

Robust Multicriteria Risk-Averse Stochastic Programming Models

Xiao Liu

Department of Integrated Systems Engineering, The Ohio State University, U.S.A.

liu.2738@osu.edu

Simge Küçükyavuz

Industrial and Systems Engineering, University of Washington, U.S.A.

simge@uw.edu

Nilay Noyan¹

Industrial Engineering Program, Sabancı University, Turkey

nnoyan@sabanciuniv.edu

May 20, 2017

ABSTRACT: In this paper, we study risk-averse models for multicriteria optimization problems under uncertainty. We use a weighted sum-based scalarization and take a robust approach by considering a set of scalarization vectors to address the ambiguity and inconsistency in the relative weights of each criterion. We model the risk aversion of the decision makers via the concept of *multivariate conditional value-at-risk (CVaR)*. First, we introduce a model that optimizes the *worst-case multivariate CVaR* and show that its optimal solution lies on a particular type of stochastic efficient frontier. To solve this model, we develop a finitely convergent delayed cut generation algorithm for finite probability spaces. We also show that the proposed model can be reformulated as a compact linear program under certain assumptions. In addition, for the cut generation problem, which is in general a mixed-integer program, we give a stronger formulation than the existing ones for the equiprobable case. Next, we observe that similar polyhedral enhancements are also useful for a related class of *multivariate CVaR-constrained* optimization problems that has attracted attention recently. In our computational study, we use a budget allocation application to benchmark our proposed maximin type risk-averse model against its risk-neutral counterpart and a related multivariate CVaR-constrained model. Finally, we illustrate the effectiveness of the proposed solution methods for both classes of models.

Keywords: stochastic programming; risk aversion; robust optimization; multicriteria optimization; conditional value-at-risk; cut generation; mixed-integer programming; McCormick envelopes; RLT technique

1. Introduction For many decision making problems under uncertainty, it may be essential to consider multiple possibly conflicting stochastic performance criteria. Stochastic multicriteria decision making problems arise in a wide range of areas, including financial portfolio optimization, humanitarian relief network design, production scheduling, public debt management, and homeland security budget allocation (see, e.g., Balibek and Köksalan, 2010; Hu et al., 2011; Köksalan and Şakar, 2016; Noyan, 2012; Noyan et al., 2016). In such problems, we can represent the stochastic outcomes of interest by a random vector, each dimension of which corresponds to a particular decision criterion. Then, comparing the potential decisions requires specifying preference relations among random vectors. It is also crucial to compare the random outcomes based on the decision makers' risk preferences. These concerns call for optimization models that incorporate multivariate risk-averse preference relations into *constraints* and/or *objectives*. The class of models, which incorporates the multivariate risk preferences into the *constraints* using benchmarking relations, has received some attention in the recent literature. Alternatively, in this study, we introduce a new class of models with an *objective* of optimizing a multivariate risk measure.

First, we review the existing literature on risk-averse multicriteria optimization models that feature benchmarking preference relations. In this line of research initiated by Dentcheva and Ruszczyński (2009), two types of benchmarking relations are modeled as constraints: multivariate risk-averse relations based on second-order

stochastic dominance (SSD) and conditional value-at-risk (CVaR). These models assume that a benchmark random outcome vector is available and extend univariate (scalar-based) preference rules to the multivariate (vector-based) case by using linear scalarization functions. The linear scalarization corresponds to the *weighted-sum* approach, which is widely used in multicriteria decision making (Steuer, 1986; Ehrgott, 2005); the scalarization coefficients are interpreted as the weights representing the relative (subjective) importance of each decision criterion.

In many decision making situations, especially those involving multiple decision makers, it can be difficult to determine a single weight vector. There are many alternative methods to elicit relative weights of each criterion, including multiattribute weighting, swing weighting and the analytic hierarchy process (for a review, see von Winterfeldt and Edwards, 1986; Saaty, 2000). However, the relative weights of even a single expert could be very different depending on which elicitation approach is used as shown in Schoemaker and Waid (1982) and Borchering et al. (1991). The problem of choosing a single weight vector is further exacerbated if multiple experts are involved. To address these ambiguity and inconsistency issues, a so-called *robust approach* considers a collection of weight vectors within a prescribed *scalarization set* instead of a single weight vector. Various scalarization sets are considered in the literature such as the set of all non-negative coefficients, arbitrary polyhedral and arbitrary convex sets (see, e.g., Dentcheva and Ruszczyński, 2009; Homem-de-Mello and Mehrotra, 2009; Hu et al., 2012, respectively).

While the majority of existing studies focuses on enforcing multivariate SSD relations (see, e.g., Dentcheva and Ruszczyński, 2009; Homem-de-Mello and Mehrotra, 2009; Hu et al., 2012; Dentcheva and Wolfhagen, 2013), this modeling approach can be overly conservative in practice and leads to very demanding constraints that sometimes cannot be satisfied. For example, due to this infeasibility issue, Hu et al. (2011) solve such an optimization problem with relaxed SSD constraints. As an alternative, Noyan and Rudolf (2013) propose to use a multivariate preference relation based on CVaR; their approach is motivated by the fact that the univariate SSD relation is equivalent to a continuum of CVaR inequalities (Dentcheva and Ruszczyński, 2006). The authors consider polyhedral scalarization sets and show that their CVaR-based methodology can be extended to optimization problems featuring benchmarking constraints based on a wider class of coherent risk measures. In our study, we follow the line of research of Noyan and Rudolf (2013), which provides sufficient flexibility to obtain feasible problem formulations and capture a wide range of risk preferences, including risk-neutral and worst-case approaches.

Optimization models under both types of multivariate preference relations (SSD and CVaR) are challenging, since they require introducing infinitely many univariate risk constraints associated with all possible weight vectors in the scalarization set. For polyhedral scalarization sets, Homem-de-Mello and Mehrotra (2009) and Noyan and Rudolf (2013) show that enforcing the corresponding univariate risk constraint for a finite (exponential) subset of weight vectors is sufficient to model the multivariate SSD and CVaR relations, respectively. These finite representation results allow them to develop finitely convergent delayed cut generation algorithms, where each cut is obtained by solving a mixed-integer programming (MIP) problem. Since solving these MIP formulations is the main computational bottleneck, Küçükyavuz and Noyan (2016) develop computationally effective solution methods for the cut generation problems arising in both types of optimization models.

As outlined earlier, the existing literature on risk-averse multicriteria optimization problems mainly focuses on *multivariate risk-constrained models*, where a benchmark random vector is available and the goal is to find a

solution with a multivariate outcome vector that is preferable to the benchmark (with respect to the multivariate SSD or CVaR relation). In this paper, we propose a novel model which *does not require a given benchmark* and aims to optimize the risk associated with the decision-based random vector of outcomes. In this sense, the problem we consider can be seen as a *risk-averse stochastic multiobjective* optimization. There are, in general, two types of approaches to solve stochastic multiobjective problems: 1) to eliminate the stochastic nature of the problem by replacing each random objective function with one of its summary statistics; 2) to eliminate the multiobjective structure of the problem by aggregating the multiple objectives and obtaining a single random objective function. For recent surveys on these two types of approaches we refer to [Gutjahr and Pichler \(2016\)](#) and [Ben Abdelaziz \(2012\)](#). The first (non-aggregation based) approach results in a traditional deterministic multiobjective problem and requires the identification of multiple (typically exponential) non-dominated solutions in the efficient frontier. Ultimately, however, the decision makers need to specify the weights for each criterion to choose among the non-dominated solutions. In the second (aggregation-based) approach, one can consider a weighted sum of the multiple objectives and solve the resulting stochastic problem to obtain a solution. However, the weights to be used in either approach can be *ambiguous* and *inconsistent* due to the presence of conflicting criteria and lack of consensus among multiple experts. Alternatively, in the second approach of aggregating multiple objectives into one, one can use an aggregated (but non-scalarized) single objective using stochastic goal programming. This approach considers random and/or deterministic goals (benchmark values) for the different objectives and constructs a single objective based on a function of the deviations from the goals. However, a benchmark goal may not be immediately available in all practical applications. For problems where the relative importance of the criteria is ambiguous and a benchmark performance vector is not available, we propose to focus on the worst-case CVaR with respect to the prescribed scalarization set and employ a recent notion of CVaR robustness in the context of stochastic multicriteria optimization.

In a related line of work, to address the ambiguity and inconsistency in the weights used to scalarize the multiple criteria in the objective function of a *deterministic* optimization problem, [Ogryczak \(2010\)](#) and [Hu and Mehrotra \(2012\)](#) consider minimax type robustness with respect to a given weight set. Note that such existing *robust weighted-sum models* assume that either the problem parameters are deterministic or the decision makers are risk-neutral. For an overview on minimax robustness for multiobjective optimization problems we refer to [Ehrgott et al. \(2014\)](#). However, some multicriteria decision-making problems of recent interest, such as disaster preparedness ([Hu and Mehrotra, 2012](#)) and homeland security ([Hu et al., 2011](#)), involve uncertain events with small probabilities but dire consequences. Therefore, it is crucial to incorporate risk aversion into multicriteria optimization models, which is the main focus of our study. Note that the risk-averse model we propose in this paper features the risk-neutral version as a special case.

In the recent literature, another type of CVaR robustness appears in the univariate case stemming from the distributional robustness. [Zhu and Fukushima \(2009\)](#) and [Wozabal \(2014\)](#) consider optimizing the worst-case CVaR and a wider class of convex risk measures (of a scalar-based random variable), respectively. However, this line of work assumes that there is ambiguity in the underlying probability distribution and express the worst-case with respect to a specified set of distributions. In contrast, we assume that the underlying probability distribution is known but there is ambiguity in the scalarization vector (i.e., relative importance of multiple criteria) within a polyhedral set; this leads to a worst-case multivariate CVaR measure. For robust optimization in general, the interested reader may refer to [Ben-Tal et al. \(2009\)](#) and [Bertsimas et al. \(2011\)](#).

1.1 Our contributions We incorporate risk aversion into multicriteria optimization models using the concept of multivariate CVaR. We propose a maximin type model optimizing the worst-case CVaR over a scalarization set. While the worst-case multivariate CVaR measure was recently introduced in the finance literature to assess the risk of given portfolio vectors (see, e.g., Rüschemdorf, 2013), to the best of our knowledge, there is no model or method to assist decision makers in choosing a portfolio vector that optimizes this risk measure. In this paper, we fill this gap, and give an optimization model that maximizes the worst-case multivariate CVaR. To demonstrate the adequacy of the proposed model, we show that the risk measure of interest is coherent in an appropriate multivariate sense, and an optimal solution of the model is not dominated in an appropriate stochastic sense. These two properties are highly desirable in risk-averse optimization and multicriteria optimization, respectively.

Unlike the risk-neutral version with a polyhedral weight set, in the risk-averse case, the inner optimization problem involves a concave minimization. Hence, the problem in general can no longer be solved as a compact linear program (as in Hu and Mehrotra, 2012). Therefore, we propose a delayed cut generation-based solution algorithm and show that the cut generation problem can be modeled as a bilinear program that contains the multiplication of the scalarization variables and some binary variables used for representing CVaR. We demonstrate that the assumptions on the scalarization set allow us to employ the reformulation-linearization technique (RLT) (Sherali and Adams, 1994; Sherali et al., 1998) to strengthen the resulting MIP formulations of the cut generation problem. We observe that the cut generation subproblems in the proposed algorithm have similar structure with those encountered in solving the related multivariate CVaR-constrained optimization model. The cut generation problem for this related model has been identified as the major bottleneck in earlier work and progress has been made in strengthening the corresponding formulations under the general probability case. Recognizing the importance of reducing the cut generation solution times, we utilize the RLT technique to obtain further stronger and computationally more efficient cut generation formulations for optimization under multivariate CVaR constraints, especially for the equal probability case. Note that the equiprobable case is of particular importance for the Monte Carlo sampling-based stochastic optimization models. While similar bilinear formulations and their weak linearizations exist in the literature for the CVaR-constrained optimization model, our work is a first in observing that these formulations are amenable to significant strengthening using the RLT reformulations and other polyhedral developments we provide. This observation in turn speeds up the overall solution time considerably as we show in our extensive computational study.

1.2 Outline The rest of the paper is organized as follows. In Section 2, we introduce the new worst-case CVaR optimization model and provide some analytical results to highlight the appropriateness of the proposed modeling approach. This section also presents a cut generation algorithm and effective mathematical programming formulations of the original optimization problem and the corresponding cut generation problems for some special cases. We describe how to apply some of these algorithmic features to the multivariate CVaR-constrained models in Section 3. Section 4 gives a hybrid model that includes both the multivariate CVaR-based constraints and objective. We give a unified methodology that solves the hybrid model, integrating the algorithmic developments in Sections 2 and 3. Section 5 is dedicated to the computational study, while Section 6 contains concluding remarks.

2. Worst-case CVaR Optimization Model In our study, we consider a multicriteria decision making problem where d random performance measures of interest associated with the decision vector \mathbf{z} are represented

by the random outcome vector $\mathbf{G}(\mathbf{z}) = (G_1(\mathbf{z}), \dots, G_d(\mathbf{z}))$. All random variables in this paper are assumed to be defined on some finite probability spaces; we simplify our exposition accordingly. Let $(\Omega, 2^\Omega, \mathbb{P})$ be a finite probability space with $\Omega = \{\omega_1, \dots, \omega_n\}$ and $\mathbb{P}(\omega_i) = p_i, i = 1, \dots, n$. In particular, denoting the set of feasible decisions by Z , the random outcomes are determined according to the outcome mapping $\mathbf{G} : Z \times \Omega \rightarrow \mathbb{R}^d$, and the random outcome vector $G(\mathbf{z}) : \Omega \rightarrow \mathbb{R}^d$ is defined by $G(\mathbf{z})(\omega) = G(\mathbf{z}, \omega)$ for all $\omega \in \Omega$. For a given elementary event ω_i the mapping $\mathbf{g}_i : Z \rightarrow \mathbb{R}^d$ is defined by $\mathbf{g}_i(\mathbf{z}) = \mathbf{G}(\mathbf{z}, \omega_i)$. Let $C \subset \mathbb{R}_+^d$ be a polyhedron of scalarization vectors, each component of which corresponds to the relative importance of each criterion. We naturally assume, without loss of generality, that C is a subset of the unit simplex, C_f , i.e., $C \subseteq C_f := \{\mathbf{c} \in \mathbb{R}_+^d \mid \sum_{i \in [d]} c_i = 1\}$.

Before proceeding to give our definitions and models, we need to make a note of some conventions used throughout this paper, and recall a basic definition. The set of the first n positive integers is denoted by $[n] = \{1, \dots, n\}$, while the positive part of a number $x \in \mathbb{R}$ is denoted by $[x]_+ = \max(x, 0)$. We assume that *larger values of random variables* are preferred. We quantify the risk associated with a random variable via a risk measure (specifically, CVaR) where *higher values correspond to less risky* random outcomes. In this context, risk measures are often referred to as *acceptability functionals*. Our presentation follows along the lines of Pflug and Römisch (2007) and Noyan and Rudolf (2013). Recall that for a *univariate* random variable X with (not necessarily distinct) realizations x_1, \dots, x_n and corresponding probabilities p_1, \dots, p_n , the *conditional value-at-risk* at confidence level $\alpha \in (0, 1]$ is given by (Rockafellar and Uryasev, 2000)

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

$$= \max \left\{ \eta - \frac{1}{\alpha} \sum_{i \in [n]} p_i w_i : w_i \geq \eta - x_i, \forall i \in [n], \mathbf{w} \in \mathbb{R}_+^n, \eta \in \mathbb{R} \right\} \quad (2)$$

$$= \max_{k \in [n]} \left\{ x_k - \frac{1}{\alpha} \sum_{i \in [n]} p_i [x_k - x_i]_+ \right\}, \quad (3)$$

where the last equality follows from the well known result that the maximum in definition (2) is attained at the α -quantile, which is known as the *value-at-risk* (VaR) at confidence level α (denoted by $\text{VaR}_\alpha(X)$) and that $\text{VaR}_\alpha(X) = x_k$ for at least one $k \in [n]$. For risk-averse decision makers typical choices for the confidence level are small values such as $\alpha = 0.05$. Note that $\text{CVaR}_\alpha(X)$, as defined in (1), is concave in X . For example, suppose that the random variable X represents the return of an investment and its realizations are equally likely. In this context, larger values are preferred, and $\text{VaR}_\alpha(X)$ provides a lower bound on the return that is exceeded with a high probability of $1 - \alpha$ while $\text{CVaR}_\alpha(X)$ measures the severity of the return if it is no larger than $\text{VaR}_\alpha(X)$. When $\alpha = k/n$ for some $k \in [n]$, $\text{VaR}_\alpha(X)$ is equal to the k th smallest realization and $\text{CVaR}_\alpha(X)$ is the average of the k smallest (least favorable) realizations no larger than $\text{VaR}_\alpha(X)$.

The significance of modeling robustness against the ambiguity and inconsistency in relative weights motivates us to introduce a new robust optimization model for the stochastic multicriteria decision making problem of interest. To model the risk aversion of the decision makers, we use CVaR as the acceptability functional. In particular, we focus on the recently introduced worst-case multivariate CVaR (Rüschendorf, 2013) with respect to the specified scalarization set C , which we review next.

DEFINITION 2.1 (WORST-CASE MULTIVARIATE POLYHEDRAL CVaR) *Let \mathbf{X} be a d -dimensional random vector and $C \subseteq C_f$ a set of scalarization vectors. The worst-case multivariate polyhedral CVaR (WCVaR) at*

confidence level $\alpha \in (0, 1]$ with respect to C is defined as

$$\text{WCVaR}_{C,\alpha}(\mathbf{X}) = \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}). \quad (4)$$

Following a risk-averse approach, we propose to optimize $\text{WCVaR}_{C,\alpha}$ for a given confidence level $\alpha \in (0, 1]$ and a scalarization set C , and introduce a new class of robust multicriteria optimization problems of the general form

$$(\mathbf{W} - \mathbf{CVaR}) : \quad \max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})). \quad (5)$$

We note that the proposed risk-averse $(\mathbf{W} - \mathbf{CVaR})$ problem features the risk-neutral version, proposed in [Hu and Mehrotra \(2012\)](#), as a special case when $\alpha = 1$. Another special case appears in the literature ([Ehrgott, 2005](#)) for a sufficiently small value of α (corresponding to the worst-case); it optimizes the worst value of a particular weighted sum over the set of scenarios. This robust version of the weighted sum scalarization problem is clearly a special case of $(\mathbf{W} - \mathbf{CVaR})$ if we assume that all scenarios are equally likely, $\alpha = 1/n$, and there is a single scalarization vector in the scalarization set C .

It is important to note that the major difficulty of the proposed optimization problem $(\mathbf{W} - \mathbf{CVaR})$, and the related models in the literature, stems from the presence of the *joint* acceptability functional $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}))$. One might wonder why an alternative model that maximizes the scalarization of component-wise acceptability functionals, i.e., $\max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \sum_{i \in [d]} c_i \text{CVaR}_\alpha(G_i(\mathbf{z}))$ is not preferred. After all, this approach would lead to more tractable reformulations; for example, the alternative model can be formulated as a linear program when there is no integrality restriction on the decision vector \mathbf{z} , Z is a polyhedral set, and the mapping $\mathbf{g}_i(\mathbf{z})$ is linear in \mathbf{z} for all $i \in [n]$. However, such a model completely ignores the correlation between the random variables $G_i(\mathbf{z})$, $i \in [d]$. The worst α proportion of scenarios with respect to one criterion would most likely not coincide with the worst α proportion of scenarios with respect to the other criteria, except for the very trivial case when $G_i(\mathbf{z})$, $i \in [d]$, are comonotone random variables. Therefore, using the aforementioned alternative modeling approach could only be justified to capture the multivariate risk in the trivial case when the worst-case scenarios of the multiple random outcomes coincide, which does not appear to be the typical situation in optimization with conflicting criteria. In all other cases, it would be a conservative approximation. In this paper, we are interested in exact models and methods that optimize a multivariate risk measure based on the joint behavior of the random outcomes of interest.

In the remainder of this section, we first provide some analytical results to highlight the appropriateness of the proposed model (Section 2.1). Then, in Section 2.2, we develop methods to solve this new class of problems.

2.1 Coherence and Stochastic Pareto Optimality We first analyze the properties of $\text{WCVaR}_{C,\alpha}$ as a risk measure and then show that an optimal solution of $(\mathbf{W} - \mathbf{CVaR})$ is Pareto optimal according to a certain stochastic dominance relation.

Desirable properties of risk measures have been axiomatized starting with the work of [Artzner et al. \(1999\)](#), in which the concept of coherent risk measures for scalar-valued random variables is introduced. There are several approaches to define the concept of coherency for the vector-valued random variables (see, e.g., [Jouini et al., 2004](#); [Burgert and Rüschendorf, 2006](#); [Rüschendorf, 2013](#); [Hamel et al., 2013](#)). For example, [Hamel et al. \(2013\)](#) introduce set-valued conditional value-at-risk for multivariate random variables; using such set-valued functionals as risk measures is appropriate for financial market models with transaction costs

(see, e.g., [Jouini et al., 2004](#)). Our approach is more aligned with the studies which consider multivariate risk measures that map a random vector to a scalar value; in particular, we consider the following definition of coherence in the multivariate case ([Ekeland and Schachermayer, 2011](#)).

We say that a functional $\rho : L^\infty(\Omega, 2^\Omega, \mathbb{P}; \mathbb{R}^d) \rightarrow \mathbb{R}$ mapping a d -dimensional random vector to a real number is *coherent in dimension d* (in other words, ρ is a *coherent acceptability functional* in dimension d , equivalently, that $-\rho$ is a *coherent risk measure* in dimension d), if ρ has the following properties (for all d -dimensional random vectors $\mathbf{V}, \mathbf{V}_1, \mathbf{V}_2$):

- (i) *Normalized*: $\rho(\mathbf{0}) = 0$.
- (ii) *Monotone*: $\mathbf{V}_1 \leq \mathbf{V}_2 \Rightarrow \rho(\mathbf{V}_1) \leq \rho(\mathbf{V}_2)$.
- (iii) *Positive homogeneous*: $\rho(\lambda \mathbf{V}) = \lambda \rho(\mathbf{V})$ for all $\lambda > 0$.
- (iv) *Superadditive*: $\rho(\mathbf{V}_1 + \mathbf{V}_2) \geq \rho(\mathbf{V}_1) + \rho(\mathbf{V}_2)$.
- (v) *Translation invariant (equivariant)*: $\rho(\mathbf{V} + \lambda \mathbf{e}) = \rho(\mathbf{V}) + \lambda$ for all $\lambda \in \mathbb{R}$.

The constant vector \mathbf{e} denotes the vector of ones $(1, 1, \dots, 1)$. It is easy to see that for $d = 1$ the definition coincides with the notion of coherence for scalar-valued random variables ([Artzner et al., 1999](#)); we remind the reader that we provide the definition for acceptability functionals, along the lines of [Pflug and Römisch \(2007\)](#). In the monotonicity property (ii), we consider the usual component-wise ordering; i.e., $\mathbf{V}_1 \leq \mathbf{V}_2$ if $\mathbf{V}_1(j) \leq \mathbf{V}_2(j)$ for all $j \in [d]$.

The next result indicates that the proposed risk measure is of particular importance since it satisfies the desirable properties axiomatized in the above definition of coherence.

PROPOSITION 2.1 *Consider a one-dimensional mapping ρ and a scalarization set $C \subseteq C_f$, and let $\rho_C(\mathbf{X}) = \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top \mathbf{X})$ for a d -dimensional random vector \mathbf{X} . If ρ is a coherent acceptability functional ($-\rho$ is a coherent risk measure), then $\rho_C(\mathbf{X})$ denoting the worst-case functional in dimension d (with respect to C) is also coherent.*

PROOF. It is easy to verify that ρ_C is normalized, monotone, and positive homogeneous. To show that ρ_C is superadditive, let us consider two d -dimensional random vectors \mathbf{V}_1 and \mathbf{V}_2 . Then, by the superadditivity of ρ and the minimum operator, we have $\rho_C(\mathbf{V}_1 + \mathbf{V}_2) = \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top (\mathbf{V}_1 + \mathbf{V}_2)) \geq \min_{\mathbf{c} \in C} (\rho(\mathbf{c}^\top \mathbf{V}_1) + \rho(\mathbf{c}^\top \mathbf{V}_2)) \geq \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top \mathbf{V}_1) + \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top \mathbf{V}_2) = \rho_C(\mathbf{V}_1) + \rho_C(\mathbf{V}_2)$. The translation invariance of ρ_C follows from the assumptions that $\sum_{j \in [d]} c_j = 1$ and ρ is translation invariant: for any constant λ , $\rho_C(\mathbf{V} + \lambda \mathbf{e}) = \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top (\mathbf{V} + \lambda \mathbf{e})) = \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top \mathbf{V} + \lambda) = \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top \mathbf{V}) + \lambda = \rho_C(\mathbf{V}) + \lambda$. \square

We note that one can also consider a stronger notion of translation invariance in condition (v) of the above definition of coherence; for example, [Burgert and Rüschendorf \(2006\)](#) state it as follows: $\rho(\mathbf{V} + \lambda \mathbf{e}_j) = \rho(\mathbf{V}) + \lambda$ for all $j \in [d]$ and $\lambda \in \mathbb{R}$, where \mathbf{e}_j is the standard basis vector (1 in the j th component, 0 elsewhere). [Rüschendorf \(2013\)](#) claims that $\rho_C(\mathbf{X})$ is coherent when ρ is a coherent acceptability functional, even with the above mentioned stronger translation invariance property. However, this claim is not correct even for the unit simplex ($C = C_f$), as we explain next. Since $\rho(\mathbf{c}^\top \mathbf{V})$ is concave in \mathbf{c} , the minimum in the definition of $\rho_C(\mathbf{V})$ is attained at an extreme point of C , i.e., $\rho_C(\mathbf{V}) = \min\{\rho(V_1), \rho(V_2), \dots, \rho(V_d)\}$ if C is a unit simplex. Suppose that $\rho(V_j)$, $j \in [d]$, are not all equal, which implies that there exists an index $j^* \in [d]$ such that $\rho_C(\mathbf{V}) < \rho(V_{j^*})$. Then, for any $\lambda > 0$, by the monotonicity of ρ , we have $\rho(\mathbf{V} + \lambda \mathbf{e}_{j^*}) =$

$\min\{\min_{j \in [d] \setminus \{j^*\}} \rho(V_j), \rho(V_{j^*} + \lambda)\} = \min_{j \in [d] \setminus \{j^*\}} \rho(V_j) = \rho_C(\mathbf{V}) < \rho(V_{j^*}) < \rho(V_{j^*} + \lambda) = \rho(V_{j^*}) + \lambda$. This provides an example where $\rho_C(\mathbf{V} + \lambda \mathbf{e}_j) \neq \rho_C(\mathbf{V}) + \lambda$ for all $j \in [d]$ and $\lambda \in \mathbb{R}$.

We next discuss the *Pareto efficiency/optimality* of the solutions of $(\mathbf{W} - \mathbf{CVaR})$. For *deterministic* multiobjective optimization problems, the concept of Pareto optimality is well-known and it defines a dominance relation to compare the solutions with respect to the multiple criteria. It is natural to consider the “non-dominated” solutions as potentially good solutions. Here, we recall two widely-used Pareto optimality concepts:

- A point $\mathbf{z}^* \in Z$ is called *Pareto optimal* if there exists no point $\mathbf{z} \in Z$ such that

$$G_j(\mathbf{z}) \geq G_j(\mathbf{z}^*) \text{ for all } j \in [d] \text{ and } G_j(\mathbf{z}) > G_j(\mathbf{z}^*) \text{ for at least one index } j \in [d]. \quad (6)$$

- A point $\mathbf{z}^* \in Z$ is called *weakly Pareto optimal* if there exists no point $\mathbf{z} \in Z$ such that

$$G_j(\mathbf{z}) > G_j(\mathbf{z}^*) \text{ for all } j \in [d]. \quad (7)$$

In contrast to the deterministic case, in a stochastic context there is no single widely-adopted concept of Pareto optimality. The challenge stems from the stochasticity of the criteria; $G_1(\mathbf{z}), \dots, G_d(\mathbf{z})$ are in general random variables for any decision vector $\mathbf{z} \in Z$, and one should specify how to compare solutions in terms of these random objective criteria. To this end, in this paper, we use the stochastic dominance rules and introduce stochastic dominance-based Pareto optimality concepts below for stochastic multiobjective optimization problems. For $k \in \mathbb{N}_0 = \{0, 1, \dots\}$, let us denote the k th degree stochastic dominance (k SD) relation by $\succeq_{(k)}$; we refer the reader to Appendix A for a brief review of these relations (see, also, [Ogryczak and Ruszczyński, 2001](#)).

DEFINITION 2.2 (STOCHASTIC DOMINANCE-BASED PARETO OPTIMALITY) *A point $\mathbf{z}^* \in Z$ is called k SD-based Pareto optimal for some $k \in \mathbb{N}_0$ if there exists no point $\mathbf{z} \in Z$ such that*

$$G_j(\mathbf{z}) \succeq_{(k)} G_j(\mathbf{z}^*) \text{ for all } j \in [d] \text{ and } G_j(\mathbf{z}) \succ_{(k)} G_j(\mathbf{z}^*) \text{ for at least one index } j \in [d]. \quad (8)$$

DEFINITION 2.3 (STOCHASTIC DOMINANCE-BASED WEAK PARETO OPTIMALITY) *A point $\mathbf{z}^* \in Z$ is called weakly k SD-based Pareto optimal for some $k \in \mathbb{N}_0$ if there exists no point $\mathbf{z} \in Z$ such that*

$$G_j(\mathbf{z}) \succ_{(k)} G_j(\mathbf{z}^*) \text{ for all } j \in [d]. \quad (9)$$

These stochastic Pareto optimality concepts are based on comparing the random variables $G_j(\mathbf{z})$ and $G_j(\mathbf{z}^*)$ (in relations (6) and (7)) using a univariate stochastic dominance rule for each criterion $j \in [d]$. Such a component-wise dominance relation provides a natural and an intuitive approach for extending the concept of traditional Pareto optimality to the stochastic case. A closely related but slightly different notion of efficiency based on the realizations under each scenario is presented in [Ben Abdelaziz \(2012\)](#). Alternatively, one can consider a multivariate stochastic dominance relation as in [Ben Abdelaziz et al. \(1995\)](#). However, multivariate stochastic dominance relations are very restrictive (see, e.g., [Müller and Stoyan, 2002](#)) and finding a non-dominated solution according to such a multivariate relation may not even be possible. For other generalizations of the Pareto efficiency concept to multiobjective stochastic problems we refer to [Ben Abdelaziz \(2012\)](#).

We next focus on the zeroth-order stochastic dominance (ZSD) rule (also known as statewise dominance) defined in Appendix A, and present a close analogue of Theorem 2.2 in [Hu and Mehrotra \(2012\)](#), which provides some managerial insights about our new model $(\mathbf{W} - \mathbf{CVaR})$.

PROPOSITION 2.2 Let $C \subseteq C_f$ and \mathbf{z}^* be an optimal solution of $(\mathbf{W} - \mathbf{CVaR})$.

- (i) \mathbf{z}^* is a weakly ZSD-based Pareto optimal solution of $(\mathbf{W} - \mathbf{CVaR})$.
- (ii) If for every $\mathbf{c} \in C$ we have $c_j > 0$ for all $j \in [d]$, then \mathbf{z}^* is a ZSD-based Pareto optimal solution of $(\mathbf{W} - \mathbf{CVaR})$.
- (iii) If \mathbf{z}^* is a unique optimal solution of $(\mathbf{W} - \mathbf{CVaR})$, then it is a ZSD-based Pareto optimal solution of $(\mathbf{W} - \mathbf{CVaR})$.

PROOF. Let us assume for contradiction that \mathbf{z}^* is not a weakly ZSD-based Pareto optimal solution of $(\mathbf{W} - \mathbf{CVaR})$. Then there exists $\hat{\mathbf{z}} \in Z$ such that $G_j(\hat{\mathbf{z}}, \omega_i) > G_j(\mathbf{z}^*, \omega_i)$ for all $i \in [n]$ and $j \in [d]$. By the non-negativity of $\mathbf{c} \in C$ and the observation that $c_k > 0$ for at least one index $k \in [d]$ for every $\mathbf{c} \in C$, we have $\sum_{j \in [d]} c_j G_j(\hat{\mathbf{z}}, \omega_i) > \sum_{j \in [d]} c_j G_j(\mathbf{z}^*, \omega_i)$ for all $i \in [n]$ and $\mathbf{c} \in C$. Then, by the monotonicity of CVaR it is easy to see that $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\hat{\mathbf{z}})) > \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*))$ holds for any $\alpha \in (0, 1]$ and $\mathbf{c} \in C$. Therefore, the following inequalities hold and result in a contradiction:

$$\begin{aligned} \max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) &\geq \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\hat{\mathbf{z}})) \\ &> \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)) = \max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})). \end{aligned}$$

This completes the proof of part (i). The proofs of parts (ii) and (iii) follow from similar arguments. \square

We would like to emphasize that the $(\mathbf{W} - \mathbf{CVaR})$ model keeps the stochastic nature of the weighted-sum, and is novel in terms of incorporating the risk associated with the inherent randomness. Therefore, it calls for the development of stochastic Pareto efficiency concepts discussed above. In contrast, in some of the existing stochastic multiobjective optimization models, summary statistics such as expected value, CVaR or variance are used as the multiple criteria (see, for example, Köksalan and Şakar, 2016, for a stochastic portfolio optimization problem with three criteria: expected return, CVaR and a liquidity measure). Using these summary statistics, the resulting problem becomes a deterministic multicriteria optimization problem for which the well-defined deterministic Pareto optimality concepts can be applied. One method of obtaining Pareto optimal solutions is to scalarize these multiple criteria using a single weight vector in the scalarization set C . By heuristically searching over C using the weighted Tchebycheff program, multiple solutions in the deterministic efficient frontier are generated, and then an interactive method is employed for the decision makers to choose among these solutions. To illustrate this approach, consider a modification of the portfolio optimization problem in Köksalan and Şakar (2016), where $G_1(\mathbf{z})$ is the uncertain return of the portfolio and $G_2(\mathbf{z})$ is a random liquidity measure. Suppose that two criteria are considered: $\text{CVaR}_\alpha(G_1(\mathbf{z}))$ and $\text{CVaR}_\alpha(G_2(\mathbf{z}))$. Thus, for a fixed $\mathbf{c} \in C$, the objective is to maximize the scalarization of the component-wise acceptability functionals, leading to the following problem: $\max_{\mathbf{z} \in Z} \{c_1 \text{CVaR}_\alpha(G_1(\mathbf{z})) + c_2 \text{CVaR}_\alpha(G_2(\mathbf{z}))\}$. In contrast, in our model, we search over $\mathbf{c} \in C$, such that the worst-case multivariate CVaR is maximized while considering the joint behavior of the random outcomes of interest: $\max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \{\text{CVaR}_\alpha(c_1 G_1(\mathbf{z}) + c_2 G_2(\mathbf{z}))\}$. First, note that the term in the minimization in the $(\mathbf{W} - \mathbf{CVaR})$ model is different from the objective of the interactive approach, because the order of CVaR and scalarization operations cannot be changed. Only for the special case that the decision makers are risk-neutral (i.e., $\alpha = 1$), the order of CVaR (expectation) and scalarization operations can be changed. The stochastic multiobjective modeling approaches based on component-wise risk measures (or acceptability functionals), such as the one proposed by Köksalan and Şakar (2016), have the advantage of providing a set of non-dominated solutions to choose from. In cases when there is a conflict

or ambiguity in the decision makers' preferences and the joint behavior of the random performance measures is of particular interest, our model can be used to directly determine a single solution which optimizes the worst-case CVaR performance of the scalarized random outcomes with respect to all the possible choices of the preference weights; this set of possible preference weights of the multiple criteria is closely related to the well-known uncertainty set in robust optimization (see, e.g., Ben-Tal et al., 2009). The resulting single solution could be seen as a worst-case compromise decision that could be used as a benchmark or an additional input to support the decision making process.

2.2 Solution Methods In this section, we give reformulations and solution methods for $(\mathbf{W} - \mathbf{CVaR})$. We also provide improved formulations for the important special case when each scenario has an equal probability. Before proceeding to describe the solution methods we first show that $(\mathbf{W} - \mathbf{CVaR})$ is a convex program under certain conditions.

PROPOSITION 2.3 *If Z is a convex set and $G_j(\mathbf{z})$ is concave in $\mathbf{z} \in Z$ for all $j \in [d]$, then $(\mathbf{W} - \mathbf{CVaR})$ is a convex program.*

PROOF. It is sufficient to prove that the mapping $\mathbf{z} \mapsto \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}))$ is concave. Recall that by our assumptions c_j is non-negative and $G_j(\mathbf{z})$ is concave in $\mathbf{z} \in Z$ for all $j \in [d]$ and $\mathbf{c} \in C$. Since any non-negative combination of concave functions is also concave, the mapping $\mathbf{z} \mapsto \mathbf{c}^\top \mathbf{G}(\mathbf{z})$ is concave for any $\mathbf{c} \in C$. Then, by the monotonicity and concavity of CVaR, the mapping $\mathbf{z} \mapsto \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}))$ is concave, and the claim follows from the superadditivity of the minimum operator. \square

2.2.1 General Probability Case We first observe that, for the concave outcome mappings G_j , $j \in [d]$, the inner optimization problem in (5) is a concave minimization over a convex set, which implies that an optimal solution of the inner problem occurs at an extreme point of C . Let $\hat{\mathbf{c}}^1, \dots, \hat{\mathbf{c}}^N$ be the extreme points of C . Then, using the definition of CVaR given in (2), we can formulate (5) as follows:

$$\max \psi \tag{10a}$$

$$\text{s.t. } \psi \leq \eta_\ell - \frac{1}{\alpha} \sum_{i \in [n]} p_i w_{\ell i}, \quad \forall \ell \in [N] \tag{10b}$$

$$w_{\ell i} \geq \eta_\ell - (\hat{\mathbf{c}}^\ell)^\top \mathbf{g}_i(\mathbf{z}), \quad \forall \ell \in [N], i \in [n] \tag{10c}$$

$$\mathbf{z} \in Z, \quad \mathbf{w} \in \mathbb{R}_+^{N \times n}, \quad \boldsymbol{\eta} \in \mathbb{R}^N, \quad \psi \in \mathbb{R}. \tag{10d}$$

Note that if the mapping $\mathbf{g}_i(\mathbf{z})$ is linear in \mathbf{z} for all $i \in [n]$, Z is a polyhedral set, and \mathbf{z} is a continuous decision vector, then the formulation (10) is a linear program. Under certain assumptions on the scalarization set, the number of extreme points of C may be polynomial (we will discuss these cases in Section 2.2.2), and hence the resulting formulation (10) is compact. However, in general, the number of extreme points, N , is exponential. Therefore, we propose a delayed cut generation algorithm to solve (10). We start with an initial subset of scalarization vectors $\hat{\mathbf{c}}^1, \dots, \hat{\mathbf{c}}^L$ and solve an intermediate relaxed master problem (RMP), which is obtained by replacing N with L in (10). Solving the RMP provides us with a candidate solution denoted by $(\mathbf{z}^*, \psi^*, \mathbf{w}^*, \boldsymbol{\eta}^*)$. At each iteration, we solve a cut generation problem:

$$(\mathbf{CutGen} - \mathbf{Robust}) : \quad \min_{\mathbf{c} \in C} \text{CVaR}(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)).$$

If the optimal objective function value of the cut generation problem is not smaller than ψ^* , then the current solution $(\mathbf{z}^*, \psi^*, \mathbf{w}^*, \boldsymbol{\eta}^*)$ is optimal. Otherwise, the optimal solution \mathbf{c}^t at iteration t gives a violated inequality of the form $\psi \leq \text{CVaR}_\alpha((\mathbf{c}^t)^\top \mathbf{G}(\mathbf{z}))$. We augment the RMP by setting $L \leftarrow L + 1$, and $\hat{\mathbf{c}}^{L+1} \leftarrow \mathbf{c}^t$.

Observe that in the multivariate CVaR-constrained problems studied in [Noyan and Rudolf \(2013\)](#) and [Küçükyavuz and Noyan \(2016\)](#), given a random benchmark vector \mathbf{Y} , the cut generation problems are given by $\min_{\mathbf{c} \in C} \text{CVaR}(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)) - \text{CVaR}(\mathbf{c}^\top \mathbf{Y})$ (we will revisit this cut generation problem in Section 3). Due to the similar structure, we can use the formulations and enhancements given in [Noyan and Rudolf \(2013\)](#) and [Küçükyavuz and Noyan \(2016\)](#) to formulate the cut generation problem (**CutGen – Robust**) as a mixed-integer program. Let $\mathbf{X} = \mathbf{G}(\mathbf{z}^*)$ with the realizations $\mathbf{x}_1, \dots, \mathbf{x}_n$ (i.e., $\mathbf{x}_i = \mathbf{g}_i(\mathbf{z}), i \in [n]$). The representation of CVaR in (3) leads to the following formulation of (**CutGen – Robust**):

$$\min \quad \mu \tag{11a}$$

$$\text{s.t.} \quad \mu \geq \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha} \sum_{i \in [n]} p_i [\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+, \quad \forall k \in [n], \tag{11b}$$

$$\mathbf{c} \in C, \quad \mu \in \mathbb{R}. \tag{11c}$$

The shortfall terms $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$ in inequalities (11b) present a computational challenge. Introducing additional variables and constraints, we can linearize these terms using big-M type of constraints, and obtain an equivalent MIP formulation similar to the one proposed by [Noyan and Rudolf \(2013\)](#) for the cut generation problems arising in optimization under multivariate polyhedral CVaR constraints. However, the big-M type constraints may lead to weak LP relaxation bounds and computational difficulties. In order to deal with these difficulties, [Küçükyavuz and Noyan \(2016\)](#) propose an improved model based on a new representation of VaR_α , which we describe next. Let

$$M_{ik} = \max_{\mathbf{c} \in C} \{\max\{\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i, 0\}\} \text{ and } M_{ki} = \max_{\mathbf{c} \in C} \{\max\{\mathbf{c}^\top \mathbf{x}_i - \mathbf{c}^\top \mathbf{x}_k, 0\}\}.$$

Also let $M_{i*} = \max_{k \in [n]} M_{ik}$ and $M_{*i} = \max_{k \in [n]} M_{ki}$ for $i \in [n]$, and $\tilde{M}_j = \max\{c_j : \mathbf{c} \in C\}$ for $j \in [d]$. Then, the following inequalities hold for any $\mathbf{c} \in C$:

$$z \leq \mathbf{c}^\top \mathbf{x}_i + \beta_i M_{i*}, \quad \forall i \in [n] \tag{12a}$$

$$z \geq \mathbf{c}^\top \mathbf{x}_i - (1 - \beta_i) M_{*i}, \quad \forall i \in [n] \tag{12b}$$

$$z = \sum_{i \in [n]} \xi_i^\top \mathbf{x}_i, \tag{12c}$$

$$\xi_{ij} \leq \tilde{M}_j u_i, \quad \forall i \in [n], j \in [d] \tag{12d}$$

$$\sum_{i \in [n]} \xi_{ij} = c_j, \quad \forall j \in [d] \tag{12e}$$

$$\sum_{i \in [n]} p_i \beta_i \geq \alpha, \tag{12f}$$

$$\sum_{i \in [n]} p_i \beta_i - \sum_{i \in [n]} p_i u_i \leq \alpha - \epsilon, \tag{12g}$$

$$\sum_{i \in [n]} u_i = 1, \tag{12h}$$

$$u_i \leq \beta_i, \quad \forall i \in [n] \tag{12i}$$

$$\boldsymbol{\beta}, \mathbf{u} \in \{0, 1\}^n, \quad \boldsymbol{\xi} \in \mathbb{R}_+^{n \times d}, \quad z \in \mathbb{R}, \tag{12j}$$

if and only if $z = \text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Here ϵ is a sufficiently small positive constant to ensure that the constraint (12g) is equivalent to the strict inequality $\sum_{i \in [n]} p_i \beta_i - \sum_{i \in [n]} p_i u_i < \alpha$. Denoting the finite set of all non-zero probabilities of events by $\mathcal{K} = \{\mathbb{P}(S) : S \in 2^\Omega, \mathbb{P}(S) > 0\}$ it is easy to see that ϵ can be taken as any number that satisfies $0 < \epsilon < \min \{\alpha - \kappa : \kappa \in \mathcal{K} \cup \{0\}, \kappa < \alpha\}$. For example, for the equiprobable case and $\alpha = k/n$ for some $k \in [n]$, we let $0 < \epsilon < \frac{1}{n}$. The logical variable $u_i = 1$ only if the i -th scenario corresponds to $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, the logical variable $\beta_i = 1$ only if $\mathbf{c}^\top \mathbf{x}_i \leq \text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, and the variable $\xi_{ij} = c_j$ only when $u_i = 1$ for all $i \in [n]$ and $j \in [d]$.

Based on the representation of $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ given in (12), we propose an alternative formulation for **(CutGen – Robust)**:

$$\min \quad z - \frac{1}{\alpha} \sum_{i \in [n]} p_i v_i \quad (13a)$$

$$\text{s.t. (12a) – (12i),} \quad (13b)$$

$$v_i - \delta_i = z - \mathbf{c}^\top \mathbf{x}_i, \quad \forall i \in [n] \quad (13c)$$

$$v_i \leq M_{i*} \beta_i, \quad \forall i \in [n] \quad (13d)$$

$$\delta_i \leq M_{*i} (1 - \beta_i), \quad \forall i \in [n] \quad (13e)$$

$$\boldsymbol{\beta}, \mathbf{u} \in \{0, 1\}^n, \quad \boldsymbol{\xi} \in \mathbb{R}_+^{n \times d}, \quad z \in \mathbb{R}, \quad (13f)$$

$$\mathbf{c} \in C, \quad \mathbf{v}, \boldsymbol{\delta} \in \mathbb{R}_+^n. \quad (13g)$$

In this formulation, it is guaranteed that $v_i = [z - \mathbf{c}^\top \mathbf{x}_i]_+$ and $\delta_i = [\mathbf{c}^\top \mathbf{x}_i - z]_+$ for $i \in [n]$.

2.2.2 Equal Probability Case To keep our exposition simple, we consider confidence levels of the form $\alpha = k/n$ for some $k \in [n]$, and refer to [Noyan and Rudolf \(2013\)](#) for an extended MIP formulation with an arbitrary confidence level. In this case, an alternative formulation of **(CutGen – Robust)**, adapted from [Noyan and Rudolf \(2013\)](#), is given by the bilinear program

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} c_j \beta_i$$

$$\text{s.t.} \quad \sum_{i \in [n]} \beta_i = k,$$

$$\boldsymbol{\beta} \in [0, 1]^n, \quad \mathbf{c} \in C.$$

Note that we can relax the integrality of $\boldsymbol{\beta}$ in this formulation, which follows from the observation that in the special case of equal probabilities and $\alpha = k/n$, $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ corresponds to the weighted sum of the smallest k out of n realizations ($\mathbf{c}^\top \mathbf{x}_i$, $i \in [n]$). Using McCormick envelopes ([McCormick, 1976](#)), we can linearize the bilinear terms $c_j \beta_i$ in the objective function. Introducing the additional variables $\gamma_{ij} = c_j \beta_i$, for all $i \in [n]$ and $j \in [d]$, an equivalent MIP formulation is stated as:

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij} \quad (14a)$$

$$\text{s.t.} \quad \gamma_{ij} \leq c_j, \quad \forall i \in [n], j \in [d] \quad (14b)$$

$$\gamma_{ij} \geq c_j - \tilde{M}_j (1 - \beta_i), \quad \forall i \in [n], j \in [d] \quad (14c)$$

$$\gamma_{ij} \leq \tilde{M}_j \beta_i, \quad \forall i \in [n], j \in [d] \quad (14d)$$

$$\sum_{i \in [n]} \beta_i = k, \quad (14e)$$

$$\boldsymbol{\beta} \in \{0, 1\}^n, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^{n \times d}, \quad \mathbf{c} \in C. \quad (14f)$$

For $i \in [n]$, if $\beta_i = 1$, then constraint (14b) together with (14c) enforces that $\gamma_{ij} = c_j$, for all $j \in [d]$. For $i \in [n]$, if $\beta_i = 0$, then constraint (14d) enforces γ_{ij} to be 0.

Let $P := \{(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{c}) \in \mathbb{R}_+^{n \times d} \times \{0, 1\}^n \times C \mid \boldsymbol{\gamma} = \boldsymbol{\beta} \mathbf{c}^\top, \sum_{i \in [n]} \beta_i = k\}$. Then we have $\min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \min_{(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{c}) \in P} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij}$. Note that the structure of P also appears in pooling problems (c.f., Gupte et al., 2017). The next proposition gives the convex hull of P for a special choice of C using the reformulation-linearization technique (RLT) (Sherali and Adams, 1994).

PROPOSITION 2.4 (Sherali et al. (1998); Gupte et al. (2017)) *If C is a unit simplex (i.e., $C = C_f$), then the convex hull of P is described by:*

$$\text{conv}(P) = \{(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{c}) \in \mathbb{R}_+^{n \times d} \times [0, 1]^n \times C \mid \gamma_{ij} \leq c_j, i \in [n], j \in [d], \sum_{j \in [d]} \gamma_{ij} = \beta_i, i \in [n], \sum_{i \in [n]} \gamma_{ij} = kc_j, j \in [d]\}.$$

Using the fact that $C \subseteq C_f$ and Proposition 2.4, we can strengthen the formulation (14) as follows:

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij} \quad (15a)$$

$$\text{s.t. } \gamma_{ij} \leq c_j, \quad \forall i \in [n], j \in [d] \quad (15b)$$

$$\sum_{j \in [d]} \gamma_{ij} = \beta_i, \quad \forall i \in [n] \quad (15c)$$

$$\sum_{i \in [n]} \gamma_{ij} = kc_j, \quad \forall j \in [d] \quad (15d)$$

$$(14c) - (14d), \quad (15e)$$

$$\mathbf{c} \in C, \quad \boldsymbol{\beta} \in \{0, 1\}^n, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^{n \times d}. \quad (15f)$$

Note also that if C is the unit simplex ($C = C_f$), then the integrality restrictions on $\boldsymbol{\beta}$ can be relaxed in (15) and the cut generation problem is an LP. However, recall that if C is the unit simplex, then the extreme points of C are polynomial, given by $\hat{\mathbf{c}}^\ell = \mathbf{e}_\ell$ for $\ell \in [d]$. Hence, in this case, the overall problem formulation (10) itself is a compact LP when the mapping $\mathbf{g}_i(\mathbf{z})$ is linear in \mathbf{z} for all $i \in [n]$, and Z is a polyhedral set without integrality restrictions, even under general probabilities.

Furthermore, using the additional information on the structure of the scalarization polytope C and the RLT technique, we can obtain stronger formulations. Suppose that $C = \{\mathbf{c} \in \mathbb{R}_+^d \mid B\mathbf{c} \geq \mathbf{b}\}$, for a given $r \times d$ matrix B and $\mathbf{b} = (b_1, \dots, b_r)$. Let B_ℓ be the ℓ th row of B . Then, we can strengthen the formulation (14) as follows:

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij} \quad (16a)$$

$$\text{s.t. } \sum_{j \in [d]} B_{\ell j} \gamma_{ij} - b_\ell \beta_i \leq B_\ell \mathbf{c} - b_\ell, \quad \forall i \in [n], \ell \in [r] \quad (16b)$$

$$\sum_{j \in [d]} \sum_{i \in [n]} B_{\ell j} \gamma_{ij} - b_\ell \beta_i \geq 0, \quad \forall i \in [n], \ell \in [r] \quad (16c)$$

$$\sum_{i \in [n]} (\sum_{j \in [d]} B_{\ell j} \gamma_{ij} - b_\ell \beta_i) = k(B_\ell \mathbf{c} - b_\ell), \quad \forall \ell \in [r] \quad (16d)$$

$$\mathbf{c} \in C, \quad \boldsymbol{\beta} \in \{0, 1\}^n, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^{n \times d}. \quad (16e)$$

It is known that if $C = \{\mathbf{c} \in \mathbb{R}_+^d \mid B\mathbf{c} \geq \mathbf{b}\}$ is a d -simplex, then $\text{conv}(P) = \{(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{c}) \in \mathbb{R}_+^{n \times d} \times [0, 1]^n \times C \mid (16b) - (16d)\}$ (Gupte et al., 2017). Therefore, the LP relaxation of (16) is integral in this case.

REMARK 2.1 Note that if $\tilde{M}_j = 1$ for all $j \in [d]$ (as is the case when C is the unit simplex), then constraints (14c)-(14d) are implied by (15c)-(15d), and can be dropped from the formulation. However, for the situations where $\tilde{M}_j < 1$ for some $j \in [d]$, the constraints (14c)-(14d), obtained by applying the RLT technique to the constraints $c_j \leq \tilde{M}_j, j \in [d]$, can be useful to reduce the solution time.

REMARK 2.2 It is also possible to obtain stronger formulations of (12) by applying the RLT technique for the general probability case. In particular, the RLT procedure based on the constraint $\sum_{i \in [d]} c_i = 1$ provides the following valid inequality

$$\sum_{j \in [d]} \xi_{ij} = u_i, \quad (17)$$

which can be added to the formulation (12).

Next we consider an important special case of C that applies to multicriteria optimization when certain criteria are deemed more important than others. In particular, we study the case where C contains ordered preference constraints that take the form

$$C = \{\mathbf{c} \in \mathbb{R}_+^d \mid \sum_{j \in [d]} c_j = 1, c_j \geq c_{j+1}, \quad \forall j \in [d-1]\}. \quad (18)$$

If the set C has the ordered preference structure (18), then we are able to obtain the convex hull of P , which is stated in the following result.

PROPOSITION 2.5 If C is given by (18), then the convex hull of P is described by:

$$\text{conv}(P) = \{(\boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{c}) \in \mathbb{R}_+^{n \times d} \times [0, 1]^n \times C \mid (15c), (15d), \gamma_{ij} \geq \gamma_{i,j+1}, \gamma_{ij} - \gamma_{i,j+1} \leq c_j - c_{j+1}, i \in [n], j \in [d-1]\}.$$

PROOF. First, we show that the extreme points of C are given by

$$\begin{aligned} \hat{\mathbf{c}}^1 &= (1, 0, 0, \dots, 0) \\ \hat{\mathbf{c}}^2 &= \left(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0\right) \\ \hat{\mathbf{c}}^3 &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \dots, 0\right) \\ &\vdots \\ \hat{\mathbf{c}}^d &= \left(\frac{1}{d}, \frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right). \end{aligned}$$

Let $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_d)$ be a feasible point of C , by definition, we have $\tilde{c}_1 \geq \tilde{c}_2 \geq \dots \geq \tilde{c}_d$. First, we show that $\tilde{c}_j \leq \frac{1}{j}$, for all $j \in [d]$. Suppose that there exists $j \in [d]$ such that $\tilde{c}_j > \frac{1}{j}$, then we have $\sum_{i=1}^j \tilde{c}_i \geq j\tilde{c}_j > 1$ since $\tilde{c}_i \geq c_j$, for all $i \in [j-1]$; this results in a contradiction. Hence, for any feasible point, we have $\tilde{c}_j \leq \frac{1}{j}$, for all $j \in [d]$. Next, let $\lambda_j = j(\tilde{c}_j - \tilde{c}_{j+1})$, for all $j \in [d]$, where $\tilde{c}_{d+1} = 0$. Note that $0 \leq \lambda_j \leq 1$, for all $j \in [d]$, and $\sum_{j=1}^d \lambda_j = 1$. We have $\tilde{\mathbf{c}} = \sum_{j=1}^d \lambda_j \hat{\mathbf{c}}^j$, which indicates that any feasible point $\tilde{\mathbf{c}}$ can be represented as a convex combination of the points $\hat{\mathbf{c}}^j$, for all $j \in [d]$. As a result, C is a $(d-1)$ -simplex, and the proposition follows similarly from Gupte et al. (2017). \square

2.3 Finite Convergence In this section, we study the convergence of the proposed cut generation algorithm.

PROPOSITION 2.6 *The delayed cut generation algorithm described in Section 2.2 to solve $(\mathbf{W} - \mathbf{CVaR})$ is finitely convergent.*

PROOF. We show that given a solution to RMP we can find an optimal solution to the cut generation subproblem, which is an extreme point of C . As a result, the proposed cut generation algorithm is finitely convergent, because there are finitely many extreme points of C . For the general probability case, we can obtain such a vertex optimal solution by using the following method: suppose that we solve one of the MIP formulations of $(\mathbf{CutGen} - \mathbf{Robust})$ and obtain an optimal solution \mathbf{c}^* . Let π be a permutation describing a non-decreasing ordering of the realizations of the random vector $\mathbf{c}^{*\top} \mathbf{X}$, i.e., $\mathbf{c}^{*\top} \mathbf{x}_{\pi(1)} \leq \dots \leq \mathbf{c}^{*\top} \mathbf{x}_{\pi(n)}$, and define

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

Then, we can obtain the desired vertex solution $\hat{\mathbf{c}}$ by finding a vertex optimal solution of the following linear program:

$$\min_{\mathbf{c} \in C} \frac{1}{\alpha} \left[\sum_{i \in K^*} p_i \mathbf{c}^\top \mathbf{x}_i + \left(\alpha - \sum_{i \in K^*} p_i \right) \mathbf{c}^\top \mathbf{x}_{\pi(k^*)} \right].$$

This LP relies on the subset-based representation of CVaR (Theorem 1, [Noyan and Rudolf, 2013](#)). The feasible set is the polytope C , so there exists a vertex optimal solution $\hat{\mathbf{c}}$. It is easy to show that $\hat{\mathbf{c}}$ is also an optimal solution of $(\mathbf{CutGen} - \mathbf{Robust})$. \square

Furthermore, when equal probability is assumed, by solving the alternative cut generation formulation (16) using a branch-and-bound (B&B) method, we are guaranteed to obtain a desired vertex optimal solution \mathbf{c} without solving an additional LP. To see this, note that once the LP relaxation at a B&B node results in an integral β , the only remaining constraints enforce $\mathbf{c} \in C$.

3. Multivariate CVaR-constrained Optimization Model In this section, we consider a related class of multicriteria decision making problems, where the decision vector \mathbf{z} is selected from a feasible set Z and associated random outcomes are determined by the outcome mapping $\mathbf{G} : Z \times \Omega \rightarrow \mathbb{R}^d$. We consider an arbitrary objective function $f : Z \mapsto \mathbb{R}$ and assume that a d -dimensional benchmark random vector \mathbf{Y} is available. We aim to find the best decision vector \mathbf{z} for which the random outcome vector $\mathbf{G}(\mathbf{z})$ is preferable to the benchmark \mathbf{Y} with respect to the multivariate polyhedral CVaR preference relation. Such multivariate CVaR-constrained optimization problems are introduced in [Noyan and Rudolf \(2013\)](#). Given a polyhedron of scalarization vectors $C \subseteq C_f$ and a confidence level $\alpha \in (0, 1]$, the problem is of the general form:

$$\max f(\mathbf{z}) \tag{19a}$$

$$\text{s.t. } \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \quad \forall \mathbf{c} \in C \tag{19b}$$

$$\mathbf{z} \in Z. \tag{19c}$$

The benchmark random vector can be defined on a different probability space, but in practical applications it often takes the form $\mathbf{Y} = \mathbf{G}(\bar{\mathbf{z}})$, where $\bar{\mathbf{z}}$ is a benchmark decision.

Observe that (19b) contains infinitely many inequalities. [Noyan and Rudolf \(2013\)](#) show that these inequalities can be replaced with those for a finite subset of scalarization vectors corresponding to the vertices of a higher dimensional polyhedron. The authors propose a delayed cut generation algorithm, which involves the

solution of a relaxed master problem (RMP-B) to obtain a candidate solution $\hat{\mathbf{z}} \in Z$, and the following cut generation subproblem:

$$\text{(CutGen – Benchmark)} : \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) - \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \quad (20)$$

where $\mathbf{X} = \mathbf{G}(\hat{\mathbf{z}})$. If the optimal objective function value of **(CutGen – Benchmark)** is non-negative, then $\hat{\mathbf{z}}$ is optimal, otherwise we obtain a solution $\mathbf{c}^* \in C$ such that the corresponding CVaR inequality in (19b) is violated. We augment RMP-B by adding this violated CVaR constraint and resolve it. According to Noyan and Rudolf (2013), the main bottleneck of this delayed cut generation algorithm is solving the cut-generation problem (20), since it is generally nonconvex. Therefore, the main focus of this section is the cut generation problem. Throughout the rest of this paper, we assume that \mathbf{Y} is a random vector with (not necessarily distinct) realizations $\mathbf{y}_1, \dots, \mathbf{y}_m$ and corresponding probabilities q_1, \dots, q_m . As before, we let $\mathbf{g}_i(\hat{\mathbf{z}}) = \mathbf{x}_i = (x_{i1}, \dots, x_{id})$ for all $i \in [n]$.

To solve (20), we first need to represent $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ and $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$ appropriately. Using the LP representation (2) for $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$, we can reformulate **(CutGen – Benchmark)** as

$$\begin{aligned} \min \quad & \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \\ \text{s.t.} \quad & w_l \geq \eta - \mathbf{c}^\top \mathbf{y}_l, \quad \forall l \in [m] \end{aligned} \quad (21a)$$

$$\mathbf{w} \in \mathbb{R}_+^m, \quad \eta \in \mathbb{R}, \quad \mathbf{c} \in C. \quad (21b)$$

The minimization of the concave term $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ causes computational difficulties. For this cut generation problem, Küçükyavuz and Noyan (2016) introduce a MIP formulation based on the VaR representation of CVaR (see (12)), which is given by

$$\min \quad z - \frac{1}{\alpha} \sum_{i \in [n]} p_i v_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \quad (22a)$$

$$\text{s.t.} \quad (13b) - (13e), \quad (21a), \quad (22b)$$

$$\boldsymbol{\beta}, \mathbf{u} \in \{0, 1\}^n, \quad \boldsymbol{\xi} \in \mathbb{R}_+^{n \times d}, \quad \mathbf{z}, \eta \in \mathbb{R}, \quad (22c)$$

$$\mathbf{c} \in C, \quad \mathbf{v}, \boldsymbol{\delta} \in \mathbb{R}_+^n, \quad \mathbf{w} \in \mathbb{R}_+^m. \quad (22d)$$

The authors demonstrate that this formulation, which we refer to as **(MIP – CVaR)**, along with computational enhancements, outperforms existing models for **(CutGen – Benchmark)** under general probabilities. In this section, we consider the special case of equal probabilities, and propose strengthened MIP formulations for the cut generation problems using the RLT technique.

As in Section 2.2.2, to keep our exposition simple, we consider confidence levels of the form $\alpha = k/n$ and assume that all the outcomes of \mathbf{X} are equally likely. For this special case, similar to the development in Section 2.2.2, Noyan and Rudolf (2013) give the equivalent formulation below:

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \gamma_i^\top \mathbf{x}_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \quad (23a)$$

$$\text{s.t.} \quad (14b) - (14e), \quad (21a), \quad (23b)$$

$$\boldsymbol{\beta} \in \{0, 1\}^n, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^{n \times d}, \quad \mathbf{c} \in C, \quad \mathbf{w} \in \mathbb{R}_+^m, \quad \eta \in \mathbb{R}. \quad (23c)$$

As before, $\tilde{M}_j = \max\{c_j : \mathbf{c} \in C\}$. Suppose that the vertices of the polytope C is known and given as $\{\hat{\mathbf{c}}_1, \dots, \hat{\mathbf{c}}_N\}$. Then, we can simply set $\tilde{M}_j = \max_{\ell \in [N]} \hat{c}_{\ell j}$. Furthermore, we can use the RLT-based strengthening

for (14b)-(14e) and obtain the following MIP formulation:

$$\text{(MIP - Special) :} \quad \min \quad \frac{1}{k} \sum_{i \in [n]} \gamma_i^\top \mathbf{x}_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \quad (24a)$$

$$\text{s.t.} \quad (15b) - (15e), \quad (21a), \quad (24b)$$

$$\boldsymbol{\beta} \in \{0, 1\}^n, \quad \boldsymbol{\gamma} \in \mathbb{R}_+^{n \times d}, \quad \mathbf{c} \in C, \quad \mathbf{w} \in \mathbb{R}_+^m, \quad \eta \in \mathbb{R}. \quad (24c)$$

In addition, we can use the RLT technique to further strengthen this formulation using any additional constraints in C as in (16); we provide some numerical results on the performance of such strengthened versions in the computational study (Section 5.2).

From Proposition 2.4, we can obtain the minimum of $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ by solving a linear program when C is a d -simplex. However, even for the special case of unit simplex, constraints (15b)–(15d) are not sufficient to describe the convex hull of the set of feasible solutions to (24), due to the additional constraints (21a)–(21b) representing $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$. To show this and develop potentially stronger MIP formulations, we derive a class of valid inequalities that describes facets of the convex hull of feasible solutions to (24c). Let

$$\mathcal{S} := \{(\boldsymbol{\gamma}, \mathbf{c}, \boldsymbol{\beta}, \eta, \mathbf{w}) \in \mathbb{R}_+^{n \times d} \times \mathbb{R}_+^d \times \{0, 1\}^n \times \mathbb{R} \times \mathbb{R}_+^m \mid \boldsymbol{\gamma} = \boldsymbol{\beta} \mathbf{c}^\top, \sum_{j \in [d]} c_j = 1, \sum_{i \in [n]} \beta_i = k, \mathbf{c}^\top \mathbf{y}_l \geq \eta - w_l, \forall l \in [m]\}.$$

PROPOSITION 3.1 *For any $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, the inequality*

$$\mathbf{c}^\top \mathbf{y}_s - \sum_{j \in [d]} (y_{sj} - y_{tj}) \gamma_{ij} \geq \eta - w_s - w_t, \quad (25)$$

is valid for \mathcal{S} . In addition, inequality (25) defines a facet of $\text{conv}(\mathcal{S})$ if and only if $s \in [m]$, $t \in [m] \setminus \{s\}$ are such that $y_{sj} < y_{tj}$ and $y_{si} > y_{ti}$ for some $i, j \in [d]$.

PROOF. Suppose that $\beta_i = 0$, then $\gamma_{ij} = 0$ for all $j \in [d]$. Hence, inequality (25) reduces to

$$\mathbf{c}^\top \mathbf{y}_s \geq \eta - w_s - w_t,$$

which is valid since $w_t \geq 0$. Otherwise, suppose that $\beta_i = 1$, then $\gamma_{ij} = c_j$ for all $j \in [d]$, and inequality (25) reduces to

$$\mathbf{c}^\top \mathbf{y}_t \geq \eta - w_t - w_s,$$

which is valid, because $w_s \geq 0$, for all $s \in [m]$. We provide the facet proof in Appendix B (see Proposition B.2). \square

Note that applying the RLT procedure directly to the additional constraints

$$\mathbf{c}^\top \mathbf{y}_l \geq \eta - w_l, \quad \forall l \in [m], \quad (26)$$

in the set \mathcal{S} , would lead to additional bilinear terms $\eta\beta_i$ and $w_l\beta_i$ that will need to be linearized by introducing additional variables and big-M constraints. The proposed inequalities (25) can also be obtained by an indirect application of the RLT procedure as follows. Given $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, multiply constraint (26) for $l = s$ with $(1 - \beta_i)$, constraint (26) for $l = t$ with β_i , constraint $0 \geq -w_s$ with β_i and constraint $0 \geq -w_t$ with $(1 - \beta_i)$, and sum the resulting inequalities up to obtain inequality (25) (the undesirable nonlinear terms cancel out with this selection of multipliers). It is interesting to note that such an application of RLT yields facet-defining inequalities as claimed in Proposition 3.1.

Alternative VaR-based formulations. Here, without loss of generality, we assume that all the realizations of $\mathbf{c}^\top \mathbf{X}$ are non-negative (or equivalently, \mathbf{x}_i is non-negative for all $i \in [n]$). Then, it is easy to show that **(CutGen – Benchmark)** can be formulated as follows:

$$\begin{aligned} \min \quad & \frac{1}{k} \sum_{i \in [n]} \theta_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \\ \text{s.t.} \quad & \theta_i \geq \mathbf{c}^\top \mathbf{x}_i - (1 - \beta_i) M_i, \quad \forall i \in [n] \end{aligned} \quad (27a)$$

$$\sum_{i \in [n]} \beta_i = k, \quad (27b)$$

$$(21a), \quad (27c)$$

$$\mathbf{c} \in C, \quad \beta \in \{0, 1\}^n, \quad \boldsymbol{\theta} \in \mathbb{R}_+^n, \quad \mathbf{w} \in \mathbb{R}_+^m, \quad \eta \in \mathbb{R}. \quad (27d)$$

In this formulation, M_i is the largest possible value of θ_i (e.g., $M_i = \max_{\mathbf{c} \in C} \mathbf{c}^\top \mathbf{x}_i$). This new formulation again follows from the observation that in the special case of equal probabilities and $\alpha = k/n$, $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ corresponds to the weighted sum of the smallest k realizations of $\mathbf{c}^\top \mathbf{X}$. In this special case, $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ corresponds to the k th smallest realization, and the model guarantees that $\theta_i = \mathbf{c}^\top \mathbf{x}_i$ if $\mathbf{c}^\top \mathbf{x}_i \leq \text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, and $\theta_i = 0$ otherwise. However, this MIP formulation is weak due to the big- M constraints (27a). Hence, we can take advantage of the new representation of VaR given in (12) to develop a stronger MIP formulation:

$$\text{(MIP_VaR_Special)} : \quad \min \quad \frac{1}{k} \sum_{i \in [n]} \theta_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \quad (28a)$$

$$\text{s.t.} \quad z \leq \mathbf{c}^\top \mathbf{x}_i + \beta_i M_{i*}, \quad \forall i \in [n] \quad (28b)$$

$$\theta_i \geq \mathbf{c}^\top \mathbf{x}_i - (1 - \beta_i) M_i, \quad \forall i \in [n] \quad (28c)$$

$$z \geq \theta_i, \quad \forall i \in [n] \quad (28d)$$

$$z = \sum_{i \in [n]} \xi_i^\top \mathbf{x}_i, \quad (28e)$$

$$\sum_{i \in [n]} \xi_{ij} = c_j, \quad \forall j \in [d] \quad (28f)$$

$$\sum_{j \in [d]} \xi_{ij} = u_i, \quad \forall i \in [n] \quad (28g)$$

$$\sum_{i \in [n]} \beta_i = k, \quad (28h)$$

$$(12d), (12h) - (12i), (21a), \quad (28i)$$

$$\mathbf{c} \in C, \quad \beta, \mathbf{u} \in \{0, 1\}^n, \quad \mathbf{w} \in \mathbb{R}_+^m, \quad \eta, z \in \mathbb{R}, \quad (28j)$$

$$\boldsymbol{\xi} \in \mathbb{R}_+^{n \times d}, \quad \boldsymbol{\theta} \in \mathbb{R}_+^n. \quad (28k)$$

In this formulation, the variable $z = \text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ is represented by $\sum_{i \in [n]} \xi_i^\top \mathbf{x}_i = \sum_{i \in [n]} u_i \mathbf{c}^\top \mathbf{x}_i$, and it is guaranteed that $\xi_{ij} = c_j u_i$ for all $i \in [n]$ and $j \in [d]$. These bilinear terms are linearized by using the McCormick envelopes and their RLT strengthening based on only the information that C is a subset of the unit simplex. Additional constraints on the scalarization set C can be used to further strengthen the above formulation. Notice that different from (12), this formulation includes the RLT strengthening equality (17) (or (28g)).

Finally, we note that **(MIP_VaR_Special)** can also be applied to solve **(CutGen – Robust)** by dropping the variables and constraints associated with $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$; leading to enhanced versions of (13) for the equal probability case. We test its computational performance in Section 5.2.

4. Hybrid Model In this section, we present a hybrid model that includes both the multivariate CVaR constraints and the robust objective based on the worst-case CVaR. We show that the algorithms in Sections 2 and 3 can be integrated into a unified methodology to solve the hybrid model of the form

$$\begin{aligned} \text{(Hybrid)} : \quad & \max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \\ & \text{s.t.} \quad \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \quad \forall \mathbf{c} \in C. \end{aligned} \quad (29)$$

For a given subset of scalarization vectors $\tilde{C} := \{\tilde{\mathbf{c}}^1, \dots, \tilde{\mathbf{c}}^{\tilde{L}}\} \subset C$ a relaxed master problem (RMP-H) is given by

$$\max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \quad (30a)$$

$$\text{s.t.} \quad \text{CVaR}_\alpha((\tilde{\mathbf{c}}^\ell)^\top \mathbf{G}(\mathbf{z})) \geq \text{CVaR}_\alpha((\tilde{\mathbf{c}}^\ell)^\top \mathbf{Y}), \quad \forall \ell \in [\tilde{L}]. \quad (30b)$$

We can represent the constraints (30b) by linear inequalities, leading to the following equivalent reformulation of RMP-H:

$$\begin{aligned} \max \quad & \min_{\mathbf{c} \in C} \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \\ \text{s.t.} \quad & \tilde{\eta}_r - \frac{1}{\alpha} \sum_{i \in [n]} p_i \tilde{w}_{ri} \geq \text{CVaR}_\alpha((\tilde{\mathbf{c}}^r)^\top \mathbf{Y}), \quad \forall r \in [\tilde{L}] \\ & \tilde{w}_{ri} \geq \tilde{\eta}_r - (\tilde{\mathbf{c}}^r)^\top \mathbf{g}_i(\mathbf{z}), \quad \forall r \in [\tilde{L}], i \in [n] \\ & \tilde{\mathbf{w}} \in \mathbb{R}_+^{\tilde{L} \times n}, \quad \tilde{\boldsymbol{\eta}} \in \mathbb{R}_+^{\tilde{L}}, \quad \mathbf{z} \in Z. \end{aligned}$$

As discussed in Section 2.2, we can handle the maximin type objective function of interest using a finitely convergent delayed cut generation algorithm. In this spirit, suppose now that $\hat{C} = \{\hat{\mathbf{c}}^1, \dots, \hat{\mathbf{c}}^L\} \subset C$ is a given subset of scalarization vectors used to calculate the worst-case CVaR. In line with the formulation given in (10), RMP-H takes the following form:

$$\max \quad \psi \quad (31a)$$

$$\text{s.t.} \quad \tilde{\eta}_r - \frac{1}{\alpha} \sum_{i \in [n]} \tilde{p}_i \tilde{w}_{ri} \geq \text{CVaR}_\alpha((\tilde{\mathbf{c}}^r)^\top \mathbf{Y}), \quad \forall r \in [\tilde{L}] \quad (31b)$$

$$\tilde{w}_{ri} \geq \tilde{\eta}_r - (\tilde{\mathbf{c}}^r)^\top \mathbf{g}_i(\mathbf{z}), \quad \forall r \in [\tilde{L}], i \in [n] \quad (31c)$$

$$\psi \leq \eta_\ell - \frac{1}{\alpha} \sum_{i \in [n]} p_i w_{\ell i}, \quad \forall \ell \in [L], i \in [n] \quad (31d)$$

$$w_{\ell i} \geq \eta_\ell - (\hat{\mathbf{c}}^\ell)^\top \mathbf{g}_i(\mathbf{z}), \quad \forall \ell \in [L], i \in [n] \quad (31e)$$

$$\tilde{\mathbf{w}} \in \mathbb{R}_+^{\tilde{L} \times n}, \quad \mathbf{w} \in \mathbb{R}_+^{L \times n}, \quad \tilde{\boldsymbol{\eta}} \in \mathbb{R}_+^{\tilde{L}}, \quad \psi \in \mathbb{R}, \quad \mathbf{z} \in Z. \quad (31f)$$

Given a solution to the RMP-H (31), two types of cut generation problems are solved to identify if the current solution is optimal or if there is a scalarization vector $\mathbf{c} \in C$ for which at least one of the following constraints is violated: (10b) and (29). As discussed in Section 2.2, for minimizing the worst-case CVaR, it is sufficient to consider the extreme points of C . On the other hand, for the multivariate CVaR relation, it is sufficient to consider the finitely many \mathbf{c} vectors obtained as the projections of the vertices of the higher dimensional polyhedron $P(C, \mathbf{Y})$ given by (Noyan and Rudolf, 2013)

$$P(C, \mathbf{Y}) = \{(\mathbf{c}, \boldsymbol{\eta}, \mathbf{w}) \in C \times \mathbb{R} \times \mathbb{R}_+^m : w_l \geq \eta - \mathbf{c}^\top \mathbf{y}_l, \quad l \in [m]\}. \quad (32)$$

Thus, generating the violated constraints associated with those particular vertex scalarization vectors at each iteration guarantees the finite convergence of the delayed cut generation algorithm of (Hybrid). In

other words, the provable finite convergence depends on finding a solution to the cut generation problems (**CutGen – Robust**) and (**CutGen – Benchmark**), which is an extreme point of C and the projection of a vertex of $P(C, \mathbf{Y})$, respectively. In Section 2.3, we discuss how to obtain a vertex optimal solution of (**CutGen – Robust**) from an optimal solution obtained by solving one of its MIP formulations (such as (13)). For obtaining the desired vertex optimal solution of (**CutGen – Benchmark**), we refer to Noyan and Rudolf (2013).

REMARK 4.1 *The finitely convergent delayed cut generation algorithms for (**W – CVaR**) and (**Hybrid**) are also valid even if we relax the assumption that $G_j(\mathbf{z})$ is concave in $\mathbf{z} \in Z$ for all $j \in [d]$. Under the more general setting, the claim that it is sufficient to consider the extreme points of C for minimizing the worst-case CVaR follows from the finite representation of the multivariate CVaR relation (see (32)) with $\mathbf{Y} \equiv \mathbb{0}$. In this special case, it is easy to see that the projections of the vertices of the polyhedron $P(C, \mathbf{Y})$ in (32) coincide with the vertices of the set C , as desired.*

5. Computational study In the first part of our computational study, we investigate the value of the proposed (**W – CVaR**) model with respect a robust risk-neutral model and a multivariate CVaR-constrained model. We also report on the performance of the cut generation algorithm for the (**W – CVaR**) model. In the second part, we demonstrate the computational effectiveness of the MIP formulations developed (in Section 3) for the cut generation problem arising in multivariate CVaR-constrained optimization models.

5.1 Worst-case Multivariate CVaR Optimization We explore the effectiveness of the proposed (**W – CVaR**) model by applying it to a homeland security budget allocation (HSBA) problem (Hu et al., 2011). This problem studies the allocation of a fixed budget to ten urban areas, which are classified in three groups: 1) *higher risk*: New York; 2) *medium risk*: Chicago, San Francisco Bay Area, Washington DC-MD-VA-WV, and Los Angeles-Long Beach; 3) *lower risk*: Philadelphia PA-NJ, Boston MA-NH, Houston, Newark, and Seattle-Bellevue-Everett. The risk share of each area is measured based on four criteria: 1) *property losses*, 2) *fatalities*, 3) *air departures* and 4) *average daily bridge traffic*. To represent the inherent randomness a random risk share matrix $A : \Omega \rightarrow \mathbb{R}_+^{4 \times 10}$ is considered, where A_{ij} denotes the proportion of losses in urban area j relative to the total losses for criterion i . The set $Z = \{\mathbf{z} \in \mathbb{R}_+^{10} : \sum_{j \in [10]} z_j = 1\}$ represents all the feasible allocations and the associated random performance measures of interest are specified based on a particular type of penalty function for allocations under the risk share. The negatives of these budget misallocations associated with each criterion are used to construct the random outcome vector $\mathbf{G}(\mathbf{z}) = (G_1(\mathbf{z}), \dots, G_4(\mathbf{z}))$, as given below, in order to be consistent with our setup where the larger values of the random variables are preferred:

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

Hu et al. (2011) model this HSBA problem using optimization under multivariate polyhedral SSD constraints based on two benchmarks: one based on average government allocations (Department of Homeland Security’s Urban Areas Security Initiative) - denoted by $\mathbf{G}(\mathbf{z}^G)$, and one based on the suggestions in the RAND report (Willis et al., 2005) - denoted by $\mathbf{G}(\mathbf{z}^R)$. On the other hand, Noyan and Rudolf (2013) replace the SSD constraints with CVaR-based ones, leading to the following optimization model:

$$\max \min_{\mathbf{c} \in C} \mathbb{E}(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \tag{33a}$$

$$\text{s.t. } \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^R)), \quad \forall \mathbf{c} \in C \quad (33b)$$

$$\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^G)), \quad \forall \mathbf{c} \in C \quad (33c)$$

$$\mathbf{z} \in Z. \quad (33d)$$

We benchmark our model ($\mathbf{W} - \mathbf{CVaR}$), defined in (5), against two relevant existing models: the first one, which we refer to as ($\mathbf{B} - \mathbf{CVaR}$), is obtained from (33) by dropping (33c) (the government benchmark is ignored for simplicity), and the second one is the risk-neutral counterpart of our model (Hu and Mehrotra, 2012):

$$(\mathbf{W} - \mathbf{Exp}) : \max_{\mathbf{z} \in Z} \min_{\mathbf{c} \in C} \mathbb{E}(\mathbf{c}^\top \mathbf{G}(\mathbf{z})).$$

We follow the data generation scheme described in Noyan and Rudolf (2013) and consider their “base case” scalarization set given by $C = C_{\text{Base}} := \{\mathbf{c} \in \mathbb{R}_+^4 : \sum_{i \in [4]} c_j = 1, c_j \geq c_j^* - \frac{\theta}{3}, j \in [4]\}$, where $\mathbf{c}^* = (1/4, 1/4, 1/4, 1/4)$ and $\theta = 0.25$. Additionally, we also consider a second choice of C , which involves the so-called *ordered preferences* as follows: $C = C_{\text{Ord}} := \{\mathbf{c} \in \mathbb{R}_+^4 : \sum_{i \in [4]} c_j = 1, c_2 \geq c_1 \geq c_3 \geq c_4\}$. This choice relies on the assumption that the second criterion (based on fatalities) is the most important one, followed by the first criterion (based on property losses), the third criterion (based on air departures) and the fourth criterion (based on average daily bridge traffic). For further details on data generation, we refer to Hu et al. (2011) and Noyan and Rudolf (2013).

In our benchmarking analysis, we consider the equal probability case, set $n = 500$ and obtain the results for three models ($\mathbf{W} - \mathbf{CVaR}$), ($\mathbf{W} - \mathbf{Exp}$), and ($\mathbf{B} - \mathbf{CVaR}$) under each value of $\alpha \in \{0.05, 0.1, 0.15\}$. The results on allocation decisions - averaged over ten randomly generated instances - are reported in Table 1. As seen from these results, for each setting, ($\mathbf{B} - \mathbf{CVaR}$) provides solutions that allocate most of the budget (at least 51%) to the area with the highest risk (New York). This is primarily due to that fact that New York has a large (58.61%) allocation in the RAND benchmark. On the other hand, the budget percentage allocated to the five urban areas with lower risk cities is less than 20 and 12 for the scalarization sets C_{Base} and C_{Ord} , respectively. Since the set C_{Ord} involves the scalarization vectors giving more priority to the second criterion (based on fatalities), ($\mathbf{B} - \mathbf{CVaR}$) suggests to allocate even more budget to New York, the most populated area with a significantly higher risk share associated with fatalities; for the raw data for fatalities and the remaining three criterion see Table 1 in Hu et al. (2011). As expected, the allocation decisions obtained by ($\mathbf{B} - \mathbf{CVaR}$) with benchmarking constraints are sensitive to the particular benchmark allocations. On the other hand, the robust risk-neutral model ($\mathbf{W} - \mathbf{Exp}$) provides a more “averaged” solution compared to ($\mathbf{B} - \mathbf{CVaR}$) and ($\mathbf{W} - \mathbf{CVaR}$). For both choices of the scalarization set, ($\mathbf{W} - \mathbf{Exp}$) always allocates more budget to the areas with medium risk compared to the other models. For example, for the instances with C_{Ord} and $\alpha = 0.05$, it allocates over five percent more budget to such areas than ($\mathbf{W} - \mathbf{CVaR}$), and this behavior is also observed under the other settings. The results of ($\mathbf{W} - \mathbf{Exp}$) are consistent with its “risk-neutral” nature.

Finally, we would like to emphasize that ($\mathbf{W} - \mathbf{CVaR}$) allocates more budget to the areas with lower risk compared to the other models. In particular, for the instances with the scalarization set C_{Base} , ($\mathbf{W} - \mathbf{CVaR}$) allocates on average close to three percent more budget to such areas than ($\mathbf{W} - \mathbf{Exp}$). These results are consistent with the risk-averse perspective of ($\mathbf{W} - \mathbf{CVaR}$). Moreover, it is much less conservative than ($\mathbf{B} - \mathbf{CVaR}$) with respect to its allocation to New York.

In summary, in the presence of multiple criteria with ambiguous weights, we recommend the use of the risk-neutral model ($\mathbf{W} - \mathbf{Exp}$) if the decision makers are interested in the average performance. This expectation-

based decision making approach is justified if the uncertain environment does not feature extreme events, or by the Law of Large Numbers if the same decisions are made repeatedly under similar conditions. Otherwise, when it is essential to hedge against a potentially high level of random variability, we recommend that the decision makers use $(\mathbf{B} - \mathbf{CVaR})$ or $(\mathbf{W} - \mathbf{CVaR})$ depending on the availability of a reasonable benchmark random outcome vector – often based on a benchmark decision. More specifically, $(\mathbf{B} - \mathbf{CVaR})$ should be used in the presence of a benchmark to be outperformed, while $(\mathbf{W} - \mathbf{CVaR})$ is the model of preference if the decision makers do not have an alternative decision to benchmark against.

We next provide some insights about the solution times of our model $(\mathbf{W} - \mathbf{CVaR})$ for the instances under consideration. All computations in this study are performed on a 64-bit Windows Server 2012 R2 Datacenter with 2.40GHz Intel Xeon CPU E5-2630 processor with 32 GB RAM, unless otherwise stated. The vertices of both types of scalarization sets are known and there are only four of them. Thus, we could easily solve $(\mathbf{W} - \mathbf{CVaR})$ using the compact LP formulation (10). For the HSBA instances with C_{Base} , $\alpha = 0.1$, and $n = 500$, it takes at most 20 seconds to obtain an optimal solution; even for $n = 5000$ it takes at most 60 seconds. We observe that while the cut generation algorithm we propose is only essential for cases where the number of extreme points of C is exponential, it could also be useful in cases where the number of extreme points is small. For example, for C_{Base} , the compact LP takes 200 seconds on average for the three hardest instances with $n = 5000$ and $\alpha = 0.15$, whereas the cut generation algorithm takes on average 20 seconds, and generates only three extreme points of C_{Base} . This difference in solution times can be due to the large number of scenario dependent constraints and variables introduced in (10b)-(10c) for each extreme point of C .

		Allocations (%) for Areas with Higher Risk Medium Risk Lower Risk			Allocations (%) for Areas with Higher Risk Medium Risk Lower Risk		
		Base Polytope (C_{Base})			Ordered Preferences (C_{Ord})		
$\alpha = 0.05$	$(\mathbf{W} - \mathbf{CVaR})$	31.49	36.31	32.20	52.72	28.71	18.57
	$(\mathbf{B} - \mathbf{CVaR})$	51.50	29.00	19.50	56.92	31.29	11.79
$\alpha = 0.10$	$(\mathbf{W} - \mathbf{CVaR})$	31.01	36.58	32.41	51.71	30.02	18.27
	$(\mathbf{B} - \mathbf{CVaR})$	51.98	30.30	17.72	56.89	31.44	11.67
$\alpha = 0.15$	$(\mathbf{W} - \mathbf{CVaR})$	32.20	34.24	33.56	51.50	30.68	17.82
	$(\mathbf{B} - \mathbf{CVaR})$	51.50	31.16	17.34	56.82	31.35	11.83
$(\mathbf{W} - \mathbf{Exp})$		32.15	37.82	30.03	49.18	34.15	16.67
RAND Benchmark		58.61	34.31	7.07	58.61	34.31	7.07

Table 1: Model benchmarking results for the HSBA data with $n = 500$

5.2 Multivariate Polyhedral CVaR-Constrained Optimization In this section, we perform a detailed analysis on comparing the computational performance of the alternative MIP formulations of $(\mathbf{CutGen} - \mathbf{Benchmark})$ under equal probabilities. Note that for the HSBA instances, \mathbf{Y} is already well-defined since the benchmark allocations are given. To obtain the realizations of the random vector \mathbf{X} , we solve the corresponding RMP-B once, and use its optimal solution to calculate the realizations of the associated 4-dimensional random vector \mathbf{X} .

All the optimization problems are modeled with the AMPL mathematical programming language. All runs were executed 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 unless otherwise stated. CPLEX 12.2 is invoked with its default set of options and parameters. If optimality is not proven within the time allotted, we record both the best lower bound on the optimal objective value (retrieved from CPLEX and denoted by LB) and the best available objective value (denoted by UB). Since the optimal objective function can take any value including 0, we report the following relative optimality gap: $ROG = |LB - UB| / (|LB|)$. We report the results averaged over two instances with different benchmarks (based on Government and RAND benchmarks) for each combination of α and n . In all the tables in this section, the “Time” column reports the average solution time and the “B&B Nodes” column collects the average number of nodes used during the branch-and-cut process.

One can obtain slightly different versions of the presented MIP formulations by applying the RLT techniques for different types of available information (such as the valid lower and upper bounds on the scalarization vectors). We next provide the alternative MIP formulations of (**CutGen – Benchmark**) for which we report results in Tables 2-3.

- (**MIP – CVaR**): The best available benchmark model proposed by Küçükyavuz and Noyan (2016); it is based on the VaR representation (12) and its formulation is given by (22). For further computational enhancements, we added the valid inequality (17), and deleted the set of Big-M constraints (12d).
- (**MIP_VaR_Special**): This new formulation is also based on the VaR representation (12) but it is valid for the case of equal probabilities. Its formulation is given in (28); (12d) is deleted as in (**MIP – CVaR**).
- (**MIP – Special**): This new model is obtained by using the RLT-based strengthening for (23). The formulation (24) involves the inequalities obtained by applying the RLT procedure based on the unit simplex condition and the upper bounding constraints. We also apply the RLT procedure based on the lower bounding information ($c_j \geq L_j^c$, $j \in [d]$), which provides the following valid inequalities:

$$\gamma_{ij} \geq L_j^c \beta_i, \quad \forall i \in [n], j \in [d], \quad (34)$$

$$-\gamma_{ij} + c_j \geq L_j^c (1 - \beta_i), \quad \forall i \in [n], j \in [d]. \quad (35)$$

Unless stated otherwise, (**MIP – Special**) refers to the formulation obtained by adding the constraints (34)-(35) to (24).

From Remark 2.1 for the unit simplex case, we drop the redundant constraints (those obtained from the upper and lower bounding information). In Table 4, “Base Special” refers to the model obtained from (**MIP – Special**) by deleting the constraints (14c)-(14d) and (34)-(35); it only involves the most effective constraints (obtained from the unit simplex condition).

From Table 2, we can see that (**MIP – Special**) solves a majority of the test instances in the shortest amount of time. However, there are some instances (for example, for HSBA data, unit simplex, $\alpha = 0.01, n = 1000, 1500$) for which (**MIP – Special**) only solves one out of the four instances within the time limit as opposed to (**MIP_VaR_Special**) which solves three of the instances within the limit. In addition, both of the new formulations we propose significantly outperform the existing formulation (**MIP – CVaR**) for the equal probability case.

Furthermore, we can apply the computational enhancements proposed in Küçükyavuz and Noyan (2016) to the proposed formulations, namely variable fixing, bounding and a class of valid inequalities referred to as the

		(MIP – CVaR)		(MIP_VaR_Special)		(MIP – Special)	
		Time;	B&B	Time;	B&B	Time;	B&B
		[ROG]	Nodes	[ROG]	Nodes	[ROG]	Nodes
Base Polytope & $\alpha = 0.01$							
n	1500	705.8	1524.2	447.1	1781.9	3.1	0.1
	2000	1225.1	3095.9	†1840 [50]	3510.0	6.4	0.0
	2500	†2313 [50]	5439.5	†1846 [50]	2995.9	9.1	0.0
	3000	†2275 [50.5]	3712.3	†1970 [50]	2658.1	10.1	0.0
	5000	†[1415.8]	4594.0	†[56.2]	6170.4	34.7	0.0
Base Polytope & $\alpha = 0.05$							
n	500	109.5	1422.6	109.6	1514.2	0.7	0.0
	1000	1667.5	11976.4	†1829 [50]	8823.9	2.9	0.0
	1500	†[100.3]	14627.9	†2316 [50]	7430.1	9.2	0.5
	2000	†[451.9]	9696.0	†3071 [50]	8230.4	8.5	1.3
	2500	†[174]	5998.7	†[57.2]	6928.7	†1837 [50]	2028.8
	3000	◊	1008.0	◊[21.4]	4073.8	95.6	18.7
	5000	†[145.5]	2901.7	†[77]	1332.6	†[50.2]	1047.9
Unit Simplex & $\alpha = 0.01$							
n	500	96.1	546.5	60.5	441.8	4.7	73.9
	1000	1122.2	4388.1	1130.6	4139.8	†2048 [42.6]	26341.8
	1500	†[129.6]	6424.9	†3245 [50]	7059.8	†[81.3]	29985.7
	2000	†[108.5]	5910.1	†[159.7]	5327.7	†[215.9]	22330.7
	2500	†[106.2]	1704.0	†[118.2]	4784.6	†[188]	19352.6
Unit Simplex & $\alpha = 0.05$							
n	300	211.4	2796.1	186.3	2833.4	†2001 [42.2]	71177.5
	500	†2425 [145.3]	17459.6	1854.4	15871.0	†[125.1]	54251.9

†: Time limit with integer feasible solution.

◊: Time limit with no integer feasible solution.

B&B Nodes are reported in hundreds.

Table 2: Computational performance of the alternative MIPs for (CutGen – CVaR) under equal probabilities

ordering inequalities (on the β variables). The variable fixing method recognizes scenarios which are guaranteed to be larger than VaR, and fixes the corresponding β variables to zero. In particular, for each $k \in [n]$, a set $L_k := \{i \in [n] \setminus k : \max_{\mathbf{c} \in C} \{\mathbf{c}^\top (\mathbf{x}_i - \mathbf{x}_k)\} < 0\}$ (resp., $H_k := \{i \in [n] \setminus k : \max_{\mathbf{c} \in C} \{\mathbf{c}^\top (\mathbf{x}_k - \mathbf{x}_i)\} < 0\}$) is defined to denote the scenarios that are guaranteed to return lower (resp., higher) scalarized outcomes than scenario k . Note the correction in the definition of the set L_k (resp., H_k) compared to that in Küçükyavuz and Noyan (2016), where we exclude scenario i with $\max_{\mathbf{c} \in C} \{\mathbf{c}^\top (\mathbf{x}_i - \mathbf{x}_k)\} = 0$ (resp., $\max_{\mathbf{c} \in C} \{\mathbf{c}^\top (\mathbf{x}_k - \mathbf{x}_i)\} = 0$) from the set L_k (resp., H_k). In addition, for the existing MIP (23) and (MIP – Special), we introduce upper and lower bounds on $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, for the others which involve the z decision variable (representing the VaR) we introduce upper and lower bounds on $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Table 3 summarizes our computational experience with using these enhancements. The ‘Remaining Bin. Var.’ column reports remaining percentage of binary variables after the preprocessing, and the ‘# of Order. Ineq.’ column represents the number of ordering inequalities added to the formulations. Observe that there is a significant reduction in the number of binary variables. Furthermore, many ordering inequalities are added to strengthen the formulation. As a result,

instances that were not solvable to optimality by any of the methods (reported in Table 2) can now be solved to optimality with at least one of the new formulations. We would also like to note that the total time spent on preprocessing (for calculating the Big-M coefficients and handling all the enhancements - fixing, bounding, and ordering inequalities), which is not included in the times reported, is negligible.

n	Existing-Special (23)		(MIP – CVaR)		(MIP_VaR_Special)		(MIP – Special)		Remaining Binary Var. (%)	# of Order. Ineq.
	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes		
n	Base Polytope & $\alpha = 0.01$									
1500	141.4	2580.9	2.9	23.4	1.8	9.8	0.3	0.0	3.6	353.0
2000	†[820.36]	19786.1	16.0	112.5	5.2	33.3	0.7	0.0	4.5	745.0
2500	†[290.89]	14933.2	23.0	160.8	15.2	74.2	1.0	0.0	4.3	1168.5
3000	†[422.16]	13630.5	42.3	245.3	24.3	99.4	1.0	0.0	4.4	1908.5
5000	†[221.54]	5731.1	366.3	1707.8	149.5	521.9	6.7	0.1	4.3	4960.0
n	Base Polytope & $\alpha = 0.05$									
300	106.5	4128.0	1.1	23.2	0.7	5.5	0.1	0.0	19.2	325.0
500	†2175 [71.3]	43351.7	5.3	58.5	1.4	9.7	0.3	0.0	16.2	818.0
1000	†[739.6]	28977.7	55.6	544.3	16.2	122.4	1.4	0.1	15.0	2959.0
1500	†[441]	17373.6	470.9	3461.1	90.9	415.8	2.6	0.0	15.7	7620.5
2000	†[351]	9858.6	768.7	3538.1	171.9	510.9	5.3	0.0	14.9	12656.5
2500	†[269.9]	3967.2	†3021 [9.2]	9700.8	813.0	1932.5	13.6	0.6	15.8	22194.5
3000	†[253.1]	2399.3	†[66.5]	8618.1	1029.1	1809.3	20.7	0.3	15.9	30857.0
5000	†[250.9]	632.4	†[138.5]	2248.4	†[74.7]	1671.8	42.7	0.3	15.5	84272.0
n	Unit Simplex & $\alpha = 0.01$									
500	788.6	32724.9	5.8	83.5	2.4	41.0	0.6	11.3	22.2	170
1000	†[106]	49324.4	105.1	788.0	81.8	640.7	†1847 [14.7]	37653.9	23.9	771.5
1500	†[105.7]	29495.4	374.5	2054.1	290.1	1631.1	†2002 [49.3]	21375.9	21.9	1837.5
2000	†[106.3]	19970.0	1706.1	7224.5	1172.6	4874.4	†[233.3]	27470.6	22.2	3237.5
2500	†[105]	14320.9	†[393.8]	9854.0	†3469 [50]	10137.6	†[1834.8]	15418.6	22.9	5386.5
n	Unit Simplex & $\alpha = 0.05$									
300	†[107]	72202.6	55.2	946.3	39.9	720.5	132.7	4143.7	52.8	761.5
500	†[107.1]	35273.8	473.6	5888.6	352.4	4430.1	†2756 [44.2]	33873.7	53.3	2049

†: Time limit with integer feasible solution.

B&B Nodes are reported in hundreds.

Table 3: Computational performance of the alternative enhanced MIPs (fixing, bounding, ordering inequalities) for (CutGen – CVaR) under equal probabilities

Next, we use the additional information on C to obtain stronger RLT formulations. Our experiments are reported in Table 4, for the scalarization sets C_{Base} and C_{Ord} . We observe that the RLT-based strengthening using only the unit simplex information (15b)-(15d), reported in the column titled Base Simplex, is not very effective. Recall (Remark 2.1) that when there exists an index $j \in [d]$ such that $\tilde{M}_j = \max\{c_j : \mathbf{c} \in C\} < 1$, the constraints (14c)-(14d) are not redundant for (MIP – Special). In fact, for the HSBA instances, including these inequalities in (MIP – Special) leads to a significant reduction in the computational time as reported in the second column of Table 4. It is surprising to observe that (MIP – Special) could solve some instances

in very short CPU time, while it reaches the time limit when (14c)-(14d) are dropped.

When we have the extreme points of C , we can easily obtain the upper and lower bounds on the components of \mathbf{c} . For C_{Ord} including the ordered preference constraints $c_j \geq c_{j+1}$, we obtain the corresponding inequalities obtained by using the RLT (see Proposition 2.5):

$$\gamma_{ij} \geq \gamma_{ij+1}, \quad \forall i \in [n], j \in [d-1], \quad (36)$$

$$\gamma_{ij} - \gamma_{ij+1} \leq c_j - c_{j+1}, \quad \forall i \in [n], j \in [d-1]. \quad (37)$$

In addition, for this case, $\tilde{\mathbf{M}} = (1, 1/2, 1/3, 1/4)$ and $\mathbf{L}^c = (1/4, 0, 0, 0)$. In our computational experiments reported in Table 4, we use the RLT strengthening of the upper bounding inequalities and the ordered preference constraints defining C_{Base} .

Table 4 demonstrates that the most effective solution method for cut generation under equal probabilities is to use the formulation (**MIP – Special**) with all enhancements: fixing, bounding, ordering inequalities on β , and the RLT-based strengthening using the additional inequalities defining C .

6. Conclusions In this paper, we study risk-averse models for multicriteria optimization problems under uncertainty. First, we introduce a model that optimizes the *worst-case multivariate CVaR*, and develop a finitely convergent delayed cut generation algorithm for finite probability spaces. In addition, for the cut generation problem, which is in general a mixed-integer program, we give a stronger formulation for the equiprobable case using the reformulation linearization technique. Next, we observe that similar polyhedral enhancements are also useful for a related class of *multivariate CVaR-constrained* optimization problems that has attracted attention recently. Our computational study demonstrates the effectiveness of the proposed solution methods for both classes of models.

Acknowledgments We thank the two referees and the associate editor for their valuable comments that improved the presentation. Simge Küçükyavuz and Xiao Liu are supported, in part, by National Science Foundation Grants 1055668 and 1537317. Nilay Noyan acknowledges the support from Bilim Akademisi - The Science Academy, Turkey, under the BAGEP program.

Appendix A. Stochastic dominance In this section, we review the well-known stochastic dominance relations, which are essential for the stochastic Pareto optimality definitions presented in Section 2.

The stochastic dominance relations are fundamental concepts in comparing random variables (Mann and Whitney, 1947; Lehmann, 1955) and have been widely used in economics and finance (see, e.g., Levy, 1992). Different from the approaches based on risk measures, in a stochastic dominance based approach, the random variables are compared by a point-wise comparison of some performance functions (constructed from their distribution functions when the order is greater than zero). We note that the lower order dominance relations ($k = 0, 1$, and 2) are the most common ones (referred to as ZSD, FSD, and SSD, respectively). We provide the formal definitions below and refer the reader to Müller and Stoyan (2002) and Shaked and Shanthikumar (1994) for further details.

- We say that a random variable X *dominates* another random variable Y in the *zeroth order* if $X \geq Y$ everywhere, i.e., $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$.
- An integrable random variable X *dominates* another integrable Y in the *first order* (or X is stochas-

	Base Special	Base & UB Ineq.	Base & UB & LB or RLT Ord. Ineq.	Base Special	Base & UB Ineq.	Base & UB & LB or RLT Ord. Ineq.	Remaining Binary Var. (%)	# of Order. Ineq.
	Time; [ROG] Nodes	Time; [ROG] Nodes	Time; [ROG] Nodes	F&B&Order. Ineq. Time; [ROG] Nodes	F&B&Order. Ineq. Time; [ROG] Nodes	F&B&Order. Ineq. Time; [ROG] Nodes	Binary Var. (%)	# of Order. Ineq.
Base Polytope & $\alpha = 0.01$ (Time limit=1800 sec.)								
n								
1500	†921.7 [6.6] 2832.6	3.6 0.2	3.4 0.1	1.2 4.3	0.6 0.3	0.2 0.0	3.6	353
2000	†942.5 [18.7] 2225.0	6.8 0.0	7.2 0.0	18.3 102.4	1.3 0.2	0.4 0.0	4.5	745
2500	†1642 [29.6] 2234.7	7.7 0.0	10.1 0.0	5.1 8.2	1.5 0.0	0.7 0.0	4.3	1168.5
3000	†[87.5] 2052.9	15.5 0.0	12.8 0.0	57.9 247.4	2.2 0.0	0.7 0.0	4.4	1908.5
5000	†[115.1] 832.5	146.3 12.2	37.4 0.0	†908 [4.2] 1848.6	10.3 0.3	5.7 0.1	4.3	4960
Base Polytope & $\alpha = 0.05$ (Time limit=1800 sec.)								
n								
500	†902 [5.3] 3734.6	1.0 0.0	0.8 0.0	1.3 3.2	0.4 0.0	0.2 0.0	16.2	818
1000	†1241 [15.9] 3432.8	7.9 12.3	3.8 0.0	169.9 681.4	2.5 0.2	1.3 0.1	15.0	2959
1500	†[72.5] 2531.5	28.3 12.1	10.3 0.5	†903 [1.9] 1583.6	8.0 4.0	2.7 0.0	15.7	7620.5
2000	†930 [33.1] 758.9	56.8 12.2	12.3 1.3	†906 [2.2] 790.7	11.0 0.1	6.0 0.0	14.9	12656.5
2500	†[83.5] 931.8	104.6 27.7	†932 [50] 674.0	†929 [5.4] 714.3	29.0 2.0	14.7 0.6	15.8	22194.5
3000	†[87.8] 786.5	317.0 120.7	83.6 18.7	†920.8 [7.3] 431.1	43.9 1.2	22.5 0.3	15.9	30857
5000	†[91.3] 441.5	†[50.3] 390.6	†[50.2] 421.6	†1104 [10.4] 74.8	141.6 1.6	44.2 0.3	15.5	84272
Ordered Preference & $\alpha = 0.01$ (Time limit=3600 sec.)								
n								
1000	†1809 [50] 21268.5	995.8 4017.4	242.4 746.5	7.8 219.0	4.7 70.1	3.8 61.9	7.6	314.0
1500	†[49.6] 26700.0	†1843 [45.9] 4786.3	†1802 [37.9] 3687.5	93.0 1605.9	36.4 444.6	42.5 521.8	7.2	687.5
2000	†1832 [49.5] 10935.4	†1873 [49] 3325.9	†1803 [48.5] 3643.8	1035.1 15729.2	271.2 2602.5	395.4 3575.9	7.2	1192.5
2500	†[54.4] 14035.7	†1831 [50] 2562.0	†1835 [50] 1063.8	†[50.5] 31553.8	†1914 [50] 7503.4	†1802 [50] 14374.4	7.0	1899.5
3000	† [50.2] 9335.9	†1842 [50] 1244.5	†1820 [50] 940.0	†[50.9] 29419.5	†1929 [50] 10192.9	†1803 [50] 10223.6	6.8	2398.0

†: Time limit with integer feasible solution. B&B Nodes are reported in hundreds.

UB inequalities: (14c)-(14d). LB inequalities: (34)-(35). RLT Ord. Ineq.: (36)-(37); used only for the instances with ordered preference structure - see Proposition 2.5.

Table 4: Computational performance of the RLT procedure for (CutGen - CVaR) under equal probabilities

tically larger than Y) if $F_1(X, \eta) := \mathbb{P}(X \leq \eta) \leq F_1(Y, \eta) := \mathbb{P}(Y \leq \eta)$ for all $\eta \in \mathbb{R}$.

- For $k \geq 2$ we say that a k -integrable random variable X (i.e., $\in \mathcal{L}^k$) *dominates* another k -integrable random variable Y *in the k th order* if $F_k(X, \eta) \leq F_k(Y, \eta)$ for all $\eta \in \mathbb{R}$, where $F_k(X, \eta) = \int_{-\infty}^{\eta} F_{k-1}(X, t) dt$ for all $\eta \in \mathbb{R}$.
- For $k = 0$, if $X(\omega) > Y(\omega)$ for all $\omega \in \Omega$, we will refer to the relation as “*strong ZSD*” and denote it by $X \succ_{(0)} Y$. For $k \geq 1$, if all the inequalities $F_k(X, \eta) \leq F_k(Y, \eta)$ are strict, then we refer to the relation as “*strong k SD*” and denote it by $X \succ_{(k)} Y$. We remark that the notion of “*strong k SD*” is not analogous to the notion of strict k SD, which requires that at least one of the inequalities defining the dominance relation is strict.

Appendix B. A class of facets of $\text{conv}(\mathcal{S})$. Recall that

$$\mathcal{S} := \{(\gamma, \mathbf{c}, \boldsymbol{\beta}, \eta, \mathbf{w}) \in \mathbb{R}_+^{n \times d} \times \mathbb{R}_+^d \times \{0, 1\}^n \times \mathbb{R} \times \mathbb{R}_+^m \mid \gamma = \boldsymbol{\beta} \mathbf{c}^\top, \sum_{j \in [d]} c_j = 1, \sum_{i \in [n]} \beta_i = k, \mathbf{c}^\top \mathbf{y}_l \geq \eta - w_l, \forall l \in [m]\}.$$

Before we study the facets of $\text{conv}(\mathcal{S})$, we first need to establish its dimension.

PROPOSITION B.1 *Conv*(\mathcal{S}) is a polyhedron with dimension $n + d + m - 1$.

PROOF. Note that $\text{conv}(\mathcal{S})$ is a polyhedron, because $\boldsymbol{\beta} \in \{0, 1\}^n$. Next, we show that the dimension of $\text{conv}(\mathcal{S})$ is $n + d + m - 1$. Clearly, in the original constraints defining \mathcal{S} , there are two linearly independent equalities: $\sum_{j \in [d]} c_j = 1$, $\sum_{i \in [n]} \beta_i = k$. In addition, there are nd equalities: $\gamma_{ij} = c_j \beta_i$, for all $i \in [n]$ and $j \in [d]$. Hence, $\dim(\text{conv}(\mathcal{S})) \leq n + m + d - 1$.

Consider the following set of points:

$$\begin{aligned} (\mathbf{u}_v \mathbf{e}_1^\top, \mathbf{e}_1, \mathbf{u}_v, 0, 0) & \quad \forall v \in [n], \\ (\mathbf{u}_1 \mathbf{e}_j^\top, \mathbf{e}_j, \mathbf{u}_1, 0, 0) & \quad \forall j \in [d] \setminus \{1\}, \\ (\mathbf{u}_1 \mathbf{e}_1^\top, \mathbf{e}_1, \mathbf{u}_1, 0, e_l) & \quad \forall l \in [m], \\ (\mathbf{u}_1 \mathbf{e}_1^\top, \mathbf{e}_1, \mathbf{u}_1, -1, 0), & \end{aligned}$$

where \mathbf{u}_v , for all $v \in [n]$ are any affinely independent vectors with k elements equal to 1 and the remaining elements equal to 0. These vectors exist because the dimension of the following system:

$$\boldsymbol{\beta} \in \{0, 1\}^n, \quad \sum_{i \in [n]} \beta_i = k, \quad (38)$$

is $n - 1$. Clearly, this set of points is feasible and affinely independent. In addition, the cardinality of this set is $n + m + d$. Hence, $\dim(\text{conv}(\mathcal{S})) \geq n + m + d - 1$, which completes the proof. \square

PROPOSITION B.2 For any $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, inequality (25) is facet-defining for $\text{conv}(\mathcal{S})$ if and only if $s \in [m]$, $t \in [m] \setminus \{s\}$ are such that $y_{sj} < y_{tj}$ and $y_{si} > y_{ti}$ for some $i, j \in [d]$.

PROOF. To show the necessity, we first note that if there exists a pair $s \in [m]$, $t \in [m] \setminus \{s\}$ such that $y_{sj} \geq y_{tj}$ or $y_{sj} \leq y_{tj}$ for all $j \in [d]$, in other words, when the realizations under a scenario are *dominated* by the realizations under another scenario, then the corresponding inequality (25) is dominated. To see this,

suppose that $y_{sj} \leq y_{tj}$ for all $j \in [d]$ for some pair $\forall s \in [m], \forall t \in [m] \setminus \{s\}$. Then the corresponding inequality (25) is dominated by the original inequality $\mathbf{c}^\top \mathbf{y}_s \geq \eta - w_s$, because the coefficients of γ_{ij} are $y_{tj} - y_{sj} \geq 0$, and $\gamma_{ij}, w_t \geq 0$. Now consider the case that $y_{sj} \geq y_{tj}$ for all $j \in [d]$ for some pair $\forall s \in [m], \forall t \in [m] \setminus \{s\}$. Then the corresponding inequality (25) is dominated by the original inequality $\mathbf{c}^\top \mathbf{y}_t \geq \eta - w_t$. To see this, observe that we can rewrite inequality (25) for this choice of s and t as, $\mathbf{c}^\top \mathbf{y}_t + \sum_{j \in [d]} (y_{sj} - y_{tj})(c_j - \gamma_{ij}) \geq \eta - w_t - w_s$. It is now easy to see that the inequality is dominated, because $y_{sj} - y_{tj} \geq 0$, $c_j \geq \gamma_{ij}$ and $w_s \geq 0$.

To show sufficiency, we need to show that for any given $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, there are $n + m + d - 1$ affinely independent points that satisfy (25) at equality. From the necessity condition, we only need to consider the cases for which there exists an index $j_1 \in [d]$, such that $y_{sj_1} < y_{tj_1}$, and there exists an index $j_2 \in [d]$, such that $y_{sj_2} > y_{tj_2}$. In order to simplify the notation, and without loss of generality, throughout the rest of the proof, we let $j_1 = 1$, and $j_2 = 2$, or equivalently, $y_{s1} < y_{t1}$, and $y_{s2} > y_{t2}$.

First, we construct a set of points:

$$\mathbf{PT}_v^1 = (\mathbf{u}_v \tilde{\mathbf{e}}_v^\top, \tilde{\mathbf{e}}_v, \mathbf{u}_v, \rho_v^1, \boldsymbol{\xi}_v^1), \quad \forall v \in [n], \quad (39)$$

where if $u_{vi} = 0$, then $\tilde{\mathbf{e}}_v = \mathbf{e}_1$ and $\rho_v^1 = y_{s1}$, else if $u_{vi} = 1$, then $\tilde{\mathbf{e}}_v = \mathbf{e}_2$ and $\rho_v^1 = y_{t2}$. In addition, $\xi_{vs}^1 = \xi_{vt}^1 = 0$, and $\xi_{vl}^1 = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $v \in [n]$ and $l \in [m] \setminus \{s, t\}$. Clearly, the set of points defined in (39) are affinely independent feasible points, and satisfy (25) at equality. Next, we construct a set of points:

$$\mathbf{PT}_j^2 = (\tilde{\mathbf{u}}_j \mathbf{e}_j^\top, \mathbf{e}_j, \tilde{\mathbf{u}}_j, \rho_j^2, \boldsymbol{\xi}_j^2), \quad \forall j \in [d] \setminus \{1, 2\}, \quad (40)$$

where $\tilde{\mathbf{u}}_j$ is any feasible point of (38) with $\tilde{u}_{ji} = 0$ if $y_{sj} \leq y_{tj}$, and $\tilde{u}_{ji} = 1$ otherwise (i.e., if $y_{sj} \geq y_{tj}$), for all $j \in [d] \setminus \{1, 2\}$. In addition, $\rho_j^2 = \min\{y_{sj}, y_{tj}\}$, for all $j \in [d] \setminus \{1, 2\}$. Furthermore, $\xi_{js}^2 = \xi_{jt}^2 = 0$, and $\xi_{jl}^2 = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $j \in [d] \setminus \{1, 2\}$ and $l \in [m] \setminus \{s, t\}$. It is easy to see that the set of points defined in (40) are feasible, affinely independent from (39), and satisfy (25) at equality.

Furthermore, we construct the following set of points:

$$\mathbf{PT}_s^3 = (\bar{\mathbf{u}}_1 \mathbf{e}_1^\top, \mathbf{e}_1, \bar{\mathbf{u}}_1, y_{t1}, \boldsymbol{\xi}_s^3) \quad (41a)$$

$$\mathbf{PT}_t^3 = (\bar{\mathbf{u}}_2 \mathbf{e}_2^\top, \mathbf{e}_2, \bar{\mathbf{u}}_2, y_{s2}, \boldsymbol{\xi}_t^3) \quad (41b)$$

$$\mathbf{PT}_l^3 = \mathbf{PT}_s^3 + (0, 0, 0, 0, \mathbf{e}_l), \quad \forall l \in [m] \setminus \{s, t\}, \quad (41c)$$

where $\bar{\mathbf{u}}_1$ is any feasible point of (38) with $\bar{u}_{1i} = 0$, and $\bar{\mathbf{u}}_2$ is any feasible point of (38) with $\bar{u}_{2i} = 1$. In addition, $\xi_{ss}^3 = y_{t1} - y_{s1}$, $\xi_{st}^3 = 0$, and $\xi_{sl}^3 = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $l \in [m] \setminus \{s, t\}$. Similarly, $\xi_{ts}^3 = 0$, $\xi_{tt}^3 = y_{s2} - y_{t2}$, and $\xi_{tl}^3 = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $l \in [m] \setminus \{s, t\}$. Clearly, the set of points defined by (41) are affinely independent feasible points which satisfy (25) at equality.

Finally, we construct the single point:

$$\mathbf{PT}^4 = (\mathbf{u}_1 \mathbf{c}^{*\top}, \mathbf{c}^*, \mathbf{u}_1, \eta^*, \boldsymbol{\xi}^4), \quad (42)$$

where $\mathbf{c}^* = (c_1^*, c_2^*, 0, \dots, 0)$, and the parameters (c_1^*, c_2^*, η^*) are uniquely defined by the following linear system:

$$\begin{aligned} c_1^* + c_2^* &= 1 \\ y_{s1} c_1^* + y_{s2} c_2^* &= \eta^* \\ y_{t1} c_1^* + y_{t2} c_2^* &= \eta^*, \end{aligned}$$

or equivalently, $c_1^* = \frac{y_{s2} - y_{t2}}{y_{s2} - y_{t2} + y_{t1} - y_{s1}}$, $c_2^* = 1 - c_1^*$, and $0 < c_1^*, c_2^* < 1$. In addition, $\xi_s^4 = \xi_t^4 = 0$, and $\xi_l^4 = \max\{\tilde{M}_s, \tilde{M}_t\}$, for all $l \in [m] \setminus \{s, t\}$.

Clearly, \mathbf{PT}^4 is affinely independent from the points defined by (39), since the following matrix:

$$\begin{bmatrix} 1 & 0 & y_{s1} \\ 0 & 1 & y_{t2} \\ c_1^* & c_2^* & \eta^* = y_{s1}c_1^* + y_{s2}c_2^* \end{bmatrix}, \quad (43)$$

has full rank (due to $y_{t2} < y_{s2}$). In addition, it is easy to check that (42) is affinely independent from the points defined by (40) and (41). Furthermore, it is also a feasible point which satisfies (25) at equality.

From (39)-(42), we obtain $n + m + d - 1$ affinely independent feasible points which satisfy (25) at equality. Hence, inequalities (25) are facet defining. □

References

- Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Balibek, E. and Köksalan, M. (2010). A multi-objective multi-period stochastic programming model for public debt management. *European Journal of Operational Research*, 205(1):205 – 217.
- Ben Abdelaziz, F. (2012). Solution approaches for the multiobjective stochastic programming. *European Journal of Operational Research*, 216(1):1–16.
- Ben Abdelaziz, F., Lang, P., and Nadeau, R. (1995). Distributional efficiency in multiobjective stochastic linear programming. *European Journal of Operational Research*, 85(2):399 – 415.
- Ben-Tal, A., Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press, Princeton, NJ.
- Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501.
- Borcherding, K., Eppel, T., and von Winterfeldt, D. (1991). Comparison of weighting judgments in multiattribute utility measurement. *Management Science*, 37(12):1603–1619.
- 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. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking and Finance*, 30(2):433–451.
- Dentcheva, D. and Ruszczyński, A. (2009). Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117(1):111–127.
- Dentcheva, D. and Wolfhagen, E. (2013). Optimization with multivariate dominance constraints. In Deodatis, G., Ellingwood, B., and Frangopol, D., editors, *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*. CRC Press LLC.
- Ehrgott, M. (2005). *Multicriteria Optimization*. Springer, Berlin.
- Ehrgott, M., Ide, J., and Schöbel, A. (2014). Minmax robustness for multi-objective optimization problems. *European Journal of Operational Research*, 239(1):17–31.
- Ekeland, I. and Schachermayer, W. (2011). Law invariant risk measures on $L^\infty(\mathbb{R}^d)$. *Statistics and Risk Modeling with Applications in Finance and Insurance*, 28(3):195–225.

- Gupte, A., Ahmed, S., Dey, S. S., and Cheon, M. S. (2017). Relaxations and discretizations for the pooling problem. *Journal of Global Optimization*, 67(3):631–669.
- Gutjahr, W. J. and Pichler, A. (2016). Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research*, 236(2):475–499.
- Hamel, A. H., Rudloff, B., and Yankova, M. (2013). Set-valued average value at risk and its computation. *Mathematics and Financial Economics*, 7(2):229–246.
- 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.
- Jouini, E., Meddeb, M., and Touzi, N. (2004). Vector-valued coherent risk measures. *Finance and Stochastics*, 8(4):531–552.
- Köksalan, M. and Şakar, C. T. (2016). An interactive approach to stochastic programming-based portfolio optimization. *Annals of Operations Research*, 245(1):47–66.
- Küçükyavuz, S. and Noyan, N. (2016). Cut generation for optimization problems with multivariate risk constraints. *Mathematical Programming*, 159(1):165–199.
- Lehmann, E. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, 26(3):399–419.
- Levy, H. (1992). Stochastic dominance and expected utility: Survey and analysis. *Management Science*, 38(4):555–593.
- Mann, H. and Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1):50–60.
- McCormick, G. (1976). Computability of global solutions to factorable nonconvex programs: Part I – convex underestimating problems. *Mathematical Programming*, 10(1):147–175.
- Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester.
- 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., Balcik, B., and Atakan, S. (2016). A stochastic optimization model for designing last mile relief networks. *Transportation Science*, 50(3):1092–1113.
- Noyan, N. and Rudolf, G. (2013). Optimization with multivariate conditional value-at-risk-constraints. *Operations Research*, 61(4):990–1013.
- Ogryczak, W. (2010). On robust solutions to multi-objective linear programs. In Trzaskalik, T. and Wachowicz, T., editors, *Multiple Criteria Decision Making '09*, pages 197–212.
- Ogryczak, W. and Ruszczyński, A. (2001). On consistency of stochastic dominance and mean–semideviation models. *Mathematical Programming*, 89(2):217–232.
- 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.
- Rüschendorf, L. (2013). Risk measures for portfolio vectors. In *Mathematical Risk Analysis*, Springer Series in Operations Research and Financial Engineering, pages 167–188. Springer Berlin Heidelberg.
- Saaty, T. (2000). *Decision Making for Leaders; the Analytical Hierarchy Process for Decisions in a Complex World*. RWS Publications.
- Schoemaker, P. J. H. and Waid, C. C. (1982). An experimental comparison of different approaches to determining weights in additive utility models. *Management Science*, 28(2):182–196.
- Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and their Applications*. Associated Press, Boston.
- Sherali, H. D. and Adams, W. P. (1994). A hierarchy of relaxations and convex hull representations for mixed-integer zero-one programming problems. *Discrete Appl. Math.*, 52(1):83–106.
- Sherali, H. D., Adams, W. P., and Driscoll, P. J. (1998). Exploiting special structures in constructing a hierarchy of relaxations for 0-1 mixed integer problems. *Operations Research*, 46(3):396–405.
- Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation, and Application*. John Wiley & Sons, New York.
- von Winterfeldt, D. and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge University Press.
- Willis, H. H., Morral, A. R., Kelly, T. K., and Medby, J. J. (2005). Estimating terrorism risk. Technical report, The RAND Corporation, Santa Monica, CA.
- Wozabal, D. (2014). Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research*, 62(6):1302–1315.
- Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168.