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

September 15, 2015

Abstract: We consider a class of stochastic optimization problems that features benchmarking preference relations among random vectors representing multiple random performance measures (criteria) of interest. Given a benchmark random performance vector, preference relations are incorporated into the model as constraints, which require the decision-based random vector to be preferred to the benchmark according to a relation based on multivariate conditional value-at-risk (CVaR) or second-order stochastic dominance (SSD). We develop alternative mixed-integer programming formulations and solution methods for cut generation problems arising in optimization under such multivariate risk constraints. The cut generation problems for CVaR- and SSD-based models involve the epigraphs of two distinct piecewise linear concave functions, which we refer to as reverse concave sets. We give the complete linear description of the linearization polytopes of these two non-convex substructures. We present computational results that show the effectiveness of our proposed models and methods.

Keywords: stochastic programming; multivariate risk-aversion; conditional value-at-risk; stochastic dominance; cut generation; convex hull; reverse concave set

---

1. Introduction In many decision making problems, such as those arising in relief network design, homeland security budget allocation, and financial management, there are multiple random performance measures of interest. In such problems, comparing the potential decisions requires specifying preference relations among random vectors, where each dimension of a vector corresponds to a performance measure (or decision criterion). Moreover, it is often crucial to take into account decision makers' risk preferences. Incorporating 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 multivariate risk-averse preference relations based on SSD or CVaR.

The SSD relation has received significant attention due to 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. 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). In such decision-making situations, enforcing a preference relation over a family of scalarization vectors allows the representation of a wider range of views and differing opinions of multiple experts (for motivating discussions see, e.g., Hu and Mehrotra, 2012). Dentcheva and Ruszczyński (2009) consider linear scalarization with all non-negative coefficients (this set can be equivalently truncated to a unit simplex), 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 polyhedral and convex scalarization sets, respectively.

Optimization models with univariate SSD constraints can be formulated as linear programs with a potentially large number of scenario-dependent variables and constraints (see, e.g., Dentcheva and Ruszczyński, 2006; Noyan et al., 2008; Luedtke, 2008). While efficient cut generation methods can be employed to solve such large-scale linear programs (Rudolf and Ruszczyński, 2008; Dentcheva and Ruszczyński, 2010; Fábián et al., 2011), enforcing these constraints for infinitely many scalarization vectors causes additional challenges. For finite probability spaces, Homem-de-Mello and Mehrotra (2009) show that infinitely many risk constraints (associated with polyhedral scalarization sets) reduce to finitely (typically exponentially) many scalar-based risk constraints for the SSD case, naturally leading to a finitely convergent cut generation algorithm. However, such an algorithm is computationally demanding as it requires the iterative solution of non-convex (difference of convex functions) cut generation subproblems. The authors formulate the cut generation problem as a binary mixed-integer program (MIP) by linearizing the piecewise linear shortfall terms, and develop a branch-and-cut algorithm. They also propose concavity and convexity inequalities, and a big-M improvement method within the branch-and-cut tree to strengthen the 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). In another line of work, Dentcheva and Wolfhagen (2015) use methods from difference of convex (DC) programming to perform cut generation for the multivariate SSD-constrained problem. The authors also provide a finite representation of the multivariate SSD relation if the decisions are taken in a finite dimensional space, even if the probability space is not finite.

A few studies (Armbruster and Luedtke, 2015; Haskell et al., 2013) consider the multivariate SSD relation based on multidimensional utility functions instead of using scalarization functions. The resulting models enforce stricter dominance relations (than those based on the scalarization approach) but they can be formulated as linear programs, and hence, are computationally more tractable. On the other hand, the scalarization approach allows us to use univariate SSD constraints, which are less conservative than the multivariate version, and also offers the flexibility to control the degree of conservatism by varying the scalarization sets. However, the scalarization-based multivariate SSD relation can still be overly conservative in practice and leads to infeasible formulations. 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 multivariate SSD relation by a collection of multivariate CVaR constraints at various confidence levels. This is a very natural relaxation due to the well-known fact that the univariate SSD relation is equivalent to a continuum of CVaR inequalities (Dentcheva and Ruszczyński, 2006); we note that a similar idea also led to a cutting plane algorithm for the optimization models with univariate SSD constraints (Dentcheva et al., 2010). Noyan and Rudolf (2013) define the multivariate CVaR constraints based on the polyhedral scalarization sets; as a result, their modeling approach strikes a good balance between tractability and flexibility. They show that, similar to the SSD-constrained counterpart, it is sufficient to consider finitely many scalarization vectors, and propose a finitely convergent cut generation algorithm. The corresponding cut generation problem has the DC programming structure, as in the SSD case, with similar MIP reformulations involving big-M type constraints. In addition, the authors utilize alternative optimization representations of CVaR to develop 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 problems can increasingly become a computational bottleneck as the number of scenarios increases. According to the results presented in Hu et al. (2011) and Noyan and Rudolf (2013), the cut generation generally takes no less than 90% to 95% of the total time spent. The DC functions encountered in the cut generation problems have polyhedral structure that can be exploited to devise enhanced and easy-to-implement models. In line with these discussions, this paper contributes to the literature by providing more effective and easy-to-implement methods to solve the cut generation problems arising in optimization under multivariate polyhedral SSD and CVaR constraints. For SSD-constrained problems, the cut generation problems naturally decompose by scenarios, and the main difficulty is due to the weakness of the MIP formulation involving big-M type constraints. A similar difficulty arises in CVaR-constrained problems. However, in this case, an additional challenge stems from the combinatorial structure required to identify the α-quantile of the decision-based random variables. Therefore, this study is mainly dedicated to developing computationally efficient methods for the multivariate CVaR-constrained models. However, we also describe how our results can be applied in the SSD case. As in the previous studies, we focus on finite probability spaces, and our approaches can naturally be used in a framework based on sample average approximation.

In the next section, we present the general forms of the optimization models featuring the multivariate polyhedral risk preferences as constraints. In Section 3, we study the cut generation problem 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. The cut generation problem involves the epigraph of a piecewise linear concave function, which we refer to as a reverse concave set. We give the complete linear description of this non-convex substructure. In Section 4, we give analogous results for SSD-constrained models. We emphasize that the reverse concave sets featured in CVaR and SSD cut generation problems are fundamental sets that may appear in other problems. In Section 5, we present our computational experiments on two data sets: a previously studied budget allocation problem and a set of randomly generated test instances. Our results show that the proposed methods lead to more effective cut generation-based algorithms to solve the multivariate risk-constrained optimization models. We conclude the paper in Section 6.

2. Optimization with multivariate risk constraints In this section, we present the general forms of the 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 functionals, since higher values indicate less risky random outcomes. 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 $[x]_+ = \max\{x, 0\}$. Throughout this paper, we assume that all random variables are defined on some finite probability spaces, and simplify our exposition accordingly when possible.

We consider a decision making problem where the multiple random performance measures associated with the decision vector $\mathbf{z}$ 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. Considering the interpretation of the scalarization vectors and the fact that larger outcomes are preferred, we naturally assume that $C \subseteq \{\mathbf{c} \in \mathbb{R}^d_+ \ : \ \sum_{i\in[d]} c_i =$

1}. Given the benchmark (reference) random outcome vector $\mathbf{Y}$ and the confidence level $\alpha \in (0,1]$, the optimization problems involving the multivariate polyhedral CVaR and SSD constraints take, respectively, the following forms:

$$(\mathbf{G-MCVaR}) \qquad \begin{aligned} \max \quad & f(\mathbf{z}) \\ \text{s.t.} \quad & \mathrm{CVaR}_\alpha(\mathbf{c}^\top G(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \qquad \forall\, \mathbf{c} \in C, \\ & \mathbf{z} \in Z. \end{aligned} \tag{1}$$

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

where $X \succcurlyeq_{(2)} Y$ denotes that the univariate random variable X dominates Y in the second order. While $\mathbf{Y}$ is allowed to be defined on a probability space different from Ω, 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.

According to the results on finite representations of the scalarization polyhedra, it is sufficient to consider finitely many scalarization vectors in (1) and (2). However, these vectors correspond to the vertices of some higher dimensional 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 (1) or (2)) is violated. The corresponding cut generation problem 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 discuss the existing mathematical programming formulations of this cut generation problem, 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 (1) is violated or to show that such a vector does not exist. In this regard, we solve the cut generation problem at confidence level $\alpha \in (0,1]$ of the 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 $(\mathbf{CutGen\_CVaR})$ involves the minimization of the difference of two concave functions, because $\mathrm{CVaR}_\alpha(X)$, given by (Rockafellar and Uryasev, 2000)

$$\mathrm{CVaR}_\alpha(X) = \max \left\{ \eta - \frac{1}{\alpha} \mathbb{E}\left([\eta - X]_+\right) \; : \; \eta \in \mathbb{R} \right\}, \tag{3}$$

is a concave function of a scalar-based random variable X. It is well known that the maximum in definition (3) is attained at the α-quantile, also known as the *value-at-risk* at confidence level α denoted by $\mathrm{VaR}_\alpha(X)$. If the optimal objective value of (**CutGen_CVaR**) is non-negative, it follows that $\mathbf{z}^*$ is an optimal solution of $(\mathbf{G}-\mathbf{MCVaR})$. Otherwise, there exists an optimal solution $\mathbf{c}^* \in C$ for which the corresponding constraint $\mathrm{CVaR}_\alpha(\mathbf{c}^{*\top}\mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^{*\top}\mathbf{Y})$ is 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.

3.1 Existing mathematical programming formulations In this section, we present one of the existing mathematical programming formulations of (**CutGen_CVaR**). The second nonlinear term $(-\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{Y}))$ in (**CutGen_CVaR**) can be expressed with linear inequalities and continuous variables because it involves the maximization of a piecewise linear concave function (see (3)). What makes it difficult to solve (**CutGen_CVaR**) is the minimization of the first concave term $(\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{X}))$. 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.

Note that for finite probability spaces $\mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X}) = \mathbf{c}^\top\mathbf{x}_k$ for at least one $k \in [n]$, implying

$$\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{X}) = \mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X}) - \frac{1}{\alpha}\sum_{i\in[n]} p_i[\mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X}) - \mathbf{c}^\top\mathbf{x}_i]_+ \tag{4}$$

$$= \max_{k\in[n]}\left\{\mathbf{c}^\top\mathbf{x}_k - \frac{1}{\alpha}\sum_{i\in[n]} p_i[\mathbf{c}^\top\mathbf{x}_k - \mathbf{c}^\top\mathbf{x}_i]_+\right\}. \tag{5}$$

This key observation leads to the following formulation of (**CutGen_CVaR**) (Noyan and Rudolf, 2013):

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

$$\text{s.t.} \quad w_l \geq \eta - \mathbf{c}^\top\mathbf{y}_l, \qquad \forall\, l \in [m], \tag{6}$$

$$\mathbf{c} \in C, \;\; \mathbf{w} \in \mathbb{R}^m_+, \tag{7}$$

$$\mu \geq \mathbf{c}^\top\mathbf{x}_k - \frac{1}{\alpha}\sum_{i\in[n]} p_i v_{ik}, \qquad \forall\, k \in [n], \tag{8}$$

$$v_{ik} - \delta_{ik} = \mathbf{c}^\top\mathbf{x}_k - \mathbf{c}^\top\mathbf{x}_i, \qquad \forall\, i \in [n],\; k \in [n], \tag{9}$$

$$v_{ik} \leq M_{ik}\beta_{ik}, \qquad \forall\, i \in [n],\; k \in [n], \tag{10}$$

$$\delta_{ik} \leq \hat{M}_{ik}(1-\beta_{ik}), \qquad \forall\, i \in [n],\; k \in [n], \tag{11}$$

$$\beta_{ik} \in \{0,1\}, \qquad \forall\, i \in [n],\; k \in [n], \tag{12}$$

$$\mathbf{v} \in \mathbb{R}^{n\times n}_+, \quad \boldsymbol{\delta} \in \mathbb{R}^{n\times n}_+. \tag{13}$$

Here, the continuous variables η and $\mathbf{w}$ together with the linear inequalities (6) are used to express $\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{Y})$ according to (3). On the other hand, μ represents $\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{X})$ according to the relation

(5), which can be incorporated into the model using the following non-convex constraint

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

This non-convex constraint corresponds to the epigraph of a piecewise linear concave function, and the variables v_{ik} and δ_{ik} are introduced to linearize the shortfall terms $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$. In addition, M_{ik} and $\hat{M}_{ik}$ are sufficiently large constants (big-M coefficients) to make constraints (10) and (11) redundant whenever the right-hand side is positive. Due to constraints (10)-(13) only one of the variables v_{ik} and δ_{ik} is positive. Then, constraint (9) 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*

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

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 the homeland security problem in our computational study* $d = 4$*. Suppose that the vertices of the polytope* C *are known and given as* $\{\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_N\}$*. Then,* $M_{ik} = \max\{\max_{j \in [N]} \hat{\mathbf{c}}_j^\top (\mathbf{x}_k - \mathbf{x}_i), 0\}$.

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 the existing paper for the complete formulation of this special MIP, which is referred to as (**MIP_Special**) in our study. 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 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**).

Bounds on $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Suppose that we have a lower bound L_μ and an upper bound U_μ for $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Then, (**MIP_CVaR**) can be strengthened using the following valid inequalities:

$$L_\mu \leq \mu \leq U_\mu. \tag{14}$$

For example, consider two discrete random variables $X_{\min}$ and $X_{\max}$ with realizations $\min_{\mathbf{c} \in C}\{\mathbf{c}^\top \mathbf{x}_i\}$, $i \in [n]$, and $\max_{\mathbf{c} \in C}\{\mathbf{c}^\top \mathbf{x}_i\}$, $i \in [n]$, respectively. The random variable $X_{\min}$ is no larger than $\mathbf{c}^\top \mathbf{X}$ with probability one for any $\mathbf{c} \in C$. Similarly, $X_{\max}$ is no smaller than $\mathbf{c}^\top \mathbf{X}$ with probability one for any $\mathbf{c} \in C$. Therefore, we can set L_μ and U_μ as $\mathrm{CVaR}_\alpha(X_{\min})$ and $\mathrm{CVaR}_\alpha(X_{\max})$, respectively. Note that the calculation of the realizations of $X_{\min}$ and $X_{\max}$ requires solving n small (d-dimensional) LPs.

Variable reduction using symmetry. We observe the symmetric relation between the $\boldsymbol{\delta}$ and $\mathbf{v}$ variables ($\delta_{ik} = v_{ki}$ for all pairs of $i \in [n]$ and $k \in [n]$), and substitute v_{ki} for δ_{ik} to obtain a more compact formulation. In this regard, we only need to define β_{ik} for $i, k \in [n]$ such that $i < k$, and write constraints (9)-(11) for $i, k \in [n]: \ i < k$. Furthermore, we substitute M_{ki} for $\hat{M}_{ik}$, and let $v_{kk} = 0$ in (8). We refer to the resulting simplified MIP as (**SMIP_CVaR**); the number of binary variables and constraints (9)-(11) associated with the shortfall terms is reduced by half. Furthermore, the linearization polytope defined by (9)-(13) can be strengthened using valid inequalities. In Section 4.2, we study the linearization polytope corresponding to $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$ for a given pair $i, k \in [n]$. This substructure also arises in the cut generation problems with multivariate SSD constraints.

Preprocessing. Let K be a set of scenarios for which $\mathbf{c}^\top \mathbf{x}_k$ cannot be equal to $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ for any $\mathbf{c} \in C$. Preprocessing methods can be used to identify the set K, which would allow us to enforce constraint (8) for a reduced set of scenarios $k \in \bar{K} := [n] \setminus K$. This would also result in reduced number of variables and constraints (9)-(13) 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 $i \in \bar{K}, k \in K$. In addition, we define variables β_{ik} and constraints (9)-(11) for $i, k \in \bar{K}, i < k$ and for $k \in \bar{K}, i \in K$ (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 (14), as (**RSMIP_CVaR**).

Next, we elaborate on how to identify K that yields a reduced set of scenarios $\bar{K}$. Recall that we focus on the left tail of the probability distributions; for example, under equal probabilities, $\mathrm{VaR}_{b/n}(\mathbf{c}^\top \mathbf{X})$ is the bth 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 $\mathrm{VaR}_{b/n}(\mathbf{c}^\top \mathbf{X})$. In line with these discussions, we use the next proposition to identify the set $\bar{K} = [n] \setminus K$.

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 $k \in K$. Moreover, $i \in K$ for all $i \in H_k$.*

Proof. Note that for any $k \in [n]$ and $i \in L_k$, $M_{ki} = 0$ implies that $\mathbf{c}^\top \mathbf{x}_i \leq \mathbf{c}^\top \mathbf{x}_k$ for all $\mathbf{c} \in C$. Thus, the first claim immediately follows from the following VaR definition: Let $\mathbf{c}^\top \mathbf{x}_{(1)} \leq \mathbf{c}^\top \mathbf{x}_{(2)} \leq \cdots \leq \mathbf{c}^\top \mathbf{x}_{(n)}$ denote an ordering of the realizations of $\mathbf{c}^\top \mathbf{X}$ for a given $\mathbf{c}$. Then, for a given confidence level $\alpha \in (0, 1]$,

$$\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \mathbf{c}^\top \mathbf{x}_{(k)}, \ \text{where } k = \min \left\{ j \in [n] : \sum_{i \in [j]} p_{(i)} \geq \alpha \right\}. \tag{15}$$

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 $X_{\min}$ and $X_{\max}$: $L = \mathrm{VaR}_\alpha(X_{\min})$ and $U = \mathrm{VaR}_\alpha(X_{\max})$.

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

3.2.2 An alternative model 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 computationally more efficient 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. Before proceeding, we first introduce some big-M coefficients. Throughout the paper, we use the notation, M, to emphasize that the associated parameter is used in a big-M type variable upper bounding (VUB) constraint (see, e.g., M_{ik} defined in Remark 3.1 as the maximum possible value of $v_{ik} = [\mathbf{c}^\top(\mathbf{x}_k - \mathbf{x}_i)]_+$ over all $\mathbf{c} \in C$, used in the VUB constraint (10)). Let $M_{i*} = \max_{k\in[n]} M_{ik}$, be the maximum possible value of $[\mathbf{c}^\top(\mathbf{x}_k - \mathbf{x}_i)]_+$ taken over all $k \in [n]$ for a given $i \in [n]$. Similarly, let $M_{*i} = \max_{k\in[n]} M_{ki}$ for $i \in [n]$. Finally, let $\tilde{M}_\ell = \max\{c_\ell : \ \mathbf{c} \in C\}$ for $\ell \in [d]$ be the maximum possible value of c_ℓ (note that $\tilde{M}_\ell \leq 1$ because C is a subset of the unit simplex).

Theorem 3.1 *Suppose that $\mathbf{X}$ is a random vector with realizations $\mathbf{x}_1, \dots, \mathbf{x}_n$ and corresponding probabilities p_i, $i \in [n]$. For a given confidence level α 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 $(z, \boldsymbol{\beta}, \boldsymbol{\zeta}, \mathbf{u})$ satisfying the following system:*

$$z \leq \mathbf{c}^\top\mathbf{x}_i + \beta_i M_{i*}, \quad i \in [n], \tag{16}$$

$$z \geq \mathbf{c}^\top\mathbf{x}_i - (1-\beta_i) M_{*i}, \quad i \in [n], \tag{17}$$

$$\sum_{i\in[n]} p_i\beta_i \geq \alpha, \tag{18}$$

$$\sum_{i\in[n]} p_i\beta_i - \sum_{i\in[n]} p_i u_i \leq \alpha - \epsilon, \tag{19}$$

$$z = \sum_{i\in[n]} \boldsymbol{\zeta}_i^\top \mathbf{x}_i, \tag{20}$$

$$\zeta_{i\ell} \leq \tilde{M}_\ell u_i, \quad i \in [n], \ \ell \in [d], \tag{21}$$

$$\sum_{i\in[n]} \zeta_{i\ell} = c_\ell, \quad \ell \in [d], \tag{22}$$

$$\sum_{i\in[n]} u_i = 1, \tag{23}$$

$$u_i \leq \beta_i, \quad i \in [n], \tag{24}$$

$$\boldsymbol{\beta} \in \{0,1\}^n, \quad \boldsymbol{\zeta} \in \mathbb{R}_+^{n\times d}, \quad \mathbf{u} \in \{0,1\}^n. \tag{25}$$

In constraint (19), ϵ is a very small constant to ensure that the left-hand side is strictly smaller than α.

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

$$k^* = \min\left\{k \in [n] \ : \ \sum_{i\in[k]} p_{\pi(i)} \geq \alpha\right\} \quad \text{and} \quad K^* = \{\pi(1), \dots, \pi(k^*)\}, \tag{26}$$

and using (15) we have $z = \mathbf{c}^\top\mathbf{x}_{\pi(k^*)}$. Then, a feasible solution of (16)-(25) can be obtained as follows:

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

For the reverse implication, let us consider a feasible solution $(z, \boldsymbol{\beta}, \boldsymbol{\zeta}, \mathbf{u})$ of (16)-(25) and let $\bar{K} = \{i \in [n] : \beta_i = 1\}$. To prove our claim, it is sufficient to show that there exists a permutation π 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 (26)).

We first focus on the intermediate set of linear inequalities (16)-(19), (23)-(24), and the quadratic equality

$$z = \sum_{i \in [n]} u_i \mathbf{c}^\top \mathbf{x}_i. \tag{27}$$

By the definition of $\bar{K}$ and inequalities (16)-(17) 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 $i \in [n] \setminus \bar{K}$, (24) ensures that $u_i = 0$ for all $i \in [n] \setminus \bar{K}$. Then, (23) and (24) guarantee that $z = \sum_{i \in \bar{K}} u_i \mathbf{c}^\top \mathbf{x}_i = \mathbf{c}^\top \mathbf{x}_{\bar{k}}$ 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 (18) and (19), $\mathcal{P}(\mathbf{c}^\top \mathbf{X} \leq z) = \sum_{i \in \bar{K}} p_i \geq \alpha$ and $\sum_{i \in \bar{K} \setminus \bar{k}} p_i < \alpha$. It follows that, according to the definition in (15), $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \mathbf{c}^\top \mathbf{x}_{\bar{k}} = z$.

Since $\mathbf{c}$ is a decision vector, equality (27) 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 (16)-(19), (23)-(25), and (27). Therefore, it is easy to show that we can linearize the $u_i c_\ell$ terms by replacing them with the new decision variables $\zeta_{i\ell} \in \mathbb{R}_+$ in (27) to obtain (20), and enforcing the additional constraints (21)-(22). 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**)*:*

$$\min \quad z - \frac{1}{\alpha} \sum_{i \in [n]} p_i v_i - \eta + \frac{1}{\alpha} \sum_{l \in [m]} q_l w_l \tag{28}$$

$$s.t. \quad (6) - (7), (16) - (25),$$

$$v_i - \delta_i = z - \mathbf{c}^\top \mathbf{x}_i, \qquad i \in [n], \tag{29}$$

$$v_i \leq M_{i*} \beta_i, \qquad i \in [n], \tag{30}$$

$$\delta_i \leq M_{*i}(1 - \beta_i), \qquad i \in [n], \tag{31}$$

$$\mathbf{v} \in \mathbb{R}^n_+, \quad \boldsymbol{\delta} \in \mathbb{R}^n_+, \tag{32}$$

$$L \leq z \leq U. \tag{33}$$

Proof. We represent $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$ in (**CutGen_CVaR**) using the classical formulation (3). On the other hand, we express $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ using the formula (4), i.e., $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = z - \frac{1}{\alpha} \sum_{i \in [n]} p_i [z - \mathbf{c}^\top \mathbf{x}_i]_+$, where $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, and ensure the exact calculation of z for any $\mathbf{c} \in C$ by enforcing (16)-(25), from Theorem 3.1. Then, by simple manipulation and linearizing the terms $[z - \mathbf{c}^\top \mathbf{x}_i]_+ =: v_i$ using (29)-(32), 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**)*:*

$$\beta_k \le \beta_i, \quad k \in [n],\ i \in L_k, \tag{34}$$

or equivalently,

$$\beta_i \le \beta_k, \quad k \in [n],\ i \in H_k. \tag{35}$$

Proof. If $i \in L_k$, then $M_{ki} = \max_{\mathbf{c}\in C}[\mathbf{c}^\top(\mathbf{x}_i - \mathbf{x}_k)]_+ = 0$. 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 (34) is valid. The validity proof of inequality (35) follows similarly. □

Introducing inequalities (34) or (35) 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 β_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 $M_{i*} = 0$ or $M_{*i} = 0$, 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 $\max_{\mathbf{c}\in C} \mathbf{c}^\top\mathbf{x}_i < L$ 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 β_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 (4) with $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X})$. Using formula (4) in (**CutGen_CVaR**) leads to a concave minimization. Therefore, we study the linearization of the set (referred to as a *reverse concave set*) corresponding to the epigraph of $-g(z,\mathbf{c})$, given by (29)-(32) in (**NewMIP_CVaR**). We propose valid inequalities that give a complete linear description of this linearization set for a given $i \in [n]$. As a result, these valid inequalities can be used to strengthen the formulation (**NewMIP_CVaR**) (as will be shown in our computational study in Section 5).

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, $x_j \ge 0, j \in [d]$. This implies the nonnegativity of $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X})$, since $\mathbf{c} \ge \mathbf{0}$. 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

$$v - \delta = z - \sum_{j\in[d]} x_j c_j, \tag{36}$$

$$v \le M_v\beta, \tag{37}$$

$$\delta \le M_\delta(1-\beta), \tag{38}$$

$$\sum_{j\in[d]} c_j = 1, \tag{39}$$

$$\mathbf{c}, v, \delta \ge 0, \tag{40}$$

$$\beta \in \{0,1\}, \tag{41}$$

$$0 \leq z \leq \bar{U}. \tag{42}$$

At this time, we let $\bar{U} = \max_{s\in[n],k\in[d]}\{x_{sk}\}$, i.e, the largest component of $\mathbf{x}_s$ over all $s \in [n]$, which is a trivial upper bound on $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Also let $M_v = \bar{U} - \min_{k\in[d]}\{x_k\}$ be the big-M coefficient for the variable $v = [z - \sum_{j\in[d]} x_j c_j]_+$, and $M_\delta = \max_{k\in[d]}\{x_k\}$ be the big-M coefficient for the variable $\delta = [\sum_{j\in[d]} x_j c_j - z]_+$. Let $\mathcal{Q} = \{(\mathbf{c}, v, \delta, \beta, z) : (36) - (42)\}$.

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 kth entry and zeroes elsewhere.

Proposition 3.3 *The extreme points (c, v, δ, β, 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$_k$**:** $(e_k, \bar{U} - x_k, 0, 1, \bar{U})$ *for all* $k \in [d]$ *with* $x_k < \bar{U}$.

Proof. First, note that, from the definitions of $\bar{U}$, M_v, and M_δ, we have $x_k \leq M_\delta \leq \bar{U}$, and $0 \leq \bar{U} - x_k \leq M_v$ for all $k \in [d]$. Hence, points **QEP1**$_k$–**QEP4**$_k$ are feasible and they cannot be expressed as a convex combination of any other feasible 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 $x_k = \bar{U}$ for some $k \in [d]$, then **QEP4**$_k$ is equivalent to **QEP3**$_k$. Therefore, we only define **QEP4**$_k$ for $k \in [d]$ with $x_k < \bar{U}$.

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

Proposition 3.4 *For $k \in [d]$, the inequality*

$$v \leq \sum_{j\in[d]} [x_k - x_j]_+ c_j + (\bar{U} - x_k)\beta \tag{43}$$

is valid for $\mathcal{Q}$. Similarly, for $k \in [d]$, the inequality

$$\delta \leq \sum_{j\in[d]} [x_j - x_k]_+ c_j + x_k(1 - \beta) \tag{44}$$

is valid for $\mathcal{Q}$.

Proof. First, we prove the validity of inequality (43). If $\beta = 0$, then $v = 0$ from (37). Because $\mathbf{c} \geq \mathbf{0}$, inequality (43) holds trivially. If $\beta = 1$, then $\delta = 0$ from (38). Thus, for any $k \in [d]$,

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

where the last inequality follows from (42). Thus, inequality (43) is valid.

Next, we prove the validity of inequality (44). If $\beta = 1$, then $\delta = 0$ from (38). Because $\mathbf{c} \geq \mathbf{0}$, inequality (44) holds trivially. If $\beta = 0$, then $v = 0$ from (38). Thus, for any $k \in [d]$,

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

Hence, inequality (44) is valid. □

Theorem 3.2 *conv($\mathcal{Q}$) is completely described by equalities (36) and (39), and inequalities (40), (43), and (44).*

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 (40), (43), and (44) 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 (40), (43), and (44) 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 (36) 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 (39), the subtracted term is a constant (γ_{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 $k \in [d]$,

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

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 $x_k < \bar{U}$ 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 $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$.

(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 $C(\mathbf{QEP4}_k) = C(\mathbf{QEP3}_k) + (\gamma^z + \gamma^v)(\bar{U} - x_k), k \in [d]$. 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 $\phi_k := \gamma^z \bar{U} + \gamma^\beta + \gamma^v(\bar{U} - x_k)$ 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 $x_1 \leq x_2 \leq \cdots \leq x_d \leq \bar{U}$ 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 (43) 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 (43) 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 $x_{k^*} < \bar{U}$. 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 $x_k = \bar{U}$ for all k with $\gamma_k = 0$. In this case, inequality (43) 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:

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

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 $\bar{U} > x_k$.) So the extreme points $\mathbf{QEP4}_k, k \in [d]$ are never optimal under these cost assumptions. Hence, $v^\kappa = 0$ for all $\kappa \in O(\boldsymbol{\gamma}, \gamma^v, \gamma^\delta, \gamma^\beta, \gamma^z)$.

(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, $\min_{k \in [d]}\{C(\mathbf{QEP4}_k)\} = C(\mathbf{QEP4}_{k^*}) = \gamma^z \bar{U} + \gamma^\beta$. If $\gamma^\delta x_d > \gamma^z \bar{U} + \gamma^\beta$, 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 $\gamma^\delta x_d \leq \gamma^z \bar{U} + \gamma^\beta$. If $\gamma^\delta x_1 < \gamma^z \bar{U} + \gamma^\beta$, 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 $\gamma^\delta x_1 \geq \gamma^z \bar{U} + \gamma^\beta$. Let $\bar{k} := \min\{j \in [d] : \gamma^\delta x_j \leq \gamma^z \bar{U} + \gamma^\beta\}$. If there exists $j \geq \bar{k}$ such that $C(\mathbf{QEP1}_j) = \gamma_j + \gamma^\delta x_j < \gamma^z \bar{U} + \gamma^\beta = C(\mathbf{QEP4}_{k^*}) \leq C(\mathbf{QEP4}_k)$ 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 $\gamma_j + \gamma^\delta x_j = \gamma^z \bar{U} + \gamma^\beta$ 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 (44) 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.

□

Note that in the definition of the set $\mathcal{Q}$, we used weaker bounds on v, δ 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, $U < \bar{U}$. Then, we simply update inequality (43) as

$$v \leq \sum_{j \in [d]} [x_k - x_j]_+ c_j + (U - x_k)\beta. \tag{45}$$

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 $M'_z = \bar{U} - L$ 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 (44) in the original space becomes

$$\delta \leq \sum_{j \in [d]} [x_j - x_k]_+ c_j + (x_k - L)(1 - \beta). \tag{46}$$

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

Finally, from Section 3, we know that $v \leq M_{i*}\beta$ and $\delta \leq M_{*i}(1-\beta)$ for the given scenario $i \in [n]$ for which the linearization polytope is written. Again, in most cases, $M_{i*} \leq M_v$ and $M_{*i} \leq M_\delta$. In this case, we cannot have $c_k = 1$ and $z = L$ for k such that $x_k - L > M_{*i}$, because otherwise $\delta = [\sum_{j \in [d]} c_j x_j - z]_+ = x_k - L > M_{*i}$, which violates the constraint $\delta \leq \mathcal{M}_{*i}(1-\beta)$. Hence for all k with $x_k - L > M_{*i}$, if $c_k > 0$ and $z = L$, then we must have $c_\ell = 1 - c_k$ for some $\ell \in [d]$ with $x_\ell - L < M_{*i}$. Then, $\delta = M_{*i}$ in such an extreme point solution. In this case, we can construct an equivalent polyhedron where we let $x_k^\ell = M_{*i} + L$ for all $k \in [d]$ such that $x_k - L > M_{*i}$ and $\ell \in [d]$ such that $x_\ell - L < M_{*i}$. Similarly, we cannot have $c_k = 1$ and $z = U$ for k such that $U - x_k > M_{i*}$, 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 $U - x_k > M_{i*}$, then we must have $c_\ell = 1 - c_k$ for some $\ell \in [d]$ with $U - x_\ell < M_{i*}$. Then $v = M_{i*}$ in such an extreme point solution. In this case, we can construct an equivalent polyhedron where we let $\bar{x}_k^\ell = U - M_{i*}$ for all $k \in [d]$ such that $U - x_k > M_{i*}$ and $\ell \in [d]$ such that $U - x_\ell < M_{i*}$. 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 (45) for $k \in [d]$ such that $U - x_k \leq M_{i*}$, and inequality (46) for $k \in [d]$ such that $x_k - L \leq M_{*i}$. Note that after this update inequalities (45) for $k \in [d]$ such that $U - x_k = M_{i*}$ reduces to $v \leq M_{i*}\beta$, and inequality (46) for $k \in [d]$ such that $x_k - L = M_{*i}$ reduces to $\delta \leq \mathcal{M}_{*i}(1-\beta)$. 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 $\mathcal{Q}'_i = \{(\mathbf{c}, v_i, \delta_i, \beta_i, z) \in \mathbb{R}_+^{d+4} : (29)-(31), (33), (39), \beta_i \in \{0,1\}\}$. Then conv$(\mathcal{Q}'_i)$ is completely described by adding inequalities*

$$v_i \leq \sum_{j\in[d]} [x_{ik} - x_{ij}]_+ c_j + (U - x_{ik})\beta_i, \quad \forall\, k \in [d] \;:\; U - x_{ik} < M_{i*}, \tag{47}$$

$$\delta_i \leq \sum_{j\in[d]} [x_{ij} - x_{ik}]_+ c_j + (x_{ik} - L)(1-\beta_i), \quad \forall\, k \in [d] : x_{ik} - L < M_{*i} \tag{48}$$

to the original constraints (29)-(31),(33), *and* (39).

In this section and in Section 4.2, we derive valid inequalities and convex hull descriptions using only the condition that C is a unit simplex. However, we note that the unit simplex condition applies, without loss of generality, to all scalarization sets of interest, and therefore, the presented inequalities are valid even if there are additional constraints on the scalarization vectors, i.e., even if C is a strict subset of the unit simplex.

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 $(\mathbf{G-MSSD})$. 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 $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 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 be a polytope of scalarization vectors.

The random vector $\mathbf{X}$ is said to dominate $\mathbf{Y}$ in *polyhedral linear second order* with respect to C if and only if

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

$$\sum_{i\in[n]} p_i [\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i]_+ \leq \sum_{k\in[m]} q_k [\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{y}_k]_+, \quad \forall\, l \in [m],\ \mathbf{c} \in C. \tag{49}$$

As discussed in Section 2, 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 $P_l = \{w_k \geq \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{y}_k,\ k \in [m],\ \mathbf{c} \in C,\ \mathbf{w} \in \mathbb{R}_+^m\}$ for $l \in [m]$. Thus, $(\mathbf{G-MSSD})$ can be reformulated as an optimization problem with exponentially many constraints, and solved using a delayed constraint generation algorithm (Homem-de-Mello and Mehrotra, 2009). The SSD constraints corresponding to a subset of the scalarization vectors are initially present in the formulation. Then given a solution to this intermediate relaxed problem, a cut generation problem is solved to identify whether there is a constraint violated by the current solution.

Due to the structure of the SSD relation (49), 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 lth realization of the benchmark vector $\mathbf{Y}$ is given by

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

4.1 Existing mathematical programming approaches Note that (**CutGen_SSD**) involves a minimization of the difference of convex functions. Dentcheva and Wolfhagen (2015) use methods from DC programming to solve this problem directly. Similar to the case of univariate SSD constraints (Dentcheva and Ruszczyński, 2003), we can easily linearize the first type of shortfalls featured in the objective function:

$$\min \left\{ \sum_{k\in[m]} q_k w_k - \sum_{i\in[n]} p_i[\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i]_+ \ : \ (\mathbf{c},\mathbf{w}) \in P_l \right\}, \tag{50}$$

which results in a concave minimization with potentially many local minima. If the optimal objective function value of (50) is smaller than 0, then there is a scalarization vector for which the SSD condition associated with the lth realization is violated. Note that it is crucial to solve the cut generation problem exactly for the correct execution of the solution method for $(\mathbf{G-MSSD})$. Otherwise, if we obtain a local minimum and the objective is positive, then we might prematurely stop the algorithm.

The methods based on DC programming and concave minimization may not fully utilize the polyhedral nature of the objective and the constraints. In addition, DC methods can only guarantee local optimality. The main challenge in the cut generation problem (50) 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 lth realization of the benchmark vector $\mathbf{Y}$:

$$
\begin{aligned}
(\mathbf{MIP\_SSD}_l) \quad \min \quad & \sum_{k\in[m]} q_k w_k - \sum_{i\in[n]} p_i v_i \\
\text{s.t.} \quad & w_k \geq \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{y}_k, && \forall\, k \in [m], && (51)\\
& \mathbf{w} \in \mathbb{R}^m_+, && && (52)\\
& v_i - \delta_i = \mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i, && \forall\, i \in [n], && (53)\\
& v_i \leq M_i \beta_i, && \forall\, i \in [n], && (54)\\
& \delta_i \leq \hat{M}_i (1-\beta_i), && \forall\, i \in [n], && (55)\\
& \mathbf{c} \in C, \quad \mathbf{v} \in \mathbb{R}^n_+, \quad \boldsymbol{\delta} \in \mathbb{R}^n_+, \quad \boldsymbol{\beta} \in \{0,1\}^n. && && (56)
\end{aligned}
$$

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 $\hat{M}_i = -\min\{\min_{\mathbf{c}\in C}\{\mathbf{c}^\top \mathbf{y}_l - \mathbf{c}^\top \mathbf{x}_i\}, 0\}$. 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.

4.2 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 (53)-(56) 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 (34)–(35) *are also valid for* ($\mathbf{MIP\_SSD}_l$) *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 (29) (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 linear convex shortfall terms.

Linearization of $[\mathbf{a}^\top \mathbf{c}]_+$ *in* (**CutGen_SSD**). For a given vector $\mathbf{a} \in \mathbb{R}^d$, consider the convex function $h(c) = [\mathbf{a}^\top \mathbf{c}]_+ := \max\{0, \mathbf{a}^\top \mathbf{c}\}$ 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 (50), 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 ($\mathbf{MIP\_SSD}_l$). Therefore, we study the linearization of the set (also a *reverse concave set*) corresponding to the epigraph of $-h(\mathbf{c})$. (Note that this structure also appears in the cut generation problem for CVaR (9)–(13), where we let $\mathbf{a} = \mathbf{x}_k - \mathbf{x}_i$, for $i, k \in [n]$.) We propose valid inequalities that give a complete linear description of this linearization set for a given $i \in [n]$. As a result, these valid inequalities can be used to strengthen the formulations involving such linearization terms.

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

$$v - \delta = \sum_{j\in[d]} a_j c_j, \tag{57}$$

$$v \leq \bar{M}_v \beta, \tag{58}$$

$$\delta \leq \bar{M}_\delta (1 - \beta), \tag{59}$$

$$\sum_{j\in[d]} c_j = 1, \tag{60}$$

$$\mathbf{c}, v, \delta \geq 0, \tag{61}$$

$$\beta \in \{0, 1\}, \tag{62}$$

where $\bar{M}_v = a_{d_1}$ is the big-M coefficient associated with the variable $v = [\sum_{j\in[d]} a_j c_j]_+$, and $\bar{M}_\delta = -a_d$ is the big-M coefficient associated with the variable $\delta = [\sum_{j\in[d]} -a_j c_j]_+$.

Let $\mathcal{S} = \{(\mathbf{c}, v, \delta, \beta) : (57) - (62)\}$. First, we characterize the extreme points of conv($\mathcal{S}$). Recall that e_k denotes the d-dimensional unit vector with 1 in the kth 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 and they cannot be expressed as a convex combination of any other feasible 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*

$$v \leq \sum_{j=1}^{d_1} [a_j - a_k]_+ c_j + a_k \beta \tag{63}$$

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

$$\delta \leq \sum_{j=d_1+1}^{d} [a_k - a_j]_+ c_j - a_k (1 - \beta) \tag{64}$$

is valid for $\mathcal{S}$.

Proof. If $\beta = 0$, then $v = 0$ from (58). Because $\mathbf{c} \geq \mathbf{0}$, inequality (63) holds trivially. If $\beta = 1$, then $\delta = 0$ from (59). Thus, for any $k = 1, \ldots, d_1$,

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

where the last inequality follows from (60).

To see the validity of inequality (64), note that equality (57) can be rewritten as $\delta - v = \sum_{j \in [d]} (-a_j) c_j$. Thus, we obtain an equivalent set where v and δ, and D^+ and D^- are interchanged.

□

Remark 4.1 *Inequality (58) is a special case of (63) with* $k = d_1$, *and inequality (59) is a special case of (64) with* $k = d$.

Remark 4.2 *Note that* $\beta \geq 0$ *is implied by inequality (58), and* $\beta \leq 1$ *is implied by (59).*

Remark 4.3 *Consider a related set,* $\mathcal{T}$, *where constraint (60) 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 (63) 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 (57) and (60), and inequalities (61), (63), and (64).*

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 (61), (63), and (64) 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 (61), (63), and (64) 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 (57) 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 (60), 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 (63) 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 γ^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 (63) 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 (63) 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 (63) 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 (57), 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 (64), equality (57) can be rewritten as $\delta - v = \sum_{j\in[d]}(-a_j)c_j$. Thus, we obtain an equivalent set where v and δ, and D^+ and D^- are interchanged. Thus, the proof is complete, using the same arguments as in Case A and inequalities (64).

□

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 $i \in [n]$. Then, an enhanced MIP formulation of (**CutGen_SSD**) for the lth realization of $\mathbf{Y}$ is obtained by replacing (54)-(55) in ($\mathbf{MIP\_SSD}_l$) with the following constraints:

$$v_i \leq \sum_{j\in D_i^+} [a_{ij} - a_{ik}]_+ c_j + a_{ik}\beta_i, \qquad \forall\ i \in [n], k \in D_i^+, \tag{65}$$

$$\delta_i \leq \sum_{j\in D_i^-} [a_{ik} - a_{ij}]_+ c_j - a_{ik}(1-\beta_i), \qquad \forall\ i \in [n], k \in D_i^-. \tag{66}$$

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: $\mathrm{G}_1 = |\,\mathrm{LB}-\mathrm{UB}\,|/(|\,\mathrm{UB}\,|)$ and $\mathrm{G}_2 = |\,\mathrm{LB}-\mathrm{UB}\,|/(|\,\mathrm{LB}\,|)$. It is easy to see that the maximum of G_1 and G_2 is an upper bound on the actual relative optimality gap; we do not report G_1 when $|\,\mathrm{UB}\,| = 0$ or `CPLEX` yields a trivial lower bound of $-\infty$.

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.

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 larger 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 polytope 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 $\frac{\theta}{3} \leq \min\limits_{i \in \{1,\dots,4\}} c_i^*$ 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.

For this set of instances, Noyan and Rudolf (2013) report computational results with the formulation (**MIP_Special**) – developed for the multivariate CVaR-constrained problem under the special case of equal probabilities. For example, for the largest problem instances with 500 scenarios and $\alpha = 0.05$ (resp., $\alpha = 0.01$), on average, two (resp., 1.6) cut generation problems need to be solved taking 14386 (resp., 11507) seconds (around 99.8% of overall solution time). We note that in the initialization step of the algorithm, four risk constraints are additionally generated based on the vertices of C. Similarly, for the multivariate SSD-constrained problems, Hu et al. (2011) report that for the largest test problems with 300 scenarios, only one cut generation problem is solved taking 1,318 seconds (96% of overall solution time). Since the cut generation is the main bottleneck, in our computational study we only focus on solving the cut generation problems. Hence, different from the existing studies, we also explain how we obtain the realizations of the random vector $\mathbf{X}$. In accordance with the existing studies, the risk constraints associated with the vertices of the scalarization polytope C are initially added to the intermediate relaxed problem. In the base case, the polytope C is a three-dimensional simplex with the vertices $\hat{\mathbf{c}}_1, \dots, \hat{\mathbf{c}}_4$, where the ith 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, 2015):

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

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 and 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 $\mathbf{X} = \mathbf{Rz}$ 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 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 α 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 (HSBA data, $n = 1000, \alpha = 0.05$), 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 the HSBA data set, (**MIP_Special**) has inferior performance when compared to (**SMIP_CVaR**) for problems with 300 or more scenarios. On the other hand, for the random data set (described in Section 5.1.2) (**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 the parameters $L, U, M_{ik}, M_{i*}, M_{*i}, H_k$), which is not included in the times reported, is negligible. Therefore, we can conclude that (**NewMIP_CVaR**) is a better formulation than the existing formulations (**MIP_Special**), (**MIP_CVaR**) and its enhancement (**SMIP_CVaR**).

Next we study the effectiveness of various classes of valid inequalities and preprocessing strategies described in Sections 3.2.2 and 3.2.3. Note that when we test the performance of a class of inequalities, we add all inequalities a priori to the formulation, because there are polynomially many of them. We consider two sets of data as before, one with HSBA data (Table 2), and one with the randomly generated data (Table 3). In Tables 2 and 3, 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 2, 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 2, 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 (34). Comparing the first two columns of Table 2, 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

	(MIP_Special) Time; [G$_1$,G$_2$]	B&B Node	(MIP_CVaR) Time; [G$_1$,G$_2$]	B&B Node	(SMIP_CVaR) Time; [G$_1$,G$_2$]	B&B Node	(NewMIP_CVaR) Time; [G$_1$,G$_2$]	B&B Node
n	$\alpha = 0.01$ & Base polytope: HSBA instances							
200	6.8	22,507	78.2	41	40.6	0	1.0	419
300	752.4	1,510,614	746.5	494	171.9	39	5.2	2556
500	†[-,632.7]	5,980,878	†[15.8,23.1]	1	1232.6	473	40.0	14,806
1000	†[-,163.6]	2,351,513	⋄[*,100]	0	⋄[105.5,▲]	3	325.7	48,326
n	$\alpha = 0.05$ & Base polytope: HSBA instances							
200	†[▲,164.1]	10,755,872	437.8	6121	76.7	2668	5.2	4154
300	†[-,160.4]	6,408,832	259.7	266	237.7	1607	46.5	42,419
500	†[-,135.4]	2,592,061	†⋄[233.5,555.7]	0	2727.4	1306	189.2	92,627
1000	†[-,126.1]	1,915,464	⋄[*,100]	0	⋄[*,593.5]	0	2034.3	749,132
n	$\alpha = 0.01$ & Unit simplex: Random instances							
200	15.0	40,314	†[635.2,103.8]	11,629	3913.7	46,358	5.9	3008
300	3892.8	8,388,555	†[▲,101.3]	575	†[▲,101.2]	2390	26.2	10,331
500	†[▲,102.2]	10,505,307	⋄[134.5,201.4]	0	⋄[104.3,208.7]	0	165.3	60,581
n	$\alpha = 0.05$ & Unit simplex: Random instances							
200	18.1	40,493	†[▲,120.2]	15,419	†[254.6,41.8]	33,675	8.7	6446
300	3822.6	8,235,087	†[▲,103.0]	1703	†[▲,102.0]	7574	51.5	21,668
500	†[▲,102.2]	9,960,451	†⋄[▲,83.9]	0	†⋄[▲,80.2]	0	221.0	58,687

Table 1: Computational performance of the alternative MIPs for (**CutGen_CVaR**)

G$_1$ and G$_2$ values (%) are respectively reported in [ ] and the values above **3500%** are indicated with ▲.
†: Time limit with integer feasible solution and ⋄: Time limit with no integer feasible solution.
-: $|\,\text{UB}\,| = 0$ and *: `CPLEX` yields a trivial LB of $-\infty$.

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 a result of the observation in Proposition 3.1. We did not observe any additional fixing based on the bounds on VaR in our experiments. Finally, from the last column we see that ordering inequalities are highly effective and have the best 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 and a relatively large number of ordering relations (34) across scenarios exist in these instances, we did not see much benefit of inequalities (47)-(48). We note that this behavior is highly data-dependent as we see in Table 3. In this table, we compare different settings in the first three columns: (i) (**NewMIP_CVaR**) without any enhancements, (ii) (**NewMIP_CVaR**) with fixing, bounding, and ordering inequalities (34), and (iii) (**NewMIP_CVaR**) with fixing, bounding, and all classes of cuts ((34) and (47)-(48)). We do not report our detailed results for (**NewMIP_CVaR**) with fixing and bounding, because the conclusions are similar to Table 2. 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 yields the best performance in most cases, with close to 50% reduction in solution time for several instances. The inequalities (47)-(48) are useful when added on to the setting with all other improvements, in the most difficult cases. Overall, with this setting, all instances are solved within the time limit with much fewer branch-and-bound (B&B) nodes explored.

	(**RSMIP_CVaR**) Fix&Bound		(**NewMIP_CVaR**)		(**NewMIP_CVaR**) Fix&Bound (F&B)		(**NewMIP_CVaR**) F&B& Order. Ineq.			Remaining Binary Var (%) (Fixing)	
	Time; $[G_1,G_2]$	B&B Node	Time; $[G_1,G_2]$	B&B Node	Time; $[G_1,G_2]$	B&B Node	Time; $[G_1,G_2]$	B&B Node	# Ineq (34)	NewMIP	RSMIP
n	Equal Probability Case & $\alpha = 0.01$										
500	5.1	0.2	40.0	14.8	0.7	0.8	0.5	0.3	46	6.4	6.2
1000	14.7	0.1	325.7	48.3	3.2	4.2	1.5	0.9	164	4.5	4.4
2000	105.3	0	2951.8† [-,50]	308.8	48.2	44.5	19.3	10.8	710	4.1	4.0
3000	452.0	1.4	3371.9† [▲,58]	202.9	194.9	172.6	71.3	30.7	1787	4.2	4.1
5000	†[124.6,▲]	0	†[▲,▲]	230.2	1780.2	793.5	404.6	167.9	4903	4.1	4.0
n	Equal Probability Case & $\alpha = 0.05$										
500	43.2	0.1	189.2	92.6	24.8	37.7	8.3	7.2	818	16.2	14.9
1000	440.3	0.5	2034.3	749.1	202.8	338.7	63.5	52.3	2959	15.0	13.9
2000	†[65.1,246.8]	1.3	†[-,▲]	676.4	†[-,50]	3333.1	1023.7	403.3	12656	14.9	13.7
n	General Probability Case & $\alpha = 0.01$										
500	4.0	0.0	62.3	17.4	0.7	0.9	0.6	0.5	40	6.2	6.0
1000	15.6	0.0	353.6	46.3	2.3	3.1	1.5	1.0	171	4.5	4.4
2000	191.1	0.6	3513.5	156.0	49.7	33.9	17.5	10.6	1001	5.1	5.0
3000	3620.0	2.0	[-,50]	172.5	208.6	122.4	60.2	21.1	2474	5.0	4.9
5000	†[4.9,5.4]	1.9	[-,▲]	112.8	1000.9	299.1	352.8	94.4	4279	4.0	4.0

Table 2: Computational performance of the enhanced MIPs for (**CutGen_CVaR**) - Base polytope: HSBA instances

G_1 and G_2 values (%) are respectively reported in [ ] and the values above **300%** are indicated with ▲.
†: Time limit with integer feasible solution, -: $|\,\mathrm{UB}\,| = 0$, B&B Nodes are reported in thousands.
The ordering inequalities (34) are added for the binary variables that could not be fixed.

5.3 Computational performance - cut generation for (G − MSSD) In Table 4, 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 average and the standard deviation of the solution times taken over all tested benchmark realizations for a given setting. We compare the performance of two formulations: (**MIP_SSD**$_l$) and (**MIP_SSD**$_l$) with variable fixing and ordering inequalities. In the first column, we report the elapsed time statistics (in seconds) for (**MIP_SSD**$_l$) without any computational enhancements. From the standard deviation columns, we observe a high variability in the solution times. In fact, 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 under the column "# Unslvd".

Note that unlike the CVaR case, which benefits from additional information on VaR for fixing variables, in the SSD case not many binary variables can be fixed. On average, over 65% of the binary variables remain in the formulation. Next, we analyze the performance of ordering inequalities (34), in addition to fixing, reported in the second column. In the last column of Table 4, we report the average number of ordering inequalities added to the formulation (**MIP_SSD**$_l$). We recognize that the ordering inequalities are highly

	(**NewMIP_CVaR**)		(**NewMIP_CVaR**) F&B& Order. Ineq.		(**NewMIP_CVaR**) F&B& All Cuts		Remaining Binary	# of Inequalities		
	Time; $[G_1,G_2]$	B&B Node	Time; $[G_1,G_2]$	B&B Node	Time; $[G_1,G_2]$	B&B Node	Var. (%) (Fixing)	Order.	Ineqs. (47)	(48)
n	Equal Probability Case & $d=4$									
500	125.6	81.6	2.7	3.7	3.6	3.5	17.4	146	348	334
1000	1487.4	313.9	43.7	28.8	37.6	19.5	14.9	547	594	585
2000	†[▲,96.9]	805.5	1067.5	491.2	986.9	367.0	15.2	2288	1218	1218
2500	†[▲,98.2]	328.9	4316.9	1382.0	3011.8	867.8	17.2	4238	1716	1716
n	Equal Probability Case & $d=6$									
300	79.2	62.5	5.6	5.6	7.3	4.7	30.0	52	540	490
500	716.3	345.2	55.7	50.5	76.3	52.0	31.1	191	933	894
1000	†[▲,90.2]	948.2	2369.9	856.2	1621.5	522.4	30.5	954	1827	1809
n	Equal Probability Case& $d=8$									
300	192.4	133.6	31.8	26.0	27.0	11.0	49.0	72	1176	1021
500	3735.4	701.3	384.5	165.1	330.8	133.2	43.0	255	1720	1612
n	General Probability Case & $d=4$									
500	174.3	93.0	9.2	12.0	9.6	8.2	19.4	210	388	388
1000	1273.8	319.4	34.6	27.1	37.5	17.5	15.7	646	626	616
2000	†[▲,96.2]	296.8	1284.4	457.5	971.5	282.4	15.7	2498	1254	1254

Table 3: Effectiveness of the valid inequalities for (**NewMIP_CVaR**) - Unit simplex: Random instances ($\alpha = 0.01$)

G_1 and G_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. "All cuts" refers to the valid inequalities (34), (47) and (48).

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 with inequalities (65)-(66) on these instances, but observed that it does not perform better than the version with ordering inequalities. In our experience, ordering inequalities, when a large number of them exist, are preferable because they are sparse and they provide information on the realizations under multiple scenarios. In contrast, inequalities (65)-(66) are denser with very small coefficients for the instances tested, and they provide information on the correct calculation of the nonlinear shortfall term for a single scenario at a time. As a result, if a much larger number of ordering relations (34) across scenarios exist than the number of inequalities (65)-(66) (given by the multiplication of remaining number of scenarios and d), then it is preferable to use only the ordering inequalities in a brute force method that adds all inequalities a priori to the formulation. Alternatively, a branch-and-cut method can be implemented, with a more elaborate cut management system so as to benefit from both types of cuts. Furthermore, inequalities (65)-(66) can be strengthened using the ordering relation information for a scenario under which the realization is known to be smaller than the realization under another scenario. On the other hand, when the number of ordering relations is relatively small, the additional information provided by inequalities (65)-(66) could be more useful (see Table 3 for the performance of the analogue of inequalities (65)-(66) for the CVaR case).

6. Conclusions In this paper, we develop alternative mixed-integer programming formulations and solution methods for cut generation problems arising in a class of stochastic optimization problems that fea-

	(**MIP_SSD**$_l$) Time Aveg.	Std.	# Unslvd	(**MIP_SSD**$_l$) Fixing & Order. Ineq. Time Aveg.	Std.	Remaining Binary Var. (%) (Fixing) Aveg.	# of Order. Ineqs. (34) Aveg.
n	Equal Probability Case & $d = 4$						
200	14.9	12.6		5.9	5.8	73.6	2252
300	49.5	56.0		18.6	14.8	68.8	3844
500	†1147.2	1194.4	0.5	219.4	256.7	73.4	13621
n	Equal Probability Case & $d = 6$						
200	†517.2	991.5	0.5	41.0	87.1	85.9	1716
300	†3788.7	2202.9	8.5	559.6	663.5	84.7	3598
n	General Probability Case & $d = 4$						
200	12.4	9.7		5.6	4.8	73.6	2252
300	41.5	42.9		18.3	15.8	68.8	3844
500	460.9	490.6		89.5	86.7	73.4	13621
n	General Probability Case & $d = 6$						
200	270.6	550.2		28.4	52.8	85.9	1716
300	†2507.1	2189.4	4	165.5	189.0	84.7	3598

Table 4: Effectiveness of fixing and ordering inequalities for (**MIP_SSD**$_l$) - Unit simplex: Random instances
†: There exist instances with time limit, the number of these instances are reported under # Unslvd.

tures benchmarking constraints based on multivariate 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. In addition, we show that the proposed computational enhancements can be adapted to cut generation problems for multivariate polyhedral SSD-constrained optimization. We give the convex hull description of a polytope describing the linearization of a non-convex substructure arising in the SSD 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 the random vector of interest.

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

References

Armbruster, B. and Luedtke, J. (2015). Models and formulations for multivariate dominance constrained stochastic programs. *IIE Transactions*, 47:1–14.

Dentcheva, D., Penev, S., and Ruszczyński, 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 Ruszczyński, A. (2010). Inverse cutting plane methods for optimization problems with second-order stochastic dominance constraints. *Optimization*, 59(3):323–338.

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

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

Fábián, C., Mitra, G., and Roman, D. (2011). Processing second-order stochastic dominance models using cutting-plane representations. *Mathematical Programming*, 130(1):33–57.

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.

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

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.

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

Rudolf, G. and Ruszczyński, A. (2008). Optimization problems with second order stochastic dominance constraints: duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19:1326–1343.

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.