Convolution Bounds on Quantile Aggregation

Jose Blanchet* Henry Lam† Yang Liu‡ Ruodu Wang§

Abstract

Quantile aggregation with dependence uncertainty has a long history in probability theory with wide applications in finance, risk management, statistics, and operations research. Using a recent result on inf-convolution of quantile-based risk measures, we establish new analytical bounds for quantile aggregation which we call convolution bounds. Convolution bounds both unify every analytical result available in quantile aggregation and enlighten our understanding of these methods. These bounds are the best available in general. Moreover, convolution bounds are easy to compute, and we show that they are sharp in many relevant cases. They also allow for interpretability on the extremal dependence structure. The results directly lead to bounds on the distribution of the sum of random variables with arbitrary dependence. We discuss relevant applications in risk management and economics.

Keywords: Range-Value-at-Risk, convolution, model uncertainty, dependence structure, duality

1 Introduction

The problem of quantile aggregation with dependence uncertainty refers to finding possible values of quantiles of an aggregate variable $S = X_1 + \cdots + X_n$ (often representing a total risk, but it can also represent the completion time of a task). The random variables $X_1, \dots, X_n$ have given marginal distributions, but unspecified dependence structure. More precisely, given marginal distributions $\mu_1, \dots, \mu_n$ on $\mathbb{R}$, the following quantities are of interest:

$$\sup\{q_t(X_1 + \cdots + X_n) : X_i \sim \mu_i,\ i = 1, \dots, n\} \tag{1}$$

and

$$\inf\{q_t(X_1 + \cdots + X_n) : X_i \sim \mu_i,\ i = 1, \dots, n\}, \tag{2}$$

where $q_t(X)$ stands for a (left or right) quantile of a random variable X at probability level $t \in [0, 1]$. The optimization problems (1) and (2) are, respectively, referred to as the worst-case and the best-case quantile aggregation. An equivalent problem is to find the maximum and the minimum values of $\mathbb{P}(S \leqslant x)$ for a given $x \in \mathbb{R}$. This problem has a long history in probability theory; see Makarov (1981) and Rüschendorf (1982) for

*Department of Management Science and Engineering, Stanford University, USA. Email: jose.blanchet@stanford.edu
†Department of Industrial Engineering and Operations Research, Columbia University, USA. Email: khl2114@columbia.edu
‡School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China. Email: yangliu16@cuhk.edu.cn
§Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: wang@uwaterloo.ca

early results. It has also been studied in combinatorial optimization with applications in statistical testing (e.g., Vovk and Wang (2020); Vovk et al. (2022)) and risk management (e.g., Embrechts et al. (2013, 2015)); see Section 2 and Appendix G for several applications. A key feature of quantile aggregation is the "arbitrary dependence" structure imposed. Naturally, this level of generality leads to robust estimates, although these can be conservative in some situations.

Because of the level of generality imposed both in marginal distributions and dependence, the quantile aggregation problems (1) and (2) rarely have analytical tractability. In the literature, some analytical bounds for the homogeneous setting (i.e., identical marginal distributions) are obtained by Embrechts and Puccetti (2006), Wang et al. (2013) and Puccetti and Rüschendorf (2013), and approximating algorithms are available such as the rearrangement algorithm (RA) in Puccetti and Rüschendorf (2012) and Embrechts et al. (2013). The sharpness of these bounds is rarely obtained with the exception of Wang et al. (2013) and Puccetti and Rüschendorf (2013) under some strong conditions. The RA only gives a lower bound on the quantile aggregation, and its convergence is not guaranteed. As a variant of optimal transport problem, discrete versions of problems (1) and (2) admit a linear programming reformulation, which involves exponentially many variables, and is computationally difficult for moderate dimensions (e.g., $n \geqslant 6$); details are explained in Appendix F. So, it is sensible to discuss bounds that can be shown to be sharp in continuous relaxations as the ones that we consider here.

In this paper, we propose a class of bounds on Range-Value-at-Risk (RVaR) based on the inf-convolution formulas introduced by Embrechts et al. (2018). We will call them *convolution bounds*. Since RVaR includes the two regulatory risk measures, Value-at-Risk (VaR) and the Expected Shortfall (ES, also known as CVaR), as special cases, the results on RVaR give rise to useful bounds on quantile aggregation problems (1) and (2).

As our main contributions, convolution bounds can provide by far the most convenient and sharpest theoretical results on quantile aggregation in a wide range of practical settings, and they can be applied to any marginal distributions, discrete, continuous, or mixed. As such, convolution bounds enjoy multifaceted advantages. They can be applied to both the quantile and RVaR aggregations (Theorems 1 and A.1); they combine different existing sharpness results of quantile aggregation and some new cases into a unified form (Theorems 2 and A.2); they lead to tractable extremal dependence structures for interpretation or approximation (Theorem 3), and they are computationally convenient and efficient. To the best of our knowledge, there is no other theoretical result on quantile aggregation which cannot be covered by our convolution bounds. Moreover, our results provide novel bounds on RVaR aggregation and establish sharpness for the dual bound (Theorem 4.17 of Rüschendorf (2013)). Although the sharpness of convolution bounds requires some conditions, their numerical performance suggests that they are generally very accurate even in cases where sharpness cannot be theoretically proved. As we mentioned above, our results on quantile aggregation can be directly applied to compute bounds on the distribution of the sum of random variables with arbitrary dependence. Our technical development builds on some results on risk sharing in Embrechts et al. (2018). Our target problem and theoretical contributions are very different from those on risk sharing, which aim to optimally allocate a fixed total risk (random variable X) to different agents (several random

variables that sum to X). Our objective, on the contrary, aims to solve the max/min values of the quantile of the sum random variable given known marginals (the total risk is not fixed). This problem is called the problem of robust risk aggregation in the literature; see Section 2.1.

We can relate and contrast our investigation to the recent fast-growing literature on distributionally robust optimization (DRO) (e.g., Goh and Sim (2010); Delage and Ye (2010); Wiesemann et al. (2014)) and chance constrained optimization (e.g., Nemirovski and Shapiro (2007), Chen et al. (2010) and Chapter 4 of Shapiro et al. (2021)). Viewed as a distributional analog to (deterministic) robust optimization (Bertsimas et al. (2011); Ben-Tal et al. (2009)), this literature tackles decision-making where the underlying parameter in a stochastic problem is uncertain. This leads to the optimization of decision under the worst-case scenario, where the worst case is over a region in which the uncertain parameter is believed to lie in, often known as the uncertainty set or ambiguity set. In DRO, the uncertain parameter, and hence the decision variable in the inner maximization, is the underlying probability distribution. Common constraints to characterize the belief on uncertain distributions include neighborhood balls formed by statistical distances such as the Wasserstein distance (Esfahani and Kuhn (2018); Gao and Kleywegt (2016); Blanchet and Murthy (2019)) and ϕ-divergence (Bertsimas et al. (2018); Ben-Tal et al. (2013)), moments and supports (Bertsimas and Popescu (2005)), geometric shape (Popescu (2005)), and marginal information (Doan and Natarajan (2012); Mishra et al. (2014)).[1] The RVaR and quantile aggregation considered in this paper can be regarded as an optimization over distributions having a marginal information constraint (i.e., the latest class listed above). When placed as a constraint, quantile or percentile criterion can be converted into a chance or probabilistic constraint (e.g., Delage and Mannor (2010)). The worst-case VaR under various settings of model uncertainty is also popular in robust portfolio optimization; see e.g., El Ghaoui et al. (2003), Zhu and Fukushima (2009) and Zymler et al. (2013). However, in contrast to the DRO literature which often focuses on solution methods via convex reformulations, here our problem is knowingly computationally intractable, and our goal is to obtain tractable analytical bounds that are provably tight in important cases.

The rest of the paper is organized as follows. Section 2 presents two motivating examples of robust risk management and the O-ring model in economics (Kremer (1993)), and Section 3 contains technical preliminaries. The (upper) convolution bounds on the quantile and RVaR aggregations are established in Sections 4-5. A general extremal dependence structure and some explicit approximations are presented in Section 6. The dual formulation of the quantile aggregation problems is studied in Section 7 (Theorem 4). The numerical advantages of the new bounds are carefully examined in Section 8. The two motivating examples are revisited in Section 9, where we apply our main results and discuss their implications. Section 10 concludes the paper. To better illustrate our main ideas, the lower convolution bounds and related discussions are postponed to Appendix A (in particular, Theorems A.1 and A.2). Appendices B-G include

[1] There is a large literature on DRO problems with various formulations, in addition to the few papers mentioned. We refer to Blanchet et al. (2019a,b) and the references therein for recent developments on DRO with Wasserstein distance, and to Hu and Hong (2013); Glasserman and Xu (2014); Jiang and Guan (2016); Lam (2016); Blanchet et al. (2020) for DRO problems with ϕ-divergence. DRO formulated by moments and supports are also studied by, e.g., Delage and Ye (2010); Goh and Sim (2010); Ghosh and Lam (2019). See also Van Parys et al. (2016); Lam and Mottet (2017); Li et al. (2019) for various settings of DRO with geometric shape.

all proofs, the counter-examples, technical discussions and other operations research applications.

2 Motivating examples

In this section, we list two examples where the quantile aggregation problems (1) and (2) become natural in various contexts relevant to modern operations research. We will revisit these examples with our theoretical results and numerical illustrations in Section 9.

2.1 Robust risk management

The worst-case value of a risk measure ρ evaluating an aggregate risk is extensively studied in the risk management literature, known as the problem of robust risk aggregation. It is motivated by the context in which data from different correlated products are separately collected and hence their dependence information is not available; see Embrechts et al. (2013, 2015) and the references therein. As a specific example,[2] the European Union has established a solidarity fund since 2002 to help member states in case of some catastrophic events. While the loss curves are well estimated in each country, the fund has to pay for the sum of all losses. An independence assumption cannot be justified for climate-related events, which may affect several countries. Since an estimate of the copula is not available, the worst-case analysis in this section provides some bounds on the total losses.

Suppose that there are n random losses $X_1, \dots, X_n$ with known marginal distributions $\mu_1, \dots, \mu_n$ and unknown dependence structure. To calculate the regulatory margin conservatively, one relies on the worst-case aggregate risk, that is,

$$\sup\{\rho(X_1 + \cdots + X_n) : X_i \sim \mu_i,\ i = 1, \dots, n\}.$$

If ρ is a convex risk measure such as an ES, then its worst-case value is easy to compute due to convexity; see Rüschendorf (2013). In case ρ is the VaR at level t, this quantity is (1), which is highly non-trivial because of non-convexity of the quantile. Due to the connection of quantiles to risk measures like VaR, quantile aggregation is a popular problem in risk management (Section 8.4 of McNeil et al. (2015)), and many useful technical results were developed in this literature, e.g., Embrechts and Puccetti (2006), Wang et al. (2013), Embrechts et al. (2015) and Jakobsons et al. (2016). Our main results directly address this problem for the cases that ρ is a VaR or RVaR; some numerical illustrations are presented in Section 8.2.

Next, we bring the worst-case risk calculation to the context of portfolio selection. The traditional problem of VaR-based portfolio selection (e.g., Basak and Shapiro (2001)) is formulated as

$$\text{maximize } \mathbb{E}[u(\boldsymbol{\lambda} \cdot \mathbf{X})] \text{ over } \boldsymbol{\lambda} \in \overline{\Delta}_{n-1}, \quad \text{subject to } q_t(\boldsymbol{\lambda} \cdot (-\mathbf{X})) \leqslant x,$$

[2] We thank an anonymous referee for providing the context in this example.

where $\boldsymbol{\lambda}$ represents a portfolio weight vector, $\mathbf{X}$ represents future asset values, x is a constant risk limit, $u : \mathbb{R} \to \mathbb{R}$ is a strictly concave and increasing utility function, $t \in (0,1)$ is close to 1, and $\overline{\Delta}_{n-1}$ is the standard n-simplex, that is,

$$\overline{\Delta}_{n-1} = \left\{ (\lambda_1, \dots, \lambda_n) \in [0,1]^n : \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

Note that the quantile constraint can be equivalently formulated as a chance (exceedance probability) constraint, popular in the literature of stochastic programming. This formulation requires a full specification on the joint distribution of all the assets $(X_1, \dots, X_n)$ which can be difficult to obtain. In the presence of dependence uncertainty, we consider the following robust optimization problem, for a given tuple $\boldsymbol{\mu}$ of marginal distributions,

$$\text{maximize } \inf_{\mathbf{X} \sim \boldsymbol{\mu}} \mathbb{E}[u(\boldsymbol{\lambda} \cdot \mathbf{X})] \text{ over } \boldsymbol{\lambda} \in \overline{\Delta}_{n-1}, \quad \text{subject to } \sup_{\mathbf{X} \sim \boldsymbol{\mu}} q_t(\boldsymbol{\lambda} \cdot (-\mathbf{X})) \leqslant x, \tag{3}$$

where $\mathbf{X} \sim \boldsymbol{\mu}$ represents the marginal conditions $X_i \sim \mu_i$, $i = 1, \dots, n$. The problem (3) has robustness implications. Assume that the marginal distribution is well-specified and we solve (3). Denote the optimal solution by $\boldsymbol{\lambda}^*$ and the optimal value by v^*. They satisfy the guarantee that $q_t(\boldsymbol{\lambda}^* \cdot (-\mathbf{X}_0)) \leqslant x$ and $\mathbb{E}[u(\boldsymbol{\lambda}^* \cdot \mathbf{X}_0)] \geqslant v^*$, where $\mathbf{X}_0$ follows the unknown true distribution. In other words, we guarantee that the quantile constraint under the true distribution is satisfied, while the attained objective value under the true distribution has at least a performance level v^*.

The problem (3) is challenging due to the non-convexity of VaR. Portfolio optimization with dependence uncertainty has been studied by, e.g., Pflug and Pohl (2018), but there are no results on the case of VaR. Our results on quantile aggregation can be applied to address this problem. As our analysis in Section 9 shows, although this worst-case approach is generally conservative, the obtained optimal strategies are quite intuitive.

2.2 The O-ring model

The O-ring theory of economic development was proposed by Kremer (1993); see also the recent work of Boerma et al. (2021) and the references therein. The O-ring model can be formulated in a stochastic context. Assume that there are continuums of firms and n types of workers. Each firm requires n workers, one in each type, to format a team in production. Let ω be a firm in the continuum. The product value of the firm ω is denoted by $Z(\omega) \in (0, \infty)$. For a type-$i$ worker matched with the firm ω, the probability to successfully complete his/her task is denoted by $X_i(\omega) \in (0,1)$. A high-skilled worker has a higher value of X_i. Among all firms, the value Z has a distribution μ_Z. Among all workers of type i, the value X_i has a distribution μ_i. The product of a firm is considered successful if all n workers in the firm complete their individual tasks (this explains the name of the O-ring model). It is customary as in Kremer (1993) to assume that n individual events, in which the i-th worker completes his/her task, $i = 1, \dots, n$, are independent for a

fixed firm. Hence, the production function of the firm ω is the product value times the probability of success, that is,

$$y(X_1(\omega),\dots,X_n(\omega),Z(\omega)) = Z(\omega)\cdot\prod_{i=1}^{n} X_i(\omega). \tag{4}$$

A classic problem is to seek a global matching between multiple heterogeneous workers into teams at heterogeneous firms in order to maximize $\mathbb{E}[y(X_1,\dots,X_n,Z)]$ among all kinds of dependence structures with the given marginal distributions. The solution of the optimal sorting is positively dependent; more precisely, $Z, X_1,\dots,X_n$ are comonotonic. The interpretation is that the good workers (X_i all have a higher value) should work together in a good firm (Z also has a higher value). This partially explains the assignment of global economic industries between the developed and developing countries as argued by Kremer (1993).

As argued by Boerma et al. (2021, 2023), labour matching observed in the labour market does not show the comonotonic pattern as implied by the classic O-ring theory. Below, we explain that a quantile aggregation problem leads to a richer matching pattern which can be solved using the results in this paper.[3] There is a recently increasing interest in quantiles as decision criteria in economics; see Rostek (2010) and de Castro and Galvao (2019) for theoretical advances and de Castro et al. (2022) for experimental analysis.

Instead of optimizing the expected production in (4) across firms, one may be concerned about how many productions have low values below a certain threshold $y_0 > 0$, e.g., a level that is unacceptable by the society. That is, one investigates the deficiency proportion minimization problem

$$\min\left\{\mathbb{P}(y(X_1,\dots,X_n,Z)\leqslant y_0) : Z\sim\mu_Z,\ X_i\sim\mu_i,\ i=1,\dots,n\right\}, \tag{5}$$

where the probability $\mathbb{P}$ measures the proportion of productions that falls below the deficiency threshold. Since the problems of quantile aggregation and probability bounds translate to each other, for (5) it suffices to solve the problem of quantile aggregation on $\log(Z)+\sum_{i=1}^{n}\log(X_i)$. The extremal dependence structure attaining (5) illustrates the optimal matching pattern, which is the topic of Section 6. Section 9 contains a detailed illustration. Our results can also be applied to the model of Boerma et al. (2021), where the product is considered successful if at least one worker, instead of all, is able to complete the task.

3 Notation and preliminaries

Let $\mathcal{M}$ be the set of (Borel) probability measures on $\mathbb{R}$ and $\mathcal{M}_1$ be the set of probability measures on $\mathbb{R}$ with finite mean. For $\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)\in\mathcal{M}^n$, let $\Gamma(\boldsymbol{\mu})$ be the set of probability measures on $\mathbb{R}^n$ that have one-dimensional marginals $\mu_1,\dots,\mu_n$. For a probability measure μ on $\mathbb{R}^n$, define $\lambda_\mu\in\mathcal{M}$ by

$$\lambda_\mu(-\infty,x] = \mu(\{(x_1,\dots,x_n)\in\mathbb{R}^n : x_1+\cdots+x_n\leqslant x\}),\ x\in\mathbb{R}.$$

[3] It is not our intention to say that the real labour market follows such a model; this issue would require a separate study. Our model provides a way to generate rich matching patterns. This is also the approach taken by Boerma et al. (2021, 2023) for different settings.

In other words, λ_μ is the distribution measure of $\sum_{i=1}^n X_i$ where the random vector $(X_1, \dots, X_n)$ follows μ. Moreover, let $\Lambda(\boldsymbol{\mu}) = \{\lambda_\mu : \mu \in \Gamma(\boldsymbol{\mu})\}$. Thus, $\Lambda(\boldsymbol{\mu})$ is the set of the aggregate distribution measures with specified marginals $\boldsymbol{\mu}$. For $t \in (0,1]$, define the left quantile functional

$$q_t^-(\mu) = \inf\{x \in \mathbb{R} : \mu(-\infty, x] \geqslant t\}, \quad \mu \in \mathcal{M},$$

and for $t \in [0,1)$, define the right quantile functional

$$q_t^+(\mu) = \inf\{x \in \mathbb{R} : \mu(-\infty, x] > t\}, \quad \mu \in \mathcal{M}.$$

The two extreme cases q_0^+ and q_1^- correspond to the essential infimum and the essential supremum. Note that $q_t^\pm$ is defined on $\mathcal{M}$ instead of on the set of random variables as in the introduction. The most important objects in this paper are the average quantile functionals which we define next. For $0 \leqslant \beta < \beta + \alpha \leqslant 1$, define

$$R_{\beta,\alpha}(\mu) = \frac{1}{\alpha} \int_\beta^{\beta+\alpha} q_{1-t}^+(\mu) \mathrm{d}t, \quad \mu \in \mathcal{M}. \tag{6}$$

By definition, $R_{\beta,\alpha}(\mu)$ is the average of the quantile[4] of μ over $[1-\beta-\alpha, 1-\beta]$. The functional $R_{\beta,\alpha}$, introduced originally by Cont et al. (2010), is called an RVaR by Wang et al. (2015). The value $R_{\alpha,\beta}(\mu)$ in (6) is always finite for $\beta > 0$ and $\alpha + \beta < 1$, and it may take the value ∞ or $-\infty$ in case one of $\beta = 0$ or $\alpha + \beta = 1$. For the special case in which $\beta = 0$ and $\alpha = 1$, $R_{0,1}$ is precisely the mean, and it is only well defined on the set $\mathcal{M}_1$ of distributions with finite mean. The left and right quantiles can be obtained as limiting cases of $R_{\beta,\alpha}$ for $\beta \in (0,1)$ via

$$\lim_{\alpha \downarrow 0} R_{\beta,\alpha}(\mu) = q_{1-\beta}^-(\mu) \quad \text{and} \quad \lim_{\alpha \downarrow 0} R_{\beta-\alpha,\alpha}(\mu) = q_{1-\beta}^+(\mu), \quad \mu \in \mathcal{M}. \tag{7}$$

Two other useful special cases are ES and the left-tail ES (LES), defined, respectively, at level $\alpha \in (0,1)$ via

$$\mathrm{ES}_\alpha(\mu) = R_{0,\alpha}(\mu) = \frac{1}{\alpha} \int_{1-\alpha}^1 q_u^+(\mu) \mathrm{d}u, \quad \mu \in \mathcal{M},$$

and

$$\mathrm{LES}_\alpha(\mu) = R_{1-\alpha,\alpha}(\mu) = \frac{1}{\alpha} \int_0^\alpha q_u^+(\mu) \mathrm{d}u, \quad \mu \in \mathcal{M}.$$

As explained by Embrechts et al. (2018), the RVaR functional R bridges the gap between quantiles (VaR) and ES, the two most popular risk measures in banking and insurance.

It is sometimes convenient to slightly abuse the notation by using $R_{\beta,\alpha}(X)$ or $q_t(X)$ for $R_{\beta,\alpha}(\mu)$ or $q_t(\mu)$ where $X \sim \mu$. All random variables appearing in the paper live in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$.[5] We use $\bigvee_{i=1}^n \alpha_i$ for the maximum of real numbers $\alpha_1, \dots, \alpha_n$.

[4] We can use either q^+ or q^- in the integral, as the two quantities are the same almost everywhere on $[0,1]$.

[5] A probability space is atomless if there exists a continuously distributed random variable on this space.

4 Convolution bounds on RVaR aggregation

Our starting point is that an upper bound on RVaR aggregation, which we shall refer to as convolution bounds, can be obtained from an inequality on RVaR from Embrechts et al. (2018). More precisely, Theorem 2 of Embrechts et al. (2018) gives the following inf-convolution formula, for any integrable random variable X and $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \in [0,1]$ with $\beta + \alpha \leqslant 1$ where $\beta = \sum_{i=1}^n \beta_i$ and $\alpha = \bigvee_{i=1}^n \alpha_i$,

$$R_{\beta,\alpha}(X) = \inf\left\{ \sum_{i=1}^n R_{\beta_i,\alpha_i}(X_i) : X_1 + \cdots + X_n = X \right\}, \tag{8}$$

where the infimum is taken over all random variables $X_1, \dots, X_n$. As a consequence of (8), we have an RVaR aggregation inequality

$$R_{\beta,\alpha}\left(\sum_{i=1}^n X_i \right) \leqslant \sum_{i=1}^n R_{\beta_i,\alpha_i}(X_i) \tag{9}$$

for all $X_1, \dots, X_n$, provided the right-hand side of (9) is well defined (not "$\infty - \infty$").[6] The objective of Embrechts et al. (2018) is the risk sharing problem where the aggregate risk X and the preferences of the agents are known (thus, $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ are given) and one optimizes $\sum_{i=1}^n R_{\beta_i,\alpha_i}(X_i)$ over possible allocations $X_1, \dots, X_n$ satisfying $X_1 + \cdots + X_n = X$.

In this paper, we use the reverse direction of (9): we fix $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$ and t, s with $0 \leqslant t < t + s \leqslant 1$, and aim to find the worst-case value of the aggregate risk $R_{t,s}(\nu)$ over $\nu \in \Lambda(\boldsymbol{\mu})$ using (9). For any $0 \leqslant t < t + s \leqslant 1$, $\beta_0 \in [s, t+s]$, $\nu \in \Lambda(\boldsymbol{\mu})$, noting that $R_{t,s} \leqslant R_{t+s-\beta_0,\beta_0}$, (9) leads to

$$R_{t,s}(\nu) \leqslant R_{\sum_{i=1}^n \beta_i, \beta_0}(\nu) \leqslant \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i), \tag{10}$$

where $\sum_{i=1}^n \beta_i = t + s - \beta_0$. Taking a supremum among all $\nu \in \Lambda(\boldsymbol{\mu})$ and an infimum among all feasible $(\beta_0, \beta_1, \dots, \beta_n)$ in (10), we get, for any fixed (t, s) with $0 \leqslant t < t + s \leqslant 1$,

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) \leqslant \inf_{\substack{\sum_{i=0}^n \beta_i = t+s \\ \beta_0 \geqslant s > 0}} \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i). \tag{11}$$

The right-hand side of (11) depends only on the marginal distributions $\mu_1, \dots, \mu_n$ and (t, s), and thus we obtain a novel upper bound on the worst-case RVaR aggregation. We shall refer to the bound in (11) as a *convolution bound*, since it is obtained from the inf-convolution formula in (8). To simplify notation, for each $n \in \mathbb{N}$, let

$$\Delta_n = \left\{ (\beta_0, \beta_1, \dots, \beta_n) \in (0,1) \times [0,1)^n : \sum_{i=0}^n \beta_i = 1 \right\},$$

which is the set of vectors in the standard $(n+1)$-simplex with positive first component. In all results, $\boldsymbol{\beta}$

[6] The inequality in (9) is essentially Theorem 1 of Embrechts et al. (2018), which requires a condition on integrability. We slightly generalize this result to probability measures without finite means, which will be useful for the generality of results offered in this paper; see Lemma EC.1 in the appendix. Also note that our parameterization is slightly different from Embrechts et al. (2018).

represents $(\beta_0, \beta_1, \dots, \beta_n)$.

We formally present the convolution bound in Theorem 1 below. More importantly, we show that this bound is indeed sharp under a few sets of conditions, and hence the convolution bounds are useful in calculating worst-case values in risk aggregation problems. As far as we are aware of, Theorem 1 is the only result in the literature on RVaR aggregation with given marginal distributions. The practically relevant case of quantiles ($s \downarrow 0$) will be discussed in detail in Section 5.

Throughout, by "admitting a decreasing density" we mean that the distribution has a left-bounded support and it has a decreasing probability density function with respect to the Lebesgue measure on its support. The case for "admitting an increasing density" is analogous.

Theorem 1. *Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$. For any t, s with $0 \leqslant t < t + s \leqslant 1$,*

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) \leqslant \inf_{\substack{\boldsymbol{\beta} \in (t+s)\Delta_n \\ \beta_0 \geqslant s > 0}} \sum_{i=1}^{n} R_{\beta_i, \beta_0}(\mu_i). \tag{12}$$

Moreover, (12) holds as an equality in the following cases:

(i) $t = 0$;

(ii) each of $\mu_1, \dots, \mu_n$ admits a decreasing density beyond its $(1 - t - s)$-quantile;

(iii) $\sum_{i=1}^{n} \mu_i \left(q^+_{1-t-s}(\mu_i), q^-_1(\mu_i)\right] \leqslant t + s$.

In Theorem 1, case (i) corresponds to the aggregation of ES, which is well known in the literature, e.g., Chapter 8 of McNeil et al. (2015). Case (ii) in Theorem 1 is the most useful as decreasing densities are common in many areas of applications, including but not limited to finance and insurance. The proof of this case is quite technical, and it relies on advanced results on robust risk aggregation established in Wang and Wang (2016) and Jakobsons et al. (2016). Case (iii) corresponds to an assumption which allows for a lower mutually exclusive (see Puccetti and Wang (2015) and also Definition EC.1 in Appendix B.1) random vector following marginal distributions $\mu_1, \dots, \mu_n$. Such a situation is not common, but it may happen in the context of credit portfolio analysis, where each μ_i represents the distribution of loss from a defaultable security which has a small probability of being positive. For instance, take $t = s = 0.05$, $n = 50$ and let μ_i be Bernoulli distributions with $\mu_i(\{1\}) = 0.001$ for $i = 1, \dots, n$. In this example, the aggregate risk represents the loss from a portfolio of defaultable bonds with default probability 0.001, and the condition in case (iii) is satisfied because $\sum_{i=1}^{n} \mu_i \left(q^+_{1-t-s}(\mu_i), q^-_1(\mu_i)\right] = \sum_{i=1}^{n} \mu_i(\{1\}) = 0.05 \leqslant t + s$. The proof for case (iii) is based on convenient properties of a mutually exclusive random vector. Moreover, we will show in Figure 2 (right panel) in Section 8 that the bound (12) is not sharp for marginals with increasing densities, even for homogeneous marginals; however for quantiles (limits of RVaR), the bound becomes sharp for increasing densities (Theorem 2).

Results that are symmetric to the upper convolution bounds are collected in Appendix A. For instance, a lower bound on $\inf_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu)$, which is symmetric to Theorem 1, is given in Theorem A.1.

Case (ii) in Theorem 1 involves conditional distributions above a certain quantile. For $\mu \in \mathcal{M}$ and $t \in [0,1)$, let μ^{t+} be the probability measure given by

$$\mu^{t+}(-\infty, x] = \max\left\{\frac{\mu(-\infty, x] - t}{1-t}, 0\right\}, \quad x \in \mathbb{R}.$$

The probability measure μ^{t+} is called the t-tail distribution of μ by Rockafellar and Uryasev (2002). In other words, μ^{t+} is the distribution measure of the random variable $q_U(\mu)$ where U is a uniform random variable on $[t,1]$. Equivalently, μ^{t+} is the distribution measure of μ restricted beyond its t-quantile. For example, the statement in case (ii) that μ admits a decreasing density beyond its $(1-t-s)$-quantile is equivalent to the one that $\mu^{(1-t-s)+}$ admits a decreasing density. Moreover, by direct computation, for fixed $\mu \in \mathcal{M}$ and $t \in [0,1)$, we have

$$\begin{aligned} R_{\beta,\alpha}(\mu^{t+}) &= R_{(1-t)\beta,(1-t)\alpha}(\mu), \quad \text{for all } 0 \leqslant \beta < \beta + \alpha \leqslant 1; \\ q_u^-(\mu^{t+}) &= q^-_{t+(1-t)u}(\mu), \quad \text{for all } u \in (0,1]. \end{aligned} \tag{13}$$

Using (13), we obtain Proposition 1 below based on Theorem 4.1 of Liu and Wang (2021). This result is useful in the proof of Theorem 1. For $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$ and $t \in [0,1)$, denote by $\boldsymbol{\mu}^{t+} = (\mu_1^{t+}, \dots, \mu_n^{t+})$.

Proposition 1. *For $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$, $t \in [0,1)$ and $s \in (0, 1-t]$, we have*

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{(1-t-s)+})} \mathrm{LES}_{\frac{s}{t+s}}(\nu)$$

and

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{t+})} q_0^+(\nu).$$

Proposition 1 suggests that for the worst-case problems of RVaR aggregation, it suffices to consider the one started from quantile level 0, i.e. the LES aggregation. In particular, for the worst-case problems of quantile aggregation, it suffices to consider the one at quantile level 0, i.e. the problems $\sup_{\nu \in \Lambda(\boldsymbol{\mu}^{t+})} q_0^+(\nu)$ for generic choices of $\boldsymbol{\mu}$. This result will be used repeatedly in our discussions, and it will be the general approach taken in the proof of our main results.

5 Convolution bounds on quantile aggregation

5.1 Convolution bounds

In Theorem 2 below we summarize bounds on $\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu)$. Most cases can be obtained by sending s to 0 and replacing t with $(1-t)$ in Theorem 1, but a notable difference is that the convolution bounds are sharp for both decreasing and increasing densities, for $n = 2$, and for two types of mutual exclusivity (see Appendix B.1). This is in drastic contrast to the RVaR convolution bounds which are only sharp for decreasing densities or upper mutual exclusivity (see Figure 2). Results on lower bounds on $q_t^-(\nu)$ are put in Appendix A. In particular, Theorem A.2 is symmetric to Theorem 2.

Theorem 2. *For $\boldsymbol{\mu} \in \mathcal{M}^n$ and $t \in [0,1)$, we have*

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu) \leqslant \inf_{\boldsymbol{\beta} \in (1-t)\Delta_n} \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i). \tag{14}$$

Moreover, (14) holds as an equality in the following cases:

(i) $n \leqslant 2$;

(ii) each of $\mu_1, \dots, \mu_n$ admits a decreasing density beyond its t-quantile;

(iii) each of $\mu_1, \dots, \mu_n$ admits an increasing density beyond its t-quantile;

(iv) $\sum_{i=1}^n \mu_i \left(q_t^+(\mu_i), q_1^-(\mu_i)\right] \leqslant 1-t$;

(v) $\sum_{i=1}^n \mu_i \left[q_t^+(\mu_i), q_1^-(\mu_i)\right) \leqslant 1-t$.

Remark 1. If $\mu_1, \dots, \mu_n$ have positive densities on their supports, then $\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^-(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu)$ for all $t \in (0,1)$; see Lemma 4.5 of Bernard et al. (2014). Hence, using $q_t^-(\nu)$ or $q_t^+(\nu)$ in Theorem 2 is not essential to our discussions.

Remark 2. The classic probability bound $\mathbb{P}(\sum_{i=1}^n X_i \geqslant \sum_{i=1}^n z_i) \leqslant \sum_{i=1}^n \mathbb{P}(X_i \geqslant z_i)$ for all $z_1, \dots, z_n \in \mathbb{R}$, is a special case of Theorem 2 by converting quantile bounds into probability bounds. To see this, let μ_i be the distribution of X_i and $t_i = \mathbb{P}(X_i \geqslant z_i)$ for $i \in \{1, \dots, n\}$, and let ν be the distribution of $\sum_{i=1}^n X_i$. The bound (14) gives $q^+_{1-\sum_{i=1}^n t_i}(\nu) \leqslant \sum_{i=1}^n q^-_{1-t_i}(\mu_i) \leqslant \sum_{i=1}^n z_i$. This implies $\mathbb{P}(\sum_{i=1}^n X_i \geqslant \sum_{i=1}^n z_i) \leqslant \sum_{i=1}^n t_i$.

In the literature, some sharp bounds on quantile aggregation for decreasing densities are obtained by Wang et al. (2013) and Puccetti and Rüschendorf (2013) in the homogeneous case ($\mu_1 = \cdots = \mu_n$) and Jakobsons et al. (2016) in the heterogeneous case. For the heterogeneous case, the method of Jakobsons et al. (2016) involves solving a system of $(n+1)$-dimensional implicit ODE (equations (E1) and (E2) of Jakobsons et al. (2016)), which requires a highly complicated calculation. In contrast, our result in Theorem 2 gives sharp bounds based on the minimum or maximum of an $(n+1)$-dimensional function.

In the homogeneous case $\mu_1 = \cdots = \mu_n$, as an immediate consequence of Theorem 2, we obtain the following reduced bounds in which one replaces $\inf_{\boldsymbol{\beta} \in (1-t)\Delta_n} \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i)$ by a one-dimensional optimization problem. We show that, in some homogeneous case, the sharp result in Theorem 2 can be achieved by the reduced bound. A proof of this result follows from a combination of Theorem 2 and Proposition 1 of Embrechts et al. (2014). In what follows, $\Lambda_n(\mu) = \Lambda(\mu, \dots, \mu)$ is the set of the aggregate distribution measures with the homogeneous marginal μ.

Proposition 2 (Reduced convolution bounds)**.** *For $\mu \in \mathcal{M}$ and $t \in [0,1)$, we have*

$$\sup_{\nu \in \Lambda_n(\mu)} q_t^+(\nu) \leqslant \inf_{\alpha \in (0,(1-t)/n)} nR_{\alpha,1-t-n\alpha}(\mu) = \inf_{\alpha \in (0,(1-t)/n)} \frac{n}{1-t-n\alpha} \int_{t+(n-1)\alpha}^{1-\alpha} q_u^-(\mu) \mathrm{d}u. \tag{15}$$

Moreover, (15) holds as an equality if μ admits a decreasing density beyond its t-quantile.

First, it is clear that the convolution bound (14) is better (smaller) than the reduced one (15), while the latter is easier to compute. They are not generally equal. Second, in case μ admits a decreasing density, Proposition 8.32 of McNeil et al. (2015) (reformulated from Wang et al. (2013, Theorem 3.4)) gives

$$\sup_{\nu\in\Lambda_n(\mu)} q_t^+(\nu) = \frac{n}{1-t-n\alpha}\int_{t+(n-1)\alpha}^{1-\alpha} q_u^-(\mu)\mathrm{d}u$$

for some $\alpha \in [0,(1-t)/n)$. Together with (14), we get the sharpness of (15).

Since quantiles commute with strictly increasing transforms, Theorem 2 leads to a multiplicative version of the convolution bounds, which can be useful for some applications, in particular, the O-ring theory in Section 2.2 and Section 9. Recall that for any random variable X following distribution μ and any Borel function f, the random variable $f(X)$ has distribution $\mu\circ f^{-1}$ where f^{-1} is the set-valued inverse of f.

Proposition 3. *For $\mu_1,\dots,\mu_n\in\mathcal{M}$ with support included in $(0,\infty)$, we have*

$$\sup_{X_i\sim\mu_i,i=1,\dots,n} q_t^+\left(\prod_{i=1}^n X_i\right) \leqslant \exp\left(\inf_{\boldsymbol{\beta}\in(1-t)\Delta_n}\sum_{i=1}^n R_{\beta_i,\beta_0}\left(\mu_i\circ\exp\right)\right), \quad t\in[0,1). \tag{16}$$

Moreover, (16) holds as an equality in the following cases (denote by $f_1,\dots,f_n$ the densities of $\mu_1,\dots,\mu_n$):

(i) $n\leqslant 2$;

(ii) *for each $i=1,\dots,n$, $x\mapsto xf_i(x)$ is decreasing beyond the t-quantile of μ_i;*

(iii) *for each $i=1,\dots,n$, $x\mapsto xf_i(x)$ is increasing beyond the t-quantile of μ_i;*

(iv) $\sum_{i=1}^n \mu_i\left(q_t^+(\mu_i),q_1^-(\mu_i)\right]\leqslant 1-t$;

(v) $\sum_{i=1}^n \mu_i\left[q_t^+(\mu_i),q_1^-(\mu_i)\right)\leqslant 1-t$;

5.2 Technical discussions

We do not expect that the formula (14) always gives sharp bounds, and this is a situation similar to Theorem 1. A counter-example of non-sharpness of the bounds in Theorem 2 is presented in Section 8.2 with some discrete marginal distributions (see also Example EC.1 in Appendix C). Nevertheless, in most cases, the bounds in Theorem 2 work quite well, as illustrated by the numerical examples later. In some special cases, the reduced bounds in Proposition 2 are equivalent to those in Theorem 2. We shall show this does not generally hold (e.g., for some distribution with increasing density) later in Figure 3 (right panel).

In the following proposition, we note that $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_t^+(\nu)$ is always attainable as a maximum, which is implied by Lemma 4.2 of Bernard et al. (2014).

Proposition 4. *For $\boldsymbol{\mu}\in\mathcal{M}^n$ and $t\in[0,1)$, there exists $\nu_+\in\Lambda(\boldsymbol{\mu})$ such that*

$$\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_t^+(\nu) = q_t^+(\nu_+).$$

We next turn to the right-hand side of (14). Because of the continuity of $R_{\alpha,\beta}$ in $\alpha, \beta \in [0,1]$, the infimum in $\inf_{\boldsymbol{\beta} \in (1-t)\Delta_n} \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i)$ for any $t \in [0,1)$ is attainable in the closure $\overline{\Delta}_n$ of Δ_n; see Appendix D for details.

To address computational efficiency, we first focus on the case of monotone densities which are sufficient for (ii) and (iii) in Theorem 2 for any t. These two assumptions will be used repeatedly later.

(DD) each of $\mu_1, \dots, \mu_n$ admits a decreasing density;

(ID) each of $\mu_1, \dots, \mu_n$ admits an increasing density.

Under condition (DD) or (ID), we can formally argue that the convolution bound is easy to compute. For an illustration, consider the infimum problem in (14) with the condition (DD); here we take $t = 0$ without loss of generality due to Proposition 1. For a fixed $\beta_0 \in (0,1)$, note that the mapping

$$\phi_i : \beta_i \mapsto \frac{1}{\beta_0} \int_{1-\beta_i-\beta_0}^{1-\beta_i} q_u^-(\mu_i) \mathrm{d}u$$

is convex, because $u \mapsto q_u^-(\mu_i)$ is convex under (DD) which implies that $\beta_i \mapsto q_{1-\beta_i-\beta_0}^-(\mu_i) - q_{1-\beta_i}^-(\mu_i)$ is increasing. Therefore, for fixed β_0,

$$(\beta_1, \dots, \beta_n) \mapsto \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i) = \sum_{i=1}^n \frac{1}{\beta_0} \int_{1-\beta_i-\beta_0}^{1-\beta_i} q_u^-(\mu_i) \mathrm{d}u$$

is convex since it is the sum of convex functions in each component. The full optimization can be converted to an n-dimensional convex minimization problem over $(\beta_1, \dots, \beta_n)$ and a one-dimensional problem of optimization over β_0, which is not necessarily convex. The objective is continuous in β_0, so that the one-dimensional problem is computable by suitable discrete approximation up to any specified accuracy. In case (ID) holds, the objective is concave in $(\beta_1, \dots, \beta_n)$, and its solution always lies on the boundary of the simplex $(1-\beta_0)\overline{\Delta}_{n-1}$. When (DD) and (ID) do not hold, the above optimization may be more complicated, but in our numerical experiments in Section 8, they are always solved quite fast and produce results that are consistent with other methods. The convolution bound is also compared with a discrete linear programming formulation in Appendix F, showing its advantages in computational time and feasibility in high dimensions.

The next proposition concerns the truncation of the marginal distributions. When calculating the supremum of q_0^+ for the aggregation of non-negative risks, one can safely truncate the marginal distributions at a high threshold. This result is convenient when applying several results in the literature formulated for distributions with finite mean or a compact support, including Theorem 1 of Embrechts et al. (2018). For a probability measure $\mu \in \mathcal{M}$ and a constant $m \in \mathbb{R}$, let $\mu^{[m]}$ be the distribution of $X \wedge m$ where $X \sim \mu$ and $x \wedge y$ stands for the minimum of two numbers x and y. Further denote that $\boldsymbol{\mu}^{[m]} = (\mu_1^{[m]}, \dots, \mu_n^{[m]})$ for $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$.

Proposition 5. *For any distributions $\mu_1, \dots, \mu_n$ on $[0,\infty]$, $t \in [0,1)$, and $m \geqslant \sum_{i=1}^n q^+_{1-(1-t)/n}(\mu_i)$, we have*

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{[m]})} q_t^+(\nu). \tag{17}$$

5.3 Quantile aggregation at levels 0 and 1

Now we restate the specific cases of quantile aggregation q_0^+ and q_1^-, where an analogous result to Theorem 2 is used; see Appendix A.

Proposition 6 (Convolution bounds at levels 0 and 1)**.** *For $\boldsymbol{\mu} \in \mathcal{M}^n$, we have*

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) \leqslant \inf_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^n R_{\beta_i, \beta_0}(\mu_i), \tag{18}$$

and

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_1^-(\nu) \geqslant \sup_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0, \beta_0}(\mu_i). \tag{19}$$

The two bounds are both sharp if $n \leqslant 2$, or each of $\mu_1, \dots, \mu_n$ admits a decreasing (respectively, increasing) density on its support.

If $\mu_1, \dots, \mu_n$ have finite means, the inequalities in (18) and (19) can be combined into a chain of inequalities.

Proposition 7. *For $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}_1^n$, we have*

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_1^-(\nu) \geqslant \sup_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0, \beta_0}(\mu_i) \geqslant \sum_{i=1}^n R_{0,1}(\mu_i) \geqslant \inf_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^n R_{\beta_i, \beta_0}(\mu_i) \geqslant \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu). \tag{20}$$

The tuple of distributions $\boldsymbol{\mu} \in \mathcal{M}^n$ is said to be *jointly mixable* (JM) if $\delta_C \in \Lambda(\boldsymbol{\mu})$ for some $C \in \mathbb{R}$; see Appendix E. Proposition 7 implies that (18) and (19) become sharp if $\boldsymbol{\mu} \in \mathcal{M}_1^n$ is JM. If $\mu_1, \dots, \mu_n$ do not have finite means, the relationships in (20) may not hold generally, which is illustrated by Example EC.3 in Appendix C.

6 Approximation of the extremal dependence

A significant advantage of the convolution bounds on the quantile aggregation problem is that we are able to visualize, in certain cases, the extremal dependence structure corresponding to the convolution bounds. In view of Proposition 1, for the problems of worst-case quantile aggregation, it suffices to consider the one at quantile level 0, i.e., $\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu)$. Similarly, for the problems of the best-case quantile aggregation, it suffices to consider the one at quantile level 1, i.e., $\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_1^-(\nu)$ as in Proposition A.1. The supremum and the infimum can be replaced by a maximum and a minimum, respectively, as implied by Proposition 4.

We will describe a dependence structure, which approximately solves $\max_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ and $\min_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu)$ in certain cases. If the marginal distributions all have decreasing densities as in Theorem 2 (ii) and Proposition 6, then this dependence structure precisely attains both the maximum and the minimum above.

6.1 Extremal dependence structures: Monotone densities

We first focus on the case of monotone densities. To describe the optimal dependence structure, we divide the sample space Ω into $(n+1)$ disjoint events $A_1, \dots, A_n, B$; in other words, $\Omega = A_1 \cup \cdots \cup A_n \cup B$. These sets have the following interpretations:

(B) "body": the event that all individual random variables take the "medium value" of their distributions and the sum of them is a constant;

(A_i) "i-th right tail": the event that the i-th individual random variable takes a "large value" and the other $(n-1)$ random variables take "small values".

Intuitively, the dependence structure is summarized as "joint mix" (Wang and Wang (2016); see Appendix E) and "(approximate) mutual exclusivity". Moreover, A_i is a tail event of the i-th random variable X_i by Wang and Zitikis (2021). The above dependence structure is not completely specified, as one further needs to properly specify what we meant by "large value", "medium value" and "small value", and on each event how the random variables are constructed and dependent. Unfortunately, it is in general not possible to provide an analytical description, if the marginal distributions are heterogeneous. In the homogeneous case with a decreasing density, an analytical description is possible, as discussed in Wang and Wang (2011). More explicit formulas of this dependence structure will be discussed in Section 6.2.

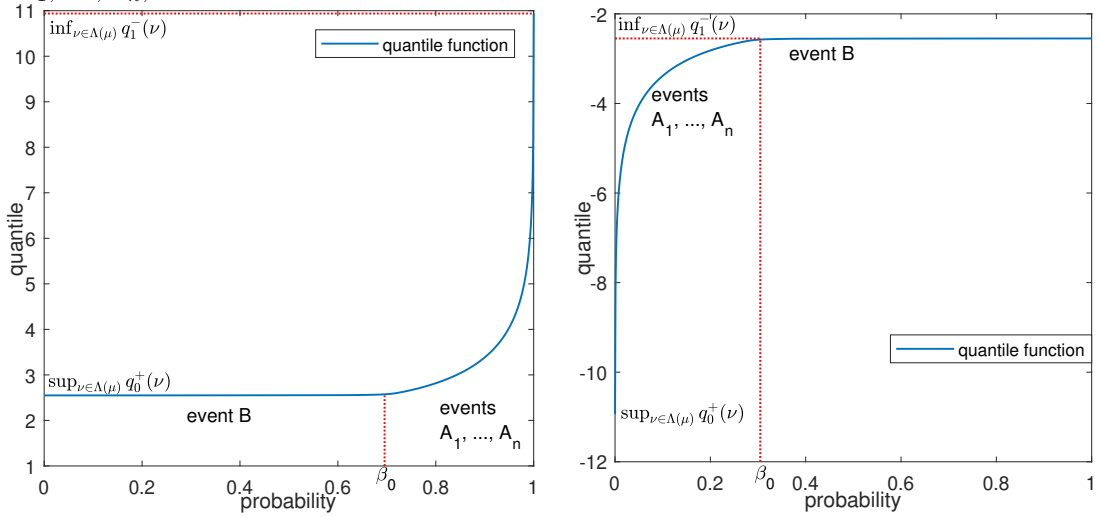
An optimal structure for the problem of both $\max_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ and $\min_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu)$ admits the above dependence structure when (DD) holds, as observed by Jakobsons et al. (2016); a formal and more general result on this observation is Theorem 3 below. The optimality comes from a result of Jakobsons et al. (2016) where it is shown that the sum under this dependence structure is the minimum with respect to convex order given marginal distributions. Moreover, this dependence structure leads to an approximation to optimality in many relevant situations that (DD) does not hold; some numerical results will be shown in Section 8.3.

We note that the optimal structure for $\max_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ or $\min_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu)$ is not unique in general, and in this section we only describe one such candidate. In all our follow-up discussions, we will focus on this candidate.

For now, assume that all marginal distributions have decreasing densities, i.e., (DD) holds. Our method of convolution bounds allows us to determine the existence of the above events $A_1, \dots, A_n, B$ from the optimizing vector $\boldsymbol{\beta}$. We explain this below. For $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n) \in \overline{\Delta}_n$, we denote by

$$R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{i=1}^{n} R_{\beta_i,\beta_0}(\mu_i). \tag{21}$$

Figure 1: Quantile functions for the sum. Left panel: decreasing densities ($n = 3$, quantile functions are $\frac{6}{5}r(t)$, $\frac{4}{5}r(t)$ and $\frac{4}{5}r(t)$, where $r(t) = -\log(\varepsilon + (1-\varepsilon)(1-t)), t \in [0,1]$) and $\varepsilon = 0.0001$); Right panel: increasing densities ($n = 3$, quantile functions are $-\frac{6}{5}r(1-t)$, $-\frac{4}{5}r(1-t)$ and $-\frac{4}{5}r(1-t)$, $t \in [0,1]$.). The events $A_1, \dots, A_n, B$ are described in Theorem 3.



The bound (14) is sharp in our setting by Theorem 2. Suppose that $\boldsymbol{\beta} \in \overline{\Delta}_n$ is the optimizer to (14), that is,

$$\max_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) = \inf_{\boldsymbol{\beta}' \in \Delta_n} R^+_{\boldsymbol{\beta}'}(\boldsymbol{\mu}) = R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}).$$

We describe a classification on the existence of $A_1, \dots, A_n, B$ based on the obtained value of $\boldsymbol{\beta}$.

1. If $\beta_1 = \cdots = \beta_n = 0$, then the optimal dependence structure is "a full joint mix"; that is, the individual random variables add up to a constant on the whole probability space. Only the event B occurs; all events A_i are of zero probability.

2. If $\beta_i \neq 0$ for $i \in I$ where the index set I is a non-empty proper subset of $\{1, \dots, n\}$, then the events of "body" and "i-th right tail" occur; i.e., the possible events are $\{B, A_i : i \in I\}$.

3. If $\beta_i \neq 0$ for all $i = 1, \dots, n$, then all $(n+1)$ events occur; i.e., the possible events are $\{B, A_1, \dots, A_n\}$.

To show the above classification statement, note that β_i indicates the maximum value i-th random variable X_i takes on the event B. More precisely, the largest value of X_i takes on B is $q_{1-\beta_i}(\mu_i)$. Hence, $\beta_i = 0$ means that there is no "large" values of X_i that is considered as a "tail", and thus A_i does not occur. The quantile function of the corresponding sum is illustrated in an example by the left panel of Figure 1.

For a general $t \in (0,1)$, to build a corresponding dependence structure for $\max_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu)$, we need to build the above events B and $A_1, \dots, A_n$ for the conditional distributions $\mu_1^{t+}, \dots, \mu_n^{t+}$. These events will take up probability $1-t$ in total. The remaining event C has probability t, which can specified as

(C) "feet": the event that all individual random variables take small values below their t-quantile.

Conditional on the event C, the dependence structure of $(X_1, \dots, X_n)$ no longer matters, as it does not

contribute to the quantile of the sum. For the optimality, it suffices to require (DD) to hold for $\mu_1^{t+}, \dots, \mu_n^{t+}$. Similarly we can deal with the case of $\min_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^-(\nu)$.

The above discussions also apply to the setting with (ID) in place of (DD) by replacing each involved random variable X_i with its negation $-X_i$. In this case, a similar dependence structure can be constructed based on an event B of "body" and n events $A_1, \dots, A_n$ of "left tail". An example is presented in the right panel of Figure 1. We omit the details here.

In general, if (DD) and (ID) do not hold, but the densities are approximately increasing or decreasing, then we can still use the above construction, and obtain an approximately optimal structure. This will be discussed next.

6.2 The general case and an approximation

Section 6.1 contains a description of a class of dependence structures that leads to the optimized value of $\max_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu)$ under the assumption (DD) or (ID). In this section, we discuss more on this class of dependence structures and show that some further specifications may be used as an approximation for the cases without (DD) and (ID).

In the following, for $0 \leqslant \alpha < \beta \leqslant 1$ and any probability measure μ, we let $\mu^{[\alpha,\beta]}$ be the probability measure given by

$$\mu^{[\alpha,\beta]}(-\infty, x] = \frac{(\min\{\mu(-\infty, x], \beta\} - \alpha)_+}{\beta - \alpha}, \quad x \in \mathbb{R}.$$

Equivalently, $\mu^{[\alpha,\beta]}$ is the distribution measure of the random variable $q_V(\mu)$ where $V \sim \mathrm{U}[\alpha, \beta]$, a uniform random variable on $[\alpha, \beta]$. In particular, $\mu^{[\alpha,1]} = \mu^{\alpha+}$ is the α-tail distribution of μ in Section 4.

We say that a random vector $(X_1^*, \dots, X_n^*)$ attains the maximum of q_0^+ for $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$ if $X_1^* \sim \mu_1, \dots, X_n^* \sim \mu_n$ and $q_0^+(X_1^* + \cdots + X_n^*) = \max_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu)$. The existence of the maximizer $\nu_+ \in \Lambda(\boldsymbol{\mu})$ is guaranteed by Proposition 4.

Next, we introduce a special class of dependence structures in a way similar to Section 6.1. Fix $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n) \in \overline{\Delta}_n$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$. Let the random vector $(X_1^*, \dots, X_n^*)$ satisfy

$$\begin{cases} X_i^* = Z_i \mathbb{1}_{A_i} + W_i \mathbb{1}_{B^c \setminus A_i} + Y_i \mathbb{1}_B, \ i = 1, \dots, n, \\ \text{where } (A_1, \dots, A_n, B) \text{ is a partition of } \Omega \text{ independent of all others, and } \mathbb{P}(A_i) = \beta_i \text{ for each } i, \\ Z_i \sim \mu_i^{[1-\beta_i, 1]}, \ W_i \sim \mu_i^{[0, 1-\beta_0-\beta_i]}, \ Y_i \sim \mu_i^{[1-\beta_0-\beta_i, 1-\beta_i]}, \text{ and } \sum_{i=1}^n Y_i = R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) \text{ almost surely.} \end{cases} \tag{22}$$

The existence of $(X_1^*, \dots, X_n^*)$ satisfying (22) requires some conditions, which will be clear from Theorem 3 below. The construction in (22) is not unique. In particular, the dependence among $(Z_1, \dots, Z_n, W_1, \dots, W_n)$ is not specified. A specification may be given by

$$\begin{cases} X_i^* = q^-_{1-\frac{\beta_i}{1-\beta_0} U}(\mu_i) \mathbb{1}_{\{U \in [0, 1-\beta_0), K = i\}} + q^-_{\frac{1-\beta_0-\beta_i}{1-\beta_0} U}(\mu_i) \mathbb{1}_{\{U \in [0, 1-\beta_0), K \neq i\}} + Y_i \mathbb{1}_{\{U \in [1-\beta_0, 1]\}} \\ \text{for each } i = 1, \dots, n, \text{ where } U \sim \mathrm{U}[0,1], \ \mathbb{P}(K = i) = \frac{\beta_i}{1-\beta_0} \text{ for } i = 1, \dots, n, \\ \text{the random vector } (Y_1, \dots, Y_n) \text{ is coupled by (22), and } U, K, (Y_1, \dots, Y_n) \text{ are independent.} \end{cases} \tag{23}$$

Theorem 3. *Suppose that* $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$ *and* $\max_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) = R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ *for some* $\boldsymbol{\beta} \in \overline{\Delta}_n$. *There exists a random vector* $(X_1^*, \dots, X_n^*)$ *of the form* (22) *attaining the maximum of* q_0^+ *for* $\boldsymbol{\mu}$. *Moreover, if* $\beta_0 = 1$, *then* $\boldsymbol{\mu}$ *is jointly mixable; if* $\beta_0 \neq 1$, $\beta_1, \dots, \beta_n > 0$ *and the minimum of each of the functions* $h_i : (0, 1-\beta_0] \to \mathbb{R}$,

$$h_i(u) = q^-_{1-\frac{\beta_i}{1-\beta_0}u}(\mu_i) + \sum_{j \neq i} q^-_{\frac{1-\beta_0-\beta_j}{1-\beta_0}u}(\mu_j), \quad i = 1, \dots, n, \tag{24}$$

is attained at $u = 1 - \beta_0$, *then* $(X_1^*, \dots, X_n^*)$ *in* (23) *attains the maximum of* q_0^+ *for* $\boldsymbol{\mu}$.

Theorem 3 gives useful information on the worse-case dependence structure attaining $\max_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ based on our knowledge of minimizer $\boldsymbol{\beta}$. The form (23) gives a specification of all n "right tail" events: for each $i = 1, \dots, n$, on the "i-th right tail" event A_i, the $(n-1)$ other random variables are all comonotonic, while they are counter-monotonic to the i-th individual random variable. Theorem 3 can be applied to arbitrary quantile levels t by considering the conditional distributions $\mu_1^{t+}, \dots, \mu_n^{t+}$.

In the homogeneous case ($\mu_1 = \cdots = \mu_n$), the condition for optimality of the dependence structure (23) holds for any distribution with a decreasing density if $\beta_0 \neq 1$ and $\beta_1 = \cdots = \beta_n = \frac{1-\beta_0}{n}$. In this case, $h_1 = \cdots = h_n$ on $(0, 1-\beta_0]$. According to Theorem 3.2 and Proposition 3.4 of Bernard et al. (2014), h_1 is decreasing on $(0, 1-\beta_0]$. Theorem 3 then shows that the corresponding measure ν_+ attains the worst-case quantile aggregation. In the heterogeneous case, we give some numerical examples to show the performance of (23) in Section 8.

The dependence structure (22) is motivated by the discussions in Section 6.1 on the setting (DD), and hence it performs well for distributions with approximately decreasing densities. For the setting (ID), the optimal $\boldsymbol{\beta}$ in (14) for $\max_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ is often given by $\beta_0 = 0$, $\beta_i = 1$ for some i, $\beta_j = 0$ for other $j \neq i$. Hence, in (22), we have $\mathbb{P}(A_i) = 1$ for some i and $\mathbb{P}(A^c) = \mathbb{P}(A_j) = 0$ for other $j \neq i$. In this case, although Theorem 3 holds true, the dependence in (22) is completely unspecified. Nevertheless, for approximately increasing densities, an alternative explicit dependence structure can be similarly designed based on an event B of "body" and n events $A_1, \dots, A_n$ of "left tail". We omit the details.

The only unspecified part in (23) is the design of $Y_1, \dots, Y_n$ which add up to a constant. Such random variables are known to exist under some conditions of joint mixability, but they are not easy to explicitly construct or to simulate except for some very simple cases such as uniform marginal distributions. Below, we give an explicit suboptimal dependence structure as an approximation of (23) without the vector $(Y_1, \dots, Y_n)$:

$$X_i^* = \begin{cases} q^-_{1-\frac{\beta_i}{1-\beta_0}U}(\mu_i)\mathbb{1}_{\{K=i\}} + q^-_{\frac{1-\beta_0-\beta_i}{1-\beta_0}U}(\mu_i)\mathbb{1}_{\{K\neq i\}}, & \text{if } \beta_0 \neq 1, \\ q^-_{1-\frac{1}{n}U}(\mu_i)\mathbb{1}_{\{K=i\}} + q^-_{\frac{n-1}{n}U}(\mu_i)\mathbb{1}_{\{K\neq i\}}, & \text{if } \beta_0 = 1, \end{cases} \quad i = 1, \dots, n, \tag{25}$$

where U, K are given in (23) and we further set $\mathbb{P}(K = i) = \frac{1}{n}$, $i = 1, \dots, n$ in case $\beta_0 = 1$ (i.e., set $\beta_i/(1-\beta_0) = 1/n$). For $(X_1^*, \dots, X_n^*)$ in (25), it is easy to see that $X_i^* \sim \mu_i$ for each $i = 1, \dots, n$. If $\beta_0 \neq 1$,

using h_i in (24), the essential infimum of $\sum_{i=1}^n X_i^*$ is given by

$$\min_{1\leqslant i\leqslant n}\min_{0\leqslant x\leqslant 1} h_i(x)=\min_{1\leqslant i\leqslant n}\min_{0\leqslant x\leqslant 1}\left\{q^-_{1-\frac{\beta_i}{1-\beta_0}x}(\mu_i)+\sum_{j\neq i} q^-_{\frac{1-\beta_0-\beta_j}{1-\beta_0}x}(\mu_j)\right\}. \tag{26}$$

Since $X_1^*,\dots,X_n^*$ are obtained by explicit construction, the above infimum (26) serves as a lower bound for $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$. If $\beta_0\neq 1$, the first-order condition in the optimality of $\boldsymbol{\beta}$ gives $h_i(1-\beta_0)=R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ for $i=1,\dots,n$ satisfying $\beta_i\neq 0$; see (EC.17) in Appendix B.

The dependence structure in (25) has an explicit formula as soon as β_0 is computed, so it has at most the same computational complexity as computing the convolution bound; see the explanations of the computational issues in the numerical results in Section 8.

This construction can be further improved as follows. For any $\boldsymbol{\beta}'\in\Delta_n$, define

$$H(\boldsymbol{\beta}')=\min_{1\leqslant i\leqslant n}\min_{0\leqslant x\leqslant 1} h_i(x;\boldsymbol{\beta}')=\min_{1\leqslant i\leqslant n}\min_{0\leqslant x\leqslant 1}\left\{q^-_{1-\frac{\beta'_i}{1-\beta'_0}x}(\mu_i)+\sum_{j\neq i} q^-_{\frac{1-\beta'_0-\beta'_j}{1-\beta'_0}x}(\mu_j)\right\}.$$

We then solve another n-dimensional optimization problem

$$\sup_{\boldsymbol{\beta}'\in\Delta_n} H(\boldsymbol{\beta}'). \tag{27}$$

The maximum point is denoted by $\boldsymbol{\gamma}=(\gamma_0,\gamma_1,\dots,\gamma_n)\in\overline{\Delta}_n$. Hence, we have the suboptimal dependence structure

$$X_i^*=\begin{cases} q^-_{1-\frac{\gamma_i}{1-\gamma_0}U}(\mu_i)\mathbb{1}_{\{K=i\}}+q^-_{\frac{1-\gamma_0-\gamma_i}{1-\gamma_0}U}(\mu_i)\mathbb{1}_{\{K\neq i\}}, & \text{if } \gamma_0\neq 1,\\ q^-_{1-\frac{1}{n}U}(\mu_i)\mathbb{1}_{\{K=i\}}+q^-_{\frac{n-1}{n}U}(\mu_i)\mathbb{1}_{\{K\neq i\}}, & \text{if } \gamma_0=1,\end{cases} \qquad i=1,\dots,n, \tag{28}$$

where U,K are independent, $U\sim \mathrm{U}[0,1]$ and $\mathbb{P}(K=i)=\frac{\gamma_i}{1-\gamma_0}$, $i=1,\dots,n$ if $\gamma_0\neq 1$ and $\mathbb{P}(K=i)=\frac{1}{n}$, $i=1,\dots,n$ if $\gamma_0=1$. For $(X_1^*,\dots,X_n^*)$ in (28), it is easy to see that $X_i^*\sim\mu_i$ for each $i=1,\dots,n$ and the essential infimum of $\sum_{i=1}^n X_i^*$ is $H(\boldsymbol{\gamma})$.

It turns out that (25) gives a good approximation for the maximum value of q_0^+ in many cases and (28) does even better. The numerical performance will be illustrated in Section 8. Note that

$$H(\boldsymbol{\beta})\leqslant H(\boldsymbol{\gamma})\leqslant \sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)\leqslant R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}). \tag{29}$$

As a result, we provide two-side approximation intervals $[H(\boldsymbol{\beta}),R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})]$ or $[H(\boldsymbol{\gamma}),R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})]$ for true value of $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$. If only $\boldsymbol{\beta}$ is provided, the former interval can be adopted to approximate the worst-case quantile aggregation; if it is convenient to conduct another optimization (27), the latter one would be more accurate in approximation.

7 Dual formulation

In this section, we investigate the dual formulation of the quantile aggregation problem. In Theorem 2, the convolution bound (14) is obtained by an n-dimensional optimization problem. The main result in this section is that under continuity conditions the convolution bound (14) is equal to a dual bound (30), with a convenient correspondence between the minimizers of both problems.

The following proposition gives a dual bound on quantile aggregation, which is essentially Theorem 4.17 of Rüschendorf (2013) that is expressed in terms of probability instead of quantiles.

Proposition 8. *For $t \in [0,1)$, it holds that*

$$\text{[dual bound]} \qquad \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu) \leqslant D_n^{-1}(1-t), \tag{30}$$

where $D_n^{-1}(\alpha) = \inf\{x \in \mathbb{R} : D_n(x) < \alpha\}$, $\alpha \in (0,1]$ and the function $D_n : \mathbb{R} \to \mathbb{R}$ is defined by

$$D_n(x) = \inf_{\mathbf{r} \in \Delta_n(x)} \left\{ \sum_{i=1}^n \frac{1}{x-r} \int_{r_i}^{x-r+r_i} \mu_i(y,\infty) \mathrm{d}y \right\}, \ x \in \mathbb{R}, \tag{31}$$

where $\mathbf{r} = (r_1, \dots, r_n)$, $r = \sum_{i=1}^n r_i$ and $\Delta_n(x) = \{(r_1, \dots, r_n) \in \mathbb{R}^n : \sum_{i=1}^n r_i < x\}$.

Below we always write $\mathbf{r} = (r_1, \dots, r_n)$ and $r = \sum_{i=1}^n r_i$. We find that the dual bound (30) is equal to our convolution bound (14) if the marginal distribution and quantile functions are continuous.

Theorem 4. *For fixed $t \in [0,1)$, let $x = D_n^{-1}(1-t)$. Suppose that each of $\mu_1, \dots, \mu_n$ has continuous distribution and quantile functions. The convolution bound (14) and the dual bound (30) share the same value x. Moreover, the correspondence between the minimizers $\boldsymbol{\beta} \in \overline{\Delta}_n$ of (14) and $\mathbf{r}$ in the closure of $\Delta_n(x)$ of (31) is given by:*

$$\mu_i(-\infty, r_i] = 1 - \beta_0 - \beta_i, \quad \mu_i(-\infty, x-r+r_i] = 1 - \beta_i, \quad i = 1, \dots, n. \tag{32}$$

As far as we are aware of, there are no sharpness results on the dual bound in the setting of heterogeneous marginals. Therefore, our main results on convolution bounds also contribute to the literature by establishing the sharpness of the dual bounds in several situations, as the convolution bound and the dual bound are usually equal. Moreover, we note that the convolution bound is applicable to RVaR aggregation problems, whereas the dual bound based on probability is specific to quantile aggregation. On the computational side, as the set $(1-t)\Delta_n$ is bounded and the set $\Delta_n(x)$ is unbounded, optimization of the convolution bound (14) is often easier than that of the dual bound (30). Moreover, (30) needs to additionally compute an inverse function from D_n. Further, the equivalence in Theorem 4 may fail for discrete distributions; see Table 2 of Section 8.

In the homogeneous case $\mu_1 = \cdots = \mu_n$, Embrechts and Puccetti (2006) derived a (reduced) dual bound

for the worst-case quantile aggregation based on a one-dimensional optimization problem:

$$\text{[reduced dual bound]} \qquad D^{-1}(1-t) = \inf\left\{x \in \mathbb{R} : D(x) < 1-t\right\}, \tag{33}$$

where

$$D(x) = \inf_{a<\frac{x}{n}} \frac{n}{x-na} \int_a^{x-(n-1)a} \mu\left((y,\infty)\right) \mathrm{d}y, \quad x \in \mathbb{R}.$$

This dual bound is a special case of (30) by letting $r_1 = \cdots = r_n$ in (31). Thus, the reduced dual bound (33) is larger than or equal to the dual bound (30), as well as our convolution bound (14) by Theorem 4. Similarly to the discussion in Section 5, the dual bound (30) and the reduced one (33) are not generally equal.

Similarly to Theorem 4, one can show that the reduced dual bound (33) is the same as the reduced convolution bound (15) if the marginal distribution and quantile functions are continuous. In Figure 3 (right panel) of Section 8, we give out examples that (14) is strictly smaller than (33).

8 Numerical illustration

In this section, the convolution bounds in Theorems 1-2 are computed and compared with the existing bounds by numerical examples, including the dual bound of Embrechts and Puccetti (2006) and the rearrangement algorithm (RA) of Puccetti and Rüschendorf (2012) and Embrechts et al. (2013). We give some numerical examples to show the performance of the candidate and suboptimal dependence structures (23) and (25) in the heterogeneous case.

We briefly explain the output of RA. If a tuple $\boldsymbol{\mu}$ of marginal distributions is given as quantile functions or distribution functions, then RA involves discretization of the marginal distributions by N steps, where N is chosen as 10^5 in our implementations. If the marginal distributions are given as empirical distributions of data, then discretization is not needed. Running RA on $\boldsymbol{\mu}$ returns an interval $[\underline{s}_N, \bar{s}_N]$, whose left and right end-points are close when N is sufficiently large, providing an approximation for $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_t^+(\nu)$. The left end-point is always a (numerical) lower bound, whereas the right end-point is not a lower or upper bound. Although RA is a popular algorithm, there is no guarantee that its produced lower bound converges to the true value of $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_t^+(\nu)$, and there are no theoretical results on the time for RA to converge. For the above claims and a detailed explanation on implementing RA, see Embrechts et al. (2013). Consistently with the literature, we treat the upper value produced by RA as a good approximation of the true value of the worst-case RVaR, although no convergence result is established.

Next, we explain how we compute the convolution bound, denoted by B_{conv}. We use the built-in function `fmincon` in MATLAB to numerically compute B_{conv}, where the input are the marginal quantile functions. No discretization is needed for this computation, as long as marginal quantiles can be specified.[7] All computations are performed on MATLAB R2017b with Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz. In

[7]Note that many distributions, including empirical distributions from data, have their quantile functions as built-in functions in most computational softwares. For those that do not have a built-in quantile function, computing a numerical quantile function is a standard and simple task.

our implementations, we use the default optimization method, the interior-point algorithm. The convergence criterion is a termination tolerance scalar value 10^{-6} on the first-order optimality, which is the default choice, and this convergence criterion is met in all computations. Note, however, that the optimization problem required in computing B_{conv} is generally non-convex, and so global optimality may not always be attainable by `fmincon`. Nonetheless, in theory, as the left end-point $\underline{s}_N$ of RA is a lower bound and B_{conv} is an upper bound, we have

$$\text{RA output } \underline{s}_N \leqslant \text{true } \sup_{\nu\in\Lambda_n(\mu)} q_t^+(\nu) \leqslant \text{true } B_{\mathrm{conv}} \leqslant \text{computation of } B_{\mathrm{conv}}.$$

As far as we know, there is no theoretical guarantee that the global optimum in the computation of B_{conv} is attained. Nevertheless, in most results, $\underline{s}_N$ and computed values of B_{conv} coincide almost perfectly, and therefore convergence is practically verified in these cases.

With discretization, the quantile aggregation problem can also be formulated into a linear programming (LP) problem with an exponential number of variables. The LP formulation approximates to the true optimal value of the problem, but it is difficult if the dimension n or the number of points in the discretization is high. In a real risk management problem where loss distributions are typically continuous (such as asset prices or insurance losses), a fine discretization is required to ensure good approximation, making LP very slow. In Appendix F, we provide a detailed comparison among LP, RA and the convolution bound to compute the quantile aggregation problem.

8.1 Convolution bounds on RVaR aggregation

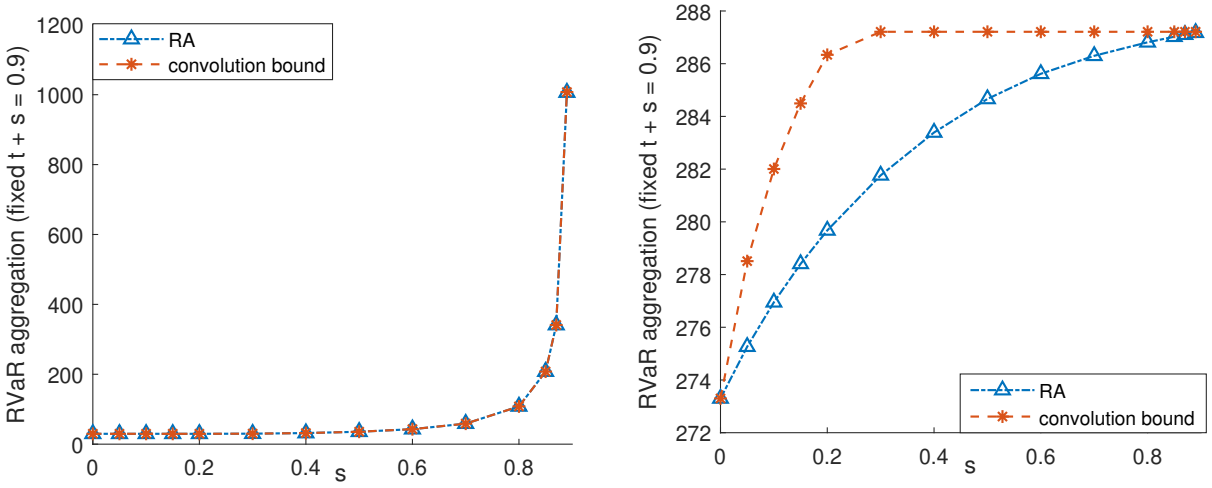
For any t, s with $0 \leqslant t < t + s \leqslant 1$, we numerically compute the RVaR aggregation value $R_{t,s}(\nu)$, $\nu \in \Lambda_n(\mu)$ with different methods in the homogeneous case, where the marginal distribution is identical and denoted by μ. The convolution bound is given by (12) and the true value is approximated by RA.

We fix $t + s = 0.9$ and change $s \in (0, 0.9)$ to simulate values of $\sup_{\nu\in\Lambda_n(\mu)} R_{t,s}(\nu)$. In Figure 2 (left panel), we check Theorem 1 that the convolution bound (12) is sharp for marginals with decreasing densities. In Figure 2 (right panel), we see that the convolution bound (12) is not sharp for marginals with increasing densities. Although this bound is not sharp for increasing densities, the difference is small and it performs quite well numerically. Moreover, in Figure 2, the convolution bound (12) is sharp if $t = 0$ (Theorem 1) and $s \downarrow 0$ (Theorem 2).

8.2 Numerical comparison with existing results

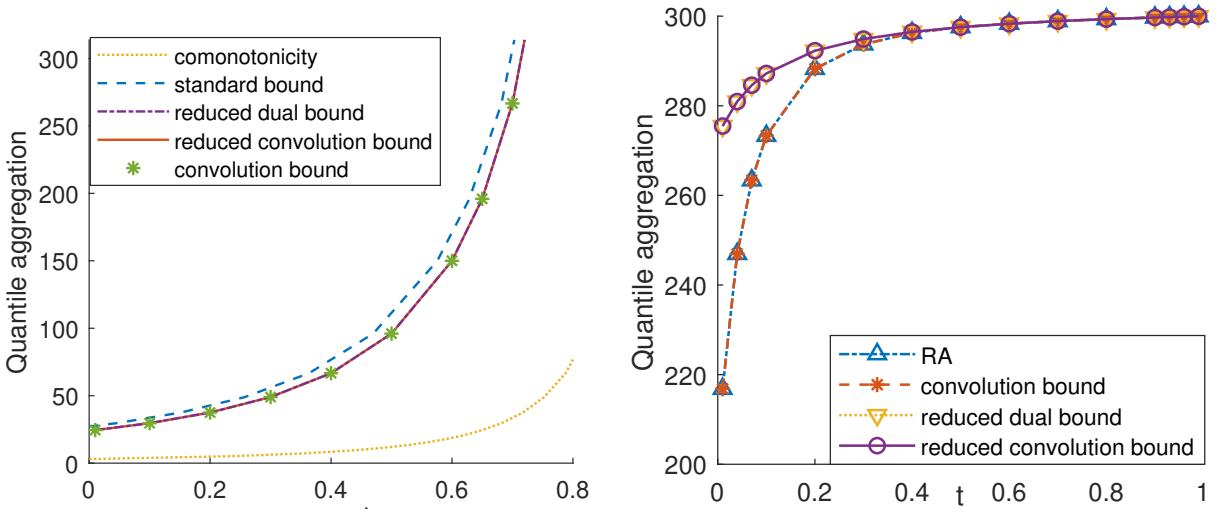
For $t \in [0, 1)$, we numerically compare the quantile aggregation value $q_t^+(\nu)$, $\nu \in \Lambda_n(\mu)$ with analytical bounds obtained in the homogeneous case, where the marginal distribution is identical and denoted by μ. Recall that the convolution bound is given by (14), the (reduced) dual bound derived in Embrechts and Puccetti (2006) is given by (33) and the reduced convolution bound is given by (15). The standard bound is

Figure 2: Bounds for $\sup_{\nu\in\Lambda_3(\mu)} R^+_{0.9-s,s}(\nu)$. Left panel: μ = Pareto$(1, 1/2)$ with a decreasing density $\frac{1}{2}x^{-3/2}$, $x \in [1,\infty)$. Right panel: μ has an increasing density $\frac{5}{9}(101-x)^{-\frac{3}{2}}$, $x \in [1, 100]$.



derived from the lower Fréchet-Hoeffding bound (see Remark A.29 of Föllmer and Schied (2016)). We also give the quantile aggregation value under a comonotonic scenario for comparison.

Figure 3: Bounds for $\sup_{\nu\in\Lambda_3(\mu)} q^+_t(\nu)$. Left panel: $\mu = $ Pareto$(1, 1/2)$ with a decreasing density $\frac{1}{2}x^{-3/2}$, $x \in [1,\infty)$. Right panel: μ has an increasing density $\frac{5}{9}(101-x)^{-3/2}$, $x \in [1, 100]$. In the left panel, "reduced dual bound", "reduced convolution bound", "convolution bound" and "RA" have the same curve, and for better visibility the "RA" curve is not plotted. In the right panel, "RA" and "convolution bound" have the same curve and "reduced dual bound" and "reduced convolution bound" have the same curve.



This section also serves as a numerical illustration for the worst-case risk aggregation in Section 2.1. In Figure 3, we compute (1) in the setting that losses in a portfolio follow some given marginal distributions. Figure 3 (left panel) illustrates that the convolution bound (14), the reduced convolution bound (15) and the reduced dual bound (33) share the same value of quantile aggregation for a Pareto distribution. The standard bound performs worst as an upper bound for $\sup_{\nu\in\Lambda_n(\mu)} q^+_t(\nu)$. The comonotonic scenario serves

Table 1: RA (with $N = 10^5$) and the convolution bound B_{conv} to compute $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ for two different settings of heterogeneous marginal distributions. RA produces an interval whose left-end point is a lower bound, and the convolution bound B_{conv} is an upper bound which is sharp in the first setting of Pareto distributions.

	$X_i \sim \mathrm{Pareto}(1,\alpha_i)$, $\alpha_i = 2+i$, $i = 1,\dots,20$	time $n = 20$	$X_i \sim \mathrm{Pareto}(1, i+2)$, $i = 1,\dots,20$, $X_{20+i} \sim \mathrm{LogN}(5-i, (i/2)^2)$, $i = 1,\dots,20$, $X_{40+i} \sim \Gamma(i+1, \frac{10}{i})$, , $i = 1,\dots,20$	time $n = 60$
RA	[22.5966, 22.5971]	111s	[539.5141, 539.6205]	639s
B_{conv} (14)	22.5968	46s	539.5611	672s
RA minus (14)	$[-2.04 * 10^{-4}, 3.02 * 10^{-4}]$		[-0.0470, 0.0594]	

as a lower bound. Results for other distributions such as Lognormal and Gamma distribution are similar and we omit them.

In Figure 3 (right panel), we plot analytical bounds of the maximum possible quantile aggregation value $\sup_{\nu\in\Lambda_3(\mu)} q_t^+(\nu)$, where the convolution bound (14) achieves a strictly smaller value than the reduced dual bound (33). It means that our bound (14) is an analytically better bound for quantile aggregation. Figure 3 (right panel) further shows that (14) is better than the reduced convolution bound (15); see also Example EC.2.

In Table 1, we numerically check the performance of the bound (14) against RA in more detail.

Concerning performance, Figure 3 (right panel) and Table 1 both indicate that the convolution bound and RA have a similar value for most cases. We discuss three aspects. First, we emphasize again that the true value of $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ is generally unavailable. It is available in cases with monotone densities, where the true value equals to the convolution bound according to Theorem 2. It is the case of Figure 3 and the first model in Table 1. Second, if the true value of $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$ is unknown, then we can use the (upper) convolution bound together with the lower bound provided by the RA to approximately target the true value. As shown in the second model of in Table 1, the difference between the two bounds is quite small and we can approximately know the true value. Third, we show that in some cases RA does not perform well while the convolution bound provides a sharp result; see Example 1.

Example 1. Let μ be a triatomic uniform distribution on $\{1,2,3\}$. By constructing a random vector uniformly distributed on $\{(1,2,3),(2,3,1),(3,1,2)\}$, we get $\sup_{\nu\in\Lambda(\mu,\mu,\mu)} q_0^+(\nu) = 6$. As a result, (14) provides a sharp upper bound $\inf_{\boldsymbol{\beta}\in\Delta_3} R_{\boldsymbol{\beta}}^+(\mu,\mu,\mu) = 6$ with the optimal $\boldsymbol{\beta} = (1,0,0,0)$. However, the interval provided by the RA is $[5,5]$; see Example EC.4 for details.

Concerning computation time, we find that the convolution bound (14) is computed quicker than or similarly to RA. In conclusion, (14) is not only a good analytical upper bound, but also performs quickly in the numerical calculation for the maximum possible lower end-point $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu)$.

Theorem 4 assumes continuous distribution and quantile functions. Generally, equivalence between quantile methods and probability methods can be troublesome when dealing with discrete distributions. We illustrate in a simple example that the equivalence in Theorem 4 may fail. Define $\mu = \mathrm{Bernoulli}(0.5)$ and $n = 3$. We numerically compute the values of RA, the convolution bound and the dual bound on

Table 2: RA (with $N = 10^5$), true value, convolution bound B_{conv} and dual bound B_{dual} on $\sup_{\nu\in\Lambda_3(\mu)} q_t^+(\nu)$.

t	0.1	0.2	0.3	0.4	0.5	0.6	0.7
RA (left end-point)	1	1	2	2	**2**	3	3
true value	1	1	**2**	**2**	**3**	3	3
B_{conv}	1.6667	1.875	**2**	**2**	**3**	3	3
B_{dual}	1.6667	1.875	**2.1429**	**2.5**	3	3	3

Table 3: Numerical values of lower end-points in Figure 4.

	μ_1 = Pareto(1,3) μ_2 = LogN(0,1) $\mu_3 = \Gamma(1,2)$	μ_1 = Pareto(1,1/3) μ_2 = LogN(0,1) $\mu_3 = \Gamma(1,2)$	μ_1 = Pareto(1,3) μ_2 = LogN(-1,1) $\mu_3 = \Gamma(1,2)$	μ_1 = Pareto(1,3) μ_2 = LogN(0,1) $\mu_3 = \Gamma(3,2)$
Mean	5.1487	∞	4.1065	9.1487
RA $[H(\boldsymbol{\gamma}), R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})]$	[4.2856,4.2857] [4.1185, 4.2857]	[8.5933,8.5936] [8.055, 8.5936]	[3.2545,3.2545] [3.1254,3.2545]	[7.6338,7.6341] [7.3653,7.634]
Bound (14)	4.2857	8.5936	3.2545	7.634
Candidate (23)	4.2855	8.4995	3.2545	7.5415
Suboptimum (25)	4.0739	7.7835	3.0587	7.2889
Suboptimum (28)	4.1185	8.055	3.1254	7.3653

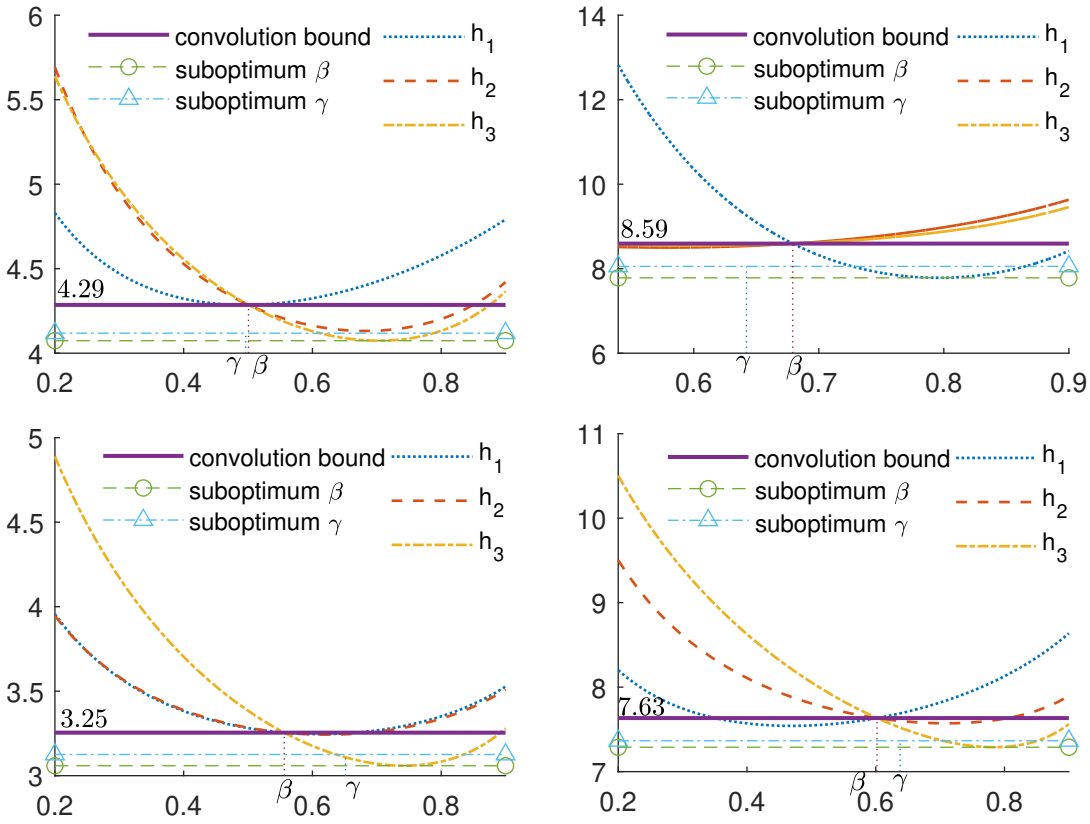
$\sup_{\nu\in\Lambda_3(\mu)} q_t^+(\nu)$ in Table 2. The true values are available in this simple setting for comparison. We make two observations. First, as we mentioned above, the true value of $\sup_{\nu\in\Lambda_3(\mu)} q_t^+(\nu)$ is bounded from below by the RA left end-point and from above by the convolution bound and is exactly obtained if these two bounds are equal. Second, for $t = 0.3$ and 0.4, the values of the convolution bound B_{conv} and the dual bound B_{dual} are not the same. In these cases, we observe that B_{conv} is closer (equal) to the true value than B_{dual}.

8.3 Performance of extremal dependence structures

Recall that in Section 6.2 we propose a candidate dependence structure (23) for the worst-case quantile aggregation. We also state a suboptimal structure (25) without involving $Y_1, \dots, Y_n$. A better suboptimum (28) is obtained by solving another optimization problem and a two-side approximation interval $[H(\boldsymbol{\gamma}), R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})]$ is established. We now give some numerical examples to compare their corresponding lower end-points in the heterogeneous case with $n = 3$. As shown in Theorem 3, possible values of the aggregation variable in (23) are those of the functions h_1, h_2, h_3 on $[0, 1-\beta_0]$, while the corresponding values in (25) are those of h_1, h_2, h_3 on $[0, 1]$. Thus, the lower end-point derived from (23) is attained at the minimal values of all h_1, h_2, h_3 on $[0, 1-\beta_0]$, while that from (25) is attained at those on $[0, 1]$.

In Figure 4, according to the sufficient condition in Theorem 3, the third subfigure shows that (23) gives out the worst-case quantile aggregation. Even in the other subfigures, the essential infimum of $\sum_{i=1}^n X_i^*$ of (23), which is the minimal value of h_1 on $[0, 1-\beta_0]$, is just slightly lower than the corresponding $q_0^+(\nu_+)$. We further show the numerical values in Table 3, including the convolution bound, the RA results, and values from the suboptimal structures (25) and (28). Recall that (25) is based on $\boldsymbol{\beta}$, while (28) requires solving $\boldsymbol{\gamma}$ in another optimization problem. The suboptimal methods give explicit random vectors, and hence it is useful in visualizing the worst case of quantile aggregation. In Table 3, both (25) and (28) produce numbers close to the convolution bound in many cases, while (28) is always better but requires more computation.

Figure 4: Performance of extremal dependence structures with settings in Table 3. In each panel, we plot the function of h_i and the values of quantile aggregation provided by the convolution bound, the suboptimum $\boldsymbol{\beta}$ and the suboptimum $\boldsymbol{\gamma}$.



9 Two applications

9.1 Robust risk management

The convolution bounds can be directly applied to compute the worst-case or best-case risk aggregation problem in Section 2.1 for the risk measure being VaR or RVaR. In this section, we solve the robust portfolio selection problem (3) also presented in Section 2.1. Using a standard Lagrangian technique, for (3), it is equivalent to solve

$$\text{maximize} \quad \inf_{\mathbf{X}\sim\boldsymbol{\mu}} \mathbb{E}[u(\boldsymbol{\lambda}\cdot\mathbf{X})] - \xi \sup_{\mathbf{X}\sim\boldsymbol{\mu}} q_t^+(\boldsymbol{\lambda}\cdot(-\mathbf{X})) \quad \text{over } \boldsymbol{\lambda}\in\overline{\Delta}_{n-1}, \tag{34}$$

where $\xi \geqslant 0$ is a Lagrangian multiplier. After obtaining the optimizer $\boldsymbol{\lambda}_\xi^*$ of (34) for varying $\xi \geqslant 0$ we can calibrate ξ with the risk constraint $\sup_{\mathbf{X}\sim\boldsymbol{\mu}} q_t^+(\boldsymbol{\lambda}_\xi^*\cdot(-\mathbf{X})) = x$ to solve (3) whenever the risk constraint is binding. In what follows, we will focus on (34). Recall that u is a strictly concave and increasing function.

Table 4: Different optimal portfolios under different risk constraints, where $n = 3$, $t = 0.99$, $X_1 \sim \mathrm{N}(0.9, 1.8^2)$, $X_2 \sim \mathrm{N}(0.1, 0.2^2)$, $X_3 = Y - 4$ with $Y \sim \mathrm{LogN}(1.5, 1.2^2)$, and $u(x) = 5(1 - \exp(-x/5))$, $x \in \mathbb{R}$

Lagrangian ξ	0	0.01	0.1	1	10
risk constraint x	$\geqslant 3.075$	2.489	1.626	0.459	0.365
optimal portfolio $\boldsymbol{\lambda}^*$	$(0.41, 0.23, 0.36)$	$(0.23, 0.38, 0.39)$	$(0, 0.64, 0.36)$	$(0, 0.97, 0.03)$	$(0, 1, 0)$
utility $U(\boldsymbol{\lambda}^*)$	0.704	0.702	0.678	0.200	0.095

For $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$, the first term in the problem (34) admits a simple formula

$$\inf_{\mathbf{X} \sim \boldsymbol{\mu}} \mathbb{E}[u(\boldsymbol{\lambda} \cdot \mathbf{X})] = \int_0^1 u\left(\sum_{i=1}^n \lambda_i q_v^+(\mu_i) \right) \mathrm{d}v =: U(\boldsymbol{\lambda}), \tag{35}$$

because the worst-case portfolio is comonotonic (e.g., Corollary 3.29 of Rüschendorf (2013)). Since $\boldsymbol{\lambda} \mapsto u(\sum_{i=1}^n \lambda_i q_v^+(\mu_i))$ is concave, so is U. The concavity of U implies that a maximizer for (35) may favour some diversification. On the other hand, as shown by Proposition 7.1 of Chen et al. (2022), a minimizer for $\sup_{\mathbf{X} \sim \boldsymbol{\mu}} q_t^+(\boldsymbol{\lambda} \cdot (-\mathbf{X}))$ favours no diversification if the marginal distributions are identical and satisfy (DD) or (ID). Therefore, intuitively, there is a trade-off between diversification and concentration in (34). Applying the convolution bound in the form of Theorem A.2 (the symmetric version of Theorem 2) and using the positive homogeneity of RVaR, we have

$$\begin{aligned} \inf_{\mathbf{X} \sim \boldsymbol{\mu}} \mathbb{E}[u(\boldsymbol{\lambda} \cdot \mathbf{X})] - \xi \sup_{\mathbf{X} \sim \boldsymbol{\mu}} q_t^+(\boldsymbol{\lambda} \cdot (-\mathbf{X})) &= U(\boldsymbol{\lambda}) + \xi \inf_{\mathbf{X} \sim \boldsymbol{\mu}} q_{1-t}^-(\boldsymbol{\lambda} \cdot \mathbf{X}) \\ &\geqslant U(\boldsymbol{\lambda}) + \xi \sup_{\boldsymbol{\beta} \in (1-t)\Delta_n} \sum_{i=1}^n R_{1-\beta_0-\beta_i, \beta_0}(\lambda_i X_i) \\ &= \sup_{\boldsymbol{\beta} \in (1-t)\Delta_n} \left\{ U(\boldsymbol{\lambda}) + \xi \sum_{i=1}^n \lambda_i R_{1-\beta_0-\beta_i, \beta_0}(\mu_i) \right\}. \end{aligned}$$

The above objective will be maximized over $\boldsymbol{\lambda}$. For a fixed $\boldsymbol{\beta} \in (1-t)\Delta_n$, the above objective is concave in $\boldsymbol{\lambda}$, which is easy to maximize. As we discussed in Section 5.2, the optimization of $\boldsymbol{\beta}$ is also often simple. The inequality above becomes an equality when the convolution bound is sharp. This is guaranteed if $\mu_1, \dots, \mu_n$ have increasing (or decreasing) tail densities below level $1 - t$ (Theorem A.2). Since t is close to 1, this requirement is very weak and it is satisfied by portfolio models in practice. We will assume that the convolution bound is sharp from now on.

For a simple illustration, we consider three heterogeneous assets with normal and log-normal distributions and different parameters. The parameters of these distributions are chosen such that the problem is non-trivial in the sense that the three marginal distributions do not dominate each other. We take an exponential utility function to characterize the preference of the decision maker. The numerical results of the following setting are given in Table 4. If the Lagrangian multiplier $\xi = 0$, then the problem is robust utility maximization under uncertainty without risk constraint (or, the risk constraint is not binding), and

Table 5: Different cases of the optimal matching

threshold y	$(0, 0.046)$	$[0.046, 0.353)$	$[0.353, 1)$
corresponding probability t	$(0, 0.077)$	$[0.077, 0.420)$	$[0.420, 1)$
optimal $\boldsymbol{\beta}/(1-t)$ of (16)	$(0, 1, 0, 0)$	$(0, 1, 0, 0)$	$(1, 0, 0, 0)$
optimal $\boldsymbol{\beta}$ of (36)	$(<1, >0, >0, >0)$	$(<1, >0, 0, 0)$	$(1, 0, 0, 0)$
possible events	C, B, A_Z, A_X	C, B, A_Z	C, B

the optimizer is a diversified portfolio $(0.41, 0.23, 0.36)$. As the Lagrangian multiplier ξ increases, the role of the risk constraint is getting more important, and the optimal portfolio becomes more concentrated. In case $\xi = 10$, the optimal portfolio is to only invest in the second asset, which has the smallest expected return and the smallest variance (thus, the safest choice). This is consistent to our intuition that when the penalty on the worst-case dependence is large, the decision maker prefers a concentrated portfolio, which does not have dependence uncertainty. A similar phenomenon is also observed by Pflug and Pohl (2018) and Chen et al. (2022) in different settings without the utility term $U(\boldsymbol{\lambda})$.

9.2 The O-ring model

We proceed to analyze the O-ring model presented in Section 2.2. Our goal is to find the minimum value of (5) as well as its optimizing dependence structure. This optimizing dependence structure will yield matching patterns in the label market. For this, we use Proposition 3 and the arguments in Section 6.1 on the extremal dependence.

For an illustration, we will use the following simple setting: $\mu_Z = \text{Beta}(5/6, 1)$, $\mu_1 = \mu_2 = \text{Beta}(5/4, 1)$. That is, there are two workers in each firm, where the product value of the firm and the successful probabilities of the workers follow the Beta distributions. Note that all Beta distributions of the form $\text{Beta}(\alpha, 1)$ satisfy the condition in Proposition 3 (iii); equivalently, (ID) holds for the distributions of $\log(Z)$, $\log(X_1)$ and $\log(X_2)$. Hence, the right-hand side of (16) is the true minimum value in (5). For values of the threshold $y \in [0, 1)$, we obtain the corresponding minimum probability t as well as the optimal $\boldsymbol{\beta}/(1-t)$ of (16) in Table 5. We will explain this table in more detail below.

For a given threshold $y \in (0, 1)$ and its corresponding probability level $t \in (0, 1)$, by Proposition 1, one needs to consider an optimal matching of the conditional distributions μ_Z^{t+}, μ_1^{t+} and μ_2^{t+} on an event with total probability $1 - t$, and the matching on the remaining event C with probability t can be arranged arbitrarily. As $\mu_1 = \mu_2$, by symmetry, the overall optimal dependence structure includes four possible events ($\Omega = A_Z \cup A_X \cup B \cup C$):

(B) two medium-skilled workers work together as a team in a medium-value firm;

(A_Z) a low-value firm hires two high-skilled workers as a team;

(A_X) a low-skilled worker works with a high-skilled coworker in a high-value firm;

(C) two very low-skilled workers work in a very low-value firm.

In the above construction of (Z, X_1, X_2), we have

$$B \cup A_Z \cup A_X = \{Z \geqslant q_t^+(\mu_Z)\} = \{X_1 \geqslant q_t^+(\mu_1)\} = \{X_2 \geqslant q_t^+(\mu_2)\};$$

$$C = \{Z < q_t^+(\mu_Z)\} = \{X_1 < q_t^+(\mu_1)\} = \{X_2 < q_t^+(\mu_2)\}.$$

Such a structure is called t-concentration by Wang and Zitikis (2021).

For a given threshold $y \in (0,1)$ and its corresponding probability level $t \in (0,1)$, by the arguments in Section 6.1, since (ID) holds, to determine the possible events in the dependence structure, one should compute the optimal $\boldsymbol{\beta}$ from the lower convolution bound (A.4), i.e.,

$$\sup_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=Z,1,2} R_{1-\beta_i-\beta_0,\beta_0} \left(\mu_i^{t+} \circ \exp\right). \tag{36}$$

We denote the optimal $\boldsymbol{\beta}$ in Table 5 by $(1-\beta_Z-\beta_1-\beta_2, \beta_Z, \beta_1, \beta_2)$. As in Section 6.1, we have the following classification:

Case 1. If $y \in [0.353, 1)$, then $\beta_Z = \beta_1 = \beta_2 = 0$, implying that the events C and B occur.

Case 2. If $y \in [0.046, 0.353)$, then $\beta_Z > 0$ and $\beta_1 = \beta_2 = 0$, implying that the events C, B and A_Z occur.

Case 3. If $y \in (0, 0.046)$, then $\beta_Z, \beta_1, \beta_2 > 0$, implying that all the events C, B, A_Z and A_X occur.

The event C corresponds to the proportion of firms and workers that are given up by the matching problem. Since our goal is to obtain as many project values above y as possible, some projects have to be left behind, and they are composed of low-value firms and low-skilled workers. The event B corresponds to the proportion of medium-value firms which hire medium-skilled workers. This reflects the majority of firms and workers and they are matched together. The event A_Z means that the low-value firm has to hire high-skilled workers to minimize the global deficiency proportion of production.

The event A_X matches a high-value firm and a high-skilled worker with a low-skilled coworker. If the firm value is high enough and the threshold y is low enough, then there is no point for this firm to hire two high-skilled workers anymore; in fact, the firm can hire one high-skilled worker and reduce its cost by hiring a low-skilled coworker if the goal is only to bypass the threshold $y = 0.046$. This may be realistic in settings where robots or automated machines are cheaper and less effective than human workers, but they are sufficient to pass a threshold of interest (e.g., quality control) for the firm, so the firm would use robots or automated machines. However, this situation does not happen if the threshold is high enough.

The optimal matching, featured with events A_Z and A_X, for problem (5) is quite different from the classic result in Kremer (1993), where high-skilled workers are always matched with high-value firms. Certainly, the objectives in the two settings are different. To explain this from the perspective of dependence, the product function $(z, x_1, \dots, x_n) \mapsto z \prod_{i=1}^n x_i$ is a supermodular function, and its expected value is maximized by

positive matching, that is, comonotonicity; see e.g., Puccetti and Wang (2015, Section 2). On the other hand, $(z, x_1, \dots, x_n) \mapsto \mathbb{1}_{\{z \prod_{i=1}^n x_i \leqslant y\}}$ is neither supermodular nor submodular, and its minimization (or maximization) is highly complicated and involves both positive and negative matching; see Puccetti and Wang (2015, Section 3) for extremal negative dependence. Translating this into the O-ring theory, to minimize the percentage of production values under a threshold, one needs to assign high-skilled workers and high-value firms to assist less-performed workers or firms. Such a matching policy is quite common in socially relevant real situations, e.g., team tournaments, help groups, and financial assignments, to name a few. The appearance of negative matching is getting increasing attention in various economic contexts; see e.g., the recent work of Boerma et al. (2021, 2023).

10 Conclusion

Using the RVaR convolution result of Embrechts et al. (2018), we establish new (semi-analytical) bounds for the problem of quantile aggregation, and show that these bounds are sharp in many cases with analytical formulas in the literature. We can interpret the corresponding worst-case dependence structure and give explicit construction for the complicated optimization problem. The convolution bounds cover all existing theoretical results on quantile aggregation. Moreover, the proposed bound has advantages in its tractability, interpretability, and computation.

The level of theoretical difficulty in quantile aggregation leaves ample room for future adventures and challenges. For instance, the sharpness of convolution bounds under general conditions, other than those in Theorems 1, 2, A.1 and A.2, is an open question. For the interested reader, we connect our results to the theory of joint mixability in Appendix E, where many questions remain to be open. Additional information on the dependence structure, other than the marginal distributions, can be incorporated in the quantile aggregation problem, and it usually leads to highly challenging questions; see e.g., Bernard et al. (2017a,b) and Bartl et al. (2022). In view of the broad appearance of quantile aggregation, its application domain includes many problems in economics, finance, risk management, statistics, and scheduling, in addition to the two applications discussed Section 2. We mention some applications in Appendix G, on which many relevant questions warrant thorough future investigation.

A Lower convolution bounds

In this appendix, we quickly collect results on lower convolution bounds for $\inf_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu)$ and $\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^-(\nu)$, and some related results. The proofs of these results are symmetric to those on the upper convolution bounds, and they are omitted.

Theorem A.1 (RVaR aggregation). *Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$. For any t, s with $0 \leqslant t < t + s \leqslant 1$,*

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) \geqslant \sup_{\substack{\boldsymbol{\beta} \in (1-t)\Delta_n \\ \beta_0 \geqslant s > 0}} \sum_{i=1}^{n} R_{1-\beta_i-\beta_0,\beta_0}(\mu_i). \tag{A.1}$$

Moreover, (A.1) holds as an equality in the following cases:

(i) $t + s = 1$;

(ii) each of $\mu_1, \dots, \mu_n$ *admits an increasing density below its* $(1-t)$*-quantile;*

(iii) $\sum_{i=1}^{n} \mu_i \left[q_0^+(\mu_i), q_{1-t}^-(\mu_i) \right) \leqslant 1 - t$.

Let μ^{t-} be the probability measure given by

$$\mu^{t-}(-\infty, x] = \min\left\{ \frac{\mu(-\infty, x]}{t}, 1 \right\}, \quad x \in \mathbb{R}.$$

That is, μ^{t-} is the distribution measure of the random variable $q_V(\mu)$ where V is a uniform random variable on $[0, t]$. In the case of Theorem A.1 (iii), it equivalently means that each of $\mu_1^{(1-t)-}, \dots, \mu_n^{(1-t)-}$ admits an increasing density. We denote by $\boldsymbol{\mu}^{t-} = (\mu_1^{t-}, \dots, \mu_n^{t-})$. Proposition A.1 (symmetric to Proposition 1) shows relevant results.

Proposition A.1. *For $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$, for $0 \leqslant t < t + s \leqslant 1$, we have*

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) = \inf_{\nu \in \Lambda(\boldsymbol{\mu}^{(1-t)-})} \mathrm{ES}_{s/(1-t)}(\nu)$$

and

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^-(\nu) = \inf_{\nu \in \Lambda(\boldsymbol{\mu}^{t-})} q_1^-(\nu).$$

Similarly to the worst-case values, for the best-case values of RVaR aggregation, it suffices to consider the one ended at quantile level 1, i.e. the ES aggregation. In particular, for the worst-case problems of quantile aggregation, it suffices to consider the one at quantile level 1, i.e. the problems $\inf_{\nu \in \Lambda(\boldsymbol{\mu}^{t-})} q_1^+(\nu)$ for generic choices of $\boldsymbol{\mu}$.

Theorem A.2 (Quantile aggregation). *For $\boldsymbol{\mu} \in \mathcal{M}^n$, for $t \in (0, 1]$, we have*

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^-(\nu) \geqslant \sup_{\boldsymbol{\beta} \in t\Delta_n} \sum_{i=1}^{n} R_{1-\beta_i-\beta_0,\beta_0}(\mu_i). \tag{A.2}$$

Moreover, (A.2) holds as an equality in the following cases:

(i) $n \leqslant 2$;

(ii) each of $\mu_1, \dots, \mu_n$ *admits an increasing density below its* t*-quantile;*

(iii) each of $\mu_1, \dots, \mu_n$ admits a decreasing density below its t-quantile;

(iv) $\sum_{i=1}^n \mu_i \left[q_0^+(\mu_i), q_t^-(\mu_i)\right) \leqslant t$;

(v) $\sum_{i=1}^n \mu_i \left(q_0^+(\mu_i), q_t^-(\mu_i)\right] \leqslant t$.

Proposition A.2 (symmetric to Proposition 2) concerns a reduced lower convolution bound.

Proposition A.2. *For $\mu \in \mathcal{M}$ and $t \in (0,1]$, we have*

$$\inf_{\nu \in \Lambda_n(\mu)} q_t^-(\nu) \geqslant \sup_{\alpha \in (0, t/n)} nR_{1-t+(n-1)\alpha, t-n\alpha}(\mu) = \sup_{\alpha \in (0,t/n)} \frac{n}{t-n\alpha} \int_\alpha^{t-(n-1)\alpha} q_s^-(\mu) \mathrm{d}s. \tag{A.3}$$

Moreover, (A.3) holds as an equality if μ admits an increasing density below its t-quantile.

Proposition A.3 (symmetric to Proposition 4) shows that $\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_1^-(\nu)$ is always attainable and the infimum can be replaced by a minimum.

Proposition A.3. *For $\boldsymbol{\mu} \in \mathcal{M}^n$ and $t \in (0,1]$, there exists $\nu_- \in \Lambda(\boldsymbol{\mu})$ such that $\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^-(\nu) = q_t^-(\nu_-)$.*

Proposition A.4 (symmetric to Proposition 3) presents a lower convolution bound for multiplicative risks.

Proposition A.4. *For $\mu_1, \dots, \mu_n \in \mathcal{M}$ with support included in $(0, \infty)$, we have*

$$\inf_{X_i \sim \mu_i, i=1,\dots,n} q_t^- \left(\prod_{i=1}^n X_i \right) \geqslant \exp \left\{ \sup_{\boldsymbol{\beta} \in t\Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0, \beta_0} \left(\mu_i \circ \exp\right) \right\}, \quad t \in (0,1]. \tag{A.4}$$

Moreover, (A.4) holds as an equality in the following cases (denote by $f_1, \dots, f_n$ the densities of $X_1, \dots, X_n$):

(i) $n \leqslant 2$;

(ii) for each $i = 1, \dots, n$, $x \mapsto x f_i(x)$ is decreasing beyond the t-quantile of μ_i;

(iii) for each $i = 1, \dots, n$, $x \mapsto x f_i(x)$ is increasing beyond the t-quantile of μ_i.

(vi) $\sum_{i=1}^n \mu_i \left[q_0^+(\mu_i), q_t^-(\mu_i)\right) \leqslant t$;

(vii) $\sum_{i=1}^n \mu_i \left(q_0^+(\mu_i), q_t^-(\mu_i)\right] \leqslant t$.

Acknowledgements

We thank an Editor, an Associate Editor, three anonymous referees, Peng Liu, and Giovanni Puccetti for many helpful comments that have greatly improved the paper. JB and YL gratefully acknowledge financial support from the Air Force Office of Scientific Research under award number FA9550-20-1-0397, and additional support is gratefully acknowledged from NSF 1915967, 2118199, 2229011. YL acknowledges support from The Chinese University of Hong Kong, Shenzhen research startup fund (No.

UDF01003336) and Shenzhen Excellent Science and Technology Innovation Talents Development Plan (No. RCBS20231211090814028) and is partly supported by the Guangdong Provincial Key Laboratory of Mathematical Foundations for Artificial Intelligence (Grant No. 2023B1212010001). HL acknowledges support from the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280. RW acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2024-03728, CRC-2022-00141).

References

Aas, K. and Puccetti, G. (2014). Bounds on total economic capital: the DNB case study. *Extremes*, **17**, 693–715.

Altschuler, J. M. and Boix-Adserà, E. (2021). Hardness results for multimarginal optimal transport problems. *Discrete Optimization*, **42**, 100669.

Altschuler, J. M. and Boix-Adserà, E. (2023). Polynomial-time algorithms for multimarginal optimal transport problems with structure. *Mathematical Programming*, **199**, 1107–1178.

Bai, Y., Huang, Z. and Lam, H. (2020). A distributionally robust optimization approach to the NASA Langley uncertainty quantification challenge. In *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*.

Balci, O. and Sargent, R. G. (1982). Some examples of simulation model validation using hypothesis testing. In *Proceedings of the 14th Conference on Winter Simulation*, **2**, 621–629.

Bartl, D., Kupper, M., Lux, T. and Papapantoleon, A. (2022). Marginal and dependence uncertainty: bounds, optimal transport, and sharpness. *SIAM Journal on Control and Optimization*, **60**(1), 410–434.

Basak, S. and Shapiro, A. (2001). Value-at-Risk-based risk management: Optimal policies and asset prices. *Review of Financial Studies*, **14**(2), 371–405.

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B. and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, **59**(2), 341–357.

Ben-Tal, A., El Ghaoui, L. and Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press.

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

Bernard, C., Rüschendorf, L. and Vanduffel, S. (2017a). VaR bounds with variance constraint. *Journal of Risk and Insurance*, **84**(3), 923–959.

Bernard, C., Rüschendorf, L., Vanduffel, S. and Wang, R. (2017b). Risk bounds for factor models. *Finance and Stochastics*, **21**(3), 631–659.

Bertsimas, D., Brown, D. B. and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, **53**(3), 464–501.

Bertsimas, D., Gupta, V. and Kallus, N. (2018). Robust sample average approximation. *Mathematical Programming*, **171**(1-2), 217–282.

Bertsimas, D. and Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, **15**(3), 780–804.

Besser, K. and Jorswieck, E. A. (2020). Reliability bounds for dependent fading wireless channels. *IEEE Transactions*

on Wireless Communications, **59**(9), 5833–5845.

Blanchet, J., Glynn, P. W., Yan, J. and Zhou, Z. (2019a). Multivariate distributionally robust convex regression under absolute error loss. In *Advances in Neural Information Processing Systems*, 11817–11826.

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

Blanchet, J., Kang, Y. and Murthy, K. (2019b). Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, **56**(3), 830–857.

Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, **44**(2), 565–600.

Boerma, J., Tsyvinski, A. and Zimin, A. P. (2021). Sorting with team formation. *arXiv*: `2109.02730`.

Boerma, J., Tsyvinski, A., Wang, R. and Zhang, Z. (2023). Composite sorting. *arXiv*: `2303.06701`.

Boudt, K., Jakobsons, E., and Vanduffel, S. (2018). Block rearranging elements within matrix columns to minimize the variability of the row sums. *4OR - A Quarterly Journal of Operations Research*, **16**, 31–50.

Chen, W., Sim, M., Sun, J. and Teo, C. P. (2010). From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, **58**(2), 470–485.

Chen, Y., Liu, P., Liu, Y. and Wang, R. (2022). Ordering and inequalities of mixtures on risk aggregation. *Mathematical Finance*, **32**(1), 421–451.

Coffman, E. G. and Sethi, R. (1976). Algorithms minimizing mean flow time: schedule-length properties. *Acta Informatica*, **6**(1), 1–14.

Coffman, E. G. and Yannakakis, M. (1984). Permuting elements within columns of a matrix in order to minimize maximum row sum. *Mathematics of Operations Research*, **9**(3), 384–390.

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

de Castro, L. and Galvao, A. F. (2019). Dynamic quantile models of rational behavior. *Econometrica*, **87**, 1893–1939.

de Castro, L., Galvao, A. F., Noussair, C. and Qiao, L. (2022). Do people maximize quantiles? *Games and Economic Behavior*, **132**, 22–40.

Delage, E. and Mannor, S. (2010). Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research*, **58**(1), 203–213.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, **58**(3), 595–612.

Doan, X. V. and Natarajan, K. (2012). On the complexity of nonoverlapping multivariate marginal bounds for probabilistic combinatorial optimization problems. *Operations Research*, **60**(1), 138–149.

El Ghaoui, L., Oks, M. and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, **51**(4), 543–556.

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

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics*, **10**, 341–352.

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

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.

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

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.

Frey, J. C. and Kaplan, E. H. (2010). Queue inference from periodic reporting data. *Operations Research Letters*, **38**(5), 420–426.

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv*: 1604.02199.

Ghosh, S. and Lam, H. (2019). Robust analysis in stochastic simulation: Computation and performance guarantees. *Operations Research*, **67**(1), 232–249.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, **14**(1), 29–58.

Goeva, A., Lam, H., Qian, H. and Zhang, B. (2019). Optimization-based calibration of simulation input models. *Operations Research*, **67**(5), 1362–1382.

Goh, J. and Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations Research*, **58**(4-part-1), 902–917.

Grötschel, M., Lovász, L. and Schrijver, A. (1981). The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, **1**, 169–197.

Grötschel, M., Lovász, L. and Schrijver, A. (2012). *Geometric Algorithms and Combinatorial Optimization*, Vol. 2, Springer Science and Business Media.

Haus, U. (2015). Bounding stochastic dependence, complete mixability of matrices, and multidimensional bottleneck assignment problems. *Operations Research Letters*, **43**(1), 74–79.

Hsu, W.-L. (1984). Approximation algorithms for the assembly line crew scheduling problem. *Mathematics of Operations Research*, **9**(3), 376–383.

Hu, Z. and Hong, L. J. (2013). Kullback-leibler divergence constrained distributionally robust optimization. Available at *Optimization Online*. http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf

Jakobsons, E., Han, X. and Wang, R. (2016). General convex order on risk aggregation. *Scandinavian Actuarial Journal*, **2016**(8), 713–740.

Jiang, R. and Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming*, **158**(1-2), 291–327.

Kleijnen, J. P. C. (1995). Verification and validation of simulation models. *European Journal of Operational Research*, **82**(1), 145–162.

Kremer, M. (1993). The O-Ring theory of economic development. *Quarterly Journal of Economics*, **108**(3), 551–575.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.

Lam, H. (2016). Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, **41**(4), 1248–1275.

Lam, H. and Mottet, C. (2017). Tail analysis without parametric models: A worst-case perspective. *Operations Research*, **65**(6), 1696–1711.

Law, A. M. and Kelton, W. D. (2000). *Simulation Modeling and Analysis*, McGraw-Hill New York.

Li, B., Jiang, R. and Mathieu, J. L. (2019). Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming*, **173**(1-2), 151–192.

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

Makarov, G. D. (1981). Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and its Applications*, **26**, 803–806.

Mandelbaum, A. and Zeltyn, S. (1998). Estimating characteristics of queueing networks using transactional data. *Queueing Systems*, **29**(1), 75–127.

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

Mishra, V. K., Natarajan, K., Padmanabhan, D., Teo, C.-P. and Li, X. (2014). On theoretical and empirical aspects of marginal distribution choice models. *Management Science*, **173**(1-2), 151–192.

Nemirovski, A. and Shapiro, A. (2007). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, **17**(4), 969–996.

Papadimitriou, C. H. and Steiglitz, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Dover.

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

Popescu, I. (2005). A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, **30**(3), 632–657.

Puccetti, G., Rigo, P., Wang, B. and Wang, R. (2019). Centers of probability measures without the mean. *Journal of Theoretical Probability*, **32**(3), 1482–1501.

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

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

Puccetti, G. and Wang R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.

Ramdas, A. K., Barber, R. F., Wainwright, M. J. and Jordan, M. I. (2019). A unified treatment of multiple testing with prior knowledge using the p-filter. *Annals of Statistics*, **47**(5), 2790–2821.

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

Rostek, M. (2010). Quantile maximization in decision theory. *Review of Economic Studies*, **77**, 339–371.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, **14**(3), 623–632.

Rüschendorf, L. (2013). *Mathematical Risk Analysis*. Springer Series in Operations Research and Financial Engineering, Springer-Verlag Berlin Heidelberg.

Sargent, R. G. (2010). Verification and validation of simulation models. In *Proceedings of the 2010 winter simulation conference*, IEEE, 166–183.

Shapiro, A., Dentcheva, D. and Ruszczynski, A. (2021). *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics.

Van Parys, B. P. G., Goulart, P. J. and Kuhn, D. (2016). Generalized Gauss inequalities via semidefinite programming.

Mathematical Programming, **156**(1-2), 271–302.

Vrbik, J. (2018). Small-sample corrections to Kolmogorov–Smirnov test statistic. *Pioneer Journal of Theoretical and Applied Statistics*, **15**(1-2), 15–23.

Vovk, V., Wang, B. and Wang, R. (2022). Admissible ways of merging p-values under arbitrary dependence. *Annals of Statistics*, **50**(1), 351–375.

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

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, **102**(10), 1344–1360.

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

Wang, R., Bignozzi, V. and Tsakanas, A. (2015). How superadditive can a risk measure be? *SIAM Journal on Financial Mathematics*, **6**, 776–803.

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

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

Whitt, W. (1981). Approximating a point process by a renewal process: The view through a queue, an indirect approach. *Management Science*, **27**(6), 619–636.

Wiesemann, W., Kuhn, D. and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, **62**(6), 1358–1376.

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

Zymler, S., Kuhn, D. and Rustem, B. (2013). Worst-case value-at-risk of nonlinear portfolios. *Management Science*, **59**(1), 172–188.

Proposed E-Companion: Technical Appendices B-G

B Proofs of main results

B.1 Proofs in Section 4

We first present a lemma slightly generalizing the RVaR inequalities in Theorem 1 of Embrechts et al. (2018) to include distributions possibly with no finite mean.

Lemma EC.1. *Let $\alpha_1,\dots,\alpha_n,\beta_1,\dots,\beta_n\in[0,1]$. Denote by $b=\sum_{i=1}^n\beta_i$ and $a=\bigvee_{i=1}^n\alpha_i$. If $b+a\leqslant 1$, then for all $\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)\in\mathcal{M}^n$ and $\nu\in\Lambda(\boldsymbol{\mu})$,*

$$R_{b,a}(\nu)\leqslant\sum_{i=1}^n R_{\beta_i,\alpha_i}(\mu_i), \tag{EC.1}$$

provided the right-hand side of (EC.1) is well-defined (no "$\infty-\infty$").

Proof of Lemma EC.1. Theorem 1 of Embrechts et al. (2018) with the notation $\mathrm{RVaR}_{\beta,\alpha}(\mu)=R_{\beta,\alpha}(\mu)$ for $\alpha,\beta\geqslant 0$, $\alpha+\beta\leqslant 1$ gives (EC.1) if $\mu_1,\dots,\mu_n\in\mathcal{M}_1$. For $\mu_1,\dots,\mu_n$ that do not necessarily have finite means, we always assume that the right-hand side of (EC.1) is well-defined (no "$\infty-\infty$").

If there exists some i such that $R_{\beta_i,\alpha_i}(\mu_i)=\infty$, (EC.1) holds trivially. Now we assume $R_{\beta_i,\alpha_i}(\mu_i)<\infty$, $i=1,\dots,n$. There are four cases:

1. Suppose $b+a<1$ and $b>0$. In this case, $R_{b,a}$ and R_{β_i,α_i} are continuous with respect to weak convergence on $\mathcal{M}$ (see e.g. Cont et al. (2010)). For $\mu\in\Gamma(\mu_1,\dots,\mu_n)$ such that $\nu=\lambda_\mu$, we can find a sequence $\mu^{(k)}$, $k\in\mathbb{N}$ such that all one-dimensional margins of $\mu^{(k)}$ are in $\mathcal{M}_1$, and $\mu^{(k)}\to\mu$ weakly as $k\to\infty$. As a consequence, all one-dimensional margins of $\mu^{(k)}$, as well as its projection λ_{μ_k}, converge weakly. Since (EC.1) holds for probability measures in $\mathcal{M}_1$, using the continuity of $R_{b,a}$ and R_{β_i,α_i}, we know (EC.1) holds in this case.

2. Suppose $b+a=1$ and $b>0$. If $R_{1-a,a}(\nu)=-\infty$, (EC.1) holds trivially. If $R_{1-a,a}(\nu)>-\infty$, then

$$\lim_{\varepsilon\downarrow 0}R_{1-a,a-\varepsilon}(\nu)=R_{1-a,a}(\nu)$$

since $R_{1-a,a-\varepsilon}(\nu)$ is monotone for $\varepsilon\in(0,a)$. In the first case, we have shown, for $\varepsilon\in(0,\bigwedge_{i=1}^n\alpha_i)$,

$$R_{1-a,a-\varepsilon}(\nu)\leqslant\sum_{i=1}^n R_{\beta_i,\alpha_i-\varepsilon}(\mu_i).$$

Taking a limit as $\varepsilon\downarrow 0$ establishes (EC.1).

3. Suppose $b+a<1$ and $b=0$. It implies that $\beta_1=\cdots=\beta_n=0$. Because $R_{0,\alpha_i}(\mu_i)<\infty$, $i=1,\dots,n$,

we have

$$\lim_{\varepsilon \downarrow 0} R_{\varepsilon,\alpha_i}(\mu_i) = R_{0,\alpha_i}(\mu_i),$$

since $R_{\varepsilon,\alpha_i}(\mu_i)$ is monotone for $\varepsilon \in (0, 1-\alpha_i)$, $i = 1, \dots, n$. In the first case, we have shown, for $\varepsilon \in (0, 1-a)$,

$$R_{n\varepsilon,a}(\nu) \leqslant \sum_{i=1}^{n} R_{\varepsilon,\alpha_i}(\mu_i).$$

Taking a limit as $\varepsilon \downarrow 0$ establishes (EC.1).

4. Suppose $b + a = 1$ and $b = 0$. It implies that $\beta_1 = \cdots = \beta_n = 0$. Because $R_{0,\alpha_i}(\mu_i) < \infty$, $i = 1, \dots, n$, we know $\sum_{i=1}^{n} R_{0,1}(\mu_i)$ is well defined. By the linearity of $R_{0,1}$, we have

$$R_{0,1}(\nu) = \sum_{i=1}^{n} R_{0,1}(\mu_i) \leqslant \sum_{i=1}^{n} R_{0,\alpha_i}(\mu_i),$$

which establishes (EC.1). □

Proof of Theorem 1. The inequality (12) is shown in the text above Theorem 1. We proceed to prove the sharpness under the following cases.

(i) If $t = 0$, then $R_{t,s} = \mathrm{ES}_s$ and $\{\boldsymbol{\beta} \in (1-t)\Delta_n : \beta_0 \geqslant s\} = \{(s, 0, \dots, 0)\}$. It is well known (e.g., Kusuoka (2001)) that ES_s is subadditive and comonotonic additive, which gives

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} R_{0,s}(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu})} \mathrm{ES}_s(\nu) = \sum_{i=1}^{n} \mathrm{ES}_s(\mu_i) = \inf_{\substack{\boldsymbol{\beta} \in (0+s)\Delta_n \\ \beta_0 \geqslant s}} \sum_{i=1}^{n} R_{\beta_i,\beta_0}(\mu_i).$$

(ii) **Step 1**: Using Proposition 1 (which will be shown later), we have

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{(1-t-s)+})} \mathrm{LES}_{\frac{s}{t+s}}(\nu). \tag{EC.2}$$

Hence, it suffices to consider the problem of the right-hand side of (EC.2).

Step 2: Since each of $\mu_1, \dots, \mu_n$ admits a decreasing density beyond its $(1-t-s)$-quantile, each of the measures $\mu_1^{(1-t-s)+}, \dots, \mu_n^{(1-t-s)+}$ admits a decreasing density on its support. We can define an aggregate random variable T_{s_n} by (see Equation (3.4) of Jakobsons et al. (2016))

$$T_{s_n} = h(U)\mathbb{1}_{\{U \in (0,s_n)\}} + d(s_n)\mathbb{1}_{\{U \in [s_n,1]\}},$$

which will be explained below.

(a) We can write $T_{s_n} = \sum_{i=1}^{n} X_i$ where $X_i \sim \mu_i^{(1-t-s)+}$, $i = 1, \dots, n$. Let ν_0 be the distribution measure of T_{s_n}. Lemma 3.4 (c) of Jakobsons et al. (2016) gives $\nu_0 \in \Lambda(\boldsymbol{\mu}^{t+})$.

(b) U is a uniform random variable on $[0,1]$, $h, d : [0,1] \to \mathbb{R}$ are functions and $s_n \in [0,1]$ is a constant. They are given by:

$$h(x) = \sum_{i=1}^{n} y_i(x) - (n-1)y(x), \quad x \in (0,1),$$
$$d(x) = \frac{1}{1-x} \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{y_i(x)-y(x) \leqslant X_i \leqslant y_i(x)\}}\right], \quad x \in (0,1),$$
$$s_n = \inf\{x \in (0,1) : h(x) \leqslant d(x)\},$$

where $y, y_1, \ldots, y_n$ are functions on $(0,1)$ satisfying (see Equations (E1)-(E2) of Jakobsons et al. (2016))

$$\text{(E1)}: \quad \sum_{i=1}^{n} \mathbb{P}(X_i > y_i(x)) = x,$$
$$\text{(E2)}: \quad \mathbb{P}(y_i(x) - y(x) < X_i \leqslant y_i(x)) = 1 - x, \quad i = 1, \ldots, n.$$

(c) According to Lemma 3.2 of Jakobsons et al. (2016), h is a decreasing function on $(0, s_n)$. Hence, for all $u \in (0, s_n)$, we have $h(u) \geqslant d(s_n)$, and further $d(s_n) = q_0^+(\nu_0)$.

Step 3: Denote by $a = \min\{\frac{t}{t+s}, s_n\}$. We proceed to show

$$\text{LES}_{\frac{s}{t+s}}(\nu_0) = d(a). \tag{EC.3}$$

We verify this by direct computation. If $t/(t+s) \geqslant s_n$, then

$$\text{LES}_{\frac{s}{t+s}}(\nu_0) = \frac{1}{\frac{s}{t+s}} \mathbb{E}\left[T_{s_n} \mathbb{1}_{\{U \in [\frac{t}{t+s}, 1]\}}\right] = d(s_n);$$

if $t/(t+s) < s_n$, then

$$\begin{aligned}
\text{LES}_{\frac{s}{t+s}}(\nu_0) &= \frac{1}{\frac{s}{t+s}} \mathbb{E}\left[T_{s_n} \mathbb{1}_{\{U \in [\frac{t}{t+s}, 1]\}}\right] \\
&= \frac{1}{\frac{s}{t+s}} \left(\mathbb{E}\left[T_{s_n}\right] - \mathbb{E}\left[h(U) \mathbb{1}_{\{U \in (0, \frac{t}{t+s})\}}\right]\right) \\
&= \frac{1}{\frac{s}{t+s}} \left(\sum_{i=1}^{n} \mathbb{E}\left[X_i\right] - \sum_{i=1}^{n} \mathbb{E}\left[X_i (\mathbb{1}_{\{X_i > y_i(\frac{t}{t+s})\}} + \mathbb{1}_{\{X_i < y_i(\frac{t}{t+s}) - y(\frac{t}{t+s})\}})\right]\right) \\
&= \frac{1}{\frac{s}{t+s}} \sum_{i=1}^{n} \mathbb{E}\left[X_i \mathbb{1}_{\{y_i(\frac{t}{t+s}) - y(\frac{t}{t+s}) \leqslant X_i \leqslant y_i(\frac{t}{t+s})\}}\right] = d\left(\frac{t}{t+s}\right),
\end{aligned}$$

where the third equality is due to Lemma 3.3 of Jakobsons et al. (2016).

Step 4: We now show

$$d(a) = R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}), \tag{EC.4}$$

for some $\boldsymbol{\beta} \in (t+s)\Delta_n$ satisfying $\beta_0 \geqslant s$, which is defined by

$$\beta_0 = (t+s)(1-a), \quad \beta_i = (t+s)\mu_i^{(1-t-s)+}(y_i(a), \infty) = \mu_i(y_i(a), \infty), \ i = 1, \dots, n.$$

According to (E1), $\sum_{i=1}^n \beta_i = (t+s)a$. We have $(\beta_0, \beta_1, \dots, \beta_n) \in (t+s)\Delta_n$, and $\beta_0 \geqslant s$. Hence,

$$\begin{aligned}
d(a) &= \sum_{i=1}^n \frac{1}{1-a} \mathbb{E}\left[X_i \mathbb{1}_{\{y_i(a)-y(a) \leqslant X_i \leqslant y_i(a)\}}\right] \\
&= \sum_{i=1}^n \frac{1}{1-a} \int_{y_i(a)-y(a)}^{y_i(a)} x \mu_i^{(1-t-s)+}(\mathrm{d}x) \\
&= \sum_{i=1}^n \frac{1}{1-a} \int_{a-\frac{\beta_i}{t+s}}^{1-\frac{\beta_i}{t+s}} q_u^-(\mu_i^{(1-t-s)+}) \mathrm{d}u \\
&= \sum_{i=1}^n \frac{1}{1-a} \int_{a-\frac{\beta_i}{t+s}}^{1-\frac{\beta_i}{t+s}} q^-_{1-t-s+(t+s)u}(\mu_i) \mathrm{d}u \\
&= \sum_{i=1}^n \frac{1}{(1-a)(t+s)} \int_{1-\beta_i-\beta_0}^{1-\beta_i} q_v^-(\mu_i) \mathrm{d}v \\
&= \sum_{i=1}^n \frac{1}{\beta_0} \int_{1-\beta_i-\beta_0}^{1-\beta_i} q_v^-(\mu_i) \mathrm{d}v = \sum_{i=1}^n R_{\beta_i, \beta_0}(\mu_i),
\end{aligned}$$

where the third equality is due to the fact that $\mu_i^{(1-t-s)+}(-\infty, y_i(a) - y(a)] = a - \frac{\beta_i}{t+s}$ derived from (E2).

Step 5: Combining (EC.2), (EC.3) and (EC.4), we have

$$\begin{aligned}
\sup_{\nu \in \Lambda(\boldsymbol{\mu})} R_{t,s}(\nu) = \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{(1-t-s)+})} \mathrm{LES}_{\frac{s}{t+s}}(\nu) &\geqslant \mathrm{LES}_{\frac{s}{t+s}}(\nu_0) \\
&= d(a) = \sum_{i=1}^n R_{\beta_i, \beta_0}(\mu_i) \geqslant \inf_{\substack{\boldsymbol{\beta}' \in (t+s)\Delta_n \\ \beta_0' \geqslant s}} \sum_{i=1}^n R_{\beta_i', \beta_0'}(\mu_i).
\end{aligned}$$

Thus, the bound (12) is sharp.

(iii) It suffices to prove the statement for $t = 1 - s$. The assumption $\sum_{i=1}^n \mu_i(q_0^+(\mu_i), q_1^-(\mu_i)] \leqslant 1$ allows for the existence of a lower mutually exclusive (see Definition EC.1 below) random vector $(X_1, \dots, X_n)$ where $X_i \sim \mu_i$, $i = 1, \dots, n$. Hence, the desired result follows from Lemma EC.2 below by checking that the bound (12) is attained by such a vector. □

Definition EC.1 (Mutually exclusivity)**.** We say that a random vector $(X_1, \dots, X_n)$ where $X_i \sim \mu_i$, $i = 1, \dots, n$ is *lower mutually exclusive* if $\mathbb{P}(X_i > q_0^+(\mu_i), X_j > q_0^+(\mu_j)) = 0$ for all $i \neq j$ and it is *upper mutually exclusive* if $\mathbb{P}(X_i < q_1^-(\mu_i), X_j < q_1^-(\mu_j)) = 0$ for all $i \neq j$.

Lemma EC.2. *If random variables $X_1, \dots, X_n$ are lower mutually exclusive and bounded from below, then*

for $\alpha \in (0,1)$,

$$R_{1-\alpha,\alpha}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} R_{\beta_i,\alpha}(X_i), \tag{EC.5}$$

for some $\beta_1, \dots, \beta_n \in [0,1)$ *with* $\sum_{i=1}^{n} \beta_i = 1-\alpha$.

Proof. Without loss of generality, we assume $q_0^+(\mu_i) = 0$ for each i. If $q_\alpha^+\left(\sum_{i=1}^n X_i\right) = 0$, then $\sum_{i=1}^n \mathbb{P}(X_i > 0) = \mathbb{P}\left(\sum_{i=1}^n X_i > 0\right) \leqslant 1-\alpha$. Hence, we can choose $\beta_i \geqslant \mathbb{P}(X_i > 0)$ for each i, and both sides of (EC.5) are 0. Below we assume $q_\alpha^+(\sum_{i=1}^n X_i) > 0$.

First, we assume that the distribution μ_i of X_i is continuous on $\{X_i > 0\}$ for each $i = 1, \dots, n$, and so is the conditional distribution of $\sum_{i=1}^n X_i$ on $\{\sum_{i=1}^n X_i > 0\}$.

Let $y = q_\alpha^+\left(\sum_{i=1}^n X_i\right)$ and $A = \{\sum_{i=1}^n X_i \leqslant y\}$. We have $\mathbb{P}(A) = \alpha$. For each $i = 1, \dots, n$, let $\alpha_i = \mathbb{P}(A \cap \{X_i > 0\}) = \mathbb{P}(0 < X_i \leqslant y)$ and $t_i = \mathbb{P}(X_i > 0)$. By direct calculation

$$\begin{aligned}
R_{1-\alpha,\alpha}\left(\sum_{i=1}^{n} X_i\right) = \mathbb{E}\left[\sum_{i=1}^{n} X_i \mid A\right] &= \sum_{i=1}^{n} \mathbb{E}\left[X_i \mid A\right] \\
&= \frac{1}{\alpha}\sum_{i=1}^{n}\left(0 + \mathbb{E}\left[X_i \mathbb{1}_{A\cap\{X_i>0\}}\right]\right) \\
&= \frac{1}{\alpha}\sum_{i=1}^{n}\left(\int_{1-t_i+\alpha_i-\alpha}^{1-t_i} q_u^+(\mu_i)\mathrm{d}u + \int_{1-t_i}^{1-t_i+\alpha_i} q_u^+(\mu_i)\mathrm{d}u\right) \\
&= \sum_{i=1}^{n} R_{t_i-\alpha_i,\alpha}(X_i).
\end{aligned}$$

We can check, by lower mutual exclusivity and the continuity assumption, that

$$\sum_{i=1}^{n}(t_i - \alpha_i) = \sum_{i=1}^{n}\left(\mathbb{P}(X_i > 0) - \mathbb{P}(0 < X_i \leqslant y)\right) = \sum_{i=1}^{n}\mathbb{P}(X_i > y) = \mathbb{P}\left(\sum_{i=1}^{n} X_i > y\right) = 1-\alpha.$$

By (12), we have

$$R_{1-\alpha,\alpha}\left(\sum_{i=1}^{n} X_i\right) \leqslant \sum_{i=1}^{n} R_{t_i-\alpha_i,\alpha}(X_i).$$

Therefore, (EC.5) holds by choosing $\beta_i = t_i - \alpha_i$, $i = 1, \dots, n$. In case the conditional distributions of $X_1, \dots, X_n$ are positive and not continuous, we can approximate (by convergence in distribution) $X_1, \dots, X_n$ by conditionally continuous distributions while fixing $\mathbb{P}(X_i > 0)$ for each i. The compactness of the set $(1-\alpha)\overline{\Delta}_{n-1}$ on which $(\beta_1, \dots, \beta_n)$ takes values and the continuity of $R_{\beta,\alpha}$ with respect to weak convergence (e.g., Cont et al. (2010)) yields the desirable result. □

Proof of Proposition 1. The first equality is a direct consequence of Theorem 4.1 and Example 6.3 of Liu and Wang (2021), and the second equality follows from Remark 4.1 of Liu and Wang (2021). □

B.2 Proofs in Section 5

Proof of Theorem 2. The convolution bound (14) is obtained by taking a limit of (12) in Theorem 2 using (7). Similarly, based on Theorem 2 and the fact that R_{β_i,β_0} is continuous in β_0, this limit argument also gives sharpness in (i), (ii) and (iv). Next we proceed to show sharpness in (iii) and (v).

(iii) First, we note that

$$\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_t^+(\nu) = -\inf_{\tilde{\nu}\in\Lambda(\tilde{\boldsymbol{\mu}})} q_{1-t}^-(\tilde{\nu}), \tag{EC.6}$$

where $\tilde{\mu}_i$ is the distribution measure of the random variable $-X_i$ with $X_i \sim \mu_i$, $i = 1,\dots,n$ and $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1,\dots,\tilde{\mu}_n)$. The fact that each of $\mu_1,\dots,\mu_n$ admits an increasing density beyond its t-quantile implies that each of $\tilde{\mu}_1,\dots,\tilde{\mu}_n$ admits an decreasing density below its $(1-t)$-quantile. Note that a distribution that has a decreasing density below its $(1-t)$-quantile is supported in either a finite interval $[a,b]$ or a half real line $[a,\infty)$ for some $a,b\in\mathbb{R}$. Hence, without loss of generality, we can assume $q_0^+(\tilde{\mu}_i) = 0$, $i = 1,\dots,n$.

For sharpness of (14), we need to show

$$\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_t^+(\nu) \geqslant \inf_{\boldsymbol{\beta}\in(1-t)\Delta_n} \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i).$$

By (EC.6) and the definition of $R_{\beta,\alpha}$, it suffices to show

$$\inf_{\tilde{\nu}\in\Lambda(\tilde{\boldsymbol{\mu}})} q_{1-t}^-(\tilde{\nu}) \leqslant \sup_{\boldsymbol{\beta}\in(1-t)\Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0,\beta_0}(\tilde{\mu}_i). \tag{EC.7}$$

Fix $j\in\{1,\dots,n\}$ and $\beta_j\in(0,1-t)$. By taking $\beta_i = 0$ for $i\in\{1,\dots,n\}\backslash\{j\}$ and $\beta_0 = 1-t-\beta_j$, we get

$$\sup_{\boldsymbol{\beta}\in(1-t)\Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0,\beta_0}(\tilde{\mu}_i) \geqslant R_{t,1-t-\beta_j}(\tilde{\mu}_j) + \sum_{i\neq j} R_{t+\beta_j,1-t-\beta_j}(\tilde{\mu}_i) \geqslant R_{t,1-t-\beta_j}(\tilde{\mu}_j).$$

Taking a supremum over $\beta_j\in(0,1-t)$ and $j\in\{1,\dots,n\}$ yields

$$\sup_{\boldsymbol{\beta}\in(1-t)\Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0,\beta_0}(\tilde{\mu}_i) \geqslant \bigvee_{j=1}^n \sup_{\beta_j\in(0,1-t)} R_{t,1-t-\beta_j}(\tilde{\mu}_j) = \bigvee_{j=1}^n q_{1-t}^-(\tilde{\mu}_j). \tag{EC.8}$$

If $\bigvee_{j=1}^n q_{1-t}^-(\tilde{\mu}_j) = \infty$, then the right-hand side of (EC.7) is ∞, which holds automatically. If $\bigvee_{j=1}^n q_{1-t}^-(\tilde{\mu}_j) < \infty$, we can apply Corollary 4.7 of Jakobsons et al. (2016), using the condition that each of $\mu_1,\dots,\mu_n$ admits a decreasing density below its $(1-t)$-quantile. This gives

$$\inf_{\tilde{\nu}\in\Lambda(\tilde{\boldsymbol{\mu}})} q_{1-t}^+(\tilde{\nu}) = \max\left\{\bigvee_{i=1}^n q_{1-t}^-(\tilde{\mu}_i), \sum_{i=1}^n R_{t,1-t}(\tilde{\mu}_i)\right\}. \tag{EC.9}$$

Also note that in this case, $R_{t,1-t}(\tilde{\mu}_i) < \infty$, $i = 1, \dots, n$, and hence

$$\sup_{\boldsymbol{\beta} \in (1-t)\Delta_n} \sum_{i=1}^{n} R_{1-\beta_i-\beta_0,\beta_0}(\tilde{\mu}_i) \geqslant \sum_{i=1}^{n} R_{t,1-t}(\tilde{\mu}_i). \tag{EC.10}$$

Combining (EC.8)-(EC.10), we get (EC.7).

(v) It suffices to prove the case $t = 0$. The assumption $\sum_{i=1}^{n} \mu_i[q_0^+(\mu_i), q_1^-(\mu_i)) \leqslant 1$ allows for the existence of an upper mutually exclusive (see Definition EC.1) random vector $(X_1, \dots, X_n)$ where $X_i \sim \mu_i$, $i = 1, \dots, n$. Hence, we have

$$q_0^+\left(\sum_{i=1}^{n} X_i\right) = \min_{1 \leqslant i \leqslant n} \left(q_0^+(\mu_i) + \sum_{j \neq i} q_1^-(\mu_j)\right) \geqslant \inf_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^{n} R_{\beta_i,\beta_0}(\boldsymbol{\mu}).$$

Hence, the desired result follows as the bound (14) is attained by such a vector. □

Proof of Proposition 2. Letting $\beta_1 = \cdots = \beta_n = \alpha$ in Theorems 2 and A.2, we immediately get (15). We show (15) holds as an equality in this case of decreasing density. Note that the second equality in (15) is simply the definition. By Proposition 1 of Embrechts et al. (2014), $\sup_{\nu \in \Lambda_n(\mu)} q_t^+(\nu)$ is equal to n times the conditional mean of μ on an interval $[t + (n-1)\alpha, 1 - \alpha]$ for some $\alpha \in [0, \frac{1-t}{n}]$. Therefore,

$$\sup_{\nu \in \Lambda_n(\mu)} q_t^+(\nu) \geqslant \inf_{\alpha \in (0, \frac{1-t}{n})} nR_{\alpha,1-t-n\alpha}(\mu).$$

Also note that the "$\leqslant$" sign in (15) is implied by Proposition 2. Hence, (15) holds as an equality in this case. □

Proof of Proposition 3. For any fixed $X_i \sim \mu_i$, $i = 1, \dots, n$, we define $Y_i = \log(X_i)$ and hence have

$$q_t^+\left(\prod_{i=1}^{n} X_i\right) = q_t^+\left(\exp\left\{\sum_{i=1}^{n} Y_i\right\}\right) = \exp\left\{q_t^+\left(\sum_{i=1}^{n} Y_i\right)\right\}.$$

We obtain the desired results by investigating the corresponding quantile problem $q_t^+\left(\sum_{i=1}^{n} Y_i\right)$. We only prove the cases (ii)-(iii). Denote by f_i the density of X_i. The density of Y_i at $y \in \mathbb{R}$ is $\exp(y) \cdot f_i(\exp(y))$. According to Theorem 2, (16) is sharp if $\exp(y) \cdot f_i(\exp(y))$ are all decreasing (resp. increasing) beyond the t-quantile of Y_i for all $i = 1, \dots, n$. With a change of variables (and the fact that log is strictly increasing), the condition is translated into that the functions $x \cdot f_i(x)$ are all decreasing (resp. increasing) beyond the t-quantile of X_i for all $i = 1, \dots, n$. □

Proof of Proposition 4. The statement for q_t^+, $t \in (0, 1)$, is shown in Lemma 4.2 of Bernard et al. (2014). The case of $t = 0$ follows from the same argument by noting the upper semicontinuity of q_0^+. □

Proof of Proposition 5. To show the "$\geqslant$" direction of (17), we note that for any $\nu_Y \in \Lambda(\boldsymbol{\mu}^{[\boldsymbol{m}]})$ such that $Y_1 \sim \mu_1^{[m]}, \dots, Y_n \sim \mu_n^{[m]}$ and $\sum_{i=1}^{n} Y_i \sim \nu_Y$, by letting $X_i = q_{U_{Y_i}}^-(\mu_i)$, $i = 1, \dots, n$, we get $X_i \sim \mu_i$ and

$X_i \geqslant Y_i$ for each $i = 1, \dots, n$. Denote by ν_X the distribution measure of $\sum_{i=1}^n X_i$. It follows that $\nu_X \in \Lambda(\boldsymbol{\mu})$ and $q_t^+(\nu_X) \geqslant q_t^+(\nu_Y)$, which gives

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu) \geqslant \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{[m]})} q_t^+(\nu).$$

To show the "$\leqslant$" direction of (17), for any $\nu_X \in \Lambda(\boldsymbol{\mu})$ such that random variables $X_1 \sim \mu_1, \dots, X_n \sim \mu_n$ and $\sum_{i=1}^n X_i \sim \nu_X$, let $Y_i = X_i \wedge m$, $i = 1, \dots, n$. Write $S_X = \sum_{i=1}^n X_i$ and $S_Y = \sum_{i=1}^n Y_i$. Denote by ν_Y the distribution measure of S_Y. We have $\nu_Y \in \Lambda(\boldsymbol{\mu}^{[m]})$. By Corollary 1 of Embrechts et al. (2018), we have, for $\varepsilon > 0$,

$$q^-_{t+\varepsilon}(\nu_X) \leqslant \sum_{i=1}^n q^-_{1-(1-t-\varepsilon)/n}(\mu_i).$$

Taking a limit of the above equation as $\varepsilon \downarrow 0$, we obtain

$$q_t^+(\nu_X) \leqslant \sum_{i=1}^n q^+_{1-(1-t)/n}(\mu_i) \leqslant m. \tag{EC.11}$$

It is clear that $(S_X \wedge m) \leqslant S_Y$ because the real function $x \mapsto x \wedge m$ is subadditive. Denote by $\tilde{\nu}$ the distribution of $S_X \wedge m$. Hence, (EC.11) implies

$$q_t^+(\nu_X) = q_t^+(\tilde{\nu}) \leqslant q_t^+(\nu_Y). \tag{EC.12}$$

Taking a supremum of (EC.12) over all possible choices of $\nu_X \in \Lambda(\boldsymbol{\mu})$, we get

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_t^+(\nu) \leqslant \sup_{\nu \in \Lambda(\boldsymbol{\mu}^{[m]})} q_t^+(\nu).$$

This completes the proof. □

Proof of Proposition 6. It is a direct corollary by letting $t \downarrow 0$ in Theorem 2 and $t \uparrow 1$ in Theorem A.2 respectively. □

Proof of Proposition 7. Note that

$$\begin{aligned}\sup_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0,\beta_0}(\mu_i) &\geqslant \lim_{\varepsilon \downarrow 0} \sum_{i=1}^n R_{(n-1)\varepsilon, 1-n\varepsilon}(\mu_i) \\ &= \sum_{i=1}^n R_{0,1}(\mu_i) = \lim_{\varepsilon \downarrow 0} \sum_{i=1}^n R_{\varepsilon, 1-n\varepsilon}(\mu_i) \geqslant \inf_{\boldsymbol{\beta} \in \Delta_n} \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i).\end{aligned}$$

Hence, the second and the third inequalities in (20) hold. The first and the last inequalities are due to Proposition 6. □

B.3 Proofs in Section 6.2

Proof of Theorem 3. By assumption and Proposition 4, there exists $\nu_+ \in \Lambda(\boldsymbol{\mu})$ such that

$$q_0^+(\nu_+) = \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) = R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}).$$

Take $X_i^* \sim \mu_i$, $i = 1, \dots, n$ such that $\sum_{i=1}^n X_i^* \sim \nu_+$ and $\sum_{i=1}^n X_i^* \geqslant q_0^+(\nu_+)$ almost surely. We divide the proof into several steps. We first prove the properties of X_i^* in (22) in Steps 1-3 and the feasibility of (23) and its optimality given the above sufficient condition in Steps 4-5. In Steps 1-3, we will show that the probability space Ω is divided into $\Omega = A_1 \cup \cdots \cup A_n \cup B$ where A_i is defined by $A_i = \{X_i^* > q_{1-\beta_i}^-(\mu_i)\}$ ("right-tail" parts of $X_1, \dots, X_n$) and $B^c = \bigcup_{i=1}^n A_i$, with the following properties:

(a) on the set B, $X_i^* \sim \mu_i^{[1-\beta_0-\beta_i, 1-\beta_i]}$ for all $i = 1, \dots, n$ and $\sum_{i=1}^n X_i^* = q_0^+(\nu_+)$ almost surely;

(b) for any fixed $i = 1, \dots, n$, on the set A_i, $X_i^* \sim \mu_i^{(1-\beta_i, 1]}$ and $X_j^* \sim \mu_j^{[0, 1-\beta_0-\beta_j)}$ for all $j \neq i$.

Step 1: We show that the set $\left\{\sum_{i=1}^n X_i^* = q_0^+(\nu_+)\right\}$ has probability no less than β_0. By (10), we have

$$q_0^+(\nu_+) \leqslant R_{1-\beta_0, \beta_0}(\nu_+) \leqslant R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) = q_0^+(\nu_+), \tag{EC.13}$$

and hence all inequalities in (EC.13) are equalities. The fact that $q_0^+(\nu_+) = R_{1-\beta_0, \beta_0}(\nu_+)$ implies that $q_t^-(\nu_+) = q_0^+(\nu_+)$ for all $t \in (0, \beta_0]$ and the set $\left\{\sum_{i=1}^n X_i^* = q_0^+(\nu_+)\right\}$ has probability no less than β_0.
Step 2: We proceed to show that the events ("body" parts of $X_1, \dots, X_n$)

$$\{q_{1-\beta_0-\beta_i}^-(\mu_i) \leqslant X_i^* \leqslant q_{1-\beta_i}^-(\mu_i)\},\ i = 1, \dots, n, \tag{EC.14}$$

are identical and $\sum_{i=1}^n X_i^* = q_0^+(\nu_+)$ almost surely on this set.

As the events $A_i = \{X_i^* > q_{1-\beta_i}^-(\mu_i)\}$, $i = 1, \dots, n$, and $B^c = \cup_{i=1}^n A_i$, we have $\mathbb{P}(B^c) \leqslant \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) = \sum_{i=1}^n \beta_i = 1 - \beta_0$. Denote by $\kappa_i \in \mathcal{M}$ the distribution measure of $T_i = X_i^* \mathbb{1}_{A_i^c} + m \mathbb{1}_{A_i}$, $i = 1, \dots, n$, where m is a real number satisfying that $m < \min_{1 \leqslant i \leqslant n} q_{1-\beta_0-\beta_i}^-(\mu_i)$. Denote by τ the distribution measure of the sum variable $\sum_{i=1}^n T_i$. It is verified that κ_i has a finite mean and $R_{\beta_i, \beta_0}(\mu_i) = \mathrm{ES}_{\beta_0}(\kappa_i)$, $i = 1, \dots, n$. We first prove that

$$q_t^-(\tau) \geqslant q_{t-1+\beta_0}^-(\nu_+), \quad t \in (1 - \beta_0, 1]. \tag{EC.15}$$

Fix $t \in (1 - \beta_0, 1]$. We have $\sum_{i=1}^n T_i \mathbb{1}_B = \sum_{i=1}^n X_i^* \mathbb{1}_B$ and for any $x \in \mathbb{R}$,

$$\begin{aligned}\tau(x, \infty) = \mathbb{P}\left(\sum_{i=1}^n T_i > x\right) &\geqslant \mathbb{P}\left(\sum_{i=1}^n X_i^* > x, B\right) \geqslant \mathbb{P}\left(\sum_{i=1}^n X_i^* > x\right) - \mathbb{P}(B^c) \\ &\geqslant \mathbb{P}\left(\sum_{i=1}^n X_i^* > x\right) - 1 + \beta_0 = \nu_+(x, \infty) - 1 + \beta_0.\end{aligned}$$

For any $x < q_{t-1+\beta_0}^-(\nu_+)$, we have $\nu_+(-\infty, x] < t - 1 + \beta_0$, and $\tau(-\infty, x] \leqslant \nu_+(-\infty, x] + 1 - \beta_0 < t$ and

then $x < q_t^-(\tau)$. Hence we have $q_t^-(\tau) \geqslant q_{t-1+\beta_0}^-(\nu_+)$ and prove (EC.15). Thus, it follows from sharpness of (EC.13) that

$$\begin{aligned}\sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i) = \sum_{i=1}^n \mathrm{ES}_{\beta_0}(\kappa_i) &\geqslant \mathrm{ES}_{\beta_0}(\tau)\\ &= \frac{1}{\beta_0}\int_{1-\beta_0}^1 q_t^-(\tau)\mathrm{d}t\\ &\geqslant \frac{1}{\beta_0}\int_{1-\beta_0}^1 q_{t-1+\beta_0}^-(\nu_+)\mathrm{d}t\\ &= \frac{1}{\beta_0}\int_0^{\beta_0} q_t^-(\nu_+)\mathrm{d}t = R_{1-\beta_0,\beta_0}(\nu_+) = \sum_{i=1}^n R_{\beta_i,\beta_0}(\mu_i),\end{aligned} \tag{EC.16}$$

where the first inequality is the well-known subadditivity of ES_{β_0}. Thus, all inequalities in (EC.16) are sharp.

The fact that the first inequality in (EC.16) is sharp implies that T_i, $i = 1,\dots,n$ share the same tail event with probability β_0 according to Theorem 5 in Wang and Zitikis (2021), i.e., the sets $\{T_i \geqslant q_{1-\beta_0}^-(\kappa_i)\} = \{q_{1-\beta_0-\beta_i}^-(\mu_i) \leqslant X_i^* \leqslant q_{1-\beta_i}^-(\mu_i)\}$, $i = 1,\dots,n$, (also in (EC.14)) are identical and have probability β_0. We denote this set by B'. Furthermore, B' does not intersect any A_i, $i = 1,\dots,n$.

We write $Y_i = X_i^*|_{B'}$. Hence $Y_i \sim \mu^{[1-\beta_0-\beta_i,1-\beta_i]}$ and $\sum_{i=1}^n Y_i = \sum_{i=1}^n X_i^*$ on the set B' and $q_0^+\left(\sum_{i=1}^n Y_i\right) = \mathbb{E}\left[\sum_{i=1}^n Y_i\right] = R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu})$. Thus, $\sum_{i=1}^n Y_i = R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu})$ almost surely.

Step 3: We proceed to show that the events A_i, $i = 1,\dots,n$, are mutually disjoint. We can calculate

$$\frac{\partial}{\partial \beta_i'} R_{\boldsymbol{\beta}'}^+(\boldsymbol{\mu}) = \frac{1}{\beta_0'}\left(R_{\boldsymbol{\beta}'}^+(\boldsymbol{\mu}) - q_{1-\beta_i'}^-(\mu_i) - \sum_{j\neq i} q_{1-\beta_0'-\beta_j'}^-(\mu_j)\right),\ \boldsymbol{\beta}' \in \Delta_n.$$

The first-order condition from the optimality of $\boldsymbol{\beta}$ reads as

$$\begin{cases} R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) - q_{1-\beta_i}^-(\mu_i) - \sum_{j\neq i} q_{1-\beta_0-\beta_j}^-(\mu_j) = 0, \text{ if } \beta_0 > 0 \text{ and } i \in \{1,\dots,n\} \text{ satisfying } \beta_i \neq 0;\\ R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) - q_1^-(\mu_i) - \sum_{j\neq i} q_{1-\beta_0-\beta_j}^-(\mu_j) \geqslant 0, \text{ if } \beta_0 > 0 \text{ and } i \in \{1,\dots,n\} \text{ satisfying } \beta_i = 0;\\ R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) - \sum_{j=1}^n q_{1-\beta_j}^-(\mu_j) = 0, \text{ if } \beta_0 = 0. \end{cases} \tag{EC.17}$$

Denote the sets (the "left-tail" parts of $X_1,\dots,X_n$) by $C_i = \{X_i < q_{1-\beta_0-\beta_i}^-(\mu_i)\}$, $i = 1,\dots,n$. We have a partition $\Omega = A_i \cup B' \cup C_i$ and $\mathbb{P}(C_i) = 1-\beta_0-\beta_i$, $i = 1,\dots,n$. (EC.17) shows that $\mathbb{P}\left(\cap_{j=1}^n C_j\right) = 0$ because for any $\omega \in \cap_{j=1}^n C_j$, for any fixed $i \in \{1,\dots,n\}$,

$$\sum_{j=1}^n X_j^*(\omega) < q_{1-\beta_i}^-(\mu_j) + \sum_{j\neq i} q_{1-\beta_0-\beta_j}^-(\mu_j) \leqslant R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) = q_0^+\left(\sum_{j=1}^n X_j^*\right).$$

Arguing by contradiction that there exists $1 \leqslant k < l \leqslant n$ such that $\mathbb{P}(A_k \cap A_l) > 0$. For any fixed

$i \in \{1,\dots,n\}\backslash\{k,l\}$, we have

$$\begin{aligned}
\mathbb{P}(C_i) &= \mathbb{P}\left(\cap_{j=1}^n C_j\right) + \mathbb{P}\left(\cup_{j\neq i}(C_i\cap A_j)\right)\\
&= \mathbb{P}\left(\cap_{j=1}^n C_j\right) + \mathbb{P}\left(\cup_{j\neq i,k,l}(C_i\cap A_j)\cup(C_i\cap A_k\cap A_l^c)\cup(C_i\cap A_k^c\cap A_l)\cup(C_i\cap A_k\cap A_l)\right)\\
&\leqslant \mathbb{P}\left(\cap_{j=1}^n C_j\right) + \sum_{j\neq i,k,l}\mathbb{P}(A_j) + \mathbb{P}(A_k\cap A_l^c) + \mathbb{P}(A_k\cap A_l^c) + \mathbb{P}(A_k\cap A_l)\\
&= \mathbb{P}\left(\cap_{j=1}^n C_j\right) + \sum_{j\neq i,k,l}\mathbb{P}(A_j) + \mathbb{P}(A_k) + \mathbb{P}(A_l) - \mathbb{P}(A_k\cap A_l)\\
&= \mathbb{P}\left(\cap_{j=1}^n C_j\right) + 1 - \beta_0 - \beta_i - \mathbb{P}(A_k\cap A_l)\\
&= \mathbb{P}\left(\cap_{j=1}^n C_j\right) + \mathbb{P}(C_i) - \mathbb{P}(A_k\cap A_l).
\end{aligned}$$

Hence $\mathbb{P}\left(\cap_{i=1}^n C_i\right) \geqslant \mathbb{P}(A_k\cap A_l) > 0$, which leads to a contradiction. Thus, $A_1,\dots,A_n$ are mutually disjoint and $\mathbb{P}(B^c) = \mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i) = 1-\beta_0$. As the set B' does not intersect B^c and $\mathbb{P}(B') = \beta_0$, we know $B' = B$ and thus the partition $\Omega = A_1\cup\cdots\cup A_n\cup B$. This completes the first statement in the theorem on the properties of X_i^* in (22).

Step 4: If $\beta_0 = 1$, we have that $\boldsymbol{\mu}$ is jointly mixable. If $\beta_0 < 1$, we check that the corresponding X_i^* given by (23) has distribution μ_i, $i = 1,\dots,n$. For each $i = 1,\dots,n$, if $x < q^-_{1-\beta_0-\beta_i}(\mu_i)$, we have

$$\begin{aligned}
\mathbb{P}(X_i^* \leqslant x) &= \mathbb{P}(U < 1-\beta_0, K\neq i, q^-_{\frac{1-\beta_0-\beta_i}{1-\beta_0}U}(\mu_i) \leqslant x)\\
&= \mathbb{P}(K\neq i)\mathbb{P}(q^-_{\frac{1-\beta_0-\beta_i}{1-\beta_0}U}(\mu_i) \leqslant x)\\
&= \frac{1-\beta_0-\beta_i}{1-\beta_0}\frac{1-\beta_0}{1-\beta_0-\beta_i}\mu_i(-\infty,x] = \mu_i(-\infty,x].
\end{aligned}$$

One can similarly check that $\mathbb{P}(X_i^* > x) = \mu_i(x,\infty)$ if $x > q^-_{1-\beta_i}(\mu_i)$ and $\mathbb{P}(X_i^* \leqslant x) = \mu_i(-\infty,x]$ if $q^-_{1-\beta_0-\beta_i}(\mu_i) \leqslant x \leqslant \mu^-_{1-\beta_i}(\mu_i)$. Hence $X_i^* \sim \mu_i$, $i = 1,\dots,n$.

Step 5: We finally show that if $\beta_0 < 1$, $\beta_1,\dots,\beta_n > 0$ and the minimum of each of the functions $h_1,\dots,h_n$ is attained at $x = 1-\beta_0$, then $(X_1^*,\dots,X_n^*)$ in (23) attains the maximum of q_0^+ for $\boldsymbol{\mu}$. According to the first-order condition (EC.17), we have $h_1(1-\beta_0) = \cdots = h_n(1-\beta_0) = R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$. For all $i = 1,\dots,n$, we have $h_i(x) \geqslant R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ for all $x \in (0, 1-\beta_0]$, i.e., $\sum_{i=1}^n X_i^* \geqslant R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ almost surely on $\{U \in [0,1-\beta_0)\}$. Since $\sum_{i=1}^n X_i^* = \sum_{i=1}^n Y_i = R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ on $\{U\in[1-\beta_0,1]\}$, we have $q_0^+\left(\sum_{i=1}^n X_i^*\right) = \max_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu) = R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$. □

B.4 Proofs in Section 7

Proof of Proposition 8. Theorem 4.17 of Rüschendorf (2013) gives

$$\inf_{\nu\in\Lambda(\boldsymbol{\mu})} \nu(-\infty, s] \geqslant 1 - D_n(s). \tag{EC.18}$$

Standard argument inverting (EC.18) gives (30). □

Proof of Theorem 4.

1. For fixed $t \in [0,1)$, denote by $x_1 = R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ the right-hand side (14). We proceed to show $D_n^{-1}(1-t) \leqslant x_1$ and thus the dual bound (30) is not greater than the convolution bound.

   Case 1: if the infimum in (14) is attained at $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n) \in (1-t)\Delta_n$ with $\beta_0, \dots, \beta_n > 0$, the first-order condition reads as the first equation in (EC.17). Because $\beta_0 > 0$ and $1 - \beta_0 - \beta_i < 1 - \beta_i$, we have $q^-_{1-\beta_0-\beta_i}(\mu_i) < q^-_{1-\beta_i}(\mu_i)$. Define $r_i = q^-_{1-\beta_0-\beta_i}(\mu_i)$ for $i = 1, \dots, n$. One can check from the first-order condition that $r_i = q^-_{1-\beta_i}(\mu_i) + r - x_1$, and hence $r < x_1$ and $\mathbf{r} \in \Delta_n(x_1)$. We have

$$\begin{aligned}
x_1 = \frac{1}{\beta_0} \sum_{j=1}^n \int_{1-\beta_0-\beta_j}^{1-\beta_j} q^-_u(\mu_j) \mathrm{d}u &= \frac{1}{\beta_0} \sum_{j=1}^n \int_{q^-_{1-\beta_0-\beta_j}(\mu_j)}^{q^-_{1-\beta_j}(\mu_j)} y \mu_j(\mathrm{d}y) \\
&= \frac{1}{\beta_0} \sum_{j=1}^n \int_{r_j}^{x_1 - r + r_j} y \mu_j(\mathrm{d}y) \\
&= \frac{1}{\beta_0} \sum_{j=1}^n \left((x_1 - r + r_j)(1-\beta_j) - r_j(1-\beta_0-\beta_j) - \int_{r_j}^{x_1-r+r_j} \mu_j(-\infty, y] \mathrm{d}y \right) \\
&= \frac{1}{\beta_0} \sum_{j=1}^n \left((x_1 - r)(1-\beta_j) + r_j \beta_0 - (x_1 - r) + \int_{r_j}^{x_1-r+r_j} \mu_j(y, \infty) \mathrm{d}y \right) \\
&= \frac{1}{\beta_0} \left(x_1 \beta_0 - (1-t)(x_1 - r) + \sum_{j=1}^n \int_{r_j}^{x_1-r+r_j} \mu_j(y, \infty) \mathrm{d}y \right).
\end{aligned}$$

   It follows that $1 - t = \sum_{j=1}^n \frac{1}{x_1 - r} \int_{r_j}^{x_1-r+r_j} \mu_j(y, \infty) \mathrm{d}y \geqslant D_n(x_1)$.

   Case 2: Suppose that the infimum in (14) is attained at $\boldsymbol{\beta}$ with some $\beta_i = 0$ for some $i \in \{1, \dots, n\}$ and $\beta_0 = 1 - t - \sum_{i=1}^n \beta_i > 0$. For $i = 1, \dots, n$, we have $q^-_1(\mu_i) > q^-_t(\mu_i)$ because $t < 1$ and define $r_i = q^-_{1-\beta_0-\beta_i}(\mu_i) < q^-_1(\mu_i)$. For $i \in \{1, \dots, n\}$ satisfying $\beta_i \neq 0$, the first-order condition reads as the first equation in (EC.17) and gives that $r_i = q^-_{1-\beta_i}(\mu_i) + r - x_1$. For i satisfying $\beta_i = 0$, the first-order condition reads as the second equation in (EC.17) and gives $q^-_1(\mu_i) \leqslant x_1 - r + r_i$ and $r \leqslant x_1 - (q^-_1(\mu_i) - r_i) < x_1$, which implies $\mu_i(-\infty, x_1 - r + r_i] = 1$ and $\mathbf{r} \in \Delta_n(x_1)$. Similarly,

$$x_1 = \frac{1}{\beta_0} \left(x_1 \beta_0 - (1-t)(x_1 - r) + \sum_{j=1}^n \int_{r_j}^{x_1-r+r_j} \mu_j(y, \infty) \mathrm{d}y \right).$$

   Therefore,

$$1 - t = \frac{1}{x_1 - r} \sum_{j=1}^n \int_{r_j}^{x_1-r+r_j} \mu_j(y, \infty) \mathrm{d}y \geqslant D_n(x_1).$$

   Case 3: If the infimum in (14) is attained at some $\boldsymbol{\beta}$ with $\beta_0 = 1 - t - \sum_{i=1}^n \beta_i = 0$, then from (14) we have the third equation in (EC.17). Define $r_i = q^-_{1-\beta_0-\beta_i}(\mu_i)$, $i = 1, \dots, n$. Then $r = \sum_{i=1}^n r_i = x_1$ and

$$1 - t = \sum_{i=1}^n \beta_i = \sum_{i=1}^n \mu_i(r_i, +\infty) = \lim_{\substack{\mathbf{r}' \in \Delta_n(x_1) \\ \mathbf{r}' \to \mathbf{r}}} \frac{1}{x_1 - r'} \sum_{i=1}^n \int_{r'_i}^{x_1 - r' + r'_i} \mu_i(y, \infty) \mathrm{d}y \geqslant D_n(x_1).$$

In all three cases, $1-t \geqslant D_n(x_1)$. Since D_n is decreasing, $D_n^{-1}(1-t) \leqslant x_1$, and thus the dual bound is not greater than the convolution bound.

2. For fixed $t \in [0,1)$, we proceed to show that the dual bound $D_n^{-1}(1-t)$ is not smaller than the convolution bound. We first claim that if quantile functions of $\mu_1, \dots, \mu_n$ are continuous, then D_n is strictly decreasing on $\left(-\infty, \sum_{i=1}^n q_1^-(\mu_j)\right)$ and is constant 0 on $\left[\sum_{i=1}^n q_1^-(\mu_j), \infty\right)$. Indeed, for any $x_1 < x_2$, we have

$$\begin{aligned} D_n(x_1) &= \inf_{\mathbf{r}\in\Delta_n(x_1)} \left\{ \sum_{i=1}^n \frac{1}{x_1 - r} \int_{r_i}^{x_1-r+r_i} \mu_i(y,\infty)\mathrm{d}y \right\} \\ &\geqslant \inf_{\mathbf{r}\in\Delta_n(x_2)} \left\{ \sum_{i=1}^n \frac{1}{x_1 - r} \int_{r_i}^{x_1-r+r_i} \mu_i(y,\infty)\mathrm{d}y \right\} \\ &\geqslant \inf_{\mathbf{r}\in\Delta_n(x_2)} \left\{ \sum_{i=1}^n \frac{1}{x_2 - r} \int_{r_i}^{x_2-r+r_i} \mu_i(y,\infty)\mathrm{d}y \right\} = D_n(x_2). \end{aligned}$$

We prove that if "=" holds, it must be $D_n(x_1) = D_n(x_2) = 0$. Since $D_n(x_1) = D_n(x_2) \in [0,n]$ is bounded, the infimum is attained at some $\mathbf{r}$ with $r \leqslant x_1$. Because $\mu_j(\cdot,\infty)$ is decreasing, we have for $j = 1, \dots, n$, $\mu_j(\cdot,\infty)$ is a constant on $[r_j, x_2 - r + r_j]$. The fact that quantile functions of $\mu_1, \dots, \mu_n$ are continuous implies that these constants can only be 0 or 1, and they cannot be 1 since it is an infimum. Hence for $j = 1, \dots, n$, $\mu_j(y,\infty) \equiv 0$ for $y \in [r_j, x_2 - r + r_j]$, which implies $D_n(x_1) = D_n(x_2) = 0$. It is straightforward to check that $D_n(x) = 0$ implies $x \geqslant \sum_{i=1}^n q_1^-(\mu_j)$. Thus we prove the claim. One can further verify that D_n is continuous.

Now we continue to prove the main result. For fixed $t \in [0,1)$, we have $D_n^{-1}(1-t) < \sum_{i=1}^n q_1^-(\mu_j)$ and $D_n(D_n^{-1}(1-t)) > 0$. As D_n is strictly decreasing and continuous on $\left(-\infty, \sum_{i=1}^n q_1^-(\mu_j)\right)$, we have $D_n(D_n^{-1}(1-t)) = 1-t$. Denote by $x_2 = D_n^{-1}(1-t)$ the value of the dual bound.

Case 1: Suppose that the infimum of $D_n(x_2)$ is attained at $\mathbf{r} = (r_1, \dots, r_n) \in \Delta_n(x_2)$. Its first-order condition reads as, for any $i = 1, \dots, n$,

$$\mu_i(r_i,\infty) + \sum_{j\neq i} \mu_j(x_2 - r + r_j, \infty) = \frac{1}{x_2 - r} \sum_{j=1}^n \int_{r_j}^{x_2-r+r_j} \mu_j(y,\infty)\mathrm{d}y = D_n(x_2) = 1 - t.$$

Define $\beta_i = \mu_i(x_2 - r + r_i, \infty)$, $i = 1, \dots, n$ and $\beta_0 = 1 - t - \sum_{i=1}^n \beta_i$. One can check $\beta_i = 1 - \beta_0 -$

$\mu_i(-\infty, r_i]$ and $\boldsymbol{\beta} \in (1-t)\Delta_n$ because $r < x_2$. We have

$$\begin{aligned}
1-t &= \frac{1}{x_2 - r} \sum_{j=1}^{n} \int_{r_j}^{x_2 - r + r_j} \mu_j(y, \infty) \mathrm{d}y \\
&= \frac{1}{x_2 - r} \sum_{j=1}^{n} \int_{1-\beta_0-\beta_j}^{1-\beta_j} (1-u) \mathrm{d} q_u^-(\mu_j) \\
&= \frac{1}{x_2 - r} \sum_{j=1}^{n} \left((x_2 - r + r_j)\beta_j - r_j(\beta_0 + \beta_j) + \int_{1-\beta_0-\beta_j}^{1-\beta_j} q_u^-(\mu_j) \mathrm{d}u \right) \\
&= \frac{1}{x_2 - r} \left((x_2 - r)(1-t) - x_2\beta_0 + \sum_{j=1}^{n} \int_{1-\beta_0-\beta_j}^{1-\beta_j} q_u^-(\mu_j) \mathrm{d}u \right).
\end{aligned}$$

Therefore,

$$x_2 = \frac{1}{\beta_0} \sum_{j=1}^{n} \int_{1-\beta_0-\beta_j}^{1-\beta_j} q_u^-(\mu_j) \mathrm{d}u,$$

which implies that the value of the dual bound x_2 is not smaller than that of the convolution bound.

Case 2: If the infimum of $D_n(x_2)$ is attained at some $\mathbf{r}$ with $r = x_2$, then

$$D_n(x_2) = \lim_{\substack{\mathbf{r}' \in \Delta_n(x_2) \\ \mathbf{r}' \to \mathbf{r}}} \sum_{i=1}^{n} \frac{1}{x_2 - r'} \int_{r_i'}^{x_2 - r' + r_i'} \mu_i(y, \infty) \mathrm{d}y = \sum_{i=1}^{n} \mu_i(r_i, \infty) = 1 - t.$$

Define $\beta_i = \mu_i(r_i, \infty)$, $i = 1, \dots, n$ and $\beta_0 = 0$. We have $r_i = q_{1-\beta_i}^-(\mu_i)$, and

$$x_2 = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} q_{1-\beta_i}^-(\mu_i) = \lim_{\substack{\boldsymbol{\beta}' \in \Delta_n \\ \boldsymbol{\beta}' \to \boldsymbol{\beta}}} \sum_{i=1}^{n} \frac{1}{\beta_0'} \int_{1-\beta_0'-\beta_i'}^{1-\beta_i'} q_u^-(\mu_i) \mathrm{d}u \geqslant \inf_{\boldsymbol{\beta}' \in \Delta_n} \sum_{i=1}^{n} \frac{1}{\beta_0'} \int_{1-\beta_0'-\beta_i'}^{1-\beta_i'} q_u^-(\mu_i) \mathrm{d}u,$$

which implies that the value of the dual bound x_2 is not smaller than that of the convolution bound. The statement on the correspondence is shown in the above steps. □

C Counter-examples

Example EC.1 (Non-sharpness in Theorem 2)**.** Without loss of generality, we consider the case $t = 0$. Let μ be a bi-atomic uniform distribution on $\{-1, 1\}$. It is easy to see that $\sup_{\nu \in \Lambda_3(\mu)} q_0^+(\nu) = -1$ since any $\nu \in \Lambda_3(\mu)$ is supported in $\{-3, -1, 1, 3\}$ with mean 0. On the other hand, for $(\beta_0, \beta_1, \beta_2, \beta_3) \in \Delta_3$ with $\beta_1 \geqslant \beta_2 \geqslant \beta_3$, by symmetry, and the fact that $R_{1-\beta,\beta-\alpha}$ is increasing in α and increasing in β, we have

$$R_{\beta_1,\beta_0}(\mu) = -R_{1-\beta_0-\beta_1,\beta_0}(\mu) = -R_{\beta_2+\beta_3,\beta_0}(\mu) \geqslant -R_{\beta_2,\beta_0+\beta_3}(\mu),$$

$$R_{\beta_2,\beta_0}(\mu) \geqslant R_{\beta_2,\beta_0+\beta_3}(\mu),$$

and

$$R_{\beta_3,\beta_0}(\mu) \geqslant R_{\beta_3,\beta_0+\beta_2}(\mu) \geqslant R_{\beta_3,1-2\beta_3}(\mu) = 0.$$

Combining the above three inequalities, we have $\sum_{i=1}^{3} R_{\beta_i,\beta_0}(\mu) \geqslant 0$. Hence,

$$\sup_{\nu\in\Lambda_3(\mu)} q_0^+(\nu) = -1 < 0 \leqslant \inf_{\boldsymbol{\beta}\in\Delta_n} \sum_{i=1}^{3} R_{\beta_i,\beta_0}(\mu),$$

showing that (18) is not an equality.

Example EC.2 ((15) does not hold as an equality for an increasing density)**.** Without loss of generality, we consider the case $t = 0$. Suppose that $\mu \in \mathcal{M}$ has an increasing density on its support. Then, the cdf of μ is convex, and hence the left quantile $q_u^-(\mu)$ is a concave function of $u \in (0,1)$. For the concave and increasing function $q_u^-(\mu)$, we have

$$\frac{1}{1-n\alpha}\int_{(n-1)\alpha}^{1-\alpha} q_u^-(\mu)\mathrm{d}u \geqslant \frac{1}{1-2\alpha}\int_{\alpha}^{1-\alpha} q_u^-(\mu)\mathrm{d}u \geqslant \int_0^1 q_u^-(\mu)\mathrm{d}u.$$

Therefore,

$$\inf_{\alpha\in(0,\frac{1}{n})} nR_{\alpha,1-n\alpha}(\mu) = nR_{0,1}(\mu).$$

Note that if (15) holds as an equality, then $\inf_{\nu\in\Lambda_n(\mu)} q_1^-(\nu) = nR_{0,1}(\mu)$, which, by Proposition EC.3 below, implies that μ is n-CM. There are distributions μ with a decreasing density that are not n-CM, and an equivalent condition is obtained by Wang and Wang (2011); see Appendix E for further explanation. Therefore, (15) does not hold as an equality for some distributions with an increasing density. A specific example is shown in Figure 3 (right panel).

Example EC.3 ((20) does not hold without a finite mean)**.** By Theorem 4.2 of Puccetti et al. (2019), for standard Cauchy probability measures $\mu_1,\dots,\mu_n$, there exists $\nu_1,\nu_2 \in \Gamma(\boldsymbol{\mu})$ such that

$$q_0^+(\nu_1) = q_1^-(\nu_1) = \sup_{\boldsymbol{\beta}\in\Delta_n} \sum_{i=1}^{n} R_{1-\beta_0-\beta_i,\beta_0}(\mu_i) = -\frac{n\log(n-1)}{\pi}$$

and

$$q_0^+(\nu_2) = q_1^-(\nu_2) = \inf_{\boldsymbol{\beta}\in\Delta_n} \sum_{i=1}^{n} R_{\beta_i,\beta_0}(\mu_i) = \frac{n\log(n-1)}{\pi}.$$

Hence, we have

$$\begin{aligned}\inf_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu) = \sup_{\boldsymbol{\beta}\in\Delta_n} \sum_{i=1}^{n} R_{1-\beta_0-\beta_i,\beta_0}(\mu_i) &= -\frac{n\log(n-1)}{\pi} \\ &< \frac{n\log(n-1)}{\pi} = \inf_{\boldsymbol{\beta}\in\Delta_n} \sum_{i=1}^{n} R_{\beta_i,\beta_0}(\mu_i) = \sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu).\end{aligned}$$

Example EC.4 (RA fails)**.** In Example 1, μ is a triatomic uniform distribution on $\{1,2,3\}$, and the convolution bound gives $\sup_{\nu\in\Lambda(\mu,\mu,\mu)} q_0^+(\nu) \leqslant 6$ which is attainable. With the initial matrix below, we see that RA does not provide 6, but instead it gives the interval [5,5].

$$\text{RA:} \quad \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix} \implies \begin{pmatrix} 3 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 3 & 3 \end{pmatrix} \implies \text{termination}$$

For how RA runs, see Embrechts et al. (2013).

D Well-posedness

Similarly to (21), for $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}^n$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n) \in \Delta_n$, we denote by

$$R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{i=1}^{n} R_{1-\beta_i-\beta_0,\beta_0}(\mu_i). \tag{EC.19}$$

We discuss the attainability of the infimum in $\inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ and the supremum in $\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu})$. Note that $R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ and $R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ are well defined for $\boldsymbol{\beta} \in \Delta_n$. Now we discuss cases with β_i taking boundary values of $0, 1$. We discuss whether $R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ and $R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ are well defined on the closure $\overline{\Delta}_n$.

1. For $\boldsymbol{\beta} \in \Delta_n \subset \overline{\Delta}_n$, there is no undefined form "$\infty - \infty$" in $R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ and $R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu})$, which are hence always well defined.

2. For $\boldsymbol{\beta} \in \overline{\Delta}_n$ with $\beta_i = 0$ for some $i \in \{1, \dots, n\}$ and $\beta_0 \in (0, 1]$, we define $R^-_{\boldsymbol{\beta}}$ and $R^+_{\boldsymbol{\beta}}$ similarly:

$$R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{j\neq i} R_{\beta_j,\beta_0}(\mu_j) + R_{0,\beta_0}(\mu_i), \quad R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{j\neq i} R_{1-\beta_j-\beta_0,\beta_0}(\mu_j) + R_{1-\beta_0,\beta_0}(\mu_i),$$

   except "$\infty - \infty$" cases that the integral of $q_t^-(\mu_i)$ at the neighbour of 0 is negative infinite and that of $q_t^-(\mu_j)$ at the neighbour of 1 is infinite for some $i, j \in \{1, \dots, n\}$, i.e., $R_{0,\varepsilon}(\mu_j) = \infty$ and $R_{1-\varepsilon,\varepsilon}(\mu_i) = -\infty$ for some $\varepsilon \in (0, 1)$. $R^+_{\boldsymbol{\beta}}$ and $R^-_{\boldsymbol{\beta}}$ are always well defined if $\boldsymbol{\mu} \in \mathcal{M}^n_1$.

3. For $\boldsymbol{\beta} \in \overline{\Delta}_n$ with $\beta_0 = 0$, we define

$$R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{i=1}^{n} q^-_{1-\beta_i}(\mu_i), \qquad R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{i=1}^{n} q^+_{\beta_i}(\mu_i),$$

   except "$\infty - \infty$" cases that $q_1^-(\mu_i) = \infty$ and $q_0^-(\mu_j) = -\infty$ for some $i, j \in \{1, \dots, n\}$ and $i \neq j$. They are always well defined if $\mu_1, \dots, \mu_n$ are all bounded from the positive or negative side.

Because of the continuity of $R_{\beta,\alpha}$ in $\beta, \alpha \in [0, 1]$, it can be proved that the infimum of $\inf_{\boldsymbol{\beta}\in(1-t)\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ of cases $t \in [0, 1)$ and the supremum of $\sup_{\boldsymbol{\beta}\in t\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ of cases $t \in (0, 1]$ are attained in the well-defined

part of $\overline{\Delta}_n$.

E Connection to joint mixability

Joint mixability is closely related to quantile aggregation. The tuple of distributions $\boldsymbol{\mu} \in \mathcal{M}^n$ is said to be *jointly mixable* (JM, Wang et al. (2013)) if $\delta_C \in \Lambda(\boldsymbol{\mu})$ for some $C \in \mathbb{R}$. Such C is called a center of $\boldsymbol{\mu}$. The name JM means that the marginal distributions $(\mu_1, \dots, \mu_n)$ is *able* to support a *joint mix* dependence (i.e., a random vector with a constant sum). Similarly, a probability measure μ on $\mathbb{R}$ is *n-completely mixable* (n-CM, Wang and Wang (2011)) if the n-tuple $(\mu, \dots, \mu)$ is JM. Obviously, if $\boldsymbol{\mu} \in \mathcal{M}_1^n$ is JM, then its center is unique and equal to the sum of the means of its components. If $\boldsymbol{\mu} \in \mathcal{M}^n$ is JM but it is not in $\mathcal{M}_1^n$, then its center may not be unique (Puccetti et al. (2019)). The determination of joint mixability for a given $\boldsymbol{\mu} \in \mathcal{M}^n$ is well known to be a challenging problem and analytical results are limited. The main results of this appendix are a sufficient condition on the sharpness of convolution bounds and some conditions on the determination of JM.

We first see that JM is a sufficient condition for the bounds in Proposition 6 to be sharp for probability measures with finite means.

Proposition EC.1. *If $\boldsymbol{\mu} \in \mathcal{M}_1^n$ is JM, then the bounds in Proposition 6 are sharp, and their values are equal to the unique center of $\boldsymbol{\mu}$.*

Proof. Note that since $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ is JM, we know $\delta_C \in \Lambda(\boldsymbol{\mu})$ where $C = \sum_{i=1}^n R_{0,1}(\mu_i)$. Hence, by Proposition 6,

$$\inf_{\boldsymbol{\beta} \in \Delta_n} R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) \geqslant \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) \geqslant q_0^+(\delta_C) \geqslant C \geqslant \inf_{\boldsymbol{\beta} \in \Delta_n} R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}).$$

The case for (19) is similar. □

Proposition EC.1 supports Proposition 6 by giving further conditions for the bounds in Proposition 6 to be sharp, which can be checked through existing results on joint mixability in Wang and Wang (2016). However, unlike Theorem A.2, Proposition EC.1 itself does not offer new ways to calculate quantile aggregation, since the convolution bounds in (18) and (19) are all trivially equal to the center if we know $(\mu_1, \dots, \mu_n) \in \mathcal{M}_1^n$ is JM.

Next, we look in the converse direction: implications of Theorem 2 and Theorem A.2 on conditions for JM. Proposition 6 directly implies the following necessary condition for JM, which is also noted by Proposition 3.3 of Puccetti et al. (2019) with a similar argument. If $\boldsymbol{\mu}$ is JM with center C, then $C = q_0^+(\nu_0) = q_1^-(\nu_0)$ for some $\nu_0 \in \Lambda(\boldsymbol{\mu})$. Hence,

$$\inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_1^-(\nu) \leqslant C \leqslant \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu).$$

Using Proposition 6, we arrive at (where $R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu})$ is defined at (EC.19))

$$\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) \leqslant C \leqslant \inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}).$$

If the means of $\mu_1,\dots,\mu_n$ are finite, then by Proposition 7, we have $\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) \geqslant \inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})$. Therefore, a necessary condition for $\boldsymbol{\mu}\in\mathcal{M}_1^n$ to be JM is

$$\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}).$$

We summarize the above simple findings in the following proposition. We use the convention that the closed interval $[a,b]$ is empty if $a>b$.

Proposition EC.2. *The possible center C of $\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)\in\mathcal{M}^n$ satisfies*

$$C \in \left[\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}), \inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu})\right]. \tag{EC.20}$$

In particular, if $\boldsymbol{\mu}$ is JM, then

$$\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) \leqslant \inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}), \tag{EC.21}$$

and further if $\boldsymbol{\mu}\in\mathcal{M}_1^n$, then

$$\sup_{\boldsymbol{\beta}\in\Delta_n} R^-_{\boldsymbol{\beta}}(\boldsymbol{\mu}) = \sum_{i=1}^n R_{0,1}(\mu_i) = \inf_{\boldsymbol{\beta}\in\Delta_n} R^+_{\boldsymbol{\beta}}(\boldsymbol{\mu}). \tag{EC.22}$$

The set Δ_n appeared in Proposition EC.2 may be replaced by $\overline{\Delta}_n$ if $(\mu_1,\dots,\mu_n)\in\mathcal{M}_1^n$. We next verify that for many classes distributions known in the literature, (EC.21)-(EC.22) actually are sufficient for JM, and all centers are identified with Proposition EC.2. We first present a convenient result which is useful for the determination of JM for distributions with finite means.

Proposition EC.3. *For $\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)\in\mathcal{M}_1^n$, the following statements are equivalent.*

(i) $\boldsymbol{\mu}$ *is JM.*

(ii) $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu) = \sum_{i=1}^n R_{0,1}(\mu_i)$.

(iii) $\inf_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu) = \sum_{i=1}^n R_{0,1}(\mu_i)$.

(iv) $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu) = \inf_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu)$.

Proof. Let $C=\sum_{i=1}^n R_{0,1}(\mu_i)$. By Proposition 7,

$$\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu) \leqslant C \leqslant \inf_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu). \tag{EC.23}$$

As a consequence, (iv)⇒(ii)-(iii).

If $\boldsymbol{\mu}$ is JM, then there exists $\nu_0 \in \Lambda(\boldsymbol{\mu})$ such that $q_0^+(\nu_0) = \sum_{i=1}^n R_{0,1}(\mu_i) = q_1^-(\nu_0)$. This, together with (EC.23), shows the implication (i)⇒(ii)-(iv).

If $\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu) = C$, then, noting that $R_{0,1}(\nu) = C$ for all $\nu \in \Lambda(\boldsymbol{\mu})$, we have $\delta_C \in \Lambda(\boldsymbol{\mu})$, since $\Lambda(\boldsymbol{\mu})$ is closed under weak convergence (Theorem 2.1 of Bernard et al. (2014)). This shows (ii)⇒(i). Similarly, (iii)⇒(i). □

Next, in view of Theorem 2, we show in Proposition EC.4 that it can be checked through convolution bounds whether some distributions are JM if they have monotone densities.

Proposition EC.4. *$\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \mathcal{M}_1^n$ is JM if and only if (EC.22) holds, in the following cases:*

(i) Each of $\mu_1, \dots, \mu_n$ admits a decreasing density on its support.

(ii) Each of $\mu_1, \dots, \mu_n$ admits an increasing density on its support.

(iii) $\mu_1, \dots, \mu_n$ are from the same location-scale family with unimodal and symmetric densities on their supports.

Proof. The necessity of (EC.22) is stated in Proposition EC.2, and hence we only show its sufficiency.

(i) By Theorem 2 and (EC.22), we know

$$\sup_{\nu\in\Lambda(\boldsymbol{\mu})} q_0^+(\nu) = \inf_{\boldsymbol{\beta}\in\Delta_n} R_{\boldsymbol{\beta}}^+(\boldsymbol{\mu}) = \sum_{i=1}^n R_{0,1}(\mu_i).$$

By Proposition EC.3 (ii)⇒(i), $\boldsymbol{\mu}$ is JM.

(ii) This is symmetric to (i).

(iii) Without loss of generality, we may assume that $\mu_1, \dots, \mu_n$ all have mean zero and they have scale parameters $a_1 \geqslant \dots \geqslant a_n > 0$, respectively. By Corollary 3.6 of Wang and Wang (2016), we know that $(\mu_1, \dots, \mu_n)$ is JM if and only if $2\bigvee_{i=1}^n a_i \leqslant \sum_{i=1}^n a_i$. Take $\boldsymbol{\beta} = (1-\varepsilon, \varepsilon, 0, 0, \dots, 0) \in \overline{\Delta}_n$ for some $\varepsilon \in (0,1)$. Since $\mu_1, \dots, \mu_n$ are from the same location-scale family with symmetric densities, we have $-R_{\varepsilon,1-\varepsilon}(\mu_i)/a_i = -R_{0,1-\varepsilon}(\mu_i)/a_i = R_{0,1-\varepsilon}(\mu_1)/a_1 > 0$ for $i = 1, \dots, n$. By (EC.22), we have

$$\begin{aligned} 0 = \sum_{i=1}^n R_{0,1}(\mu_i) \geqslant \sum_{i=1}^n R_{1-\beta_0-\beta_i,\beta_0}(\mu_i) &= R_{0,1-\varepsilon}(\mu_1) + \sum_{i=2}^n R_{\varepsilon,1-\varepsilon}(\mu_i) \\ &= \frac{R_{0,1-\varepsilon}(\mu_1)}{a_1}\left(a_1 - \sum_{i=2}^n a_i\right). \end{aligned}$$

Therefore, $a_1 - \sum_{i=2}^n a_i \leqslant 0$, which implies $2\bigvee_{i=1}^n a_i \leqslant \sum_{i=1}^n a_i$.

□

Remark EC.1. By Theorem 3.2 of Wang and Wang (2016), for $\mu_1, \dots, \mu_n \in \mathcal{M}$ with decreasing densities, $(\mu_1, \dots, \mu_n)$ is JM if and only if

$$\bigvee_{i=1}^{n} \left(q_1^-(\mu_i) - q_0^+(\mu_i)\right) \leqslant \sum_{i=1}^{n} \left(R_{0,1}(\mu_i) - q_0^+(\mu_i)\right). \tag{EC.24}$$

We already know that (EC.22) is necessary for $(\mu_1, \dots, \mu_n)$ to be JM. One can directly check that (EC.24) is implied by (EC.22), thus showing the equivalence of (EC.22) and (EC.24).

Remark EC.2. A similar situation of Proposition EC.4 is obtained for distributions without the mean: if each of $\mu_1, \dots, \mu_n$ is a standard Cauchy distribution, the set of all centers of $(\mu_1, \dots, \mu_n)$ is precisely given by (EC.20). This statement is based on Example 4.1 and Theorem 4.2 of Puccetti et al. (2019).

F Comparison of different methods in computation

In this appendix, we compare three potential ways of computing the quantile aggregation problem or its approximations. The first two approaches, RA and linear program, require a discretization, whereas the third approach, the convolution bound, can be applied with either discrete input or functional input.

(a) **Original problem**: Let $[n] := \{1, \dots, n\}$. Consider the quantile aggregation problem

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) = \sup\{q_0^+(X_1 + \cdots + X_n) : X_i \sim \mu_i,\ i \in [n]\}, \tag{EC.25}$$

where $\mu_1, \dots, \mu_n$ are distributions on $\mathbb{R}$, with supports bounded from below, and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$. Note that $\mu_1, \dots, \mu_n$ are assumed to have supports bounded from below because otherwise q_0^+ may be infinite. The probability level 0 is chosen here without loss of generality, because any risk aggregation problem for q_t^+ can be equivalently formulated as one for q_0^+ as shown in Proposition 1.

(b) **Discretization**: To tackle problem (EC.25) numerically for given distributions $\mu_1, \dots, \mu_n$, a common step is to discretize using their quantiles, that is, to consider a number m (ideally large) and for each $i \in [n]$, a distribution μ_i^m over m points (some may be equal) each with probability $1/m$:

$$z_1^i = q_0^+(\mu_i), \quad \dots, \quad z_m^i = q_{(m-1)/m}^+(\mu_i). \tag{EC.26}$$

The input values of the discrete problem are these z_j^i for $i \in [n]$, $j \in [m]$. This discretization is asymptotically consistent in the following sense: Let $\boldsymbol{\mu}^m = (\mu_1^m, \dots, \mu_n^m)$. As $m \to \infty$, since $\mu_i^m \to \mu_i$ weakly, $\mu_i^m \leqslant_{\mathrm{st}} \mu_i$ (where $\leqslant_{\mathrm{st}}$ stands for stochastic order), and q_0^+ is upper semi-continuous, we have

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu}^m)} q_0^+(\nu) \to \sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu).$$

(c) **Rearrangement algorithm (RA)**: We first write the input values of the discrete problem into a matrix

$$\begin{pmatrix} z_1^1 & z_1^2 & \dots & z_1^n \\ \vdots & \vdots & \ddots & \vdots \\ z_m^1 & z_m^2 & \dots & z_m^n \end{pmatrix}. \tag{EC.27}$$

The RA, introduced by Embrechts et al. (2013), tries to maximize the minimal row sum of the matrix resulting from rotating the elements within each column of (EC.27). Note that rotating elements within each column corresponds to changing the dependence structure of a discrete random vector while maintaining its marginals, and the row-sum vector corresponds to the distribution of the sum of components of the random vector; see Section G.2 for this problem in a different context. The problem of finding the maximum of the minimal row sum is NP-hard (e.g., Haus (2015)), but RA can compute a suboptimal answer very quickly, which typically has good accuracy. The theoretical computational complexity of RA is unknown in the literature. For discrete distributions, as in our setting here, the output of RA is $\underline{s}_N$, which is a lower bound on the true value of (EC.25). For continuous distributions, RA outputs an interval $[\underline{s}_N, \bar{s}_N]$ with $\underline{s}_N$ again being a lower bound for the original (continuous) problem, but there is no guarantee for $\bar{s}_N$ to be either a lower bound or an upper bound for the original problem. RA is the most popular and standard method in the risk management literature to compute the quantile aggregation problem; see Embrechts et al. (2013, 2014).

(d) **Linear program (LP)**: Recall that $\Gamma(\boldsymbol{\mu})$ represents the set of all distributions on $\mathbb{R}^n$ with marginals $\boldsymbol{\mu}$. The problem (EC.25) can be equivalently formulated as

$$\sup_{\nu \in \Lambda(\boldsymbol{\mu})} q_0^+(\nu) = \sup\left\{ x \in \mathbb{R} : \int_{\mathbb{R}^n} \mathbb{1}_{\{x_1+\cdots+x_n \leqslant x\}} \Pi(\mathrm{d}x_1, \dots, \mathrm{d}x_n) = 0 \text{ for some } \Pi \in \Gamma(\boldsymbol{\mu}) \right\}. \tag{EC.28}$$

Writing $\Pi_{\mathbf{k}} = \Pi(\{(z_{k_1}^1, \dots, z_{k_n}^n)\})$ for $\mathbf{k} = (k_1, \dots, k_n) \in [m]^n$, which represents the probability

$$\mathbb{P}(X_1 = z_{k_1}^1, \dots, X_n = z_{k_n}^n)$$

for a random vector $(X_1, \dots, X_n)$ with the given marginals. The above problem can be formulated as the following program

$$\begin{aligned} \sup \ & x \\ \text{subject to } \ & \Pi_{\mathbf{k}} \in [0,1] \text{ for } \mathbf{k} = (k_1, \dots, k_n) \in [m]^n \\ & \sum_{\mathbf{k} \in [m]^n} \mathbb{1}_{\{z_{k_1}^1 + \cdots + z_{k_n}^n \leqslant x\}} \Pi_{\mathbf{k}} = 0; \\ & \sum_{\mathbf{k}: k_i = j} \Pi_{\mathbf{k}} = \frac{1}{m} \quad \text{for each } i \in [n] \text{ and } j \in [m]. \end{aligned} \tag{EC.29}$$

Practically, we need to try to solve for discrete values of x. The method has two steps.

Step 1: We specify a real interval T that covers the range of values for the optimal value of (EC.29). For each fixed $x \in T$, we define

$$\begin{aligned} V(x) = \min \quad & \sum_{\mathbf{k}\in[m]^n} y_{\mathbf{k}} \Pi_{\mathbf{k}} \\ \text{subject to} \quad & \Pi_{\mathbf{k}} \geqslant 0 \text{ for } \mathbf{k} = (k_1, \dots, k_n) \in [m]^n, \\ & \sum_{\mathbf{k}: k_i = j} \Pi_{\mathbf{k}} = \frac{1}{m} \quad \text{for each } i \in [n] \text{ and } j \in [m], \end{aligned} \tag{EC.30}$$

where $y_{\mathbf{k}} = \mathbb{1}_{\{z^1_{k_1} + \cdots + z^n_{k_n} \leqslant x\}}$ are parameters. Problem (EC.30) is a linear program (LP) with m^n variables and $n \times m$ equality constraints.

Step 2: The optimal value output by the algorithm is given by

$$\sup\{x \in T : \text{the optimal value } V(x) \text{ in Problem (EC.30)} \leqslant 0\}. \tag{EC.31}$$

For each fixed $x \in T$, this method has m^n variables and $n \times m$ constraints. Note that instead of solving for each $x \in T$, to search for the optimal x, a bisection approach can replace the specification of T. The complexity of this problem is discussed in Remark EC.3.

(e) **Convolution bound (CB)**: We directly take the quantile functions of $\mu_1, \dots, \mu_n$ as input, let $t = 0$, and solve for

$$B_{\mathrm{conv}} = \inf_{\boldsymbol{\beta} \in (1-t)\Delta_n} \sum_{i=1}^{n} R_{\beta_i, \beta_0}(\mu_i).$$

Each term $R_{\beta_i,\beta_0}(\mu_i)$ is an integral of the corresponding quantile function. If the marginal distributions are given as discrete data points, then the input values for CB are the empirical quantile functions. Although this minimization is not convex and we do not know the theoretical computational complexity of CB, in all numerical examples we find that it can be computed very fast and accurately.

Next, we provide numerical results to compare the three methods, RA, LP and CB, described above. For a comparison, we assume that each marginal distribution μ_i is uniform on m points, denoted by $z^i_1, \dots, z^i_m$ as in the discussion above. Note that CB can also take quantile functions as input, whereas RA and LP can only take discrete input. These m points are specified in two different ways.

(i) They are the values of the quantile functions at different levels as in (EC.26).

(ii) They are randomly sampled from some distributions.

For this specification of the marginal distributions (there is no discretization involved, as the marginal distributions are themselves discrete), LP produces a true value of (EC.25), RA produces a lower bound on (EC.25), and CB produces an upper bound on (EC.25).

The numerical results are reported in Table 6. We make the following observations from the results.

1. All methods become slower when either m or n increases, as expected.

2. LP produces the true value for (EC.25), but it has some drawbacks for implementation. As each LP problem involves m^n variables, the applicability seems to be very limited. For $m \geqslant 180$ with $n = 3$, or $m \geqslant 20$ with $n = 5$, the computation is either unavailable in MATLAB or costs more than 24 hours. In a real risk management problem where loss distributions are typically continuous (such as asset prices or insurance losses), the value of m needs to be relatively large to ensure good approximation (typically at least 10^5; see Embrechts et al. (2013)).

3. RA is fast in most cases and simple in coding. It can handle $m = 10^6$ and $n = 200$ as demonstrated by Embrechts et al. (2013). However, it only provides a lower bound, and sometimes this lower bound may not be close to the true value provided by LP. There are no theoretical results on the convergence of RA.

4. CB is much faster than LP, but slower than RA. CB can handle dimensions up to $n = 200$. In some cases studied in this paper, such as continuous distributions with monotone densities, it is theoretically proved that it produces an exact true value. In our numerical results, the CB value often coincides with, or is very close to that of LP. It also enjoys the interpretability of the dependence structure (see Section 6). It can directly handle continuous distributions often encountered in risk management without the need to discretize (for such setting, m is practically infinity). We comment on two disadvantages of CB. First, in the computation of CB, the convergence is based on the optimization function in the software (`fmincon` in MATLAB in our case). There is no theoretical guarantee that this function finds the global optimum, due to lack of convexity. Nevertheless, in all numerical results where CB and LP agree, we know that global optimum is reached. Second, in case the conditions in Theorem 2 do not hold, we do not know whether CB is equal to the original problem (EC.25). Despite the gap due to the non-convexity in computing CB and the gap between CB and (EC.25), if we obtain a solution $\boldsymbol{\beta}$ from the optimization software, the objective value evaluated at $\boldsymbol{\beta}$ is guaranteed to give an upper bound for (EC.25).

5. Combining RA and CB gives a theoretically proven interval in which the true value of (EC.25) lies. This gives a fast and reliable way of finding the range of (EC.25). While RA is practically fast and commonly used, it only provides a one-sided bound (and arguably the less important side), and CB essentially closes the other side.

Remark EC.3. This binary search on the values of x in the LP (EC.30) inside (EC.31) can be done efficiently as $V(x)$ is monotone (although not necessarily strictly monotone). This implies that the number of binary queries is $O(\log(\text{length}(T)/\varepsilon)$, where length$(T)$ is the length of the interval T and ε is the (additive) error tolerance to which we want to compute x. Hence, Problem (EC.29) has a poly(m, n) complexity if Problem (EC.30) with any fixed $x \in T$ has; see Chapter 8.7 of Papadimitriou and Steiglitz (1998) for details.

Table 6: Comparison of the rearrangement algorithm (RA), the linear program (LP), and the convolution bound (CB). In (a)-(e), the distribution μ_i is uniform on a set $\{z_1^i, \ldots, z_m^i\}$ for each $i \in [n]$. In (f), the marginal distributions are continuous, and RA produces an interval $[\underline{s}_N, \bar{s}_N]$ using a discretization with $N = 10^5$ steps.

	$z_j^i = i \times j$	$m = 120$	$z_j^i = i \times j$	$m = 160$	$z_j^i = i \times j$	$m = 180$
	value	time	value	time	value	time
LP	363	11988s	483	24960s	NA	>24h
RA	358	66s	478	140s	534	212s
CB	363	1447s	483	1687s	543	2315s

(a) $n = 3$ and $z_1^i, \ldots, z_m^i$ are equidistant for each $i \in [n]$

	z_j^i iid $\sim \exp(100 \times i)$	$m = 60$	z_j^i iid $\sim \mathrm{U}[0, 100 \times i^2]$	$m = 100$
	value	time	value	time
LP	589.6	104.7s	499.3	2865.3s
RA	583.2	0.02s	499.3	35.3s
CB	608.0	1.8s	499.3	692.1s

(b) $n = 3$ and $z_1^i, \ldots, z_m^i$ are randomly generated for each $i \in [n]$

	$z_j^i = i^2 \times j$	$m = 30$	$z_j^i = j^2$	$m = 30$
	value	time	value	time
LP	436	5280s	1260	6707s
RA	436	0.02s	1230	0.01s
CB	446	0.9s	1260.7	0.5s

(c) $n = 4$ and $z_1^i, \ldots, z_m^i$ are deterministic for each $i \in [n]$

	z_j^i iid $\sim \exp(100 \times i)$	$m = 30$	z_j^i iid $\sim \mathrm{Binomial}(300, 0.2 \times i)$	$m = 30$
	value	time	value	time
LP	845.3	2937s	1932	3091s
RA	828.8	0.03s	1924	0.02s
CB	881.5	1.7s	1935.5	0.9s

(d) $n = 4$ and $z_1^i, \ldots, z_m^i$ are randomly generated for each $i \in [n]$

	z_j^i iid $\sim \exp(100 \times i)$	$m = 15$	z_j^i iid $\sim \mathrm{U}[0, 100 \times i^2]$	$m = 20$
	value	time	value	time
LP	1198.0	4988.9s	NA	>24h
RA	1158.7	0.02s	2715.8	0.03s
CB	1214.0	1.3s	2760.0	1.8s

(e) $n = 5$ and $z_1^i, \ldots, z_m^i$ are randomly generated for each $i \in [n]$

	$\mu_i = \mathrm{Gamma}(3, 1)$	$n = 200$
	value	time
RA	[599.9, 600.0]	1710s
CB	600	2021s

(f) $n = 200$ with continuous distribution; LP cannot handle such large n

Next we focus on the complexity of Problem (EC.30). While it has an exponential in n, namely m^n, number of decision variables, its dual problem, given by

$$\begin{aligned} &\max_{p_{ij}\in\mathbb{R}, i=1,\dots,n, j=1,\dots,m} \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\frac{1}{m}, \\ &\text{subject to } \ y_{\mathbf{k}} - \textstyle\sum_{i=1}^{n} p_{i,k_i} \geqslant 0 \text{ for any } \mathbf{k}=(k_1,\dots,k_n)\in[m]^n, \end{aligned} \tag{EC.32}$$

is an LP with $n\times m$ variables and m^n constraints. The polynomial number of variables in the dual problem (EC.32) suggests the potential of yielding poly(m,n)-time algorithms for the primal problem (EC.30). In particular, via the ellipsoid method (e.g., Lemma 3.2 of Grötschel et al. (1981)), the LP (EC.30) is polynomial-time solvable if there is a separation oracle for (EC.32) that runs in polynomial time; see Definition 6.2.2 of Grötschel et al. (2012). However, deducing the availability of such a polynomial-time separation oracle appears challenging. In fact, (EC.30) belongs to the multi-marginal optimal transport (MOT) problem (Altschuler and Boix-Adserà (2021, 2023)) with a so-called set-optimization structure (Section 6.1 of Altschuler and Boix-Adserà (2023)). More precisely, given the fixed matrix $\mathbf{z}$ defined by (EC.27) and x, we define the set $S=\{\mathbf{k}\in[m]^n : z^1_{k_1}+\cdots+z^n_{k_n} > x\}$. Based on Definition 6.5 and Theorem 6.8 of Altschuler and Boix-Adserà (2023), Problem (EC.30) for fixed x has a polynomial complexity if the problem $\min_{\mathbf{k}\in S} -\sum_{i=1}^n p_{i,k_i}$, for any arbitrary matrix $\mathbf{p}\in\mathbb{R}^{m\times n}$, has. To this end, neither Altschuler and Boix-Adserà (2023) or any other works to our best knowledge has worked out the polynomial complexity of the problem $\min_{\mathbf{k}\in S} -\sum_{i=1}^n p_{i,k_i}$ with our considered S. With this, it appears that the question of whether (EC.29) has a polynomial complexity remains open.

G Two further applications

We illustrate the convolution bounds in two additional applications. Section G.1 constructs a new robust test for simulation calibration. Section G.2 discusses the classic assembly line crew scheduling problem.

G.1 Simulation calibration

In multiple statistical hypothesis testing, quantile aggregation gives critical values for various methods to combine p-values from different tests among which, most often, no dependence information is available; see e.g., Ramdas et al. (2019), Vovk and Wang (2020) and Vovk et al. (2022). This problem also arises in operations research, especially in the context of stochastic simulation model calibration (Kleijnen (1995); Sargent (2010)).

In simulation analysis, calibration refers to the search for parameters of simulation models to best match real data. These models are constructed to resemble the hidden dynamics of a system, which are often complex and not amenable to closed-form analysis. Instead, by running Monte Carlo, we can obtain the model outputs for prediction and other downstream decision-making tasks such as sensitivity analysis

and optimization (see, e.g., Law and Kelton (2000) for a range of applications in production and operations management). However, to ensure that the conclusions of these analyses are reliable, it is critical that the input parameters in the hidden dynamics are correctly tuned. This calls for the need for calibration, where the outputs from the simulation model are matched against the real data in order to locate the parameter values. In a setting of one-dimensional (continuous) output, one could rely on a two-sample goodness-of-fit test such as the Kolmogorov-Smirnov (KS) test, which looks at the KS statistic

$$\mathrm{KS} = \sup_{x\in\mathbb{R}} |\hat{F}_{\mathrm{sim}}(x) - \hat{F}_{\mathrm{real}}(x)|.$$

Here $\hat{F}_{\mathrm{sim}}$ and $\hat{F}_{\mathrm{real}}$ are the empirical distributions of the simulated and real output data, respectively. Under the null hypothesis that the parameters are correctly calibrated, the asymptotic distribution of KS is equal to the supremum difference between two independent scaled Brownian bridges. In the multi-dimensional case, one can look at multiple KS-statistics, one for each dimension, and further use a Bonferroni correction to adjust the critical value. More precisely, when the dimension is K, we would look at the KS-statistic KS_k for each dimension of output $k = 1, \ldots, K$. If any of the KS_k is above the adjusted critical value $q_{1-\gamma/K}$, then we conclude that the simulation model is different from the real data. Put another way, if $\max_{k=1,\ldots,K} \mathrm{KS}_k > q_{1-\gamma/K}$, then we reject the hypothesis that the model is the same as reality.

It is known that the Bonferroni correction is conservative, especially when different dimensions of the outputs are highly dependent. The question is whether one can improve it without losing validity. This resembles the problem of so-called p-value aggregation (Ramdas et al. (2019); Vovk and Wang (2020)), which aims to construct tight family-wise p-values from merging multiple p-values in individual experiments. In the considered case, a natural alternative way to construct an aggregated statistic over all dimensions is the sum of individual KS-statistics, $\sum_{k=1}^{K} \mathrm{KS}_k$. Note that each KS_k has the same marginal distribution (with quantile $q_{1-\gamma}$), thus we can use Proposition 2 to derive a new critical value given by

$$\inf_{\alpha\in(0,\gamma/K)} \frac{K}{\gamma - K\alpha} \int_{1-\gamma+(K-1)\alpha}^{1-\alpha} q_u \mathrm{d}u.$$

The Kolmogorov probability density function is decreasing at its tail part. For γ sufficiently small (e.g., $\gamma < 0.3$), this critical value is sharp among all possible dependence structures of the KS_k, since the marginal densities are monotonically decreasing beyond $(1-\gamma)$-quantile. Particularly, for $K = 5$, Table 7 lists this new critical value[8] for different γ.

We illustrate our sum-of-KS statistic and newly derived critical values, and compare them with using the Bonferroni correction, in a multi-class queueing model (M/M/1/∞). In this model, there are 5 types of customers ($k = 1, \ldots, 5$), each with its own exponential service rate (μ_k) and Poisson arrival rate (λ_k). The

[8] It is known that if $\hat{F}_{\mathrm{real}}$ is continuous, then under the null hypothesis, the distribution of each $\sqrt{M}\mathrm{KS}_k$ converges to the Kolmogorov distribution as the sample size M goes to infinity. But the convergence rate is slow. We use the method in Vrbik (2018) for the asymptotic approximation (replacing x by $x + \frac{1}{6\sqrt{M}} + \frac{x-1}{4M}$ in the Kolmogorov cumulative distribution function $F(x)$).

Table 7: Critical values for $\sum_{k=1}^{5} \mathrm{KS}_k$ in the two-sample test (all sample sizes are M).

M \ γ	0.2	0.1	0.05	0.02	0.01
100	0.8875	0.9801	**1.0645**	1.1667	1.2384
1000	0.2833	0.3127	0.3394	0.3718	0.3945

Table 8: Parameters (μ_k, λ_k) of classes $k = 1, \ldots, 5$ in each configuration

Class \ Configuration	Both True	False μ	False λ	Both False
1. Slow-service-small-arrival	(6, ρ)	(6.05, ρ)	(6, 0.95ρ)	(6.05, 0.95ρ)
2. Slow-service-large-arrival	(6, 2ρ)	(6.05, 2ρ)	(6, 1.95ρ)	(6.05, 1.95ρ)
3. Medium-service-medium-arrival	(8, 1.6ρ)	(8.05, 1.6ρ)	(8, 1.45ρ)	(8.05, 1.45ρ)
4. Quick-service-small-arrival	(10, ρ)	(10.05, ρ)	(10, 0.95ρ)	(10.05, 0.95ρ)
5. Quick-service-large-arrival	(10, 2ρ)	(10.05, 2ρ)	(10, 1.95ρ)	(10.05, 1.95ρ)

system is first-come-first-served and starts from empty. Suppose we do not know the arrival and service time parameters in the model. On the other hand, suppose we have real output data on the average waiting times for each class of customers among 1000 total arrivals. Such a setting where only output- but not input-level data are observed can arise due to various administrative or operational constraints; see, e.g., Mandelbaum and Zeltyn (1998); Frey and Kaplan (2010); Whitt (1981); Goeva et al. (2019). Then, to validate a given set of parameter values in Table 8, we can generate simulation outputs from several conjectured configurations (four in our example) and run the aggregated KS tests described above, which treats the average waiting time of each customer class as one output dimension.

More precisely, we consider four parameter configurations that are listed in Table 8. The first column shows the true configuration and the rest are incorrectly conjectured. To facilitate the presentation and to test our approach on several ground-truth models, we define a model parameter $\rho = \sum_{k=1}^{5} \frac{\lambda_k}{\mu_k}$, which can be viewed as a summary of the traffic intensity. We experiment on three ground-truth settings: $\rho = 1.1, 1, 0.9$, representing scenarios with respectively long, medium and short waiting times. For each model parameter ρ, we conduct 1000 experimental repetitions, where in each repetition we independently generate a synthetic data set of size $M = 100$, run simulation with the same size on each of the four configurations depicted in Table 8, and then use our sum statistic and Bonferroni correction on KS to do multiple hypothesis tests. The results are summarized in Table 9.

Our sum-of-KS statistic is shown to be useful in the model where data across dimensions are highly and complicatedly dependent. We find in Table 9 that compared to Bonferroni correction, the new sum statistic has consistently slightly greater statistical power under various close-to-critical traffic intensities in this example. The basic reason is that the waiting times of different classes are highly dependent, and hence the sum statistic takes advantage from the bound (15) on quantile aggregation with dependence uncertainty. In case of near independence (e.g., small ρ), the Bonferroni correction is known to have a very good power, and it outperforms the sum-of-KS method. The drawback of this sum-of-KS statistic may be an overemphasis on the worst-case scenario. There are other methods of multivariate goodness-of-fit tests adapting well to

Table 9: Testing outputs of two methods. Here, the number of different classes is $K = 5$, the significant level is $\gamma = 0.05$, the sample size on the synthetic data and simulation runs for each configuration is $M = 100$ and the number of total arrivals is 1000. By computation, the critical value for the sum statistic is 1.0645 and that for Bonferroni correction is 0.2302. We use 1000 experimental repetitions. Type-I Error means the percentage of cases that the configuration with both true parameters in Table 8 is mistakenly rejected, a number set by us. Power records the percentage of cases that the wrong configuration is successfully rejected in our experiment.

Method \ Model	Long waiting ($\rho = 1.1$)		Medium waiting ($\rho = 1$)		Short waiting ($\rho = 0.9$)	
	Type-I Error	Power	Type-I Error	Power	Type-I Error	Power
Sum statistic	0.0140	**0.5907**	0.0100	**0.6460**	0.0120	**0.6353**
Bonferroni correction	0.0140	0.5887	0.0080	0.6327	0.0100	0.6247

the environment of dependence, such as Peacock's test and its later variants.[9]

G.2 Assembly line crew scheduling

The quantile aggregation problem is closely related to the problem of assembly line crew scheduling, which we explain in this section.

A manufacturing facility produces items which require the completion of n tasks in series. Suppose that there are m assembly lines (rows) and n operations (columns). Each operation has m crews to be assigned to each line. If the i-th operation is put in the j-th assembly line, it costs z_j^i units of time (or another type of resource) to complete the task. Therefore, we can use the $m \times n$ matrix in (EC.27), where the number z_j^i at (i, j)-position represents the processing time of the i-th crew in the j-th operation. The objective is to appropriately assign crews in each operation to the lines in order to minimize the makespan, that is, the maximum total processing time of all assembly lines. For $j = 1, \dots, n$, denote by μ_j the distribution measure for a discrete uniform distribution on the j-th column (m elements). The objective is to find an optimal arrangement of elements in each column to minimize the maximum row sum (the makespan). We denote the minimal makespan by $s_{\min}$. A similar problem appears in many other fields, e.g., in healthcare operations where usage of operating rooms among all types of elective surgeries is to be optimized. This problem is essentially the same as the matrix rotation problem explained in Appendix F, but minimizing the maximum row sum instead of maximizing the minimum row sum. These two problems can be converted into each other by simply putting a negative sign in front of all values z_j^i.

A few remarks on the problem of assembly line crew scheduling and that of quantile aggregation are needed. Denote by $q_{\min} = \inf_{\nu \in \Lambda(\boldsymbol{\mu})} q_1^-(\nu)$, the infimum value of the quantile aggregation problem with the same marginals. First, the problem of assembly line crew scheduling is know to be NP-complete; see Hsu (1984) and Haus (2015). Second, since each arrangement induces a dependence structure among random variables with marginal distributions $\mu_1, \dots, \mu_n$, the two problems are closely connected with one difference: the assembly line crew scheduling problem only allows for discrete dependence structures taking m different values in $\mathbb{R}^n$ (indeed, it can be seen as a quantile aggregation problem restricted on a discrete probability space of m states), whereas the quantile aggregation also allows for other dependence structures, such as

[9] We thank an anonymous referee for pointing out Peacock's test in this application.

independence. Third, $q_{\min}$ is a lower bound for $s_{\min}$, and RA can produce an upper bound for $s_{\min}$ (RA can be used to compute both the maximum of minimum row sum or minimum of maximum row sum). CB in (19) in Proposition 6 produces a lower bound B_{conv} on $q_{\min}$ and hence also on $s_{\min}$. Fourth, the difference between $q_{\min}$ and $s_{\min}$ is relatively small. These two values often coincide: as seen from our numerical examples, often RA coincides with CB, making the inequalities in

$$\mathrm{RA} \geqslant s_{\min} \geqslant q_{\min} \geqslant B_{\mathrm{conv}}$$

all exact. Fifth, one can also use the LP method in Appendix F to compute $q_{\min}$ when the values of m and n are small (see the numerical experiments in Appendix F). This yields a lower bound for $s_{\min}$.

Suppose that a matrix of representation is given by the left-hand side of (EC.33), with a makespan of $87+60+83=230$. Let X_i be a uniform discrete random variable valued on the i-th column of the matrix and μ_i be the corresponding distribution. For example ($i=1$), X_1 takes each value of the first column $\{44, 66, 67, 71, 87\}$ with probability $1/5$. For the discrete distributions μ_1, μ_2, μ_3, the minimal makespan is at least $\inf_{\nu\in\Lambda(\boldsymbol{\mu})} q_1^-(\nu)$. According to Proposition 6, $\sup_{\boldsymbol{\beta}\in\Delta_n} \sum_{i=1}^n R_{1-\beta_i-\beta_0,\beta_0}(\mu_i)$ serves as a lower bound for the minimal makespan and the maximizer $\boldsymbol{\beta}$ provides a hint to the optimal scheduling rule. In this example, the explicit bound is attainable: $\sup_{\boldsymbol{\beta}\in\Delta_3} \sum_{i=1}^3 R_{1-\beta_i-\beta_0,\beta_0}(\mu_i) = 160$ with the maximizer $\boldsymbol{\beta} = (0, 0.2, 0.6, 0.2)$. If an arrangement yields a minimal makespan of 160, it must optimal. Indeed, one optimal arrangement is given by the right-hand side of (EC.33).

$$\text{Convolution bound: } \begin{pmatrix} 44 & 10 & 24 \\ 66 & 32 & 37 \\ 67 & 48 & 41 \\ 71 & 57 & 43 \\ 87 & 60 & 83 \end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix} 87 & 10 & 43 \\ 71 & 60 & 24 \\ 67 & 48 & 41 \\ 44 & 32 & 83 \\ 66 & 57 & 37 \end{pmatrix} \tag{EC.33}$$

There are several algorithms in the literature for the problem of assembly line crew scheduling. Coffman and Sethi (1976) and Hsu (1984) naturally adopted greedy-type (largest first) methods. Coffman and Yannakakis (1984) improved and developed an algorithm to approximate the problem. Embrechts et al. (2013) proposed the RA in the context of risk management. These numerical algorithms provide an upper bound for the minimal makespan because they always return to a plausible scheduling rule.

The k-partitioning problem is a similar problem to the one in this section, as it can be solved by finding the minimal maximum row sum of a matrix; see Boudt et al. (2018). From a different perspective, our convolution bound provides analytical assistance for this type of problem. As it is a lower bound for the minimal makespan, if it is equal to the RA results, we can guarantee that the scheduling rule is optimal.