# Cut Generation for Optimization Problems with Multivariate Risk Constraints

Simge Küçükyavuz
Department of Integrated Systems Engineering, The Ohio State University, kucukyavuz.2@osu.edu
Nilay Noyan[1]
Manufacturing Systems/Industrial Engineering Program, Sabancı University, Turkey, nnoyan@sabanciuniv.edu

August 14, 2014

Abstract: We consider a class of multicriteria stochastic optimization problems that features benchmarking constraints based on conditional value-at-risk and second-order stochastic dominance. We develop alternative mixed-integer programming formulations and solution methods for cut generation problems arising in optimization under such multivariate risk constraints. We give the complete linear description of two non-convex substructures appearing in these cut generation problems. We present computational results that show the effectiveness of our proposed models and methods.

*Keywords*: stochastic programming; integer programming; multicriteria optimization; multivariate risk-aversion; conditional value-at-risk; stochastic dominance; cut generation

**1. Introduction** Multicriteria/multiobjective decision making problems in the presence of uncertainty arise in a wide range of areas, including relief network design, homeland security budget allocation, and finance. In such decision problems with multiple random performance measures of interest, comparing the potential decisions requires specifying preference relations among random vectors, where each dimension of a vector corresponds to a decision criterion. Moreover, it is often crucial to take into account decision makers' risk preferences. Incorporating such stochastic multivariate preference relations into optimization models is a fairly recent research area. The existing models feature benchmarking preference relations as constraints, requiring the decision-based random vectors to be preferred (according to the specified preference rules) to some benchmark random vectors. The literature mainly focuses on two types of multivariate risk-averse preference relations: multivariate relations based on second-order stochastic dominance (SSD) and conditional value-at-risk (CVaR).

The SSD relation has received significant attention due its correspondence with risk-averse preferences (Hadar and Russell, 1969). In this regard, the majority of existing studies on optimization models with multivariate risk constraints extend the univariate SSD rule to the multivariate case (Dentcheva and Ruszczyński, 2009; Homem-de-Mello and Mehrotra, 2009; Hu et al., 2012; Dentcheva and Wolfhagen, 2013). In this line of research, scalar-based preferences are extended to vector-valued random variables by considering a family of linear scalarization functions and requiring that all scalarized versions of the random vectors conform to the specified univariate preference relation. Scalarization coefficients can be interpreted as weights representing the subjective importance of each decision criterion. Thus, the scalarization approach is closely related to the weighted sum method, which is widely used in multicriteria decision making (see, e.g., Ehrgott, 2005). It is also important to point out that enforcing a preference relation over a family of scalarization vectors allows us to represent a wider range of views and involve differing opinions of multiple experts (for motivating discussions see, e.g., Hu and Mehrotra, 2012). Dentcheva and Ruszczyński (2009) consider linear scalarization with all nonnegative coefficients[2], and provide a theoretical background for the multivariate SSD-constrained problems. On the other hand, Homem-de-Mello and Mehrotra (2009) and Hu et al. (2012) allow arbitrary

[2]This set of scalarization vectors can be truncated to a unit simplex.

polyhedral and arbitrary convex scalarization sets, respectively. There are a few studies (Armbruster and Luedtke, 2014; Haskell et al., 2013) which consider the multivariate SSD relation based on multi-dimensional utility functions instead of using scalarization functions.

While the existing literature mainly focuses on enforcing multivariate stochastic dominance relations, this approach can be overly conservative in practice and often leads to infeasible formulations; for detailed discussions we refer the reader to Noyan and Rudolf (2013). As an alternative, Noyan and Rudolf (2013) propose the use of constraints based on coherent risk measures, which provide sufficient flexibility to lead to feasible problem formulations while still being able to capture a broad range of risk preferences. In particular, they focus on the widely applied risk measure CVaR and replace the SSD relation by a collection of CVaR constraints at various confidence levels. This is a very natural relaxation (as also pointed out in Fábián et al., 2011), due to the well-known fact that the univariate SSD relation is equivalent to a continuum of CVaR inequalities (Dentcheva and Ruszczyński, 2006). Other attempts to weaken the univariate stochastic dominance relations have resulted in the concept of almost stochastic dominance (Leshno and Levy, 2002; Lizyayev and Ruszczyński, 2011). In terms of the scalarization set, Noyan and Rudolf (2013) follow the line of research of Homem-de-Mello and Mehrotra (2009), and define the multivariate CVaR constraints based on polyhedral sets; this modeling approach strikes a good balance between tractability and flexibility.

Optimization models with univariate stochastic dominance constraints are generally known to be computationally challenging, which can partially be attributed to the potentially large number of scenario-dependent variables and constraints (see, e.g., Luedtke, 2008; Noyan and Ruszczyński, 2008; Dentcheva et al., 2010; Gollmer et al., 2011). Enforcing such constraints for infinitely many scalarization vectors further complicates solving such large-scale models. For finite probability spaces, Homem-de-Mello and Mehrotra (2009) and Noyan and Rudolf (2013) show that infinitely many risk constraints (associated with polyhedral scalarization sets) reduce to finitely (typically exponentially) many scalar-based risk constraints for the SSD and CVaR cases, respectively. These results naturally lead to finitely convergent cut generation-based solution algorithms. However, such algorithms are computationally demanding as they typically require solving several non-convex cut generation subproblems. For both polyhedral SSD-constrained and polyhedral CVaR-constrained optimization models, the cut generation problem can be modeled as a difference of convex (DC) programming problem. Homem-de-Mello and Mehrotra (2009) exploit the polyhedral structure of this problem to present a branch-and-cut algorithm that incorporates concepts from concave minimization and binary mixed-integer programming (MIP). The authors also propose concavity and convexity inequalities, and a big-M improvement method within the branch-and-bound tree to strengthen this MIP. However, it appears that for the practical applications, the authors directly solve the MIP formulation of the cut generation problem (Hu et al., 2011; 2012). These studies also use a sample average approximation-based approach. In another line of work, Dentcheva and Wolfhagen (2013) use methods from DC programming to perform cut generation for the multivariate SSD-constrained problem. On the other hand, Noyan and Rudolf (2013) utilize alternative optimization representations of CVaR to develop computationally tractable MIP formulations for the cut generation problem for the polyhedral CVaR-constrained problem.

Despite the existing algorithmic developments, solving the MIP formulations of the cut generation problem can increasingly become a computational bottleneck as the number of scenarios increases. For example, Noyan and Rudolf (2013) found the homeland security allocation problem instances featuring up to 500 scenarios tractable, but solving the cut generation MIPs in reasonable times does not seem to be possible for a larger number of scenarios. Moreover, the cut generation is by far the most time consuming step of the algorithms to solve the multivariate risk-constrained optimization models of interest. According to the results presented in Hu et al. (2011) and Noyan and Rudolf (2013), cut generation generally takes no less than 90% to 95% of the total time spent by the cut generation-based algorithms. In line with these discussions, this paper contributes to the literature by providing more effective methods to solve the cut generation problems arising in optimization under polyhedral SSD and CVaR constraints. As in Homem-de-Mello and Mehrotra (2009) and Noyan and Rudolf (2013) we focus on finite probability spaces, and our approaches can naturally be used in a framework based on sample average approximation (SAA).

As observed in previous studies, the cut generation problems arising in multivariate CVaR or SSD-constrained optimization involve minimization of the difference of convex functions, or alternatively minimization of concave functions. However, these functions have polyhedral structure that can be exploited to devise enhanced and easy to implement models. In particular, the cut generation problems can be reformulated as mixed-integer linear programs (MIP) involving big-M terms for the linearization of the nonlinear shortfall terms, and can be solved with black-box MIP solvers. However, as noted in the earlier studies, such MIP formulations have weak linear relaxation bounds, and thus a branch-and-cut-based method to solve them becomes the bottleneck of an algorithm that requires the solution of cut generation problems iteratively. For polyhedral SSD-constrained problems, the cut generation problems naturally decompose by scenarios, and the main difficulty is due to the weakness of the linearized MIP. A similar difficulty arises in polyhedral CVaR-constrained problems. However, in this case, an additional challenge stems from the combinatorial structure required to identify the $\alpha$-quantile of the decision-based random vectors. Therefore, this study is mainly dedicated to develop computationally efficient methods for the multivariate CVaR-constrained models. However, we also describe how our results can be applied in the SSD case.

In this paper, we develop alternative mixed-integer programming formulations and algorithms for cut generation problems arising in optimization under polyhedral constraints based on CVaR and SSD. We give the complete linear description of two non-convex substructures arising in these cut generation problems. Our computational experiments show that the proposed models lead to more effective cut generation-based algorithms to solve the multivariate risk-constrained optimization models. We perform a computational study of a budget allocation problem, previously studied in Hu et al. (2011) and Noyan and Rudolf (2013), and also obtain additional results for another set of randomly generated problem instances, to evaluate the effectiveness of our solution methods.

In the next section, we first review fundamental concepts and results related to CVaR, SDD, and multivariate polyhedral risk preferences, and then present the general forms of the optimization models featuring these preference relations as constraints. In Section 3, we study the cut generation problems arising in CVaR-constrained models. We give a new MIP formulation, and several classes of valid inequalities that improve this formulation. In addition, we propose variable fixing methods that are highly effective in certain classes of problems. In Section 4, we give analogous results for SSD-constrained models. We present our computational experiments with the proposed models and methods in Section 5. Finally, we conclude the paper in Section 6.

## 2. Basic Definitions and Fundamental Results

Comparing uncertain outcomes is one of the fundamental interests of decision theory. We first review some basic definitions and results related to the univariate stochastic preference rules based on CVaR and SSD. Then, we recall the multivariate polyhedral CVaR and SSD relations and some relevant results to make the paper self-contained. To conclude this section, we present a general form of optimization models featuring multivariate CVaR and SSD constraints based on polyhedral scalarization.

Before proceeding, we need to make a note of some conventions used throughout the paper. *Larger values of random variables*, as well as *larger values of risk measures*, are considered to be preferable. In this context, risk measures are often referred to as acceptability functional, since higher values indicate less risky random outcomes.

The cumulative distribution function (CDF) of a random variable $X$ is denoted by $F_X$. The set of the first $n$ positive integers is denoted by $[n] = \{1, \ldots, n\}$, while the positive part of a number $x \in \mathbb{R}$ is denoted by $[x]_+ = \max(x, 0)$. All random variables in this paper are assumed to be defined on some finite probability spaces. Since the focus of this study is on finite probability spaces, we simplify our exposition accordingly when possible.

**2.1 Risk measures - CVaR and VaR** Risk measures are functionals that represent the risk associated with a random variable by a scalar value, and provide a direct way to compare random outcomes based on the decision makers' risk preferences. Desirable properties of risk measures, such as law invariance and coherence, are axiomatized in Artzner et al. (1999). CVaR, introduced by Rockafellar and Uryasev (2000), satisfies these axioms, and moreover, it serves as a fundamental building block for other law invariant coherent risk measures (Kusuoka, 2001). Due to its useful properties, CVaR has recently been widely used in decision making problems under uncertainty. We next present some relevant definitions and relations; for more detailed discussions, see Pflug and Römisch (2007), Rockafellar (2007), and Noyan and Rudolf (2013).

Definition 2.1 (Rockafellar and Uryasev (2000; 2002)) *For a random variable $X$, the* conditional value-at-risk *at confidence level $\alpha \in (0, 1]$ is given by*

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

For risk-averse decision makers typical choices for the confidence level are small values such as $\alpha = 0.05$.

Suppose $X$ is a random variable with (not necessarily distinct) realizations $x_1, \ldots, x_n$ and corresponding probabilities $p_1, \ldots, p_n$.

- The optimization problem in (1) can equivalently be formulated as the following linear program:
$$\max\{\eta - \frac{1}{\alpha}\sum_{i \in [n]} p_i w_i \ : \ w_i \geq \eta - x_i, \ \forall\, i \in [n], \quad \mathbf{w} \in \mathbb{R}^n_+\}. \tag{2}$$
- It is well known that the maximum in definition (2) is attained at the $\alpha$-quantile (also known as the *value-at-risk* (VaR) at confidence level $\alpha$) given by:
$$\mathrm{VaR}_\alpha(X) = \min\{\eta \ : \ F_X(\eta) \geq \alpha\}. \tag{3}$$
- Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ denote an ordering of the realizations. Then, for a given confidence level $\alpha \in (0, 1]$ we can show that
$$\mathrm{VaR}_\alpha(X) = x_{(k)}, \ \text{ where } k = \min\left\{ j \in [n] : \sum_{i \in [j]} p_i \geq \alpha \right\} \tag{4}$$

and

$$\text{CVaR}_\alpha(X) = \text{VaR}_\alpha(X) - \frac{1}{\alpha}\mathbb{E}\left([\text{VaR}_\alpha(X) - X]_+\right) \tag{5}$$

$$= \frac{1}{\alpha}\left[\sum_{i\in K} p_i x_{(i)} + \left(\alpha - \sum_{i\in K} p_i\right) x_{(k)}\right], \text{ where } K = [k-1]. \tag{6}$$

Due to its latter representation as the weighted sum of the realizations in (6), CVaR is also known in the literature as *average value-at-risk* and *tail value-at-risk*.

- For decision making problems, the ordering of the realizations of a decision-based random variable cannot be known in advance. In this context, the subset-based representation (6) can still be utilized to formulate CVaR as an optimization problem (as an alternative to (2)). That particular optimization representation of CVaR, along with the representation of VaR in (4), proves useful in developing new methods in our study. For alternative optimization representations of CVaR we refer the reader to Noyan and Rudolf (2013).

**2.2 Univariate second-order stochastic dominance** The preference relations based on risk measures provide comparisons with respect to a single scalar-valued functional. When the distribution of a random variable is of significant interest, a single measure may not be sufficient to model the decision makers' preferences. In this regard, stochastic dominance relations compare random variables with respect to a collection of measures, which correspond to pointwise values of some performance functions constructed from probability distribution functions. In particular, the SSD relation is based on the *second-order distribution function* $F_X^{(2)} : \mathbb{R} \to \mathbb{R}$ of the random variable $X$ defined by

$$F_X^{(2)}(\eta) = \int_{-\infty}^{\eta} F_X(a)\,\mathrm{d}a = \mathbb{E}\left([\eta - X]_+\right).$$

Definition 2.2 *We say that* $X$ dominates $Y$ in the second order, *denoted by* $X \succcurlyeq_{(2)} Y$, *if*

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

- It is well-known that the SSD relation is equivalent to the continuum of expected utility inequalities $\mathbb{E}(u(X)) \ge \mathbb{E}(u(Y))$ for all concave non-decreasing (i.e., risk-averse) utility functions $u$. We refer the reader to Müller and Stoyan (2002) for a modern review on stochastic dominance relations.
- Suppose $Y$ is a benchmark random variable with realizations $y_1, \dots, y_m$. Then, the SSD relation $X \succcurlyeq_{(2)} Y$ is equivalent to (see Dentcheva and Ruszczyński, 2003):

$$\mathbb{E}([y_l - X]_+) \le \mathbb{E}([y_l - Y]_+), \quad \text{for all } l \in [m]. \tag{7}$$

- It is well-known that the SSD relation (7) can be represented by finitely many linear inequalities. Suppose that $X$ is a random variable with realizations $x_1, \dots, x_n$ and corresponding probabilities $p_i,\ i \in [n]$. For a given benchmark random variable $Y$ with realizations $y_1 \le y_2 \le \dots \le y_m$ and corresponding probabilities $q_l,\ l \in [m]$, the inequalities in (7) hold if and only if there exist $\mathbf{v} \in \mathbb{R}_+^{nm}$ satisfying the following system:

$$\sum_{i\in[n]} p_i v_{il} \le \mathbb{E}([y_l - Y]_+) = \sum_{k\in[l-1]} (y_l - y_k) q_k, \quad \forall\, l \in [m], \tag{8}$$

$$v_{il} \ge y_l - x_i, \quad \forall\, i \in [n],\ l \in [m]. \tag{9}$$

In the decision making context, $x_i,\ i \in [n]$, would be decision variables representing the realizations of the decision-based random outcome of interest. Utilizing the above formulation (8)-(9), the SSD relation can be incorporated into stochastic optimization problems as constraints (see, e.g., Dentcheva and Ruszczyński, 2006; Noyan et al., 2008). For other types of SSD formulations we refer to Strassen (1965) and Luedtke (2008).

**2.3 Multivariate polyhedral risk preferences** Along the lines of Homem-de-Mello and Mehrotra (2009) and Noyan and Rudolf (2013), we consider the CVaR and SSD relations based on polyhedral linear scalarization. We now present the formal definitions of these multivariate stochastic preference rules.

DEFINITION 2.3 (MULTIVARIATE SSD RELATION, HOMEM-DE-MELLO AND MEHROTRA (2009)) *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two d-dimensional random vectors,* $C \subset \mathbb{R}^d$ *a set of scalarization vectors. We say that* $\mathbf{X}$ dominates $\mathbf{Y}$ in the (polyhedral linear) second order *with respect to* $C$*, denoted as* $\mathbf{X} \succcurlyeq^{C}_{(2)} \mathbf{Y}$*, if*

$$\mathbf{c}^\top \mathbf{X} \succeq_{(2)} \mathbf{c}^\top \mathbf{Y} \quad \textit{for all } \mathbf{c} \in C. \tag{10}$$

DEFINITION 2.4 (MULTIVARIATE CVAR RELATION, NOYAN AND RUDOLF (2013)) *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two d-dimensional random vectors,* $C \subset \mathbb{R}^d$ *a set of scalarization vectors, and* $\alpha \in (0,1]$ *a specified confidence level. We say that* $\mathbf{X}$ *is* CVaR-preferable *to* $\mathbf{Y}$ *at confidence level* $\alpha$ *with respect to* $C$*, denoted as* $\mathbf{X} \succcurlyeq^{C}_{\mathrm{CVaR}_\alpha} \mathbf{Y}$*, if*

$$\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}) \quad \textit{for all } \mathbf{c} \in C. \tag{11}$$

For any nontrivial polyhedron $C$ of scalarization vectors, the multivariate rules (10) and (11) are equivalent by definition to a collection of infinitely many univariate preference relations. Homem-de-Mello and Mehrotra (2009) and Noyan and Rudolf (2013) show that for finite probability spaces it is sufficient to consider a finite subset of scalarization vectors, obtained as projections of the vertices of higher dimensional polyhedra. Noyan and Rudolf (2013) also obtain such finite representations of scalarization polyhedra for a wider class of coherent risk measures. Homem-de-Mello and Mehrotra (2009) identify the sufficient set of finitely many scalarization vectors using $m$ lower-dimensional polyhedra, where $m$ is the number of realizations of the benchmark random vector $\mathbf{Y}$. Noyan and Rudolf (2013) instead use a single polyhedron, and their finiteness result also extends to the multivariate SSD case, providing an alternative to the representation in Homem-de-Mello and Mehrotra (2009). These finiteness results are essential to prove the finite convergence of the cut generation-based algorithms to solve the optimization problems featuring the multivariate polyhedral CVaR and SSD relations as constraints.

REMARK 2.1 *For both types of multivariate risk relations, we can assume* without loss of generality that the polyhedron $C$ is a polytope *(see Homem-de-Mello and Mehrotra (2009) and Noyan and Rudolf (2013) for the corresponding proofs: Proposition 1 and Proposition 2, respectively). In particular, any nonempty convex set* $C$ *can equivalently be replaced by* $\tilde{C} = \{\mathbf{c} \in \mathrm{cl\ cone}(C) \ : \ \|\mathbf{c}\|_1 \leq 1\}$*, where* $\mathrm{cl\ cone}(C)$ *denotes the closure of the conical hull of the set* $C$*.*

*Given the interpretation of the scalarization vectors and the fact that larger outcomes are considered to be preferable, we naturally assume that* $C \subseteq \{\mathbf{c} \in \mathbb{R}^d_+ \ : \ \sum_{i \in [d]} c_i = 1\}$*.*

**2.4 Optimization with multivariate risk constraints** We consider a multicriteria decision making problem where the multiple random performance measures associated with the decision variable $\mathbf{z}$ are represented by the random outcome vector $G(\mathbf{z})$. Let $(\Omega, 2^{\Omega}, \mathcal{P})$ be a finite probability space with $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $\mathcal{P}(\omega_i) = p_i$. The set of feasible decisions is denoted by $Z$ and the random outcomes are determined according to the mapping $G : Z \times \Omega \to \mathbb{R}^d$. Let $f : Z \to \mathbb{R}$ be a continuous objective function and $C \subset \mathbb{R}^d_+$ be a polytope of scalarization vectors. Given the benchmark (reference) random outcome vector $\mathbf{Y}$ and the confidence level $\alpha \in (0, 1]$, we obtain general forms of the optimization problems involving multivariate polyhedral CVaR and SSD constraints as follows:

$$(\mathbf{G-MCVaR}) \qquad \max \ f(\mathbf{z})$$
$$\text{s.t.} \ \mathrm{CVaR}_\alpha(\mathbf{c}^\top G(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \qquad \forall\, \mathbf{c} \in C, \tag{12}$$
$$\mathbf{z} \in Z.$$

$$(\mathbf{G-MSSD}) \qquad \max \ f(\mathbf{z})$$
$$\text{s.t.} \ \mathbf{c}^\top G(\mathbf{z}) \succeq_{(2)} \mathbf{c}^\top \mathbf{Y}, \qquad \forall\, \mathbf{c} \in C, \tag{13}$$
$$\mathbf{z} \in Z.$$

While $\mathbf{Y}$ is allowed to be defined on a probability space different from $\Omega$, it is often constructed from a benchmark decision $\bar{\mathbf{z}} \in Z$, i.e., $\mathbf{Y} = G(\bar{\mathbf{z}})$. For ease of exposition, we present the formulations with a single multivariate risk constraint. However, we can also consider multiple benchmarks, multiple confidence levels, and varying scalarization sets. For example, as pointed out in Noyan and Rudolf (2013), the constraints in (12) can take the form of

$$\mathrm{CVaR}_{\alpha_{tk}}(\mathbf{c}^\top G(\mathbf{z})) \geq \mathrm{CVaR}_{\alpha_{tk}}(\mathbf{c}^\top \mathbf{Y}_t), \quad \forall\, t \in [T],\ k \in [K_t],\ \mathbf{c} \in C_{tk}. \tag{14}$$

As will be discussed in the following sections, cut generation problem is defined given a single benchmark and a scalarization set (and a single confidence level in CVaR case); therefore, our proposed solution methods for cut generation can directly be applied to the more general problems (such as those involving (14)).

According to the results on finite representations of the scalarization polyhedra, it is sufficient to consider finitely many scalarization vectors. However, they correspond to the vertices of some polyhedra, and therefore, there are still potentially exponentially many scalarization-based risk constraints. A natural approach is to solve some relaxations of the above problems obtained by replacing the set $C$ with a finite subset (can be even empty). This subset is augmented by adding the scalarization vectors generated in an iterative fashion. In this spirit, at each iteration of such a cut generation algorithm, given a current decision vector we attempt to find a scalarization vector for which the corresponding risk constraint (of the form (12) or (13)) is violated. To do this, we solve the corresponding cut generation problem, which is the main focus of our study.

**3. Cut Generation for Optimization with Multivariate CVaR Constraints** In this section, we first briefly describe the cut generation problem arising in optimization problems of the form $(\mathbf{G-MCVaR})$. Then we proceed to present several existing mathematical programming formulations of (**CutGen_CVaR**), which constitute a basis for our new developments. The rest of the section is dedicated to the proposed, computationally more effective formulations and methods.

Consider an iteration of the cut generation-based algorithm (proposed in Noyan and Rudolf (2013)), and let $\mathbf{X} = G(\mathbf{z}^*)$ be the random outcome vector associated with the decision vector $\mathbf{z}^*$ obtained by solving the current relaxation of $(\mathbf{G-MCVaR})$. The aim is to either find a vector $\mathbf{c} \in C$ for which the corresponding univariate CVaR constraint (12) is violated or to show that such a vector does not exist. In this regard, we solve the cut generation problem at confidence level $\alpha \in (0, 1]$, which can be presented in the following general form:

$$(\mathbf{CutGen\_CVaR}) \qquad \min_{\mathbf{c}\in C} \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}).$$

Observe that (**CutGen_CVaR**) involves the minimization of the difference of two concave functions, because CVaR is a concave function given by (1). If the optimal objective value of (**CutGen_CVaR**) is non-negative, it follows that $\mathbf{z}^*$ is an optimal solution of $(\mathbf{G-MCVaR})$. Otherwise, there exists an optimal solution $\mathbf{c}^*$ for which the corresponding constraint $\mathrm{CVaR}_\alpha(\mathbf{c^*}^\top \mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c^*}^\top \mathbf{Y})$ is a valid inequality violated by the current solution.

Note that we can easily calculate the realizations of the random outcome $\mathbf{X} = G(\mathbf{z}^*)$ given the decision vector $\mathbf{z}^*$. In the rest of the paper, we focus on solving the cut generation problems given two $d$-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ with realizations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_m$, respectively. Let $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ denote the corresponding probabilities, and let $C \subseteq \{\mathbf{c} \in \mathbb{R}^d_+ \ : \ \sum_{i\in[d]} c_i = 1\}$ be a polytope of scalarization vectors.

**3.1 Existing mathematical programming formulations** In this section, we present two mathematical programming formulations of (**CutGen_CVaR**), which are proposed in Noyan and Rudolf (2013). It is easy to see that the second nonlinear term $(-\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}))$ in (**CutGen_CVaR**) can be expressed with linear inequalities and continuous variables as in (2). What makes it difficult to solve (**CutGen_CVaR**) is the first term $(\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}))$ in the minimization. Using two alternative optimization representations of CVaR, Noyan and Rudolf (2013) first formulate (**CutGen_CVaR**) as a (generally nonconvex) quadratic program. Then instead of dealing with the quadratic problem, the authors propose MIP formulations which are considered to be potentially more tractable.

According to (4) and (5), $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \mathbf{c}^\top \mathbf{x}_k$ for at least one $k \in [n]$, implying

$$\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \max_{k\in[n]} \left\{ \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha} \sum_{i\in[n]} p_i [\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+ \right\}.$$

This key observation leads to the following formulation of (**CutGen_CVaR**):

$$\min \quad \mu - \eta + \frac{1}{\alpha} \sum_{l\in[m]} q_l w_l \tag{15}$$

$$\text{s.t.} \quad \mu \geq \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha} \sum_{i\in[n]} p_i [\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+, \qquad \forall\, k \in [n], \tag{16}$$

$$w_l \geq \eta - \mathbf{c}^\top \mathbf{y}_l, \qquad \forall\, l \in [m], \tag{17}$$

$$\mathbf{c} \in C, \ \mathbf{w} \in \mathbb{R}^m_+. \tag{18}$$

The shortfall terms $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$ in inequalities (16) present a computational challenge. Introducing additional variables and constraints, Noyan and Rudolf (2013) linearize these terms and obtain the following equivalent MIP formulation:

$$(\mathbf{MIP\_CVaR}) \qquad \min \quad \mu - \eta + \frac{1}{\alpha} \sum_{l\in[m]} q_l w_l$$

$$\begin{aligned}
\text{s.t.}\quad & (17)-(18) \\
& \mu \geq \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha}\sum_{i\in[n]} p_i v_{ik}, && \forall\ k\in[n], && (19)\\
& v_{ik} - \delta_{ik} = \mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i, && \forall\ i\in[n],\ k\in[n], && (20)\\
& v_{ik} \leq M_{ik}\beta_{ik}, && \forall\ i\in[n],\ k\in[n], && (21)\\
& \delta_{ik} \leq \hat{M}_{ik}(1-\beta_{ik}), && \forall\ i\in[n],\ k\in[n], && (22)\\
& \beta_{ik}\in\{0,1\}, && \forall\ i\in[n],\ k\in[n], && (23)\\
& \mathbf{v}\in\mathbb{R}_+^{n\times n},\quad \boldsymbol{\delta}\in\mathbb{R}_+^{n\times n}. && && (24)
\end{aligned}$$

Here $M_{ik}$ and $\hat{M}_{ik}$ are sufficiently large constants (Big-M coefficients) to make constraints (21) and (22) redundant whenever the right-hand side is positive. By constraints (21)-(24) only one of the variables $v_{ik}$ and $\delta_{ik}$ is positive. Then, constraint (20) ensures that $v_{ik} = [\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$ for all pairs of $i$ and $k$. A similar linearization is used for the SSD case described in Section 4.

Remark 3.1 *[Big-M Coefficients] It is well-known that the choice of the Big-M coefficients is crucial in obtaining stronger MIP formulations. In* (**MIP_CVaR**) *, we can set*

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

*These parameters can easily be obtained by solving very simple LPs. Furthermore, in practical applications, the dimension of the decision vector* $\mathbf{c}$ *and the number of vertices of the polytope* $C$ *would be small; e.g., in our computational study* $d = 4$*. Suppose that the vertices of the polytope* $C$ *is known and given as* $\{\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_L\}$*. Then,*

$$M_{ik} = \max\{\max_{j\in[L]} \hat{\mathbf{c}}_j^\top(\mathbf{x}_k - \mathbf{x}_i), 0\}.$$

Remark 3.2 *In Section 3.2.3, we present polyhedral analysis of the linearization polytope* (20)-(24)*, which appears in* (**MIP_CVaR**) *to deal with the term* $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$*; this substructure also arises in the cut generation problems with multivariate SSD constraints.*

In the special case when all the outcomes of $\mathbf{X}$ are equally likely, Noyan and Rudolf (2013) propose an alternate MIP formulation which involves only $O(n)$ binary variables instead of $O(n^2)$. We refer to this formulation as (**MIP_Special**), and for completeness present it in the appendix. Following the rationale of Noyan and Rudolf (2013), in the next section we develop new formulations and methods based on integer programming approaches. We only focus on the general probability case; it turns out that even these general formulations perform better than (**MIP_Special**) as we show in Section 5.

**3.2 New developments** In this section, we first propose several simple improvements to the existing MIP formulations. Then, we introduce a MIP formulation based on a new representation of VaR. We propose valid inequalities that strengthen the resulting MIPs. We also give the complete linear description of the linearization polytope of a non-convex substructure appearing in the new formulation.

**3.2.1 Computational enhancements** We first present trivial valid inequalities based on the bounds for $\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{X})$, and then describe two approaches to reduce the number of variables and constraints of (**MIP_CVaR**).

- Suppose that we have a lower bound $L_\mu$ and an upper bound $U_\mu$ for $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Then, (**MIP_CVaR**) can be strengthened using the following valid inequalities:

$$L_\mu \leq \mu \leq U_\mu. \tag{25}$$

  It is easy to see that the random variable $\min_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{X}\}$ is *stochastically smaller* than $\mathbf{c}^\top \mathbf{X}$, i.e., $\mathbf{c}^\top \mathbf{X}$ dominates $\min_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{X}\}$ in the first order, for any $\mathbf{c} \in C$. Similarly, $\max_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{X}\}$ is *stochastically larger* than $\mathbf{c}^\top \mathbf{X}$, i.e., $\max_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{X}\}$ dominates $\mathbf{c}^\top \mathbf{X}$ in the first order, for any $\mathbf{c} \in C$. Therefore, we can set $L_\mu$ and $U_\mu$ as $\text{CVaR}_\alpha(\min_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{X}\})$ and $\text{CVaR}_\alpha(\max_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{X}\})$, respectively. A similar bounding scheme is proposed by Bülbül et al. (2014) in a different context.

- It is easy to observe the following symmetric relation between the $\boldsymbol{\delta}$ and $\mathbf{v}$ decisions: $\delta_{ik} = v_{ki}$ for all pairs of $i \in [n]$ and $k \in [n]$. According to this observation, we can drop the $\delta_{ik}$ variables and replace (19)-(24) by the following set of inequalities:

$$\mu \geq \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha}\sum_{i\in[n]\setminus\{k\}} p_i v_{ik}, \qquad \forall\, k \in [n], \tag{26}$$

$$v_{ik} - v_{ki} = \mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i, \qquad \forall i,k \in [n] : i < k, \tag{27}$$

$$v_{ik} \leq M_{ik}\beta_{ik}, \qquad \forall i,k \in [n] : i < k, \tag{28}$$

$$v_{ki} \leq M_{ki}(1-\beta_{ik}), \qquad \forall i,k \in [n] : i < k, \tag{29}$$

$$\beta_{ik} \in \{0,1\}, \qquad \forall i,k \in [n] : i < k, \tag{30}$$

$$v_{ik} \geq 0 \qquad \forall i,k \in [n] : i \neq k. \tag{31}$$

  We refer to the resulting compact MIP as (**SMIP_CVaR**); the number of binary variables and constraints (20)-(22) associated with the shortfall terms is reduced by half. Furthermore, the linearization polytope defined by (27)-(31) can be strengthened using valid inequalities. In Section 4, we study the linearization polytope corresponding to $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$ for a given pair $i,k \in [n]$.

- Preprocessing methods can be used to identify the scenarios for which $\mathbf{c}^\top \mathbf{x}_k$ cannot be equal to $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Such methods would allow us to enforce constraint (26) for a smaller set of $k$ indices. Let us denote the remaining potentially relevant indices by $\bar{K} \subseteq [n]$, then (26) becomes

$$\mu \geq \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha}\sum_{i\in[n]\setminus\{k\}} p_i v_{ik}, \quad \forall\, k \in \bar{K}.$$

  This would also result in reduced number of variables and constraints (27)-(31) that are used to represent the shortfall terms. In particular, we need to define the variables $v_{ik}$ only for all $k \in \bar{K}, i \in [n]$ and for $i \in \bar{K}, k \in [n] \setminus \bar{K}$. In addition, we define variables $\beta_{ik}$ and constraints (27)-(29) for $i,k \in \bar{K}, i < k$ and for $k \in \bar{K}, i \in [n] \setminus \bar{K}$ (note that due to the elimination of some of the $v$ variables, the symmetry argument does not hold for the latter condition, so we do not have the restriction that $i < k$ unless $i,k \in \bar{K}$). We refer to the resulting more compact MIP, which also involves (25), as (**RSMIP_CVaR**). We next elaborate on how to identify such a reduced set of scenarios $\bar{K}$.

Recall that we focus on the left tail of the probability distributions; for example, under equal probabilities, $\text{VaR}_{b/n}(\mathbf{c}^\top \mathbf{X})$ is the $b$th smallest realization of $\mathbf{c}^\top \mathbf{X}$ where $b$ is a small integer. Thus, $\mathbf{c}^\top \mathbf{x}_k$ values which definitely take relatively larger values cannot correspond to $\text{VaR}_{b/n}(\mathbf{c}^\top \mathbf{X})$. In line with these discussions, we obtain the result presented in the next proposition and use it to identify the set $\bar{K}$.

Proposition 3.1 *Suppose that the parameters $M_{ki}$ are calculated as described in Remark 3.1. For a scenario index $k \in [n]$, let $L_k = \{i \in [n] \setminus k \ : \ M_{ki} = 0\}$ and $H_k = \{i \in [n] \setminus k \ : \ M_{ik} = 0\}$. If $\sum_{i \in L_k} p_i \geq \alpha$ then $\mathbf{c}^\top \mathbf{x}_k = \mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ cannot hold for any $\mathbf{c} \in C$, implying $k \notin \bar{K}$. Moreover, $i \notin \bar{K}$ for all $i \in H_k$.*

Proof. Since $\mathbf{c}^\top \mathbf{x}_i \leq \mathbf{c}^\top \mathbf{x}_k$ for all $i \in L_k$ and for any $\mathbf{c} \in C$, the first claim immediately follows from the VaR definition (4). Similarly, the second claim holds because $L_k \subseteq L_i$ for all $i \in H_k$. □

Note that if for some $k \in [n]$, we have non-empty sets $L_k$ or $H_k$, we can employ ***variable fixing*** by letting $\beta_{ik} = 1$, $\beta_{ki} = 0$ for $i \in L_k$ and $\beta_{ik} = 0$, $\beta_{ki} = 1$ for $i \in H_k$.

Another method can utilize the bounds on $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ while identifying the set $\bar{K}$. Suppose that we have a lower bound $L$ and an upper bound $U$ for $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. If $\max_{\mathbf{c} \in C} \mathbf{c}^\top \mathbf{x}_k < L$ or $\min_{\mathbf{c} \in C} \mathbf{c}^\top \mathbf{x}_k > U$, then $k \notin \bar{K}$. Similar to the case of CVaR, we can calculate the bounds $L$ and $U$ using the random variables $\min_{\mathbf{c} \in C}\{\mathbf{c}^\top \mathbf{X}\}$ and $\max_{\mathbf{c} \in C}\{\mathbf{c}^\top \mathbf{X}\}$: $L = \mathrm{VaR}_\alpha(\min_{\mathbf{c} \in C}\{\mathbf{c}^\top \mathbf{X}\})$ and $U = \mathrm{VaR}_\alpha(\max_{\mathbf{c} \in C}\{\mathbf{c}^\top \mathbf{X}\})$.

In our numerical study, we have observed that the above methods can significantly impact the computational performance (see Section 5).

**3.2.2 Alternative approaches based on a new representation of VaR** When the realizations are based on a decision, we cannot know their ordering in advance. While the structure of the objective function makes it easy to express VaR in the context of VaR or CVaR maximization, in our cut generation problem we need a new representation of VaR. Recall that we can use the classical definition of CVaR in the second CVaR term appearing in the objective function of (**CutGen_CVaR**), but for the first CVaR term we need alternative representations of CVaR to develop new, computational more tractable solution methods. The main challenge is to express $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, which depends on $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. The next theorem provides a set of inequalities to calculate $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ when $\mathbf{c}$ is a decision vector.

Theorem 3.1 *Suppose that $X$ is a random variable with realizations $x_1, \ldots, x_n$ and corresponding probabilities $p_i$, $i \in [n]$. Let $M_{i\cdot} = \max_{k \in [n]} M_{ik}$, $M_{\cdot i} = \max_{k \in [n]} M_{ki}$ for $i \in [n]$ and $\tilde{M}_\ell = \max\{c_\ell : \ \mathbf{c} \in C\}$ for $\ell \in [d]$. Then, for a given confidence level $\alpha$ and any decision vector $\mathbf{c} \in C$, the equality $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ holds if and only if there exists a vector $(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \mathbf{u})$ satisfying the following system:*

$$z \leq \mathbf{c}^\top \mathbf{x}_i + \beta_i M_{i\cdot}, \qquad i \in [n], \tag{32}$$

$$z \geq \mathbf{c}^\top \mathbf{x}_i - (1-\beta_i) M_{\cdot i}, \qquad i \in [n], \tag{33}$$

$$\sum_{i \in [n]} p_i \beta_i \geq \alpha, \tag{34}$$

$$\sum_{i \in [n]} p_i \beta_i - \sum_{i \in [n]} p_i u_i \leq \alpha - \epsilon, \tag{35}$$

$$z = \sum_{i \in [n]} \boldsymbol{\zeta}_i^\top \mathbf{x}_i, \tag{36}$$

$$\zeta_{i\ell} \leq \tilde{M}_\ell u_i, \qquad i \in [n], \ \ell \in [d], \tag{37}$$

$$\sum_{i \in [n]} \zeta_{i\ell} = c_\ell \qquad \ell \in [d], \tag{38}$$

$$\sum_{i \in [n]} u_i = 1, \tag{39}$$

$$u_i \leq \beta_i, \qquad i \in [n], \tag{40}$$

$$\boldsymbol{\beta} \in \{0,1\}^n, \quad \boldsymbol{\zeta} \in \mathbb{R}_+^{n \times d}, \quad \mathbf{u} \in \{0,1\}^n. \tag{41}$$

*Here $\epsilon$ is a very small constant to ensure that the left-hand side is strictly smaller than $\alpha$.*

Proof. Suppose that $z = \text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ for a decision vector $\mathbf{c} \in C$. Let $\pi$ be a permutation describing a non-decreasing ordering of the realizations of the random vector $\mathbf{c}^\top \mathbf{X}$, i.e., $\mathbf{c}^\top \mathbf{x}_{\pi(1)} \leq \cdots \leq \mathbf{c}^\top \mathbf{x}_{\pi(n)}$. Defining

$$k^* = \min\left\{k \in [n] \ : \ \sum_{i \in [k]} p_{\pi(i)} \geq \alpha\right\} \quad \text{and} \quad K^* = \{\pi(1), \ldots, \pi(k^*)\}, \tag{42}$$

and using (4) we have $z = \mathbf{c}^\top \mathbf{x}_{\pi(k^*)}$. Then, it is easy to see that a feasible solution of (32)-(41) can be obtained as follows:

$$\beta_i = \begin{cases} 1 & i \in K^* \\ 0 & \text{otherwise} \end{cases}, \quad u_i = \begin{cases} 1 & i = k^* \\ 0 & \text{otherwise} \end{cases}, \quad \zeta_{i\ell} = \begin{cases} c_\ell & i = k^*, \ell \in [d] \\ 0 & \text{otherwise.} \end{cases}$$

For the reverse implication, let us consider a feasible solution $(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \mathbf{u})$ of (32)-(41) and let $\bar{K} = \{i \in [n] \ : \ \beta_i = 1\}$. To prove our claim, it is sufficient to show that there exists a permutation $\pi$ where $\bar{K} = K^*$ and $z = \mathbf{c}^\top \mathbf{x}_{\pi(k^*)} = \mathbf{c}^\top \mathbf{x}_{\bar{k}}$ for a scenario index $\bar{k} \in \arg\max_{i \in \bar{K}} \{\mathbf{c}^\top \mathbf{x}_i\}$ ($K^*$ and $k^*$ are defined as in (42)).

We first focus on the intermediate set of linear inequalities and quadratic equalities (32)-(35), (39)-(40) and

$$z = \sum_{i \in [n]} u_i \mathbf{c}^\top \mathbf{x}_i, \tag{43}$$

$$\boldsymbol{\beta} \in \{0,1\}^n, \quad \mathbf{u} \in \{0,1\}^n. \tag{44}$$

By the definition of $\bar{K}$ and inequalities (32)-(33) we have $z \leq \mathbf{c}^\top \mathbf{x}_i$, $i \in [n] \setminus \bar{K}$, and $z \geq \mathbf{c}^\top \mathbf{x}_i$, $i \in \bar{K}$. Since $\beta_i = 0$ for all $i \in [n] \setminus \bar{K}$, (40) ensures that $u_i = 0$ for all $i \in [n] \setminus \bar{K}$. Then, (39) and (40) guarantee that $z = \sum\limits_{i \in \bar{K}} u_i \mathbf{c}^\top \mathbf{x}_i = \mathbf{c}^\top \mathbf{x}_{\bar{k}}$ for a scenario index $\bar{k}$ such that $\mathbf{c}^\top \mathbf{x}_{\bar{k}} = \max_{i \in \bar{K}} \{\mathbf{c}^\top \mathbf{x}_i\}$. Thus, $u_i = 1$ for $i = \bar{k}$, and 0, otherwise. Then, from (34) and (35), $\mathcal{P}(\mathbf{c}^\top \mathbf{X} \leq z) = \sum\limits_{i \in \bar{K}} p_i \beta_i \geq \alpha$ and $\sum\limits_{i \in \bar{K} \setminus \bar{k}} p_i \beta_i < \alpha$. It follows that, according to the definition in (3), $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \mathbf{c}^\top \mathbf{x}_{\bar{k}} = z$.

Since $\mathbf{c}$ is a decision vector, equality (43) involves quadratic terms of the form $u_i c_\ell$. First observe that $u_i c_\ell = c_\ell$, $\ell \in [d]$, for exactly one scenario index $i$, implying $\sum_{i \in [n]} u_i c_\ell = c_\ell$, $\ell \in [d]$, at any feasible solution satisfying (32)-(35), (39)-(40) and (43)-(44). Therefore, it is easy to show that we can linearize the $u_i c_\ell$ terms by replacing them with the new decision variables $\boldsymbol{\zeta}_{i\ell} \in \mathbb{R}_+$ in (43) to obtain (36), and enforcing the additional constraints (37)-(38). This completes our proof. □

Corollary 3.1 *The cut generation problem* (**CutGen_CVaR**) *is equivalent to the following optimization problem, referred to as* (**NewMIP_CVaR**)*:*

$$\min \quad z - \frac{1}{\alpha} \sum_{i \in [n]} p_i v_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \tag{45}$$

$$\text{s.t.} \quad (17) - (18), (32) - (41),$$

$$v_i - \delta_i = z - \mathbf{c}^\top \mathbf{x}_i, \qquad i \in [n], \tag{46}$$

$$v_i \leq M_{i \cdot} \beta_i, \qquad i \in [n], \tag{47}$$

$$\delta_i \le M_{\cdot i}(1-\beta_i), \qquad i \in [n], \tag{48}$$

$$\mathbf{v} \in \mathbb{R}^n_+, \quad \boldsymbol{\delta} \in \mathbb{R}^n_+, \tag{49}$$

$$L \le z \le U. \tag{50}$$

Proof. We represent $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$ in (**CutGen_CVaR**) using the classical formulation (1). On the other hand, we express $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ using the formula (5) and the representation of $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ provided in Theorem 3.1. In particular, for any decision vector $c \in C$, we ensure that $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ by (32)-(41), and use the following formulation:

$$\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = z - \frac{1}{\alpha}\sum_{i\in[n]} p_i[z - \mathbf{c}^\top \mathbf{x}_i]_+. \tag{51}$$

According to this approach, for a decision vector $c \in C$, we can represent $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$ by

$$\left\{ z - \frac{1}{\alpha}\sum_{i\in[n]} p_i[z - \mathbf{c}^\top \mathbf{x}_i]_+ \;:\; (32)-(41) \right\} - \max\left\{ \eta - \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l \;:\; (17),\ \mathbf{w} \in \mathbb{R}^m_+ \right\}.$$

Then, by simple manipulation and linearizing the terms $[z - \mathbf{c}^\top \mathbf{x}_i]_+ =: v_i$ using (46)-(49), we obtain the desired formulation. □

Note that there are $O(n)$ binary variables in (**NewMIP_CVaR**) compared to $O(n^2)$ binary variables in (**RSMIP_CVaR**). We next describe valid inequalities, which we refer to as *ordering inequalities*, to strengthen the formulation (**NewMIP_CVaR**).

Proposition 3.2 *Suppose that the parameters $M_{ki}$ are calculated as described in Remark 3.1. For a scenario index $k \in [n]$, let $L_k = \{i \in [n] \setminus k \;:\; M_{ki} = 0\}$ and $H_k = \{i \in [n] \setminus k \;:\; M_{ik} = 0\}$. Then the following sets of inequalities are valid for* (**NewMIP_CVaR**)*:*

$$\beta_k \le \beta_i, \quad k \in [n],\ i \in L_k, \tag{52}$$

*or equivalently,*

$$\beta_i \le \beta_k, \quad k \in [n],\ i \in H_k. \tag{53}$$

Proof. If $i \in L_k$, then $\max_{\mathbf{c}\in C}[\mathbf{c}^\top(\mathbf{x}_i - \mathbf{x}_k)]_+ = 0$. In other words, $\mathbf{c}^\top \mathbf{x}_k \ge \mathbf{c}^\top \mathbf{x}_i$ for all $\mathbf{c} \in C$. Now if $z > \mathbf{c}^\top \mathbf{x}_k$ for some $\mathbf{c} \in C$, then $\beta_k = 1$. Because $\mathbf{c}^\top \mathbf{x}_k \ge \mathbf{c}^\top \mathbf{x}_i$, we also have $\beta_i = 1$. On the other hand, if $z < \mathbf{c}^\top \mathbf{x}_i$ for some $\mathbf{c} \in C$, then $\beta_i = 0$. Because $z < \mathbf{c}^\top \mathbf{x}_i \le \mathbf{c}^\top \mathbf{x}_k$, we also have $\beta_k = 0$. Thus, inequality (52) is valid. The validity proof of inequality (53) follows similarly. □

Introducing inequalities (52) or (53) to (**NewMIP_CVaR**) provides us with a stronger formulation. When the number of such inequalities is considered to be large, we may opt to introduce them only for a selected set of scenarios. For example, we fix the values of a subset of $\beta_i$ variables using preprocessing methods when possible, and introduce the ordering inequalities for those that cannot be fixed. The trivial variable fixing sets $\beta_i = 0$ or $\beta_i = 1$ for all $i \in [n]$ such that $M_{i\cdot} = 0$ or $M_{\cdot i} = 0$, respectively. In addition, we propose a more elaborate *variable fixing*, which relies on Proposition 3.1 to identify the scenarios for which the corresponding realizations are too large to be equal to $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Suppose we show that $k$ is among such scenarios, i.e., $k \notin \bar{K}$. Then, at any feasible solution we have $\beta_k = 0$, and consequently, $\beta_i = 0$ for all $i \in H_k$. One can also employ variable fixing by using the bounds on $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. In particular, let $\beta_i = 1$ if $\max_{\mathbf{c}\in C} \mathbf{x}_i < L$ and let $\beta_i = 0$ if $\min\limits_{\mathbf{c}\in C} \mathbf{c}^\top \mathbf{x}_i > U$. We note that the proposed ordering inequalities and variable fixing methods can also be applied to other relevant MIP formulations involving $\beta_i$ decisions. In such MIPs, e.g., (**MIP_Special**), the set $\{k \in [n] \ : \ \beta_k = 1\}$ corresponds to the realizations which are less than or equal to $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$.

**3.2.3 Linearization of $(z - \mathbf{x}^\top \mathbf{c})_+$ in (CutGen_CVaR)** Consider the convex function $g(z, \mathbf{c}) = [z - \mathbf{x}_i^\top \mathbf{c}]_+ := \max\{0, z - \mathbf{x}_i^\top \mathbf{c}\}$ for $(z, \mathbf{c}) \in \mathbb{R}_+^{d+1}$ and $i \in [n]$ such that $\sum_{j\in[d]} c_j = 1$, which appears in (51). Using formula (51) in (**CutGen_CVaR**) leads to a concave minimization. Therefore, we study the linearization given by (46)-(49).

Throughout this subsection, we drop the scenario indices and focus on the linearization of one term of the form $[z - \mathbf{x}^\top \mathbf{c}]_+$. Due to the translation invariance of CVaR, we assume without loss of generality that all the realizations of $\mathbf{X}$ are non-negative. Therefore, $x_j \geq 0, j \in [d]$. This implies the nonnegativity of $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ and $z$, since $z$ represents $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ and $\mathbf{c} \geq \mathbf{0}$. In addition, to avoid trivial cases, we assume that $x_j > 0$ for some $j \in [d]$, because otherwise, we can let $z = v$ and $\delta = 0$. We are interested in the polytope defined by

$$v - \delta = z - \sum_{j\in[d]} x_j c_j, \tag{54}$$

$$v \leq M_v \beta, \tag{55}$$

$$\delta \leq M_\delta (1 - \beta), \tag{56}$$

$$\sum_{j\in[d]} c_j = 1, \tag{57}$$

$$\mathbf{c}, v, \delta \geq 0, \tag{58}$$

$$\beta \in \{0, 1\}, \tag{59}$$

$$0 \leq z \leq M_z. \tag{60}$$

At this time, we let $M_z = \max_{s\in[n],k\in[d]}\{x_{sk}\}$, i.e, the largest component of $\mathbf{x}_s$ over all $s \in [n]$, $M_v = M_z - \min_{k\in[d]}\{x_k\}$ and $M_\delta = \max_{k\in[d]}\{x_k\}$. Let $\mathcal{Q} = \{(\mathbf{c}, v, \delta, \beta, z) : (54) - (60)\}$.

First, we characterize the extreme points of $\mathrm{conv}(\mathcal{Q})$. Throughout, we let $e_k$ denote the $d$-dimensional unit vector with 1 in the $k$th entry and zeroes elsewhere.

Proposition 3.3 *The extreme points $(c, v, \delta, \beta, z)$ of conv$(\mathcal{Q})$ are as follows:*

**QEP1$_k$:** $(e_k, 0, x_k, 0, 0)$ *for all* $k \in [d]$ *with* $x_k > 0$,

**QEP2$_k$:** $(e_k, 0, 0, 0, x_k)$ *for all* $k \in [d]$,

**QEP3$_k$:** $(e_k, 0, 0, 1, x_k)$ *for all* $k \in [d]$,

**QEP4$_k$:** $(e_k, M_z - x_k, 0, 1, M_z)$ *for all* $k \in [d]$ *with* $x_k < M_z$.

Proof. First, note that, from the definitions of $M_z$, $M_v$, and $M_\delta$, we have $x_k \leq M_\delta \leq M_z$, and $0 \leq M_z - x_k \leq M_v$ for all $k \in [d]$. Hence, points **QEP1$_k$**–**QEP4$_k$** are feasible. It is also easy to see that they are extreme points of conv$(\mathcal{Q})$. Finally, observe that any other feasible point with $0 < c_j < 1$ for some $j \in [d]$ cannot be an extreme point, because it can be written as a convex combination of **QEP1$_k$**–**QEP4$_k$**. □

Note that if $x_k = 0$ for some $k \in [d]$, then **QEP1$_k$** is equivalent to **QEP2$_k$**. Therefore, we only define **QEP1$_k$** for $k \in [d]$ with $x_k > 0$. Similarly, if $x_k = M_z$ for some $k \in [d]$, then **QEP4$_k$** is equivalent to **QEP3$_k$**. Therefore, we only define **QEP4$_k$** for $k \in [d]$ with $x_k < M_z$.

Next we give valid inequalities for $\mathcal{Q}$.

Proposition 3.4 *For $k = 1, \ldots, d$, the inequality*

$$v \leq \sum_{j\in[d]} [x_k - x_j]_+ c_j + (M_z - x_k)\beta \tag{61}$$

*is valid for $\mathcal{Q}$. Similarly, for $k = 1, \ldots, d$, the inequality*

$$\delta \leq \sum_{j\in[d]} [x_j - x_k]_+ c_j + x_k(1-\beta) \tag{62}$$

*is valid for $\mathcal{Q}$.*

Proof. First, we prove the validity of inequality (61). If $\beta = 0$, then $v = 0$ from (55). Because $\mathbf{c} \geq \mathbf{0}$, inequality (61) holds trivially. If $\beta = 1$, then $\delta = 0$ from (56). Thus, for any $k = 1, \ldots, d$,

$$\begin{aligned} v - \delta = v = z - \sum_{j\in[d]} x_j c_j + x_k(\sum_{j\in[d]} c_j - 1) = z + \sum_{j\in[d]} (x_k - x_j)c_j - x_k \\ \leq \sum_{j\in[d]} [x_k - x_j]_+ c_j + M_z - x_k = \sum_{j\in[d]} [x_k - x_j]_+ c_j + (M_z - x_k)\beta, \end{aligned}$$

where the last inequality follows from (60). Thus, inequality (61) is valid.

Next, we prove the validity of inequality (62). If $\beta = 1$, then $\delta = 0$ from (56). Because $\mathbf{c} \geq \mathbf{0}$, inequality (62) holds trivially. If $\beta = 0$, then $v = 0$ from (56). Thus, for any $k = 1, \ldots, d$,

$$\begin{aligned} \delta = \sum_{j\in[d]} x_j c_j - z \leq \sum_{j\in[d]} (x_j - x_k)c_j + x_k \\ \leq \sum_{j\in[d]} [x_j - x_k]_+ c_j + x_k(1-\beta). \end{aligned}$$

Hence, inequality (62) is valid.

□

Theorem 3.2 *conv($\mathcal{Q}$) is completely described by equalities (54) and (57), and inequalities (58), (61), and (62).*

Proof. Let $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$, denote the index set of extreme point optimal solutions to the problem $\min\{\boldsymbol{\gamma}^\top \mathbf{c} + \gamma^v v + \gamma^\delta \delta + \gamma^\beta \beta + \gamma^z z : (\mathbf{c}, v, \delta, \beta, z) \in \text{conv}(\mathcal{Q})\}$, where $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z) \in \mathbb{R}^{d+4}$ is an arbitrary objective vector, not perpendicular to the smallest affine subspace containing conv($\mathcal{Q}$). In other words, $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z) \neq \lambda(-\mathbf{x}, -1, 1, 0, 1)$ and $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z) \neq \lambda(\mathbf{1}, 0, 0, 0, 0)$ for $\lambda \in \mathbb{R}$. Therefore, the set of optimal solutions is not conv($\mathcal{Q}$) (conv($\mathcal{Q}$) $\neq \emptyset$). We prove the theorem by giving an inequality among (58), (61), and (62) that is satisfied at equality by $(\mathbf{c}^\kappa, v^\kappa, \delta^\kappa, \beta^\kappa, z^\kappa)$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$ for the given objective vector. Then, since $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$ is arbitrary, for every facet of conv($\mathcal{Q}$), there is an inequality among (58), (61), and (62) that defines it. Throughout the proof, without loss of generality, we assume that $x_1 \leq x_2 \leq \cdots \leq x_d$. We consider all possible cases.

**Case A.** Suppose that $\gamma^\beta \geq 0$. Without loss of generality we can assume that $\gamma^\delta = 0$ by adding $\gamma^\delta(v - \delta - z + \sum_{j\in[d]} x_j c_j)$ to the objective. From equation (54) the added term is equal to zero, and so this operation does not change the set of optimal solutions. Furthermore, we can also assume that

$\gamma_j \geq 0$ for all $j \in [d]$ without loss of generality by subtracting $\gamma_{k^*}(\sum_{j\in[d]} c_j)$ from the objective, where $k^* := \arg\min\{\gamma_j, j \in [d]\}$. From equation (57), the subtracted term is a constant $(\gamma_{k^*})$, and so this operation does not change the set of optimal solutions. Therefore, for the case that $\gamma^\beta \geq 0$, we assume that $\gamma^\delta = 0$, $\gamma_j \geq 0$ for all $j \in [d]$, and $\gamma_{k^*} = 0$. Under these assumptions, we can express the cost of each extreme point solution (denoted by $C(\cdot)$) given in Proposition 3.3:

$C(\mathbf{QEP1}_k) = \gamma_k$ for $k \in [d]$ with $x_k > 0$,

$C(\mathbf{QEP2}_k) = \gamma_k + \gamma^z x_k$ for $k \in [d]$,

$C(\mathbf{QEP3}_k) = \gamma_k + \gamma^z x_k + \gamma^\beta$ for $k \in [d]$,

$C(\mathbf{QEP4}_k) = \gamma_k + \gamma^z M_z + \gamma^\beta + \gamma^v(M_z - x_k)$ for $k \in [d]$ with $x_k < M_z$.

Note that $\mathbf{QEP1}_k$ for $k \in [d]$ with $x_k > 0$ are the only extreme points with $\delta > 0$, and $\mathbf{QEP4}_k$ for $k \in [d]$ with $x_k < M_z$ are the only extreme points with $v > 0$. We use this observation in the following cases we consider.

(i) $\gamma^z < 0$. In this case, $C(\mathbf{QEP2}_k) < C(\mathbf{QEP1}_k)$ for all $k \in [d]$ with $x_k > 0$. Therefore, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. So we can assume that $\gamma^z \geq 0$.

(ii) $\gamma^z \geq 0$. In this case, $C(\mathbf{QEP1}_k) \leq C(\mathbf{QEP2}_k) \leq C(\mathbf{QEP3}_k)$ for all $k \in [d]$. Note that $C(\mathbf{QEP4}_k) = C(\mathbf{QEP3}_k) + (\gamma^z + \gamma^v)(M_z - x_k), k \in [d]$. Therefore, if $\gamma^z + \gamma^v > 0$, then $C(\mathbf{QEP4}_k) > C(\mathbf{QEP3}_k)$ for all $k \in [d]$, and hence extreme points $\mathbf{QEP4}_k, k \in [d]$ are never optimal. As a result, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. So we can assume that $\gamma^z + \gamma^v \leq 0$. Because $\gamma^z \geq 0$, we must have $\gamma^v \leq 0$. Let $\phi_k := \gamma^z M_z + \gamma^\beta + \gamma^v(M_z - x_k)$ for $k \in [d]$. Therefore, $C(\text{QEP4}_k) = \gamma_k + \phi_k$. Note that $\phi_1 \leq \phi_2 \leq \cdots \leq \phi_d$ because $x_1 \leq x_2 \leq \cdots \leq x_d \leq M_z$ and $\gamma^v \leq 0$ by assumption. If $\phi_1 > 0$, then $\phi_k > 0$ and so $C(\mathbf{QEP4}_k) > C(\mathbf{QEP1}_k)$ for all $k \in [d]$. Therefore, extreme points $\mathbf{QEP4}_k, k \in [d]$ are never optimal. Hence, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. Similarly, if $\phi_d < 0$, then $\phi_k < 0$ for all $k \in [d]$. Therefore, extreme points $\mathbf{QEP1}_k, k \in [d]$ are never optimal. Hence, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. As a result, we can assume that $\phi_1 \leq 0$ and $\phi_d \geq 0$. If there exists $j \in [d]$ such that $\gamma_j > 0$ and $\gamma_j + \phi_j > 0$, then $C(\mathbf{QEP1}_{k^*}) = 0 < C(\mathbf{QEP1}_j) \leq C(\mathbf{QEP2}_j) \leq C(\mathbf{QEP3}_j) < C(\mathbf{QEP4}_j)$. Hence, $c_j^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. As a result, we can assume that either $\gamma_k = 0$ or $\gamma_k + \phi_k \leq 0$ for all $k \in [d]$. If there exists $j \in [d]$ such that $\gamma_j > 0$ and $\gamma_j + \phi_j < 0 = C(\mathbf{QEP1}_{k^*})$, then extreme points $\mathbf{QEP1}_k, k \in [d]$ are never optimal. Hence, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. As a result, we can assume that for every $k \in [d]$, either $\gamma_k = 0$ or $\gamma_k + \phi_k = 0$.

(a) If $\gamma^\beta > 0$, then the optimal extreme point solutions are $\mathbf{QEP1}_j$ for all $j \in [d]$ such that $\gamma_j = 0$; $\mathbf{QEP2}_j$ for all $j \in [d]$ such that $\gamma_j = 0$ if $\gamma^z = 0$; and $\mathbf{QEP4}_k$ for all $k \in [d]$ such that $\gamma_k + \phi_k = 0$. Let $k' := \max\{j \in [d] : \phi_j \leq 0\}$. Note that $\phi_j > 0$ for $j > k'$ by definition, which implies that $\gamma_j + \phi_j > 0$. Therefore, we must have $\gamma_j = 0$ for $j > k'$. Then inequality (61) for $k'$ holds at equality for all optimal solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$.

(b) If $\gamma^\beta = 0$ and $\gamma^z > 0$, then the optimal extreme point solutions are $\mathbf{QEP1}_j$ for all $j \in [d]$ such that $\gamma_j = 0$ and $\mathbf{QEP4}_k$ for all $k \in [d]$ such that $\gamma_k + \phi_k = 0$. Then inequality (61) for $k'$ holds at equality for all optimal solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$.

(c) The only case left to consider is if $\gamma^\beta = \gamma^z = 0$. In this case, because we assume that $\gamma^v \leq 0$, there are two cases to consider. If $\gamma^v = 0$, then $\phi_k = 0$ for all $k \in [d]$ and we must have

$\gamma_k = 0$ for all $k \in [d]$, which contradicts our initial assumption that $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z) \neq \lambda(\mathbf{1}, 0, 0, 0, 0)$ for any $\lambda \in \mathbb{R}$. Therefore, we must have $\gamma^v < 0$. In this case, $\phi_k < 0$ for all $k \in [d]$. Suppose there exists $k^* \in [d]$ (with $\gamma_{k^*} = 0$) such that $x_{k^*} < M_z$. Then, $C(\mathbf{QEP4}_{k^*}) < 0 = C(\mathbf{QEP1}_{k^*})$. Because $C(\mathbf{QEP1}_{k^*}) \leq C(\mathbf{QEP1}_j)$ for all $j \in [d]$, extreme points $\mathbf{QEP1}_j, j \in [d]$ are never optimal. Hence, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. The only case left to consider is when $x_k = M_z$ for all $k$ with $\gamma_k = 0$. In this case, inequality (61) for $k^*$ holds at equality for all optimal solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. This completes the proof of Case A.

**Case B.** Suppose that $\gamma^\beta < 0$. As before, we can assume that $\gamma_j \geq 0$ for all $j \in [d]$, and that $\gamma_{k^*} = 0$ for some $k^* \in [d]$. Finally, we can assume that $\gamma^v = 0$ by subtracting $\gamma^v(v - \delta - z + \sum_{j\in[d]} x_j c_j)$ from the objective. Under these assumptions, we can express the cost of each extreme point solution (denoted by $C(\cdot)$) given in Proposition 3.3:

$$\begin{aligned}
C(\mathbf{QEP1}_k) &= \gamma_k + \gamma^\delta x_k \text{ for } k \in [d] \text{ with } x_k > 0,\\
C(\mathbf{QEP2}_k) &= \gamma_k + \gamma^z x_k \text{ for } k \in [d],\\
C(\mathbf{QEP3}_k) &= \gamma_k + \gamma^z x_k + \gamma^\beta \text{ for } k \in [d],\\
C(\mathbf{QEP4}_k) &= \gamma_k + \gamma^z M_z + \gamma^\beta \text{ for } k \in [d] \text{ with } x_k < M_z.
\end{aligned}$$

Note that due to the assumption that $\gamma^\beta < 0$, $C(\mathbf{QEP2}_k) > C(\mathbf{QEP3}_k)$ for all $k \in [d]$. So the extreme points $\mathbf{QEP2}_k, k \in [d]$ are never optimal under these cost assumptions. We use this observation in the following cases we consider.

(i) $\gamma^z > 0$. In this case, $C(\mathbf{QEP4}_k) > C(\mathbf{QEP3}_k)$ for all $k \in [d]$. (Recall that $\mathbf{QEP4}_k$ exists for some $k \in [d]$ only if $M_z > x_k$.) So the extreme points $\mathbf{QEP4}_k, k \in [d]$ are never optimal under these cost assumptions. Hence, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. Therefore, we can assume that $\gamma^z \leq 0$.

(ii) $\gamma^z \leq 0$. If $\gamma^z \leq \gamma^\delta$, then $C(\mathbf{QEP1}_k) > C(\mathbf{QEP3}_k)$ for all $k \in [d]$. Therefore, extreme points $\mathbf{QEP1}_k, k \in [d]$ are never optimal. Hence, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. As a result, we can assume that $\gamma^\delta < \gamma^z \leq 0$ and $C(\mathbf{QEP4}_k) \leq C(\mathbf{QEP3}_k)$ for all $k \in [d]$. Note that because $\gamma^\delta < 0$, $0 > \gamma^\delta x_1 \geq \gamma^\delta x_2 \geq \cdots \geq \gamma^\delta x_d$. In addition, $\min_{k\in[d]}\{C(\mathbf{QEP4}_k)\} = C(\mathbf{QEP4}_{k^*}) = \gamma^z M_z + \gamma^\beta$. If $\gamma^\delta x_d > \gamma^z M_z + \gamma^\beta$, then extreme points $\mathbf{QEP1}_k, k \in [d]$ are never optimal. Hence, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. So we can assume that $\gamma^\delta x_d \leq \gamma^z M_z + \gamma^\beta$. If $\gamma^\delta x_1 < \gamma^z M_z + \gamma^\beta$, then extreme points $\mathbf{QEP4}_k, k \in [d]$ are never optimal. Hence, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. So we can assume that $\gamma^\delta x_1 \geq \gamma^z M_z + \gamma^\beta$. Let $\bar{k} := \min\{j \in [d] : \gamma^\delta x_j \leq \gamma^z M_z + \gamma^\beta\}$. If there exists $j \geq \bar{k}$ such that $C(\mathbf{QEP1}_j) = \gamma_j + \gamma^\delta x_j < \gamma^z M_z + \gamma^\beta = C(\mathbf{QEP4}_{k^*}) \leq C(\mathbf{QEP4}_k)$ for all $k \in [d]$, then extreme points $\mathbf{QEP4}_k, k \in [d]$ are never optimal. Hence, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. Therefore, we have $\gamma_j + \gamma^\delta x_j = \gamma^z M_z + \gamma^\beta$ for all $j \geq \bar{k}$. Under these assumptions, the optimal solutions are $\mathbf{QEP1}_j$ for $j \geq \bar{k}$; $\mathbf{QEP4}_k$ for $k \in [d]$ such that $\gamma_k = 0$; and $\mathbf{QEP3}_k$ for $k \in [d]$ such that $\gamma_k = 0$ if $\gamma^z = 0$. Then inequality (62) for $\bar{k}$ holds at equality for all optimal solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$. This completes the proof.

□

Remark 3.3 *Note that in the definition of the set $\mathcal{Q}$, we used weaker bounds on $v, \delta$ and $z$ than are available using the improvements proposed in Section 3. In particular, we can let $z \leq U$, where $U$ is the upper bound* *on VaR obtained by using the quantile information (as described in Section 3.2.1); in most cases, $U < M_z$. Then, we simply update inequality (61) as*

$$v \leq \sum_{j\in[d]} [x_k - x_j]_+ c_j + (U - x_k)\beta. \tag{63}$$

*In addition, we can let $z \geq L$, using the lower bound information on VaR, and typically $L > 0$. If this is the case, then we can define new variables $z' = z - L$ and $\delta' = \delta - L$, and let $M'_z = M_z - L$ and $M'_\delta = M_\delta - L$, and obtain a linearization polytope of the same form as $\mathcal{Q}$ in the $(\mathbf{c}, v, \delta', \beta, z')$ space. The updated inequality (62) in the original space becomes*

$$\delta \leq \sum_{j\in[d]} [x_j - x_k]_+ c_j + (x_k - L)(1 - \beta). \tag{64}$$

*Therefore, our results hold for $L > 0$ with this translation of variables.*

*Finally, from Section 3, we know that $v \leq M_{i\cdot}\beta$ and $\delta \leq M_{\cdot i}(1-\beta)$ for the given scenario $i \in [n]$ for which the linearization polytope is written. Again, in most cases, $M_{i\cdot} \leq M_v$ and $M_{\cdot i} \leq M_\delta$. In this case, we cannot have $c_k = 1$ and $z = L$ for $k$ such that $x_k - L > M_{\cdot i}$, because otherwise $\delta = [\sum_{j\in[d]} c_j x_j - z]_+ = x_k - L > M_{\cdot i}$, which violates the constraint $\delta \leq \mathcal{M}_{\cdot i}(1-\beta)$. Hence for all $k$ with $x_k - L > M_{\cdot i}$, if $c_k > 0$ and $z = L$, then we must have $c_\ell = 1 - c_k$ for some $\ell \in [d]$ with $x_\ell - L < M_{\cdot i}$. Then, $\delta = M_{\cdot i}$ in such an extreme point solution. In this case, we can construct an equivalent polyhedron where we let $x_k^\ell = M_{\cdot i} + L$ for all $k \in [d]$ such that $x_k - L > M_{\cdot i}$ and $\ell \in [d]$ such that $x_\ell - L < M_{\cdot i}$. Similarly, we cannot have $c_k = 1$ and $z = U$ for $k$ such that $U - x_k > M_{i\cdot}$, because otherwise $v = [z - \sum_{j\in[d]} c_j x_j]_+ = U - x_k > M_{i\cdot}$, which violates the constraint $v \leq \mathcal{M}_{i\cdot}\beta$. If $c_k > 0$ for $k$ with $U - x_k > M_{i\cdot}$, then we must have $c_\ell = 1 - c_k$ for some $\ell \in [d]$ with $U - x_\ell < M_{i\cdot}$. Then $v = M_{i\cdot}$ in such an extreme point solution. In this case, we can construct an equivalent polyhedron where we let $\bar{x}_k^\ell = U - M_{i\cdot}$ for all $k \in [d]$ such that $U - x_k > M_{i\cdot}$ and $\ell \in [d]$ such that $U - x_\ell < M_{i\cdot}$. The resulting polyhedron satisfies the bound assumptions in the definition of $\mathcal{Q}$, and the non-trivial inequalities that define its convex hull are given by (63) for $k \in [d]$ such that $U - x_k \leq M_{i\cdot}$, and inequality (64) for $k \in [d]$ such that $x_k - L \leq M_{\cdot i}$. Note that after this update inequalities (63) for $k \in [d]$ such that $U - x_k = M_{i\cdot}$ reduces to $v \leq M_{i\cdot}\beta$, and inequality (64) for $k \in [d]$ such that $x_k - L = M_{\cdot i}$ reduces to $\delta \leq \mathcal{M}_{\cdot i}(1-\beta)$.*

Translating back to the original space of variables and re-introducing the scenario indices we have the following corollary.

Corollary 3.2 *For $i \in [n]$, consider the polyhedron $\mathcal{Q}'_i = \{(\mathbf{c}, v_i, \delta_i, \beta_i, z) \in \mathbb{R}_+^{d+4} : (46)-(48), (50), (57), \beta_i \in \{0, 1\}\}$. Then conv$(\mathcal{Q}'_i)$ is completely described by adding inequalities*

$$v_i \leq \sum_{j\in[d]} [x_{ik} - x_{ij}]_+ c_j + (U - x_{ik})\beta_i, \quad \forall\ k \in [d]\ :\ U - x_{ik} < M_{i\cdot}, \tag{65}$$

$$\delta_i \leq \sum_{j\in[d]} [x_{ij} - x_{ik}]_+ c_j + (x_{ik} - L)(1 - \beta_i), \quad \forall\ k \in [d] : x_{ik} - L < M_{\cdot i} \tag{66}$$

*to the original constraints (46)-(48),(50), and (57).*

**4. Cut Generation for Optimization with Multivariate SSD Constraints** In this section, we study the cut generation problem arising in optimization problems of the form $(\mathbf{G}-\mathbf{MSSD})$. As in Section 3, we focus on solving the cut generation problems given two $d$-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ with realizations $\mathbf{x}_1,\dots,\mathbf{x}_n$ and $\mathbf{y}_1,\dots,\mathbf{y}_m$, respectively. Let $p_1,\dots,p_n$ and $q_1,\dots,q_m$ denote the corresponding probabilities, and let $C\subset\mathbb{R}^d_+$ be a polytope of scalarization vectors.

By Definition 2.3 and relation (7), $\mathbf{X}$ is said to dominate $\mathbf{Y}$ in *polyhedral linear second order* with respect to $C$ if and only if

$$\mathbb{E}([\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{X}]_+)\le\mathbb{E}([\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{Y}]_+),\quad l\in[m],\ \mathbf{c}\in C,\ \text{or equivalently,}$$

$$\sum_{i\in[n]}p_i[\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{x}_i]_+\le\sum_{k\in[m]}q_k[\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{y}_k]_+,\quad l\in[m],\ \mathbf{c}\in C. \tag{67}$$

As discussed in Section 2.3, Homem-de-Mello and Mehrotra (2009) show that for finite probability spaces it is sufficient to consider a finite subset of scalarization vectors, obtained as projections of the vertices of $m$ polyhedra. Specifically, each polyhedron corresponds to a realization of the benchmark random vector $\mathbf{Y}$ and is given by $P_l=\{w_k\ge\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{y}_k,\ k\in[m],\ \mathbf{c}\in C,\ \mathbf{w}\in\mathbb{R}^m_+\}$ for $l\in[m]$. Thus, $(\mathbf{G}-\mathbf{MSSD})$ can be reformulated as an optimization problem with exponentially many constraints. Clearly, we cannot expect to solve this problem with exponentially many constraints, hence a delayed constraint generation algorithm is proposed (Homem-de-Mello and Mehrotra, 2009) such that the SSD constraints corresponding to a subset of the scalarization vectors are initially present in the formulation. Then given a solution to this intermediate problem, a cut generation problem is solved to identify if there is a constraint violated by the current solution, or if the current solution is optimal.

Due to the structure of the SSD relation (67), a separate cut generation problem is defined for each realization of the benchmark random vector. Thus, in contrast to the CVaR-constrained models, the number of cut generation problems depends on the number of benchmark realizations. The cut generation problem associated with the $l$th realization of the benchmark vector $\mathbf{Y}$ reads:

$$(\mathbf{CutGen\_SSD})\qquad \min_{\mathbf{c}\in C}\ \sum_{k\in[m]}q_k[\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{y}_k]_+-\sum_{i\in[n]}p_i[\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{x}_i]_+.$$

Note that this is a minimization of the difference of convex (DC) functions. Dentcheva and Wolfhagen (2013) use methods from DC programming to solve this problem directly. Similar to the univariate case in (8)-(9), we can easily linearize the first type of shortfalls featured in the objective function:

$$\min\left\{\sum_{k\in[m]}q_kw_k-\sum_{i\in[n]}p_i[\mathbf{c}^\top\mathbf{y}_l-\mathbf{c}^\top\mathbf{x}_i]_+\ :\ (\mathbf{c},\mathbf{w})\in P_l\right\}, \tag{68}$$

which results in a concave minimization with potentially many local minima.

If the optimal objective function value is smaller than 0; then there is a scalarization vector for which the SSD condition associated with the $l$th realization is violated. Note that it is crucial to solve the cut generation problem exactly for the correct execution of the solution method for $(\mathbf{G}-\mathbf{MSSD})$. If we obtain a local minimum and the objective is positive, we can prematurely stop the algorithm. Also, methods from DC programming and concave minimization may not fully utilize the polyhedral nature of the objective and the constraints.

The main challenge in the cut generation problem (68) is to linearize the second type of shortfalls appearing in the objective function. In this regard, Homem-de-Mello and Mehrotra (2009) introduce additional variables and constraints, and obtain the following MIP formulation of (**CutGen_SSD**) associated with the $l$th realization of the benchmark vector $\mathbf{Y}$:

$$
\begin{aligned}
(\mathbf{MIP\_SSD}_l) \quad \min \quad & \sum_{k\in[m]} q_k w_k - \sum_{i\in[n]} p_i v_i \\
\text{s.t.} \quad & w_k \geq \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{y}_k, && \forall\, k \in [m], \\
& \mathbf{w} \in \mathbb{R}^m_+, \\
& v_i - \delta_i = \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i, && \forall\, i \in [n], \quad (69)\\
& v_i \leq M_i \beta_i, && \forall\, i \in [n], \quad (70)\\
& \delta_i \leq \hat{M}_i(1-\beta_i), && \forall\, i \in [n], \quad (71)\\
& \mathbf{c} \in C, \quad \mathbf{v} \in \mathbb{R}^n_+, \quad \boldsymbol{\delta} \in \mathbb{R}^n_+, \quad \boldsymbol{\beta} \in \{0,1\}^n. && (72)
\end{aligned}
$$

Here we can set $M_i = \max\{\max_{\mathbf{c}\in C} \{\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i\}, 0\}$ and $\hat{M}_i = -\min\{\min_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i\}, 0\}$. It is easy to see that this formulation guarantees that $v_i = [\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i]_+$ for all $i \in [n]$.

The authors also propose concavity and convexity cuts to strengthen the formulation ($\mathbf{MIP\_SSD}_l$). However, the concavity cuts require the complete enumeration of a set of edge directions (may be exponential), and solving a system of linear equations based on this enumeration. Hence, this may not be practicable. In addition, the convexity cuts require the solution of another cut generation LP in higher dimension. Indeed, in their computational study, Hu et al. (2011) do not utilize these cuts and solve ($\mathbf{MIP\_SSD}_l$) directly. They also note that this step is the bottleneck taking over 90% of the total solution time, and it needs to be improved.

**4.1 New developments** We begin by presenting an analogue of Proposition 3.2, which provides valid ordering inequalities that strengthen the formulation ($\mathbf{MIP\_SSD}_l$). Then, we study the structure of a generalization of the linearization polytope defined by (69)-(72) for a given $l \in [m]$ and $i \in [n]$. We give two classes of valid inequalities analogous to those in Proposition 3.4 for this polytope. Furthermore, we show that these inequalities are enough to give the complete linear description when added to the formulation with $C = \{\mathbf{c} \in \mathbb{R}^d_+ : \sum_{j\in[d]} c_j = 1\}$.

Lemma 4.1 *The* ordering inequalities (52)–(53) *are also valid for* ($\mathbf{MIP\_SSD}_\mathbf{l}$) *given $l$th realization of the benchmark random vector* $\mathbf{Y}$.

This claim immediately follows from the trivial observation that $z$ can be replaced by $\mathbf{c}^\top \mathbf{y}_l$ in (46) (and also in the proof of Proposition 3.2) for any $l \in [m]$. Next we give a polyhedral study of the set defining the linearization of the piecewise convex shortfall terms.

**4.1.1 Linearization of $[\mathbf{a}^\top \mathbf{c}]_+$ in (CutGen_SSD)** For a given vector $\mathbf{a} \in \mathbb{R}^d$, consider the convex function $f(c) = [\mathbf{a}^\top \mathbf{c}]_+ := \max\{0, \mathbf{a}^\top \mathbf{c}\}$ for $\mathbf{c} \in \mathbb{R}^d_+$ such that $\sum_{j\in[d]} c_j = 1$. This function appears in the cut generation problems for optimization under multivariate risk given in (68), where $\mathbf{a} = \mathbf{y}_l - \mathbf{x}_i$ for some $l \in [m]$ and $i \in [n]$. An MIP linearizing this term is given in ($\mathbf{MIP\_SSD}_l$). (Note that this structure also appears in the cut generation problem for CVaR (27)-(31), where we let $\mathbf{a} = \mathbf{x}_k - \mathbf{x}_i$, for $i, k \in [n] : i < k$.) Let $D^+ = \{j \in [d] : a_j \geq 0\}$ and $D^- = \{j \in [d] : a_j < 0\}$. Due to the nature of the cut generation problems, we can assume that $D^+ \neq \emptyset$ and $D^- \neq \emptyset$ (otherwise, we can fix the corresponding binary variables). Without loss of generality, we assume that $D^+ = \{1, \ldots, d_1\}$ with $a_1 \leq a_2 \leq \cdots \leq a_{d_1}$, and $D^- = \{d_1+1, \ldots, d\}$ with $-a_{d_1+1} \leq -a_{d_1+2} \leq \cdots \leq -a_d$.

In this subsection, we drop the scenario indices, and study the polytope given by

$$v - \delta = \sum_{j \in [d]} a_j c_j, \tag{73}$$

$$v \leq \bar{M}_v \beta, \tag{74}$$

$$\delta \leq \bar{M}_\delta (1-\beta), \tag{75}$$

$$\sum_{j \in [d]} c_j = 1 \tag{76}$$

$$\mathbf{c}, v, \delta \geq 0, \tag{77}$$

$$\beta \in \{0,1\}, \tag{78}$$

where $\bar{M}_v = a_{d_1}$ and $\bar{M}_\delta = -a_d$.

Let $\mathcal{S} = \{(\mathbf{c}, v, \delta, \beta) : (73) - (78)\}$. First, we characterize the extreme points of conv$(\mathcal{S})$. Recall that $e_k$ denotes the $d$-dimensional unit vector with 1 in the $k$th entry and zeroes elsewhere.

Proposition 4.1 *The extreme points* $(\mathbf{c}, v, \delta, \beta)$ *of conv*$(\mathcal{S})$ *are as follows:*

**EP1**$_k$**:** $(e_k, a_k, 0, 1)$ *for all* $k \in D^+$,

**EP2**$_\ell$**:** $(e_\ell, 0, -a_\ell, 0)$ *for all* $\ell \in D^-$,

**EP3**$_{k,\ell}$**:** $(\frac{-a_\ell}{a_k - a_\ell} e_k + \frac{a_k}{a_k - a_\ell} e_\ell, 0, 0, 1)$ *for all* $k \in D^+$ *and* $\ell \in D^-$,

**EP4**$_{k,\ell}$**:** $(\frac{-a_\ell}{a_k - a_\ell} e_k + \frac{a_k}{a_k - a_\ell} e_\ell, 0, 0, 0)$ *for all* $k \in D^+$ *and* $\ell \in D^-$.

Proof. First, note that, from the definition of $\bar{M}_v$, $\bar{M}_\delta$, $D^+$ and $D^-$, we have $0 \leq a_k \leq \bar{M}_v$ for all $k \in D^+$ and $0 < -a_\ell \leq \bar{M}_\delta$ for $\ell \in D^-$. Hence, points **EP1**$_k$ and **EP2**$_\ell$ are feasible. It is also easy to see that they are extreme points of conv$(\mathcal{S})$. Finally, observe that any other feasible point with $0 < c_k < 1$ for some $k \in D^+$, we must have $c_\ell = 1 - c_k$ for some $\ell \in D^-$ in any extreme point of conv$(\mathcal{S})$ such that $c_k a_k + c_\ell a_\ell = 0 = v = \delta$. In this case, we can have either $\beta = 0$ or $\beta = 1$. As a result, we obtain the extreme points **EP3**$_{k,\ell}$ and **EP4**$_{k,\ell}$. This completes the proof. □

Next we give valid inequalities for $\mathcal{S}$.

Proposition 4.2 *For* $k = 1, \ldots, d_1$, *the inequality*

$$v \leq \sum_{j=1}^{d_1} [a_j - a_k]_+ c_j + a_k \beta \tag{79}$$

*is valid for* $\mathcal{S}$. *Similarly, for* $k = d_1+1, \ldots, d$, *the inequality*

$$\delta \leq \sum_{j=d_1+1}^{d} [a_k - a_j]_+ c_j - a_k(1-\beta) \tag{80}$$

*is valid for* $\mathcal{S}$.

Proof. If $\beta = 0$, then $v = 0$ from (74). Because $\mathbf{c} \geq \mathbf{0}$, inequality (79) holds trivially. If $\beta = 1$, then $\delta = 0$ from (75). Thus, for any $k = 1, \ldots, d_1$,

$$v - \delta = v = \sum_{j\in[d]} a_j c_j \leq \sum_{j=1}^{d_1} a_j c_j = \sum_{j=1}^{d_1}(a_j - a_k)c_j + a_k \sum_{j=1}^{d_1} c_j$$
$$\leq \sum_{j=1}^{d_1}[a_j - a_k]_+ c_j + a_k = \sum_{j=1}^{d_1}[a_j - a_k]_+ c_j + a_k\beta,$$

where the last inequality follows from (76).

To see the validity of inequality (80), note that equality (73) can be rewritten as $\delta - v = \sum_{j\in[d]}(-a_j)c_j$. Thus, we obtain an equivalent set where $v$ and $\delta$, and $D^+$ and $D^-$ are interchanged.

□

Remark 4.1 *Inequality (74) is a special case of (79) with $k = d_1$, and inequality (75) is a special case of (80) with $k = d$.*

Remark 4.2 *Note that $\beta \geq 0$ is implied by inequality (74), and $\beta \leq 1$ is implied by (75).*

Remark 4.3 *Consider a related set, $\mathcal{T}$, where constraint (76) is relaxed to $\sum_{j\in[d]} c_j \leq 1$. This set can be written in the form of the set $\mathcal{S}$ with $\mathbf{c} \in \mathbb{R}^{d+1}$, where $D = \{0, \ldots, d\}$, and $a_0 = 0$. In this case, inequality (79) for $k = 0$ is given by $v \leq \sum_{j=1}^{d_1} a_j c_j$.*

Theorem 4.1 *conv$(\mathcal{S})$ is completely described by equalities (73) and (76), and inequalities (77), (79), and (80).*

Proof. Let $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$, denote the index set of extreme point optimal solutions to the problem $\min\{\boldsymbol{\gamma}^\top \mathbf{c} + \gamma^v v + \gamma^\delta \delta + \gamma^\beta \beta : (\mathbf{c}, v, \delta, \beta) \in \text{conv}(\mathcal{S})\}$, where $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta) \in \mathbb{R}^{d+3}$ is an arbitrary objective vector, not perpendicular to the smallest affine subspace containing conv$(\mathcal{S})$. In other words, $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta) \neq \lambda(\mathbf{a}, -1, 1, 0)$ and $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta) \neq \lambda(\mathbf{1}, 0, 0, 0)$ for $\lambda \in \mathbb{R}$. Therefore, the set of optimal solutions is not conv$(\mathcal{S})$ (conv$(\mathcal{S}) \neq \emptyset$). We prove the theorem by giving an inequality among (77), (79), and (80) that is satisfied at equality by $(\mathbf{c}^\kappa, v^\kappa, \delta^\kappa, \beta^\kappa)$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$ for the given objective vector. Then, since $(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$ is arbitrary, for every facet of conv$(\mathcal{S})$, there is an inequality among (77), (79), and (80) that defines it. We consider all possible cases.

**Case A.** Suppose that $\gamma^\beta \geq 0$. Without loss of generality we can assume that $\gamma^\delta = 0$ by adding $\gamma^\delta(v - \delta - \sum_{j\in[d]} a_j c_j)$ to the objective. From equation (73) the added term is equal to zero, and so this operation does not change the set of optimal solutions. Furthermore, we can also assume that $\gamma_j \geq 0$ for all $j \in D$ without loss of generality by subtracting $\gamma_{\min}(\sum_{j\in[d]} c_j)$ from the objective, where $\gamma_{\min} := \min_{j\in[d]}\{\gamma_j\}$. From equation (76), the added term is a constant $(-\gamma_{\min})$, and so this operation does not change the set of optimal solutions. Note also that after this update $\gamma_{\min} = 0$. Therefore, for the case that $\gamma^\beta \geq 0$, we assume that $\gamma^\delta = 0$ and $\gamma_{\min} = 0$. Under these assumptions, we can express the cost of each extreme point solution (denoted by $C(\cdot)$) given in Proposition 4.1:

$C(\mathbf{EP1}_k)$ $= \gamma_k + \gamma^v a_k + \gamma^\beta$ for $k \in D^+$,

$C(\mathbf{EP2}_\ell)$ $= \gamma_\ell$ for $\ell \in D^-$,

$C(\mathbf{EP3}_{k,\ell}) \; = \gamma_k \frac{-a_\ell}{a_k - a_\ell} + \gamma_\ell \frac{a_k}{a_k - a_\ell} + \gamma^\beta$ for $k \in D^+$ and $\ell \in D^-$,

$C(\mathbf{EP4}_{k,\ell}) \; = \gamma_k \frac{-a_\ell}{a_k - a_\ell} + \gamma_\ell \frac{a_k}{a_k - a_\ell}$ for $k \in D^+$ and $\ell \in D^-$.

Let $k^* \in \arg\min\{\gamma_j, j \in D^+\}$ and $\ell^* \in \arg\min\{\gamma_j, j \in D^-\}$. Note that $\min\{\gamma_{k^*}, \gamma_{\ell^*}\} = \gamma_{\min} = 0$. Observe that $C(\mathbf{EP2}_\ell) < C(\mathbf{EP4}_{k,\ell})$ for $k \in D^+$ and $\ell \in D^-$ if $\gamma_\ell < \gamma_k$. On the other hand, if $\gamma_\ell > \gamma_k$, then $C(\mathbf{EP2}_\ell) > C(\mathbf{EP4}_{k,\ell})$ for $k \in D^+$ and $\ell \in D^-$. Also, the only extreme points for which $\delta > 0$ are $\mathbf{EP2}_\ell$ for $\ell \in D^-$ with $-a_\ell > 0$, and the only extreme points for which $v > 0$ are $\mathbf{EP1}k$ for $k \in D^+$ with $a_k > 0$. We use these observations in the following cases we consider.

(i) $\gamma_{\ell^*} = 0 < \gamma_{k^*}$. In this case, $\mathbf{EP4}_{k,\ell}$ cannot be an optimal solution for any $k \in D^+$ and $\ell \in D^-$. Furthermore, because of the assumption that $\gamma^\beta \geq 0$, $\mathbf{EP3}_{k,\ell}$ cannot be an optimal solution for any $k \in D^+$ and $\ell \in D^-$ either.

   (a) If there exists $j \in D^+$ such that $C(\mathbf{EP1}_j) = \gamma_j + \gamma^v a_j + \gamma^\beta > 0 = C(\mathbf{EP2}_{\ell^*})$, then $c_j^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma_k + \gamma^v a_k + \gamma^\beta \leq 0$ for all $k \in D^+$. Now suppose that $\gamma_j + \gamma^v a_j + \gamma^\beta < 0$ for some $j \in D^+$. In this case, $C(\mathbf{EP1}_j) < C(\mathbf{EP2}_\ell)$ for all $\ell \in D^-$. Therefore, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma_k + \gamma^v a_k + \gamma^\beta = 0$ for all $k \in D^+$.

   (b) If there exists $j \in D^-$ such that $C(\mathbf{EP2}_j) = \gamma_j > 0 = C(\mathbf{EP2}_{\ell^*})$, then $c_j^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma_\ell = 0$ for all $\ell \in D^-$. In summary, for the case that $\gamma^\beta \geq 0$ and $\gamma_{\ell^*} = 0 < \gamma_{k^*}$, we have $\gamma_k + \gamma^v a_k + \gamma^\beta = 0$ for all $k \in D^+$ and $\gamma_\ell = 0$ for all $\ell \in D^-$. In this case, the set $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$ is given by $\mathbf{EP1}_k$ for all $k \in D^+$ and $\mathbf{EP2}_\ell$ for all $\ell \in D^-$. Inequality (79) for $k = 1$ is tight for all these extreme point optimal solutions. Hence, the proof is complete for this case.

(ii) $\gamma_{\ell^*} > \gamma_{k^*} = 0$. Recall that, in this case, $C(\mathbf{EP4}_{k^*,\ell}) < C(\mathbf{EP2}_\ell)$ for all $\ell \in D^-$. Therefore, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. Hence, the proof is complete for this case.

(iii) $\gamma_{\ell^*} = \gamma_{k^*} = 0$.

   (a) If there exists $j \in D^-$ such that $\gamma_j > 0$, then $c_j^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma_\ell = 0$ for all $\ell \in D^-$.

   (b) Suppose that $\gamma_j + \gamma^v a_j + \gamma^\beta < 0$ for some $j \in D^+$. In this case, $\mathbf{EP1}_j$ has a strictly better objective value than $\mathbf{EP2}_\ell$, $\mathbf{EP3}_{k,\ell}$, and $\mathbf{EP4}_{k,\ell}$ for all $k \in D^+$ and $\ell \in D^-$. Therefore, $\delta^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma_k + \gamma^v a_k + \gamma^\beta \geq 0$ for all $k \in D^+$. If there exists $j \in D^+$ such that $\gamma_j > 0$ and $\gamma_j + \gamma^v a_j + \gamma^\beta > 0$, then $c_j^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that at least one of the conditions $\gamma_k = 0$ or $\gamma_k + \gamma^v a_k + \gamma^\beta = 0$ holds for all $k \in D^+$. Let $D_0^+ = \{j \in D^+ : \gamma_j = 0\}$ and $D_1^+ = D^+ \setminus D_0^+$. Note that $k^* \in D_0^+$ and $\gamma_k + \gamma^v a_k + \gamma^\beta = 0$ for all $k \in D_1^+$.

   (c) Suppose that $\gamma_k = 0$ for all $k \in D^+$ (i.e., $D_1^+ = \emptyset$). Recall that we also have $\gamma_\ell = 0$ for all $\ell \in D^-$, $\gamma^\delta = 0$ and $\gamma^\beta \geq 0$. If $\gamma^\beta = 0$, then $\gamma^v$ cannot equal to 0 (then all solutions are optimal). Suppose that $\gamma^\beta = 0$, then $\gamma^v > 0$ (because we showed that $\gamma_k + \gamma^v a_k + \gamma^\beta \geq 0$ for all $k \in D^+$). Then $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma^\beta > 0$. If $\gamma^v \geq 0$, then $\mathbf{EP1}_k$ is not optimal for any $k \in D^+$. Therefore, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. So we can assume that $\gamma^v < 0$. Because we showed that $\gamma_k + \gamma^v a_k + \gamma^\beta \geq 0$ for all $k \in D^+$, and we assume that $\gamma_k = 0$ for all $k \in D^+$, we have $\gamma^\beta \geq -\gamma^v a_{d_1}$. If $\gamma^v a_{d_1} + \gamma^\beta > 0$, then $\mathbf{EP1}_k$ is not optimal for any $k \in D^+$. Therefore, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$, and we can assume that $\gamma^v a_{d_1} + \gamma^\beta = 0$. In this case, inequality (79) for $k = d_1$ holds at equality for the set of all optimal extreme solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$ (namely, $\mathbf{EP1}_k$ for $k \in D^+$ with $a_k = a_{d_1}$, $\mathbf{EP2}_\ell$ and $\mathbf{EP4}_{j,\ell}$ for all $j \in D^+$ and $\ell \in D^-$).

(d) There exists $k \in D^+$ such that $\gamma_k > 0$ (i.e., $D_1^+ \neq \emptyset$). In this case, for $k \in D_1^+$, $\gamma_k = -\gamma^v a_k - \gamma^\beta > 0$. Because $\gamma^\beta \geq 0$, we must have $\gamma^v < 0$ and $a_k > 0$ for $k \in D_1^+$. In this case, we cannot have $\gamma^\beta = 0$ (unless $a_j = 0$ for all $j \in D_0^+$), because otherwise $\gamma_j + \gamma^v a_j + \gamma^\beta < 0$ for $j \in D_0^+$ with $a_j > 0$ violating the condition in part (b) that $\gamma_k + \gamma^v a_k + \gamma^\beta \geq 0$ for all $k \in D^+$. So $\gamma^\beta > 0$ and $\mathbf{EP3}_{j,\ell}$ is not optimal for any $j \in D^+, \ell \in D^-$. Let $k_1 = \min\{j \in D_1^+\}$, then we must have $k \in D_1^+$ for all $k \in D^+$ with $k > k_1$. In this case, the set of all optimal solutions is given by $\mathbf{EP1}_k$ for $k \in D_1^+$, $\mathbf{EP2}_\ell$ and $\mathbf{EP4}_{j,\ell}$ for all $j \in D_0^+$ and $\ell \in D^-$, where the optimal objective value is zero. Then inequality (79) for $k = k_1$ holds at equality for the set of all optimal extreme solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$. The last case to consider is that $a_j = 0$ for all $j \in D_0^+$ and hence $\gamma^\beta = 0$. In this case, inequality (79) for $k = k^*$ holds at equality for the set of all optimal extreme solutions $O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta)$ (namely, $\mathbf{EP1}_k$ for $k \in D^+$, $\mathbf{EP2}_\ell$, $\mathbf{EP3}_{j,\ell}$ and $\mathbf{EP4}_{j,\ell}$ for all $j \in D_0^+$ and $\ell \in D^-$).

**Case B.** Suppose that $\gamma^\beta < 0$. Without loss of generality we can assume that $\gamma^v = 0$ by subtracting $\gamma^v(v - \delta - \sum_{j\in[d]} a_j c_j)$ from the objective. From equation (73), the subtracted term is equal to zero, and so this operation does not change the set of optimal solutions. As argued in the proof of the validity of (80), equality (73) can be rewritten as $\delta - v = \sum_{j\in[d]}(-a_j)c_j$. Thus, we obtain an equivalent set where $v$ and $\delta$, and $D^+$ and $D^-$ are interchanged. Thus, the proof is complete, using the same arguments as in Case A and inequalities (80).

□

In line with the above analysis, we introduce $a_{ij} = (\mathbf{y}_l - \mathbf{x}_i)_j$, $D_i^+ = \{j \in [d] : a_{ij} \geq 0\}$ and $D_i^- = \{j \in [d] : a_{ij} < 0\}$ for all $i \in [n]$, and conclude this section by presenting the resulting enhanced MIP formulation of (**CutGen_SSD**) for the $l$th realization of $\mathbf{Y}$:

$$
\begin{aligned}
(\mathbf{NewMIP\_SSD}_l) \quad \min \quad & \sum_{k\in[m]} q_k w_k - \sum_{i\in[n]} p_i v_i \\
\text{s.t.} \quad & w_k \geq \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{y}_k, && \forall\, k \in [m], \\
& \mathbf{w} \in \mathbb{R}^m_+, \\
& v_i - \delta_i = \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i, && \forall\, i \in [n], \\
& v_i \leq \sum_{j\in D_i^+} [a_{ij} - a_{ik}]_+ c_j + a_{ik}\beta_i, && \forall\, i \in [n], k \in D_i^+, \\
& \delta_i \leq \sum_{j\in D_i^-} [a_{ik} - a_{ij}]_+ c_j - a_{ik}(1 - \beta_i), && \forall\, i \in [n], k \in D_i^-, \\
& \mathbf{c} \in C, \quad \mathbf{v} \in \mathbb{R}^n_+, \quad \boldsymbol{\delta} \in \mathbb{R}^n_+, \quad \boldsymbol{\beta} \in \{0,1\}^n.
\end{aligned}
$$

**5. Computational Study** The goals of our computational study are two-fold. In the first part, we demonstrate that the methods developed in Section 3.2 – including variable fixing, bounding, and incorporating valid inequalities – are effective in solving (**CutGen_CVaR**). In the second part, we perform a similar analysis for the methods presented in Section 4 for (**CutGen_SSD**).

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 5400 seconds. `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). In cut generation problems, the optimal objective function can take any value including 0, and so in order to provide more insight, we calculate two types of relative optimality gap: $\mathrm{ROG}_1 = |\,\mathrm{LB}-\mathrm{UB}\,|/(|\,\mathrm{UB}\,|)$ and $\mathrm{ROG}_2 = |\,\mathrm{LB}-\mathrm{UB}\,|/(|\,\mathrm{LB}\,|)$. It is easy to see that the maximum of $\mathrm{ROG}_1$ and $\mathrm{ROG}_2$ is an upper bound on the actual relative optimality gap; we do not report $\mathrm{ROG}_1$ when $|\,\mathrm{UB}\,| = 0$ or `CPLEX` yields a trivial lower bound of $-\infty$.

**5.1 Generation of the problem instances** In this section, we describe two sets of data used for our computational experiments.

**5.1.1 Homeland security budget allocation** We test the computational effectiveness of our proposed methods on a homeland security budget allocation (HSBA) problem presented in Hu et al. (2011) for optimization under multivariate polyhedral SSD constraints. We follow the related data generation scheme described in Noyan and Rudolf (2013), where the polyhedral SSD constraints are replaced by the CVaR-based ones. The main problem is to allocate a fixed budget to ten urban areas in order to prevent, respond to, and recover from national disasters. The risk share of each area is based on four criteria: property losses, fatalities, air departures, and average daily bridge traffic. The penalty for allocations under the risk share is expressed by a budget misallocation function associated with each criterion, and these functions are used as the multiple random performance measures of interest. In order to be consistent with our convention of preferring large values, we construct random outcome vectors of interest from the negative of the budget misallocation functions associated with four criteria. Two different benchmarks are considered: one based on average government allocations by the Department of Homeland Security's Urban Areas Security Initiative, and one based on suggestions in the RAND report by Willis et al. (2005). The scalarization polyhedron is of the form $C = \left\{\mathbf{c} \in \mathbb{R}^4 \ : \ \|\mathbf{c}\|_1 = 1,\ c_i \geq c_i^* - \frac{\theta}{3}\right\}$, where $\mathbf{c}^* \in \mathbb{R}^4$ is a center satisfying $\|\mathbf{c}^*\|_1 = 1$, and $\theta \in [0,1]$ is a constant for which $\frac{\theta}{3} \leq \min\limits_{i\in\{1,\ldots,4\}} c_i^*$ holds. We consider the "base case" with $\theta = 0.25$ and $\mathbf{c}^* = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, unless otherwise stated. We refer the reader to Hu et al. (2011) and Noyan and Rudolf (2013) for more details on the data generation.

Since we only focus on solving the cut generation problems, different from the existing studies we also explain how we obtain the realizations of the random vector $\mathbf{X}$. In accordance with Hu et al. (2011) and Noyan and Rudolf (2013), the risk constraints associated with the vertices of the scalarization polytope $C$ are initially added to the intermediate relaxed problem, which is solved at each iteration of the main cut generation-based algorithm. In the base case, the polytope $C$ is a three-dimensional simplex with the vertices $\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_4$, where the $i$th element of $\hat{\mathbf{c}}_i$ is equal to 0.5, and other elements are 0.5/3. We solve the master problem once, and use its optimal solution to calculate the realizations of the associated 4-dimensional random vector $\mathbf{X}$. Note that it is clear how to obtain the realizations of the random vector $\mathbf{Y}$, since the benchmark allocations are given.

**5.1.2 Randomly generated data** To further analyze the computational performance of the proposed methods, we consider a different type of problem (inspired by Dentcheva and Wolfhagen, 2013):

$$\max\{f(\mathbf{z}) \ : \ \mathbf{R}\mathbf{z} \succcurlyeq \mathbf{Y}, \quad \mathbf{z} \in \mathbb{R}^{100}_{+}\},$$

where $\mathbf{R} : \Omega \mapsto [0,1]^{d\times 100}$ is a random matrix and the relation $\succcurlyeq$ represents a stochastic multivariate preference relation. In our setup, the relation $\succcurlyeq$ represents $\succcurlyeq^{C}_{\text{CVaR}_\alpha}$ and $\succcurlyeq^{C}_{(2)}$ for the multivariate polyhedral CVaR or SSD relation, respectively. We assume that the benchmark vector $\mathbf{Y}$ takes the form of $\bar{\mathbf{R}}\bar{\mathbf{z}}$, where $\bar{\mathbf{R}}$ is also a $d \times 100$-dimensional random matrix and $\bar{\mathbf{z}} \in \mathbb{R}^{100}_{+}$ is a given benchmark decision. The entries of the matrices $\mathbf{R}$ and $\bar{\mathbf{R}}$ are independently generated from the uniform distribution on the interval $[0,1]$. Since we directly focus on solving the associated cut generation problems, we also randomly generated the decision variables $\mathbf{z}$ and $\bar{\mathbf{z}}$; in particular, they are independently and uniformly generated from the interval $[100, 500]$. This data generation scheme directly provides us with the realizations of two $d$-dimensional random vectors $\mathbf{X} = \mathbf{R}\mathbf{z}$ and $\mathbf{Y} = \bar{\mathbf{R}}\bar{\mathbf{z}}$.

**5.2 Computational performance - cut generation for (G − MCVaR)** First, we study the effectiveness of alternative MIP formulations for (**CutGen_CVaR**). In these experiments, we assume that each scenario is equally likely, and consider confidence levels of the form $\alpha = k/n$. For an arbitrary confidence level $\bar{\alpha}$, we calculate $k$ as $\lceil \bar{\alpha} * n \rceil$. In Table 1, we present our experiments on the HSBA data described in Section 5.1.1. We compare the performances of four alternative formulations: (i) the MIP model – (**MIP_Special**) – developed for the special case of equal probabilities (Noyan and Rudolf, 2013), (ii) the MIP model – (**MIP_CVaR**) – for general probabilities presented in Noyan and Rudolf (2013), (iii) the more compact model – (**SMIP_CVaR**) – proposed in Section 3.2.1, and (iv) the new model – (**NewMIP_CVaR**) – proposed in Section 3.2.2. We report the results averaged over two instances (based on Government and RAND benchmarks) for each combination of $\alpha$ and $n$. We see that the new formulation using the VaR representation is highly effective in reducing the solution time for these instances. Problem instances that are not solvable within the time limit of 5400 seconds with the existing formulation (**MIP_CVaR**) and its enhancement (**SMIP_CVaR**), is now solvable in six minutes for all instances but one ($n = 1000, \alpha = 0.05$), which is also solved well within the time limit. We observe that (**MIP_CVaR**) terminates at the root node for large instances with no integer feasible solution available. This may be due to the large size of the formulation (quadratic number of binary variables). In contrast, (**NewMIP_CVaR**) contains a linear number of binary variables. What is also surprising is that even the formulation (**MIP_Special**), which uses more information due to the equal probability assumption, is not able to solve many of the instances. For this data set, (**MIP_Special**) has inferior performance when compared to (**SMIP_CVaR**) for problems with 300 or more scenarios. Table 2 compares these formulations on the random data set described in Section 5.1.2. For these instances (**MIP_Special**) performs better than (**MIP_CVaR**) and (**SMIP_CVaR**). However, it still cannot solve larger instances with 500 or more scenarios. In contrast, (**NewMIP_CVaR**) solves these problems within a few minutes. We would also like to note that the total time spent on preprocessing for (**NewMIP_CVaR**) (calculation of $L, U, M_{ik}, M_{i\cdot}, M_{\cdot i}, H_k$), which is not included in the times reported, is negligible (18 seconds on average). Therefore, we can conclude that (**NewMIP_CVaR**) is a better formulation than the existing formulations (**MIP_Special**), (**MIP_CVaR**) and its enhancement (**SMIP_CVaR**).

We would like to remind the reader that during a cut generation-based algorithm, the solution procedure of the cut generation problem is allowed to terminate early without finding the most violated cut. However, when such a heuristic procedure cannot find a violated cut, it is still required to prove that the optimal objective function value is non-negative. Therefore, in our experiments we opt for solving the cut generation problem to optimality.

| | (**MIP_Special**) Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes | (**MIP_CVaR**) Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes | (**SMIP_CVaR**) Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes | (**NewMIP_CVaR**) Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes |
|---|---|---|---|---|---|---|---|---|
| $n$ | $\alpha = 0.01$ | | | | | | | |
| 100 | 0.24 | 98 | 6.91 | 0 | 1.58 | 0 | 0.15 | 0 |
| 200 | 6.75 | 22,507 | 78.17 | 41 | 40.61 | 0 | 0.97 | 419 |
| 300 | 752.39 | 1,510,614 | 746.46 | 494 | 171.85 | 39 | 5.20 | 2556 |
| 500 | †[-,632.72] | 5,980,878 | †[15.82,23.13] | 1 | 1232.56 | 473 | 39.96 | 14,806 |
| 1000 | †[-,163.56] | 2,351,513 | ⋄[*,100] | 0 | ⋄[105.45,▲] | 3 | 325.71 | 48,326 |
| $n$ | $\alpha = 0.05$ | | | | | | | |
| 100 | 292.76 | 1,301,663 | 12.06 | 55 | 3.35 | 0 | 0.94 | 1313 |
| 200 | †[▲,164.10] | 10,755,872 | 437.79 | 6121 | 76.70 | 2668 | 5.17 | 4154 |
| 300 | †[-,160.43] | 6,408,832 | 259.65 | 266 | 237.71 | 1607 | 46.50 | 42,419 |
| 500 | †[-,135.36] | 2,592,061 | †⋄[233.48,555.65] | 0 | 2727.42 | 1306 | 189.17 | 92,627 |
| 1000 | †[-,126.05] | 1,915,464 | ⋄[*,100] | 0 | ⋄[*,593.52] | 0 | 2034.32 | 749,132 |

Table 1: Computational performance of the alternative MIPs for (**CutGen_CVaR**) - Base polytope: HSBA instances

$\text{ROG}_1$ and $\text{ROG}_2$ values (%) are respectively reported in [ ] and the values above **1000%** are indicated with ▲.

†: Time limit with integer feasible solution.

⋄: Time limit with no integer feasible solution.

-: $|\,\text{UB}\,| = 0$ and *: `CPLEX` yields a trivial LB of $-\infty$.

Next we study the effectiveness of various classes of valid inequalities and preprocessing strategies described in Sections 3.2.2 and 3.2.3. We consider two sets of data as before, one with HSBA data (Table 3), and one with the randomly generated data (Table 4). In Tables 3 and 4, the relative improvements and optimality gaps are given as percentages and all presented results are averaged over the two instances with different benchmarks. In the first two columns of Table 3, we compare the performance of (**RSMIP_CVaR**), which is the original formulation enhanced with variable reduction due to symmetry, variable fixing and bounding, against the new formulation (**NewMIP_CVaR**) without any enhancements. In the third column of Table 3, we report the performance of (**NewMIP_CVaR**) with variable fixing and bounding. Finally, in the fourth column, we report the performance of (**NewMIP_CVaR**) with variable fixing, bounding and ordering inequalities (52). Comparing the first two columns of Table 3, we see that fixing and bounding the variables are highly effective strategies, and as a result (**RSMIP_CVaR**) outperforms (**NewMIP_CVaR**). However, it cannot solve the larger instances within the time limit, and in general stops with a large relative optimality gap. On the other hand, when these strategies are also applied to (**NewMIP_CVaR**), all test instances are solved within the time limit, as observed from the third column. The reduction in solution time comparing columns 2 and 3 can be attributed to the large reduction in the binary variables due to variable fixing, fewer than 7% and 17% of the binary variables remain in the formulation for instances with $\alpha = 0.01$ and $\alpha = 0.05$, respectively. The reduction in binary variables is primarily due to fixing of the variables $\beta_k = 0$ for $k \notin \bar{K}$, because $\sum_{i \in L_k} p_i \geq \alpha$, and $\beta_s = 0$ for all $s \in H_k$. We did not observe any additional fixing based on the bounds on VaR in our experiments. Finally, we see that ordering inequalities are highly effective and have the best

| | (MIP_Special) | | (MIP_CVaR) | | (SMIP_CVaR) | | (NewMIP_CVaR) | |
|---|---|---|---|---|---|---|---|---|
| | Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes | Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes | Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes | Elap. Time; $[\text{ROG}_1,\text{ROG}_2]$ | B&B Nodes |
| $n$ | $\alpha = 0.01$ | | | | | | | |
| 100 | 0.20 | 199 | 1220.98 | 43,131 | 362.18 | 26,023 | 0.31 | 0 |
| 200 | 15.00 | 40,314 | $^\dagger$[635.19,103.77] | 11,629 | 3913.65 | 46,358 | 5.90 | 3008 |
| 300 | 3892.79 | 8,388,555 | $^\dagger$[▲,101.25] | 575 | $^\dagger$[▲,101.21] | 2390 | 26.17 | 10,331 |
| 500 | $^\dagger$[▲,102.23] | 10,505,307 | $^\diamond$[134.49,201.36] | 0 | $^\diamond$[104.30,208.65] | 0 | 165.27 | 60,581 |
| $n$ | $\alpha = 0.05$ | | | | | | | |
| 100 | 0.23 | 199 | 2659.79 | 72,200 | 799.98 | 44,953 | 2.44 | 1361 |
| 200 | 18.13 | 40,493 | $^\dagger$[▲,120.17] | 15,419 | $^\dagger$[254.57,41.79] | 33,675 | 8.72 | 6446 |
| 300 | 3822.56 | 8,235,087 | $^\dagger$[▲,103.01] | 1703 | $^\dagger$[▲,102.04] | 7574 | 51.50 | 21,668 |
| 500 | $^\dagger$[▲,102.15] | 9,960,451 | $^{\dagger\diamond}$[▲,83.88] | 0 | $^{\dagger\diamond}$[▲,80.15] | 0 | 221.03 | 58,687 |

Table 2: Computational performance of the alternative MIPs for (**CutGen_CVaR**) - Unit simplex: Random instances

The ROG values above **3500%** are indicated with ▲.

performance, when used in addition to fixing and bounding, compared to the other settings that do not use these inequalities. Because a large number of variables are fixed in these instances, we did not see much benefit of inequalities (65)-(66). We note that this behavior is highly data-dependent as we see in Table 4. In this table, we compare different settings in the first five columns: (i) (**NewMIP_CVaR**) without any enhancements, (ii) (**NewMIP_CVaR**) with fixing and bounding, (iii) (**NewMIP_CVaR**) with fixing, bounding, and ordering inequalities (52), (iv) (**NewMIP_CVaR**) with fixing, bounding, and inequalities (65)-(66), and finally (v) (**NewMIP_CVaR**) with fixing, bounding, and all classes of cuts ((52) and (65)-(66)). For these instances, while a significant number of binary variables can be fixed, the percentage of remaining variables is higher than that for the HSBA data. In this case, the setting with all enhancements and valid inequalities yield the best performance in most cases, with close to 50% reduction in solution time for several instances. Overall, all instances are solved within the time limit with much fewer branch&bound nodes explored.

**5.3 Computational performance - cut generation for (G − MSSD)** In Table 5, we report our computational experiments with the randomly generated data described in Section 5.1.2 to illustrate the effectiveness of the strategies proposed for multivariate SSD-constrained optimization problems. Recall that the cut generation problems decompose by benchmark realizations for SSD. In these experiments, we solve the cut generation problem for $\lceil m/20 \rceil$ of the benchmark realizations. Because we solve multiple cut generation problems for each setting, we let $n \in \{200, 300, 500\}$. For each setting, we generate two instances and report their average statistics. We report the minimum, maximum and average statistics taken over all tested benchmark realizations for a given setting. We compare the performance of the formulation ($\mathbf{MIP\_SSD}_l$), ($\mathbf{MIP\_SSD}_l$) with variable fixing, and ($\mathbf{MIP\_SSD}_l$) with variable fixing and ordering inequalities. In the first column, we report the elapsed time statistics (in seconds) for ($\mathbf{MIP\_SSD}_l$) without any computational enhancements. We observe a high variability in the solution times; the minimum solution times are in a few seconds, whereas the maximum solution times are at the time limit of 5400 seconds. We also report the number of instances that were not solved within the time limit. The next column shows the effect of variable fixing in the reduction of the solution time (as percentages). We observe that while there is a significant (35-65%)

| | (**RSMIP_CVaR**) Fix&Bound | | (**NewMIP_CVaR**) | | (**NewMIP_CVaR**) Fix&Bound (F&B) | | (**NewMIP_CVaR**) F&B&Order. Ineq. | | Remaining Binary Var. (%) (Fixing) | | # of Ordering Inequalities for Remaining Sce. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | NewMIP | RSMIP | |
| $n$ | Equal Probability Case & $\alpha = 0.01$ | | | | | | | | | | |
| 500 | 5.10 | 0.2 | 39.96 | 14.8 | 0.71 | 0.8 | 0.46 | 0.3 | 6.40 | 6.19 | 46.5 |
| 1000 | 14.66 | 0.1 | 325.71 | 48.3 | 3.21 | 4.2 | 1.48 | 0.9 | 4.45 | 4.35 | 164 |
| 2000 | 105.34 | 0 | 2951.841† [-,50] | 308.8 | 48.17 | 44.5 | 19.31 | 10.8 | 4.10 | 4.01 | 710 |
| 3000 | 451.96 | 1.4 | 3371.93† [▲,58.54] | 202.9 | 194.90 | 172.6 | 71.29 | 30.7 | 4.17 | 4.08 | 1787 |
| 5000 | † [124.56,▲] | 0 | † [▲,▲] | 230.2 | 1780.15 | 793.5 | 404.56 | 167.9 | 4.12 | 4.03 | 4903 |
| $n$ | Equal Probability Case & $\alpha = 0.05$ | | | | | | | | | | |
| 500 | 43.18 | 0.1 | 189.17 | 92.6 | 24.77 | 37.7 | 8.33 | 7.2 | 16.20 | 14.87 | 818 |
| 1000 | 440.33 | 0.5 | 2034.32 | 749.1 | 202.80 | 338.7 | 63.48 | 52.3 | 15.00 | 13.87 | 2959 |
| 2000 | † [65.14,246.84] | 1.3 | † [-,▲] | 676.4 | † [-,50] | 3333.1 | 1023.72 | 403.3 | 14.85 | 13.74 | 12656.5 |
| $n$ | General Probability Case & $\alpha = 0.01$ | | | | | | | | | | |
| 500 | 3.98 | 0.0 | 62.25 | 17.4 | 0.74 | 0.9 | 0.59 | 0.5 | 6.20 | 6.00 | 40.5 |
| 1000 | 15.57 | 0.0 | 353.57 | 46.3 | 2.27 | 3.1 | 1.49 | 1.0 | 4.50 | 4.40 | 171.5 |
| 2000 | 191.09 | 0.6 | 3513.47 | 155.98 | 49.67 | 33.9 | 17.51 | 10.6 | 5.10 | 4.97 | 1001.5 |
| 3000 | 3619.97 | 2.0 | [-,50] | 172.5 | 208.57 | 122.4 | 60.15 | 21.1 | 4.98 | 4.86 | 2474.5 |
| 5000 | † [4.90,5.44] | 1.9 | [-,▲] | 112.8 | 1000.86 | 299.1 | 352.77 | 94.4 | 4.03 | 3.95 | 4279 |

Table 3: Computational performance of the enhanced MIPs for (**CutGen_CVaR**) - Base polytope: HSBA instances

$ROG_1$ and $ROG_2$ values (%) are respectively reported in [ ] and the values above **300%** are indicated with ▲.

†: Time limit with integer feasible solution.

-: $|UB| = 0$.

B&B Nodes are reported in thousands.

| | (**NewMIP_CVaR**) Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | (**NewMIP_CVaR**) F&B Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | (**NewMIP_CVaR**) F&B&Ord. Ineq. Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | (**NewMIP_CVaR**) F&B&Ineq. (65)-(66) Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | (**NewMIP_CVaR**) F&B&All Cuts Elap. Time; [$ROG_1$,$ROG_2$] | B&B Nodes | Remaining Binary Var. (%) (Fixing) | # of Inequalities Order. | # of Inequalities Ineqs. (65) | # of Inequalities Ineqs. (66) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Equal Probability Case & $d = 4$ | | | | | | | | | | | | | |
| 100 | 0.27 | 0.0 | 0.13 | 0.0 | 0.13 | 0.0 | 0.09 | 0.0 | 0.09 | 0.0 | 12.0 | 0 | 48 | 33 |
| 300 | 21.89 | 22.8 | 1.15 | 2.7 | 1.10 | 1.6 | 1.19 | 2.0 | 1.34 | 1.4 | 20.3 | 38 | 244 | 218 |
| 500 | 125.63 | 81.6 | 6.96 | 8.6 | 2.71 | 3.7 | 9.79 | 12.4 | 3.64 | 3.5 | 17.4 | 146 | 348 | 334 |
| 1000 | 1487.39 | 313.9 | 65.35 | 58.2 | 43.74 | 28.8 | 58.92 | 62.6 | 37.58 | 19.5 | 14.9 | 547 | 594 | 585 |
| 2000 | †[▲,96.89] | 805.5 | 1883.87 | 745.2 | 1067.50 | 491.2 | 1773.72 | 806.0 | 986.89 | 367.0 | 15.2 | 2288 | 1218 | 1218 |
| 2500 | †[▲,98.21] | 328.9 | †[▲,82.62] | 1982.3 | 4316.92 | 1382.0 | †[38.46,21.74] | 1777.3 | 3011.77 | 867.8 | 17.2 | 4238 | 1716 | 1716 |
| $n$ | Equal Probability Case & $d = 6$ | | | | | | | | | | | | | |
| 100 | 0.37 | 0.0 | 0.18 | 0.0 | 0.19 | 0.0 | 0.14 | 0.0 | 0.14 | 0.0 | 24.5 | 0 | 147 | 118 |
| 300 | 79.21 | 62.5 | 9.14 | 10.7 | 5.62 | 5.6 | 10.43 | 7.8 | 7.34 | 4.7 | 30.0 | 52 | 540 | 490 |
| 500 | 716.28 | 345.2 | 59.79 | 74.8 | 55.70 | 50.5 | 99.20 | 100.3 | 76.32 | 52.0 | 31.1 | 191 | 933 | 894 |
| 1000 | †[▲,90.21] | 948.2 | 3025.31 | 1351.1 | 2369.93 | 856.2 | 3128.48 | 1709.0 | 1621.50 | 522.4 | 30.5 | 954 | 1827 | 1809 |
| $n$ | Equal Probability Case& $d = 8$ | | | | | | | | | | | | | |
| 100 | 0.82 | 0.1 | 0.36 | 0.0 | 0.32 | 0.0 | 0.32 | 0.0 | 0.36 | 0.0 | 38.0 | 0 | 304 | 221 |
| 300 | 192.35 | 133.6 | 37.56 | 48.1 | 31.75 | 26.0 | 25.11 | 11.9 | 26.96 | 11.0 | 49.0 | 72 | 1176 | 1021 |
| 500 | 3735.36 | 701.3 | 438.07 | 281.7 | 384.47 | 165.1 | 468.69 | 257.8 | 330.75 | 133.2 | 43.0 | 255 | 1720 | 1612 |
| $n$ | General Probability Case & $d = 4$ | | | | | | | | | | | | | |
| 100 | 1.38 | 1.0 | 0.16 | 0.0 | 0.19 | 0.0 | 0.15 | 0.1 | 0.18 | 0.1 | 20.0 | 9 | 80 | 63 |
| 300 | 37.00 | 37.6 | 3.38 | 7.2 | 1.80 | 3.6 | 3.63 | 6.0 | 2.90 | 3.5 | 23.0 | 67 | 276 | 238 |
| 500 | 174.33 | 93.0 | 12.96 | 21.6 | 9.16 | 12.0 | 15.69 | 24.1 | 9.64 | 8.2 | 19.4 | 210 | 388 | 388 |
| 1000 | 1273.82 | 319.4 | 47.88 | 64.0 | 34.57 | 27.1 | 51.85 | 44.7 | 37.45 | 17.5 | 15.7 | 646 | 626 | 616 |
| 2000 | †[▲,96.23] | 296.8 | 1903.03 | 769.1 | 1284.37 | 457.5 | 2296.41 | 723.3 | 971.45 | 282.4 | 15.7 | 2498 | 1254 | 1254 |

Table 4: Effectiveness of the valid inequalities for (**NewMIP_CVaR**) - Unit simplex: Random instances

$ROG_1$ and $ROG_2$ values (%) are respectively reported in [ ] and the values above **300%** are indicated with ▲.

†: Time limit with integer feasible solution.

B&B Nodes are reported in thousands.

reduction in the maximum solution time for instances that can be solved within the time limit, the number of unsolved instances increases by one (in the last setting) when variable fixing is employed. We would like to note that for almost all the instances that reached the time limit, the best solution found was optimal; however, optimality was not proved. However, a provably optimal solution to the cut generation problem is required for the convergence of the overall delayed cut generation algorithm to the correct optimal solution (in cases when the current solution does not give a violated cut). There is also up to 53% increase in the minimum solution time for (**MIP_SSD**$_l$) with fixing. Overall, we conclude that fixing is effective as it reduces the overall average solution time in most cases (except for the last setting). Note that unlike the CVaR case, not many binary variables can be fixed. On average, over 65% of the binary variables remain in the formulation in most cases. Next, we analyze the performance of ordering inequalities (52), in addition to fixing, reported in the third column. In the last column of Table 5, we report the maximum, minimum and average number of ordering inequalities added to the formulation (**MIP_SSD**$_l$). We recognize that the ordering inequalities are highly effective, as they reduce the average solution time significantly, enabling the solution of all instances within the time limit. We also tested the performance of the formulation (**NewMIP_SSD**$_l$) on these instances, but observed that this formulation does not perform better than the version with ordering inequalities. Hence, we do not report these experiments.

**6. Conclusions** In this paper, we develop alternative mixed-integer programming formulations and solution methods for cut generation problems arising in a class of multicriteria stochastic optimization problems that features benchmarking constraints based on polyhedral conditional value-at-risk. We propose a mixed-integer programming formulation of the cut generation problem that involves a new representation of value-at-risk. We show that this new formulation is highly effective in solving the cut generation problems. In addition, we describe computational enhancements involving variable fixing and bounding. Furthermore, we give a class of valid inequalities, which establish a relative order between scenario-dependent binary variables when possible. Finally, we give the convex hull description of a polytope describing the linearization of a non-convex substructure arising in this cut generation problem. Our computational results illustrate the effectiveness of our proposed models and methods for the CVaR-constrained optimization problems. We also show that the proposed computational enhancements can be adapted to cut generation problems for polyhedral SSD-constrained optimization. In this case, variable fixing and ordering inequalities continue to be effective. We also give the convex hull description of a polytope describing the linearization of a non-convex substructure arising in this cut generation problem for each benchmark realization. However, these inequalities need to be further strengthened to improve their practical performance. One possible area of future research is to study the intersection of these linearization polytopes for two or more different realizations of **X**.

**Appendix A. A MIP Formulation of (CutGen_CVaR) for the Equal Probability Case** In the special case when all the outcomes of **X** are equally likely, Noyan and Rudolf (2013) propose a MIP formulation alternative to (**MIP_CVaR**), which involves fewer binary variables than (**MIP_CVaR**). The underlying idea relies on the following representation of CVaR under equal probabilities:

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

In this case, the quadratic vector $\beta_i\mathbf{c}^T$ presents a challenge. Introducing the decision vector $\boldsymbol{\zeta}_i^T = (\zeta_{i1},\dots,\zeta_{id})$ and the constraints ensuring $\boldsymbol{\zeta}_i = \beta_i\mathbf{c}$ for all $i \in [n]$ leads to the equivalent MIP formulation of

| | (**MIP_SSD**$_l$) Elapsed Time | | | | (**MIP_SSD**$_l$) Fixing Relative Red.(%) in Time | | | | (**MIP_SSD**$_l$) Fixing & Ord. Ineq. Relative Red.(%) in Time wrt Fixing | | | Remaining Binary Variable (%) (Fixing) | | | # of Ordering Inequalities | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Max | Min | Ave | Ave # Inst. | Max | Min | Ave | Ave # Inst. | Max | Min | Ave | Max | Min | Ave | Max | Min | Ave |
| $n$ | Equal Probability Case & $d=4$ | | | | | | | | | | | | | | | | |
| 200 | 37.08 | 1.09 | 14.95 | | 64.74 | -7.93 | 20.97 | | 76.26 | -113.90 | 24.29 | 90.0 | 37.5 | 73.6 | 3814.0 | 433.5 | 2252.3 |
| 300 | 202.10 | 5.47 | 49.49 | | 47.60 | -14.95 | 15.64 | | 77.91 | -41.90 | 29.43 | 90.7 | 36.5 | 68.8 | 8228.5 | 933.0 | 3843.9 |
| 500 | 3557.48 | 3.67 | 1147.18 | †0.5 | 42.54 | -28.02 | 6.27 | †0.5 | 92.29 | -31.94 | 68.44 | 96.1 | 31.4 | 73.4 | 29990.5 | 1070.5 | 13621.1 |
| $n$ | Equal Probability Case & $d=6$ | | | | | | | | | | | | | | | | |
| 200 | 2985.08 | 1.95 | 517.18 | †0.5 | 39.52 | -21.08 | 7.82 | †0.5 | 98.17 | -1.88 | 53.08 | 98.0 | 55.5 | 85.9 | 2607.0 | 585.0 | 1715.9 |
| 300 | †5410.07 | 9.03 | 3788.74 | †8.5 | 35.83 | -9.08 | 4.97 | †8.5 | 97.84 | -0.35 | 79.25 | 99.0 | 47.7 | 84.7 | 6169.0 | 574.5 | 3597.6 |
| $n$ | General Probability Case & $d=4$ | | | | | | | | | | | | | | | | |
| 200 | 30.56 | 0.79 | 12.43 | | 36.97 | -22.57 | 12.38 | | 65.46 | -12.97 | 39.35 | 90.0 | 37.5 | 73.6 | 3814.0 | 433.5 | 2252.3 |
| 300 | 139.35 | 3.85 | 41.52 | | 49.73 | -31.81 | 14.60 | | 71.89 | -94.24 | 10.32 | 90.7 | 36.5 | 68.8 | 8228.5 | 933.0 | 3843.9 |
| 500 | 1979.72 | 0.40 | 460.90 | | 38.76 | -32.31 | 7.27 | | 91.47 | -59.14 | 59.85 | 96.1 | 31.4 | 73.4 | 33955.0 | 1413.0 | 14252.5 |
| $n$ | General Probability Case & $d=6$ | | | | | | | | | | | | | | | | |
| 200 | 1668.88 | 2.11 | 270.59 | | 51.16 | -40.88 | 11.30 | | 92.96 | -32.58 | 34.40 | 48.5 | 14.5 | 39.5 | 2607.0 | 585.0 | 1715.9 |
| 300 | †5410.63 | 3.13 | 2507.08 | †4 | 43.18 | -52.79 | -1.14 | †5 | 98.17 | -24.53 | 80.91 | 99.0 | 47.7 | 84.7 | 6169.0 | 574.5 | 3597.6 |

Table 5: Effectiveness of fixing and ordering inequalities for (**NewMIP_SSD**$_l$) - Unit simplex: Random instances

†: Time limit with integer feasible solution.

(**CutGen_CVaR**) below:

$$
\begin{aligned}
(\textbf{MIP\_Special}) \qquad \min \quad & \frac{1}{k}\sum_{i\in[n]} \boldsymbol{\zeta}_i^T \mathbf{x}_i - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l \\
\text{s.t.} \quad & \sum_{i\in[n]} \beta_i = k, \\
& 0 \le \zeta_{i\ell} \le c_\ell, && \forall\, i \in [n],\ \ell \in [d], \\
& \zeta_{i\ell} \le \tilde{M}_\ell \beta_i, && \forall\, i \in [n],\ \ell \in [d], \\
& -\zeta_{i\ell} + c_\ell \le \tilde{M}_\ell (1-\beta_i), && \forall\, i \in [n],\ \ell \in [d], \\
& \boldsymbol{\beta} \in \{0,1\}^n, \\
& (17)-(18).
\end{aligned}
$$

To keep our exposition simple, we consider confidence levels of the form $\alpha = k/n$ and refer to Noyan and Rudolf (2013) for the extended version of the above MIP formulation with an arbitrary confidence level.

**Acknowledgments** Simge Küçükyavuz is supported, in part, by National Science Foundation Grants 1055668 and 1100383. Nilay Noyan acknowledges the support from Bilim Akademisi - The Science Academy, Turkey, under the BAGEP program.

## References

Armbruster, B. and Luedtke, J. (2014). Models and formulations for multivariate dominance constrained stochastic programs. *IIE Transactions*. DOI: 10.1080/0740817X.2014.889336.

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

Bülbül, K., Küçükyavuz, S., Noyan, N., and Şen, H. (2014). Chance-constrained two-stage mean-risk stochastic programming. Working Paper.

Dentcheva, D., Penev, S., and Ruszczyski, A. (2010). Kusuoka representation of higher order dual risk measures. *Annals of Operations Research*, 181:325–335.

Dentcheva, D. and Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566.

Dentcheva, D. and Ruszczyński, A. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking and Finance*, 30(2):433–451.

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

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.

Fábián, C. I., Mitra, G., Roman, D., and Zverovich, V. (2011). An enhanced model for portfolio choice with SSD criteria: a constructive approach. *Quantitative Finance*, 11(10):1525–1534.

Gollmer, R., Gotzes, U., and Schultz, R. (2011). A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming*, 216(1):179–190.

Hadar, J. and Russell, W. R. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, 59(1):25–34.

Haskell, W., Shanthikumar, J., and Shen, Z. (2013). Optimization with a class of multivariate integral stochastic order constraints. *Annals of Operations Research*, 206(1):147–162.

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.

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

Leshno, M. and Levy, H. (2002). Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. *Management Science*, 48(8):1074–1085.

Lizyayev, A. and Ruszczyński, A. (2011). Tractable almost stochastic dominance. *European Journal of Operational Research*, 218(2):448–455.

Luedtke, J. (2008). New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450.

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

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

Noyan, N., Rudolf, G., and Ruszczyński, A. (2008). Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters*, 34(6):653–659.

Noyan, N. and Ruszczyński, A. (2008). Valid inequalities and restrictions for stochastic programming problems with first order stochastic dominance constraints. *Mathematical Programming, Ser. A*, 114(2):249–275.

Pflug, G. C. and Römisch, W. (2007). *Modelling, managing and measuring risk*. World Scientific publishing, Singapore.

Rockafellar, R. (2007). Coherent approaches to risk in optimization under uncertainty. In *Tutorials in Operations Research*, pages 38–61. INFORMS.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41.

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

Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439.

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.