Risk Adjusted Budget Allocation Models with Application in Homeland Security*

Jian Hu

Department of Industrial Engineering and Management Sciences

Northwestern University

jianhu@northwestern.edu

Tito Homem-de-Mello

Department of Mechanical and Industrial Engineering

University of Illinois at Chicago

thmello@uic.edu

Sanjay Mehrotra

Department of Industrial Engineering and Management Sciences

Northwestern University

mehrotra@iems.northwestern.edu

January 25, 2011

*The previous version of this paper is titled Multi-Criterion Robust and Stochastic Dominance-Constrained Models with Application to Budget Allocation in Homeland Security

Abstract

This paper presents and studies models for multi-criterion budget allocation problems under uncertainty. The proposed models incorporate uncertainties in decision maker's weights using a robust weighted sum approach. The risk averseness of the decision maker in satisfying random risk related constraints is ensured by using stochastic dominance. A sample average approximation approach together with a cutting surface method is used to solve this model. An analysis for the computation of statistical lower and upper bounds is also given. We use the proposed models to study the budget allocation to ten urban areas in the United States under the Urban Areas Security Initiative (UASI). Here the decision maker considers property losses, fatalities, air departures, and average daily bridge traffic as separate criteria. The properties of the proposed modeling and solution methodology are discussed using a RAND Corporation proposed allocation policy and the current government budget allocation as two benchmarks. The budget results are discussed under several parameter scenarios.

Key Words: Multi-Criteria, Resource Allocation, Homeland Security, Stochastic Programming, Stochastic Dominance, Robust Optimization, Risk

1 Introduction

Budget allocation problems arise frequently in the real world. Examples include finance, education, construction, transportation, ecology, telecommunication, etc. (see Phillips and Bana e Costa (2007); Paxson and Schady (2002); Karlaftis et al. (2007); Chauvenet et al. (2010); Lin et al. (2009)). Phillips and Bana e Costa (2007) describe five characteristics of the decision dilemma in such problems: (1) benefits are typically expressed by multiple objectives often in conflict; (2) decision makers cannot know every detail about a large number of given alternatives to make informed decisions; (3) individually optimal decisions rarely lead to the collectively best use of the available resources; (4) involved stakeholders cause dispute and competition; (5) those who disagree with the decisions overemphasize their own opinions in implementation. Uncertainty resulting from incomplete information is a major reason of misallocation of resources (see Dougherty and Psacharopoulos (1977); Al-Najjar et al. (2008)). An important example is the cost overrun risk in project management (see Lerche and Paleologos (2001); Flyvbjerg et al. (2003); Tseng et al. (2005); Miller et al. (2010)). The Association for Project Management for the United Kingdom defines the cost overrun as "the amount by which a contractor exceeds or expects to exceed the estimated costs and/or the final limitations (the ceiling) of a contract". Phillips and Bana e Costa (2007) highlight the need for multi-criterion approaches that enable decision makers to tradeoff the costs, risks, and multiple benefits, and to construct good investment portfolios across different areas so as to make the best use of limited resources. Typical methods applied to capital budgeting include the goal programming approach, the weighted sum approach, the multi-criterion decision analysis, and the analytic hierarchy process (see Steuer and Na (2003) for a survey).

In this paper we consider a class of budget allocation problems where the goal is to allocate funds to multiple sites according to multiple decision criteria. In its simplest form if there is only one decision criterion with a deterministic performance indicator for each site and a fixed overall budget, the problem is simple — the budget allocated to a site is proportional to its performance indicator. In the general case, though, there may be multiple evaluation criteria. In this situation, typically there is discrepancy (uncertainty) among multiple experts in giving relative weights to each evaluation criteria (see Hu and Mehrotra (2010) for a detailed discussion). Moreover, for each criterion, the indicators measuring a site performance may be subject to uncertainty. In this situation there is an additional uncertainty of specifying decision maker's risk aversion when making stochastic comparisons. Thus we have a multi-objective problem involving three types of uncertainties: (1) uncertainty in weights that would balance multiple decision criteria; (2) uncertainty in the performance indicators; and (3) uncertainty in decision maker's risk aversion.

Stochastic dominance (see e.g., Levy (2006), Sriboonchita et al. (2009), Shaked and Shanthikumar (1994) and Müller and Stoyan (2002)) provides a way to incorporate risk aversion by imposing constraints that ensure that any risk averse decision maker will prefer the outcome of the model depending on the solution to a benchmark. Optimization problems with stochastic dominance constraints are introduced in Dentcheva and Ruszczyński (2003, 2004). Recently, univariate stochastic dominance in the optimization models have been used to model finance, energy and transportation problems Karoui and Meziou (2006), Roman et al. (2006), Dentcheva and Ruszczyński (2006), Gollmer et al. (2007), Luedtke (2008), Nie et al. (2009). The notions of multivariate positive linear second order stochastic dominance is studied in Dentcheva and Ruszczyński (2009). The concept of multivariate positive linear second order stochastic dominance is extended to that of a multivariate polyhedral second order stochastic dominance ($\mathcal{P}$-dominance) in Homem-de-Mello and Mehrotra (2009). Homem-de-Mello and Mehrotra (2009) also develop a cutting surface method for solving the corresponding optimization problems. A sample average approximation technique for solving such problems having continuously distributed random parameters is developed in Hu et al. (2010).

The budget models presented in this paper use a multi-criteria robust weighted sum technique (McRow) (see Hu and Mehrotra (2010)) together with the concept of $\mathcal{P}$-dominance to address the three types of uncertainties described above. More specifically, the models have the following features: (1) a robust weighted sum objective that allows the decision makers to suggest a set of trade-off weights of criteria and then optimizes the worst case over the set; (2) risk-averse constraints using the concept of $\mathcal{P}$-dominance into the robust model to address uncertainty in describing a decision maker's utility function (degree of risk aversion). We study the characteristics of these modeling features in the application to a budget allocation problem in homeland security.

The problem of budgeting for homeland security is a decision problem with multiple objectives. For instance, Haimes (2004) lists four objectives for these problems, involving protection of (1) critical cyber-physical infrastructures, (2) economic structures, (3) organizational-societal infrastructure and the continuity of the government operations, as well as (4) human lives and individual property, liberty, and freedom. The decision maker is faced with the dilemma of balancing monetary expenditure (losses) for threat prevention with potential threat to loss of human life and social values. The models presented in the literature often measure consequences of the disasters in monetary terms. However, it is difficult to convert human losses into monetary terms precisely, and these estimates may vary widely depending on the perspective one takes. A further complicating fact is that the assessment of each of these criteria separately may be inaccurate, which leads to a risk of budget misallocation due to parameter uncertainty. Hence, we have a situation where the budgeting decision is to be made for multiple objectives where each of the objectives is stochastic, which fits the framework of our models described above.

The remainder of the paper is organized as follows. In Section 2, we propose prototypes of our models. In Section 3, we present an approach for solving these models using the Sample Average Approximation (SAA) technique. Section 4 applies the models to a budget allocation problem in homeland security. We study these models for the budget allocation to ten main urban areas in the United States under the Urban Areas Security Initiative (UASI). The models are studied with a RAND corporation proposed allocation policy as a benchmark, and the average government budget allocation (over five years) as a benchmark. In Section 5, we numerically analyze the performance of the algorithms and methods by which we solve these models. Section 6 discusses various cases with different parameter settings in these models and studies the characteristics of these models. The results suggest that the government current budget allocation are consistent with those suggested by our models. Appendix A has severed tables and figures discussed in Section 5 and 6. Appendix B provides proofs for several results stated in the text. Appendix C further analyzes the sensitivity of important parameters in these models.

2 Definitions and Models

2.1 The Budget Allocation Problem

We introduce some notation for the class of budget allocation problems we consider. Let n denote the number of sites among which we need to allocate the budget, and let m denote the number of criteria used to evaluate the performance of each site. The performance indicator of site j with respect to criterion i is given by a random quantity A_{ij}, which is subject to uncertainty. Thus, we have an $(m \times n)$ random matrix A such that $A_{ij}'s$ are (possibly dependent) nonnegative random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ with the property that $\sum_j A_{ij} = 1$ a.s.

for all i.

We denote the percentage of the budget allocated to site j by x_j, which is a decision variable. More generally, the decision vector is represented by $x \in \mathcal{X} \subseteq \mathbb{R}^n$. Given a particular allocation x, the *budget misallocation* with respect to criterion i is measured by a function $M_i(x, A)$. Examples of such functions are the absolute deviation $\sum_{j=1}^{n} |A_{ij} - x_j|$, the semi-deviation $\sum_{j=1}^{n} (A_{ij} - x_j)_+$ (where $(\cdot)_+ := \max\{\cdot, 0\}$), etc. Other nonlinear examples are also possible. We will restrict our selection to the semi-deviation metric because of the limitation of the current optimization solver technology available. The function $M_i(x, A)$ may also be viewed as a regret (penalty) function. If the budget is allocated to site j is in proportion to or greater than the performance indicator A_{ij}, then there is no regret. Otherwise, the regret increases linearly with amount of shortfall in proportional allocation with respect to this performance indicator value. The semi-deviation metric is also called the cost overrun risk in project management (e.g. see Tseng et al. (2005) for discussions on the cost overrun). Note that the quantity $M_i(x, A)$ is random, since A is random. We assume that $M_i(x, A)$ is integrable, so the multiple decision criteria are given by a random vector $M(x, A) \in \mathcal{L}_1^m$ ($\mathcal{L}_1^m$ is the space of integrable mappings from the underlying probability space to $\mathbb{R}^m$) . For simplicity, we also assume that we want to minimize the misallocation corresponding to each of the decision criteria. Thus, it is important to define appropriate ways to formulate this stochastic optimization problem, which we do next.

2.2 Multi-Objective Robust Models with Dominance

2.2.1 Stochastic Dominance

Before we describe the model, we review notions of stochastic dominance, which is used to compare risk in our model. A random variable $X \in \mathcal{L}_1$ stochastically dominates another random variable $Y \in \mathcal{L}_1$ in the concave second order, written $X \trianglerighteq_{(\mathrm{icv})} Y$, if $E[u(X)] \geq E[u(Y)]$ holds for all increasing concave utility functions $u(.)$ for each $E[u(X)]$ and $E[u(Y)]$ are finite. This notion is appropriate to model risk averseness when larger values are preferable to smaller ones (e.g. gains). When the opposite holds — for example, if the random variables measure losses — then risk averseness corresponds to increasing convex functions. In that case X is preferable to Y (denoted $X \trianglelefteq_{(\mathrm{icx})} Y$) if $E[u(X)] \leq E[u(Y)]$ for all increasing convex u. It is easy to show that $X \trianglelefteq_{(\mathrm{icx})} Y$ if and only if $-X \trianglerighteq_{(\mathrm{icv})} -Y$. In words, $X \trianglelefteq_{(\mathrm{icx})} Y$ if any risk-averse decision maker prefers the outcomes given by X to the outcomes given by Y.

We use the concept of $\mathcal{P}$-dominance as follows. A random vector $X \in \mathcal{L}_1^m$ is preferable to another random vector $Y \in \mathcal{L}_1^m$ with respect to a given non-empty polyhedral set $\mathcal{P} \subseteq \mathbb{R}^m$ if

$$w^T X \trianglelefteq_{(\text{icx})} w^T Y \quad \text{for all } w \in \mathcal{P}. \tag{2.1}$$

No generality is lost if we assume $\|w\|_1 = 1$. We will also assume that $w \geq 0$ since the components of w represent the importance of each criterion. In our computational study we assume that $\mathcal{P}$ represents the convex hull of a collection of weights. In our context the assumption that $\mathcal{P}$ is given by a finite number of extreme points is reasonable as specific weights may correspond to opinion weights of different decision makers. It is possible to consider other convex sets, such as an ellipsoid (see Hu et al. (2010); Hu and Mehrotra (2010)), however it adds to the complexity of the model.

2.2.2 A Robust Optimization Model with Stochastic Dominance Constraints

Let v^d, $d = 1, \dots a$, be the weights suggested by the decision makers, and $\mathcal{P}$ be the convex hull of $v^1, \dots, v^a$. Let $y^k = (y_1^k, \dots, y_n^k)$, $k = 1, \dots, q$ be the given benchmarks. The robust risk adjusted budget allocation model (Model I) is given as follows:

$$\min_{x \in \mathcal{X}} \max_{w \in \mathcal{P}} \; E[w^T M(x, A)], \tag{2.2a}$$

$$\text{s.t. } w^T M(x, A) \trianglelefteq_{(\text{icx})} w^T M(y^k, A), \quad \forall w \in \mathcal{P},\; k = 1, \dots, q. \tag{2.2b}$$

The objective function in Model I provides a way to handle multi-objective problems with uncertainty. While this objective ensures that the optimal solution given by the model is better than the benchmark in terms of the expectation (assuming the benchmark is a feasible solution) of the weighted sum of misallocation values, it is possible that the proposed solution is riskier than an available benchmark. Decision makers often have implicit or explicit benchmarks that they would like to compare against the solutions given by a model. Using $\mathcal{P}$-dominance, constraints (2.2b) ensure the prerequisite that a feasible solution be less risky than the provided benchmarks. Nevertheless, Model I without dominance constraints (2.2b) is a valid optimization model for budget allocation which may have independent value. We represent this model as (Model I′):

$$\min_{x \in \mathcal{X}} \max_{w \in \mathcal{P}} \; E[w^T M(x, A)]. \tag{2.3}$$

The increasing convex dominance requirement in (2.2b) is consistent with the situation where the degree of risk aversion of decision makers is unknown. The model ensures that the expected negative impact (utility) of the budget misallocation proposed by our model is less than that of the benchmarks for all increasing convex functions. The requirement of $\mathcal{P}$-dominance above is weaker than a requirement that $E[u(M(x, A))] \leq E[u(M(y^k, A))]$ for all increasing convex functions $u : \mathbb{R}^m \to \mathbb{R}$, which is thought as very conservative in practice (see Zaras and Martel (1994) and Nowak (2004)). In fact, the homeland security case study model in Sections 4-6 becomes infeasible if the notion of positive linear second order dominance is used. Note that the positive linear second order dominance is a relaxation of general second order dominance just described. Hence, the homeland security case study also has no feasible solution when constrained by general second order dominance. Alternative approaches of defining vector dominance are also given in Zaras and Martel (1994) and Nowak (2004) under an additive and probability independence condition. Under a component-wise probability independence assumption, Huang et al. (1978) show that the multivariate dominance reduces to the univariate dominance between corresponding components. If additive independence condition holds, Keeney and Raiffa (1976) prove that a multidimensional utility function can be formulated as a weighted sum of one-dimensional utility functions. Our proposed approach here does not make any such assumptions, thus weakening the dominance requirements. Moreover, our approach appears algorithmically and computationally more tractable as seen from the computational results in Section 5. Note that although we use the same $\mathcal{P}$ in defining the objective function and constraints in Model I, it would be possible to use different sets for objective and constraints if desired.

3 Model Reformulations and Algorithmic Analysis

In this section we study solution methods for Models I and I$'$. Throughout we will assume that $\mathcal{P}$ is given by its extreme points. This allows us to produce more practical bounds in our analysis when the given number of extreme points is small. The situations where $\mathcal{P}$ is given by linear equality and inequality constraints is more general. The analysis in Hu et al. (2010) is applicable for such situations, and it can be adapted for the models in this paper.

3.1 Model I Reformulation

We first present a reformulation of Model I that is used for later algorithmic development and analysis. From (Müller and Stoyan, 2002, Theorem 1.5.7) constraints (2.2b) are equivalent to

$$E[(w^T M(x, A) - \eta)_+] \leq E[(w^T M(y^k, A) - \eta)_+], \quad (\eta, w) \in \mathbb{R} \times \mathcal{P},\ k = 1, \ldots, q. \tag{3.1}$$

Hence, the following model is equivalent to Model I.

$$\min_{x \in \mathcal{X}} z \tag{3.2a}$$

$$\text{s.t. } E[v^{d^T} M(x, A)] \leq z, \quad d = 1, \ldots, a, \tag{3.2b}$$

$$E[(w^T M(x, A) - \eta)_+ - (w^T M(y^k, A) - \eta)_+] \leq 0, \quad (\eta, w) \in \mathbb{R} \times \mathcal{P},\ k = 1, \ldots, q. \tag{3.2c}$$

Note that (3.2) has uncountably many expected value constraints. This presents some difficulties in developing algorithms for computing statistical lower and upper bounds and establishing global convergence properties. Building upon our previous work (Hu et al. (2010)) we present an algorithm and convergence results later in this section for such problems. Our analysis uses several assumptions which are discussed below. In order to successfully generate an upper bound in our method we need to assume that the feasible set of (3.2) has a nonempty interior. Unfortunately it can be shown that model (3.2) never has a nonempty interior. Hence, we perturb the constraints in (3.2c) by a small constant $\iota > 0$. If no upper bound is required by the user, then no such perturbation is needed. With this in mind, we consider the model:

$$\min_{x \in \mathcal{X}} z \tag{3.3a}$$

$$\text{s.t. } E[v^{d^T} M(x, A)] \leq z, \quad d = 1, \ldots, a, \tag{3.3b}$$

$$E[(w^T M(x, A) - \eta)_+ - (w^T M(y^k, A) - \eta)_+] \leq \iota, \quad (\eta, w) \in \mathbb{R} \times \mathcal{P},\ k = 1, \ldots, q. \tag{3.3c}$$

In the following we develop an algorithm for (3.3) under the following additional assumptions:

(A1). $\mathcal{X}$ is a compact set and there exists a constant $D \geq 0$ such that $\mathcal{X} \subseteq \{x \in \mathbb{R}^n : \|x\|_\infty \leq D\}$.

(A2). There exists a constant $K \geq 0$ such that, for all $x \in \mathcal{X}$, $\|M(x, A)\|_\infty \leq K$, a.s..

(A3). $M(., A)$ is convex and uniformly Lipschitz continuous on $\mathcal{X}$ a.s.. In particular, there exists a constant $L \geq 0$ such that $\|M(x, A) - M(y, A)\|_\infty \leq L\|x - y\|_\infty$ holds for all $x,\ y \in \mathcal{X}$ a.s..

Later in Section 5 we will see that in our budget allocation problem the constants of D, K and L in Assumptions (A1) - (A3) are known. Assumption (A2) implies that $|w^T M(x, A)| \leq \|w\|_1 \|M(x, A)\|_\infty \leq K$ for all $x \in \mathcal{X}$. Thus, constraints (3.3c) are further equivalent to

$$E[(w^T M(x, A) - \eta)_+] \leq E[(w^T M(y^k, A) - \eta)_+] + \iota, \quad (\eta, w) \in [-K, K] \times \mathcal{P},\ k = 1, \ldots, q. \quad (3.4)$$

Also since $\|v^d\|_1 \leq 1 (d = 1, \ldots, a)$, we can restrict z to the interval $[-K, K]$, hence the feasible region of (3.3) is compact.

3.2 Sample Average Approximation

We now study the convergence properties of the sample average approximation (SAA) of Model 3.3. The SAA method approximates the expected value function with an average of Monte Carlo samples to formulate an approximation of the original problem. The analysis presented here follows similar steps to the one in Hu et al. (2010). However, here the results are presented under the assumption that $\mathcal{P}$ is given by a set of vertices instead of a set of linear inequalities. We now discuss the SAA approximation of (3.3). Let

$$\begin{aligned} F^d(x, A) &:= {v^d}^T M(x, A), \\ G^k(x, \eta, w, A) &:= (w^T M(x, A) - \eta)_+ - (w^T M(y^k, A) - \eta)_+ - \iota, \end{aligned}$$

for $d = 1, \ldots, a$ and $k = 1, \ldots, q$. Generate N independent and identically distributed samples as $A^1, \ldots, A^N$. Using the integrands, denote the expected value functions and their sample average approximations as $f^d(x) := E[F^d(x, A)]$, $f_N^d(x) := \frac{1}{N}\sum_{j=1}^N F^d(x, A^j)$, $g^k(x, \eta, w) := E[G^k(x, \eta, w, A)]$, and $g_N^k(x, \eta, w) := \frac{1}{N}\sum_{j=1}^N G^k(x, \eta, w, A^j)$. Given $\epsilon \geq 0$, denote

$$\begin{aligned} S^\epsilon := &\{(z, x) \in [-K, K] \times \mathcal{X} : f^d(x) - z \leq \epsilon,\ g^k(x, \eta, w) \leq \epsilon, \\ &d = 1, \ldots, a,\ k = 1, \ldots, q,\ (\eta, w) \in [-K, K] \times \mathcal{P}\}, \quad (3.5) \\ S_N^\epsilon := &\{(z, x) \in [-K, K] \times \mathcal{X} : f_N^d(x) - z \leq \epsilon,\ g_N^k(x, \eta, w) \leq \epsilon, \\ &d = 1, \ldots, a,\ k = 1, \ldots, q,\ (\eta, w) \in [-K, K] \times \mathcal{P}\}. \quad (3.6) \end{aligned}$$

The sample average approximation of (3.3) is written as

$$\min_{(z,x) \in S_N^0} z. \quad (3.7)$$

3.2.1 Convergence

We now study the asymptotic convergence of the feasible region of model (3.7) with increasing sample size. Let us consider the probability of an event $\{S^{-\epsilon} \subseteq S_N^0 \subseteq S^{\epsilon}\}$ for a given $\epsilon > 0$. The following theorem presents a large deviation type result for our problem. This theorem shows that as the sample size N increases, the sample average approximation set S_N^0 is contained in, and it contains ϵ perturbations of the original feasible set (i.e., $S^{-\epsilon}$ and S^{ϵ}) with a probability that converges to 1 exponentially fast. The result is very similar to Theorem 3.2 in Hu et al. (2010) but there are some important differences due to the different assumptions made — for example the inequalities in Theorem 1 hold for any $\epsilon > 0$ whereas the corresponding results in Hu et al. (2010) are valid only for sufficiently small ϵ. A proof is presented in Appendix B.2.

Theorem 1 *Suppose Assumptions (A1) - (A3) hold. Given $\epsilon > 0$, it follows that*

$$P(S^{-\epsilon} \subseteq S_N^0 \subseteq S^{\epsilon}) \geq 1 - 2(a+q)\left(\frac{\max\{D,K,1\}\max\{2mK,4L\}}{\epsilon}\right)^{m+n+1} e^{-\frac{N\epsilon^2}{8(K+\iota)^2}}.$$

In particular, given $\beta \in [0,1]$, if

$$N \geq \frac{8(K+\iota)^2}{\epsilon^2}\ln\left\{\frac{2(a+q)}{\beta}\left(\frac{\max\{D,K,1\}\max\{2mK,4L\}}{\epsilon}\right)^{m+n+1}\right\},$$

then $P(S^{-\epsilon} \subseteq S_N^0 \subseteq S^{\epsilon}) \geq 1-\beta$.

3.2.2 Cut Generation Algorithm

SAA model (3.7) is a semi-infinite problem. Its feasible region S_N^0 is an intersection of constraints, $g_N^k(x,\eta,w) \leq 0$, which must be satisfied for all (η, w) in an uncountable set $[-K,K]\times\mathcal{P}$. Homem-de-Mello and Mehrotra (2009) show that there exists a finite subset of $[-K,K]\times\mathcal{P}$ with which a finite number of constraints can be constructed to equivalently represent the uncountable stochastic dominance constraints.

Algorithm 1 extends the cutting surface approach proposed by Homem-de-Mello and Mehrotra (2009) to solve model (3.7). Step 1 of this algorithm solves a relaxation of model (3.7), over a subset of the feasible region S_N^0. We build a heuristic method in Step 2. Let V be a finite subset of $\mathcal{P}$. At a solution of $(\hat{z},\hat{x})$, we denote

$$i^{kv} := \arg\min_{i=1,\dots,N} -g_N^k(\hat{x}, v^T M(y^k, A^i), v) \tag{3.9}$$

Algorithm 1 Cut-Generation Algorithm

0. Let $l = 0$ and choose arbitrary finite sets $V^l \subset P$ and $W^l \subset \{1, \ldots, q\} \times [-1, 0] \times \mathcal{P}$. Denote

$$\begin{aligned}\widetilde{S}_N(W^l) :=& \{(z, x) \in [-K, K] \times \mathcal{X} : f_N^d(x) \le z,\ d = 1, \ldots, a,\\ & g_N^k(x, \eta, w) \le 0,\ (k, \eta, w) \in W^l\}.\end{aligned}$$

1. Find an optimal solution $(\hat{z}, \hat{x})$ of
$$\min_{(z,x)\in\widetilde{S}_N(W^l)} z, \tag{3.8}$$
or if (3.8) is infeasible, exit.

2. For $k = 1, \ldots, q$ and each $v \in V^l$,
   if optimal value of (3.9) $\phi^{kv} < 0$, $W^l = W^l \cup \{k, v^T M(y^k, A^{i^{kv}}), v\}$.
   If W^l is unchanged, go to Step 3; otherwise, go to Step 1.

3. For $i = 1, \ldots, N$ and $k = 1, \ldots, q$,
   solve the problems (3.10), let w^{ik} and ψ^{ik} be a optimal solution and objective function value;
   if $\psi^{ik} < 0$, $W^l = W^l \cup \{(k, {w^{ik}}^T M(y^k, A^i), w^{ik})\}$ and $V^l = V^l \cup \{w^{ik}\}$.

4. If W^l is unchanged, exit; otherwise, let $W^{l+1} = W^l$, $V^{l+1} = V^l$, and $l = l + 1$, go to Step 1.

for $k = 1, \ldots, q$ and each $v \in V$. It is known that the constraints $g_N^k(x, \eta, v) \le 0$ for some $v \in \mathcal{P}$ and all $\eta \in [-K, K]$ are equivalent to the constraints $g_N^k(x, v^T M(y^k, A^i), v) \le 0$ for $i = 1, \ldots, N$ and $k = 1, \ldots, q$ (see Proposition 3.2 in Dentcheva and Ruszczyński (2003)). Therefore, if $\phi^{kv} := -g_N^k(\hat{x}, v^T M(y^k, A^{i_{kv}}), v) < 0$, we generate a valid cut $g_N(x, v^T M(y^k, A^{i^{k,v}}), v) \le 0$ for $(\hat{z}, \hat{x})$. If the heuristic method generates some new cuts, this algorithm returns back to Step 1; otherwise, it goes on with the next step. In Step 3, we consider the separation problems ($i = 1, \ldots, N$ and $k = 1, \ldots, q$) for the solution $(\hat{z}, \hat{x})$

$$\min_{w\in\mathcal{P}} -g_N^k(\hat{x}, w^T M(y^k, A^i), w). \tag{3.10}$$

If all the subproblems (3.10) have a non-negative objective function value, the algorithm stops with an optimal solution of model (3.7). Otherwise, we have a solution $\hat{w}$ of model (3.10), for some i and k, with a negative value. Using this solution we generate a valid cut $g_N(x, \hat{w}^T M(y^k, A^i), \hat{w}) \le 0$ for $(\hat{z}, \hat{x})$.

Theorem 2 shows that this algorithm will stop after a finite number of iterations. The theorem follows directly from the results in Homem-de-Mello and Mehrotra (2009) but we include it here for completeness and give the proof in Appendix B.3.

Theorem 2 *Algorithm 1 terminates after a finite number of steps with either an optimal solution* *to model* (3.7) *or a proof of infeasibility of model* (3.7)*.*

Algorithm 1 needs to solve N subproblems (3.10), which are polyhedral DC (Difference of Convex Functions) programming problems minimizing a difference of two convex polyhedral functions over a polyhedral set. Homem-de-Mello and Mehrotra (2009) also present an integer programming formulation of (3.10). We use this reformulation (see Problem (4.5)) in our computation.

3.2.3 Statistical Bounds

Since solving a large scale stochastic program to optimality is expensive and often not necessary, it is useful to have techniques for computing statistical lower and upper bounds for such problems. For optimization problems with stochastic dominance constraints Hu et al. (2010) developed such techniques in a general setting. Below we adapt their ideas for computing statistical lower and upper bounds for the $\mathcal{P}$-dominance constrained model (3.3).

Statistical Lower Bound

Using an initial sample group with a small sample size N_0, Algorithm 1 stops with a master program with a finite subset of $\{1,\ldots,q\}\times[-K,K]\times\mathcal{P}$ generated in Steps 2 and 3. Let $\widetilde{W}_{N_0}$ denote that subset. Let $\hat{\nu}(k,\eta,w)$ $((k,\eta,w)\in\widetilde{W}_{N_0})$ be the optimal dual solutions for the dominance constraints in $\widetilde{S}_{N_0}(\widetilde{W}_{N_0})$ and $\hat{\mu}_d$, $d=1,\ldots,a$, be those of the non-dominance constraints $f^d_{N_0}(x)-z\le 0$. We now generate additional i.i.d. M_l groups of samples of size N_l each. Using these samples, we define

$$\varphi^j_{N_0,N_l} := \min_{x\in\mathcal{X}} \sum_{d=1}^{a}\hat{\mu}_d f^{d,j}_{N_l}(x) + \sum_{(k,\eta,w)\in\widetilde{W}_{N_0}} \hat{\nu}(k,\eta,w) g^{k,j}_{N_l}(x,\eta,w), \tag{3.11}$$

for $j=1,\ldots,M_l$. Here $f^{d,j}_{N_l}(\cdot)$ and $g^{k,j}_{N_l}(\cdot)$ are the sample average approximation of $f^d(\cdot)$ and $g^k(\cdot)$ with the j^{th} sample group. (3.11) excludes z since the sum of all $\hat{\mu}_d$ is equal 1 by KKT condition. Following Hu et al. (2010) we now compute a statistical lower bound for (3.3) in the following proposition.

Proposition 1 *Denote* $\bar{\varphi}_{N_0,N_l,M_l} := \frac{1}{M_l}\sum_{j=1}^{M_l}\varphi^j_{N_0,N_l}$ *and* $\bar{\sigma}^2_{N_0,N_l,M_l} := \frac{1}{M_l(M_l-1)}\sum_{j=1}^{M_l}(\varphi^j_{N_0,N_l} - \bar{\varphi}_{N_0,N_l,M_l})^2$*. Then,*

$$L_{N_0,N_l,M_l} := \bar{\varphi}_{N_0,N_l,M_l} - t_{\alpha,M_l-1}\bar{\sigma}_{N_0,N_l,M_l}$$

is a $100(1-\alpha)\%$ *confidence lower bound for the optimal value of* (3.3)*.*

Statistical Upper Bound

We now present a method for generating an upper bound for (3.3). The method extends the method in Hu et al. (2010). Note that constraints $f^d(x) - z \le 0$ and $f_N^d(x) - z \le 0$ have no contribution to the feasible regions of Model I and its SAA problem. Hence, only considering the $\mathcal{P}$-dominance constraints, we can first use the line search algorithm in Hu et al. (2010) with an initial sample size N_0 to obtain a solution x_{N_0} which is feasible to (3.3) with probability $100(1-\alpha_1)\%$ for a given $\alpha_1 \in (0,1)$. We state that algorithm in Appendix B.1 for completeness. Next, we construct statistical upper bound for $\max_{d\in\{1,\ldots,a\}} f^d(x)$ at x_{N_0}, which is also a probabilistic upper bound for (3.3). With N_u independent samples, we denote $(\tilde{\sigma}^d_{N_u}(x))^2 := \frac{1}{N_u-1}\sum_{j=1}^{N_u}(F^d(x,A^j) - f^d_{N_u}(x))^2$. Using the method in Mak et al. (1999), we obtain

$$U^d_{N_0,N_u} := f^d_{N_u}(x_{N_0}) + t_{\alpha_2,N_u-1}\widetilde{\sigma}^d_{N_u}(x_{N_0}), \tag{3.12}$$

which is a $100(1-\alpha_2)\%$ confidence upper bound for $f^d(x_{N_0})$ for $\alpha_2 \in (0,1)$. Thus, we have a statistical upper bound for (3.3) in the following proposition.

Proposition 2

$$U_{N_0,N_u} := \max_{d\in\{1,\ldots,a\}} U^d_{N_0,N_u}$$

is a $100(1-\alpha_1)(1-\alpha_2)\%$ *confidence upper bound for the optimal value of* (3.3).

Given a desired significance level $100(1-\alpha_1)\%$ for the upper bound, we can choose α_1 and α_2 such that $1-\alpha = (1-\alpha_1)(1-\alpha_2)$. Note that the effort required to obtain a probabilistic feasible solution is much bigger than the effort to compute the approximation of the objective function value. Since higher confidence levels yield wider confidence intervals — which in turn requires the use of larger sample sizes to keep the interval width under a desired tolerance — we recommend choosing $1-\alpha_2$ close to one in order to keep $1-\alpha_1$ as low as possible (although of course $1-\alpha_1 > 1-\alpha$).

4 Homeland Security Budget Allocation Case Study

Budget allocation problems are fundamentally multi-objective problems where multiple views have to be reconciled to achieve organizational goals. In this section we present an application of our proposed modeling techniques in the context of budgeting for homeland security, an area that has become centrally important since the terrorist attacks of September 11, 2001. As catastrophes are unpredictable and lead to unwanted consequences, the Department of Homeland Security (DHS) has endeavored to use risk management to help guide the spending on prevention, response, and recovery from such national catastrophes (see Caudle (2005), Wright et al. (2006), and Reifel (2006)). DHS funds national urban areas through the Urban Area Security Initiative (UASI). The total budget under this initiative for 2009 was $799 million. This budget was allocated to 62 urban areas with 60% of the money allocated to the 10 highest risk urban areas. The problem of risk adjusted budget allocation is central to terrorism and security risk reduction under the Urban Areas Security Initiative (Brody 2010). The Federal Emergency Management Agency collects information on area risks (Metropolitan Statistical Areas and States/Territories) to evaluate their individual risk profiles. This risk assessment information together with project proposals are used to make urban area funding decisions. Unfortunately, currently these decisions are ad hoc and lead to dissatisfaction among those affected by the allocated budgets (Dahl 2010).

An approach to the budget allocation problems that mitigates the impact of disasters within the limited resources is desirable. Ideally, appropriate budget allocation should be made to achieve best possible risk reduction. The process of target risk estimation (for example, see Azaiez and Bier (2007) for a reliability based approach) is beyond the scope of this paper. For our purposes we take the target risk estimators as given but assume that there is some imprecision in the values. Note that, since vulnerable targets have different types of loss profiles such as population density (resulting in greater fatalities) versus infrastructure value (resulting in greater monetary losses), it is difficult to equate such criteria with a consensus weight vector.

Decision methodologies such as scenario analysis (see Willis and Kleinmuntz (2007)), reliability and failure analysis (see Azaiez and Bier (2007))), game theory (see Powell (2007) and Zhuang and Bier (2007)), and cost-benefit analysis (see Farrow and Stuart (2009)) have been proposed for homeland security budget allocation (see also Parnell et al. (2006)). Willis et al. (2005) presented an optimization model to minimize the expected potential error resulting from underestimating funding share for the urban areas under UASI. Bier et al. (2008) proposed a game theoretic approach for allocation strategy among the 10 highest risk urban areas. All of these proposed techniques do not take into account the multi-objective nature of the problem, and assume that the risk parameters are estimated accurately.

We now give a prototype model for UASI budget allocation. Recall that in our context we will take A to be a $(m \times n)$ stochastic matrix such that $A_{ij} \geq 0$, $\sum_{j=1}^{n} A_{ij} = 1$. The (i, j) entry of the matrix A represents, for each criterion i, the proportion of losses for urban area j relative to the total losses for all urban areas. The randomness of A reflects the fact that the losses cannot be predicted exactly. The matrix A is called the *risk share matrix*. In the RAND Corporation report Willis et al. (2005) constructs a budget misallocation function as $\sum_{j=1}^{n}(A_{ij} - x_j)_+^2$. for criterion i. Under the i^{th} criterion the goal is to allocate budget to the j^{th} target according to risk share A_{ij}. Allocations under the risk share are penalized, but are not penalized if they are over the risk share.

Similar to Willis et al. (2005), our budget misallocation function is constructed by the cost overrun risk

$$M_i(x, A) := \sum_{j=1}^{n}(A_{ij} - x_j)_+. \tag{4.1}$$

Note that we are not squaring the budget overrun function. This allows us to generate linear subproblems. Using the function $M(\cdot)$ defined in (4.1), we solve two versions of the homeland security budgeting problem — a robust unconstrained model (Model I′), and a stochastic dominance-constrained version (Model I) that uses the government and RAND allocations as benchmarks.

4.1 Homeland Security Budgeting Model Analysis

We now study the characteristics of the budget misallocation function $M(.)$. The special form of (4.1) allows us to get practical values of algorithmic parameters for our case here. Since budget allocation problems can be formulated in terms of percentages, we can take the feasibility set $\mathcal{X}$ as $\{x \in \mathbb{R}_+^n : \|x\|_1 \leq 1\}$, hence Assumption (A1) holds with $D = 1$. We also have the following proposition (see Appendix B.4 for proof), showing that Assumptions (A2) and (A3) hold with $K = 1$ an $L = n$.

Proposition 3 *The budget misallocation function* (4.1) *has the following properties:*

(1) For all $w \in \mathcal{P}$ (where $\mathcal{P}$ is the set defined in (2.1)*) and $x \in \mathbb{R}_+^n$, $0 \leq w^T M(x, A) \leq 1$, and $\|M(x, A)\|_\infty \leq 1$ a.s.*

(2) $M(., A)$ is convex a.e and uniformly Lipschitz continuous on $\mathbb{R}_+^n$. In particular, for any risk share matrix A,

$$\|M(x, A) - M(y, A)\|_\infty \leq n\|x - y\|_\infty$$

holds for all $x,\ y \in \mathbb{R}_+^n$.

By Theorem 1 in Homem-de-Mello and Mehrotra (2009), we rewrite model (3.7) as a linear program by introducing the intermediate variables r^{ijlk} and t^{jsk}:

$$\min z \tag{4.2a}$$

$$\text{s.t. } z \geq \frac{1}{N}\sum_{j=1}^{N}\sum_{s=1}^{m}\sum_{h=1}^{n} v_s^d t^{jsh}, \quad d = 1, \ldots, a \tag{4.2b}$$

$$\frac{1}{N}\sum_{j=1}^{N} r^{ijlk} \leq \frac{1}{N}\sum_{j=1}^{N}({w^{ilk}}^T M(y^k, A^j) - {w^{ilk}}^T M(y^k, A^i))_+ + \iota, \tag{4.2c}$$

$$i = 1, \ldots, N,\ k = 1, \ldots, q,\ l = 1, \ldots, b_{ik},$$

$$r^{ijlk} \geq \sum_{s=1}^{m}\sum_{h=1}^{n} w_s^{ilk} t^{jsh} - {w^{ilk}}^T M(y^k, A^i), \tag{4.2d}$$

$$i = 1, \ldots, N,\ j = 1, \ldots, N,\ k = 1, \ldots, q,\ l = 1, \ldots, b_{ik},$$

$$t^{jsh} \geq A_{sh}^j - x_h, \quad j = 1, \ldots, N,\ s = 1, \ldots, m,\ h = 1, \ldots, n, \tag{4.2e}$$

$$r^{ijlk} \geq 0, \quad i = 1, \ldots, N,\ j = 1, \ldots, N,\ k = 1, \ldots, q,\ l = 1, \ldots, b_{ik}, \tag{4.2f}$$

$$t^{jsh} \geq 0, \quad j = 1, \ldots, N,\ s = 1, \ldots, m,\ h = 1, \ldots, n, \tag{4.2g}$$

$$\sum_{h=1}^{n} x_h \leq 1, \tag{4.2h}$$

$$x_h \geq 0, \quad h = 1, \ldots, n. \tag{4.2i}$$

Here, w^{ilk} ($l = 1, \ldots, b_{ik}$) are the w-components of the vertex solutions of P_{ik} in (B.7) which are finite in number, i.e., b_{ik} is finite. We solve problem (3.8) in Algorithm 1 by its linear reformulation given in (4.2). Recall that we need the optimal dual solutions in the construction of the statistical lower bound. Let us now consider the relationship between the optimal dual solutions of (3.7) and (4.2). Denote vectors μ, ν, and τ as the nonnegative multipliers of constraints (4.2b) — (4.2g). μ and ν respectively correspond to constraints (4.2b) and (4.2c) and τ is for other constraints. Thus, we construct the Lagrangian function of (4.2) on the feasible region $\mathcal{X}$ as:

$$\begin{aligned}\phi(z,x,t,r,\mu,\nu,\tau) := z &+ \mu^T \left[\frac{1}{N}\sum_{j=1}^{N}\sum_{s=1}^{m}\sum_{h=1}^{n} v_s^d t^{jsh} - z\right] \\ &+ \nu^T \left[\frac{1}{N}\sum_{j=1}^{N} r^{ijlk} - (w^{ilk^T} M(y^k, A^j) - w^{ilk^T} M(y^k, A^i))_+ - \iota\right] \\ &+ \tau^T \begin{bmatrix} \sum_{s=1}^m \sum_{h=1}^n w_s^{ilk} t^{jsh} - w^{ilk^T} M(y^k, A^i) - r^{ijlk} \\ A_{sh}^j - x_h - t^{jsh} \\ -r^{ijlk} \\ -t^{jsh} \end{bmatrix}. \end{aligned} \tag{4.3}$$

Also, define the Lagrangian of (3.7) as:

$$\psi(z,x,\mu,\nu) := z + \mu^T \left[f_N^d(x) - z\right] + \nu^T \left[g_N^k(x, w^{ilk^T} M(y^k, A^i), w^{ilk})\right]. \tag{4.4}$$

The following theorem links the Lagrangian functions $\phi(\cdot)$ and $\psi(\cdot)$. A proof is given in Appendix B.5.

Theorem 3 *If $(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau})$ is a saddle point for $\phi(\cdot)$, then $(\hat{z}, \hat{x}, \hat{\mu}, \hat{\nu})$ is a saddle point for $\psi(\cdot)$.*

If $(\hat{z}, \hat{x}, \hat{t}, \hat{r})$ is an optimal solution of linear program (4.2) and $(\hat{\mu}, \hat{\nu}, \hat{\tau})$ is the corresponding optimal dual solution, then $(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau})$ is a saddle point of $\phi(\cdot)$. Theorem 3 shows that $(\hat{\mu}, \hat{\nu})$ is an optimal dual solution of (3.7).

We solve the separation subproblems (3.10) in Algorithm 1 by an equivalent binary mixed integer program (see Homem-de-Mello and Mehrotra (2009)):

$$\min \; \frac{1}{N}\sum_{j=1}^{N} u_j - g_j \tag{4.5a}$$

$$\text{s.t. } (w,u) \in \mathcal{P}_{ik}, \tag{4.5b}$$

$$g_j - h_j = w^T M(y^k, A^j) - w^T M(\hat{x}, A^i), \quad j = 1, \ldots, N, \tag{4.5c}$$

$$\alpha_j b_j \geq g_j, \; \beta_j(1 - b_j) \geq h_j, \quad j = 1, \ldots, N, \tag{4.5d}$$

$$g_j \geq 0, \; h_j \geq 0, \; b_j \in \{0, 1\}, \quad j = 1, \ldots, N, \tag{4.5e}$$

where $\alpha_j := \big(\max_{w\in\mathcal{P}} w^T M(y^k, A^j) - w^T M(\hat{x}, A^i)\big)_+$ and $\beta_j := \big(-\min_{w\in\mathcal{P}} w^T M(y^k, A^j) - w^T M(\hat{x}, A^i)\big)_+$. The coefficients α_j, β_j together with the binary variable b_j are introduced in (4.5) to ensure that only one of the variables g_j and h_j is positive at a feasible solution of (4.5). Note that Algorithm 1 requires solving $2qN^2$ linear programs in each iteration to obtain all α_j and β_j. As the sample size increases, this procedure is costly. Using Proposition 3, we have $-1 \leq w^T M(y^k, A^j) - w^T M(\hat{x}, A^i) \leq 1$ for $i = 1, \ldots, N$, $j = 1, \ldots, N$, and $k = 1, \ldots, q$. Thus, it suffices to let α_j and β_j be 1 for all j in (4.5). Although this relaxation may result in an increase of computation time to solve each mixed integer problem, it brings net savings by avoiding all the calculations of α_j's and β_j's.

4.2 Data Generation for the Model

We now describe the data we use in the computational testing of our model. The raw data is obtained from Willis et al. (2005) and Bier et al. (2008). It involves four risk indicators: property losses, fatalities, air departures, and average daily bridge traffic. The air departures and daily bridge traffic are considered critical infrastructure. Although Willis et al. (2005) also provide data for injuries, we did not include it here since the data for fatalities and injuries are strongly correlated. The results presented here are for the 10 major urban areas of the United States (see Table 2). The same 10 areas were used in Bier et al. (2008). These areas received 40% of the total UASI budget in 2004 and 60% of the total UASI budget in 2009. The number of air departures in these urban areas is roughly a third of the total departures in the nation (see Bier et al. (2008)).

Construction of Risk Share Matrix

Using the RMS (Risk Management Solutions) Terrorism Risk Model, Willis et al. (2005) produce three perspectives on terrorist threats: a standard, reduced, and enhanced threat outlook. The details of the RMS model are not available. However, it is suggested that this model considers foreign terrorist organizations and domestic threat groups and different beliefs about the terrorist motivations and capabilities. The data from RMS terrorism risk model is given in Table 2 under columns "standard", "reduced", and "increased" risk.

Using the property losses, fatalities, air departures, and average daily bridge traffic data the risk share matrix A is estimated as follows. Denote a random risk index matrix as R for which element R_{ij} is the risk value for j^{th} city under i^{th} risk metric. We regard property losses R_{1j} and fatalities R_{2j} as random variables with equally likely outcomes, i.e. a probability of 1/3 is assigned to each scenario. The air departures are denoted by R_{3j} and the average daily bridge traffic by R_{4j}. In practice, these are random quantities as they change from one day to the other. We consider the number in Table 2 as the expected value of air departures and bridge traffic. In our computation we assume that they are random variables with log-uniform distributions. The log-uniform distribution has been used in Willis et al. (2005) to simulate uncertain parameters. We describe R_{ij} (for $i = 3$ and 4) as follows:

$$R_{ij} = t_{ij}\gamma^U, \tag{4.6}$$

where U is a random variable uniformly distributed on $[-1, 1]$, $\gamma > 1$, and t_{ij} are some constants. By taking the expectation of R_{ij}, t_{ij} can be expressed as

$$t_{ij} = \frac{2E[R_{ij}]\gamma \ln \gamma}{\gamma^2 - 1}. \tag{4.7}$$

The parameter γ controls the volatilities of the random air departure and bridge traffic. The random R_{ij} are converted to risk shares A_{ij} as

$$A_{ij} = \frac{R_{ij}}{\sum_{j=1}^{n} R_{ij}}. \tag{4.8}$$

Construction of Benchmarks.

We use two benchmarks in Model I for the DHS UASI budget allocation problem. The first benchmark is based on previous DHS UASI budget allocations. The second benchmark is based on the allocation suggested in the RAND report by Willis et al. (2005). The goal here is to find a budget allocation that stochastically dominates the budget allocations from the government and the RAND report. These benchmark budget allocations were constructed as follows. Table 3 shows the annual percentage of the budget allocations of the UASI Funding in the latest five fiscal years. Their average gives the government benchmark. We choose the allocation recommendation from Willis et al. (2005) according to the property loss criterion. For the 10 areas this allocation was computed using the property loss data in Table 2, and the budgets were allocated using optimal solution of $\min\{\sum_{j=1}^{n} E[(A_{1j} - x_j)_+^2] : \sum_{j=1}^{n} x_j \leq 1,\ 0 \leq x_j \leq 1,\ j = 1, \ldots, n.\}$ as suggested by Willis et al. (2005). Allocations to some urban areas in these two benchmarks differ significantly. The biggest difference is for New York, which is 58.61% - 31.93% = 26.68% of the budget amounting to \$2.364 billion over 5 years. The average difference across the urban areas in the government and RAND allocations is 6.74% of the budget.

Construction of Weight Region.

We now describe the approach we use in our numerical tests for constructing the weight region $\mathcal{P}$. Let the number of vertices of $\mathcal{P}$ be equal to the number of criteria m. For a given center $c \in \mathbb{R}^m_+$ with $\|c\|_1 = 1$, choose $\theta \in [0, 1]$ such that $\frac{\theta}{m-1} \leq \min_{j=1,\dots,m} c_j$ and then assign v^d_d $(d = 1, \dots, m)$, the d^{th} element of vertex v^d of $\mathcal{P}$, by $c_d + \theta$ and v^d_j by $c_j - \frac{\theta}{m-1}$ for any j not equal to d. The center represents an initial choice of the importance of each criterion and our choice for $\mathcal{P}$ allows for perturbations of c while keeping the norm equal to 1. In the case of the homeland security problem, there are four criteria, hence $m = 4$. In our tests, we use three centers: (1/4, 1/4, 1/4, 1/4) equally treating each criterion, (1/8, 1/8, 3/8, 3/8) emphasizing the air departures and bridge traffic, and (3/8, 3/8, 1/8, 1/8) giving more consideration to the property losses and fatalities. To simplify the statement, the first center is called Equality-Center, the second is called Infrastructure-Center, and the last is called Property-Fatality-Center. A larger θ means a larger weight region. For Model I′ this results in a solution that is best for the worst possible weight from a larger uncertainty set. For Model I the use of a larger weight region requires finding a solution that is preferable to the benchmarks for risk averse decision makers.

5 Numerical Performance of the Stochastic Dominance Budget Allocation Model

In this section, we discuss the numerical performance of Models I′ and I. The tests are developed with Visual C++ 2008 and CPLEX 12 on Window Vista. The test computer is a ThinkPad laptop with 2.20 GHz Inter T7500 Duo Core CPU. We fix the constant ι in (3.3c) as 0.005. Our test cases first analyze the complexity of the cut generation algorithm. We give the number of iterations and generated cuts at different sample sizes. Also, running times are recorded for the entire procedure and particularly for Step 3 of Algorithm 1. Next, we discuss the results for the statistical lower and upper bounds. The small gaps show that the sample average approximation (3.7) has a good convergence behavior on these problems. We finally study a batch sampling method, which is a good way to reduce the variance of the optimal value and solution of (3.7).

5.1 Computing Time

We ran Model I′ with sample sizes 100, 200, and 300 respectively, locating the center at Equality-Center and choosing the variability parameter γ and the weight region size parameter θ as (γ, θ) =(3, 0.15). The corresponding computing times are 0.534, 2.262, and 5.758 seconds suggesting that Model I′ is easy to solve for a reasonable sample size. Model I is solved by Algorithm 1. Table 4 records the numbers of iterations, generated cuts in Steps 2 and 3, and nodes used by CPLEX 12 to solve all mixed-integer linear programs (4.5) — reformulation of subproblems (3.10) — and computing time in seconds. Column 'Total Time' gives the total time used by Algorithm 1 and Column 'Time of Step 3' gives the running time for Step 3 which entails solving (4.5).

In the test, we set W^0 (the set of generated cuts) to be empty and V^0 (the set used in the heuristic method of Step 2) to include all the vertices of weight region $\mathcal{P}$. The time required to perform Steps 1 and 2 of Model I is less than one minute, even for the largest sample size. However, the total computational time of Model I is significantly greater and it grows quickly with the sample size. In particular, these times are dominated by the time required to solve the mixed integer programs in Step 3. We also observe that while solving the cut generation sub-problem the number of nodes required by CPLEX 12 increase with the sample size. Interestingly, however, in our test problems all added cuts are generated using the vertices of $\mathcal{P}$, which are known to us ahead of the time. The availability of these cuts can be explored without solving the separation problems (Step 2). It is known, however, that such a property is not true in general, as pointed out by the example in Homem-de-Mello and Mehrotra (2009). Therefore, Step 3 is needed to verify that no additional cuts are available. It is worthwhile observing that, as long as there is at least one binding constraint in the original problem for some $w \in \mathcal{P}$, the optimal value of subproblem (3.11) is less than or equal to zero. This explains why Step 3 is so time consuming - the mixed integer problem (4.5) must be solved to optimality to ensure that no $w \in \mathcal{P}$ yields a strictly negative objective function value in (3.11).

The observation that CPLEX 12 uses 92 percent of the total time to solve the mixed integer programs in (4.5) suggests the importance of developing efficient methods for generating cuts from (3.10), and quickly verifying when no such cuts are available. Such an investigation is currently in progress.

5.2 Statistical Lower and Upper Bounds

Let us now discuss the results in Table 5 for obtaining the statistical lower and upper bounds for Model I which we described in Section 3.1. The sampling parameters for building the lower and upper bounds are given as $N_0 = 50$, $N_l = 1000$, $M_l = 20$, and $N_u = 500,000$. We use Equality-Center for the test results discussed in this section. All the gaps between the lower and upper bounds are very small. The biggest difference is only 0.0084 when comparing 99% bounds for the case that (γ, θ) is equal to $(3, 0.25)$. This case also provides the maximum relative difference 2.51% relative to the value of the lower bound. The small gaps show that the SAA problem (3.7) has a good convergence behavior for Model I. Note that the method does not calculate bounds in some cases which appear to be either infeasible or have very small feasible regions. These are shown by '—' in the table. We also give the times in seconds for computing the 95% confident lower and upper bounds. The computing times for these lower bounds are relatively stable. The computing times for these upper bounds seem to increase with the variability (γ) and the size of the weight region (θ).

Let x_N be the x-component of the optimal solution of (3.7) when $(\gamma, \theta) = (3, 0.15)$. We estimate $\max_{w \in \mathcal{P}} E[w^T M(x_N, A)]$, the objective function value of Model I at x_N, with independent samples of large size 500,000. Table 6 lists the estimated objective values for five independent sample groups with sample size N. When N is 25 or 50, two of the five values are out of the range between lower and upper bounds in Table 5, which suggests that in these cases the optimal solution is not achieved. Moreover, the estimated objective values show some fluctuation. When N is up to 75 or 100, the five values become stable and all fall in the range between the bounds, indicating that the true optimal solution has likely been found.

5.3 Batch Sampling for Optimal Solution Estimators

We now discuss the use of Batch Sampling Method (BSM) to obtain more accurate estimators of optimal solutions. Solving (3.7) for M independent sample groups with size N each, we obtain the respective optimal solutions (z_N^j, x_N^j) for $j = 1, \ldots, M$. Let $\bar{z}_N := \frac{1}{M}\sum_{j=1}^{M} z_N^j$ and $\bar{x}_N := \frac{1}{M}\sum_{j=1}^{M} x_N^j$. BSM takes the batch mean $(\bar{z}_N, \bar{x}_N)$ as the optimal solution of (3.7). Note that $\bar{z}_N$ is also an estimator of the optimal value. Birge (2009) uses numerical tests to show that BSM can give better results with a smaller sample size, i.e. M batches of N samples each give better solutions than a single batch of MN samples.

We choose $M = 20$ and $N = 100$ in the test. In the case $(\gamma, \theta) = (3, 0.25)$, compared to the sample means (budget allocation values in the range [2.42% - 49.27%]), the sample variances are very small ($\leq 10^{-4}$). More specifically, the ratios of sample standard deviation over the sample mean are (1.39%, 3.70%, 3.70%, 5.92%, 4.26%, 2.73%, 7.82%, 7.12%, 12.39%, 7.20%). This result suggests two conclusions: (i) the optimal solution of (3.7) appears to be unique; (ii) it is possible and reasonable to analyze model (3.3) by the batch means of the optimal values and solutions of (3.7) for reducing the variance.

6 Discussion on Budget Allocations

We now study the budget allocations recommended by Models I′ and I. Recall that Model I′ is same as Model I but it has no constraints. The behaviors of these models are analyzed at different settings of the parameters: the center location, the size of weight region, the variability of air departures and bridge traffic, and the conditional probability relating property losses and fatalities. Recall that the center determines the levels of importance for property losses, fatalities, air departures, and bridge traffic; θ means the size of weight region $\mathcal{P}$; γ contributes to the variability of the uncertain air departures and bridge traffic; π affects the correlation between property losses and fatalities. We let $\pi = 1/2$ in the following analyses. except for the results in the sensitive analysis for π in Section C.3.

6.1 A Base Case

Let us first discuss the case with Equality-Center and $(\gamma, \theta) = (3, 0.25)$. We call this our base case. Figure 1a gives the optimal values and solutions of Models I′ and I. We also give the government and RAND benchmarks in Columns (i) and (ii). When choosing the two benchmarks as the solutions, the corresponding objective values are 0.3454 and 0.3921 respectively. In comparison, Model I′ has the lowest objective function value 0.3163 and Model I has an objective function value between Model I′ and the benchmarks. Let $\hat{x}$ be the optimal solution of Model I. By the definition of $\mathcal{P}$-dominance and convexity of M, we have $E[w^T M(\hat{x}, A)] \leq E[w^T M(y^k, A)]$ for all $w \in \mathcal{P}$ and $k = 1$, 2. It follows that $\max_{w \in \mathcal{P}} E[w^T M(\hat{x}, A)] \leq \max_{w \in \mathcal{P}} E[w^T M(y^k, A)]$, which verifies that the optimal objective function value of Model I is below those values given by the benchmarks.

The optimal allocation of Model I′ is close to the government benchmark. Indicated by Column (iii) of Figure 1a and Table 1, the maximum difference is 6.77% in Los Angeles-Long Beach and the

Table 1: Comparison of the benchmarks and Outputs of Models I′ and I

Urban Area	The Government Benchmark (%)	The RAND Benchmark (%)	Model I′* (%) Solution	Model I′* (%) Absolute Variation	Model I′* (%) Relative Variation	Model I** (%) Solution	Model I** (%) Absolute Variation	Model I** (%) Relative Variation
New York	31.93%	58.61%	33.06%	1.14%	3.56%	49.27%	9.35%	15.95%
Chicago	10.42%	16.39%	14.36%	3.94%	37.79%	12.82%	3.57%	21.78%
Bay Area	6.83%	7.91%	8.00%	1.17%	17.17%	6.89%	1.02%	12.88%
Washington, DC-MD-VA-WV	12.94%	5.06%	7.56%	5.38%	41.58%	6.63%	1.58%	31.25%
Los Angeles-Long Beach	15.24%	4.95%	8.48%	6.77%	44.39%	6.62%	1.67%	33.63%
Philadelphia, PA-NJ	4.19%	2.42%	4.36%	0.17%	4.07%	3.32%	0.90%	37.28%
Boston, MA-NH	3.74%	2.11%	7.04%	3.31%	88.55%	4.39%	2.28%	108.02%
Houston	5.88%	1.21%	6.29%	0.41%	6.99%	3.77%	2.56%	210.69%
Newark, NJ	6.60%	0.66%	6.69%	0.10%	1.48%	3.86%	3.20%	485.52%
Seattle-Bellevue-Everett	2.24%	0.67%	4.15%	1.91%	85.46%	2.42%	1.75%	259.81%

* The absolute variation of Model I′ measures the difference between the optimal solution of Model I′ and the government benchmark. The relative variation is the ratio of the absolute variation over the government benchmark.
** The absolute variation of Model I measures the difference between the optimal solution of Model I and the RAND benchmark. The relative variation is the ratio of the absolute variation over the RAND benchmark.

average difference is 2.43%. Note that the absolute and relative variations of Model I′ in Table 1 measure the difference between the optimal solution of Model I′ and the government benchmark and those of Model I depend on the RAND benchmark. The maximum difference between Model I′ and the RAND benchmark is 25.55% for New York and the average difference is 5.51%.

The maximum difference between the recommendation of Model I and the RAND benchmark is 9.35% for New York and 2.79% on average. It implies that constraints (2.2b) with the government benchmark are much weaker than those with the RAND benchmark, which pulls the final decision away from the solution of Model I′ toward the RAND benchmark. This suggests that Model I′ does not give sufficient attention to the risk control that the decision makers require in the RAND benchmark.

Moreover, Table 1 gives the relative variations of Models I′ and I with respect to the government and RAND benchmarks respectively. The two largest relative differences happen in Boston, MA-NH (88.55%) and Seattle-Bellevue-Everett (85.46%) for Model I′. The two areas together are allocated 6.98% of the total budget by the government benchmark. For Model I, they are Newark, NJ (485.52%) and Seattle-Bellevue-Everett (259.81%). They are only allocated 1.33% of the total budget by the RAND benchmark.

6.2 Role of the Center

We now compare the results of Models I′ and I for different centers. Figure 1b gives the optimal values and solutions of the models at Infrastructure-Center and Figure 1c shows those at Property-Fatality-Center. We have mentioned that Equality-Center equally treats the four criteria, Infrastructure-Center emphasizes protecting the infrastructure, and Property-Fatality-Center gives more emphasis to property losses and fatalities.

At Infrastructure-Center, Model I′ still gives an optimal solution closer to the government benchmark. The maximum difference is 6.75% in Los Angeles-Long Beach and the average difference is 2.36%. However, Model I′ supports the RAND benchmark at Property-Fatality-Center. Compared to the RAND benchmark, the maximum difference is 4.52% in New York and the average difference is 1.78%. These results suggest that the government benchmark is somehow inclined to defending the infrastructure against terrorist attacks. As we mentioned, the RAND benchmark is constructed with property losses. Model I′ obviously agrees to the RAND benchmark when Property-Fatality-Center is used to underline the importance of property losses and fatalities. When we are seeking a robust solution over a sufficiently large $\mathcal{P}$, Model I$'$ is insensitive to the centers. Let us further consider some cases in which we fix $\mathcal{P} \equiv \{w \in \mathbb{R}^m_+ : \|w\|_1 = 1\}$, i.e., the cases provide robust optimal solutions over the largest possible weight region. To illustrate the robustness with respect to the weight region, Table 7 gives the optimal solutions of Model I$'$ for different values of the variability factor γ and the maximum and average differences between the optimal solutions and the government benchmark. The results are rather stable at about 6.45% for the biggest differences and 2.52% for the average.

Infrastructure-Center reduces the impact of property losses and fatalities by giving more emphasis on air departures and bridge traffic. As a result, the dominance constraints lose their impact, and consequently Model I yields the same solution as Model I$'$ (see Figure 1b). A small weight for property loss directly relaxes the dominance constraints from the RAND benchmark. Also, the dominance constraints function weakly at Property-Fatality-Center (see Figure 1c). The optimal allocation of Model I is a little closer to the RAND benchmark than that of Model I$'$; however, the maximum difference of the allocations given by Models I$'$ and I is only 1.11% in New York. Since Model I$'$ for Property-Fatality-Center has highlighted property losses and fatalities, the RAND benchmark obtained from property losses does provide much additional contribution. On the other hand, it is for the same reason that the emphasis on property losses and fatalities weakens the dominance constraints from the government benchmark.

It is important to note that the functions of the objective and the dominance constraints are partially overlapped in their purpose in Model I. Both of them work to control the budget misallocation.

7 Conclusions

This paper has proposed a modeling approach for multi-criterion budget allocation problems under uncertainty, using recently developed concepts of multi-objective robust optimization and multivariate linear second order stochastic dominance constraints. We have applied our modeling approach to real data to determine a budget allocation to the ten main urban areas in the United States. This approach produces solutions that dominate the benchmarks from a risk-averse perspective. We have also analyzed the sensitivity of the results with respect to the input parameters. The results demonstrate the applicability and validity of the proposed models. Using our model(s), we obtained allocations that are less risky than the benchmarks, which indicates that the proposed models can be very effective. We also performed limited sensitivity analysis by correlating property losses with fatalities to understand the impact of this correlation on the model solution. No significant changes were observed in the budget allocations with various degree of correlations (see Appendix C.4).

The numerical tests have also shown the need for an efficient and fast algorithm to solve the large-scale sample average approximation of the stochastic dominance model. The bottleneck of the algorithm used in this paper is the mixed-integer program that must be solved every iteration. The numbers presented in the paper correspond to the solutions times obtained with a general MIP solver; conceivably, the computational times can be greatly reduced by implementing strategies that exploit the structure of the subproblems. Research on that topic is underway.

Acknowledgements

This research has been supported by NSF Grant CMMI-0727532 and ONR N00014-09-10518.

References

N. Al-Najjar, S. Baliga, and B. David. Market forces meet behavioral biases: cost misallocation and irrational pricing. *RAND Journal of Economics*, 39(1):214–237, 2008.

M. Azaiez and V. Bier. Optimal resource allocation for security in reliability systems. *European Journal of Operational Research*, 181:773–786, 2007.

M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming Theory and Algorithms (3rd Edition)*. Wiley, New Jersy, 2006.

V. M. Bier, N. Haphuriwat, J. Meonoyo, R. Zimmerman, and A. M. Culpen. Optimal resource allocation for defense of targets based on differing measures of attractiveness. *Risk Analysis*, 28: 763–770, 2008.

J. Birge. Uses of sub-sample estimates to reduce errors in optimization models. The University of Chicago Booth School of Business, 2009.

M. H. Brody. Private communication, 2010. Federal Emergency Management Agency.

S. Caudle. Homeland security capabilities-based planning: Lessons from the defense community. *Homeland Security Affairs*, 1, 2005.

A. L. M. Chauvenet, P. W. J. Baxter, E. McDonald-Madden, and H. P. Possingham. Optimal allocation of conservation effort among subpopulations of a threatened species: How important is patch quality? *Ecological Applications*, 20(3):789–797, 2010.

J. Dahl. Private communications, 2010. Director, Grants Unit, New York Police Department.

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

D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Math. Programming*, 99:329–350, 2004.

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

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

C. Dougherty and G. Psacharopoulos. Measuring the cost of misallocation of investment in education. *The Journal of Human Resources*, 12(4):446–459, 1977.

S. Farrow and S. Stuart. The benefit-cost analysis of security focused regulations. *Journal of Homeland Security and Emergency Management*, 6(1), 2009.

B. Flyvbjerg, N. Bruzelius, and W. Rothengatter. *Megaprojects and risk: an anatomy of ambition.* Cambridge University Press, 2003.

R. Gollmer, U. Gotzes, F. Neise, and R. Schultz. Risk modeling via stochastic dominance in power systems with dispersed generation. Manuscript, Department of Mathematics, University of Duisburg-Essen, Duisburg, Germany, 2007.

Y. Y. Haimes. *Risk Modeling, Assessment, and Management.* Wiley-interscience, New York, 2004.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

T. Homem-de-Mello and S. Mehrotra. A cutting surface method for linear programs with polyhedral stochastic dominance constraints. *SIAM Journal on Optimization*, 20(3):1250–1273, 2009.

J. Hu and S. Mehrotra. Robust and stochastically weighted multi-objective optimization models and reformulations. Manuscript, Northwestern University, available at `http://www.optimization-online.org/DB_HTML/2010/10/2762.html`, 2010.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Sample average approximation of stochastic dominance constrained programs. To appear in *Mathematical Programming*, 2010.

C. Huang, D. Kira, and I. Vertinsky. Stochastic dominance rules for multiattribute utility functions. *Review of Economic Studies*, 41:611–616, 1978.

M. G. Karlaftis, K. L. Kepaptsoglou, and S. Lambropoulos. Fund allocation for transportation network recovery following natural disasters. *Journal of Urban Planning and Development*, 133 (1):82–89, 2007.

N. E. Karoui and A. Meziou. Constrained optimization with respect to stochastic dominance: Application to portfolio insurance. *Mathematical Finance*, 16(1):103–117, 2006.

R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value tradeoffs*. John Wiley & Sons, New York, 1976.

I. Lerche and E. K. Paleologos. *Environmental risk analysis*. McGraw-Hill, 2001.

H. Levy. *Stochastic Dominance : Investment Decision Making under Uncertainty*. Springer, New York, 2006.

M. H. Lin, J. F. Tsai, and Y. Ye. Budget allocation in a competitive communication spectrum economy. *EURASIP Journal on Advances in Signal Processing*, 2009. Article ID 963717, 12 pages.

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

W. K. Mak, D. P. Morton, and R. K. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.

F. P. Miller, A. F. Vandome, and J. McBrewster, editors. *Cost Overrun: Cost, Budget, Cost Escalation, Infrastructure, Building, Project, Information Technology, Telescopic Sight, Sydney Opera House, Concorde.* Alphascript Publishing, 2010.

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

Y. Nie, X. Wu, and T. Homem-de-Mello. Optimal path problems with second-order stochastic dominance constraints. Manuscript, Northwestern University, 2009.

M. Nowak. Interactive approach in multicriteria analysis based on stochastic dominance. *Control and Cybernetics*, 33(3):463–476, 2004.

G. Parnell, R. Dillon-Merrill, and T. Bresnick. Integrating risk management with security and antiterrorism resource allocation decision making. In *The McGraw-Hill Homeland Security Handbook*, pages 431–461. McGraw-Hill, New York, 2006.

C. Paxson and N. R. Schady. The allocation and impact of social funds: spending on school infrastructure in Peru. *The World Bank Economic Review*, 16(2):297–319, 2002.

L. D. Phillips and C. A. Bana e Costa. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*, 154(1):51–68, 2007.

R. Powell. Defending against terrorist attacks with limited resources. *American Political Science Review*, 101:527–541, 2007.

C. Reifel. Quantitative risk analysis for homeland security resource allocation. Master's thesis, Naval Postgraduate School, Monterey, CA, 2006.

D. Roman, K. Darby-Dowman, and G. Mitra. Portfolio construction based on stochastic dominance and target return distributions. *Math. Programming*, 108:541–569, 2006.

M. Shaked and J. G. Shanthikumar. *Stochastic Orders and their Applications.* Academic Press, Boston, 1994.

S. Sriboonchita, W. Wong, S. Dhompongsa, and H. T. Nguyen. *Stochastic Dominance and Applications to Finance, Risk and Economics*. Chapman & Hall, 2009.

R. E. Steuer and P. Na. Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of Operational Research*, 150(3):496–515, 2003.

C.-L. Tseng, K. Y. Lin, and S. K. Sundararajan. Managing cost overrn risk in project funding allocation. *Annals of Operations Research*, 135:127–153, 2005.

H. H. Willis and D. Kleinmuntz. Applying optimal capital allocation methods to homeland security resources:a case study of californias allocation of the buffer zone protection plan grants. DHS University Network Summit on Research and Education, 2007.

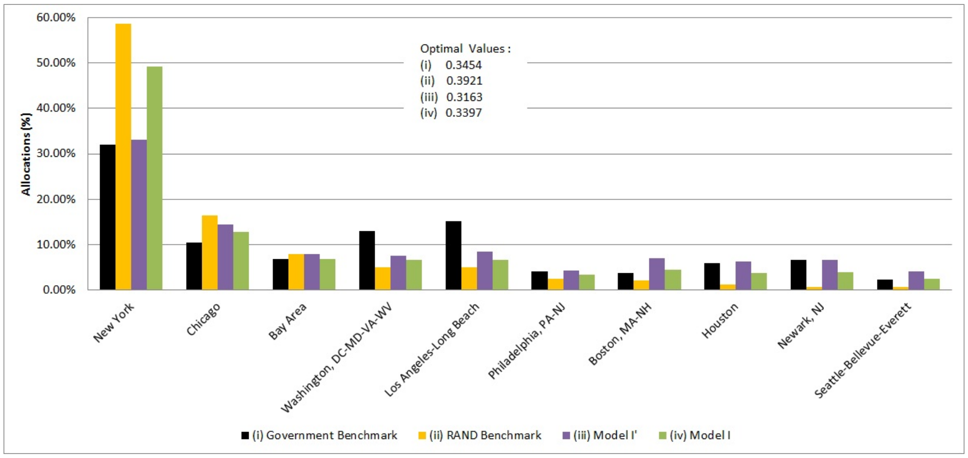
H. H. Willis, A. R. Morral, T. K. Kelly, and J. J. Medby. *Estimating Terrorism Risk*. RAND Corporation, Santa Monica, 2005. `http://www.rand.org/pubs/monographs/2005/RAND_MG388.pdf`.

D. P. Wright, M. J. Liberatore, and R. L. Nydick. A survey of operations research models and applciations in homeland security. *Interfaces*, 36(6):514–529, 2006.

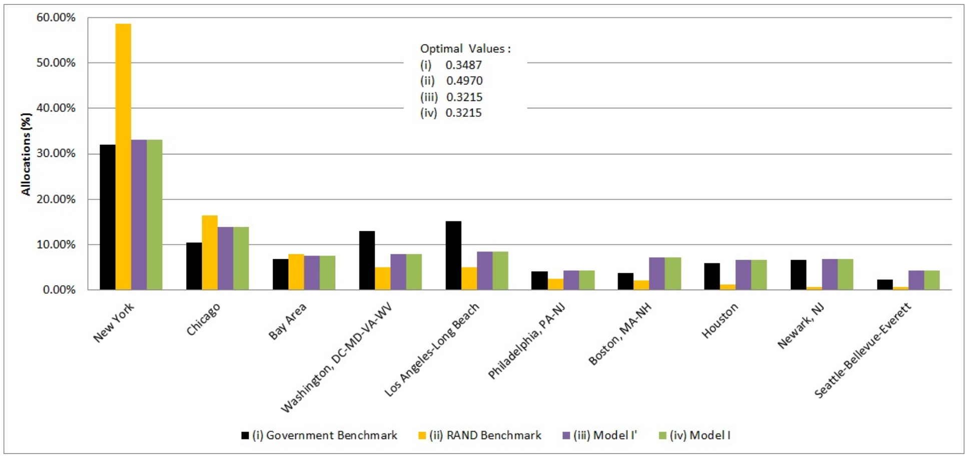
K. Zaras and J. Martel. *Multiattribute analysis based on stochastic dominance*. Kluwer Academic Publishers, Dordrecht, 1994.

J. Zhuang and V. Bier. Balancing terrorism and natural disasters - defensive strategy with endogenous attacker effort. *Operations Research*, 55:976–991, 2007.

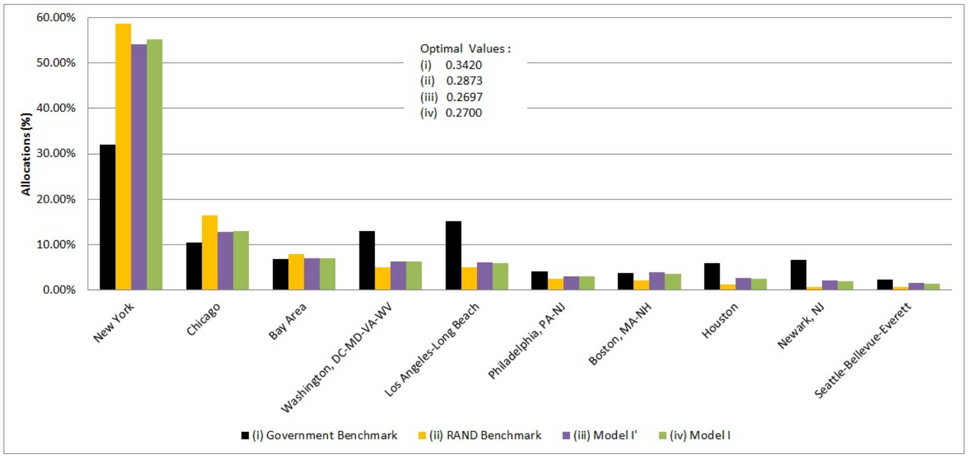
A Figures and Tables



(a) Equality-Center: (1/4, 1/4, 1/4, 1/4)

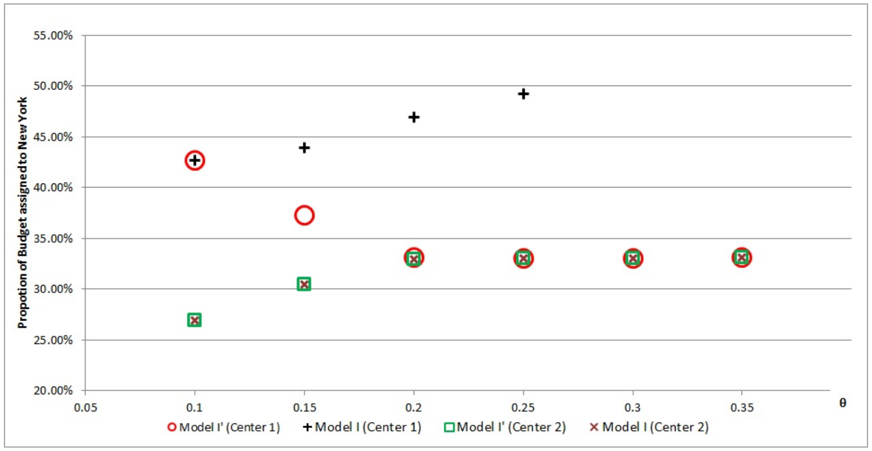


(b) Infrastructure-Center: (1/8, 1/8, 3/8, 3/8)

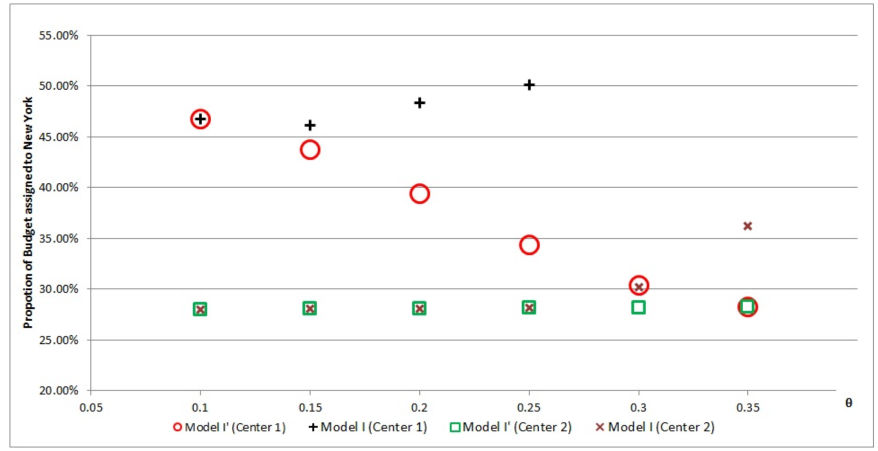


(c) Property-Fatality-Center: (3/8, 3/8, 1/8, 1/8)

Figure 1: Allocation Strategies of Models I′ and I

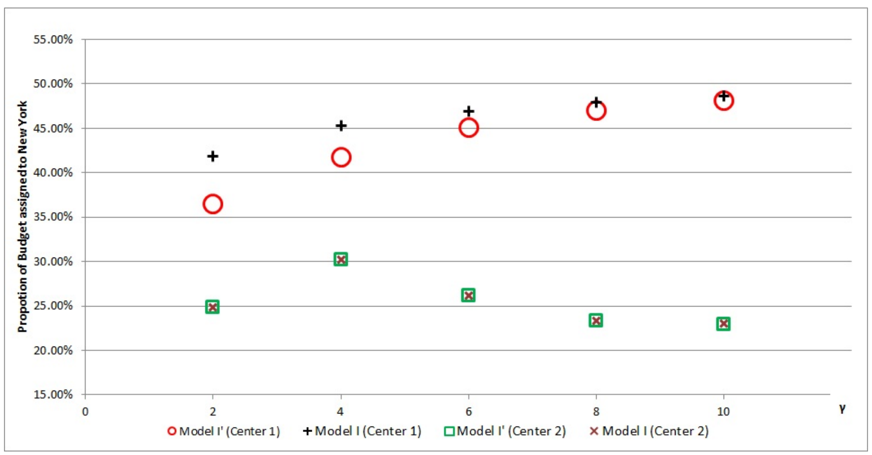


(a) $\gamma = 3$

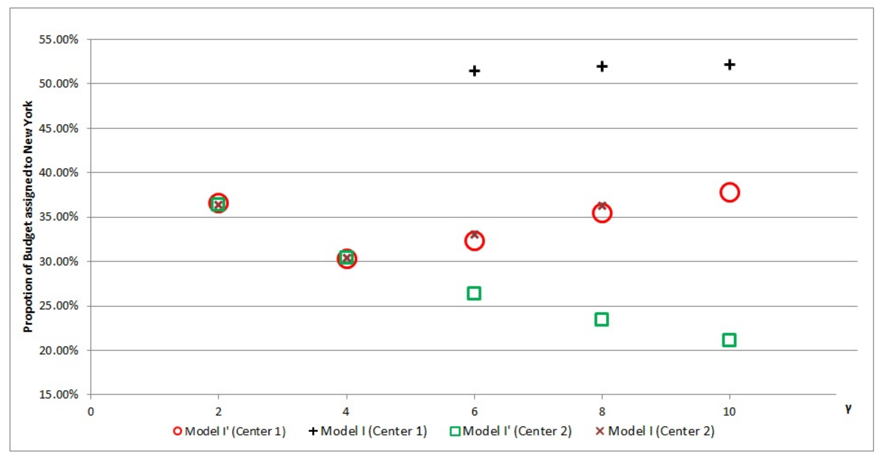


(b) $\gamma = 5$

Figure 2: Sensitive Analysis on θ (Center 1: Equality-Center; Center 2: Infrastructure-Center)



(a) $\theta = 0.15$



(b) $\theta = 0.3$

Figure 3: Sensitive Analysis on γ (Center 1: Equality-Center; Center 2: Infrastructure-Center)

Table 2: Terrorism Losses, Air Departure, and Average Daily Bridge Traffic

Urban Area	Property Losses ($ million) Standard	Reduced	Increased	Fatalities Standard	Reduced	Increased	Air Departures	Average Daily Bridge Traffic
New York	413	265	550	304	221	401	23,599	596,400
Chicago	115	77	150	54	38	73	39,949	318,800
San Francisco	57	38	81	24	16	36	19,142	277,700
Washington, DC-MD-VA-WV	36	21	59	29	16	48	17,253	254,975
Los Angeles-Long Beach	34	16	58	17	7	31	28,816	336,000
Philadelphia, PA-NJ	21	8	28	9	5	13	13,640	192,204
Boston, MA-NH	18	8.3	26	12	8	17	11,625	669,000
Houston	11	6.7	15	9	6	12	20,979	308,060
Newark	7.3	0.8	12	4	0.1	9	12,827	518,100
Seattle-Bellevue-Everett	6.7	4	10	4	3	6	13,578	212,000

Table 3: Percentage of Budget Allocations of FY 2005-2009 UASI and RAND Corporation

Urban Area	UASI (%) FY 2005	FY 2006	FY 2007	FY 2008	FY 2009	Average	RAND Corp. (%) Property	Fatalities
New York	40.76	28.90	29.45	30.51	30.02	31.93	58.61	67.58
Chicago	9.14	12.12	10.33	9.75	10.77	10.42	16.39	11.06
Bay Area	3.81	6.53	7.47	7.84	8.49	6.83	7.91	4.81
Washington, DC-MD-VA-WV	15.62	10.72	13.63	12.71	12.01	12.94	5.06	6.27
Los Angeles-Long Beach	12.38	18.88	16.04	14.83	14.08	15.24	4.95	3.75
Philadelphia, PA-NJ	4.57	4.66	4.18	3.81	3.73	4.19	2.42	1.54
Boston, MA-NH	5.33	4.20	3.08	2.97	3.11	3.74	2.11	2.23
Houston	3.62	3.96	5.49	8.05	8.28	5.88	1.21	1.53
Newark	2.48	7.93	7.91	7.42	7.25	6.60	0.66	0.64
Seattle-Bellevue-Everett	2.29	2.10	2.42	2.12	2.28	2.24	0.67	0.60

Table 4: Computational Performance of the Cut-Generation Algorithm

Sample Size	Outer Iteration	Cuts from Step 2	Cuts from Step 3	Nodes Explored by CPLEX	Time of Step 3	Total Time	$\frac{\text{Step 3}}{\text{Total}}$
100	1	3	0	5,163,678	46.304	51.189	90.46%
150	1	3	0	13,137,699	108.948	121.145	89.93%
200	1	3	0	27,280,477	242.168	262.333	92.31%
250	1	3	0	49,465,356	397.265	435.253	91.27%
300	1	3	0	128,487,464	1318.620	1373.550	96.00%

Table 5: Statistical Lower and Upper Bounds for Optimum Objective Function Values of Model I

		Lower Bound			Upper Bound			Time (95% Bounds)	
γ	θ	90%	95%	99%	90%	95%	99%	Lower	Upper
2	0.15	0.2890	0.2889	0.2886	0.2916	0.2916	0.2918	288.692	387.517
2	0.2	0.2990	0.2988	0.2985	0.3024	0.3026	0.3029	345.658	543.777
2	0.25	0.3118	0.3117	0.3113	—	—	—	406.458	115.083
2	0.3	—	—	—	—	—	—	—	—
3	0.15	0.3134	0.3132	0.3129	0.3174	0.3175	0.3176	306.762	502.274
3	0.2	0.3227	0.3225	0.3221	0.3285	0.3287	0.3289	337.366	666.456
3	0.25	0.3347	0.3345	0.3342	0.3420	0.3422	0.3426	320.750	799.058
3	0.3	—	—	—	—	—	—	—	—
4	0.15	0.3319	0.3318	0.3314	0.3355	0.3355	0.3356	288.938	529.218
4	0.2	0.3422	0.3420	0.3416	0.3472	0.3474	0.3475	349.034	689.030
4	0.25	0.3549	0.3547	0.3543	0.3602	0.3603	0.3606	334.973	875.651
4	0.3	—	—	—	—	—	—	—	—

Table 6: Estimated Optimum Objective Function Values using Batch Sampling

Sample Size	Estimated Objective Function Value Group 1	Group 2	Group 3	Group 4	Group 5	Number of the Estimated Objective Values out of the Range of Statistical Bounds in Table 5
25	0.3183	0.3159	0.3191	0.3165	0.3169	2
50	0.3168	0.3173	0.3178	0.3162	0.3193	2
75	0.3164	0.3161	0.3174	0.3164	0.3169	0
100	0.3165	0.3161	0.3172	0.3163	0.3162	0

Table 7: Outputs of Model I′ over the Largest Weight Region

Urban Area	The government Benchmark (%)	Model I′ (%) $\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$
New York	31.93	36.66	33.19	30.48	28.27
Chicago	10.42	12.64	13.77	14.05	14.14
Bay Area	6.83	7.12	7.35	7.56	7.72
Washington, DC-MD-VA-WV	12.94	7.61	8.04	8.29	8.42
Los Angeles-Long Beach	15.24	8.13	8.59	9.10	9.36
Philadelphia, PA-NJ	4.19	4.13	4.39	4.64	4.88
Boston, MA-NH	3.74	6.21	6.83	7.22	7.71
Houston	5.88	6.47	6.74	7.08	7.34
Newark, NJ	6.60	6.73	6.65	6.92	7.32
Seattle-Bellevue-Everett	2.24	4.29	4.45	4.66	4.83
Maximum Difference		7.11	6.65	6.14	5.88
Average Difference		2.50	2.31	2.45	2.81

B Proofs

B.1 A Line Search Algorithm for the Statistical Upper Bound

We now adapt the line search algorithm in Hu et al. (2010) to our case. Let

$$h(x) := \max_{(k,\eta,w)\in\{1,\dots,q\}\times[-K,K]\times\mathcal{P}} g^k(x,\eta,w), \tag{B.1}$$

$$h_N(x) := \max_{(k,\eta,w)\in\{1,\dots,q\}\times[-K,K]\times\mathcal{P}} g^k_N(x,\eta,w). \tag{B.2}$$

We construct two sets as:

$$T^\epsilon := \{x \in \mathcal{X} : h(x) \le \epsilon\}, \tag{B.3}$$

$$T^\epsilon_N := \{x \in \mathcal{X} : h_{N_g}(x) \le \epsilon\}, \tag{B.4}$$

for $\epsilon \in [-\iota, 0]$, where ι is the given constant in the relaxed version of $\mathcal{P}$-dominance constraints (3.3c). Note that T^ϵ and T^ϵ_N coincide with S^ϵ and S^ϵ_N respectively except for constraints $f^d(x) - z \le \epsilon$ and $f^d_N(x) - z \le \epsilon$. Using T^ϵ_N, we denote a sample problem as

$$\min_{x\in T^\epsilon_N} \max_{d\in\{1,\dots,a\}} f^d_N(x). \tag{B.5}$$

For $\epsilon = 0$, Problem (B.5) is actually an alternative formulation of (3.7). We rewrite it here for clearly describing our method to build the statistical upper bound. The feasible region of problem (B.5) shrinks as ϵ decreases. When $\epsilon < -\iota$, (B.5) is an infeasible problem.

We obtain an optimal solution $x_{N_0}(\epsilon)$ of (B.5) given an initial sample group of size N_0. Then we statistically test whether $x_{N_0}(\epsilon)$ belongs to T^ϵ. In the affirmative case, $x_{N_0}(\epsilon)$ is feasible to (3.3) with a given confidence. We now choose a statistical test method. Using M_g i.i.d sample groups of size N_g each, the method in Mak et al. (1999) gives a $100(1-\alpha_1)\%$ confidence upper bound for $h(x_{N_0}(\epsilon))$ as

$$V_{N_g,M_g}(x_{N_0}(\epsilon)) := \overline{h}_{N_g,M_g}(x_{N_0}(\epsilon)) + t_{\alpha_g,M_g-1}\delta_{N_g,M_g}(x_{N_0}(\epsilon)), \tag{B.6}$$

where $\overline{h}_{N_g,M_g}(x_{N_0}(\epsilon)) := \frac{1}{M_g}\sum_{j=1}^{M_g} h^j_{N_g}(x_{N_0}(\epsilon))$ and $\delta^2_{N_g,M_g}(x_{N_0}(\epsilon)) := \frac{1}{M_g(M_g-1)}\sum_{j=1}^{M_g}(h^j_{N_g}(x_{N_0}(\epsilon)) - \overline{h}_{N_g,M_g}(x_{N_0}(\epsilon)))^2$. If $V_{N_g,M_g}(x_{N_0}(\epsilon)) \le 0$, we claim that $x_{N_0}(\epsilon)$ is an $100(1-\alpha_1)\%$ feasible solution of (3.3).

Based on the above idea, Algorithm 2 proposes the line search algorithm for $\epsilon \in [-\iota, 0]$ to find an upper bound. For each ϵ, if problem (B.5) is infeasible, we increase ϵ to expand the feasible region of (B.5); otherwise, we obtain an optimal solution $x_{N_0}(\epsilon)$. If $V_{N_g,M_g}(x_{N_0}(\epsilon)) > 0$, we decrease ϵ in order to seek a new optimal solution of (B.5) over a smaller feasible region, which is more probably feasible to Model I. On the contrary, If $V_{N_g,M_g}(x_{N_0}(\epsilon)) \le 0$, we increase ϵ for a tighter upper bound.

Algorithm 2 A Line Search Algorithm for Upper Bound

---

0. Choose a small sample size N_0. Let $k := 1,\ a_1 := -\iota,\ b_1 := 0$.
   Choose a tolerance $\gamma > 0$ and let n be the smallest positive integer such that $(1/2)^n \le \gamma/\iota$.

1. Let $\epsilon_k := (a_k + b_k)/2$ and compute an optimal solution $x_{N_0}(\epsilon_k)$ of (B.5); if (B.5) is infeasible, go to step 3.
   Evaluate $V_{N_g,M_g}(x_{N_0}(\epsilon_k))$ by (B.6).
   Go to step 2 if $V_{N_g,M_g}(x_{N_0}(\epsilon_k)) > 0$; otherwise, $x_{N_0}(\epsilon_k)$ is a wanted statistical feasible solution of (3.3). Then, stop or go to step 3 for a tighter upper bound.

2. Let $a_{k+1} := a_k$ and $b_{k+1} := \epsilon_k$; go to step 4.

3. Let $a_{k+1} := \epsilon_k$ and $b_{k+1} := b_k$; go to step 4.

4. If $k = n$, stop; otherwise, replace k by $k + 1$ and go back to step 1.

B.2 Proof of Theorem 1

We first study the properties of the expected value and sample average functions by the following lemma.

Lemma 1 *Suppose Assumptions (A2) and (A3) hold. Then,*

(1) Let $f^d(\cdot)$ be defined as in Section 3.2. For all $t = (x, \eta, w) \in \mathcal{X} \times [-K, K] \times \mathcal{P}$, $|G^k(t, A)| \le 2K + \iota$ a.s. for $k = 1, \ldots, q$.

(2) Let $g^d(\cdot)$ be defined as in Section 3.2. For all $x,\ x' \in \mathcal{X}$, it follows that

$$|f^d(x) - f^d(x')| \le L\|x - x'\|_\infty$$

$$|f_N^d(x) - f_N^d(x')| \le L\|x - x'\|_\infty \ a.s.$$

for $d = 1, \ldots, a$.

(3) For all $t = (x, \eta, w),\ t' = (x', \eta', w') \in \mathcal{X} \times [-K, K] \times \mathcal{P}$, it follows that

$$|g^k(t) - g^k(t')| \leq \max\{2mK, L\}\|t - t'\|_\infty$$
$$|g_N^k(t) - g_N^k(t')| \leq \max\{2mK, L\}\|t - t'\|_\infty$$

for $k = 1, \ldots, q$.

Proof : (1) Using Assumption (A2), for $(x, \eta, w) \in \mathcal{X} \times [-K, K] \times \mathcal{P}$, we have

$$|G^k(x, \eta, w, A)| \leq |w^T M(x, A) - w^T M(y^k, A) - \iota| \leq \|M(x, A)\|_\infty + \|M(y^k, A)\|_\infty + \iota \leq 2K + \iota.$$

(2) By Assumption (A3), we obtain

$$|f^d(x) - f^d(x')| \leq E[\|v^d\|_1 \|M(x, A) - M(x', A)\|_\infty] \leq L\|x - x'\|_\infty,$$

and the same proof also applies to $f_N^d(.)$.

(3) Fixing η and w, we have

$$\begin{aligned} |g^k(x, \eta, w) - g^k(x', \eta, w)| \leq & E|w^T M(x, A) - w^T M(x'A)| \\ \leq & E[\|w\|_1 \|M(x, A) - M(x', A)\|_\infty] \\ \leq & L\|x - x'\|_\infty, \end{aligned}$$

for all $x,\ x' \in \mathcal{X}$. Analogously, it follows that

$$|g^k(x, \eta, w) - g^k(x, \eta', w)| \leq 2|\eta - \eta'|,$$

and

$$|g^k(x, \eta, w) - g^k(x, \eta, w')| \leq \|w - w'\|_\infty (\|M(x, A)\|_1 + \|M(y^k, A)\|_1) \leq 2mK\|w - w'\|_\infty.$$

Thus, the first inequality holds by

$$
\begin{aligned}
|g^k(t) - g^k(t')| \leq & |g^k(x,\eta,w) - g^k(x',\eta,w)| + |g^k(x,\eta,w) - g^k(x,\eta',w)| \\
& + |g^k(x,\eta,w) - g^k(x,\eta,w')| \\
& \leq \max\{2mK, L\}(\|x - x'\|_\infty + |\eta - \eta'| + \|w - w'\|_\infty) \\
& \leq \max\{2mK, L\}\|t - t'\|_\infty
\end{aligned}
$$

Using the same argument, we can prove that the second inequality also holds. □

We now prove Theorem 1. Note that

$$
P(S^{-\epsilon} \subseteq S_N^0 \subseteq S^\epsilon) \geq 1 - P(S^{-\epsilon} \nsubseteq S_N^0) - P(S_N^0 \nsubseteq S^\epsilon).
$$

We only need to calculate the probability of $\{S^{-\epsilon} \nsubseteq S_N^0\}$. It is clear to see that the same approach also applies to $\{S_N^0 \nsubseteq S^\epsilon\}$. Then, it follows

$$
\begin{aligned}
& P(S^{-\epsilon} \nsubseteq S_N^0) \\
& \leq P(\exists (z,x) \in [-K,K] \times \mathcal{X}, d \in \{1,\ldots,a\}, \text{ s.t. } f^d(x) - z \leq -\epsilon \text{ and } f_N^d(x) - z > 0) \\
& \quad + P(\exists t = (x,\eta,w) \in \mathcal{X} \times [-K,K] \times \mathcal{P}, k \in \{1,\ldots,q\}, \text{ s.t. } g^k(t) \leq -\epsilon \text{ and } g_N^k(t) > 0) \\
& \leq \sum_{d=1}^{a} P(\exists x \in \mathcal{X}, \text{ s.t. } f_N^d(x) - f^d(x) > \epsilon) \\
& \quad + \sum_{k=1}^{q} P(\exists t \in \mathcal{X} \times [-K,K] \times \mathcal{P}, \text{ s.t. } g_N^k(t) - g^k(t) > \epsilon)
\end{aligned}
$$

By Assumption (A1), compactness of $\mathcal{X}$ implies that there exists a finite set $\Phi \subset \mathcal{X}$ with cardinality $|\Phi| \leq (4LD/\epsilon)^n$ such that, for all $x \in \mathcal{X}$, we have $x' \in \Phi$ satisfying $\|x - x'\|_\infty \leq \epsilon/4L$. It follows from Lemma 1 (1) that

$$
\begin{aligned}
& |f^d(x) - f^d(x')| \leq L\|x - x'\|_\infty \leq \epsilon/4, \\
& |f_N^d(x) - f_N^d(x')| \leq \epsilon/4.
\end{aligned}
$$

Recall that Assumption (A2) implies that $-K \leq F^d(x, A) \leq K$ a.s. Using Hoeffding's Inequality (see Theorem 2 in Hoeffding (1963)), we have

$$\begin{aligned} P(\exists x \in [-K,K]^n, \text{ s.t. } f_N^d(x) - f^d(x) > \epsilon) &\leq P(\exists x \in \Phi, \text{ s.t. } f_N^d(x) - f^d(x) > \epsilon/2) \\ &\leq |\Phi| e^{-\frac{N\epsilon^2}{8K^2}} \\ &\leq \left(\frac{4LD}{\epsilon}\right)^n e^{-\frac{N\epsilon^2}{8K^2}}. \end{aligned}$$

Note that $\mathcal{P} \subset [0,1]^m$ and $|G^k(t, A)| \leq K + \iota$ a.s. for $t \in \mathcal{X}[-K,K] \times \mathcal{P}$ by Lemma 1 (1). Similarly, we construct a finite set $\Psi \subset \mathcal{X} \times [-K,K] \times \mathcal{P}$ with $|\Psi| \leq (\max\{D, K, 1\} \max\{2mK, 4L\}/\epsilon)^{m+n+1}$ so that

$$\begin{aligned} P(\exists t \in \mathcal{X} \times [-K,K] \times \mathcal{P}, \text{ s.t. } g_N^k(t) - g^k(t) > \epsilon) \leq & P(\exists t \in \Psi, \text{ s.t. } g_N^k(t) - g^k(t) > \epsilon/2) \\ &\leq \left(\frac{\max\{D,K,1\}\max\{2mK,4L\}}{\epsilon}\right)^{m+n+1} e^{-\frac{N\epsilon^2}{8(K+\iota)^2}}. \end{aligned}$$

Therefore, it follows that

$$P(S^{-\epsilon} \not\subseteq S_N^0) \leq (a+q)\left(\frac{\max\{D,K,1\}\max\{2mK,4L\}}{\epsilon}\right)^{m+n+1} e^{-\frac{N\epsilon^2}{8(K+\iota)^2}}.$$

Thus, considering the probability of $\{S_N^0 \not\subseteq S^\epsilon\}$ together, we finally have

$$P(S^{-\epsilon} \subseteq S_N^0 \subseteq S^\epsilon) \geq 1 - 2(a+q)\left(\frac{\max\{D,K,1\}\max\{2mK,4L\}}{\epsilon}\right)^{m+n+1} e^{-\frac{N\epsilon^2}{8(K+\iota)^2}}.$$

Also, the second part follows by imposing $2(a+q)\left(\frac{\max\{D,K,1\}\max\{2mK,4L\}}{\epsilon}\right)^{m+n+1} e^{-\frac{N\epsilon^2}{8(K+\iota)^2}} \leq \beta$.

B.3 Proof of Theorem 2

Let

$$\mathcal{P}_{ik} := \{(w,u) \in \mathbb{R}^{m+N} : u_j \geq w^T(M(y^k, A^j) - M(y^k, A^i)), u_j \geq 0, w \in \mathcal{P}, j = 1, \ldots, N\}, \quad \text{(B.7)}$$

for $i = 1, \ldots, N$ and $k = 1, \ldots, q$. It follows from Theorem 1 in Homem-de-Mello and Mehrotra (2009) that all w^{ik} generated in step 3 are the w-components of vertex solutions of $\mathcal{P}_{ik}$. It is well known that $\mathcal{P}_{ik}$ has finite number of vertex solutions.

B.4 Proof of Proposition 3

(1) Recall that A is a stochastic matrix and $\|w\|_1 = 1$ for all $w \in \mathcal{P}$. It follows that

$$\|M(x,A)\|_\infty = \max_{i\in\{1,\dots,m\}} \sum_{j=1}^{n} (A_{ij} - x_j)_+ \leq \max_{i\in\{1,\dots,m\}} \sum_{j=1}^{n} A_{ij} = 1.$$

and

$$|w^T M(x,A)| \leq \|w\|_1 \|M(x,A)\|_\infty \leq 1 \quad \text{a.s..}$$

Since $w \in \mathbb{R}^+$ and $M_i(x,A) \geq 0$ for each i, we have $0 \leq w^T M(x,A) \leq 1$.

(2) It is clear that $M(.,A)$ is convex a.s. We now prove that it is Lipschitz continuous a.s. Choose $x,\ y \in R^n_+$. It follows that

$$\begin{aligned}\|M(x,A) - M(y,A)\|_\infty &\leq \max_{i\in\{1,\dots,m\}} \sum_{j=1}^{n} |(A_{ij} - x_j)_+ - (A_{ij} - y_j)_+| \\ &\leq \sum_{j=1}^{n} |x_j - y_j| \\ &\leq n\|x - y\|_\infty.\end{aligned}$$

B.5 Proof of Theorem 3

For convenience in the proof, let Y be the feasible region described by constraints (4.2d) - (4.2i). Also denote constants $K^{ilk} := 1/N \sum_{j=1}^{N} (w^{ilk^T} M(y^k, A^j) - w^{ilk^T} M(y^k, A^i))_+ + \iota$.

Clearly, $(\hat{z}, \hat{x}, \hat{t}, \hat{r})$ is an optimal solution of (4.2) and hence $(\hat{z}, \hat{x})$ is an optimal solution of (3.7). Thus, it is easy to see that

$$\hat{t}^{jsh} = (A^j_{sh} - \hat{x}^h)_+, \tag{B.8}$$

$$\hat{r}^{ijlk} = \left(\sum_{s=1}^{m} \sum_{h=1}^{n} w_s^{ilk}\, \hat{t}^{jsh} - w^{ilk^T} M(y^k, A^i) \right)_+ . \tag{B.9}$$

By (B.8), for $d = 1, \dots, a$, we have that

$$\begin{aligned} \hat{\mu}^d(f_N^d(\hat{x}) - \hat{z}) &= \hat{\mu}^d \left(\frac{1}{N} \sum_{j=1}^{N} \sum_{s=1}^{m} \sum_{h=1}^{n} v_s^d (A_{sh}^j - \hat{x}_h)_+ - \hat{z} \right) \\ &= \hat{\mu}^d \left(\frac{1}{N} \sum_{j=1}^{N} \sum_{s=1}^{m} \sum_{h=1}^{n} v_s^d \, \hat{t}^{jsh} - \hat{z} \right) \\ &= 0. \end{aligned} \tag{B.10}$$

The last equality holds since $(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau})$ is a saddle point for $\phi(\cdot)$. Analogously, for $i = 1, \dots, N$, $k = 1, \dots, q$, and $l = 1, \dots, b_{ik}$, it follows by (B.8) and (B.9) that

$$\hat{\nu}^{ilk} g_N^k(\hat{x}, w^{ilk^T} M(y^k, A^i), w^{ilk}) \;=\; \hat{\nu}^{ilk} \left(\frac{1}{N} \sum_{j=1}^{N} \hat{r}^{ijlk} - K^{ilk} \right) \;=\; 0. \tag{B.11}$$

Let us now prove that $(\hat{z}, \hat{x})$ is a minimizer of $\psi(z, x, \hat{\mu}, \hat{\nu})$ over $[0, 1] \times \mathcal{X}$. Note that it follows from KKT that $\phi'_z(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau}) = 0$. Then, we have $\|\hat{\mu}\|_1 = 1$. On the same way as (4.2) to introduce the immediate variables r^{ijlk} and t^{jsk}, the problem $\min\{\psi(z, x, \hat{\mu}, \hat{\nu}) : (z, x) \in [0, 1] \times \mathcal{X}\}$ can be reformulated as

$$\min_{(x,t,r)\in Y} \sum_{d=1}^{a} \frac{\hat{\mu}^d}{N} \sum_{j=1}^{N} \sum_{s=1}^{m} \sum_{h=1}^{n} v_s^d t^{jsh} + \sum_{i=1}^{N} \sum_{k=1}^{q} \sum_{l=1}^{b_{ik}} \hat{\nu}^{ilk} \left(\frac{1}{N} \sum_{j=1}^{N} r^{ijlk} - K^{ilk} \right). \tag{B.12}$$

Recall that vector τ is the multiplier for constraints (4.2d) — (4.2g). Let τ consist of four components, τ_1, τ_2, τ_3, and τ_4, which respectively respond to constraints (4.2d) — (4.2g). Let

$$\begin{aligned} \Upsilon(\hat{x}, \hat{t}, \hat{r}) :=& \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{q} \sum_{l=1}^{b_{ik}} \hat{\tau}_1^{ijlk} \left[\sum_{s=1}^{m} \sum_{h=1}^{n} w_s^{ilk} \hat{t}^{jsh} - w^{ilk^T} M(y^k, A^i) - \hat{r}^{ijlk} \right] \\ &+ \sum_{j=1}^{N} \sum_{s=1}^{m} \sum_{h=1}^{n} \hat{\tau}_2^{jsh} (A_{sh}^j - \hat{x}_h - \hat{t}^{jsh}) - \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{q} \sum_{l=1}^{b_{ik}} \hat{\tau}_3^{ijlk} \hat{r}^{ijlk} - \sum_{j=1}^{N} \sum_{s=1}^{m} \sum_{h=1}^{n} \hat{\tau}_4^{jsh} \hat{t}^{jsh}. \end{aligned}$$

Further (B.12) can be equivalently written as

$$\min_{(x,t,r)\in Y} \sum_{d=1}^{a} \frac{\hat{\mu}^d}{N} \sum_{j=1}^{N} \sum_{s=1}^{m} \sum_{h=1}^{n} v_s^d t^{jsh} + \sum_{i=1}^{N} \sum_{k=1}^{q} \sum_{l=1}^{b_{ik}} \hat{\nu}^{ilk} \left(\frac{1}{N} \sum_{j=1}^{N} r^{ijlk} - K^{ilk} \right) + \Upsilon(\hat{x}, \hat{t}, \hat{r}), \tag{B.13}$$

since $\Upsilon(\hat{x}, \hat{t}, \hat{r})$ is 0 for $(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau})$ is a saddle point. Now compare (B.13) with (4.3). If we replace $\Upsilon(\hat{x}, \hat{t}, \hat{r})$ with $\Upsilon(x, t, r)$, the objective function of (B.13) changes to $\phi(z, x, t, r, \hat{\mu}, \hat{\nu}, \hat{\tau})$. We use the properties of a saddle point again. The new problem over the region Y obviously provides an optimal value, $\phi(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau})$, no more than the optimal value of (B.13). For all $(z, x) \in \mathbb{R} \times X$, it follows that

$$\psi(z, x, \hat{\mu}, \hat{\nu}) \geq \phi(\hat{z}, \hat{x}, \hat{t}, \hat{r}, \hat{\mu}, \hat{\nu}, \hat{\tau}) = \psi(\hat{z}, \hat{x}, \hat{\mu}, \hat{\nu}). \tag{B.14}$$

Note that the last equality is from the same substitution for $\hat{x}$ as above by (B.8) and (B.9).

Finally, by (B.10), (B.11), (B.14), and the fact that $(\hat{z}, \hat{x})$ is feasible to (3.7), our conclusion directly follows from Theorem 6.2.5 in Bazaraa et al. (2006).

C Sensitivity Analyses of Parameters

C.1 Budget Allocated to New York in the Models

New York is the most important region in all the urban area budgets. While New York has the largest discrepancy between the government benchmark and the RAND benchmark, it is indisputably given most of the budget by all policies. The average recommendation from Models I′ and I gives New York 41% of the total budget in Figure 1a, 33% of the total budget in Figure 1b, and 55% of the total budget in Figure 1c. Chicago, the second important area in our cases, gets about 14%. Moveover, we realize that, for different models and cases, the recommendations for New York show a big fluctuation. In Figure 1a, the difference is 16%. We compare the average results in the two cases for Equality-Center and Infrastructure-Center. This value at New York decrease from 41% to 33%. In comparison, the allocations are rather stable at the other nine cities. The differences are at most 3% for all the policies in Figure 1a. Also, the biggest difference between those cases at the two centers is only 2%. The above observation indicates that we can use the difference at New York as a measure to compare results among the models and cases. Thus, our following discussion focuses on the model behavior for New York. Let the proportion of the total budget given to New York be called NY-Budget in short.

C.2 The Role of the Weight Region Size

We now discuss the impact of different values of the weight region parameter θ on the allocations suggested by Models I′ and I. It is clear that the effect of dominance constraints get magnified by increasing θ. Testing the two cases for different γ's, Figure 2 labels NY-Budgets of the models at Equality-Center and Infrastructure-Center as θ increases. Note that the RAND benchmark and the government benchmark are two constants, 58.61% and 31.93% in Figure 2.

Given a small θ, Model I′ has different NY-Budgets for the two centers; however, the NY-Budgets become closer as θ increases and are identical at large values of θ. Actually, both of them are gradually approaching the robust optimal solutions over the largest weight region in Table 7. Recall that Model I′ optimizes the worst case with respect to the weight region. Thus, as the size of the weight region increases, the center becomes unimportant.

Dominance constraints are unbinding for the initial small weight regions so that Models I′ and I have the same NY-Budgets when θ is 0.1. As θ increases, these constraints compel the NY-Budgets of Model I to deviate from those of Model I′ and go toward the dominating benchmarks. In the two figures, Model I for Equality-Center is feasible only if $\mathcal{P}$ is a small region. We have mentioned the conflict of the two benchmarks in Section 4.2, i.e., the corresponding allocations are very different. Because of this conflict, Model I becomes infeasible as θ grows. It follows that the constraints from the government benchmark prevent the solution of Model I from significantly departing from the government benchmark.

C.3 Role of the Parameter Uncertainty

We now discuss the NY-Budgets of Models I$'$ and I for different values of the variability parameter γ in Figure 3. The NY-Budgets of Model I$'$ show monotonic behaviors for the Equality-Center in Figure 3a and for the Infrastructure-Center in Figure 3b. For the other two cases, the values fluctuate in the beginning; however, NY-Budgets still monotonically increase or decrease when γ is large. The air departure and bridge traffic have higher volatilities as γ increases. For Equality-Center, it is a good policy that Model I$'$ emphasizes the relatively stable criteria — property losses and fatalities — when deciding how to allocate the resource. The NY-Budgets correspondingly go up. On the contrary, the NY-Budgets decrease in these cases for Infrastructure-Center, where we are required to highlight the air departure and bridge traffic even if they are highly uncertain. As a result, we have to give these two criteria more consideration to weaken the impact of their uncertainty.

Model I for Equality-Center shows a monotonically increasing NY-Budget toward the dominating benchmark, the RAND benchmark, along with γ. The observation also applies to Model I for Infrastructure-Center. When the dominance constraints are not binding in the beginning, Models I$'$ and I overlap NY-Budgets. However, as γ increases some dominance constraints become binding and NY-Budgets exhibit a monotonic behavior. The rule verifies that the dominance constraints from the RAND benchmark are emphasized as γ increases and, on the other hand, those from the government benchmark becomes weaker. The same reason also applies to the feasibility of Model I. For the case with $\theta = 3$, Model I for Equality-Center is feasible only for a large enough γ, which relaxes the dominance constraints from government benchmark. On the contrary, Model I for Infrastructure-Center loses its feasibility as γ increases, since in that case it becomes difficult to satisfy the dominance constraints from the RAND benchmark.

C.4 Sensitivity Analysis with Property Loss and Fatality Risk Correlation

Since property losses and fatalities are often correlated, here we summarize our experience of their correlation on proposed budgets. We assume that risk values of property losses and fatalities are correlated with each other, but both are independent of air departures and bridge traffic. Since we do not have data on the correlation for each urban area, we adopt the following approach. Let s, t be one of the three scenarios: "standard", "reduced", and "increased" risks. Denote p_{st} as the conditional probability of fatality in scenario t given the property loss in scenario s. Let the conditional probability be $\pi := p_{ss} \in [1/3, 1]$ and $p_{st} = \frac{1-\pi}{2}$ for $s \neq t$. For $\pi = 1/3$ the property losses and fatalities are independent and if $\pi = 1$ they are fully correlated. Note that we first fix $\pi = 1/2$ in our computation and then discuss its impact on the optimal solutions of Models I$'$ and I by a sensitive analysis. For $(\gamma, \theta) = (3, 0.25)$ and different values of π in the above range we found no significant difference in allocated budgets.