

Optimization with multivariate conditional value-at-risk constraints

Nilay Noyan and Gábor Rudolf

Manufacturing Systems/Industrial Engineering Program, Sabancı University, 34956 Istanbul, Turkey,
nnoyan@sabanciuniv.edu and grudolf@sabanciuniv.edu

ABSTRACT: For many decision making problems under uncertainty, it is crucial to develop risk-averse models and specify the decision makers' risk preferences based on multiple stochastic performance measures (or criteria). Incorporating such multivariate preference rules into optimization models is a fairly recent research area. Existing studies focus on extending univariate stochastic dominance rules to the multivariate case. However, enforcing multivariate stochastic dominance constraints can often be overly conservative in practice. As an alternative, we focus on the widely-applied risk measure conditional value-at-risk (CVaR), introduce a multivariate CVaR relation, and develop a novel optimization model with multivariate CVaR constraints based on polyhedral scalarization. To solve such problems for finite probability spaces we develop a cut generation algorithm, where each cut is obtained by solving a mixed integer problem. We show that a multivariate CVaR constraint reduces to finitely many univariate CVaR constraints, which proves the finite convergence of our algorithm. We also show that our results can be naturally extended to a wider class of coherent risk measures. The proposed approach provides a flexible, and computationally tractable way of modeling preferences in stochastic multi-criteria decision making. We conduct a computational study for a budget allocation problem to illustrate the effect of enforcing multivariate CVaR constraints and demonstrate the computational performance of the proposed solution methods.

Keywords: multivariate risk-aversion; conditional value-at-risk; multiple criteria; cut generation; coherent risk measures; stochastic dominance; Kusuoka representation

1. Introduction The ability to compare random outcomes based on the decision makers' risk preferences is crucial to modeling decision making problems under uncertainty. In this paper we focus on optimization problems that feature risk preference relations as constraints. Risk measures are functionals that represent the risk associated with a random variable by a scalar value, and provide a direct way to define such preferences. Popular risk measures include semi-deviations, quantiles (under the name *value-at-risk*), and *conditional value-at-risk* (CVaR). Desirable properties of risk measures, such as *law invariance* and *coherence*, have been axiomatized starting with the work of Artzner et al. (1999). CVaR, introduced by Rockafellar and Uryasev (2000), is a risk measure of particular importance which not only satisfies these axioms, but also serves as a fundamental building block for other law invariant coherent risk measures (as demonstrated by Kusuoka, 2001). Due to these attractive properties, univariate risk constraints based on CVaR have been widely incorporated into optimization models, primarily in a financial context (see, e.g., Uryasev, 2000; Rockafellar and Uryasev, 2002; Fábíán and Veszprémi, 2008).

Relations derived from risk measures use a single scalar-valued functional to compare random outcomes. In contrast, stochastic dominance relations provide a well-established (Mann and Whitney, 1947; Lehmann, 1955) basis for more sophisticated comparisons; for a review on these and other comparison methods we refer the reader to Shaked and Shanthikumar (1994), Müller and Stoyan (2002), and the references therein. In particular, the second-order stochastic dominance (SSD) relation has been receiving significant attention due to its correspondence with risk-averse preferences. Dentcheva and Ruszczyński (2003) have proposed to incorporate such relations into optimization problems as constraints, requiring the decision-based random outcome to stochastically dominate some benchmark random outcome. Recently, such optimization models with univariate stochastic dominance constraints have been studied, among others, by Luedtke (2008); Noyan et al. (2008); Noyan and Ruszczyński (2008); Rudolf and Ruszczyński (2008); Gollmer et al. (2011), and they have been applied to various areas including financial portfolio optimization (see, e.g., Dentcheva and Ruszczyński, 2006), emergency service system design (Noyan, 2010), power planning (see, e.g., Gollmer et al., 2008), and optimal path problems (Nie et al., 2011).

For many decision making problems, it may be essential to consider multiple random outcomes of interest. In

contrast to the scalar-based comparisons mentioned above, such a multi-criteria (or multi-objective) approach requires specifying preference relations among random vectors, where each dimension of a vector corresponds to a decision criterion. This is usually accomplished by extending scalar-based preferences to vector-valued random variables. Incorporating multivariate preference rules as constraints into optimization models is a fairly recent research area, focusing on problems of the general form

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succcurlyeq \mathbf{Y} \\ & \mathbf{z} \in Z. \end{aligned}$$

Here $G(\mathbf{z})$ is the random *outcome vector* associated with the decision variable \mathbf{z} according to some *outcome mapping* G , the relation \succcurlyeq represents multivariate preferences, and \mathbf{Y} is a *benchmark* (or *reference*) random outcome vector. A key idea in this line of research, initiated by the work of Dentcheva and Ruszczyński (2009), is to consider a family of scalarization functions and require that the scalarized versions of the random variables conform to some scalar-based preference relation. In case of linear scalarization, one can interpret scalarization coefficients as the weights representing the subjective importance of each criterion. This ‘weighted sum’ approach is widely used in multi-criteria decision making (see, e.g., Steuer, 1986; Ehrgott, 2005), and there is a rich literature on methods to elicit the subjective importance that decision makers place on each criterion (see, e.g., the *analytical hierarchy process* in Saaty, 1980). However, in many decision making situations, especially those involving multiple decision makers, it can be difficult to exactly specify a single scalarization. In such cases one can enforce the preference relation over a given set of weights representing a wider range of views.

Dentcheva and Ruszczyński (2009) consider linear scalarization with positive coefficients and apply a univariate SSD dominance constraint to all nonnegative weighted combinations of random outcomes, leading to the concept of *positive linear SSD*. They provide a solid theoretical background and develop duality results for this problem, while Homem-de-Mello and Mehrotra (2009) propose a cutting surface method to solve a related class of problems. The latter study considers only finitely supported random variables under certain linearity assumptions, but the set of scalarization coefficients is allowed to be an arbitrary polyhedron. However, their method is computationally demanding as it typically requires solving a large number of non-convex cut generation problems. Hu et al. (2010) introduce an even more general concept of dominance by allowing arbitrary convex scalarization sets, and apply a sample average approximation-based solution method. Not all notions of multivariate stochastic dominance rely on scalarization functions. Armbruster and Luedtke (2010) consider optimization problems constrained by first and second order stochastic dominance relations based on multi-dimensional utility functions (see, e.g., Müller and Stoyan, 2002).

As we have seen, the majority of existing studies on optimization models with multivariate risk-averse preference relations focus on extending univariate stochastic dominance rules to the multivariate case. However, this approach typically results in very demanding constraints that can be excessively hard to satisfy in practice, and sometimes even lead to infeasible problems. For example, Hu et al. (2011b) solve a multivariate SSD-constrained homeland security budget allocation problem, and ensure feasibility by introducing a tolerance parameter into the SSD constraints. Other attempts to weaken stochastic dominance relations in order to extend the feasible region have resulted in concepts such as *almost stochastic dominance* (Leshno and Levy, 2002; Lizyayev and Ruszczyński, 2011) and *stochastically weighted stochastic dominance* (Hu et al., 2011a).

In this paper we propose an alternative approach, where stochastic dominance relations are replaced by a collection of *conditional value-at-risk* (CVaR) constraints at various confidence levels. This is a very natural relaxation, due to the well known fact that the univariate SSD relation is equivalent to a continuum of CVaR inequalities (Dentcheva and Ruszczyński, 2006). Furthermore, compared to methods directly based on dominance concepts, the ability to specify confidence levels allows a significantly higher flexibility to express decision makers’ risk preferences. At the extreme ends of the spectrum CVaR-based constraints

can express both risk-neutral and worst case-based decision rules, while SSD relations can be approximated (and even exactly modeled) by simultaneously enforcing CVaR inequalities at multiple confidence levels. Comparison between random vectors is achieved by means of a polyhedral scalarization set, along the lines of [Homem-de-Mello and Mehrotra \(2009\)](#), leading to multivariate *polyhedral CVaR constraints*. We remark that this concept is not directly related to the risk measure introduced under the name “multivariate CVaR” by [Prékopa \(2012\)](#), defined as the conditional expectation of a scalarized random vector. To the best of our knowledge, incorporating the risk measure CVaR is a first for optimization problems with multivariate preference relations based on a set of scalarization weights.

The contributions of this study are as follows.

- We introduce a new multivariate risk-averse preference relation based on CVaR and linear scalarization.
- We develop a modeling approach for multi-criteria decision making under uncertainty featuring multivariate CVaR-based preferences.
- We develop a finitely convergent cut generation algorithm to solve polyhedral CVaR-constrained optimization problems. Under linearity assumptions we provide explicit formulations of the master problem as a linear program, and of the cut generation problem as a mixed integer linear program.
- We provide a theoretical background to our formulations, including duality results. We also show that on a finite probability space a polyhedral CVaR constraint can be reduced to a finite number of univariate CVaR inequalities.
- We extend our CVaR-based methodology to optimization problems featuring polyhedral constraints based on a wider class of coherent risk measures.
- We adapt and extend some existing results from the theory of risk measures to fit the framework of our problems, as necessary. In particular, we prove the equivalence of relaxed SSD relations to a continuum of relaxed CVaR constraints, and show that for finite probability spaces this continuum can be reduced to a finite set. We also present a form of Kusuoka’s representation theorem for coherent risk measures which does not require the underlying probability space to be atomless.
- In a small-scale numerical study we examine the feasible regions associated with various polyhedral CVaR constraints, and compare them to their SSD-based counterparts. We also conduct a comprehensive computational study of a budget allocation problem, previously explored in [Hu et al. \(2011b\)](#), to evaluate the effectiveness of our proposed model and solution methods.

The rest of the paper is organized as follows. In [Section 2](#) we review fundamental concepts related to CVaR, SSD, and linear scalarization. Then we define multivariate CVaR relations, and present a general form of optimization problems involving such relations as constraints. [Section 3](#) contains theoretical results including optimization representations of CVaR, and finite representations of polyhedral CVaR and SSD constraints. In [Section 4](#) we provide a linear programming formulation and duality results under certain linearity assumptions. In [Section 5](#) we present a detailed description of a cut generation-based solution method, and prove its correctness and finite convergence. In [Section 6](#) we apply our methodology to a more general class of problems, featuring multivariate preference constraints based on coherent risk measures. [Section 7](#) is dedicated to numerical results, while [Section 8](#) contains concluding remarks.

2. Basic concepts and fundamental results In this section we aim to introduce a stochastic optimization framework for multi-objective (multi-criteria) decision making problems where the decision leads to a vector of random outcomes which is required to be preferable to a reference random outcome vector. We begin by discussing some widely used risk measures and associated relations which can be used to establish preferences between scalar-valued random variables. We also recall and generalize some fundamental results on the connections between these relations. Next, we extend these relations to vector-valued random variables, and present a general form of optimization problems involving them as constraints.

REMARK 2.1 Throughout our paper larger values of random variables are considered to be preferable. In this context, risk measures are often referred to as acceptability functionals, since higher values indicate less risky, i.e., more acceptable random outcomes. In the literature the opposite convention (where small values are preferred) is also widespread, especially when dealing with loss functions. When citing such sources, the definitions and formulas are altered to reflect this difference.

2.1 VaR, CVaR, and second order stochastic dominance We now present some basic definitions and results related to the risk measure CVaR. Unless otherwise specified, all random variables in this paper are assumed to be integrable (i.e., in \mathcal{L}^1), which ensures that the following definitions and formulas are valid. For a more detailed exposition on the concepts described below we refer to Pflug and Römisch (2007) and Rockafellar (2007).

- Let V be a random variable with a cumulative distribution function (CDF) denoted by F_V . The *value-at-risk* (VaR) at confidence level $\alpha \in (0, 1]$, also known as the α -quantile, is defined as

$$\text{VaR}_\alpha(V) = \inf\{\eta : F_V(\eta) \geq \alpha\}. \quad (1)$$

- The *conditional value-at-risk* at confidence level α is defined (Rockafellar and Uryasev, 2000; 2002) as

$$\text{CVaR}_\alpha(V) = \sup \left\{ \eta - \frac{1}{\alpha} \mathbb{E}([\eta - V]_+) : \eta \in \mathbb{R} \right\}, \quad (2)$$

where $[x]_+ = \max(x, 0)$ denotes the *positive part* of a number $x \in \mathbb{R}$.

- It is well known (Rockafellar and Uryasev, 2002) that if $\text{VaR}_\alpha(V)$ is finite, the supremum in the above definition is attained at $\eta = \text{VaR}_\alpha(V)$, i.e.,

$$\text{CVaR}_\alpha(V) = \text{VaR}_\alpha(V) - \frac{1}{\alpha} \mathbb{E}([\text{VaR}_\alpha(V) - V]_+). \quad (3)$$

- CVaR is also known in the literature as *average value-at-risk* and *tail value-at-risk*, due to the following expression:

$$\text{CVaR}_\alpha(V) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma(V) d\gamma. \quad (4)$$

The term $\mathbb{E}([\eta - V]_+)$ introduced in (2) is known as the *expected shortfall*¹, and it is closely related to the *second order distribution function* $F_V^{(2)} : \mathbb{R} \rightarrow \mathbb{R}$ of the random variable V defined by

$$F_V^{(2)}(\eta) = \int_{-\infty}^\eta F_V(\xi) d\xi.$$

Using integration by parts we obtain the following well-known equality:

$$F_V^{(2)}(\eta) = \int_{-\infty}^\eta F_V(\xi) d\xi = \eta F_V(\eta) - \int_{-\infty}^\eta \xi dF_V(\xi) = \int_{-\infty}^\eta (\eta - \xi) dF_V(\xi) = \mathbb{E}([\eta - V]_+). \quad (5)$$

CVaR is a widely used risk measure with significant advantages over VaR, due to a number of useful properties. For example, in contrast to VaR, the risk measure CVaR_α is *coherent* (Pflug, 2000), and serves as a fundamental building block for other coherent measures (see Section 6 for more details). Furthermore, for a given random variable V the mapping $\alpha \mapsto \text{CVaR}_\alpha$ is continuous and non-decreasing. CVaR can be used to express a wide range of risk preferences, including risk neutral (for $\alpha = 1$) and pessimistic worst-case (for sufficiently small values of α) approaches. We now introduce notation to express some risk preference relations associated with CVaR.

- Let V_1 and V_2 be two random variables with respective CDFs F_{V_1} and F_{V_2} . We say that V_1 is *CVaR-preferable* to V_2 at confidence level α , denoted as $V_1 \succ_{\text{CVaR}_\alpha} V_2$, if

$$\text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2). \quad (6)$$

¹ Acerbi (2002) uses the phrase *expected shortfall* to refer to CVaR itself.

- We say that V_1 is *second-order stochastically dominant* over V_2 (or that V_1 *dominates* V_2 in the second order), denoted as $V_1 \succ_{(2)} V_2$, if $F_{V_1}^{(2)}(\eta) \leq F_{V_2}^{(2)}(\eta)$ holds for all $\eta \in \mathbb{R}$.

We proceed by examining the close connection between CVaR-preferability and second-order stochastic dominance (SSD) relations. It is well known (Ogryczak and Ruszczyński, 2002; Dentcheva and Ruszczyński, 2006; Pflug and Römisch, 2007) that an SSD constraint is equivalent to the continuum of CVaR constraints for all confidence levels $\alpha \in (0, 1]$, i.e.,

$$V_1 \succ_{(2)} V_2 \iff \text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2) \text{ for all } \alpha \in (0, 1]. \quad (7)$$

Part (i) of the next proposition generalizes this result, while part (iii) shows that when the probability space is finite, SSD constraints can be reduced to a finite number of CVaR inequalities. The proof can be found in Appendix A.1.

PROPOSITION 2.1 *Let V_1 and V_2 be two random variables on the probability space $(\Omega, \mathcal{A}, \Pi)$, with respective CDFs F_{V_1} and F_{V_2} . For a tolerance parameter $\iota \in \mathbb{R}$ we define the relaxed ($\iota \geq 0$) or tightened ($\iota \leq 0$) SSD relation $V_1 \succ_{(2),\iota} V_2$ by*

$$F_{V_1}^{(2)}(\eta) \leq F_{V_2}^{(2)}(\eta) + \iota \quad \text{for all } \eta \in \mathbb{R}. \quad (8)$$

- (i) *The relation $V_1 \succ_{(2),\iota} V_2$ holds if and only if we have*

$$\text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2) - \frac{\iota}{\alpha} \quad \text{for all } \alpha \in (0, 1]. \quad (9)$$

- (ii) *Let $\mathcal{K} = \{\Pi(S) : S \in \mathcal{A}, \Pi(S) > 0\}$ denote the set of all non-zero probabilities of events. Then the relation $V_1 \succ_{(2),\iota} V_2$ holds if and only if we have*

$$\text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2) - \frac{\iota}{\alpha} \quad \text{for all } \alpha \in \mathcal{K}. \quad (10)$$

- (iii) *If the probability space is finite, then so is the set \mathcal{K} . In addition, if all elementary events in $\Omega = \{\omega_1, \dots, \omega_n\}$ have equal probability, then the relation $V_1 \succ_{(2),\iota} V_2$ holds if and only if we have*

$$\text{CVaR}_{\frac{k}{n}}(V_1) \geq \text{CVaR}_{\frac{k}{n}}(V_2) - \frac{n\iota}{k} \quad \text{for all } k = 1, \dots, n.$$

REMARK 2.2 *We have introduced relaxed SSD relations of the form (8) since these appear in the study Hu et al. (2011b), which forms the basis of our numerical experiments in Section 7.2. However, equation (9) shows that these relaxations carry little information about the tails of the distributions, as the tolerance term $\frac{\iota}{\alpha}$ becomes excessively large for confidence levels near zero. Relations of the form $V_1 \succ_{(2)} V_2 - \vartheta$ provide a more natural relaxation (see, e.g., the ‘scaled tails’ approach in Fábíán et al., 2011), and can also be easily formulated in terms of CVaR as follows: $\text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2) - \vartheta$ for all $\alpha \in (0, 1]$.*

To conclude this section, we briefly discuss some connections with utility theory. It is well-known (Müller and Stoyan, 2002) that the SSD relation $V_1 \succ_{(2)} V_2$ is equivalent to the continuum of expected utility inequalities $\mathbb{E}(u(V_1)) \leq \mathbb{E}(u(V_2))$ for all concave non-decreasing (i.e., risk-averse) utility functions u . On the other hand, according to (3) one can view $\text{CVaR}_\alpha(V)$ as the expected value of $U_V(V)$, where $U_V(t) = \text{VaR}_\alpha(V) - \frac{1}{\alpha}[\text{VaR}_\alpha(V) - t]_+$ is a *probability-dependent utility function* (Street, 2009). In this context the relation (6) can also be interpreted in terms of expected utilities as

$$\mathbb{E}(U_{V_1}(V_1)) \geq \mathbb{E}(U_{V_2}(V_2)).$$

We can now view both sides of the equivalence (7) as continuums of expected utility inequalities.

2.2 Comparing random vectors via scalarization To be able to tackle multiple criteria we need to extend scalar-based preferences to vector-valued random variables. The key concept is to consider a family of *scalarization functions* and require that all scalarized versions of the random variables conform to some preference relation. In order to eventually obtain computationally tractable formulations, we restrict ourselves to linear scalarization functions.

DEFINITION 2.1 *Let \preceq be a preordering of scalar-valued random variables, and let $C \subset \mathbb{R}^d$ be a set of scalarization vectors. Given two d -dimensional random vectors \mathbf{X} and \mathbf{Y} we say that \mathbf{X} is \preceq -preferable to \mathbf{Y} with respect to C , denoted as $\mathbf{X} \succ^C \mathbf{Y}$, if*

$$\mathbf{c}^T \mathbf{X} \succ \mathbf{c}^T \mathbf{Y} \text{ holds for all } \mathbf{c} \in C.$$

REMARK 2.3 *A natural way to compare two random vectors $\mathbf{X} = (X_1, \dots, X_d)$ and $\mathbf{Y} = (Y_1, \dots, Y_d)$ is by coordinate-wise preference: we say that \mathbf{X} is preferable to \mathbf{Y} if $X_l \succ Y_l$ for all $l = 1, \dots, d$. It is easy to see that this is a special case of Definition 2.1 obtained with the choice $C = \{\mathbf{e}_1, \dots, \mathbf{e}_d\}$, where $\mathbf{e}_l = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^d$ is the unit vector with the 1 in the l th position. In addition, whenever we have $\{\mathbf{e}_1, \dots, \mathbf{e}_d\} \subset C$, preference with respect to C implies coordinate-wise preference. Notably, this is the case for the positive linear SSD relation mentioned below.*

An example of the type of preference rule introduced in Definition 2.1 has been suggested under the name *positive linear SSD* by Dentcheva and Ruszczyński (2009), with the choice $C = \mathbb{R}_+^d$, and \preceq representing the SSD relation $\preceq_{(2)}$. Homem-de-Mello and Mehrotra (2009) generalize this approach by allowing $C \subset \mathbb{R}^d$ to be an arbitrary polyhedron, leading to the concept of *polyhedral linear SSD*. Their idea is motivated by the observation that, by taking C to be a proper subset of the positive orthant, polyhedral dominance can be a significantly less restrictive constraint than positive linear dominance. This reflects a wider trend in recent literature suggesting that in a practical optimization context stochastic dominance relations are often excessively hard to satisfy. Attempts to weaken stochastic dominance relations in order to extend the feasible region have resulted in the study of concepts such as *almost stochastic dominance* and *stochastically weighted stochastic dominance* (Leshno and Levy, 2002; Lizyayev and Ruszczyński, 2011; Hu et al., 2011b). Recalling Proposition 2.1, another natural way to relax the stochastic dominance relation is to require CVaR-preferability only at certain confidence levels, as opposed to the full continuum of constraints. This motivates us to introduce a special case of Definition 2.1.

DEFINITION 2.2 (MULTIVARIATE CVaR RELATION) *Let \mathbf{X} and \mathbf{Y} be two d -dimensional random vectors, $C \subset \mathbb{R}^d$ a set of scalarization vectors, and $\alpha \in (0, 1]$ a specified confidence level. We say that \mathbf{X} is CVaR-preferable to \mathbf{Y} at confidence level α with respect to C , denoted as $\mathbf{X} \succ_{\text{CVaR}_\alpha}^C \mathbf{Y}$, if*

$$\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) \geq \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y}) \quad \text{for all } \mathbf{c} \in C. \quad (11)$$

In our following analysis we focus on CVaR-preferability with respect to polyhedral scalarization sets. We begin by proving a close analogue of Proposition 1 in Homem-de-Mello and Mehrotra (2009), which shows that in these cases we can assume without loss of generality that the polyhedron C is compact, i.e., a polytope.

PROPOSITION 2.2 *Let C be a nonempty convex set, and let $\tilde{C} = \{\mathbf{c} \in \text{cl cone}(C) : \|\mathbf{c}\|_1 \leq 1\}$, where $\text{cl cone}(C)$ denotes the closure of the conical hull of the set C . Then, given any integrable random vectors \mathbf{X} and \mathbf{Y} the relations $\mathbf{X} \succ_{\text{CVaR}_\alpha}^C \mathbf{Y}$ and $\mathbf{X} \succ_{\text{CVaR}_\alpha}^{\tilde{C}} \mathbf{Y}$ are equivalent for all confidence levels $\alpha \in (0, 1]$.*

PROOF. For any non-zero vector $\mathbf{c} \in C$ we have $\frac{\mathbf{c}}{\|\mathbf{c}\|_1} \in \tilde{C}$. Since CVaR is positive homogenous it immediately follows that the relation $\mathbf{X} \succ_{\text{CVaR}_\alpha}^{\tilde{C}} \mathbf{Y}$ implies $\mathbf{X} \succ_{\text{CVaR}_\alpha}^C \mathbf{Y}$. On the other hand, let us assume that $\mathbf{X} \succ_{\text{CVaR}_\alpha}^C \mathbf{Y}$ and consider a non-zero vector $\tilde{\mathbf{c}} = \sum_{i=1}^k \lambda_i \mathbf{c}_i \in \text{cone}(C)$, where $\lambda_i > 0$ and $\mathbf{c}_i \in C$ for all $i = 1, \dots, k$. Since C is convex, we have $\frac{\tilde{\mathbf{c}}}{\sum_{i=1}^k \lambda_i} \in C$, implying

$$\text{CVaR}_\alpha(\tilde{\mathbf{c}}^T \mathbf{X}) \geq \text{CVaR}_\alpha(\tilde{\mathbf{c}}^T \mathbf{Y}) \quad \text{for all } \tilde{\mathbf{c}} \in \text{cone}(C). \quad (12)$$

Finally, let $\bar{\mathbf{c}}$ be a vector in \tilde{C} . Since $\tilde{C} \subset \text{cl cone}(C)$, there exists a sequence $\{\mathbf{c}_k\} \subset \text{cone}(C)$ such that $\mathbf{c}_k \rightarrow \bar{\mathbf{c}}$, which also implies $\|\mathbf{c}^T \mathbf{X} - \mathbf{c}_k^T \mathbf{X}\|_1 \rightarrow 0$ and $\|\mathbf{c}^T \mathbf{Y} - \mathbf{c}_k^T \mathbf{Y}\|_1 \rightarrow 0$. As CVaR_α is continuous in the \mathcal{L}^1 -norm (Ruszczyński and Shapiro, 2006), we now have $\text{CVaR}_\alpha(\mathbf{c}_k^T \mathbf{X}) \rightarrow \text{CVaR}_\alpha(\bar{\mathbf{c}}^T \mathbf{X})$ and $\text{CVaR}_\alpha(\mathbf{c}_k^T \mathbf{Y}) \rightarrow \text{CVaR}_\alpha(\bar{\mathbf{c}}^T \mathbf{Y})$. Therefore, (12) implies the inequality $\text{CVaR}_\alpha(\bar{\mathbf{c}}^T \mathbf{X}) \geq \text{CVaR}_\alpha(\bar{\mathbf{c}}^T \mathbf{Y})$, which proves our claim. \square

2.3 Optimization with multivariate CVaR constraints Let $(\Omega, 2^\Omega, \Pi)$ be a finite probability space with $\Omega = \{\omega_1, \dots, \omega_n\}$ and $\Pi(\omega_i) = p_i$. Consider a multi-criteria decision making problem where the decision variable \mathbf{z} is selected from a feasible set Z , and associated random outcomes are determined by the outcome mapping $G : Z \times \Omega \rightarrow \mathbb{R}^d$. We introduce the following additional notation:

- For a given decision $\mathbf{z} \in Z$ the random outcome vector $G(\mathbf{z}) : \Omega \rightarrow \mathbb{R}^d$ is defined by $G(\mathbf{z})(\omega) = G(\mathbf{z}, \omega)$.
- For a given elementary event ω_i the mapping $g_i : Z \rightarrow \mathbb{R}^d$ is defined by $g_i(\mathbf{z}) = G(\mathbf{z}, \omega_i)$.

Let $f : Z \rightarrow \mathbb{R}$ be an objective function, \mathbf{Y} a d -dimensional *benchmark random vector*, $C \subset \mathbb{R}^d$ a polytope of scalarization vectors, and $\alpha \in (0, 1]$ a confidence level. Our goal is to provide an explicit mathematical programming formulation and, in some cases, a computationally tractable solution method to problems of the following form.

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succ_{C\text{VaR}_\alpha}^C \mathbf{Y} \\ & \mathbf{z} \in Z \end{aligned} \tag{GeneralP}$$

While the benchmark random vector can be defined on a probability space different from Ω , in practical applications it often takes the form $\mathbf{Y} = G(\bar{\mathbf{z}})$, where $\bar{\mathbf{z}} \in Z$ is a *benchmark decision*. For risk-averse decision makers typical choices for the confidence level are small values such as $\alpha = 0.05$.

In order to keep our exposition simple, in (GeneralP) we only consider a single CVaR constraint. However, all of our results and methods remain fully applicable for problems of the more general form

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succ_{C\text{VaR}_{\alpha_{ij}}}^{C_{ij}} \mathbf{Y}_i \quad i = 1, \dots, M, \quad j = 1, \dots, K_i \\ & \mathbf{z} \in Z, \end{aligned} \tag{13}$$

with CVaR constraints enforced for M multiple benchmarks, multiple confidence levels, and varying scalarization sets. In addition, constraints can be replaced by the relaxed versions introduced in (9). In Section 7.2.2 we present numerical results for a budget allocation problem featuring relaxed constraints on two benchmarks, enforced at up to nine confidence levels for each.

3. Main theoretical results In this section we provide the theoretical background necessary to develop, and prove the finite convergence of, our solution methods. We begin by expressing CVaR as the optimum of various minimization and maximization problems, then proceed to prove that in finite probability spaces one can replace scalarization polyhedra by a finite set of scalarization vectors. To conclude the section, we show that this finiteness result extends to multivariate SSD constraints, providing an alternative to the representation in Homem-de-Mello and Mehrotra (2009).

3.1 Alternative expressions of CVaR By definition, CVaR can be obtained as a result of a maximization problem. On the other hand, CVaR is also a *spectral risk measure* (Acerbi, 2002) and thus can be viewed as a weighted sum of the least favorable outcomes. This allows us to express CVaR as the optimum of minimization problems.

THEOREM 3.1 *Let V be a random variable with (not necessarily distinct) realizations v_1, \dots, v_n and corresponding probabilities p_1, \dots, p_n . Then, for a given confidence level $\alpha \in (0, 1]$ the optimum values of the following optimization problems are all equal to $\text{CVaR}_\alpha(V)$.*

(i)

$$\begin{aligned}
\max \quad & \eta - \frac{1}{\alpha} \sum_{i=1}^n p_i w_i \\
\text{s.t.} \quad & w_i \geq \eta - v_i \quad i = 1, \dots, n \\
& w_i \geq 0 \quad i = 1, \dots, n
\end{aligned} \tag{14}$$

(ii)

$$\begin{aligned}
\min \quad & \frac{1}{\alpha} \sum_{i=1}^n \gamma_i v_i \\
\text{s.t.} \quad & \sum_{i=1}^n \gamma_i = \alpha \\
& 0 \leq \gamma_i \leq p_i \quad i = 1, \dots, n
\end{aligned} \tag{15}$$

(iii)

$$\begin{aligned}
\min \quad & \Psi_\alpha(V, K, k) \\
\text{s.t.} \quad & K \subset \{1, \dots, n\} \\
& k \in \{1, \dots, n\} \setminus K \\
& \sum_{i \in K} p_i \leq \alpha \\
& \alpha - \sum_{i \in K} p_i \leq p_k,
\end{aligned} \tag{16}$$

$$\text{where } \Psi_\alpha(V, K, k) = \frac{1}{\alpha} \left[\sum_{i \in K} p_i v_i + \left(\alpha - \sum_{i \in K} p_i \right) v_k \right].$$

PROOF. It is easy to see that at an optimal solution of (14) we have $w_i = \max(\eta - v_i, 0) = [\eta - v_i]_+$. Therefore, by the definition given in (2), the optimum value equals $\text{CVaR}_\alpha(V)$. Problem (15) is equivalent to the linear programming dual of (14), therefore its optimum also equals $\text{CVaR}_\alpha(V)$.

Without loss of generality assume $v_1 \leq v_2 \leq \dots \leq v_n$, and let $k^* = \min \left\{ k \in \{1, \dots, n\} : \sum_{i=1}^k p_i \geq \alpha \right\}$. Since (15) is a continuous knapsack problem, the greedy solution given by the following formula is optimal.

$$\gamma_i^* = \begin{cases} p_i & i = 1, \dots, k^* - 1 \\ \alpha - \sum_{i=1}^{k^*-1} p_i & i = k^* \\ 0 & i = k^* + 1, \dots, n \end{cases}$$

Setting $K^* = \{1, \dots, k^* - 1\}$, the pair (K^*, k^*) is a feasible solution of (16) with objective value $\Psi_\alpha(V, K^*, k^*) = \text{CVaR}_\alpha(V)$. On the other hand, for any feasible solution (K, k) of (16) we can construct a feasible solution

$$\gamma_i = \begin{cases} p_i & i \in K \\ \alpha - \sum_{i \in K} p_i & i = k \\ 0 & i \notin K \cup \{k\} \end{cases}$$

of (15) with objective value $\Psi_\alpha(V, K, k)$. This implies that the optimum values of (15) and (16) coincide, which completes our proof. \square

REMARK 3.1 The minimization problem in (15) is equivalent to the well-known risk envelope-based dual representation of CVaR (see, e.g., [Rockafellar, 2007](#)), while the objective function in (16) is similar to the CVaR formula for an ordered set of realizations in [Rockafellar and Uryasev \(2002\)](#). We also mention that an alternative subset-based representation of CVaR can be found in [Künzi-Bay and Mayer \(2006\)](#).

COROLLARY 3.1 *A simple consequence of claim (i) in Theorem 3.1 is the well known fact that CVaR-relations can be represented by linear inequalities. For a benchmark value $b \in \mathbb{R}$ the inequality $\text{CVaR}_\alpha(V) \geq b$ holds if and only if there exist $\eta \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n$ satisfying the following system.*

$$\begin{aligned} \eta - \frac{1}{\alpha} \sum_{i=1}^n p_i w_i &\geq b \\ w_i &\geq \eta - v_i & i = 1, \dots, n, \\ w_i &\geq 0 & i = 1, \dots, n \end{aligned}$$

When realizations of the random variable V are equally likely, CVaR has alternative closed form representations, presented below. These results prove useful in developing tractable solution methods (see Section 5.4).

PROPOSITION 3.1 *Let V be a random variable with (not necessarily distinct) realizations v_1, \dots, v_n and corresponding equal probabilities $p_1 = \dots = p_n = \frac{1}{n}$.*

(i) *Let $v_{(1)} \leq v_{(2)} \leq \dots \leq v_{(n)}$ denote an ordering of the realizations. Then*

$$\text{CVaR}_{\frac{k}{n}}(V) = \frac{1}{k} \sum_{i=1}^k v_{(i)}$$

holds for all $k = 1, \dots, n$.

(ii) *For a confidence level $\alpha \in [\frac{k}{n}, \frac{k+1}{n})$, $k \in [n-1]$, we have*

$$\text{CVaR}_\alpha(V) = \lambda_\alpha \text{CVaR}_{\frac{k}{n}}(V) + (1 - \lambda_\alpha) \text{CVaR}_{\frac{k+1}{n}}(V),$$

where $\lambda_\alpha = \frac{k(k+1-\alpha n)}{\alpha n}$. Note that $0 < \lambda_\alpha \leq \lambda_{\frac{k}{n}} = 1$.

PROOF. Since $\text{VaR}_{\frac{k}{n}}(V) = v_{(k)}$, by (3) we have

$$\text{CVaR}_{\frac{k}{n}}(V) = v_{(k)} - \frac{n}{k} \sum_{i=1}^n p_i [v_{(k)} - v_{(i)}]_+ = v_{(k)} - \frac{1}{k} \sum_{i=1}^k (v_{(k)} - v_{(i)}) = \frac{1}{k} \sum_{i=1}^k v_{(i)},$$

proving (i). For $\alpha = \frac{k}{n}$ claim (ii) trivially holds. Now suppose that $\alpha \in (\frac{k}{n}, \frac{k+1}{n})$. Then $\text{VaR}_\alpha(V) = v_{(k+1)}$, and using (i) we have

$$\begin{aligned} \lambda_\alpha \text{CVaR}_{\frac{k}{n}}(V) + (1 - \lambda_\alpha) \text{CVaR}_{\frac{k+1}{n}}(V) &= \frac{k(k+1-\alpha n)}{\alpha n} \frac{1}{k} \sum_{i=1}^k v_{(i)} + \frac{(k+1)(\alpha n - k)}{\alpha n} \frac{1}{k+1} \sum_{i=1}^{k+1} v_{(i)} \\ &= v_{(k+1)} - \frac{1}{\alpha n} \sum_{i=1}^k (v_{(k+1)} - v_{(i)}) = v_{(k+1)} - \frac{1}{\alpha} \sum_{i=1}^n p_i [v_{(k+1)} - v_{(i)}]_+ = \text{CVaR}_\alpha(V). \end{aligned}$$

□

3.2 Finite representations of scalarization polyhedra For any nontrivial polyhedron C of scalarization vectors the corresponding CVaR-preferability constraint is equivalent by definition to a collection of infinitely many scalar-based CVaR constraints, one for each scalarization vector $\mathbf{c} \in C$. The next theorem shows that for finite probability spaces it is sufficient to consider a finite subset of these vectors, obtained as projections of the vertices of a higher dimensional polyhedron. Before formally stating this result, we introduce a simple geometric notion. Let us call a vector $\mathbf{c} \in \mathbb{R}^d$ a *d-vertex* of a polyhedron $P \subset \mathbb{R}^d \times \mathbb{R}^h$ if it can be extended into a vertex, i.e., if there exists some $\mathbf{y} \in \mathbb{R}^h$ such that (\mathbf{c}, \mathbf{y}) is a vertex of P .

THEOREM 3.2 *Let \mathbf{X} and \mathbf{Y} be d -dimensional random vectors with realizations $\mathbf{x}_1, \dots, \mathbf{x}_n$ and $\mathbf{y}_1, \dots, \mathbf{y}_m$, respectively. Let p_1, \dots, p_n and q_1, \dots, q_m denote the corresponding probabilities, and let $C \subset \mathbb{R}^d$ be a polytope of scalarization vectors. \mathbf{X} is CVaR-preferable to \mathbf{Y} at confidence level α with respect to C if and only if*

$$\text{CVaR}_\alpha(\mathbf{c}_\ell^T \mathbf{X}) \geq \text{CVaR}_\alpha(\mathbf{c}_\ell^T \mathbf{Y}) \quad \text{for all } \ell = 1, \dots, N,$$

where $\mathbf{c}_{(1)}, \dots, \mathbf{c}_{(N)}$ are the d -vertices of the (line-free) polyhedron

$$P(C, \mathbf{Y}) = \{(\mathbf{c}, \eta, \mathbf{w}) \in C \times \mathbb{R} \times \mathbb{R}_+^m : w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j, \quad j = 1, \dots, m\}. \quad (17)$$

PROOF. If \mathbf{X} is preferable to \mathbf{Y} , the condition trivially holds, since $\mathbf{c}_{(\ell)} \in C$ for all $\ell = 1, \dots, N$. Now assume that \mathbf{X} is not preferable to \mathbf{Y} . Then the optimal objective value Δ of the following problem is negative:

$$\min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) - \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y}). \quad (18)$$

Using Theorem 3.1 we can reformulate this problem as

$$\begin{aligned} \min \quad & \Psi_\alpha(\mathbf{c}^T \mathbf{X}, K, k) - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \\ \text{s.t.} \quad & K \subset \{1, \dots, n\} \\ & k \in \{1, \dots, n\} \setminus K \\ & \sum_{i \in K} p_i \leq \alpha \\ & \alpha - \sum_{i \in K} p_i \leq p_k \\ & w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \\ & w_j \geq 0 \quad j = 1, \dots, m \\ & \mathbf{c} \in C. \end{aligned} \quad (\text{SetBased})$$

Let $(K^*, k^*, \mathbf{c}^*, \eta^*, \mathbf{w}^*)$ be an optimal solution of (SetBased). Then, by fixing $K = K^*$ and $k = k^*$ we obtain the following problem, which clearly has the same optimal objective value Δ .

$$\begin{aligned} \min \quad & \Psi_\alpha(\mathbf{c}^T \mathbf{X}, K^*, k^*) - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \\ \text{s.t.} \quad & w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \\ & w_j \geq 0 \quad j = 1, \dots, m \\ & \mathbf{c} \in C. \end{aligned} \quad (\text{FixedSet})$$

Since $\Psi_\alpha(\mathbf{c}^T \mathbf{X}, K^*, k^*)$ is a linear function of \mathbf{c} , (FixedSet) is a linear program with feasible set $P(C, \mathbf{Y})$. Therefore, problem (FixedSet) has an optimal solution which is a vertex of $P(C, \mathbf{Y})$, i.e., of the form $(\mathbf{c}_{(\ell)}, \bar{\eta}, \bar{\mathbf{w}})$ for some $\ell \in \{1, \dots, N\}$. Let $V = \mathbf{c}_{(\ell)}^T \mathbf{X}$; then Theorem 3.1 implies that $\text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}) = \text{CVaR}_\alpha(V)$ is equal to the optimal objective value of the minimization problem (16). Since (K^*, k^*) is a feasible solution of (16), we have

$$\Psi_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}, K^*, k^*) \geq \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}). \quad (19)$$

Observe that if we fix $\mathbf{c} = \mathbf{c}_{(\ell)}$ in problem (FixedSet), it becomes

$$\Psi_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}, K^*, k^*) - \max \left\{ \eta - \frac{1}{\alpha} \mathbf{q}^T \mathbf{w} : w_j \geq \eta - \mathbf{c}_{(\ell)}^T \mathbf{y}_j, \quad j = 1, \dots, m, \quad \mathbf{w} \in \mathbb{R}_+^m \right\},$$

where by (2) the maximization term equals $\text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y})$. Consequently, taking into account (19) we have

$$0 > \Delta = \Psi_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}, K^*, k^*) - \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y}) \geq \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}) - \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y}), \quad (20)$$

which completes our proof. \square

COROLLARY 3.2 *Under the conditions of the previous theorem there exists an index $\ell \in \{1, \dots, N\}$ such that the d -vertex $\mathbf{c}_{(\ell)}$ of $P(C, \mathbf{Y})$ is an optimal solution of problem (18).*

PROOF. Let $\mathbf{c}_{(\ell)}$ be the d -vertex obtained as part of a vertex optimal solution to (FixedSet) like in the previous proof. By (20) we have $\text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}) - \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y}) \leq \Delta$, where Δ denotes the optimal objective value of the minimization problem (18). On the other hand, $\mathbf{c}_{(\ell)}$ is a feasible solution, which proves our claim. \square

REMARK 3.2 In Theorem 3.2 the confidence levels applied to the two sides coincide. However, this is not a necessary condition, as it is easy to verify that the same proof is valid for the following asymmetric relation with any $\alpha_1, \alpha_2 \in (0, 1]$:

$$\text{CVaR}_{\alpha_1}(\mathbf{c}^T \mathbf{X}) \geq \text{CVaR}_{\alpha_2}(\mathbf{c}^T \mathbf{Y}) \quad \text{for all } \mathbf{c} \in C.$$

An even more general form of this result, featuring a wider class of risk measures, will be presented in Section 6.2.2.

COROLLARY 3.3 Using our notation from Theorem 3.2, the random vector \mathbf{X} dominates \mathbf{Y} in polyhedral linear second order with respect to C if and only if

$$\mathbf{c}_{(\ell)}^T \mathbf{X} \succ_{(2)} \mathbf{c}_{(\ell)}^T \mathbf{Y} \quad \text{for all } \ell = 1, \dots, N.$$

PROOF. We show that the following statements are equivalent:

- (i) $\mathbf{c}^T \mathbf{X} \succ_{(2)} \mathbf{c}^T \mathbf{Y}$ for all $\mathbf{c} \in C$.
- (ii) $\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) \geq \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y})$ for all $\alpha \in (0, 1]$, $\mathbf{c} \in C$.
- (iii) $\text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{X}) \geq \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y})$ for all $\alpha \in (0, 1]$, $\ell = 1, \dots, N$.
- (iv) $\mathbf{c}_{(\ell)}^T \mathbf{X} \succ_{(2)} \mathbf{c}_{(\ell)}^T \mathbf{Y}$ for all $\ell = 1, \dots, N$.

Equivalences (i) \Leftrightarrow (ii) and (iii) \Leftrightarrow (iv) follow from the fact that, by Proposition 2.1, the SSD constraint is equivalent to the continuum of CVaR constraints for all $\alpha \in (0, 1]$. On the other hand, Theorem 3.2 implies the equivalence of (ii) and (iii). \square

REMARK 3.3 The previous result is closely related to Theorem 1 of Homem-de-Mello and Mehrotra (2009), where the continuous variable η in (17) is replaced by the finite set of terms $\mathbf{c}^T \mathbf{y}_j$ for $j = 1, \dots, m$, leading to a set of m lower-dimensional polyhedra instead of our single polyhedron $P(C, \mathbf{Y})$.

4. Linear programming formulation and duality From a practical perspective it is interesting to consider the case when the mappings f and G are linear, the set Z is polyhedral, and the probability space is finite. In this section we present a linear programming formulation and duality results for problem (GeneralP) under these assumptions. Let us introduce the following notation:

- $Z = \{\mathbf{z} \in \mathbb{R}^{r_1} : A\mathbf{z} \leq \mathbf{b}\}$ for some $A \in \mathbb{R}^{r_2 \times r_1}$ and $\mathbf{b} \in \mathbb{R}^{r_2}$.
- $f(\mathbf{z}) = \mathbf{f}^T \mathbf{z}$ for some vector $\mathbf{f} \in \mathbb{R}^{r_1}$.
- $G(\mathbf{z}, \omega) = \Gamma(\omega)\mathbf{z}$ for a random matrix $\Gamma : \Omega \rightarrow \mathbb{R}^{d \times r_1}$. In addition, let $\Gamma_i = \Gamma(\omega_i)$ for $i = 1 \dots, n$.

Using the above notation, problem (GeneralP) becomes

$$\begin{aligned} \max \quad & \mathbf{f}^T \mathbf{z} \\ \text{s.t.} \quad & \text{CVaR}_\alpha(\mathbf{c}^T \Gamma \mathbf{z}) \geq \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y}) \quad \text{for all } \mathbf{c} \in C \\ & A\mathbf{z} \leq \mathbf{b}. \end{aligned} \tag{LinearP}$$

By Corollary 3.1 scalar-based CVaR-relations can be represented using linear inequalities. Working under the assumption that C is a polytope, this allows us to formulate (LinearP) as a linear program (LP). For a finite

set $\tilde{C} = \{\tilde{\mathbf{c}}_{(1)}, \dots, \tilde{\mathbf{c}}_{(L)}\}$ we consider the following LP.

$$\begin{aligned}
\max \quad & \mathbf{f}^T \mathbf{z} \\
\text{s.t.} \quad & \eta_\ell - \frac{1}{\alpha} \sum_{i=1}^n p_i w_{i\ell} \geq \text{CVaR}_\alpha(\tilde{\mathbf{c}}_{(\ell)}^T \mathbf{Y}) & \ell = 1, \dots, L \\
& w_{i\ell} \geq \eta_\ell - \tilde{\mathbf{c}}_{(\ell)}^T \Gamma_i \mathbf{z} & i = 1, \dots, n, \ell = 1, \dots, L \\
& w_{i\ell} \geq 0 & i = 1, \dots, n, \ell = 1, \dots, L \\
& \mathbf{A} \mathbf{z} \leq \mathbf{b}
\end{aligned} \tag{FiniteP}(\tilde{C})$$

The next proposition is an easy consequence of Theorem 3.2 and Corollary 3.1.

PROPOSITION 4.1 *Let \hat{C} denote the set consisting of the d -vertices $\mathbf{c}_{(1)}, \dots, \mathbf{c}_{(N)}$ as defined in Theorem 3.2, and assume that the finite set \tilde{C} satisfies $\hat{C} \subset \tilde{C} \subset C$. Then a vector $\mathbf{z} \in \mathbb{R}^{r_1}$ is a feasible (optimal) solution of (LinearP) if and only if $(\mathbf{z}, \boldsymbol{\eta}^{\mathbf{z}}, \mathbf{w}^{\mathbf{z}})$ is a feasible (optimal) solution of (FiniteP)(\tilde{C}), where $\eta_\ell^{\mathbf{z}} = \text{VaR}_\alpha(\tilde{\mathbf{c}}_{(\ell)}^T \Gamma \mathbf{z})$ and $w_{i\ell}^{\mathbf{z}} = [\eta_\ell^{\mathbf{z}} - \tilde{\mathbf{c}}_{(\ell)}^T \Gamma_i \mathbf{z}]_+$.*

The LP formulation (FiniteP)(\tilde{C}) serves as the basis for our cut generation-based solution algorithm in Section 5. We now present a strong duality result and corresponding optimality conditions, which can provide a theoretical foundation for developing dual (column generation-type) or primal-dual solution methods. Let $\mathcal{M}_+^F(C)$ denote the set of all finitely supported finite non-negative measures on the scalarization polyhedron C , and consider the following dual problem to (LinearP):

$$\begin{aligned}
\min \quad & \boldsymbol{\lambda}^T \mathbf{b} - \int_C \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y}) \mu(d\mathbf{c}) \\
\text{s.t.} \quad & \mathbb{E}(\nu) = \mu \\
& \nu(\omega_i) \leq \frac{1}{\alpha} \mu & i = 1, \dots, n \\
& \mathbb{E} \left(\int_C \mathbf{c}^T \Gamma \nu(d\mathbf{c}) \right) = \boldsymbol{\lambda}^T \mathbf{A} - \mathbf{f}^T \\
& \boldsymbol{\lambda} \in \mathbb{R}_+^{r_2}, \mu \in \mathcal{M}_+^F(C), \nu : \Omega \rightarrow \mathcal{M}_+^F(C)
\end{aligned} \tag{LinearD}$$

THEOREM 4.1 *The problem (LinearP) has a finite optimum value if and only if (LinearD) does, in which case the two optimum values coincide. In addition, a feasible solution \mathbf{z} of (LinearP) and a feasible solution $(\boldsymbol{\lambda}, \mu, \nu)$ of (LinearD) are both optimal for their respective problems if and only if the following complementary slackness conditions hold:*

$$\begin{aligned}
\text{support}(\mu) & \subset \{ \mathbf{c} : \text{CVaR}_\alpha(\mathbf{c}^T \Gamma \mathbf{z}) = \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y}) \} \\
\text{support}(\nu(\omega_i)) & \subset \{ \mathbf{c} : \text{VaR}_\alpha(\mathbf{c}^T \Gamma \mathbf{z}) \geq \mathbf{c}^T \Gamma_i \mathbf{z} \} & i = 1, \dots, n \\
\text{support}(\frac{1}{\alpha} \mu - \nu(\omega_i)) & \subset \{ \mathbf{c} : \text{VaR}_\alpha(\mathbf{c}^T \Gamma \mathbf{z}) \leq \mathbf{c}^T \Gamma_i \mathbf{z} \} & i = 1, \dots, n \\
\boldsymbol{\lambda}^T (\mathbf{A} \mathbf{z} - \mathbf{b}) & = 0.
\end{aligned}$$

PROOF. We first recall the simple facts that for a finitely supported measure $\mu \in \mathcal{M}_+^F(C)$ and a function $h : C \rightarrow \mathbb{R}$ we have

$$\text{support}(\mu) = \{ \mathbf{c} \in C : \mu(\{\mathbf{c}\}) > 0 \} \quad \text{and} \quad \int_C h(\mathbf{c}) \mu(d\mathbf{c}) = \sum_{\mathbf{c} \in \text{support}(\mu)} h(\mathbf{c}) \mu(\{\mathbf{c}\}). \tag{21}$$

Let us now consider the linear programming dual of (FiniteP(\tilde{C})) for an arbitrary finite set $\tilde{C} = \{\tilde{\mathbf{c}}_{(1)}, \dots, \tilde{\mathbf{c}}_{(L)}\} \subset C$:

$$\begin{aligned}
\min \quad & \boldsymbol{\lambda}^T \mathbf{b} - \sum_{\ell=1}^L \mu_{\ell} \text{CVaR}_{\alpha}(\tilde{\mathbf{c}}_{(\ell)}^T \mathbf{Y}) \\
\text{s.t.} \quad & \sum_{i=1}^n p_i \nu_{i\ell} = \mu_{\ell} & \ell = 1, \dots, L \\
& \nu_{i\ell} \leq \frac{1}{\alpha} \mu_{\ell} & i = 1, \dots, n, \ell = 1, \dots, L \\
& \sum_{i=1}^n p_i \sum_{\ell=1}^L \nu_{i\ell} \tilde{\mathbf{c}}_{(\ell)}^T \Gamma_i = \boldsymbol{\lambda}^T A - \mathbf{f}^T \\
& \boldsymbol{\lambda} \in \mathbb{R}_+^{r_2}, \quad \boldsymbol{\mu} \in \mathbb{R}_+^L, \quad \boldsymbol{\nu} \in \mathbb{R}_+^{n \times L}.
\end{aligned} \tag{FiniteD}(\tilde{C})$$

Note that the above formulation slightly differs from the usual LP dual, since a scaling factor of p_i has been applied to each dual variable $\nu_{i\ell}$. The dual variable $\boldsymbol{\mu}$ naturally defines a measure $\mu \in \mathcal{M}_+^F(C)$ supported on the finite set \tilde{C} with $\mu(\{\tilde{\mathbf{c}}_{(\ell)}\}) = \mu_{\ell}$. Similarly, the dual variable $\boldsymbol{\nu}$ defines a random measure $\nu : \Omega \rightarrow \mathcal{M}_+^F(C)$, where each $\nu(\omega_i)$ is supported on \tilde{C} with $\nu(\omega_i)(\{\tilde{\mathbf{c}}_{(\ell)}\}) = \nu_{i\ell}$. Keeping in mind (21), it follows that for any feasible solution $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$ of (FiniteD(\tilde{C})) we have a corresponding feasible solution $(\boldsymbol{\lambda}, \mu, \nu)$ of (LinearD), with the same objective value. Conversely, for a feasible solution $(\boldsymbol{\lambda}, \mu, \nu)$ of (LinearD) and a finite set $\tilde{C} = \{\tilde{\mathbf{c}}_{(1)}, \dots, \tilde{\mathbf{c}}_{(L)}\}$ that contains $\text{support}(\nu) = \bigcup_{i=1}^n \text{support}(\nu(\omega_i))$ we can define a feasible solution $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$ of (FiniteD(\tilde{C})) with the same objective value by setting $\mu_{\ell} = \mu(\{\tilde{\mathbf{c}}_{(\ell)}\})$ and $\nu_{i\ell} = \nu(\omega_i)(\{\tilde{\mathbf{c}}_{(\ell)}\})$.

We now establish weak duality. Let \mathbf{z} and $(\boldsymbol{\lambda}, \mu, \nu)$ be feasible solutions of (LinearP) and (LinearD), respectively, and denote their corresponding objective values by OBF_P and OBF_D . Then $(\mathbf{z}, \boldsymbol{\eta}^{\mathbf{z}}, \mathbf{w}^{\mathbf{z}})$ and $(\boldsymbol{\lambda}, \mu, \nu)$ are feasible solutions of the LPs (FiniteP($\text{support}(\nu)$)) and (FiniteD($\text{support}(\nu)$)), again with corresponding objective values OBF_P and OBF_D . Since these LPs form a primal-dual pair, the inequality $\text{OBF}_P \leq \text{OBF}_D$ follows from the weak duality theorem of linear programming.

To prove strong duality, let us first assume that (LinearP) has a finite optimum OPT_P . Then, by Proposition 4.1 and linear programming duality, both of the LPs (FiniteP(\hat{C})) and (FiniteD(\hat{C})) have the same optimum OPT_P . For an optimal solution $(\boldsymbol{\lambda}, \mu, \nu)$ of the latter problem, $(\boldsymbol{\lambda}, \mu, \nu)$ is a feasible solution of (LinearD) with the same objective value of OPT_P . Since weak duality implies that the objective value for any feasible solution of (LinearD) is greater than or equal to OPT_P , the dual solution $(\boldsymbol{\lambda}, \mu, \nu)$ is necessarily optimal. Similarly, let us consider an arbitrary optimal solution $(\boldsymbol{\lambda}, \mu, \nu)$ of (LinearD), and let $\tilde{C} = \text{support}(\nu) \cup \hat{C}$. Then $(\boldsymbol{\lambda}, \mu, \nu)$ is an optimal solution of (FiniteD(\tilde{C})), which (again by LP duality) has the same optimum value as (FiniteP(\tilde{C})). According to Proposition 4.1, the problem (FiniteP(\tilde{C})) has the same optimum value as (LinearP), which proves our claim.

Finally, consider a feasible solution \mathbf{z} of (LinearP) and a feasible solution (λ, μ, ν) of (LinearD), and again let $\tilde{C} = \text{support}(\nu) \cup \hat{C}$. Then these solutions are simultaneously optimal if and only if $(\mathbf{z}, \boldsymbol{\eta}^{\mathbf{z}}, \mathbf{w}^{\mathbf{z}})$ and $(\boldsymbol{\lambda}, \mu, \nu)$ are optimal solutions of the LPs (FiniteP(\tilde{C})) and (FiniteD(\tilde{C})), respectively. This in turn is equivalent to the following set of linear programming complementary slackness conditions:

$$\begin{aligned}
\mu_{\ell} > 0 & \Rightarrow \eta_{\ell}^{\mathbf{z}} - \frac{1}{\alpha} \sum_{i=1}^n p_i w_{i\ell}^{\mathbf{z}} = \text{CVaR}_{\alpha}(\tilde{\mathbf{c}}_{(\ell)}^T \mathbf{Y}) & \ell = 1, \dots, L \\
\nu_{i\ell} > 0 & \Rightarrow \eta_{\ell}^{\mathbf{z}} \geq \tilde{\mathbf{c}}_{(\ell)}^T \Gamma_i \mathbf{z} & i = 1, \dots, n, \ell = 1, \dots, L \\
\nu_{i\ell} < \frac{1}{\alpha} \mu_{\ell} & \Rightarrow w_{i\ell}^{\mathbf{z}} = 0 & i = 1, \dots, n, \ell = 1, \dots, L \\
\boldsymbol{\lambda}^T (A\mathbf{z} - \mathbf{b}) & = 0.
\end{aligned}$$

By equation (3) and the definitions of $\boldsymbol{\eta}^{\mathbf{z}}$ and $\mathbf{w}^{\mathbf{z}}$, for all $\ell \in \{1, \dots, L\}$ we have

$$\eta_{\ell}^{\mathbf{z}} - \frac{1}{\alpha} \sum_{i=1}^n p_i w_{i\ell}^{\mathbf{z}} = \text{VaR}_{\alpha}(\tilde{\mathbf{c}}_{\ell}^T \Gamma \mathbf{z}) - \frac{1}{\alpha} \sum_{i=1}^n p_i [\text{VaR}_{\alpha}(\tilde{\mathbf{c}}_{\ell}^T \Gamma \mathbf{z}) - \tilde{\mathbf{c}}_{\ell}^T \Gamma_i \mathbf{z}]_+ = \text{CVaR}_{\alpha}(\tilde{\mathbf{c}}_{\ell}^T \Gamma \mathbf{z}).$$

In accordance with (21) we can now equivalently rewrite the first complementary slackness condition as $\text{support}(\mu) \subset \{\mathbf{c} : \text{CVaR}_\alpha(\mathbf{c}^T \Gamma \mathbf{z}) = \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y})\}$. Since the second and third conditions can be rewritten in a similar fashion, our claim follows. \square

REMARK 4.1 *The finite representation guaranteed by Theorem 3.2 allowed us to derive a strong duality result directly from linear programming duality, without relying on additional tools. We mention here that the dual problem (LinearD) essentially corresponds to Haar’s dual in the duality theory of linear semi-infinite programs. For an overview of Haar-type dual problems we refer to Bonnans and Shapiro (2000); numerical solution methods have also been explored by Goberna and Jornet (1996). In this more general framework strong duality can, under appropriate conditions, also be extended to the case of non-polyhedral scalarization sets (while still assuming that f and G are linear, and Z is polyhedral). Furthermore, Lagrangian duality theory of semi-infinite programs (see, e.g., Shapiro, 2005) can be applied directly to (GeneralP) for general C , f , G , and Z .*

5. Solution methods Here we develop methods to solve the multivariate CVaR-constrained optimization problem (GeneralP) in the case when the probability space is finite and the scalarization set C is polyhedral. We begin by briefly discussing a naive “brute force” approach based on vertex enumeration, which is made possible by the finite representation developed in Section 3.2.

According to Theorem 3.2, we can replace a scalarization polytope C by the finite set of the d -vertices of the (unbounded) polyhedron $P(C, \mathbf{Y})$ without affecting the set of feasible decisions. Accordingly, if we have access to the set of d -vertices, we can attempt to directly solve (GeneralP). However, enumerating the vertices of a polyhedron is an NP-hard problem (Khachiyan et al., 2008) with potentially exponential output size. Therefore, the usefulness of this approach is limited to small-scale instances, where it allows us to explicitly describe the feasible region (as seen on the examples in Section 7.1).

We now proceed to present a cut generation algorithm which avoids many of the pitfalls associated with an enumeration-based approach. After proving finite convergence, we provide a detailed discussion on implementing various steps of the algorithm.

5.1 A cut-generation algorithm In this section we present an iterative algorithm which solves our original problem (GeneralP) in the case when the objective function f is continuous, the outcome mapping $\mathbf{z} \mapsto G(\mathbf{z})$ is continuous in the \mathcal{L}^1 -norm, the scalarization set C is a non-empty polytope, and the feasible set Z is compact².

Each iteration consists of two steps: first we find an optimal solution \mathbf{z}^* of a relaxed problem obtained by replacing the scalarization set C with a finite subset $\tilde{C} \subset C$ (see the formulation (Master) in Algorithm 1). Then, given the associated outcome vector $\mathbf{X} = G(\mathbf{z}^*)$ we attempt to find a scalarization vector $\mathbf{c}^* \in C$ for which the corresponding condition

$$\text{CVaR}_\alpha(\mathbf{c}^{*T} \mathbf{X}) \geq \text{CVaR}_\alpha(\mathbf{c}^{*T} \mathbf{Y}) \quad (22)$$

is violated. We accomplish this by solving the *cut generation problem* (18). If the optimal objective value is non-negative, it follows that \mathbf{z}^* is an optimal solution of (GeneralP). Otherwise, by Corollary 3.2 there exists an optimal solution \mathbf{c}^* which is a d -vertex of the polyhedron $P(C, \mathbf{Y})$ introduced in (17). We find such a vector and add it to the set \tilde{C} , which creates a tighter relaxation to be solved in the next iteration. This corresponds to introducing the constraint (22), which is a valid cut for the current solution \mathbf{z}^* . Note that introducing the new constraint requires calculating the parameter $\text{CVaR}_\alpha(\mathbf{c}^{*T} \mathbf{Y})$. This simple calculation is automatically performed as a byproduct of solving the optimization problems presented in Sections 5.3-5.4. Algorithm 1 provides a formal description of our solution method.

²While the assumption of having a polyhedral scalarization set is essential to proving finite convergence, the compactness assumptions on C and Z are adopted for the ease of exposition only.

Algorithm 1 Cut-Generation Algorithm

1: Initialize a set of scalarization vectors $\tilde{C} = \{\tilde{\mathbf{c}}_{(1)}, \dots, \tilde{\mathbf{c}}_{(L)}\} \subset C$.

2: Solve the master problem

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & \text{CVaR}_\alpha(\tilde{\mathbf{c}}_{(\ell)}^T G(\mathbf{z})) \geq \text{CVaR}_\alpha(\tilde{\mathbf{c}}_{(\ell)}^T \mathbf{Y}) \quad \ell = 1, \dots, L \\ & \mathbf{z} \in Z. \end{aligned} \tag{Master}$$

3: **if** the master problem is infeasible **then**

4: Stop.

5: **else**

6: Let \mathbf{z}^* be an optimal solution.

7: Given the optimal decision vector \mathbf{z}^* set $\mathbf{X} = G(\mathbf{z}^*)$, and solve the cut generation problem

$$\min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) - \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y}). \tag{CutGen}$$

8: **if** the optimal objective value of the cut generation problem is nonnegative **then**

9: Stop.

10: **else**

11: Find an optimal solution $\tilde{\mathbf{c}}_{(L+1)}$ of the cut generation problem which is a d -vertex of $P(C, \mathbf{Y})$. Set $\tilde{C} = \tilde{C} \cup \{\tilde{\mathbf{c}}_{(L+1)}\}$ and $L = L + 1$, then go to Step 2.

12: **end if**

13: **end if**

REMARK 5.1 *A trivial way to perform the initialization in Step 1 is by setting $L = 0$ and $\tilde{C} = \emptyset$. However, since the cut generation problem often presents a computational bottleneck, more aggressive initialization strategies can improve the performance of the algorithm. When the master problem is comparatively easier to solve, considering a large initial scalarization set does not result in a significant burden. For instance, if the vertices $\hat{\mathbf{c}}_{(1)}, \dots, \hat{\mathbf{c}}_{(k)}$ of the scalarization polyhedron are known, setting $L = k$ and $\tilde{C} = \{\hat{\mathbf{c}}_{(1)}, \dots, \hat{\mathbf{c}}_{(k)}\}$ can provide a suitable initialization.*

While Algorithm 1 is presented for the case of a single CVaR constraint, it can naturally be extended to problems of the more general form (13). In this case a separate cut generation problem is defined for each pair of a benchmark vector and an associated confidence level. Note that, in contrast to the method proposed in Homem-de-Mello and Mehrotra (2009) to solve SSD-constrained models, the number of cut generation problems does not depend on the number of benchmark realizations.

THEOREM 5.1 *Algorithm 1 terminates after a finite number of iterations, and provides either an optimal solution of (GeneralP), or a proof of infeasibility.*

PROOF. We first recall that CVaR_α , as all coherent risk measures, is continuous in the \mathcal{L}^1 -norm (Ruszczyński and Shapiro, 2006). Therefore, under our assumptions both the master problem and the cut generation problem involve the optimization of a continuous function over a compact set. It follows that the master problem either has an optimal solution or it is infeasible, while the cut generation problem always has an optimal solution since its feasible set C is non-empty. In addition, Corollary 3.2 states that at least one of the optimal solutions of the cut generation problem is a d -vertex of $P(C, \mathbf{Y})$. Therefore, the cut generation algorithm operates as described, and can terminate in one of two ways:

- The master problem is infeasible. Since the master problem is formally a relaxation of (GeneralP), this constitutes a proof of infeasibility for our original problem.

- The optimum of the cut generation problem is non-negative. This implies that the current optimal solution \mathbf{z}^* of the master problem is a feasible, and therefore optimal, solution of (GeneralP).

It remains to show that the algorithm always terminates in a finite number of iterations. This follows from the fact that every non-terminating iteration introduces a distinct d -vertex of the polyhedron $P(C, \mathbf{Y})$, and the number of d -vertices is finite. \square

5.2 Solving the master problem Corollary 3.1 allows us to represent CVaR constraints by linear inequalities, leading to the following formulation of (Master).

$$\begin{aligned}
& \max f(\mathbf{z}) \\
& \text{s.t. } \eta_\ell - \frac{1}{\alpha} \sum_{i=1}^n p_i w_{i\ell} \geq \text{CVaR}_\alpha(\mathbf{c}_{(\ell)}^T \mathbf{Y}) \quad \ell = 1, \dots, L \\
& \quad w_{i\ell} \geq \eta_\ell - \mathbf{c}_{(\ell)}^T g_i(\mathbf{z}) \quad i = 1, \dots, n, \ell = 1, \dots, L \\
& \quad w_{i\ell} \geq 0 \quad i = 1, \dots, n, \ell = 1, \dots, L \\
& \quad \mathbf{z} \in Z
\end{aligned} \tag{23}$$

In the general case we can attempt to solve this problem using non-linear programming techniques, or, with appropriate assumptions on f and Z , a convex programming approach. Under the linearity assumptions of Section 4 the master problem becomes the linear program (FiniteP(\tilde{C})), providing a computationally tractable formulation.

5.3 Solving the cut generation problem In this section we consider two d -dimensional random vectors \mathbf{X} and \mathbf{Y} with realizations $\mathbf{x}_1, \dots, \mathbf{x}_n$ and $\mathbf{y}_1, \dots, \mathbf{y}_m$, respectively. Let p_1, \dots, p_n and q_1, \dots, q_m denote the corresponding probabilities, and let $C = \{\mathbf{c} \in \mathbb{R}^d : B\mathbf{c} \leq \mathbf{h}\}$ be a polytope of scalarization vectors for some matrix B and vector \mathbf{h} of appropriate dimensions. The cut generation problem at confidence level $\alpha \in (0, 1]$ involves either finding a vector $\mathbf{c} \in C$ such that $\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) < \text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y})$ or showing that such a vector does not exist. To accomplish this, we aim to solve the optimization problem (CutGen). Recalling Theorem 3.1, we represent $\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X})$ and $\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y})$ using formulations (15) and (14), respectively. This allows us to restate (CutGen) as a quadratic program:

$$\begin{aligned}
& \min \frac{1}{\alpha} \sum_{i=1}^n \gamma_i \mathbf{c}^T \mathbf{x}_i - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \\
& \text{s.t. } \sum_{i=1}^n \gamma_i = \alpha \\
& \quad 0 \leq \gamma_i \leq p_i \quad i = 1, \dots, n \\
& \quad w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \\
& \quad \mathbf{c} \in C, \mathbf{w} \in \mathbb{R}_+^m.
\end{aligned} \tag{24}$$

Note that this quadratic problem is not necessarily convex, and therefore can present a significant computational challenge. This motivates us to introduce an alternate *mixed integer linear programming* (MIP) formulation which is potentially more tractable.

According to (3) the supremum in the classical definition of CVaR_α is attained at VaR_α . Since the probability space is finite, $\text{VaR}_\alpha(\mathbf{c}^T \mathbf{X}) = \mathbf{c}^T \mathbf{x}_k$ for at least one $k \in \{1, \dots, n\}$, implying

$$\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) = \max \left\{ \mathbf{c}^T \mathbf{x}_k - \frac{1}{\alpha} \sum_{i=1}^n p_i [\mathbf{c}^T \mathbf{x}_k - \mathbf{c}^T \mathbf{x}_i]_+ : k \in \{1, \dots, n\} \right\}.$$

Representing $\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{Y})$ as before, we obtain the following intermediate formulation of (CutGen).

$$\begin{aligned}
\min \quad & z - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \\
\text{s.t.} \quad & z \geq \mathbf{c}^T \mathbf{x}_k - \frac{1}{\alpha} \sum_{i=1}^n p_i [\mathbf{c}^T \mathbf{x}_k - \mathbf{c}^T \mathbf{x}_i]_+ \quad k = 1, \dots, n \\
& w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \\
& \mathbf{c} \in C, \quad \mathbf{w} \in \mathbb{R}_+^m
\end{aligned} \tag{25}$$

The term $[\mathbf{c}^T \mathbf{x}_k - \mathbf{c}^T \mathbf{x}_i]_+$ is not linear. To obtain a MIP formulation we linearize it by introducing additional variables and constraints (a similar linearization is used in Homem-de-Mello and Mehrotra, 2009).

$$\min \quad z - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \tag{26}$$

$$\text{s.t.} \quad z \geq \mathbf{c}^T \mathbf{x}_k - \frac{1}{\alpha} \sum_{i=1}^n p_i v_{ik} \quad i = 1, \dots, n, \quad k = 1, \dots, n \tag{27}$$

$$v_{ik} - \delta_{ik} = \mathbf{c}^T \mathbf{x}_k - \mathbf{c}^T \mathbf{x}_i \quad i = 1, \dots, n, \quad k = 1, \dots, n \tag{28}$$

$$M \beta_{ik} \geq v_{ik} \quad i = 1, \dots, n, \quad k = 1, \dots, n \tag{29}$$

$$M(1 - \beta_{ik}) \geq \delta_{ik} \quad i = 1, \dots, n, \quad k = 1, \dots, n \tag{30}$$

$$\beta_{ik} \in \{0, 1\} \quad i = 1, \dots, n, \quad k = 1, \dots, n \tag{31}$$

$$v \in \mathbb{R}_+^{n \times n}, \quad \delta \in \mathbb{R}_+^{n \times n} \tag{32}$$

$$w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \tag{33}$$

$$B\mathbf{c} \leq \mathbf{h} \tag{34}$$

$$\mathbf{w} \in \mathbb{R}_+^m \tag{35}$$

Here M is a sufficiently large constant to make constraints (29) and (30) redundant whenever the left-hand side is positive. Constraints (29)-(32) ensure that only one of the variables v_{ik} and δ_{ik} is positive. Then by constraint (28) we have $v_{ik} = [\mathbf{c}^T \mathbf{x}_k - \mathbf{c}^T \mathbf{x}_i]_+$ for all pairs of i and k . The equivalence of the MIP (26)-(35) to (25) follows immediately.

REMARK 5.2 *The choice of the constant M can significantly impact computational performance. In order to achieve tighter bounds, M in constraints (29) and (30) can be replaced by $M_{ki} = \max_{\mathbf{c} \in C} [\mathbf{c}^T \mathbf{x}_k - \mathbf{c}^T \mathbf{x}_i]_+$ and $\hat{M}_{ki} = \max_{\mathbf{c} \in C} [\mathbf{c}^T \mathbf{x}_i - \mathbf{c}^T \mathbf{x}_k]_+$, respectively.*

The above formulation (26)-(35) contains $O(n^2)$ binary variables. In the next section we show that, in the special case when scalarization vectors are non-negative and all the outcomes of \mathbf{X} are equally likely, this can be reduced to $O(n)$.

5.4 Solving the cut generation problem in the equal probability case In this section we consider a polytope $C = \{\mathbf{c} \in \mathbb{R}_+^d : B\mathbf{c} \leq \mathbf{h}\}$ of non-negative scalarization vectors. Since we consider larger outcomes to be preferable, the assumption of non-negativity is justified. In addition, we assume that each realization of \mathbf{X} has probability $\frac{1}{n}$, and at first consider confidence levels of the form $\alpha = \frac{k}{n}$ for some $k \in \{1, \dots, n\}$. Recalling formula (15) in Theorem 3.1 and introducing the scaled variables $\beta_i = n\gamma_i$ we have

$$\text{CVaR}_{\frac{k}{n}}(\mathbf{c}^T \mathbf{X}) = \min \left\{ \frac{1}{k} \sum_{i=1}^n \beta_i \mathbf{c}^T \mathbf{x}_i \quad : \quad \sum_{i=1}^n \beta_i = k, \quad \beta \in [0, 1]^n \right\}.$$

The cut generation problem (24) now reads

$$\begin{aligned}
\min \quad & \frac{1}{k} \sum_{i=1}^n \beta_i \mathbf{c}^T \mathbf{x}_i - \eta + \frac{1}{\alpha} \sum_{j=1}^n q_j w_j \\
\text{s.t.} \quad & \sum_{i=1}^n \beta_i = k \\
& \boldsymbol{\beta} \in [0, 1]^n \\
& w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \\
& \mathbf{c} \in C, \mathbf{w} \in \mathbb{R}_+^m.
\end{aligned} \tag{36}$$

We linearize the quadratic terms $\beta_i \mathbf{c}^T \mathbf{x}_i$ that appear in the objective function of problem (36) by introducing some additional variables and constraints. Using the notation $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{id})^T$ we obtain a MIP formulation with n binary variables.

$$\min \quad \frac{1}{k} \sum_{i=1}^n \boldsymbol{\delta}_i^T \mathbf{x}_i - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \tag{37}$$

$$\text{s.t.} \quad \sum_{i=1}^n \beta_i = k \tag{38}$$

$$\boldsymbol{\beta} \in \{0, 1\}^n \tag{39}$$

$$0 \leq \delta_{il} \leq c_l \quad i = 1, \dots, n, \quad l = 1, \dots, d \tag{40}$$

$$\delta_{il} \leq M \beta_i \quad i = 1, \dots, n, \quad l = 1, \dots, d \tag{41}$$

$$-\delta_{il} + c_l \leq M(1 - \beta_i) \quad i = 1, \dots, n, \quad l = 1, \dots, d \tag{42}$$

$$w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j \quad j = 1, \dots, m \tag{43}$$

$$B\mathbf{c} \leq \mathbf{h} \tag{44}$$

$$\mathbf{c} \in \mathbb{R}_+^d, \quad \mathbf{w} \in \mathbb{R}_+^m, \tag{45}$$

where M is again a sufficiently large constant to make constraints (41) and (42) redundant whenever the right-hand side is positive. It is easy to see that constraints (39)-(42) guarantee that

$$\delta_{il} = \begin{cases} c_l & \text{if } \beta_i = 1 \\ 0 & \text{if } \beta_i = 0 \end{cases} \quad \text{for all } i = 1, \dots, n, \quad l = 1, \dots, d.$$

Therefore, we have $\sum_{i=1}^n \boldsymbol{\delta}_i^T \mathbf{x}_i = \sum_{i=1}^n \beta_i \mathbf{c}^T \mathbf{x}_i$ which shows the equivalence of (36) and the MIP (37)-(45).

We proceed by extending the above formulation (37)-(45) to allow arbitrary confidence levels. The key observation is that for a given $\alpha \in [\frac{k}{n}, \frac{k+1}{n})$ Proposition 3.1 allows us to express $\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X})$ as a convex combination of $\text{CVaR}_{\frac{k}{n}}(\mathbf{c}^T \mathbf{X})$ and $\text{CVaR}_{\frac{k+1}{n}}(\mathbf{c}^T \mathbf{X})$:

$$\text{CVaR}_\alpha(\mathbf{c}^T \mathbf{X}) = \lambda_\alpha \text{CVaR}_{\frac{k}{n}}(\mathbf{c}^T \mathbf{X}) + (1 - \lambda_\alpha) \text{CVaR}_{\frac{k+1}{n}}(\mathbf{c}^T \mathbf{X}),$$

where $\lambda_\alpha = \frac{k(k+1-\alpha n)}{\alpha n}$. Analogously to the previous formulation, we express $\text{CVaR}_{\frac{k}{n}}$ and $\text{CVaR}_{\frac{k+1}{n}}$ using the

binary vectors $\beta^{(1)}$ and $\beta^{(2)}$, respectively. This leads to an alternate MIP representation of (CutGen):

$$\begin{aligned}
\min \quad & \frac{\lambda_\alpha}{k} \sum_{i=1}^n \delta_i^{(1)T} \mathbf{x}_i + \frac{(1-\lambda_\alpha)}{k} \sum_{i=1}^n \delta_i^{(2)T} \mathbf{x}_i - \eta + \frac{1}{\alpha} \sum_{j=1}^m q_j w_j \\
\text{s.t.} \quad & \sum_{i=1}^n \beta_i^{(1)} = k \\
& \delta_{il}^{(1)} \leq M \beta_i^{(1)} & i = 1, \dots, n, \quad l = 1, \dots, d \\
& 0 \leq \delta_{il}^{(1)} \leq c_l & i = 1, \dots, n, \quad l = 1, \dots, d \\
& -\delta_{il}^{(1)} + c_l \leq M(1 - \beta_i^{(1)}) & i = 1, \dots, n, \quad l = 1, \dots, d \\
& \beta^{(1)} \in \{0, 1\}^n \\
& \sum_{i=1}^n \beta_i^{(2)} = k + 1 \\
& \delta_{il}^{(2)} \leq M \beta_i^{(2)} & i = 1, \dots, n, \quad l = 1, \dots, d \\
& 0 \leq \delta_{il}^{(2)} \leq c_l & i = 1, \dots, n, \quad l = 1, \dots, d \\
& -\delta_{il}^{(2)} + c_l \leq M(1 - \beta_i^{(2)}) & i = 1, \dots, n, \quad l = 1, \dots, d \\
& \beta^{(2)} \in \{0, 1\}^n \\
& w_j \geq \eta - \mathbf{c}^T \mathbf{y}_j & j = 1, \dots, m \\
& B\mathbf{c} \leq \mathbf{h} \\
& \mathbf{c} \in \mathbb{R}_+^d, \quad \mathbf{w} \in \mathbb{R}_+^m.
\end{aligned} \tag{46}$$

REMARK 5.3 Similarly to the general case in Section 5.3, the parameter M in constraints (41) and (42), as well as in their counterparts in (46), can be replaced by $M_l = \max\{c_l : \mathbf{c} \in C\}$.

To conclude this section, we present a set of valid inequalities to strengthen the MIP formulation in (46).

PROPOSITION 5.1 There exists an optimal solution to the problem (46) satisfying the relations below.

$$\begin{aligned}
\sum_{i=1}^n \beta_i^{(2)} - \beta_i^{(1)} &= 1 \\
\beta^{(1)} &\leq \beta^{(2)}
\end{aligned}$$

PROOF. Keeping in mind the knapsack structure explored in the proof of Theorem 3.1, note that in the above formulation $\text{CVaR}_{\frac{k}{n}}(\mathbf{c}^T \mathbf{X})$ and $\text{CVaR}_{\frac{k+1}{n}}(\mathbf{c}^T \mathbf{X})$ are expressed as the mean of k and $k+1$ smallest realizations of the random variable $\mathbf{c}^T \mathbf{X}$, respectively. The selection of realizations to be featured in these means is encoded by the binary variables $\beta^{(1)}$ and $\beta^{(2)}$. While some of the realizations $\mathbf{c}^T \mathbf{x}_1, \dots, \mathbf{c}^T \mathbf{x}_n$ might coincide, our claim immediately follows from the trivial observation that a set of k smallest realizations can always be extended to a set of $k+1$ smallest realizations by adding to it a single new realization. For example, the choice of the lexicographically smallest optimal vectors $\beta^{(1)}$ and $\beta^{(2)}$ provides a solution with the desired properties. \square

5.5 Finding a d -vertex solution The provable finite convergence of Algorithm 1 depends on finding a solution to the cut generation problem which is d -vertex of the polyhedron $P(C, \mathbf{Y})$. Let \mathbf{c}^* be an optimal solution obtained using one of the methods outlined in Sections 5.3 and 5.4, and let π be a permutation describing a non-decreasing ordering of the realizations of the random vector $\mathbf{c}^{*T} \mathbf{X}$, i.e., $\mathbf{c}^{*T} \mathbf{x}_{\pi(1)} \leq \dots \leq \mathbf{c}^{*T} \mathbf{x}_{\pi(n)}$. Defining

$$k^* = \min \left\{ k \in \{1, \dots, n\} : \sum_{i=1}^k p_{\pi(i)} \geq \alpha \right\} \quad \text{and} \quad K^* = \{\pi(1), \dots, \pi(k^* - 1)\}, \tag{47}$$

we can obtain the desired d -vertex solution $\hat{\mathbf{c}}$ by finding a vertex optimal solution $(\hat{\mathbf{c}}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}})$ of the linear program ([FixedSet](#)). According to Corollary 3.2 the vector $\hat{\mathbf{c}}$ is also an optimal solution of ([CutGen](#)). We remark that this step is often redundant in practice, since MIP solvers typically provide vertex solutions.

6. Coherent risk measures In this section we outline how our methodology for handling multivariate preference constraints based on the risk measure CVaR can be extended to the wider class of coherent risk measures, introduced in the seminal paper by [Artzner et al. \(1999\)](#).

Consider the set $\mathcal{V} = \mathcal{V}(\Omega, 2^\Omega, \Pi)$ of all random variables on a finite probability space. We say that a mapping $\rho : \mathcal{V} \rightarrow \mathbb{R}$ is a *coherent risk measure* if it has the following properties (for all $V, V_1, V_2 \in \mathcal{V}$):

- *Monotone*: $V_1 \leq V_2 \Rightarrow \rho(V_1) \leq \rho(V_2)$.
- *Superadditive*: $\rho(V_1 + V_2) \geq \rho(V_1) + \rho(V_2)$.
- *Positive homogeneous*: $\rho(\lambda V) = \lambda \rho(V)$ for all $\lambda \geq 0$.
- *Translation invariant*: $\rho(V + \lambda) = \rho(V) + \lambda$.

Risk measures that depend only on the distributions of random variables are of particular importance, in a large part because they can be consistently estimated from empirical data. Denoting the family of CDFs for all random variables by $\mathcal{F} = \mathcal{F}(\Omega, 2^\Omega, \Pi) = \{F_V : V \in \mathcal{V}(\Omega, \mathcal{A}, \Pi)\}$ we say that a mapping $\rho : \mathcal{V} \rightarrow \mathbb{R}$ is *law invariant* if the value $\rho(V)$ depends only on the distribution of the random variable V , i.e., if there exists a mapping $\varphi_\rho : \mathcal{F} \rightarrow \mathbb{R}$ such that $\rho(V) = \varphi_\rho(F_V)$ holds for all $V \in \mathcal{V}$. Note that in this case φ_ρ is uniquely determined by ρ .

We now mention an important subclass of law invariant coherent risk measures. A mapping $\rho : \mathcal{V} \rightarrow \mathbb{R}$ that has a representation of the form

$$\rho(V) = \int_0^1 \phi(\alpha) \text{VaR}_\alpha(V) d\alpha \quad \text{for all } V \in \mathcal{V}$$

for some non-increasing function $\phi : [0, 1] \rightarrow \mathbb{R}_+$ satisfying $\|\phi\|_1 = 1$ is called a *spectral risk measure* ([Acerbi, 2002](#)). Spectral risk measures have received significant attention in the recent literature (see, e.g., [Acerbi, 2004](#)) due to their connection to risk aversion, and the fact that they have the so-called *comonotone additive* property. This property is often desirable in a financial context because it states that ρ does not reward risk pooling for “worst-case” dependence structures.

Similarly to the notion of CVaR-preferability introduced in (6), we can consider coherent risk measure-based preference relations among random variables. We say that V_1 is ρ -preferable to V_2 , denoted by $V_1 \succ_\rho V_2$, if the inequality $\rho(V_1) \geq \rho(V_2)$ holds. Utilizing Definition 2.1, the relation \succ_ρ gives rise to *multivariate coherent risk preferences* of the form \succ_ρ^C among random vectors. We can then consider the following more general variant of problem ([GeneralP](#)), featuring multivariate coherent risk preference constraints:

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & G(\mathbf{z}) \succ_\rho^C \mathbf{Y} \\ & \mathbf{z} \in Z. \end{aligned} \tag{GeneralP}_\rho$$

Paralleling our developments in Section 3, in Section 6.1 we describe some representations of coherent risk measures. In Section 6.2 we use these representations to obtain finite representations of scalarization polyhedra for a rich class of risk measures. Finally, in Section 6.3 we briefly discuss how our solution methods from Section 5 can be adapted to tackle optimization problems of the form ([GeneralP](#) $_\rho$).

6.1 Representations of coherent risk measures The fact that CVaR can be expressed as the optimum of a minimization problem proved very useful in the development of our theoretical results and solution methods. In this section we present similar representations of coherent risk measures as infimums.

It is well known (Pflug, 2000) that CVaR is a law invariant coherent risk measure. Moreover, CVaR forms a basic building block of coherent risk measures, as the fundamental theorem of Kusuoka (2001) shows. According to this theorem, in atomless probability spaces every law invariant coherent risk measure is an infimum of spectral risk measures, while spectral risk measures can be expressed as a mixture (integral convex combination) of CVaR measures. However, in this study our main focus is on finite probability spaces, and in such spaces Kusuoka representations do not always exist (for a constructive example see Pflug and Römisch, 2007). Nevertheless, law invariant coherent risk measures that do not have Kusuoka representations can be viewed as “pathological” in the sense that they cannot be coherently extended to other probability spaces. We now formalize this intuitive notion.

DEFINITION 6.1 *Consider a not necessarily atomless probability space $(\Omega, \mathcal{A}, \Pi)$, and a law invariant mapping $\rho : \mathcal{V}(\Omega, \mathcal{A}, \Pi) \rightarrow \mathbb{R}$. We say that ρ is a functionally coherent risk measure if there exists a law invariant coherent risk measure $\bar{\rho} : \mathcal{L}^p(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\Pi}) \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ on p -integrable random variables in an atomless probability space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\Pi})$ for some value $p \in [1, \infty]$ such that φ_ρ is a restriction of $\varphi_{\bar{\rho}}$, i.e., we have $\varphi_\rho = \varphi_{\bar{\rho}}|_{\mathcal{F}(\Omega, \mathcal{A}, \Pi)}$.*

Functionally coherent risk measures allow us to use Kusuoka representations even in probability spaces that are not atomless, as the next proposition shows. The proof can be found in Noyan and Rudolf (2012).

PROPOSITION 6.1 *Consider a finite probability space $(\Omega, 2^\Omega, \Pi)$, and the set $\mathcal{K} = \{\Pi(S) : S \subset \Omega, \Pi(S) > 0\}$ introduced in Proposition 2.1.*

- (i) *A mapping $\rho : \mathcal{V}(\Omega, \mathcal{A}, \Pi) \rightarrow \mathbb{R}$ is a spectral risk measure if and only if it can be written as a convex combination of finitely many CVaR measures, i.e., if it has a representation of the form*

$$\rho(V) = \sum_{i=1}^M \mu_i \text{CVaR}_{\alpha_i}(V) \quad \text{for all } V \in \mathcal{V}, \quad (48)$$

for some integer M , confidence levels $\alpha_1, \dots, \alpha_M \in \mathcal{K}$, and corresponding weights $\mu_1, \dots, \mu_M \in \mathbb{R}_+$ that satisfy $\sum_{i=1}^M \mu_i = 1$.

- (ii) *A mapping $\rho : \mathcal{V}(\Omega, \mathcal{A}, \Pi) \rightarrow \mathbb{R}$ is a functionally coherent risk measure if and only if it has a representation of the form*

$$\inf_{\mu \in \mathcal{M}} \sum_{i=1}^M \mu_i \text{CVaR}_{\alpha_i}(V) \quad \text{for all } V \in \mathcal{V}, \quad (49)$$

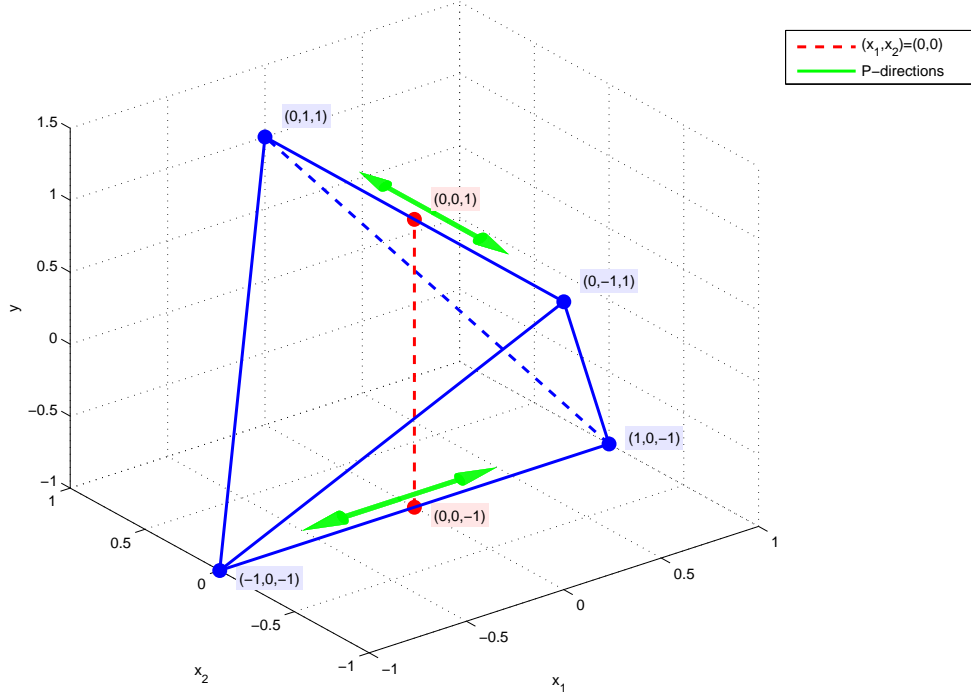
for some integer M , confidence levels $\alpha_1, \dots, \alpha_M \in \mathcal{K}$, and a family $\mathcal{M} \subset \left\{ \mu \in \mathbb{R}_+^M : \sum_{i=1}^M \mu_i = 1 \right\}$ of weight vectors.

We mention that the assumption of functional coherence is not particularly restrictive in our context, as law invariant risk measures that can be coherently extended to a family containing all finite discrete distributions are functionally coherent (Noyan and Rudolf, 2012), and thus have Kusuoka representations of the form (49).

To conclude this section, we present an alternative dual representation for coherent risk measures on finite probability spaces, due to Artzner et al. (1999). We mention that analogous results exist for more general probability spaces (see, e.g., Pflug and Römisch, 2007).

THEOREM 6.1 *Let $(\Omega, 2^\Omega, \Pi)$ be a finite probability space. For every coherent risk measure $\rho : \mathcal{V}(\Omega, 2^\Omega, \Pi) \rightarrow \mathbb{R}$ there exists a risk envelope $\mathcal{Q} \subset \{Q \in \mathcal{V} : Q \geq 0, \mathbb{E}(Q) = 1\}$ such that*

$$\rho(V) = \inf_{Q \in \mathcal{Q}} \mathbb{E}(QV) \quad \text{holds for all } V \in \mathcal{V}. \quad (50)$$

Figure 1: Tetrahedron P in Example 6.1

6.2 Finite representations of scalarization polyhedra For finite probability spaces, Theorem 3.2 shows that when the set of scalarization vectors is polyhedral, the multivariate CVaR constraints given in (11) can be reduced to finitely many univariate CVaR constraints. In this section we will extend this important finiteness result to constraints based on a rich class of coherent risk measures.

If a risk measure has a representation of the form (49) where the set \mathcal{M} is finite, then we say that it is a *finitely representable coherent risk measure*. The next result shows that, over bounded families of random variables, these risk measures are dense among functionally coherent risk measures; the proof can be found in Noyan and Rudolf (2012). Thus, finitely representable coherent risk measures can be used to closely approximate functionally coherent risk preferences in our proposed decision problems.

PROPOSITION 6.2 *Consider a family $\bar{\mathcal{V}}$ of random variables on a finite probability space, and assume that $\bar{\mathcal{V}}$ is bounded in the \mathcal{L}^1 -norm. Then, for any functionally coherent risk measure ρ and $\epsilon > 0$ there exists a finitely representable coherent risk measure $\bar{\rho}$ such that $|\rho(V) - \bar{\rho}(V)| < \epsilon$ holds for all $V \in \bar{\mathcal{V}}$.*

Before we proceed to prove the main result of this section, we need to provide some geometric background.

6.2.1 Geometric preliminaries The notation for this subsection is largely independent from that used for the rest of the paper. Let us recall that Theorem 3.2 provided a finite representation of the scalarization polyhedron C via the d -vertices of the polyhedron $P(C, \mathbf{Y})$. In order to extend this theorem to a more general class of risk measures, it will be necessary to consider more complicated polyhedra in place of $P(C, \mathbf{Y})$. Given a polyhedron $P = P^{(1)} \subset \mathbb{R}^n \times \mathbb{R}^m$ we introduce the following series of “liftings”:

$$P^{(k)} = \left\{ (\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}) \in \mathbb{R}^n \times \mathbb{R}^m \times \dots \times \mathbb{R}^m : (\mathbf{x}, \mathbf{y}^{(i)}) \in P \text{ for all } i = 1, \dots, k \right\} \quad (51)$$

The following example shows that lifting a polyhedron in the above manner can introduce additional n -vertices.

EXAMPLE 6.1 Let $P \subset \mathbb{R}^2 \times \mathbb{R}^1$ be the tetrahedron depicted in Figure 1 with vertices $(-1, 0, -1)$, $(1, 0, -1)$, $(0, -1, 1)$, $(0, 1, 1)$. In accordance with (51), let

$$P^{(2)} = \left\{ (x_1, x_2, y^{(1)}, y^{(2)}) : (x_1, x_2, y^{(1)}) \in P, (x_1, x_2, y^{(2)}) \in P \right\}.$$

In Appendix A.2 we show that the point $(0, 0)$ is not a 2-vertex of P , but it is a 2-vertex of $P^{(2)}$.

Even though the lifting procedure can introduce new n -vertices, the set of n -vertices of the series of polyhedra P_1, P_2, \dots eventually stabilizes. The proof of the next theorem can be found in Appendix A.2.

THEOREM 6.2 Let $P \subset \mathbb{R}^n \times \mathbb{R}^m$ be an arbitrary polyhedron, and let k be a positive integer. Then every n -vertex of the lifted polyhedron $P^{(k)}$ is also an n -vertex of $P^{(n)}$.

6.2.2 General finiteness proof We are now going to prove an analogue of the finiteness result in Theorem 3.2 for finitely representable coherent risk measures.

THEOREM 6.3 Let ρ be a finitely representable coherent risk measure on a finite probability space with elementary events $\omega_1, \dots, \omega_n$ and corresponding probabilities p_1, \dots, p_n . Consider a polytope $C \subset \mathbb{R}^d$ of scalarization vectors and a d -dimensional benchmark random vector \mathbf{Y} with realizations $\mathbf{y}_i = \mathbf{Y}(\omega_i)$ for $i = 1, \dots, n$.

- (i) There exists a polyhedron P such that, for any random d -dimensional random vector \mathbf{X} with realizations $\mathbf{x}_i = \mathbf{X}(\omega_i)$, the relation $\mathbf{X} \succ_{\rho}^C \mathbf{Y}$ is equivalent to the condition

$$\rho(\mathbf{c}_{(\ell)}^T \mathbf{X}) \geq \rho(\mathbf{c}_{(\ell)}^T \mathbf{Y}) \quad \text{for all } \ell = 1, \dots, N, \quad (52)$$

with $\mathbf{c}_{(1)}, \dots, \mathbf{c}_{(N)}$ denoting the d -vertices of P .

- (ii) If the risk measure ρ is spectral, then the above equivalence holds with the choice of the lifted polyhedron $P = P^{(d)}(C, \mathbf{Y})$, where the polyhedron $P(C, \mathbf{Y})$ is defined as in (17).

PROOF. Let us first assume that the relation $\mathbf{X} \succ_{\rho}^C \mathbf{Y}$ does not hold, implying

$$\inf_{\mathbf{c} \in C} \rho(\mathbf{c}^T \mathbf{X}) - \rho(\mathbf{c}^T \mathbf{Y}) < 0. \quad (53)$$

According to Theorem 6.1 there exists a risk envelope \mathcal{Q} that provides a dual representation of the form (50). Substituting into (53), we obtain

$$\inf \{ \mathbb{E}(Q \mathbf{c}^T \mathbf{X}) - \rho(\mathbf{c}^T \mathbf{Y}) \mid Q \in \mathcal{Q}, \mathbf{c} \in C \} < 0.$$

It follows that there exists a random variable $\hat{Q} \in \mathcal{Q}$ such that

$$\inf_{\mathbf{c} \in C} \mathbb{E}(\hat{Q} \mathbf{c}^T \mathbf{X}) - \rho(\mathbf{c}^T \mathbf{Y}) < 0 \quad (54)$$

holds. Since ρ is finitely representable, recalling part (ii) of Proposition 6.1 there exist confidence levels $\alpha_1, \dots, \alpha_M \in (0, 1]$ and weight vectors $\boldsymbol{\mu}^{(1)}, \dots, \boldsymbol{\mu}^{(H)}$ such that

$$\rho(V) = \min_{h \in \{1, \dots, H\}} \sum_{j=1}^M \mu_j^{(h)} \text{CVaR}_{\alpha_j}(V) \quad \text{for all } V \in \mathcal{V}. \quad (55)$$

Using Corollary 3.1, we can therefore express the infimum in (54) as the optimum of the following linear program:

$$\begin{aligned} \min \quad & \sum_{i=1}^n p_i \hat{Q}(\omega_i) \mathbf{c}^T \mathbf{x}_i - z \\ \text{s.t.} \quad & z \leq \sum_{j=1}^M \mu_j^{(h)} \left(\eta^{(j)} - \frac{1}{\alpha_j} \sum_{i=1}^n p_i w_i^{(j)} \right) & h = 1, \dots, H \\ & w_i^{(j)} \geq \eta^{(j)} - \mathbf{c}^T \mathbf{y}_i & i = 1, \dots, n, j = 1, \dots, M \\ & w_i^{(j)} \geq 0 & i = 1, \dots, n, j = 1, \dots, M \\ & \mathbf{c} \in C. \end{aligned} \quad (56)$$

Let P denote the feasible set of the above problem. Since P is a polyhedron, there exists an optimal solution $(\mathbf{c}^*, \boldsymbol{\eta}^*, \mathbf{w}^*, z^*)$ of (56) which is a vertex. Recalling (54) we now have

$$\rho(\mathbf{c}^{*T} \mathbf{X}) - \rho(\mathbf{c}^{*T} \mathbf{Y}) = \inf_{\mathbf{Q} \in \mathcal{Q}} \mathbb{E}(\mathbf{Q} \mathbf{c}^{*T} \mathbf{X}) - \rho(\mathbf{c}^{*T} \mathbf{Y}) \leq \mathbb{E}(\hat{\mathbf{Q}} \mathbf{c}^{*T} \mathbf{X}) - \rho(\mathbf{c}^{*T} \mathbf{Y}) = \inf_{\mathbf{c} \in C} \mathbb{E}(\hat{\mathbf{Q}} \mathbf{c}^T \mathbf{X}) - \rho(\mathbf{c}^T \mathbf{Y}) < 0.$$

As the vector \mathbf{c}^* is a d -vertex of P , the relation (52) is violated. On the other hand, notice that for every vector $(\mathbf{c}, \boldsymbol{\eta}, \mathbf{w}, z) \in P$ we have $\mathbf{c} \in C$, therefore the d -vertices of the polyhedron P form a subset of C . Thus, the relation $\mathbf{X} \succ_{\rho}^C \mathbf{Y}$ trivially implies (52), which completes the proof of part (i).

To show part (ii), let us consider a spectral risk measure ρ , and recall that it has a representation of the form (48). Therefore, in this case the infimum in (54) can be expressed as the optimum of the following linear program:

$$\begin{aligned} \min \quad & \sum_{i=1}^n p_i \hat{Q}(\omega_i) \mathbf{c}^T \mathbf{x}_i - \sum_{j=1}^M \mu_j \left(\eta^{(j)} - \frac{1}{\alpha_j} \sum_{i=1}^n p_i w_i^{(j)} \right) \\ \text{s.t.} \quad & w_i^{(j)} \geq \eta^{(j)} - \mathbf{c}^T \mathbf{y}_i \quad i = 1, \dots, n, \quad j = 1, \dots, M \\ & w_i^{(j)} \geq 0 \quad i = 1, \dots, n, \quad j = 1, \dots, M \\ & \mathbf{c} \in C. \end{aligned}$$

Using the notation introduced in (51), the feasible set of this problem is the lifted polyhedron $P^{(M)}(C, \mathbf{Y})$. Thus, there exists an optimal solution $(\mathbf{c}^*, \boldsymbol{\eta}^*, \mathbf{w}^*)$ which is a vertex of $P^{(M)}(C, \mathbf{Y})$. By Theorem 6.2 the vector \mathbf{c}^* is a d -vertex of $P^{(d)}(C, \mathbf{Y})$. The rest of the proof is analogous to that of part (i). \square

REMARK 6.1 *To keep our exposition simple we stated Theorem 6.3 for relations of the form $\mathbf{X} \succ_{\rho}^C \mathbf{Y}$, where \mathbf{X} and \mathbf{Y} are random vectors over the same probability space. However, a more general form of the statement can be proved in essentially the same fashion for preference relations of the form*

$$\rho_1(\mathbf{c}^T \mathbf{X}) \geq \rho_2(\mathbf{c}^T \mathbf{Y}) \quad \text{for all } \mathbf{c} \in C,$$

where the risk measures ρ_1 and ρ_2 can be defined on different finite probability spaces.

6.3 Solution methods We now examine how the cut generation algorithm introduced in Section 5 can be adapted to solve (GeneralP $_{\rho}$) for a functionally coherent risk measure ρ . As in Section 5, let us assume that f is continuous, $\mathbf{z} \mapsto G(\mathbf{z})$ is continuous in the \mathcal{L}^1 -norm, C is a non-empty polytope, and Z is compact.

The master problem is again a non-linear program with finitely many constraints, and becomes a convex program under the appropriate assumptions. In addition, when ρ is finitely representable and given in the form (55), under the linearity assumptions established in Section 4 we can formulate the master problem as a linear program by introducing multiple copies of our auxiliary variables, and replacing the first constraint in (23) with

$$\sum_{j=1}^M \mu_j^{(h)} \left(\eta_{\ell}^{(j)} - \frac{1}{\alpha_j} \sum_{i=1}^n p_i w_{i\ell}^{(j)} \right) \geq \rho(\mathbf{c}_{\ell}^T \mathbf{Y}) \quad \ell = 1, \dots, L, \quad h = 1, \dots, H.$$

The cut generation problem takes the form

$$\min_{\mathbf{c} \in C} \rho(\mathbf{c}^T \mathbf{X}) - \rho(\mathbf{c}^T \mathbf{Y}). \quad (57)$$

Due to the superadditive property of the risk measure ρ , the mapping $\mathbf{c} \mapsto \rho(\mathbf{c}^T \mathbf{W})$ is concave for any random vector \mathbf{W} . Therefore, (57) is a *difference of convex* (DC) programming problem for any functionally coherent risk measure ρ , and can be solved using methods available in the literature (see, e.g., An and Tao, 2005). We remark that Homem-de-Mello and Mehrotra (2009) take a similar DC-based approach in the context of multivariate SSD-constrained optimization, while Wozabal et al. (2010) express univariate VaR as the difference of two CVaRs, again leading to a DC formulation. Similarly to the case of the master problem, for a spectral risk measure given in the form (48) we can introduce multiple copies of our auxiliary variables

in the formulations (26)-(35), (37)-(45), or (46) to obtain a cut generation MIP. For finitely representable coherent risk measures given as the minimum of spectral risk measures, we can use a disjunctive approach, and solve a separate cut generation MIP for each spectral measure in the representation.

For finitely representable coherent risk measures Theorem 6.3 implies the finite convergence of the cut generation algorithm analogously to Theorem 5.1, assuming that a d -vertex solution of (57) can always be found. When ρ is spectral, given an optimal scalarization vector \mathbf{c}^* produced by the cut generation problem we can find a d -vertex solution as follows (analogously to Section 5.5). We first define indices k_j^* and corresponding index sets K_j^* as in (47), for each confidence level α_j featured in the representation (48). Then with the usual replication of the auxiliary variables, we expand (FixedSet) to obtain a linear program that finds a d -vertex solution. A similar approach can be taken for finitely representable coherent risk measures.

While the finite convergence of the algorithm is not guaranteed for the entire class of functionally coherent risk measures, according to Proposition 6.2 every such risk measure can be approximated arbitrarily closely by a finitely representable one. We note that Proposition 6.2 is applicable here because the set $\bar{\mathcal{V}} = \{\mathbf{c}^T G(\mathbf{z}) : \mathbf{c} \in C, \mathbf{z} \in Z\} \cup \{\mathbf{c}^T \mathbf{Y} : \mathbf{c} \in C\}$ of scalar-valued random variables that feature in our problems is compact. Therefore, the methods described above can be applied to solve suitably close approximations of the problem (GeneralP $_\rho$).

As the above discussions show, the computational tractability of optimization problems with multivariate coherent risk constraints depends on the available representations of the underlying risk measure ρ . In particular, in the equal probability case spectral risk measures have a representation of at most the same size as the sample space (i.e., at most n), which makes it possible to formulate the cut generation problem as a MIP with $O(n^2)$ binary variables. Since the other subproblems of the cut generation algorithm can be cast as linear programs, this leads to a tractable solution method comparable to what we have found for CVaR-constrained problems.

To conclude this section we mention that when ρ is finitely representable, Theorem 6.3 allows us to replace the risk constraint $G(\mathbf{z}) \succ_\rho^C \mathbf{Y}$ in (GeneralP $_\rho$) by finitely many constraints of the form $\rho(\mathbf{c}_{(\ell)}^T G(\mathbf{z})) \geq \rho(\mathbf{c}_{(\ell)}^T \mathbf{Y})$. Therefore, in simple instances we can attempt to generate the d -vertices $\mathbf{c}_{(1)}, \mathbf{c}_{(2)}, \dots$ of the polyhedron P , and use nonlinear programming techniques to solve our problem; moreover, under linearity assumptions the problem (GeneralP $_\rho$) becomes a linear program. We point out that for spectral risk measures the polyhedron $P = P^{(d)}(C, \mathbf{Y})$ has some favorable properties: it does not depend on the choice of ρ , and has a simple linear description of small size.

7. Computational Study In this section we demonstrate the effectiveness of our CVaR-based methods by presenting two numerical studies. First we examine feasible regions associated with various multivariate risk constraints on an illustrative example. Then we evaluate the effectiveness of our optimization models and solution methods by applying them to a homeland security budget allocation problem.

We used MATLAB® 7.11.0 to generate data and perform supporting calculations, AMPL (Fourer et al., 2003) to formulate models and implement solution methods, and CPLEX 11.2 (ILOG, 2008) to solve optimization problems. All experiments were carried out on a single core of an HP Linux workstation with two Intel® Xeon® W5580 3.20 GHz CPUs and 32 GB of memory.

7.1 A small-scale study of feasibility regions We now present a simple two-dimensional problem to illustrate feasible regions associated with multivariate CVaR constraints, along with the effects of various parameter choices. The problem originally appeared in Hu et al. (2011a), where the authors compare the feasible regions associated with various multivariate SSD constraints: *positive linear dominance*, *weak stochastically weighted dominance*, *stochastically weighted dominance with chance*, and *relaxed strong stochastically weighted*

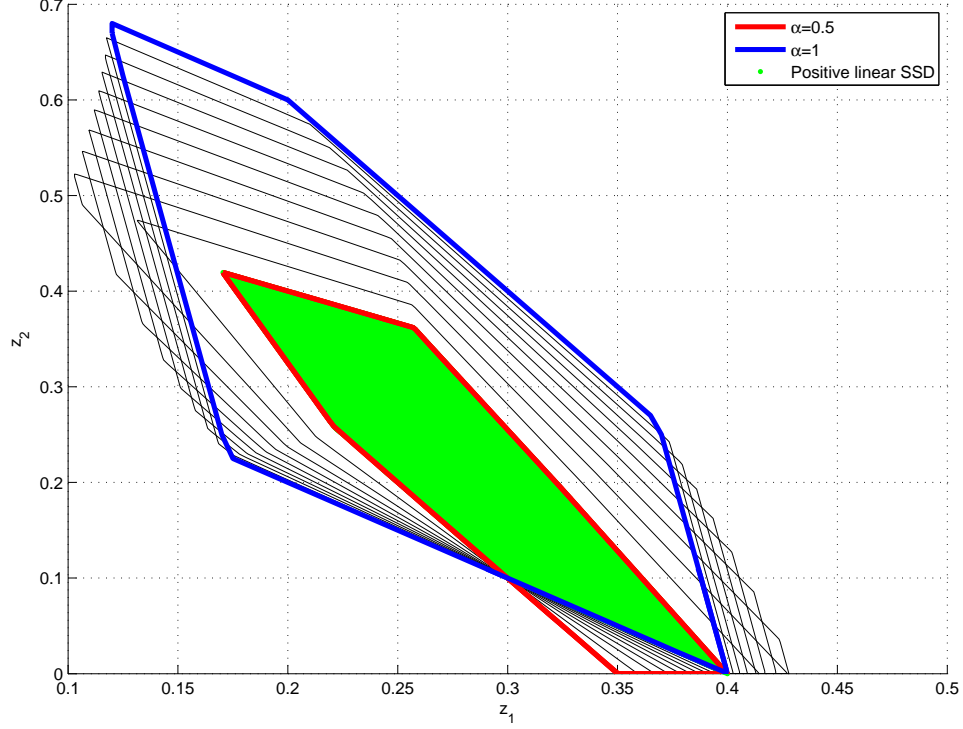


Figure 2: Feasible regions for $\alpha^{(1)} = \alpha^{(2)} \in [0.5, 1]$ and $\vartheta = 0$

dominance. We chose to explore the same numerical example, as this allows a direct comparison between our CVaR constraints and the dominance concepts mentioned above.

Consider the probability space $(\Omega, 2^\Omega, \Pi)$ where $\Omega = \{\omega_1, \omega_2\}$ and $\Pi(\omega_1) = \Pi(\omega_2) = \frac{1}{2}$. Let $\Delta : \Omega \rightarrow \mathbb{R}$ denote the random variable with realizations $\Delta(\omega_1) = 1$, $\Delta(\omega_2) = -1$, and let

$$\Gamma = \begin{bmatrix} 1 + 0.25\Delta & 0.5 \\ 0.5 & 0.5 - 0.25\Delta \\ 0.25 & 0.03 \end{bmatrix}, \quad \mathbf{Y}^{(1)} = \begin{bmatrix} 0.5 - 0.0025\Delta \\ 0.4 \\ 0.1 + 0.013\Delta \end{bmatrix}, \quad \mathbf{Y}^{(2)} = \begin{bmatrix} 0.05 \\ 0.2 - 0.025\Delta \\ 0.01 + 0.013\Delta \end{bmatrix}.$$

In addition, we define the scalarization polyhedra

$$C_\vartheta = \{(c_1, c_2, c_3) \in \mathbb{R}^3 : c_1 + c_2 + c_3 = 1, c_1 \geq \vartheta, c_2 \geq \vartheta, c_3 \geq \vartheta\}, \quad \vartheta \in \left[0, \frac{1}{3}\right].$$

Note that C_0 is the simplex used to define positive linear dominance, while $C_{\frac{1}{3}}$ consists of the single scalarization vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. We are interested in feasibility regions defined by constraints of the form

$$\begin{aligned} -\Gamma \mathbf{z} &\succ_{C_{\vartheta}^{\text{CVaR}_{\alpha^{(1)}}}} -\mathbf{Y}^{(1)} \\ \Gamma \mathbf{z} &\succ_{C_{\vartheta}^{\text{CVaR}_{\alpha^{(2)}}}} \mathbf{Y}^{(2)} \\ \mathbf{z} &\in \mathbb{R}_+^2, \end{aligned}$$

where $\mathbf{z} = (z_1, z_2)$ is a decision vector. Figure 2 shows the feasible regions associated with the scalarization polyhedron C_0 and confidence levels $\alpha^{(1)} = \alpha^{(2)}$ changing between 0.5 and 1. Note that these regions are not nested, i.e., CVaR-preferability at a certain confidence level does not imply preferability at other levels. In accordance with part (iii) of Proposition 2.1, the intersection of these regions (filled area) corresponds to the region associated with the *positive linear SSD constraint* (compare with Figure 2(a) in Hu et al., 2011b). Figure 3 illustrates shapes of feasible regions obtained by various combinations of $\alpha^{(1)}$ and $\alpha^{(2)}$, for a range of

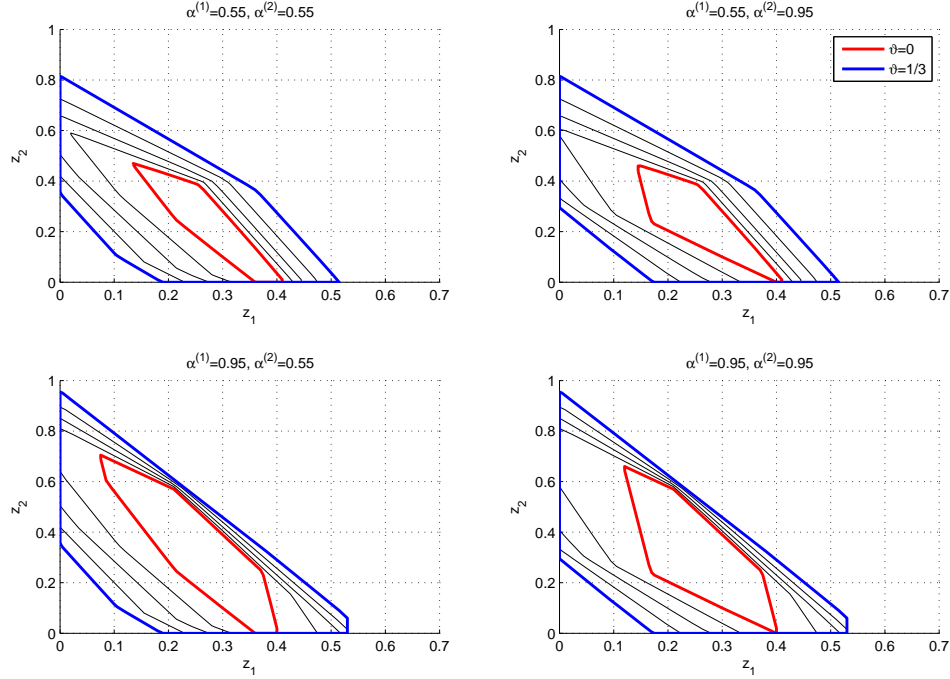


Figure 3: Feasible regions for various choices of $\alpha^{(1)}$ and $\alpha^{(2)}$ with $\vartheta \in [0, \frac{1}{3}]$

ϑ values between 0 and $\frac{1}{3}$. Note that $\vartheta_1 \leq \vartheta_2$ implies $C_{\vartheta_1} \supset C_{\vartheta_2}$, therefore CVaR-preferability with respect to C_{ϑ_1} implies preferability with respect to C_{ϑ_2} . This results in a nested structure between the corresponding feasible regions.

Further customization of the feasible region can be achieved by requiring CVaR constraints to hold at multiple different confidence levels, and with respect to different corresponding scalarization polyhedra, for each reference variable.

7.2 Homeland security budget allocation To explore the computational performance of our methods, along with the impact of various polyhedral CVaR constraints, we examine a budget allocation problem. This problem was presented in [Hu et al. \(2011b\)](#) with polyhedral SSD constraints in a homeland security context, and also inspired the numerical study in [Armbruster and Luedtke \(2010\)](#). Our exposition below closely follows that in [Hu et al. \(2011b\)](#), replacing the SSD constraints with CVaR-based ones. The model concerns the allocation of a fixed budget to ten urban areas (New York, Chicago, etc.). The budget is used for prevention, response, and recovery from national catastrophes. The risk share of each area is defined based four criteria: property losses, fatalities, air departures, and average daily bridge traffic. Accordingly, we consider a random *risk share matrix* $A : \Omega \rightarrow \mathbb{R}_+^{4 \times 10}$, where the entry $A_{ij} : \Omega \rightarrow \mathbb{R}$ denotes, for criterion i , the proportion of losses in urban area j relative to the total losses. The penalty for allocations under the risk share is expressed by the *budget misallocation functions* $\mathcal{M}_i : Z \rightarrow \mathcal{V}(\Omega, 2^\Omega, \Pi)$ defined as

$$\mathcal{M}_i(\mathbf{z}) = \sum_{j=1}^{10} [A_{ij} - z_j]_+ \quad \text{for each criterion } i = 1, \dots, 4,$$

where $Z = \{\mathbf{z} \in \mathbb{R}_+^{10} : \|\mathbf{z}\|_1 = 1\}$ denotes the set of all feasible allocations. Let us also introduce the notation $\mathbf{M} = (M_1, M_2, M_3, M_4)^T$.

We consider two benchmark solutions: one based on average government allocations by the Department of Homeland Security's Urban Areas Security Initiative, and one based on suggestions in the RAND report by [Willis et al. \(2005\)](#). These benchmark allocations are denoted by \mathbf{z}^G and \mathbf{z}^R , respectively. The scalarization

polyhedron is of the form $C = \{\mathbf{c} \in \mathbb{R}^4 : \|\mathbf{c}\|_1 = 1, c_i \geq c_i^* - \frac{\theta}{3}\}$, where $\mathbf{c}^* \in \mathbb{R}^4$ is a *center* satisfying $\|\mathbf{c}^*\|_1 = 1$, and $\theta \in [0, 1]$ is a constant for which $\frac{\theta}{3} \leq \min_{i \in \{1, \dots, 4\}} c_i^*$ holds. It is easy to see that if θ is positive, the polyhedron C is a 3-dimensional simplex. Denoting the vertices of C by $\hat{\mathbf{c}}_{(1)}, \dots, \hat{\mathbf{c}}_{(4)}$ the objective function of the budget allocation problem, based on a robust approach, is defined as

$$f(\mathbf{z}) = \max_{k \in \{1, \dots, 4\}} \mathbb{E} \left(\hat{\mathbf{c}}_{(k)}^T \mathbf{M}(\mathbf{z}) \right).$$

Selecting two finite sets of confidence levels $\mathcal{A}^G, \mathcal{A}^R \subset (0, 1]$ we introduce an optimization problem with multivariate CVaR constraints:

$$\begin{aligned} \min \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & -\mathbf{M}(\mathbf{z}) \succ_{\text{CVaR}_\alpha}^C -\mathbf{M}(\mathbf{z}^G) & \alpha \in \mathcal{A}^G \\ & -\mathbf{M}(\mathbf{z}) \succ_{\text{CVaR}_\alpha}^C -\mathbf{M}(\mathbf{z}^R) & \alpha \in \mathcal{A}^R \\ & \mathbf{z} \in Z. \end{aligned} \tag{58}$$

Note that the negative signs were added in order to be consistent with our convention of preferring large values. To keep the exposition concise, we refer the reader to [Hu et al. \(2011b, Section 4\)](#) for a description of how the objective function f can be linearized, along with the explicit construction of the benchmarks \mathbf{z}^G , \mathbf{z}^R and the realizations of the risk share matrix A . Unless otherwise specified, we consider the “base case” with the choices of $\theta = 0.25$, the *equality center* $\mathbf{c}^* = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, and additional correlation and volatility parameters $\pi = \frac{1}{2}$, $\gamma = 3$.

7.2.1 Computational performance We use the cut generation method outlined in Section 5.1 to solve problem (58) in the case when each scenario in $\Omega = \{\omega_1, \dots, \omega_n\}$ has probability $\frac{1}{n}$, and all confidence levels in \mathcal{A}^G and \mathcal{A}^R are chosen from the set $\{\frac{1}{n}, \dots, \frac{n}{n}\}$. When necessary, the confidence levels in our tables were rounded up to a multiple of $\frac{1}{n}$ during computation. Note that we have a separate cut generation problem for each pair of a benchmark and an associated confidence level. Under our assumptions all of these cut generation problems take the form of the MIP (37)-(45). All numerical results in Sections 7.2.1-7.2.2 were obtained using batch sampling, averaging over 5 samples.

Table 1 shows the computational performance of our implementation when solving problem (58) with a single CVaR constraint based on the RAND benchmark ($\mathcal{A}^G = \emptyset$, $\mathcal{A}^R = \{\alpha\}$). We report the total number of cuts, including those introduced in the initialization step (associated with the four vertices of the scalarization polyhedron). Additional cuts are generated in each iteration except the final one, at which the algorithm terminates by proving optimality. While the master problem can be solved nearly instantaneously, solving the cut generation MIP can become a computational bottleneck as the number of scenarios increases. It is interesting to note that CPU times are typically higher for $\alpha = 0.05$ than for $\alpha = 0.01$ when solving otherwise identical problems. The reason lies in the increased combinatorial complexity of the cut generation MIP, which involves selecting $\lceil \alpha n \rceil$ binary variables (out of a total of n) to take value 1. This point is further illustrated by Figure 4(a), which shows that CPU times are significantly lower for α values near the endpoints of the interval $[0, 1]$ despite generating a similar number of cuts. By contrast, for a fixed value of α , considering larger scalarization sets by increasing θ results in a higher number of cuts and a proportional increase of CPU time; see Figure 4(b).

7.2.2 Numerical study on the effect of risk constraints We now look at optimal solutions of problem (58) and its SSD-constrained counterpart, along with an “unconstrained” variant of the problem which features no risk constraints. To keep our presentation simple, for the purposes of discussing allocation results we have divided the set of urban areas into three groups:

- New York (highest risk);
- Chicago, Bay Area, Washington DC-MD-VA-WV, and Los Angeles-Long Beach (medium risk);

α	n	Number of Cuts			Number of		CPU Time (sec)		
		Total	Initial	CutGen	Iterations	MIP solved	CutGen	Total	CutGen/Total
0.01	50	5.2	4	1.2	2.2	2.2	0.14	0.34	40.356%
	100	5.2	4	1.2	2.2	2.2	0.36	1.10	33.809%
	150	5	4	1	2	2	23.60	25.15	93.889%
	200	5.2	4	1.2	2.2	2.2	56.48	58.82	96.054%
	250	5	4	1	2	2	6312.35	6315.66	99.948%
	500	4.6	4	0.6	1.6	1.6	11507.10	11528.40	99.792%
0.05	50	5	4	1	2	2	12.19	12.41	98.213%
	100	4.8	4	0.8	1.8	1.8	5244.96	5245.48	99.990%
	150	4	4	0	1	1	3921.22	3922.16	99.976%
	200	4	4	0	1	1	5004.41	5006.03	99.968%
	250	4	4	0	1	1	6021.99	6024.31	99.962%
	500	5	4	1	2	2	14386.69	14413.13	99.817%

Table 1: Computational performance of the cut generation algorithm for a single CVaR constraint

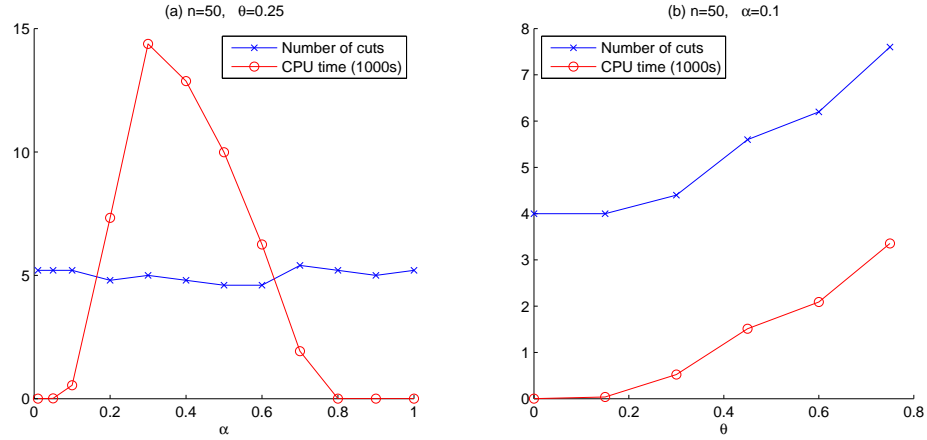
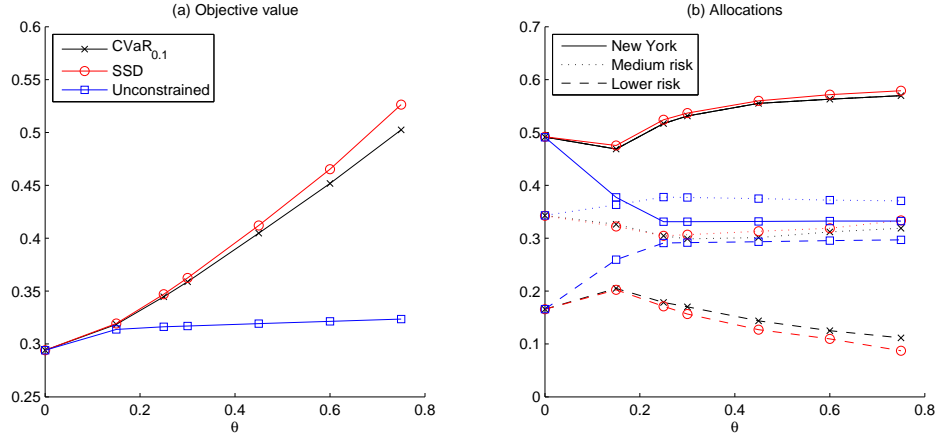


Figure 4: Computational performance of cut generation algorithm for a single benchmark

- Philadelphia PA-NJ, Boston MA-NH, Houston, Newark, and Seattle-Bellevue-Everett (lower risk).

Figure 5 shows optimal results for problem (58) with CVaR preferability required over the benchmark \mathbf{z}^R at a single confidence level of 0.1, along with solutions of SSD-constrained and unconstrained versions of the problem. As the parameter θ increases, the scalarization set becomes larger, leading to more restrictive constraints. Accordingly, as illustrated in Figure 5(a), optimal objective values of the CVaR- and SSD-constrained problems diverge sharply from that of the unconstrained version. We observe that while the budget allocated to urban areas with medium risk remains relatively unchanged in all three models, under CVaR and SSD constraints there is a significant tradeoff between allocations to New York and areas with lower risk. It is interesting to note that enforcing the CVaR constraint at a single confidence level yields results very close to those obtained under SSD constraints, although the difference between the two models becomes more pronounced for larger values of θ .

We next present results for problem (58) with CVaR constraints on both benchmarks \mathbf{z}^G and \mathbf{z}^R , enforced at multiple common confidence levels ($\mathcal{A}^G = \mathcal{A}^R = \mathcal{A}$). While problems requiring (weak) preference over a single benchmark solution are always feasible, this is not necessarily the case when considering multiple benchmarks. A natural approach to overcome this issue is to relax risk constraints by introducing a tolerance parameter ι , as described in part (i) of Proposition 2.1. In accordance with Hu et al. (2011b), we set $\iota = 0.005$. We

Figure 5: Optimal results for benchmark \mathbf{z}^R and $n = 100$

remark that smaller values of ι typically result in infeasible SSD-constrained problems, and at some confidence level settings we encounter infeasibility in certain problem instances even under CVaR constraints. Table 2 contains our results for the relaxed two-benchmark problem. We can see that enforcing CVaR constraints at low confidence levels yields solutions close to the unconstrained allocations, while requiring them to hold at both ends of the spectrum results in convergence to the SSD-constrained solution. Although the latter fact is not surprising given the equivalence established in Proposition 2.1, it is interesting to note that simply requiring CVaR to hold at the lowest and highest levels (corresponding to worst case- and expectation-based constraints) already leads to a close approximation of the SSD constraint (this observation is consistent with the findings of Fábian et al., 2011). In line with the conclusions reached by Hu et al. (2011b) we finally observe that the budget allocated to New York, the area with the highest risk, gradually increases with the introduction of additional risk constraints (from 32.9% in the unconstrained case to a maximum of 49.3% under SSD). The reason for this behavior is that New York has a large (58.6%) allocation in the RAND benchmark and a high volatility in the corresponding risk share. As a consequence, solutions with allocations to New York that are significantly smaller than in the RAND benchmark cannot produce a stochastically preferable outcome in the presence of more demanding risk constraints.

8. Conclusion and future research We have introduced new multivariate risk-averse preference relations based on CVaR and linear scalarization, referred to as *polyhedral CVaR constraints*. We have demonstrated that they provide an efficient and computationally tractable way of relaxing multivariate stochastic dominance constraints. Additionally, we have illustrated that the flexibility of our approach allows for modeling a wide range of risk preferences. In particular, unlike existing SSD-based relations, the ability to specify confidence levels allows us to focus on various aspects of the distribution (including the tails, expectation, and worst case behavior) separately or in arbitrary combinations. We have shown that our framework can be extended from CVaR to a wider class of coherent risk measures, including spectral risk measures.

We have incorporated polyhedral CVaR constraints into optimization problems, providing a novel way of modeling risk preferences in stochastic multi-criteria decision making. We have developed a finitely convergent cut generation algorithm to solve such problems on finite probability spaces. Under certain linearity assumptions we have formulated the master problem as a linear program, and the cut generation problem as a MIP, solvable by off-the-shelf software such as CPLEX. We have applied our solution methods to a budget allocation problem featuring CVaR constraints at multiple confidence levels for two benchmark solutions, and compared our results to those obtained by an SSD-based model. As this computational study shows, our approach can naturally be used in a framework based on sample average approximation (SAA). Performing a detailed study of convergence behavior and developing related statistical bounds for the SAA method (as it has successfully

Confidence levels		Objective	New York	Medium risk	Lower risk
Unconstrained model	$(\mathcal{A} = \emptyset)$	0.316245	0.331189	0.377809	0.291002
0.01		0.316245	0.331189	0.377809	0.291002
0.01,0.05		0.316245	0.331189	0.377809	0.291002
0.01,0.05,0.1		0.321030	0.388341	0.355589	0.256069
0.01,0.05,0.1,0.2		0.329228	0.447436	0.332685	0.219879
0.01,0.05,0.1,0.2,0.3		0.332944	0.467037	0.326971	0.205992
0.01,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,0.2,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,0.2,0.3,	1	0.338968	0.490263	0.327498	0.182239
0.01,0.05,0.1,0.2,0.3,	0.9, 1	0.339043	0.490795	0.326833	0.182372
0.01,0.05,0.1,0.2,0.3,	0.8, 0.9, 1	0.339101	0.490891	0.327012	0.182097
0.01,0.05,0.1,0.2,0.3,	0.7, 0.8, 0.9, 1	0.339101	0.490891	0.327012	0.182097
SSD-constrained model	$(\mathcal{A} = (0, 1])$	0.339167	0.491158	0.327025	0.181817
Government benchmark		0.345505	0.319300	0.454300	0.226500
RAND benchmark		0.393543	0.586100	0.343100	0.070700

Table 2: Optimal objective and allocations for two benchmarks, $n = 100$, $\theta = 0.25$ and $\iota = 0.005$

been done for multivariate SSD-constrained problems) forms part of our future research plans.

While problem instances featuring up to 500 scenarios were found to be tractable, solving our MIP formulations increasingly became a computational bottleneck. Developing valid inequalities and heuristics which lead to more efficient solution of these MIPs is the topic of future research. In addition, utilizing CVaR-based Kusuoka representations, such advances could also be crucial in the development of efficient solution methods for large-scale problems with multivariate coherent risk constraints.

Acknowledgment. The first author has been supported by The Scientific and Technological Research Council of Turkey (TUBITAK) Career Award # 111M543 and the second author has been funded by TUBITAK-2216 Research Fellowship Programme. We would also like to thank the anonymous referees for their valuable comments and suggestions.

References

- Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518.
- Acerbi, C. (2004). Coherent representations of subjective risk aversion. In Szegö, G., editor, *Risk Measures for the 21st Century*, pages 147–207. Wiley, New York.
- An, L. T. H. and Tao, P. D. (2005). The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems. *Annals of Operations Research*, 133(1):23–46.
- Armbruster, B. and Luedtke, J. (2010). Models and formulations for multivariate dominance constrained stochastic programs. http://www.optimization-online.org/DB_FILE/2010/05/2613.pdf.
- Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Bonnans, J. F. and Shapiro, A. (2000). *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York.
- Dentcheva, D. and Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566.
- Dentcheva, D. and Ruszczyński, A. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of*

- Banking and Finance*, 30(2):433–451.
- Dentcheva, D. and Ruszczyński, A. (2009). Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117:111–127.
- Ehrgott, M. (2005). *Multicriteria Optimization*. Springer, Berlin.
- Fábián, C. I., Mitra, G., Roman, D., and Zverovich, V. (2011). An enhanced model for portfolio choice with SSD criteria: a constructive approach. *Quantitative Finance*, 11(10):1525–1534.
- Fábián, C. I. and Veszprémi, A. (2008). Algorithms for handling CVaR-constraints in dynamic stochastic programming models with applications to finance. *The Journal of Risk*, 10:111–131.
- Fourer, R., Gay, D., and Kernighan, B. (2003). *AMPL: A Modeling Language for Mathematical Programming Second Edition*. Duxbury Press Brooks Cole Publishing Co.
- Goberna, M. and Jornet, V. (1996). On Haar’s dual problem. *Operations-Research-Spektrum*, 18:209–217.
- Gollmer, R., Gotzes, U., and Schultz, R. (2011). A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming*, 216(1):179–190.
- Gollmer, R., Neise, F., and Schultz, R. (2008). Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse. *SIAM Journal on Optimization*, 19(2):552–571.
- Homem-de-Mello, T. and Mehrotra, S. (2009). A cutting surface method for uncertain linear programs with linear stochastic dominance constraints. *SIAM Journal on Optimization*, 20(3):1250–1273.
- Hu, J., Homem-de-Mello, T., and Mehrotra, S. (2010). Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming, Online First (October 2010)*. [dx.doi.org/10.1007/s10107-010-0428-9](https://doi.org/10.1007/s10107-010-0428-9).
- Hu, J., Homem-de-Mello, T., and Mehrotra, S. (2011a). Concepts and applications of stochastically weighted stochastic dominance. http://www.optimization-online.org/DB_FILE/2011/04/2981.pdf.
- Hu, J., Homem-de-Mello, T., and Mehrotra, S. (2011b). Risk-adjusted budget allocation models with application in homeland security. *IIE Transactions*, 43(12):819–839.
- ILOG (2008). *ILOG AMPL CPLEX System Version 11 User’s Guide*. ILOG CPLEX Division.
- Khachiyan, L., Boros, E., Borys, K., Elbassioni, K. M., and Gurvich, V. (2008). Generating all vertices of a polyhedron is hard. *Discrete & Computational Geometry*, 39(1-3):174–190.
- Künzi-Bay, A. and Mayer, J. (2006). Computational aspects of minimizing conditional value-at-risk. *Computational Management Science*, 3(1):3–27.
- Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95.
- Lehmann, E. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, 26(3):399–419.
- Leshno, M. and Levy, H. (2002). Preferred by “all” and preferred by “most” decision makers: Almost stochastic dominance. *Management Science*, 48(8):1074–1085.
- Lizyayev, A. and Ruszczyński, A. (2011). Tractable almost stochastic dominance. *European Journal of Operational Research*, 218(2):448–455.
- Luedtke, J. (2008). New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450.
- Mann, H. and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60.
- Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester.
- Nie, Y., Wu, X., and Homem-de-Mello, T. (2011). Optimal path problems with second-order stochastic dominance constraints. *Networks and Spatial Economics, Online First (October 2011)*. [dx.doi.org/10.1007/s11067-011-9167-6](https://doi.org/10.1007/s11067-011-9167-6).
- Noyan, N. (2010). Alternate risk measures for emergency medical service system design. *Annals of Operations Research*, 181(1):559–589.
- Noyan, N. and Rudolf, G. (2012). Kusuoka representations of coherent risk measures in finite probability spaces. Technical report, RUTCOR-Rutgers Center for Operations Research, RRR 33-2012. http://rutcor.rutgers.edu/pub/rrr/reports2012/33_2012.pdf.
- Noyan, N., Rudolf, G., and Ruszczyński, A. (2008). Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters*, 34(6):653–659.
- Noyan, N. and Ruszczyński, A. (2008). Valid inequalities and restrictions for stochastic programming problems with first order stochastic dominance constraints. *Mathematical Programming, Ser. A*, 114(2):249–275.
- Ogryczak, W. and Ruszczyński, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM journal of*

- optimization, 13(2):60–78.
- Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In Uryasev, S., editor, *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, Dordrecht.
- Pflug, G. C. and Römisch, W. (2007). *Modelling, managing and measuring risk*. World Scientific publishing, Singapore.
- Prékopa, A. (2012). Multivariate value at risk and related topics. *Annals of Operations Research*, 193:49–69.
- Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press, Princeton.
- Rockafellar, R. (2007). Coherent approaches to risk in optimization under uncertainty. In *Tutorials in Operations Research*, pages 38–61. INFORMS.
- Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41.
- Rockafellar, R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471.
- Rudolf, G. and Ruszczyński, A. (2008). Optimization problems with second order stochastic dominance constraints: duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19:1326–1343.
- Ruszczynski, A. and Shapiro, A. (2006). Optimization of risk measures. In Calafiore, G. and Dabbene, F., editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 119–157. Springer London.
- Saaty, T. L. (1980). *The analytic hierarchy process*. McGraw-Hill, New York.
- Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and their Applications*. Associated Press, Boston.
- Shapiro, A. (2005). On duality theory of convex semi-infinite programming. *Optimization*, 54(6):535–543.
- Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation, and Application*. John Wiley & Sons, New York.
- Street, A. (2009). On the conditional value-at-risk probability-dependent utility function. *Theory and Decision*, 68(1–2):49–68.
- Uryasev, S. (2000). Conditional value-at-risk: Optimization algorithms and applications. *Financial Engineering News*.
- Willis, H. H., Morral, A. R., Kelly, T. K., and Medby, J. J. (2005). Estimating terrorism risk. Technical report, The RAND Corporation, Santa Monica, CA.
- Wozabal, D., Hochreiter, R., and Pflug, G. C. (2010). A difference of convex formulation of value-at-risk constrained optimization. *Optimization*, 59(3):377–400.

Appendix A. Omitted proofs

A.1 Proof of Proposition 2.1 The following proof of part (i) is a straightforward extension of the arguments in (Dentcheva and Ruszczyński, 2006), and it utilizes some basic concepts and results from the theory of conjugate duality (for a good overview see Rockafellar, 1970). Denoting the *extended real line* by $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, the *Fenchel conjugate* of a function $f : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ is the mapping $f^* : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ defined by $f^*(\alpha) = \sup\{\alpha\eta - f(\eta) : \eta \in \mathbb{R}\}$. It is easy to see that for any constant ι the conjugate of $f + \iota$ is given by $f^* - \iota$.

The second order distribution function $F_V^{(2)}$ of a random variable V is the integral of a monotone non-decreasing function, therefore it is continuous and convex. By the Fenchel-Moreau theorem it follows that both of the functions $F_{V_1}^{(2)}$ and $F_{V_2}^{(2)} + \iota$ are equal to their respective biconjugates. This implies, due to the order reversing property of conjugation, that the condition (8) is equivalent to

$$[F_{V_1}^{(2)}]^*(\alpha) \geq [F_{V_2}^{(2)}]^*(\alpha) - \iota \quad \text{for all } \alpha \in \mathbb{R}. \quad (59)$$

According to (5) we have $F_V^{(2)}(\eta) = \mathbb{E}([\eta - V]_+)$. Taking into account (2) it is easy to verify that

$$[F_V^{(2)}]^*(\alpha) = \begin{cases} \infty & \alpha < 0 \\ 0 & \alpha = 0 \\ \alpha \text{ CVaR}_\alpha(V) & \alpha \in (0, 1] \\ \infty & \alpha > 1 \end{cases} \quad (60)$$

holds for any random variable V .³ If we substitute into (59), part (i) immediately follows.

³The function $[F_V^{(2)}]^*$ is equal to the so-called *second quantile function* of V , also known in the literature as the *generalized Lorenz curve*, and the *absolute Lorenz curve* (see, e.g., Ogryczak and Ruszczyński, 2002).

To prove the non-trivial implication in part (ii), it is sufficient to show that the condition (10) implies (9). Accordingly, let us assume that the relation $\text{CVaR}_\alpha(V_1) \geq \text{CVaR}_\alpha(V_2) - \frac{\iota}{\alpha}$ holds for all $\alpha \in \mathcal{K}$, and consider an arbitrary confidence level $\hat{\alpha} \in (0, 1]$. Since the level sets of random variables V_1 and V_2 are measurable, the values

$$\alpha_- = \max_{i \in \{1,2\}} \Pi(V_i < \text{VaR}_{\hat{\alpha}}(V_i)) \quad \text{and} \quad \alpha_+ = \min_{i \in \{1,2\}} \Pi(V_i \leq \text{VaR}_{\hat{\alpha}}(V_i))$$

both belong to the set $\mathcal{K} \cup \{0\}$. In the case $\alpha_- > 0$ by our assumption we have

$$\alpha_- \text{CVaR}_{\alpha_-}(V_1) \geq \alpha_- \text{CVaR}_{\alpha_-}(V_2) - \iota \quad \text{and} \quad \alpha_+ \text{CVaR}_{\alpha_+}(V_1) \geq \alpha_+ \text{CVaR}_{\alpha_+}(V_2) - \iota. \quad (61)$$

Furthermore, by the definition of VaR the inequalities $\alpha_- \leq \hat{\alpha} \leq \alpha_+$ hold, and for any $\gamma \in (\alpha_-, \alpha_+]$, $i \in \{1, 2\}$ we have $\text{VaR}_\gamma(V_i) = \text{VaR}_{\hat{\alpha}}(V_i)$. It follows that, according to the formulas (4) and (60), the functions $[F_{V_1}^{(2)}]^*$ and $[F_{V_2}^{(2)}]^* - \iota$ are both affine on the interval $[\alpha_-, \alpha_+]$, with respective slopes $\text{VaR}_{\hat{\alpha}}(V_1)$ and $\text{VaR}_{\hat{\alpha}}(V_2)$. Since (61) states that the inequality $[F_{V_1}^{(2)}]^*(\alpha) \geq [F_{V_2}^{(2)}]^*(\alpha) - \iota$ holds at the endpoints of the interval $[\alpha_-, \alpha_+]$, it must also hold at the intermediate point $\hat{\alpha}$ due to the affine property, which proves our claim. To see that the claim also holds when we have $\alpha_- = 0$, note that in this case $[F_{V_1}^{(2)}]^*$ and $[F_{V_2}^{(2)}]^* - \iota$ are linear on the interval $[0, \alpha_+]$, therefore the inequality at the upper endpoint α_+ implies the inequality at the intermediate point $\hat{\alpha}$.

Finally, to prove part (iii) we simply observe that for a finite probability space with $|\Omega| = n$ we have $|\mathcal{K}| < 2^n$, and in the equal probability case we have $\mathcal{K} = \{\frac{1}{n}, \dots, \frac{n}{n}\}$. \square

A.2 Proofs for Section 6.2.1 Let us consider a polyhedron $P \subset \mathbb{R}^n$. We say that a vector $\mathbf{d} \in \mathbb{R}^n$ is a P -direction of a point $\mathbf{p} \in P$ if there exists $\epsilon > 0$ such that both $\mathbf{p} + \epsilon \mathbf{d}$ and $\mathbf{p} - \epsilon \mathbf{d}$ belong to P . It is easy to see that the P -directions of a point $\mathbf{p} \in P$ always constitute a linear space. We can use P -directions to characterize the vertices of the polyhedron:

$$\mathbf{p} \text{ is a vertex of } P \iff \mathbf{p} \text{ has no non-zero } P\text{-directions.} \quad (62)$$

For the lifted polyhedra introduced in (51) we can easily characterize $P^{(k)}$ -directions in terms of P -directions:

OBSERVATION A.1 A vector $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k)})$ is a $P^{(k)}$ -direction of a point $(\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}) \in P^{(k)}$ if and only if $(\mathbf{d}^{(0)}, \mathbf{d}^{(i)})$ is a P -direction of $(\mathbf{x}, \mathbf{y}^{(i)})$ for all $i = 1, \dots, k$.

PROOF OF EXAMPLE 6.1. The fact that $(0, 0)$ is not a 2-vertex of P can be verified by simply looking at the list of the vertices of P . We now show that $(0, 0, -1, 1)$ is a vertex of $P^{(2)}$, which proves our claim. Assume that $(d_1^{(0)}, d_2^{(0)}, d_1^{(1)}, d_2^{(2)})$ is a $P^{(2)}$ -direction of $(0, 0, -1, 1)$. Then, by Observation A.1, the vector $(d_1^{(0)}, d_2^{(0)}, d_1^{(1)})$ is a P -direction of the point $(0, 0, -1)$. Since this point lies in the relative interior of the edge $[(-1, 0, -1), (1, 0, -1)] = \{(\lambda, 0, -1) : \lambda \in [-1, 1]\}$ of P , it is easy to see that $d_2^{(0)} = d_1^{(1)} = 0$. Analogously, $(d_1^{(0)}, d_2^{(0)}, d_2^{(2)})$ is a P -direction of the point $(0, 0, 1)$, which lies in the relative interior of the edge $[(0, -1, 1), (0, 1, 1)]$, implying $d_1^{(0)} = d_2^{(2)} = 0$. Therefore $(0, 0, -1, 1)$ has no non-zero $P^{(2)}$ -directions, so according to (62) it is a vertex. \square

PROOF OF THEOREM 6.2. We prove our theorem by showing that the following two statements hold:

- (i) For an integer $k < n$ any n -vertex of $P^{(k)}$ is also an n -vertex of $P^{(k+1)}$.
- (ii) For an integer $k > n$ any n -vertex of $P^{(k)}$ is also an n -vertex of $P^{(n)}$.

Let us first assume $k < n$, and let $\mathbf{v}^{(k)} = (\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)})$ be a vertex of $P^{(k)}$. We prove (i) by showing that $\mathbf{v}^{(k+1)} = (\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}, \mathbf{y}^{(k)})$ is a vertex of $P^{(k+1)}$. Indeed, if $\mathbf{d} = (\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k+1)})$ is a $P^{(k+1)}$ -direction of $\mathbf{v}^{(k+1)}$, then by Observation A.1 both $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k-1)}, \mathbf{d}^{(k)})$ and $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k-1)}, \mathbf{d}^{(k+1)})$ are $P^{(k)}$ -directions of $\mathbf{v}^{(k)}$. According to (62), the vertex $\mathbf{v}^{(k)}$ has no non-zero $P^{(k)}$ -directions. Therefore every component of \mathbf{d} is zero, which implies that $\mathbf{v}^{(k+1)}$ is a vertex.

Now assume $k > n$, and again let $\mathbf{v}^{(k)} = (\mathbf{x}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)})$ be a vertex of $P^{(k)}$. The P -directions of the vector $(\mathbf{x}, \mathbf{y}^{(i)})$ form a linear subspace in $\mathbb{R}^n \times \mathbb{R}^m$; we denote the projection of this subspace to its first n coordinates by

$$L_i = \left\{ \mathbf{d}^{(0)} \in \mathbb{R}^n \mid \exists \mathbf{d}^{(i)} \in \mathbb{R}^n : (\mathbf{d}^{(0)}, \mathbf{d}^{(i)}) \text{ is a } P\text{-direction of } (\mathbf{x}, \mathbf{y}^{(i)}) \right\}, \quad i = 1, \dots, k.$$

We next show that $\bigcap_{i=1}^k L_i = \{\mathbf{0}\}$. Indeed, for any vector $\mathbf{d}^{(0)} \in \bigcap_{i=1}^k L_i$ there exist $\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k)}$ such that $(\mathbf{d}^{(0)}, \mathbf{d}^{(i)})$ is a P -direction of $(\mathbf{x}, \mathbf{y}^{(i)})$ for all i . As $\mathbf{v}^{(k)}$ is a vertex and $(\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \dots, \mathbf{d}^{(k)})$ is a $P^{(k)}$ -direction of $\mathbf{v}^{(k)}$, it follows that $\mathbf{d}^{(0)} = \mathbf{0}$.

Since the family $\{L_1, \dots, L_k\}$ consists of linear subspaces in \mathbb{R}^n , it has a subfamily of size n that intersects only in $\mathbf{0}$, i.e., there exist $\{L_{i_1}, \dots, L_{i_n}\} \subset \{L_1, \dots, L_k\}$ such that $\bigcap_{j=1}^n L_{i_j} = \{\mathbf{0}\}$ holds. To prove our claim, we show that $\mathbf{v}^{(n)} = (\mathbf{x}, \mathbf{y}^{(i_1)}, \dots, \mathbf{y}^{(i_n)})$ is a vertex of $P^{(n)}$. Assume that $(\mathbf{d}^{(0)}, \mathbf{d}^{(i_1)}, \dots, \mathbf{d}^{(i_n)})$ is a $P^{(n)}$ -direction of $\mathbf{v}^{(n)}$. Then, as $(\mathbf{d}^{(0)}, \mathbf{d}^{(i_j)})$ is a P -direction of $(\mathbf{x}, \mathbf{y}^{(i_j)})$ for every j , we have $\mathbf{d}^{(0)} \in \bigcap_{j=1}^n L_{i_j}$, implying $\mathbf{d}^{(0)} = \mathbf{0}$. Notice that in this case each vector of the form $(\mathbf{0}, \mathbf{0}, \dots, \mathbf{d}^{(i_j)}, \dots, \mathbf{0}) \in \mathbb{R}^n \times \mathbb{R}^m \times \dots \times \mathbb{R}^m$, with $\mathbf{d}^{(i_j)}$ in the i_j th \mathbb{R}^m -component and $\mathbf{0}$ everywhere else, is a $P^{(k)}$ -direction of the vertex $\mathbf{v}^{(k)}$, thus we have $\mathbf{d}^{(i_j)} = \mathbf{0}$. As $\mathbf{v}^{(n)}$ has no non-zero $P^{(n)}$ -directions, by (62) it is a vertex, which completes our proof. \square