Inf-convolution, Optimal Allocations, and Model Uncertainty for Tail Risk Measures

Fangda Liu* Tiantian Mao† Ruodu Wang‡ Linxiao Wei§

January 16, 2022

Abstract

Inspired by the recent developments in risk sharing problems for the Value-at-Risk (VaR), the Expected Shortfall (ES), or the Range-Value-at-Risk (RVaR), we study the optimization of risk sharing for general tail risk measures. Explicit formulas of the inf-convolution and Pareto-optimal allocations are obtained in the case of a mixed collection of left and right VaRs, and in that of a VaR and another tail risk measure. The inf-convolution of tail risk measures is shown to be a tail risk measure with an aggregated tail parameter, a phenomenon very similar to the cases of VaR, ES and RVaR. Optimal allocations are obtained in the settings of elliptical models and model uncertainty. In particular, several results are established for tail risk measures in the presence of model uncertainty, which may be of independent interest outside the framework of risk sharing. The technical conclusions are quite general without assuming any form of convexity of the tail risk measures. Our analysis generalizes in several directions the recent literature on quantile-based risk sharing.

Key-words: Risk sharing, Pareto optimality, Value-at-Risk, Range-Value-at-Risk, non-convex optimization

1 Introduction

Over the past two decades, tail risk measures, in particular the Value-at-Risk (VaR) and the Expected Shortfall (ES), play a prominent role as the standard risk metrics in global banking and

*Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ fangda.liu@uwaterloo.ca.

†International Institute of Finance, School of Management, University of Science and Technology of China. ✉ tmao@ustc.edu.cn.

‡Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ wang@uwaterloo.ca.

§College of Science, Wuhan University of Technology, China. ✉ lxwei@whut.edu.cn.

insurance regulatory frameworks, such as Basel III/IV (BCBS (2019)) and Solvency II (EIOPA (2011)). Recently, Embrechts et al. (2018) developed analytical formulas and equilibrium allocations for risk sharing games with quantile-based risk measures, which are special cases of tail risk measures, via the mathematical tool of inf-convolution. The inf-convolution of risk measures $\rho_1, \dots, \rho_n$ on a domain $\mathcal{X}$ is defined as

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \inf \left\{ \sum_{i=1}^{n} \rho_i(X_i) : X_1, \dots, X_n \in \mathcal{X}, \ \sum_{i=1}^{n} X_i = X \right\}, \quad X \in \mathcal{X}. \tag{1}$$

See Section 2 for precise definitions of risk measures (including RVaR and VaR below) and inf-convolution. Inf-convolution (or sup-convolution) is closely related to the problem of risk sharing; see e.g., the monographs Starr (2011), Delbaen (2012) and Rüschendorf (2013). In particular, its minimizers correspond to Pareto-optimal allocations in a risk sharing problem for finite monetary risk measures, which are often also equilibrium allocations in non-cooperative games (see Remark 2).

A main result in Embrechts et al. (2018) is the RVaR inf-convolution formula

$$\mathop{\square}_{i=1}^{n} \mathrm{RVaR}_{\alpha_i,\beta_i} = \mathrm{RVaR}_{\alpha,\beta}, \tag{2}$$

where $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \geqslant 0$, $\alpha = \sum_{i=1}^n \alpha_i$, and $\beta = \bigvee_{i=1}^n \beta_i := \max\{\beta_1, \dots, \beta_n\}$; see Theorem 2 of Embrechts et al. (2018). A special case of (2), where $\beta_1 = \dots = \beta_n = 0$, is the following (left) VaR inf-convolution formula (Corollary 2 of Embrechts et al. (2018)),

$$\mathop{\square}_{i=1}^{n} \mathrm{VaR}^L_{\alpha_i} = \mathrm{VaR}^L_{\alpha}. \tag{3}$$

In short, both RVaR and (left) VaR have the nice feature that their inf-convolution is again in the same class, with the parameter being an aggregation (either sum or max) of the individual parameters. Moreover, an optimal allocation always exists in explicit form if $\alpha + \beta < 1$.

In this paper, we generalize the formulas (2) and (3) in several directions within the context of tail risk measures developed by Liu and Wang (2021). In particular, we aim to answer the following natural questions arising from (2) and (3).

1. Embrechts et al. (2018) defined the risk measure VaR as the left-quantile in (3), denoted by VaR^L in this paper (we omit the probability level here). One naturally wonders whether the nice formula (3) also holds for right-quantiles, which shall be denoted by VaR^R, and more generally, for a mixed collection of VaR^L and VaR^R. For the role of left and right quantiles as risk measures, see the discussion in Acerbi and Tasche (2002). Moreover, the existence of

an optimal allocation needs to be discussed, as neither VaR^L nor VaR^R is convex. Note that even the simple case of $\mathrm{VaR}^L \square \mathrm{VaR}^R$ or $\mathrm{VaR}^R \square \mathrm{VaR}^R$ is not studied in the literature.

2. Is there an explicit formula for the inf-convolution of (left or right) VaR and a general tail risk measure, and does an optimal allocation always exist? The inf-convolution of VaR^L and a tail distortion risk measure (defined in Appendix A) is obtained explicitly by Wang and Wei (2020, Theorem 5.3), and it is not clear whether this result holds with greater generality.

3. Suppose that $\rho_1, \dots, \rho_n$ are generic tail risk measures. Is their inf-convolution $\square_{i=1}^n \rho_i$ again a tail risk measure? If yes, is the corresponding tail parameter (defined in Section 2) an aggregation of the individual tail parameters, like in the cases of (2) and (3)?

4. The above three questions do not use any convexity assumptions on the risk measures, noting that VaR and RVaR are generally not convex. Would some results above on the inf-convolution of tail risk measures be improved if the underlying risk measures are convex, or the inf-convolution is constrained to be comonotonic? Note that an unconstrained optimal allocation is comonotonic if the risk measures are convex (e.g., Rüschendorf (2013, Theorem 10.52)), and hence these two questions should share the same answer.

5. In risk management practice, a risk allocation often has some concrete structures for modeling tractability. Would the additional imposed structure yield explicit formulas for the inf-convolution and the optimal allocations?[1]

6. Model uncertainty is prevalent in risk management. If model uncertainty, in some form, is incorporated in the risk sharing problem, meaning that the agents are uncertain about the distributions of the random losses allocated to them, how would the optimal allocations and the inf-convolution change?

This paper is dedicated to theoretical results which answer the above six questions. For the economic interpretation of inf-convolution and risk sharing problems for risk measures, we refer the reader to Embrechts et al. (2018), Baes et al. (2020), Ghamami and Glasserman (2019) and the references therein. Some recent work on quantile-based risk sharing includes Weber (2018), Embrechts et al. (2020), and Wang and Wei (2020). An application to insurance networks is studied by Hamm et al. (2020).

[1] We are grateful to an anonymous referee for raising questions 5 and 6.

Our main results in Theorems 1-6 provide answers to the above six questions, respectively. Most notably, in the case of a mixed collection of VaR^L and VaR^R (Theorem 1), and the case of VaR^L or VaR^R and another tail risk measure (Theorem 2), we obtain explicit forms of the inf-convolution as well as the corresponding optimal allocations. In particular, for $\Lambda_1, \dots, \Lambda_n \in \{L, R\}$, we obtain the simple formula

$$\mathop{\square}_{i=1}^{n} \mathrm{VaR}_{\alpha_i}^{\Lambda_i} = \mathrm{VaR}_{\alpha}^{\Lambda}, \quad \text{where } \Lambda = L \text{ if } \Lambda_1 = \dots = \Lambda_n = L, \text{ and } \Lambda = R \text{ otherwise,}$$

and an optimal allocation of X has the form

$$X_i = (X - \mathrm{VaR}_{\alpha}^{\Lambda}(X))\mathbb{1}_{A_i} + \frac{1}{n}\mathrm{VaR}_{\alpha}^{\Lambda}(X), \quad i = 1, \dots, n,$$

for some partition $(A_1, \dots, A_n)$ of the sample space. Moreover, an optimal allocation always exists for VaR^L and a tail risk measure, but generally not so for VaR^R and a tail risk measure. From these results, we discover the somewhat surprising fact that the roles of VaR^L and VaR^R are asymmetric in risk sharing, and hence a separate analysis in this paper is necessary to fully understand risk sharing problems involving quantiles.

There are many tail risk measures, as one can generate a tail risk measure from any law-invariant risk measure. With such generality, explicit forms of the inf-convolution, or existence results of the optimal allocation for generic tail risk measures are not available. Nevertheless, we show that the inf-convolution of tail risk measures exhibits nice properties similar to those in (2) and (3). Precisely, the tail parameter of the inf-convolution is an aggregation of those of individual risk measures, and the aggregation is a summation (Theorem 3) in the case of unconstrained inf-convolution, like the α-parameter in (2), whereas it is a maximum (Theorem 4) in the case of constrained (comonotonic) inf-convolution, like the β-parameter in (2).

In case allocations have the additional structure of multivariate elliptical models, optimal allocations for tail risk measures are obtained (Theorem 5). Two popular settings of model uncertainty modeled by either bounds on likelihood ratios or Wasserstein metrics are analyzed with a worst-case approach. In Table 2, we summarize how the two types of model uncertainty affect tail risk measures. Model uncertainty induced by likelihood ratios can be analyzed with existing methods. However, for non-convex risk measures (such as VaR), model uncertainty induced by Wasserstein metrics is difficult to analyze. We obtain optimal allocations in a specific setting of comonotonic allocations among several VaR agents (Theorem 6).

The rest of the paper is organized in a straightforward manner. In Section 2, we collect necessary definitions and notation. Sections 3-8 contain our theoretical results which provide answers to the six questions posed above. Section 9 contains some concluding remarks. Some background information on risk measures is put in Appendix A and the proofs of all lemmas, propositions and corollaries are relegated to Appendix B.

2 Preliminaries

2.1 Notation

We work with an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let L^q be the set of all random variables in $(\Omega, \mathcal{F}, \mathbb{P})$ with finite q-th moment, $q \in (0, \infty)$, L^0 be the set of all random variables, and L^∞ be the set of essentially bounded random variables. Throughout, for any $X \in L^0$, a positive (negative) value of X represents a financial loss (profit), F_X represents the distribution function of X, and its left-continuous inverse is given by

$$F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geqslant p\}, \ p \in (0,1], \quad \text{and} \quad F_X^{-1}(0) = \inf\{x \in \mathbb{R} : F_X(x) > 0\}.$$

Let U_X be a uniform random variable on $[0,1]$ such that $F_X^{-1}(U_X) = X$ almost surely (a.s.). The existence of such uniform random variable U_X for any X is given, for instance, in Lemma A.32 of Föllmer and Schied (2016). The mappings ess-inf$(\cdot)$ and ess-sup$(\cdot)$ on L^0 stand for the essential infimum and the essential supremum of a random variable, respectively. We denote by $X \stackrel{\mathrm{d}}{=} Y$ if the random variables X and Y have the same distribution. For $x, y \in \mathbb{R}$, $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$.

2.2 Risk measures

Let $\mathcal{X}$ be a convex cone of random variables containing L^∞. A *risk measure* ρ is a mapping from $\mathcal{X}$ to $[-\infty, \infty)$.[2] Whenever a risk measure appears in this paper, its domain is $\mathcal{X}$ unless otherwise specified. We assume that for $X \in \mathcal{X}$, if $Y \stackrel{\mathrm{d}}{=} X$, then $Y \in \mathcal{X}$. This is certainly satisfied by commonly used domains of risk measures, such as $\mathcal{X} = L^q$, $q \in [0, \infty]$.

For a risk measure $\rho : \mathcal{X} \to [-\infty, \infty)$, we say that ρ is (i) *law-invariant* if $\rho(X) = \rho(Y)$ for all $X, Y \in \mathcal{X}$ with $X \stackrel{\mathrm{d}}{=} Y$; (ii) *monotone* if $\rho(X) \leqslant \rho(Y)$ for all $X, Y \in \mathcal{X}$ with $X \leqslant Y$ a.s.; (iii) *translation-invariant* if $\rho(X - m) = \rho(X) - m$ for all $m \in \mathbb{R}$ and $X \in \mathcal{X}$. Moreover, a risk measure ρ is a *monetary risk measure* if it is monotone and translation-invariant. Some other commonly used properties for risk measures are collected in Appendix A. For economic interpretations of these

[2]Typically, we are only interested in risk measures that take finite values. The reason that we include $-\infty$ in the range of a risk measure is to allow for an inf-convolution to be considered as a risk measure; see Section 2.3.

properties, we refer to Artzner et al. (1999), Föllmer and Schied (2016) and Delbaen (2012). In this paper, we would like to impose as little requirement on the risk measures as possible (except that they are tail risk measures) to aim for great generality. In particular, we will not assume any form of convexity or quasi-convexity in the majority of the paper.

Remark 1. Some researchers argue that "law dependence" is a more accurate term for "law invariance", and "translation equivariance" is a more accurate term for "translation invariance"; see e.g., Remark 39 of Delbaen (2012). In this paper, we stick to the most commonly used terminology in the risk measure literature, bearing in mind that they may not be perfect.

The two most popular classes of risk measures used in banking and insurance practice are the *Value-at-Risk (VaR)* and the *Expected Shortfall (ES)*; for a general treatment of these risk measures in risk management and optimization, we refer to Pflug and Römisch (2007) and McNeil et al. (2015). For a confidence level $\alpha \in [0,1]$, VaR has two versions: the right-quantile (VaR^R_α) and the left-quantile (VaR^L_α). The left-VaR at level $\alpha \in [0,1]$ is defined as

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

and the right-VaR at level $\alpha \in [0,1]$ is defined as

$$\mathrm{VaR}^R_\alpha(X) = \inf\{x \in \mathbb{R} : F_X(x) > 1-\alpha\}, \quad X \in L^0,$$

where $\inf(\varnothing) = \infty$. In addition, let ess-sup $= \mathrm{VaR}^L_0$ and ess-inf $= \mathrm{VaR}^R_1$. In risk management practice, one typically does not distinguish between VaR^R_α and VaR^L_α as they are identical for random variables with a continuous inverse distribution function at α. However, their subtle difference leads to interesting observations; see Remark 5 after Theorem 1. Both VaR^R_α and VaR^L_α will be referred to as $(1-\alpha)$-quantiles or VaRs in this paper.

Next we define the family of *Range-Value-at-Risk (RVaR)*, introduced by Cont et al. (2010) as a family of robust risk measures. Following the notation in Embrechts et al. (2018, 2020), the RVaR at level $(\alpha,\beta) \in [0,1]^2$ with $\alpha+\beta \leqslant 1$ is defined as

$$\mathrm{RVaR}_{\alpha,\beta}(X) = \begin{cases} \frac{1}{\beta}\int_\alpha^{\alpha+\beta} \mathrm{VaR}^L_q(X)\,\mathrm{d}q, & \text{if } \beta > 0, \\ \mathrm{VaR}^L_\alpha(X), & \text{if } \beta = 0, \end{cases} \qquad X \in L^1.$$

For $\beta \in [0,1)$, the Expected Shortfall is defined as $\mathrm{ES}_\beta = \mathrm{RVaR}_{0,\beta}$. Both VaR^R_α and VaR^L_α are connected to ES_α via an optimization formula; see Rockafellar and Uryasev (2000) and Pflug (2000). Whenever RVaR and ES appear in this paper, their domain is set to $\mathcal{X} = L^1$ to guarantee finiteness. In Embrechts et al. (2018), the parameters α and β of RVaR are allowed to be greater than 1 (leading to $\mathrm{RVaR}_{\alpha,\beta} = -\infty$). For the results in our paper, it is sufficient to consider only the case $\alpha+\beta \leqslant 1$.

2.3 Risk sharing and inf-convolution

In risk management and game theory, a risk sharing (or risk allocation) problem is typically referred to distributing a given aggregate risk to multiple agents so that their own risk measures (or utilities) are optimized. A popular tool for an analysis of the above problem is through minimizing the aggregate risk value. Mathematically, given a random variable $X \in \mathcal{X}$ representing the total random loss in the future, and a total number n of agents in this risk sharing game, we define the set of *allocations* of X as

$$\mathbb{A}_n(X) = \left\{ (X_1, \dots, X_n) \in \mathcal{X}^n : \sum_{i=1}^{n} X_i = X \right\}. \tag{4}$$

For $i = 1, \dots, n$, agent i is equipped with a risk measure $\rho_i : \mathcal{X} \to \mathbb{R}$. Each allocation from $\mathbb{A}_n(X)$ represents a certain way of splitting the aggregate risk X among n agents, and the associated *aggregate risk value* is $\sum_{i=1}^{n} \rho_i(X_i)$. Using the notation (4), the inf-convolution of risk measures (e.g. Delbaen (2012) and Rüschendorf (2013)) in (1) reads as

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \inf \left\{ \sum_{i=1}^{n} \rho_i(X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n(X) \right\}, \quad X \in \mathcal{X}. \tag{5}$$

Since the risk measures $\rho_1, \dots, \rho_n$ do not take the value ∞ on $\mathcal{X}$, the infimum in (5) is well posed. Note that $\square_{i=1}^{n} \rho_i$ may take the value $-\infty$; see e.g., the case of VaR agents in Corollary 2 of Embrechts et al. (2018). An n-tuple $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is called an *optimal allocation* of X for $(\rho_1, \dots, \rho_n)$ if $\sum_{i=1}^{n} \rho_i(X_i) = \square_{i=1}^{n} \rho_i(X)$. If risk measures are interpreted as the capital charge for a financial institution to take risky positions, as in Artzner et al. (1999) and BCBS (2019), then $\square_{i=1}^{n} \rho_i(X)$ represents the smallest possible aggregate capital for the total risk X in the financial system.

Remark 2. Fix risk measures $\rho_1, \dots, \rho_n$ and a total risk $X \in \mathcal{X}$. An allocation $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$ is *Pareto-optimal* with respect to $(\rho_1, \dots, \rho_n)$ if for any allocation $(Y_1, \dots, Y_n) \in \mathbb{A}_n(X)$, $\rho_i(Y_i) \leqslant \rho_i(X_i)$ for all $i = 1, \dots, n$ implies $\rho_i(Y_i) = \rho_i(X_i)$ for all $i = 1, \dots, n$. Pareto-optimal allocations are closely related to the inf-convolution defined in (5). In particular, assuming that each of $\rho_i(X_i)$, $i = 1, \dots, n$ is finite, $(X_1, \dots, X_n)$ is a Pareto-optimal allocation of X if and only if $(X_1, \dots, X_n)$ is an optimal allocation of X (Proposition 1 of Embrechts et al. (2018)). Moreover, the above optimal allocations are often equilibrium allocations in non-cooperative games. In particular, such an equivalence is generally true in a convex setting, and also true in some non-convex settings; see e.g., Xia and Zhou (2016) and Embrechts et al. (2018, 2020).

2.4 Tail risk measures

We follow the mathematical framework of tail risk measures developed by Liu and Wang (2021), where the following terminology for tail risks and tail risk measures is introduced. For a random variable $X \in \mathcal{X}$ and $p \in (0,1]$, X_p presents the *tail risk of* X beyond its $(1-p)$-quantile, defined by

$$X_p = F_X^{-1}(1 - p + pU_X).$$

One can easily check

$$\mathbb{P}(X_p \leqslant x) = \mathbb{P}(X \leqslant x | U_X \geqslant 1-p) = \frac{(\mathbb{P}(X \leqslant x) - (1-p))_+}{p}, \quad x \in \mathbb{R}. \tag{6}$$

We assume that $X \in \mathcal{X}$ implies $X_p \in \mathcal{X}$. This assumption is satisfied by common choices of $\mathcal{X}$, such as $\mathcal{X} = L^q$, $q \in [0, \infty]$.

Definition 1 (Liu and Wang (2021))**.** For $p \in (0,1)$, a risk measure ρ is a *p-tail risk measure* if $\rho(X) = \rho(Y)$ for all $X, Y \in \mathcal{X}$ satisfying $X_p \overset{\mathrm{d}}{=} Y_p$. In this case, we simply say that ρ is a *tail risk measure*, and the value p is a *tail parameter* of ρ.[3]

Tail risk measures are risk measures that depend solely on the tail distribution. For a p-tail risk measure ρ, there exists a law-invariant risk measure ρ^* such that $\rho(X) = \rho^*(X_p)$ for all $X \in \mathcal{X}$. Note that every number q in $[p, 1)$ is a tail parameter of ρ, and thus the concept of p-tail risk measure gets weaker as p increases. Therefore, we sometimes look for the smallest tail parameter of a tail risk measure, if it exists.

It is immediate from Definition 1 that VaRs, ES and RVaR are generally tail risk measures. In particular, for $\alpha, \beta > 0$ and $\alpha + \beta < 1$, VaR^R_α has a smallest tail parameter α, VaR^L_α has a tail parameter γ for all $\gamma > \alpha$ but not $\gamma = \alpha$, ES_β has a smallest tail parameter β, and $\mathrm{RVaR}_{\alpha,\beta}$ has a smallest tail parameter $\alpha + \beta$. In addition to VaRs, ES and RVaR, many other tail risk measures are studied in the recent risk management literature, such as the Glue-VaR (Belles-Sampera et al. (2014)) and the Gini Shortfall (Furman et al. (2017)); see the examples in Liu and Wang (2021). The notions of tail risk measures and the VaR-type risk measures introduced by Weber (2018) are complementary; see Appendix A.

[3] Here, we use a "small p" convention, which is different from Liu and Wang (2021), that a p-tail risk measure is a $(1-p)$-tail risk measure using the terminology of Liu and Wang (2021). The choice of "small p" convention is in agreement with Embrechts et al. (2018), and it makes many results in the paper more concise. This choice is suggested by two anonymous referees.

3 Inf-convolution of several VaRs

In this section, we study the inf-convolution of several VaRs, which can be a mixed collection of left- and right-quantiles. For this special case, explicit forms of the inf-convolution and the optimal allocation are available.

First, Corollary 2 of Embrechts et al. (2018) gives, for $\alpha_1, \alpha_2 > 0$ with $\alpha_1 + \alpha_2 < 1$,

$$\mathrm{VaR}^L_{\alpha_1} \square \mathrm{VaR}^L_{\alpha_2} = \mathrm{VaR}^L_{\alpha_1+\alpha_2}. \tag{7}$$

The inf-convolution of $\mathrm{VaR}^L_{\alpha_1}$ and $\mathrm{VaR}^R_{\alpha_2}$ and that of $\mathrm{VaR}^R_{\alpha_1}$ and $\mathrm{VaR}^R_{\alpha_2}$ are not available in the literature, and they are technically trickier. In the next result, we will obtain a general formula for the inf-convolution of several VaRs, which implies

$$\mathrm{VaR}^L_{\alpha_1} \square \mathrm{VaR}^R_{\alpha_2} = \mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^L_{\alpha_2} = \mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^R_{\alpha_2} = \mathrm{VaR}^R_{\alpha_1+\alpha_2}. \tag{8}$$

Moreover, a corresponding optimal allocation can be constructed explicitly. Note from (8) that the roles of left- and right-quantiles are indeed asymmetric in the problem of inf-convolution, and our analysis on the general mixed case completes the full picture of this problem.

Theorem 1. *Suppose that $X \in \mathcal{X}$, $\alpha_1, \dots, \alpha_n > 0$ with $\alpha = \sum_{i=1}^n \alpha_i < 1$, and $\Lambda_1, \dots, \Lambda_n \in \{L, R\}$.*

(i) $\square_{i=1}^n \mathrm{VaR}^{\Lambda_i}_{\alpha_i} = \mathrm{VaR}^{\Lambda}_{\alpha}$, *where $\Lambda = L$ if $\Lambda_1 = \dots = \Lambda_n = L$, and $\Lambda = R$ otherwise.*

(ii) There exists an optimal allocation of X for $(\mathrm{VaR}^{\Lambda_1}_{\alpha_1}, \dots, \mathrm{VaR}^{\Lambda_n}_{\alpha_n})$ which has the form

$$X_i = (X - \mathrm{VaR}^{\Lambda}_{\alpha}(X))\mathbb{1}_{A_i} + \frac{1}{n}\mathrm{VaR}^{\Lambda}_{\alpha}(X), \quad i = 1, \dots, n, \tag{9}$$

for some partition $(A_1, \dots, A_n)$ of Ω.

(iii) For any partition $(A_1, \dots, A_n)$ of Ω independent of X with $\mathbb{P}(A_i) = \alpha_i/\alpha$, $i = 1, \dots, n$, (9) gives an optimal allocation of X.

Proof. We first note that for any random variable Y and $\beta \in (0, 1)$,

$$\mathrm{VaR}^R_{\beta}(Y) \leqslant 0 \iff \mathbb{P}(Y > \varepsilon) < \beta \text{ for all } \varepsilon > 0, \tag{10}$$

and

$$\mathrm{VaR}^L_{\beta}(Y) \leqslant 0 \iff \mathbb{P}(Y > 0) \leqslant \beta. \tag{11}$$

Moreover, for any $\delta \in (0, \alpha_1)$, we have, by definition

$$\mathrm{VaR}^L_{\alpha_1-\delta} \geqslant \mathrm{VaR}^R_{\alpha_1} \geqslant \mathrm{VaR}^L_{\alpha_1}. \tag{12}$$

These facts will be used repeatedly in the proof.

(i) We first show (8). Using (7) and (12), we have

$$\mathrm{VaR}^R_{\alpha_1+\alpha_2-\delta} \geqslant \mathrm{VaR}^L_{\alpha_1+\alpha_2-\delta} = \mathrm{VaR}^L_{\alpha_1-\delta} \square \mathrm{VaR}^L_{\alpha_2} \geqslant \mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^L_{\alpha_2}.$$

Since the right-quantile VaR^R_β is left-continuous in β, we have

$$\mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^L_{\alpha_2} \leqslant \lim_{\delta \downarrow 0} \mathrm{VaR}^R_{\alpha_1+\alpha_2-\delta} = \mathrm{VaR}^R_{\alpha_1+\alpha_2}. \tag{13}$$

Next we show the inverse direction of the inequality in (13). Take any random variables $Y, Z \in \mathcal{X}$. For any $\varepsilon > 0$, using (10), we have

$$\begin{aligned}
\mathbb{P}(Y + Z > \mathrm{VaR}^R_{\alpha_1}(Y) + \mathrm{VaR}^L_{\alpha_2}(Z) + \varepsilon) &\leqslant \mathbb{P}(Y \geqslant \mathrm{VaR}^R_{\alpha_1}(Y) + \frac{\varepsilon}{2} \text{ or } Z > \mathrm{VaR}^L_{\alpha_2}(Z) + \frac{\varepsilon}{2}) \\
&\leqslant \mathbb{P}(Y \geqslant \mathrm{VaR}^R_{\alpha_1}(Y) + \frac{\varepsilon}{2}) + \mathbb{P}(Z > \mathrm{VaR}^L_{\alpha_2}(Z) + \frac{\varepsilon}{2}) \\
&= \mathbb{P}(Y - \mathrm{VaR}^R_{\alpha_1}(Y) \geqslant \frac{\varepsilon}{2}) + \mathbb{P}(Z - \mathrm{VaR}^L_{\alpha_2}(Z) > \frac{\varepsilon}{2}) \\
&< \alpha_1 + \alpha_2.
\end{aligned}$$

Using (10) again, we obtain

$$\mathrm{VaR}^R_{\alpha_1+\alpha_2}(Y + Z) \leqslant \mathrm{VaR}^R_{\alpha_1}(Y) + \mathrm{VaR}^L_{\alpha_2}(Z).$$

This, together with (13), implies our desired statement

$$\mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^L_{\alpha_2} = \mathrm{VaR}^R_{\alpha_1+\alpha_2}. \tag{14}$$

Next, using (12) and (14), we obtain

$$\mathrm{VaR}^R_{\alpha_1+\alpha_2-\delta} = \mathrm{VaR}^L_{\alpha_1-\delta} \square \mathrm{VaR}^R_{\alpha_2} \geqslant \mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^R_{\alpha_2} \geqslant \mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^L_{\alpha_2} = \mathrm{VaR}^R_{\alpha_1+\alpha_2}.$$

Using again the fact that the right-quantile VaR^R_β is left-continuous in β, we get

$$\mathrm{VaR}^R_{\alpha_1} \square \mathrm{VaR}^R_{\alpha_2} = \mathrm{VaR}^R_{\alpha_1+\alpha_2}.$$

Finally, by noting that inf-convolution is commutative, we have

$$\mathrm{VaR}^L_{\alpha_1} \square \mathrm{VaR}^R_{\alpha_2} = \mathrm{VaR}^R_{\alpha_2} \square \mathrm{VaR}^L_{\alpha_1} = \mathrm{VaR}^R_{\alpha_1+\alpha_2}.$$

Therefore, all quantities in (8) are equal. By Lemma 2 of Liu et al. (2020),

$$\mathop{\square}_{i=1}^{n} \mathrm{VaR}^{\Lambda_i}_{\alpha_i} = \mathrm{VaR}^{\Lambda_1}_{\alpha_1} \square \dots \square \mathrm{VaR}^{\Lambda_n}_{\alpha_n}.$$

The statement $\square_{i=1}^n \mathrm{VaR}^{\Lambda_i}_{\alpha_i} = \mathrm{VaR}^{\Lambda}_{\alpha}$, is obtained via a repeated application of (7) and (8).

(ii) If $\Lambda = L$, in which case $\Lambda_1 = \cdots = \Lambda_n = L$, the statement is shown by Embrechts et al. (2018). It suffices to consider the case $\Lambda = R$.

Take a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ with $\varepsilon_k \downarrow 0$ as $k \to \infty$. Let $B_k = \{X - \mathrm{VaR}^R_\alpha(X) > \varepsilon_k\}$ for $k \in \mathbb{N}$, and $B_0 = \varnothing$. By (10), we have $\mathbb{P}(B_k) < \alpha$ for all k, and $\{B_k\}_{k\in\mathbb{N}}$ is an increasing sequence of sets. Let $(A^k_1, \dots, A^k_n)$ be a partition of $B_k \setminus B_{k-1}$ for each $k \in \mathbb{N}$, satisfying

$$\mathbb{P}(A^k_i) = \frac{\alpha_i}{\alpha}\mathbb{P}(B_k \setminus B_{k-1}), \quad i = 1, \dots, n,\ k \in \mathbb{N}.$$

The existence of such a sequence $\{(A^k_1, \dots, A^k_n)\}_{k\in\mathbb{N}}$ is guaranteed since the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless. Note that for each $K \in \mathbb{N}$,

$$\mathbb{P}\left(\bigcup_{k=1}^{K} A^k_i\right) = \frac{\alpha_i}{\alpha}\mathbb{P}(B_K) < \alpha_i.$$

Let $(C_1, \dots, C_n)$ be an arbitrary partition of $\{X \leqslant \mathrm{VaR}^R_\alpha(X)\}$, and $A_i = (\bigcup_{k\in\mathbb{N}} A^k_i) \cup C_i$ for $i = 1, \dots, n$. Note that by construction, $(A_1, \dots, A_n)$ is a partition of Ω. For all $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that $\varepsilon_K < \varepsilon$, and we have, for $i = 1, \dots, n$,

$$\begin{aligned}\mathbb{P}\left((X - \mathrm{VaR}^R_\alpha(X))\mathbb{1}_{A_i} > \varepsilon\right) &\leqslant \mathbb{P}\left((X - \mathrm{VaR}^R_\alpha(X))\mathbb{1}_{A_i} > \varepsilon_K\right)\\ &= \mathbb{P}(B_K \cap A_i) = \mathbb{P}\left(\bigcup_{k=1}^{K} A^k_i\right) < \alpha_i.\end{aligned}$$

Thus, we obtain

$$\mathbb{P}\left((X - \mathrm{VaR}^R_\alpha(X))\mathbb{1}_{A_i} > \varepsilon\right) < \alpha_i. \tag{15}$$

Using (10), we get

$$\mathrm{VaR}^R_{\alpha_i}((X - \mathrm{VaR}^R_\alpha(X))\mathbb{1}_{A_i}) \leqslant 0. \tag{16}$$

Next, define $X_1, \dots, X_n$ according to (9). We can easily check

$$\sum_{i=1}^{n} X_i = (X - \mathrm{VaR}^R_\alpha(X))\sum_{i=1}^{n}\mathbb{1}_{A_i} + \mathrm{VaR}^R_\alpha(X) = X.$$

Thus $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$. Using (16), we have

$$\sum_{i=1}^{n}\mathrm{VaR}^R_{\alpha_i}(X_i) \leqslant \sum_{i=1}^{n}\frac{1}{n}\mathrm{VaR}^R_\alpha(X) = \mathrm{VaR}^R_\alpha(X).$$

Using the result in part (i), we know that $(X_1, \dots, X_n)$ is an optimal allocation of X.

(iii) If $\Lambda = L$, in which case $\Lambda_1 = \cdots = \Lambda_n = L$, the conclusion follows from the fact that $\mathrm{VaR}^L_{\alpha_i}(X_i) = \frac{1}{n}\mathrm{VaR}^L_\alpha(X)$, which is easy to check. Next we suppose that $\Lambda = R$. For each $i = 1, \ldots, n$, using (10), we obtain, for all $\varepsilon > 0$,

$$\mathbb{P}\left((X - \mathrm{VaR}^R_\alpha(X))\mathbb{1}_{A_i} > \varepsilon\right) = \mathbb{P}\left(X - \mathrm{VaR}^R_\alpha(X) > \varepsilon\right)\mathbb{P}(A_i) < \alpha \times \frac{\alpha_i}{\alpha} = \alpha_i.$$

Hence, (15) holds. Following the same argument as in part (ii), we conclude that $(X_1, \ldots, X_n)$ is an optimal allocation of X. □

Remark 3. In Theorem 1 (iii), the existence of $(A_1, \ldots, A_n)$ independent of X is guaranteed if there exists a uniform random variable independent of X. The assumption that the probability space is atomless is not necessary, as long as $(A_1, \ldots, A_n)$ in the statement exists.

Theorem 1 (iii) directly gives the existence and the form of an optimal allocation under a technical condition that there exists a uniform random variable independent of X, which we call condition (E) below for simplicity. From its proof, we can see that Theorem 1 (ii) requires a much more complicated construction of $(A_1, \ldots, A_n)$ than (iii). Certainly, condition (E) is very weak, and it is satisfied in all practical situations. In particular, if one is allowed to extend the underlying probability space after X is specified (e.g., "throwing a die"), then (E) always holds.

Remark 4. We note that, as discussed in Liu et al. (2020), (E) cannot be taken for granted in general. Indeed, in a standard probability space (i.e., a probability space isomorphic to $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ is the Lebesgue measure), one can always find a (very special) random variable X such that (E) does not hold; see Example 7 of Liu et al. (2020). In case that (E) is not satisfied, the partition $(A_1, \ldots, A_n)$ in the optimal allocation (9) may not be explicit.

Remark 5. As we mentioned above, the distinction between two α-quantiles is usually ignored in risk management practice. Below we provide a motivation which leads to an interesting interpretation of Theorem 1. The distinction between VaR^L_α and VaR^R_α arises when the distribution function of the underlying risk is strictly flat across a certain range of outcomes at level α. Such a situation is unlikely to arise in practice, but as risks are estimated from statistical models and data, it may well happen that a risk distribution is approximately flat around α. In such a situation, taking model uncertainty into account, there would be a fairly large interval (a "confidence interval") of possible values for the quantile, even if in theory it is unique. In this case, VaR^L_α can be thought of as an idealized representation of the left end-point of the confidence interval, whereas VaR^R_α represents the right end-point. A strict regulator can be expected to impose VaR^R_α, whereas a lenient regulator might accept VaR^L_α. The result of Theorem 1 can be interpreted as that, if all agents in a risk sharing game are under a lenient regime, then effectively the same applies to the representative agent, but if at least one of the agents has a strict regulator, then the strict regime effectively holds for the representative as well.

Although an optimal allocation for a mixed collection of VaRs always exists, such an existence result cannot be expected for general tail risk measures. This will be discussed in the next section, where we will see the intriguing fact that an optimal allocation always exists for $(\mathrm{VaR}^L_\alpha, \rho)$ but not necessarily for $(\mathrm{VaR}^R_\alpha, \rho)$, where ρ is a generic monetary tail risk measure.

4 Inf-convolution of VaR and another tail risk measure

In this section, we analyze the inf-convolution of VaR and another monetary tail risk measure. Similar to the case in Section 3, we obtain explicit formulas for the inf-convolution and the optimal allocation.

Theorem 2. *Suppose that $\varepsilon \in (0,1)$ and ρ is a monetary ε-tail risk measure. For $X \in \mathcal{X}$ and $\alpha \in (0, 1-\varepsilon)$, write $X^{[\alpha]} = X\mathbb{1}_{\{U_X \leqslant 1-\alpha\}} + \mathrm{VaR}^R_{\alpha+\varepsilon}(X)\mathbb{1}_{\{U_X > 1-\alpha\}}$, and the following statements hold.*

(i) $\mathrm{VaR}^L_\alpha \,\Box\, \rho(X) = \rho(X^{[\alpha]})$.

(ii) $(X - X^{[\alpha]}, X^{[\alpha]})$ *is an optimal allocation of X for* $(\mathrm{VaR}^L_\alpha, \rho)$.

(iii) $\mathrm{VaR}^R_\alpha \,\Box\, \rho(X) = \lim_{\delta\downarrow 0} \rho(X^{[\alpha-\delta]})$.

(iv) Both $\mathrm{VaR}^R_\alpha \,\Box\, \rho$ *and* $\mathrm{VaR}^L_\alpha \,\Box\, \rho$ *are monetary $(\alpha+\varepsilon)$-tail risk measures.*

Proof. We note that, for $x \in \mathbb{R}$,

$$\mathbb{P}(X^{[\alpha]} \leqslant x) = \begin{cases} F_X(x), & x < \mathrm{VaR}^R_{\alpha+\varepsilon}(X) \\ F_X(x) + \alpha, & \mathrm{VaR}^R_{\alpha+\varepsilon}(X) \leqslant x < \mathrm{VaR}^L_\alpha(X) \\ 1, & x \geqslant \mathrm{VaR}^L_\alpha(X). \end{cases} \tag{17}$$

Moreover, by construction,

$$\mathrm{VaR}^R_\varepsilon(X^{[\alpha]}) = \mathrm{VaR}^R_{\alpha+\varepsilon}(X). \tag{18}$$

We will use these formulas in several parts of the proof.

(i) Denote by $X_1 = X - X^{[\alpha]} = (X - \mathrm{VaR}^R_{\alpha+\varepsilon}(X))\mathbb{1}_{\{U_X > 1-\alpha\}}$. Since

$$\mathbb{P}(X_1 > 0) \leqslant \mathbb{P}(U_X > 1-\alpha) = \alpha,$$

we have $\mathrm{VaR}^L_\alpha(X_1) \leqslant 0$. It follows that

$$\mathrm{VaR}^L_\alpha \,\square\, \rho(X) \leqslant \mathrm{VaR}^L_\alpha(X_1) + \rho(X^{[\alpha]}) \leqslant \ \rho(X^{[\alpha]}).$$

Next, we show $\rho(X^{[\alpha]}) \leqslant \mathrm{VaR}^L_\alpha \,\square\, \rho(X)$. For this, it suffices to prove $\rho(X^{[\alpha]}) \leqslant \rho(X - Y)$ for all $Y \in \mathcal{X}$ with $\mathrm{VaR}^L_\alpha(Y) = 0$ due to translation-invariance of ρ. Since $\mathrm{VaR}^L_\alpha(Y) = 0$ implies $\mathbb{P}(Y > 0) \leqslant \alpha$, we have

$$\mathbb{P}(X - Y > x) \geqslant (\mathbb{P}(X > x) - \mathbb{P}(Y > 0))_+ \geqslant (\mathbb{P}(X > x) - \alpha)_+, \qquad \text{for } x \in \mathbb{R}.$$

As a consequence, for $x \geqslant \mathrm{VaR}^R_{\alpha+\varepsilon}(X)$, using (17), we have

$$\begin{aligned}\mathbb{P}(X - Y \leqslant x) &\leqslant 1 - (\mathbb{P}(X > x) - \alpha)_+ \\ &= (1 - \mathbb{P}(X > x) + \alpha) \wedge 1 \\ &= (\mathbb{P}(X \leqslant x) + \alpha) \wedge 1 = \mathbb{P}(X^{[\alpha]} \leqslant x),\end{aligned}$$

that is, $F_{X-Y}(x) \leqslant F_{X^{[\alpha]}}(x)$ for all $x \geqslant \mathrm{VaR}^R_{\alpha+\varepsilon}(X)$. Since ρ is a monotone ε-tail risk measure and $\mathrm{VaR}^R_\varepsilon(X^{[\alpha]}) = \mathrm{VaR}^R_{\alpha+\varepsilon}(X)$ by (18), we have $\rho(X^{[\alpha]}) \leqslant \mathrm{VaR}^L_\alpha \,\square\, \rho(X)$. Therefore, we conclude that $\rho(X^{[\alpha]}) = \mathrm{VaR}^L_\alpha \,\square\, \rho(X)$.

(ii) The optimality of $(X - X^{[\alpha]}, X^{[\alpha]})$ is obtained by $\rho(X^{[\alpha]}) = \mathrm{VaR}^L_\alpha \,\square\, \rho(X)$ in part (i).

(iii) Note that $\mathrm{VaR}^L_\alpha \leqslant \mathrm{VaR}^R_\alpha \leqslant \mathrm{VaR}^L_{\alpha-\delta}$ for all $\delta \in (0, \alpha)$. Therefore, using part (i), we have

$$\mathrm{VaR}^R_\alpha \,\square\, \rho(X) \leqslant \mathrm{VaR}^L_{\alpha-\delta} \,\square\, \rho(X) = \rho(X^{[\alpha-\delta]}).$$

Noting that $X^{[\alpha-\delta]} \downarrow X^{[\alpha]}$ a.s. as $\delta \downarrow 0$, and ρ is monotone, we know that the limit of $\rho(X^{[\alpha-\delta]})$ as $\delta \downarrow 0$ exists, which gives the inequality $\mathrm{VaR}^R_\alpha \,\square\, \rho(X) \leqslant \lim_{\delta\downarrow 0} \rho(X^{[\alpha-\delta]})$. On the other hand, we have

$$\begin{aligned}\lim_{\delta\downarrow 0} \rho(X^{[\alpha-\delta]}) = \lim_{\delta\downarrow 0} \mathrm{VaR}^L_{\alpha-\delta} \,\square\, \rho(X) &= \lim_{\delta\downarrow 0} \inf_{X_1 \in \mathcal{X}} \left(\mathrm{VaR}^L_{\alpha-\delta}(X_1) + \rho(X - X_1)\right) \\ &\leqslant \inf_{X_1 \in \mathcal{X}} \left(\lim_{\delta\downarrow 0} \mathrm{VaR}^L_{\alpha-\delta}(X_1) + \rho(X - X_1)\right) \\ &= \inf_{X_1 \in \mathcal{X}} \left(\mathrm{VaR}^R_\alpha(X_1) + \rho(X - X_1)\right) \\ &= \mathrm{VaR}^R_\alpha \,\square\, \rho(X).\end{aligned}$$

Therefore, we conclude that $\mathrm{VaR}^R_\alpha \,\square\, \rho(X) = \lim_{\delta\downarrow 0} \rho(X^{[\alpha-\delta]})$.

(iv) It is easy to check that the inf-convolution of monetary risk measures is still monetary. Take $Y, Z \in \mathcal{X}$ such that $Y_{\alpha+\varepsilon} \stackrel{\mathrm{d}}{=} Z_{\alpha+\varepsilon}$ (but $Y \stackrel{\mathrm{d}}{=} Z$ may not hold). For $\delta \in [0, \alpha)$, since

$$(F_Y(x) - (1-\varepsilon-\alpha+\delta))_+ = (F_Z(x) - (1-\varepsilon-\alpha+\delta))_+, \quad x \in \mathbb{R},$$

it is easy to check $(Y^{[\alpha-\delta]})_\varepsilon \stackrel{\mathrm{d}}{=} (Z^{[\alpha-\delta]})_\varepsilon$ through the following argument

$$\begin{aligned}
\mathbb{P}((Y^{[\alpha-\delta]})_\varepsilon \leqslant x) &= \frac{1}{\varepsilon}(\mathbb{P}(Y^{[\alpha-\delta]} \leqslant x) - (1-\varepsilon))_+ \\
&= \begin{cases} 0, & x < \mathrm{VaR}^R_{\alpha+\varepsilon-\delta}(Y) \\ \frac{1}{\varepsilon}(F_Y(x) - (1-\varepsilon-\alpha+\delta))_+, & \mathrm{VaR}^R_{\alpha+\varepsilon-\delta}(Y) \leqslant x < \mathrm{VaR}^L_{\alpha-\delta}(Y) \\ 1, & x \geqslant \mathrm{VaR}^L_{\alpha-\delta}(Y), \end{cases} \\
&= \mathbb{P}((Z^{[\alpha-\delta]})_\varepsilon \leqslant x).
\end{aligned}$$

Since ρ is an ε-tail risk measure, we have $\rho(Y^{[\alpha-\delta]}) = \rho(Z^{[\alpha-\delta]})$ for all $\delta \in [0, \alpha)$. Using part (i) and (iii), respectively, we know that $\mathrm{VaR}^L_\alpha \square \rho$ and $\mathrm{VaR}^R_\alpha \square \rho$ are $(\alpha+\varepsilon)$-tail risk measures. $\square$

Theorem 2 generalizes Theorem 5.3 of Wang and Wei (2020) where the inf-convolution of VaR^L_α and another distortion risk measure is studied (see Appendix A for the definition of distortion risk measures). In Theorem 2, we do not need to assume that ρ is a distortion risk measure. As a direct consequence of (i) and (iii), if ρ is continuous from above with respect to a.s. convergence, then

$$\mathrm{VaR}^R_\alpha \square \rho(X) = \mathrm{VaR}^L_\alpha \square \rho(X) = \rho(X^{[\alpha]}). \tag{19}$$

Remark 6. In Theorem 2 (iii), in addition to being $(\alpha+\varepsilon)$-tail risk measures, $\mathrm{VaR}^R_\alpha \square \rho$ and $\mathrm{VaR}^L_\alpha \square \rho$ are also VaR-type risk measures with parameter α according to the terminology of Weber (2018); see Appendix A.

From Theorem 2, we see that an optimal allocation always exists for $(\mathrm{VaR}^L_\alpha, \rho)$ where ρ is an ε-tail risk measure and $\alpha < 1-\varepsilon$. However, this is not the case for $(\mathrm{VaR}^R_\alpha, \rho)$, as we discuss below.

Let $\alpha, \beta > 0$ with $\alpha + \beta < 1$. By direct calculation, Theorem 2 gives the following formula

$$\mathrm{VaR}^L_\alpha \square \mathrm{ES}_\beta(X) = \mathrm{ES}_\beta(X^{[\alpha]}) = \mathrm{RVaR}_{\alpha,\beta}(X), \qquad X \in \mathcal{X},$$

and an optimal allocation of X for $(\mathrm{VaR}^L_\alpha, \mathrm{ES}_\beta)$; this result is also implied by Theorem 2 of Embrechts et al. (2018). Further, in the next proposition, we obtain a similar formula

$$\mathrm{VaR}^R_\alpha \square \mathrm{ES}_\beta(X) = \mathrm{RVaR}_{\alpha,\beta}(X), \qquad X \in \mathcal{X}. \tag{20}$$

Notably, an optimal allocation for $(\mathrm{VaR}^R_\alpha, \mathrm{ES}_\beta)$ often does not exist, in sharp contrast to the case of $(\mathrm{VaR}^L_\alpha, \mathrm{ES}_\beta)$, for which an optimal allocation always exists.

Proposition 1. *For $\alpha, \beta > 0$ with $\alpha + \beta < 1$, we have $\mathrm{VaR}^R_\alpha \Box \mathrm{ES}_\beta = \mathrm{RVaR}_{\alpha,\beta}$. Moreover, an optimal allocation of $X \in \mathcal{X}$ for $(\mathrm{VaR}^R_\alpha, \mathrm{ES}_\beta)$ exists if and only if $\mathrm{VaR}^R_{\alpha+\beta}(X) = \mathrm{VaR}^R_\alpha(X)$. In particular, if X is continuously distributed, then no optimal allocation exists.*

The non-existence result in Proposition 1 requires a characterization (42) of all optimal allocations for $(\mathrm{VaR}^L_\alpha, \mathrm{ES}_\beta)$, which is technically quite complicated. The proof of Proposition 1 is put in Appendix B. Below we illustrate the non-existence with a simple example.

Example 1. Take a Bernoulli random variable X with $\mathbb{P}(X = 1) = \alpha \in (0, 1)$, and $\beta \in (0, 1 - \alpha)$. Suppose for the purpose of contradiction that X has an optimal allocation $(X_1, X_2) \in \mathbb{A}_2(X)$ such that $\mathrm{VaR}^R_\alpha(X_1) + \mathrm{ES}_\beta(X_2) = \mathrm{RVaR}_{\alpha,\beta}(X) = 0$. Without loss of generality, we can assume $\mathrm{VaR}^R_\alpha(X_1) = \mathrm{ES}_\beta(X_2) = 0$. Note that $\mathrm{VaR}^R_\alpha(X_1) = 0$ implies $\mathbb{P}(X_1 > 0) \leqslant \alpha$, which in turn implies $\mathbb{P}(X_2 < 0) \leqslant \alpha$ as $\{X_2 < 0\} \subseteq \{X_1 > 0\}$. Hence, $\mathrm{VaR}^R_\beta(X_2) \geqslant \mathrm{VaR}^R_{1-\alpha}(X_2) \geqslant 0$, which, together with $\mathrm{ES}_\beta(X_2) = 0$, yields $\mathrm{VaR}^R_\gamma(X_2) = 0$ for all $\gamma \in (0, \beta]$, and thus $\mathbb{P}(X_2 \leqslant 0) = 1$. Hence, $X_1 = X - X_2 \geqslant X$ a.s., which implies $\mathrm{VaR}^R_\alpha(X_1) \geqslant \mathrm{VaR}^R_\alpha(X) = 1$, a contradiction.

Proposition 1 and Example 1 suggest that the existence of an optimal allocation for tail risk measures cannot be taken for granted, especially when VaR^R_α is involved. This makes the existence result in Theorem 1 somewhat surprising, as the inf-convolution of a mixed collection of VaRs always has an optimal allocation. This existence relies strongly on the specific functional form of VaRs, and it is not guaranteed even in the simple case of VaR^R_α and ES_β as shown in Proposition 1 and Example 1.

Remark 7. Similarly to (20), we can get, for $\alpha, \beta, \gamma > 0$ with $\alpha + \beta + \gamma < 1$,

$$\mathrm{VaR}^R_\alpha \Box \mathrm{RVaR}_{\gamma,\beta} = \mathrm{RVaR}_{\alpha+\gamma,\beta} = \mathrm{VaR}^L_\alpha \Box \mathrm{RVaR}_{\gamma,\beta}.$$

Therefore, we obtain formulas for the inf-convolution of all possible combinations of left-VaR, right-VaR, RVaR and ES, and this completes the results in Embrechts et al. (2018) which exclude right-VaR from risk sharing problems. Similarly to the case of $(\mathrm{VaR}^R_\alpha, \mathrm{ES}_\beta)$, an optimal allocation for $(\mathrm{VaR}^R_\alpha, \mathrm{RVaR}_{\gamma,\beta})$ does not necessarily exist.

Remark 8. A risk measure ρ is *normalized* if $\rho(0) = 0$. Theorem 3.3 of Liu and Wang (2021) shows that VaR^R_α is the smallest normalized α-tail monetary risk measure. Combining this result with Theorem 2, we have

$$\mathrm{VaR}^L_\alpha \Box \rho(X) \geqslant \mathrm{VaR}^R_{\alpha+\varepsilon}(X)$$

for all normalized ε-tail monetary risk measures ρ. The above inequality holds as an equality for the choice $\rho = \mathrm{VaR}^R_\varepsilon$, as we have seen in Theorem 1.

5 Inf-convolution of tail risk measures

In this section, we analyze the unconstrained inf-convolution of generic tail risk measures. As shown by Liu et al. (2020), the inf-convolution of law-invariant risk measures is law-invariant under some weak conditions. The following theorem shows that the inf-convolution of tail risk measure is still a tail risk measure under some weak conditions, and the tail parameter is an aggregation of the individual tail parameters. This result shows that the tail risk measures exhibit a similar property to the cases of RVaR and VaR in (2) and (3).

In Theorem 3 below, continuity is defined with respect to the sup-norm, that is, for a sequence $(X_n)_{n\in\mathbb{N}}$ and a random variable X in $\mathcal{X}$, $\rho(X_n) \to \rho(X)$ if ess-sup$|X_n - X| \to 0$. Note that all monetary risk measures on arbitrary domains are continuous with respect to the sup-norm, and this requirement is very weak. Moreover, a 1-tail risk measure below simply means a law-invariant risk measure, which is a natural extension of Definition 1 to $p = 1$.

Theorem 3. *Suppose that ρ_i is a ε_i-tail risk measure for some $\varepsilon_i \in (0,1)$, $i = 1, \dots, n$. If one of $\rho_1, \dots, \rho_n$ is monotone and (sup-norm) continuous from above, then $\square_{i=1}^n \rho_i$ is a monotone ε-tail risk measure, where $\varepsilon = \min\{\sum_{i=1}^n \varepsilon_i, 1\}$.*

Proof. The proof of Theorem 3 requires the following technical lemma. The proof of the lemma is given in Appendix B.

Lemma 1. *For $X \in \mathcal{X}$ and $\varepsilon \in (0,1)$, we have $X'_\varepsilon \overset{\mathrm{d}}{=} X_\varepsilon$, where $X' = \max\{X, m\}$ and $m \in [\mathrm{VaR}^L_\varepsilon(X), \mathrm{VaR}^R_\varepsilon(X)]$.*

We continue the proof of Theorem 3. Take $X \in \mathcal{X}$ and $(X_1, \dots, X_n) \in \mathbb{A}_n(X)$. Let $X'_i = \max\{X_i, x_i\}$, where $x_i = \mathrm{VaR}^L_{\varepsilon_i}(X_i)$ for $i = 1, \dots, n$, and $X' = \max\{X, x\}$, where $x = \mathrm{VaR}^L_\varepsilon(X)$ (here x may be $-\infty$ if $\varepsilon = 1$). We have $(X'_i)_{\varepsilon_i} \overset{\mathrm{d}}{=} (X_i)_{\varepsilon_i}$ by Lemma 1, and thus $\rho_i(X_i) = \rho_i(X'_i)$, $i = 1, \dots, n$. Corollary 1 of Embrechts et al. (2018) implies

$$X'_1 + \cdots + X'_n \geqslant x_1 + \cdots + x_n \geqslant x.$$

Together with the observation that $X_1' + \cdots + X_n' \geqslant X$, we know $X_1' + \cdots + X_n' \geqslant X'$. Thus

$$\begin{aligned}
\mathop{\square}_{i=1}^{n} \rho_i(X) &= \inf\left\{\sum_{i=1}^{n} \rho_i(X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n(X)\right\} \\
&= \inf\left\{\sum_{i=1}^{n} \rho_i(X_i') : (X_1, \dots, X_n) \in \mathbb{A}_n(X)\right\} \\
&\geqslant \inf\left\{\sum_{i=1}^{n} \rho_i(Y_i) : Y_1, \dots, Y_n \in \mathcal{X},\ Y_1 + \cdots + Y_n \geqslant X'\right\} \\
&= \inf\left\{\inf\left\{\sum_{i=1}^{n} \rho_i(Y_i) : (Y_1, \dots, Y_n) \in \mathbb{A}_n(Z)\right\} : Z \geqslant X',\ Z \in \mathcal{X}\right\} \\
&= \inf\left\{\mathop{\square}_{i=1}^{n} \rho_i(Z) : Z \geqslant X',\ Z \in \mathcal{X}\right\}.
\end{aligned}$$

By Lemma 1 of Liu et al. (2020), as long as one of $\rho_1, \dots, \rho_n$ is monotone, $\square_{i=1}^n \rho_i$ is monotone. Hence,

$$\mathop{\square}_{i=1}^{n} \rho_i(X) \geqslant \inf\left\{\mathop{\square}_{i=1}^{n} \rho_i(Z) : Z \geqslant X',\ Z \in \mathcal{X}\right\} = \mathop{\square}_{i=1}^{n} \rho_i(X').$$

Using the monotonicity of $\square_{i=1}^n \rho_i$ again, we have $\square_{i=1}^n \rho_i(X) \leqslant \square_{i=1}^n \rho_i(X')$, which eventually gives

$$\mathop{\square}_{i=1}^{n} \rho_i(X) = \mathop{\square}_{i=1}^{n} \rho_i(X'). \tag{21}$$

Take $Y \in \mathcal{X}$ and $Y_\varepsilon \overset{\mathrm{d}}{=} X_\varepsilon$, we have $Y' \overset{\mathrm{d}}{=} X'$ where $Y' = \max\{Y, \mathrm{VaR}_\varepsilon^L(Y)\}$. Since ρ_i is ε_i-tail risk measure, it is a law-invariant risk measure. Next, using Theorem 2 of Liu et al. (2020), we know $\square_{i=1}^n \rho_i$ is law-invariant. By (21), we have

$$\mathop{\square}_{i=1}^{n} \rho_i(Y) = \mathop{\square}_{i=1}^{n} \rho_i(Y') = \mathop{\square}_{i=1}^{n} \rho_i(X') = \mathop{\square}_{i=1}^{n} \rho_i(X).$$

Therefore, $\square_{i=1}^n \rho_i$ is an ε-tail risk measure. $\square$

The tail parameter in Theorem 3 is generally sharp, as shown by the formula

$$\mathop{\square}_{i=1}^{n} \mathrm{VaR}_{\alpha_i}^R = \mathrm{VaR}_{\sum_{i=1}^n \alpha_i}^R$$

obtained from Theorem 1 (i). Recall that for $\alpha \in (0,1)$, VaR_α^R has a (smallest possible) tail parameter of α. Therefore, the smallest tail parameter of the inf-convolution in this case is indeed the sum of those of the individual risk measures, and the tail parameter obtained from Theorem 3 cannot be improved without specifying the tail risk measures; recall that a smaller tail parameter is a stronger property.

Example 2. Let revisit the formula (2) in the introduction, that is,

$$\mathop{\square}_{i=1}^{n} \mathrm{RVaR}_{\alpha_i,\beta_i} = \mathrm{RVaR}_{\sum_{i=1}^n \alpha_i, \bigvee_{i=1}^n \beta_i},$$

where for simplicity we assume $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n > 0$ and $\sum_{i=1}^n \alpha_i + \sum_{i=1}^n \beta_i < 1$. Note that for each $i = 1, \ldots, n$, $\mathrm{RVaR}_{\alpha_i,\beta_i}$ is a $(\alpha_i + \beta_i)$-tail risk measure. Applying Theorem 3 to the risk measures $\mathrm{RVaR}_{\alpha_1,\beta_1}, \ldots, \mathrm{RVaR}_{\alpha_n,\beta_n}$, we conclude that $\square_{i=1}^n \mathrm{RVaR}_{\alpha_i,\beta_i}$ has a tail parameter $\sum_{i=1}^n \alpha_i + \sum_{i=1}^n \beta_i$. By using the formula (2) directly, we know that $\square_{i=1}^n \mathrm{RVaR}_{\alpha_i,\beta_i}$ actually has a tail parameter $\sum_{i=1}^n \alpha_i + \bigvee_{i=1}^n \beta_i$, suggesting that the tail parameter obtained in Theorem 3 can be improved in this case.

Remark 9. Let ρ be a monetary ε-tail risk measure. Theorem 3 implies that $\mathrm{VaR}_\alpha^R \square \rho$ is an $(\alpha+\varepsilon)$-tail risk measure for $\alpha \in (0, 1-\varepsilon)$. Theorem 2 (iii) further gives that $\mathrm{VaR}_\alpha^L \square \rho$ is also a $(\alpha+\varepsilon)$-tail risk measure, which is not covered by the result of Theorem 3, noting that VaR_α^L is not a α-tail risk measure.

6 Comonotonic inf-convolution of tail risk measures

In this section, we consider inf-convolution of tail risk measures constrained to comonotonic allocations. In some situations, especially in an insurance context, it may be preferred or mandatory to allocate the aggregate risk in a comonotonic way. Two random variables X and Y are *comonotonic*, denoted by $X /\!\!/ Y$, if there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ and

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0 \quad \text{for all } \omega, \omega' \in \Omega_0.$$

Comonotonicity of more than two random variables are defined via pair-wise comonotonicity. For a review on comonotonicity and risk measures, see Dhaene et al. (2006). We further define the set of *comonotonic allocations* as

$$\mathbb{A}_n^+(X) = \{(X_1, \ldots, X_n) \in \mathbb{A}_n(X) : \; X_i /\!\!/ X, \; i = 1, \ldots, n\}. \tag{22}$$

The *constrained inf-convolution* of risk measures $\rho_1, \ldots, \rho_n$ is defined as

$$\mathop{\boxplus}_{i=1}^{n} \rho_i(X) = \inf\left\{\sum_{i=1}^n \rho_i(X_i) : (X_1, \ldots, X_n) \in \mathbb{A}_n^+(X)\right\}. \tag{23}$$

An n-tuple $(X_1, \ldots, X_n) \in \mathbb{A}_n^+(X)$ is called an *optimal constrained allocation* of X for $(\rho_1, \ldots, \rho_n)$ if $\sum_{i=1}^n \rho_i(X_i) = \boxplus_{i=1}^n \rho_i(X)$.

There are two main reasons to consider comonotonic allocations. First, comonotonic allocations are closely connected with the concept of *convex-order consistency*. A risk measure ρ is convex-order consistent if $\rho(X) \leqslant \rho(Y)$ for $X \prec_{\mathrm{cx}} Y$, $X, Y \in \mathcal{X}$, where $X \prec_{\mathrm{cx}} Y$ means that $\mathbb{E}[f(X)] \leqslant \mathbb{E}[f(Y)]$ for all convex functions f, provided that both expectations exist. The convex-order consistency is equivalent to the notion of strong risk aversion in decision theory.[4] Due to the well-known result of *comonotone improvement* (see e.g. Carlier et al. (2012)), if the underlying risk measures respect convex order, Pareto-optimal risk allocations can be constrained in the set of comonotonic allocations. Second, the study on the constrained inf-convolution is especially useful in an insurance context because an aggregate insurance risk is often distributed among insurers and the insured in a comonotonic way. Comonotonicity of allocations is known as the *no-sabotage* or *no-moral-hazard* condition in insurance; see e.g., Huberman et al. (1983) and Carlier and Dana (2003).

It is obvious by definition that $\square_{i=1}^{n} \rho_i(X) \leqslant \boxplus_{i=1}^{n} \rho_i(X)$. Hence, if an optimal allocation of X is comonotonic, then it is also an optimal constrained allocation and $\square_{i=1}^{n} \rho_i(X) = \boxplus_{i=1}^{n} \rho_i(X)$. For convex-order consistent risk measures, including all law-invariant convex risk measures (see Appendix A for definition), optimal constrained allocations are also optimal allocations. However, for risk measures that do not respect the convex order, such as VaR, optimal constrained allocations are generally not optimal in the unconstrained case; for instance, this is the case for RVaR as in Theorem 2 of Embrechts et al. (2018).

The next result shows that in the comonotonic risk sharing problem, the constrained inf-convolution of tail risk measures is again a tail risk measure, similarly to the unconstrained case in Theorem 3, although its tail parameter is not the summation, but the maximum, of those of individual risk measures. An explicit formula of the constrained inf-convolution of tail risk measures is available through the concept of p-generator. For $p \in (0, 1)$ and a p-tail risk measure ρ, its *p-generator* is a law-invariant risk measure ρ^* satisfying $\rho(X) = \rho^*(X_p)$ for all $X \in \mathcal{X}$, which always exists and is unique on the set of random variables in $\mathcal{X}$ bounded from below (Proposition 3.1 of Liu and Wang (2021)). We also say that ρ is generated by ρ^*.

Theorem 4. *Suppose that for $i = 1, \dots, n$, ρ_i is an ε_i-tail risk measure for some $\varepsilon_i \in (0, 1)$, and let $\varepsilon = \bigvee_{i=1}^{n} \varepsilon_i$.*

(i) $\boxplus_{i=1}^{n} \rho_i(X) = \boxplus_{i=1}^{n} \rho_i^*(X_\varepsilon)$ for all $X \in \mathcal{X}$, where ρ_i^* is the ε-generator of ρ_i, $i = 1, \dots, n$.

(ii) $\boxplus_{i=1}^{n} \rho_i$ is an ε-tail risk measure.

[4] Convex order is also known as *mean-preserving spreads* in the decision theoretic literature.

(iii) If $\rho_1, \dots, \rho_n$ are convex-order consistent, then $\square_{i=1}^n \rho_i = \boxplus_{i=1}^n \rho_i$, and hence $\square_{i=1}^n \rho_i$ is an ε-tail risk measure.

Proof. The proof of Theorem 4 relies on the following lemma, which connects an allocation of the tail risk to the tail risks of components in an allocation. The proof of the lemma is given in Appendix B.

Lemma 2. *Suppose $X \in \mathcal{X}$ and $\varepsilon \in (0,1)$.*

(a) For $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$, there exists $(Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X_\varepsilon)$ such that $Y_i \stackrel{\mathrm{d}}{=} (X_i)_\varepsilon$ for $i = 1, \dots, n$.

(b) For $(Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X_\varepsilon)$, there exists $(X_1, \dots, X_n) \in \mathbb{A}_n^+(X)$ such that $(X_i)_\varepsilon \stackrel{\mathrm{d}}{=} Y_i$ for $i = 1, \dots, n$.

We continue the proof of Theorem 4 using Lemma 2. Since each ρ_i is an ε_i-tail risk measure, it is an ε-tail risk measure. By (a) of Lemma 2, we know

$$\begin{aligned}\boxplus_{i=1}^n \rho_i(X) &= \inf\left\{\sum_{i=1}^n \rho_i^*((X_i)_\varepsilon) : (X_1, \dots, X_n) \in \mathbb{A}_n^+(X)\right\} \\ &\geqslant \inf\left\{\sum_{i=1}^n \rho_i^*(Y_i) : (Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X_\varepsilon)\right\} = \boxplus_{i=1}^n \rho_i^*(X_\varepsilon).\end{aligned}$$

Similarly, by (b) of Lemma 2,

$$\begin{aligned}\boxplus_{i=1}^n \rho_i^*(X_\varepsilon) &= \inf\left\{\sum_{i=1}^n \rho_i^*(Y_i) : (Y_1, \dots, Y_n) \in \mathbb{A}_n^+(X_\varepsilon)\right\} \\ &\geqslant \inf\left\{\sum_{i=1}^n \rho_i^*((X_i)_\varepsilon) : (X_1, \dots, X_n) \in \mathbb{A}_n^+(X)\right\} = \boxplus_{i=1}^n \rho_i(X).\end{aligned}$$

Therefore, $\boxplus_{i=1}^n \rho_i(X) = \boxplus_{i=1}^n \rho_i^*(X_\varepsilon)$, thus statement (i) holds. Statement (ii) follows directly from statement (i). For the statement (iii), it suffices to note that according to the comonotonic improvement, for convex-order consistent risk measures, we have $\square_{i=1}^n \rho_i = \boxplus_{i=1}^n \rho_i$. $\square$

Theorem 4 implies that if risk sharing is constrained to be comonotonic, then the corresponding inf-convolution has a tail parameter equal to the maximum of those of the individual risk measures. Comparing this result with Theorem 3, where the (unconstrained) inf-convolution has a tail parameter equal to the sum of those of the individual risk measures, the constraint on comonotonicity indeed reduces the tail parameter.

Example 3. Let $\beta_1, \dots, \beta_n \in (0,1)$ and $\beta = \bigvee_{i=1}^n \beta_i$. According to Theorem 4, the inf-convolution of $\mathrm{ES}_{\beta_1}, \dots, \mathrm{ES}_{\beta_n}$ is a β-tail risk measure, since each of $\mathrm{ES}_{\beta_1}, \dots, \mathrm{ES}_{\beta_n}$ is convex-order consistent. Indeed, Theorem 2 of Embrechts et al. (2018) gives $\square_{i=1}^n \mathrm{ES}_{\beta_i} = \mathrm{ES}_\beta$, which is clearly a β-tail risk measure, and this example suggests that the tail parameter obtained in Theorem 4 is sharp.

Remark 10. Wang and Zitikis (2020) obtained inf-convolution formulas for VaR^L (with a slightly different setting) under a spectrum of *weak comonotonicity* constraints, which range from imposing comonotonicity to imposing no constraints. Indeed, the results obtained by Wang and Zitikis (2020) allow for the tail parameter of the inf-convolution of $\mathrm{VaR}^L_{\varepsilon_1}, \dots, \mathrm{VaR}^L_{\varepsilon_n}$ to be anything between $\bigvee_{i=1}^n \varepsilon_i$ (comonotonicity) and $\sum_{i=1}^n \varepsilon_i$ (no constraints).

Remark 11. For distortion risk measures $\rho_1, \dots, \rho_n$, their comonotonic inf-convolution can be written explicitly; see Proposition 5 of Embrechts et al. (2018). From that result, one can easily check that $\square_{i=1}^n \rho_i$ is indeed a $(\bigvee_{i=1}^n \varepsilon_i)$-tail risk measure if the distortion risk measure ρ_i is an ε_i-tail risk measure for $i = 1, \dots, n$. Therefore, Theorem 4 generalizes the above observation beyond distortion risk measures.

7 Risk sharing for elliptical models

In this section, we consider risk sharing within the class of elliptical models. Elliptical models are the most popular parametric risk models, and they appear prominently in both asset pricing models (e.g., the classic Bachelier and Black-Scholes models) and time series analysis (e.g., GARCH models); see McNeil et al. (2015) for a general background. Common special cases of the elliptical models are the multivariate normal and the multivariate t-distributions; we refer to Chapter 6 of McNeil et al. (2015) for details. Let $\mathcal{E}_n(\psi)$ be the class of all n-dimensional elliptically distributed random vectors with a fixed characteristic generator ψ, i.e., for $\mathbf{X} \in \mathcal{E}_n(\psi)$, there exist $\boldsymbol{\mu} \in \mathbb{R}^n$ and a positive semi-definite matrix $\Sigma \in \mathbb{R}^{n \times n}$ such that the characteristic function of $\mathbf{X}$ is

$$\mathbf{t} \mapsto e^{i\mathbf{t}^\top \boldsymbol{\mu}} \psi(\mathbf{t}^\top \Sigma \mathbf{t})$$

where $\mathbf{t}^\top$ is the transpose of $\mathbf{t} \in \mathbb{R}^n$. Assume throughout this section that $\mathcal{X}$ contains $\mathcal{E}_1(\psi)$ and $X \in \mathcal{E}_1(\psi)$. We consider the risk sharing problem confined to the elliptical class $\mathcal{E}_n(\psi)$, namely,

$$\mathop{\oplus}_{i=1}^{n} \rho_i(X) = \min \left\{ \sum_{i=1}^n \rho_i(X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n(X) \cap \mathcal{E}_n(\psi) \right\}. \tag{24}$$

In other words, the allocated risk positions in (24) have to be jointly elliptical. For instance, if X is Gaussian (or t-distributed), then we require an allocation $(X_1, \dots, X_n)$ of X to be jointly Gaussian (or t-distributed). It turns out that for tail risk measures with tail parameter no larger than 1/2, such a risk sharing problem admits a simple solution, due to the nice structure of elliptically distributed random vectors.

Theorem 5. *Suppose that for $i = 1, \dots, n$, ρ_i is a monetary ε_i-tail risk measure, $\bigvee_{i=1}^n \varepsilon_i \leqslant 1/2$ and $\oplus_{i=1}^n \rho_i(X) > -\infty$. An optimal allocations to (24) is comonotonic and given by $X_i^* = \mu_i + c_i^* X$, $i = 1, \dots, n$, where $\mu_1 + \cdots + \mu_n = 0$, and $(c_1^*, \dots, c_n^*)$ is an optimizer to the following problem*

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^n \rho_i(c_i X) \\ \text{subject to} \quad & \sum_{i=1}^n c_i = 1,\ c_i \geqslant 0,\ i = 1, \dots, n. \end{aligned} \tag{25}$$

Proof. Let $X \in \mathcal{E}_1(\psi)$ with characteristic function

$$t \mapsto e^{it\mu}\psi(\sigma t^2)$$

for some $\mu \in \mathbb{R}$ and $\sigma^2 \geqslant 0$. Note that for any $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{E}_n(\psi)$, there exist $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$, $k \geqslant 1$, and $A = (\mathbf{a}_1, \dots, \mathbf{a}_n) \in \mathbb{R}^{k \times n}$ such that

$$\mathbf{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + A^\top \boldsymbol{S},$$

where $\boldsymbol{S} = (S_1, \dots, S_k)$ follows the standard k-dimensional spherical distribution with characteristic generator ψ.

Note that $X_i = \mu_i + \mathbf{a}_i^\top \boldsymbol{S} \overset{\mathrm{d}}{=} \mu_i + \|\mathbf{a}_i\| S_1$ where $\|\cdot\|$ is the Euclidean norm (see e.g., Theorem 6.18 of McNeil et al. (2015)). By translation-invariance of ρ_i, we have

$$\rho_i(X_i) = \mu_i + \rho_i\left(\|\mathbf{a}_i\| S_1\right), \quad i = 1, \dots, n,$$

Note that by requiring $X_1 + \cdots + X_n = X$, we have $\sum_{i=1}^n \mu_i = \mu$. Hence, the problem (24) is equivalent to

$$\text{minimize} \quad \sum_{i=1}^n \left(\mu_i + \rho_i(\mathbf{a}_i^\top \boldsymbol{S})\right) = \mu + \sum_{i=1}^n \rho_i\left(\|\mathbf{a}_i\| S_1\right) \tag{26}$$

$$\text{subject to} \quad \mu + \sum_{i=1}^n \mathbf{a}_i^\top \boldsymbol{S} \overset{\mathrm{d}}{=} X. \tag{27}$$

We have the following observations:

(i) The constraint (27) is equivalent to $\|\sum_{i=1}^n \mathbf{a}_i\|^2 = \sigma^2$.

(ii) The objective function $\rho_i(\mathbf{a}_i^\top \boldsymbol{S})$ is increasing in $\|\mathbf{a}_i\|$. To see it, note that S_1 has symmetric distribution and the tail risk of S_1 beyond its $(1-\varepsilon_i)$-quantile, denoted by $(S_1)_{\varepsilon_i}$, is non-negative for $\varepsilon_i \leqslant 1/2$. It then follows that $x(S_1)_{\varepsilon_i}$ is increasing in $x \geqslant 0$, which implies that $\rho_i(xS_1)$ is increasing in $x \geqslant 0$ since ρ_i is monotone. Hence, $\rho_i(\mathbf{a}_i^\top \boldsymbol{S})$ is increasing in $\|\mathbf{a}_i\|$ for $i = 1, \ldots, n$.

(iii) For any $\mathbf{b} \in \mathbb{R}^k$ and $c \in \mathbb{R}_+$, the following problem

$$\text{minimize} \quad \|\mathbf{a}\| \quad \text{subject to } \|\mathbf{a}+\mathbf{b}\|^2 = c \tag{28}$$

admits an optimal solution $\mathbf{a}^*$ satisfying $\mathbf{a} = c_0\mathbf{b}$. This implies that each $\mathbf{a}_i$ minimizing (26) is a multiple of each other.

Summarizing the above three observations, we conclude that an optimal allocation to the problem (24), denoted by $(X_1^*, \ldots, X_n^*)$ must be comonotonic and the problem (24) is equivalent to

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^n \rho_i(X_i) \\ &\text{subject to} \quad X_i = \mu_i + c_i X, \quad \sum_{i=1}^n c_i = 1, \ c_i \geqslant 0, \ i = 1, \ldots, n. \end{aligned} \tag{29}$$

Hence, we complete the proof. $\square$

If the risk measures $\rho_1, \ldots, \rho_n$ are positively homogeneous, then $\oplus_{i=1}^n \rho_i(X)$ boils down to a simple form, which follows directly from (25) and $\rho_i(c_i X) = c_i \rho_i(X)$.

Corollary 1. *Suppose that for $i = 1, \ldots, n$, ρ_i is a monetary and positively homogeneous ε_i-tail risk measure, $\bigvee_{i=1}^n \varepsilon_i \leqslant 1/2$, and $\oplus_{i=1}^n \rho_i(X) > -\infty$. We have*

$$\mathop{\oplus}_{i=1}^n \rho_i(X) = \min_{i=1,\ldots,n} \rho_i(X),$$

and an optimal allocation $(X_1^*, \ldots, X_n^*)$ for (24) is given by $X_{i^*}^* = X$ and $X_j^* = 0$ for all $j \neq i^*$ for any $i^* \in \arg\min_{i=1,\ldots,n} \rho_i(X)$.

We make the following further observations about Theorem 5.

(i) The comonotonicity of the optimal allocation in Theorem 5 is remarkable, because comonotonicity is often a consequence of convex-order consistency (see e.g., Theorem 4). Note that convex-order consistency is not assumed in Theorem 5. The main reason for this is that the tail risk with level $\varepsilon \leqslant 1/2$ of the elliptical distribution becomes larger when scaled up, i.e.,

$\lambda(S_1)_\varepsilon \geqslant (S_1)_\varepsilon$ for $\lambda \geqslant 1$ and a spherically distributed random variable S_1. This argument also illustrates that the assumptions that ρ_i is monetary and $\varepsilon_i \leqslant 1/2$ are important for the result in Theorem 5.

(ii) From (25), the optimal solution does not depend on the generator ψ.

(iii) If $\rho_i(cX)$ is strictly increasing in $c \geqslant 0$, then all optimal allocations to (24) are in the form of Theorem 5.

(iv) Unlike $\square_{i=1}^n \rho_i$ and $\boxplus_{i=1}^n \rho_i(X)$, which are well defined on $\mathcal{X}$, the mapping $\oplus_{i=1}^n \rho_i$ is only properly defined on $\mathcal{E}_1(\psi)$. For this reason, we fix X in this section, and do not discuss the tail parameter of $\oplus_{i=1}^n \rho_i$.

In the setting of Theorem 5, optimal allocations can be confined to $\mathbb{A}_n^+(X)$. Hence, for risk measures satisfying the conditions in Theorem 5, we have the general inequalities

$$\mathop{\square}_{i=1}^{n} \rho_i(X) \leqslant \mathop{\boxplus}_{i=1}^{n} \rho_i(X) \leqslant \mathop{\oplus}_{i=1}^{n} \rho_i(X). \tag{30}$$

In the following example, we compare the three inf-convolutions in (30) for risk measures being VaR, ES and RVaR. To use the result in Theorem 5, we assume that all risk measures have tail parameters less than 1/2.

Example 4. Since ES, VaR and RVaR are all positively homogenous, we can use Corollary 1 for the value of $\oplus_{i=1}^n \rho_i(X)$. The values of $\square_{i=1}^n \rho_i(X)$ and $\boxplus_{i=1}^n \rho_i(X)$ are obtained by Theorem 2 and Proposition 5 of Embrechts et al. (2018), and we omit the detailed calculations.

(i) For ES, it holds

$$\mathop{\square}_{i=1}^{n} \mathrm{ES}_{\alpha_i}(X) = \mathop{\boxplus}_{i=1}^{n} \mathrm{ES}_{\alpha_i}(X) = \mathop{\oplus}_{i=1}^{n} \mathrm{ES}_{\alpha_i}(X) = \mathrm{ES}_{\bigvee_{i=1}^n \alpha_i}(X),$$

thus we have equalities in (30).

(ii) For VaR, it holds

$$\mathop{\square}_{i=1}^{n} \mathrm{VaR}^L_{\alpha_i}(X) = \mathrm{VaR}^L_{\sum_{i=1}^n \alpha_i}(X) \leqslant \mathop{\boxplus}_{i=1}^{n} \mathrm{VaR}^L_{\alpha_i}(X) = \mathop{\oplus}_{i=1}^{n} \mathrm{VaR}^L_{\alpha_i}(X) = \mathrm{VaR}^L_{\bigvee_{i=1}^n \alpha_i}(X),$$

and generally the inequality above is not an equality.

(iii) For RVaR, using Theorem 2 and Proposition 5 of Embrechts et al. (2018), we see that the inequalities in (30), i.e.,

$$\mathop{\square}_{i=1}^{n} \mathrm{RVaR}_{\alpha_i,\beta_i}(X) \leqslant \mathop{\boxplus}_{i=1}^{n} \mathrm{RVaR}_{\alpha_i,\beta_i}(X) \leqslant \mathop{\oplus}_{i=1}^{n} \mathrm{RVaR}_{\alpha_i,\beta_i}(X),$$

are generally not equalities. Some numerical results are presented in Table 1.

	$(\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3)$	$\square_{i=1}^3 \mathrm{RVaR}_{\alpha_i,\beta_i}$	$\boxplus_{i=1}^3 \mathrm{RVaR}_{\alpha_i,\beta_i}$	$\oplus_{i=1}^3 \mathrm{RVaR}_{\alpha_i,\beta_i}$
$X \sim \mathrm{N}(0,1)$	$(.02,.08,.1,.2,.12,.08)$	0.5319	1.0577	1.0863
	$(.01,.03,.1,.15,.13,.02)$	0.7982	1.1928	1.2271
$X \sim \mathrm{t}_2(0,1)$	$(.02,.08,.1,.2,.12,.08)$	0.6357	1.4413	1.4882
	$(.01,.03,.1,.15,.13,.02)$	1.0067	1.6974	1.7650

Table 1: Comparison of the three inf-convolutions. The total risk X has either a standard normal distribution or a standard t-distribution with 2 degrees of freedom.

Example 5. For $\alpha,\beta \in (0,1/2]$, $\lambda \in (0,1)$, $\gamma_i > 0$, $i=1,\dots,n$, let $\mathcal{R}_i$ be a α-tail risk measure with generator being the entropic risk measure $\mathrm{ER}_{\gamma_i}(X) = \gamma_i \log \mathbb{E}[e^{X/\gamma_i}]$, and ρ_β be a positively homogeneous β-tail risk measure. We define

$$\rho_i(X) := \lambda \mathcal{R}_i(X) + (1-\lambda)\rho_\beta(X) = \lambda \mathrm{ER}_{\gamma_i}(X_\alpha) + (1-\lambda)\rho_\beta(X).$$

For instance, if $\rho_\beta = \mathrm{VaR}_\beta^L$ and ρ_i is used by a financial institution, then ρ_i may be interpreted as a mixture of a common regulatory risk measure VaR_β^L and an internal utility-based risk measure $\mathcal{R}_i$. Note that $\mathcal{R}_i$ is not convex in general, and the unconstrained or comonotonic inf-convolution is difficult to solve. The inf-convolution confined to the elliptical class defined by (24) is

$$\begin{aligned}\mathop{\oplus}_{i=1}^{n} \rho_i(X) &= \min\left\{\sum_{i=1}^n \lambda \mathrm{ER}_{\gamma_i}(c_i X_\alpha) + (1-\lambda)\rho_\beta(X) : \sum_{i=1}^n c_i = 1, c_i \geqslant 0\right\} \\ &= \lambda \mathrm{ER}_{\sum_{i=1}^n \gamma_i}(X_\alpha) + (1-\lambda)\rho_\beta(X),\end{aligned}$$

and an optimal allocation is $X_i^* = \gamma_i X/(\gamma_1 + \cdots + \gamma_n)$, $i=1,\dots,n$. The second equality above is due to Theorem 3.9 of Barrieu and El Karoui (2005).

8 Risk sharing under model uncertainty

In this section, we bring model uncertainty into risk sharing. Instead of a fixed and known probability measure $\mathbb{P}$, we assume that each agent i has some uncertainty about the distribution of the risk she is allocated to. Our framework can be seen as a natural extension of the framework of Embrechts et al. (2020) where agents have heterogeneous beliefs modelled by different probability measures. The set describing model uncertainty for each agent will be induced either by the likelihood ratio between probability measures (i.e., the Radon-Nikodym derivative) or by the Wasserstein metric between random variables. We will treat these two cases separately.

8.1 Model uncertainty induced by likelihood ratios

A popular way to incorporate model uncertainty into decisions is through a worst-case approach axiomatized by Gilboa and Schmeidler (1989); see e.g., Zhu and Fukushima (2009) for worst-case risk measures in the context of optimization. Let $\mathcal{P}$ be the set of all probability measures that are absolutely continuous with respect to $\mathbb{P}$, a pre-specified probability measure representing a common benchmark for all agents. Let ρ be a law-invariant risk measure, such as VaR^L_α, VaR^R_α or ES_α, where law-invariance is defined with respect to distributions under $\mathbb{P}$. Define $\rho^Q(X) = \rho(X_Q)$, where the distribution of X_Q (under $\mathbb{P}$) is the distribution of X under Q. In other words ρ^Q is the risk measure ρ evaluated under the probability measure Q instead of $\mathbb{P}$. We consider the worst-case risk measure

$$\overline{\rho}^{\mathcal{Q}} := \sup_{Q\in\mathcal{Q}} \rho^Q,$$

where $\mathcal{Q}$ is a subset of $\mathcal{P}$. Optimization problems involving different uncertainty sets $\mathcal{Q}$ and different tail risk measures such as VaR and ES are extensively studied in the literature; we refer to the recent work of Blanchet et al. (2020), Chen et al. (2018), Xie (2021) and Ho-Nguyen et al. (2021).

For tractability, we will consider a particular choice of $\mathcal{Q}$. For $\lambda \in (0, 1]$, define the set $\mathcal{P}_\lambda = \{Q \in \mathcal{P} : \mathrm{d}Q/\mathrm{d}\mathbb{P} \leqslant 1/\lambda\}$. Note that $\mathbb{P} \in \mathcal{P}_\lambda$. Here, λ is a parameter representing the degree of uncertainty faced by an agent.[5] In particular, $\lambda = 1$ corresponds to $\mathcal{P}_1 = \{\mathbb{P}\}$, that is, there is no uncertainty.

Proposition 2. *For $\varepsilon, \lambda \in (0, 1)$, let ρ be a monotone ε-tail risk measure generated by ρ^*.*

(i) $\overline{\rho}^{\mathcal{P}_\lambda}$ is a monotone $(\varepsilon\lambda)$-tail risk measure generated by ρ^*.

(ii) $\overline{\rho}^{\mathcal{P}_\lambda}$ is a monotone λ-tail risk measure generated by ρ.

(iii) If ρ is continuous from above with respect to a.s. convergence, then so is $\overline{\rho}^{\mathcal{P}_\lambda}$.

(iv) If ρ or ρ^* is convex-order consistent, then so is $\overline{\rho}^{\mathcal{P}_\lambda}$.

Next, let us turn to the special case of VaR and ES. For $\mathcal{Q} \subset \mathcal{P}$, write

$$\overline{\mathrm{VaR}}^{R,\mathcal{Q}}_\alpha = \sup_{Q\in\mathcal{Q}} (\mathrm{VaR}^R_\alpha)^Q, \quad \overline{\mathrm{VaR}}^{L,\mathcal{Q}}_\alpha = \sup_{Q\in\mathcal{Q}} (\mathrm{VaR}^L_\alpha)^Q, \quad \text{and} \quad \overline{\mathrm{ES}}^{\mathcal{Q}}_\alpha = \sup_{Q\in\mathcal{Q}} \mathrm{ES}^Q_\alpha.$$

Proposition 2 (ii) immediately gives the following formulas.

[5] Generally, we may consider the set induced by the ϕ-divergence, $\mathcal{P}_\phi = \{Q \in \mathcal{P} : \mathbb{E}[\phi(\mathrm{d}Q/\mathrm{d}\mathbb{P})] \leqslant 1\}$ where $\phi : [0, \infty] \to [0, \infty]$ is a convex function, which includes $\mathcal{P}_\lambda$ via $\phi(x) = \infty \times \mathbb{1}_{\{x>1/\lambda\}}$ and the set induced by the Kullback-Leibler divergence via $\phi(x) = \beta x \log x$ for some $\beta > 0$. However, explicit results on $\overline{\rho}^{\mathcal{P}_\phi}$ are not available for general choices of ϕ.

Corollary 2. *For $\lambda \in (0,1]$ and $\alpha \in (0,1)$,*

$$\overline{\mathrm{VaR}}^{R,\mathcal{P}_\lambda}_\alpha = \mathrm{VaR}^R_{\alpha\lambda}, \quad \overline{\mathrm{VaR}}^{L,\mathcal{P}_\lambda}_\alpha = \mathrm{VaR}^L_{\alpha\lambda}, \quad \textit{and} \quad \overline{\mathrm{ES}}^{\mathcal{P}_\lambda}_\alpha = \mathrm{ES}_{\alpha\lambda}. \tag{31}$$

Using Proposition 2 and Corollary 2, we obtain parallel results to Theorems 1-4 in the setting of the model uncertainty. For instance, Theorems 1 and 2 can be restated via

$$\mathop{\square}_{i=1}^{n} \overline{\mathrm{VaR}}^{\Lambda_i,\mathcal{P}_{\lambda_i}}_{\alpha_i} = \mathop{\square}_{i=1}^{n} \mathrm{VaR}^{\Lambda_i}_{\lambda_i\alpha_i} \quad \text{and} \quad \overline{\mathrm{VaR}}^{\Lambda,\mathcal{P}_{\lambda_2}}_\alpha \,\square\, \overline{\rho}^{\mathcal{P}_{\lambda_1}} = \mathrm{VaR}^{\Lambda}_{\lambda_2\alpha} \,\square\, \rho_0,$$

for an ε-tail risk measure ρ, where $\rho_0 = \overline{\rho}^{\mathcal{P}_{\lambda_1}}$ is a monetary $(\varepsilon\lambda_1)$-tail risk measure. We omit similar corollaries of Theorems 3 and 4 in this setting, which use (iii) and (iv) of Proposition 2. To summarize, using model uncertainty sets $\mathcal{P}_\lambda$ for each agent endowed with a tail risk measure, the risk sharing problems can be converted to the classic one without model uncertainty by adjusting the tail parameters.

8.2 Model uncertainty induced by Wasserstein metrics

The Wasserstein metric is a popular notion used in mass transportation and distributionally robust optimization; see e.g., Esfahani and Kuhn (2018). In the one-dimensional setting, the Wasserstein metric has an explicit formula. For two random variables X and Y with respective distributions F and G, the Wasserstein metric of order $k \geqslant 1$ is given by

$$W_k(X,Y) = W_k(F,G) = \left(\int_0^1 \left|F^{-1}(x) - G^{-1}(x)\right|^k \,\mathrm{d}x\right)^{1/k}.$$

For a risk measure ρ and a constant $\delta > 0$, we define its robust version $[\rho]^k_\delta$ via the Wasserstein metric as

$$[\rho]^k_\delta(X) := \sup\{\rho(Y) : W_k(Y,X) \leqslant \delta\}, \quad X \in \mathcal{X}. \tag{32}$$

For simplicity, we assume $\mathcal{X} = L^\infty$, so that W_k is well defined for each $k \geqslant 1$. It is straightforward to verify that $[\rho]^k_\delta$ is a monetary risk measure if ρ is a monetary risk measure. The following result illustrates that the robust version of a tail risk measure is again a tail risk measure with the same tail parameter, and its generator is a robust version of the original generator.

Proposition 3. *Let $\varepsilon \in (0,1)$ and $\delta > 0$. If ρ is a monotone ε-tail risk measure with generator ρ^*, then*

$$[\rho]^k_\delta(X) = [\rho^*]^k_{\varepsilon^{-1/k}\delta}(X_\varepsilon), \quad X \in \mathcal{X},$$

where X_ε is the ε-tail risk of X. Moreover, $[\rho]^k_\delta$ is an ε-tail risk measure.

Section	model uncertainty	tail parameter	generator	Proposition
-	none	ε	ρ^*	-
8.1	likelihood ratios	$\lambda\varepsilon$	ρ^*	2
8.2	Wasserstein metrics	ε	$[\rho^*]^k_{\varepsilon^{-1/k}\delta}$	3

Table 2: Two formulations of model uncertainty for a monotone ε-tail risk measure ρ with generator ρ^*.

Proposition 3 illustrates that, unlike the model uncertainty described in Section 8.1 via likelihood ratios, the model uncertainty induced by Wasserstein metrics does not change the tail parameter of a tail risk measure. Instead, it changes the generator of the tail risk measure. We summarize these findings in Table 2.

For the rest of the section, model uncertainty and robustness always refer to those induced by Wasserstein metrics.

For a coherent distortion risk measure (see Appendix A for details), we obtain an explicit formula for its robust version. Recall that a distortion function is a function $h : [0,1] \to [0,1]$ which is increasing and satisfies $h(0) = 0$ and $h(1) = 1$, and in the case where h is continuous, the distortion risk measure ρ_h can be written as

$$\rho_h(X) = \int_0^1 \mathrm{VaR}^\Lambda_u(X)\,\mathrm{d}h(u), \quad X \in \mathcal{X}, \tag{33}$$

where $\Lambda \in \{L, R\}$. We will assume that h is concave, implying that ρ_h is a coherent risk measure. In this case,

$$\rho_h(X) = \int_0^1 \mathrm{VaR}^\Lambda_u(X) h'(u)\,\mathrm{d}u, \quad X \in \mathcal{X},$$

where h' is the left-derivative of h on $(0,1)$.

Proposition 4. *Suppose that $X \in \mathcal{X}$, $k \geqslant 1$ and $\delta > 0$.*

(i) For a continuous and concave distortion function h,

$$[\rho_h]^k_\delta(X) = \rho_h(X) + \delta \|h'\|_q,$$

where $q = (1 - 1/k)^{-1}$ with the convention $0^{-1} = \infty$, and $||\cdot||_q$ is the ℓ_q-norm.

(ii) For $\alpha \in (0,1)$,

$$[\mathrm{ES}_\alpha]^k_\delta(X) = \mathrm{ES}_\alpha(X) + \frac{\delta}{\alpha^{1/k}}.$$

(iii) For $\alpha \in (0,1)$, we have $[\mathrm{VaR}^L_\alpha]^k_\delta(X) = [\mathrm{VaR}^R_\alpha]^k_\delta(X) = x$, where x is the unique number satisfying the equation

$$\int_0^\alpha \left(x - \mathrm{VaR}^R_u(X)\right)_+^k \,\mathrm{d}u = \delta^k. \tag{34}$$

Moreover,

$$[\mathrm{VaR}^R_\alpha]^k_\delta(X) \geqslant \mathrm{VaR}^R_\alpha(X) + \frac{\delta}{\alpha^{1/k}}. \tag{35}$$

By Proposition 4 (i), we can obtain explicit formulas for the inf-convolution of robust coherent distortion risk measures. For n concave distortion functions $h_1, \dots, h_n$, we have

$$\begin{aligned}\mathop{\square}_{i=1}^n [\rho_{h_i}]^{k_i}_{\delta_i}(X) &:= \inf\left\{\sum_{i=1}^n [\rho_{h_i}]^{k_i}_{\delta_i}(X_i) : (X_1, \dots, X_n) \in \mathbb{A}_n(X)\right\} \\ &= \mathop{\square}_{i=1}^n \rho_{h_i}(X) + \sum_{i=1}^n \delta_i \|h_i'\|_{(1-1/k_i)^{-1}},\end{aligned}$$

and the optimal allocations for robust coherent distortion risk measures are the same as those without model uncertainty specified by Proposition 5 of Embrechts et al. (2018). In particular, we have

$$\mathop{\square}_{i=1}^n [\mathrm{ES}_{\alpha_i}]^{k_i}_{\delta_i}(X) = \mathrm{ES}_{\bigvee_{i=1}^n \alpha_i}(X) + \sum_{i=1}^n \frac{\delta_i}{\alpha_i^{1/k_i}}.$$

In contrast, the optimal allocations become different for VaR when considering model uncertainty. In general, it is challenging to obtain inf-convolution for robust VaR under model uncertainty induced by Wasserstein metrics, since VaR lacks desirable convexity. We can obtain the following bounds for inf-convolution of robust VaR using results in Theorem 1 and Proposition 4:

$$\begin{aligned}\mathrm{VaR}^R_{\sum_{i=1}^n \alpha_i}(X) + \sum_{i=1}^n \frac{\delta_i}{\alpha_i^{1/k_i}} &= \mathop{\square}_{i=1}^n \left(\mathrm{VaR}^R_{\alpha_i} + \frac{\delta_i}{\alpha_i^{1/k_i}}\right)(X) \\ &\leqslant \mathop{\square}_{i=1}^n [\mathrm{VaR}^R_{\alpha_i}]^{k_i}_{\delta_i}(X) \\ &\leqslant \mathop{\square}_{i=1}^n [\mathrm{ES}_{\alpha_i}]^{k_i}_{\delta_i}(X) = \mathrm{ES}_{\bigvee_{i=1}^n \alpha_i}(X) + \sum_{i=1}^n \frac{\delta_i}{\alpha_i^{1/k_i}}.\end{aligned}$$

The above inequalities are generally not sharp, unless X has some special distribution; for instance, both inequalities are equalities if $\mathrm{VaR}^R_u(X)$ is constant for $u \in (0, \sum_{i=1}^n \alpha_i)$.

Next, we look at a very specific setting, in which each agent uses $[\mathrm{VaR}^\Lambda_{\alpha_i}]^1_{\delta_i}$, and the allowed allocations are comonotonic as in the setting of Section 6. We further assume that either the parameters $\delta_1, \dots, \delta_n$ are identical or the parameters $\alpha_1, \dots, \alpha_n$ are identical. In this special case, we can obtain an explicit formula for the inf-convolution of VaR and the optimal allocation.

Theorem 6. *Suppose that $X \in \mathcal{X}$ and $\Lambda \in \{L, R\}$.*

(i) If $0 < \delta_1 \leqslant \ldots \leqslant \delta_n$ and $\alpha \in (0,1)$, then

$$\mathop{\boxplus}_{i=1}^{n} [\mathrm{VaR}_{\alpha}^{\Lambda}]_{\delta_i}^{1}(X) = [\mathrm{VaR}_{\alpha}^{\Lambda}]_{\delta_1}^{1}(X) + \sum_{i=2}^{n} \frac{\delta_i}{\alpha}. \tag{36}$$

(ii) If $1 > \alpha_1 \geqslant \ldots \geqslant \alpha_n > 0$ and $\delta > 0$, then

$$\mathop{\boxplus}_{i=1}^{n} [\mathrm{VaR}_{\alpha_i}^{\Lambda}]_{\delta}^{1}(X) = [\mathrm{VaR}_{\alpha_1}^{\Lambda}]_{\delta}^{1}(X) + \sum_{i=2}^{n} \frac{\delta}{\alpha_i}. \tag{37}$$

In either setting, an optimal allocation is given by $(X, 0, \ldots, 0)$.

Proof. (i) By Proposition 4, in the setting of robust VaR, $\Lambda \in \{L, R\}$ is irrelevant, and we will assume $\Lambda = L$. For simplicity, denote by

$$V_\delta(Y) = [\mathrm{VaR}_{\alpha}^{L}]_{\delta}^{1}(Y), \quad \delta > 0,\ Y \in \mathcal{X}.$$

It is straightforward to see that the allocation $(X_1^*, \ldots, X_n^*)$ satisfies

$$\sum_{i=1}^{n} V_{\delta_i}(X_i^*) = V_{\delta^*}(X) + \frac{\sum_{i=1}^{n} \delta_i - \delta^*}{\alpha}.$$

Therefore, it remains to show (36). Indeed, it suffices to show (36) for $n = 2$ and the general case follows by induction. We assume $\delta_1 \leqslant \delta_2$ and aim to show

$$V_{\delta_1} \boxplus V_{\delta_2}(X) = V_{\delta_1}(X) + \frac{\delta_2}{\alpha}. \tag{38}$$

From now on, assume $X_1^* = X + \delta_2/\alpha$ and $X_2^* = -\delta_2/\alpha$; there is a minor clash of notation with the last statement of the theorem in which $(X_1^*, X_2^*) = (X, 0)$, but clearly the constant shift does not affect optimality. We have $V_{\delta_1}(X_1^*) = V_{\delta_1}(X) + \delta_2/\alpha$ and $V_{\delta_2}(X_1^*) = 0$, which imply

$$V_{\delta_1} \boxplus V_{\delta_2}(X) \leqslant V_{\delta_1}(X) + \frac{\delta_2}{\alpha}. \tag{39}$$

It remains to show the opposite direction of (39). Since all risk measures involved are translation-invariant, so is their inf-convolution; see e.g., Section 4 of Mao and Wang (2020). Therefore, from now on, we will assume that $V_{\delta_1} \boxplus V_{\delta_2}(X) \leqslant 0$, and it suffices to show $V_{\delta_1}(X) + \delta_2/\alpha \leqslant 0$.

Take $(X_1, X_2) \in \mathbb{A}_n^+(X)$ such that $V_{\delta_1}(X_1) \leqslant 0$, $V_{\delta_2}(X_2) \leqslant 0$ and $\mathbb{P}(X_2 > 0) \leqslant \mathbb{P}(X_1 > 0)$. Let $t = \mathbb{P}(X_2 > 0)$ and $s = \mathbb{P}(X_1 > 0)$. We have $0 \leqslant t \leqslant s \leqslant \alpha$ as $\mathrm{VaR}_{\alpha}^{R}(X_1) < V_{\delta_1}(X_1) \leqslant 0$. By Proposition 4, for any $Y \in \mathcal{X}$ and $\delta > 0$,

$$V_\delta(Y) \leqslant 0 \iff \int_0^\alpha (-\mathrm{VaR}_u^L(Y))_+ \, \mathrm{d}u \geqslant \delta. \tag{40}$$

By comonotonic additivity of VaR^L and $X_2^* \equiv -\delta_2/\alpha$, we have we know

$$\mathrm{VaR}_u^L(X_1) - \mathrm{VaR}_u^L(X_1^*) = \mathrm{VaR}_u^L(X_2^*) - \mathrm{VaR}_u^L(X_2) = -\frac{\delta_2}{\alpha} - \mathrm{VaR}_u^L(X_2),$$

which is increasing in $u \in [0, \alpha]$. Note that

$$\int_t^\alpha (-\mathrm{VaR}_u^L(X_2^*))\,\mathrm{d}u = \frac{\alpha - t}{\alpha}\delta_2 \leqslant \delta_2 \leqslant \int_0^\alpha (-\mathrm{VaR}_u^L(X_2))_+\,\mathrm{d}u = \int_t^\alpha (-\mathrm{VaR}_u^L(X_2))\,\mathrm{d}u,$$

where the second inequality is due to (40) and $V_{\delta_2}(X_2) \leqslant 0$. Therefore,

$$\begin{aligned}\int_t^\alpha (-\mathrm{VaR}_u^L(X_1^*))\,\mathrm{d}u &= \int_t^\alpha (\mathrm{VaR}_u^L(X_2^*) - \mathrm{VaR}_u^L(X))\,\mathrm{d}u \\ &\geqslant \int_t^\alpha (\mathrm{VaR}_u^L(X_2) - \mathrm{VaR}_u^L(X))\,\mathrm{d}u = \int_t^\alpha (-\mathrm{VaR}_u^L(X_1))\,\mathrm{d}u.\end{aligned}$$

Since $\int_t^\alpha (\mathrm{VaR}_u^L(X_1) - \mathrm{VaR}_u^L(X_1^*))\,\mathrm{d}u \geqslant 0$ and $\mathrm{VaR}_u^L(X_1) - \mathrm{VaR}_u^L(X_1^*)$ is increasing in u, we have

$$\int_x^\alpha (-\mathrm{VaR}_u^L(X_1^*))\,\mathrm{d}u \geqslant \int_x^\alpha (-\mathrm{VaR}_u^L(X_1))\,\mathrm{d}u$$

for all $x \in [t, \alpha]$. Hence,

$$\begin{aligned}\int_0^\alpha (-\mathrm{VaR}_u^L(X_1^*))_+\,\mathrm{d}u &\geqslant \int_s^\alpha (-\mathrm{VaR}_u^L(X_1^*))_+\,\mathrm{d}u \\ &\geqslant \int_s^\alpha (-\mathrm{VaR}_u^L(X_1^*))\,\mathrm{d}u \\ &\geqslant \int_s^\alpha (-\mathrm{VaR}_u^L(X_1))\,\mathrm{d}u = \int_0^\alpha (-\mathrm{VaR}_u^L(X_1))_+\,\mathrm{d}u.\end{aligned}$$

Using (40), we know that $V_{\delta_1}(X_1^*) \leqslant 0$. Hence, if $V_{\delta_1}(X_1) \leqslant 0$ and $V_{\delta_2}(X_2) \leqslant 0$, then $V_1(X_1^*) \leqslant 0$ and $V_2(X_2^*) \leqslant 0$. Using translation-invariance of V_{δ_1} and V_{δ_2}, we obtain

$$V_{\delta_1}(Y_1) + V_{\delta_2}(Y_2) \geqslant V_1(X_1^*) + V_2(X_2^*) = V_{\delta_1}(X) + \frac{\delta_2}{\alpha}$$

for any $(Y_1, Y_2) \in \mathbb{A}_2^+(X)$ with $\mathbb{P}(Y_2 > 0) \leqslant \mathbb{P}(Y_1 > 0)$. By symmetry,

$$V_{\delta_1}(Y_1) + V_{\delta_2}(Y_2) \geqslant V_1(X_1^*) + V_2(X_2^*) = V_{\delta_2}(X) + \frac{\delta_1}{\alpha}$$

for any $(Y_1, Y_2) \in \mathbb{A}_2^+(X)$ with $\mathbb{P}(Y_2 > 0) \geqslant \mathbb{P}(Y_1 > 0)$. Therefore,

$$V_{\delta_1} \boxplus V_{\delta_2}(X) \geqslant \min\left\{V_{\delta_1}(X) + \frac{\delta_2}{\alpha}, V_{\delta_2}(X) + \frac{\delta_1}{\alpha}\right\} = V_{\delta_1}(X) + \frac{\delta_2}{\alpha},$$

where the last equality is due to Proposition 4 and the assumption $\delta_1 \leqslant \delta_2$. Together with (39) we obtain (38) which holds for X satisfying $V_{\delta_1} \boxplus V_{\delta_2}(X) \leqslant 0$. Using translation-invariance of $V_{\delta_1} \boxplus V_{\delta_2}$, we know that (38) holds for all $X \in \mathcal{X}$.

(ii) We first show that for $\lambda \geqslant 1$, we have

$$[\mathrm{VaR}^{\Lambda}_{\alpha}]^{1}_{\delta}(Y) \geqslant [\mathrm{VaR}^{\Lambda}_{\lambda\alpha}]^{1}_{\lambda\delta}(Y), \quad Y \in \mathcal{X}. \tag{41}$$

This follows from (34) by noting that for any $x \in \mathbb{R}$,

$$\int_0^{\lambda\alpha} \left(x - \mathrm{VaR}^R_u(Y)\right)_+ \mathrm{d}u = \lambda\delta \implies \int_0^{\alpha} \left(x - \mathrm{VaR}^R_u(Y)\right)_+ \mathrm{d}u \leqslant \delta.$$

Using (41), by letting $\alpha = \alpha_1$ and $\delta_i = \alpha\delta/\alpha_i \geqslant \delta$ for each $i = 1, \dots, n$, we have

$$[\mathrm{VaR}^{\Lambda}_{\alpha}]^{1}_{\delta_i} \leqslant [\mathrm{VaR}^{\Lambda}_{\alpha_i}]^{1}_{\delta} \quad \text{on } \mathcal{X}.$$

Therefore, (36) yields

$$[\mathrm{VaR}^{\Lambda}_{\alpha}]^{1}_{\delta_1}(X) + \sum_{i=2}^{n} \frac{\delta_i}{\alpha} = \mathop{\boxplus}_{i=1}^{n} [\mathrm{VaR}^{\Lambda}_{\alpha}]^{1}_{\delta_i}(X) \leqslant \mathop{\boxplus}_{i=1}^{n} [\mathrm{VaR}^{\Lambda}_{\alpha_i}]^{1}_{\delta}(X).$$

On the other hand, the allocation $(X, 0, \dots, 0)$ attains the lower bound above, and hence (37) holds. $\square$

Assume that in a risk sharing problem with model uncertainty, each agent i uses $[\mathrm{VaR}_{\alpha_i}]^{1}_{\delta_i}$ as her risk measure. The parameter α_i represents the agent's attitude towards risk, and the parameter δ_i represents the agent's attitude towards uncertainty. More specifically, a smaller α_i represents more sensitivity towards risk, and a larger δ_i represents more sensitivity towards uncertainty. The results in Theorem 6 illustrate the following observations in a comonotonic risk sharing problem for these agents.

1. If all agents have the same risk attitude (i.e., identical α_i), then an optimal allocation is to allocate all random loss to the agent who is the least sensitive to uncertainty (the smallest δ_i). Agents may make side-payments in cash to maintain individual rationality; see e.g., Embrechts et al. (2018).

2. If all agents have the same uncertainty attitude (i.e., identical δ_i), then an optimal allocation is to allocate all random loss to the agent who is the least sensitive to risk (the largest α_i).

3. In the case all agents have the same risk parameter α and uncertainty parameter δ, the equally weighted allocation $(X/n, \dots, X/n)$ is generally not optimal. Indeed, one can verify that $(X, 0, \dots, 0)$, which is optimal by Theorem 6, usually strictly outperforms $(X/n, \dots, X/n)$. On the other hand, if all agents use the same convex risk measure, then $(X/n, \dots, X/n)$ is

always optimal due to convexity; note that the robust version of a convex risk measure is again convex. Hence, in the context of model uncertainty, VaR still exhibits the "all or nothing" feature observed by Embrechts et al. (2018, 2020), in sharp contrast to the optimal allocations for convex risk measures, which are often proportional (see e.g., Theorem 3.9 of Barrieu and El Karoui (2005)).

4. If agents have neither identical uncertainty attitude nor identical risk attitude, then the problem becomes quite difficult to analyze and this seems to require different techniques to be developed in the future.

9 Concluding remarks

Motivated by several questions related to the RVaR and VaR formulas in (2) and (3) discovered by Embrechts et al. (2018), the inf-convolution of tail risk measures is analyzed in detail. Because every tail risk measure corresponds one-to-one a law-invariant risk measure (its generator), analytical results for the inf-convolution or its optimal allocation cannot be expected in general without specifying the form of the tail risk measures. For the special setting of several VaRs, or that of a VaR and another tail risk measure, we are able to obtain the inf-convolution and the corresponding optimal allocation in explicit forms. Moreover, we found that tail risk measures exhibit similar properties to the RVaR and VaR formula, and in particular their inf-convolution is still a tail risk measure. Explicit allocations are found in the setting of elliptical models, and several results are obtained for tail risk measures and risk sharing problems in the presence of model uncertainty. Results in this paper complement the work by Embrechts et al. (2018) on risk sharing problems for VaR, ES and RVaR.

Risk sharing with model uncertainty, studied in Section 8, gives rise to many open questions that are not addressed with current techniques. For instance, in the setting of model uncertainty induced by general ϕ-divergences or Wasserstein metrics, optimal allocations are unclear in either the constrained or the unconstrained setting, even for VaR agents. These challenging questions require future work.

Acknowledgements

The authors thank the Editors, two anonymous referees, Taizhong Hu, and Haiyan Liu for constructive comments on an earlier version of the paper. In particular, results in Sections 7 and 8 are largely inspired by the received comments. F. Liu acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2020-04717, DGECR-2020-00340). T. Mao would like to thank the financial support from National Natural Science Foundation of China (Nos. 71871208, 71921001, 71671176). 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. L. Wei acknowledges financial support from the China Scholarship Council and National Natural Science Foundation of China (No. 12001411). All authors contributed equally to the paper. Tiantian Mao and Linxiao Wei are corresponding authors.

A Some background on risk measures

In this section we collect some common terminology and results on risk measures, which are briefly mentioned in the text of the paper, but not essential to the presentation of our main results.

First, some standard properties of risk measures, in addition to (i) law-invariance, (ii) monotonicity and (iii) translation-invariance, are listed below.

(iv) *Convexity*: $\rho(\lambda X + (1-\lambda)Y) \leqslant \lambda\rho(X) + (1-\lambda)\rho(Y)$ for $\lambda \in [0,1]$ and $X, Y \in \mathcal{X}$.

(v) *Positive homogeneity*: $\rho(\lambda X) = \lambda\rho(X)$ for $\lambda > 0$ and $X \in \mathcal{X}$.

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

(vii) *Comonotonic additivity*: $\rho(X+Y) = \rho(X) + \rho(Y)$ if $X, Y \in \mathcal{X}$ are comonotonic.

Using the standard terminology in Föllmer and Schied (2016), a risk measure ρ is a *convex risk measure* if it is monetary and convex, and it is a *coherent risk measure* if it is monetary, convex, and positively homogeneous. It is well known that a law-invariant convex risk measure on L^∞ is convex-order consistent; see e.g., Föllmer and Schied (2016, Corollary 4.65). The risk measures ES_α, $\alpha \in [0,1)$ belong to the RVaR family, and they are the only ones in that family that are coherent, convex, or convex-order consistent (see e.g., Wang et al. (2020)).

The class of *distortion risk measures* is defined by

$$\rho_h(X) = \int_0^\infty h \circ \mathbb{P}(X > x)\,\mathrm{d}x - \int_{-\infty}^0 (1 - h \circ \mathbb{P}(X > x))\,\mathrm{d}x, \quad X \in \mathcal{X},$$

for some non-decreasing function $h : [0,1] \to [0,1]$ satisfying $h(0) = \lim_{x\downarrow 0} h(x) = 0$ and $h(1) = \lim_{x\uparrow 0} h(x) = 1$, such that the above integral is finite. Here h is called a *distortion function.* Via a quantile representation (e.g., Dhaene et al. (2012)), all of VaR^L, VaR^R, ES and RVaR belong to the class of distortion risk measures; indeed, all comonotonic-additive monetary risk measures are distortion risk measures (e.g., Theorem 1 of Wang et al. (2020)). For $p \in (0,1)$, a distortion risk measure ρ_h is a p-tail risk measure if and only if it dominates VaR_p^R, which is equivalent to $h(p) = 1$. If h is continuous, then ρ_h has a representation (33).

Using the terminology of Weber (2018), a risk measure ρ is VaR-type with parameter $\alpha \in (0,1)$ if it satisfies

$$\rho(X) = \rho\left(X\mathbb{1}_{\{X \leqslant \mathrm{VaR}_\alpha^L(X)\}} + \mathrm{VaR}_\alpha^L(X)\mathbb{1}_{\{X > \mathrm{VaR}_\alpha^L(X)\}}\right), \quad X \in \mathcal{X}.$$

In other words, such a risk measure ρ ignores the tail part of the distribution with probability α. It is easy to check that ρ is VaR-type with parameter α if and only if $X \mapsto \rho(-X)$ is a $(1-\alpha)$-tail risk measure.

B Proofs of all propositions, lemmas, and corollaries

Proof of Proposition 1. Noting that ES_β is continuous from above with respect to a.s. convergence, (19) leads to

$$\mathrm{VaR}_\alpha^R \square \mathrm{ES}_\beta(X) = \mathrm{ES}_\beta(X^{[\alpha]}) = \mathrm{RVaR}_{\alpha,\beta}(X), \quad X \in \mathcal{X}.$$

Next we analyze the existence of an optimal allocation. First assume $\mathrm{VaR}_{\alpha+\beta}^R(X) = \mathrm{VaR}_\alpha^R(X)$. As in Theorem 2, we take

$$X^{[\alpha]} = X\mathbb{1}_{\{U_X \leqslant 1-\alpha\}} + \mathrm{VaR}_{\alpha+\beta}^R(X)\mathbb{1}_{\{U_X > 1-\alpha\}}.$$

It is easy to check that $\mathrm{VaR}_\alpha^L(X - X^{[\alpha]}) = 0$, $\mathrm{ES}_\beta(X^{[\alpha]}) = \mathrm{RVaR}_{\alpha,\beta}(X)$, and by Theorem 2, $(X - X^{[\alpha]}, X^{[\alpha]})$ is an optimal allocation of X for $(\mathrm{VaR}_\alpha^L, \mathrm{ES}_\beta)$. For any $\delta > 0$, (10) leads to

$$\mathbb{P}\left(X - X^{[\alpha]} > \delta\right) \leqslant \mathbb{P}\left(X - \mathrm{VaR}_{\alpha+\beta}^R(X) > \delta\right) = \mathbb{P}\left(X > \mathrm{VaR}_\alpha^R(X) + \delta\right) < \alpha,$$

which implies $\mathrm{VaR}_\alpha^R(X - X^{[\alpha]}) \leqslant 0$. It follows that

$$\mathrm{VaR}_\alpha^R(X - X^{[\alpha]}) + \mathrm{ES}_\beta(X^{[\alpha]}) \leqslant \mathrm{RVaR}_{\alpha,\beta}(X) = \mathrm{VaR}_\alpha^R \square \mathrm{ES}_\beta(X).$$

Therefore, $(X - X^{[\alpha]}, X^{[\alpha]})$ is an optimal allocation of X for $(\mathrm{VaR}_\alpha^R, \mathrm{ES}_\beta)$.

In the following, we show that, if $\mathrm{VaR}_{\alpha+\beta}^R(X) < \mathrm{VaR}_\alpha^R(X)$, then there is no optimal allocation for $(\mathrm{VaR}_\alpha^R, \mathrm{ES}_\beta)$. Suppose that (X_1, X_2) is an optimal allocation of X for $(\mathrm{VaR}_\alpha^R, \mathrm{ES}_\beta)$. We have

$$\mathrm{RVaR}_{\alpha,\beta}(X) = \mathrm{VaR}_\alpha^R \square \mathrm{ES}_\beta(X) = \mathrm{VaR}_\alpha^R(X_1) + \mathrm{ES}_\beta(X_2) \geqslant \mathrm{VaR}_\alpha^L(X_1) + \mathrm{ES}_\beta(X_2).$$

Noting that

$$\mathrm{RVaR}_{\alpha,\beta}(X) = \mathrm{VaR}^L_\alpha \square \mathrm{ES}_\beta(X) \leqslant \mathrm{VaR}^L_\alpha(X_1) + \mathrm{ES}_\beta(X_2),$$

we obtain $\mathrm{RVaR}_{\alpha,\beta}(X) = \mathrm{VaR}^L_\alpha(X_1) + \mathrm{ES}_\beta(X_2)$, and therefore, (X_1, X_2) is an optimal allocation of X for $(\mathrm{VaR}^L_\alpha, \mathrm{ES}_\beta)$. We use Theorem 4.8 of Wang and Wei (2020), which gives a full characterization of all optimal allocations (X_1, X_2) for $(\mathrm{VaR}^L_\alpha, \mathrm{ES}_\beta)$ as follows:

$$\begin{cases} X_1 = Y\mathbb{1}_B - Z + c, \quad X_2 = X - X_1 \\ \text{where } B \in \mathcal{F} \text{ satisfies } \{X > \mathrm{VaR}^L_\alpha(X)\} \subset B,\ \mathbb{P}(B) = \alpha, \\ \text{and moreover } B \subset \{X \geqslant \mathrm{VaR}^L_\alpha(X)\} \text{ if } \mathrm{VaR}^R_{\alpha+\beta}(X) \neq \mathrm{VaR}^L_\alpha(X), \\ Y \geqslant X - \mathrm{VaR}^R_{\alpha+\beta}(X), \quad 0 \leqslant Z \leqslant (\mathrm{VaR}^R_{\alpha+\beta}(X) - X + Y\mathbb{1}_B)_+, \text{ and } c \in \mathbb{R}. \end{cases} \tag{42}$$

Since both VaR^L_α and ES_β are monetary risk measures, the constant c does not matter for the optimality of (X_1, X_2), and we can set $c = 0$ for simplicity. It is easy to verify that $\mathrm{VaR}^L_\alpha(X_1) = 0$ and $\mathrm{ES}_\beta(X_2) = \mathrm{RVaR}_{\alpha,\beta}(X)$. Consequently, we have $\mathrm{VaR}^R_\alpha(X_1) = \mathrm{RVaR}_{\alpha,\beta}(X) - \mathrm{ES}_\beta(X_2) = 0$, which implies

$$\mathbb{P}(X_1 > \varepsilon) < \alpha, \qquad \text{for all } \varepsilon > 0. \tag{43}$$

Note that, for X_1 in (42) and $\varepsilon > 0$, we have

$$\begin{aligned} \mathbb{P}(X_1 > \varepsilon) &\geqslant \mathbb{P}\left(Y\mathbb{1}_B - (\mathrm{VaR}^R_{\alpha+\beta}(X) - X + Y\mathbb{1}_B)_+ > \varepsilon\right) \\ &= \mathbb{P}\left(\min\left\{Y\mathbb{1}_B, X - \mathrm{VaR}^R_{\alpha+\beta}(X)\right\} > \varepsilon\right) \\ &\geqslant \mathbb{P}\left((X - \mathrm{VaR}^R_{\alpha+\beta}(X))\mathbb{1}_B > \varepsilon\right). \end{aligned} \tag{44}$$

If $\mathrm{VaR}^R_{\alpha+\beta}(X) = \mathrm{VaR}^L_\alpha(X) < \mathrm{VaR}^R_\alpha(X)$, we can take $\varepsilon_1 = \mathrm{VaR}^R_\alpha(X) - \mathrm{VaR}^L_\alpha(X) - \delta$ for some $0 < \delta < \mathrm{VaR}^R_\alpha(X) - \mathrm{VaR}^L_\alpha(X)$. Plugging $\varepsilon = \varepsilon_1$ in (44), and noting that $\{X > \mathrm{VaR}^L_\alpha(X) + \varepsilon_1\} \subset B$, we obtain

$$\mathbb{P}(X_1 > \varepsilon_1) \geqslant \mathbb{P}\left(X > \mathrm{VaR}^L_\alpha(X) + \varepsilon_1\right) = \mathbb{P}\left(X > \mathrm{VaR}^R_\alpha(X) - \delta\right) \geqslant \alpha,$$

which contradicts (43). If $\mathrm{VaR}^R_{\alpha+\beta}(X) < \mathrm{VaR}^L_\alpha(X)$, we take $0 < \varepsilon_2 < \mathrm{VaR}^L_\alpha(X) - \mathrm{VaR}^R_{\alpha+\beta}(X)$. Plugging $\varepsilon = \varepsilon_2$ in (44), and noting that $B \subset \{X \geqslant \mathrm{VaR}^L_\alpha(X)\}$, we obtain

$$\mathbb{P}(X_1 > \varepsilon_2) \geqslant \mathbb{P}\left(B \cap \{X - \mathrm{VaR}^R_{\alpha+\beta}(X) \geqslant \mathrm{VaR}^L_\alpha(X) - \mathrm{VaR}^R_{\alpha+\beta}(X)\}\right) = \mathbb{P}(B) = \alpha,$$

which contradicts (43). Therefore, as long as $\mathrm{VaR}^R_{\alpha+\beta}(X) < \mathrm{VaR}^R_\alpha(X)$, we can conclude that $(\mathrm{VaR}^R_\alpha, \mathrm{ES}_\beta)$ does not have an optimal allocation. $\square$

Proof of Lemma 1. We first prove the case $X' = \max\{X, \mathrm{VaR}^R_\varepsilon(X)\}$. It is easy to check that $\mathbb{P}(X \leqslant \mathrm{VaR}^R_\varepsilon(X)) \geqslant 1-\varepsilon$, and the distribution functions of X_ε and X'_ε are respectively given by, for $t \in \mathbb{R}$,

$$\mathbb{P}(X_\varepsilon \leqslant t) = \frac{(F_X(t) - (1-\varepsilon))_+}{\varepsilon} \quad \text{and} \quad \mathbb{P}(X'_\varepsilon \leqslant t) = \frac{(F_{X'}(t) - (1-\varepsilon))_+}{\varepsilon}.$$

For $t < \mathrm{VaR}^R_\varepsilon(X)$, we have $\mathbb{P}(X' \leqslant t) = 0$ and $\mathbb{P}(X \leqslant t) \leqslant 1-\varepsilon$, which imply that both $\mathbb{P}(X_{1-\varepsilon} \leqslant t)$ and $\mathbb{P}(X'_{1-\varepsilon} \leqslant t)$ are zero. For $t \geqslant \mathrm{VaR}^R_\varepsilon(X)$, we also have

$$\mathbb{P}(X' \leqslant t) = \mathbb{P}(\max\{X, \mathrm{VaR}^R_\varepsilon(X)\} \leqslant t) = \mathbb{P}(X \leqslant t).$$

Therefore $X'_\varepsilon \stackrel{\mathrm{d}}{=} X_\varepsilon$.

Next, consider the case $X' = \max\{X, m\}$ for some $m \in [\mathrm{VaR}^L_\varepsilon(X), \mathrm{VaR}^R_\varepsilon(X))$. Note that $X \leqslant X' \leqslant X''$ where $X'' = \max\{X, \mathrm{VaR}^R_\varepsilon(X)\}$. Using the result $X_\varepsilon \stackrel{\mathrm{d}}{=} X''_\varepsilon$ obtained above and (6), we know that $X'_\varepsilon \stackrel{\mathrm{d}}{=} X_\varepsilon$. □

Proof of Lemma 2. (a) Let $Y_i = F^{-1}_{X_i}(1-\varepsilon+\varepsilon U_X)$ for $i = 1, \dots, n$. Note that $Y_1, \dots, Y_n$ are comonotonic. Comonotonicity of $X_1, \dots, X_n$ implies $F^{-1}_X = \sum_{i=1}^n F^{-1}_{X_i}$ (this is because quantiles are comonotonic-additive; see e.g., Proposition 7.20 of McNeil et al. (2015)), and hence we have $\sum_{i=1}^n Y_i = X_\varepsilon$. Moreover, it is clear from the definition that $Y_i \stackrel{\mathrm{d}}{=} (X_i)_\varepsilon$ for $i = 1, \dots, n$.

(b) Note again from comonotonicity that $F^{-1}_{X_\varepsilon} = \sum_{i=1}^n F^{-1}_{Y_i}$. This implies that $F^{-1}_{Y_i}(0) > -\infty$ for each $i = 1, \dots, n$, since $F^{-1}_{X_\varepsilon}(0) = \mathrm{VaR}^R_\varepsilon(X) > -\infty$. For $i = 1, \dots, n$, let

$$X_i = F^{-1}_{Y_i}\left(\frac{U_X - (1-\varepsilon)}{\varepsilon}\right) \mathbb{1}_{\{U_X > 1-\varepsilon\}} + \left(\frac{X - \mathrm{VaR}^R_\varepsilon(X)}{n} + F^{-1}_{Y_i}(0)\right) \mathbb{1}_{\{U_X \leqslant 1-\varepsilon\}}.$$

It is easy to check that $X_i /\!\!/ X$ for $i = 1, ..., n$. Moreover,

$$\begin{aligned}
\sum_{i=1}^n X_i &= F^{-1}_{X_\varepsilon}\left(\frac{U_X - (1-\varepsilon)}{\varepsilon}\right) \mathbb{1}_{\{U_X > 1-\varepsilon\}} + \left(X - \mathrm{VaR}^R_\varepsilon(X) + F^{-1}_{X_{1-\varepsilon}}(0)\right) \mathbb{1}_{\{U_X \leqslant 1-\varepsilon\}} \\
&= F^{-1}_X(U_X) \mathbb{1}_{\{U_X > 1-\varepsilon\}} + \left(X - \mathrm{VaR}^R_\varepsilon(X) + \mathrm{VaR}^R_\varepsilon(X)\right) \mathbb{1}_{\{U_X \leqslant 1-\varepsilon\}} \\
&= F^{-1}_X(U_X) \mathbb{1}_{\{U_X > 1-\varepsilon\}} + X \mathbb{1}_{\{U_X \leqslant 1-\varepsilon\}} = X.
\end{aligned}$$

Thus, $(X_1, \dots, X_n) \in \mathbb{A}^+_n(X)$. By definition $(X_i)_\varepsilon = F^{-1}_{X_i}((1-\varepsilon) + \varepsilon U_{X_i})$. Also note that, by construction, $\mathrm{VaR}^R_\varepsilon(X_i) = F^{-1}_{Y_i}(0)$. For $x \geqslant F^{-1}_{Y_i}(0)$, we have

$$\begin{aligned}
F_{(X_i)_{1-\varepsilon}}(x) = \frac{F_{X_i}(x) - (1-\varepsilon)}{\varepsilon} &= \frac{1}{\varepsilon}\mathbb{P}\left(F^{-1}_{Y_i}\left(\frac{U_X - (1-\varepsilon)}{\varepsilon}\right) \leqslant x,\ U_X > 1-\varepsilon\right) \\
&= \frac{1}{\varepsilon}\mathbb{P}\left(1-\varepsilon \leqslant U_X \leqslant \varepsilon F_{Y_i}(x) + (1-\varepsilon)\right) = F_{Y_i}(x).
\end{aligned}$$

That is $(X_i)_\varepsilon \stackrel{\mathrm{d}}{=} Y_i$ for $i = 1, \dots, n$. □

Proof of Proposition 2. (i) First note that, given a random variable X and $Q \in \mathcal{P}_\lambda$, we have $Q(A) \leqslant \mathbb{P}(A)/\lambda$ for $A \in \mathcal{F}$ and thus,

$$\mathbb{P}((X_Q)_\varepsilon \leqslant x) = \frac{(\varepsilon - Q\,(X > z))_+}{\varepsilon} \geqslant \frac{\left(\varepsilon - \frac{1}{\lambda}\mathbb{P}\,(X > z)\right)_+}{\varepsilon} = \mathbb{P}(X_{\lambda\varepsilon} \leqslant z).$$

It follows from the law-invariance and monotonicity of ρ^* inherited from ρ that

$$\overline{\rho}^{\mathcal{P}_\lambda}(X) = \sup_{Q \in \mathcal{P}_\lambda} \rho^*((X_Q)_\varepsilon) \leqslant \rho^*(X_{\lambda\varepsilon}). \tag{45}$$

On the other hand, for any X, define a probability measure Q' as $\mathrm{d}Q'/\,\mathrm{d}\mathbb{P} = \mathbb{1}_{\{U_X > 1-\lambda\}}/\lambda$. We have

$$\begin{aligned}
\mathbb{P}((X_{Q'})_\varepsilon > z) &= \frac{(\varepsilon - Q'(X > z))_+}{\varepsilon} \\
&= \frac{(\varepsilon - \mathbb{P}(X > z, U_X > 1-\lambda)/\lambda)_+}{\varepsilon} \\
&= \frac{(\lambda\varepsilon - \mathbb{P}(X > z))_+}{\varepsilon\lambda} = \mathbb{P}(X_{\varepsilon\lambda} > z),
\end{aligned}$$

where the third equality is due to $\mathbb{P}(X > z, U_X > 1-\lambda) = \min\{\mathbb{P}(X > z), \lambda\}$ and $(\varepsilon - \lambda/\lambda)_+ = 0$. Therefore,

$$\rho^*(X_{\lambda\varepsilon}) = \rho^*((X_{Q'})_\varepsilon) = \rho^{Q'}(X) \leqslant \sup_{Q \in \mathcal{P}_\lambda} \rho^Q(X) = \overline{\rho}^{\mathcal{P}_\lambda}(X). \tag{46}$$

Combining (45) and (46), we have that $\overline{\rho}^{\mathcal{P}_\lambda}$ is an $(\varepsilon\lambda)$-tail risk measure generated by ρ^*. Finally, the monotonicity of ρ^* implies the monotonicity of $\overline{\rho}^{\mathcal{P}_\lambda}$.

(ii) It follows immediately from (i) that

$$\overline{\rho}^{\mathcal{P}_\lambda}(X) = \rho^*(X_{\lambda\varepsilon}) = \rho^*((X_\lambda)_\varepsilon) = \rho(X_\lambda).$$

(iii) Suppose ρ is continuous from above with respect to a.s. convergence. Take $Y_n \downarrow X$ a.s. with respect to $\mathbb{P}$ as $n \to \infty$. Then $(Y_n)_\lambda \downarrow X_\lambda$ a.s. with respect to $\mathbb{P}$. By (ii), we have

$$\overline{\rho}^{\mathcal{P}_\lambda}(Y_n) = \rho((Y_n)_\lambda) \downarrow \rho(X_\lambda) = \overline{\rho}^{\mathcal{P}_\lambda}(X), \qquad n \to \infty,$$

i.e., $\overline{\rho}^{\mathcal{P}_\lambda}$ is also continuous from above with respect to a.s. convergence under $\mathbb{P}$. If ρ is continuous from above with respect to a.s. convergence, then so is $\overline{\rho}^{\mathcal{P}_\lambda}$.

(iv) We only consider the case that ρ is convex-order consistent. Note that for any pair of random variables X and Y, $X \prec_{\mathrm{cx}} Y$ implies $X_\lambda \prec_{\mathrm{icx}} Y_\lambda$, where $X \prec_{\mathrm{icx}} Y$ means that $\mathbb{E}[f(X)] \leqslant$

$\mathbb{E}[f(Y)]$ for all increasing convex functions f, provided that both expectations exist. By Theorem 4.A.6 of Shaked and Shanthikumar (2007), there exists Z such that $X_\lambda \leqslant Z \prec_{\mathrm{cx}} Y_\lambda$. It then follows from the monotonicity of ρ that

$$\overline{\rho}^{\mathcal{P}_\lambda}(X) = \rho(X_\lambda) \leqslant \rho(Z) \leqslant \rho(Y_\lambda) = \overline{\rho}^{\mathcal{P}_\lambda}(Y).$$

This completes the proof. $\square$

Proof of Corollary 2. The case $\lambda = 1$ is trivial. For $\lambda < 1$, the first and the third equalities in (31) follow directly from Proposition 2. To show $\overline{\mathrm{VaR}}_\alpha^{L,\mathcal{P}_\lambda} = \mathrm{VaR}_{\alpha\lambda}^L$, note that $(\mathrm{VaR}_{\alpha+\delta}^R)^Q \leqslant (\mathrm{VaR}_\alpha^L)^Q \leqslant (\mathrm{VaR}_\alpha^R)^Q$ for all $\delta \in (0, 1-\alpha)$ and probability measure Q. Using the first equality in (31), we have

$$\overline{\mathrm{VaR}}_\alpha^{L,\mathcal{P}_\lambda} \geqslant \overline{\mathrm{VaR}}_{\alpha+\delta}^{R,\mathcal{P}_\lambda} = \mathrm{VaR}_{(\alpha+\delta)\lambda}^R \geqslant \mathrm{VaR}_{(\alpha+\delta)\lambda}^L.$$

Since VaR_q^L is right-continuous with respect to q, taking $\delta \downarrow 0$, we have

$$\overline{\mathrm{VaR}}_\alpha^{L,\mathcal{P}_\lambda} \geqslant \lim_{\delta \downarrow 0} \mathrm{VaR}_{(\alpha+\delta)\lambda}^L = \mathrm{VaR}_{\alpha\lambda}^L.$$

On the other hand, if a constant z satisfies $\mathbb{P}(X > z) \leqslant \alpha\lambda$, then $Q(X > z) \leqslant \alpha$ for all $Q \in \mathcal{P}_\lambda$. Thus, $(\mathrm{VaR}_\alpha^L)^Q \leqslant \mathrm{VaR}_{\alpha\lambda}^L$ for all $Q \in \mathcal{P}_\lambda$. This implies $\overline{\mathrm{VaR}}_\alpha^{L,\mathcal{P}_\lambda} \leqslant \mathrm{VaR}_{\alpha\lambda}^L$. $\square$

Proof of Proposition 3. Let $\mathcal{R}$ be a monotone risk measure. In this proof, we employ the notation $\mathcal{R}(F) = \mathcal{R}(X)$, where F is the distribution of X, and denote by F_ε the distribution function of X_ε. Note that for any distribution G such that $W_k(F, G) \leqslant \delta$, let G^* be a distribution function satisfying $(G^*)^{-1} := \max\{G^{-1}, F^{-1}\}$. One can verify that $W_k(G^*, F) \leqslant W_k(G, F) \leqslant \delta$ and $\mathcal{R}(G^*) \geqslant \mathcal{R}(G)$. Therefore, we have

$$[\mathcal{R}]_\delta^k(F) = \sup\left\{\mathcal{R}(G) : W_k(G, F) \leqslant \delta,\ F \geqslant G\right\}. \tag{47}$$

It follows that

$$\begin{aligned}
[\rho^*]_{\varepsilon^{-1/k}\delta}^k(F_\varepsilon) &= \sup\{\rho^*(G) : W_k(G, F_\varepsilon) \leqslant \varepsilon^{-1/k}\delta\} \\
&= \sup\{\rho^*(G) : W_k(G, F_\varepsilon) \leqslant \varepsilon^{-1/k}\delta,\ F_\varepsilon \geqslant G\} \\
&= \sup\{\rho^*(G_\varepsilon) : W_k(G_\varepsilon, F_\varepsilon) \leqslant \varepsilon^{-1/k}\delta,\ F_\varepsilon \geqslant G_\varepsilon\} \\
&\geqslant \sup\{\rho(G) : W_k(G, F) \leqslant \delta,\ F \geqslant G\} \qquad (48) \\
&= [\rho]_\delta^k(F),
\end{aligned}$$

where the third equality follows from that any random variable bounded from below can be viewed as an ε-tail risk of some random variable, and the inequality (48) follows from that $W_k(F,G) \leqslant \delta$ and $F \geqslant G$ imply $W_k(F_\varepsilon, G_\varepsilon) \leqslant \varepsilon^{-1/k}\delta$ and $F_\varepsilon \geqslant G_\varepsilon$.

We next show $[\rho^*]^k_{\varepsilon^{-1/k}\delta}(F_\varepsilon) \leqslant [\rho]^k_\delta(F)$. It suffices to show the inequality (48) can be reversed. Note that for any distribution G with $W_k(F_\varepsilon, G_\varepsilon) \leqslant \varepsilon^{-1/k}\delta$ and $F_\varepsilon \geqslant G_\varepsilon$, define G^* as

$$(G^*)^{-1}(x) = G^{-1}(x)\mathbb{1}_{\{x>1-\varepsilon\}} + F^{-1}(x)\mathbb{1}_{\{x\leqslant 1-\varepsilon\}}, \quad x \in (0,1),$$

which is a well-defined distribution as $F^{-1} \leqslant G^{-1}$. One can verify that $F \geqslant G^*$, $G_\varepsilon = G^*_\varepsilon$, and $W_k(G^*, F) = \varepsilon^{1/k} W_k(G_\varepsilon, F_\varepsilon) \leqslant \delta$. As a consequence,

$$\sup\{\rho^*(G_\varepsilon) : W_k(G_\varepsilon, F_\varepsilon) \leqslant \varepsilon^{-1/k}\delta,\ F_\varepsilon \geqslant G_\varepsilon\} \leqslant \sup\{\rho(G) : W_k(G,F) \leqslant \delta,\ F \geqslant G\}$$

and thus the inequality in (48) is an equality. The other statement of the proposition is straightforward from (48). $\square$

Proof of Proposition 4. (i) Note that for any random variable $Y \sim G$ such that $W_k(Y,X) \leqslant \delta$, let Y^* be a random variable with distribution function G^* given by $(G^*)^{-1} = \max\{G^{-1}, F^{-1}\}$. We can verify that $W_k(Y^*, X) \leqslant W_k(Y,X) \leqslant \delta$ and $\rho_h(Y^*) \geqslant \rho_h(Y)$. Therefore, we have

$$[\rho_h]^k_\delta(X) = \sup\left\{\rho_h(Y) : \mathbb{E}[(Y-X)^k] \leqslant \delta^k,\ Y \geqslant X \text{ a.s.}\right\}. \tag{49}$$

By subadditivity and comonotonic additivity of ρ_h, we have $[\rho_h]^k_\delta(X) = \rho_h(X) + c$, where

$$c = \sup\left\{\rho_h(Z) : \mathbb{E}[Z^k] \leqslant \delta^k,\ Z \geqslant 0 \text{ a.s.}\right\}. \tag{50}$$

It suffices to show $c = \delta\|h'\|_q$, which will follow from Hölder's inequality. More specifically, we consider two cases separately.

(a) If $k=1$, then $\rho_h(Z) \leqslant \|h'\|_\infty \int_0^1 \mathrm{VaR}^L_u(Z)\,\mathrm{d}u = \delta\|h'\|_\infty$ for any Z satisfying (50) which implies $c \leqslant \delta\|h'\|_\infty$. For $\varepsilon \in (0,1)$, take Z_ε satisfying $\mathbb{P}(Z_\varepsilon = \delta/\varepsilon) = \varepsilon = 1 - \mathbb{P}(Z_\varepsilon = 0)$. We have $\lim_{\varepsilon\downarrow 0} \rho_h(Z_\varepsilon) = \lim_{\varepsilon\downarrow 0} \int_0^\varepsilon \mathrm{VaR}^L_u(Z_\varepsilon) h'(u)\,\mathrm{d}u = \delta\|h'\|_\infty$. Hence, $c = \delta\|h'\|_\infty$.

(b) If $k>1$, then by Hölder's inequality, we have $c \leqslant \delta\|h'\|_q$. Take a random variable Z^* with quantile function given by $\mathrm{VaR}_u(Z^*) = \delta(h'(u))^{q-1}/\|h'\|_q^{q/k}$, $u \in [0,1]$. We have $\rho_h(Z^*) = \delta\|h'\|_q$ which implies $c = \delta\|h'\|_q$.

Statement (ii) follows immediately from (i).

(iii) Note that for any random variable $Y \sim G$ such that $W_k(Y, X) \leqslant \delta$, define $Y^* \sim G^*$ such that its quantile function is

$$(G^*)^{-1}(x) := F^{-1}(x)1_{\{x<1-\alpha\}} + \left(G^{-1}(1-\alpha) \vee F^{-1}(x)\right) 1_{\{x\geqslant 1-\alpha\}}, \quad x \in (0,1). \tag{51}$$

It is straightforward to verify that $W_k(Y^*, X) \leqslant W_k(Y, X) \leqslant \delta$ and $\mathrm{VaR}_\alpha^\Lambda(Y^*) \geqslant \mathrm{VaR}_\alpha^\Lambda(Y)$ with $\Lambda \in \{L, R\}$. Hence, an optimal distribution for the problem $\sup\{\mathrm{VaR}_\alpha^\Lambda(Y) : W_k(Y, X) \leqslant \delta\}$ satisfies (51), and thus the optimal value is the solution to (34).

To show (35), it suffices to see that by plugging $x^* = \mathrm{VaR}_\alpha^R(X) + \delta/\alpha^{1/k}$ in the left-hand side of (34), we can see

$$\int_0^\alpha (x^* - \mathrm{VaR}_u^R(X))_+^k \,\mathrm{d}u \leqslant \int_0^\alpha (x^* - \mathrm{VaR}_\alpha^R(X))_+^k \,\mathrm{d}u = \delta^k,$$

and hence $[\mathrm{VaR}_\alpha^R]_\delta^k(X) \geqslant x^*$. This completes the proof. □

References

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, **26**(7), 1487–1503.

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

Baes, M., Koch-Medina, P. and Munari, C. (2020). Existence, uniqueness, and stability of optimal payoffs of eligible assets. *Mathematical Finance*, **30**(1), 128–166.

Barrieu, P. and El Karoui, N. (2005). Inf-convolution of risk measures and optimal risk transfer. *Finance and stochastics*, **9**(2), 269-298.

Blanchet, J., He, F. and Murthy, K. (2020). On distributionally robust extreme value analysis. *Extremes*, **23**, 317–347.

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

Belles-Sampera, J., Guillén, M. and Santolino, M. (2014). Beyond value-at-risk: GlueVaR distortion risk measures. *Risk Analysis*, **34**(1), 121–134.

Carlier, G. and Dana, R.-A. (2003). Pareto efficient insurance contracts when the insurer's cost function is discontinuous. *Economic Theory*, **21**, 871–893.

Carlier, G., Dana, R.-A. and Galichon, A. (2012). Pareto efficiency for the concave order and multivariate comonotonicity. *Journal of Economic Theory*, **147**, 207–229.

Xie, W. (2021). On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, **186**(1), 115–155.

Chen, Z., Kuhn, D. and Wiesemann, W. (2018). Data-driven chance constrained programs over Wasserstein balls. *arXiv preprint arXiv:1809.00210.*

Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.

Delbaen, F. (2012): *Monetary Utility Functions.* Osaka University Press, Osaka.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., Tang, Q. and Vynche, D. (2006). Risk measures and comonotonicity: a review. *Stochastic Models*, **22**, 573–606.

Dhaene, J., Kukush, A., Linders, D. and Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, **2**(2), 319–328.

Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.

Embrechts, P., Liu, H., Mao, T. and Wang, R. (2020). Quantile-based risk sharing with heterogeneous beliefs. *Mathematical Programming*, **181**(2), 319–347.

EIOPA (2011). Equivalence assessment of the Swiss supervisory system in relation to articles 172, 227 and 260 of the Solvency II Directive, *EIOPA-BoS-11-028.*

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, **171**(1-2), 115–166.

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

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocation and heavy-tailed risks. *Journal of Banking and Finance*, **83**, 70–84.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**, 141–153.

Ghamami, S. and Glasserman, P. (2019). Submodular risk allocation. *Management Science*, **65**(10), 4451–4949.

Hamm, A. M., Knispel, T. and Weber, S. (2020). Optimal risk sharing in insurance networks. *European Actuarial Journal*, **10**(1), 203–234.

Ho-Nguyen, N., Kılınç-Karzan, F., Küçükyavuz, S. and Lee, D. (2021). Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity. *Mathematical Programming*, published online.

Huberman, G., Mayers, D., and Smith, C. W. (1983). Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, **14**(2), 415–288.

Liu, P., Wang, R. and Wei, L. (2020). Is the inf-convolution of law-invariant preferences law-invariant? *Insurance: Mathematics and Economics*, **91**, 144–154

Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.

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

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

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization* (Ed. Uryasev, S.), pp. 272–281, Springer, Dordrecht.

Pflug, G. C. and Römisch, W. (2007). *Modeling, Measuring and Managing Risk.* World Scientific, Singapore.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21–42.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios.* Springer, Heidelberg.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic orders.* Springer, New York.

Starr, R. M. (2011). *General Equilibrium Theory: An Introduction.* Second Edition, Cambridge University Press.

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. and Wei, Y. (2020). Characterizing optimal allocations in quantile-based risk sharing. *Insurance: Mathematics and Economics*, **93**, 288–300.

Wang, R., and Zitikis, R. (2020). Weak comonotonicity. *European Journal of Operational Research*, **282**, 386–397.

Weber, S. (2018). Solvency II, or how to sweep the downside risk under the carpet. *Insurance: Mathematics and Economics*, **82**, 191–200.

Xia, J. and Zhou, X. (2016). Arrow-Debreu equilibria for rank-dependent utilities. *Mathematical Finance*, **26**, 558–588.

Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155-1168.