Optimization with stochastic preferences based on a general class of scalarization functions

Nilay Noyan
Industrial Engineering Program, Sabancı University, Istanbul, Turkey, nnoyan@sabanciuniv.edu
Gábor Rudolf
Department of Industrial Engineering, Koc University, Istanbul, Turkey, grudolf@ku.edu.tr

September 17, 2016

Abstract: It is of crucial importance to develop risk-averse models for multicriteria decision making under uncertainty. A major stream of the related literature studies optimization problems that feature multivariate stochastic benchmarking constraints. These problems typically involve a univariate stochastic preference relation, often based on stochastic dominance or a coherent risk measure such as conditional value-at-risk (CVaR), which is then extended to allow the comparison of random vectors by the use of a family of scalarization functions: All scalarized versions of the vector of the uncertain outcomes of a decision are required to be preferable to the corresponding scalarizations of the benchmark outcomes. While this line of research has been dedicated almost entirely to linear scalarizations, the corresponding deterministic literature uses a wide variety of scalarization functions that, among other advantages, offer a high degree of modeling flexibility. In this paper we aim to incorporate these scalarizations into a stochastic context by introducing the general class of min-biaffine functions. We study optimization problems in finite probability spaces with multivariate stochastic benchmarking constraints based on min-biaffine scalarizations. We develop duality results, optimality conditions, and a cut generation method to solve these problems. We also introduce a new characterization of the risk envelope of a coherent risk measure in terms of its Kusuoka representation as a tool towards proving the finite convergence of our solution method. The main computational challenge lies in solving cut generation subproblems; we develop several mixed-integer programming formulations by exploiting the min-affine structure and leveraging recent advances for solving similar problems with linear scalarizations. We conduct a computational study on a well-known homeland security budget allocation problem to examine the impact of the proposed scalarizations on optimal solutions, and illustrate the computational performance of our solution methods.

Keywords: stochastic programming; multicriteria; multivariate risk; coherent risk measures; conditional value-at-risk; stochastic dominance; cut generation

---

1. Introduction Multicriteria stochastic optimization problems, which involve decisions leading to uncertain outcomes that can be evaluated according to multiple stochastic performance measures of interest, find natural applications in a wide variety of fields including humanitarian logistics, medical (radiation) treatment planning, agriculture revenue management (see, e.g., Noyan, 2012; Hu and Mehrotra, 2012; Armbruster and Luedtke, 2015, respectively). In such decision making problems it is often desirable to find a solution that results in a random outcome which is preferable to an existing reference, or benchmark. Optimization with the arising multivariate stochastic benchmarking constraints has been receiving increasing attention in the literature. These constraints are based on preference relations between random vectors, and thus generalize two widely studied notions: the stochastic ordering of scalar-valued random variables, and preference relations between deterministic vectors of multiple performance criteria.

The literature on optimization with multicriteria stochastic preference constraints primarily focuses on two types of benchmarking relations: the multivariate second-order stochastic dominance (SSD) relation, and multivariate CVaR relations. While dominance constraints are intuitively appealing and have a well-developed theoretical background (see, e.g., Dentcheva and Ruszczyński, 2009; Homem-de-Mello and Mehrotra, 2009; Hu et al., 2012; Dentcheva and Wolfhagen, 2015; 2016), it is well-known that they are often overly demanding and conservative in practice, and can even lead to infeasible problems. As a remedy, one may instead enforce more relaxed preference relations based on risk measures. In particular, Noyan and Rudolf (2013) highlight the use of CVaR-based preferences as a way to provide flexible, meaningful, and tractable relaxations for SSD relations; this approach is further developed in Kucukyavuz and Noyan (2016) and Liu et al. (2015). Moreover, Noyan and Rudolf (2013) extend their CVaR-based methodology to optimization problems with benchmarking constraints that feature a wider class of coherent risk measures. Following suit, while in this paper we detail optimization with the multivariate preference relations based both on SSD and coherent risk measures, our primary focus is the latter problem (with a special emphasis on CVaR).

A large proportion of existing studies extend univariate SSD and CVaR-based relations to the multivariate case by utilizing *scalarization functions* (for an overview of non-scalarizing methods see, e.g., Gutjahr and Pichler, 2013). This scalarization approach typically considers a family of linear (weighted-sum) scalarization functions, parametrized by a *scalarization set* of weight vectors, and requires all scalarized versions of the random vectors to conform to the univariate preference relation of interest. Introducing the scalarization set proves to be a useful way to address ambiguities and inconsistencies in the weight vectors which represent the relative subjective importance of the decision criteria; for detailed discussions we refer the reader to Hu and Mehrotra (2012) and Liu et al. (2015). One common approach (which the present paper will also follow), is to allow arbitrary polyhedra as scalarization sets, providing a good balance between flexibility and computational tractability.

While the aforementioned methods feature a variety of underlying univariate stochastic preference relations and scalarization sets, the scalarization functions are almost exclusively of the linear type (one exception, albeit not in an optimization context, is the work of Burgert and Rüschendorf, 2006). This is in sharp contrast to the deterministic multicriteria optimization literature, which has seen a proliferation of scalarization methods due to several important factors. Firstly, the wide variety of available scalarizations offers a high degree of modeling flexibility, with many functions (such as the achievement scalarization functions introduced in Wierzbicki, 1980) having natural interpretations for decision makers. Secondly, scalarization functions with favorable analytical properties such as Lipschitz continuity can enable the use of efficient numerical methods. Thirdly, different scalarization functions typically lead to different optimal solutions, and the use of appropriate scalarizations often makes it possible to capture the Pareto-efficient frontier. We note that in a deterministic context scalarization is commonly employed in the objective, in contrast to our benchmarking approach where the choice of a suitable scalarization function can be used to shape the feasible region.

The goal of the present paper is to incorporate the variety of non-linear scalarization functions that exist in the deterministic literature into multicriteria optimization problems with stochastic preference constraints. The primary challenge in doing so lies in the fact that all available solution methods to these highly computationally demanding problems depend heavily on linear formulations. Our main contribution is to introduce a novel class of scalarization functions that is sufficiently general to include popularly used scalarizations, yet still allows tractable solution methods in a stochastic context.

Analogously to the well-known notion of min-affine and max-affine functions, we propose the use of *min-biaffine* functions with two vector variables (where the scalarization vector is the first variable, and the outcome vector the second). These are the functions that can be expressed as the minimum of finitely many mappings that are affine in both of the variables. Two important factors make min-biaffine functions uniquely suited for use in our context. Firstly, they form a sufficiently rich family containing a wide range of scalarizations such as linear scalarization (see, e.g., Ehrgott, 2005), weighted worst-case scalarization and weighted Chebyshev scalarization (see,e.g., Kaya and Maurer, 2014), and a variety of achievement scalarizing functions (ASFs) including classical ASFs (see, e.g., Wierzbicki, 1999), two-slope ASFs (Luque et al., 2012), and parameterized ASFs (Nikulin et al., 2012)). Secondly, their affine structure makes it possible to incorporate them into optimization problems in a fashion that allows us to adapt and extend many of the tools and formulations that are available for linear scalarization functions.

In addition to providing a tractable solution method we also develop a theoretical background to optimization problems with stochastic preference constraints that involve our new class of min-biaffine scalarizations.

We obtain duality formulations and optimality conditions which generalize the results of Dentcheva and Ruszczyński (2009) for the SSD-based constraints, and the results of Noyan and Rudolf (2013) for the risk measure-based constraints. Some of the arising side results also prove to be of independent interest, including a characterization of the maximal risk envelope in terms of the Kusuoka representation for coherent risk measures. In addition, our solution algorithms are made more efficient by a new method to exploit scenario proximity in the data.

The rest of the paper is organized as follows. In Section 2, we recall some fundamental definitions and results. Section 3 focuses on our new class of min-biaffine scalarization functions, and establishes corresponding multivariate preference relations based on coherent risk measures and SSD. We provide basic theoretical results, including a finite representation theorem, and introduce optimization problems with benchmarking constraints using our new class of preference relations. Section 4 is dedicated to the important special case where the outcome mappings are linear, presenting duality results and optimality conditions. This is followed in Sections 5 and 6 by finitely convergent cut generation-based solution methods for risk measure- and SSD-based models, respectively. Section 7 presents a brief computational study, while Section 8 contains our concluding remarks.

2. Preliminaries

We begin by establishing some notation and conventions used throughout the paper. We consider larger values of random variables to be preferable and quantify the risk associated with a random variable either via risk measures (where higher values correspond to riskier random outcomes) or via *acceptability functionals* (where higher values indicate less risky outcomes).

The cumulative distribution function of a random variable X is denoted by F_X. All random variables in this paper are assumed to be defined on a finite probability space $(\Omega, 2^{\Omega}, \mathcal{P})$ with $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $\mathcal{P}(\omega_i) = p_i > 0,\ i \in \{1, \ldots, n\}$. We denote the (finite) set of all non-zero probabilities of events by

$$\mathcal{K} = \left\{\mathcal{P}(S)\ :\ S \in 2^{\Omega},\ \mathcal{P}(S) > 0\right\}. \tag{1}$$

The set of the first n positive integers is denoted by $[n] = \{1, \ldots, n\}$, while the positive part of a number $x \in \mathbb{R}$ is denoted by $[x]_+ = \max(x, 0)$. Given an optimization problem PR we denote the objective function value at a decision vector $\mathbf{v}$ by $\mathrm{OBF}_{\mathrm{PR}}(\mathbf{v})$.

2.1 Coherent risk measures

Unless specified otherwise, the definitions and results in this section are presented specifically for finite probability spaces. For a more general treatment of these topics we refer to Pflug and Römisch (2007) and Shapiro et al. (2009).

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

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

The definition of a coherent risk measure given above was first introduced in the influential work of Artzner et al. (1999); our presentation follows along the lines of Pflug and Römisch (2007) and Pflug and Wozabal (2009).

We now introduce an important family of coherent acceptability functionals. The *conditional value-at-risk* (Rockafellar and Uryasev, 2000; 2002) at confidence level $\alpha \in (0, 1]$ for a random variable X is defined as

$$\mathrm{CVaR}_{\alpha}(X) = \max\left\{\eta - \frac{1}{\alpha}\mathbb{E}\left([\eta - X]_+\right)\ :\ \eta \in \mathbb{R}\right\}. \tag{2}$$

The optimization problem in (2) can equivalently be formulated as the following linear program

$$\max\{\eta - \frac{1}{\alpha}\sum_{i\in[n]} p_i w_i \ : \ w_i \geq \eta - x_i \quad \forall\, i \in [n],\ \mathbf{w}\in\mathbb{R}^n_+,\ \eta\in\mathbb{R}\}, \tag{3}$$

where $x_1,\dots,x_n$ are the (not necessarily distinct) realizations of X with corresponding probabilities $p_1,\dots,p_n$. It is well known that the maximum in these formulations is attained when $\eta = \mathrm{VaR}_\alpha(X)$, where $\mathrm{VaR}_\alpha(X) = \min\{\gamma \ : \ F_X(\gamma)\geq\alpha\}$ is the *value-at-risk* at confidence level α (also known as the α-quantile). Accordingly, $\mathrm{CVaR}_\alpha(X)$ can also be expressed via the following trivial minimization:

$$\min\{\mathrm{VaR}_\alpha(X) - \frac{1}{\alpha}\sum_{i\in[n]} p_i w_i \ : \ w_i \geq [\mathrm{VaR}_\alpha(X) - x_i]_+ \quad \forall\, i\in[n]\}. \tag{4}$$

CVaR is of particular importance because it is coherent (Pflug, 2000) and serves as a fundamental building block for other coherent risk measures. Kusuoka (2001) has shown that (under very general conditions) coherent acceptability functionals can be represented as the infimum of continuous convex combinations of CVaR at various confidence levels. The functionals which have finite such representations form a rich and computationally tractable class: We say that $-\rho$ is a *finitely representable coherent risk measure* if it has a representation of the form

$$\rho(V) = \min_{h\in[H]} \sum_{j\in[K]} \mu_j^{(h)}\, \mathrm{CVaR}_{\alpha_j}(V) \quad \text{for all} \quad V\in\mathcal{V}, \tag{5}$$

for some integers H and K, confidence levels $\alpha_1,\dots,\alpha_K \in (0,1]$, and weight vectors $\boldsymbol{\mu}^{(h)} \in \left\{\boldsymbol{\mu}\in\mathbb{R}^K_+ \ : \ \sum_{j\in[K]}\mu_j = 1\right\}$.

We can assume without loss of generality that the confidence levels in the above representation are selected from the set $\mathcal{K}$, see, e.g., Noyan and Rudolf (2015), who also prove that finitely representable measures can provide an arbitrarily close approximation to any coherent risk measure that can be evaluated on all finite probability spaces. Additional details about finding such approximations, along with explicit bounds on the approximation error, can be found in Haskell et al. (2016). It was also shown in Noyan and Rudolf (2015) that the class of risk measures $-\rho$ that have a representation of the form (5) with $H=1$, i.e., can be expressed as

$$\rho(V) = \sum_{j\in[K]} \mu_j\, \mathrm{CVaR}_{\alpha_j}(V) \quad \text{for all} \quad V\in\mathcal{V}, \tag{6}$$

coincides with the class of *spectral risk measures*. For the original definition of spectral risk measures we refer to Acerbi (2004). These risk measures have received significant attention in a financial context due to the so-called *comonotone additive* property, which states that risk pooling is not rewarded for "worst-case" dependence structures.

To conclude this section, we present an alternative dual representation for coherent risk measures on finite probability spaces (Artzner et al., 1999).

THEOREM 2.1 *For every coherent risk measure $-\rho$ there exists a* risk envelope $\mathcal{Q} \subset \{Q\in\mathcal{V} \ : \ Q\geq 0,\ \mathbb{E}(Q)=1\}$ *such that*

$$\rho(V) = \inf_{Q\in\mathcal{Q}} \mathbb{E}(QV) \quad \textit{holds for all } V\in\mathcal{V}. \tag{7}$$

2.2 Univariate stochastic dominance Stochastic dominance relations compare random variables with respect to the pointwise values of a performance function constructed from the distribution function. In particular, in the first order dominance (FSD) relation the performance function is the CDF itself, while the SSD relation, which has received significant attention due its correspondence with risk-averse preferences, is based on the so-called *second-order distribution function*. We refer the reader to Müller and Stoyan (2002) for a detailed review on stochastic dominance relations.

Definition 2.1 *We say that a random variable X dominates another random variable Y in the first order, denoted by $X \succcurlyeq_{(1)} Y$, if the relation $F_X(\eta) \le F_Y(\eta)$ holds for all $\eta \in \mathbb{R}$. We say that X dominates Y in the second order, denoted by $X \succcurlyeq_{(2)} Y$, if*

$$\mathbb{E}([\eta - X]_+) \le \mathbb{E}([\eta - Y]_+) \quad \text{for all} \quad \eta \in \mathbb{R}. \tag{8}$$

The definition remains valid if in (8) the inequality is only required for possible values $\eta \in \text{Range}(Y)$ of Y; see Dentcheva and Ruszczyński (2003). It is well-known (see, e.g., Ogryczak and Ruszczyński, 2002) that the SSD relation can be equivalently defined by the following inverse formulation:

$$\text{CVaR}_\alpha(X) \ge \text{CVaR}_\alpha(Y) \quad \text{for all} \quad \alpha \in (0,1]. \tag{9}$$

The definition remains valid if in (9) the inequality is only required whenever α is the probability of an event, i.e., for $\alpha \in \mathcal{K}$; see Noyan and Rudolf (2013).

2.3 Geometric preliminaries. Let us call a vector $\mathbf{c} \in \mathbb{R}^d$ a *d-vertex* of a polyhedron $P \subset \mathbb{R}^d \times \mathbb{R}^n$ if it can be extended into a vertex, i.e., if there exists some $\mathbf{v} \in \mathbb{R}^h$ such that $(\mathbf{c}, \mathbf{v})$ is a vertex of P. Given a polyhedron $P = P^{(1)} \subset \mathbb{R}^d \times \mathbb{R}^n$ we introduce the following series of "liftings":

$$P^{(k)} = \left\{ (\mathbf{c}, \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)}) \in \mathbb{R}^d \times \mathbb{R}^n \times \dots \times \mathbb{R}^n \;:\; (\mathbf{c}, \mathbf{v}^{(i)}) \in P \text{ for all } i \in [k] \right\}. \tag{10}$$

The next theorem shows that this lifting procedure can only introduce a finite number of new d-vertices.

Theorem 2.2 *(Noyan and Rudolf, 2013) Let $P \subset \mathbb{R}^d \times \mathbb{R}^n$ be an arbitrary polyhedron, and let k be a positive integer. Then every d-vertex of the lifted polyhedron $P^{(k)}$ is also a d-vertex of $P^{(d)}$.*

3. Multivariate Stochastic Preference Relations Preference relations among scalar-valued random variables, such as the SSD relation or the relations induced by risk measures, can be extended to vector-valued variables by the use of scalarization functions. In multiobjective optimization, scalarization functions are introduced in the context of either minimization or maximization problems (in an essentially equivalent fashion). In this paper we use the latter convention; all definitions and notations below have been adapted accordingly whenever necessary.

Definition 3.1 *Let $\mathbf{X}, \mathbf{Y} \in \mathcal{V}^d$ be two d-dimensional random vectors, $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ a scalarization function, $C \subset \mathbb{R}^d$ a convex set of scalarization vectors, and $\rho : \mathcal{V} \to \mathbb{R}$ a coherent risk measure. We say that $\mathbf{X}$ is ρ-*preferable *to Y with respect to φ and C, denoted as $\mathbf{X} \succcurlyeq_\rho^{\varphi,C} \mathbf{Y}$, if*

$$\rho(\varphi(\mathbf{c}, \mathbf{X})) \ge \rho(\varphi(\mathbf{c}, \mathbf{Y})) \quad \text{for all } \mathbf{c} \in C. \tag{11}$$

Similarly, we say that the random vector $\mathbf{X}$ dominates $\mathbf{Y}$ in second order with respect to φ and C, denoted as $\mathbf{X} \succcurlyeq_{(2)}^{\varphi,C} \mathbf{Y}$, if

$$\varphi(\mathbf{c}, \mathbf{X}) \succcurlyeq_{(2)} \varphi(\mathbf{c}, \mathbf{Y}) \quad \text{for all } \mathbf{c} \in C. \tag{12}$$

Remark 3.1 *For the sake of simplicity, in the above definition we assumed that scalarization vectors and the realizations of the random vectors belong to the same space $\mathbb{R}^d$. This is indeed the case for all of the scalarization functions featured in this paper, with the single exception of the* two-slope scalarization functions *discussed below, which take the form $\varphi : \mathbb{R}^{2d} \times \mathbb{R}^d \to \mathbb{R}$. We point out that all of our results and methods remain valid for the case of unequal dimensions, i.e., for functions $\varphi : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$, since vectors of lower dimension can be artificially extended with irrelevant "dummy" coordinates without affecting problem formulations.*

The majority of mathematical and methodological difficulties in this line of research arise from the presence of joint risk expressions such as $\rho(\varphi(\mathbf{c}, \mathbf{X}))$. A natural alternative idea would be to replace these by $\varphi(\mathbf{c}, \rho(\mathbf{X}))$, i.e., by the scalarization of the component-wise risk measures $\rho(\mathbf{X}) = (\rho(X_1), \dots, \rho(X_d))$. In this case even large-scale problems typically become immediately tractable via simple linear programming formulations. However, the expression $\varphi(\mathbf{c}, \rho(\mathbf{X}))$ only depends on the random vector $\mathbf{X}$ through the marginal distributions of $X_1, \dots, X_d$, and thus ignores entirely their joint behavior. This is highly undesirable because the canonical coherent risk measures CVaR_α, which can be viewed as the fundamental building blocks of *all* coherent risk measures, are designed to describe the behavior of a system in the "worst" α proportion of possible scenarios. Therefore using the aforementioned alternative formulation would only be intuitively justified in the trivial case when the random outcomes $X_1, \dots, X_n$ are comonotone, and thus their worst-case scenarios coincide. Similar considerations apply for the case of SSD-based preference relations.

3.1 Min-biaffine scalarization functions In this paper we focus on the class of scalarization functions that can be represented as a minimum of finitely many biaffine functions:

$$\varphi(\mathbf{c}, \mathbf{x}) = \min_{t \in [T]} A_t(\mathbf{c}, \mathbf{x}), \text{ where } T \in \mathbb{N} \text{ and } A_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \text{ is affine in both variables for all } t \in [T]. \quad (13)$$

If a function φ has a representation of the form (13), we say that it is a *min-biaffine* function. The class of such functions, denoted by Φ, is quite general and includes most scalarization functions that are commonly used in the deterministic multiobjective optimization literature, as the following representations show:

- Linear (weighted-sum) scalarization (see, e.g., Steuer, 1986; Ehrgott, 2005):
  - Definition: $\varphi(\mathbf{c}, \mathbf{x}) = \mathbf{c}^\top \mathbf{x}$.
  - Min-affine representation: $T = 1$, $A_1(\mathbf{c}, \mathbf{x}) = \mathbf{c}^\top \mathbf{x}$.
- Weighted worst-case scalarization (note that this function is based on the Chebyshev norm, as we have $\varphi(\mathbf{c}, \mathbf{X}) = -\| -(c_1 X_1, \dots, c_d X_d)\|_\infty$):
  - $\varphi(\mathbf{c}, \mathbf{X}) = \min_{j \in [d]} c_j X_j$.
  - $T = d$, $A_t(\mathbf{c}, \mathbf{x}) = c_t x_t$.
- Weighted Chebyshev scalarization function (see, e.g., Kaya and Maurer, 2014):
  - $\varphi(\mathbf{c}, \mathbf{X}) = \min_{j \in [d]} c_j (X_j - r_j)$, where $\mathbf{u} = (r_1, \dots, r_d)$ is a fixed reference vector (often chosen to be the so-called *utopian point*).
  - $T = d$, $A_t(\mathbf{c}, \mathbf{x}) = c_t x_t - r_t$.
- The class of *achievement scalarizing functions* (ASFs) includes a variety of similar functions (see, e.g., Wierzbicki, 1999, for a detailed overview); the following is a typical example:
  - $\varphi(\mathbf{c}, \mathbf{X}) = \min_{j \in [d]} c_j \min(X_j - r_j, 0) + \delta \sum_{i \in [d]} c_i (X_i - r_i)$, where $\delta > 0$ is a small constant.
  - For $\mathbf{c} \geq \mathbb{0}$: $T = 2d$,
  $$A_t(\mathbf{c}, \mathbf{X}) = \begin{cases} c_t (X_t - r_t) + \delta \sum_{i \in [d]} c_i (X_i - r_i) & \text{if } t \in [d] \\ \delta \sum_{i \in [d]} c_i (X_i - r_i) & \text{if } t = d + 1. \end{cases}$$
- Two-slope ASFs introduced by Luque et al. (2012) use one of two different scalarization weights for each decision criterion, depending on whether or not the objective value exceeds the corresponding reference (cf. Remark 3.1):
  - $\varphi\left((\mathbf{c}^{(1)}, \mathbf{c}^{(2)}), \mathbf{X}\right) = \min\limits_{j \in [d]} c_j^{(1)} \min(X_j - r_j, 0) + c_j^{(2)} \max(X_j - r_j, 0) + \delta \sum\limits_{i \in [d]} (X_i - r_i)$.
  - For $\mathbf{c} \geq \mathbb{0}$: $T = d + 1$, $A_t(\mathbf{c}, \mathbf{X}) = \begin{cases} c_t (X_t - r_t) + \delta_{i \in [d]} c_i (X_i - r_i) & \text{if } t \in [d] \\ \delta_{i \in [d]} c_i (X_i - r_i) & \text{if } t = d + 1. \end{cases}$
- The parameterized ASFs introduced by Nikulin et al. (2012) for parameters $q \in [d]$:
  - $\varphi(\mathbf{c}, \mathbf{X}) = \min\limits_{J \subset [d],\ |J| = q} \sum\limits_{j \in J} c_j \min(X_j - r_j, 0)$.
  - For $\mathbf{c} \geq \mathbb{0}$: $T = \sum_{k=0}^{q} \binom{d}{k}$, $A_t(\mathbf{c}, \mathbf{x}) = \sum_{i \in Q_t} c_i (X_i - r_i)$, where $Q_1, \dots, Q_T$ is a numbered list of the subsets of $[d]$ with at most q elements.

REMARK 3.2 *Min-biaffine functions are naturally suited to provide computationally tractable approximations to jointly concave functions $f(\mathbf{c},\mathbf{x})$ of two vector variables (i.e., concave as a function of the vector $(\mathbf{c},\mathbf{x})$), since such functions are superdifferentiable in the interior of their domain, and thus can be expressed as the infimum of biaffine functions. On the other hand, all min-biaffine functions are (separately) concave in both variables, because for any fixed $\mathbf{c}_0$ and $\mathbf{x}_0$ the mappings $\mathbf{x} \mapsto f(\mathbf{c}_0,\mathbf{x})$ and $\mathbf{c} \mapsto f(\mathbf{c},\mathbf{x}_0)$ are the minimum of affine functions and therefore concave. However, not all separately concave functions can be expressed as the infimum of biaffine functions. For example, it is easy to verify that while the function $f:\mathbb{R}^2 \to \mathbb{R}$ given by $f(c,x) = c^2x^2 - (c-1)^2x^2 - c^2(x-1)^2$ is concave in both of its variables, it does not have tight biaffine majorants at $(c,x) = (0,0)$.*

In a deterministic multiobjective context scalarization functions are often discussed in relation to the efficient frontier. In particular, it is well-known that linear scalarization functions can be used to capture the efficient frontier of a convex region, and under appropriate conditions ASFs can be shown to produce all (properly or weakly) Pareto-optimal solutions (see,e.g., Miettinen, 1999) even in the non-convex case. While there is no unique widely accepted definition of an efficient frontier for sets of random vectors (see, e.g., the notion of stochastic dominance-based Pareto optimality in Liu et al. (2015) and the earlier related work of Ben Abdelaziz (2012)), it remains true that stochastic preference relations based on different scalarization functions give rise to different sets of non-dominated points. Accordingly, choosing an appropriate scalarization function plays an important role in shaping the feasible regions of problems with the multivariate stochastic preference constraints. We now present a small-scale example to illustrate the impact of using different scalarization methods.

EXAMPLE 3.1 *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 $\Delta : \Omega \to \mathbb{R}$ denote the random variable with realizations $\Delta(\omega_1) = 0$, $\Delta(\omega_2) = 1$, and let*

$$\mathbf{X}^1 = \begin{bmatrix} 1.1-\Delta \\ 0.1+\Delta \end{bmatrix}, \quad \mathbf{X}^2 = \begin{bmatrix} 0.1+\Delta \\ 1.1-\Delta \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 1.2-\Delta \\ 1.2+\Delta \end{bmatrix}.$$

In addition, we define the scalarization polyhedron $C = \{(c_1,c_2) \in \mathbb{R}^2 \ : \ c_1+c_2 = 1,\ 0.2 \le c_1 \le 0.8\}$.

Figure 1 shows feasibility regions of the form $\{(z^1,z^2) \mid z^1\mathbf{X}^1 + z^2\mathbf{X}^2 \succcurlyeq \mathbf{Y}\}$, for the following choices of the multivariate stochastic preference relation $\succcurlyeq$, where $\succcurlyeq_{(i)}$ denotes either the univariate FSD relation $\succcurlyeq_{(1)}$ or the univariate SSD relation $\succcurlyeq_{(2)}$:

- *Component-wise preference: $\mathbf{X} \succcurlyeq \mathbf{Y}$ if $X_1 \succcurlyeq_{(i)} Y_1$ and $X_2 \succcurlyeq_{(i)} Y_2$. Note that this corresponds to using a linear scalarization approach with the discrete scalarization set $\{(0,1),\ (1,0)\}$.*
- *Preference based on linear scalarizations: $\mathbf{X} \succcurlyeq \mathbf{Y}$ if $\mathbf{c}^\top\mathbf{X} \succcurlyeq_{(i)} \mathbf{c}^\top\mathbf{Y}$ for all $\mathbf{c} \in C$.*
- *Preference based on weighted worst-case (Chebyshev norm) scalarizations: $\mathbf{X} \succcurlyeq \mathbf{Y}$ if $\min(c_1X_1, c_2X_2) \succcurlyeq_{(i)} \min(c_1Y_1, c_2Y_2)$ for all $\mathbf{c} \in C$.*

3.2 Finiteness results If the scalarization function is continuous in $\mathbf{c}$, we can assume without loss of generality that the scalarization set C is compact. For the case of polyhedral scalarization sets this implies that we can assume C is a polytope. The proof of the following result is analogous to the proof of Proposition 2 in Noyan and Rudolf (2013) for the case of linear scalarization functions.

PROPOSITION 3.1 *Assume that the mapping $\varphi(\cdot,\mathbf{x}) : \mathbf{c} \mapsto (\mathbf{c},\mathbf{x})$ is continuous for all $\mathbf{x} \in \mathbb{R}^d$. Let C be a nonempty convex set, and let $\tilde{C} = \{\mathbf{c} \in \mathrm{cl\ cone}(C) \ : \ \|\mathbf{c}\|_1 \le 1\}$, where $\mathrm{cl\ cone}(C)$ denotes the closure of the conical hull of the set C. Then, given any integrable random vectors $\mathbf{X}$ and $\mathbf{Y}$ the relations $\mathbf{X} \succcurlyeq_\rho^{\varphi,C} \mathbf{Y}$ and $\mathbf{X} \succcurlyeq_\rho^{\varphi,\tilde{C}} \mathbf{Y}$ are equivalent for any coherent risk measure ρ. Similarly, the relations $\mathbf{X} \succcurlyeq_{(2)}^{\varphi,C} \mathbf{Y}$ and $\mathbf{X} \succcurlyeq_{(2)}^{\varphi,\tilde{C}} \mathbf{Y}$ are also equivalent.*

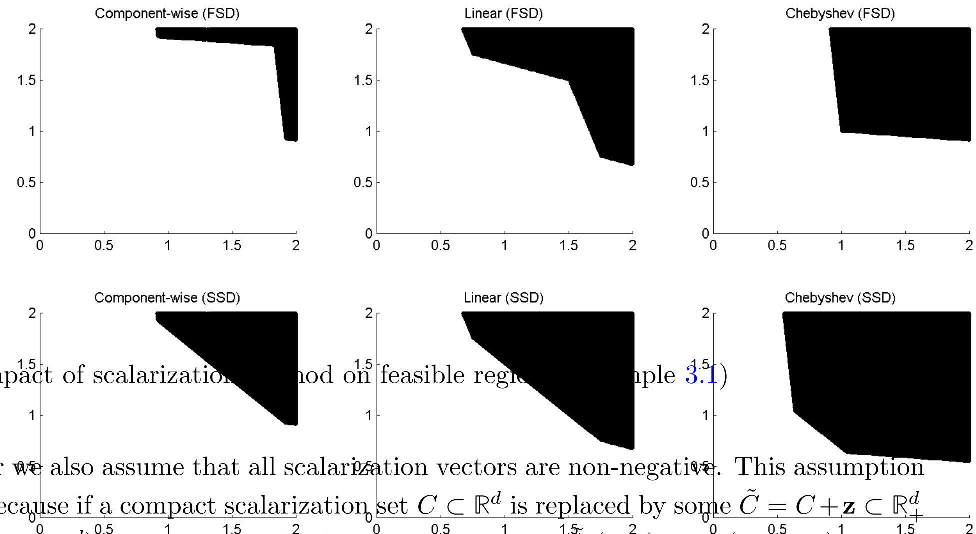

Figure 1: Impact of scalarization method on feasible region (Example 3.1)

In the remainder of this paper we also assume that all scalarization vectors are non-negative. This assumption is without loss of generality because if a compact scalarization set $C \subset \mathbb{R}^d$ is replaced by some $\tilde{C} = C + \mathbf{z} \subset \mathbb{R}^d_+$ (with a suitably large vector $\mathbf{z} \in \mathbb{R}^d$), the correspondingly altered mappings $\tilde{A}_t(\tilde{\mathbf{c}}, \mathbf{x}) = A_t(\tilde{\mathbf{c}} - \mathbf{z}, \mathbf{x})$ remain biaffine.

For any nontrivial polyhedron C of scalarization vectors the corresponding ρ-preferability constraint (11) is equivalent by definition to a collection of infinitely many scalar-based constraints, one for each scalarization vector $\mathbf{c} \in C$. The next theorem shows that, for the case of finitely representable coherent risk measures in finite probability spaces, it is sufficient to consider a finite subset of these vectors. These vectors can be obtained as projections of the vertices of a higher dimensional polyhedron.

Theorem 3.1 *Let $-\rho$ be a finitely representable coherent risk measure (equivalently, let ρ be a finitely representable coherent acceptability functional). Consider a scalarization function $\varphi(\mathbf{c}, \mathbf{x}) \in \Phi$, a scalarization polytope $C \subset \mathbb{R}^d_+$ and a d-dimensional benchmark random vector $\mathbf{Y}$ with realizations $\mathbf{y}_i = \mathbf{Y}(\omega_i)$ for $i \in [n]$.*

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

$$\rho(\varphi(\mathbf{c}^{(\ell)}, \mathbf{X})) \geq \rho(\varphi(\mathbf{c}^{(\ell)}, \mathbf{Y})) \quad \text{for all } \ell \in [N], \tag{14}$$

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

(ii) *If the risk measure $-\rho$ is spectral, then the above equivalence holds with the choice of the lifted polyhedron $P = P^{(d)}(\varphi, C, \mathbf{Y})$, with the polyhedron $P(\varphi, C, \mathbf{Y})$ defined as follows:*

$$P(\varphi, C, \mathbf{Y}) = \left\{ (\mathbf{c}, \eta, \mathbf{w}) \in C \times \mathbb{R} \times \mathbb{R}^n_+ \ : \ w_i \geq \eta - A_t(\mathbf{c}, \mathbf{y}_i), \quad i \in [n], \ t \in [T] \right\}, \tag{15}$$

where $A_1, \ldots, A_T$ are the affine mappings in the representation (13) of φ.

(iii) *In the case $\rho = \mathrm{CVaR}_\alpha$ the equivalence holds with $P = P(\varphi, C, \mathbf{Y})$ for any confidence level $\alpha \in (0, 1]$.*

Proof. Let us first assume that the relation $\mathbf{X} \succcurlyeq_\rho^{\varphi, C} \mathbf{Y}$ does not hold, implying that the optimal objective value of the following *cut generation problem*

$$\min_{\mathbf{c} \in C} \rho(\varphi(\mathbf{c}, \mathbf{X})) - \rho(\varphi(\mathbf{c}, \mathbf{Y})) \tag{16}$$

is negative. Since ρ is finitely representable, it can be written in the form (5) for some confidence levels $\alpha_1, \ldots, \alpha_K \in (0,1]$ and weight vectors $\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(H)}$. Using the representation (3) of CVaR and the representation (13) of φ, we can therefore express the minimum in (16) as the optimum of the following problem:

$$
\begin{aligned}
\min \quad & \rho(\varphi(\mathbf{c}, \mathbf{X})) - z && && \text{(17a)}\\
\text{s.t.} \quad & z \leq \sum_{j \in [K]} \mu_j^{(h)} \left(\eta^{(j)} - \frac{1}{\alpha_j} \sum_{i \in [n]} p_i w_i^{(j)} \right), && \forall\, h \in [H] && \text{(17b)}\\
& w_i^{(j)} \geq \eta^{(j)} - A_t(\mathbf{c}, \mathbf{y}_i), && \forall\, i \in [n],\ j \in [K],\ t \in [T] && \text{(17c)}\\
& w_i^{(j)} \geq 0, && \forall\, i \in [n],\ j \in [K] && \text{(17d)}\\
& \mathbf{c} \in C. && && \text{(17e)}
\end{aligned}
$$

We first observe that the objective function in (17a) is concave. Indeed, as ρ is coherent, it is monotone and concave. Furthermore, by our assumptions the mapping $c \mapsto \varphi(\mathbf{c}, \mathbf{X})$ is a minimum of affine functions, and therefore also concave. It immediately follows that the mapping $c \mapsto \rho(\varphi(\mathbf{c}, \mathbf{X}))$ is concave for any random vector $\mathbf{X}$.

Let P denote the feasible set of the above problem. As P is a polyhedron, the concave minimization problem (17) has a vertex optimal solution, i.e., a solution of the form $(\mathbf{c}^{(\ell)}, z^*, \boldsymbol{\eta}^*, \mathbf{w}^*)$ for some $\ell \in [N]$. Since the vector $\mathbf{c}^{(\ell)}$ is a d-vertex of P, the relation (14) 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 $\mathbf{X} \succcurlyeq_{\rho}^{\varphi, C} \mathbf{Y}$ trivially implies (14), which completes the proof of part (i).

To show part (ii), let us consider a spectral risk measure $-\rho$, and recall that it has a representation of the form (6). Therefore, in this case the minimum in (16) can be expressed as the optimum of the following problem:

$$
\begin{aligned}
\min \quad & \rho(\varphi(\mathbf{c}, \mathbf{X})) - \sum_{j \in [K]} \mu_j \left(\eta^{(j)} - \frac{1}{\alpha_j} \sum_{i \in [n]} p_i w_i^{(j)} \right)\\
\text{s.t.} \quad & (17c) - (17e).
\end{aligned}
$$

Since the second term of the objective function is linear, we again have a concave minimization problem. The feasible set is the polyhedron $P^{(K)}(C, \mathbf{Y})$, so there exists a vertex optimal solution $(\mathbf{c}^*, \eta^*, \mathbf{w}^*)$. By Theorem 2.2 the vector $\mathbf{c}^*$ is a d-vertex of $P^{(d)}(C, \mathbf{Y})$.

Part (iii) follows similarly by observing that for $\rho = \mathrm{CVaR}_\alpha$ the representation (6) holds with $K = \mu_1 = 1$ and $\alpha_1 = \alpha$. □

Remark 3.3 *To keep our exposition simple we stated Theorem 3.1 for relations of the form* $\mathbf{X} \succcurlyeq_{\rho}^{\varphi, C} \mathbf{Y}$*, 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*

$$\rho_1(\varphi(\mathbf{c}, \mathbf{X})) \geq \rho_2(\varphi(\mathbf{c}, \mathbf{Y})) \quad \textit{for all } \mathbf{c} \in C, \tag{18}$$

where the acceptability functionals ρ_1 *and* ρ_2 *can be defined on different finite probability spaces.*

Since, as discussed in Section 2.2, in finite probability spaces the SSD relation is equivalent to a finite collection of CVaR inequalities, part (iii) of Theorem 3.1 immediately implies that the results extend to the SSD case.

Corollary 3.1 *Using our notation from Theorem 3.1, the relation* $\mathbf{X} \succcurlyeq_{(2)}^{\varphi, C} \mathbf{Y}$ *holds if and only if*

$$\varphi(\mathbf{c}^{(\ell)}, \mathbf{X}) \succeq_{(2)} \varphi(\mathbf{c}^{(\ell)}, \mathbf{Y}) \quad \textit{for all } \ell \in [N]. \tag{19}$$

3.3 Optimization models We are now ready to introduce the class of multicriteria stochastic decision making problems that are the main object of our study. In these problems a decision $\mathbf{z}$ is selected from some feasible set Z, leading to a random outcome vector $G(\mathbf{z})$. This outcome vector represents the multiple random performance measures associated with the decision $\mathbf{z}$, and is determined according to some mapping $G : Z \times \Omega \to \mathbb{R}^d$, where $G(\mathbf{z}) : \Omega \to \mathbb{R}^d$ is the random vector given by $[G(\mathbf{z})](\omega) = G(\mathbf{z}, \omega)$. Our goal is to maximize the continuous objective function $f : Z \to \mathbb{R}$ while ensuring that the random outcome vector $G(\mathbf{z})$ is preferable to some benchmark random outcome vector $\mathbf{Y}$ according to a scalarization-based multivariate relation (as introduced in Definition 3.1). Let us fix a scalarization function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and a scalarization polytope $C \subset \mathbb{R}^d_+$. Then for a coherent risk measure $\rho : \mathcal{V} \to \mathbb{R}$ we consider the following problem:

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

Similarly, we introduce the SSD-constrained variant

$$\begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & \varphi(\mathbf{c}, G(\mathbf{z})) \succeq_{(2)} \varphi(\mathbf{c}, \mathbf{Y}), \quad \forall\, \mathbf{c} \in C \\ & \mathbf{z} \in Z. \end{aligned} \tag{GeneralP$_{\text{SSD}}$}$$

We remark that the benchmark $\mathbf{Y}$ is often constructed as $\mathbf{Y} = G(\bar{\mathbf{z}})$ from some benchmark decision $\bar{\mathbf{z}} \in Z$. For ease of exposition, we present GeneralP$_\rho$ and GeneralP$_{\text{SSD}}$ with a single multivariate preference relation. However, our methods and results extend naturally to the case of multiple constraints featuring different benchmarks and scalarization sets. Furthermore, in GeneralP$_\rho$ one can replace preference constraints by those of the more general form (18).

4. Linear Programming Formulations and Duality In this section we focus on the important special case where the mappings f and G are linear, the scalarization function φ is min-biaffine, and the set Z is polyhedral. Let us introduce the following notation:

- $Z = \{\mathbf{z} \in \mathbb{R}^{r_1} \ : \ U\mathbf{z} \leq \mathbf{u}\}$ for some $U \in \mathbb{R}^{r_2 \times r_1}$ and $\mathbf{u} \in \mathbb{R}^{r_2}$.
- $f(\mathbf{z}) = \mathbf{f}^\top \mathbf{z}$ for some vector $\mathbf{f} \in \mathbb{R}^{r_1}$.
- $G(\mathbf{z}, \omega) = \Gamma(\omega)\mathbf{z}$ for a random matrix $\Gamma : \Omega \to \mathbb{R}^{d \times r_1}$.
- $\varphi(\mathbf{c}, \mathbf{x}) = \inf\limits_{t \in [T]} A_t(\mathbf{c}, \mathbf{x}) = \inf\limits_{t \in [T]} \{\mathbf{a}_t^\top(\mathbf{c})\mathbf{x} + b_t(\mathbf{c})\}$ for some $\mathbf{a}_t : \mathbb{R}^d \to \mathbb{R}^d,\ b_t : \mathbb{R}^d \to \mathbb{R},\ t \in [T]$.

Using the above notation, (GeneralP$_\rho$) becomes

$$\begin{aligned} \max \quad & \mathbf{f}^\top \mathbf{z} \\ \text{s.t.} \quad & \rho(\varphi(\mathbf{c}, \Gamma\mathbf{z})) \geq \rho(\varphi(\mathbf{c}, \mathbf{Y})), \quad \forall\, \mathbf{c} \in C \\ & U\mathbf{z} \leq \mathbf{u}. \end{aligned} \tag{LinearP$_\rho$}$$

Working under the assumption that C is a polytope we can formulate (LinearP$_\rho$) as a linear program (LP). For a finite set $\tilde{C} = \{\tilde{\mathbf{c}}^{(1)}, \ldots, \tilde{\mathbf{c}}^{(L)}\}$ consider the following LP:

$$\begin{aligned} \max \quad & \mathbf{f}^\top \mathbf{z} \\ \text{s.t.} \quad & \sum_{j \in [K]} \mu_j^{(h)} \left(\eta_\ell^{(j)} - \frac{1}{\alpha_j} \sum_{i \in [n]} p_i w_{\ell i}^{(j)} \right) \geq \rho(\varphi(\tilde{\mathbf{c}}^{(\ell)}, \mathbf{Y})), && \forall\, \ell \in [L],\ h \in [H] \\ & w_{\ell i}^{(j)} \geq \eta_\ell^{(j)} - (\mathbf{a}_t^\top(\tilde{\mathbf{c}}^{(\ell)})\Gamma(\omega_i)\mathbf{z} + b_t(\tilde{\mathbf{c}}^{(\ell)})), && \forall\, j \in [K],\ \ell \in [L],\ t \in [T],\ i \in [n] \\ & w_{\ell i}^{(j)} \geq 0, && \forall\, j \in [K],\ \ell \in [L],\ i \in [n] \\ & U\mathbf{z} \leq \mathbf{u}. \end{aligned}$$

(FiniteP$_\rho(\tilde{C})$)

The next proposition, which shows that for suitable choices of $\tilde{C}$ the above LP is equivalent to (LinearP$_\rho$), is an easy consequence of Theorem 3.1.

PROPOSITION 4.1 *Let $\hat{C}$ denote the set consisting of the d-vertices $\mathbf{c}_{(1)}, \ldots, \mathbf{c}_{(N)}$ as defined in Theorem 3.1, 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$_\rho$) if and only if $(\mathbf{z}, \boldsymbol{\eta}^{(1)}(\mathbf{z}), \cdots, \boldsymbol{\eta}^{(K)}(\mathbf{z}), \mathbf{w}^{(1)}(\mathbf{z}), \cdots, \mathbf{w}^{(K)}(\mathbf{z}))$ is a feasible (optimal) solution of (FiniteP$_\rho(\tilde{C})$), where $\eta_\ell^{(j)}(\mathbf{z}) = \mathrm{VaR}_\alpha(\varphi(\tilde{\mathbf{c}}^{(\ell)}, \Gamma\mathbf{z}))$ and $w_{\ell i}^{(j)}(\mathbf{z}) = [\eta_\ell^{(j)}(\mathbf{z}) - \varphi(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z})]_+$.*

We next discuss how to obtain similar linear formulations for (GeneralP$_{\mathrm{SSD}}$). Keeping in mind the inverse formulation (9) of the SSD relation, (GeneralP$_{\mathrm{SSD}}$) can be written as

$$\begin{aligned} \max \quad & \mathbf{f}^\top \mathbf{z} \\ \text{s.t.} \quad & \mathrm{CVaR}_\alpha\left(\varphi(\mathbf{c}, \Gamma\mathbf{z})\right) \geq \mathrm{CVaR}_\alpha\left(\varphi(\mathbf{c}, \mathbf{Y})\right), \quad \forall\, \mathbf{c} \in C,\ \alpha \in (0,1] \\ & U\mathbf{z} \leq \mathbf{u}. \end{aligned} \tag{LinearP$_{\mathrm{SSD}}$}$$

When C is a polytope, (LinearP$_{\mathrm{SSD}}$) can also be formulated as a finite LP. It is an easy consequence of Corollary 3.1 that for finite sets $\tilde{C}$ satisfying $\hat{C} \subset \tilde{C} \subset C$ the following LP is equivalent to (LinearP$_{\mathrm{SSD}}$) in the sense established by Proposition 4.1.

$$\begin{aligned} \max \quad & \mathbf{f}^\top \mathbf{z} \\ \text{s.t.} \quad & \eta_\ell^{(j)} - \frac{1}{\alpha_j} \sum_{i \in [n]} p_i w_{\ell i}^{(j)} \geq \mathrm{CVaR}_{\alpha_j}(\varphi(\tilde{\mathbf{c}}^{(\ell)}, \mathbf{Y})), && \forall\, j \in [K],\ \ell \in [L] \\ & w_{\ell i}^{(j)} \geq \eta_\ell^{(j)} - (\mathbf{a}_t^\top(\tilde{\mathbf{c}}^{(\ell)})\Gamma(\omega_i)\mathbf{z} + b_t(\tilde{\mathbf{c}}^{(\ell)})), && \forall\, j \in [K],\ \ell \in [L],\ t \in [T],\ i \in [n] \\ & w_{\ell i}^{(j)} \geq 0, && \forall\, j \in [K],\ \ell \in [L],\ i \in [n] \\ & U\mathbf{z} \leq \mathbf{u}. \end{aligned} \tag{FiniteP$_{\mathrm{SSD}}(\tilde{C})$}$$

REMARK 4.1 *Alternatively, it is also possible to use the shortfall representation (8) instead of (9) to rewrite the SSD constraint in (GeneralP$_{\mathrm{SSD}}$), in which case the dominance constraint in (LinearP$_{\mathrm{SSD}}$) is replaced by the inequalities $\mathbb{E}\left([\eta - \varphi(\mathbf{c}, \Gamma\mathbf{z})]_+\right) \leq \mathbb{E}\left([\eta - \varphi(\mathbf{c}, \mathbf{Y})]_+\right)$ for all $\mathbf{c} \in C$ and $\eta \in \mathbb{R}$. Similar to the* CVaR*-based case, Corollary 3.1 implies that the arising problem can be equivalently formulated as a finite linear program.*

In practice linear programming formulations such as (FiniteP$_\rho(\tilde{C})$) and (FiniteP$_{\mathrm{SSD}}(\tilde{C})$) do not immediately offer a tractable way to solve to our optimization problems, as they can feature an exponential number of constraints. We now proceed to show that these formulations do, however, provide an important direct way to derive strong duality results and optimality conditions.

4.1 Duality – risk measure-based models Before we can formulate the dual of the optimization problem (LinearP$_\rho$) we need to establish some additional notation.

- The family of finitely supported non-negative measures on a set S is denoted by $\mathcal{M}_+^F(S)$.
- If μ_1 and μ_2 are measures on the sets S_1 and S_2, respectively, then $\mu_1 \times \mu_2$ denotes their product measure on the set $S_1 \times S_2$, satisfying $[\mu_1 \times \mu_2](D_1 \times D_2) = \mu_1(D_1)\mu_2(D_2)$ for measurable sets $D_1 \subset S_1$ and $D_2 \subset S_2$.
- If ν is a measure on $S_1 \times S_2$ then we let $\Pi_{S_2}\nu$ denote its marginal measure on S_2, defined by the equality $[\Pi_{S_2}\nu](D) = \nu(S_1 \times D)$ for measurable sets $D \subset S_2$.
- For a measure μ on a set S and an integrable function $f : S \to \mathbb{R}$ the measure $\int f(s)\, \mu(\mathrm{d}s)$ with Radon-Nikodym derivative f is given by $\left[\int f(s)\, \mu(\mathrm{d}s)\right](D) = \int_D f(s)\, \mu(\mathrm{d}s)$ for measurable sets $D \subset S$.

We now present a strong duality result and the corresponding optimality conditions. Assume that the risk measure $-\rho$ is given by its Kusuoka representation (5) and the scalarization function is given in the form (13). Let us consider the following dual to problem (LinearP$_\rho$):

$$
\begin{aligned}
\min \quad & -\sum_{h\in[H]}\int_C \rho(\varphi(\mathbf{c},\mathbf{Y}))\,\lambda_h(\mathrm{d}\mathbf{c}) + \mathbb{E}\left(\sum_{t\in[T]}\int_C \mathbf{b}_t(\mathbf{c})\,[\Pi_C\delta_t](\mathrm{d}\mathbf{c})\right) + \boldsymbol{\zeta}^\top\mathbf{u} \\
\text{s.t.} \quad & \mathbb{E}\left(\sum_{t\in[T]}\delta_t\right) = \sum_{h\in[H]}\lambda_h\times\mu_h, \\
& \sum_{t\in[T]}\delta_t(\omega_i) \le \sum_{h\in[H]}\left[\int\frac{1}{\alpha}\,\mu_h(\mathrm{d}\alpha)\right]\times\lambda_h, && \forall\, i\in[n] \\
& \mathbb{E}\left(\sum_{t\in[T]}\int_C \mathbf{a}_t^\top(\mathbf{c})\Gamma\,[\Pi_C\delta_t](\mathrm{d}\mathbf{c})\right) = \boldsymbol{\zeta}^\top U - \mathbf{f}^\top, \\
& \boldsymbol{\zeta}\in\mathbb{R}_+^{r_2} \\
& \lambda_h\in\mathcal{M}_+^F(C), && \forall\, h\in[H] \\
& \delta_t : \Omega\to\mathcal{M}_+^F((0,1]\times C) && \forall\, t\in[T].
\end{aligned}
\tag{LinearD$_\rho$}
$$

The proof of the next duality theorem, which is similar to that of Theorem 3 in Noyan and Rudolf (2013), is relegated to Appendix A. We remark that while our dual formulation is analogous to Haar's dual for semi-infinite linear programs (see, e.g., Bonnans and Shapiro, 2000), we obtain our strong duality result without any constraint qualification.

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

$$
\begin{aligned}
\mathrm{support}(\lambda_h) &\subset \{\mathbf{c} &&: \rho(\varphi(\mathbf{c},\Gamma\mathbf{z})) = \textstyle\int_0^1 \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\,\mu_h(\mathrm{d}\alpha), \\
& && \quad \rho(\varphi(\mathbf{c},\Gamma\mathbf{z})) = \rho(\varphi(\mathbf{c},\mathbf{Y}))\}, \quad h\in[H] \\
\mathrm{support}(\delta_t(\omega_i)) &\subset \{(\alpha,\mathbf{c}) &&: \varphi(\mathbf{c},\Gamma(\omega_i)\mathbf{z}) = A_t(\mathbf{c},\Gamma(\omega_i)\mathbf{z}), \\
& && \quad \varphi(\mathbf{c},\Gamma(\omega_i)\mathbf{z}) \le \mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\}, \quad i\in[n],\ t\in[T] \\
\mathrm{support}\Big(\sum_{h\in[H]}\left[\textstyle\int\frac{1}{\alpha}\,\mu_h(\mathrm{d}\alpha)\right]\times\lambda_h - \sum_{t\in[T]}\delta_t(\omega_i)\Big) &\subset \{(\alpha,\mathbf{c}) &&: \varphi(\mathbf{c},\Gamma(\omega_i)\mathbf{z}) \ge \mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\}, \quad i\in[n] \\
\boldsymbol{\zeta}^\top(U\mathbf{z}-\mathbf{u}) &= 0.
\end{aligned}
$$

We mention here that for the special case $\rho = \mathrm{CVaR}_\alpha$ the above theorem generalizes previous duality results for problems with linear scalarization functions (given in Noyan and Rudolf, 2013) to problems with min-affine scalarization functions.

4.1.1 Lagrangian Duality The duality results established in Theorem 4.1 admit a simple Lagrangian interpretation. As the preference constraints in (GeneralP$_\rho$) are indexed by the scalarization set C, measures on C are a natural choice to use as Lagrange multipliers. Accordingly, let us introduce the Lagrangian function $L : Z\times\mathcal{M}_+^F(C)\to\mathbb{R}$ defined by $L(\mathbf{z},\lambda) = \mathbf{f}^\top\mathbf{z} + \int_C \rho(\varphi(\mathbf{c},\Gamma\mathbf{z}))\,\lambda(\mathrm{d}\mathbf{c}) - \int_C \rho(\varphi(\mathbf{c},\mathbf{Y}))\,\lambda(\mathrm{d}\mathbf{c})$. Then the problem (LinearP$_\rho$) is equivalent to $\max\limits_{\mathbf{z}\in Z}\ \min\limits_{\lambda\in\mathcal{M}_+^F(C)} L(\mathbf{z},\lambda)$, while the corresponding Lagrangian dual problem is

$$
\min_{\lambda\in\mathcal{M}_+^F(C)}\ \max_{\mathbf{z}\in Z} L(\mathbf{z},\lambda). \tag{LagrangianD}
$$

The following strong duality result is a direct consequence of Theorem 4.1. The proof, which follows the same pattern as the upcoming derivation of Theorem 4.4 from Theorem 4.3, is omitted here for the sake of brevity.

THEOREM 4.2 *If* (LinearP$_\rho$) *has an optimal solution, then* (LagrangianD) *also has an optimal solution, and the optimal objective values coincide. The optimal solutions of* (LagrangianD) *are those measures* $\lambda^* \in \mathcal{M}_+^F(C)$ *that, for some optimal solution* $\mathbf{z}^*$ *of* (LinearP$_\rho$)*, satisfy the equations*

$$\int_C \rho(\varphi(\mathbf{c}, \Gamma\mathbf{z}^*))\,\lambda(\mathrm{d}\mathbf{c}) = \int_C \rho(\varphi(\mathbf{c}, \mathbf{Y}))\,\lambda(\mathrm{d}\mathbf{c}), \tag{20}$$

$$L(\mathbf{z}^*, \lambda^*) = \max_{\mathbf{z}\in Z} L(\mathbf{z}, \lambda^*). \tag{21}$$

4.2 Duality – SSD-based models Recalling Definition 2.1, an SSD constraint for scalar-valued random variables can be viewed either as a continuum of shortfall constraints or as a continuum of CVaR constraints. Dentcheva and Ruszczyński (2003) showed that, for SSD-constrained optimization problems formulated via the shortfall representation, a dual optimal solution can be interpreted as a risk-averse utility function (the so-called "*implied utility function*" of the benchmark). Analogously, the CVaR-based representation of the SSD relation leads to dual solutions interpreted as mixed-CVaR risk measures. A form of this result appears in Dentcheva and Ruszczyński (2006), where the dual solutions are viewed as rank-dependent dual utility functions (Yaari, 1987); a correspondence between Yaari's dual utility functions and spectral representations of mixed-CVaR risk measures is established, for example, in Rockafellar et al. (2002).

Dentcheva and Ruszczyński (2009) extend the above shortfall representation-based duality result to the case of multivariate SSD constraints with linear scalarization, showing that dual optimal solutions arise from certain operators that assign utility functions to scalarization vectors. In this section we generalize this result to SSD constraints with min-affine scalarizations, and provide an analogous statement for formulations that use the CVaR-representation of SSD. We point out that the previous strong duality results mentioned above require additional technical conditions such as the *uniform dominance condition* in Dentcheva and Ruszczyński (2003) and the *uniform inverse dominance condition* in Dentcheva and Ruszczyński (2006). In contrast, for the case of finite probability spaces our approach will establish strong duality without the need for constraint qualification. Let us begin by considering the following dual problem to (LinearP$_{\mathrm{SSD}}$):

$$\min \quad - \int\limits_{(0,1]\times C} \mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{Y}))\,\lambda(\mathrm{d}(\alpha, \mathbf{c})) + \mathbb{E}\left(\sum_{t\in[T]} \int\limits_C \mathbf{b}_t(\mathbf{c})\,[\Pi_C\delta_t](\mathrm{d}\mathbf{c})\right) + \boldsymbol{\zeta}^\top \mathbf{u} \tag{LinearD$_{\mathrm{SSD}}$}$$

$$\text{s.t.} \quad \mathbb{E}\left(\sum_{t\in[T]} \delta_t\right) = \lambda, \tag{22}$$

$$\sum_{t\in[T]} \delta_t(\omega_i) \le \int \frac{1}{\alpha}\,\lambda(\mathrm{d}(\alpha, \mathbf{c})), \qquad \forall\ i \in [n] \tag{23}$$

$$\mathbb{E}\left(\sum_{t\in[T]} \int\limits_C \mathbf{a}_t^\top(\mathbf{c})\Gamma\,[\Pi_C\delta_t](\mathrm{d}\mathbf{c})\right) = \boldsymbol{\zeta}^\top U - \mathbf{f}^\top, \tag{24}$$

$$\boldsymbol{\zeta} \in \mathbb{R}_+^{r_2}, \tag{25}$$

$$\lambda \in \mathcal{M}_+^F((0,1]\times C), \tag{26}$$

$$\delta_t : \Omega \to \mathcal{M}_+^F((0,1]\times C), \qquad \forall\ t \in [T]. \tag{27}$$

The proof of the next theorem essentially replicates that of Theorem 4.1, replacing the LP dual of (FiniteP$_\rho(\tilde{C})$) with that of (FiniteP$_{\mathrm{SSD}}(\tilde{C})$) to establish strong duality.

THEOREM 4.3 *The problem* (LinearP$_{\mathrm{SSD}}$) *has a finite optimum value if and only if* (LinearD$_{\mathrm{SSD}}$) *does, in which case the two optimum values coincide. In addition, a feasible solution* $\mathbf{z}$ *of* (LinearP$_{\mathrm{SSD}}$) *and a feasible solution* $(\lambda, \delta, \boldsymbol{\zeta})$ *of* (LinearD$_{\mathrm{SSD}}$) *are both optimal for their respective problems if and only if the following* *complementary slackness conditions hold:*

$$\begin{aligned}
\text{support}(\lambda) &\subset \{(\alpha,\mathbf{c}) &:& \quad \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z})) = \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{Y}))\}, \\
\text{support}(\delta_t(\omega_i)) &\subset \{(\alpha,\mathbf{c}) &:& \quad \varphi(\mathbf{c},\Gamma(\omega_i)\mathbf{z}) = A_t(\mathbf{c},\Gamma(\omega_i)\mathbf{z}), \\
&&& \quad \varphi(\mathbf{c},\Gamma(\omega_i)\mathbf{z}) \le \mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\}, \qquad i\in[n],\ t\in[T] \\
\text{support}(\textstyle\int \frac{1}{\alpha}\lambda(\mathrm{d}(\alpha,\mathbf{c})) - \sum_{t\in[T]}\delta_t(\omega_i)) &\subset \{(\alpha,\mathbf{c}) &:& \quad \varphi(\mathbf{c},\Gamma(\omega_i)\mathbf{z}) \ge \mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\}, \qquad i\in[n] \\
\boldsymbol{\zeta}^\top(U\mathbf{z}-\mathbf{u}) &= 0.
\end{aligned}$$

Remark 4.2 *While the shortfall-based primal formulation detailed in Remark 4.1 is equivalent to* (LinearP$_{\mathrm{SSD}}$)*, it gives rise to a different dual problem. In the shortfall-based dual the roles that the confidence levels* $\alpha \in (0,1]$ *and the corresponding functions* CVaR_α *play in* (LinearD$_{\mathrm{SSD}}$) *are assumed by the threshold levels* $\eta \in \mathbb{R}$ *and the corresponding negative expected shortfall functions* $X \mapsto -\mathbb{E}([\eta - X]_+)$*. The constraints* (22)-(23) *are replaced by the inequality* $\sum_t \delta_t(\omega_i) \le \lambda$ *for all* $i \in [n]$*, and an additional term* $\mathbb{E}\left(\sum_{t\in[T]} \int_{\mathbb{R}} \eta\,[\Pi_{\mathbb{R}}\delta_t](\mathrm{d}\eta)\right)$ *appears in the dual objective function.*

4.2.1 Lagrangian duality Along the lines of the framework established in Dentcheva and Ruszczyński (2009), given a scalarization function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and a scalarization set $C \subset \mathbb{R}^d_+$ we introduce two generators for the order $\preceq^{\varphi,C}_{(2)}$:

- Let $\mathcal{R}$ be the family of operators $\rho : \mathbf{c} \mapsto \rho_{\mathbf{c}}$ that assign a spectral acceptability functional $\rho_{\mathbf{c}}$ to each scalarization vector $\mathbf{c} \in \mathcal{C}$. For a finitely supported measure γ on C we define $\phi^R_{\rho,\gamma} : \mathcal{V}^d \to \mathbb{R}$ by $\phi^R_{\rho,\gamma}(\mathbf{X}) = \int_C \rho_{\mathbf{c}}(\varphi(\mathbf{c},\mathbf{X}))\,\gamma(\mathrm{d}\mathbf{c})$. Finally, let $\Phi^R = \{\phi^R_{\rho,\gamma}(\mathbf{X}) \ :\ \rho \in \mathcal{R}, \gamma \in \mathcal{M}^F_+(C)\}$.
- Let $\mathcal{U}$ be the family of operators $u : \mathbf{c} \mapsto u_{\mathbf{c}}$ that assign a concave non-decreasing utility function $u_{\mathbf{c}}$ to each scalarization vector $\mathbf{c} \in \mathcal{C}$. For a finitely supported measure γ on C we define $\phi^U_{u,\gamma} : \mathcal{V}^d \to \mathbb{R}$ by $\phi^U_{u,\gamma}(\mathbf{X}) = \mathbb{E}\left(\int_C u_{\mathbf{c}}(\varphi(\mathbf{c},\mathbf{X}))\,\gamma(\mathrm{d}\mathbf{c})\right)$. Finally, let $\Phi^U = \{\phi^U_{u,\gamma}(\mathbf{X}) \ :\ u \in \mathcal{U}, \gamma \in \mathcal{M}^F_+(C)\}$.

We next prove that the families Φ^R and Φ^U are indeed generators for $\preceq^{\varphi,C}_{(2)}$.

Proposition 4.2 *The three conditions below are equivalent for any two random vectors* $\mathbf{X}, \mathbf{Y} \in \mathcal{V}^d$*:*

(i) $\mathbf{X} \succeq^{\varphi,C}_{(2)} \mathbf{Y}$.

(ii) $\phi(\mathbf{X}) \ge \phi(\mathbf{Y})$ *for all* $\phi \in \Phi^R$.

(iii) $\phi(\mathbf{X}) \ge \phi(\mathbf{Y})$ *for all* $\phi \in \Phi^U$.

Proof. We outline the proof for (i) $\Leftrightarrow$ (ii); the proof for (i) $\Leftrightarrow$ (iii) is essentially identical. To show (ii) $\Rightarrow$ (i), for a scalarization vector $\mathbf{c} \in C$ and $\alpha \in (0,1]$ let $\delta_{\mathbf{c}}$ be the Dirac measure concentrated on $\mathbf{c}$, and let $\rho_\alpha \equiv \mathrm{CVaR}_\alpha$. Then (ii) implies that $\mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{X})) = \phi^R_{\rho_\alpha,\delta_{\mathbf{c}}}(\mathbf{X}) \ge \phi^R_{\rho_\alpha,\delta_{\mathbf{c}}}(\mathbf{Y}) \ge \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{Y}))$ holds for all $\mathbf{c} \in C$ and $\alpha \in (0,1]$, which is equivalent to the SSD condition $\mathbf{X} \succeq^{\varphi,C}_{(2)} \mathbf{Y}$ in (i). On the other hand, if the SSD condition holds, then for any γ and any $\rho \in \mathcal{R}$ with mixed-CVaR representations $\rho_{\mathbf{c}} = \int_0^1 \mathrm{CVaR}_\alpha\,\mu_{\mathbf{c}}(\mathrm{d}\alpha)$ the relation $\phi^R_{\rho,\gamma}(\mathbf{X}) = \int_C \int_0^1 \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{X}))\,\mu_{\mathbf{c}}(\mathrm{d}\alpha)\,\gamma(\mathrm{d}\mathbf{c}) \ge \int_C \int_0^1 \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{Y}))\,\mu_{\mathbf{c}}(\mathrm{d}\alpha)\,\gamma(\mathrm{d}\mathbf{c}) = \phi^R_{\rho,\gamma}(\mathbf{Y})$ follows directly by integrating the inequalities $\mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{X})) \ge \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\mathbf{Y}))$. □

To conclude this section we show that, for finite probability spaces, the strong duality established in Theorem 4.3 and Remark 4.2 fits naturally into the Lagrangian framework of Dentcheva and Ruszczyński (2009). Let us introduce the Lagrangian function $L(\mathbf{z},\phi) = \mathbf{f}^\top\mathbf{z} + \phi(\Gamma\mathbf{z}) - \phi(\mathbf{Y})$ and the following dual problems to (LinearP$_{\mathrm{SSD}}$):

$$\min_{\phi\in\Phi^R} \max_{\mathbf{z}\in Z} L(\mathbf{z},\phi), \qquad (\text{LagrangianD}^R)$$

$$\min_{\phi\in\Phi^U} \max_{\mathbf{z}\in Z} L(\mathbf{z},\phi). \qquad (\text{LagrangianD}^U)$$

According to Proposition 4.2 we have $L(\mathbf{z},\phi)\geq \mathbf{f}^\top\mathbf{z}$ for any feasible solution $\mathbf{z}$ of (LinearP$_{\text{SSD}}$) and any ϕ in the generator Φ^R or Φ^U. It immediately follows that weak duality holds: the optimal values of (LagrangianD^R) and (LagrangianD^U) are greater than or equal to the optimal value of (LinearP$_{\text{SSD}}$). The next theorem establishes strong duality and provide optimality conditions.

THEOREM 4.4 *If the primal problem* (LinearP$_{\text{SSD}}$) *has an optimal solution, then the dual problems* (LagrangianD^R) *and* (LagrangianD^U) *also have optimal solutions, and the optimal objective values coincide. A primal feasible solution* $\mathbf{z}^*$ *and a dual feasible solution* $\hat{\phi}$ *are simultaneously optimal if and only if they satisfy the equations*

$$\hat{\phi}(\Gamma\mathbf{z}^*)=\hat{\phi}(\mathbf{Y}), \tag{28}$$

$$L(\mathbf{z}^*,\hat{\phi})=\max_{\mathbf{z}\in Z}L(\mathbf{z},\hat{\phi}). \tag{29}$$

PROOF. Below we present the proof of the theorem for the dual problem (LagrangianD^R) using the strong duality between the CVaR-based formulations (LinearP$_{\text{SSD}}$) and (LinearD$_{\text{SSD}}$). The claim for (LagrangianD^U) follows analogously from the shortfall-based primal-dual pair detailed in Remarks 4.1 and 4.2.

We prove the theorem by showing the following three claims:

(i) If the equations (28)-(29) hold for some primal primal feasible $\mathbf{z}^*$ and dual feasible $\hat{\phi}$, then these solutions are simultaneously optimal, and the objective values coincide.

(ii) For any given optimal solution $\mathbf{z}^*$ of (LinearP$_{\text{SSD}}$) there exists a corresponding feasible solution ϕ^* of (LagrangianD^R) such that the equations (28) and (29) are satisfied for the choice $\hat{\phi}=\phi^*$.

(iii) If $\mathbf{z}^*$ is primal optimal and $\hat{\phi}$ is dual optimal, then they satisfy the equations (28)-(29).

Let us denote the optimum values of (LinearP$_{\text{SSD}}$) and (LagrangianD^R) by OPT_P and OPT_D, respectively. Since weak duality holds, we have $\text{OPT}_P\leq\text{OPT}_D$. On the other hand, (28)-(29) immediately imply

$$\text{OPT}_P\geq f^\top\mathbf{z}^*=L(\mathbf{z}^*,\hat{\phi})=\max_{\mathbf{z}\in Z}L(\mathbf{z},\hat{\phi})\geq\min_{\phi\in\Phi^R}\max_{\mathbf{z}\in Z}L(\mathbf{z},\phi)=\text{OPT}_D,$$

which proves claim (i).

To prove claim (ii), let us consider an optimal solution $\mathbf{z}^*$ of (LinearP$_{\text{SSD}}$). According to Theorem 4.3 there exists an optimal solution $(\lambda^*,\delta^*,\zeta^*)$ of (LinearD$_{\text{SSD}}$) with the same optimal value. We introduce the following notation and definitions:

- For a vector $\mathbf{c}\in C$ let the measure $\lambda^*_{\mathbf{c}}\in\mathcal{M}_+^F((0,1])$ be defined by $\lambda^*_{\mathbf{c}}(S)=\lambda^*(S\times\{\mathbf{c}\})$. Note that $\int_C\int_{(0,1]}h(\alpha,\mathbf{c})\,\lambda^*_{\mathbf{c}}(\mathrm{d}\alpha)\,[\Pi_C\lambda^*]\,(\mathrm{d}\mathbf{c})=\int_{(0,1]\times C}h(\alpha,\mathbf{c}),\lambda^*\,(\mathrm{d}(\alpha,\mathbf{c}))$ holds for all $h:(0,1]\times C\to\mathbb{R}$.
- Let us define the measure $\gamma^*\in\mathcal{M}_+^F(C)$ as $\gamma^*=\int\lambda^*_{\mathbf{c}}\,((0,1])\,[\Pi_C\lambda^*]\,(\mathrm{d}\mathbf{c})$.
- Note that $\mathbf{c}\in\text{support}(\gamma^*)$ holds if and only if $\lambda^*_{\mathbf{c}}\,((0,1])>0$; in this case we define the spectral risk measure $\rho^*_{\mathbf{c}}$ by $\rho^*_{\mathbf{c}}(V)=\frac{1}{\lambda^*_{\mathbf{c}}((0,1])}\int_{(0,1]}\text{CVaR}_\alpha(V)\,\lambda^*_{\mathbf{c}}(\mathrm{d}\alpha)$ for $V\in\mathcal{V}$.
- Finally, for the generator $\phi^*=\phi^R_{\rho^*,\gamma^*}\in\Phi^R$ and a random vector $\mathbf{X}\in\mathcal{V}^d$ we have

$$\begin{aligned}\phi^*(\mathbf{X})&=\int_C\rho^*_{\mathbf{c}}\left(\varphi(\mathbf{c},\mathbf{X})\right)\gamma^*(\mathrm{d}\mathbf{c})=\int_C\rho^*_{\mathbf{c}}\left(\varphi(\mathbf{c},\mathbf{X})\right)\lambda^*_{\mathbf{c}}\left((0,1]\right)\,[\Pi_C\lambda^*]\,(\mathrm{d}\mathbf{c})\\&=\int_C\int_{(0,1]}\text{CVaR}_\alpha(\phi(\mathbf{c},\mathbf{X}))\,\lambda^*_{\mathbf{c}}(\mathrm{d}\alpha)\,[\Pi_C\lambda^*]\,(\mathrm{d}\mathbf{c})=\int_{(0,1]\times C}\text{CVaR}_\alpha(\phi(\mathbf{c},\mathbf{X}))\,\lambda^*\left(\mathrm{d}(\alpha,\mathbf{c})\right).\end{aligned} \tag{30}$$

Using the above expression for ϕ^*, the condition (28) is an immediate consequence of the fact that, according to the first complementary slackness condition in Theorem 4.1, the equality $\text{CVaR}_\alpha(\phi(\mathbf{c},\Gamma\mathbf{z}^*))=\text{CVaR}_\alpha(\phi(\mathbf{c},\mathbf{Y}))$ is valid on the support of the measure λ^*. It now only remains to show that the condition (29) also holds. Noting that the third term of the Lagrangian function $L(\mathbf{z},\phi)=\mathbf{f}^\top\mathbf{z}+\phi(\Gamma\mathbf{z})-\phi(\mathbf{Y})$

does not depend on the decision vector $\mathbf{z}$, when we substitute $\hat{\phi} = \phi^*$ into (29), we can replace the expression $L(\mathbf{z}, \phi^*)$ by the truncated form $L(\mathbf{z}, \phi^*) + \phi^*(\mathbf{Y}) = \mathbf{f}^\top \mathbf{z} + \phi^*(\Gamma \mathbf{z})$. Furthermore, as $U\mathbf{z} \leq \mathbf{u}$ holds for all $\mathbf{z} \in Z$, and the Lagrange multiplier $\boldsymbol{\zeta}^*$ is nonnegative, the final complementary slackness condition in Theorem 4.1 implies $(\boldsymbol{\zeta}^*)^\top U\mathbf{z} \leq (\boldsymbol{\zeta}^*)^\top \mathbf{u} = (\boldsymbol{\zeta}^*)^\top U\mathbf{z}^*$. Therefore it suffices to show that the inequality $L(\mathbf{z}, \phi^*) + \phi^*(\mathbf{Y}) - (\boldsymbol{\zeta}^*)^\top U\mathbf{z} \leq L(\mathbf{z}^*, \phi^*) + \phi^*(\mathbf{Y}) - (\boldsymbol{\zeta}^*)^\top U\mathbf{z}^*$, which can be equivalently written as $\mathbf{f}^\top \mathbf{z} - (\boldsymbol{\zeta}^*)^\top U\mathbf{z} + \phi^*(\Gamma\mathbf{z}) \leq \mathbf{f}^\top \mathbf{z}^* - (\boldsymbol{\zeta}^*)^\top U\mathbf{z}^* + \phi^*(\Gamma\mathbf{z}^*)$, is valid for all $\mathbf{z} \in Z$. To this end, we will prove that the following chain of inequalities holds for any $\mathbf{z} \in Z$, and that all inequalities hold with equality for $\mathbf{z} = \mathbf{z}^*$.

$$
\begin{aligned}
&\mathbf{f}^\top \mathbf{z} - (\boldsymbol{\zeta}^*)^\top U\mathbf{z} + \phi^*(\Gamma\mathbf{z}) \\
&= -\,\mathbb{E}\left(\sum_{t\in[T]} \int_C \mathbf{a}_t^\top(\mathbf{c})\Gamma\mathbf{z}\,[\Pi_C \delta_t^*](\mathrm{d}\mathbf{c})\right) + \int_{(0,1]\times C} \mathrm{CVaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\,\lambda^*(\mathrm{d}(\alpha,\mathbf{c})) && (31)\\
&= -\sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} \mathbf{a}_t^\top(\mathbf{c})\Gamma_i\mathbf{z}\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) \\
&\quad + \int_{(0,1]\times C} \left(\mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z})) - \frac{1}{\alpha}\sum_{i\in[n]} p_i[\mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z})) - \varphi(\mathbf{c},\Gamma_i\mathbf{z})]_+\right) \lambda^*(\mathrm{d}(\alpha,\mathbf{c})) && (32)\\
&\leq \sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} b_t(\mathbf{c}) - \varphi(\mathbf{c},\Gamma_i\mathbf{z})\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) \\
&\quad + \sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} \mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z}))\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) \\
&\quad - \sum_{i\in[n]} p_i \int_{(0,1]\times C} [\mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z})) - \varphi(\mathbf{c},\Gamma_i\mathbf{z}))]_+ \frac{1}{\alpha}\lambda^*(\mathrm{d}(\alpha,\mathbf{c})) && (33)\\
&\leq \sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} b_t(\mathbf{c})\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) \\
&\quad + \sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} \mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z})) - \varphi(\mathbf{c},\Gamma_i\mathbf{z})\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) \\
&\quad - \sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} [\mathrm{VaR}_\alpha(\varphi(\mathbf{c},\Gamma\mathbf{z})) - \varphi(\mathbf{c},\Gamma_i\mathbf{z}))]_+\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) && (34)\\
&\leq \sum_{i\in[n]} p_i \sum_{t\in[T]} \int_{(0,1]\times C} b_t(\mathbf{c})\,[\delta_t^*(\omega_i)](\mathrm{d}(\alpha,\mathbf{c})) && (35)
\end{aligned}
$$

The equality (31) follows from (24) and (30). We then expand the expected value and projection operators in the first term, and substitute definition (4) of CVaR into the second term to arrive at (32). To verify (33), we apply the equality (22) to replace λ^* in the second term and note that by definition we have $\varphi(\mathbf{c}, \Gamma_i\mathbf{z}) \leq \mathbf{a}_t^\top(\mathbf{c})\Gamma_i\mathbf{z} + b_t(\mathbf{c})$. Furthermore, the second complementary slackness condition in Theorem 4.3 implies that equality holds for $\mathbf{z} = \mathbf{z}^*$. We obtain (34) by rearranging the first two terms and applying the inequality (23); equality for $\mathbf{z} = \mathbf{z}^*$ is guaranteed by the third complementary slackness condition. Finally, (35) is a consequence of the trivial inequality $x \leq [x]_+$. Equality will again hold for $\mathbf{z} = \mathbf{z}^*$ because the second complementary slackness constraint implies that on the support of $\delta_t^*(\omega_i)$ we have $\mathrm{VaR}_\alpha(\varphi(\mathbf{c}, \Gamma\mathbf{z})) - \varphi(\mathbf{c}, \Gamma_i\mathbf{z})) \geq 0$.

Finally, to prove claim (iii), let us consider a primal optimal solution $\mathbf{z}^*$ and a dual optimal solution $\hat{\phi}$. According to claim (ii) there exists some $\phi^* \in \Phi^R$ such that $\hat{\phi}(\Gamma\mathbf{z}^*) = \phi^*(\mathbf{Y})$ and $L(\mathbf{z}^*, \phi^*) = \max_{\mathbf{z}\in Z} L(\mathbf{z}, \hat{\phi})$. As the set Z is compact and the Lagrangian function L is continuous, there also exists some $\hat{\mathbf{z}}$ such that

$L(\hat{\mathbf{z}}, \hat{\phi}) = \max_{\mathbf{z}\in Z} L(\mathbf{z}, \hat{\phi})$. Since $\mathbf{z}^*$ is primal feasible, $\hat{\phi}(\Gamma\mathbf{z}^*) \geq \hat{\phi}(\mathbf{Y})$ holds. On the other hand, we have

$$\hat{\phi}(\mathbf{Y}) - \hat{\phi}(\Gamma\mathbf{z}^*) = f^\top\mathbf{z}^* - L(\mathbf{z}^*, \hat{\phi}) = L(\hat{\mathbf{z}}, \hat{\phi}) - L(\mathbf{z}^*, \hat{\phi}) = \max_{\mathbf{z}\in Z} L(\mathbf{z}, \hat{\phi}) - L(\mathbf{z}^*, \hat{\phi}) \geq 0,$$

which implies (28). Therefore, keeping in mind that $\mathbf{z}^*$ and $\hat{\phi}$ are both optimal and have the same objective value, (29) follows:

$$L(\mathbf{z}^*, \hat{\phi}) = f^\top\mathbf{z}^* + \hat{\phi}(\Gamma\mathbf{z}^*) - \hat{\phi}(\mathbf{Y}) = f^\top\mathbf{z}^* = \min_{\phi\in\Phi^R} \max_{\mathbf{z}\in Z} L(\mathbf{z}, \phi) = \max_{\mathbf{z}\in Z} L(\mathbf{z}, \hat{\phi}).$$

□

5. Solution Algorithm In this section we present an algorithm to solve the optimization problem (GeneralP$_\rho$). Throughout the section we assume that the risk measure $-\rho$ is given by a finite Kusuoka representation (5), while the scalarization function φ is given in the form (13). According to Theorem 3.1 it is sufficient to consider finitely many scalarization vectors to ensure that the multivariate preference relation (11) holds for $\mathbf{X} = G(\mathbf{z})$. However, as these scalarization vectors correspond to the vertices of a higher dimensional polyhedron, enumerating them is a hard problem and potentially leads to an exponential number of constraints. To address these concerns, we propose using a cut generation method.

We begin by solving the *relaxed master problem*, a relaxation of (GeneralP$_\rho$) obtained by replacing the set C with a (possibly empty) finite subset $\tilde{C}$. Under the linearity assumptions of Section 4 this master problem takes the form of the LP (LinearP$_\rho$), while in the general case the linear outcome term $\Gamma(\omega_i)\mathbf{z}$ and the linear constraint $U\mathbf{z} \leq \mathbf{u}$ can be replaced, respectively, by an arbitrary outcome mapping $G(\mathbf{z}, \omega_i)$ and a feasibility constraint of the form $\mathbf{z} \in Z$. We then iteratively augment the subset $\tilde{C}$ by adding to it scalarization vectors for which the corresponding risk constraint is violated. At each iteration, given an optimal solution $\mathbf{z}^*$ of the relaxed master problem, we attempt to find such a violating scalarization vector by solving the corresponding cut generation problem of the form

$$\min_{\mathbf{c}\in C} \rho(\varphi(\mathbf{c}, G(\mathbf{z}^*))) - \rho(\varphi(\mathbf{c}, \mathbf{Y})). \tag{CutGen}$$

If the optimal objective value is non-negative, it follows that $\mathbf{z}^*$ is an optimal solution of (GeneralP$_\rho$). Otherwise, Theorem 3.1 guarantees that there exists an optimal solution $\mathbf{c}^*$ which is a d-vertex of a certain polyhedron P. We provide a formal description of our proposed solution method in Algorithm 1.

Algorithm 1: Cut Generation Algorithm

`// Initialization`

Initialize a (possibly empty) set of scalarization vectors $\tilde{C} = \{\tilde{\mathbf{c}}_{(1)}, \ldots, \tilde{\mathbf{c}}_{(L)}\} \subset C$;

`// Main Loop`

Solve the *relaxed master problem*, obtained by replacing C with $\tilde{C}$ in (GeneralP$_\rho$).

if *the relaxed master problem is feasible* **then**

- Let $\mathbf{z}^*$ be an optimal solution;
- Given the random vector $\mathbf{X} = G(\mathbf{z}^*)$ solve the cut generation problem (CutGen);
- **if** *the optimal objective value of* (CutGen) *is negative* **then**
    - Let $\mathbf{c}^*$ be an optimal solution;
    - Given $\mathbf{c}^*$ find a d-vertex optimal solution $\tilde{\mathbf{c}}_{(L+1)}$ of (CutGen);
    - Set $\tilde{C} = \tilde{C} \cup \{\tilde{\mathbf{c}}_{(L+1)}\}$ and $L = L + 1$.
- **else** Optimality detected, break;

else Infeasibility detected, break;

The finite convergence of our algorithm follows directly from Theorem 3.1 and the continuity of coherent risk measures in the $\mathcal{L}^1$-norm (Ruszczyński and Shapiro, 2006). The proof of the following theorem is essentially identical to that of the analogous result for optimization with the multivariate polyhedral CVaR constraints (Theorem 4, Noyan and Rudolf, 2013).

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

As we have seen, the master problem is a non-linear program, which under appropriate linearity assumptions becomes the LP (FiniteP$_\rho(\tilde{C})$). The remainder of this section is dedicated to solving the typically more challenging cut generation problem. To keep our presentation accessible, we initially provide formulations for the case $\rho = \mathrm{CVaR}_\alpha$ in Section 5.1, and introduce some novel computationally efficient enhancements in Section 5.1.2. Using Kusuoka representations these results extend in a straightforward fashion from CVaR to all finitely representable measures, as detailed in Section 5.1.3.

5.1 Cut generation In this section we assume $\rho = \mathrm{CVaR}_\alpha$, and focus on solving (CutGen) given two d-dimensional random vectors $\mathbf{X} = G(\mathbf{z}^*)$ and $\mathbf{Y}$. As CVaR_α is defined for all finite probability spaces, we do not require the random vectors $\mathbf{X}$ and $\mathbf{Y}$ to be defined on the same space. Accordingly, let us denote their realizations and the corresponding probabilities by $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{y}_1, \dots, \mathbf{y}_m$ and $p_1, \dots, p_n$, $q_1, \dots, q_m$, respectively. The problem (CutGen) then takes the following form:

$$\min \quad \mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X})) - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \tag{36a}$$

$$\text{s.t.} \quad w_l \geq \eta - A_t(\mathbf{c}, \mathbf{y}_l), \qquad \forall\, l \in [m],\ t \in [T] \tag{36b}$$

$$\mathbf{w} \in \mathbb{R}^m_+, \quad \mathbf{c} \in C. \tag{36c}$$

Since $\mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$ is concave in $\mathbf{c}$, the cut generation problem requires the minimization of a concave objective function. The challenge therefore lies in expressing the risk $\mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$ associated with the decision-based random outcome $\mathbf{X}$. A similar structure appears when tackling optimization with the multivariate CVaR constraints based on linear (weighted-sum) scalarization; Noyan and Rudolf (2013) and Kucukyavuz and Noyan (2016) develop computationally tractable mixed integer programming (MIP) formulations for the arising cut generation problems. In particular, the computational study in Kucukyavuz and Noyan (2016) shows that their formulation, which utilizes an alternative representation of VaR, generally outperforms other approaches. We therefore adapt some ideas from the aforementioned papers to formulate (CutGen) as a MIP. We will also frequently make use of the following simple LP representations of minima and maxima:

OBSERVATION 5.1 *Given n values $q_1, \dots, q_n \in \mathbb{R}$ we have $v = \max_{i \in [n]} q_i$ if and only if the system* (37) *or, equivalently, the system* (39) *is feasible. Similarly, $v = \min_{i \in [n]} q_i$ can be represented either by* (38) *or by* (40).

- *Disjunction-based representations:*

$$\textit{For } \max \quad : \quad v \leq q_i + (1 - a_i)M, \quad v \geq q_i,\ \forall i \in [n], \quad \sum_{i \in [n]} a_i = 1, \quad \mathbf{a} \in \{0,1\}^n \tag{37}$$

$$\textit{For } \min \quad : \quad v \geq q_i - (1 - a_i)M, \quad v \leq q_i,\ \forall i \in [n], \quad \sum_{i \in [n]} a_i = 1, \quad \mathbf{a} \in \{0,1\}^n, \tag{38}$$

  where M is a sufficiently large constant ($M \geq \max_{i \in [n]} q_i - \min_{i \in [n]} q_i$). In both cases we have $v = q_{i^*}$, where $i^* \in [n]$ is the unique index with $a_{i^*} = 1$.

- *Convex combination-based representations:*

$$\textit{For } \max \quad : \quad v = \sum_{i \in [n]} q_i a_i, \quad v \geq q_i,\ \forall i \in [n], \quad \sum_{i \in [n]} a_i = 1, \quad \mathbf{a} \in [0,1]^n \tag{39}$$

$$\textit{For}\ \min \quad : \quad v = \sum_{i\in[n]} q_i a_i, \quad v \le q_i,\ \forall i \in [n], \quad \sum_{i\in[n]} a_i = 1, \quad \mathbf{a} \in [0,1]^n. \tag{40}$$

These representations remain valid if the vector $\mathbf{a}$ *is required to be binary. In this case we again have* $v = q_{i^*}$, *where* $i^* \in [n]$ *is the unique index with* $a_{i^*} = 1$.

We will initially use the disjunction-based representations, as these lead to a simpler (although not necessarily stronger) MIP formulation than their convex combination-based counterparts. We note that when q_i are fixed parameters, it is natural to prefer a convex combination-based representation as it does not introduce any additional binary variables. However, this advantage does not apply to our study because we are interested in calculating maxima and minima of decision variables, not parameters. In fact, we will see that it becomes necessary to enforce the binary restriction on the a_i variables in order to be able to linearize the bilinear terms $q_i a_i$. For the remainder of this section we assume that the values $A_t(\mathbf{c}, \mathbf{x}_i)$, and consequently the realizations $\varphi(\mathbf{c}, \mathbf{x}_i)$, are nonnegative for all $i \in [n]$, $t \in [T]$, $\mathbf{c} \in C$. This assumption is without loss of generality because adding a sufficiently large common constant term to all of the biaffine functions A_t does not affect our optimization problems (due to the translation equivariance of coherent risk measures).

Let us represent $\mathrm{VaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$ by the variable ϑ, the realization of $\varphi(\mathbf{c}, \mathbf{X})$ under scenario $i \in [n]$ by the variable λ_i, and the corresponding shortfall $[\vartheta - \lambda_i]_+$ by the variable v_i. We can now formulate (CutGen) as the following MIP:

$$\min \quad \vartheta - \frac{1}{\alpha}\sum_{i\in[n]} p_i v_i - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l \tag{41a}$$

$$\text{s.t.} \quad w_l \ge \eta - A_t(\mathbf{c}, \mathbf{y}_l), \qquad \forall\, l \in [m],\ t \in [T] \tag{41b}$$

$$\lambda_i \le A_t(\mathbf{c}, \mathbf{x}_i), \qquad \forall\, i \in [n],\ t \in [T] \tag{41c}$$

$$\lambda_i \ge A_t(\mathbf{c}, \mathbf{x}_i) - (1 - a_{it})\tilde{M}_{it}, \qquad \forall\, i \in [n],\ t \in [T] \tag{41d}$$

$$\sum_{t\in[T]} a_{it} = 1, \qquad \forall\, i \in [n] \tag{41e}$$

$$\sum_{i\in[n]} p_i \beta_i \ge \alpha, \tag{41f}$$

$$\sum_{i\in[n]} p_i \beta_i - \sum_{i\in[n]} p_i u_i \le \alpha - \epsilon, \tag{41g}$$

$$\vartheta \le \lambda_i + \beta_i M_{i\cdot}, \qquad \forall\, i \in [n] \tag{41h}$$

$$\vartheta \ge \lambda_i - (1 - \beta_i) M_{\cdot i}, \qquad \forall\, i \in [n] \tag{41i}$$

$$u_i \le \beta_i, \qquad \forall\, i \in [n] \tag{41j}$$

$$\sum_{i\in[n]} u_i = 1, \tag{41k}$$

$$\vartheta \le \lambda_i + (1 - u_i) M_{i\cdot}, \qquad \forall\, i \in [n] \tag{41l}$$

$$v_i \le \beta_i M_{i\cdot}, \qquad \forall\, i \in [n] \tag{41m}$$

$$v_i \le \vartheta - \lambda_i + (1 - \beta_i) M_{\cdot i}, \qquad \forall\, i \in [n] \tag{41n}$$

$$v_i \ge \vartheta - \lambda_i, \qquad \forall\, i \in [n] \tag{41o}$$

$$\mathbf{c} \in C, \quad \boldsymbol{\beta}, \mathbf{u} \in \{0,1\}^n, \quad \mathbf{a} \in \{0,1\}^{n\times T}, \tag{41p}$$

$$\mathbf{w} \in \mathbb{R}^m_+, \quad \mathbf{v}, \boldsymbol{\lambda} \in \mathbb{R}^n_+, \quad \vartheta, \eta \in \mathbb{R}. \tag{41q}$$

This formulation involves several new constants. Throughout this section we assume that all Big-M coefficients are sufficiently large to ensure that, whenever they are multiplied by a non-zero term, the constraints they appear in become redundant; we discuss the appropriate choice for the parameters $M_{i\cdot}$, $M_{\cdot i}$, and $\tilde{M}_{it}$ in Section 5.1.2 and Appendix B. Also, the positive constant ϵ should be sufficiently small to ensure that the constraint (41g) is equivalent to the strict inequality $\sum_{i\in[n]} p_i\beta_i - \sum_{i\in[n]} p_i u_i < \alpha$; this is clearly the case if $\epsilon < \min\{\alpha - \kappa : \kappa \in \mathcal{K}\cup\{0\}, \kappa < \alpha\}$ holds.

We now verify that the formulation (41) is indeed equivalent to the original (CutGen). To this end let us first note that, by Observation 5.1, constraints (41c)-(41e) mean that we have

$$\lambda_i = \min_{t\in[T]} A_t(\mathbf{c}, \mathbf{x}_i) = \varphi(\mathbf{c}, \mathbf{x}_i) \qquad \text{for all } i \in [n]. \tag{42}$$

Similarly, constraints (41i)-(41l) simply state that $\vartheta = \max\{\lambda_i : \beta_i = 1\} = \lambda_{i^*}$ holds, where i^* is the unique index with $u_{i^*} = 1$. If in place of the disjunctive representation (37) of the maximum we were to use the convex combination-based (39), the constraint (41l) would equivalently be replaced by $\vartheta = \sum_{i\in[n]} u_i\lambda_i$. Introducing the new variables $\zeta_i \in \mathbb{R}_+$ to represent the bilinear terms $u_i\lambda_i$, this new constraint can in turn be linearized as follows (keeping in mind that $\boldsymbol{\lambda}$ is assumed to be non-negative):

$$\vartheta = \sum_{i\in[n]} \zeta_i, \tag{43a}$$

$$\zeta_i \le u_i\tilde{M}_i, \qquad \forall\, i \in [n] \tag{43b}$$

$$\zeta_i \le \lambda_i, \qquad \forall\, i \in [n] \tag{43c}$$

$$\zeta_i \ge \lambda_i - (1 - u_i)\tilde{M}_i, \qquad \forall\, i \in [n]. \tag{43d}$$

We again point to Appendix B regarding the appropriate choice of the Big-M coefficients $\tilde{M}_i$. Theorem 3.1 in Kucukyavuz and Noyan (2016) immediately implies that constraints (41f)-(41k) in conjunction with (43) ensure $\vartheta = \mathrm{VaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$. Moreover, by (41h) we have $\lambda_i \ge \vartheta$ whenever $\beta_i = 0$, and similarly by (41i) we have $\lambda_i \le \vartheta$ whenever $\beta_i = 1$. Therefore, constraints (41m)-(41n) imply $v_i \le [\vartheta - \lambda_i]_+$. On the other hand, by (41o) and the non-negativity of $\mathbf{v}$ we have $v_i \ge [\vartheta - \lambda_i]_+$, therefore these inequalities hold with equality. Recalling (4), it follows that the term $\vartheta - \frac{1}{\alpha}\sum_{i\in[n]} p_i v_i$ in the objective function expresses $\mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$ as required.

Remark 5.1 *In order to highlight the structure of our problem and the equivalence to* (CutGen)*, the above MIP formulation features some redundant constraints. As the variables v_i are non-negative,* (41i) *is an immediate consequence of* (41n)*. The constraint* (41o) *is also redundant: we have seen that given* (41h) *and* (41i) *the constraints* (41m)-(41n) *are equivalent to the inequalities $v_i \le [\vartheta - \lambda_i]_+$. Since there are no other upper bounds on the v_i variables, these inequalities can be assumed to hold with equality at optimal solutions because all v_i appear in the objective with non-positive coefficients.*

Our implementation in Section 7 is based on a direct adaptation of the ideas in Kucukyavuz and Noyan (2016), which leads to a slightly more complex model where the constraints (41n)-(41o) *are equivalently expressed by introducing the auxiliary variables $\delta_i = \lambda_i + v_i - \vartheta$ along with the inequalities $0 \le \delta_i \le (1 - \beta_i)M_{\cdot i}$ for all $i \in [n]$.*

We mention that the MIP formulation (41) can be strengthened by adding the valid inequality

$$\vartheta \le \lambda_i + (\beta_i - u_i)M_i. \tag{44}$$

to replace the constraints (41h) and (41l), which are trivially implied by (44).

5.1.1 Alternative formulations In the MIP (41) the maximum $\vartheta = \max\{\lambda_i : \beta_i = 1\}$ and the minima $\lambda_i = \min_{t\in[T]} A_t(\mathbf{c}, \mathbf{x}_i)$ are both expressed using the disjunction-based representations in Observation 5.1, therefore we refer to this formulation as "*disjunctive-disjunctive*". It is of course also possible to use the equivalent convex combination-based representations (39) and (40) in place of (37) and (38), leading to four equivalent alternative formulations of our problem. More precisely, as we have seen above, one can replace the constraint (41l) by (43) to obtain a "*convex-disjunctive*" formulation. Similarly, to obtain a "*disjunctive-convex*" formulation from (41) we replace the constraint (41d) with $\lambda_i = \sum_{t\in[T]} a_{it} A_t(\mathbf{c}, \mathbf{x}_i),\ i \in [n]$, which, introducing new variables $\varrho_{it} \in \mathbb{R}_+$ to represent the bilinear terms $a_{it}A_t(\mathbf{c}, \mathbf{x}_i)$, can in turn be linearized as follows (under our non-negativity assumptions for the $A_t(\mathbf{c}, \mathbf{x}_i)$ values):

$$\lambda_i = \sum_{t\in[T]} \varrho_{it}, \qquad \forall\ i \in [n] \tag{45a}$$

$$\varrho_{it} \le A_t(\mathbf{c}, \mathbf{x}_i), \qquad \forall\ i \in [n],\ t \in [T] \tag{45b}$$

$$\varrho_{it} \le a_{it} \tilde{M}_i, \qquad \forall\ i \in [n],\ t \in [T] \tag{45c}$$

$$\varrho_{it} \ge A_t(\mathbf{c}, \mathbf{x}_i) - (1 - a_{it})\tilde{M}_{it}, \qquad \forall\ i \in [n],\ t \in [T]. \tag{45d}$$

Finally, the "*convex-convex*" formulation is obtained by replacing both (41l) with (43), and (41d) with (45). We will compare the computational performance of these four alternative formulations in Section 7.

We briefly mention here that, for the special case where each scenario has equal probability, it is possible to obtain simpler formulations that utilize the well-known risk envelope-based dual representation of CVaR (cf. Noyan and Rudolf, 2013), and, as proposed in Liu et al. (2015), the VaR representation by Kucukyavuz and Noyan (2016).

5.1.2 Enhancements to the cut generation method It is well-known that the choice of Big-M coefficients is crucial in obtaining stronger MIP formulations. For problem (41) one could simply set $M_{i\cdot} = M_{\cdot i} = \max_{k\in[n]} \tilde{M}_k$, where $\tilde{M}_k = \max_{\mathbf{c}\in C} \varphi(\mathbf{c}, \mathbf{x}_k)$; each of these values is easily calculated by solving an LP. On the other hand, the following more detailed formulas find the tightest possible coefficient values. We introduce the quantities

$$M_{ik} = \max_{\mathbf{c}\in C}\{\varphi(\mathbf{c}, \mathbf{x}_k) - \varphi(\mathbf{c}, \mathbf{x}_i)\} = \max_{\mathbf{c}\in C}\ \{\min_{t\in[T]} A_t(\mathbf{c}, \mathbf{x}_k) - \min_{t\in[T]} A_t(\mathbf{c}, \mathbf{x}_i)\} \quad \text{for } i, k \in [n], \tag{46}$$

and for $i \in [n]$ we set

$$M_{i\cdot} = \max_{k\in[n]} [M_{ik}]_+, \qquad M_{\cdot i} = \max_{k\in[n]} [M_{ki}]_+. \tag{47}$$

An added benefit of this approach lies in the fact that, as Kucukyavuz and Noyan (2016) demonstrate, when M_{ik} values are known to be non-positive, it is often possible to fix a large proportion of binary variables during the preprocessing stage, and further strengthen the MIP formulations by the addition of ordering constraints. These enhancements can dramatically improve the performance of the cut generation subproblem, which is confirmed by our computational study in Section 7. The convex maximization involved in calculating an exact M_{ik} value can be carried out by solving T linear programs (see Appendix B). However, it is possible to obtain high quality bounds with significantly lower computational effort whenever multiple scenarios lead to similar outcomes. Notably, this is often the case both for scenarios constructed from historical data, and for scenario sets generated by sampling from a continuous distribution. The idea that will allow us to exploit scenario proximity is the following: when multiple scenarios lead to similar outcome vectors, the scalarizations of these outcome vectors will also be close to one another. By consequence, if one replaces the scenarios i and k in (46) with a similar pair, it is possible to bound the resulting change in the Big-M values by using an appropriate measure of distance between the original and the new scenarios.

In order to keep our presentation compact, during the remainder of this section we restrict our attention to the case of the Chebyshev norm-based weighted worst-case scalarization where $T = d$ and $A_t(\mathbf{c}, \mathbf{x}) = c_t x_t$. Using this scalarization function the formula (46) becomes

$$M_{ik} = \max_{\mathbf{c}\in C}\ \{\min_{j\in[d]} c_j x_{jk} - \min_{j\in[d]} c_j x_{ji}\} \quad \text{for } i, k \in [n], \tag{48}$$

where x_{ji} denotes the realization of the outcome X_j under scenario i. For $j \in [d]$ let $\hat{M}_j = \max\{c_j\ :\ \mathbf{c} \in C\}$ and $\check{M}_j = \min\{c_j\ :\ \mathbf{c} \in C\}$ denote the maximal and minimal j-coordinates in the scalarization set, respectively.

Since C is a polyhedron, $\hat{M}_j$ and $\check{M}_j$ can be calculated by solving an LP, or, if the vertices of C are known, by simple enumeration. Given a random vector X we now introduce the following (non-symmetric) notion of distance between its realizations: $\Delta(i_1, i_2) = \max_{\ell\in[d]} \hat{M}_\ell [x_{\ell i_2} - x_{\ell i_1}]_+ - \check{M}_t [x_{\ell i_2} - x_{\ell i_1}]_-$. Let us observe that we have

$$c_j(x_{ji_2} - x_{ji_1}) \leq \Delta(i_1, i_2) \qquad \text{for all } i_1, i_2 \in [n],\ d \in [j],\ \mathbf{c} \in C. \tag{49}$$

If upper and lower bounds are available for a parameter M_{i_1,k_1}, the following lemma allows us to obtain bounds on another parameter M_{i_2,k_2}. The closer the pairs of realizations $\mathbf{x}_{i_1}, \mathbf{x}_{i_2}$ and $\mathbf{x}_{k_1}, \mathbf{x}_{k_2}$ are according to Δ, the tighter our new bounds will be.

Lemma 5.1 *The following inequalities hold for any indices $i_1, k_1, i_2, k_2 \in [n]$:*

$$M_{i_2k_2} \leq M_{i_1k_1} + \Delta(i_2, i_1) + \Delta(k_1, k_2), \tag{50}$$

$$M_{i_2k_2} \geq M_{i_1k_1} - \Delta(i_1, i_2) - \Delta(k_2, k_1). \tag{51}$$

Proof. We only prove (50), because by switching the roles of (i_1, k_1) and (i_2, k_2) the inequality (51) will immediately follow. Let $\mathbf{c}^*$ be maximizer in the defining equation (48) for $M_{i_2k_2}$, and let $j^* \in \arg\min_{j\in[d]} c^*_j x_{ji_2}$ be a corresponding minimizer of the second term. Furthermore, let $\hat{j} \in \arg\min_{j\in[d]} c^*_j x_{jk_1}$. Then we have:

$$M_{i_1k_1} = \max_{\mathbf{c}\in C}\{\min_{j\in[d]} c_j x_{jk_1} - \min_{j\in[d]} c_j x_{ji_1}\} \geq \min_{j\in[d]} c^*_j x_{jk_1} - c^*_{j^*} x_{j^*i_1} = c^*_{\hat{j}} x_{\hat{j}k_1} - c^*_{j^*} x_{j^*i_1},$$

$$M_{i_2k_2} = \max_{\mathbf{c}\in C}\{\min_{j\in[d]} c_j x_{jk_2} - \min_{j\in[d]} c_j x_{ji_2}\} = \min_{j\in[d]} c^*_j x_{jk_2} - c^*_{j^*} x_{j^*i_2} \leq c^*_{\hat{j}} x_{\hat{j}k_2} - c^*_{j^*} x_{j^*i_2}.$$

Combining the two inequalities above we finally obtain (50) as required:

$$M_{i_2k_2} - M_{i_1k_1} \leq c^*_{j^*}(x_{j^*i_1} - x_{j^*i_2}) + c^*_{\hat{j}}(x_{\hat{j}k_2} - x_{\hat{j}k_1}) \leq \Delta(i_2, i_1) + \Delta(k_1, k_2),$$

where the final inequality follows from the observation (49). □

For the purposes of our enhanced methods we need to determine, for every pair of scenarios $i, k \in [d]$, whether $M_{ik} \leq 0$ or $M_{ik} > 0$ holds; in the latter case we also require an upper bound on M_{ik}. To accomplish this, we first calculate the distances $\Delta(i, k)$ and initialize upper and lower bounds $M^+_{ik} = \infty$, $M^-_{ik} = -\infty$ for all $i, k \in [n]$. Then we repeat the following procedure until the answer is known for all pairs:

- Select a pair $i_1, k_1 \in [n]$ for which the answer is not yet known (i.e., for which $M^+_{ik} > 0$ and $M^-_{ik} \leq 0$), calculate $M_{i_1k_1}$ [solving the T separate LPs], and set $M^+_{i_1k_1} = M^-_{i_1k_1} = M_{i_1k_1}$.
- Use the newly calculated value $M_{i_1k_1}$ to update the bounds for all other pairs $i_2, k_2 \in [d]$ in accordance with (50) and (51):

$$M^+_{i_2k_2} \leftarrow \min\left(M^+_{i_2k_2},\ M^+_{i_1k_1} + \Delta(i_2, i_1) + \Delta(k_1, k_2)\right) \tag{52}$$

$$M^-_{i_2k_2} \leftarrow \max\left(M^-_{i_2k_2},\ M^-_{i_1k_1} - \Delta(i_1, i_2) - \Delta(k_2, k_1)\right). \tag{53}$$

5.1.3 Cut generation for finitely representable risk measures We now briefly outline how to modify the cut generation MIP (41) when the acceptability functional ρ is not CVaR_α, but is instead given by a representation of the form (5).

Let us first consider a spectral risk measure $-\rho$ given in the form (6). We introduce K copies of the variables ϑ and $\mathbf{v}$, with $\vartheta^{(j)}$ representing $\mathrm{VaR}_{\alpha_j}(\varphi(\mathbf{c}, \mathbf{X}))$, and $v^{(j)}_i$ representing the shortfall $[\vartheta^{(j)} - \lambda_i]_+$ for $j \in [K]$, $i \in [n]$. Similarly, $\eta^{(j)}$ and $\mathbf{w}^{(j)}$ are used in relation to the benchmark outcome Y, and we also introduce copies $\boldsymbol{\beta}^{(j)}$, $\mathbf{u}^{(j)}$ of our auxiliary binary variables. All constraints that involve the aforementioned variables are also replaced by K corresponding copies. Finally, to express $\rho(\varphi(\mathbf{c}, \mathbf{X})) - \rho(\varphi(\mathbf{c}, \mathbf{Y}))$, we replace the objective function in (41a) by

$$\sum_{j\in[K]} \mu_j \left[\vartheta^{(j)} - \frac{1}{\alpha} \sum_{i\in[n]} p_i v^{(j)}_i\right] - \sum_{j\in[K]} \mu_j \left[\eta^{(j)} - \frac{1}{\alpha} \sum_{l\in[m]} q_l w^{(j)}_l\right].$$

Let us now consider an arbitrary finitely representable risk measure of the form (5). We can rewrite the representation (5) as $\rho(V) = \min_{h\in[H]} \rho_h(V)$, where $-\rho_h(V) = -\sum_{j\in[K]} \mu_i^{(h)} \operatorname{CVaR}_{\alpha_j}(V)$ is a spectral risk measure for all $h \in [H]$. It is easy to see that the problem (CutGen) has a negative optimum if and only if at least one of the H problems $\min_{\mathbf{c}\in C} \rho_{h^*}(\varphi(\mathbf{c},\mathbf{X})) - \rho(\varphi(\mathbf{c},\mathbf{Y}))$ does, where $h^* \in [H]$. Therefore we can solve these H problems one by one until we either find a negative optimum (which provides a cut) or verify that they all have non-negative optima (which proves the optimality of our current decision). To solve the problem for any given $h^* \in H$, we again introduce copies of the variables ϑ, $\mathbf{v}$, η, $\mathbf{w}$, $\boldsymbol{\beta}$ and $\mathbf{u}$, along with the constraints involving them, in the same fashion as before. The objective function will now take the form

$$\sum_{j\in[K]} \mu_j^{(h^*)} \left[\vartheta^{(j)} - \frac{1}{\alpha} \sum_{i\in[n]} p_i v_i^{(j)} \right] - z,$$

with the additional constraint

$$z \le \sum_{j\in[K]} \mu_j^{(h)} \left[\eta^{(j)} - \frac{1}{\alpha} \sum_{l\in[m]} q_l w_l^{(j)} \right] \qquad \forall h \in [H]$$

to ensure that the new variable z represents $\rho(\varphi(\mathbf{c},\mathbf{Y}))$.

Remark 5.2 *Burgert and Rüschendorf (2006) introduce a class of scalar-valued risk measures for random vectors given by the representation*

$$-\hat{\rho}(\mathbf{X}) = -\inf_{\delta\in\Delta} \int_{\Delta} \rho_\delta(\varphi_\delta(\mathbf{X}))\, \nu(\mathrm{d}\delta), \tag{54}$$

where $-\rho_\delta : \mathcal{V} \to \mathbb{R}$ and $\varphi_\delta : \mathbb{R}^d \to \mathbb{R}$ are parametric families of risk measures and scalarization functions, respectively; the authors' primary focus is on instances featuring either linear or weighted worst-case scalarizations. Our methods can be naturally adapted to solve problems with benchmark constraints of the form $\hat{\rho}(G(\mathbf{z})) \ge \hat{\rho}(\mathbf{Y})$ for the important special case when $\rho_\delta \equiv \rho$ is a finitely representable risk measure, $\Delta = C$ is a scalarization polyhedron, and for $\delta = \mathbf{c} \in C$ we have $\varphi_\delta(\mathbf{x}) = \varphi(\mathbf{c},\mathbf{x})$ where φ is a min-biaffine mapping. In particular, the arising cut generation problems $\min_{\mathbf{c}\in C} \rho(\varphi(\mathbf{c},\mathbf{X})) \ge \hat{\rho}(\mathbf{Y})$ are a simpler variant of the ones discussed in this section, with the same left-hand side but a constant on the right-hand side. A similar approach can be taken for constraints featuring $\hat{\rho}$ of the general form (54) when the index set Δ is finite, all ρ_δ are finitely representable, and all φ_δ are min-affine.

5.1.4 Finding a *d*-vertex optimal solution The MIP formulations we have discussed so far provide a way to obtain an optimal solution $\mathbf{c}^*$ to the cut generation problem (16). We recall that Theorem 3.1 guarantees the existence of an optimal solution which is a d-vertex of a certain polyhedron P. We now describe how to use $\mathbf{c}^*$ to construct such a d-vertex optimal solution. Our approach, which highlights the connection between the risk envelope representation and the Kusuoka representation of coherent risk measures, provides an alternative to the subset-based method outlined in Noyan and Rudolf (2013).

According to Theorem 2.1 there exists a risk envelope $\mathcal{Q}$ that provides a dual representation of the form (7) for the acceptability functional ρ. We can assume without loss of generality that $\mathcal{Q} = \{Q \in \mathcal{V} \,:\, Q \ge 0,\ \mathbb{E}(Q) = 1,\ \mathbb{E}(QV) \ge \rho(V) \text{ for all } V \in \mathcal{V}\}$ is the *maximal* risk envelope. Since $\mathcal{Q}$ is a compact set, for any random variable V there exists some $\hat{Q} \in \mathcal{Q}$ such that

$$\rho(V) = \inf_{Q\in\mathcal{Q}} \mathbb{E}(QV) = \mathbb{E}(\hat{Q}V) \tag{55}$$

holds. The following results provide a way to obtain such a 'tight' element $\hat{Q}$ of the risk envelope for $V = \varphi(\mathbf{c}^*,\mathbf{X})$, as a first step toward finding the desired d-vertex solution.

Lemma 5.2 *Let Q be a random variable on a finite probability space, and let $-\rho$ be a finitely representable coherent risk measure given by its Kusuoka representation (5). The inequalities*

$$\mathbb{E}(QV) \ge \rho(V) \quad \textit{for all } V \in \mathcal{V} \tag{56}$$

hold if and only if we have $Q = \sum_{j\in[K]} R^{(j)}$ *for some non-negative random vector* $\mathbf{R} = (R^{(1)}, \ldots, R^{(K)})$ *satisfying the following conditions:*

$$\mathbb{E}(\mathbf{R}) \in \operatorname{conv}\left(\left\{\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(H)}\right\}\right), \tag{57}$$

$$R^{(j)} \leq \frac{1}{\alpha_j}\mathbb{E}(R^{(j)}) \quad \textit{for all } j \in [K], \tag{58}$$

where $\boldsymbol{\mu}^{(h)} = \left(\mu_1^{(h)}, \ldots, \mu_K^{(h)}\right)$ *for all* $h \in [H]$.

Proof. Let us denote the probabilities of the elementary events by $p_1, \ldots, p_n$. For random variables Q and V let us denote their corresponding realizations by $q_1, \ldots, q_n$ and $v_1, \ldots, v_n$. Taking into account the representation (5) it is easy to see that (56) holds if and only if the system

$$\sum_{i\in[n]} p_i q_i v_i < \sum_{j\in[K]} \mu_j^{(h)} \operatorname{CVaR}_{\alpha_j}(V) \quad \text{for all } h \in [H] \tag{59}$$

is infeasible. In accordance with (3) we can rewrite (59) as

$$\begin{aligned}
&\sum_{i\in[n]} p_i q_i v_i + \epsilon \leq \sum_{j\in[K]} \mu_j^{(h)} \left(\eta^{(j)} - \frac{1}{\alpha_j}\sum_{i\in[n]} p_i w_i^{(j)}\right), && \forall\ h \in [H]\\
&w_i^{(j)} \geq \eta^{(j)} - v_i, && \forall\ i \in [n],\ j \in [H]\\
&w_i^{(j)} \geq 0, && \forall\ i \in [n],\ j \in [H]\\
&\epsilon > 0.
\end{aligned}$$

By Farkas' lemma the above system is infeasible if and only if the following system is feasible:

$$\begin{aligned}
&p_i q_i \sum_{h\in[H]} a_h = \sum_{j\in[H]} b_{ij}, && \forall\ i \in [n]\\
&\sum_{i\in[n]} b_{ij} = \sum_{h\in[H]} \mu_j^{(h)} a_h, && \forall\ j \in [H]\\
&b_{ij} \leq \frac{p_i}{\alpha_j} \sum_{h\in[H]} \mu_j^{(h)} a_h, && \forall\ i \in [n],\ j \in [H]\\
&\sum_{h\in[H]} a_h = 1,\\
&\mathbf{a} \in \mathbb{R}_+^H, \quad \mathbf{b} \in \mathbb{R}_+^{n\times H}.
\end{aligned} \tag{60}$$

We introduce the random variables $R^{(j)}$ with realizations $r_i^{(j)} = \frac{b_{ij}}{p_i}$ for $i \in [n],\ j \in [H]$. Observing that $\mathbb{E}(R^{(j)}) = \sum_{i\in[n]} b_{ij}$ holds for all $j \in [H]$ we can now equivalently rewrite (60) as follows, which immediately proves our claim.

$$\begin{aligned}
&q_i = \sum_{j\in[H]} r_i^{(j)}, && \forall\ i \in [n]\\
&\mathbb{E}(R^{(j)}) = \sum_{h\in[H]} a_h \mu_j^{(h)}, && \forall\ j \in [H]\\
&r_i^{(j)} \leq \frac{1}{\alpha_j}\mathbb{E}(R^{(j)}), && \forall\ i \in [n],\ j \in [H]\\
&\sum_{h\in[H]} a_h = 1,\\
&\mathbf{a} \in \mathbb{R}_+^H,\\
&\mathbf{R} \in \mathcal{V}^H, \quad \mathbf{R} \geq 0.
\end{aligned}$$

□

Corollary 5.1 *A random variable $\hat{Q}$ belongs to the maximal risk envelope $\mathcal{Q}$ and attains the infimum in (55) for $V = \varphi(\mathbf{c}^*, \mathbf{X})$ if and only if its realizations $\hat{q}_1, \ldots, \hat{q}_n$ satisfy the following system of linear equations:*

$$\begin{aligned}
&\sum_{i\in[n]} p_i \hat{q}_i = 1, \\
&\sum_{i\in[n]} p_i \hat{q}_i \varphi(\mathbf{c}^*, \mathbf{x}_i) = \rho(\varphi(\mathbf{c}^*, \mathbf{X})), \\
&\hat{q}_i = \sum_{j\in[H]} r_i^{(j)}, && \forall\, i \in [n] \\
&\sum_{i\in[n]} p_i r_i^{(j)} = \sum_{h\in[H]} a_h \mu_j^{(h)}, && \forall\, j \in [H] \\
&r_i^{(j)} \le \frac{1}{\alpha_j} \sum_{i\in[n]} p_i r_i^{(j)}, && \forall\, i \in [n],\ j \in [H] \\
&\sum_{h\in[H]} a_h = 1, \\
&\mathbf{a} \in \mathbb{R}_+^H, \quad \mathbf{r} \in \mathbb{R}_+^{n\times H}, \quad \hat{\mathbf{q}} \in \mathbb{R}_+^n.
\end{aligned} \tag{61}$$

We can now find a d-vertex optimal solution to the cut generation problem (16) by solving an LP.

Proposition 5.1 *Suppose that $\mathbf{c}^*$ is an optimal solution to (16), and $\hat{Q}$ is a solution to the linear system (61). Let us define $\tau_i^* \in \arg\min_{t\in[T]} A_t(\mathbf{c}^*, \mathbf{x}_i)$ for $i \in [n]$, and consider the following LP:*

$$\begin{aligned}
\min \quad & \sum_{i\in[n]} p_i \hat{Q}(\omega_i) A_{\tau_i^*}(\mathbf{c}, \mathbf{x}_i) - z \\
\text{s.t.} \quad & z \le \sum_{j\in[K]} \mu_j^{(h)} \left(\eta^{(j)} - \frac{1}{\alpha_j} \sum_{i\in[n]} p_i w_i^{(j)} \right), && \forall\, h \in [H] \\
& w_i^{(j)} \ge \eta^{(j)} - A_t(\mathbf{c}, \mathbf{y}_i), && \forall\, i \in [n],\ j \in [K],\ t \in [T] \\
& w_i^{(j)} \ge 0, && \forall\, i \in [n],\ j \in [K] \\
& \mathbf{c} \in C.
\end{aligned} \tag{62}$$

If $(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}})$ is a vertex optimal solution to (62), then $\hat{\mathbf{c}}$ is also an optimal solution to (16), and it is a d-vertex of the polyhedron P introduced in Theorem 3.1.

Proof. Keeping in mind (55) we can reformulate the cut generation problem (16) as

$$\min_{\mathbf{c}\in C} \left\{ \mathbb{E}(\hat{Q}\varphi(\mathbf{c}, \mathbf{X})) - \rho(\varphi(\mathbf{c}, \mathbf{Y})) \right\}. \tag{63}$$

It is easy to see that $\mathbf{c}^*$ is an optimal solution of (63), and all optimal solutions of (63) are also optimal for (16). Using the representations (3), (5), and (13) we can express the minimum in (63) as the optimum of the following problem:

$$\min \quad \sum_{i\in[n]} p_i \hat{Q}(\omega_i) A_{\tau_i}(\mathbf{c}, \mathbf{x}_i) - z \tag{64a}$$

$$\text{s.t.} \quad z \le \sum_{j\in[K]} \mu_j^{(h)} \left(\eta^{(j)} - \frac{1}{\alpha_j} \sum_{i\in[n]} p_i w_i^{(j)} \right), \qquad \forall\, h \in [H] \tag{64b}$$

$$w_i^{(j)} \ge \eta^{(j)} - A_t(\mathbf{c}, \mathbf{y}_i), \qquad \forall\, i \in [n],\ j \in [K],\ t \in [T] \tag{64c}$$

$$w_i^{(j)} \ge 0, \qquad \forall\, i \in [n],\ j \in [K] \tag{64d}$$

$$A_{\tau_i}(\mathbf{c}, \mathbf{x}_i) \le A_t(\mathbf{c}, \mathbf{x}_i), \qquad \forall\, i \in [n],\ t \in [T] \tag{64e}$$

$$\tau_i \in [T], \qquad \forall\, i \in [n] \tag{64f}$$

$$\mathbf{c} \in C. \tag{64g}$$

As $\mathbf{c}^*$ is an optimal solution of (63), the above problem has an optimal solution of the form $(\mathbf{c}^*, z^*, \boldsymbol{\eta}^*, \mathbf{w}^*, \boldsymbol{\tau}^*)$. Note that, due to the presence of the index variables τ_i, (64) is not a linear program. However, by fixing the indices $\boldsymbol{\tau} = \boldsymbol{\tau}^*$ and dropping the constraints (64e)-(64f) we obtain the LP (62). Since the feasible set of this LP is the polyhedron P, for a vertex optimal solution $(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}})$ the vector $\hat{\mathbf{c}}$ is a d-vertex of P (i.e., $\hat{\mathbf{c}} = \mathbf{c}^{(\ell)}$ for some $\ell \in [N]$).

If we set $\hat{\boldsymbol{\tau}}$ such that $\hat{\tau}_i \in \arg\min_{t\in[T]} A_t(\hat{\mathbf{c}}, \mathbf{x}_i)$ for all $i \in [n]$, then $(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}, \hat{\boldsymbol{\tau}})$ satisfies the constraint (64e) and is thus a feasible solution of (64). We now show that it is in fact an optimal solution, which in turn implies that $\hat{\mathbf{c}}$ is an optimal solution of (16), completing our proof. Our claim follows from the chain of inequalities below.

$$\mathrm{OBF}_{(64)}(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}, \hat{\boldsymbol{\tau}}) \le \mathrm{OBF}_{(64)}(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}, \boldsymbol{\tau}^*) \tag{65}$$

$$= \mathrm{OBF}_{(62)}(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}) \tag{66}$$

$$\le \mathrm{OBF}_{(62)}(\mathbf{c}^*, z^*, \boldsymbol{\eta}^*, \mathbf{w}^*) \tag{67}$$

$$= \mathrm{OBF}_{(64)}(\mathbf{c}^*, z^*, \boldsymbol{\eta}^*, \mathbf{w}^*, \boldsymbol{\tau}^*). \tag{68}$$

To see that the inequality (65) holds, we first observe that by the definition of $\hat{\boldsymbol{\tau}}$ we have $A_{\tau_i^*}(\hat{\mathbf{c}}, \mathbf{x}_i) - A_{\hat{\tau}_i}(\hat{\mathbf{c}}, \mathbf{x}_i) \ge 0$ for all $i \in [n]$. Taking into account that both $\mathbf{p}$ and $\hat{Q}$ are non-negative, we obtain

$$\mathrm{OBF}_{(64)}(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}, \boldsymbol{\tau}^*) - \mathrm{OBF}_{(64)}(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}, \hat{\boldsymbol{\tau}}) = \sum_{i\in[n]} p_i \hat{Q}(\omega_i) \left(A_{\tau_i^*}(\hat{\mathbf{c}}, \mathbf{x}_i) - A_{\hat{\tau}_i}(\hat{\mathbf{c}}, \mathbf{x}_i) \right) \ge 0.$$

The equalities (66) and (68) reflect the fact that problem (62) is obtained from (64) by fixing $\boldsymbol{\tau} = \boldsymbol{\tau}^*$. Finally, (67) holds because $(\hat{\mathbf{c}}, \hat{z}, \hat{\boldsymbol{\eta}}, \hat{\mathbf{w}}, \hat{\boldsymbol{\tau}})$ and $(\mathbf{c}^*, z^*, \boldsymbol{\eta}^*, \mathbf{w}^*)$ are, respectively, an optimal and a feasible solution of (62). □

To summarize, given an optimal solution $\mathbf{c}^*$ of the cut generation problem (16), in step 8 of Algorithm 1 we can obtain a solution $\hat{\mathbf{c}}$ that is a d-vertex of P as follows:

1. Solve the linear system (61) to find a random variable $\hat{Q}$.
2. Set the indices $\tau_i^* \in \arg\min_{t\in[T]} A_t(\mathbf{c}^*, \mathbf{x}_i)$ for all $i \in [n]$.
3. Solve the LP (62) to obtain $\hat{\mathbf{c}}$.

Remark 5.3 *It is not necessary to start the procedure with an* optimal *vector* $\mathbf{c}^*$. *More precisely, it is easy to verify that our arguments in this section also imply the following slightly stronger statement: If the above steps are performed starting with an arbitrary scalarization vector* $\mathbf{c}^* \in C$, *then the resulting vector* $\hat{\mathbf{c}} \in C$ *is a d-vertex of* P *that satisfies* $\mathrm{OBF}_{(16)}(\hat{\mathbf{c}}) \le \mathrm{OBF}_{(16)}(\mathbf{c}^*)$. *In particular, if the risk constraint associated with* $\mathbf{c}^* \in C$ *is violated, then so is the risk constraint associated with* $\hat{\mathbf{c}} \in C$.

6. Optimization with Multivariate SSD Constraints In this section we briefly discuss how to solve SSD-constrained optimization problems of the form ($\mathrm{GeneralP}_{\mathrm{SSD}}$). Recalling Definition 2.1, the SSD relation $X \succeq_{(2)} Y$ can be expressed either via a collection of CVaR constraints, or via a collection of expected shortfall constraints. While both involve similar shortfall expressions, the latter approach is significantly more efficient because it utilizes fixed shortfall thresholds without the need to identify the value-at-risk. Accordingly, we can replace the SSD constraint in ($\mathrm{GeneralP}_{\mathrm{SSD}}$) by

$$\mathbb{E}([\varphi(\mathbf{c}, \mathbf{y}_l) - \varphi(\mathbf{c}, G(\mathbf{z}))]_+) \le \mathbb{E}([\varphi(\mathbf{c}, \mathbf{y}_l) - \varphi(\mathbf{c}, \mathbf{Y})]_+), \quad \forall\, l \in [m],\ \mathbf{c} \in C, \text{ or equivalently, by}$$

$$\sum_{i\in[n]} p_i \left[\varphi(\mathbf{c}, \mathbf{y}_l) - \varphi\left(\mathbf{c}, [G(\mathbf{z})](\omega_i)\right)\right]_+ \le \sum_{k\in[m]} q_k \left[\varphi(\mathbf{c}, \mathbf{y}_l) - \varphi(\mathbf{c}, \mathbf{y}_k)\right]_+, \quad \forall\, l \in [m],\ \mathbf{c} \in C, \tag{69}$$

and again employ a cut generation algorithm as outlined in Section 5 to solve the arising problem. We note that under our usual linearity assumptions the master problem becomes an LP, because (69) can be expressed with linear constraints by introducing auxiliary variables $\mathbf{v} \geq 0$, that are required to satisfy the inequalities $v_{il} \geq \varphi(\mathbf{c}, \mathbf{y}_l) - \varphi(\mathbf{c}, \Gamma_i \mathbf{z})$ for $i \in [n]$, $l \in [m]$, to represent the shortfalls on the left-hand side.

6.1 Cut generation The SSD constraint in (GeneralP$_{\text{SSD}}$) is violated if and only if the shortfall inequality in (69) is violated for some scalarization vector $\mathbf{c} \in C$ and at least one $l \in [m]$. Accordingly, to find such a violating $\mathbf{c} \in C$ we define a separate cut generation subproblem for each realization of the benchmark vector. The cut generation problem associated with the lth realization of the benchmark $\mathbf{Y}$ is given by

$$\min_{\mathbf{c}\in C} \sum_{k\in[m]} q_k[\varphi(\mathbf{c},\mathbf{y}_l) - \varphi(\mathbf{c},\mathbf{y}_k)]_+ - \sum_{i\in[n]} p_i[\varphi(\mathbf{c},\mathbf{y}_l) - \varphi(\mathbf{c},\mathbf{x}_i)]_+, \tag{CutGenSSD}$$

where $\mathbf{X} = G(\mathbf{z}^*)$ again denotes the random outcome associated with the current optimal solution $\mathbf{z}^*$ of the relaxed master problem. If *all* of the m cut generation subproblems have a non-negative optimum, then $\mathbf{z}^*$ is an optimal solution of (GeneralP$_{\text{SSD}}$), otherwise the optimal solution to *any* of the cut generation problems with a negative optimum will provide a new cut. As the scalarization function φ is min-biaffine, the problem (CutGenSSD) involves the minimization of a difference of convex functions. However, difference-of-convex programming methods only guarantee locally optimal solutions, which might lead to premature termination of the cut generation algorithm. We therefore develop a MIP formulation instead, along the lines of the work done for the case of linear scalarization in Homem-de-Mello and Mehrotra (2009) and Kucukyavuz and Noyan (2016).

Let us consider the cut generation problem associated with the lth realization of the benchmark $\mathbf{Y}$. Similarly to Section 5.1 we represent the ith realization of $\varphi(\mathbf{c}, \mathbf{X})$ by the variable λ_i, and we represent the kth realization of $\varphi(\mathbf{c}, \mathbf{Y})$ by μ_k (noting that only μ_l needs to be assigned as a decision variable). The corresponding shortfall quantities $[\mu_l - \lambda_i]_+$ and $[\mu_l - \mu_k]_+$ are represented by v_i and w_k, respectively. Then we can formulate the cut generation problem (CutGenSSD) as the following MIP:

$$\min \quad \sum_{k\in[m]} q_k w_k - \sum_{i\in[n]} p_i v_i$$

$$\text{s.t.} \quad \lambda_i \leq A_t(\mathbf{c}, \mathbf{x}_i), \qquad \forall\, i \in [n],\ t \in [T] \tag{70a}$$

$$\lambda_i \geq A_t(\mathbf{c}, \mathbf{x}_i) - (1 - a_{it})\tilde{M}_{it}, \qquad \forall\, i \in [n],\ t \in [T] \tag{70b}$$

$$\sum_{t\in[T]} a_{it} = 1, \qquad \forall\, i \in [n] \tag{70c}$$

$$\mu_l \leq A_t(\mathbf{c}, \mathbf{y}_l), \qquad \forall\, t \in [T] \tag{70d}$$

$$\mu_l \geq A_t(\mathbf{c}, \mathbf{y}_l) - (1 - \bar{a}_t)\bar{M}_{lt}, \qquad \forall\, t \in [T] \tag{70e}$$

$$\sum_{t\in[T]} \bar{a}_t = 1, \tag{70f}$$

$$v_i \leq \beta_i M_i, \qquad \forall\, i \in [n] \tag{70g}$$

$$v_i \leq \mu_l - \lambda_i + (1 - \beta_i)\hat{M}_i, \qquad \forall\, i \in [n] \tag{70h}$$

$$v_i \geq \mu_l - \lambda_i, \qquad \forall\, i \in [n] \tag{70i}$$

$$w_k \geq \mu_l - A_t(\mathbf{c}, \mathbf{y}_k), \qquad \forall\, k \in [m],\ t \in [T] \tag{70j}$$

$$\mathbf{c} \in C, \quad \boldsymbol{\beta} \in \{0,1\}^n, \quad \mathbf{a} \in \{0,1\}^{n\times T}, \quad \bar{\mathbf{a}} \in \{0,1\}^T, \tag{70k}$$

$$\boldsymbol{\lambda}, \mathbf{v} \in \mathbb{R}^n_+, \quad \mu_l \in \mathbb{R}_+, \quad \mathbf{w} \in \mathbb{R}^m_+. \tag{70l}$$

The newly introduced Big-M coefficients function similarly to their counterparts in earlier sections, and should be set to satisfy $\hat{M}_i \geq \max_{\mathbf{c}\in C} [\varphi(\mathbf{c},\mathbf{x}_i) - \varphi(\mathbf{c},\mathbf{y}_l)]_+$ and $\bar{M}_{lt} \geq \max_{\mathbf{c}\in C} A_t(\mathbf{c},\mathbf{y}_l)$. To verify that (70) is indeed equivalent to (CutGenSSD) it is sufficient to observe that the values of the variables $\boldsymbol{\lambda}$, μ_l, $\mathbf{v}$, and $\mathbf{w}$ are determined as intended by constraints (70a)-(70c), (70d)-(70f), (70g)-(70i), and (70j), respectively.

We remark that the minima in (70) were expressed via the disjunctive representation in Observation 5.1. Similarly to our arguments in Section 5.1.1, it is possible to obtain alternative formulations by utilizing convex combination-based representations instead.

6.2 Finding a d-vertex optimal solution We have seen in Section 5 that Theorem 3.1 implies the finite convergence of our cut generation method for risk measure-constrained problems. To see that Corollary 3.1 similarly implies finite convergence for the SSD-constrained problems, we will show (analogously to our arguments in Section 5.1.4) that, given a solution $\mathbf{c}^*$ of (CutGenSSD) with a negative objective value, it is possible to find another such solution $\hat{\mathbf{c}}$ which is also the d-vertex of the polyhedron $P(\varphi, C, \mathbf{Y})$ introduced in (15).

If we have $\mathrm{OBF}_{(\mathrm{CutGenSSD})}(\mathbf{c}^*) < 0$, then the relation $\varphi(\mathbf{c}^*, \mathbf{X}) \succeq_{(2)} \varphi(\mathbf{c}^*, \mathbf{Y})$ does not hold. According to Definition 2.1 the SSD relation is equivalent to the collection of CVaR_α-inequalities for all $\alpha \in \mathcal{K}$. Therefore, there exists some confidence level $\alpha^* \in \mathcal{K}$ for which the CVaR_{α^*}-constraint associated with the scalarization vector $\mathbf{c}^*$ is violated. If such an α^* value was known, then according to Remark 5.3 we could find a desired d-vertex solution $\hat{\mathbf{c}}$ as described in Section 5.1.4. However, the size of the set $\mathcal{K}$ of candidate α values is potentially exponential. We now show that we can restrict the candidate set to probabilities of level sets of $\varphi(\mathbf{c}^*, \mathbf{X})$ and $\varphi(\mathbf{c}^*, \mathbf{Y})$. The next lemma follows from the same argument as Proposition 1 part (ii) in Noyan and Rudolf (2013), therefore the proof is omitted.

Lemma 6.1 *Given two real-valued random variables V and W with respective realizations $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, the inequality $\mathrm{CVaR}_\alpha(V) \geq \mathrm{CVaR}_\alpha(W)$ holds for all $\alpha \in (0, 1]$ if and only if it holds for all $\alpha \in \hat{\mathcal{K}}$, where $\hat{\mathcal{K}} = (\mathrm{Range}(F_V) \cup \mathrm{Range}(F_W)) \setminus 0 = \{\mathcal{P}(V \leq v_i) \ : \ i \in [n]\} \cup \{\mathcal{P}(W \leq w_l) \ : \ l \in [m]\}$ denotes the family of the probabilities of level sets.*

Since the number of level sets is trivially bounded by the number of realizations, we can now find a suitable α^* by checking at most $n + m$ inequalities of the form $\mathrm{CVaR}_\alpha(\varphi(\mathbf{c}^*, \mathbf{X})) < \mathrm{CVaR}_\alpha(\varphi(\mathbf{c}^*, \mathbf{Y}))$.

7. Computational Study We designed our computational study with two main goals in mind. From a modeling point of view, we would like to demonstrate the value of our proposed approach. To this end, we analyze the effect of linear and weighted worst-case scalarizations on the solutions of a budget allocation problem. From a computational point of view, we aim to investigate the performance of our cut generation MIP formulations, and of the overall cut-generation-based algorithm.

All the optimization problems are modeled with the AMPL mathematical programming language, and solved using `CPLEX` 12.2 with its default set of options and parameters. All experiments were carried out on 4 threads of a Lenovo(R) workstation with two Intel® Xeon® 2.30 GHz CE5-2630 CPUs and 64 GB memory running on Microsoft Windows Server 8.1 Pro x64 Edition. All reported times are elapsed times, and the time limit is set to 3600 seconds.

7.1 An illustrative example - Homeland security budget allocation We test the impact of alternative scalarization functions, along with the computational effectiveness of our proposed methods, on a homeland security budget allocation (HSBA) problem presented in Hu et al. (2011). The description, given below, is adopted from the relevant existing studies (see, e.g., Kucukyavuz and Noyan, 2016). The main problem is to allocate a fixed budget to ten urban areas, which are classified in three groups: *higher risk* (New York), *medium risk* (Chicago, San Francisco Bay Area, Washington DC-MD-VA-WV, and Los Angeles-Long Beach); *lower risk* (Philadelphia PA-NJ, Boston MA-NH, Houston, Newark, and Seattle-Bellevue-Everett). The risk share of each area is based on four criteria: property losses, fatalities, air departures, and average daily bridge traffic. The penalty for allocations under the risk share is expressed by a budget misallocation function associated with each criterion, and these functions are used as the multiple random performance measures of interest.

In order to be consistent with our convention of preferring large values, we construct random outcome vectors of interest from the negative of the budget misallocation functions. More precisely, the random outcome vector $\mathbf{G}(\mathbf{z}) = (G_1(\mathbf{z}), \ldots, G_4(\mathbf{z}))$ associated with the allocation decision $\mathbf{z} \in Z = \{\mathbf{z} \in \mathbb{R}^{10}_+ : \sum_{j\in[10]} z_j = 1\}$ is given by

$$G_i(\mathbf{z}) = -\sum_{j\in[10]} [A_{ij} - z_j]_+, \quad i \in [4],$$

where $A : \Omega \to \mathbb{R}^{4\times 10}_+$ is a random risk share matrix with A_{ij} denoting the proportion of losses in urban area j relative to the total losses for criterion i.

Hu et al. (2011) model this HSBA problem using optimization under multivariate polyhedral SSD constraints based on linear (weighted-sum) scalarization. Two different benchmarks are considered, including one based on suggestions in the RAND report by Willis et al. (2005) and denoted by $\mathbf{G}(\mathbf{z}^R)$. As an alternative to restrictive SSD constraints, Noyan and Rudolf (2013) replace it with CVaR-based ones. We follow this line of research, but also consider the use of the weighted worst-case scalarization function, which leads to the following optimization model:

$$\max \quad \min_{\mathbf{c}\in C} \mathbb{E}(\varphi(\mathbf{c}, \mathbf{G}(\mathbf{z}))) \tag{71a}$$

$$\text{s.t.} \quad \mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{G}(\mathbf{z}))) \geq \mathrm{CVaR}_\alpha(\varphi(\mathbf{c}, \mathbf{G}(\mathbf{z}^R))), \quad \forall\, \mathbf{c} \in C \tag{71b}$$

$$\mathbf{z} \in Z, \tag{71c}$$

where we have either $\varphi(\mathbf{c}, \mathbf{G}(\mathbf{z})) = \sum_{j\in[d]} c_j G_j(\mathbf{z})$ or $\varphi(\mathbf{c}, \mathbf{G}(\mathbf{z})) = \min_{j\in[d]} c_j G_j(\mathbf{z})$. For the sake of brevity, in the remainder of Section 7 we restrict our attention to these CVaR-constrained problems, as they prove to be significantly more challenging computationally than their SSD-constrained counterparts. We follow the scheme described in Hu et al. (2011) to generate sets of equal-probability scenarios, and focus on the base case described in Noyan and Rudolf (2013). The scalarization polyhedron is of the form $C = \{\mathbf{c} \in \mathbb{R}^4_+ : \sum_{i\in[4]} c_j = 1,\ c_j \geq c^*_j - \frac{\theta}{3},\ j \in [4]\}$, where $\mathbf{c}^* = (1/4, 1/4, 1/4, 1/4)$ and $\theta = 0.25$, unless otherwise stated.

7.2 Effect of alternative scalarizations The choice of scalarization function has a significant impact on the optimal solutions. In Table 1 we compare the benchmark provided by the RAND corporation with four different solutions to the HSBA problem outlined above: The optimal solutions that are $\mathrm{CVaR}_{0.1}$-preferable to the RAND benchmark, using either linear or weighted worst-case scalarization (both in the objective and in the risk constraint), along with the optimal solutions when no risk constraints are enforced. All four models are solved for a variety of scalarization polyhedra parametrized by θ, ranging from a single scalarization vector at $\theta = 0$ to the unit simplex (i.e., all possible scalarizations) at $\theta = 0.75$.

As Figure 2 shows, for smaller scalarization sets the use of a risk constraint with weighted worst-case scalarization leads to more balanced allocations than linear scalarization. For large scalarization sets the two risk-constrained models both arrive at allocations close to the benchmark solution, which heavily favors regions with higher risk. The models without risk constraints provide similar solutions to their constrained counterparts for small scalarization sets. As the value of θ increases, the allocations by the unconstrained model with weighted worst-case objective remain essentially unchanged, and the solutions from the variant with linear scalarization in the objective approach these balanced allocations.

7.3 Computational performance As Table 1 shows, the use of weighted worst-case scalarization (and, by extension, a min-biaffine scalarization involving non-trivial minimization) leads to significantly more challenging optimization problems than linear scalarization. This can be seen clearly from the "Time" column, which gives the running times when using the *disjunctive-disjunctive* cut generation MIP (41) in conjunction with the enhancements outlined in Section 5.1.2. As the case of linear scalarization has been extensively investigated in Noyan and Rudolf (2013) and Kucukyavuz and Noyan (2016), in the remainder of this section we restrict our attention to problems involving weighted worst-case scalarization. It is worth noting that in

	Objective	New York	Medium risk	Lower risk	Time	# Cuts
θ	Unconstrained model ($C=\emptyset$) - Linear scalarization					
0	0.294269	0.486714	0.346263	0.167023	0.09	-
0.15	0.313312	0.375411	0.371391	0.253199	0.09	-
0.3	0.316379	0.330795	0.386874	0.282332	0.11	-
0.45	0.318490	0.330754	0.386604	0.282643	0.11	-
0.6	0.320476	0.330894	0.385346	0.283760	0.11	-
0.75	0.322364	0.331520	0.383328	0.285152	0.10	-
	Unconstrained model ($C=\emptyset$) - Weighted worst-case scalarization					
0	0.080591	0.331520	0.383328	0.285152	0.10	-
0.15	0.128946	0.331520	0.383328	0.285152	0.11	-
0.3	0.177300	0.331520	0.383328	0.285152	0.12	-
0.45	0.225655	0.331520	0.383328	0.285152	0.12	-
0.6	0.274010	0.331520	0.383328	0.285152	0.13	-
0.75	0.322364	0.331520	0.383328	0.285152	0.11	-
	RAND benchmark constraint - Linear scalarization					
0	0.294269	0.486941	0.346565	0.166494	0.7	1
0.15	0.317431	0.460309	0.334090	0.205602	0.4	4
0.3	0.359942	0.532964	0.303637	0.163399	1.1	5.2
0.45	0.406532	0.556044	0.306557	0.137399	1.5	6
0.6	0.457026	0.566205	0.317794	0.116001	2.2	6.4
0.75	0.507398	0.573105	0.322279	0.104615	2.8	7.6
	RAND benchmark constraint - Weighted worst-case scalarization					
0	0.080591	0.331520	0.383328	0.285152	0.2	1
0.15	0.143973	0.401400	0.340858	0.257742	4.4	4
0.3	0.262964	0.556136	0.308092	0.135773	5.2	4.2
0.45	0.352066	0.569477	0.322712	0.107811	35.6	4.8
0.6	0.427785	0.569783	0.322799	0.107418	328.2	4.8
0.75	0.503276	0.569783	0.322799	0.107418	323.8	4.8
RAND benchmark		0.586100	0.343100	0.070700		

Table 1: Optimal objective and allocations for alternative scalarizations, $n = 50$ nd $\alpha = 0.1$

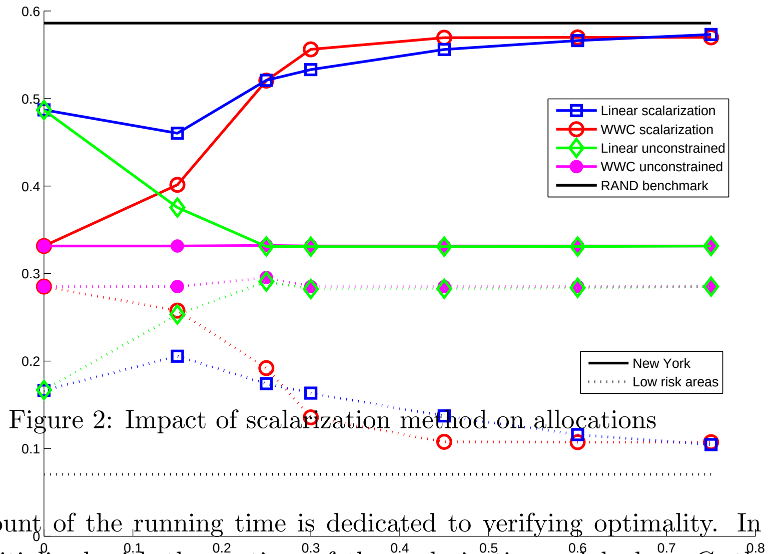

Figure 2: Impact of scalarization method on allocations

practice a significant amount of the running time is dedicated to verifying optimality. In particular, when the scalarization set is initialized with the vertices of the scalarization polyhedron C, the master problem often immediately finds an optimal solution, but the cut generation problem cannot always verify optimality within the prescribed time limit for instances with a large number of scenarios. As the last column of Table 1 illustrates, the four initial scalarization vectors (which collapse to a single vertex when $\theta = 0$) account for the majority of the cuts reported in the last column. This observation is confirmed by the performance data in Table 2; in fact, in the second half of the table (for $\theta = 0.25$) the number of cuts is not reported because the four initial cuts are always sufficient. Table 2 also shows the effect that the parameters n and α have on the computational performance, which is not detailed here as the findings are consistent with those drawn from the more extensive Tables 3-4 below.

In order to obtain apples-to-apples comparisons between the various MIP formulations, in Tables 3 and 4 we provide performance data when solving single instances of the cut generation subproblem (as opposed to executing the entirety of Algorithm 1). Optimality gaps are not reported, because the optimal objective value was zero for all instances that reached the time limit. Table 3 details the performance of the four alternative MIP formulations of (CutGen) as outlined in Section 5.1.1, for the fundamental case $\rho = \mathrm{CVaR}_\alpha$. While there is no clear-cut "winner" among the four cut generation MIPs, the *convex-disjunctive* (CD) and *disjunctive-disjunctive* (DD) formulations (i.e., those that use a disjunctive representation to express scalarization function) appear to consistently outperform the other two on the more challenging instances. Similarly to what was observed in Noyan and Rudolf (2013), confidence values α that are farther from the extreme values of 0 and 1 lead to increased combinatorial complexity when identifying VaR_α. This fact accounts for the significant performance decrease seen when the value of α is increased from 0.01 to 0.05. It is also clear that the main computational bottleneck for all formulations is the number of scenarios, which is again consistent with previous studies.

Table 4 shows the performance improvements that result from the enhanced preprocessing methods proposed in Kucukyavuz and Noyan (2016): bounding, variable fixing, and a class of valid *ordering inequalities* on the β variables. The bounding method introduces upper and lower bounds on the decision variable ϑ that represents $\mathrm{VaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$. The variable fixing method identifies realizations of $\mathbf{X}$ for which $\varphi(\mathbf{c}, \mathbf{x}_i)$ is guaranteed to be larger than $\mathrm{VaR}_\alpha(\varphi(\mathbf{c}, \mathbf{X}))$ for every $\mathbf{c} \in C$, and fixes the corresponding β_i variables to zero (a similar method to fix a_{it} variables is discussed in Appendix B). In addition, we include a set of ordering inequalities for the remaining β variables. Both the fixing of the β variables and the ordering inequalities rely on the bounding of big-M parameters; in particular, we need to identify pairs (i, k) of scenarios for which $M_{ik} \leq 0$ holds. While the preprocessing procedure is computationally demanding (unlike in the case of linear scalarizations), it benefits significantly from using Lemma 5.1 to bound the M_{ik} values. Among all pairs (i, k) for which the sign of M_{ik} is known, the proportion of those pairs that were identified using this lemma is shown in the final column of Table 4. The total preprocessing time increases polynomially with the number of scenarios, and, as the table confirms, constitutes only a small proportion of the total running time when solving the most challenging instances. We remark here that the preprocessing time is taken up almost entirely by the bounding of the M_{ik} values as described in Appendix B; in order to isolate the effects of the enhancements mentioned above, the same big-M values were also used for the non-enhanced variants of the model.

As discussed before, the problems become more complex for higher confidence levels. Table 4 shows that while for $\alpha = 0.01$ a significant majority of the binary β_i variables could be fixed during the preprocessing stage, this proportion drops to around 50% for $\alpha = 0.05$. However, this effect is partially compensated for by the larger number of valid ordering inequalities that can be found for the remaining variables. The previous observation that (CD) and (DD) are preferable to the other MIP formulations remains valid when we use enhanced preprocessing, with (DD) providing the best overall performance on larger instances. A comparison between Tables 3 and 4 shows that the enhancements lead to significant improvements in computational efficiency, and make problems with up to 300 scenarios consistently tractable under an hour on a single workstation.

$\theta = 0.45$				$\theta = 0.25$			
α	n	Time	# cuts	n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
0.025	50	2.6	4.6	50	1.0	2.0	5.6
	100	122.4	5.0	100	7.2	18.0	64.0
	150	971.8	4.6	150	34.0	219.8	326.4
0.05	50	7.2	4.8	200	98.2	665.6	972.6
	100	892.6	6.0	250	425.4	2791.2	3438.8
	150	4944.4	6.0	300	943.8	3972.4	4034.0

Table 2: Computational performance of the cut generation-based algorithm for a single CVaR constraint

8. Conclusions We introduced the class of min-biaffine functions, and showed for an extensive variety of scalarization functions from the deterministic multicriteria optimization literature that they belong to this new class. We then examined multicriteria stochastic optimization problems with benchmarking constraints based on min-biaffine scalarizations. Under appropriate conditions both SSD- and risk measure-constrained variants of such problems can be expressed as semi-infinite linear programs. We proved finiteness results guaranteeing that only a finite number of constraints are needed in these formulations, which in turn led to Haar-type strong duality results and optimality conditions without the need for constraint qualifications. We also showed that the dual problems have a natural Lagrangian interpretation. Under this interpretation, in the SSD-constrained case, the dual variables were seen to establish an assignment of either risk-averse utility functions or coherent risk measures to the scalarization vectors. This generalizes important previous duality results for the multivariate SSD-constrained stochastic optimization.

The finiteness result also provides the basis of a cut generation algorithm to solve our problems. Although the algorithm itself is largely straightforward, it is non-trivial to ensure that the cuts produced by cut generation subproblems are of the type indicated in the finiteness theorem. To overcome this difficulty we proved, and made use of, a lemma of independent interest that characterizes the elements of a risk envelope in terms of

	Convex-convex		Convex-disjunctive		Disjunctive-convex		Disjunctive-disjunctive	
	Time	B&B Nodes	Time	B&B Nodes	Time	B&B Nodes	Time	B&B Nodes
n	$\alpha = 0.01$							
50	1.0	0.9	1.0	1.5	1.0	1.4	0.9	1.6
100	3.5	2.1	9.5	8.9	2.8	2.5	11.7	17.9
150	43.9	31.9	78.1	54.6	5.8	3.6	36.1	25.7
200	874.0	597.4	180.2	92.5	760.1	596.3	259.9	232.6
250	†	1363.5	†2377.6	1377.9	†	2005.8	†3493.4	2697.1
300	†	1331.0	†	1812.2	†	1800.2	†	2440.7
n	$\alpha = 0.05$							
50	10.9	13.6	6.1	6.5	12.5	19.6	5.1	7.0
100	649.2	817.7	66.7	59.3	1002.3	1514.1	70.3	79.0
150	1837.8	1322.2	1181.5	1115.9	1816.7	1741.0	1177.8	1482.1
200	†	1518.5	†	3067.3	†	2351.6	†3471.8	3435.7
250	†	1708.2	†	2918.0	†	1550.8	†	3808.5
300	†	556.7	†	1884.9	†	1230.6	†	2216.9

Table 3: Computational performance of the alternative MIPs for (CutGen), $\theta = 0.25$

†: At least one instance hits time limit with integer feasible solution
B&B Nodes are reported in thousands.

the Kusuoka representation of a coherent risk measure. The main challenge of our algorithm lies in formulating and solving cut generation subproblems. By utilizing the structure of min-biaffine functions we were able to build on recent advances made for similar problems with linear scalarizations, leading to a number of alternative cut generation MIP formulations. In addition to adapting some existing enhancements to these MIPs, we further improved computational performance by using a novel idea to exploit similarities between outcome scenarios. We implemented and tested our methods on a well-known homeland security budget allocation problem, and found tractable solutions for moderate-sized instances with several hundred scenarios. Finally, we explored the effect of various parameters on the computational difficulty, and examined the impact of the chosen scalarization methods on the optimal solutions.

Acknowledgments Both authors acknowledge the support from Bilim Akademisi - The Science Academy, Turkey, under the BAGEP program. The first author has also been supported in part by The Scientific and Technological Research Council of Turkey (TUBITAK) under grant #115M560.

References

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.

Armbruster, B. and Luedtke, J. (2015). Models and formulations for multivariate dominance constrained stochastic programs. *IIE Transactions*, 47(1).

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

Ben Abdelaziz, F. (2012). Solution approaches for the multiobjective stochastic programming. *European Journal of Operational Research*, 216(1):1–16.

Bonnans, J. F. and Shapiro, A. (2000). *Perturbation analysis of optimization problems.* Springer Series in

	Convex-convex w/ enhancements		Convex-disjunctive w/ enhancements		Disjunctive-convex w/ enhancements		Disjunctive-disjunctive w/ enhancements		Binary variables remaining after fixing (%)		Number of ordering inequalities	Preprocessing time	Additional (%) identified pairs
	Time	B&B Nodes	Time	B&B Nodes	Time	B&B Nodes	Time	B&B Nodes	β_i	a_{it}			
n	$\alpha = 0.01$												
50	0.3	0	0.2	0	0.3	0	0.2	0	25.0	73.8	0	0.5	10.7
100	0.4	0	0.3	0	0.4	0	0.2	0	9.0	77.6	0	7.5	12.4
150	2.4	0.6	3.6	4.1	2.4	0.4	2.5	2.3	30.7	80.5	29.5	28.0	10.6
200	2.7	0.7	2.4	1.4	2.6	0.3	2.7	2.4	20.3	76.4	34.0	104.5	13.3
250	†1804	1465.8	183.2	163.1	†1978	1674.3	161.5	162.9	25.8	74.9	40.5	317.5	11.8
300	626.9	397.7	683.0	577.2	†1805	1283.9	308.9	241.8	24.8	73.7	52.0	771.5	11.0
n	$\alpha = 0.05$												
50	3.1	4.6	1.8	4.7	1.1	2.3	1.1	2.5	51.0	76.8	17.0	0.5	10.4
100	18.6	23.8	24.1	35.8	18.6	21.0	12.9	18.6	41.5	76.0	64.5	8.0	12.6
150	508.9	428.7	193.5	227.6	838.0	829.7	172.5	226.9	46.7	77.7	135.0	31.0	11.1
200	†2786	1701.7	439.0	376.9	1364.3	969.8	312.2	296.6	52.8	76.5	421.0	121.5	13.3
250	†	1665	1492.4	1339.3	†	1966	1300.6	1175.1	51.8	77.2	468.0	331.0	11.5
300	†	1557	†2691	1990.2	†2862	1448.4	2497.2	2196.0	52.3	74.6	664.0	624.5	10.9

Table 4: Computational performance of the enhanced MIPs (with fixing, bounding, ordering inequalities) for (CutGen), $\theta = 0.25$

B&B Nodes are reported in thousands.

Operations Research. Springer-Verlag, New York.

Burgert, C. and Rüschendorf, L. (2006). Consistent risk measures for portfolio vectors. *Insurance: Mathematics and Economics*, 38(2):289 – 297.

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). Inverse stochastic dominance constraints and rank dependent expected utility theory. *Mathematical Programming*, 108(2):297–311.

Dentcheva, D. and Ruszczyński, A. (2009). Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117:111–127.

Dentcheva, D. and Wolfhagen, E. (2015). Optimization with multivariate stochastic dominance constraints. *SIAM Journal on Optimization*, 25(1):564–588.

Dentcheva, D. and Wolfhagen, E. (2016). Two-stage optimization problems with multivariate stochastic order constraints. *Mathematics of Operations Research*, 41(1):1–22.

Ehrgott, M. (2005). *Multicriteria Optimization*. Springer, Berlin.

Gutjahr, W. J. and Pichler, A. (2013). Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research*, 236(2):475–499.

Haskell, W., Xu, H., Chao, Q., and Zhiyue, Y. (2016). Online risk-aware optimization. Technical report. https://www.ise.nus.edu.sg/staff/haskell/pdf/Haskell_RiskOnline.pdf.

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. (2011). Risk-adjusted budget allocation models with application in homeland security. *IIE Transactions*, 43(12):819–839.

Hu, J., Homem-de Mello, T., and Mehrotra, S. (2012). Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming*, 133(1-2):171–201.

Hu, J. and Mehrotra, S. (2012). Robust and stochastically weighted multiobjective optimization models and reformulations. *Operations Research*, 60(4):936–953.

Kaya, C. and Maurer, H. (2014). A numerical method for nonconvex multi-objective optimal control problems. *Computational Optimization and Applications*, 57(3):685–702.

Kucukyavuz, S. and Noyan, N. (2016). Cut generation for optimization problems with multivariate risk constraints. *Mathematical Programming*, 159(1):165–199.

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

Liu, X., Kucukyavuz, S., and Noyan, N. (2015). Robust multicriteria risk-averse stochastic programming models. http://www.optimization-online.org/DB_FILE/2015/10/5157.pdf.

Luque, M., Miettinen, K., Ruiz, A. B., and Ruiz, F. (2012). A two-slope achievement scalarizing function for interactive multiobjective optimization. *Computers & Operations Research*, 39(7):1673–1681.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*, volume 12 of *International Series in Operations Research and Management Science*. Kluwer Academic Publishers, Dordrecht.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks.* John Wiley & Sons, Chichester.

Nikulin, Y., Miettinen, K., and Mäkelä, M. M. (2012). A new achievement scalarizing function based on parameterization in multiobjective optimization. *OR Spectrum*, 34(1):69–87.

Noyan, N. (2012). Risk-averse two-stage stochastic programming with an application to disaster management. *Computers and Operations Research*, 39(3):541–559.

Noyan, N. and Rudolf, G. (2013). Optimization with multivariate conditional value-at-risk-constraints. *Operations Research*, 61(4):990–1013.

Noyan, N. and Rudolf, G. (2015). Kusuoka representations of coherent risk measures in general probability spaces. *Annals OR*, 229(1):591–605.

Ogryczak, W. and Ruszczyński, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM journal of optimization*, 13(2):60–78.

Pflug, G. and Wozabal, D. (2009). Ambiguity in portfolio selection. In Dempster, A. . H. ., Pflug, G., and Mitra, G., editors, *Quantitative Fund Management*, Chapman & Hall/CRC Financial Mathematics Series, pages 377–391. Boca Raton, FL: CRC Press.

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.

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.

Rockafellar, R. T., Uryasev, S. P., and Zabarankin, M. (2002). Deviation measures in risk analysis and optimization. *University of Florida, Department of Industrial & Systems Engineering Working Paper*, (2002-7).

Ruszczyński, 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.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on stochastic programming: modeling and theory.* The society for industrial and applied mathematics and the mathematical programming society, Philadelphia, USA.

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation, and Application.* John Wiley & Sons, New York.

Wierzbicki, A. (1999). Reference point approaches. In Gal, T., Stewart, T., and Hanne, T., editors, *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*, pages 9.1–9.39. Kluwer Academic Publishers, Boston.

Wierzbicki, A. P. (1980). The use of reference objectives in multiobjective optimization. In Fandel, G. and

Gal, T., editors, *Multiple Criteria Decision Making Theory and Application: Proceedings of the Third Conference Hagen/Königswinter, West Germany, August 20–24, 1979*, pages 468–486. Springer Berlin Heidelberg, Berlin, Heidelberg.

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.

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

Appendix A. Proof of Theorem 4.1 We first recall the simple facts that for a finitely supported measure $\lambda \in \mathcal{M}_+^F(C)$ and a function $g : C \to \mathbb{R}$ we have

$$\text{support}(\lambda) = \{\mathbf{c} \in C \ : \ \lambda(\{\mathbf{c}\}) > 0\} \qquad \text{and} \qquad \int_C g(\mathbf{c})\,\lambda(\mathrm{d}\mathbf{c}) = \sum_{\mathbf{c}\in\text{support}(\lambda)} g(\mathbf{c})\lambda(\{\mathbf{c}\}). \tag{72}$$

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

$$\begin{aligned}
\min \quad & -\sum_{h\in[H]}\sum_{\ell\in[L]} \lambda_\ell^{(h)} \rho(\varphi(\tilde{\mathbf{c}}^{(\ell)}, \mathbf{Y})) + \sum_{j\in[K]}\sum_{\ell\in[L]}\sum_{t\in[T]}\sum_{i\in[n]} p_i \delta_{\ell,t,i}^{(j)} b_t(\tilde{\mathbf{c}}^{(\ell)}) + \mathbf{u}^\top \boldsymbol{\zeta} \\
\text{s.t.} \quad & \sum_{t\in[T]}\sum_{i\in[n]} p_i \delta_{\ell,t,i}^{(j)} = \sum_{h\in[H]} \mu_j^{(h)} \lambda_\ell^{(h)}, && \forall\, j \in [K],\ \ell \in [L] \\
& \sum_{t\in[T]} \delta_{\ell,t,i}^{(j)} \le \frac{1}{\alpha_j} \sum_{h\in[H]} \mu_j^{(h)} \lambda_\ell^{(h)}, && \forall\, j \in [K],\ \ell \in [L],\ i \in [n] \\
& \sum_{j\in[K]}\sum_{\ell\in[L]}\sum_{t\in[T]}\sum_{i\in[n]} p_i \mathbf{a}_t^\top(\tilde{\mathbf{c}}^{(\ell)}) \Gamma(\omega_i) = \boldsymbol{\zeta}^\top U - \mathbf{f}^\top, \\
& \boldsymbol{\lambda} \in \mathbb{R}_+^{H\times L}, \quad \boldsymbol{\delta} \in \mathbb{R}_+^{K\times L\times T\times n}, \quad \boldsymbol{\zeta} \in \mathbb{R}_+^{r_2}.
\end{aligned} \tag{FiniteD$_\rho(\tilde{C})$}$$

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 $\delta_{\ell,t,i}^{(j)}$. The dual variable $\boldsymbol{\lambda}$ naturally defines a measure $\lambda_h \in \mathcal{M}_+^F(C)$ supported on the finite set $\tilde{C}$ with $\lambda_h\left(\{\tilde{\mathbf{c}}^{(\ell)}\}\right) = \lambda_\ell^{(h)}$ for all $h \in H$. Similarly, the dual variable $\boldsymbol{\delta}$ defines a random measure $\delta_t : \Omega \to \mathcal{M}_+^F((0,1] \times C)$ for all $t \in [T]$. The measures δ_t are given by $\delta_t(\omega_i)\left(\{(\alpha_j, \tilde{\mathbf{c}}_{(\ell)})\}\right) = \delta_{\ell,t,i}^{(j)}$ and are supported on $\mathcal{K} \times \tilde{C}$, where $\mathcal{K}$ is the finite set of probabilities introduced in (1).

Keeping in mind (72), it follows that for any feasible solution $(\boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\zeta})$ of (FiniteD$_\rho(\tilde{C})$) we have a corresponding feasible solution $(\boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\zeta})$ of (LinearD$_\rho$), with the same objective value. Conversely, for a feasible solution $(\lambda, \delta, \boldsymbol{\zeta})$ of (LinearD$_\rho$) and a finite set $\tilde{C} = \{\tilde{\mathbf{c}}^{(1)}, \dots, \tilde{\mathbf{c}}^{(L)}\}$ that contains $\text{support}(\lambda) = \bigcup_{h\in[H]} \text{support}(\lambda_h)$ we can define a feasible solution $(\boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\zeta})$ of (FiniteD$_\rho(\tilde{C})$) with the same objective value by setting $\lambda_\ell^{(h)} = \lambda_h\left(\{\tilde{\mathbf{c}}^{(\ell)}\}\right)$ and $\delta_{\ell,t,i}^{(j)} = \delta_t(\omega_i)\left(\{(\alpha_j, \tilde{\mathbf{c}}^{(\ell)})\}\right)$.

We now establish weak duality. Let $\mathbf{z}$ and $(\lambda, \delta, \boldsymbol{\zeta})$ be feasible solutions of (LinearP$_\rho$) and (LinearD$_\rho$), respectively, and denote their corresponding objective values by OBF$_P$ and OBF$_D$. Then $(\mathbf{z}, \boldsymbol{\eta}^{(1)}(\mathbf{z}), \cdots, \boldsymbol{\eta}^{(K)}(\mathbf{z}), \mathbf{w}^{(1)}(\mathbf{z}), \cdots, \mathbf{w}^{(K)}(\mathbf{z}))$, as defined in Proposition 4.1, is a feasible solution of the LP (FiniteP$_\rho$ (support(λ))) with a corresponding objective value OBF$_P$. On the other hand, $(\boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\zeta})$ is a feasible solution of the LP (FiniteD ρ (support(λ))), with objective value OBF$_D$. Since these LPs form a primal-dual pair, the inequality OBF$_P \le$ OBF$_D$ follows from the weak duality theorem of linear programming.

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

Finally, consider a feasible solution $\mathbf{z}$ of (LinearP$_\rho$) and a feasible solution $(\lambda, \delta, \boldsymbol{\zeta})$ of (LinearD$_\rho$), and again let $\tilde{C} = \text{support}(\lambda) \cup \hat{C}$. Then these solutions are simultaneously optimal if and only if $(\mathbf{z}, \boldsymbol{\eta}^{(1)}(\mathbf{z}), \cdots, \boldsymbol{\eta}^{(K)}(\mathbf{z}), \mathbf{w}^{(1)}(\mathbf{z}), \cdots, \mathbf{w}^{(K)}(\mathbf{z}))$ and $(\boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\zeta})$ are optimal solutions of the LPs (FiniteP$_\rho(\tilde{C})$) and (FiniteD$_\rho(\tilde{C})$), respectively. This in turn is equivalent to the following set of linear programming complementary slackness conditions:

$$\begin{array}{rcll} \lambda_\ell^{(h)} > 0 & \Rightarrow & \sum\limits_{j\in[K]} \mu_j^{(h)} \left(\eta_\ell^{(j)}(\mathbf{z}) - \frac{1}{\alpha_j} \sum\limits_{i\in[n]} p_i w_{\ell i}^{(j)}(\mathbf{z}) \right) = \rho(\varphi(\tilde{\mathbf{c}}^{(\ell)}, \mathbf{Y})), & \forall\ \ell \in [L],\ h \in [H] \\ \delta_{\ell,t,i}^{(j)} > 0 & \Rightarrow & w_{\ell i}^{(j)} = \eta_\ell^{(j)}(\mathbf{z}) - A_t(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z}), & \forall\ j \in [K],\ \ell \in [L],\ t \in [T],\ i \in [n] \\ \sum\limits_{t\in[T]} \delta_{\ell,t,i}^{(j)} < \frac{1}{\alpha_j} \sum\limits_{h\in[H]} \mu_j^{(h)} \lambda_\ell^{(h)} & \Rightarrow & w_{\ell i}^{(j)}(\mathbf{z}) = 0, & \forall\ j \in [K],\ \ell \in [L],\ i \in [n] \\ \boldsymbol{\zeta}^\top (U\mathbf{z} - \mathbf{u}) & = & 0. & \end{array} \tag{73}$$

In order to reformulate the first condition we consider the following chain of inequalities for $\ell \in [L]$ and $h \in [H]$:

$$\sum_{j\in[K]} \mu_j^{(h)} \left(\eta_\ell^{(j)}(\mathbf{z}) - \frac{1}{\alpha_j} \sum_{i\in[n]} p_i w_{\ell i}^{(j)}(\mathbf{z}) \right)$$

$$= \sum_{j\in[K]} \mu_j^{(h)} \left(\text{VaR}_{\alpha_j}(\varphi(\tilde{c}^{(\ell)}, \Gamma\mathbf{z})) - \frac{1}{\alpha_j} \sum_{i\in[n]} p_i \left[\text{VaR}_{\alpha_j}(\varphi(\tilde{c}^{(\ell)}, \Gamma\mathbf{z})) - \varphi(\tilde{c}^{(\ell)}, \Gamma(\omega_i)\mathbf{z}) \right]_+ \right) \tag{74}$$

$$= \sum_{j\in[K]} \mu_j^{(h)} \left(\text{CVaR}_{\alpha_j}(\varphi(\tilde{c}^{(\ell)}, \Gamma\mathbf{z})) \right) \tag{75}$$

$$\geq \rho(\varphi(\tilde{c}^{(\ell)}, \Gamma\mathbf{z})) \tag{76}$$

$$\geq \rho(\varphi(\tilde{c}^{(\ell)}, Y)). \tag{77}$$

Here equality (74) directly follows from the definitions of $\boldsymbol{\eta}^{(j)}(\mathbf{z})$ and $\mathbf{w}^{(j)}(\mathbf{z})$, while (75) holds since the maximum in the formulas (2) and (3) is attained when η equals the value-at-risk. Inequality (76) reflects the fact that $\boldsymbol{\mu}^{(h)}$ is featured in the Kusuoka representation (5). Finally, inequality (76) is valid because $\mathbf{z}$ is a feasible solution of (LinearP$_\rho$).

The equality in the first condition of (73) is satisfied if and only if inequalities (76) and (77) both hold with equality. Therefore, in accordance with (72) we can rewrite the first complementary slackness condition as

$$\begin{array}{rcl} \text{support}(\lambda_h) & \subset \{\mathbf{c} \quad : & \rho(\varphi(\mathbf{c}, \Gamma\mathbf{z})) = \int_0^1 \text{CVaR}_\alpha(\varphi(\mathbf{c}, \Gamma\mathbf{z}))\, \mu_h(\mathrm{d}\alpha), \\ & & \rho(\varphi(\mathbf{c}, \Gamma\mathbf{z})) = \rho(\varphi(\mathbf{c}, \mathbf{Y}))\}. \end{array}$$

To rewrite the next condition, observe that the following inequalities hold for all j, ℓ, i, and t:

$$w_{\ell i}^{(j)} \geq \eta_\ell^{(j)}(\mathbf{z}) - \varphi\left(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z}\right) = \eta_\ell^{(j)}(\mathbf{z}) - \min_{\tau\in[T]} A_\tau\left(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z}\right) \geq \eta_\ell^{(j)}(\mathbf{z}) - A_t\left(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z}\right). \tag{78}$$

The equality in the second condition of (73) is satisfied if and only if the both inequalities in (78) are satisfied with equality. Keeping in mind the definition of shortfall values $w_{\ell i}^{(j)}$, this is the case exactly when both of the conditions $\eta_\ell^{(j)}(\mathbf{z}) \geq \varphi(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z})$ and $\varphi(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z}) = A_t(\tilde{\mathbf{c}}^{(\ell)}, \Gamma(\omega_i)\mathbf{z})$ hold. Therefore, in accordance with (72) and the definition of $\eta_\ell^{(j)}(\mathbf{z})$, we can equivalently rewrite the second complementary slackness condition as

$$\begin{array}{rcl} \text{support}(\delta_t(\omega_i)) & \subset \{(\alpha, \mathbf{c}) \quad : & \varphi(\mathbf{c}, \Gamma(\omega_i)\mathbf{z}) = A_t(\mathbf{c}, \Gamma(\omega_i)\mathbf{z}), \\ & & \varphi(\mathbf{c}, \Gamma(\omega_i)\mathbf{z}) \leq \text{VaR}_\alpha(\varphi(\mathbf{c}, \Gamma\mathbf{z}))\}. \end{array}$$

It is easy to verify that the third condition can also be rewritten in an analogous fashion, and our claim follows.

Appendix B. Big-M Constants and Related Calculations The cut generation MIPs in Section 5 involve several big-M coefficients. On the one hand, these parameters need to be sufficiently high to ensure the correctness of the formulations. On the other hand, avoiding excessively high big-M values is known to improve computational performance, and, as mentioned in Sections 5.1.2 and 7.3, bounds on the minimal acceptable values can be used for variable fixing and to provide valid ordering inequalities. We now briefly outline the bounding scheme for big-M variables that was used in our implementation in place of exact calculations.

- We first set the big-M coefficients $\tilde{M}_{it}$ for $i \in [n]$, $t \in [T]$ at their lowest possible values as $\tilde{M}_{it} = \max_{\mathbf{c} \in C} A_t(\mathbf{c}, \mathbf{x}_i)$. When the vertices of the scalarization polyhedron C are known, as is the case in our HSBA test problems, these values can be calculated by simple enumeration (without the need to directly solve an LP).
- The smallest possible value of $\tilde{M}_i$ is $\max_{\mathbf{c} \in C} \min_{t \in [T]} A_t(\mathbf{c}, \mathbf{x}_i)$. We instead obtain a trivial upper bound by changing the order of maximization and minimization, and set $\tilde{M}_i = \min_{t \in [T]} \max_{\mathbf{c} \in C} A_t(\mathbf{c}, \mathbf{x}_i) = \min_{t \in [T]} \tilde{M}_{it}$ for all $i \in [n]$.
- We establish an initial upper bound on the M_{ik} values introduced in (46) by setting $M^+_{ik} := \min_{t \in [T]} \max_{\mathbf{c} \in C} A_t(\mathbf{c}, \mathbf{x}_k) - \min_{t \in [T]} \min_{\mathbf{c} \in C} A_t(\mathbf{c}, \mathbf{x}_i) = \tilde{M}_i - \min_{t \in [T]} \min_{\mathbf{c} \in C} A_t(\mathbf{c}, \mathbf{x}_i)$. Here the second term can again be calculated by enumeration when the vertices of C are known (and in the general case by solving T simple LPs).
- The expression $M^-_{ik} := \max_{\mathbf{c} \in \tilde{C}} \{\min_{t \in [T]} A_t(\mathbf{c}, \mathbf{x}_k) - \min_{t \in [T]} A_t(\mathbf{c}, \mathbf{x}_i)\}$ provides an initial lower bound on M_{ik} for any finite subset $\tilde{C}$ of C; in our implementation $\tilde{C}$ is chosen as the vertex set of C.
- We use a modified version of the algorithm at the end of Section 5.1.2 to bound the M_{ik} values. First, we calculate the initial bounds M^+_{ik}, M^-_{ik} for all pairs (i, k), then use steps (52)-(53) to update these bounds. In order to reduce the computational burden, we do not perform updates for pairs of indices with positive lower bounds or non-positive upper bounds.
- Finally, we calculate $M_{i\cdot}$ and $M_{\cdot i}$ using formulas (47), with our upper bounds in place of the exact M_{ik} values.

We note that, as an alternative to the above bounding scheme, it is also possible to calculate M_{ik} exactly by solving the following LP for all $t^* \in T$ and selecting the highest optimum objective:

$$\begin{aligned} \max \quad & \eta - A_{t^*}(\mathbf{c}, \mathbf{x}_i) & \\ \text{s.t.} \quad & \eta \leq A_t(\mathbf{c}, \mathbf{x}_k), & \forall\, t \in [T] \\ & A_{t^*}(\mathbf{c}, \mathbf{x}_i) \leq A_t(\mathbf{c}, \mathbf{x}_i), & \forall\, t \in [T]. \end{aligned}$$

However, solving T separate LPs for each pair of scenarios at every iteration of the cut generation algorithm can be overly demanding computationally, although parallelization provides a potential way to improve this situation.

Similarly to the way bounds on the M_{ik} values can be used to fix the value of some binary variables β_i, it is also possible to fix the value of certain a_{it} variables and simplify the relevant constraints (41d)-(41e) by using the following simple observation. If the inequality $\min_{c \in C} A_{t_1}(\mathbf{c}, \mathbf{x}_i) > \max_{c \in C} A_{t_2}(\mathbf{c}, \mathbf{x}_i)$ holds for some indices $i \in [n]$, $t_1, t_2 \in [T]$, then we can set $a_{it_1} = 0$, because t_1 clearly cannot be the minimizing index in the calculation (42) for $i \in [n]$. We note that when the vertices of C are known, the inequalities in question can again be checked by simple enumeration instead of linear programming.