ is a *p-variable* if $\mathbb{P}(P \leqslant \varepsilon) \leqslant \varepsilon$ for all $\varepsilon \in (0,1)$, and its realization is called a p-value. In multiple hypothesis testing, one natural problem is to merge individual p-values into one p-value. More specifically, with $n$ p-variables $P_1, \dots, P_n$, one needs to choose an increasing Borel function $F : [0,1]^n \to [0,\infty)$ as a *merging function* such that $F(P_1,\dots,P_n)$ is a p-variable. $F$ is a *precise merging function* if for each $\varepsilon \in (0,1)$, $\mathbb{P}(F(P_1,\dots,P_n) \leqslant \varepsilon) = \varepsilon$ for some p-variables $P_1,\dots,P_n$.

As explained in Vovk and Wang (2020), an advantage of using averaging methods to combine p-values, compared to classic methods on order statistics, is that we can introduce weights to p-values in an intuitive way. Without imposing any dependence assumption on the individual p-variables, an averaging method uses, for $r \in [-\infty,\infty]$ ($r \in \{-\infty, 0, \infty\}$ are interpreted as limits),

$$F : [0,1]^n \to [0,\infty),\ (p_1,\dots,p_n) \mapsto a_{r,\mathbf{w}}(w_1 p_1^r + \cdots + w_n p_n^r)^{\frac{1}{r}},$$

as the merging function, where $a_{r,\mathbf{w}}$ is a constant multiplier and $\mathbf{w} = (w_1,\dots,w_n) \in \Delta_n$. The constant $a_{r,\mathbf{w}}$ is chosen so that $F$ is a precise merging function, thus the most powerful choice of the constant multiplier. Let $\mathcal{U}$ be the set of uniform random variables distributed on [0,1]. Lemma 1 in Vovk and Wang (2020) gives

$$a_{r,\mathbf{w}} = \begin{cases} -\sup\left\{q_0(-\sum_{i=1}^n w_i P_i^r) \mid P_1,\dots,P_n \in \mathcal{U}\right\}^{-1/r}, & r>0; \\ \exp\left(\sup\{q_0(\sum_{i=1}^n w_i \log(1/P_i)) \mid P_1,\dots,P_n \in \mathcal{U}\}\right), & r=0; \\ \sup\left\{q_0\left(\sum_{i=1}^n w_i P_i^r\right) \mid P_1,\dots,P_n \in \mathcal{U}\right\}^{-1/r}, & r<0, \end{cases}$$

where $q_0 : X \mapsto \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) > 0\}$ is the essential infimum. Clearly, $a_{r,\mathbf{w}}$ involves calculating $\overline{\mathrm{VaR}}_p(\mathbf{F})$ for Pareto, exponential or Beta distributions, and letting $p \downarrow 0$.

Denote $a_{r,\mathbf{w}}$ by $a_{r,n}$ where $\mathbf{w} = (1/n,\dots,1/n)$. Analytical results for $a_{r,n}$ has been well studied in Vovk and Wang (2020) whereas results for $a_{r,\mathbf{w}}$ are limited since there are no analytical formulas of $\overline{\mathrm{VaR}}_p(\mathbf{F})$ in general for heterogeneous marginal distributions. Although the rearrangement algorithm of Puccetti and Rüschendorf (2012) and Embrechts et al. (2013) can be used to calculate $a_{r,\mathbf{w}}$ numerically, the calculation burden becomes quite heavy in high-dimensional situation, which is unfortunately very common in multiple hypothesis testing. It turns out that our Theorem 4.4 is helpful to provide a convenient upper bound on $a_{r,\mathbf{w}}$.

**Proposition 7.2.** *For $r \in \mathbb{R}$, we have $a_{r,\mathbf{w}} \leqslant a_{r,n}$.*

*Proof.* Note that for $r<0$, $P_i^r$, $i=1,\dots,n$, has a decreasing density, and $(1/n,\dots,1/n) \prec (w_1,\dots,w_n)$ in majorization order. By letting $p \downarrow 0$ in Proposition 5.3, we have

$$\sup\left\{q_0\left(\sum_{i=1}^n w_i P_i^r\right) \mid P_1,\dots,P_n \in \mathcal{U}\right\} \leqslant \sup\left\{q_0\left(\sum_{i=1}^n \frac{1}{n} P_i^r\right) \mid P_1,\dots,P_n \in \mathcal{U}\right\}.$$

Therefore $a_{r,\mathbf{w}} \leqslant a_{r,n}$ for $r < 0$. If $r \geqslant 0$, the argument can be proved similarly using Corollary 4.6. □

The interpretation of Proposition 7.2 is that, when using a weighted p-merging method, one can safely rely on the same coefficient obtained from a symmetric p-merging method. This is particularly convenient when validity of the test is more important than the quality of an approximation; see Vovk and Wang (2020) for more discussions on such applications.

# 8 Some further technical discussions

## 8.1 Location shifts for distribution and quantile mixtures

In this section we discuss the difference between distribution and quantile mixtures when location shifts are applied. Let $V_x = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 + \dots + x_n = x\}$ for $x \in \mathbb{R}$. For $\mathbf{F} \in \mathcal{M}^n$ and $\mathbf{x} \in V_x$, we have the invariance relation

$$\mathcal{D}_n(\mathbf{T}_\mathbf{x}(\mathbf{F})) = T_x(\mathcal{D}_n(\mathbf{F})). \tag{11}$$

The aggregation set of quantile mixture is invariant under location shifts of the marginal distributions, in sharp contrast to the case of distribution mixture. For $\mathbf{F} \in \mathcal{M}^n$ and $\mathbf{x} \in V_x$, it holds that for $\Lambda \in \mathcal{Q}_n$,

$$\mathcal{D}_n(\Lambda \otimes \mathbf{T}_\mathbf{x}(\mathbf{F})) = T_x\left(\mathcal{D}_n(\Lambda \otimes \mathbf{F})\right).$$

That means, $\mathcal{D}_n(\Lambda \otimes \mathbf{T}_\mathbf{x}(\mathbf{F}))$ is the same for all $\mathbf{x} \in V_x$. However, this does not hold for the distribution mixture, that is, generally, $\mathcal{D}_n(\Lambda\mathbf{T}_\mathbf{x}(\mathbf{F}))$ is not the same for $\mathbf{x} \in V_x$, and

$$\mathcal{D}_n(\Lambda\mathbf{T}_\mathbf{x}(\mathbf{F})) \neq T_x\left(\mathcal{D}_n(\Lambda\mathbf{F})\right).$$

In particular, for $x \neq 0$ and $F_1 \neq F_2$,

$$\mathcal{D}_2\left(\frac{1}{2}(T_x(F_1) + F_2), \frac{1}{2}(T_x(F_1) + F_2)\right) \neq \mathcal{D}_2\left(\frac{1}{2}(F_1 + T_x(F_2)), \frac{1}{2}(F_1 + T_x(F_2)))\right).$$

The above example shows that distribution mixture and quantile mixtures treat location shifts differently.

Inspired by the above observation, we slightly generalize Theorem 2.2 by including location shifts. For $\mathbf{F} \in \mathcal{M}^n$, we define the set $\mathcal{A}_n(\mathbf{F})$ of averaging and location shifts of $\mathbf{F}$ as

$$\mathcal{A}_n(\mathbf{F}) = \{\Lambda\mathbf{T}_\mathbf{x}(\mathbf{F}) : \Lambda \in \mathcal{Q}_n,\ \mathbf{x} \in \mathbb{R}^n,\ x_1 + \dots + x_n = 0\},$$

and denote by $\overline{\mathcal{A}_n(\mathbf{F})}$ the closure of the convex hull of $\mathcal{A}_n(\mathbf{F})$ with respect to weak convergence. It is straightforward to check

$$\overline{\mathcal{A}_n(\mathbf{T}_\mathbf{y}(\mathbf{F}))} = \mathbf{T}_\mathbf{y}\left(\overline{\mathcal{A}_n(\mathbf{F})}\right), \quad \mathbf{y} = (y, \dots, y) \in \mathbb{R}^n.$$

**Proposition 8.1.** *For $\mathbf{F} \in \mathcal{M}^n$ and $\mathbf{G} \in \overline{\mathcal{A}_n(\mathbf{F})}$, we have $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$.*

*Proof.* First, by Theorem 2.2 and (11), $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$ for each $\mathbf{G} \in \mathcal{A}_n(\mathbf{F})$. Denote by $\mathrm{cx}(\mathcal{A}_n(\mathbf{F}))$ the convex hull of $\mathcal{A}_n(\mathbf{F})$. By Lemma 2.1(ii-b), for each $\mathbf{G} \in \mathrm{cx}(\mathcal{A}_n(\mathbf{F}))$, we have $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$. Take $\mathbf{G} \in \overline{\mathcal{A}_n(\mathbf{F})}$, and write it as the limit of $\{\mathbf{G}_k\}_{k=1}^{\infty} \subset \mathrm{cx}(\mathcal{A}_n(\mathbf{F}))$. It follows that for any $F \in \mathcal{D}_n(\mathbf{F})$, $F$ is also in $\mathcal{D}_n(\mathbf{G}_k)$. This implies $F$ is also in $\mathcal{D}_n(\mathbf{G})$ by the compactness property in Theorem 2.1(vii-b) of Bernard et al. (2014). $\square$

## 8.2 Connection to joint mixability

Joint mixability (Wang et al. (2013) and Wang and Wang (2016)) is a central concept in the study of risk aggregation with dependence uncertainty, and analytical results are quite limited. In this section, we study the implication of our results on conditions for joint mixability. We denote by $\delta_x$ the point mass at $x \in \mathbb{R}$.

**Definition 8.2** (Joint mixability)**.** An $n$-tuple of distributions $\mathbf{F} \in \mathcal{M}^n$ is *jointly mixable* (JM) if $\mathcal{D}_n(\mathbf{F})$ contains a point mass distribution $\delta_x$, where $x \in \mathbb{R}$ is called a *center* of $\mathbf{F}$.

Example 2.5 implies a conclusion on the joint mixability of Bernoulli distributions.

**Proposition 8.3.** *For $p_1, \dots, p_n \in [0, 1]$, $(B_{p_1}, \dots, B_{p_n})$ is jointly mixable if and only if $\sum_{i=1}^n p_i$ is an integer.*

*Proof.* The "only-if" part is trivial since the sum of Bernoulli random variables takes value in integers. To show the "if" part, let $k = \sum_{i=1}^n p_i$ and $\mathbf{1}_k \in \{0, 1\}^n$ be a vector whose first $k$ entries are 1 and the remaining entries are 0. It is clear that $\mathbf{p} \prec \mathbf{1}_k$ (see Section 1.A.3 of Marshall et al. (2011)). Hence, from Example 2.5,

$$\{\delta_k\} = \mathcal{D}_n(\underbrace{B_1, \dots, B_1}_{k}, \underbrace{B_0, \dots, B_0}_{n-k}) \subset \mathcal{D}_n(B_{p_1}, \dots, B_{p_n}).$$

Therefore $(B_{p_1}, \dots, B_{p_n})$ is jointly mixable. $\square$

The set $\overline{\mathcal{A}_n(\mathbf{F})}$ can also be used to obtain joint mixability of some tuples of distributions. In particular, we shall see in the following proposition that $\overline{\mathcal{A}_n(\delta_0, \dots, \delta_0)}$ is the set of all jointly mixable tuples with center 0.

**Proposition 8.4.** *For $\mathbf{G} \in \mathcal{M}^n$, the following statements are equivalent.*

*(i)* *$\mathbf{G}$ is jointly mixable.*

*(ii)* *$\mathbf{G} \in \overline{\mathcal{A}_n(\delta_c, \dots, \delta_c)}$ for some $c \in \mathbb{R}$.*

*(iii)* *$\mathbf{G} \in \overline{\mathcal{A}_n(\mathbf{F})}$ for some $\mathbf{F} \in \mathcal{M}^n$ which is jointly mixable.*

*Proof.* (ii)⇒(iii) is trivial. (iii)⇒(i): Suppose that $\mathbf{G} \in \overline{\mathcal{A}_n(\mathbf{F})}$ and $\mathbf{F}$ is jointly mixable with center $x \in \mathbb{R}$. By Proposition 8.1, we have $\{\delta_x\} \subset \mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\mathbf{G})$. This shows $\mathbf{G}$ is jointly mixable. Next, we show (i)⇒(ii). Suppose that $\mathbf{G}$ is jointly mixable, and without loss of generality we can assume it has center 0. By definition, there exists a random vector $\mathbf{X} = (X_1, \dots, X_n)$ such that $X_i \sim G_i$ and $X_1 + \cdots + X_n = 0$. Denote by $H$ the distribution measure of $\mathbf{X}$. For $A \in \mathcal{B}(\mathbb{R})$ and $i = 1, 2, \dots, n$,

$$G_i(A) = \mathbb{P}(X_i \in A) = \int_{\mathbb{R}^n} \mathbb{P}(X_i \in A | \mathbf{X} = \mathbf{y}) H(\mathrm{d}\mathbf{y}) = \int_{\mathbb{R}^n} \delta_{y_i}(A) H(\mathrm{d}\mathbf{y}),$$

and as a consequence,

$$\mathbf{G}(A) = (G_1(A), \dots, G_n(A)) = \int_{\mathbb{R}^n} (\delta_{y_1}(A), \dots, \delta_{y_n}(A)) H(\mathrm{d}\mathbf{y}).$$

Noting that $H$ is supported in $V_0 = \{(y_1, \dots, y_n) \in \mathbb{R}^n : y_1 + \cdots + y_n = 0\}$, we have

$$\mathbf{G}(A) = \int_{V_0} (\delta_{y_1}(A), \dots, \delta_{y_n}(A)) H(\mathrm{d}\mathbf{y}) = \int_{V_0} \mathbf{T}_{\mathbf{y}}(\delta_0(A), \dots, \delta_0(A)) H(\mathrm{d}\mathbf{y}).$$

Hence, we conclude that $\mathbf{G} \in \overline{\mathcal{A}_n(\delta_0, \dots, \delta_0)}$. □

The set $\overline{\mathcal{A}_n(\delta_c, \dots, \delta_c)}$ is quite rich and cannot be analytically characterized. The simple example of uniform distributions might be helpful to understand Proposition 8.4. Suppose that $F_i = \mathrm{U}[0, a_i]$, $a_i > 0$, $i = 1, \dots, n$, and $\sum_{i=1}^n a_i \geqslant 2 \bigvee_{i=1}^n a_i$. By Theorem 3.1 of Wang and Wang (2016), we know that $\mathbf{F}$ is jointly mixable. Then, Proposition 8.4 implies that every tuple in the set $\overline{\mathcal{A}_n(\mathbf{F})}$ is jointly mixable.

It remains an open question whether it is possible to characterize the set $\overline{\mathcal{A}_n(\mathbf{F})}$ for uniform random variables. This would lead to many classes of jointly mixable distributions including those with monotone densities and symmetric densities; see Wang and Wang (2016).

# 9 Concluding remarks and open questions

This paper studies the ordering relationship for aggregation sets where the marginal distributions for different sets are connected by either a distribution mixture or a quantile mixture. For general marginal distributions, the aggregation set becomes larger after making a distribution mixture on the marginal risks, whereas the aggregation sets are not necessarily comparable in general by a quantile mixture on the marginal risks. Nevertheless, we obtain several useful results especially on the comparison of VaR aggregation, which has applications in and outside financial risk management.

Although the marginal distributions are assumed known in our main setting, this assumption is not essential for the interpretation of our results in practical situations. In case both marginal uncertainty and dependence uncertainty are present, our results can be directly applied to obtain ordering relationships, as we explain below. Suppose that $\Lambda \in \mathcal{Q}_n$ and $\mathcal{F} \subset \mathcal{M}^n$ is a set of possible marginal models, representing uncertainty on the marginal distributions. In this case, the set of all possible distributions of aggregate risk is $\bigcup_{\mathbf{F}\in\mathcal{F}}\mathcal{D}_n(\mathbf{F})$, and the worst-case value of a risk measure $\rho$ is $\sup\{\rho(G) : G \in \mathcal{D}_n(\mathbf{F}),\ \mathbf{F} \in \mathcal{F}\} = \sup_{\mathbf{F}\in\mathcal{F}} \overline{\rho}(\mathbf{F})$. Using Theorem 2.2, Proposition 4.3 and Theorem 4.4, we have

$$\bigcup_{\mathbf{F}\in\mathcal{F}}\mathcal{D}_n(\mathbf{F}) \subset \bigcup_{\mathbf{F}\in\mathcal{F}}\mathcal{D}_n(\Lambda\mathbf{F}), \qquad \sup_{\mathbf{F}\in\mathcal{F}}\overline{\rho}(\mathbf{F}) \leqslant \sup_{\mathbf{F}\in\mathcal{F}}\overline{\rho}(\Lambda\mathbf{F}),$$

and, if $\mathcal{F} \subset \mathcal{M}_D^n \cup \mathcal{M}_I^n$,

$$\sup_{\mathbf{F}\in\mathcal{F}}\overline{\mathrm{VaR}}_p(\mathbf{F}) \leqslant \sup_{\mathbf{F}\in\mathcal{F}}\overline{\mathrm{VaR}}_p(\Lambda\otimes\mathbf{F}).$$

Thus, our results on set inclusion and risk measure inequalities remain valid in the presence of marginal uncertainty.

Many questions on quantile mixtures are still open, and we conclude the paper with four of them. The first question concerns whether $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\Lambda\otimes\mathbf{F})$ holds for cases other than the uniform distributions in Proposition 3.4. As we have seen from Example 3.2, for $\mathbf{F} \in \mathcal{M}^n$ and $\Lambda \in \mathcal{Q}_n$, $\mathcal{D}_n(\mathbf{F})$ and $\mathcal{D}_n(\Lambda\otimes\mathbf{F})$ are generally not comparable. It remains open whether $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\Lambda\otimes\mathbf{F})$ under some conditions. For instance, Proposition 3.4 requires $n \geqslant 3$ and $\Lambda$ being a constant times the identity, to use the characterization of $\mathcal{D}_n(\mathbf{F})$ from Mao et al. (2019). It remains unclear whether the same conclusion holds for $n = 2$ or other choices of $\Lambda$.

The second question concerns decreasing densities (or increasing densities). A concrete conjecture is presented below, which is inspired by Theorem 4.4. It is unclear how to formulate natural classes of distributions other than $\mathcal{M}_D$ (or $\mathcal{M}_I$) such that similar statements can be expected.

**Conjecture 9.1.** *For $\Lambda \in \mathcal{Q}_n$ and $\mathbf{F} \in \mathcal{M}_D^n$, we have $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(\Lambda\otimes\mathbf{F})$. Weaker versions of this conjecture are:*

*(i) For $F \in \mathcal{M}_D$, and $\boldsymbol{\lambda}, \boldsymbol{\gamma} \in \mathbb{R}_+^n$, if $\boldsymbol{\gamma} \prec \boldsymbol{\lambda}$, then $\mathcal{D}_n(F^{\lambda_1}, \dots, F^{\lambda_n}) \subset \mathcal{D}_n(F^{\gamma_1}, \dots, F^{\gamma_n})$.*

*(ii) For $F_1, \dots, F_n \in \mathcal{M}_D$, $\mathcal{D}_n(F_1, \dots, F_n) \subset \mathcal{D}_n(F, \dots, F)$ where $F^{-1} = \frac{1}{n}\sum_{i=1}^n F_i^{-1}$.*

*(iii) For $F \in \mathcal{M}_D$ and $(\lambda_1, \dots, \lambda_n) \in \Delta_n$, $\mathcal{D}_n(F^{n\lambda_1}, \dots, F^{n\lambda_n}) \subset \mathcal{D}_n(F, \dots, F)$.*

It is obvious that the main statement in Conjecture 9.1 implies (i) by noting that one can choose $\Lambda$ such that $\boldsymbol{\gamma} = \Lambda\boldsymbol{\lambda}$ and it implies (ii) by choosing $\Lambda = (\frac{1}{n})_{n\times n}$. Both (i) and (ii) imply (iii). An example is provided below to illustrate the connection of Conjecture 9.1 to joint mixability.

**Example 9.2.** We make a connection of Conjecture 9.1 to Theorem 3.2 of Wang and Wang (2016), which says that for $F_i \in \mathcal{M}_D$ with essential support $[0, b_i]$, $i = 1, \dots, n$, $\mathcal{D}_n(F_1, \dots, F_n)$ contains a point mass if and only if the *mean-length condition* holds, that is,

$$\sum_{i=1}^n \mu_i \geqslant \max_{i=1,\dots,n} b_i$$

where $\mu_i$ is the mean of $F_i$, $i = 1, \ldots, n$. For $\Lambda \in \mathcal{Q}_n$ and $\mathbf{F} \in \mathcal{M}_D^n$, let $(\hat{\mu}_1, \ldots, \hat{\mu}_n)$ be the mean vector of $\Lambda \otimes \mathbf{F}$. Note that

$$\sum_{i=1}^{n} \hat{\mu}_i = \mathbf{1}_n^\top \Lambda \boldsymbol{\mu} = \mathbf{1}_n^\top \boldsymbol{\mu} = \sum_{i=1}^{n} \mu_i,$$

where $\mathbf{1}_n = (1, \ldots, 1) \in \mathbb{R}^n$. On the other hand, each component of $\Lambda \otimes \mathbf{F}$ has a shorter or equal length of support than the maximum length of $\mathbf{F}$. As a consequence, if the mean-length condition holds for $\mathbf{F}$, then it also holds for $\Lambda \otimes \mathbf{F}$. Therefore, if $\mathcal{D}_n(\mathbf{F})$ contains a point mass, then so does $\mathcal{D}_n(\Lambda \otimes \mathbf{F})$; on the contrary, if $\mathcal{D}_n(\Lambda \otimes \mathbf{F})$ contains a point mass, $\mathcal{D}_n(\mathbf{F})$ does not necessarily contains a point mass, since it may have a longer length of the maximum support. This, at least intuitively, suggests that $\mathcal{D}_n(\mathbf{F}) \subsetneq \mathcal{D}_n(\Lambda \otimes \mathbf{F})$ may hold, as in Conjecture 9.1.

The third question is about the order of VaR for quantile mixture. Our numerical results in Figure 4 suggest that the VaR relation

$$\overline{\mathrm{VaR}}_p(\mathbf{F}) \leqslant \overline{\mathrm{VaR}}_p(\Lambda \otimes \mathbf{F})$$

holds for more general choices of $\mathbf{F}$ than the ones in Theorem 4.4. We are not sure what general conditions on $\mathbf{F}$ will guarantee this relation to hold.

The last question concerns a cross comparison of distribution and quantile mixtures. As we see from Proposition 5.3,

$$\overline{\mathrm{VaR}}_p(\Lambda \mathbf{F}) \leqslant \overline{\mathrm{VaR}}_p(\Lambda \otimes \mathbf{F})$$

holds for $\mathbf{F}$ being a vector of Pareto distributions with the same shape parameter and infinite mean. We wonder whether the same relationship holds for other distributions without a finite mean. Note that for the case of finite mean, the relationship may be reversed, as illustrated in Figure 1; however we do not have a proof for the reverse inequality (assuming finite mean) either. Generally, it is unclear to us whether and in which situation $\mathcal{D}_n(\Lambda \mathbf{F})$ and $\mathcal{D}_n(\Lambda \otimes \mathbf{F})$ are comparable.

## Acknowledgements

We thank an Associate Editor and two anonymous referees for helpful comments on an earlier version of the paper. Y. Liu is financially supported by the China Scholarship Council. R. Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2018-03823, RGPAS-2018-522590) and from the Center of Actuarial Excellence Research Grant from the Society of Actuaries.

## AUTHOR CONTRIBUTION

All authors contributed equally to the paper.

## DATA AVAILABILITY STATEMENT

Data sharing not applicable - no new data generated.

# A Some proofs and further technical results

## A.1 A lemma used in the proof of Theorem 4.4

The following lemma is rephrased from Theorem 2 of Blanchet et al. (2020).

**Lemma A.1.** *For $p \in (0,1)$ and any $\mathbf{F} = (F_1, \ldots, F_n) \in \mathcal{M}^n$,*

$$\overline{\mathrm{VaR}}_p^*(\mathbf{F}) \leqslant \inf_{\boldsymbol{\beta} \in \mathbb{B}_n} \sum_{i=1}^n \frac{1}{(1-p)(1-\beta)} \int_{p+(1-p)(\beta-\beta_i)}^{1-(1-p)\beta_i} \mathrm{VaR}_u(F_i) \mathrm{d}u, \tag{12}$$

*where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$, $\beta = \sum_{i=1}^n \beta_i$ and $\mathbb{B}_n = \{\boldsymbol{\beta} \in [0,1)^n : \beta < 1\}$, and the above inequality is an equality if $\mathbf{F} \in \mathcal{M}_D^n \cup \mathcal{M}_I^n$.*

## A.2 Proof of Proposition 4.2

*Proof.* We first focus on (i). We will show (a) $\Leftrightarrow$ (c). (c) $\Rightarrow$ (a) is trivial by the definition of stochastic order. For (a) $\Rightarrow$ (c), note that $\Lambda \mathbf{F} \prec_{\mathrm{st}} \mathbf{F}$ with $\Lambda = (\Lambda_{ij})$ implies

$$\sum_{j=1}^n \Lambda_{ij} F_j(x) \geqslant F_i(x), \; x \in \mathbb{R}, i = 1, \ldots, n. \tag{13}$$

Adding all the inequalities in (13) yields

$$\sum_{i=1}^n \sum_{j=1}^n \Lambda_{ij} F_j(x) \geqslant \sum_{i=1}^n F_i(x), \; x \in \mathbb{R}.$$

Due to the fact that $\Lambda$ is a doubly stochastic matrix, we have

$$\sum_{i=1}^n \sum_{j=1}^n \Lambda_{ij} F_j(x) = \sum_{i=1}^n F_i(x), \; x \in \mathbb{R}.$$

Hence all the inequalities in (13) are essentially equalities. This proves (c). We can analogously show that (b) $\Leftrightarrow$ (c). This establishes the claims in (i). We will omit the proof of (ii) since it is similar to the proof of (i).

We next focus on (iii). Trivially, (c) $\Rightarrow$ (a) and (c) $\Rightarrow$ (b). Next, we will only show (a) $\Rightarrow$ (c) since (b) $\Rightarrow$ (c) is similar. Denote by $\mathbf{G} = (G_1, \ldots, G_n) = \Lambda \otimes \mathbf{F}$. Hence

$$G_i^{-1} = \sum_{j=1}^n \Lambda_{ij} F_j^{-1}.$$

By definition, $\Lambda \otimes \mathbf{F} \prec_{\mathrm{cx}} \mathbf{F}$ implies $G_i \prec_{\mathrm{cx}} F_i, i = 1, \ldots, n$. It is well known (see e.g., Theorem 3.A.5 of Shaked and Shanthikumar (2007)) that for any two distributions $F$ and $G$ in $\mathcal{M}_1$,

$$F \prec_{\mathrm{cx}} G \quad \Leftrightarrow \quad \mathrm{ES}_p(F) \leqslant \mathrm{ES}_p(G) \text{ for all } p \in (0,1). \tag{14}$$

Moreover, by the comonotonic-additivity of $\mathrm{ES}_p$, we have

$$\mathrm{ES}_p(G_i) = \sum_{j=1}^{n} \Lambda_{ij} \mathrm{ES}_p(F_j), \; i = 1, \ldots, n.$$

Consequently,

$$\mathrm{ES}_p(G_i) = \sum_{j=1}^{n} \Lambda_{ij} \mathrm{ES}_p(F_j) \leqslant \mathrm{ES}_p(F_i), \; p \in (0,1), i = 1, \ldots, n. \tag{15}$$

Noting that $\Lambda$ is a doubly stochastic matrix, similarly as in the proof of (i), adding all the inequalities in (15) leads to

$$\sum_{i=1}^{n} \mathrm{ES}_p(G_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} \Lambda_{ij} \mathrm{ES}_p(F_j) = \sum_{i=1}^{n} \mathrm{ES}_p(F_i), \; p \in (0,1).$$

This implies that the inequalities in (15) are equalities, which means that $\Lambda \otimes \mathbf{F} = \mathbf{F}$ by (14). We complete the proof of (iii).

Finally, we consider (iv). (b) $\Rightarrow$ (a) is trivial. We will show (a) $\Rightarrow$ (b). By (14), $\Lambda \mathbf{F} \prec_{\mathrm{cx}} \mathbf{F}$ is equivalent to

$$\mathrm{ES}_p(F_i) \geqslant \mathrm{ES}_p\left(\sum_{j=1}^{n} \Lambda_{ij} F_j\right), \quad i = 1, \ldots, n. \tag{16}$$

Moreover, by the concavity of $\mathrm{ES}_p$ on mixtures (e.g., Theorem 3 of Wang et al. (2020)), we have

$$\mathrm{ES}_p\left(\sum_{j=1}^{n} \Lambda_{ij} F_j\right) \geqslant \sum_{j=1}^{n} \Lambda_{ij} \mathrm{ES}_p(F_j).$$

Therefore, we have

$$\mathrm{ES}_p(F_i) \geqslant \mathrm{ES}_p\left(\sum_{j=1}^{n} \Lambda_{ij} F_j\right) \geqslant \sum_{j=1}^{n} \Lambda_{ij} \mathrm{ES}_p(F_j), \quad i = 1, \ldots, n. \tag{17}$$

Adding the inequalities in (17) with noting that $\Lambda$ is a doubly stochastic matrix yields

$$\sum_{i=1}^{n} \mathrm{ES}_p(F_i) \geqslant \sum_{i=1}^{n} \mathrm{ES}_p\left(\sum_{j=1}^{n} \Lambda_{ij} F_j\right) \geqslant \sum_{i=1}^{n} \sum_{j=1}^{n} \Lambda_{ij} \mathrm{ES}_p(F_j) = \sum_{i=1}^{n} \mathrm{ES}_p(F_i).$$

Hence

$$\sum_{i=1}^{n} \mathrm{ES}_p(F_i) = \sum_{i=1}^{n} \mathrm{ES}_p\left(\sum_{j=1}^{n} \Lambda_{ij} F_j\right),$$

which implies that inequalities in (16) are all equalities. We establish the claim by (14). □

## A.3 Proof of Proposition 5.1

*Proof.* (i) Note that $(P_{\alpha,\theta})^{-1} = \theta(P_{\alpha,1})^{-1}$ for $\theta, \alpha > 0$. Hence we prove (i) by showing that

$$(\Lambda \otimes \mathbf{P}_{\alpha,\boldsymbol{\theta}})^{-1} = \Lambda(\mathbf{P}_{\alpha,\boldsymbol{\theta}})^{-1} = \mathbf{P}^{-1}_{\alpha,\Lambda\boldsymbol{\theta}}.$$

(ii) Let $\mathcal{U}$ be the set of uniform random variables on $[0,1]$. By monotonicity of $\rho$, we have, for $0 < \alpha_1 < \alpha_2$,

$$\begin{aligned}\overline{\rho}(\mathbf{P}_{\alpha_1,\boldsymbol{\theta}}) &= \sup\left\{\rho\left(\theta_1 U_1^{-1/\alpha_1} + \cdots + \theta_n U_n^{-1/\alpha_1}\right) \mid U_1,\ldots,U_n \in \mathcal{U}\right\} \\ &\geqslant \sup\left\{\rho\left(\theta_1 U_1^{-1/\alpha_2} + \cdots + \theta_n U_n^{-1/\alpha_2}\right) \mid U_1,\ldots,U_n \in \mathcal{U}\right\} = \overline{\rho}(\mathbf{P}_{\alpha_2,\boldsymbol{\theta}}).\end{aligned}$$

This implies that $\overline{\rho}(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ is decreasing in $\alpha$.

(iii) By monotonicity of $\rho$, we can establish the claim of (iii) similarly as the proof of (ii). □

## A.4 Some further properties of $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$

Properties of $\overline{\rho}(P_{\alpha,\boldsymbol{\theta}})$ in Proposition 5.1 can be strengthened for $\rho = \mathrm{VaR}_p$.

**Proposition A.2.** *For $p \in (0,1)$, $\alpha > 0$ and $\boldsymbol{\theta} \in (0,\infty)^n$,*

*(i)* $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ *is increasing and continuous in $p$;*

*(ii)* $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ *is decreasing and continuous in $\alpha$;*

*(iii)* $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ *is increasing and continuous in each component of $\boldsymbol{\theta}$;*

*(iv)* $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ *is homogeneous in $\boldsymbol{\theta}$, that is, for $\lambda > 0$,*

$$\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\lambda\boldsymbol{\theta}}) = \lambda\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}});$$

*(v) If $\alpha > 1$, then*

$$\frac{\mathbf{1}\cdot\boldsymbol{\theta}}{(1-p)^{1/\alpha}} \leqslant \overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}}) \leqslant \frac{\alpha}{\alpha-1} \times \frac{\mathbf{1}\cdot\boldsymbol{\theta}}{(1-p)^{1/\alpha}}. \tag{18}$$

*Proof.* (i) As the quantile of Pareto distribution is continuous, by Lemma 4.4 and 4.5 of Bernard et al. (2014), $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ is continuous in $p$ on $(0,1)$.

(ii) Let $\mathcal{U}$ be the set of uniform random variables distributed on $(0,1)$. We note that

$$\begin{aligned}\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}}) &= \sup\left\{\mathrm{VaR}_p\left(\theta_1 U_1^{-1/\alpha}+\cdots+\theta_n U_n^{-1/\alpha}\right) : U_1,\ldots,U_n\in\mathcal{U}\right\}\\ &= \sum_{i=1}^n \theta_i \sup\left\{\mathrm{VaR}_{1-p}\left(M_{\alpha,\boldsymbol{\theta}}(U_1,\ldots,U_n)\right) : U_1,\ldots,U_n\in\mathcal{U}\right\}^{-\frac{1}{\alpha}},\end{aligned}$$

where $M_{\alpha,\boldsymbol{\theta}}(u_1,\ldots,u_n)=\left(\theta_1 u_1^{-1/\alpha}+\cdots+\theta_n u_n^{-1/\alpha}\right)^{-\alpha}/\left(\sum_{i=1}^n\theta_i\right)^{-\alpha}$, $u_i\in(0,1)$ for $i=1,\ldots,n$. Let $\underline{\boldsymbol{\theta}}=\min(\boldsymbol{\theta}/\left(\sum_{i=1}^n\theta_i\right))$. With the classic averaging inequalities, for $0<\alpha_1<\alpha_2$, $M_{\alpha_1,\boldsymbol{\theta}}\leqslant M_{\alpha_2,\boldsymbol{\theta}}$ (Hardy et al. (1934), Theorem 16) and $\underline{\boldsymbol{\theta}}^{\alpha_1}M_{\alpha_1,\boldsymbol{\theta}}\geqslant\underline{\boldsymbol{\theta}}^{\alpha_2}M_{\alpha_2,\boldsymbol{\theta}}$ (Hardy et al. (1934), Theorem 23). We note that $0<M_{\alpha,\boldsymbol{\theta}}<1$ and these two inequalities are directly translated to

$$\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha_2,\boldsymbol{\theta}})^{\alpha_2/\alpha_1}\leqslant\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha_1,\boldsymbol{\theta}})\leqslant\underline{\boldsymbol{\theta}}^{1-\alpha_2/\alpha_1}\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha_2,\boldsymbol{\theta}})^{\alpha_2/\alpha_1}.$$

By letting $\alpha_1\uparrow\alpha_2$ and $\alpha_2\downarrow\alpha_1$, we get the continuity of $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$ in $\alpha>0$.

(iii) Without loss of generality, we assume $\boldsymbol{\theta}_1=(\theta_1,\ldots,\theta_n)$ and $\boldsymbol{\theta}_2=(\lambda\theta_1,\ldots,\theta_n)$, $\lambda>0$. The monotonicity relative to $\boldsymbol{\theta}$ follows directly from Proposition 5.1. Using the homogeneity of $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})$, which is proved in (iv), and the monotonicity with respect to $\boldsymbol{\theta}$ if $0<\lambda<1$,

$$\lambda\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}_1})\leqslant\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}_2})\leqslant\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}_1}),$$

otherwise

$$\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}_1})\leqslant\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}_2})\leqslant\lambda\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}_1}).$$

By letting $\lambda\uparrow 1$ and $\lambda\downarrow 1$, we get the desired result.

(iv) For $\lambda>0$,

$$\begin{aligned}\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\lambda\boldsymbol{\theta}}) &= \sup\{\mathrm{VaR}_p(G) : G\in\mathcal{D}_n(\mathbf{P}_{\alpha,\lambda\boldsymbol{\theta}})\}\\ &= \sup\left\{\mathrm{VaR}_p\left(G\left(\frac{\cdot}{\lambda}\right)\right) : G\in\mathcal{D}_n(\mathbf{P}_{\alpha,\boldsymbol{\theta}})\right\}\\ &= \lambda\sup\{\mathrm{VaR}_p(G) : G\in\mathcal{D}_n(\mathbf{P}_{\alpha,\boldsymbol{\theta}})\}=\lambda\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}}).\end{aligned}$$

(v) For $\alpha>1$, $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})\leqslant\overline{\mathrm{ES}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})=\sum_{i=1}^n\mathrm{ES}_p(P_{\alpha,\theta_i})=\alpha\sum_{i=1}^n\theta_i/\left((\alpha-1)(1-p)^{1/\alpha}\right)$, and $\overline{\mathrm{VaR}}_p(\mathbf{P}_{\alpha,\boldsymbol{\theta}})\geqslant\sum_{i=1}^n\mathrm{VaR}_p(P_{\alpha,\theta_i})=\sum_{i=1}^n\theta_i/(1-p)^{1/\alpha}$. □

# References

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements. `https://www.bis.org/bcbs/publ/d457.htm`

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.

Blanchet, J., Lam, H., Liu, Y. and Wang, R. (2020). Convolution bounds on quantile aggregation. *arXiv:* 2007.09320.

Cai, J., Liu, H. and Wang, R. (2018). Asymptotic equivalence of risk measures under dependence uncertainty. *Mathematical Finance*, **28**(1), 29–49.

Delbaen, F., Bellini, F., Bignozzi, V. and Ziegel, J. (2016). Risk measures with convex level sets. *Finance and Stochastics*, **20**(2), 433–453.

Eckstein, S., Kupper, M. and Pohl, M. (2020). Robust risk aggregation with neural networks. *Mathematical Finance*, published online at `doi.org/10.1111/mafi.12280`.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Heidelberg.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* Fourth Edition. Walter de Gruyter, Berlin.

Hardy, G. H., Littlewood, J. E., Pólya, G. (1934). *Inequalities.* Cambridge University Press.

Mao, T., Wang, B. and Wang, R. (2019). Sums of uniform random variables. *Journal of Applied Probability*, **56**(3), 918–936.

Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

Marshall, A. W., Olkin, I. and Arnold, B. (2011). *Inequalities: Theory of Majorization and Its Applications.* Springer, 2nd edition.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools.* Revised Edition. Princeton, NJ: Princeton University Press.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Statistical Models and Risks.* Wiley, England.

Pflug, G. C. and Pohl, M. (2018). A review on ambiguity in stochastic portfolio optimization. *Set-Valued and Variational Analysis*, **26**(4), 733–757.

Puccetti, G. and Rüschendorf, L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, **236**(7), 1833–1840.

Puccetti, G. and Rüschendorf, L. (2013). Sharp bounds for sums of dependent risks. *Journal of Applied Probability,* **50**(1), 42–53.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders.* Springer Series in Statistics.

Vovk, V. and Wang, R. (2020). Combining p-values via averaging. *Biometrika*, **107**(4), 791–808.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.

Wang, R., Wei, Y. and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.

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. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**(3), 1413–1429.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

Ziegel, J. (2016). Coherence and elicitability. *Mathematical Finance*, **26**, 901–918.