Models and Formulations for Multivariate Dominance Constrained Stochastic Programs

Benjamin Armbruster* James Luedtke†

June 22, 2011

Abstract

Dentcheva and Ruszczyński recently proposed using a stochastic dominance constraint to specify risk preferences in a stochastic program. Such a constraint requires the random outcome resulting from one's decision to stochastically dominate a given random comparator. These ideas have been extended to problems with multiple random outcomes, using the notion of positive linear stochastic dominance. We propose a constraint using a different version of multivariate stochastic dominance. This version is natural due to its connection to expected utility maximization theory and is relatively tractable. In particular, we show that such a constraint can be formulated with linear constraints for the second-order dominance relation, and with mixed-integer constraints for the first-order relation. This is in contrast to a constraint on second-order positive linear dominance, for which no efficient algorithms are known. We tested these formulations in the context of two applications: budget allocation in a setting with multiple objectives and finding radiation treatment plans in the presence of organ motion.

Keywords: Stochastic dominance, stochastic programming

1 Introduction

A major challenge in using optimization to make risk-averse decisions in the face of uncertainty is how to specify an acceptable level of risk. Recently, Dentcheva and Ruszczyński (2003) have introduced and studied the concept of a *stochastic dominance constraint*, which offers a flexible approach to specifying risk preferences in an optimization model. The idea is that a decision-maker can specify an acceptable random outcome (e.g., based on a default decision) and then the optimization model includes a constraint that the outcome of the selected solution should stochastically dominate the given random outcome. Specifically, the model takes the form

$$\begin{aligned} &\min_{x\in C} \; f(x) \\ &\; s.t. \; G(x) \succeq Y \end{aligned} \tag{1}$$

where $f(x)$ is an objective (e.g., cost) to be minimized, $G(x)$ is a random outcome depending on x, and Y is a given reference random variable. Thus, a decision-maker is assured that the optimization model will produce a solution that is preferable in a strong sense to the specified random outcome. This approach has been applied for financial portfolio optimization (Dentcheva and Ruszczyński,

*Northwestern University, armbruster@northwestern.edu

†University of Wisconsin-Madison, jrluedt1@wisc.edu, This research has been supported in part by the National Science Foundation under grant CMMI-0952907.

2006), power system optimization (Gollmer et al., 2007), and homeland security resource planning (Hu et al., 2010).

The notion of stochastic dominance can also be extended to random vectors. This can be useful for a decision-making setting that has multiple measures of interest (e.g., objectives) that are random. Dentcheva and Ruszczyński (2009) have used the notion of *positive linear stochastic dominance* as a constraint in the multi-dimensional setting, which is based on requiring one-dimensional dominance of any nonnegative weighted combination of the different outcomes. This model was generalized in Homem-de-Mello and Mehrotra (2009) by limiting the considered weights to within a specified set. Unfortunately, the models presented in these papers are computationally demanding to solve. In particular, even for the case of second-order stochastic dominance, which induces a convex feasible region, the algorithm of Homem-de-Mello and Mehrotra (2009) requires global optimization of a nonconvex problem as a subproblem.

In this paper, we propose to use a different notion of multivariate stochastic dominance as a constraint in a stochastic optimization model. The definitions we use are based on the connection of stochastic dominance to expected utility maximization theory and, for example, can be found in Müller and Stoyan (2002). (Another resource for information on stochastic dominance relations, also known as *stochastic orders*, is Shaked and Shanthikumar (2007).) Specifically, for d-dimensional random vectors $\boldsymbol{Y}$ and $\boldsymbol{W}$ having finite expectations, the first-order stochastic dominance relation (FSD), denoted $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$, holds if $\mathbb{E}[u(\boldsymbol{W})] \geq \mathbb{E}[u(\boldsymbol{Y})]$ for all nondecreasing utility functions u and the second-order stochastic dominance relation (SSD), denoted $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$, holds if $\mathbb{E}[u(\boldsymbol{W})] \geq \mathbb{E}[u(\boldsymbol{Y})]$ for all nondecreasing *concave* utility functions u. Here nondecreasing is with respect to the component-wise order in $\mathbb{R}^d$. (First- and second-order stochastic dominance are equivalent to what Shaked and Shanthikumar (2007) call the *usual* stochastic order and the *increasing convex* order, respectively.) Our primary motivation for using this alternative model is computational tractability: using this definition we are able to derive compact linear (for the SSD case) and mixed-integer (for FSD) formulations when $\boldsymbol{W}$ and $\boldsymbol{Y}$ have finite support. Consequently, the models we present can be solved using off-the-shelf linear programming (LP) and mixed-integer programming (MIP) solvers.

Another potential advantage of using the utility-based FSD and SSD definitions is the natural interpretation: a solution that satisfies such an FSD constraint would be preferred to the reference solution by *any* utility-maximizing decision-maker, and a solution that satisfies such an SSD constraint would be preferred by *any risk-averse* utility-maximizing decision-maker. Of course, the potential drawback of such a strong definition is that it may severely limit the set of feasible solutions. However, the flexibility in setting the reference random outcome can be used to overcome this drawback by simply using a lower quality random outcome as the reference if necessary.

The first contribution of this paper, aside from the model itself, is the derivation of an LP formulation for an SSD constraint, and several MIP formulations for an FSD constraint. These formulations are similar in spirit to the approach taken in Luedtke (2008) for the scalar case, although we find that deriving good MIP formulations for the multivariate FSD case is more challenging. We also provide a specialized branch-and-bound algorithm for solving an FSD-constrained problem.

The second contribution of this paper is to conduct some numerical examples on two applications. The first application is a generic budget allocation model, inspired by the homeland security application of Hu et al. (2010), in which a limited budget must be allocated to a set of possible projects, and the allocation must stochastically dominate a given benchmark allocation. The second application is in radiation treatment planning and considers the uncertainty induced by organ movement and other factors during treatment. Here a multivariate stochastic dominance constraint is attractive because there is no natural way to trade-off the multiple objectives of providing sufficient dose to the target (tumor) while limiting the dose to nearby critical organs.

In the next section we formally define our model and state our assumptions; review some results that motivate our approach; and discuss in more detail the relationship between the stochastic dominance concept we use to that used in previous work. Then in §3 we describe our LP formulation for an SSD constraint, and in §4 we describe the MIP formulations and the specialized branch-and-bound algorithm for an FSD constraint. To illustrate the feasibility of the models we present numerical examples for our two applications in §5. §6 has our concluding remarks.

2 Preliminaries

2.1 A comparison of multivariate stochastic dominance relations

We first review the definitions of first and second-order stochastic dominance for the scalar case. The random variable W dominates Y in the first order, written $W \succeq_1 Y$, if

$$F_W(\eta) \leq F_Y(\eta) \qquad \forall\, \eta \in \mathbb{R}. \tag{2}$$

where $F_W(\eta) := \Pr[W \leq \eta]$ and $F_Y(\eta) := \Pr[Y \leq \eta]$. The random variable W dominates Y in the second order, written $W \succeq_2 Y$, if

$$\int_{-\infty}^{\eta} F_W(z)dz \leq \int_{-\infty}^{\eta} F_Y(z)dz \qquad \forall\, \eta \in \mathbb{R}. \tag{3}$$

Alternatively, for $i = 1, 2$, ith order stochastic dominance can be defined by $W \succeq_i Y$ if

$$\mathbb{E}[u(W)] \geq \mathbb{E}[u(Y)] \quad \forall u \in \mathcal{U}_i \text{ such that the expectations exist,} \tag{4}$$

where $\mathcal{U}_1$ is the set of all nondecreasing functions $u : \mathbb{R} \to \mathbb{R}$, and $\mathcal{U}_2$ is the set of all nondecreasing and *concave* functions u. These definitions are equivalent to those given by (2) and (3) for scalar random variables W and Y, see e.g., Whitmore and Findlay (1978).

Extension of these stochastic dominance concepts to random vectors $\boldsymbol{W}$ and $\boldsymbol{Y}$ can be done in many ways. Dentcheva and Ruszczyński (2009) used the notion of *positive linear* stochastic dominance for a stochastic dominance constraint, and studied the corresponding optimization model. $\boldsymbol{W}$ is said to dominate $\boldsymbol{Y}$ by ith-order positive linear stochastic dominance, denoted $\boldsymbol{W} \succeq_i^{\text{lin}} \boldsymbol{Y}$, if $\boldsymbol{v} \cdot \boldsymbol{W} \succeq_i \boldsymbol{v} \cdot \boldsymbol{Y}$ in the usual sense of scalar ith order dominance for all vectors $\boldsymbol{v} \geq 0$. Homem-de-Mello and Mehrotra (2009) and Hu et al. (2009) used a generalized version of this notion where the weights $\boldsymbol{v}$ are constrained to lie in a convex set $\mathcal{C}$.

In contrast, the extension we use is based on the natural extension of the condition (4) used for the scalar case. Specifically, for d-dimensional random vectors $\boldsymbol{W}$ and $\boldsymbol{Y}$, $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ if (4) holds with $\mathcal{U}_1$ replaced by $\mathcal{U}_1^d$, the set of all nondecreasing functions $u : \mathbb{R}^d \to \mathbb{R}$. Similarly, $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ if (4) holds with $\mathcal{U}_2$ replaced by $\mathcal{U}_2^d$, the set of all nondecreasing and concave functions $u : \mathbb{R}^d \to \mathbb{R}$.

Figure 1 shows the relationship between the different forms of multivariate stochastic dominance. Here $\succeq^{\text{comp}}$ refers to component-wise dominance. Specifically, it is clear that first-order dominance implies second-order dominance (in any version of multivariate dominance) and also that the versions of first and second-order dominance that we use imply the corresponding positive linear dominance relations. Since we have not found in the literature a demonstration that this diagram is complete, we provide minimal examples in the appendix showing that these notions of dominance are distinct and that no additional relationships among them exist.

Since $\{\boldsymbol{W} : \boldsymbol{W} \succeq_2 \boldsymbol{Y}\}$ and $\{\boldsymbol{W} : \boldsymbol{W} \succeq_2^{\text{lin}} \boldsymbol{Y}\}$ are convex and Dentcheva and Ruszczyński (2003) introduced a linear program for (scalar) second-order dominance constraints one hopes that these multivariate extensions are similarly tractable. Unfortunately $\succeq_2^{\text{lin}}$ does not appear to be: testing

$$
\begin{tikzcd}
\succeq_1 \arrow[rr] \arrow[dr] & & \succeq_1^{\text{lin}} \arrow[rr] \arrow[dr] & & \succeq_1^{\text{comp}} \arrow[dr] & \\
& \succeq_2 \arrow[rr] & & \succeq_2^{\text{lin}} \arrow[rr] & & \succeq_2^{\text{comp}}
\end{tikzcd}
$$

Figure 1: Relationship between different notions of multivariate stochastic dominance.

whether $\boldsymbol{v} \cdot \boldsymbol{W} \succeq_2 \boldsymbol{v} \cdot \boldsymbol{Y}$ for all $\boldsymbol{v} \geq 0$ when $\boldsymbol{W}$ and $\boldsymbol{Y}$ have finite support can be formulated as a nonconvex feasibility problem and the formulations for (1) in Homem-de-Mello and Mehrotra (2009) and Hu et al. (2009) also require solving a nonconvex optimization problem to check feasibility.

2.2 Problem definition and assumptions

The models with stochastic dominance constraints we consider are

$$
\begin{aligned}
&\min_{\boldsymbol{x} \in C} f(\boldsymbol{x}) \\
&\;s.t.\; \boldsymbol{G}(\boldsymbol{x}) \succeq_\ell \boldsymbol{Y}
\end{aligned} \tag{5}
$$

for $\ell = 1$ (FSD) or $\ell = 2$ (SSD). Here $\boldsymbol{x}$ represents our decision vector, C defines the deterministic constraints on $\boldsymbol{x}$, $f(\boldsymbol{x})$ is the objective to be minimized, $\boldsymbol{G}(\boldsymbol{x})$ maps the decision vector $\boldsymbol{x}$ to a random vector in $\mathbb{R}^d$, and $\boldsymbol{Y}$ is a given reference random vector. The only difference between this model and the model introduced by Dentcheva and Ruszczyński (2003) is that we use the stochastic dominance relations based on (4).

We are interested in a computational approach for finding solutions to problem (5), specifically, by devising linear and mixed-integer linear programming formulations of the stochastic dominance constraints. We therefore assume that $\boldsymbol{G}(\boldsymbol{x})$ and $\boldsymbol{Y}$ have finite support, taking values on $\boldsymbol{g}^1(\boldsymbol{x}), \ldots, \boldsymbol{g}^N(\boldsymbol{x})$ and $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M$, respectively. (For each j, $\boldsymbol{g}^j$ is a function into $\mathbb{R}^d$ and $\boldsymbol{y}^j$ is a vector in $\mathbb{R}^d$.) This allows us to write (5) as

$$
\min_{\boldsymbol{x} \in C,\ \boldsymbol{w}^j} \; f(\boldsymbol{x}) \tag{6a}
$$

$$
s.t.\; \boldsymbol{g}^j(\boldsymbol{x}) \geq \boldsymbol{w}^j \quad \forall j \in \{1, \ldots, N\} \tag{6b}
$$

$$
\boldsymbol{W} \succeq_\ell \boldsymbol{Y} \tag{6c}
$$

where the $\boldsymbol{w}^j$ are decision variables and the functions $\boldsymbol{g}^j(\boldsymbol{x})$ in (6b) describe the random mapping $\boldsymbol{G}(\boldsymbol{x})$. This allows us to create formulations of the dominance constraint (6c), $\boldsymbol{W} \succeq_\ell \boldsymbol{Y}$, where $\boldsymbol{W}$ is a discrete random variable with $\Pr[\boldsymbol{W} = \boldsymbol{w}^j] = \Pr[\boldsymbol{G}(\boldsymbol{x}) = \boldsymbol{g}^j(\boldsymbol{x})]$. Thus if (6a)–(6b) is a mixed integer program (MIP) and (6c) can be formulated with mixed-integer linear constraints (as we show in the FSD case $\succeq_1$), then the entire optimization problem is a MIP. Similarly, if (6c) can be formulated with linear constraints (as we show in the SSD case $\succeq_2$) and (6a)–(6b) can be formulated as a convex program (i.e., C is convex, $f(\boldsymbol{x})$ is convex, and $\boldsymbol{g}^j(\boldsymbol{x})$ are concave), then the entire optimization problem is a convex program.

Notation and conventions. Define $\mathcal{N} := \{1, \ldots, N\}$, $\mathcal{M} := \{1, \ldots, M\}$, $\mathcal{D} := \{1, \ldots, d\}$, $q_y(i) := \Pr[\boldsymbol{Y} = \boldsymbol{y}^i]$, and $q_w(j) := \Pr[\boldsymbol{W} = \boldsymbol{w}^j]$. Inequalities between vectors are component-wise, $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution, $(x)^+ := \max\{x, 0\}$ represents the positive part of x, and $I(\cdot)$ denotes the indicator function. Superscripts denote indices, not exponents.

3 A formulation for second-order stochastic dominance

In the scalar case, definition (3) for second-order stochastic dominance is equivalent to

$$\mathbb{E}[(y^i - W)^+] \leq \mathbb{E}[(y^i - Y)^+] \quad \forall i \in \mathcal{M}, \tag{7}$$

and Dentcheva and Ruszczyński (2003) observed that this allows an SSD constraint to be formulated as a set of linear constraints by introducing variables z_{ij} to model $(y^i - w^j)^+$. However, this approach does not extend to the multi-dimensional case because the condition (7) does not imply the second-order dominance relation we use (based on (4)). Nevertheless, as in the one-dimensional case (Luedtke, 2008), we can directly apply Strassen's Theorem (Strassen, 1965) to obtain a linear formulation of second-order stochastic dominance. We now state a convenient specialization of this theorem.

Theorem 1 (Strassen)**.** *$\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ if and only if there exists random vectors $\boldsymbol{W}' \stackrel{\mathcal{D}}{=} \boldsymbol{W}$ and $\boldsymbol{Y}' \stackrel{\mathcal{D}}{=} \boldsymbol{Y}$ such that $\boldsymbol{W}' \geq \mathbb{E}[\boldsymbol{Y}'|\boldsymbol{W}']$ almost surely.*

Corollary 2. *Suppose $\boldsymbol{W}$ and $\boldsymbol{Y}$ are d-dimensional random vectors with finite support $\boldsymbol{w}^1, \ldots, \boldsymbol{w}^N$ and $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M$, respectively. Then $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ if and only if there exists p_{ij} for $i \in \mathcal{M}, j \in \mathcal{N}$ such that*

$$\sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij} \leq \boldsymbol{w}^j q_w(j) \quad \forall j \in \mathcal{N} \tag{8a}$$

$$\sum_{j \in \mathcal{N}} p_{ij} = q_y(i) \quad \forall i \in \mathcal{M} \tag{8b}$$

$$\sum_{i \in \mathcal{M}} p_{ij} = q_w(j) \quad \forall j \in \mathcal{N} \tag{8c}$$

$$p_{ij} \geq 0 \quad \forall i \in \mathcal{M}, j \in \mathcal{N}. \tag{8d}$$

The proof of this corollary is in the appendix and follows from $p_{ij} = \Pr[\boldsymbol{Y} = \boldsymbol{y}^i, \boldsymbol{W} = \boldsymbol{w}^j]$. We refer to formulation (8) as SSDLP. This formulation has NM nonnegative decision variables (beyond the $\boldsymbol{w}^j$) and $N + M + Nd$ constraints. Thus, the number of constraints in the formulation increases linearly with the dimension d of the random vectors and with the size of the support of the random vectors $\boldsymbol{W}$ and $\boldsymbol{Y}$.

3.1 First order methods

Formulation (8) can be used directly by adding the associated constraints and variables to the model. However, if we prefer to work in the space of the original variables x, we can rewrite the stochastic dominance constraint in the form $h(\boldsymbol{x}) \leq 0$, for a suitably defined function h. Specifically, we define $h(\boldsymbol{x}) := \inf\{t : \boldsymbol{W}(\boldsymbol{x}) + t\boldsymbol{1} \succeq_2 \boldsymbol{Y}\}$ and then write the original problem ((5) or (6)) as

$$\begin{aligned} &\min_{\boldsymbol{x} \in C} f(\boldsymbol{x}) \\ &\ s.t.\ h(\boldsymbol{x}) \leq 0. \end{aligned} \tag{9}$$

Thus as an alternative to SSDLP (8) we can tackle this problem using first-order methods that only require evaluations of $h(\boldsymbol{x})$ and its subgradient $\partial h(\boldsymbol{x})$. Using (8), we obtain

$$
\begin{aligned}
h(\boldsymbol{x}) := \min_{t, p_{ij}} \; & t \\
\text{s.t.} \; & \sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij} \leq (\boldsymbol{g}_j(\boldsymbol{x}) + t\mathbf{1}) q_w(j) \quad \forall j \in \mathcal{N} \\
& \sum_{j \in \mathcal{N}} p_{ij} = q_y(i) \quad \forall i \in \mathcal{M} \\
& \sum_{i \in \mathcal{M}} p_{ij} = q_w(j) \quad \forall j \in \mathcal{N} \\
& p_{ij} \geq 0 \quad \forall i \in \mathcal{M}, j \in \mathcal{N}.
\end{aligned} \tag{10}
$$

Since this LP is feasible and the objective is bounded below, taking the dual yields

$$
\begin{aligned}
h(\boldsymbol{x}) = \max_{\boldsymbol{s}_j, \boldsymbol{u}_i, \boldsymbol{v}_j} \; & \sum_{j \in \mathcal{N}} (\boldsymbol{v}_j - \boldsymbol{s}_j \cdot \boldsymbol{g}_j(\boldsymbol{x})) q_w(j) + \sum_{i \in \mathcal{M}} \boldsymbol{u}_i q_y(i) \\
\text{s.t.} \; & \sum_{j \in \mathcal{N}} (\boldsymbol{s}_j \cdot \mathbf{1}) q_w(j) = 1 \\
& \boldsymbol{u}_i + \boldsymbol{v}_j \leq \boldsymbol{s}_j \cdot \boldsymbol{y}^i \quad \forall i \in \mathcal{M}, j \in \mathcal{N} \\
& \boldsymbol{s}_j \geq 0 \quad \forall j \in \mathcal{N}.
\end{aligned} \tag{11}
$$

The $\boldsymbol{s}_j$ variables ($\boldsymbol{s}_j \in \mathbb{R}^d$) are the shadow prices of the inequality constraints in (10). For a given $\boldsymbol{x}$, we can solve (10) or (11) to evaluate $h(\boldsymbol{x})$. Furthermore, if $\boldsymbol{s}_j^*$ optimize the dual problem, then $-\sum_{j \in \mathcal{N}} (\boldsymbol{s}_j^* \cdot \boldsymbol{g}_j'(\boldsymbol{x})) q_w(j) \in \partial h(\boldsymbol{x})$. Thus, solving a simple LP allows us to evaluate both $h(\boldsymbol{x})$ and $\partial h(\boldsymbol{x})$. This could be useful, for example, in a cutting-plane algorithm where, instead of directly introducing the p_{ij} variables and constraints (8) into the formulation, the stochastic dominance constraint is enforced by dynamically adding linear inequalities.

4 Formulations for first-order stochastic dominance

In this section, we present a series of formulations for first-order stochastic dominance. The formulations are based on the following characterization of FSD, given in Theorem 3.4.2 of Müller and Stoyan (2002).

Theorem 3. *$\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ if and only if there exists random vectors $\boldsymbol{W}' \stackrel{\mathcal{D}}{=} \boldsymbol{W}$ and $\boldsymbol{Y}' \stackrel{\mathcal{D}}{=} \boldsymbol{Y}$ such that $\boldsymbol{W}' \geq \boldsymbol{Y}'$ almost surely.*

This yields the following nonlinear, nonconvex formulation of FSD, which we subsequently use to derive a series of mixed-integer linear formulations.

Corollary 4. *Suppose $\boldsymbol{W}$ and $\boldsymbol{Y}$ are discrete d-dimensional random vectors with finite support $\boldsymbol{w}^1, \ldots, \boldsymbol{w}^N$ and $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M$, respectively. Then $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ if and only if there exists p_{ij} for $i \in \mathcal{M}, j \in \mathcal{N}$ such that (8b)–(8d) and*

$$
p_{ij}(\boldsymbol{w}^j - \boldsymbol{y}^i) \geq 0, \quad \textit{for } i \in \mathcal{M}, j \in \mathcal{N}. \tag{12}
$$

Proof. Suppose $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ and let $\boldsymbol{W}'$ and $\boldsymbol{Y}'$ be as in Theorem 3. Let $p_{ij} = \Pr[\boldsymbol{W}' = \boldsymbol{w}^j, \boldsymbol{Y}' = \boldsymbol{y}^i]$ for $i \in \mathcal{M}, j \in \mathcal{N}$. Then, $p_{ij} \geq 0$, and similar to the proof of Corollary 2, p_{ij} satisfies (8b) and (8c). For each $i \in \mathcal{M}, j \in \mathcal{N}$, if $p_{ij} = 0$, then (12) trivially holds, whereas if $p_{ij} > 0$, then because $\boldsymbol{W}' \geq \boldsymbol{Y}'$ almost surely, we must have $\boldsymbol{w}^j \geq \boldsymbol{y}^i$, and hence (12) still holds.

Now, suppose there exists p_{ij} such that (8b)–(8d) and (12) holds. As in the proof of Corollary 2, define a probability space with sample space $\Omega' = \mathcal{M} \times \mathcal{N}$ and probability measure $P'(i,j) := p_{ij}$. Then define random vectors $\boldsymbol{W}'$ and $\boldsymbol{Y}'$ by $\boldsymbol{W}'(i,j) := \boldsymbol{w}^j$ and $\boldsymbol{Y}'(i,j) := \boldsymbol{y}^i$. Then $\boldsymbol{W}' \overset{\mathcal{D}}{=} \boldsymbol{W}$, $\boldsymbol{Y}' \overset{\mathcal{D}}{=} \boldsymbol{Y}$, $\boldsymbol{w}^j \geq \boldsymbol{y}^i$ whenever $p_{ij} > 0$ and hence $\boldsymbol{W}' \geq \boldsymbol{Y}'$ almost surely. Theorem 3 then implies $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. □

Before moving to the derivation of MIP formulations for the FSD constraint, we comment on the possibility of a nonconvex formulation that includes the constraints (12) directly. First, observe that (12) is equivalent to the following set of linear complementarity conditions

$$\boldsymbol{w}^j + s_{ij}\mathbf{1} \geq \boldsymbol{y}^i, \quad s_{ij} \geq 0, \quad p_{ij}s_{ij} = 0, \quad \forall i \in \mathcal{M}, j \in \mathcal{N}.$$

While finding provable optimal solutions to problems with linear complementarity constraints is difficult in general, a number of approaches have been studied to find good solutions (which cope with some of the technical difficulties associated with the fact that "nice" constraint qualifications do not hold for this problem) (Luo et al., 1996). We conducted some preliminary experiments using this formulation in the nonlinear programming problem solver Knitro, but found that this approach was not competitive with the other formulations.

4.1 Two compact MIP formulation (FSD1 and FSD1+SSD)

We derive our first mixed-integer linear formulation by observing that the constraints (12) can equivalently be stated as

$$p_{ij} > 0 \implies \boldsymbol{w}^j \geq \boldsymbol{y}^i \quad \forall i \in \mathcal{M}, j \in \mathcal{N}. \tag{13}$$

We model this with binary variables χ_{ij}, where $\chi_{ij} = 0$ implies $p_{ij} = 0$ and $\chi_{ij} = 1$ implies $\boldsymbol{w}^j \geq \boldsymbol{y}^i$. Thus we obtain that $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ if and only if there exists p_{ij} and χ_{ij} for $i \in \mathcal{M}$ and $j \in \mathcal{N}$ such that (8b)-(8d) and

$$p_{ij} \leq \min\{q_w(j), q_y(i)\}\chi_{ij}, \quad \forall i \in \mathcal{M}, j \in \mathcal{N}, \tag{14a}$$

$$\boldsymbol{y}^{\min} + \chi_{ij}(\boldsymbol{y}^i - \boldsymbol{y}^{\min}) \leq \boldsymbol{w}^j, \quad \forall i \in \mathcal{M}, j \in \mathcal{N}, \tag{14b}$$

$$\chi_{ij} \in \{0,1\}, \quad \forall i \in \mathcal{M}, j \in \mathcal{N}. \tag{14c}$$

where $\boldsymbol{y}^{\min}$ is the component-wise minimum of $\{\boldsymbol{y}^i\}$. We refer to this MIP formulation as FSD1.

A disadvantage of FSD1 are the "big-M" type of constraints in (14b), which are likely to lead to weak linear programming relaxations. This disadvantage can be significantly alleviated by including the inequalities (8a). The following theorem justifies that this can be done.

Theorem 5. *$\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ if and only if there exists p_{ij} and χ_{ij} for $i \in \mathcal{M}$ and $j \in \mathcal{N}$ such that (8) and (14).*

Proof. Assume that $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. From formulation FSD1 there exists p_{ij} and χ_{ij} such that (8b)–(8d) and (14). We show p_{ij} that also satisfy (8a). Indeed, (14) imply (12). Summing (12) over $i \in \mathcal{M}$ yields

$$0 \leq \sum_{i \in \mathcal{M}} p_{ij}(\boldsymbol{w}^j - \boldsymbol{y}^i) = \boldsymbol{w}^j q_w(j) - \sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij}$$

by (8c), and hence (8a) is satisfied.

The reverse implication is immediate. □

We refer to the formulation of Theorem 5 as FSD1+SSD. Because this formulation includes all the variables and constraints of the SSDLP formulation, the linear programming relaxation yields a bound at least as strong as the bound obtained from the SSD relaxation, relaxing the FSD constraint $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ to an SSD constraint $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$. For scalar stochastic dominance constraints it has been shown that the SSD relaxation is often good (Noyan et al., 2006). Furthermore, for an important special case in which $\boldsymbol{W}$ and $\boldsymbol{Y}$ take on all values with equal probability, we show in § 4.4 that the SSD relaxation is in a sense as good as can be achieved.

Formulation FSD1+SSD still has two potential disadvantages. First, although the relaxation is strengthened by the addition of (8a), the "big-M" constraints (14b) are still required. This may cause slow improvement in the bound when branching. To address this concern, we present in §4.2 a MIP formulation in which the natural linear programming relaxation itself is as strong as the SSD relaxation. The second disadvantage of formulation FSD1+SSD is that it has $(d+1)NM$ constraints in (14a) and (14b), which grows large if both N and M are large. In contrast, the SSDLP formulation has NM variables but only $O(M+dN)$ constraints. One option to deal with this disadvantage is to handle the logical conditions (13) *algorithmically*, rather than reformulating them as we have done. We discuss this option in detail in §4.3.

4.2 A strong formulation in an extended space (FSD2)

We now present a formulation for $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ that successfully avoids the "big-M" inequalities (14b) and as a result has a strong LP relaxation without the need to add (8a). The disadvantage of this formulation is its large size: it has NM continuous variables, NMd binary variables, and $2Nd + NMd$ constraints.

Theorem 6. *$\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ iff there exists p_{ij} for $i \in \mathcal{M}, j \in \mathcal{N}$ and $z_{ijk} \geq 0$ for $i \in \mathcal{M}, j \in \mathcal{N}, k \in \mathcal{D}$ satisfying* (8b)–(8d)*, and*

$$\sum_{i \in \mathcal{M}} y_k^i z_{ijk} \leq w_k^j \quad \forall j \in \mathcal{N}, k \in \mathcal{D}, \tag{15a}$$

$$\sum_{i \in \mathcal{M}} z_{ijk} = 1 \quad \forall j \in \mathcal{N}, k \in \mathcal{D}, \tag{15b}$$

$$\sum_{l \in \mathcal{M}} p_{lj} I(y_k^l \geq y_k^i) \leq q_w(j) \sum_{l \in \mathcal{M}} z_{ljk} I(y_k^l \geq y_k^i) \quad \forall i \in \mathcal{M}, j \in \mathcal{N}, k \in \mathcal{D} \tag{15c}$$

Furthermore, $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ iff the above conditions hold and $z_{ijk} \in \{0,1\}$ for all $i \in \mathcal{M}, j \in \mathcal{N}, k \in \mathcal{D}$.

Proof. To prove the first claim, start by assuming that $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$. Corollary 2 then implies that there exists p_{ij} such that (8) hold. Then, choosing $z_{ijk} = p_{ij}/q_w(j)$ ensures that (15c) holds, and (15b) holds by (8c). Finally, (15a) holds due to (8a).

To prove the converse, that $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ given the assumptions, we show that (8a) holds and thus Corollary 2 applies. Fix $j \in \mathcal{N}$ and $k \in \mathcal{D}$; let σ be a permutation of $\mathcal{M}$ such that

$$y_k^{\sigma(1)} \leq y_k^{\sigma(2)} \leq \cdots \leq y_k^{\sigma(M)}; \tag{16}$$

and define $y_k^{\sigma(0)} = y_k^{\sigma(1)}$. Then we have,

$$\begin{aligned}
q_w(j)w_k^j &\geq q_w(j) \sum_{i \in \mathcal{M}} y_k^{\sigma(i)} z_{\sigma(i)jk} && \text{by (15a)} \\
&= q_w(j) \sum_{i \in \mathcal{M}} z_{\sigma(i)jk} \Big(y_k^{\sigma(0)} + \sum_{l=1}^{i} (y_k^{\sigma(l)} - y_k^{\sigma(l-1)}) \Big) \\
&= q_w(j) y_k^{\sigma(0)} + q_w(j) \sum_{l \in \mathcal{M}} (y_k^{\sigma(l)} - y_k^{\sigma(l-1)}) \sum_{i=l}^{M} z_{\sigma(i)jk} && \text{by (15b)} \\
&\geq q_w(j) y_k^{\sigma(0)} + \sum_{l \in \mathcal{M}} (y_k^{\sigma(l)} - y_k^{\sigma(l-1)}) \sum_{i=l}^{M} p_{\sigma(i),j} && \text{by (15c)} \\
&= q_w(j) y_k^{\sigma(0)} + \sum_{i \in \mathcal{M}} p_{\sigma(i),j} \sum_{l=1}^{j} (y_k^{\sigma(l)} - y_k^{\sigma(l-1)}) \\
&= q_w(j) y_k^{\sigma(0)} + \sum_{i \in \mathcal{M}} p_{\sigma(i),j} (y_k^{\sigma(i)} - y_k^{\sigma(0)}) \\
&= \sum_{i \in \mathcal{M}} p_{\sigma(i),j} y_k^{\sigma(i)} && \text{by (8c)}
\end{aligned}$$

which proves that (8a) holds.

Now we turn to the second claim. First assume that $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. Then, by Corollary 4, there exists p_{ij} such that (8b)–(8d) and (13) holds. Since $\boldsymbol{W}$ is supported at $\boldsymbol{w}^j$, (8c) and (13) imply there exists i such that $\boldsymbol{y}^i \leq \boldsymbol{w}^j$. Thus for each $j \in \mathcal{N}$ and $k \in \mathcal{D}$, we can choose $n(j,k)$ from $\arg\max_i \{y_k^i : y_k^i \leq w_k^j\}$. Set $z_{ijk} := I(i = n(j,k))$. Then (15a) and (15b) are satisfied. For each, ijk, if $\sum_{l \in \mathcal{M}} p_{lj} I(y_k^l \geq y_k^i) = 0$, then (15c) holds. If however, $\sum_{l \in \mathcal{M}} p_{lj} I(y_k^l \geq y_k^i) > 0$, then there exists l with $y_k^l \geq y_k^i$ and $p_{lj} > 0$. Then (13) implies $w_k^j \geq y_k^l$ for all $k \in \mathcal{D}$. Thus $y_k^{n(j,k)} \geq y_k^l \geq y_k^i$ and

$$q_w(j) \sum_{l \in \mathcal{M}} z_{ljk} I(y_k^l \geq y_k^i) = q_w(j) \geq \sum_{l \in \mathcal{M}} p_{lj} I(y_k^l \geq y_k^i) \tag{17}$$

where the last inequality follows from (8c). This shows that (15c) holds.

To prove the converse, that $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ given the assumptions, we show that (13) holds, thus satisfying Corollary 4. Consider an arbitrary $i \in \mathcal{M}$ and $j \in \mathcal{N}$. If $p_{ij} = 0$, then (13) holds trivially. If $p_{ij} > 0$, then (15c) implies

$$q_w(j) \sum_{l \in \mathcal{M}} z_{ljk} I(y_k^l \geq y_k^i) \geq p_{ij} > 0 \tag{18}$$

and hence by (15b) and $z_{ijk} \in \{0, 1\}$,

$$\sum_{l \in \mathcal{M}} z_{ljk} I(y_k^l \geq y_k^i) = 1. \tag{19}$$

Thus $y_k^i = \sum_{l \in \mathcal{M}} z_{ljk} y_k^i I(y_k^l \geq y_k^i) \leq \sum_{l \in \mathcal{M}} z_{ljk} y_k^l$. Then (15a) implies $w_k^j \geq y_k^i$ for all $k \in \mathcal{D}$, proving (13). □

The size of this formulation is driven primarily by the NMd inequalities in (15c), which furthermore have a total of $O(NM^2d)$ nonzeros. An equivalent formulation in which the number of nonzeros is reduced to $O(NMd)$ can be obtained by using a different set of binary variables $\hat{z}_{ijk}$ related to the variables z_{ijk} as follows

$$\hat{z}_{ijk} = \sum_{l \in \mathcal{M}} z_{ljk} I(y_k^l \geq y_k^i)$$

and also introducing continuous variables $\hat{p}_{ijk}$ that are related to the p_{ij} variables by

$$\hat{p}_{ijk} = \sum_{l \in \mathcal{M}} p_{lj} I(y_k^l \geq y_k^i).$$

These relations can be achieved by increasing the number of constraints by only a constant factor, while reducing the total number of nonzeros to $O(NMd)$; the inequalities (15c) reduce to $\hat{p}_{ijk} \leq q_w(j)\hat{z}_{ijk}$. However, in our experiments, this reformulation performed slightly worse than the original.

4.3 A specialized branch-and-bound method (BnB)

As described in §4.1, a formulation for first-order stochastic dominance is

$$\sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij} \leq \boldsymbol{w}^j q_w(j) \quad \forall j \in \mathcal{N} \tag{20a}$$

$$\sum_{j \in \mathcal{N}} p_{ij} = q_y(i) \quad \forall i \in \mathcal{M} \tag{20b}$$

$$\sum_{i \in \mathcal{M}} p_{ij} = q_w(j) \quad \forall j \in \mathcal{N} \tag{20c}$$

$$p_{ij} \geq 0 \quad \forall i \in \mathcal{M}, j \in \mathcal{N} \tag{20d}$$

and the logical conditions

$$p_{ij} > 0 \implies \boldsymbol{w}^j \geq \boldsymbol{y}^i \quad \forall i \in \mathcal{M}, j \in \mathcal{N}. \tag{21}$$

Such a logical formulation can be implemented directly in IBM Ilog Cplex using indicator constraints (specifically, with “IloIfThen” constraints), but we found this performed worse than the MIP formulations presented in §4.1. We therefore explored a specialized branch-and-bound method for solving this formulation. The idea of the branch-and-bound algorithm is to use the relaxation defined by the linear constraints (20) and then use branching to enforce the logical conditions (21) by updating bounds on the variables p_{ij} and $\boldsymbol{w}^j$.

At any node n in the branch-and-bound tree we have upper bounds $U_{ij}(n)$ on the p_{ij} variables and lower bounds $\boldsymbol{L}^j(n)$ on the $\boldsymbol{w}^j$ variables and solve the following node relaxation:

$$\begin{aligned} \mathrm{lb}(n) := \min_{\boldsymbol{x} \in C, \boldsymbol{w}^j, p_{ij}} \quad & f(\boldsymbol{x}) \\ s.t. \quad & \boldsymbol{g}^j(\boldsymbol{x}) \geq \boldsymbol{w}^j, \ \boldsymbol{w}^j \geq \boldsymbol{L}^j(n) \quad \forall j \in \mathcal{N} \\ & p_{ij} \leq U_{ij}(n) \quad \forall i \in \mathcal{M}, j \in \mathcal{N} \\ & (20). \end{aligned} \tag{22}$$

At the root node, $n = 0$, we set $U_{ij}(0) = \min\{q_y(i), q_w(j)\}$ and $\boldsymbol{L}^j(n) = \boldsymbol{y}^{\min}$ so that none of these bounds constrain p_{ij} or $\boldsymbol{w}$ beyond the relaxation (20). After solving (22) at each node n, we first check whether $\mathrm{lb}(n)$ exceeds the objective value of the best feasible solution found so far; if so, we can prune this node and choose a different node to process. Otherwise, we check whether the relaxation solution $(\bar{\boldsymbol{p}}, \bar{\boldsymbol{w}})$ satisfies the logical conditions (21). If so, the solution is feasible and the solution and its objective value $\mathrm{lb}(n)$ are saved as the current incumbent. If not, we branch by choosing an (i, j) pair that violates (21). In our implementation we selected (i^*, j^*) that maximizes the quantity $\bar{p}_{ij}(\mathbf{1} \cdot \max\{\boldsymbol{y}^i - \bar{\boldsymbol{w}}^j, \mathbf{0}\})$. We then create two new nodes, indexed by c_1 and c_2, by enforcing $p_{i^*j^*} = 0$ in c_1 (implemented by setting $U_{i^*j^*}(c_1) = 0$) and by enforcing $\boldsymbol{w}^{j^*} \geq \boldsymbol{y}^{i^*}$ in node c_2 (implemented by setting $\boldsymbol{L}^{j^*}(c_2) = \max\{\boldsymbol{L}^{j^*}(n), \boldsymbol{y}^{i^*}\}$). The algorithm terminates when there are no more nodes to explore. This branch-and-bound algorithm is correct because no feasible solution is ever excluded from the search space, and terminates with the optimal solution if one exists because there are only finitely many logical conditions that can be enforced.

Heuristic We also derived a simple heuristic that can be used in conjunction with this branch-and-bound algorithm to help it find feasible solutions more quickly. The heuristic uses the values $\bar{\boldsymbol{w}}^j$ of a relaxation solution, potentially from any node in the search tree, as an input. The first step is to solve the problem:

$$\begin{aligned} \min_{p_{ij}} \quad & \sum_{i \in \mathcal{M}, j \in \mathcal{N}} (\mathbf{1} \cdot \max\{\boldsymbol{y}^i - \bar{\boldsymbol{w}}^j, \mathbf{0}\}) p_{ij} \\ s.t. \quad & (20b) - (20d). \end{aligned} \tag{23}$$

Given an optimal solution $\hat{\boldsymbol{p}}$ of this problem, we then set lower bounds on the $\boldsymbol{w}^j$ variables that would make $\boldsymbol{w}^j$ and $\hat{p}_{ij}$ satisfy (21). Specifically, we set the lower bound for each component w^j_k to $\bar{L}^j_k = \max_i\{y^i_k : \hat{p}_{ij} > 0\}$. We then solve the problem

$$\begin{aligned} & \min_{\boldsymbol{x} \in C} f(\boldsymbol{x}) \\ & \quad s.t \ \boldsymbol{g}^j(\boldsymbol{x}) \geq \bar{\boldsymbol{L}}^j \ \forall j \in \mathcal{N}. \end{aligned} \tag{24}$$

By construction, any feasible solution to this problem satisfies the stochastic dominance constraint and is feasible in the original problem. The heuristic fails if no feasible solution to this problem is found.

4.4 A special case (FSD3)

When $N = M$ and all the outcomes are equally likely (as they would be for example when obtained from random sampling), then we have a simpler and more compact formulation for $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$, whose linear relaxation implies $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ and is equivalent to SSDLP. This formulation has only N^2 binary variables and $2N + Nd$ constraints. Furthermore, in this case of equally likely outcomes, we can show that the set of random variables dominating in second-order dominance is the convex hull of those dominating in first-order dominance. We give the proof in the appendix, since it is analogous to that of theorem 4 in Dentcheva and Ruszczyński (2004) for the scalar case.

Theorem 7. *Suppose that $N = M$ and $q_y(i) = q_w(i) = 1/N$ for all $i \in \mathcal{N}$. Then*

1. $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$ *if and only if there exists* $p_{ij} \in \{0,1\}$ $\forall i \in \mathcal{M}, j \in \mathcal{N}$ *such that*

$$\sum_{j \in \mathcal{N}} p_{ij} = 1 \quad \forall i \in \mathcal{M}, \tag{25a}$$

$$\sum_{i \in \mathcal{M}} p_{ij} = 1 \quad \forall j \in \mathcal{N}, \tag{25b}$$

$$\sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij} \le \boldsymbol{w}^j \quad \forall j \in \mathcal{N}, \tag{25c}$$

2. $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ *if and only if there exists* $p_{ij} \ge 0$ *for* $i \in \mathcal{M}, j \in \mathcal{N}$ *such that* (25).

3. $\operatorname{conv}\{\boldsymbol{W} : \boldsymbol{W} \succeq_1 \boldsymbol{Y}\} = \{\boldsymbol{W} : \boldsymbol{W} \succeq_2 \boldsymbol{Y}\}$,

5 Numerical illustration

5.1 A Budget Allocation Model

We first study the behavior of these formulations to solve a simple budget allocation problem, inspired by the model of Hu et al. (2010) for homeland security budget allocation. Given a fixed budget, the problem is to determine what fraction of the budget to allocate to a set of candidate projects, $t \in \mathcal{T}$ with $|\mathcal{T}| = T$. The quality of a budget allocation is characterized by d distinct objectives, for which larger values are preferred. Each project $t \in \mathcal{T}$ is characterized by a d-dimensional random vector of reward rates $\boldsymbol{R}_t$ for these objectives. Thus, given a feasible budget allocation $\boldsymbol{x} \in C := \{\boldsymbol{x} \in \mathbb{R}^T_+ : \boldsymbol{x} \cdot \boldsymbol{1} = 1\}$, the values of the d objectives are $\sum_{t \in \mathcal{T}} \boldsymbol{R}_t x_t$. We assume that we are given a d-dimensional random vector $\boldsymbol{Y}$ that indicates a minimal acceptable joint performance of these objectives, and we require the performance of the chosen budget allocation to stochastically dominate $\boldsymbol{Y}$. Subject to this condition, goal is to maximize a weighted combination of the expected values of the measures:

$$\begin{aligned} \max_{x \in C} \ & \sum_{t \in \mathcal{T}} \boldsymbol{v} \cdot \mathbb{E}[\boldsymbol{R}_t] x_t \\ s.t. \ & \sum_{t \in \mathcal{T}} \boldsymbol{R}_t x_t \succeq_\ell \boldsymbol{Y}, \end{aligned} \tag{26}$$

where $\boldsymbol{v} \in \mathbb{R}^d_+$ is a given weight vector and $\ell \in \{1,2\}$ determines if we wish to enforce first or second-order dominance.

For our test instances, we assumed that our reward rates $\bar{\boldsymbol{R}} := [\boldsymbol{R}_1 \ \boldsymbol{R}_2 \ \cdots \boldsymbol{R}_T]$ are one of N equally likely scenarios $\{\bar{\boldsymbol{R}}^j : j \in \mathcal{N}\}$ sampled from a joint normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. (Thus $q_w(j) = 1/N$.) The components of $\boldsymbol{\mu}$ are chosen randomly from $U[10,20]$ and the covariance matrix $\boldsymbol{\Sigma}$ was calculated as follows. The coefficient of variations were chosen from $U[0.2,1.1]$. The correlation of any two distinct elements (t,k) and (t',k') were chosen from $U[-0.2,0.4]$ if they share a project $(t = t')$ and from $U[-0.1,0.1]$ if they share an objective $(k = k')$ and were 0 otherwise. The benchmark random vector $\boldsymbol{Y}$ was determined from an allocation in which all projects are allocated an equal fraction of the budget, but to avoid being overly conservative, was then reduced by a fixed fraction δ of its mean. Specifically, given realizations $\boldsymbol{R}^j_t \in \mathbb{R}^d$, for each scenario j and project t, realization j of $\boldsymbol{Y}$ has probability $q_y(j) = 1/N$ and is given by $\boldsymbol{Y}^j = \boldsymbol{B}^j - \delta(\frac{1}{N}\sum_{k=1}^N \boldsymbol{B}^k)$ where $\boldsymbol{B}^j = \frac{1}{T}\sum_{t \in \mathcal{T}} \boldsymbol{R}^j_t$. In our experiments, we found that using $\delta = 0.1$ allowed sufficient flexibility in the stochastic dominance constraint that a solution

(d,T)	$N=100$	$N=300$	$N=500$
$(3,50)$	0.3	12.3	86.2
$(3,100)$	0.3	8.9	61.6
$(5,50)$	0.6	37.8	181.8
$(5,100)$	0.7	23.0	105.6

Table 1: Average solution times in seconds of five instances using formulation SSDLP.

significantly better than the base allocation in terms of the objective could be found. Finally, we weighted all objectives equally in the objective, $\boldsymbol{v} = \mathbf{1}$. All tests in this section were done using Cplex 12.2 on a Mac Mini running OSX 10.6.6 with a two-core 2.66 GHz processor and 8 GB RAM and with the number of threads limited to one.

Table 1 shows the computation times to solve these instances using the second-order stochastic dominance formulation SSD from §3. For these experiments, we varied the number of objectives $d \in \{3,5\}$, the number of projects $T \in \{50,100\}$, and the number of scenarios $N = M \in \{100,300,500\}$. For each combination of these parameters we display the average computation time in seconds over five instances at that size. These results indicate that with this formulation it is possible to solve instances with a relatively large number of scenarios, although the computation time does grow significantly with the number of scenarios.

For the FSD formulations, because $M = N$ and all outcomes are equally likely, we can use the formulation FSD3 of §4.4 for this special case. However, we also tested the generally applicable formulations of §4 to gain insights into the relative performance of these. For these experiments, we varied the number of objectives $d \in \{3,5\}$, the number of projects $T \in \{50,100\}$, and the number of scenarios $N = M \in \{50,100,150\}$. We solved these instances with FSD1, FSD1+SSD, FSD2, BnB (the specialized branch-and-bound algorithm of §4.3), and FSD3. In BnB, we implemented the heuristic within a CPLEX heuristic callback, which called it after the root LP relaxation was solved and occasionally throughout the branch-and-bound tree, but we were unable to control the frequency. For each set of parameters we generated five instances and used a time limit of ten minutes for these experiments.

Since the benchmark random vector is constructed based on a particular allocation, these problems are feasible by construction. However, in some cases CPLEX failed to find a feasible solution using these formulations within the time limit. To make the results more informative and to reduce variability caused by the time at which a feasible solution is found with one of the formulations, we supply the base allocation as a starting feasible solution to all the instances.

Tables 2 and 3 present the average solution times (in seconds) and the average final optimality gaps after the time limit. The average solution times are calculated only over the instances that were solved in the time limit. The optimality gap for an instance is calculated as (ub − lb)/ub, where ub is the best upper bound obtained by the method in the time limit, and lb is the value of the best feasible solution found in the time limit. Instances that were solved to optimality are included in the average gap calculation.

The first observation from these tables is that FSD3, which exploits the property that $M = N$ and all scenarios are equally likely, performs much better than the other formulations, and in particular solves most of the instances to optimality within the time limit, with the exception of the instances with $d = 5$ objectives and $T = 50$ projects, which appears to be the most difficult combination of parameters. We also see that adding inequalities (8a) to FSD1 to obtain FSD1+SSD helps significantly, in terms of the number of instances that can be solved in the time limit, the

(d, T)	N	FSD1	FSD1+SSD	FSD2	BnB	FSD3
$(3, 50)$	50	194.7 (2)	78.9 (5)	205.8 (4)	32.2 (5)	3.2 (5)
	100	– (0)	398.8 (1)	– (0)	– (0)	56.5 (5)
	150	– (0)	– (0)	– (0)	– (0)	222.9 (5)
$(3, 100)$	50	283.3 (2)	28.7 (5)	124.2 (5)	14.9 (5)	1.5 (5)
	100	– (0)	90.3 (2)	– (0)	47.8 (2)	12.5 (5)
	150	– (0)	– (0)	– (0)	– (0)	114.3 (5)
$(5, 50)$	50	– (0)	350.8 (3)	– (0)	357.2 (1)	41.1 (5)
	100	– (0)	– (0)	– (0)	– (0)	581.2 (1)
	150	– (0)	– (0)	– (0)	– (0)	– (0)
$(5, 100)$	50	– (0)	153.9 (4)	594.1 (1)	41.9 (2)	8.9 (5)
	100	– (0)	– (0)	– (0)	– (0)	304.7 (5)
	150	– (0)	– (0)	– (0)	– (0)	390.4 (1)

Table 2: Average solutions times in seconds (instances solved) of five instances for different FSD formulations.

(d, T)	N	FSD1	FSD1+SSD	FSD2	BnB	FSD3
$(3, 50)$	50	0.1%	0.0%	0.0%	0.0%	0.0%
	100	*3.3%	*0.1%	*3.8%	0.2%	0.0%
	150	*3.8%	*1.6%	–	0.2%	0.0%
$(3, 100)$	50	0.1%	0.0%	0.0%	0.0%	0.0%
	100	1.1%	0.0%	*1.6%	0.1%	0.0%
	150	*1.9%	0.3%	–	0.4%	0.0%
$(5, 50)$	50	1.9%	0.1%	2.7%	0.2%	0.0%
	100	*4.2%	*1.6%	–	1.0%	0.2%
	150	*3.7%	*3.1%	–	*2.2%	2.8%
$(5, 100)$	50	0.8%	0.0%	0.2%	0.0%	0.0%
	100	*2.5%	0.4%	–	0.2%	0.0%
	150	*3.5%	*2.9%	–	0.4%	0.1%

Table 3: Average optimality gaps of different FSD formulations after ten minute time limit over five instances when starting with equal allocation solution. * in at least one instance a feasible solution not found within ten minutes when not given a starting solution. – LP relaxation did not solve within time limit.

		Heuristic			FSD1+SSD
(d,T)	N	# feasible	Time	Gap	Time
$(3,50)$	50	5	0.12	0.19%	0.6
	100	5	0.71	0.23%	7.2
	150	5	3.56	0.18%	32.9
$(3,100)$	50	5	0.14	0.03%	0.7
	100	5	0.67	0.13%	5.0
	150	5	2.51	0.10%	30.4
$(5,50)$	50	5	0.25	0.69%	1.6
	100	5	1.90	1.13%	11.7
	150	3	8.67	0.93%	69.7
$(5,100)$	50	5	0.28	0.14%	1.9
	100	5	1.60	0.23%	15.0
	150	5	7.15	0.36%	84.0

Table 4: Heuristic (§4.3) compared to LP relaxation of FSD1+SSD for five instances. Average solution times are in seconds. Average gap compares heuristic solution to SSD relaxation.

average solution time of the solved instances, and the final average optimality gaps. FSD2 is too large to be practical for these test instances, as many fewer instances are solved to optimality and even the LP relaxation often fails to solve within the time limit. The specialized branch-and-bound algorithm performs favorably compared to FSD1+SSD for the instances with $d = 3$, but solves fewer instances to optimality when $d = 5$. However, in general this branch-and-bound algorithm yields the lowest, or close to the lowest, average optimality gaps among the generally applicable approaches.

The performance of the branch-and-bound algorithm can be understood somewhat better by looking at the results of the heuristic used to find feasible solutions in this method. Table 4 presents these results for the first time the heuristic was called, which is immedately after the linear programming relaxation is solved. We report the number instances for which the heuristic found a feasible solution (under the '# feasible' heading), the average total time to run the heuristic (including the time spent solving the initial LP relaxation), and the average gap between the value of the heuristic solution and the value of the LP relaxation. For comparison purposes, we also show in this table the average time it takes formulation FSD1+SSD just to solve the LP relaxation. This table shows that the heuristic usually finds a feasible solution, and that already after this solution is found, the optimality gap is relatively small by just comparing this solution value to the LP relaxation value (which is equal to the value from an SSD-constrained problem). The total heuristic time is also less than the time to solve the LP relaxation of FSD1+SSD, which is significantly larger due to the constraints (14a) and (14b).

We close this section by commenting on the relative difficulty of the instances, as can be observed, e.g., by looking at the results of formulation FSD3. It appears that having more objectives in the stochastic dominance constraint ($d = 5$ compared to $d = 3$) makes the instances more difficult, which is intuitive because with more objectives the problem size increases and also it may be harder to find a solution that stochastically dominates the benchmark. Somewhat less intuitively, the instances with fewer projects ($T = 50$ compared to $T = 100$) also appear more difficult. However, having more projects likely makes it easier to find a solution that stochastically dominates the benchmark, which may translate to making the problem easier to solve.

5.2 IMRT Application

We use a radiation treatment planning problem to further illustrate the idea of optimizing with a multivariate stochastic dominance constraint. (The term IMRT, the most common form of radiation treatment, short for *intensity modulated radiation treatment*, is often used interchangeably with radiation treatment.) A radiation treatment plan is a selection of beam angles, beam apertures (i.e., shape of each beam's cross-section), and beam intensities for the radiation beams that, when delivered, will create a particular radiation dose profile in the patient. A good radiation treatment plan is one that creates a dose profile delivering a sufficient dose of radiation to the tumor (the target) while minimizing the dose received by healthy organs and tissues.

The standard formulation of the planning problem assumes that the beam angles are already chosen, leaving the choice of apertures and intensities as the decision variables. Then instead of modeling the aperture shape directly, the standard approach divides the beam opening into a grid of "beamlets" or "pencil-beams" whose intensities we can set independently. (A subsequent "leaf sequencing" step then determines how to achieve the chosen set of beamlet intensities using a small set of apertures.)

Let $\mathcal{S}$ be the set of "structures", regions of the patient's body for which we will calculate the dose profile. We then discretize the volume for each structure $s \in \mathcal{S}$ into a set $\mathcal{V}(s)$ of volume elements or "voxels". Let $A_s \in \mathbb{R}^{|\mathcal{V}(s)| \times B}$ be the "dose matrix" for structure s, where B is the number of beamlets: the dose delivered to voxel $v \in \mathcal{V}(s)$ from beamlet b at intensity one is $A_s(v,b)$. Each column of A_s represents the dose profile created by having a particular beamlet on at unit intensity. The decision variable is the vector of beamlet intensities, $\boldsymbol{x} \in \mathbb{R}_+^B$. Thus the dose to structure s is $A_s\boldsymbol{x}$.

Due to radiation treatment planning's multiple objectives (delivering a prescribed uniform dose to the target structure and little dose to the remaining structures), there is some variation in how these dose profiles are turned into optimization objectives and constraints. We define the objectives to be the mean-squared deviation from a prescription dose, $\|A_s\boldsymbol{x} - d_s^*\|_2^2 \,/\, |\mathcal{V}(s)|$, where d_s^* is the prescription dose for structure s. A traditional approach then minimizes the weighted sum of these objectives:

$$\min_{\boldsymbol{x} \geq 0} \sum_s \alpha_s \|A_s\boldsymbol{x} - d_s^*\|_2^2 \,/\, |\mathcal{V}(s)| \,, \tag{27}$$

where α_s is the weight given to achieving the objective for structure s. The performance of a particular plan is then often examined using a "dose-volume histogram" such as Figures 2c-d which show the fraction of each structure receiving more than a certain dose (they are analogous to cumulative distribution functions of the *spatial* distribution of the radiation dose).

So far our description of radiation treatment planning is completely deterministic. However, there are several important sources of uncertainty for the dose distribution delivered to the patient. Organ motion (organs, especially those in the lower abdomen, have a tendency to move around); differences in the patient geometry between the radiation delivery and the image taken prior to treatment planning; patient positioning on the treatment couch; shrinkage of the target over the course of treatment (radiation treatment is usually spread over many days); and organ motion during treatment (such as lung tumors while breathing) are all important problems and major sources of uncertainty (Li and Xing, 2000). Our goal is to incorporate such uncertainty into the model and to find a treatment plan that mitigates this uncertainty. We suppose that there are N scenarios for the possible patient geometry during treatment which we index by j and with scenario j occurring with probability $q_w(j)$. The uncertainty in the patient geometry requires that we refine the definition of the voxels $v \in \mathcal{V}(s)$ to remove ambiguity: do they refer to a location in space or rather to a particular physical piece of tissue? We choose to define the voxels by the volume discretization in the reference scenario $j = 1$. In other scenarios, $j \neq 1$, voxel $v \in \mathcal{V}(s)$ refers to the location of the piece of tissue that is in voxel v in the reference scenario; in every scenario voxel v refers to the same part of the anatomy, but the voxels only form a regular grid in the reference scenario $j = 1$. We then create dose matrices A_s^j for every scenario $j \in \mathcal{N}$. For each scenario j, we then define a vector of all the performance measures (i.e., multiple objectives) $\boldsymbol{w}^j := \{w_s^j\}_s$ with $w_s^j := \left\| A_s^j \boldsymbol{x} - d_s^* \right\|^2 / |\mathcal{V}(s)|$. We combine these scenarios into a random vector $\boldsymbol{W}$ of performance measures which takes value $\boldsymbol{w}^j$ in scenario j. For our optimization problem we place the first performance measure into the objective and require that the vector of performance measures dominates some performance baseline, $\boldsymbol{Y}$:

$$\begin{aligned} \min_{\boldsymbol{x} \geq 0, \boldsymbol{w}^j} \quad & \mathbb{E}[W_1] \\ s.t. \; & w_s^j \geq \left\| A_s^j \boldsymbol{x} - d_s^* \right\|^2 / |\mathcal{V}(s)| \quad \forall j \in \mathcal{N}, \; s \in \mathcal{S} \\ & -\boldsymbol{W} \succeq -\boldsymbol{Y}. \end{aligned} \tag{28}$$

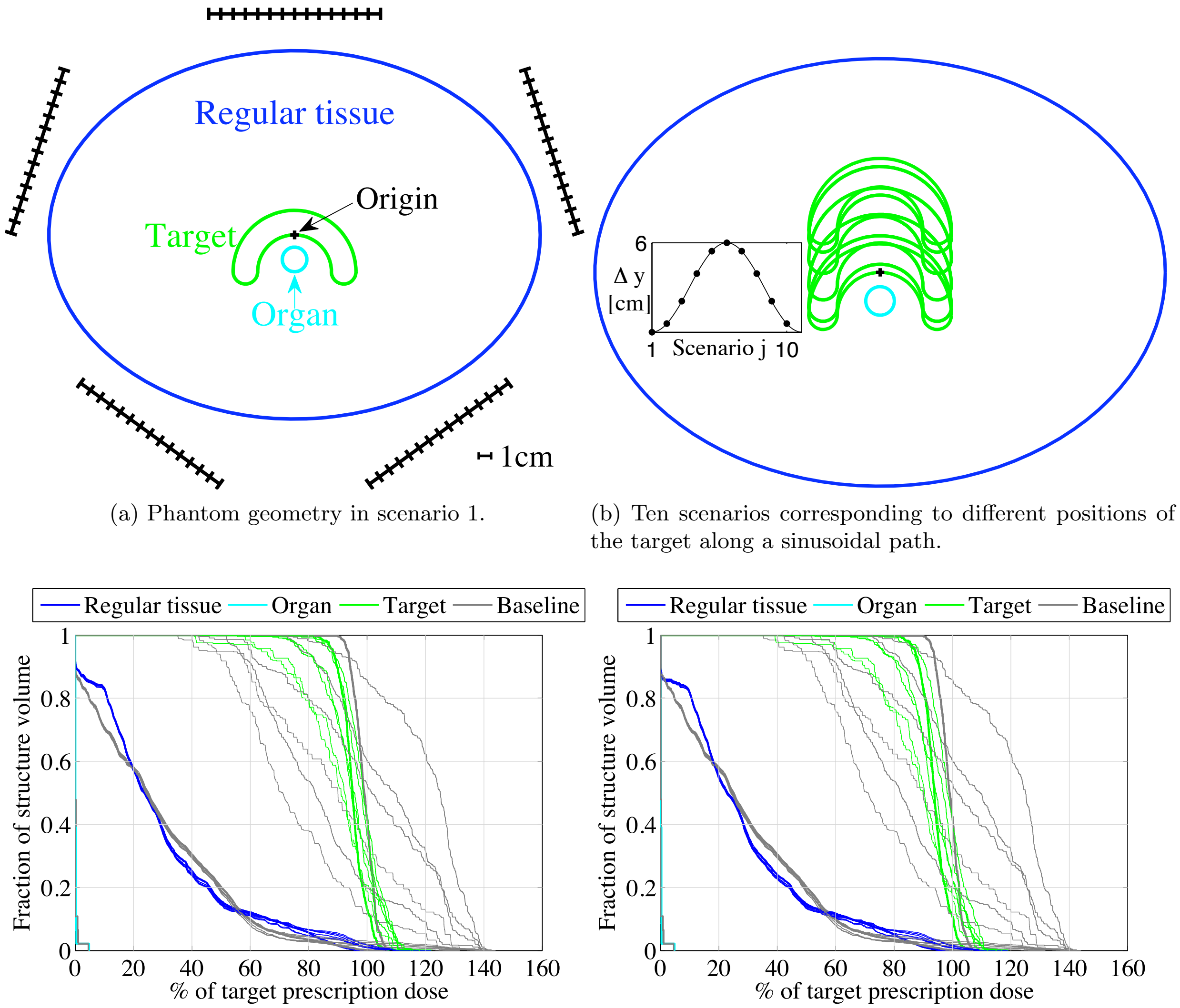
Here W_1 is the first component of $\boldsymbol{W}$, a scalar random variable, and thus a reasonable objective is to minimize $\mathbb{E}[W_1]$. The minus signs in the dominance constraint are because we seek to *minimize* the performance measures. To ensure convexity the constraint on w_s^j is relaxed, without loss of generality, to an inequality. A reasonable way to define the performance baseline $\boldsymbol{Y}$ is to define it as the performance of some baseline plan $\boldsymbol{x}_0$ under the same scenarios as we consider for the actual plan. Thus for scenario $j \in \mathcal{N}$, $\boldsymbol{Y} = \boldsymbol{y}^j$ where we define $y_s^j := \left\| A_s^j \boldsymbol{x}_0 - d_s^* \right\|^2 / |\mathcal{V}(s)|$. In our examples, we choose the baseline plan $\boldsymbol{x}_0$ to be the one that optimizes the deterministic formulation (27) of radiation treatment planning assuming the expected dose distribution; that is using the dose matrices $\mathbb{E}[A_s]$,

$$\boldsymbol{x}_0 \in \arg\min_{\boldsymbol{x} \geq 0} \sum_s (\alpha_s / |\mathcal{V}(s)|) \left\| \mathbb{E}[A_s] \boldsymbol{x} - d_s^* \right\|^2 . \tag{29}$$

In our example we work with a 2d phantom shown in Figure 2a. We are concerned about three structures: the target, an organ, and regular tissue, the tissue inside the outer blue line that is neither target nor organ. We consider $N = 10$ scenarios, all of whose probabilities are equal, where the difference among the scenarios is the location of the target. The various locations are superimposed in Figure 2b; they are equally spaced along a sinusoidal motion with a 3cm amplitude with the first scenario closest to the organ. We subdivide the phantom into a 0.5cm grid of voxels using the geometry of scenario 1. We use 5 beam directions equally spaced around the origin of the phantom as shown in Figure 2a, and we subdivide each beam direction into 14 beamlets with a 1cm cross-section. Thus there are 70 beamlets in total. The dose prescriptions and weights to weight the performance measures in the baseline (29) are $\alpha_{\text{normal}} = 1$, $\alpha_{\text{target}} = 15$, $\alpha_{\text{organ}} = 15$, $d^*_{\text{normal}} = 0$, $d^*_{\text{target}} = 100$, and $d^*_{\text{organ}} = 0$. The performance baseline $\boldsymbol{Y}$ is created as described above.

On a 2GHz MacBook with 8 GB RAM, the deterministic formulation (29) used to determine the performance baseline solved in 2 seconds using the Matlab optimization toolbox. Using Cplex 12.2, the SSDLP formulation took 3s, the FSD1+SSD formulation achieved an optimality gap of 14.49% in 20 minutes, the FSD2 formulation did not find a feasible solution in 20 minutes, and the FSD3 formulation achieved an optimality gap of 13.69% in 20 minutes. We did not try the specialized branch-and-bound algorithm for this example. The difficulty optimizing the FSD formulations may be due to the combination of binary and quadratic constraints.

Figures 2c-d show the dose-volume histograms of each structure for the solutions using an FSD constraint (solved to within 0.01% of optimal in 20,000 seconds) and using an SSD constraint,



(a) Phantom geometry in scenario 1.

(b) Ten scenarios corresponding to different positions of the target along a sinusoidal path.

(c) Using first-order stochastic dominance.

(d) Using second-order stochastic dominance.

Figure 2: An IMRT example using 5 equiangular beam directions each having 14 1cm beamlets. The dose is calculated on a 0.5cm grid. Parts (c)–(d) show the distribution of the dose: the fraction of the each structure's volume receiving more than a certain dose. The thin lines represent the different scenarios while the thicker lines correspond to the expected case.

respectively. In both cases we compare the optimal solution from the stochastic dominance constraint model, whose dose-volume histograms are colored, to the baseline plan, whose dose-volume histograms are grey. The figures use thin lines to show the dose-volume histogram for each scenario and structure, both for the solution to the stochastic formulation and for the baseline plan. To assist comparisons we also use thicker lines to show an aggregation of the various scenarios; the thicker lines are the dose-volume histograms generated by using the expected dose matrices, $\mathbb{E}[A_s]$. These dose-volume histograms are a standard way of displaying radiation treatment plans but they are not ideal for visualizing multivariate stochastic dominance (though we are not aware of *any* good way of visualizing multivariate stochastic dominance). The improvement of the optimal solutions on the baseline plan are mainly in the target where the green lines are closer to each other and fall more steeply than the grey lines of the baseline plan. This indicates that the radiation dose to the target is much more even (with less deviation from the prescription dose) under the optimal plan. For the organ, the optimized plans and the baseline plan look similar, and for regular tissue it is hard to say whether the optimized plans or the baseline plan is preferred. The solutions using first-order stochastic dominance and second-order stochastic dominance are quite similar, despite the more restrictive nature of the latter.

6 Concluding remarks

We have shown that choosing a different form of multivariate stochastic dominance to use as a constraint in a stochastic program allows us to obtain computationally tractable formulations. Specifically, we derived an LP formulation for an SSD constraint that is both of computational interest (it can be solved quickly) and theoretical interest (the constraint can be used in a convex program). For an FSD constraint we derive several MIP formulations and a specialized branch-and-bound algorithm. We also derived a MIP formulation, FSD3, for an important special case that numerical tests show can be solved quickly with a commercial MIP solver. We tested these formulations in two applications. The first application was a budget allocation problem in the face of multiple objectives, in which it is desired to find a solution that is preferred in a strong sense to a benchmark solution. In our second application we used a stochastic dominance constraint to handle an important problem in radiation treatment planning: how to account for the uncertainty resulting from organ motion. This problem has both uncertainty and multiple objectives (getting "enough" dose to the tumor and not "too much" to certain critical organs), and hence a formulation with a multivariate stochastic dominance constraint is helpful.

One limitation of our work is that the formulations we present require the random variables $\boldsymbol{W}$ and $\boldsymbol{Y}$ to have finite support, and their size grows with the size of the support. However, even if the support of $\boldsymbol{W}$ and $\boldsymbol{Y}$ are not small discrete sets, it may be possible to use sampling to obtain approximate solutions. This approach approximates the random variables $\boldsymbol{W}$ and $\boldsymbol{Y}$ by the more manageable random variables, $\boldsymbol{W}^{(n)}$ and $\boldsymbol{Y}^{(n)}$ (e.g., the empirical distribution from n i.i.d. samples). Then we solve the optimization problem (1) with $\boldsymbol{W}^{(n)} \succeq Y^{(n)}$ as the stochastic dominance constraint. This is motivated by Theorem 3.3.10 in Müller and Stoyan (2002): if $\boldsymbol{W}^{(n)} \to \boldsymbol{W}$ and $\boldsymbol{Y}^{(n)} \to \boldsymbol{Y}$ in distribution and $\boldsymbol{W}^{(n)} \succeq_1 \boldsymbol{Y}^{(n)}$ for all n, then $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. Theorem 3.4.6 in Müller and Stoyan (2002) gives a similar result applying to SSD. An important task for further research is to give a rigorous foundation for this sampling approach; specifically, to determine statistical estimates of precision when a finite sample is used and to obtain estimates of the number of samples necessary to obtain a desired accuracy. See Hu et al. (2009) for some results of this type for the case of positive linear stochastic dominance.

While the LP formulation of an SSD constraint is efficiently solved in our computational tests, the MIP formulations for an FSD constraint are significantly more difficult to solve, particularly for the formulations other than FSD3, which apply in the general case. This is partly due to the large size of these formulations. In future work, it may be interesting to seek an alternative formulation for the FSD constraint that is as strong as the ones we have proposed, but requires fewer variables. Such a formulation is likely to require exponentially many constraints, but may be efficiently solvable using a branch-and-cut approach.

Another potential drawback to our model, common to many stochastic dominance constraint formulations, is the restrictiveness of the stochastic dominance relation (i.e., requiring that the solution is preferred to the baseline under a large set of utility functions) and thus the set of feasible solutions that dominate the reference random variable may be unacceptably small (e.g., only the solution corresponding to the reference random variable). This problem is especially acute for FSD constraints and multivariate formulations because the former require dominance for all nondecreasing utility functions and the latter require dominance for many different ways of trading off the different components. However, this problem did not seem to occur with the IMRT application with the optimal solutions significantly improving on the the baseline and the solution of the more restrictive FSD constraint close to the solution with the SSD constraint. One also can view this not as an intrinsic limitation of stochastic dominance constraints but as a challenge for selecting appropriate reference random variables that reflect a decision-maker's risk avoidance preferences while still being weak enough to give a rich enough feasible set (of dominating solutions). Another possibility is to relax the strict dominance constraints to $\boldsymbol{W} \succeq \boldsymbol{Y} - \boldsymbol{c}$ by subtracting a fixed constant, $\boldsymbol{c}$, as we do in the budget allocation application, and then varying this constant to obtain a range of solutions.

Acknowledgments

The authors thank Tito Homem-de-Mello, Sanjay Mehrotra, and Jian Hu for many helpful comments.

References

D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566, 2003. URL http://link.aip.org/link/?SJE/14/548/1.

D. Dentcheva and A. Ruszczyński. Semi-infinite probabilistic optimization: first-order stochastic dominance constraints. *Optimization*, 53(5-6):583–601, 2004.

D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, 30:433, 2006.

D. Dentcheva and A. Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117(1):111–127, 2009. URL http://www.springerlink.com/content/834v6268r09575w9.

R. Gollmer, U. Gotzes, F. Neise, and R. Schultz. Risk modeling via stochastic dominance in power systems with dispersed generation. Technical Report 651, Department of Mathematics, University of Duisburg-Essen, 2007. URL http://www.uni-due.de/mathematik/preps07.shtml.

T. Homem-de-Mello and S. Mehrotra. A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *SIAM Journal on Optimization*, 20(3):1250–1273, 2009. URL http://link.aip.org/link/?SJE/20/1250/1.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Sample average approximation for stochastic dominance constrained programs. 2009. URL http://www.optimization-online.org/DB_FILE/2009/07/2337.pdf. Working paper.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Multi-criterion robust and stochastic dominance-constrained models with application to budget allocation in homeland security. 2010. URL http://www.optimization-online.org/DB_FILE/2010/04/2605.pdf. Working paper.

J. G. Li and L. Xing. Inverse planning incorporating organ motion. *Medical Physics*, 27(7):1573–1578, 2000. URL http://link.aip.org/link/?MPH/27/1573/1.

J. Luedtke. New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450, 2008. URL http://link.aip.org/link/?SJE/19/1433/1.

Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley, 2002.

N. Noyan, G. Rudolf, and A. Ruszczyński. Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters*, 34:653–659, 2006.

M. Shaked and J. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.

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

G. Whitmore and M. Findlay, editors. *Stochastic Dominance*. D.C. Heath and Company, Lexington, MA, 1978.

A Justification of Figure 1

The actual arrows are self explanatory. The diagram describes a partition of the space of pairs of random variables (W, Y). The following examples show that each piece of this partition is nonempty; that is, no arrows can be added to the diagram and all distinctions are necessary. The examples all have $N = M = d = 2$; equal probabilities on each scenario; and $\boldsymbol{y}^1 = [0; 0]$.

In the case $\boldsymbol{w}^1 = [8; 4]$, $\boldsymbol{w}^2 = [-2; 1]$, and $\boldsymbol{y}^2 = [-2; -5]$, we have $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [-4; 2]$, $\boldsymbol{w}^2 = [9; -3]$, and $\boldsymbol{y}^2 = [-8; -8]$, we have $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$, $\boldsymbol{W} \succeq_1^{lin} \boldsymbol{Y}$, but not $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [-5; 0]$, $\boldsymbol{w}^2 = [1; -1]$, and $\boldsymbol{y}^2 = [-10; -4]$, we have $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$, $\boldsymbol{W} \succeq_1^{comp} \boldsymbol{Y}$, but not $\boldsymbol{W} \succeq_1^{lin} \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [2; -1]$ $\boldsymbol{w}^2 = [7; 5]$, and $\boldsymbol{y}^2 = [9; -6]$, we have $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ but not $\boldsymbol{W} \succeq_1^{comp} \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [2; -2]$, $\boldsymbol{w}^2 = [8; 0]$, and $\boldsymbol{y}^2 = [8; -7]$, we have $\boldsymbol{W} \succeq_1^{lin} \boldsymbol{Y}$ but not $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [-7; 0]$, $\boldsymbol{w}^2 = [4; -1]$, and $\boldsymbol{y}^2 = [-7; -9]$, we have $\boldsymbol{W} \succeq_2^{lin} \boldsymbol{Y}$ and $\boldsymbol{W} \succeq_1^{comp} \boldsymbol{Y}$ but neither $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ nor $\boldsymbol{W} \succeq_1^{lin} \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [4; 1]$, $\boldsymbol{w}^2 = [5; 3]$, and $\boldsymbol{y}^2 = [9; 3]$, we have $\boldsymbol{W} \succeq_2^{lin} \boldsymbol{Y}$ but neither $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$ nor $\boldsymbol{W} \succeq_1^{comp} \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [-3; 0]$, $\boldsymbol{w}^2 = [3; 9]$, and $\boldsymbol{y}^2 = [-10; 7]$, we have $\boldsymbol{W} \succeq_1^{comp} \boldsymbol{Y}$ but not $\boldsymbol{W} \succeq_2^{lin} \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [-1; 8]$,

$\boldsymbol{w}^2 = [-2; 1]$, and $\boldsymbol{y}^2 = [-4; 6]$, we have $\boldsymbol{W} \succeq_2^{comp} \boldsymbol{Y}$ but neither $\boldsymbol{W} \succeq_2^{lin} \boldsymbol{Y}$ nor $\boldsymbol{W} \succeq_1^{comp} \boldsymbol{Y}$. In the case $\boldsymbol{w}^1 = [-5; -9]$, $\boldsymbol{w}^2 = [0; 5]$, and $\boldsymbol{y}^2 = [-6; -5]$, we don't even have $\boldsymbol{W} \succeq_2^{comp} \boldsymbol{Y}$.

B Proof of corollary 2

Suppose $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$, and let $\boldsymbol{W}'$ and $\boldsymbol{Y}'$ be as in Theorem 1. Let $p_{ij} = \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i, \boldsymbol{W}' = \boldsymbol{w}^j]$. Then, $p_{ij} \geq 0$, and

$$q_y(i) = \Pr[\boldsymbol{Y} = \boldsymbol{y}^i] = \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i] = \sum_{j \in \mathcal{N}} \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i, \boldsymbol{W}' = \boldsymbol{w}^j] = \sum_{j \in \mathcal{N}} p_{ij}$$

showing that (8b) holds. Similarly, (8c) holds. Finally, because $\boldsymbol{W}' \geq E[\boldsymbol{Y}'|\boldsymbol{W}']$ almost surely,

$$\begin{aligned}
\boldsymbol{w}^j \Pr[\boldsymbol{W} = \boldsymbol{w}^j] &\geq E[\boldsymbol{Y}'|\boldsymbol{W}' = \boldsymbol{w}^j] \Pr[\boldsymbol{W} = \boldsymbol{w}^j] \\
&= \sum_{i \in \mathcal{M}} \boldsymbol{y}^i \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i | \boldsymbol{W}' = \boldsymbol{w}^j] \Pr[\boldsymbol{W}' = \boldsymbol{w}^j] \\
&= \sum_{i \in \mathcal{M}} \boldsymbol{y}^i \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i, \boldsymbol{W}' = \boldsymbol{w}^j] = \sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij}
\end{aligned}$$

and so (8a) holds.

To prove the other direction, suppose we have p_{ij} for $i \in \mathcal{M}, j \in \mathcal{N}$ such that (8) hold. We now define a probability space with sample space $\Omega' := \mathcal{M} \times \mathcal{N}$, using the full power set of Ω' as the σ-field, and the probability measure $P'(i,j) := p_{ij}$. (This measure is well-defined since $p_{ij} \geq 0$ and summing (8b) over j shows $\sum_{ij} p_{ij} = 1$.) We then define random variables $\boldsymbol{W}'$ and $\boldsymbol{Y}'$ by $\boldsymbol{W}'(i,j) := \boldsymbol{w}^j$ and $\boldsymbol{Y}'(i,j) := \boldsymbol{y}^i$ for all $(i,j) \in \Omega'$. Then, by (8b)

$$\Pr[\boldsymbol{Y} = \boldsymbol{y}^i] = \sum_{j \in \mathcal{N}} p_{ij} = \sum_{(i,j) \in \Omega'} p_{ij} I(\boldsymbol{Y}' = \boldsymbol{y}^i) = \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i].$$

Similarly, $\Pr[\boldsymbol{W} = \boldsymbol{w}^j] = \Pr[\boldsymbol{W}' = \boldsymbol{w}^j]$ for all j, and hence $\boldsymbol{W}' \stackrel{\mathcal{D}}{=} \boldsymbol{W}$ and $\boldsymbol{Y}' \stackrel{\mathcal{D}}{=} \boldsymbol{Y}$. Finally, for any $j \in \mathcal{N}$,

$$\begin{aligned}
E[\boldsymbol{Y}'|\boldsymbol{W}' = \boldsymbol{w}^j] &= \sum_{i \in \mathcal{M}} \boldsymbol{y}^i \Pr[\boldsymbol{Y}' = \boldsymbol{y}^i | \boldsymbol{W}' = \boldsymbol{w}^j] \\
&= \sum_{i \in \mathcal{M}} \boldsymbol{y}^i \frac{\Pr[\boldsymbol{Y}' = \boldsymbol{y}^i, \boldsymbol{W}' = \boldsymbol{w}^j]}{\Pr[\boldsymbol{W}' = \boldsymbol{w}^j]} = \sum_{i \in \mathcal{M}} \boldsymbol{y}^i p_{ij} / q_w(j) \leq \boldsymbol{w}^j
\end{aligned}$$

by (8a) since $\Pr[W = \boldsymbol{w}^j] = \Pr[\boldsymbol{W}' = \boldsymbol{w}^j] > 0$. Hence $E[\boldsymbol{Y}'|\boldsymbol{W}'] \leq \boldsymbol{W}'$ almost surely. Thus we can apply Theorem 1 to prove that $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$.

C Proof of Theorem 7

Note that (25) is equivalent to (8a)–(8c). Thus claim 2 follows directly from Corollary 2. Focusing now on claim 1, assume the conditions hold. For each j, exactly one p_{ij} is nonzero, and that p_{ij} equals one. Thus (25c) implies (13) holds and hence $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. Now assume that $\boldsymbol{W} \succeq_1 \boldsymbol{Y}$. Thus, by Corollary 4, there exists p'_{ij} such that (8b)–(8d) and (13) hold. Thus, since $q_y(i) = q_w(j) = 1/N$, $P' := (Np'_{ij})$ is a doubly stochastic matrix, and the Birkhoff-von Neumann theorem states that there exists permutation matrices $Q^{(1)}, \ldots, Q^{(n)}$ and positive $\alpha_1, \ldots, \alpha_n$ such that $P' = \sum_k Q^{(k)} \alpha_k$. Since $\alpha_1 > 0$, $Q_{ij}^{(1)} > 0$ implies $p'_{ij} > 0$ and by (13) implies $\boldsymbol{w}^j \geq \boldsymbol{y}^i$. Since $Q^{(1)}$ is a permutation matrix, this implies that $\sum_{i \in \mathcal{M}} \boldsymbol{y}^i Q_{ij}^{(1)} \leq \boldsymbol{w}^j$ for all $j \in \mathcal{N}$. Thus the entries of $Q^{(1)}$ satisfy the conditions of claim 1.

Going on to claim 3, clearly $\{\boldsymbol{W} : \boldsymbol{W} \succeq_1 \boldsymbol{Y}\} \subseteq \{\boldsymbol{W} : \boldsymbol{W} \succeq_2 \boldsymbol{Y}\}$ and $\{\boldsymbol{W} : \boldsymbol{W} \succeq_2 \boldsymbol{Y}\}$ is convex. Suppose that $\boldsymbol{W} \succeq_2 \boldsymbol{Y}$. Then it remains to be shown that $\boldsymbol{W} \in \operatorname{conv}\{\boldsymbol{W} : \boldsymbol{W} \succeq_1 \boldsymbol{Y}\}$. Corollary 2 shows that there exists a doubly stochastic matrix P such that $\boldsymbol{w}^j \geq \sum_{i \in \mathcal{M}} \boldsymbol{y}^i P_{ij}$ for all $j \in \mathcal{N}$. By the Birkhoff-von Neumann theorem, there exists permutation matrices $Q^{(1)}, \ldots, Q^{(n)}$ and positive $\alpha_1, \ldots, \alpha_n$ such that $P = \sum_k Q^{(k)} \alpha_k$. Then for each k, $(\sum_{i \in \mathcal{M}} \boldsymbol{y}^i Q_{ij}^{(k)} : j)$ is a permutation of $(\boldsymbol{y}^i)$ and represents a random vector $\boldsymbol{Y}'_k \overset{\mathcal{D}}{=} \boldsymbol{Y}$. From the definition of P, $\boldsymbol{W} \geq \sum_k \alpha_k \boldsymbol{Y}'_k$, proving the claim.