



City Research Online

City, University of London Institutional Repository

Citation: Asimit, V. (2023). Modeling Risk for CVaR-Based Decisions in Risk Aggregation. *Journal of Risk and Financial Management*, 16(5), 266. doi: 10.3390/jrfm16050266

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/30419/>

Link to published version: <https://doi.org/10.3390/jrfm16050266>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

MODELING RISK FOR CVAR-BASED DECISIONS IN RISK AGGREGATION

Alexandru V. Asimit

Bayes Business School, City, University, London, United Kingdom.

E-mail: asimit@city.ac.uk.

Yuriy Zinchenko

Department of Mathematics and Statistics, University of Calgary, Calgary, Canada.

E-mail: yzinchen@ucalgary.ca

Corresponding author. Phone: 1 403 220 4044. Fax: 1 403 282 5150.

May 7, 2023

Abstract

Measuring the risk aggregation is an important exercise for any risk bearing carrier. It is not restricted to evaluation of the known portfolio risk position only, and could include complying with regulatory requirements, diversification, etc. The main difficulty of risk aggregation is creating an underlying robust probabilistic model. It is an irrefutable fact that the uncertainty of the individual risks is much lower in its complexity, as compared to modeling the dependence amongst the risks. As a result, it is often reasonable to assume that individual risks are modeled in a robust fashion, while the exact dependence remains unknown, yet some of its traits may be made available due to empirical evidence or “good practice”. Our main contribution is to propose a numerical procedure that enables to identify the worst possible dependence scenario, when the risk preferences are modeled by the Conditional Value-at-Risk in the presence of dependence uncertainty. For portfolios with two risks, it is known that CVaR ordering coincides with the lower-orthant stochastic ordering of the underlying bi-variate distributions. As a by-product of our analysis, we show that no such extensions are possible to higher dimensions.

Subject classifications: Risk management; Conditional Value-at-Risk; Uncertainty modeling; Bilinear optimization; Linear Programming; Risk aggregation.

1. INTRODUCTION

Risk aggregation is a well-known strategy to reduce the overall risk held by a financial institution, insurance company or any other risk bearing carrier. Risk portfolios are often a summation of individual risks (or lines of business) and the risk bearing carrier is usually concerned with evaluating

the risk position for this portfolio so that regulatory requirements or business targets (such as diversification, shareholder value management constraints, etc.) are met. Within the insurance and banking industries, there are regulatory requirements that financial institutions need to meet by maintaining an appropriate level of capital at all times. These calculations take into account multiple sources of risk and all other factors that contribute to changes in the company's balance sheet within a specified period of time. Examples of such regulatory requirements include international Basel II/III banking supervision guidelines (e.g., see BCBS, 2016) and Swiss Solvency Test that applies to all Swiss based insurance and reinsurance companies (e.g., see Swiss Solvency Test, 2006), where the risk measurements are performed via the well-known risk measure *Conditional Value-at-Risk* (CVaR). This risk measure is introduced in the seminal paper of Rockafellar and Uryasev (2000) and has shown clear computational advantage in OR applications. A risk aggregation application in the context of the European Union insurance regulatory known as *Solvency II* is given in Asimit *et al.* (2016).

Many practical situations show that obtaining full knowledge of the dependence amongst a group of observed random variables is a very difficult task. It is an irrefutable fact that when modeling multi-variate risks, the estimation error is weighted towards determining the dependence amongst the risks. Common practice has shown that individual risks are estimated with higher confidence as compared to the dependence model between the variables. Unlike estimating individual risks, fitting the dependence model typically presents a great challenge, especially due to data scarcity. As a result, decision-makers usually commit to a somewhat arbitrarily chosen parametric model, but these ad-hoc choices lead to inadequate evaluations of the overall risk. Therefore, it may be more preferable to use qualitative information about the dependence and use a notion of realistic weakest and strongest dependence models amongst the observed risks instead. For example, knowing that the risks are positively associated would imply that the independence represents the weakest possible dependence, etc. Thus, it is more reasonable to assume that we have reliable models for individual random variables coupled with some partial knowledge of their association.

Many attempts have been made to resolve the problem of risk valuation under uncertainty modeling and more specific under dependence uncertainty. The literature on this topic is vast and we further give only a brief account of the related work. One direction of research typically pursued in the OR literature is to adapt recent methodologies from the so-called robust optimization. For example, in robust portfolio optimization one typically assumes that a decision-maker has some partial information about the joint distribution function amongst the risks. In order to incorporate the uncertainty, several notions of the worst-case risk measure have been proposed. For example, El Ghaoui *et al.* (2003) and Zymler *et al.* (2013) discuss this problem in the context of VaR-based

optimal portfolio selection. The same problem is investigated in Zhu *et al.* (2009) and Huang *et al.* (2010), where decisions are made on the worst-case CVaR; a related insurance setting is discussed in Asimit *et al.* (2017, 2019) and Balbás *et al.* (2011). Some attention has been devoted to computing bounds on CVaR with moment information. For example, in Bertsimas *et al.* (2004) sharp explicit bounds are obtained with the first two moments and in the work of Bertsimas and Popescu (2002), where a more general numerical convex-optimization based approach is obtained. Recall that robust versions of the above moment-based models may be developed in principle, relying on the so-called robust optimization techniques (e.g., Ben-Tal *et al.*, 2009). Interesting connections between chance-constrained and robust optimization with relationship to CVaR are established in Chen *et al.* (2009). Other risk measures (beyond CVaR) are available in the literature; e.g., the higher moment risk measure that is investigated in Gómez *et al.* (2022).

The main contribution of our paper is to propose a method to evaluate sharp lower and upper bounds for the CVaR-based aggregate risk level under dependence uncertainty. Specifically, we assume that the bounds on the cumulative multi-variate distribution are available, as well as that we have the full knowledge of the individual risk distributions. Here, the partial information about dependence is given by the lower-orthant stochastic ordering type constraints. Arguably, the most practically relevant examples of such type of constraints are the so-called positive and negative quadrant dependence models. The practical advantage of using the above dependence models is that we can test the statistical significance of such properties (see Gijbels and Sznajder, 2013). In other words, the validity of restricting the range of possible dependence models may be statistically verified using the observed data. The latter plausible dependence provides us with the main motivation to include lower-orthant type restrictions in our model.

From the methodological perspective, our numerical method is based on (convex) optimization techniques and specifically, *Bilinear* and *Linear Programming (LP)*. Interestingly, despite the associated optimization problem being bilinear –and thus non-convex– in nature, we show that the problem’s objective function still retains a strong structural property, namely, it is convex in every argument, and in turn, the convexity provides the basis for efficient computations. Despite a seeming symmetry of the two problems, evaluating the sharp lower bound on CVaR appears to be more of a challenge, as compared to computing the sharp upper bound. This is as substantiated by both the complexity analysis of the proposed method, and the numerical results.

It has been known that CVaR respects the so-called lower-orthant stochastic ordering for two-dimensional portfolios (Chapter 6.2.6 of Denuit *et al.*, 2005). Yet no similar result has been established or disproved for higher dimensions. As a bi-product of our analysis, using elementary LP

techniques, we show that no such extensions are possible, and give insights as to why this is the case.

The paper is organised as follows. Section 2 presents our model for determining sharp upper and lower bounds on the CVaR-based aggregate risk level. Sections 3 and 4 describe the approach to compute the lower and upper bounds, respectively. Section 5 contains the numerical experiments and analysis, while Section 6 discusses the behavior of CVaR in multi-variate settings under the lower-orthant and other orderings. Our final comments and conclusions are summarised in Section 7.

2. MODEL SETTING

The notation relies on lower case letters t, α, x, \dots for deterministic quantities and capital letters Z, X, \dots for random variables. Likewise, we use capital letters F_i, Π, \dots to denote functions. Bold letters such as $\mathbf{x}, \mathbf{i}, \dots$ and \mathbf{X}, \dots are reserved for deterministic and random vectors, respectively; likewise, we use \mathbf{A}, \dots for vector-valued functions. Capital script letters $\mathcal{I}, \mathcal{M}, \dots$ are used for sets.

Let $\mathbf{X} = (X_1, \dots, X_n)$ denote an n -variate random vector, and let $Z = \sum_{i=1}^n X_i$ be the sum of n possibly dependent risks. The VaR of a generic loss variable Z at confidence level α , $\text{VaR}_\alpha(Z)$, represents the α -quantile of Z . Mathematically, $\text{VaR}_\alpha(Z) := \inf\{z \in \mathfrak{R} : \Pr(Z \leq z) \geq \alpha\}$, where $\inf \emptyset = \infty$. The CVaR at confidence level α , $\text{CVaR}_\alpha(Z)$, evaluates the expected loss amount incurred under the worst $100 \times (1 - \alpha)\%$ loss scenarios of Z . The CVaR has multiple formulations in the literature (Acerbi and Tasche, 2002), but in the present paper, we only refer to the following representation (Rockafellar and Uryasev, 2000),

$$\text{CVaR}_\alpha(Z) := \text{VaR}_\alpha(Z) + \frac{1}{1 - \alpha} \mathbf{E}(Z - \text{VaR}_\alpha(Z))_+ = \min_{t \in \mathfrak{R}} t + \frac{1}{1 - \alpha} \mathbf{E}(Z - t)_+,$$

where $\mathbf{E}(\cdot)$ is the expectation and $z_+ = \max\{z, 0\}$.

Let us assume that we have a portfolio consisting of n risks $\mathbf{X} = (X_1, \dots, X_n)$. The *cumulative distribution functions* (c.d.f.) of each individual risk X_i is $F_i(\cdot)$ and is assumed to be known for all $1 \leq i \leq n$, and we write $X_i \sim F_i$. Moreover, we assume that the dependence between the risks, i.e. the multi-variate distribution $F(\mathbf{x}) = \Pr(\mathbf{X} \leq \mathbf{x})$ of \mathbf{X} , is unknown, but some prior knowledge about the association amongst risks is available. Namely, the set of feasible distributions is given by

$$\mathcal{F} = \{F : \underline{F}(\mathbf{x}) \leq F(\mathbf{x}) \leq \bar{F}(\mathbf{x}), \forall \mathbf{x} \in \mathfrak{R}^n, X_i \sim F_i\}, \quad (2.1)$$

where \underline{F} and \bar{F} are some n -dimensional joint c.d.f.'s that define the set of acceptable dependence models. Note that the above assumption provides a lower and upper bound for \mathbf{X} in the lower-orthant stochastic ordering sense. Recall that two random vectors \mathbf{X} and \mathbf{Y} in \mathfrak{R}^n are *lower-orthant*

ordered, written $\mathbf{X} \preceq_{lo} \mathbf{Y}$, if $\Pr(\mathbf{X} \leq \mathbf{x}) \leq \Pr(\mathbf{Y} \leq \mathbf{x})$ for all $\mathbf{x} \in \mathfrak{R}^n$. It is known that the comonotonic¹ dependence $F^c(\mathbf{x}) := \min_{i=1, \dots, n} F_i(\mathbf{x}_i)$ gives the sharp upper bound on c.d.f. with prescribed marginals F_i . Thus, if there is no upper bound specified, without loss of generality we may set $\bar{F}(\mathbf{x}) = F^c(\mathbf{x})$. On the other hand, given the marginals, it is impossible to construct the sharp lower bound on c.d.f. for $n \geq 3$. Thus, when the lower bound \underline{F} is not known a priori, it is not so clear what should be used in place of \underline{F} , besides trivial choices.

The main aim of the paper is to compute sharp lower and upper bounds on $\text{CVaR}_\alpha(Z)$,

$$\inf_{F \in \mathcal{F}} \text{CVaR}_\alpha(Z) \quad \text{and} \quad \sup_{F \in \mathcal{F}} \text{CVaR}_\alpha(Z), \quad (2.2)$$

where $\mathbf{X} \sim F$. We approximate the solutions to (2.2) by assuming X_i 's to be discrete random variables, i.e. by considering a sample of size m^n from our population \mathbf{X} . Namely, it is assumed that \mathbf{X}_i takes the values $x_{i,1} \leq \dots \leq x_{i,m}$ with equal probability $1/m$, but we do not know the joint probability amongst the risks, represented by p.m.f.

$$\mathbf{p}_{i_1, \dots, i_n} = \Pr\left(\mathbf{X} = (x_{1, i_1}, \dots, x_{n, i_n})\right), \text{ for all } 1 \leq i_1, \dots, i_n \leq m.$$

Note that if \mathbf{X} is continuous compactly supported random vector, one can use the above discretization to approximate its distribution to within the desired accuracy by increasing m . The F_i -equivalent discrete marginal distributions are standardised and assumed to be uniform. This choice is motivated by the common sampling procedure using copulas, if parametric models for the marginal distributions are available. It is also motivated by the practical considerations on availability of historical data. The methods in the paper can be easily adapted to arbitrary marginals. However, this comes at the unnecessary expense of further complicating the notation. The c.d.f. bounds \underline{F} and \bar{F} are represented by discrete vectors $\underline{\pi}$ and $\bar{\pi}$. Likewise, we denote the aggregate risk sample by \mathbf{z} , $\mathbf{z}_i = z_{i_1, \dots, i_n} = \sum_{j=1}^n x_{j, i_j}$, where the multi-index $\mathbf{i} = (i_1, \dots, i_n)$ runs over all m^n possible values $\mathbf{i} \in \mathcal{I}$ with $i_j \in \{1, \dots, m\}$. The values of \mathbf{z}_i are only partially ordered.

¹For a multivariate vector (X_1, \dots, X_n) comonotonicity formally is defined as follows: there exists a random vector Z and non-decreasing functions f_k for all $1 \leq k \leq n$ such that $\Pr(X_k = f_k(Z)) = 1$ for all $1 \leq k \leq n$.

Thus, in order to approximate (2.2), we need to compute

$$\begin{aligned}
\underline{\text{CVaR}}_\alpha &:= \inf_{\mathbf{p}} \min_{t \in \mathfrak{R}} t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \mathbf{p}_i \\
\text{s.t.} \quad &\underline{\pi}_i \leq \sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_j \leq \bar{\pi}_i, \text{ for all } \mathbf{i} \in \mathcal{I}, \\
&\sum_{\mathbf{i}: i_j=k} \mathbf{p}_i = \frac{1}{m}, \text{ for all } j = 1, \dots, n, k = 1, \dots, m, \\
&\sum_{\mathcal{I}} \mathbf{p}_i = 1, \mathbf{p} \geq 0,
\end{aligned} \tag{2.3}$$

and

$$\begin{aligned}
\overline{\text{CVaR}}_\alpha &:= \sup_{\mathbf{p}} \min_{t \in \mathfrak{R}} t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \mathbf{p}_i \\
\text{s.t.} \quad &\underline{\pi}_i \leq \sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_j \leq \bar{\pi}_i, \text{ for all } \mathbf{i} \in \mathcal{I}, \\
&\sum_{\mathbf{i}: i_j=k} \mathbf{p}_i = \frac{1}{m}, \text{ for all } j = 1, \dots, n, k = 1, \dots, m, \\
&\sum_{\mathcal{I}} \mathbf{p}_i = 1, \mathbf{p} \geq 0,
\end{aligned} \tag{2.4}$$

where the multi-index inequalities $\mathbf{j} \leq \mathbf{i}$ are interpreted component-wise. Note that the marginal density constraints $\sum_{\mathbf{i}: i_j=k} \mathbf{p}_i = \frac{1}{m}$ are stated explicitly as part of the formulation, although, we could absorb these constraints into tighter upper and lower c.d.f. bounds.

3. COMPUTABLE LOWER BOUND

3.1. Reduction to parametric LP. Let us define the *value function* as

$$\begin{aligned}
\underline{\text{val}}(t) &:= \inf_{\mathbf{p}} t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \mathbf{p}_i \\
\text{s.t.} \quad &\underline{\pi}_i \leq \sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_j \leq \bar{\pi}_i, \text{ for all } \mathbf{i} \in \mathcal{I}, \\
&\sum_{\mathbf{i}: i_j=k} \mathbf{p}_i = \frac{1}{m}, \text{ for all } j = 1, \dots, n, k = 1, \dots, m, \\
&\sum_{\mathcal{I}} \mathbf{p}_i = 1, \mathbf{p} \geq 0,
\end{aligned} \tag{3.1}$$

and note that evaluating $\underline{\text{val}}(t)$ for a fixed t corresponds to solving an LP problem. This is critical to the design of our computational approach to solving (2.3), i.e. determining

$$\underline{\text{CVaR}}_\alpha = \inf_{t \in \mathfrak{R}} \underline{\text{val}}(t).$$

Since the solution to a moderately sized LP problem can be typically computed in reasonable time, to get an initial sense of what range $\underline{\text{CVaR}}_\alpha$ may fall into, one may simply compute a few values

$\underline{\text{val}}(t)$ for some sample values t_1, t_2, \dots . We extend on this basic idea by combining it with a few more observations that follow. Recall that evaluating $\underline{\text{CVaR}}_\alpha$ corresponds to solving the so-called bilinear optimization problem, which are notoriously difficult, due to the inherent non-convexity of the objective with potentially many local minima.

3.2. Compact support in t . We now claim that in order to compute $\underline{\text{CVaR}}_\alpha$, it is unnecessary to perform an exhaustive search over all possible values of $t \in \mathfrak{R}$.

Theorem 3.1. *Denote $\underline{t} = \min_{\mathcal{I}} \mathbf{z}_i$ and $\bar{t} = \max_{\mathcal{I}} \mathbf{z}_i$. Then, the following holds*

$$\inf_{t \in \mathfrak{R}} \underline{\text{val}}(t) = \min_{t \in [\underline{t}, \bar{t}]} \underline{\text{val}}(t).$$

Proof. Assume first that $t > \bar{t}$: since $(\mathbf{z}_i - t)_+ = 0$ for all \mathbf{i} , we have $\underline{\text{val}}(t) = t$ and thus, the value function $\underline{\text{val}}(t)$ is increasing for any $t > \bar{t}$.

Consider now the case of fixed t such that $t < \underline{t}$. Let \mathbf{p} denote an optimal probability distribution resolving $\underline{\text{val}}(t)$ at t , and $\underline{\mathbf{p}}$ denote an optimal probability distribution resolving $\underline{\text{val}}(t)$ at \underline{t} . Denoting $\Delta t = \underline{t} - t \geq 0$, we have

$$\begin{aligned} \underline{\text{val}}(t) - \underline{\text{val}}(\underline{t}) &= \left(t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \mathbf{p}_i \right) - \left(\underline{t} + \frac{1}{1-\alpha} \sum_{\mathcal{I}} (\mathbf{z}_i - \underline{t})_+ \underline{\mathbf{p}}_i \right) \\ &= -\Delta t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} \left((\mathbf{z}_i - t) \mathbf{p}_i - (\mathbf{z}_i - t - \Delta t) \underline{\mathbf{p}}_i \right) \\ &= -\Delta t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} \left((\mathbf{z}_i - t) (\mathbf{p}_i - \underline{\mathbf{p}}_i) + \Delta t \underline{\mathbf{p}}_i \right) \\ &= \frac{\alpha}{1-\alpha} \Delta t + \frac{1}{1-\alpha} \sum_{\mathcal{I}} (\mathbf{z}_i - t) (\mathbf{p}_i - \underline{\mathbf{p}}_i) \\ &\geq \frac{\alpha}{1-\alpha} \Delta t \geq 0, \end{aligned}$$

where the last identity follows from feasibility of $\underline{\mathbf{p}}$, namely, $\sum_{\mathcal{I}} \underline{\mathbf{p}}_i = 1$, while the next to last inequality follows from \mathbf{p} being optimal solution corresponding to t , which in turn implies

$$\sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \mathbf{p}_i \leq \sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \underline{\mathbf{p}}_i.$$

Therefore, $\underline{\text{val}}(t)$ is decreasing for $t < \underline{t}$. Finally, since $\underline{\text{val}}(t)$ is a continuous function minimised over a compact set, we can replace inf with min. \square

3.3. Key properties of the value function. Since evaluation of the value function can be reduced to an LP with parametric objective, we can establish the next proposition.

Proposition 3.1. *The function $\underline{\text{val}}(t)$ is a piecewise linear, continuous function, concave on every sub-interval $[\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)}]$ where $\mathbf{z}_{(\ell)}$ corresponds to re-indexing of $\mathbf{z}_{\mathbf{i}}$ values in non-decreasing order so that $\mathbf{z}_{(\ell)} \leq \mathbf{z}_{(\ell+1)}$ for all $\ell = 1, \dots, m^n$. Furthermore, $\underline{\text{val}}(t)$ has finitely many linear segments.*

Proof. Observe that restricting $t \in [\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)}]$, we can write $\underline{\text{val}}(t) = t + \frac{1}{1-\alpha} \underline{\mathbf{v}}(t)$ with

$$\begin{aligned}
\underline{\mathbf{v}}(t) &:= \inf_{\mathbf{p}, \mathbf{s}, \bar{\mathbf{s}}} \sum_{\mathbf{z}_{\mathbf{i}} \geq \mathbf{z}_{(\ell+1)}} (\mathbf{z}_{\mathbf{i}} - t) \mathbf{p}_{\mathbf{i}} \\
\text{s.t.} \quad &\sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_{\mathbf{j}} - \mathbf{s}_{\mathbf{i}} = \underline{\pi}_{\mathbf{i}}, \text{ for all } \mathbf{i} \in \mathcal{I}, \\
&\sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_{\mathbf{j}} + \bar{\mathbf{s}}_{\mathbf{i}} = \bar{\pi}_{\mathbf{i}}, \text{ for all } \mathbf{i} \in \mathcal{I}, \\
&\sum_{\mathbf{i}: i_j = k} \mathbf{p}_{\mathbf{i}} = \frac{1}{m}, \text{ for all } j = 1, \dots, n, k = 1, \dots, m, \\
&\sum_{\mathcal{I}} \mathbf{p}_{\mathbf{i}} = 1, \\
&\mathbf{p}, \mathbf{s}, \bar{\mathbf{s}} \geq 0
\end{aligned} \tag{3.2}$$

denoting the *partial value function*. In turn, determining $\underline{\mathbf{v}}(t)$ may easily be recognized as a linear optimization problem in standard minimisation form

$$\begin{aligned}
\underline{\mathbf{v}}(t) &= \min_{\mathbf{u}} (\mathbf{c} + t \Delta \mathbf{c})^T \mathbf{u} \\
\text{s.t.} \quad &\mathbf{A} \mathbf{u} = \mathbf{b}, \\
&\mathbf{u} \geq 0.
\end{aligned}$$

Note that $\mathbf{u} = (\mathbf{p}; \mathbf{s}; \bar{\mathbf{s}})$ is a vector of variables of dimension $d = 3m^n$, $\mathbf{A} : \mathbb{R}^d \rightarrow \mathbb{R}^r$ is a linear function encoded as $d \times r$ matrix with $r = 2m^n + mn + 1$ rows and $\mathbf{b} \in \mathbb{R}^r$ represents the affine equality constraints stated for $\underline{\mathbf{v}}(t)$. The t -parametric objective $\mathbf{c} + t \Delta \mathbf{c}$ corresponds to

$$\mathbf{c}_{\mathbf{i}} = \begin{cases} \mathbf{z}_{\mathbf{i}}, & \mathbf{i} : \mathbf{z}_{\mathbf{i}} \geq \mathbf{z}_{(\ell+1)}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{with} \quad \Delta \mathbf{c}_{\mathbf{i}} = \begin{cases} -1, & \mathbf{i} : \mathbf{z}_{\mathbf{i}} \geq \mathbf{z}_{(\ell+1)}, \\ 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

where we allow a slight abuse of notation when indexing \mathbf{c} and $\Delta \mathbf{c}$ by multi-index \mathbf{i} .

Clearly, $\underline{\mathbf{v}}(t)$ is a continuous piecewise linear concave function of t . By enumerating the total number of possible bases, standard LP sensitivity analysis implies that on a given subinterval $t \in [\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)}]$ function $\underline{\mathbf{v}}(t)$, and therefore $\underline{\text{val}}(t)$, consists of at most $\binom{d}{r}$ linear segments. Since we have at most $m^n - 1$ of such subintervals, we conclude that $\underline{\mathbf{v}}(t)$ consists of at most $(m^n + 1) \binom{d}{r}$ linear segments, which also includes two end subintervals $(-\infty, \mathbf{z}_{(1)}]$ and $[\mathbf{z}_{(m^n)}, \infty)$. \square

The above bound on the number of linear segments comprising $\underline{\mathbf{v}}(t)$ is very crude. Not only do we take a very pessimistic bound $\binom{d}{r}$ on the number of vertices of a very special polytope that

describes the feasible probability distributions, we also ignore a special “monotonic” structure in perturbations to the objective vector. Consequently, it is quite natural to expect the number of such segments to be much smaller.

The above proposition, based on classical sensitivity analysis for LP, albeit correct, may be misleading while designing a numerical scheme for minimising $\underline{\text{val}}(t)$. Specifically, the asserted piecewise concavity of $\underline{\text{val}}(t)$ may suggest a potential existence of several local minima, see Figure 3.1(a). We remedy this in the next theorem, which gives a complete characterization of the partial value function. Along the way, we drastically reduce an upper bound on the number of linear segments comprising $\underline{\text{val}}(t)$.

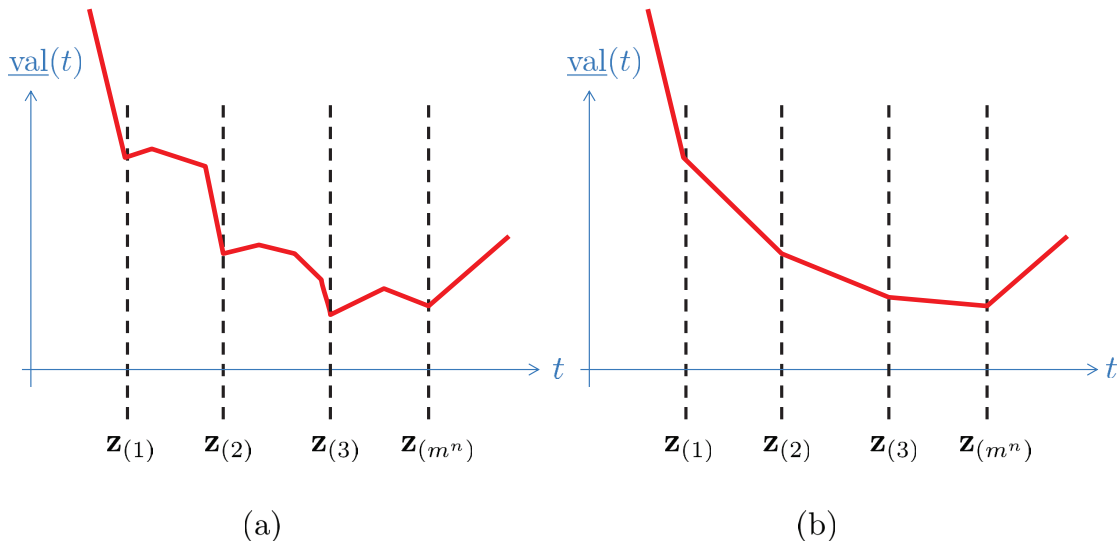


FIGURE 3.1. Perceived behaviour of $\underline{\text{val}}(t)$: (a) strict piecewise concavity, (b) convexity.

Theorem 3.2. *The function $\underline{\text{v}}(t)$ is continuous, non-negative and non-increasing satisfying $\underline{\text{v}}(t) = 0$ for $t \geq \mathbf{z}_{(m^n)}$ and $\underline{\text{v}}'(t) = -1$ for $t \leq \mathbf{z}_{(1)}$. Moreover, $\underline{\text{v}}(t)$ is convex on \mathfrak{R} and linear on every sub-interval $[\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)}]$.*

Proof. Continuity and the tail-end behaviour of $\underline{\text{v}}(t)$ are established in the proof of Proposition 3.1 and Theorem 3.1. Examining variational formulation (3.1), we easily note the non-negativity and monotonicity of the partial value function, with the latter due to the objective coefficients $(\mathbf{z}_i - t)_+$ being monotone in t . Linearity on $[\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)}]$ follows as a consequence of convexity –to be established shortly– and piecewise concavity in Proposition 3.1. It remains to show convexity.

We show the convexity property by contradiction. First, introduce

$$v_{\mathbf{p}}(t) := \sum_{\mathcal{I}} (\mathbf{z}_i - t)_+ \mathbf{p}_i$$

to be the partial value function restricted to a given feasible \mathbf{p} . Observe that $v_{\mathbf{p}}$ is convex, piecewise linear non-increasing, and its derivative $v'_{\mathbf{p}}(t)$, whenever defined, corresponds to the dot product of \mathbf{p} with the corresponding sub-vector of $\Delta \mathbf{c}$ as in (3.3). Thus, $v'_{\mathbf{p}}(t)$ is non-decreasing whenever defined. We also note that as t passes from the interval $[\mathbf{z}_{(\ell-1)}, \mathbf{z}_{(\ell)}]$ to $[\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)}]$, the number of -1 entries, i.e. the non-zeros in $\Delta \mathbf{c}$ is reduced by at least one.

If $\underline{v}(t)$ is strictly concave, there exists a cross-over point $t^=$ characterised by $t^- < t^= < t^+$ and corresponding optimal distributions \mathbf{p}^- and \mathbf{p}^+ resolving (3.2) such that $v_-(t^=) = v_+(t^=)$ with derivatives satisfying $v'_-(t^-) > v'_+(t^-)$ and $v'_-(t^+) > v'_+(t^+)$, where $v_-(t) := v_{\mathbf{p}^-}(t)$ and $v_+(t) := v_{\mathbf{p}^+}(t)$. Note that t^- and t^+ may be chosen close enough to $t^=$ to warrant differentiability of the corresponding piecewise linear v_+, v_- on $[t^-, t^=)$ and $(t^=, t^+]$. Furthermore, without loss of generality, we may assume that both v_-, v_+ have either at most one break-point at $\mathbf{z}_{(\ell)} = t^=$ for some ℓ with $t^- \in (\mathbf{z}_{(\ell-1)}, \mathbf{z}_{(\ell)})$ and $t^+ \in (\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)})$, or no break-point at all with $t^-, t^+ \in (\mathbf{z}_{(\ell)}, \mathbf{z}_{(\ell+1)})$, as can be seen in Figure 3.2. By re-scaling and shifting t we can also assume $t^= \equiv 0$ and $-t^- = t^+ = \frac{1}{2}$. With the above notation, we have $\underline{v}(t) = \min\{v_-(t), v_+(t)\}$ for $t \in [-1/2, 1/2]$ and $v'_-(t) > v'_+(t)$ for $t \in [-1/2, 0) \cup (0, 1/2]$.

Denote $\mathbf{p}(\tau) = \tau \mathbf{p}^- + (1 - \tau) \mathbf{p}^+$, $\tau \in [0, 1]$. Note that $\mathbf{p}(\tau)$ is feasible since the feasible region of (3.2) is convex. Further, let us examine $v(\tau) := v_{\mathbf{p}(\tau)}(1/2 - \tau)$. By the Fundamental Theorem of Calculus, we obtain

$$v(1/2) = v(0) + \int_0^{1/2} v'_\tau(\tau) d\tau = \underline{v}(t^+) + \int_0^{1/2} v'_{\mathbf{p}(\tau)}(\tau) d\tau < \underline{v}(t^=),$$

where the inequality is due to the fact that $|v'_{\mathbf{p}(\tau)}(\tau)| < |v'_+(\tau)|$, for all $\tau \in (0, 1/2)$, since \mathbf{p}^- ; consequently, $\mathbf{p}(\tau)$ carries less probability mass over the support of $\Delta \mathbf{c}$ at t^+ as compared to \mathbf{p}^+ . Since $\underline{v}(t^=)$ is supposed to be the smallest over all feasible \mathbf{p} at $t^=$, the contradiction is conspicuous. This completes the proof. \square

3.4. Two computational approaches. We now present two computational schemes for computing the sharp lower bound $\underline{\text{CVaR}}_\alpha$ on CVaR given the constraints on the risks' c.d.f.. The schemes are aimed at illustrating the advantages of exploiting the inherent structure of Problem 2.3 and range in order of complexity, as well as the perceived numerical efficiency. The latter is further substantiated in Section 5.

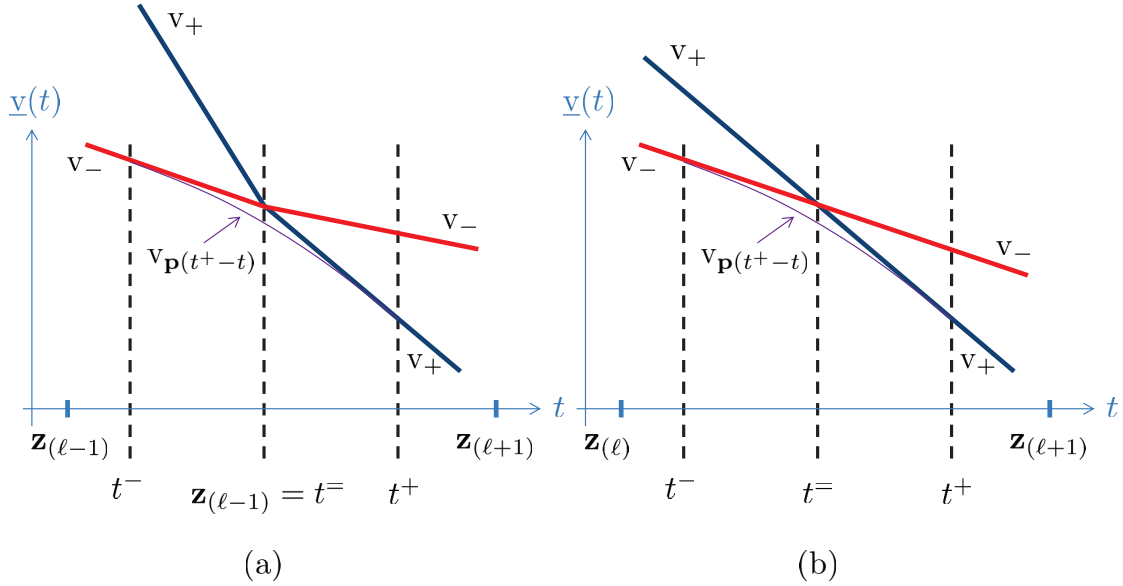


FIGURE 3.2. Hypothetical concavity of $\underline{v}(t)$: (a) $\mathbf{z}^{(\ell)} = t^=$, (b) $\mathbf{z}^{(\ell)} \neq t^=$.

3.4.1. *Naive scheme.* Observe that the piecewise concavity of $\underline{\text{val}}(t)$ established in Proposition 3.1 implies that the minimum of the value function may only occur at the end points of each interval $[\mathbf{z}^{(\ell)}, \mathbf{z}^{(\ell+1)}]$. Therefore, it suffices to compute $\underline{\text{val}}(\mathbf{z}_i)$ for all i and take the minimum value. This gives rise to the *naive scheme*.

Clearly, the naive scheme requires access to an LP solver and runs in finite time. However, it requires solving a large number –namely m^n – of (3.2)-type optimization problems, where the problem dimensions also grow proportional to m^n . As a result, the procedure may become very computationally expensive for even modest values of m and n . Further effort can be put towards reducing the computational requirements imposed by the naive scheme. For example, the LP problems for evaluating $\underline{\text{val}}(t)$ differ only in the objective function, and thus, may be well-suited for the so-called warm-start techniques as in simplex-type algorithms. In turn, the use of warm-starting may speed up solution times.

3.4.2. *Epigraph scheme.* Unlike the naive scheme, here we aim to take full advantage of the uncovered convexity of the value function. This not only allows us to greatly reduce the computational efforts required to determine the exact value $\underline{\text{CVaR}}_\alpha$, but also permits to introduce an alternative termination criteria, when only an approximate answer is required within some given absolute precision $\varepsilon > 0$.

We recall that an epigraph of a convex function can be obtained as an intersection of half-spaces. In the case of a smooth function the half-spaces correspond to tangent hyperplanes, and in the case of non-differentiable functions one may use half-spaces defined by sub-gradients. Thus, given two consecutive values $t^- < t^+$ of t with appropriately defined derivatives $\underline{\text{val}}'(t)$ of $\underline{\text{val}}(t)$, with values $v_- := \underline{\text{val}}(t^-)$, $v_+ := \underline{\text{val}}(t^+)$ and $v'_- := \underline{\text{val}}'(t^-) < 0$, $v'_+ := \underline{\text{val}}'(t^+) > 0$, we know that the minimal value $\underline{\text{val}}^*$ of $\underline{\text{val}}(t)$ corresponds to some $t^* \in [t^-, t^+]$. In addition, we have $\underline{\text{val}}^* \in [\min(v_-, v_+), \underline{\text{val}}(\tilde{t})]$, where

$$\tilde{t} = \frac{v_- - v_+ + v'_+ t^+ - v'_- t^-}{v'_+ - v'_-}$$

is the intersection of the supporting hyperplanes $v'_-(t - t^-) + v_-$ and $v'_+(t - t^+) + v_+$. This can be seen in Figure 3.3. To refine the interval $[t^-, t^+]$ and our estimate on $\underline{\text{val}}^*$, we can take the mid-point of the interval and adjust either t^-, v_-, v'_- or t^+, v_+, v'_+ accordingly.

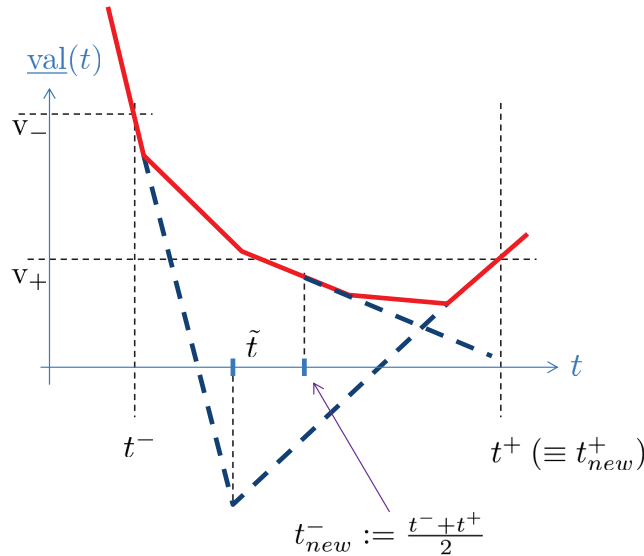


FIGURE 3.3. Epigraph scheme.

Assuming that the data are given by $\alpha, m, n, \underline{\pi}, \bar{\pi}, \mathbf{z}$, the scheme may be defined recursively as a function whose declaration is given below using MATLAB notation

$$\text{function } [t^-, t^+, v_-, v_+, v'_-, v'_+] = \text{Epigraph}(t^-, t^+, v_-, v_+, v'_-, v'_+),$$

and is defined as follows.

Input: $t^- < t^+, v'_- \leq 0, v'_+ \geq 0, v_-, v_+$, problem data.

- (0) Set $\underline{\text{val}}^* := \min\{v_-, v_+\}$,
- (1) compute \tilde{t} and $v := \underline{\text{val}}(\tilde{t})$ by solving (3.1), recovering the optimal probability distribution $\tilde{\mathbf{p}}$,
- (2) if $v = \underline{\text{val}}^*$ then return,
- (3) set $v' := 1 + \frac{1}{1-\alpha} \sum_{\mathcal{I}} \Delta \mathbf{c}_i \tilde{\mathbf{p}}_i$ with $\Delta \mathbf{c}$ corresponding to \tilde{t} as in (3.3),
- (4) if $v' \leq 0$ set $t^- := \tilde{t}, v_- := v, v'_- := v'$,
- (5) if $v' > 0$ set $t^+ := \tilde{t}, v_+ := v, v'_+ := v'$,
- (6) invoke $\text{Epigraph}(t^-, t^+, v_-, v_+, v'_-, v'_+)$.

Output: $\underline{\text{val}}^* = \min_{t \in [t^-, t^+]} \underline{\text{val}}(t)$.

Clearly, in order to get $\underline{\text{CVaR}}_\alpha$, we need to invoke $\text{Epigraph}(t^-, t^+, v_-, v_+, v'_-, v'_+)$ with initial values $t^- = \underline{t} (\equiv \mathbf{z}_{(1)})$ and $t^+ = \bar{t} (\equiv \mathbf{z}_{(m^n)})$. If one desires to terminate the procedure once the absolute precision ε is reached such that $\underline{\text{val}}^* \leq \underline{\text{CVaR}}_\alpha \leq \underline{\text{val}}^* + \varepsilon$, it suffices to replace the function termination criteria $\underline{\text{val}} - v = 0$ with $\min\{-v'_-(\tilde{t} - t^-), v_+(t^+ - \tilde{t})\} \leq \varepsilon$.

The convexity of $\underline{\text{val}}(t)$ and its tail behaviour from Theorem 3.2, we know that the fastest decrease rate of the value function does not exceed $\left|1 - \frac{1}{1-\alpha}\right| = \frac{\alpha}{1-\alpha}$ and therefore,

$$|\underline{\text{val}}(t) - \underline{\text{val}}(t + \Delta t)| \leq \frac{\alpha \Delta t}{1-\alpha}, \quad \text{for all } t, \Delta t.$$

Thus, in order to achieve an ε precision, it suffices to have $t^+ - t^- \leq \frac{\varepsilon(1-\alpha)}{\alpha}$. In turn, recalling that at every iteration of the scheme the interval $[t^-, t^+]$ gets halved, we conclude that the absolute ε precision can be attained in at most $\log_2 \left(\frac{1}{\varepsilon} \cdot \frac{\alpha(\bar{t}-\underline{t})}{(1-\alpha)} \right)$ recursive calls, where the dominant work belongs to solving an LP instance of the form (3.2).

In a nutshell, although the epigraph scheme still relies on solving multiple LP instances, in order to recover $\underline{\text{CVaR}}_\alpha$ for fixed n , its worst-case run-time is bounded from above as a polynomial function of the problem input. Also, when an approximate solution is sufficient, one would expect the number of calls to the LP solver to be dramatically less than m^n , as compared to the naive scheme. We also note that the epigraph procedure is defined recursively only in an attempt to improve clarity of exposition. Clearly, the procedure can be unrolled into `if ... else ...` statements with no recursion. Just as with the naive scheme, one may try to take advantage of warm-starting capabilities of an LP solver in an attempt to speed up computational times required.

4. COMPUTABLE UPPER BOUND

It turns out that despite apparent similarities between Problems (2.3) and (2.4), the complexity of evaluating $\underline{\text{CVaR}}_\alpha$ is quite different from that of $\overline{\text{CVaR}}_\alpha$. Namely, the calculation of $\overline{\text{CVaR}}_\alpha$ is

much simpler. We first establish an essential property that is needed for proving the main result of this section.

Proposition 4.1. *The max-value function $\overline{\text{val}}(t) = t + \frac{1}{1-\alpha} \bar{v}(t)$, where*

$$\begin{aligned} \bar{v}(t) &:= \max_{\mathbf{p}} \sum_{\mathcal{I}} (\mathbf{z}_{\mathbf{i}} - t)_+ \mathbf{p}_{\mathbf{i}} \\ \text{s.t.} \quad &\underline{\pi}_{\mathbf{i}} \leq \sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_{\mathbf{j}} \leq \bar{\pi}_{\mathbf{i}}, \text{ for all } \mathbf{i} \in \mathcal{I}, \\ &\sum_{\mathbf{i}: i_j=k} \mathbf{p}_{\mathbf{i}} = \frac{1}{m}, \text{ for all } j = 1, \dots, n, k = 1, \dots, m, \\ &\sum_{\mathcal{I}} \mathbf{p}_{\mathbf{i}} = 1, \mathbf{p} \geq 0, \end{aligned}$$

is convex in t .

Proof. For fixed nonnegative \mathbf{p} , the objective function $\sum_{\mathcal{I}} (\mathbf{z}_{\mathbf{i}} - t)_+ \mathbf{p}_{\mathbf{i}}$ is convex in t . In turn, $\bar{v}(t)$ is obtained by taking a supremum of convex functions $v_{\mathbf{p}}(t)$, indexed by \mathbf{p} , and therefore $\overline{\text{val}}(t)$ is convex as well as a positive weighted sum of t and $\bar{v}(t)$. \square

Using convexity, we note that the epigraph-based scheme from Subsection 3.4.2 can readily be adapted to computing the sharp upper bound of CVaR_{α} . Furthermore, using classical LP duality theory, finding the optimal t value corresponding to minimising $\overline{\text{val}}(t)$ may equivalently be reformulated as solving a linear optimization problem. Let

$$\mathcal{M}_{j,k} = \{\mathbf{i} : i_j = k, i_{\ell} = m \text{ for all } \ell \neq j\} \quad (4.1)$$

and $\mathcal{M} = \bigcup_{j,k} \mathcal{M}_{j,k}$ denote the set of multi-indices corresponding to sums of marginals, including the total probability mass. For simplicity, from now on, we assume that the c.d.f. bounds $\bar{\pi}, \underline{\pi}$ are consistent with marginals, that is, $\underline{\pi}_{\mathbf{i}} \leq \frac{k}{m} \leq \bar{\pi}_{\mathbf{i}}$ for all $\mathbf{i} \in \mathcal{M}_{j,k}$ where the marginal index sets are defined as in (4.1). If not, clearly, the problem of computing $\overline{\text{val}}(t)$ is infeasible.

For clarity of exposition, we first slightly modify our formulation of $\bar{v}(t)$ from above. Noting that the lower and upper bound requirements on c.d.f. are clearly redundant for $\mathbf{i} \in \mathcal{M}$, they may simply be replaced with more restrictive modified bounds $\underline{\pi}', \bar{\pi}'$, where

$$\underline{\pi}'_{\mathbf{i}} = \begin{cases} \frac{k}{m}, & \mathbf{i} \in \mathcal{M}_{j,k}, \forall j, k, \\ \underline{\pi}_{\mathbf{i}}, & \text{otherwise,} \end{cases} \quad \text{and} \quad \bar{\pi}'_{\mathbf{i}} = \begin{cases} \frac{k}{m}, & \mathbf{i} \in \mathcal{M}_{j,k}, \forall j, k, \\ \bar{\pi}_{\mathbf{i}}, & \text{otherwise.} \end{cases}$$

We are now ready to formulate the main result of this section.

Theorem 4.1. *The upper bound defined in (2.4) can be computed as follows:*

$$\begin{aligned} \overline{\text{CVaR}}_\alpha &:= \min_{t, \underline{\mathbf{y}}, \overline{\mathbf{y}}} t + \frac{1}{1-\alpha} \left(-\underline{\mathbf{y}}^T \underline{\boldsymbol{\pi}}' + \overline{\mathbf{y}}^T \overline{\boldsymbol{\pi}}' \right) \\ \text{s.t.} \quad & \sum_{j \geq i} \underline{\mathbf{y}}_j - \sum_{j \geq i} \overline{\mathbf{y}}_j \leq t - \mathbf{z}_i, \text{ for all } \mathbf{i} \in \mathcal{I}, \\ & \sum_{j \geq i} \underline{\mathbf{y}}_j - \sum_{j \geq i} \overline{\mathbf{y}}_j \leq 0, \text{ for all } \mathbf{i} \in \mathcal{I}, \\ & \underline{\mathbf{y}}, \overline{\mathbf{y}} \geq 0, t \in \mathbb{R}, \end{aligned}$$

Proof. Note that for any fixed t , the problem of computing $\bar{v}(t)$ is equivalent to solving its dual

$$\begin{aligned} \bar{v}^*(t) &:= \min_{\underline{\mathbf{y}}, \overline{\mathbf{y}}} -\underline{\mathbf{y}}^T \underline{\boldsymbol{\pi}}' + \overline{\mathbf{y}}^T \overline{\boldsymbol{\pi}}' \\ \text{s.t.} \quad & \sum_{j \geq i} \underline{\mathbf{y}}_j - \sum_{j \geq i} \overline{\mathbf{y}}_j \leq -(\mathbf{z}_i - t)_+, \text{ for all } \mathbf{i} \in \mathcal{I}, \\ & \underline{\mathbf{y}}, \overline{\mathbf{y}} \geq 0, \end{aligned}$$

where by strong LP duality, we know that $\bar{v}^*(t) = \bar{v}(t)$. Furthermore, for any dual-feasible point $(\underline{\mathbf{y}}, \overline{\mathbf{y}})$, by the weak duality property we have $-\underline{\mathbf{y}}^T \underline{\boldsymbol{\pi}}' + \overline{\mathbf{y}}^T \overline{\boldsymbol{\pi}}' \geq \bar{v}(t)$.

Noting that the dual feasible region may equivalently be rewritten as stated in the theorem, we finally observe that in order to compute the optimal t^* that satisfies $\overline{\text{CVaR}}_\alpha = \overline{\text{val}}(t^*)$, it suffices to solve the concurrent linear minimisation problem with respect to t and $(\underline{\mathbf{y}}, \overline{\mathbf{y}})$. \square

Finally, once the optimal value t^* is known, the corresponding optimal values of \mathbf{p} can easily be computed by solving for $\bar{v}(t^*)$ as a linear maximisation problem, if further desired.

5. NUMERICAL RESULTS

In this section, we provide numerical illustration to our findings from Sections 3 and 4. First, we gauge how the computational requirements scale up with problem dimensions and identify one critical bottleneck in Subsection 5.1. To do this, we compare two ways of implementing our approaches in MATLAB. One primarily relies on CVX with the embedded open-source solver SDPT3, chosen for the sake of simplicity. The other approach uses Gurobi (2023) and a direct problem formulation, as a potentially more efficient option. CVX removes the inconvenience of carefully formulating the LP (3.2) to near-standard form suitable for Gurobi, while potentially sacrificing some of the efficiencies. On the other hand, the user-provided direct specification of the underlying LP may be more of a challenge initially, but potentially gives some computational advantage when solving the problem. Next, we propose an approach that allows to circumvent one of the main computational obstacles, and illustrate the refined methodology on real-life inspired example in Subsection 5.2.

5.1. Verbatim implementation. Our first goal is to get a sense of how the performance of our method scales up with problem dimensions, as well as to gauge if the modeling environment and the LP solver play a role. For this, we use a very modest Alienware laptop with 2 core Intel i7 U640 CPU running at 1.2GHz, 4GB RAM, running Windows 7 x64, Matlab R2013b, CVX 2.0, and Gurobi 5.5.

Regardless of the approach, we rely on solving (3.1) or its variant, where the dimensions of the problem grow proportional to m^n – thus, polynomial in m and exponentially in n . Specifically, for the standard LP form of the partial value evaluation (3.2), the number of variables and constraints grow as $3m^n$ and $2m^n + mn + 1$ respectively, while the number of non-zeros in the matrix of coefficients describing affine constraint is roughly $2m^{2n} \cdot \left(\frac{m+1}{2m}\right)^n$.² Consequently, despite the fact that the fraction of non-zero entries in the matrix of affine coefficients corresponding to (3.2) decreasing exponentially in n , the number of non-zeros still grows very fast with the number of risks. For instance, in case of $m = 100$ and $n = 3$, one should expect to deal with a matrix containing more than 10^{11} non-zero entries (of one), making solving such problems on a regular compute workstation prohibitively expensive. Even with the availability of super-computing resources, one probably has to resort to very specialized algorithms and linear algebra techniques to exploit matrix sparsity structure efficiently for large m, n .

In Table 5.1 we report the average run-times for estimating sharp upper and lower bounds for problems with varying n and m . For this and the other numerical experiments, for each dimension, we generate 30 random problem instances where X_i sample values are chosen to be uniform between 0 and 1 for simplicity. `CVX` refers to only using CVX to formulate the LP sub-problem and passes it to a selected solver while tensor-like notation is used inside the CVX code. `CVX+` refers to us formulating the affine constraints of an LP in vectorized form and letting CVX only pass the data to the solver. `Direct` refers to us both formulating the problem and invoking Gurobi solver directly, bypassing CVX. When not specified, $\alpha = .95$ and $\varepsilon = 10^{-7}$.

Our first goal is to get a feel of how the proposed methods scale with dimensions. As expected, the computational cost escalates very rapidly when dimensions m , and especially n , increase. We observe that the run-time heavily depends on the LP solver. For Gurobi here we used the simplex option, while experimenting with the barrier gave inferior results on this dataset; we suspect that

²Intuitively, on average, for a uniform random integer number between 1 and m , exactly $\frac{m+1}{2m} \cdot 100\%$ of integers in $1, \dots, m$ are less or equal than the chosen number. Observing that the c.d.f. constraints have non-zeros at exactly such “lesser” sub-indices along each dimension. A proof can be established by induction.

		SDPT3				Gurobi					
		naïve		Epigraph		naïve			Epigraph		
n	m	CVX	CVX+	CVX	CVX+	CVX	CVX+	Direct	CVX	CVX+	Direct
2	2	2.59	2.53	6.71	6.77	1.41	1.38	0.10	6.80	6.66	0.27
	4	9.26	8.38	17.04	16.16	4.06	3.80	0.19	8.01	7.52	0.33
	6	21.85	18.49	18.79	16.85	8.76	8.03	0.36	8.28	7.40	0.38
	8	42.94	35.09	20.75	17.76	16.13	13.97	0.84	8.51	7.41	0.50
	10	77.72	61.51	22.35	18.25	26.60	22.32	1.76	9.20	7.70	0.73
	12	149.49	117.61	29.06	25.27	41.82	33.68	3.57	9.95	8.17	1.11
	14	264.68	200.38	35.92	31.02	63.71	49.53	7.29	10.59	8.40	1.66
3	3	16.05	13.67	19.68	17.10	6.51	6.11	0.29	8.04	7.59	0.37
	4	42.80	35.11	22.57	18.86	15.36	14.05	0.83	8.46	7.70	0.53
	5	109.92	87.24	27.32	22.98	32.05	28.60	2.57	9.18	8.23	0.95
	6	299.76	224.36	39.94	33.52	61.86	53.93	8.40	10.06	8.90	1.91

TABLE 5.1. Average run-time (in seconds) for naïve and epigraph schemes for small size problems with $\varepsilon = 10^{-10}$.

the latter is attributed to being able to take advantage of a simplex warm-start. Even when no top-of-the-line commercial solver is available, one can compute some bounds with $n < 3$ in reasonable time for small m .

We also note that in general using CVX as opposed to directly formulating the problem and feeding it into a solver, pose some processing time overhead, especially for smaller problems. While formulating the matrix of affine constraints, we rely on MATLAB loops which may potentially be sped up. Solving with $n = 2$ and $m = 50$ to within an $\varepsilon = 10^{-7}$ precision by using the epigraph scheme, MATLAB takes about 100 seconds to form a single LP matrix of the coefficients in the standard form, while solving all the subsequent LP problems takes another 150 seconds or so.

$n \setminus m$	3	6	9	12	15	20	30	40	50	60
2	0.24	0.30	0.48	0.91	1.71	4.76	28.08	93.44	240.52	543.52
3	0.29	1.62	16.87	114.93	600.53	-	-	-	-	-

TABLE 5.2. Average run-time (in seconds) for the best lower bound estimation scheme – epigraph-based with direct Gurobi – for small to medium size problems with $\varepsilon = 10^{-7}$, with a run-time limit of 15 minutes.

For estimating the lower bound on CVaR_α , between the two schemes, epigraph-based method is a clear winner over the naïve approach. The solution times grow with n, m , see Tables 5.1 and 5.2, as well as the desired precision ε , see Table 5.3(b). By comparing the results in Tables 5.2 and 5.4, we conclude that computing sharp upper bound is generally cheaper, as compared to the lower bound. When computing an exact sharp upper bound, direct LP embedding is preferred.

		(a)					(b)		
n	$m \setminus \alpha$	0.1	0.9	0.99	n	$m \setminus \epsilon$	10^{-3}	10^{-5}	10^{-7}
2	20	3.47	3.93	4.21	2	20	3.47	3.96	4.56
	30	20.36	23.36	25.61		30	19.24	22.29	27.53
	40	68.34	74.08	85.97		40	64.46	79.07	92.04
3	6	1.16	1.37	1.46	3	6	1.16	1.35	1.62
	9	11.83	13.90	14.47		9	12.22	14.14	16.84
	12	77.88	95.16	100.65		12	81.13	96.21	122.45

TABLE 5.3. Average run-time (in seconds) for Gurobi-based epigraph scheme with respect to (a) α , with $\epsilon = 10^{-5}$, and (b) ϵ .

Method	$n \setminus m$	3	6	9	12	20	30	40	50	60
Epigraph	2	0.32	0.35	0.60	1.14	5.88	35.60	139.00	354.43	688.63
LP embedding		0.05	0.07	0.13	0.33	2.64	12.05	39.19	98.43	218.13
Epigraph	3	1.94	21.72	172.48	802.23	-	-	-	-	-
LP embedding		0.67	7.41	44.57	190.86	-	-	-	-	-

TABLE 5.4. Average run-time (in seconds) for the upper bound using Gurobi-based epigraph and direct LP embedding methods with $\epsilon = 10^{-10}$, with a run-time limit of 15 minutes.

5.2. Stylized practice-inspired example. Computing CVaR sharp bounds under given marginals and lower-orthant stochastic ordering bounds on joint c.d.f., and in particular, sharp lower bound, entails solving a non-convex (bi-linear) optimization problem of potentially very high dimensionality. Namely, we seek to determine the extreme values of m^n variables representing the c.d.f. When attempting to scale up the model sizes n, m , we are faced with an obvious memory requirement issue. For instance, solving for $n = 3, m = 100$ in (2.3) entails formulating a model with over 10^{11} non-zeros that requires almost 1,000GB of RAM if we operate in standard double-precision arithmetic. The RAM requirement grows as m^{2n} and it is reasonable to expect a significant growth in the computational effort required to solve the model as well.

However, it turns out that one could produce a much sparser equivalent representation of lower and upper bound optimization models, allowing solving for sharp bounds with $n = 3, m = 100$ sized models in a reasonable time, i.e., couple of hours, on a reasonable hardware, i.e., multi-core station with enough RAM. Next, we present this refined setup along with a more practical illustration of our approach. The example is partly based on a work carried out outside of this manuscript, and has been further stylized to avoid breaching any possible Non Disclosure Agreements. We focus on the lower bound computation as it is more challenging; the upper bound evaluation can be refined in a similar manner.

Assume an insurance company with a portfolio of three risks located in 1) New York (NY), 2) Miami (FL) and 3) Houston (TX) for which the policy covers economic damages to certain buildings caused by hurricanes in these regions. The underwriter makes decisions based on the hurricane intensity estimates that in turn are predicted based on an atmospheric internal risk model. If X_k with $k \in \{1, 2, 3\}$ is the economic damage for the k -th risk in dollars, we know that X_k is Pareto(α_k, λ_k) so that the c.d.f. along with first 2 moments are

$$F(x) = 1 - \left(\frac{\lambda}{x + \lambda} \right)^\alpha, \quad x \geq 0, \quad \mathbb{E}X = \frac{\lambda}{\alpha - 1}, \quad \text{Var}X = \frac{\alpha\lambda^2}{(\alpha - 2)(\alpha - 1)^2},$$

with

$$\alpha_1 = 5, \alpha_2 = 2.1, \alpha_3 = 2.7, \lambda_1 = 7.92 \cdot 10^6, \lambda_2 = 1.11 \cdot 10^7, \lambda_3 = 7.36 \cdot 10^6$$

resulting in the expected losses of USD1.98mn, USD10.07mn and USD4.33mn respectively. Further, the *coefficient of variation (CV)*, a well-known measure of risk, is 1.29, 4.58 and 1.96, so indeed the assets are risky, as expected. A large CV is expected for coverage in more risky regions.

The underwriter has empirical evidence (based on atmospheric observational data) to identify the marginal risk distributions, but does not have the knowledge to create a spatial dependence model across the risks located in different regions. Geographical dependent ratings would be hardly available even to world leading rating agencies. Therefore, the underwriter has to rely on the available domain knowledge to come up with aggregate risk estimates $\text{CVaR}_\alpha(X_1 + X_2 + X_3)$ based on the best possible information about its risk position.

It is clear that X_k 's are not negatively associated, and thus, a lower bound on the joint distribution, in terms of the lower-orthant (LO) stochastic ordering, can be given by the independence model,

$$\underline{F}(x_1, x_2, x_3) = F_1(x_1)F_2(x_2)F_3(x_3),$$

where F_k is the c.d.f. of X_k , $k = 1, 2, 3$. The upper bound on the joint distribution, in terms of LO, assumes that the NY economic damages are independent of the other two, while economic damages from Miami and Houston could be strongly positive dependent, i.e., comonotonic, and therefore

$$\overline{F}(x_1, x_2, x_3) = F_1(x_1) \min(F_2(x_2), F_3(x_3)).$$

In terms of our CVaR lower bound formulation (2.3), the above can be encoded via discretizing the individual risks with some fixed m so that $x_{i,1} \leq \dots \leq x_{i,m}$, $i = 1, 2, 3$, correspond to Pareto distribution sample values or inverse Pareto-c.d.f. at mid-points $j = (j - 1/2)/m$, with

$$\underline{\pi}_{i_1, i_2, i_3} = \frac{i_1}{m} \times \frac{i_2}{m} \times \frac{i_3}{m}, \quad \overline{\pi}_{i_1, i_2, i_3} = \frac{i_1}{m} \times \max\left(\frac{i_2}{m}, \frac{i_3}{m}\right).$$

Our objective here is to evaluate the lower bound, specifically, $\underline{\text{CVaR}}_8$ for $n = 3, m = 100$.

Using a sparse reformulation of (2.3) which we discuss next, this objective indeed can be achieved in a reasonable compute time, here, in about 2hrs, or 7,554s to be precise, yielding $\underline{\text{CVaR}}_8 = 40.6\text{mn}$ corresponding to $t^* = 20,914,036$, with the bound computed to within the relative precision of 2.5×10^{-7} . For this set of computational experiments we move to a more powerful machine with AMD EPYC 7313P 16-core processor and 256G RAM, running Ubuntu 22.04. To solve the subsequent LPs, we use Gurobi 10.0.1, where the model was implemented using Gurobi's Python API, and benchmarked using Python 3.7. We want to emphasize that the chief enabling factor is the sparse reformulation that reduces the number of non-zeros in the model by a square root, e.g., going from 10^{11} to about 10^6 for $n = 3, m = 100$, allowing to formulate the model in RAM as well as permitting vastly faster computations, which is further improved by moving to a powerful compute server. The code can be found in GitHub as per Zinchenko (2023).

A number of further computational experiments was performed with varying sparsified model dimensions for both n, m and the run-time were recorded. A model with $n = 3, m = 50$ could now be solved in about 550s, while the $n = 3, m = 150$ is where the current computational limit is at for the above machine. The solve time scale super-linearly with problem dimensions. For lower dimensional models with $n = 2$, as before, run-times look more favorable; for instance $n = 2, m = 1,000$ could be solved in 808s.

The sparse reformulation of (2.3) is built on a pivotal observation that joint c.d.f. can be defined recursively, using inclusion-exclusion formula. Namely, if we introduce m^n auxiliary variables for the c.d.f. to represent

$$\pi_{\mathbf{i}} := \sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{P}_{\mathbf{j}},$$

we can express the c.d.f. bounds constraints as upper and lower bounds on π , and more critically, define the c.d.f. quantities recursively. Namely, for $n = 2$ we have

$$\pi_{\mathbf{i}} - \mathbf{P}_{\mathbf{i}} = \pi_{i_1-1} + \pi_{i_2-1} - \pi_{i_1-1, i_2-1},$$

and for $n = 3$,

$$\pi_{\mathbf{i}} - \mathbf{P}_{\mathbf{i}} = \pi_{i_1-1} + \pi_{i_2-1} + \pi_{i_3-1} - \pi_{i_1-1, i_2-1} - \pi_{i_1-1, i_3-1} - \pi_{i_2-1, i_3-1} + \pi_{i_1-1, i_2-1, i_3-1},$$

where $\mathbf{i} = (i_1, i_2, i_3)$ and π_{i_1-1} is a shorthand notation for $\pi_{(i_1-1, i_2, i_3)}$ and if some sub-index becomes negative, we replace the corresponding c.d.f. entry with 0. This necessitates only 5 and 9 zeros per constraint respectively, as opposed to an average of $m^n/2^n$ in the original model formulation done verbatim. To further promote sparsity, the marginals can be reformulated in terms of the c.d.f.

auxiliary variables, for instance for the first risk we can write

$$\pi_{(j,m,m)} = j/m, j = 1, \dots, m.$$

Thus, even though we gain another m^n variables in our formulation, the revised non-zero count grows as $O(2^n m^n)$ as compared to the original $O(m^{2n}/2^n)$. The construction can easily be extended and implemented to any n .

While for $n = 3$, moving beyond $m = 100$ becomes prohibitively expensive, an argument can be made that from a practical point of view perhaps this is also not so critical. It is hard to imagine a situation where the empirical marginals are known so precise that would necessitate spelling out marginal c.d.f. constraints in finer than $.01 (= 1/m)$ granularity. It could also be more important to distill the extreme dependence trends for the unknown multivariate c.d.f. rather than try to zero down on the very last digits of CVaR bounds, and as such, our approach could provide a viable exploratory tool.

6. A SPECIAL CASE AND ITS HIGHER-DIMENSIONAL VARIANTS

In this section we investigate the question of whether CVaR ordering may be consistent with the ordering of the underlying distributions for higher-dimensional portfolios, i.e. $n \geq 3$. We say that two n -dimensional random vectors $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ have identical marginals if $\Pr(\mathbf{X}_i^{(1)} \leq x) = \Pr(\mathbf{X}_i^{(2)} \leq x)$ for all $i = 1, \dots, n$ and $x \in \mathfrak{R}$. We first provide an alternative proof to a well-known result that CVaR respects the so-called lower-orthant stochastic ordering for $n = 2$ (Proposition 6.2.9. of Denuit *et al.*, 2005).

Theorem 6.1. *Let $n = 2$ and $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be two compactly-supported random vectors with identical marginals and corresponding aggregate risks $Z^{(1)}, Z^{(2)}$. If $\mathbf{X}^{(1)} \preceq_{lo} \mathbf{X}^{(2)}$ then for any $\alpha \in (0, 1)$ we have that $\text{CVaR}_\alpha(Z^{(1)}) \leq \text{CVaR}_\alpha(Z^{(2)})$.*

Although the claim may be extended to a wide class of other risk measures, the previously known proofs of the above theorem rely on the fairly exotic techniques from convex analysis. The theorem itself becomes interesting in view of the potential computational savings it may provide, when comparing the aggregate risks of bi-variate distributions with identical marginals, satisfying the lower-order stochastic ordering.

A natural question is whether such an ordering is preserved in higher dimensions, $n \geq 3$. We show that no such extension exists. In fact, one may argue that even the above result with $n = 2$ is unnatural and goes against the intuition of what should happen. To substantiate the latter point of view, we

- give an alternative and self-contained proof of the classical result from Theorem 6.1,
- state several potential extensions of such a result to higher dimensions, and
- provide counter-examples to show that no such extensions are true for $n \geq 3$.

6.1. An alternative proof for $n = 2$. We start by recalling the inclusion-exclusion type criterion (for example, see Billingsley, 1995), that characterizes a c.d.f.. The criterion ensures that the probability mass accumulated within any hypercube is non-negative, and is commonly referred to as the *rectangle inequality*.

For fixed n , consider a right-continuous non-decreasing $F : \mathfrak{R}^n \rightarrow \mathfrak{R}$ such that $\lim_{\mathbf{x}_i \rightarrow -\infty} F(\mathbf{x}) = 0$ for all $i = 1, \dots, n$, and $\lim_{\mathbf{x}_1, \dots, \mathbf{x}_n \rightarrow \infty} F(\mathbf{x}) = 1$, F is a c.d.f. if and only if

$$\sum_{j_1=1}^2 \cdots \sum_{j_n=1}^2 (-1)^{j_1 + \dots + j_n} F(\zeta_{1,j_1}, \zeta_{2,j_2}, \dots, \zeta_{n,j_n}) \geq 0,$$

for all $\zeta_{i,1} < \zeta_{i,2}$, $i = 1, \dots, n$.

In particular, the rectangle inequality guarantees the existence of a *probability mass function* (pmf) given a candidate non-decreasing step-like function on \mathfrak{R}^n . From now on, we consider an n -dimensional discrete random vector with values $(x_{1,i_1}, \dots, x_{n,i_n})$, $1 \leq i_1, \dots, i_n \leq m$, placed on $m \times m \times \dots \times m$ rectangular grid, and the corresponding pmf \mathbf{p}_i , $\mathbf{i} \in \mathcal{I}$. Thus, for $n = 2$ the above inequalities become

$$0 \leq \underbrace{F(\zeta_{1,2}, \zeta_{2,2})}_{\phi_{2,2}} - \underbrace{F(\zeta_{1,1}, \zeta_{2,2})}_{\phi_{1,2}} - \underbrace{F(\zeta_{1,2}, \zeta_{2,1})}_{\phi_{2,1}} + \underbrace{F(\zeta_{1,1}, \zeta_{2,1})}_{\phi_{1,1}},$$

and for $n = 3$ we have

$$\begin{aligned} 0 \leq & \underbrace{F(\zeta_{1,2}, \zeta_{2,2}, \zeta_{3,2})}_{\phi_{2,2,2}} - \underbrace{F(\zeta_{1,1}, \zeta_{2,2}, \zeta_{3,2})}_{\phi_{1,2,2}} - \underbrace{F(\zeta_{1,2}, \zeta_{2,1}, \zeta_{3,2})}_{\phi_{2,1,2}} - \underbrace{F(\zeta_{1,2}, \zeta_{2,2}, \zeta_{3,1})}_{\phi_{2,2,1}} \\ & + \underbrace{F(\zeta_{1,1}, \zeta_{2,1}, \zeta_{3,2})}_{\phi_{1,1,2}} + \underbrace{F(\zeta_{1,1}, \zeta_{2,2}, \zeta_{3,1})}_{\phi_{1,2,1}} + \underbrace{F(\zeta_{1,2}, \zeta_{2,1}, \zeta_{3,1})}_{\phi_{2,1,1}} - \underbrace{F(\zeta_{1,1}, \zeta_{2,1}, \zeta_{3,1})}_{\phi_{1,1,1}}, \end{aligned}$$

where ζ_{i,j_i} values correspond to the atoms on the grid, that is, $\zeta_{i,1} = x_{i,k_i}$ and $\zeta_{i,2} = x_{i,k'_i}$ with $k_i < k'_i$, $i = 1, 2, 3$. The latter expressions can be abridged to

$$0 \leq \phi_{2,2} - \phi_{1,2} - \phi_{2,1} + \phi_{1,1}, \quad (6.1)$$

and

$$0 \leq \phi_{2,2,2} - \phi_{1,2,2} - \phi_{2,1,2} - \phi_{2,2,1} + \phi_{1,1,2} + \phi_{1,2,1} + \phi_{2,1,1} - \phi_{1,1,1} \quad (6.2)$$

introducing $\phi_{j_1, j_2, \dots} = F(\zeta_{1, j_1}, \zeta_{2, j_2}, \dots)$. The summation sign pattern for values of F , or equivalently $\phi_{i_1, i_2, \dots}$, may be best illustrated graphically, as seen in Figure 6.1.

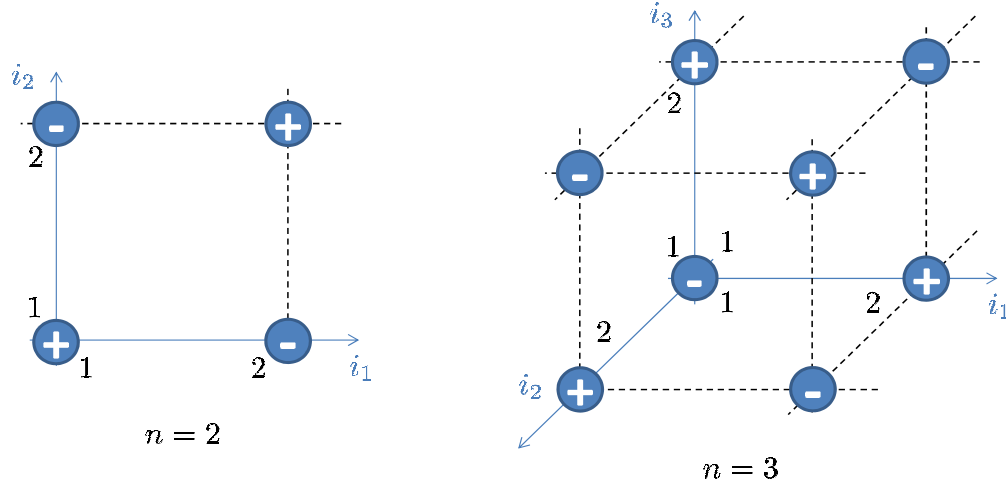


FIGURE 6.1. Rectangle inequality summation sign pattern.

The following elementary, yet critical, observation can be made and is given as Proposition 6.1.

Proposition 6.1. *Along with pmf $\mathbf{p} \in \mathfrak{R}^{m^n}$, consider the c.d.f. $\pi \in \mathfrak{R}^{m^n}$ and the survival function $\mathbf{s} \in \mathfrak{R}^{m^n}$, defined by the corresponding linear transformations*

$$\Pi : \mathbf{p} \mapsto \pi, \text{ defined as } \pi_{\mathbf{i}} = \sum_{\mathbf{j} \leq \mathbf{i}} \mathbf{p}_{\mathbf{j}} \quad \text{and} \quad \mathbf{S} : \mathbf{p} \mapsto \mathbf{s}, \text{ defined as } \mathbf{s}_{\mathbf{i}} = \sum_{\mathbf{j} \geq \mathbf{i}} \mathbf{p}_{\mathbf{j}}.$$

Then, for any $\mathbf{z} \in \mathfrak{R}^{m^n}$ we have

$$\mathbf{p}^T \mathbf{z} = \sum_{\mathcal{I}} \mathbf{p}_{\mathbf{i}} \mathbf{z}_{\mathbf{i}} = (\Pi^{-1} \mathbf{z})^T (\mathbf{S} \mathbf{p}) = (\mathbf{S}^{-1} \mathbf{z})^T (\Pi \mathbf{p}).$$

The proof of the above proposition, although somewhat tedious, simply relies on accounting for the indices in the summation $\sum_{\mathcal{I}} \mathbf{p}_{\mathbf{i}} \mathbf{z}_{\mathbf{i}}$. We also note that for $n = 2$, the c.d.f. ordering of two distributions with identical marginals is equivalent to the ordering of the survival functions.

Lemma 6.1. *Let the two bi-variate discrete random variables $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ have identical marginals and the corresponding c.d.f. $\pi^{(1)}, \pi^{(2)}$. Then, $\pi_{\mathbf{i}}^{(1)} \leq \pi_{\mathbf{i}}^{(2)}$, $\forall \mathbf{i}$, if and only if $\mathbf{s}_{\mathbf{i}}^{(1)} \leq \mathbf{s}_{\mathbf{i}}^{(2)}$, $\forall \mathbf{i}$.*

The proof is a straightforward implication of the inclusion-exclusion type fact that

$$\Pr(X_1 > x_1, X_2 > x_2) = 1 - \Pr(X_1 \leq x_1) - \Pr(X_2 \leq x_2) + \Pr(X_1 \leq x_1, X_2 \leq x_2).$$

Finally, we are now able to prove Theorem 6.1.

Proof of Theorem 6.1. The sub-problem (3.2) can be re-parameterized using the survival function \mathbf{s}_i ,

$$\begin{aligned} \underline{\mathbf{y}}(t) &= \inf_{\mathbf{s}} \left(\Pi^{-1}(\mathbf{z} - t)_+ \right)^T \mathbf{s} && \left(\equiv \inf_{\mathbf{p}} (\mathbf{z} - t)_+^T \mathbf{S}^{-1} \mathbf{S} \mathbf{p} \right) \\ \text{s.t.} & \quad \underline{\mathbf{S}}_i \leq \mathbf{s}_i \leq \bar{\mathbf{S}}_i, \text{ for all } \mathbf{i} \in \mathcal{I}, \\ & \quad \mathbf{s}_i = \frac{m - k + 1}{m}, \text{ for all } \mathbf{i} = (1, k) \text{ or } (k, 1), k = 1, \dots, m, \\ & \quad \mathbf{s}_{(1,1)} = 1, \\ & \quad \mathbf{S}^{-1} \mathbf{s} \geq 0, \end{aligned} \tag{6.3}$$

where the survival function bounds $\underline{\mathbf{S}}_i, \bar{\mathbf{S}}_i$ may easily be computed applying the inclusion-exclusion type formula similar to that in Lemma 6.1.

We now make a critical observation that $\Pi^{-1}(\mathbf{z} - t)_+ \geq 0$ for all t . From the definition of aggregate risk values \mathbf{z} , we observe that $\Pi^{-1}(\mathbf{z} - t)_+ \geq 0$ if and only if the rectangle inequality (6.1) holds for all

$$\begin{aligned} \phi_{1,1} &= (x_{1,i_1} + x_{2,i_2} - t)_+ \equiv (\mathbf{z}_{(i_1,i_2)} - t)_+, & \phi_{1,2} &= (x_{1,i_1} + x_{2,i'_2} - t)_+ \equiv (\mathbf{z}_{(i_1,i'_2)} - t)_+, \\ \phi_{2,1} &= (x_{1,i'_1} + x_{2,i_2} - t)_+ \equiv (\mathbf{z}_{(i'_1,i_2)} - t)_+, & \phi_{2,2} &= (x_{1,i'_1} + x_{2,i'_2} - t)_+ \equiv (\mathbf{z}_{(i'_1,i'_2)} - t)_+, \end{aligned}$$

with $1 \leq i_1 < i'_1 \leq m$, $1 \leq i_2 < i'_2 \leq m$ and any t . Since $x_{1,i_1} \leq x_{1,i'_1}$ and $x_{2,i_2} \leq x_{2,i'_2}$, we also have partial ordering of ζ values, namely

$$\phi_{1,1} \leq \phi_{1,2} \leq \phi_{2,2} \text{ and } \phi_{1,1} \leq \phi_{2,1} \leq \phi_{2,2}.$$

The range of all t values may clearly be partitioned into $\mathcal{T}_- = (-\infty, \mathbf{z}_{(i_1,i_2)}]$, $\mathcal{T} = (\mathbf{z}_{(i_1,i_2)}, \mathbf{z}_{(i'_1,i'_2)})$ and $\mathcal{T}_+ = [\mathbf{z}_{(i'_1,i'_2)}, \infty)$. Therefore, there are three possible cases.

(a) $t \in \mathcal{T}_-$: rectangle inequality (6.1) clearly holds as

$$0 = (x_{1,i_1} + x_{2,i_2} - t) - (x_{1,i_1} + x_{2,i'_2} - t) - (x_{1,i'_1} + x_{2,i_2} - t) + (x_{1,i'_1} + x_{2,i'_2} - t).$$

(b) $t \in \mathcal{T}$: the validity of the rectangle inequality may easily be established by assuming, without loss of generality, that $\mathbf{z}_{(i_1,i'_2)} \leq \mathbf{z}_{(i'_1,i_2)}$ and considering further sub-cases depending on where the value of t falls with respect to \mathbf{z} sub-intervals. For example, if $t \in (\mathbf{z}_{(i_1,i_2)}, \mathbf{z}_{(i_1,i'_2)}]$, then $\phi_{1,1} > \mathbf{z}_{(i_1,i_2)} - t$, and thus, the rectangle inequality results in positive mass. That is,

$$0 < (x_{1,i_1} + x_{2,i_2} - t)_+ - (x_{1,i_1} + x_{2,i'_2} - t) - (x_{1,i'_1} + x_{2,i_2} - t) + (x_{1,i'_1} + x_{2,i'_2} - t).$$

(c) $t \in \mathcal{T}_+$: clearly the inequality holds as all ϕ values are 0.

Therefore, $\Pi^{-1}(\mathbf{z} - t)_+ \geq 0$ indeed holds.

To complete the proof, consider problem (6.3) where $\underline{\mathbf{S}}, \overline{\mathbf{S}}$ correspond to $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ respectively. Due to non-negativity of the objective coefficients $\Pi^{-1}(\mathbf{z} - t)_+$ in (6.3), clearly $\underline{\text{CVaR}}_\alpha$ corresponds to $\mathbf{X}^{(1)}$. Similarly, considering a variant of (6.3) to evaluate the upper $\overline{\text{CVaR}}_\alpha$ bound, we conclude that $\overline{\text{CVaR}}_\alpha$ corresponds to $\mathbf{X}^{(2)}$. The fact that $\underline{\text{CVaR}}_\alpha \leq \overline{\text{CVaR}}_\alpha$ completes the proof. \square

6.2. A few possible generalizations and some counter-examples. We first provide the definitions of some stochastic ordering and for two multi-variate risks $\mathbf{X}^{(1)}, \mathbf{X}^{(2)} \in \mathfrak{R}^n$ we define

- (1) *upper-orthant ordering* $\mathbf{X}^{(1)} \preceq_{uo} \mathbf{X}^{(2)}$ if $\Pr(\mathbf{X}^{(1)} > \mathbf{x}) \leq \Pr(\mathbf{X}^{(2)} > \mathbf{x})$ for all $\mathbf{x} \in \mathfrak{R}^n$;
- (2) *lower-orthant ordering* $\mathbf{X}^{(1)} \preceq_{lo} \mathbf{X}^{(2)}$ if $\Pr(\mathbf{X}^{(1)} \leq \mathbf{x}) \leq \Pr(\mathbf{X}^{(2)} \leq \mathbf{x})$ for all $\mathbf{x} \in \mathfrak{R}^n$;
- (3) *concordance ordering* $\mathbf{X}^{(1)} \preceq_{co} \mathbf{X}^{(2)}$ if $\mathbf{X}^{(1)} \preceq_{uo} \mathbf{X}^{(2)}$ and $\mathbf{X}^{(1)} \preceq_{lo} \mathbf{X}^{(2)}$;
- (4) *persistent ordering* $\mathbf{X}^{(1)} \preceq_{po} \mathbf{X}^{(2)}$ if $\mathbf{X}^{(1)} \preceq_{uo} \mathbf{X}^{(2)}$ and $\mathbf{X}^{(1)} \succeq_{lo} \mathbf{X}^{(2)}$.

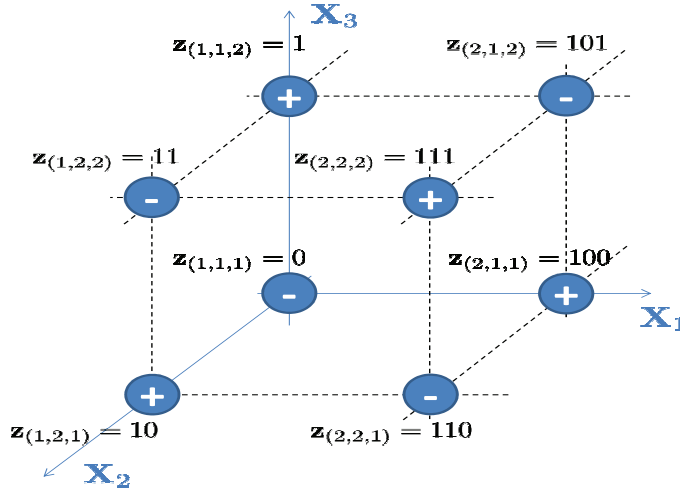
Note that due to Lemma 6.1, persistent ordering for $n = 2$ results in identical distributions, and is therefore not interesting to be investigated for the bi-variate case. Recall from the proof of Theorem 6.1 that for $n = 2$, the persistence of CVaR ordering relies on the implied upper-orthant stochastic ordering of the respective risks. Consequently, in search of an extension of such a result to $n = 3$, the following question appears to be a natural place to start: Is it true that for tri-variate distributions $\mathbf{X}^{(1)} \preceq_{uo} \mathbf{X}^{(2)}$, we have $\text{CVaR}_\alpha(Z^{(1)}) \leq \text{CVaR}_\alpha(Z^{(2)})$ for all $\alpha \in (0, 1)$, with $Z^{(1)}, Z^{(2)}$ being the corresponding aggregate risks?

From now on, we fix $n = 3$ and the marginals of $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ to be uniform. The key to constructing a counter-example to the above is the failure of the rectangle inequality (6.2) over $(\mathbf{z} - t)_+$ values. In turn, this results in a re-parameterized three-dimensional analogue of (6.3) that has both positive and negative objective coefficients $\Pi^{-1}(\mathbf{z} - t)_+$ for some suitably chosen t . Specifically, consider the only relevant CVaR estimation values of t that correspond to $\mathbf{z}_i, \mathbf{i} \in \mathcal{I}$ in accordance to Theorem 3.2. We claim that for carefully chosen \mathbf{z} , we can pick two values t_+ and t_\pm such that $\Pi^{-1}(\mathbf{z} - t_+)_+ \geq 0$, while $\Pi^{-1}(\mathbf{z} - t_\pm)_+$ contains both positive and negative entries. As a consequence to $\Pi^{-1}(\mathbf{z} - t_+)_+$ being sign-indeterminant, when estimating $\underline{\text{CVaR}}_\alpha, \overline{\text{CVaR}}_\alpha$ with bounds $\underline{\mathbf{S}}, \overline{\mathbf{S}}$ corresponding to the respective distributions $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$, it is natural to expect that we may end up having $\underline{\text{CVaR}}_\alpha < \overline{\text{CVaR}}_\alpha$ for some values of α as well as $\underline{\text{CVaR}}_\alpha > \overline{\text{CVaR}}_\alpha$ for other values of α . From here, it may simply suffice to pick ‘‘correct’’ scaling constants α in extremal characterization of CVaR.

To make this precise, consider the set of risk values for $m = 2$ and $m = 3$ in Table 6.1, with aggregate risk values depicted on the hypercube lattice in Figure 6.2. Take $t_+ = 0, t_\pm = 1$. Clearly, rectangle inequality (6.2) holds at t_+ and results in negative mass at t_\pm due to the fact that $(\phi_{1,1,1} - t)_+ > \phi_{1,1,1} - t$.

Risk	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$
$i = 1$	0	100	200
$i = 2$	0	10	20
$i = 3$	0	1	2

TABLE 6.1. Tri-variate risk sample values.

FIGURE 6.2. Tri-variate aggregate risk values with $m = 2$.

Now, we can use \mathbf{z} values to produce a desired counterexample for upper-orthant ordering. To do so, we can form an LP problem to maximize the difference between two partial value-type estimates $\sum_{\mathcal{I}}(\mathbf{z} - t)_+ \mathbf{p}^{(1)}$ and $\sum_{\mathcal{I}}(\mathbf{z} - t)_+ \mathbf{p}^{(2)}$. We subject both pmf $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ to have identical marginals, and the resulting survival functions $\mathbf{s}^{(1)} = \mathbf{S}\mathbf{p}^{(1)}$ and $\mathbf{s}^{(2)} = \mathbf{S}\mathbf{p}^{(2)}$ to satisfy the upper-orthant ordering, i.e. $\mathbf{s}^{(1)} \leq \mathbf{s}^{(2)}$. The last problem can be solved for all $t = \mathbf{z}_i$, $\mathbf{i} \in \mathcal{I}$, to extract an example.

Similar exercise can be carried out for the other stochastic orderings. Thus, for the sake of brevity, we give only a summary of our findings in Tables 6.2 and 6.3. In order to verify the results, it suffices to perform a direct calculation. For instance, with respect to the upper-orthant ordering, observe that $Z^{(1)}$ takes values 0,11,101,110 and $Z^{(2)}$ takes values 1,10,100,111 all with equal probabilities of $1/4$. Consequently, $\text{CVaR}_{.1}(Z^{(1)}) = 61\frac{2}{3}$ and $\text{CVaR}_{.1}(Z^{(2)}) = 61\frac{5}{9}$. Further, one can verify that $\text{CVaR}_{\alpha}(Z^{(1)}) > (<) \text{CVaR}_{\alpha}(Z^{(2)})$ holds for any $\alpha < (>).5$.

Interestingly, counter-examples to lower/upper-orthant and persistent orderings require a distribution supported at vertices of a single hypercube, that is, $m = 2$. On the other hand, the concordant ordering appears to require more degrees of freedom, e.g. $m = 3$. Note that in the latter

Ordering	α	i	$\mathbf{P}_{(1,1,1)}^{(i)}$	$\mathbf{P}_{(2,1,1)}^{(i)}$	$\mathbf{P}_{(1,2,1)}^{(i)}$	$\mathbf{P}_{(2,2,1)}^{(i)}$	$\mathbf{P}_{(1,1,2)}^{(i)}$	$\mathbf{P}_{(2,1,2)}^{(i)}$	$\mathbf{P}_{(1,2,2)}^{(i)}$	$\mathbf{P}_{(2,2,2)}^{(i)}$
\preceq_{uo}	.1	1	1/4	-	-	1/4	-	1/4	1/4	-
		2	-	1/4	1/4	-	1/4	-	-	1/4
\preceq_{lo}	.9	1	-	1/4	1/4	-	1/4	-	-	1/4
		2	1/4	-	-	1/4	-	1/4	1/4	-
\preceq_{po}	.1	1	1/4	-	-	1/4	-	1/4	1/4	-
		2	-	1/4	1/4	-	1/4	-	-	1/4

TABLE 6.2. Cases in which $\text{CVaR}_\alpha(Z^{(1)}) > \text{CVaR}_\alpha(Z^{(2)})$.

Ordering	α	i	$\mathbf{P}_{(1,1,1)}^{(i)}$	$\mathbf{P}_{(2,1,1)}^{(i)}$	$\mathbf{P}_{(1,2,1)}^{(i)}$	$\mathbf{P}_{(2,2,1)}^{(i)}$	$\mathbf{P}_{(1,1,2)}^{(i)}$	$\mathbf{P}_{(2,1,2)}^{(i)}$	$\mathbf{P}_{(1,2,2)}^{(i)}$	$\mathbf{P}_{(2,2,2)}^{(i)}$
\preceq_{uo}	.9	1	1/4	-	-	1/4	-	1/4	1/4	-
		2	-	1/4	1/4	-	1/4	-	-	1/4
\preceq_{lo}	.1	1	-	1/4	1/4	-	1/4	-	-	1/4
		2	1/4	-	-	1/4	-	1/4	1/4	-
\preceq_{po}	.9	1	1/4	-	-	1/4	-	1/4	1/4	-
		2	-	1/4	1/4	-	1/4	-	-	1/4

TABLE 6.3. Cases in which $\text{CVaR}_\alpha(Z^{(1)}) < \text{CVaR}_\alpha(Z^{(2)})$.

case, due to the risks being potentially supported on $m^n = 27$ vertices, we present the example in a “sparse” format, as seen Table 6.4.

\mathbf{i}	(2,1,1)	(3,1,1)	(2,3,1)	(3,3,1)	(1,2,2)	(2,1,3)	(3,1,3)	(2,3,3)	(3,3,3)
\mathbf{z}_i	100	200	120	220	11	102	202	122	222
$\mathbf{P}_i^{(1)}$	1/6	-	-	1/6	1/3	-	1/6	1/6	-
$\mathbf{P}_i^{(2)}$	-	1/6	1/6	-	1/3	1/6	-	-	1/6

TABLE 6.4. Tri-variate concordant risks with $\text{CVaR}_{.5}(Z^{(1)}) > \text{CVaR}_{.5}(Z^{(2)})$ and $\text{CVaR}_{.9}(Z^{(1)}) < \text{CVaR}_{.9}(Z^{(2)})$.

7. CONCLUSIONS

The problem of finding the entire spectrum of values for CVaR of a sum of dependent random variables under dependence uncertainty could be approached in various ways. Under restrictive assumptions, analytical approaches are implementable, but the bounds are often loose, and occasionally, not sharp. Even if the sharpness issue is not present, the lower and upper bounds are typically attained under dependence models that are difficult to justify as feasible in practice, especially for portfolios consisting of many risks, since such extreme dependence models are not realistic.

Our contribution is two-fold. Firstly, we provide a first in its class numerical method for constrained CVaR estimation, when the marginal distributions are known while only the bounds are available for the joint distribution. The latter setting is backed up by many observational data, where the dependence structure is rarely computable even if multivariate observational data are available. As a result, the lower and upper sharp bounds of the CVaR-based aggregate risk can be found. We analyze the complexity of the proposed methods for calculating these bounds as well as substantiate our findings via numerical illustrations. Our approach trivially generalizes to non-uniform marginals. Despite the fact that the computational cost increases very rapidly with the number of risks, we believe that the method may still be used as a viable exploratory tool when dealing with a relatively large risk portfolio. Finally, we show how the run-times may be significantly improved, by exploiting the very special structure of the underlying linear optimization problems at the formulation stage.

Secondly, it has been known that CVaR respects the so-called lower-orthant stochastic ordering for two-dimensional portfolios. Yet, no similar result has yet been established or disproved for higher dimensions. As a bi-product of our analysis, using elementary LP techniques, we show that no such extensions are possible. Specifically, we construct tri-variate counter-examples that demonstrate lack of aggregate risk monotonicity under upper, lower-orthant, concordant and persistent stochastic orderings. We also give a self-contained alternative proof for the bi-variate risk case, and point out the exact reason why higher-dimensional extensions are not possible.

ACKNOWLEDGMENTS

We would like to thank the authors of the free online course CVX 101 (Stephen Boyd and Lieven Vandenberghe) for the invaluable resource that is made open to public through Stanford University, that helped to sharpen our focus on convexity. We would also like to thank Michael Grant for making CVX modeling environment freely available to general public. We would also like to thank the authors of SDPT3 (Kim-Chuan Toh , Michael Todd and Reha Tutuncu) for making their solver publicly available, as well as Gurobi for providing a free academic license for their top-of-the-line LP-MIP solver. The second author would also like to express gratitude to NSERC and PIMS for supporting this piece of research.

REFERENCES

- Acerbi, C. and Tasche, D. 2002. “On the Coherence of Expected Shortfall”, *Journal of Banking and Finance*, 26(7), 1487–1503.

- Asimit, A. V., Bignozzi, V., Cheung, K.C., Hu, J. and Kim, E.-S. 2017. “Robust and Pareto Optimality of Insurance Contract”, *European Journal of Operational Research*, 262(2), 720–732.
- Asimit, A. V., Hu, J. and Xie, Y. 2019. “Optimal Robust Insurance with a Finite Uncertainty Set”, *Insurance: Mathematics and Economics*, 87, 67–81.
- Asimit, A. V., Badescu, A. M., Haberman, S. and Kim, E. S. 2016. “Efficient risk allocation within a non-life insurance group under Solvency II Regime”, *Insurance: Mathematics and Economics*, 66, 69–76.
- Balbás, A., Balbás, B. and Heras, A. 2011. “Stable Solutions for Optimal Reinsurance Problems involving Risk Measures”, *European Journal of Operational Research*, 214(3), 796–804.
- BCBS 2016. “Standards. Minimum capital requirements for Market Risk. January 2016”, Basel Committee on Banking Supervision. Basel: Bank for International Settlements.
- Ben-Tal, A., El Ghaoui, L. and Nemirovski, A. 2009. *Robust Optimization*, Princeton, New Jersey.
- 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.
- Bertsimas, D., Lauprete, G.J. and Samarov, A. 2004. “Shortfall As a Risk Measure: Properties, Optimization and Applications”, *Journal of Economic Dynamics and Control*, 28(7), 1353–1381.
- Bertsimas, D. and Popescu, I. 2002. “On the Relation between Option and Stock Prices: A Convex Optimization Approach”, *Operations Research*, 50(2), 358–374.
- Blanchet, J., Lam, H., Tang, Q. and Yuan, Z. 2017. “Applied Robust Performance Analysis for Actuarial Applications”, *Technical Report, Society of Actuaries*.
- Chen, W., Sim, M., Sun, J. and Teo, C. 2009. “From CVaR to Uncertainty Set: Implications in Joint Chance-Constrained Optimization”, *Operations Research*, 58(2), 470–485.
- Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. 2005. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, Wiley, Chichester.
- Gómez, F., Tang, Q. and Tong, Z. (2022). “The Gradient Allocation Principle based on the Higher Moment Risk Measure”, *Journal of Banking & Finance*, 143, 106544.
- Gijbels, I. and Sznajder, D. 2013. “Positive Quadrant Dependence Testing and Constrained Copula Estimation”, *Canadian Journal of Statistics*, 41(1), 36–64.
- Gurobi Optimization, LLC, 2023. “Gurobi Optimizer Reference Manual”, available at <https://www.gurobi.com>
- Huang, D., Zhu, S., Fabozzi, F.J. and Fukushima, M. 2010. “Portfolio Selection under Distributional Uncertainty: A Relative Robust CVaR Approach”, *European Journal of Operational Research*, 203(1), 185–194.

- Fabozzi, F.J., Huang, D. and Zhou, G. 2010. “Robust Portfolios: Contributions from Operations Research and Finance”, *Annals of Operations Research*, 176(1), 191–220.
- Pardo, L. 2006. *Statistical Inference Based on Divergence Measures*, Chapman & Hall, LLC, Boca Raton.
- Rockafellar, R.T. and Uryasev, S. 2000. “Optimization of Conditional Value-at-Risk”, *Journal of Risk*, 2, 21–41.
- Swiss Solvency Test 2006. FINMA SST Technisches Dokument, available at http://www.finma.ch/archiv/bpv/download/d/SST_technischesDokument_061002.pdf
- Tardos, E. 1986. “A Strongly Polynomial Algorithm to Solve Combinatorial Linear Programs”, *Operations Research*, 34(2), 250–256.
- Vavasis, S. and Ye, Y. 1996. “A Primal Dual Accelerated Interior Point Method Whose Running Time Depends Only on A ”, *Mathematical Programming*, 74(1), 79–120.
- Zhu, S. and Fukushima, M. 2009. “Worst-Case Conditional Value-at-Risk with Application to Robust Portfolio Management”, *Operations Research*, 57(5), 1155–1168.
- Zinchenko, Y. 2023. “CVaR engine for a sharp lower bound”, GitHub repository, available at <https://github.com/yzinchenko/CVaR>