ROBUST OPTIMIZATION WITH APPLICATIONS IN HEALTHCARE OPERATIONS MANAGEMENT

MEILIN ZHANG

NATIONAL UNIVERSITY OF SINGAPORE

2014

ROBUST OPTIMIZATION WITH APPLICATIONS IN HEALTHCARE OPERATIONS MANAGEMENT

MEILIN ZHANG

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF DECISION SCIENCES
NATIONAL UNIVERSITY OF SINGAPORE
2014

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the resources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Meilin Zhang

09 July 2014

Acknowledgements

As I reflect upon the past 5 years I spent for my PhD, I would like to express my deepest gratitude to all those helped me on this journey. I am fortunate to have their exceptional support and company which made my everyday.

First and foremost, I would like to thank my supervisor, Prof. Melvyn Sim, for his constant support, motivation, coaching and guidance. Melvyn is someone you will instantly like and never forget once you meet him. He's the most kind and one of the smartest people I know. He helped and guided me through the most difficult times in research and also in life, and is always a wonderful friend and mentor who is willing to help and train me from the most basic concepts and offer insightful and constructive suggestions for my research. I am always amazed by his wisdom, integrity, accessability and foremost his patience and trust. Without his support and help, I would not have the chance to reevaluate myself and truly know myself, and above all rekindle my passion for research in the right direction. It is a great honor to be his student and work with him.

I also thank Prof. Yaozhong Wu for his insightful research discussion and scientific advice. His enthusiasm and love for research and teaching is contagious. I am grateful to the help and support I have got from Prof. Mabel Chou and Prof. Chung-Piaw Teo. I would also like to thank Prof. Hanqing Zhang, Prof. Jie Sun, Prof. Jussi Keppo, Prof. Andrew Lim, Prof Lucy Chen and Prof Tong Wang from whom I learned a lot.

My PhD journey would not have been so colourful without the assistance and understanding from the staff in the PhD office and Decision Science department: Cheow Loo, Hamidah, Cythcia and specially thanks to Chwee Ming who always cheers me up. I thank Qingxia Kong, Vinit Kumar, Zhuoyu Long, Jin Qi, Zhichao Zheng, Junfei Huang, Lijian Lu, Yuanguang Zhong and Li Xiao for the learning and discussion together.

I owe my special thanks to my dear officemates Xuchuan Yuan, Rohit Nishant and Hossein Eslami who have golden hearts and love to share. We learn a lot from each other and really enjoy the each day spending together.

I will forever be thankful to my "sisters": Masia Zhiying Jiang, Qian Lu, Joecy Jie Wei for being my powerful backing and giving me the best time on this journey. Masia is just like my elder sister who took care of me with great patience and guidance. She is my best role model in my life. I entered this PhD program with Qian and Joecy the same year and we share our tears, joy, dreams and passion.

I especially thank my mom and dad. My hard-working parents have sacrificed their lives for me and provided unconditional love and care. I would not have made it this far without them. I know I always have my family to count on when times are rough.

The best outcome from these years is finding my best friend, soulmate, and husband. I believe this is the most wise decision I ever made when I determined to propose to Tianjue Lin. Tianjue is the only person who knows me, understands me, supports me and truly appreciate my work including both my research and cooking. There are no words to convey how grateful I am to have him. He has been

non-judgmental of me and instrumental in instilling confidence. He is also the most critical judge for my research and culinary skill, because he strongly believe I could always go further and further. "Uncertainty" is the most frequent scenario for the past 5 years in my life, which I was dealing with, struggling with, and frustrated with. There is no such "uncertainty" in Tianjue's "if-else" world which decomposes all possible situations and construct their respective solutions. Now I feel that we could create a better and better life together.

Singapore Meilin Zhang

July, 2014

Contents

Abstract

The combination of an increasingly complex world, the vast proliferation of data, and the pressing need to stay one step ahead of competition has sharpened focus on using analytics and optimization for decision making (see LaValle et al. (2010)). There is also a need to computationally exploit the wealth of data available in optimization problems by providing a flexible framework for modeling uncertainty that incorporates distributional information, while preserving the computational tractability for practical implementation. As motivated by the importance of such a decision making process, I investigate this procedure under robust optimization and extend the findings into real applications in health care operations management. This dissertation integrates the three aspects: theoretical foundation, software tools and applications. We developed a modular framework to obtain exact and approximate solutions to a class of linear optimization problems with recourse with the goal to minimize the worst-case expected objective over a probability distributions or *ambiguity set*. This approach extends to a multistage problem and improves upon existing variants of linear decision rules when recourse are present. We also demonstrate the practicability of our framework by developing a new algebraic modeling package named ROC, a C++ library that implements the techniques developed in theory part. In addition, we apply this methodology in two hospital applications: managing elective admission and patient flow control in emergency

department. For the two applications, we utilize the historical data from Singapore public hospitals in our numerical study. The performance of our approach could easily outperform other commonly used strategies.

List of Figures

List of Tables

1

Introduction

Decision making under uncertainty is essentially part of our daily life and business. In that setting, decision-maker needs to make some decisions even before observing the real value of underlying uncertain parameters. This process is non-trivial and costly most times, perhaps punitively to do so. Decision analysis has been deeply explored in economics, psychology, philosophy, mathematics and statistics in order to make better solutions. Traditionally, people apply the expected-value paradigm in their objective setting until mid-1960s when Dupacova (1987) pointed the practical limitations of this approach, since it requires the complete knowledge of underlying probability distribution which is hardly true for most real world problems: data is not exactly known or measured. This fact actually motivated the development of a mini-max approach (minimizing the worst-case scenario), and drew significant attention in stochastic programming literature, Scarf (1958). However, such approach usually requires finding the worst-case probability distribution. Moreover, stochastic problems, especially multistage ones, are notoriously difficult to solve either analytically or numerically. Therefore, it is important to develop an approximate model which is tractable and scalable when applied in practice. Under this circumstance, operations researchers look into robust optimization as an alternative way of dealing with uncertainty which solves the worst case optimality.

1. INTRODUCTION

Robust optimization deals with data uncertainty by finding the optimal solutions in a mini-max setting. The origins of robust optimization date back to the establishment of modern decision theory in 1950s and the use of worst case analysis as a tool for the treatment of severe uncertainty. A. L. Soyster (1973) first proposed the model which could guarantee feasibility for all possible instances within a convex set. In mid 1990s, Ben-Tal and Nemirovski (1998, 1999) further investigated the tractable robust counterparts of linear, semidefinite and other convex type optimization problems. They also tried to apply similar methodology to solve multistage stochastic programming problems which suffer from curse of dimensionality.

In either stochastic programming or robust optimization, a key modeling concept for multi-period problems is the ability to define wait and see or recourse decision variables. In reality, uncertainty will only be resolved at some known time in the future. For instance, next years interest rate and next months rainfall are unknown for now but known with certainty in future. Recourse decision variables means those decisions can be made on a wait and see basis, after the uncertainty is resolved. It is natural to connect recourse variables with the underlying uncertain variables or dependability between them. Concerning about the tractability and scalability of approximate stochastic programs, Ben-Tal et al. (2004) propose an adjustable robust counterpart to address the dynamic decision making under uncertainty. Chen et al. (2007) also suggested a tractable approximate approach for solving a class of multistage chance constrained stochastic programs. They both applied linear decision rule to ensure scalability in multistage models. Nevertheless, the resulting model usually yields very conservative solutions which are far from optimality in the nominal model of practical interests where partial information of underlying uncertainty is known. Another issue with linear decision rules is that it cannot always ensure feasibility even under simple complete recourse or the resulted solution is nonapplicable. For this reason, Chen et al. (2008) extended the linear decision rule to deflected linear decision rule and segregated linear decision rule to solve such multistage stochastic problems. The applications includes portfolio selection, inventory management, network design under uncertainty. But the price is that such decision rules are difficult and too complicated to implement in reality since we need to solve numerous sub problems in order to derive the primary one.

In addition, nearly all of these methods have been labor-intensive to transform into solvable project (tractable robust counterparts). To our knowledge, there is no general-purpose software which is of high performance and scalable to solve robust optimization problems. Existing toolboxes for robust optimization modeling include AIMMS and ROME (Goh and Sim (2009)). For AIMMS, it only covers limited functionality of robust counterpart transformation and affinely adjustable variables. For example, it does not include the expected term or support more complex decision rules if needed. For ROME, it is a algebraic modeling toolbox built in the MATLAB envirsonment which cannot solve large scale robust optimization.

Being motivated by those questions encountered above, we aim to investigate more in robust optimization both theoretically and practically, and further contribute it to decision making under various applications.

1.1 Structure of the Dissertation

This dissertation is organized as three separate topics but coherently bonded. The first topic is our theoretical foundations in distributionally robust optimization with developed software tool. In the rest two topics, we study two applications in health care operations management under robust optimization. We conclude the thesis in the last part.

- **Chapter 2: A practically efficient framework for distributionally robust linear optimization**

We developed a modular framework to obtain exact and approximate solutions to a class of linear optimization problems with recourse with the goal to minimize the worst-case expected objective over a probability distributions or *ambiguity set*. The ambiguity set is specified by linear and conic quadratic representable expectation constraints and the support set is also linear and conic quadratic representable. We propose an approach to lift the original ambiguity set to an extended one by introducing additional auxiliary random variables. We show that by replacing the recourse decision functions with generalized linear decision rules that have affine dependency on the uncertain parameters and the auxiliary random variables, we can obtain good and sometimes tight approximations to a two-stage optimization problem. This approach extends to a multistage problem and improves upon existing variants of linear decision rules. We demonstrate the practicability of our framework by developing a new algebraic modeling package named ROC, a C++ library that implements the techniques developed in this paper.

- **Chapter 3: A Robust Optimization Model for Managing Elective Admission in Hospital**

The admission of emergency inpatients in a hospital is unscheduled, urgent and takes priority over elective patients, who are usually scheduled several days in advance. Hospital beds are a critical resource and the management of elective admissions by enforcing quotas could reduce incidents of shortfall. We propose a distributionally robust optimization approach for managing elective admissions to determine these quotas. Based on an ambiguous set of probability distributions, we propose an optimized *budget of variation* approach that maximizes the level of uncertainty the admission system can withstand without violating the expected bed shortfall constraint. We solve the robust optimization model by deriving a second order conic problem (SOCP) equivalent of the model. The proposed model is tested in simulations based on real hospital admission data and we report favorable results for adopting the robust optimization models.

- **Chapter 4: Patient Flow Scheduling Study in Emergency Department with Targeted Deadlines**

  Our work examines patient flow control in the Emergency Department ED which is part of the core functionality units in hospitals. Doctors in emergency departments usually decide which patient should be seen next among all new patients and those returning patients whose prescribed tests are ready to be checked. We analyze doctors decision behaviors in practice under different workload from a large sample of historical data. In addition, we propose an optimized scheduling policy with targeted deadlines in terms of both first wait till the first consultation FW and overall length of stay LoS in hospital. Our objective is to maximize the percentage of patients who can meet those deadline constraints while keeping the extreme cases in a reasonable level. We introduce a doctors effort level (α), which deals with the uncertain service time in the optimization model. We aim to minimize this effort level and meanwhile satisfy the deadline constraints. In the numerical study, we compare 4 different policies: First Come First Serve $FCFS$, Shortest Deadline First SDF, Huang et al. (2014) heuristic policy $HeuristicPolicy$ and our optimized policy OPT. Simulation study shows our policy outperforms those commonly-used policies in terms of both FW and LoS easily.

- **Chapter 5: Conclusion and Discussion**

  In this chapter we conclude the thesis and discuss future research.

2

A practically efficient framework for distributionally robust linear optimization

Real world optimization problems are often confounded by the difficulties of addressing the issues of uncertainty. In characterizing uncertainty, Knight (1921) is among the first to establish the distinction of risk, where the probability distribution of the uncertainty is known, and ambiguity, where it is not. Ambiguity exists in practice because it is often difficult or impossible to obtain the true probability distribution due to the possibly lack of available or "good enough" empirical records associated with the uncertain parameters. However, in normative decision making, ambiguity is often ignored in favor of risk preferences over subjective probabilities. Notably, Ellsberg (1961) demonstrates that choice under the presence of ambiguity cannot be reconciled by subjective risk preferences and his findings are corroborated in later studies including the groundbreaking research of Hsu et al. (2005).

In classical stochastic optimization models, uncertainties are represented as random variables with probability distributions and the decision makers optimize the solutions according to their risk preferences (see, for instance, Birge and Louveaux (1997), Ruszczynski and Shaprio (2003)). In particular, risk neutral decision makers prefer solutions that yield optimal expected or average objectives, which are evaluated based on the given probability distributions that characterize the uncertain parameters of the models. Hence, classical stochastic optimization models do not account for ambiguity and subjective probability distributions are used in these models whenever the true distributions are unavailable.

In recent years, research on ambiguity has garnered considerable research interest in various fields including economics, mathematical finance and operations research. In the case of ambiguity aversion, robust optimization is a relatively new approach that deals with ambiguity in mathematical optimization problems. In classical robust optimization, uncertainty is distribution free described by an uncertainty set, which is typically in the form of a conic representable bounded convex set (see Ben-Tal and Nemirovski (1998, 1999, 2000), Bertsimas and Brown (2009), Bertsimas and Sim (2004), Ghaoui and Lebret (1997), El Ghaoui et al. (1998)). Both risk and ambiguity should be taken into account in modeling an optimization problem under uncertainty. From the decision theoretic perspective, Gilboa and Schmeidler (1989) propose to rank preferences based on the worst-case expected utility or disutility over an ambiguity set of distributions. Scarf (1958) is arguably the first to conjure such an optimization model when he studies a single-product newsvendor problem in which the precise demand distribution is unknown but is only characterized by its mean and variance. Indeed, such models have been discussed in the context of minimax stochastic optimization models (see Breton and EI Hachem (1995), Dupacova (1987), Shapiro and Kleywegt (2002), Shapiro and Ahmed (2004), Žáčková (1966)), and recently in the context of distributionally robust optimization models (see Chen and Sim (2009), Chen et al. (2007), Delage and Ye (2010), Popescu (2007), Wiesemann et al. (2014), Xu and Mannor (2012)).

2. A PRACTICALLY EFFICIENT FRAMEWORK FOR DISTRIBUTIONALLY ROBUST LINEAR OPTIMIZATION

Many optimization problems involve dynamic decision makings in an environment where uncertainties are progressively unfolded in stages. Unfortunately, such problems often suffer from the "curse of dimensionality" and are typically computationally intractable (see Ben-Tal et al. (2004), Dyer and Stougie (2006), Shapiro and Nemirovski (2005)). One approach to circumvent the intractability is to restrict the dynamic or recourse decisions to being affinely dependent of the uncertain parameters, an approach known as linear decision rule. Linear decision rules appear in early literatures of stochastic optimization models but are abandoned due to their lack of optimality (see Garstka and Wets (1974)). The interest in linear decision rules is rekindled by Ben-Tal et al. (2004) in their seminal work that extends classical robust optimization to encompass recourse decisions. To further motivate linear decision rules, Bertsimas et al. (2010) establish the optimality of linear decision rules in some important classes of dynamic optimization problems under full ambiguity. In more general classes of problems, Chen and Zhang (2009) improve the optimality of linear decision rules by extending linear decision rules to encompass affine dependency on the auxiliary parameters that are used to characterize the support set. Chen et al. (2007) also use linear decision rules to provide tractable solutions to a class of distributionally robust optimization problems with recourse. Henceforth, variants of linear and piecewise-linear decision rules have been proposed to improve the performance of more general classes of distributional robust optimization problems while maintaining the tractability of these problems. Such approaches include the deflected and segregated linear decision rules of Chen et al. (2008), the truncated linear decision rules of See and Sim (2009), and the bideflected and (generalized) segregated linear decision rules of Goh and Sim (2010). Interestingly, there is also a revival in decision rules for addressing stochastic optimization problems. Specifically, Kuhn et al. (2011) propose primal and dual linear decision rules techniques to solve multistage stochastic optimization problems that would also quantify the potential loss of optimality as the result of such approximations.

Despite the importance of addressing uncertainty in optimization problems, it is often ignored in practice due to the elevated complexity of modeling these problems compared to their deterministic counterparts. A useful framework for optimization under uncertainty should also translate to viable software solutions that are potentially intuitive to the users and would enable them to focus on modeling issues and relieve them from the burden of algorithm tweaking and code troubleshooting. Software that facilitates robust optimization modeling have begun to surface in recent years. Existing toolboxes for robust optimization include YALMIP[1], AIMMS[2] and ROME[3]. Of those, ROME and AIMMS have provisions for decision rules and hence, they are capable of addressing dynamic optimization problems under uncertainty. AIMMS is a commercial software package that adopts the classical robust linear optimization framework where uncertainty is only characterized by the support set without distributional information. ROME is an algebraic modeling toolbox built in the MATLAB environment that implements the distributionally robust linear optimization framework of Goh and Sim (2010). Despite the polynomial tractability, the reformulation approach of Goh and Sim (2010) can be rather demanding, which could limit the scalability potentially needed for addressing larger sized problems.

In this chapter, we develop a new modular framework to obtain exact and approximate solutions to a class of linear optimization problems with recourse with the goal to minimize the worst-case expected objective over an ambiguity set of distributions. Our contributions to this paper are as follows:

1. We propose to focus on a standard ambiguity set where the family of distributions are characterized by linear and conic representable expectation constraints and the support set is also linear and conic representable. As we will show, the standard ambiguity set has important ramifications on the tractability of the problem.

2. We adopt the approach of Wiesemann et al. (2014) to lift the original am-

biguity set to an extended one by introducing additional auxiliary random variables. We show that by replacing the recourse decision functions with generalized linear decision rules that have affine dependency on the uncertain parameters and the auxiliary random variables, we can obtain good and sometimes tight approximations to a two-stage optimization problem. This approach is easy to compute, extends to a multistage problem and improves upon existing variants of linear decision rules developed in Chen and Zhang (2009), Chen et al. (2008), Goh and Sim (2010), See and Sim (2009).

3. We demonstrate the practicality of our framework by developing a new algebraic modeling package named ROC, a C++ library that implements the techniques developed in this paper.

Notations. Given a $N \in \mathbb{N}$, we use $[N]$ to denote the set of running indices, $\{1, \dots, N\}$. We generally use bold faced characters such as $\boldsymbol{x} \in \Re^N$ and $\boldsymbol{A} \in \Re^{M \times N}$ to represent vectors and matrixes. We use $[\boldsymbol{x}]_i$ or x_i to denote the i element of the vector $\boldsymbol{x}$. We use $(x)^+$ to denote $\max\{x, 0\}$. Special vectors include $\mathbf{0}$, $\mathbf{1}$ and $\boldsymbol{e}_i$ which are respectively the vector of zeros, the vector of ones and the standard unit basis vector. Given $N, M \in \mathbb{N}$, we denote $\mathcal{R}^{N,M}$ as the space of all measurable functions from $\Re^N$ to $\Re^M$ that are bounded on compact sets. For a proper cone $\mathcal{K} \subseteq \Re^L$ (*i.e.*, a closed, convex and pointed cone with nonempty interior), we use the relations $\boldsymbol{x} \preceq_{\mathcal{K}} \boldsymbol{y}$ or $\boldsymbol{y} \succeq_{\mathcal{K}} \boldsymbol{x}$ to indicate that $\boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}$. Similarly, the relations $\boldsymbol{x} \prec_{\mathcal{K}} \boldsymbol{y}$ or $\boldsymbol{y} \succ_{\mathcal{K}} \boldsymbol{x}$ imply that $\boldsymbol{y} - \boldsymbol{x} \in \text{int}\mathcal{K}$, where $\text{int}\mathcal{K}$ represents the interior of the cone $\mathcal{K}$. Meanwhile, $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$ with $\mathcal{K}^* = \{\boldsymbol{y} : \boldsymbol{y}'\boldsymbol{x} \geq 0, \boldsymbol{x} \in \mathcal{K}\}$. We use tilde to denote an uncertain or random parameter such as $\tilde{\boldsymbol{z}} \in \Re^I$ without associating it with a particular probability distribution. We denote $\mathcal{P}_0(\Re^I)$ as the set of all probability distributions on $\Re^I$. Given a random vector $\tilde{\boldsymbol{z}} \in \Re^I$ with probability distribution $\mathbb{P} \in \mathcal{P}_0(\Re^I)$

and function $\boldsymbol{g} \in \mathcal{R}^{I,P}$, we denote $\mathbb{E}_{\mathbb{P}}(\boldsymbol{g}(\tilde{\boldsymbol{z}}))$ as the expectation of the random variable, $\boldsymbol{g}(\tilde{\boldsymbol{z}})$ over the probability distribution $\mathbb{P}$. Similarly, for a set $\mathcal{W} \subseteq \Re^I$, $\mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W})$ represents the probability of $\tilde{\boldsymbol{z}}$ being in the set $\mathcal{W}$ evaluated on the distritbution $\mathbb{P}$. Suppose $\mathbb{Q} \in \mathcal{P}_0(\Re^I \times \Re^L)$ is a joint probability distribution of two random vectors $\tilde{\boldsymbol{z}} \in \Re^I$ and $\tilde{\boldsymbol{u}} \in \Re^L$, then $\prod_{\tilde{\boldsymbol{z}}} \mathbb{Q} \in \mathcal{P}_0(\Re^I)$ denotes the marginal distribution of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}$. Likewise, for a family of distributions, $\mathbb{G} \subseteq \mathcal{P}_0(\Re^I \times \Re^L)$, $\prod_{\tilde{\boldsymbol{z}}} \mathbb{G}$ represents the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under all $\mathbb{Q} \in \mathbb{G}$, i.e., $\prod_{\tilde{\boldsymbol{z}}} \mathbb{G} = \{\prod_{\tilde{\boldsymbol{z}}} \mathbb{Q} : \mathbb{Q} \in \mathbb{G}\}$.

2.1 A two stage distributionally robust optimization problem

In this section, we focus on a two-stage optimization problem where the first stage or *here-and-now* decision is a vector $\boldsymbol{x} \in \Re^{N_1}$ chosen over the feasible set X_1. The cost incurred during the the first stage in association with the decision $\boldsymbol{x}$ is deterministic and given by $\boldsymbol{c}'\boldsymbol{x}$, $\boldsymbol{c} \in \Re^{N_1}$. In progressing to the next stage, a vector of uncertain parameters $\tilde{\boldsymbol{z}} \in \mathcal{W} \subseteq \Re^{I_1}$ is realized; thereafter, we could determine the cost incurred at the second stage. Similar to a typical stochastic programming model, for a given decision vector, $\boldsymbol{x}$ and a realization of the uncertain parameters, $\boldsymbol{z} \in \mathcal{W}$, we evaluate the second stage cost via the following linear optimization problem,

$$
\begin{aligned}
Q(\boldsymbol{x}, \boldsymbol{z}) = \quad \min \quad & \boldsymbol{d}'\boldsymbol{y} \\
\text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \\
& \boldsymbol{y} \in \Re^{N_2}
\end{aligned}
\tag{2.1}
$$

Here, $\boldsymbol{A} \in \mathcal{R}^{I_1, M \times N_1}$, $\boldsymbol{b} \in \mathcal{R}^{I_1, M}$ are functions that maps from the vector $\boldsymbol{z} \in \mathcal{W}$ to the input parameters of the linear optimization problem. Adopting the common assumptions in the robust optimization literature, these functions are affinely dependent on $\boldsymbol{z} \in \Re^{I_1}$ and are given by,

$$\boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{A}^0 + \sum_{k \in [I_1]} \boldsymbol{A}^k z_k, \boldsymbol{b}(\boldsymbol{z}) = \boldsymbol{b}^0 + \sum_{k \in [I_1]} \boldsymbol{b}^k z_k,$$

with $\boldsymbol{A}^0, \boldsymbol{A}^1, ..., \boldsymbol{A}^{I_1} \in \Re^{M \times N_1}$ and $\boldsymbol{b}^0, \boldsymbol{b}^1, ..., \boldsymbol{b}^{I_1} \in \Re^M$. The matrix $\boldsymbol{B} \in \Re^{M \times N_2}$ and the vector $\boldsymbol{d} \in \Re^{N_2}$ are unaffected by the uncertainties, which corresponds to the case of *fixed-recourse* as defined in stochastic programming literatures.

The second stage decision (*wait-and-see*) is represented by the vector $\boldsymbol{y} \in \Re^{N_2}$, which is easily determined by solving a linear optimization problem after the uncertainty is realized. However, whenever the second stage problem is infeasible, we have $Q(\boldsymbol{x}, \boldsymbol{z}) = \infty$, and the first stage solution, $\boldsymbol{x}$ would be rendered meaningless. As in the case of a standard stochastic programming model, $\boldsymbol{x}$ has to be feasible in $X_1 \cap X_2$, where

$$X_2 = \{\boldsymbol{x} \in \Re^{N_1} \ : \ Q(\boldsymbol{x}, \boldsymbol{z}) < \infty \ \forall \boldsymbol{z} \in \mathcal{W}\}.$$

Unfortunately, checking the feasibility of X_2 is already $\mathcal{NP}$-complete (see Ben-Tal et al. (2004)), hence, for simplicity, we focus on problems with relatively complete recourse, i.e.,

Assumption 1.

$$X_1 \subseteq X_2.$$

In the context of stochastic programming, complete recourse refers to the characteristics of the recourse matrix, $\boldsymbol{B}$ such that for any $\boldsymbol{t} \in \Re^M$, there exists $\boldsymbol{y} \in \Re^{N_2}$ such that $\boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{t}$. Therefore, under complete recourse we have $X_2 = \Re^{N_1}$.

Model of uncertainty

Adopting the standardized framework in Wiesemann et al. (2014), we assume that the probability distribution of $\tilde{\boldsymbol{z}}$ belongs to an *ambiguity set*, $\mathbb{F}$ as follows

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\Re^{I_1}\right) : \begin{array}{l} \tilde{\boldsymbol{z}} \in \Re^{I_1} \\ \mathbb{E}_{\mathbb{P}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}(\boldsymbol{g}(\tilde{\boldsymbol{z}})) \preceq_{\mathcal{K}_0} \boldsymbol{\sigma} \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1 \end{array} \right\} \tag{2.2}$$

with $\boldsymbol{G} \in \Re^{L_1 \times I_1}$, $\boldsymbol{\mu} \in \Re^{L_1}$, $\boldsymbol{\sigma} \in \Re^{L_2}$, $\boldsymbol{g} \in \mathcal{R}^{I_1, L_2}$ and $\mathcal{K}_0 \subseteq \Re^{L_2}$. The function $\boldsymbol{g}$ is such that the set

$$\mathcal{G} = \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \Re^{I_1} \times \Re^{I_2} : \boldsymbol{g}(\boldsymbol{z}) \preceq_{\mathcal{K}_0} \boldsymbol{u}\right\}$$

is conic representable. The support set $\mathcal{W}$ is conic representable, and we define the set

$$\bar{\mathcal{W}} = \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{G} : \boldsymbol{z} \in \mathcal{W}\right\}, \tag{2.3}$$

such that for all $\boldsymbol{z} \in \mathcal{W}$, we have $(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z})) \in \bar{\mathcal{W}}$. In particular, the explicit formulation of $\bar{\mathcal{W}}$ is given by

$$\bar{\mathcal{W}} = \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \Re^{I_1} \times \Re^{I_2} : \exists \boldsymbol{v} \in \Re^{I_3}, (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}}\right\}, \tag{2.4}$$

where we define $\hat{\mathcal{W}}$ as the extended support set,

$$\hat{\mathcal{W}} = \left\{(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3} : \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{E}\boldsymbol{v} \preceq_{\mathcal{K}} \boldsymbol{h}\right\},$$

with $\boldsymbol{C} \in \Re^{L_3 \times I_1}$, $\boldsymbol{D} \in \Re^{L_3 \times I_2}$, $\boldsymbol{E} \in \Re^{L_3 \times I_3}$, $\boldsymbol{h} \in \Re^{L_3}$ and $\mathcal{K} \subseteq \Re^{L_3}$ being a proper cone. The vector $\boldsymbol{v}$ is the new auxiliary variables associated with the conic reformulation. Hence, we can partition $[I_3]$ into two disjoint subsets $\mathcal{I}_3$, $\bar{\mathcal{I}}_3$, $\mathcal{I}_3 \cup \bar{\mathcal{I}}_3 = [I_3]$ such that v_i, $i \in \mathcal{I}_3$ are the auxiliary variables associated the representation of $\mathcal{G}$ while v_i, $i \in \bar{\mathcal{I}}_3$ are those associated with the support set $\mathcal{W}$. Note that for all $\boldsymbol{z} \in \mathcal{W}$ there exists $\boldsymbol{v} \in \Re^{I_3}$ such that $(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{v}) \in \hat{\mathcal{W}}$. Correspondingly, there also exists a function, $\boldsymbol{\nu} \in \mathcal{R}^{I_1, I_3}$ that satisfies $(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{\nu}(\boldsymbol{z})) \in \hat{\mathcal{W}}$ for all $\boldsymbol{z} \in \mathcal{W}$. We provide an explicit example as follows:

Example 2.1.1. The extended support set for

$$
\begin{aligned}
\mathcal{W} &= \left\{\boldsymbol{z} \in \Re^{I_1} \ : \ \|\boldsymbol{z}\|_1 \leq \Gamma, \|\boldsymbol{z}\|_\infty \leq 1\right\} \\
\mathcal{G} &= \left\{(\boldsymbol{z}, \boldsymbol{u}) \in \Re^{I_1} \times \Re^4 : \boldsymbol{g}(\boldsymbol{z}) \leq \boldsymbol{u}\right\} \\
g_1(\boldsymbol{z}) &= |\boldsymbol{a}_1{}'\boldsymbol{z}| \\
g_2(\boldsymbol{z}) &= (\boldsymbol{a}_2{}'\boldsymbol{z})^2 \\
g_3(\boldsymbol{z}) &= ((\boldsymbol{a}_3{}'\boldsymbol{z})^+)^3 \\
g_4(\boldsymbol{z}) &= \min\{\boldsymbol{w}'\boldsymbol{v} \ : \ \boldsymbol{H}\boldsymbol{v} \succeq_{\mathcal{K}} \boldsymbol{f} + \boldsymbol{F}\boldsymbol{z}\}
\end{aligned}
$$

is given by

$$
\hat{\mathcal{W}} = \left\{ \left(\boldsymbol{z}, \boldsymbol{u}, \begin{pmatrix} v_1 \\ v_2 \\ \boldsymbol{v}_3 \\ \boldsymbol{v}_4 \end{pmatrix} \right) \in \Re^{I_1} \times \Re^4 \times \Re^{I_3} \ : \ \begin{array}{l}
\boldsymbol{a}_1\boldsymbol{z} \leq u_1, -\boldsymbol{a}_1\boldsymbol{z} \leq u_1 \\
\sqrt{((\boldsymbol{a}_2{}'\boldsymbol{z})^2 + \left(\frac{u_2-1}{2}\right)^2} \leq \frac{u_2+1}{2} \\
v_1 \geq 0, v_1 \geq \boldsymbol{a}_3{}'\boldsymbol{z} \\
\sqrt{v_1^2 + \left(\frac{v_2-1}{2}\right)^2} \leq \frac{v_2+1}{2} \\
\sqrt{v_2^2 + \left(\frac{v_1-u_3}{2}\right)^2} \leq \frac{v_1+u_3}{2} \\
\boldsymbol{w}'\boldsymbol{v}_3 \leq u_4 \\
\boldsymbol{H}\boldsymbol{v}_3 \succeq_{\mathcal{K}} \boldsymbol{f} + \boldsymbol{F}\boldsymbol{z} \\
\boldsymbol{v}_4 \geq \boldsymbol{z}, \boldsymbol{v}_4 \geq -\boldsymbol{z} \\
\boldsymbol{1}'\boldsymbol{v}_4 \leq \Gamma, -\boldsymbol{1} \leq \boldsymbol{z} \leq \boldsymbol{1}
\end{array} \right\}
$$

Given $\boldsymbol{z} \in \mathcal{W}$, we can verify that $u_1 = |\boldsymbol{a}_1{}'\boldsymbol{z}|$, $u_2 = (\boldsymbol{a}_2{}'\boldsymbol{z})^2$, $u_3 = ((\boldsymbol{a}_3{}'\boldsymbol{z})^+)^3$, $u_4 = \min\{\boldsymbol{d}'\boldsymbol{v} \ : \ \boldsymbol{H}\boldsymbol{v} \geq \boldsymbol{f} + \boldsymbol{F}\boldsymbol{z}\}$, $v_1 = (\boldsymbol{a}_3\boldsymbol{z})^+$, $v_2 = ((\boldsymbol{a}_3\boldsymbol{z})^+)^2$, $\boldsymbol{v}_3 = \arg\min\{\boldsymbol{d}'\boldsymbol{v} \ : \ \boldsymbol{H}\boldsymbol{v} \geq \boldsymbol{f} + \boldsymbol{F}\boldsymbol{z}\}$, $\boldsymbol{v}_4 = (|z_1|, \ldots, |z_{I_1}|)'$ would be feasible in the extended support set $\hat{\mathcal{W}}$. Moreover, $v_1, v_2, \boldsymbol{v}_3$ are those associated with the set $\mathcal{G}$, while $\boldsymbol{v}_4$ is related to the support set, $\mathcal{W}$.

We refer interested readers to Wiesemann et al. (2014) for more information of the expressibility of the ambiguity set. While the ambiguity set is general to include semidefinite constraints, which can capture descriptive statistics such as covariance, we may choose to work with ambiguity sets that are linear or second order conic representation as they will lead to models that can be solved efficiently using state-of-the-art commercial solvers such as CPLEX and Gurobi. We will leave these explorations to future research as the purpose of this paper is to provide the optimization framework as well as the software that we could use to facilitate future studies.

For computational reasons, we impose the following Slater's like conditions:

Assumption 2. *There exists* $(\boldsymbol{z}^\dagger, \boldsymbol{u}^\dagger, \boldsymbol{v}^\dagger) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}$ *such that*

$$\begin{aligned} &\boldsymbol{G}\boldsymbol{z}^\dagger = \boldsymbol{\mu} \\ &\boldsymbol{u}^\dagger < \boldsymbol{\sigma} \\ &\boldsymbol{C}\boldsymbol{z}^\dagger + \boldsymbol{D}\boldsymbol{u}^\dagger + \boldsymbol{E}\boldsymbol{v}^\dagger \prec_{\mathcal{K}} \boldsymbol{h}. \end{aligned}$$

Hence, $(\boldsymbol{z}^\dagger, \boldsymbol{u}^\dagger, \boldsymbol{v}^\dagger) \in \mathrm{int}\hat{\mathcal{W}}$.

Given the ambiguity set, $\mathbb{F}$, we assume that the decision maker is ambiguity averse and the second stage cost is evaluated based on the worst case expectation over the ambiguity set given by

$$\beta(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(Q(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right). \tag{2.5}$$

Corresponding, the *here-and-now* decision is determined by minimizing the sum of the deterministic first stage cost and the worst-case expected second stage cost over the ambiguity set as follows:

$$\begin{aligned} \min \quad & \boldsymbol{c}'\boldsymbol{x} + \beta(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{x} \in X_1. \end{aligned} \tag{2.6}$$

More generally, the second stage can involve a collection of K attributes $\beta_k(\boldsymbol{x})$, $k \in [K]$, each having similar structure as $\beta(\boldsymbol{x})$ and the generalized model we solve is as follows:

$$\begin{aligned} Z^* = \quad \min \quad & \boldsymbol{c}'\boldsymbol{x} + \beta(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{c}_k{}'\boldsymbol{x} + \beta_k(\boldsymbol{x}) \leq \rho_k \quad \forall k \in [K] \\ & \boldsymbol{x} \in X_1, \end{aligned} \tag{2.7}$$

with $\boldsymbol{c}_k \in \Re^{N_1}$, $k \in [K]$ and $\boldsymbol{\rho} \in \Re^K$. For simplicity, we will focus on deriving the exact reformulation of $\beta(\boldsymbol{x})$, which could then be integrated in Problem (2.6) to obtain the optimum *here-and-now* decision, $\boldsymbol{x} \in X_1$. Naturally, similar reformulations can be extended to derive the epigraphs of $\beta_k(\boldsymbol{x})$, $k \in [K]$, which could be incorporated into Problem (2.7) to obtain a tractable optimization problem.

Observe that Problem (2.5) involves optimization of probability measures over a family of distributions and hence, it is not a finite dimensional optimization problem. Motivated from Wiesemann et al. (2014), we define the extended ambiguity set, $\mathbb{G}$ which involves auxiliary random variables over the extended support set $\hat{\mathcal{W}}$ as follows:

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}\right) : \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3} \\ \mathbb{E}_{\mathbb{Q}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma} \\ \mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \hat{\mathcal{W}}\right) = 1 \end{array} \right\}. \tag{2.8}$$

Proposition 1. *The ambiguity set $\mathbb{F}$ in (2.2) is equivalent to the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}$, for all $\mathbb{Q} \in \mathbb{G}$, i.e.,*

$$\mathbb{F} = \prod_{\tilde{\boldsymbol{z}}} \mathbb{G}.$$

In particular, for a function $\boldsymbol{\nu} \in \mathcal{R}^{I_1, I_3}$ *satisfying* $(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{\nu}(\boldsymbol{z})) \in \hat{\mathcal{W}}$ *for all* $\boldsymbol{z} \in \mathcal{W}$ *and* $\mathbb{P} \in \mathbb{F}$*, the probability distribution* $\mathbb{Q} \in \mathcal{P}_0\left(\Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}\right)$ *associated with the random variable* $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}$ *such that*

$$(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) = (\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}}), \boldsymbol{\nu}(\tilde{\boldsymbol{z}})) \ \ \mathbb{P}\text{-}a.s.$$

also lies in $\mathbb{G}$*.*

Proof. The proof is rather straightforward and a variant is presented in Wiesemann et al. (2014). We first show that $\prod_{\tilde{\boldsymbol{z}}} \mathbb{G} \subseteq \mathbb{F}$. Indeed, for any $\mathbb{Q} \in \mathbb{G}$, and $\mathbb{P} = \prod_{\tilde{\boldsymbol{z}}} \mathbb{Q}$, we have $\mathbb{E}_{\mathbb{P}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \mathbb{E}_{\mathbb{Q}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu}$. Moreover, since $\mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \hat{\mathcal{W}}\right) = 1$, we have $\mathbb{Q}\left(\tilde{\boldsymbol{z}} \in \mathcal{W}\right) = 1$ and $\mathbb{Q}(\boldsymbol{g}(\tilde{\boldsymbol{z}}) \leq \tilde{\boldsymbol{u}}) = 1$. Hence, $\mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1$ and

$$\mathbb{E}_{\mathbb{P}}(\boldsymbol{g}(\tilde{\boldsymbol{z}})) = \mathbb{E}_{\mathbb{Q}}\left(\boldsymbol{g}(\tilde{\boldsymbol{z}})\right) \leq \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma}.$$

Conversely, suppose $\mathbb{P} \in \mathbb{F}$, we observe that $\mathbb{P}\left((\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}})) \in \hat{\mathcal{W}}\right) = 1$. Since $(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{\nu}(\boldsymbol{z})) \in \hat{\mathcal{W}}$ for all $\boldsymbol{z} \in \mathcal{W}$, we can then construct a probability distribution $\mathbb{Q} \in \mathcal{P}_0\left(\Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}\right)$ associated with the random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}$ so that

$$(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) = (\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}}), \boldsymbol{\nu}(\tilde{\boldsymbol{z}})) \ \ \mathbb{P}\text{-a.s.}$$

Observe that

$$\mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) = \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{g}(\tilde{\boldsymbol{z}})\right) \leq \boldsymbol{\sigma}$$

and

$$\mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \hat{\mathcal{W}}\right) = 1.$$

Hence, $\mathbb{F} \subseteq \prod_{\tilde{\boldsymbol{z}}} \mathbb{G}$. $\square$

Exact reformulation

Before we derive an exact reformulation for evaluating $\beta(\boldsymbol{x})$, $x \in X_1$, we need to compute the worst case expectation of a piecewise linear convex function.

Proposition 2. *Let $U \in \mathcal{R}^{I_1,1}$ be a piecewise linear convex function given by*

$$U(\boldsymbol{z}) = \max_{p \in [P]} \{\boldsymbol{\zeta}_p{}'\tilde{\boldsymbol{z}} + \zeta_p^0\}$$

for some $\boldsymbol{\zeta}_p \in \Re^{I_1}, \zeta_p^0 \in \Re$, $p \in [P]$. Suppose

$$\beta^* = \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}(U(\tilde{\boldsymbol{z}}))$$

is finite, then it can be expressed as a standard robust counterpart problem

$$\begin{aligned} \beta^* = \quad \min \quad & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.} \quad & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq U(\boldsymbol{z}) \quad \forall (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \end{aligned} \tag{2.9}$$

or equivalently

$$\begin{aligned} \beta^* = \quad \min \quad & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.} \quad & r \geq \boldsymbol{\pi}_p{}'\boldsymbol{h} + \zeta_p^0 && \forall p \in [P] \\ & \boldsymbol{C}'\boldsymbol{\pi}_p = \boldsymbol{\zeta}_p - \boldsymbol{G}'s && \forall p \in [P] \\ & \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} && \forall p \in [P] \\ & \boldsymbol{E}'\boldsymbol{\pi}_p = \boldsymbol{0}, && \forall p \in [P] \\ & \boldsymbol{\pi}_p \succeq_{\mathcal{K}^*} \boldsymbol{0} && \forall p \in [P] \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\ & \boldsymbol{\pi}_p \in \Re^{L_3} && \forall p \in [P]. \end{aligned} \tag{2.10}$$

Proof. Note that a more general result can be found in Wiesemann et al.

(2014). We present an elementary proof, which would be beneficial to readers who may not be familiar with such transformation. From Proposition 1, we have equivalently

$$\beta^* = \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\max_{p\in[P]} \{ \boldsymbol{\zeta}_p{}' \tilde{\boldsymbol{z}} + \zeta_p^0 \} \right) = \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{P}} \left(\max_{p\in[P]} \{ \boldsymbol{\zeta}_p{}' \tilde{\boldsymbol{z}} + \zeta_p^0 \} \right).$$

By weak duality (referring to Isii (1962)), we have the following semi-infinite optimization problem

$$\begin{aligned} \beta^* \leq \beta_1^* = \quad & \inf \quad r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ & \text{s.t.} \quad r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \max_{p\in[P]} \{ \boldsymbol{\zeta}_p{}' \boldsymbol{z} + \zeta_p^0 \} \quad \forall (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \qquad \boldsymbol{t} \geq \boldsymbol{0} \\ & \qquad r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2}, \end{aligned}$$

where $r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2}$ are the dual variables corresponding to the expectation constraints of $\mathbb{G}$. This is also equivalent to

$$\begin{aligned} \beta_1^* = \quad & \inf \quad r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ & \text{s.t.} \quad r \geq \sup_{(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v})\in\hat{\mathcal{W}}} \left\{ (\boldsymbol{\zeta}_p - \boldsymbol{G}'\boldsymbol{s})'\boldsymbol{z} - \boldsymbol{t}'\boldsymbol{u} + \zeta_p^0 \right\} \quad \forall p \in [P] \\ & \qquad \boldsymbol{t} \geq 0 \\ & \qquad r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2}, \end{aligned} \tag{2.11}$$

By weak conic duality (see, for instance, Ben-Tal and Nemirovski (2001a)), we have for all $p \in [P]$,

$$\begin{aligned}\sup_{(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v})\in\hat{\mathcal{W}}} \left\{(\boldsymbol{\zeta}_p - \boldsymbol{G}'\boldsymbol{s})'\boldsymbol{z} - \boldsymbol{t}'\boldsymbol{u} + \zeta_p^0\right\} \leq \quad \inf \quad & \boldsymbol{\pi_p}'\boldsymbol{h} + \zeta_p^0 \\ \text{s.t.} \quad & \boldsymbol{C}'\boldsymbol{\pi}_p = \boldsymbol{\zeta}_p - \boldsymbol{G}'\boldsymbol{s} \\ & \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} \\ & \boldsymbol{E}'\boldsymbol{\pi}_p = \boldsymbol{0} \\ & \boldsymbol{\pi}_p \succeq_{\mathcal{K}^*} \boldsymbol{0} \\ & \boldsymbol{\pi}_p \in \Re^{L_3} \quad p \in [P],\end{aligned}$$

where $\boldsymbol{\pi}_p \in \Re^{L_3}, \forall p \in [P]$ are the dual variables associated with the conic constants in $\hat{\mathcal{W}}$. Hence, using standard robust counterpart techniques, we substitute the dual formulations in Problem (2.11) to yield the following compact conic optimization problem

$$\begin{aligned}\beta_2^* = \inf \quad & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.} \quad & r \geq \boldsymbol{\pi}_p{}'\boldsymbol{h} + \zeta_p^0 && \forall p \in [P] \\ & \boldsymbol{C}'\boldsymbol{\pi}_p = \boldsymbol{\zeta}_p - \boldsymbol{G}'\boldsymbol{s} && \forall p \in [P] \\ & \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} && \forall p \in [P] \\ & \boldsymbol{E}'\boldsymbol{\pi}_p = 0 && \forall p \in [P] \\ & \boldsymbol{\pi}_p \succeq_{\mathcal{K}^*} 0 && \forall p \in [P] \\ & t \geq 0 \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\ & \boldsymbol{\pi}_p \in \Re^{L_3} && \forall p \in [P].\end{aligned} \tag{2.12}$$

Observe that $\beta^* \leq \beta_1^* \leq \beta_2^*$, and our goal is to establish strong duality by showing $\beta_2^* \leq \beta^*$. Then we will next approach Problem (2.12) by taking the dual, which is

$$
\begin{aligned}
\beta_3^* = \quad \sup \quad & \sum_{p\in[P]} \left(\zeta_p^0\alpha_p + \boldsymbol{\zeta}_p{}'\bar{\boldsymbol{z}}_p\right) \\
\text{s.t.} \quad & \sum_{p\in[P]} \alpha_p = 1 \\
& \alpha_p \geq 0 && \forall p \in [P] \\
& \sum_{p\in[P]} \boldsymbol{G}\bar{\boldsymbol{z}}_p = \boldsymbol{\mu} \\
& \sum_{p\in[P]} \bar{\boldsymbol{u}}_p \leq \boldsymbol{\sigma} \\
& \boldsymbol{C}\bar{\boldsymbol{z}}_p + \boldsymbol{D}\bar{\boldsymbol{u}}_p + \boldsymbol{E}\bar{\boldsymbol{v}}_p \preceq_{\mathcal{K}} \alpha_p\boldsymbol{h} && \forall p \in [P] \\
& \alpha_p \in \Re, \bar{\boldsymbol{z}}_p \in \Re^{I_1}, && \forall p \in [P] \\
& \bar{\boldsymbol{u}}_p \in \Re^{I_2}, \bar{\boldsymbol{v}}_p \in \Re^{I_3} && \forall p \in [P].
\end{aligned}
\tag{2.13}
$$

Suppose $(\boldsymbol{z}^\dagger, \boldsymbol{u}^\dagger, \boldsymbol{v}^\dagger) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}$ satisfy the conditions in Assumption 2, then we can construct a strictly feasible solution

$$
\alpha_p = \frac{1}{P}, \bar{\boldsymbol{z}}_p = \frac{\boldsymbol{z}^\dagger}{P}, \bar{\boldsymbol{u}}_p = \frac{\boldsymbol{u}^\dagger}{P}, \bar{\boldsymbol{v}}_p = \frac{\boldsymbol{v}^\dagger}{P},
$$

for all $\forall p \in [P]$. Hence, since Problem (2.13) is strictly feasible and, as we will subsequently show, is also bounded from above, strong duality holds and $\beta_2^* = \beta_3^*$. Moreover, there exists a sequence of strictly feasible or interior solutions

$$
\left\{(\alpha_p^k, \bar{\boldsymbol{z}}_p^k, \bar{\boldsymbol{u}}_p^k, \bar{\boldsymbol{v}}_p^k)_{p\in[P]}\right\}_{k\geq 0}
$$

such that

$$
\lim_{k\to\infty} \sum_{p\in[P]} \left(\zeta_p^0\alpha_p^k + \boldsymbol{\zeta}_p{}'\bar{\boldsymbol{z}}_p^k\right) = \beta_3^*.
$$

Observe that for all k, $\alpha_p^k > 0$, $\sum_{p\in[P]} \alpha_p^k = 1$ and we can construct a sequence of discrete probability distributions $\{\mathbb{Q}_k \in \mathcal{P}_0\left(\Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}\right)\}_{k\geq 0}$ on random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3}$ such that

$$\mathbb{Q}_k \left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) = \left(\frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k}, \frac{\bar{\boldsymbol{u}}_p^k}{\alpha_p^k}, \frac{\bar{\boldsymbol{v}}_p^k}{\alpha_p^*} \right) \right) = \alpha_p^k \qquad \forall p \in [P].$$

Observe that,

$$\mathbb{E}_{\mathbb{Q}_k}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{Q}_k}(\tilde{\boldsymbol{u}}) \le \boldsymbol{\sigma}, \mathbb{Q}_k((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \hat{\mathcal{W}}) = 1,$$

and hence $\mathbb{Q}_k \in \mathbb{G}$ for all k. Moreover,

$$\begin{aligned}
\beta_3^* &= \lim_{k\to\infty} \sum_{p\in[P]} \left(\zeta_p^0 \alpha_p^k + \boldsymbol{\zeta}_p{}' \bar{\boldsymbol{z}}_p^k \right) \\
&= \lim_{k\to\infty} \sum_{p\in[P]} \alpha_p^k \left(\zeta_p^0 + \boldsymbol{\zeta}_p{}' \frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k} \right) \\
&\le \lim_{k\to\infty} \sum_{p\in[P]} \alpha_p^k \left(\max_{q\in[P]} \left\{ \zeta_q^0 + \boldsymbol{\zeta}_q{}' \frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k} \right\} \right) \\
&= \lim_{k\to\infty} \mathbb{E}_{\mathbb{Q}_k} \left(\max_{q\in[P]} \{ \zeta_q^0 + \boldsymbol{\zeta}_q{}' \tilde{\boldsymbol{z}} \} \right) \\
&\le \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(U(\tilde{\boldsymbol{z}})) \\
&= \beta^*.
\end{aligned}$$

Hence, $\beta^* \le \beta_1^* \le \beta_2^* = \beta_3^* \le \beta^*$, and strong duality holds. Since β^* is finite, Problem (2.13) is bounded from above and hence, Problem (2.12) also solvable.

□

Noting that $Q(\boldsymbol{x}, \boldsymbol{z})$, $\boldsymbol{x} \in X_1$ is also a piecewise linear convex function of $\boldsymbol{z}$, we can easily extend Proposition 2 so that the function $\beta(\boldsymbol{x})$ can be evaluated and integrated in epigraphical form to solve Problem (2.7) as a standard optimization problem.

Theorem 1. *Let $\{\boldsymbol{p}^1, ..., \boldsymbol{p}^P\}$ be the set of all extreme points of the polyhedra*

$$\mathcal{P} = \left\{ \boldsymbol{p} \in \Re^M : \begin{array}{c} \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d} \\ \boldsymbol{p} \ge \boldsymbol{0} \end{array} \right\}.$$

For a given subset of extreme points indices, $\mathcal{S} \subseteq [P]$, we define

$$
\begin{aligned}
\underline{\beta}_{\mathcal{S}}(\boldsymbol{x}) = \quad \min \quad & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.} \quad & r \geq \boldsymbol{\pi}_i{}'\boldsymbol{h} + \boldsymbol{p}_i{}'\boldsymbol{b}^0 - \boldsymbol{p}_i{}'\boldsymbol{A}^0\boldsymbol{x} & \forall i \in \mathcal{S} \\
& \boldsymbol{C}'\boldsymbol{\pi}_i = \begin{bmatrix} \boldsymbol{p}_i'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_i'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} & \forall i \in \mathcal{S} \\
& \boldsymbol{D}'\boldsymbol{\pi}_i = -\boldsymbol{t} & \forall i \in \mathcal{S} \\
& \boldsymbol{E}'\boldsymbol{\pi}_i = \boldsymbol{0} & \forall i \in \mathcal{S} \\
& \boldsymbol{\pi}_i \succeq_{\mathcal{K}^*} \boldsymbol{0} & \forall i \in \mathcal{S} \\
& \boldsymbol{t} \geq \boldsymbol{0} \\
& r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\
& \boldsymbol{\pi}_i \in \Re^{L_3} & \forall i \in \mathcal{S}.
\end{aligned}
\tag{2.14}
$$

If $\beta(\boldsymbol{x})$, $\boldsymbol{x} \in X_1$ is finite, then

$$\underline{\beta}_{\mathcal{S}}(\boldsymbol{x}) \leq \underline{\beta}_{[P]}(\boldsymbol{x}) = \beta(\boldsymbol{x}).$$

Proof. From strong linear optimization duality, we can express Problem (2.1) as

$$
\begin{aligned}
Q(\boldsymbol{x}, \boldsymbol{z}) = \quad \max \quad & \boldsymbol{p}'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{p} \in \mathcal{P}.
\end{aligned}
\tag{2.15}
$$

Since $Q(\boldsymbol{x}, \boldsymbol{z})$ is finite for all $\boldsymbol{x} \in X_1$ (Assumption 1), Problem (2.15) has an extreme point optimum solution for all $\boldsymbol{x} \in X_1$. Hence, we can express Problem (2.15) explicitly as a piecewise linear convex function of $\boldsymbol{z}$ as follows:

$$Q(\boldsymbol{x}, \boldsymbol{z}) = \max_{i \in [P]} \{\boldsymbol{p}_i{}'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x})\},$$

for all $\boldsymbol{x} \in X_1$. Since $\beta(\boldsymbol{x})$ is finite, we can use Theorem 1 to derive the exact reformulation for $\mathcal{S} = [P]$, to achieve $\underline{\beta}(\boldsymbol{x}) = \underline{\beta}_{[P]}(\boldsymbol{x})$. It is trivial to see that if

$\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq [P]$, then

$$\underline{\beta}_{\mathcal{S}_1}(\boldsymbol{x}) \leq \underline{\beta}_{\mathcal{S}_2}(\boldsymbol{x}) \leq \underline{\beta}_{[P]}(\boldsymbol{x}).$$

□

Theorem 1 suggests an approach to compute the exact value of $\beta(\boldsymbol{x})$, which may not be a polynomial sized problem due to possibly exponential number of extreme points. Unfortunately, the "separation problem" associated with finding the extreme point involves solving the following bilinear optimization problem,

$$\max_{\boldsymbol{p}\in\mathcal{P}} \left\{ \sup_{(\boldsymbol{z},\boldsymbol{u})\in\bar{\mathcal{W}}} \left\{ \left(\begin{bmatrix} \boldsymbol{p}'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \right)' \boldsymbol{z} + \boldsymbol{p}'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) - \boldsymbol{t}'\boldsymbol{u} \right\} \right\}$$

which is generally intractable. Nevertheless, Theorem (1) provides an approach to determine the lower bound of $\beta(\boldsymbol{x})$, which might be useful to determine the quality of the solution. We will next show how we can tractably compute the upper bound of $\beta(\boldsymbol{x})$ via linear decision rule approximations.

2.2 Generalized linear decision rules

Observe that any function, $\boldsymbol{y} \in \mathcal{R}^{I_1,N_2}$ satisfying

$$\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \qquad \forall \boldsymbol{z} \in \mathcal{W}$$

would be an upper bound of $\beta(\boldsymbol{x})$, $\boldsymbol{x} \in X_1$, i.e.,

$$\beta(\boldsymbol{x}) \leq \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})).$$

Moreover, equality is achieved if

$$\boldsymbol{y}(\boldsymbol{z}) \in \arg\min\{\boldsymbol{d}'\boldsymbol{y} \ : \ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z})\}$$

for all $\boldsymbol{z} \in \mathcal{W}$. Hence, we can express $\beta(\boldsymbol{x})$, $\boldsymbol{x} \in X_1$ as a minimization problem over all measurable functions as follows:

$$\begin{aligned} \beta(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y} \in \mathcal{R}^{I_1,N_2}. \end{aligned} \tag{2.16}$$

Unfortunately, Problem (2.16) is generally an intractable optimization problem as there could potentially be infinite number of constraints and variables. An upper bound of $\beta(\boldsymbol{x})$ could be computed tractably by restricting $\boldsymbol{y}$ to a smaller class of measurable functions that can be characterized by a polynomial number of decision variables such as those that are affinely dependent on $\boldsymbol{z}$ or so called linear decision rules as follows:

$$\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \sum_{j\in[I_1]} \boldsymbol{y}_j z_j,$$

for some $\boldsymbol{y}^0, \boldsymbol{y}_j \in \Re^{N_2}$, $j \in [I_1]$. However, the following example shows that linear decision rule may even be infeasible in problems with complete recourse.

Example 2.2.1. Consider the following complete recourse problem,

$$\begin{aligned} \beta = \quad \min \quad & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(y(\tilde{z})) \\ \text{s.t.} \quad & y(z) \geq z \qquad \forall z \in \Re \\ & y(z) \geq -z \qquad \forall z \in \Re \\ & y \in \mathcal{R}^{1,1} \end{aligned} \tag{2.17}$$

where

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Re) : \mathbb{E}_{\mathbb{P}}(|\tilde{z}|) \leq 1 \right\}.$$

Clearly, $y(z) = |z|$ is the optimal decision rule that yields $\beta = 1$. However, under a linear decision rule here (i.e., $y(z) = y_0 + y_1 z$ for some $y_0, y_1 \in \Re$, we would encounter the following infeasibility issue

$$\begin{aligned} y_0 + y_1 z &\geq z \quad &\forall z \in \Re \\ y_0 + y_1 z &\geq -z \quad &\forall z \in \Re. \end{aligned} \tag{2.18}$$

Using the extended ambiguity set $\mathbb{G}$, we propose the following generalized linear decision rule to encompass the auxiliary random variables $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ as well. For given subsets $\mathcal{S}_1 \subseteq [I_1]$, $\mathcal{S}_2 \subseteq [I_2]$, $\mathcal{S}_3 \subseteq [I_3]$, we define the following space of affine functions,

$$\mathcal{L}^N(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3) = \left\{ \boldsymbol{y} : \Re^{I_1} \times \Re^{I_2} \times \Re^{I_3} \to \Re^N \middle| \begin{array}{l} \exists \boldsymbol{y}^0, \boldsymbol{y}_i^1, \boldsymbol{y}_j^2, \boldsymbol{y}_k^3 \in \Re^N, \forall i \in \mathcal{S}_1, j \in \mathcal{S}_2, k \in \mathcal{S}_3 : \\ \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{y}^0 + \sum_{i \in \mathcal{S}_1} \boldsymbol{y}_i^1 z_i + \sum_{j \in \mathcal{S}_2} \boldsymbol{y}_j^2 u_j + \sum_{k \in \mathcal{S}_3} \boldsymbol{y}_k^3 v_k \end{array} \right\}.$$

This decision rule generalizes the traditional linear decision rules that depends only on the underlying uncertainty, $\tilde{\boldsymbol{z}}$, in which case, we have $\mathcal{S}_2 = \mathcal{S}_3 = \emptyset$. The segregated and extended linear decision rules found in Chen and Zhang (2009), Chen et al. (2008), Goh and Sim (2010) are special cases of having $\mathcal{S}_3 \subseteq \bar{\mathcal{I}}_3$, which incorporate auxiliary variables of the support set in the generalized linear decision rule. Based in the generalized linear decision rules, we obtain an upper bound of $\beta(\boldsymbol{x})$, $\boldsymbol{x} \in X_1$ as follows:

$$\begin{aligned} \bar{\beta}_{(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{Q} \in \mathbb{G}} \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})) \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{y} \in \mathcal{L}^{N_2}(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3). \end{aligned} \tag{2.19}$$

As the linear decision rule incorporates more auxiliary random variables, the quality of the bound improves, albeit at the expense of increased model size.

Proposition 3. *Given* $\boldsymbol{x} \in X_1$, *and* $\mathcal{S}_1 \subseteq \bar{\mathcal{S}}_1 \subseteq [I_1]$, $\mathcal{S}_2 \subseteq \bar{\mathcal{S}}_2 \subseteq [I_2]$, *and*

$\mathcal{S}_3 \subseteq \bar{\mathcal{S}}_3 \subseteq [I_3]$, *we have*

$$\beta(\boldsymbol{x}) \leq \bar{\beta}_{([I_1],[I_2],[I_3])}(\boldsymbol{x}) \leq \bar{\beta}_{(\bar{\mathcal{S}}_1,\bar{\mathcal{S}}_2,\bar{\mathcal{S}}_3)}(\boldsymbol{x}) \leq \bar{\beta}_{(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)}(\boldsymbol{x}).$$

Proof. The proof is trivial and hence omitted. □

Proposition 4. *For $\boldsymbol{x} \in X_1$, Problem (2.19) is equivalent to the following robust counterpart problem,*

$$\begin{aligned} \bar{\beta}_{(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)}(\boldsymbol{x}) = \quad \min \quad & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.} \quad & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \quad \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v})) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\ & \boldsymbol{y} \in \mathcal{L}^{N_2}(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3), \end{aligned} \tag{2.20}$$

or explicitly as

$$
\begin{array}{rll}
\bar{\beta}_{(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)}(\boldsymbol{x}) = & \min & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
& \text{s.t.} & r - \boldsymbol{d}'\boldsymbol{y}^0 \geq \boldsymbol{\pi}'\boldsymbol{h} \\
& & [\boldsymbol{C}'\boldsymbol{\pi}]_i = \boldsymbol{d}'\boldsymbol{y}_i^1 - [\boldsymbol{G}'\boldsymbol{s}]_i \qquad \forall i \in \mathcal{S}_1 \\
& & [\boldsymbol{C}'\boldsymbol{\pi}]_i = -[\boldsymbol{G}'\boldsymbol{s}]_i \qquad \forall i \in [I_1] \setminus \mathcal{S}_1 \\
& & [\boldsymbol{D}'\boldsymbol{\pi}]_j = \boldsymbol{d}'\boldsymbol{y}_j^2 - [\boldsymbol{t}]_j \qquad \forall j \in \mathcal{S}_2 \\
& & [\boldsymbol{D}'\boldsymbol{\pi}]_j = -[\boldsymbol{t}]_j \qquad \forall j \in [I_2] \setminus \mathcal{S}_2 \\
& & [\boldsymbol{E}'\boldsymbol{\pi}]_k = \boldsymbol{d}'\boldsymbol{y}_k^3 \qquad \forall k \in \mathcal{S}_3 \\
& & [\boldsymbol{E}'\boldsymbol{\pi}]_k = 0 \qquad \forall k \in [I_3] \setminus \mathcal{S}_3 \\
& & \left[\boldsymbol{A}^0\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^0 - \boldsymbol{b}^0\right]_l \geq \boldsymbol{\tau}_l'\boldsymbol{h} \qquad \forall l \in [M] \\
& & [\boldsymbol{C}'\boldsymbol{\tau}_l]_i = [\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x} - \boldsymbol{B}\boldsymbol{y}_i^1]_l \qquad \forall l \in [M], \forall i \in \mathcal{S}_1 \\
& & [\boldsymbol{C}'\boldsymbol{\tau}_l]_i = [\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x}]_l \qquad \forall l \in [M], \forall i \in [I_1] \setminus \mathcal{S}_1 \\
& & [\boldsymbol{D}'\boldsymbol{\tau}_l]_j = [-\boldsymbol{B}\boldsymbol{y}_j^2]_l \qquad \forall l \in [M], \forall j \in \mathcal{S}_2 \\
& & [\boldsymbol{D}'\boldsymbol{\tau}_l]_j = 0 \qquad \forall l \in [M], \forall j \in [I_2] \setminus \mathcal{S}_2 \\
& & [\boldsymbol{E}'\boldsymbol{\tau}_l]_k = [-\boldsymbol{B}\boldsymbol{y}_k^3]_l \qquad \forall l \in [M], \forall k \in \mathcal{S}_3 \\
& & [\boldsymbol{E}'\boldsymbol{\tau}_l]_k = 0 \qquad \forall l \in [M], \forall k \in [I_3] \setminus \mathcal{S}_3 \\
& & \boldsymbol{\pi} \succeq_{\mathcal{K}^*} \boldsymbol{0} \\
& & \boldsymbol{\tau}_l \succeq_{\mathcal{K}^*} \boldsymbol{0} \qquad \forall l \in [M] \\
& & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\
& & \boldsymbol{\pi}, \boldsymbol{\tau}_l \in \Re^{L_3}, \forall l \in [M].
\end{array}
\tag{2.21}
$$

Proof. The proof follows from Proposition 2 and hence omitted. □

In Example 2.2.1, we show that a linear decision rule that depends solely on $\tilde{\boldsymbol{z}}$ may become infeasible if the support is unbounded. Suppose, the absolute deviations of $\tilde{\boldsymbol{z}}$ are bounded, we show that there exists a generalized linear decision rule involving the axillary random variable $\tilde{\boldsymbol{u}}$ that could resolve the infeasibility issue.

Theorem 2. *Suppose Problem (2.19) has complete recourse, then exists a gen-*

eralized linear decision rule

$$\boldsymbol{y} \in \mathcal{L}^N(\emptyset, [I_1], \emptyset),$$

that is feasible in Problem (2.20) for the following family of distributions with bounded absolute deviations

$$\mathbb{F}_1 = \left\{\mathbb{P} \in \mathcal{P}_0(\Re^{I_1}) : \mathbb{E}_{\mathbb{P}}(|z_i|) \le \sigma_i \quad \forall i \in [I_1]\right\}, \quad \boldsymbol{\sigma} > \boldsymbol{0}.$$

Proof. The extended ambiguity set associate with $\mathbb{F}_1$ is

$$\mathbb{G}_1 = \left\{ \mathbb{Q} \in \mathcal{P}_0(\Re^{I_1} \times \Re^{I_1}) : \begin{array}{l} \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \le \boldsymbol{\sigma} \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \hat{\mathcal{W}}) = 1 \end{array} \right\},$$

in which the extended support set is $\hat{\mathcal{W}} = \{(\boldsymbol{z}, \boldsymbol{u}) \in \Re^{I_1} \times \Re^{I_1} \ : \ \boldsymbol{u} \ge \boldsymbol{z}, \boldsymbol{u} \ge -\boldsymbol{z}\}$. The linear decision rule $\boldsymbol{y} \in \mathcal{L}^N(\emptyset, [I_1], \emptyset)$ is given by

$$\boldsymbol{y}(\boldsymbol{u}) = \boldsymbol{y}^0 + \sum_{i \in [I_1]} \boldsymbol{y}_i^2 u_i.$$

Using these parameters, we need to show that the linear decision rule $\boldsymbol{y}(\boldsymbol{u})$ is feasible in the following problem,

$$\begin{array}{lll} \min & r + \boldsymbol{t}'\boldsymbol{\sigma} & \\ \text{s.t.} & r + \boldsymbol{t}'\boldsymbol{u} \ge \boldsymbol{d}'\boldsymbol{y}^0 + \displaystyle\sum_{i \in I_1} \boldsymbol{d}'\boldsymbol{y}_i^2 u_i & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \hat{\mathcal{W}} \\ & \boldsymbol{B}\boldsymbol{y}^0 + \displaystyle\sum_{i \in I_1} \boldsymbol{B}\boldsymbol{y}_i^2 u_i \ge \boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x} + \displaystyle\sum_{i \in I_1} (\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x}) z_i & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \hat{\mathcal{W}} \\ & \boldsymbol{t} \ge \boldsymbol{0} & \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} & \\ & \boldsymbol{y}(\boldsymbol{u}) = \boldsymbol{y}^0 + \displaystyle\sum_{i \in I_1} \boldsymbol{y}_i^2 u_i & \\ & \boldsymbol{y}^0, \boldsymbol{y}_i^2 \in \Re^{N_2} \quad i \in [I_1]. & \end{array} \tag{2.22}$$

Since $\boldsymbol{B}$ is complete recourse matrix, there exists $\bar{\boldsymbol{y}}^0, \bar{\boldsymbol{y}}_i^2$ $i \in [I_1]$, such that

$$\boldsymbol{B}\bar{\boldsymbol{y}}^0 \geq \boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x},$$

$$\boldsymbol{B}\bar{\boldsymbol{y}}_i^2 \geq (\boldsymbol{b}^j - \boldsymbol{A}^j\boldsymbol{x}), \quad \boldsymbol{B}\bar{\boldsymbol{y}}_i^2 \geq -(\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x}) \;\; \forall i \in [I_1].$$

Observe that given any $a \in \Re$, $\boldsymbol{b} \in \Re^{I_1}$

$$x + \boldsymbol{y}'\boldsymbol{u} \geq a + \boldsymbol{b}'\boldsymbol{z} \qquad \forall(\boldsymbol{z}, \boldsymbol{u}) \in \hat{\mathcal{W}}$$

if $x \geq a$, and $y_i \geq |b_i|$, $i \in [I_1]$. Hence, a feasible solution for Problem (2.22) would be

$$\begin{aligned} r &= \boldsymbol{d}'\bar{\boldsymbol{y}}^0 \\ t_i &= \max\{\boldsymbol{d}'\bar{\boldsymbol{y}}^i, 0\} \quad \forall i \in [I_1] \\ \boldsymbol{y}^j &= \bar{\boldsymbol{y}}^j \qquad\qquad\qquad \forall j \in \{0\} \cup [I_1] \end{aligned}$$

□

The generalized linear decision rule achieves the exact value of $\beta(\boldsymbol{x})$ for the following instance.

Theorem 3. *For a complete recourse problem with $N_2 = 1$ and finite $\beta(\boldsymbol{x})$, we have*

$$\beta(\boldsymbol{x}) = \bar{\beta}_{([I_1],[I_2],\emptyset)}(\boldsymbol{x}).$$

Proof. For $N_2 = 1$, the complete recourse matrix $\boldsymbol{B} \in \Re^{M\times 1}$ must satisfy either $\boldsymbol{B} > \boldsymbol{0}$ or $\boldsymbol{B} < \boldsymbol{0}$. Observe that the problem

$$\begin{aligned} Q(\boldsymbol{x}, \boldsymbol{z}) = \quad & \min \quad dy \\ & \text{s.t.} \quad \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}y \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \qquad\;\; y \in \Re, \end{aligned}$$

is unbounded below whenever $d\boldsymbol{B} < \boldsymbol{0}$. Since $\beta(\boldsymbol{x})$ is finite and the second stage decision variable y is unconstrained, we can assume without loss of generality that $\boldsymbol{B} > 0$ and $d \geq 0$. In which case,

$$Q(\boldsymbol{x},\boldsymbol{z}) = d \max_{i\in[M]} \left\{ \frac{[\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_i}{[\boldsymbol{B}]_i} \right\}.$$

Hence, applying Proposition 2, we have

$$\begin{aligned} \beta(\boldsymbol{x}) = \quad \min \quad & d(r + \boldsymbol{s}' + \boldsymbol{t}'\boldsymbol{\sigma}) \\ \text{s.t.} \quad & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \tfrac{[\boldsymbol{b}(\boldsymbol{z})-\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_i}{[\boldsymbol{B}]_i} \quad \forall i \in [M], \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2}. \end{aligned} \tag{2.23}$$

The solution derived under generalized linear decision rule is

$$\begin{aligned} \bar{\beta}_{([I_1],[I_2],\emptyset)}(\boldsymbol{x}) = \quad \min \quad & r + \boldsymbol{s}' + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.} \quad & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq dy(\boldsymbol{z},\boldsymbol{u}) \quad \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}y(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\ & y \in \mathcal{L}([I_1],[I_2],\emptyset), \end{aligned}$$

or equivalently

$$\begin{aligned} \bar{\beta}_{([I_1],[I_2],\emptyset)}(\boldsymbol{x}) = \quad \min \quad & r + \boldsymbol{s}' + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.} \quad & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq d(y^0 + \boldsymbol{y}^{1\prime}\boldsymbol{z} + \boldsymbol{y}^{2\prime}\boldsymbol{u}) \quad \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & (y^0 + \boldsymbol{y}^{1\prime}\boldsymbol{z} + \boldsymbol{y}^{2\prime}\boldsymbol{u}) \geq \tfrac{[\boldsymbol{b}(\boldsymbol{z})-\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_i}{[\boldsymbol{B}]_i} \quad \forall i \in [M], \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\ & y^0 \in \Re, \boldsymbol{y}^1 \in \Re^{I_1}, \boldsymbol{y}^2 \in \Re^{I_2}. \end{aligned} \tag{2.24}$$

Let $(r^\dagger, \boldsymbol{s}^\dagger, \boldsymbol{t}^\dagger)$ be a feasible solution of Problem (2.23). We can construct a feasible solution $(r, \boldsymbol{s}, \boldsymbol{t}, y^0, \boldsymbol{y}^1, \boldsymbol{y}^2)$ to Problem (2.24) by letting

$$y^0 = r, \boldsymbol{y}^1 = \boldsymbol{G}'\boldsymbol{s}, \boldsymbol{y}^2 = \boldsymbol{t}, r = dr^\dagger, \boldsymbol{s} = d\boldsymbol{s}^\dagger, \boldsymbol{t} = d\boldsymbol{t}^\dagger,$$

which yields the same objective as Problem (2.23). Hence, $\bar{\beta}_{([I_1],[I_2],\phi)}(\boldsymbol{x}) \leq \beta(\boldsymbol{x})$ and equality is achieved from Proposition 3. $\square$

Improvement over deflected linear decision rules

Chen et al. (2008), Goh and Sim (2010) propose a class of of piecewise linear decision rules known as deflected linear decision rules which can also circumvent the issues of infeasibility in complete recourse problems. The approach requires to solve a set of subproblems given by

$$\begin{aligned} f_i^* = \quad \min \quad & \boldsymbol{d}'\boldsymbol{y} \\ \text{s.t.} \quad & \boldsymbol{B}\boldsymbol{y} = \boldsymbol{q} \\ & \boldsymbol{q} \geq \boldsymbol{e}_i \\ & \boldsymbol{y} \in \Re^{N_2}, \boldsymbol{q} \in \Re^M, \end{aligned} \tag{2.25}$$

for all $i \in [M]$, which are not necessarily feasible optimization problems. Let $\mathcal{M} \subseteq [M]$ denote the subset of indices in which their corresponding subproblems are feasible, i.e., $\mathcal{M} = \{i \in [M] \ : \ f_i^* < \infty\}$, and $\bar{\mathcal{M}} = [M] \setminus \mathcal{M}$. Correspondingly, let $(\bar{\boldsymbol{y}}_i, \bar{\boldsymbol{q}}_i)$ be the optimal solution of Problem (2.25) for all $i \in \mathcal{M}$. Here, $f_i^* = \boldsymbol{d}'\bar{\boldsymbol{y}}_i \geq 0$, $i \in \mathcal{M}$ is assumed or otherwise, $Q(\boldsymbol{x}, \boldsymbol{z})$ would be unbounded from below. The solution to deflected linear decision is obtained by solving the following optimization problem,

$$
\begin{aligned}
\bar{\beta}_{DLDR}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) + \sum_{i\in\mathcal{M}} f_i^* \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}((-q_i(\tilde{\boldsymbol{z}}))^+) \\
\text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{b}(\boldsymbol{z}) + \boldsymbol{q}(\boldsymbol{z}) && \forall \boldsymbol{z} \in \mathcal{W} \\
& q_i(\boldsymbol{z}) \geq 0 && \forall i \in \bar{\mathcal{M}}, \forall \boldsymbol{z} \in \mathcal{W} \\
& \boldsymbol{y} \in \mathcal{L}^{N_2}([I_1], \emptyset, \emptyset) \\
& \boldsymbol{q} \in \mathcal{L}^{M}([I_1], \emptyset, \emptyset).
\end{aligned}
\tag{2.26}
$$

Suppose $(\boldsymbol{y}^*, \boldsymbol{q}^*)$ is the optimal solution of Problem (2.26), the corresponding deflected linear decision rule is given by

$$
\boldsymbol{y}_{DLDR}(\boldsymbol{z}) = \boldsymbol{y}^*(\boldsymbol{z}) + \sum_{i\in\mathcal{M}} \bar{\boldsymbol{y}}_i((-q_i^*(\boldsymbol{z}))^+).
$$

Chen et al. (2008), Goh and Sim (2010) show that $\boldsymbol{y}_{DLDR}(\tilde{\boldsymbol{z}})$ is a feasible solution of Problem (2.16). Moreover,

$$
\sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{d}'\boldsymbol{y}_{DLDR}(\tilde{\boldsymbol{z}})\right) \leq \bar{\beta}_{DLDR}(\boldsymbol{x}) \leq \bar{\beta}_{([I_1],\emptyset,\emptyset)}(\boldsymbol{x}).
$$

Our next result shows that the generalized linear decision rule can potentially improve the bound provided by the deflected linear decision rule.

Proposition 5.

$$
\bar{\beta}_{([I_1],[I_2],\emptyset)}(\boldsymbol{x}) \leq \bar{\beta}_{DLDR}(\boldsymbol{x}).
$$

Proof. From Proposition 2, we have the equivalent form of $\bar{\beta}_{DLDR}(\boldsymbol{x})$ as follows:

$$
\begin{aligned}
\bar{\beta}_{DLDR}(\boldsymbol{x}) = \quad \min \quad & r_0 + \boldsymbol{s}_0'\boldsymbol{\mu} + \boldsymbol{t}_0'\boldsymbol{\sigma} + \sum_{i\in\mathcal{M}} f_i^*(r_i + \boldsymbol{s}_i'\boldsymbol{\mu} + \boldsymbol{t}_i'\boldsymbol{\sigma}) \\
\text{s.t.} \quad & r_0 + \boldsymbol{s}_0'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_0'\boldsymbol{u} \ge \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z}) && \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\
& r_i + \boldsymbol{s}_i'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i'\boldsymbol{u} \ge -q_i(\boldsymbol{z}) && \forall i \in \mathcal{M}, \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\
& r_i + \boldsymbol{s}_i'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i'\boldsymbol{u} \ge 0 && \forall i \in \mathcal{M}, \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\
& \boldsymbol{t}_i \ge \boldsymbol{0} && \forall i \in \{0\} \cup \mathcal{M} \\
& \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{b}(\boldsymbol{z}) + \boldsymbol{q}(\boldsymbol{z}) && \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\
& q_i(\boldsymbol{z}) \ge 0 && \forall i \in \bar{\mathcal{M}} \\
& r_i \in \Re, \boldsymbol{s}_i \in \Re^{L_1}, \boldsymbol{t}_i \in \Re^{L_2} && \forall i \in \{0\} \cup \mathcal{M} \\
& \boldsymbol{y} \in \mathcal{L}^{N_2}([I_1], \emptyset, \emptyset) \\
& \boldsymbol{q} \in \mathcal{L}^{M}([I_1], \emptyset, \emptyset).
\end{aligned}
\tag{2.27}
$$

Similarly, we have the equivalent form of $\bar{\beta}_{([I_1],[I_2],\emptyset)}(\boldsymbol{x})$ as follows:

$$
\begin{aligned}
\bar{\beta}_{([I_1],[I_2],\emptyset)}(\boldsymbol{x}) = \quad \min \quad & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.} \quad & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \ge \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) && \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\
& \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \ge \boldsymbol{b}(\boldsymbol{z}) && \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\
& \boldsymbol{t} \ge \boldsymbol{0} \\
& r \in \Re, \boldsymbol{s} \in \Re^{L_1}, \boldsymbol{t} \in \Re^{L_2} \\
& \boldsymbol{y} \in \mathcal{L}^{N_2}([I_1], [I_2], \emptyset).
\end{aligned}
\tag{2.28}
$$

Let $\boldsymbol{y}^\dagger$, $\boldsymbol{q}^\dagger$, $r_i^\dagger, \boldsymbol{s}_i^\dagger, \boldsymbol{t}_i^\dagger$, $i \in \{0\} \cup \mathcal{M}$ be a feasible solution of Problem (2.27). We will show that there exists a corresponding feasible solution for Problem (2.28) with the same objective value. Let

$$\begin{aligned}
r &= r_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}\bar{\boldsymbol{y}}_i r_i^\dagger \\
\boldsymbol{s} &= \boldsymbol{s}_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{s}_i^\dagger \\
\boldsymbol{t} &= \boldsymbol{t}_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{t}_i^\dagger, \\
\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) &= \boldsymbol{y}^\dagger(\boldsymbol{z}) + \sum_{i\in\mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u} \right) \bar{\boldsymbol{y}}_i.
\end{aligned}$$

Observe that the objective value of Problem (2.28) becomes

$$\begin{aligned}
r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} &= r_0^\dagger + \boldsymbol{s}_0^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_0^{\dagger\prime}\boldsymbol{\sigma} + \sum_{i\in\mathcal{M}} (r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{\sigma})\boldsymbol{d}'\bar{\boldsymbol{y}}_i \\
&= r_0^\dagger + \boldsymbol{s}_0^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_0^{\dagger\prime}\boldsymbol{\sigma} + \sum_{i\in\mathcal{M}} f_i^*(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{\sigma}).
\end{aligned}$$

We next check the feasibility of the solution in Problem (2.28). Note that $\boldsymbol{t} \geq \boldsymbol{0}$ and for all $(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}}$,

$$\begin{aligned}
r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} &= r_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i r_i^\dagger + \left(\boldsymbol{s}_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{s}_i^\dagger \right)' (\boldsymbol{G}\boldsymbol{z}) + \left(\boldsymbol{t}_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{t}_i^\dagger \right)' \boldsymbol{u} \\
&= r_0^\dagger + \boldsymbol{s}_0^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_0^{\dagger\prime}\boldsymbol{u} + \sum_{i\in\mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u} \right) \boldsymbol{d}'\bar{\boldsymbol{y}}_i \\
&\geq \boldsymbol{d}'\boldsymbol{y}^\dagger(\boldsymbol{z}) + \sum_{i\in\mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u} \right) \boldsymbol{d}'\bar{\boldsymbol{y}}_i \\
&= \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}),
\end{aligned}$$

where the inequality follows from the first robust counterpart constraint in Problem (2.27). We now show the feasibility of second robust robust counterpart constraint in Problem (2.28). Observe that for all $(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}}$,

$$\begin{aligned}
\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) &= \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{\dagger}(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} \left(r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \right) \boldsymbol{B}\bar{\boldsymbol{y}}_i \\
&= \boldsymbol{b}(\boldsymbol{z}) + \boldsymbol{q}^{\dagger}(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} \left(r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \right) \bar{\boldsymbol{q}}_i \\
&= \boldsymbol{b}(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} q_i^{\dagger}(\boldsymbol{z})\boldsymbol{e}_i + \sum_{j \in \bar{\mathcal{M}}} q_j^{\dagger}(\boldsymbol{z})\boldsymbol{e}_j + \sum_{i \in \mathcal{M}} \left(r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \right) \bar{\boldsymbol{q}}_i \\
&\geq \boldsymbol{b}(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} q_i^{\dagger}(\boldsymbol{z})\boldsymbol{e}_i + \sum_{j \in \bar{\mathcal{M}}} q_j^{\dagger}(\boldsymbol{z})\boldsymbol{e}_j + \sum_{i \in \mathcal{M}} \left(r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \right) \boldsymbol{e}_i \\
&= \boldsymbol{b}(\boldsymbol{z}) + \sum_{j \in \bar{\mathcal{M}}} q_j^{\dagger}(\boldsymbol{z})\boldsymbol{e}_j + \sum_{i \in \mathcal{M}} \left(q_i^{\dagger}(\boldsymbol{z}) + r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \right) \boldsymbol{e}_i \\
&\geq \boldsymbol{b}(\boldsymbol{z}).
\end{aligned}$$

The first inequality holds because $\bar{\boldsymbol{q}}_i \geq \boldsymbol{e}_i$ and $r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \geq 0$ for all $i \in \mathcal{M}$, $(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}}$. The second inequality is due to $r_i^{\dagger} + {\boldsymbol{s}_i^{\dagger}}'(\boldsymbol{G}\boldsymbol{z}) + {\boldsymbol{t}_i^{\dagger}}'\boldsymbol{u} \geq -q_i^{\dagger}(\boldsymbol{z})$ for all $i \in \mathcal{M}$, $(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}}$ and $q_i^{\dagger}(\boldsymbol{z}) \geq 0$ for all $i \in \bar{\mathcal{M}}$, $(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}}$. This concludes our proof. □

On the usage of linear decision rules

We introduce linear decision rules with the goal to obtain tractable formulations, so that the optimal *here-and-now* decision $\boldsymbol{x} \in X_1$ can be determined and implemented. For a given $\boldsymbol{x} \in X_1$, let $\boldsymbol{y}^*$ be the optimal function of Problem (2.16), and $\boldsymbol{y}^*_{GLDR}$ be the optimal generalized linear decision rule of Problem (2.19). For a given function, $\boldsymbol{\nu} \in \mathcal{R}^{I_1, I_3}$ satisfying $(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{\nu}(\boldsymbol{z})) \in \hat{\mathcal{W}}$ for all $\boldsymbol{z} \in \mathcal{W}$, the function $\hat{\boldsymbol{y}}_{GLDR} \in \mathcal{R}^{I_1, N_2}$,

$$\hat{\boldsymbol{y}}_{GLDR}(\boldsymbol{z}) = \boldsymbol{y}^*_{GLDR}\left(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{\nu}(\boldsymbol{z})\right)$$

is a feasible solution to Problem (2.16). Moreover, the objective satisfies

$$\begin{aligned}\sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\hat{\boldsymbol{y}}_{GLDR}(\tilde{\boldsymbol{z}})) &= \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{d}'\boldsymbol{y}^*_{GLDR}\left(\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}}), \boldsymbol{\nu}(\tilde{\boldsymbol{z}})\right)\right) \\ &\leq \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}\left(\boldsymbol{d}'\boldsymbol{y}^*_{GLDR}\left(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}\right)\right) \\ &= \bar{\beta}_{(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)}(\boldsymbol{x}),\end{aligned}$$

where the inequality is due to Proposition (1). Suppose

$$\beta(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}})) = \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\hat{\boldsymbol{y}}_{GLDR}(\tilde{\boldsymbol{z}})) = \bar{\beta}_{(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)}(\boldsymbol{x}),$$

which is the case for complete recourse problems and $N_2 = 1$, there is a tendency to infer the optimality of $\hat{\boldsymbol{y}}_{GLDR}(\boldsymbol{z})$, such that

$$\boldsymbol{d}'\hat{\boldsymbol{y}}_{GLDR}(\boldsymbol{z}) = \boldsymbol{d}'\boldsymbol{y}^*(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W}.$$

However, this is not the case and we will demonstrate this fallacy in the following simple example.

Example 2.2.2. Consider the following complete recourse problem,

$$\begin{aligned}\beta = \quad \min \quad & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(y(\tilde{z})) \\ \text{s.t.} \quad & y(z) \geq z \qquad \forall z \in \Re \\ & y(z) \geq -z \qquad \forall z \in \Re \\ & y \in \mathcal{R}^{1,1},\end{aligned} \tag{2.29}$$

where

$$\mathbb{F} = \left\{ \ \mathbb{P} \in \mathcal{P}_0(\Re) : \mathbb{E}_{\mathbb{P}}(\tilde{z}) = 0, \mathbb{E}_{\mathbb{P}}(\tilde{z}^2) \leq 1 \ \right\}.$$

Clearly, $y^*(z) = |z|$ is the optimal solution and it is also the optimal objective value for all $z \in \Re$. However, under the generalized linear decision rule, we obtain $\hat{y}_{GLDR}(z) = \frac{1+z^2}{2}$, which is almost always greater than $y^*(z)$ except at $z = 1$ and $z = -1$. Incidentally, the worst case distribution $\mathbb{P} \in \mathbb{F}$ corresponds to the two point distributions with $\mathbb{P}(\tilde{z} = 1) = \mathbb{P}(\tilde{z} = 1) = 1/2$. Hence, this explains why the worst case expectations are the same.

Hence, from the above example, even if a generalized linear decision rule were to provide a close approximation to $\beta(\boldsymbol{x})$, $\boldsymbol{x} \in X_1$, the solution generated by the decision rule could be a far cry from the optimal function, $\boldsymbol{y}^*$. Therefore, we advise against using the generalized decision rule as a policy guide for future actions when uncertainty is realized. Instead, the second stage decision should be determined by solving a linear optimization problem after the uncertainty is resolved.

Another important feature of linear decisions rule is the ability to easily enforce non-anticipative conditions, which are necessary to capture the nature of multistage decisions where information is revealed in stages. For given subsets $\mathcal{S}_1^i \subseteq [I_1]$, that reflects information dependency of recourse decisions, y_i, $i \in [N_2]$, we can consider the generalization of Problem (2.16) as follows:

$$\begin{aligned} \gamma^*(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})) \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\ & y_i \in \mathcal{R}^{I_1,1}(\mathcal{S}_1^i) \qquad\qquad\qquad \forall i \in [N_2], \end{aligned} \tag{2.30}$$

where we define the space of restricted measurable functions as

$$\mathcal{R}^{I,N}(\mathcal{S}) = \left\{ \boldsymbol{y} \in \mathcal{R}^{I,N} \ : \ \boldsymbol{y}(\boldsymbol{v}) = \boldsymbol{y}(\boldsymbol{w}) \ \forall \boldsymbol{v}, \boldsymbol{w} \in \Re^I : v_j = w_j, j \in \mathcal{S} \right\}.$$

Problem (2.30) solves for the optimum measurable function $\boldsymbol{y} \in \mathcal{R}^{I_1,N_2}$ that minimizes the worst case expected objective taking into account of the information dependency requirement. Clearly, this problem would be much harder to solve and we are not aware of a viable approach to compute the exact solution. Yet, despite the difficulty, it is relatively simple to use generalized linear decision rules to obtain an upper bound as follows:

$$\begin{aligned} \bar{\gamma}(\boldsymbol{x}) = \quad \min \quad & \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})) \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v})) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall(\boldsymbol{z},\boldsymbol{u},\boldsymbol{v}) \in \hat{\mathcal{W}} \\ & y_i \in \mathcal{L}^1(\mathcal{S}_1^i, \mathcal{S}_2^i, \mathcal{S}_3^i) \qquad \forall i \in [N_2], \end{aligned} \tag{2.31}$$

where the subsets $\mathcal{S}_2^i \subseteq [I_2]$, $\mathcal{S}_3^i \subseteq [I_3]$, are appropriately selected to abide by the information restriction imposed by $\mathcal{S}_1^i \subseteq [I_1]$, $i \in [N_2]$. Again, we use the generalized linear decision rules to enable us to obtain a reasonably good *here-and-now* decision, $\boldsymbol{x} \in X_1$ that accounts for how decisions might be adjusted as uncertainty unfolds over the stages. Similar to the standard adjustable robust optimization technique, we propose the rolling or folding horizon implementation where we solve for the new *here-and-now* decision using the latest available information as we proceed to the next stage.

In the next section, we will briefly describe a new algebraic modeling package named ROC and show how it could be used to facilitate modeling of distributionally robust linear optimization problems.

2.3 ROC: Robust Optimization C++ package

We developed ROC as a proof of concept to provide an intuitive environment for modeling and solving distributionally robust linear optimization problems that will free the user from dealing directly with the laborious and error-prone reformulations. ROC is developed in the C++ programming language, which is fast, highly portable and well suited for deployment of robust optimization technologies in decision support system. We will briefly discuss the key aspects of ROC and provide simple examples to illustrate the algebraic modeling package. Most algebraic modeling packages for optimization are geared towards modeling deterministic optimization problems. While a robust optimization problem may be formulated as a deterministic optimization problem, it would be rather difficult for the modeler to explicitly code, say Problem (2.31) using these algebraic modeling packages.

A typical algebraic modeling package provides the standardized format for declaration of decision variables, transcription of constraints and objective functions, and interface with external solvers. ROC has additional features including declaration of uncertain parameters and linear decision rules, transcriptions of ambiguity sets and automatic reformulation of standard and distributionally robust counterparts using the techniques described in this paper. The current version of ROC solver is integrated with CPLEX and will be expanded to include other solvers. We refer readers to `http://www.meilinzhang.com/software` for more information on ROC.

Declaration of decisions, uncertain parameters and expressions

Code Segment 2.1 provides an example on how we define decision variables, uncertain parameters and linear decision rules in ROC. The code illustrates how the following deterministic decision variables are declared

$$x_1 \in \Re, x_2 \in [5, \infty), x_3 \in \{0, \dots, 100\}, x_4 \in \{0, 1\}, \boldsymbol{s} \in \Re^6, \boldsymbol{t} \in \Re^{5\times 8}.$$

By C++ convention, an array of sized N is defined on indices $0, \dots, N-1$. The variable x_2 is also associated with the name "`X2`", which would be useful in output display of the model. Note that $\tilde{z}_1, \tilde{z}_2, \tilde{z}_3$ are uncertain parameters in $\Re$ and $\tilde{\boldsymbol{u}}$ is a an array of uncertain parameters in $\Re^6$. The linear decision rules y_1, y_2, y_3 are declared. The user can selectively include the linear dependency using the `addDR` function. In this case, y_1 is affinely dependent on $\tilde{z}_1$, y_2 is affinely dependent on $(\tilde{z}_2 + \tilde{z}_3)$, and y_3 has the same dependency as y_1, i.e.,

$$
\begin{aligned}
y_1(\tilde{z}_1) &= y_1^0 + y_1^1 \tilde{z}_1 \\
y_2(\tilde{z}_2, \tilde{z}_3) &= y_2^0 + y_2^1(\tilde{z}_2 + \tilde{z}_3) \\
y_3(\tilde{z}_1) &= y_3^0 + y_1^1 \tilde{z}_3.
\end{aligned}
$$

where y_1^0,y_1^1,y_2^0, y_2^1, y_3^0 and y_3^1 are embedded decision variables that are declared in association with the linear decision rules.

```
// Declaration of decisions and uncertain parameters
ROVar x1, x2(5, ROInfinity, "X2");  // x1, x2 continuous decision variables
ROIntVar x3(0,10);  // x3 Integer variable
ROBinVar x4;        // x4 binary variables
ROVarArray s(6);    // an array of 6 decision variables
ROUn z1, z2, z3;    // three uncertain parameters
ROUnArray u(6);     // an array of 6 uncertain parameters

// Define a 2D array of 5 by 8 decision variables
ROVar2DArray t(5);
for(int i = 0; i<5 ;i++)
      t[i] = ROVarArray(8);

ROVarDR y1, y2,y3;        // two linear decision rules
// add dependency on uncertain parameters  to linear  decision rule
y1.addDR(z1);
// add dependency on uncertain parameters to linear decision rule
y2.addDR(z2+z3);
// clone dependency from y2
y3.clone(y1);
```

Code Segment 2.1: Declaration of decisions and uncertain parameters ROC.

We can also declare an expression, which is an object to contain either a quadratic function of decision variable or a biaffine function of the decision variables and uncertain parameters. An expression permits linear operations on constants, decision variables, uncertain parameters, linear decision rules and other expressions and it is useful as temporary storage. Code Segment 2.2 show some examples of expressions. Here, we have

$$
\begin{aligned}
\texttt{expr1}: &\quad x_1^2 + x_2 \\
\texttt{expr2}: &\quad s_1 + 2s_4 - x_1\tilde{z}_1 \\
\texttt{expr3}: &\quad s_1 + 2s_4 - x_1\tilde{z}_1 + y_2^0 + y_2^1(\tilde{z}_2 + \tilde{z}_3) \\
\texttt{expr4}: &\quad s_1\tilde{u}_1 + 2s_2\tilde{u}_2 + 3s_3\tilde{u}_3 + 4s_4\tilde{u}_4 + 5s_5\tilde{u}_5 + 6s_6\tilde{u}_6.
\end{aligned}
$$

```
// Simple expressions
ROExpr expr1, expr2, expr3, expr4;
expr1 = x1*x1 + x2;
expr2 = s[0] + 2 *s[3] - x1 * z1;
expr3 = expr2 + y2;

for(int i = 0; i < 6; i++)
  expr4 += (i+1)*s[i]*u[i];
```

Code Segment 2.2: The use of expressions in ROC.

Modeling Ambiguity sets

The ability to comprehensively model distribuitionally ambiguity sets ROC apart from other algebraic modeling packages. The `ambiguitySet` defined in Code Segment 2.3 describes the following ambiguity set

$$
\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\Re^3 \times \Re^6\right) : \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \Re^3 \times \Re^6 \\ \mathbb{E}_{\mathbb{Q}}(\tilde{u}_1) = 1 \\ \mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \hat{\mathcal{W}}\right) = 1 \end{array} \right\},
$$

where

$$\hat{\mathbb{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \Re^3 \times \Re^6 : \begin{array}{l} z_1 \geq z_2 \\ z_1^2 \leq z_3 \\ \|\boldsymbol{u}\|_\infty \leq 5 \\ \|\boldsymbol{u}\|_1 \leq 4 \\ \|\boldsymbol{u}\|_2 \leq 3 \end{array} \right\}.$$

```
ROConstraintSet ambiguitySet;
ambiguitySet.add(ROExpect(u[0])==1);
ambiguitySet.add(z1>=z2);
ambiguitySet.add(ROSq(z1)<= z3);
ambiguitySet.add(RONormInf(u)<=5);
ambiguitySet.add(RONorm1(u)<=4)
ambiguitySet.add(RONorm2(u)<=3);
```

Code Segment 2.3: Definition of an ambiguity set in ROC.

Note that the statement `ROSq(z1) <= z3` calls upon the function `ROSq`, which returns a newly declared uncertain parameter, say $\tilde{v}$ so that $\tilde{v} \leq \tilde{z}_3$. Internally within the function, the epigraph of $\tilde{z}_1^2 \leq \tilde{v}$ is automatically converted to a second order cone constraint, $\sqrt{(\frac{\tilde{v}-1}{2})^2 + \tilde{z}_1^2} \leq \frac{\tilde{v}+1}{2}$. Hence, the user should be disciplined in convex representation of constraints and avoid statements such as `ROSq(z1)>=7`. Likewise, the functions `RONorm1`, `RONorm2` and `RONormInf` are provided within ROC for modeling convenience. These functions return newly declared uncertain parameters and internally represent the epigraphs of these functions using linear and second order conic constraints. The functions (such as `RONorm1`) may declare other uncertain parameters that are hidden from the user. Using this approach, we can also declare other common conic quadratic representable functions within ROC including higher powers and approximations of exponential functions, among others. We have also provided functions that linearly approximates second order cones as proposed in Ben-Tal and Nemirovski (2001b), which may be useful if linearity of the model is desired. Note that decision variables are not permitted in the description of ambiguity set and that the user has the freedom to define multiple ambiguity sets.

Declaration of a model, adding constraints and the objective function

A ROC model consists of objects that represents a problem including constraints and the objective function. Deterministic constraints can be added in the model as shown in Code Segment 2.4, which models the following set of constraints

$$
\begin{aligned}
& x_1^2 + x_2 \leq t_{2,5} \\
& x_1^2 - 2x_1x_2 + x_2^2 \leq 7 \\
& |x_1 - x_3| \leq 7 \\
& (x_2 - x_3)^+ \leq x_1 \\
& \|\boldsymbol{s}\|_1 \leq 4t_{3,3} \\
& \|\boldsymbol{s}\|_2 \leq 6(x_1 + 2x_2) \\
& \|\boldsymbol{s}\|_\infty \leq -x_2^2 \\
& \boldsymbol{1}'\boldsymbol{s} \leq 10.
\end{aligned}
$$

Similar to the descriptions of ambiguity sets, the functions return newly declared decision variables and internally represent the epigraphs of these functions using linear and second order conic constraints.

```
ROModel model;              // define a robust optimization model
model.add(expr1<= t[1][4]);
model.add(x1*x1-2*x1*x2+ x2*x2 <= 7);
model.add(ROAbs(x1-x3)<= 7 );
model.add(ROPos(x2-x3)<= x1);
model.add(RONorm1(s)<= 4*t[2][2]);
model.add(RONorm2(s)<=6*(x1+2*x2) );
model.add(RONormInf(s)<= -x2*x2);
model.add(ROSum(s)<= 10);
```

Code Segment 2.4: Model declaration with deterministic constraints in ROC.

More interestingly, ROC is able to model robust counterpart constraint such as,

```
model.add(ROConstraint(expr4 <= x1, ambiguitySet));
```

which automatically reformulates the following robust counterpart,

$$s_1u_1 + 2s_2u_2 + 3s_3u_3 + 4s_4u_4 + 5s_5u_5 + 6s_6u_6 \le x_1 \quad \forall(\boldsymbol{z}, \boldsymbol{u}) \in \hat{\mathbb{W}},$$

into a set of deterministic constraints. In the process, new decision variables may be declared that are hidden away from the user. Note the ambiguity set must be specified in the robust counterpart constraint, so that ROC can extract the underlying uncertainty set $\hat{\mathbb{W}}$. Hence, different ambiguity sets can be defined for use in different robust counterpart constraints. More interestingly, a distributionally robust counterpart over the worst case expectation such as,

```
model.add(ROConstraint( (ROExpect(expr2) >= x3 + x1),
    ambiguitySet));
```

which corresponds to

$$\mathbb{E}_{\mathbb{Q}}(s_1 + 2s_4 - x_1\tilde{z}_1 + y_2^0 + y_2^1(\tilde{z}_2 + \tilde{z}_3)) \ge x_3 + x_2 \quad \forall \mathbb{Q} \in \mathbb{G},$$

or equivalently as

$$\sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(-(s_1 + 2s_4 - x_1\tilde{z}_1 + y_2^0 + y_2^1(\tilde{z}_2 + \tilde{z}_3))) \le -x_3 - x_2,$$

will be transformed to a set of deterministic constraints using Proposition 2.

The model should finally include an objective, which reflects either a minimization or maximization problem. If the objective expression contains uncertain parameters, then it must also incorporate the corresponding ambiguity set so the worst case objective can be evaluated. The following code segment illustrates an objective function that minimizes the worst case expectation of `expr2` over the ambiguity set, $\mathbb{G}$.

```
    model.add( ROMinimize(ROExpect(expr2), ambiguitySet));
```

2.4 Computation Experiment

In our experiment, we consider a multiproduct newsvendor problem with N different types of products, indexed by i. For product $i, i \in [N]$, its selling price and order cost are denoted by p_i and c_i respectively. Manager needs to decide each product's order quantity $x_i, i \in [N]$ before the demand $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_N)$ is observed. Meanwhile, the total budget for purchasing all products is Γ. After the demand becomes known, the selling quantity is decided as $\min\{x_i, z_i\}, i \in [N]$. In order to maximize the expected operating revenue, the problem could be formulated as

$$\begin{aligned} \Pi^* = \quad & \max \quad \inf_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{i \in [N]} p_i \min\{x_i, \tilde{z}_i\} \right) \\ & \text{s.t.} \quad \boldsymbol{c}'\boldsymbol{x} \leq \Gamma \\ & \qquad \boldsymbol{x} \geq \boldsymbol{0} \\ & \qquad \boldsymbol{x} \in \Re^N. \end{aligned}$$

To be consistent with the earlier framework, we formulate this as the following minimization problem

$$\begin{aligned} Z^* = -\Pi^* = \quad & \min \quad -\boldsymbol{p}'\boldsymbol{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{i \in [n]} p_i ((x_i - \tilde{z}_i)^+) \right) \\ & \text{s.t.} \quad \boldsymbol{c}'\boldsymbol{x} \leq \Gamma \\ & \qquad \boldsymbol{x} \geq \boldsymbol{0} \\ & \qquad \boldsymbol{x} \in \Re^N. \end{aligned} \tag{2.32}$$

To demonstrate the modeling power of the standardized framework for characterizing distributional ambiguity, we present the following unusual but interesting ambiguity set that is inspired by the structure of the optimization problem.

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Re^N) : \begin{array}{ll} \mathbb{E}_\mathbb{P}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_\mathbb{P}(\tilde{z}_i^2) \leq {\mu_i}^2 + \sigma_i^2 & \forall i \in [N] \\ \mathbb{E}_\mathbb{P}\left(\sum_{i \in [N]} p_i(\mu_i - \tilde{z}_i)^+ \right) \leq \psi & \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1 & \end{array} \right\}, \tag{2.33}$$

where

$$\mathcal{W} = \{ \boldsymbol{z} \in \Re^n : \boldsymbol{0} \leq \tilde{\boldsymbol{z}} \leq \bar{\boldsymbol{z}} \}.$$

Correspondingly, the extended ambiguity set of $\mathbb{F}$ is given by

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\Re^N \times \Re^{N+1} \times \Re^N) : \begin{array}{ll} \mathbb{E}_\mathbb{Q}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_\mathbb{Q}(\tilde{u}_i) \leq \mu_i^2 + \sigma_i^2 & \forall i \in [N] \\ \mathbb{E}_\mathbb{Q}(\tilde{u}_{N+1}) \leq \psi & \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \hat{\mathcal{W}}) = 1 & \end{array} \right\}, \tag{2.34}$$

where

$$\hat{\mathcal{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \Re^N \times \Re^{N+1} \times \Re^N : \begin{array}{ll} \boldsymbol{0} \leq \tilde{\boldsymbol{z}} \leq \bar{\boldsymbol{z}} & \\ {z_i}^2 \leq u_i & \forall i \in [N] \\ u_{N+1} \geq \boldsymbol{p}'\boldsymbol{v} & \\ \boldsymbol{v} \geq \boldsymbol{\mu} - \boldsymbol{z} & \\ \boldsymbol{v} \geq \boldsymbol{0} & \end{array} \right\}.$$

Using the generalized linear decision rule, we solve the following two-stage distributionally robust optimization problem,

$$
\begin{aligned}
\bar{Z}^*(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3) = \quad \min \quad & -\boldsymbol{p}'\boldsymbol{x} + \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}} \left(\boldsymbol{p}'\boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})\right) && \\
\text{s.t.} \quad & \boldsymbol{c}'\boldsymbol{x} \leq \Gamma && \\
& \boldsymbol{x} \geq \boldsymbol{0} && \\
& \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \geq \boldsymbol{0} && \forall (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\
& \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \geq \boldsymbol{x} - \boldsymbol{z} && \forall (\boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \hat{\mathcal{W}} \\
& \boldsymbol{x} \in \Re^N && \\
& \boldsymbol{y} \in \mathcal{L}^N(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3). &&
\end{aligned}
\tag{2.35}
$$

Formulating in ROC

Instead of deriving the explicit mathematical model of Problem (2.35), we present the formulation in ROC, which will automatically transform the problem and call upon a standard solver package such as CPLEX to obtain the solution. We first define the decision variables, $\boldsymbol{x} \in \Re^N$, uncertain parameters, $\tilde{\boldsymbol{z}} \in \Re^N$, $\tilde{\boldsymbol{u}} \in \Re^{N+1}$, $\tilde{\boldsymbol{v}} \in \Re^N$ and the linear decision rule $\boldsymbol{y} \in \mathcal{R}^{\cdot,N}$ as shown in Code Segment 2.5.

```
// Define Decision variables, decision rules and uncertain
    parameters
ROVarArray x(N, 0, ROInfinity, "X");
ROVarDRArray y(N);
ROUnArray z(N), u(N+1), v(N);
```

Code Segment 2.5: Defining decision variables, uncertain parameters and linear decision rule.

We next show how to characterize the dependency of the decision rule $\boldsymbol{y}$. Code Segment 2.6 presents an example where the decision rule $\boldsymbol{y}$ is defined in $\mathcal{L}^N([N], [N+1], [N])$, and hence it is fully dependent on all the uncertainty parameters including the auxiliary ones.

```
// Adding dependency to linear decision rules
for(int i = 0; i < N; i++)
{
```

```
    for(int j = 0; j < N; j++)
    {
      y[i].addDR(z[j]);
      y[i].addDR(u[j]);
      y[i].addDR(v[j]);
    }
    y[i].addDR(u[N]);
  }
```

Code Segment 2.6: Defining generalized linear decision rule in $\mathcal{L}^N([N]), [N+1]), [N])$

Next, we specify the ambiguity set $\mathbb{G}$ as shown in Code Segment 2.7.

```
// Construct the Ambiguity Set
ROConstraintSet ambiguitySet;
ROExpr unExpr;
for(int i = 0; i < N; i++)
{
  ambiguitySet.add(ROExpect(z[i]) == mu[i]);
  ambiguitySet.add(ROExpect(u[i]) <= mu[i]*mu[i] + sigma[i]*sigma
      [i]);
  ambiguitySet.add(z[i] >= 0);
  ambiguitySet.add(z[i] <= barZ[i]);
  ambiguitySet.add(ROSq(z[i]) <= u[i]);
  ambiguitySet.add(v[i] >= 0);
  ambiguitySet.add(v[i] >= mu[i] - z[i]);
  unExpr += price[i] * v[i];
}
ambiguitySet.add(ROExpect(u[N]) <= psi);
ambiguitySet.add(u[N] >= unExpr);
```

Code Segment 2.7: Constructing the ambiguity set $\mathbb{G}$.

Finally, Code Segment 2.8 show how the we model Problem (2.35) in ROC.

```
ROModel model;          // define a robust optimization model
    engine
```

```

// Adding constraints to Model
ROExpr expr1;
for(int i = 0; i < N; i++)
{
  expr1 += cost[i] * x[i];
  model.add(ROConstraint(y[i] >= 0, ambiguitySet));
  model.add(ROConstraint(y[i] >= x[i] - z[i], ambiguitySet));
}
model.add(expr1 <= budget);

// Adding objective expression
ROExpr objExpr1, objExpr2;
for(int i = 0; i < N; i++)
{
  objExpr1 -= price[i] * x[i];
  objExpr2 += price[i] * y[i];
}
model.add( ROMinimize(objExpr1 + ROExpect(objExpr2), ambiguitySet
    ) );
model.solve();
```

Code Segment 2.8: Create the robust pptimization model.

Performance of the decision rules

For the purpose of comparison, we next formulate the model to evaluate Problem (2.32) exactly. By observing that $\sum_{i\in[N]} (a_i)^+ = \max_{\mathcal{S}:\mathcal{S}\subseteq[N]} \left(\sum_{i\in\mathcal{S}} a_i\right)$, we can transform

Problem (2.32) to the following problem

$$\begin{aligned} Z^* = \quad \min \quad & -\boldsymbol{p}'\boldsymbol{x} + \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\max_{\mathcal{S}:\mathcal{S}\subseteq[N]} \left(\sum_{i\in\mathcal{S}} p_i(x_i - \tilde{z}_i) \right) \right) \\ \text{s.t.} \quad & \boldsymbol{c}'\boldsymbol{x} \le \Gamma \\ & \boldsymbol{x} \ge \boldsymbol{0} \\ & \boldsymbol{x} \in \Re^N \end{aligned} \tag{2.36}$$

Noting that the number of subsets of $[N]$ equals to 2^N, we will study a small problem so that it would be computationally variable to compare the quality of solutions obtained by linear decision rules. Hence, we restrict to $N = 10$. We solve for a particular instance with $\psi = 100$, $\Gamma = 500$ and the parameters associated with the products are shown in Table 2.1.

Product ID	`price[i]` p_i	`cost[i]` c_i	`mu[i]` μ_i	`sigma[i]` σ_i	$\bar{z}_i$
1	10.00	2.00	30.00	30.00	100
2	11.00	2.71	35.00	28.50	100
3	11.41	3.00	40.00	27.00	100
4	11.73	3.23	45.00	25.50	100
5	12.00	3.41	50.00	24.00	100
6	12.24	3.58	55.00	22.50	100
7	12.45	3.73	60.00	21.00	100
8	12.65	3.87	65.00	19.50	100
9	12.83	4.00	70.00	18.00	100
10	13.00	4.12	75.00	16.50	100

Table 2.1: Input parameters of multiproduct newsvendor problem

Table 2.2 shows the objective values of

$$\begin{aligned} \underline{\Pi}_1^* &= -\bar{Z}^*(\emptyset, \emptyset, \emptyset) \\ \underline{\Pi}_2^* &= -\bar{Z}^*([N], \emptyset, \emptyset) \\ \underline{\Pi}_3^* &= -\bar{Z}^*([N], [N+1], \emptyset) \\ \underline{\Pi}_4^* &= -\bar{Z}^*([N], [N+1], [N]) \\ \Pi^* &= -Z^* \end{aligned}$$

and also presents the corresponding optimal solutions. We observe that the

Problem	Objective	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\underline{\Pi}_1^*$	0	20.35	63.59	10.02	9.66	9.92	9.77	9.50	9.21	8.93	8.68
$\underline{\Pi}_2^*$	1172.26	0	0	0	0	0	0	0	0	21.97	100
$\underline{\Pi}_3^*$	1523.35	30	35	0	0	0	0	0	6.04	38.37	40.86
$\underline{\Pi}_4^*$	1851.17	30	35	40	45	23.47	0	0	0	0	0
Π^*	1851.17	30	35	40	45	23.47	0	0	0	0	0

Table 2.2: Computational results for multiproduct newsvendor problem

improvement in objective values as the decision rule has dependent on greater subsets of uncertain parameters. In particular, for the case of full dependency, we have $\underline{\Pi}_4^*$ achieving the optimal objective value Π^*, underscoring the potential and benefits of the generalized linear decision rule in addressing distributionlly robust linear optimization problems.

Endnotes

1. YALMIP homepage: `http://users.isy.liu.se/johanl/yalmip/`. See also Löberg (2012).
2. AIMMS homepage: `http://www.aimms.com/`.
3. ROME homepage: `http://robustopt.com`. See also Goh and Sim (2009).

3

A Robust Optimization Model for Managing Elective Admission in Hospital

Beds are a critical resource in hospital operations. Overcrowding of Accident and Emergency (A & E) is often due to availability (or rather the shortage) of hospital beds (Wardrope and Driscoll (2003)); so is cancellation of elective surgeries (Robb et al. (2004)). However, bed resources are expensive as the hospitals need highly trained personnel to manage these beds. Work has been done in the area of the acquisition and utilization of bed resources (e.g., Cochran and Roche (2008), Harper and Shahani (2002), Kao and Tung (1981), Teow and Tan (2008)). Harper and Shahani (2002) acknowledged the complexity of the internal dynamics of a hospital (especially bed management), and used a simulation model for patient flows and bed matching over time.

Typically, Day-of-Week (DoW) patterns of a hospital exhibit a wide range of variations. Emergency admissions are beyond the control of the hospital, while elective admissions are scheduled by the hospital. Nevertheless, often the relative variation is largest in elective admissions, and larger in discharges than emergency admissions (Proudlove et al. (2007)). On days with high bed occupancy, long wait time is encountered. On days with low bed occupancy, beds are under-utilized. We have tightness of usage on one hand, and looseness on the other. It is not the desired state.

Elective surgeries account for the majority of elective admissions, though medical electives (non-surgical cases) do make up for some of these admissions. Elective surgeries are procedures planned in advance and can be divided into day surgery (DS), same day surgery admission (SDA) and inpatient admission (IP). DS cases do not "consume" beds, while SDA cases require beds to accommodate patient day after surgery. IP cases require beds one day before the surgery.

In general, hospitals will admit all emergency cases. As such, in a tight bed situation, the tradeoff is to reduce the number of beds designated for elective admissions. But a more prudent and sensible approach would be to make adjustments on a dynamic basis. What this entails is that when emergency cases are fewer, then more beds could be assigned to elective cases, and vice versa. This leads to an optimal control policy, which is to maximize bed utilization on a daily basis by controlling the number of elective admissions. This requires a more prudent scheduling of operating theatre sessions. However, a higher level of complexity in planning ensues because of the high degree of uncertainty involved in bed availability and its effect on admission rates.

Various models for managing patient admissions have been proposed in the literature. Esogbue and Singh (1976) developed a method for determining optimal distribution of beds in a ward using cut-off level via shortage and holding costs. They assumed Poisson patient arrival distribution and negative exponential distribution for length of stay. Kao and Tung (1981) proposed an approach for periodically reallocating beds to services to minimize the expected overflows, using queueing models to approximate the population dynamics. In fact, queueing theory and stochastic simulation are the main methodological tools in studies of bed allocation and bed capacity (Cochran and Roche (2008), Gorunescui et al.

(2004), Lamiri et al. (2008), Vassilacopoulos (1985)). The underlying rationale for researchers relying on these methodological tools is the uncertain nature of the hospital unit vis-à-vis the number of patients as a result of random arrivals and random lengths of stay. A thorough review on OR applications in healthcare services in the United Kingdom can be found in Proudlove et al. (2007).

The admission of emergency inpatients is unscheduled and they are usually warded within hours. In contrast, admission of elective patients is less pressing and they can be warded on the day of admission or even several weeks later. The flexibility vis-à-vis elective patients allows hospitals to manage the flow of elective patients in a way as to "smooth out" the daily bed occupancy, a modus operandus known as "elective smoothing". This will ensure that on days with spikes in emergency cases, the admission rate for elective patients can be reduced. The converse applies. Some hospitals in Singapore have already incorporated this mechanism into their decision support systems and it has led to improvements when elective patient flow is high (Teow et al. (2007)). In these hospitals, the admission quotas for elective patients are obtained by solving a deterministic linear optimization problem without taking into account the variability of patient arrivals and stay durations. While this achieves smoothing in expectation, it is conceivable that efficacy would diminish when variability is high.

Due to the difficulties of obtaining true probability distributions and solving stochastic optimization problems, it is common in real world deployment of optimization technology to ignore uncertainty. A fine level of analysis would be required to obtain the distributions of patient arrivals and departure profile as a function of admission quotas, which may not necessarily lead to a computational tractable optimization problem. In recent years, robust optimization offers an attractive alternative for addressing uncertainty in optimization modeling without having to specify exact probability distributions. In many interesting cases, the approach leads to computationally tractable optimization problems; see for instance Ben-Tal and Nemirovski (1998), Bertsimas and Sim (2004), El Ghaoui et al. (1998). In classical robust optimization, uncertainty is represented by an *uncertainty set*, which is usually a simple geometric convex set such as a l-norm ball intersected with the *support set*, the minimal convex set that contains the uncertainty. The modeler requires to articulate her *ambiguity attitude*[4] by specifying the *budget of uncertainty* parameter, which relates to the size of the uncertainty set against what she seeks immunity.

While there are several proposed uncertainty sets and heuristics for specifying budgets of uncertainty, these approaches may not naturally characterize the uncertainty relating to patient movements within the hospital. In this paper, we adopt the distributionally robust optimization approach for managing elective admission in hospital, where uncertainty is characterized by the support set and a restricted ambiguous set of probability distributions (or *ambiguity set* for short); see for instance Chen et al. (2007, 2008), Delage and Ye (2010), Goh and Sim (2009, 2010). Similar to the uncertainty set in classical robust optimization, the proposed ambiguity set is adjustable via a so called *budget of variation* parameter, which is the bound on the coefficient of variation of the uncertainty parameters. The ambiguity set is enlarged by increasing the budget of variation, which leads to greater uncertainty in the patient movements.

Quite apart from the usual paradigm of robust optimization, we propose an approach to optimize the budget of variation while ensuring that the worst-case expected maximum bed requirement over the planning horizon falls below the bed capacity of the hospital. This approach is inspired by the actual problem for which we have access to the data to attest the performance. The key challenge we face is to model uncertainty in a way while keeping the computations tractable so that we can obtain consistent improvement over the static approach for which uncertainty is ignored. Interestingly, this could be achieved by solving a small collection of computationally tractable optimization problems. We also study the performance of this approach in a case study using real data.

The rest of this chapter is organized as follows. In Section 3.1, we establish a distributionally robust optimization model for managing elective admission in hospital with incomplete information of uncertainties. We then investigate deterministic formulation to this model by deriving a second order conic optimization problem (SOCP) in Section 3.2. Numerical experiments using real data are carried out in Section 3.3. Section 3.4 concludes this chapter.

Notation: We denote a random variable with a tilde sign, such as $\tilde{z}$. Matrices and vectors are represented as upper and lower case boldface characters respectively. If $\boldsymbol{x}$ is a vector, we use the notation x_i to denote the ith component of the vector. We represent uncertainty by a state-space Ω and a set (σ-algebra) $\mathcal{F}$ of events. We use the notation $\tilde{x} \geq \tilde{y}$ to denote state-wise dominance over all attributes, i.e, $\tilde{x}(\omega) \geq \tilde{y}(\omega)$ for all $\omega \in \Omega$. We use $\mathbb{P}$ to denote a probability measure on Ω and $\mathbb{E}_{\mathbb{P}}(\tilde{x})$, $\sigma_{\mathbb{P}}(\tilde{x})$ and $\mathrm{cv}_{\mathbb{P}}(\tilde{x})$ denote respectively the expectation, standard deviation and coefficient of variation of $\tilde{x}$ under $\mathbb{P}$.

3.1 Model formulation

We consider a planning horizon of T days indexed by $t = 0, 1, \ldots, T-1$. Let η_t be the decision variable representing the elective inpatients quota for the tth day within the planning horizon. For simplicity of model presentation, we assume that all inpatients are of the same type. We can easily refine the model to consider quotas for different types of inpatients that may be characterized by gender, discipline, and so forth. We detail how this is implemented in a public hospital. At the beginning of day $t = 0$ (say at 8 am when clinics open), the quotas $\boldsymbol{\eta} = (\eta_0, \ldots, \eta_{T-1})'$ will be determined and integrated within hospital decision support system for assignment of admissions. As a public hospital, the hospital does *not reject* elective admissions. During the operating hours of the elective clinics, administrators work with the patients for their admission dates, which would depend on the availability of quotas. The booking system is similar to the airline booking system. When an elective bed request is submitted, the patient could only be assigned to the days where the quota is strictly greater than the number of patients that have been assigned, which is represented as $\underline{\boldsymbol{\eta}}$. Suppose this new elective patient is scheduled to be admitted on day l, the corresponding $\underline{\eta_l}$ will be updated as $\underline{\eta_l} = \underline{\eta_l} + 1$. As we proceed to the next day, the process is repeated and a new set of quotas will be computed using the latest information on admission status.

We let $X \subseteq \mathbb{Z}^T$ be the feasible space of admissible quotas. The feasible set X should be specified accordingly to exclude trivial results such as zero assigned quotas for elective patients. For instance, since the hospital sets aside a portion of her capacity to serve elective patients, we enforce by constraining the total quotas during the planning horizon to match the desired average number of elective patients. In the rolling horizon implementation, it is also imperative to ensure that the new set of quotas is able to accommodate previously assigned elective admissions. For example, if 15 elective admissions have already been assigned on day $t = 6$, we would impose a constraint $\eta_6 \geq 15$.

We next describe the dynamics of patient flow. Let L be the maximum duration of stay for any patient. Note that by definition, inpatients are patients who are warded for at least one day. To account for the total number of inpatients on the tth day, we need to keep track of the admission status up to $L-1$ days before the planning horizon. Let $\mathcal{T}^+ = \{0, \ldots, T-1\}$, $\mathcal{T}^{--} = \{-L+1, \ldots, -1\}$ and $\mathcal{T} = \mathcal{T}^{--} \cup \mathcal{T}^+$. We denote $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ to be respectively the number of emergency and elective inpatients arriving on the tth day, $t \in \mathcal{T}$ and would be warded for at least l days, $l \in \{1, \ldots, L\}$. For instance, $\tilde{p}_{1,1}$ refers to the total number of emergency inpatients on day $t = 1$ and its value is uncertain. If $\tilde{d}$ of these patients are discharged on day $t = 2$, then $\tilde{p}_{1,2} = \tilde{p}_{1,1} - \tilde{d}$. Likewise $\tilde{a}_{-1,2}$ refers to the number of elective inpatients that arrive on the previous day ($t = -1$) and would be warded for at least 2 days. At the beginning of day $t = 0$, doctors may not have reviewed the cases for discharge. Hence, the parameter $\tilde{a}_{-1,2}$ is generally uncertain. For our purpose, we need to account for the number of inpatients during the planning horizon, i.e., on the days in $\mathcal{T}^+$. For inpatients arriving on day $t \in \mathcal{T}^{--}$, only the inpatients with the length of stay of at least l days, $l \geq 1 - t$, may remain warded in the hospital during the planning horizon. On the other hand, for patients arriving on day $t \in \mathcal{T}^+$, only the information associated with inpatients with length of stay at least l days, $l \leq \min\{L, T - t\}$ will be needed to compute the quotas. Hence, for notational convenience, we define $\mathcal{L}_t = \{\max\{1, 1-t\}, \max\{1, 1-t\}+1, \ldots, \min\{L, T-t\}\}$, $t \in \mathcal{T}$.

We now account for the total number of inpatients on the tth day during the planning horizon, $t \in \mathcal{T}^+$. For example, the total number of inpatients on day $t = 0$ can be computed as follows

$$\begin{array}{ll} \tilde{a}_{0,1} + \tilde{p}_{0,1} + & \text{(arrivals/admissions on } t = 0) \\ \tilde{a}_{-1,2} + \tilde{p}_{-1,2} + & \text{(arrivals/admissions on } t = -1 \text{ and warded for at least 2 days)} \\ \cdots + \tilde{a}_{-L+1,L} + \tilde{p}_{-L+1,L}. & \text{(arrivals/admissions on } t = -L+1 \text{ and warded for up to } L \text{ days)} \end{array}$$

In general, it follows that the total inpatients on day $t \in \mathcal{T}^+$ can be computed as

$$\sum_{(\tau,l) \in \mathcal{U}_t} (\tilde{a}_{\tau,l} + \tilde{p}_{\tau,l}),$$

where the index set $\mathcal{U}_t$ is given by

$$\mathcal{U}_t = \{(\tau, l) \ : \ \tau \in \mathcal{T}, \ l \in \mathcal{L}_\tau, l + \tau = t + 1\}.$$

A bed shortfall occurs whenever the total number of inpatients exceeds the bed capacity, which we denote by c_t, $t \in \mathcal{T}^+$. Note that for generality, we assume that bed capacity, which encompasses the physical beds and manpower availability, is time dependent. Before we could specify an optimization problem, we first need to account for the uncertainty concerning patients arrival and departure.

3.1.1 Characterizing patient arrivals and departures uncertainty

We describe a nonparametric approach for characterizing the uncertainty on patient arrivals and departures using information obtained from patient movement records. Our aim is to introduce a model of uncertainty without imposing excessive burden on the information requirement, which may otherwise deter practical implementation. Instead of ignoring variability and assuming deterministic parameters taking values at their empirical averages, which is usually done in practice, we assume that the parameters are random variables with known means. However, their precise distributions are unavailable but belong to a restricted ambiguity set. To avoid being overly conservative, we control the "size" of the ambiguity set by specifying the budget of variation, μ, which is the upper bound of the coefficients of variations of all the uncertain parameters.

We next show how the uncertain parameters $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ are interrelated, which is the basis for characterizing the support of the uncertainty. Observe that by definition, $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ are nonincreasing in l. For inpatients arriving before $t = 0$, their total admissions are known but their durations of stay may be uncertain. Let p_t^0 and a_t^0, $t \in \mathcal{T}^{--}$, be respectively the number of remaining emergency and elective inpatients who have arrived on day t and are still being warded up to the beginning of day 0. The support of the uncertain parameters $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ is given by

$$p_t^0 \geq \tilde{p}_{t,l} \geq \tilde{p}_{t,l'} \geq 0,$$
$$a_t^0 \geq \tilde{a}_{t,l} \geq \tilde{a}_{t,l'} \geq 0,$$

for all $t \in \mathcal{T}^{--}$, $l, l' \in \mathcal{L}_t, l' > l$. Similarly, for inpatients arriving during the planning horizon $t \in \mathcal{T}^{+}$, the support of the associated uncertain parameters $\tilde{p}_{t,l}$, $\tilde{a}_{t,l}$, is given by

$$p_t^0 \geq \tilde{p}_{t,l} \geq \tilde{p}_{t,l'} \geq 0,$$
$$\eta_t \geq \tilde{a}_{t,l} \geq \tilde{a}_{t,l'} \geq 0,$$

for all $t \in \mathcal{T}^{+}$, $l, l' \in \mathcal{L}_t, l' > l$. For the emergency patients, the input parameter,

p_t^0 is a prescribed upper bound of $\tilde{p}_{t,l}$. For the elective patients, according to the admission process we have described, the number of patients arriving at the tth day and be warded at least l days, $\tilde{a}_{t,l}$, is an endogenous random variable that depends on the quota, η_t. If $\eta_t = 0$, then it is clear that $\tilde{a}_{t,l} = 0$ for all $l \in \mathcal{L}_t$. We provide an example to illustrate this dependency. Suppose at $t = 1$, $\eta_1 = 10, \eta_2 = 1, \eta_3 = 15$, and the number of assigned electives, $\underline{\eta_1} = 10, \underline{\eta_2} = 0, \underline{\eta_3} = 10$, the hospital would be able to schedule new elective patients at $t = 2$ or $t = 3$, but not at $t = 1$. If every elective patient turns up at $t = 1$, then $\tilde{a}_{1,1} = 10$. Hence, $\tilde{a}_{t,l}$ is highly dependent on η_t.

Instead of assuming a probability distribution, we specify the ambiguity set such that for each distribution, $\mathbb{P}$ in the set, the uncertain parameters are random variables with known mean values and their coefficients of variations are bounded below by μ. Specifically, for inpatients arriving before $t = 0$, i.e., $t \in \mathcal{T}^{--}$, we assume that

$$\mathbb{E}_{\mathbb{P}}(\tilde{p}_{t,l}) = \bar{p}_{t,l},$$
$$\mathbb{E}_{\mathbb{P}}(\tilde{a}_{t,l}) = \bar{a}_{t,l},$$

for all $l \in \mathcal{L}_t$, where $\bar{p}_{t,l}$ and $\bar{a}_{t,l}$ are respectively the empirical averages of $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$. Since these patients are already admitted, in principle, the parameters $\bar{p}_{t,l}, \bar{a}_{t,l}$ may be inferred from the patients' likely duration of stay assessed by their doctors. If such information is unavailable, then one may also use values that are empirically estimated from historical records.

Observe that during the planning horizon, $t \in \mathcal{T}^+$, the uncertain parameters $\tilde{p}_{t,l}, t \in \mathcal{T}^+$ are associated with inpatients who have yet to arrive at the hospital. Hence, we are able to obtain the empirical averages from patient movement records as follows,

$$\mathbb{E}_{\mathbb{P}}(\tilde{p}_{t,l}) = \bar{p}_{t,l},$$

for all $t \in \mathcal{T}^+, l \in \mathcal{L}_t$. Unlike the previous case, the elective patients $\tilde{a}_{t,l}, t \in \mathcal{T}^+, l \in \mathcal{L}_t$ are associated with the quotas η_t. Clearly, the dependency on η_t

would impact on how the parameters should be estimated and how the model can be solved efficiently. Accordingly, we define new random variables $\tilde{\alpha}_{t,l}$ as

$$\tilde{\alpha}_{t,l} = \frac{\tilde{a}_{t,l}}{\eta_t},$$

for all $t \in \mathcal{T}^+$, $l \in \mathcal{L}_t$ to represent the proportion of patients who will be warded at least l days with respect to the quota η_t. From the historical data on $\tilde{a}_{t,l}$ and η_t, we could determine its empirical average as

$$\mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{t,l}) = \bar{\alpha}_{t,l},$$

for all $t \in \mathcal{T}^+, l \in \mathcal{L}_t$.

To formulate a tractable and scalable model that could be solved by commercial solvers, we impose the following assumption.

Assumption 3. *The descriptive statistics of $\tilde{\alpha}_{t,l}$ are independent on η_t.*

Assumption 3 has important ramifications on the computational tractability of the model, which we will explain in Section 3.2. It leads to simpler estimation of the descriptive statistics from data.

For notational simplicity, since we have the complete information for the quotas assigned before $t = 0$, we define

$$\begin{aligned} &\tilde{\alpha}_{t,l} = \frac{\tilde{a}_{t,l}}{\eta_t}, \quad \bar{\alpha}_{t,l} = \frac{\bar{a}_{t,l}}{\eta_t}, \quad \alpha_t^0 = \frac{a_t^0}{\eta_t}, && \forall\, t \in \mathcal{T}^{--}, l \in \mathcal{L}_t, \\ &\alpha_t^0 = 1, && \forall\, t \in \mathcal{T}^+. \end{aligned}$$

Hence we can empirically determine the values for $\{\bar{\alpha}_{t,l} : t \in \mathcal{T}, l \in \mathcal{L}_t\}$ and $\{\alpha_t^0 : t \in \mathcal{T}^{--}\}$ from the data.

Finally, the coefficients of variation of these parameters are bounded above by the budget of variation, μ as follows

$$\mathrm{cv}_{\mathbb{P}}(\tilde{p}_{t,l}) \leq \mu,$$
$$\mathrm{cv}_{\mathbb{P}}(\tilde{\alpha}_{t,l}) \leq \mu,$$

for all $t \in \mathcal{T}$, $l \in \mathcal{L}_t$. Hence, $\mu = 0$, implies that the parameters are almost surely certain and take values at their means. On the other extreme with $\mu = \infty$, then essentially the variabilities of these parameters are not constrained by μ, but could otherwise be limited by the support. We present the ambiguity set as a function of the budget of variation, μ as follows.

$$\mathbb{F}(\mu) = \left\{ \mathbb{P} \ : \begin{array}{ll} \mathbb{P}\left(\left\{ \begin{array}{ll} p_t^0 \geq \tilde{p}_{t,l} \geq \tilde{p}_{t,l'} \geq 0, & \forall\, (t,l'),(t,l) \in \mathcal{I}, l' > l \\ \alpha_t^0 \geq \tilde{\alpha}_{t,l} \geq \tilde{\alpha}_{t,l'} \geq 0, & \forall\, (t,l'),(t,l) \in \mathcal{I}, l' > l \end{array} \right\} \right) = 1 \\ \mathbb{E}_{\mathbb{P}}(\tilde{p}_{t,l}) = \bar{p}_{t,l}, \quad \forall\, (t,l) \in \mathcal{I}, \\ \mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{t,l}) = \bar{\alpha}_{t,l}, \quad \forall\, (t,l) \in \mathcal{I}, \\ \sigma_{\mathbb{P}}(\tilde{p}_{t,l}) \leq \bar{p}_{t,l}\mu, \quad \forall\, (t,l) \in \mathcal{I}, \\ \sigma_{\mathbb{P}}(\tilde{\alpha}_{t,l}) \leq \bar{\alpha}_{t,l}\mu, \quad \forall\, (t,l) \in \mathcal{I} \end{array} \right\},$$

where

$$\mathcal{I} := \{(t,l) : t \in \mathcal{T}, l \in \mathcal{L}_t\}.$$

and the parameters for characterizing the ambiguous set of distributions, $\{p_t^0, \alpha_t^0 \ : \ t \in \mathcal{T}\}$, $\{\bar{p}_{t,l} \ : \ (t,l) \in \mathcal{I}\}$, $\{\bar{\alpha}_{t,l} \ : \ (t,l) \in \mathcal{I}\}$ are values obtained from the patient movement records. Observe that the set $\mathbb{F}(\mu)$ is nondecreasing in μ, i.e.,

$$\mathbb{F}(\mu) \subseteq \mathbb{F}(\mu') \qquad \forall\, \mu' \geq \mu.$$

Remark 1. For tractability purpose, the discrete nature of the the uncertain arrivals of emergency and elective patients are not characterized in the ambiguity set. The relaxation of integer random variables to continuous ones is a common technique used in robust optimization to obtain tractable formulations. If we confine to integer random variables, we would have to enumerate exponentially many scenarios to obtain an exact formation, which would lead to intractability. This integrality gap can be significant. For instance, consider a univariate random variable, $\tilde{p}$ taking values in $\{0, 1\}$ and two distributionally ambiguity sets

$$\mathbb{F}_1 = \{\mathbb{P} \ : \ \mathbb{E}_{\mathbb{P}}(\tilde{p}) = 0.5, \mathbb{E}_{\mathbb{P}}(\tilde{p}^2) \leq 0.1^2, \mathbb{P}(\tilde{p} \in [0, 1]) = 1\}$$

and

$$\mathbb{F}_2 = \{\mathbb{P} \ : \ \mathbb{E}_{\mathbb{P}}(\tilde{p}) = 0.5, \mathbb{E}_{\mathbb{P}}(\tilde{p}^2) \leq 0.1^2, \mathbb{P}(\tilde{p} \in \{0, 1\}) = 1\}.$$

Observe that the set, $\mathbb{F}_1$ is a conservative approximation of $\mathbb{F}_2$, which is an empty set. Hence, the “integrality gap” can be arbitrarily bad. The ambiguity set we consider in the bed management problem is far more complex, and we do not see how the “integrality gap” can be eliminated in a computationally efficient manner. Likewise, there are other types of distributional ambiguity information that do not lead to computationally tractable formations. Among others, we are unable to obtain tight and tractable formulations for ambiguity information such as higher order moments with support, independence of random variables, and so forth. As in the spirit of robust optimization models, the goal here is to model uncertainty in its entire generality while keeping the model within the framework for which current state-of-the-art commercial solvers can deliver. In our computational study, we observe the importance of adjusting the conservativeness of the ambiguity set through the parameter μ. Hence, instead of fixing μ, we will propose an approach of maximizing the size of the ambiguity set (hence, the level of conservativeness) subject to a threshold constraint. In our computational studies, we observe this approach provides significant improvement over an approach with fixed ambiguity set. As it would become clearer, the ambiguity set, as we have defined, enables us to obtain the solution by solving a sequence of tractable optimization problems.

Remark 2. It would also be possible to extend our model to allow for multiple layers of uncertainty, for example, by providing confidence intervals of mean estimates in the distributional ambiguity set. The key issue we face is how we can calibrate the model of uncertainty so that we can have a reliable performance from our data. After experimenting with several distributionally robust optimization models such as incorporating variance estimates, confidence intervals of mean estimates, we observe from our available data that the current model provides consistency in performance improvement.

3.1.2 Distributionally robust optimization models

To circumvent the difficulties of obtaining probability distributions and solving the complex stochastic model, the elective smoothing approach ignores uncertainty and solves the following deterministic optimization problem:

$$Z_D = \min_{\boldsymbol{\eta} \in X} \left(\max_{t \in \mathcal{T}^+} \left\{ \sum_{(\tau,l) \in \mathcal{U}_t} (\bar{\alpha}_{\tau,l} \eta_\tau + \bar{p}_{\tau,l}) - c_t \right\} \right). \tag{3.1}$$

The decision variables here are the quotas $\boldsymbol{\eta} = (\eta_\tau)_{\tau \in \mathcal{T}^+}$. Observe that the model is essentially an attractive linear optimization problem if X is a polyhedron. The aim of the objective function is not merely to minimize bed shortages that might occur during the planning horizon, but rather to "smooth out" the daily bed occupancy by minimizing the maximum occupancy over the horizon. This is a service inspired criterion to better accommodate for fluctuations in bed demands, which is known in the literature as elective smoothing. As illustrated in Figure 3.1, even in the absence of bed shortages, the minimax criterion would favor the bed allocation in Scenario 2, which is uniformly distributed, over that of Scenario 1.

Nevertheless, despite being a tractable linear optimization problem, the model ignores the potential impact of uncertainty and could lead to severe shortfalls in hospital beds whenever bad scenarios arises. A natural extension of the elective

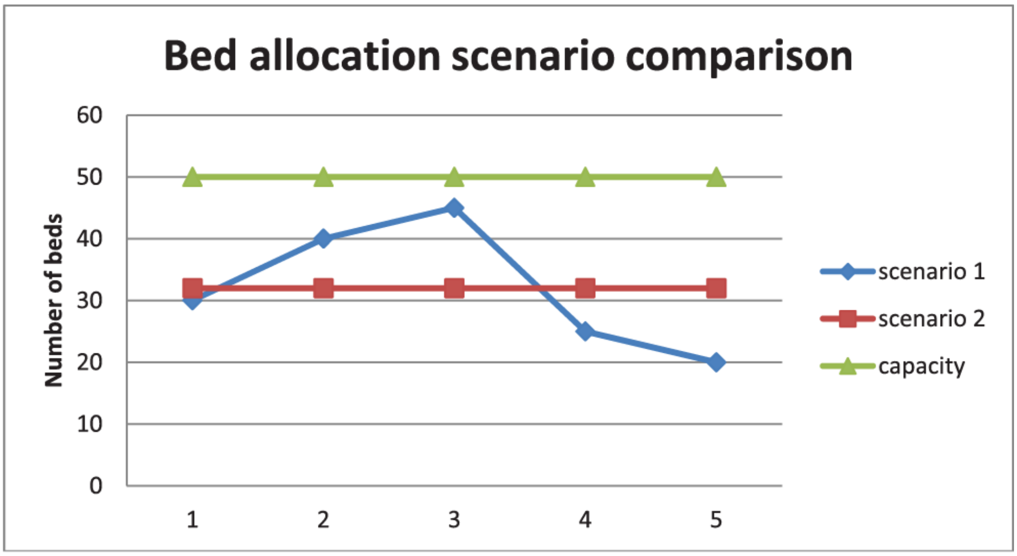

Figure 3.1: An Illustrative example of bed allocation policy. -

smoothing approach to incorporate uncertainty is to minimize the worst-case expected maximum bed excess over the planning horizon as follows

$$Z_R(\mu) = \min_{\boldsymbol{\eta} \in X} \sup_{\mathbb{P} \in \mathbb{F}(\mu)} \mathbb{E}_{\mathbb{P}} \left(\max_{t \in \mathcal{T}^+} \left\{ \sum_{(\tau,l) \in \mathcal{U}_t} (\tilde{\alpha}_{\tau,l} \eta_\tau + \tilde{p}_{\tau,l}) - c_t \right\} \right). \tag{3.2}$$

In the absence of uncertainty, i.e., $\mu = 0$, it is clear that Model (3.1) is the same as Model (3.2), hence $Z_D = Z_R(0)$. As we increase the budget of variation μ, the model takes into consideration more potential variations in the admission process. In this approach, it is the onus of the modeler's to set the budget of variations, μ. We refer to this model the fixed budget model.

Optimized budget of variation model

The main challenge of Model (3.2) is how to specify the value of μ that would yield the desired level of performance in controlling bed shortfalls. Intuitively, a meagerly or overly specified budget of variation, μ may not adequately protect against potential bed shortfalls when the actual uncertainty is realized. In practice, the parameter μ has to be tuned accordingly so that it gives the best overall performance on real data.

We note that Model (3.2) is only a means to cope with the issue of bed shortfalls. In a well managed hospital, it is imperative that beds capacity should exceed average demands, which implies $Z_D = Z_R(0) \leq 0$. Extending this notion to incorporate uncertainty, if $Z_R(\mu) \leq 0$, for $\mu > 0$, then we are guaranteed a solution that ensures that for all $\mathbb{P} \in \mathbb{F}(\mu)$, the expected maximum bed excess across the time periods is less then zero, i.e.,

$$\mathbb{E}_{\mathbb{P}}\left(\max_{t\in\mathcal{T}^+}\left\{\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau + \tilde{p}_{\tau,l}) - c_t\right\}\right) \leq 0, \qquad \forall\, \mathbb{P} \in \mathbb{F}(\mu).$$

In light of the above discussion, we propose another robust optimization approach, i.e., to find the most reliable solution that would protect against the worst uncertainty that might lead to bed shortfalls. In other words, we hope to maximize the level of uncertainty that the system can absorb without going into bed shortages. Hence, we push the boundary of uncertainty by maximizing the budget of variation, μ subject to $Z_R(\mu) \leq 0$ as follows

$$\begin{aligned} \mu^* = \max \quad & \mu \\ \text{s.t.} \quad & Z_R(\mu) \leq 0 \\ & \mu \in [0, \infty). \end{aligned} \tag{3.3}$$

Since the set $\mathbb{F}(\mu)$ is nondecreasing in μ, the function $Z_R(\mu)$ is also nondecreasing in μ. As a result, Model (3.3) is feasible and finite if and only if $Z_R(0) \leq 0$ and $Z_R(\infty) \geq 0$. Moreover, it is reasonable to assume that the inequalities are strict so that the bed capacity is sufficient to meet average demands but also not overly excessive. As opposed to the fixed budget model, we refer to this model the optimized budget model.

We note that the target value zero on the right hand side of the constraint can further be adjusted accordingly to match the service level desired by the hospital. For simplicity, we leave it at zero.

The optimal solution of Model (3.3) can easily be obtained by binary search and solving a sequence of subproblems in the form of Model (3.2) so that $Z_R(\mu^*) = 0$.

3.2 Tractable formulation

In this section, we first study the inner maximization problem of Model (3.2), and formulate it as a second order cone programming problem. Subsequently, we develop a tractable formulation of Model (3.2) in form of a deterministic SOCP. Since the problem is easy to solve when $\mu = 0$, we will focus on the case for which $\mu > 0$. We first focus on the inner maximization problem in Model (3.2), i.e.,

$$\begin{array}{lll} \sup & \mathbb{E}_{\mathbb{P}}\left(\max_{t\in\mathcal{T}^+}\left\{\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau + \tilde{p}_{\tau,l}) - c_t\right\}\right) & \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}(\tilde{p}_{\tau,l}) = \bar{p}_{\tau,l}, & \forall\,(\tau,l)\in\mathcal{I}, \\ & \mathbb{E}_{\mathbb{P}}(\tilde{p}^2_{\tau,l}) \le \bar{p}^2_{\tau,l}(1+\mu^2), & \forall\,(\tau,l)\in\mathcal{I}, \\ & \mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{\tau,l}) = \bar{\alpha}_{\tau,l}, & \forall\,(\tau,l)\in\mathcal{I}, \\ & \mathbb{E}_{\mathbb{P}}(\tilde{\alpha}^2_{\tau,l}) \le \bar{\alpha}^2_{\tau,l}(1+\mu^2), & \forall\,(\tau,l)\in\mathcal{I}, \\ & \mathbb{P}\left\{(\tilde{p}_{\tau,l},\tilde{\alpha}_{\tau,l})_{(\tau,l)\in\mathcal{I}} \in W_p\times W_a\right\} = 1, & \end{array} \tag{3.4}$$

where

$$W_p := \left\{(p_{\tau,l})_{(\tau,l)\in\mathcal{I}} \;:\; p^0_\tau \ge p_{\tau,l} \ge p_{\tau,l'} \ge 0,\ \forall\,(\tau,l),(\tau,l')\in\mathcal{I}, l'>l\right\},$$

$$W_a := \left\{(\alpha_{\tau,l})_{(\tau,l)\in\mathcal{I}} \;:\; \alpha^0_\tau \ge \alpha_{\tau,l} \ge \alpha_{\tau,l'} \ge 0,\ \forall\,(\tau,l),(\tau,l')\in\mathcal{I}, l'>l\right\}.$$

Note also that W_p and W_a are actually the cross products of a number of sets with respect to parameters $\tau \in \mathcal{T}$. In other words, W_p and W_a can be rewritten as

$$W_p = \Pi_{\tau\in\mathcal{T}} W_p^\tau, \quad W_a = \Pi_{\tau\in\mathcal{T}} W_a^\tau,$$

where

$$W_p^\tau := \left\{(p_{\tau,l})_{l\in\mathcal{L}_\tau} \;:\; p_\tau^0 \geq p_{\tau,l} \geq p_{\tau,l'} \geq 0, \forall\, l, l' \in \mathcal{L}_\tau,\, l' > l\right\}, \;\; \tau \in \mathcal{T},$$
$$W_a^\tau := \left\{(\alpha_{\tau,l})_{l\in\mathcal{L}_\tau} \;:\; \alpha_\tau^0 \geq \alpha_{\tau,l} \geq \alpha_{\tau,l'} \geq 0, \forall\, l, l' \in \mathcal{L}_\tau,\, l' > l\right\}, \;\; \tau \in \mathcal{T}.$$

Problem (3.4) is a maximization problem over a probability distribution function, which is generally an intractable optimization problem, see for instance Murty and Kabadi (1987). However, under our model of uncertainty, we will show an equivalent formulation of Problem (3.4), namely its dual problem, is a minimization problem in the form of SOCP. As a result, Problem (3.4) can be readily solved by existing commercialized SOCP solvers, such as CPLEX and MOSEK.

For convenience in description, for each $t \in \mathcal{T}^+$, let $z_{\tau,l}^t$ denote the indicator function defined by

$$z_{\tau,l}^t = \begin{cases} 1, & \text{if } \tau + l = t + 1, \\ 0, & \text{otherwise}, \end{cases}$$

for any $(\tau, l) \in \mathcal{I}$. Noticing that $\mathcal{U}_t = \{(\tau, l) \in \mathcal{I} : \tau + l = t + 1\}$, item $\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau + \tilde{p}_{\tau,l}) - c_t$ in the objective of (3.4) can then be expressed as below

$$\sum_{(\tau,l)\in\mathcal{I}} (\tilde{\alpha}_{\tau,l}\eta_\tau + \tilde{p}_{\tau,l}) z_{\tau,l}^t - c_t, \;\; \forall\, t \in \mathcal{T}^+.$$

By applying duality theory, we derive an equivalent formulation of Problem (3.4) as follows:

Theorem 4. *Problem (3.4) has the same objective as the following optimization* *problem:*

$$\inf_{\rho,(s_{\tau,l},u_{\tau,l},v_{\tau,l},w_{\tau,l})_{(\tau,l)\in\mathcal{I}}} \Bigg\{\rho + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}_{\tau,l}s_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}^2_{\tau,l}(1+\mu^2)u_{\tau,l}$$

$$+ \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}_{\tau,l}v_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}^2_{\tau,l}(1+\mu^2)w_{\tau,l}\Bigg\} \tag{3.5}$$

$$\text{s. t.} \quad \sum_{\tau\in\mathcal{T}} \pi_1^t(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau) + \sum_{\tau\in\mathcal{T}} \pi_2^t(\boldsymbol{v}_\tau - \eta_\tau \boldsymbol{z}^t_\tau, \boldsymbol{w}_\tau) + \rho + c_t \geq 0,\ \forall\, t \in \mathcal{T}^+,$$
$$u_{\tau,l},\ \ w_{\tau,l} \geq 0,\ \forall\,(\tau,l)\in\mathcal{I},$$

where for $t \in \mathcal{T}^+$, $\tau \in \mathcal{T}$,

$$\pi_1^t(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau) := \min\left\{ \sum_{l\in\mathcal{L}_\tau} \left((s_{\tau,l} - z^t_{\tau,l})p_{\tau,l} + u_{\tau,l}p^2_{\tau,l}\right) \middle| (p_{\tau,l})_{l\in\mathcal{L}_\tau} \in W^\tau_p \right\},$$

$$\pi_2^t(\boldsymbol{v}_\tau - \eta_\tau \boldsymbol{z}^t_\tau, \boldsymbol{w}_\tau) := \min\left\{ \sum_{l\in\mathcal{L}_\tau} \left((v_{\tau,l} - \eta_\tau z^t_{\tau,l})\alpha_{\tau,l} + w_{\tau,l}\alpha^2_{\tau,l}\right) \middle| (\alpha_{\tau,l})_{l\in\mathcal{L}_\tau} \in W^\tau_a \right\},$$

and $\rho \in \Re$, $\boldsymbol{s}_\tau = (s_{\tau,l})_{l\in\mathcal{L}_\tau}$, $\boldsymbol{u}_\tau = (u_{\tau,l})_{l\in\mathcal{L}_\tau}$, $\boldsymbol{v}_\tau = (v_{\tau,l})_{l\in\mathcal{L}_\tau}$, $\boldsymbol{w}_\tau = (w_{\tau,l})_{l\in\mathcal{L}_\tau}$, $\boldsymbol{z}^t_\tau = (z^t_{\tau,l})_{l\in\mathcal{L}_\tau}$.

Proof. The dual problem of Problem (3.4) can be written as

$$\begin{aligned}
\inf_{\rho,(s_{\tau,l},u_{\tau,l},v_{\tau,l},w_{\tau,l})_{(\tau,l)\in\mathcal{I}}} \quad & \Bigg\{\rho + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}_{\tau,l}s_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}^2_{\tau,l}(1+\mu^2)u_{\tau,l} \\
& + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}_{\tau,l}v_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}^2_{\tau,l}(1+\mu^2)w_{\tau,l}\Bigg\} \\
\text{s.t.} \quad & \rho + \sum_{(\tau,l)\in\mathcal{I}} s_{\tau,l}p_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} u_{\tau,l}p^2_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} v_{\tau,l}\alpha_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} w_{\tau,l}\alpha^2_{\tau,l} \\
& \geq \sum_{(\tau,l)\in\mathcal{U}_t} (\alpha_{\tau,l}\eta_\tau + p_{\tau,l}) - c_t, \ \forall\, t \in \mathcal{T}^+, \ (p_{\tau,l},\alpha_{\tau,l})_{(\tau,l)\in\mathcal{I}} \in W_p \times W_a, \\
& u_{\tau,l}, \ w_{\tau,l} \geq 0, \ \forall\, (\tau,l) \in \mathcal{I},
\end{aligned} \tag{3.6}$$

where $s_{\tau,l}, u_{\tau,l}, v_{\tau,l}, w_{\tau,l}$ and ρ are the Lagrange multipliers corresponding to the equality/inequality constraints concerning the first and second moments of $\tilde{p}_{\tau,l}$ and $\tilde{\alpha}_{\tau,l}$, $(\tau,l) \in \mathcal{I}$, together with the implicit constraint that $\mathbb{E}_{\mathbb{P}}[1] = 1$. Evidently, the multipliers, $u_{\tau,l}$ and $w_{\tau,l}$, $(\tau,l) \in \mathcal{I}$, corresponding to the inequality constraints are all nonnegative. This property, as we shall see, is very important in the subsequent analysis. Note that since the parameters of the ambiguity set are obtained empirically, there exist probability distributions that are feasible and hence, strong duality holds according to Shapiro (2001). Moreover, since $\mu > 0$, a distribution, $\mathbb{P}$ for which $\mathbb{P}\left\{(\tilde{p}_{\tau,l},\tilde{\alpha}_{\tau,l})_{(\tau,l)\in\mathcal{I}} = (\bar{p}_{\tau,l},\bar{\alpha}_{\tau,l})_{(\tau,l)\in\mathcal{I}}\right\} = 1$ would lead to expectation constraints that are strictly feasible.

Note that the system of inequality constraints in Problem (3.6) consists of infinitely many constraints. Using the notation of vector $z^t_{\tau,l}$, we can express the inequality system in the dual problem (3.6) as

$$\begin{aligned}
& \sum_{(\tau,l)\in\mathcal{I}} \left((s_{\tau,l} - z^t_{\tau,l})p_{\tau,l} + u_{\tau,l}p^2_{\tau,l}\right) + \sum_{(\tau,l)\in\mathcal{I}} \left((v_{\tau,l} - \eta_\tau z^t_{\tau,l})\alpha_{\tau,l} + w_{\tau,l}\alpha^2_{\tau,l}\right) \\
\geq \;& -\rho - c_t, \ \forall\, t \in \mathcal{T}^+, \ (p_{\tau,l},\alpha_{\tau,l})_{(\tau,l)\in\mathcal{I}} \in W_p \times W_a,
\end{aligned}$$

or equivalently,

$$\min\left\{ \sum_{(\tau,l)\in\mathcal{J}} (s_{\tau,l} - z^t_{\tau,l})p_{\tau,l} + \sum_{(\tau,l)\in\mathcal{J}} u_{\tau,l}p^2_{\tau,l} \,\middle|\, (p_{\tau,l})_{(\tau,l)\in\mathcal{J}} \in W_p \right\}$$
$$+\min\left\{ \sum_{(\tau,l)\in\mathcal{J}} (v_{\tau,l} - \eta_\tau z^t_{\tau,l})\alpha_{\tau,l} + \sum_{(\tau,l)\in\mathcal{J}} w_{\tau,l}\alpha^2_{\tau,l} \,\middle|\, (\alpha_{\tau,l})_{(\tau,l)\in\mathcal{J}} \in W_a \right\}$$
$$\geq -\rho - c_t,\ \forall\, t \in \mathcal{T}^+.$$

Note that the objectives in the above system are separable in $(p_{\tau,l})_{l\in\mathcal{L}_\tau}$, $(\alpha_{\tau,l})_{l\in\mathcal{L}_\tau}$ for $\tau \in \mathcal{T}$, respectively. Noticing that W_p and W_a can be written as the cross products of some sets with respect to the parameter $\tau \in \mathcal{T}$, thereby the "min" and "sum" operators on the left hand side are exchangeable. By recalling the definitions of W^τ_p, W^τ_a, we have

$$\sum_{\tau\in\mathcal{T}} \pi^t_1(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau) + \sum_{\tau\in\mathcal{T}} \pi^t_2(\boldsymbol{v}_\tau - \eta_\tau \boldsymbol{z}^t_\tau, \boldsymbol{w}_\tau) \geq -\rho - c_t,\ \forall\, t \in \mathcal{T}^+.$$

Thus, the desired result follows immediately. This completes the proof. □

Note that the equivalent formulation (3.5) in Theorem 4 is a deterministic counterpart of the objective function, $Z_R(\mu)$, of robust optimization model (3.2). To derive a tractable reformulation, in what follows, we investigate the underlying minimization problems in the constraints of (3.5), i.e., $\pi^t_i,\ i = 1, 2,\ t \in \mathcal{T}^+$. First, for any $\tau \in \mathcal{T}$, define an index set $\mathcal{L}^+_\tau := \mathcal{L}_\tau \cup \{1 + \min\{L, T - \tau\}\}$. We state the results as follows.

Proposition 6. *Given $\gamma \in \Re$. For $t \in \mathcal{T}^+$, the following statements hold true.*

(i) For any $\tau \in \mathcal{T}^{--}$, the system of inequality

$$\pi^t_1(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau) \geq \gamma \tag{3.7}$$

is second order cone representable in the sense that there exist $\lambda^t_{\tau,l} \geq 0,\ l \in$

$\mathcal{L}_\tau^+$, such that (3.7) is equivalent to

$$\begin{aligned}
&\sum_{l\in\mathcal{L}_\tau} y_{\tau,l}^t + p_\tau^0 \lambda_{\tau,1-\tau}^t + \gamma \le 0,\\
&4u_{\tau,l} y_{\tau,l}^t \ge \left(s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\right)^2, \; \forall\, l \in \mathcal{L}_\tau, \qquad (3.8)\\
&y_{\tau,l}^t \ge 0, \forall\, l \in \mathcal{L}_\tau.
\end{aligned}$$

(ii) For any $\tau \in \mathcal{T}^+$, the system of inequality (3.7) is second order cone representable in the sense that there exist $\lambda_{\tau,l}^t \ge 0,\ l \in \mathcal{L}_\tau^+$, such that (3.7) is equivalent to

$$\begin{aligned}
&\sum_{l\in\mathcal{L}_\tau} y_{\tau,l}^t + p_\tau^0 \lambda_{\tau,1}^t + \gamma \le 0,\\
&4u_{\tau,l} y_{\tau,l}^t \ge \left(s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\right)^2, \; \forall\, l \in \mathcal{L}_\tau, \qquad (3.9)\\
&y_{\tau,l}^t \ge 0, \forall\, l \in \mathcal{L}_\tau.
\end{aligned}$$

(iii) For any $\tau \in \mathcal{T}^{--}$, the system of inequality

$$\pi_2^t(\boldsymbol{v}_\tau - \eta_\tau \boldsymbol{z}_\tau^t, \boldsymbol{w}_\tau) \ge \gamma \qquad (3.10)$$

is second order cone representable in the sense that there exist $\lambda_{\tau,l}^t \ge 0,\ l \in \mathcal{L}_\tau^+$, such that

$$\begin{aligned}
&\sum_{l\in\mathcal{L}_\tau} y_{\tau,l}^t + \alpha_\tau^0 \lambda_{\tau,1-\tau}^t + \gamma \le 0,\\
&4w_{\tau,l} y_{\tau,l}^t \ge \left(v_{\tau,l} - \eta_\tau z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\right)^2, \; \forall\, l \in \mathcal{L}_\tau, \qquad (3.11)\\
&y_{\tau,l}^t \ge 0, \forall\, l \in \mathcal{L}_\tau.
\end{aligned}$$

(iv) For any $\tau \in \mathcal{T}^+$, the system of inequality (3.10) is second order cone rep-

resentable in the sense that there exist $\lambda^t_{\tau,l} \geq 0,\ l \in \mathcal{L}^+_\tau$, such that

$$\begin{aligned}
&\sum_{l\in\mathcal{L}_\tau} y^t_{\tau,l} + \alpha^0_\tau \lambda^t_{\tau,1} + \gamma \leq 0, \\
&4w_{\tau,l}y^t_{\tau,l} \geq \left(v_{\tau,l} - \eta_\tau z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1}\right)^2,\ \forall\, l \in \mathcal{L}_\tau, \\
&y^t_{\tau,l} \geq 0, \forall\, l \in \mathcal{L}_\tau.
\end{aligned} \tag{3.12}$$

Proof. (i). By definition, problem $\pi^t_1(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau)$ can be written as

$$\begin{aligned}
\pi^t_1(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau) = \min \quad & \sum_{l\in\mathcal{L}_\tau} \left(u_{\tau,l}p^2_{\tau,l} + (s_{\tau,l} - z^t_{\tau,l})p_{\tau,l}\right) \\
\text{s. t.} \quad & p^0_\tau \geq p_{\tau,l} \geq p_{\tau,l'} \geq 0,\ \ \forall\, l, l' \in \mathcal{L}_\tau,\ l' > l.
\end{aligned} \tag{3.13}$$

Note that the above problem is a quadratic programming in which the coefficients concerning the second degree are nonnegative as $u_{\tau,l} \geq 0$ for $l \in \mathcal{L}_\tau$ by Theorem 4. To solve this problem, we consider its dual as given below:

$$\max_{\boldsymbol{\lambda}^t_\tau \geq 0} \zeta(\boldsymbol{\lambda}^t_\tau), \tag{3.14}$$

where $\zeta(\boldsymbol{\lambda}^t_\tau)$ is the associated Lagrange dual function, $\boldsymbol{\lambda}^t_\tau \in \Re^{|\mathcal{L}_\tau|+1}$ denotes the vector of the corresponding Lagrange multipliers, and for any given set S, $|S|$ denotes the cardinality of S.

Let $\boldsymbol{p}_\tau = (p_{\tau,l})_{l\in\mathcal{L}_\tau}$. For convenience in description and without loss of generality, we assume the indices of the entries in vector $\boldsymbol{\lambda}^t_\tau$ are consistent with those of $\boldsymbol{p}_\tau$, i.e., $\boldsymbol{\lambda}^t_\tau = (\lambda^t_{\tau,l})_{l\in\mathcal{L}^+_\tau}$. Applying some basic operations, it gives the Lagrange dual function as follows:

$$\zeta(\boldsymbol{\lambda}^t_\tau) := \min_{(p_{\tau,l})_{l\in\mathcal{L}_\tau}} \left\{ \sum_{l\in\mathcal{L}_\tau} \left(u_{\tau,l}p^2_{\tau,l} + (s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1})p_{\tau,l}\right) - p^0_\tau \lambda^t_{\tau,1-\tau} \right\}.$$

Note that the Slater's condition holds true since the interior of the feasible region of Problem (3.13) is nonempty. By the strong duality theorem, the system of inequality (3.7) can then be written as what follows. There exist $\lambda^t_{\tau,l} \geq 0,\ l \in \mathcal{L}^+_\tau$, such that

$$\min_{(p_{\tau,l})_{l\in\mathcal{L}_\tau}} \left\{ \sum_{l\in\mathcal{L}_\tau} \left(u_{\tau,l} p^2_{\tau,l} + (s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1}) p_{\tau,l} \right) \right\} - p^0_\tau \lambda^t_{\tau,1-\tau} \geq \gamma,$$

which, by virtue of the separability of the above minimization problem in $p_{\tau,l}$, can be further reformulated as

$$\sum_{l\in\mathcal{L}_\tau} \left(\min_{p_{\tau,l}} \left\{ u_{\tau,l} p^2_{\tau,l} + (s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1}) p_{\tau,l} \right\} \right) - p^0_\tau \lambda^t_{\tau,1-\tau} \geq \gamma. \quad (3.15)$$

To investigate the quadratic programming problems on the left hand side of (3.15), we consider the following two cases: (a) $u_{\tau,l} > 0$ for all $l \in \mathcal{L}_\tau$; (b) $u_{\tau,l} = 0$ for some $l \in \mathcal{L}_\tau$, respectively.

For case (a), solving the optimality condition of each minimization problem involved, i.e., $2u_{\tau,l} p_{\tau,l} + s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1} = 0$, $l \in \mathcal{L}_\tau$, we immediately derive the optimal solution and the corresponding optimal value, which are denoted by $p^*_{\tau,l}$ and $f^*_{\tau,l}$ as follows:

$$\begin{aligned} p^*_{\tau,l} &= \frac{1}{2u_{\tau,l}} \left(z^t_{\tau,l} - s_{\tau,l} + \lambda^t_{\tau,l+1} - \lambda^t_{\tau,l} \right), \quad l \in \mathcal{L}_\tau, \\ f^*_{\tau,l} &= -\frac{1}{4u_{\tau,l}} \left(s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1} \right)^2, \quad l \in \mathcal{L}_\tau. \end{aligned}$$

Substituting the optimal value $f^*_{\tau,l}$ to the inequality (3.15), it yields that

$$\sum_{l\in\mathcal{L}_\tau} \frac{1}{4u_{\tau,l}} \left(s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1} \right)^2 + p^0_\tau \lambda^t_{\tau,1-\tau} \leq -\gamma. \quad (3.16)$$

To derive a second order cone representation, we introduce the additional variables $y^t_{\tau,l},\ l \in \mathcal{L}_\tau,\ t \in \mathcal{T}^+$ such that

$$\frac{1}{4u_{\tau,l}}(s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1})^2 \le y^t_{\tau,l},\ l \in \mathcal{L}_\tau.$$

Thereby, system (3.16) is equivalent to

$$\begin{aligned}
&\sum_{l \in \mathcal{L}_\tau} y^t_{\tau,l} + p^0_\tau \lambda^t_{\tau,1-\tau} \le -\gamma, \\
&4u_{\tau,l} y^t_{\tau,l} \ge \left(s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1}\right)^2,\ \forall\, l \in \mathcal{L}_\tau, \\
&y^t_{\tau,l} \ge 0, \forall\, l \in \mathcal{L}_\tau,
\end{aligned} \tag{3.17}$$

which is a second order cone representation as desired.

For case (b), the analysis is similar to case (a), but becomes much simpler, as the underlying problem reduces to a linear programming in this case. Noticing that $\pi^t_1(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau)$ is lower bounded by a constant γ, we then have $s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1} = 0$ and $p^*_{\tau,l} = 0$. Thereby, system (3.17) is valid as well.

$(ii) - (vi)$. The arguments for these cases are similar to case (i). For brevity, here we omit the details. This completes the proof. □

Using Theorem 4 and Proposition 6, we are ready to derive the following result concerning the tractability of robust optimization model (3.2), which is a main result of this paper.

Theorem 5. *Robust optimization model (3.2) is equivalent to the following*

SOCP

$$\inf_{\substack{\rho,\boldsymbol{\eta},(\boldsymbol{s}_\tau,\boldsymbol{u}_\tau,\boldsymbol{v}_\tau,\boldsymbol{w}_\tau)_{\tau\in\mathcal{T}},\\ (\boldsymbol{\lambda}^p_\tau,\boldsymbol{\lambda}^a_\tau,\boldsymbol{y}^p_\tau,\boldsymbol{y}^a_\tau)_{\tau\in\mathcal{T}}}} \left\{\rho + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}_{\tau,l}s_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}^2_{\tau,l}(1+\mu^2)u_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}_{\tau,l}v_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}^2_{\tau,l}(1+\mu^2)w_{\tau,l}\right\}$$

$$\begin{aligned}
\text{s. t.}\quad & \sum_{(\tau,l)\in\mathcal{I}} y^{t,p}_{\tau,l} + \sum_{\tau\in\mathcal{T}^{--}} p^0_\tau\lambda^{t,p}_{\tau,1-\tau} + \sum_{\tau\in\mathcal{T}^{+}} p^0_\tau\lambda^{t,p}_{\tau,1} + \sum_{(\tau,l)\in\mathcal{I}} y^{t,a}_{\tau,l} \\
& + \sum_{\tau\in\mathcal{T}^{+}} \alpha^0_\tau\lambda^{t,a}_{\tau,1} + \sum_{\tau\in\mathcal{T}^{--}} \alpha^0_\tau\lambda^{t,a}_{\tau,1-\tau} \le \rho + c_t,\ \forall\, t\in\mathcal{T}^+, \qquad (3.18)\\
& 4u_{\tau,l}y^{t,p}_{\tau,l} \ge \left(s_{\tau,l} - z^t_{\tau,l} + \lambda^{t,p}_{\tau,l} - \lambda^{t,p}_{\tau,l+1}\right)^2,\ \forall\, t\in\mathcal{T}^+,(\tau,l)\in\mathcal{I},\\
& 4w_{\tau,l}y^{t,a}_{\tau,l} \ge \left(v_{\tau,l} - \eta_\tau z^t_{\tau,l} + \lambda^{t,a}_{\tau,l} - \lambda^{t,a}_{\tau,l+1}\right)^2,\ \forall\, t\in\mathcal{T}^+,(\tau,l)\in\mathcal{I},\\
& \lambda^{t,p}_{\tau,l},\ \lambda^{t,a}_{\tau,l} \ge 0,\ \forall\, t\in\mathcal{T}^+,\tau\in\mathcal{T}, l\in\mathcal{L}^+_\tau,\\
& y^{t,p}_{\tau,l},\ y^{t,a}_{\tau,l} \ge 0,\ \forall\, t\in\mathcal{T}^+,(\tau,l)\in\mathcal{I},\\
& u_{\tau,l},\ w_{\tau,l} \ge 0,\ \forall\,(\tau,l)\in\mathcal{I},\\
& \boldsymbol{\eta}\in X.
\end{aligned}$$

Proof. First, we rewrite the system of inequality constraints of Problem (3.5) as

$$\sum_{\tau\in\mathcal{T}} \pi^t_1(\boldsymbol{s}_\tau - \boldsymbol{z}^t_\tau, \boldsymbol{u}_\tau) + \sum_{\tau\in\mathcal{T}} \pi^t_2(\boldsymbol{v}_\tau - \eta_\tau\boldsymbol{z}^t_\tau, \boldsymbol{w}_\tau) \ge -\rho - c_t,\ \forall\, t\in\mathcal{T}^+ \qquad (3.19)$$

Then according to Proposition 6 and applying some necessary operations, for each $t\in\mathcal{T}^+$, there exist some Lagrange multipliers $\lambda^{t,p}_{\tau,l}$ and $\lambda^{t,a}_{\tau,l}$, $l\in\mathcal{L}^+_\tau$, $\tau\in\mathcal{T}$, such that (3.19) is equivalent to

$$
\begin{aligned}
&\sum_{\tau\in\mathcal{T}^+}\left(\sum_{l\in\mathcal{L}_\tau} y^{t,p}_{\tau,l} + p^0_\tau\lambda^{t,p}_{\tau,1}\right) + \sum_{\tau\in\mathcal{T}^{--}}\left(\sum_{l\in\mathcal{L}_\tau} y^{t,p}_{\tau,l} + p^0_\tau\lambda^{t,p}_{\tau,1-\tau}\right) + \sum_{\tau\in\mathcal{T}^+}\left(\sum_{l\in\mathcal{L}_\tau} y^{t,a}_{\tau,l} + \alpha^0_\tau\lambda^{t,a}_{\tau,1}\right)\\
&+\sum_{\tau\in\mathcal{T}^{--}}\left(\sum_{l\in\mathcal{L}_\tau} y^{t,a}_{\tau,l} + \alpha^0_\tau\lambda^{t,a}_{\tau,1-\tau}\right) \le \rho + c_t,\ \forall\, t\in\mathcal{T}^+,\\
&4u_{\tau,l}y^{t,p}_{\tau,l} \ge \left(s_{\tau,l} - z^t_{\tau,l} + \lambda^{t,p}_{\tau,l} - \lambda^{t,p}_{\tau,l+1}\right)^2,\ \forall\, t\in\mathcal{T}^+, (\tau,l)\in\mathcal{I},\\
&4w_{\tau,l}y^{t,a}_{\tau,l} \ge \left(v_{\tau,l} - \eta_\tau z^t_{\tau,l} + \lambda^{t,a}_{\tau,l} - \lambda^{t,a}_{\tau,l+1}\right)^2,\ \forall\, t\in\mathcal{T}^+, (\tau,l)\in\mathcal{I},\\
&y^{t,p}_{\tau,l},\ y^{t,a}_{\tau,l} \ge 0,\ \ \forall\, t\in\mathcal{T}^+, (\tau,l)\in\mathcal{I}.
\end{aligned}
$$

On the other hand, according to Theorem 4, Model (3.2) is actually a “min-min” two-stage problem. Thus, Model (3.2) is equivalent to the following problem:

$$
\begin{aligned}
\inf_{\substack{\rho, \boldsymbol{\eta}, (\boldsymbol{s}_\tau, \boldsymbol{u}_\tau, \boldsymbol{v}_\tau, \boldsymbol{w}_\tau)_{\tau\in\mathcal{T}},\\ (\boldsymbol{\lambda}^p_\tau, \boldsymbol{\lambda}^a_\tau, \boldsymbol{y}^p_\tau, \boldsymbol{y}^a_\tau)_{\tau\in\mathcal{T}}}} \quad & \Bigg\{ \rho + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}_{\tau,l} s_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{p}^2_{\tau,l}(1+\mu^2) u_{\tau,l} \\
& + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}_{\tau,l} v_{\tau,l} + \sum_{(\tau,l)\in\mathcal{I}} \bar{\alpha}^2_{\tau,l}(1+\mu^2) w_{\tau,l} \Bigg\} \\
\text{s. t.} \quad & \sum_{\tau\in\mathcal{T}^+} \left(\sum_{l\in\mathcal{L}_\tau} y^{t,p}_{\tau,l} + p^0_\tau \lambda^{t,p}_{\tau,1} \right) + \sum_{\tau\in\mathcal{T}^{--}} \left(\sum_{l\in\mathcal{L}_\tau} y^{t,p}_{\tau,l} + p^0_\tau \lambda^{t,p}_{\tau,1-\tau} \right) \\
& + \sum_{\tau\in\mathcal{T}^+} \left(\sum_{l\in\mathcal{L}_\tau} y^{t,a}_{\tau,l} + \alpha^0_\tau \lambda^{t,a}_{\tau,1} \right) + \sum_{\tau\in\mathcal{T}^{--}} \left(\sum_{l\in\mathcal{L}_\tau} y^{t,a}_{\tau,l} + \alpha^0_\tau \lambda^{t,a}_{\tau,1-\tau} \right) \\
& \le \rho + c_t, \ \forall\, t \in \mathcal{T}^+, \qquad (3.20) \\
& 4u_{\tau,l} y^{t,p}_{\tau,l} \ge \left(s_{\tau,l} - z^t_{\tau,l} + \lambda^{t,p}_{\tau,l} - \lambda^{t,p}_{\tau,l+1} \right)^2, \ \forall\, t \in \mathcal{T}^+, (\tau,l) \in \mathcal{I}, \\
& 4w_{\tau,l} y^{t,a}_{\tau,l} \ge \left(v_{\tau,l} - \eta_\tau z^t_{\tau,l} + \lambda^{t,a}_{\tau,l} - \lambda^{t,a}_{\tau,l+1} \right)^2, \ \forall\, t \in \mathcal{T}^+, (\tau,l) \in \mathcal{I}, \\
& \lambda^{t,p}_{\tau,l},\ \lambda^{t,a}_{\tau,l} \ge 0, \forall\, t \in \mathcal{T}^+, \tau \in \mathcal{T}, l \in \mathcal{L}^+_\tau, \\
& y^{t,p}_{\tau,l},\ y^{t,a}_{\tau,l} \ge 0,\ \forall\, t \in \mathcal{T}^+, (\tau,l) \in \mathcal{I}, \\
& u_{\tau,l},\ w_{\tau,l} \ge 0,\ \forall\, (\tau,l) \in \mathcal{I}, \\
& \boldsymbol{\eta} \in X.
\end{aligned}
$$

This completes the proof. □

Our ability to solve the model and deploy the solution in practice critically depends on the model's computational tractability and efficiency. According to Theorem 5, we obtain a SOCP reformulation of Model (3.2). If the feasible set X is integral, then the problem becomes a SOCP problem with integrality constraints, which can be solved by state-of-the-art commercial solvers such as CPLEX. Assumption 3 allows us to obtain a tractable formation of the problem. However, if we have more elaborate models, such as the standard deviation of $\tilde{a}_{t,l}$ being a function of $\sqrt{\eta_t}$, then it would lead to a nonlinear, nonconvex optimization problem, which we do not know how to solve to optimality.

3.3 Empirical studies

In this section, we study the performance of our robust optimization models using real data from a public hospital in Singapore. Our data set consists of daily admission and length of stay of both emergency and elective patients throughout the year of 2008. For data sensitivity considerations, we scaled the original data in a proportionate manner, and all following discussions are based on the adjusted data. Emergency patients, averaging about 119 daily, account for about 82% of daily admissions. Their mean length of stay at 3.57 days exceeds that of elective patients by about 1 day.

Figure 3.2 shows the autocorrelation plot of daily emergency admissions across the year of 2008. Our investigation of seasonality in daily emergency admissions reveals volatility across the days rather than across the months. The patterns of elective admissions more or less mirror those appearing in the graphics below for emergency admissions.

Figure 3.3 shows the average daily emergency admission pattern within a week. There is an obvious weekly pattern. On average, we observe less emergency admissions during the weekend. Within a week, we see the greatest number of emergency admissions on Monday.

3.3.1 Numerical results

We have 366 days of retrospective data to evaluate and compare the performance of various models. In fact, this is very difficult for us to perform a convincing study based on those limited data points as we have to use the same data for both learning (e.g. empirical mean, coefficient of variations) and calculation. Therefore we suggest a way to impute our current data in order to provide a longer periods with more days counted. We apply the re-sampling approach

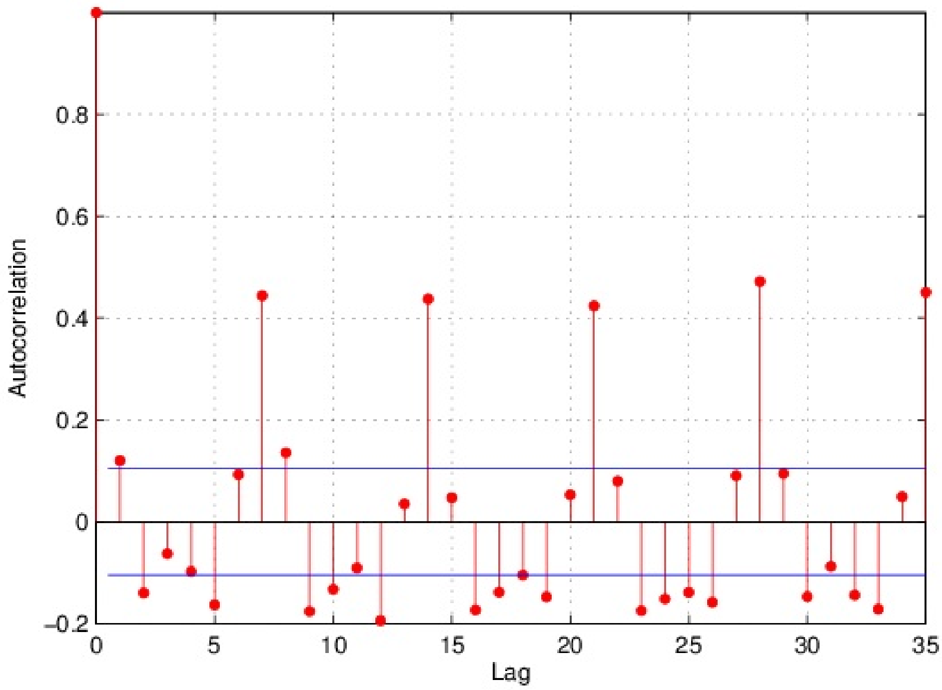

Figure 3.2: Autocorrelation of Daily Emergency Admissions. -

which only makes use of the current 366 data points we have. For example, from day 0, suppose it is Sunday, we will sample only from all those historical Sunday data points randomly with equal probability; and for day 1, suppose it is Monday, we will sample only from all those historical Monday data points randomly with equal probability and so on. In this way, we attain a sample data including 5000 days emergency and elective inpatients arrivals. Moreover, we retained the weekly arrival pattern of inpatients without knowing the exact distribution information. And this method will work best when the arrivals of inpatients are stationary which in reality may not be the case.

The numerical study commences on day T_0. From the initial part of the data, i.e. from day 1 through day $T_0 - 1$, we can establish the number of emergency and elective patients that have been warded in the hospital and their durations of stay. Given the weekly periodicity of the data, we obtain the empirical averages of the parameters $(\tilde{p}_{t,l}, \tilde{\alpha}_{t,l})$ based on the day of the week t falls into.

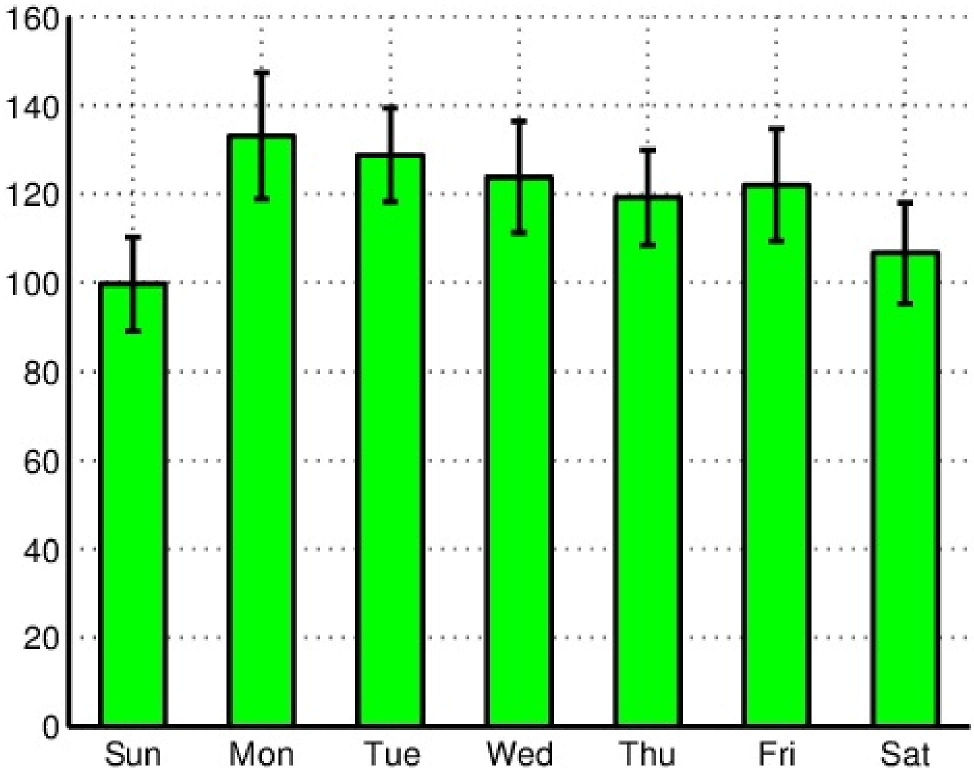

Figure 3.3: Average Daily Emergency Admissions by Weekday. - (Error Bars Indicate Standard Deviations)

We adopt a rolling horizon approach in our simulation study. Specifically, we solve the elective admission problem repeatedly every seven days over a planning horizon of T days until the end of the data is reached. In each problem we solve, we impose a daily quota within the range $[5, 80]$. Moreover, the total quota for the first seven days and the next seven days are set at 200. Hence, the feasible region X for the quota is as follows:

$$X = \left\{ \boldsymbol{\eta} \in \mathbb{Z}^{14} \ : \ 5 \le \eta_\tau \le 80, \tau \in \{0, \ldots, 13\}; \sum_{\tau=0}^{6} \eta_\tau = \sum_{\tau=7}^{13} \eta_\tau = 200 \right\}.$$

After obtaining the optimal elective admission quota $\boldsymbol{\eta}$, we simulate the patient admission process for the following seven days to evaluate the number of bed shortfalls. We use the actual emergency admission and length of stay as reflected in the data since these values are presumably independent of the quota. However, we could not directly use the elective data, since the elective admissions would be dependent on the quota imposed by our model. In the simulation study, we impute these values from the data in the following way. Given the actual quota $\hat{\eta}_t$ and $\hat{a}_{t,l}$, the actual number of elective inpatients admitted on the tth day and have stayed for at least l days, we impute the corresponding elective admission values as $a_{t,l} = \lfloor \hat{a}_{t,l} \eta_t / \hat{\eta}_t + 0.5 \rfloor$.

In our numerical study, we compare the solutions of the deterministic model (3.1), the robust model (3.2) with different budget of variations,

$$\mu \in \{0.01, 0.02, 0.05, 0.1\}$$

and the optimized budget of variation model (3.3). Note that the deterministic model (3.1) corresponds to the fixed budget robust model (3.2) with $\mu = 0$.

Under the above settings, we can obtain the solutions of our models within reasonable time. Solving the integral relaxation of the problem takes about 2 to 3 seconds on a 12-core 2.4GHz Mac Pro computer using the CPLEX solver.

For the mixed integer model, it requires about 10 to 20 seconds to obtain the optimal solution. We observe that the optimal integer solutions are close to the solution of the relaxed problem for which the integrality constraints are ignored.

We present the results under different configurations which differ in terms of length of planning horizon T, maximum duration of emergency inpatients, maximum duration of elective inpatients, hospital bed capacity, the starting period T_0 and rolling horizon days (counting periods). In Table 3.1, we list all configuration settings for our simulation study.

Configurations	1	2	3	4	5	6	7	8
Plan_Horizon	7	7	7	7	7	7	14	14
Max_Duriation_Elective	7	7	7	7	7	7	7	14
Max_Duriation_Emergent	7	14	14	14	14	14	14	14
Bed_Capacity	550	600	600	600	620	620	620	620
Start_Period	500	1000	1500	2000	2000	1000	1000	2000
Rolling_Horizon (days)	500	1000	1000	1000	1000	2000	2000	1000

Table 3.1: Configuration settings for simulation study

In Table 3.2, we report the total bed shortages of the different models under different configurations respectively. Apparently, our optimized models have a significant performance improvements over other suggest models.

Configuration \ Models	Deterministic	**Optimized**	$\mu = 0.01$	$\mu = 0.02$	$\mu = 0.05$	$\mu = 0.1$
1	49	43	48	69	102	173
2	1417	1124	1220	1301	1529	2456
3	1387	1074	1093	1089	1430	2283
4	2013	1747	1771	1763	2309	4095
5	368	289	307	297	445	1240
6	594	423	388	371	480	943
7	774	432	439	410	784	1100
8	910	685	723	695	789	1306

Table 3.2: Total bed shortages of the different models under given configurations

In Table 3.3, we report the maximum shortages on a daily base for the different models under different configurations. For most times, our optimized models yields a better results with smaller value of maximum bed shortage.

Configuration \ Models	Deterministic	**Optimized**	$\mu = 0.01$	$\mu = 0.02$	$\mu = 0.05$	$\mu = 0.1$
1	19	24	24	29	29	38
2	58	46	45	51	51	65
3	56	44	51	48	51	65
4	56	50	52	53	66	80
5	36	38	32	36	49	61
6	37	44	39	35	44	58
7	48	31	39	35	40	43
8	51	41	40	43	48	53

Table 3.3: Maximum bed shortages (daily based) of the different models under given configurations

In Table 3.4, we report the total number of days suffering bed shortage for the different models under different configurations. For most times, our optimized models yields a better results with smaller value of the total number of days suffering bed shortage.

Configuration \ Models	Deterministic	**Optimized**	$\mu = 0.01$	$\mu = 0.02$	$\mu = 0.05$	$\mu = 0.1$
1	5	4	5	5	6	11
2	89	82	92	94	100	137
3	88	81	80	81	98	133
4	145	146	147	144	161	204
5	34	31	32	34	42	87
6	56	35	32	32	40	81
7	66	37	39	35	74	96
8	78	62	66	58	72	89

Table 3.4: Total number of days suffering bed shortage of the different models under given configurations

In our computational study, we note that as we increase the budget of variation, μ the performance level of robust model (3.2) initially improves, but then deteriorates as μ increases further. Hence, this underscores the importance of adjusting the conservativeness of the ambiguity set through the parameter μ. Besides, we observe that the optimal level of performance is achieved under different μ when the model uses different starting dates, which suggests the advantage of our optimized budget of variation model, since it would be difficult to determine the parameter μ prior to the simulation. In addition, we note that the optimized budget of variation model (3.3) consistently performs better than other approaches, which suggests the superiority of this approach.

Though it is not reported in this study, we have also experimented with several other distributionally robust optimization models, such as incorporating variance estimates, confidence intervals of mean estimates. From the simulation study, the current model that we introduced provides consistently good performance without the need to have parameters' estimation beyond their first moments.

3.4 Conclusions

In this study, we present a new robust approach to manage elective admissions in hospital. Our model contributes to the methodology of robust optimization. In formulating our optimization model, instead of using the worst-case performance as the objective, we propose to maximize the level of uncertainty such that the worst-case performance meets a pre-specified target. In our problem, this method proves to provide fairly good performance without tinkering with the model parameters. We show how to solve our model efficiently and perform empirical studies based on real data. The numerical results suggest that the optimized budget of variation model (3.3) consistently generates better performance vis-à-vis the other two approaches.

This research is done as part of a project with a public hospital in Singapore. Currently, the hospital determines the quotas for different days of the week, which they obtain by solving a deterministic linear optimization problem. The approach is not dynamic and does not take into account uncertainty. Using the data provided by the hospital, we are able to show significant improvement in mitigating the bed shortfalls. The ultimate goal is to integrate our model in the decision support system, which would require us to work closely with the IT vendors so that we can obtain live updates from the system. Inspired by this work, we could apply our developed software for distributionally robust optimization written in C++, so that we can easily deploy solutions that is easily maintainable than the current approach of reformulating the robust counterpart.

4

Patient Flow Scheduling Study in Emergency Department with Targeted Deadlines

The Emergency Departments (EDs) serve as an important part of healthcare, through which 50% of non-obstetrical admissions occur (Pitts et al. (2008)). However, we note that a considerable percentage of patients experience long waiting or delay due to frequent congestions in the ED. Most hospital EDs operate near full capacity. Optimizing ED operations may have a significant effect on the overall healthcare quality and cost (Geer and Smith (2004)). In this work, we study how doctors response for system load, and examine what could be changed to improve or optimize the decision process based on historical data and hospital key performance indicators (KPI). The process used in Emergency Departments is highly complex and involves different parties, spanning the spectrum of doctors, nurses and tests. Despite the analytical challenges of ED models, we resort to simulation to appropriately address those challenges.

Most hospital EDs have characterized emergency patients into several categories with ranked priorities (Patient Acuity Category scale, or PAC scale). For example, very sick and unstable patients (PAC 1&2) are usually treated with highest priority with specially assigned resources, including radiology, lab, operating rooms, doctors and nurses. For this type of patients, the criterion are strict and there is almost no tolerance for delay (5˜10 minutes) which leaves us little space for patient flow control. Our focus here is PAC 3&4 patients (most are walk-in patients) whose symptoms are mild to moderate and there is no present threat to their lives. Those patients are treated in designated rooms and their treatments do not interfere with PAC 1&2 patients. Therefore, this context gives us more flexibility to operationally improve the process. In fact, our study of historical data shows that doctors themselves (especially those experienced ones) are behaving very differently over varying situations.

We utilize data from a Singapore public hospital covering over $200,000$ emergent patient cases with detailed information on each patient's diagnosis steps. Working with electronic data actually helps us accurately reconstruct the original operating process of ED and doctor's decision contexts. The common steps for PAC 3&4 patients in EDs are described in the Figure 4.1. Registration is directly followed by the triage process, then after triage patients are ready for doctors' consultation. During this process, a large proportion of patients (over 60%) would return to doctors several times before ultimately discharged or admitted (ED case end). Tests (POCT, Radiology, Meditation, Lab tests and Procedure) may be ordered during this period. Patients in EDs typically exhibit high uncertainty in the volume, diagnosis types and service time lengths. Their care delivery is most likely to affected by not only disease type (endogenous factor) but also hospital factors, including doctors, congestion severity caused by peer patients (exogenous factors). Take into account these recorded data, we could evaluate the doctors' behavioral response which may influence the patient selection decisions. Although those data sets are from Singapore hospitals, their ED deployment, including various operating units (radiology, lab, surgery rooms) and systematic software platforms (e.g. SAP) are common practice for most EDs worlds wide. Therefore, our proposed methodology and explored phenomena may also be applicable in other EDs beyond Singapore.

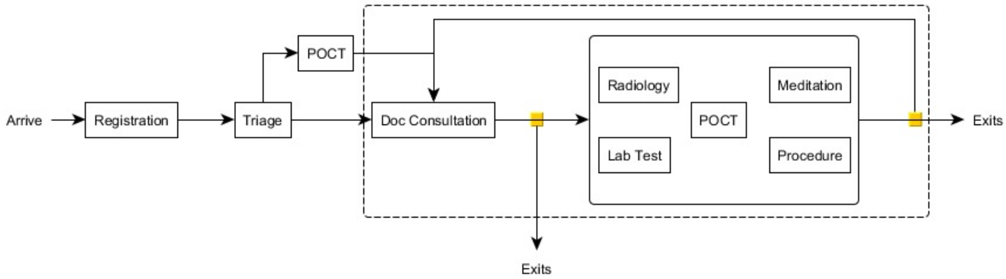

Figure 4.1: Emergency Department (ED) patient flow process. -

Hospitals often face the tradeoff among the following factors:

1. Length of Stay (LoS): the time when a patient arrives in the ED to the time s/he departs the ED;
2. First Wait (FW): the waiting time of a patient from registration till his or her first consultation;
3. Re-attending Rate: revisit rate resulting from an adverse event that occurred during the initial visit or from inappropriate care;
4. Left without being seen.

The first two factors are our major concerns here: LoS and FW. The re-attending rate is mostly caused by clinical results which may indirectly result in system congestion. As for left without being seen, we may not be able to study effectively due to lack of relevant data. Ideally the shorter the LoS and FW the better we desire. Here we use the similar requirements designed by the Singpore health care system since our objective is to maximize the percentage of patients whose FW and LoS are within targeted deadlines while maintain a reasonable limit for those patients who could not meet the targets. In this work, we propose a flow control algorithm based on dynamically solving a target based optimization model. In our model, we introduce a doctor's effort level (α), which deals with the uncertain service time. If the doctor puts more effort or work faster (high α), the service time will be reduced as a result. Another criteria is to minimize doctors' effort level which is subject to the condition that all patients should meet their deadline constraints. This selection process may incur the "fairness" (e.g. served by first come first serve) concern to some extent, but it is comparable to the current process where "unfairness" exists too.

For numerical testing, we build our simulation framework which could capture the essentials of the ED system being modeled. Our testing framework includes almost all classes of objects, e.g. patients, doctors, different functional nurses or operators (radiology) and other resources. The process framework upon which the modeling tool is developed is similar to Figure 4.1. Doctors could either select patients from the triage pool (new patient pool) or his or her own consultation pool (there is only one doctor in charge for each patient, so patients should only return to the same doctor). For radiology, lab, procedure, POCT and medication operations, those are task oriented, i.e., once the doctor or nurses order a new test task, this task will be put into its relevant task pool and operators would pick a task from their designated task pool only.

Most importantly, we compare the performance of our approach via simulation with three commonly used policies: First Come First Serve (FCFS), Shortest Deadline First (SDF) and Huang et al. (2014)'s heuristic policy (HeuristicPolicy). Due to ethical reasons, we may not be able to run a field experiment for different flow control policies. In the numerical setting, we apply the conventional heavy-traffic where the system converges to the critical load and consider only one doctor so far for simplicity. Deriving the optimal policy in such a complex system with uncertainty is both analytically and numerically challenging. Our policy may not achieve global optimality in the ED dynamics. We have determined that our proposed optimization algorithm could easily outperform the other three policies on given measures, whereas a small proportion of patients can be worse off but are of limited effects.

Our work contributes to the analytic work in healthcare operations research that studies ED patient flow control. Our approach is very different from the traditional queueing method, but based on the optimization model which utilizes historical information to deal with uncertainty.

In summary, we make the following key contributions:

- Data analysis of doctors' response for system load: Applying a large patient-level data set of over $200,000$ ED cases, we study the effects of different factors which may affect doctors' behaviors, including service acceleration, multi-tasking;

- Proposed a patient flow scheduling policy: In order to maximize the percentage of patients who could be served within targeted deadlines, we develop a robust optimization model and estimate its performance in simulation. The performance results could easily outperform other commonly recommended policies.

The rest of this chapter is structured as follows. Section 4.1 describes the system context and data sets we collected in the study. Section 4.2 is the data analysis of doctors' behavior on patient selection where we study how doctors adjust their service rate according to system load. Section 4.3 develops the target based optimization model for patient selection process and provides its analytical results. Section 4.4 is the simulation study of different policies and summarizes our work.

4.1 Clinical Setting and Data

We collected a large dataset from a major Singapore public hospital, comprising of nearly 200,000 emergency department visits over the course of two years. The average daily arrivals are around 400 with strong periodicity on different day of the week, e.g., Monday arrivals are highest, while Thursday arrivals are lowest. In our study, PAC 3&4 ("walk-ins") patients are served in separate area with dedicated resources (doctors, nurses, equipments, laboratories) as compared to PAC 1&2 (e.g. ambulance arrivals) patients. We solely focus on PAC 3&4 visits which follows a more standardized process of registration, triage, treatment (consultation and tests) while PAC 1&2 patients are required to be treated immediately. The majority of ED arrivals ($\geq 70\%$) belong to PAC 3&4 and our data sets contain $167,000$ such cases.

Most hospital EDs (Singapore, US, Europe) operate in a similar manner. Upon arrival, patients first go for registration and an electronic patient record is initiated with the current time stamp and patient basic information (gender, age, name, contact, complaint and etc.). Thereafter, triage nurses will see the patient and assesses his or her condition, measures vital signs, and record the chief complaint. Triage nurses could also order several POCT (Point-of-care testing, e.g., ECG, heart rate) test for the patient, but radiology, lab, and medication orders by nurses are not permitted in our studied hospital. The beginning and ending timestamp of triage will be marked as well as chief complaint. After triage, all patients wait in a shared room for doctors' consultation or treatment. Patients will be called for service by doctors when doctors are available.

When a new patient is firstly assigned to a doctor, the doctor will mark the start time of this consultation and this patient's in charge doctor, i.e., the patient needs to see the same doctor during his or her ED sojourn before the case ends (in rare cases patient change the doctor halfway). After doctor meets and examines the patient, the doctor usually generates a mental list of possible diagnoses and decides a treatment process. Most of the times, doctor orders several diagnostic tests or treatments, such as lab tests, medications at the end of current consultation. At this point, the patient will temporarily leave the doctor's room and wait outside for test, treatment calls or both. All tests and treatments including POCT, lab, procedure, medication and radiology are recorded electronically in the patient database. For each test and treatment order, an

Case no	Task Category	...	Status	Order Time	Ack Time	Finish Time	Close Time	...

Table 4.1: Electronic task record data fields

electronic record is stored in the system tracking system, for example, Table 4.1 shows the format of task record. Lab specimens are sent to the hospital's central lab by pneumatic tube for processing. Other tests and treatments are performed locally by ED nurses. For lab, POCT, and radiology tests which can only be closed by doctors, patients are required to return to the same doctor for assessments before the tests are closed out. Therefore, the patient may go back to his or her doctor several times before case ends. In the last consultation, the doctor may decide that whether the patient needs to be admitted or referred otherwise the patient will be discharged directly. If the patient needs to be admitted or referred, the doctor will send the "admission" or "referral" request to the responsible departments. At the end of that point, all consulations for this patient are over, but the case may not end ("admission" or "referral" may take several hours to be done). Anyway, patients will no longer be in the doctor's consultation pool, but they are still waiting in ED. This waiting time on average takes 3˜4 hours, also called "boarding" period, and relies on other departments but beyond ED's control. In our study we will exclude this "boarding" period.

Requests for tests and treatments, i.e., x-rays, medication, are generally performed in FCFS order by the in charge nurses. There are 4 time stamps (TSs) for every operation: Start TS, Acknowledge TS, Finish TS and Close TS. For example, the doctor may order an x-ray test for the patient, the ordering time is stored as the Start TS, and when the operator in radiology room receives the order, he or she will generate the Acknowledge TS. Once this x-ray test is over, a Finish TS is generated. When doctor reviews the result of x-ray which is also the point indicating the Close TS.

4.1.1 Data Processing

The data set record patient-level characteristics (gender, age, race, nationality, etc.), and complete time stamps of the progress of each visit. Since PAC 1&2 patients are treated with dedicated rooms and providers, we consider only the PAC 3&4 patients (in fact hospital do not differentiate between PAC 3 and PAC 4 type patients, all are recorded as PAC 3). We utilize patient data from 167,000 patient (PAC 3&4) visits in 2011 and 2012 from the studied Singapore hospital. The capacity of this hospital is kept at a stable level without further expansion. Since all electronic records of tests and treatments are stored in one file while consultation, registration time stamps and patient level characteristics are stored in another file, we have to derive the individual patient profile (all tests and treatment service time, consultation time and process sequence) indirectly and the only connection keys are patient ID and doctors. Service time and consultation time could be calculated from tasks' starting and ending period and sequence could be confirmed via the sorted tasks. Therefore, we have the following derived patient profile as shown in Figure 4.2.

case_no	task_seq	task_type	service_time	preparation_time
4458A	1	triage	3	N.A
4458A	2	POCT	8	N.A
4458A	3	consultation	15	N.A
4458A	4	Lab-I	15	N.A
4458A	4	Lab-II	30	N.A
4458A	5	consultation	10	N.A
4458A	6	Medication	20	2
4458A	7	consultation_end	5	N.A

Figure 4.2: A sample patient profile. -

4.2 Data Analysis of Doctors' Response to System Load

Traditionally, we thought doctors would (not strictly) select patients in first-come-first-served (FCFS) manner according to their registration time or waiting time. However, waiting time may not be consistent with registration time since patients may go through several rounds of waiting if tests or treatments are ordered in the way. In our studied hospitals, doctors can have their own decision policy and they may not necessarily follow FCFS or other commonly known criterions. In this section we will study the doctors' decision process in detail. Based on our on site observations, doctors display diverse behaviors as responses to the system load.

4.2.1 System Load Vs. Service Acceleration

We are interested in how system load may affect the actual service rate of servers. In the classical queuing theory, service rate is not subject to the system status (Wolff (1989)). Nevertheless, in reality we could easily find many examples which show that this assumption is false, especially in our hospital ED setting. From the historical operation data, we find that there is a strong dependence between system state and service rate. Similarly, Batt and Terwiesch (2012) also shows the empirical findings on this connection. In fact, in our observations we find doctors may adjust their service rate according to the ED system load. When there is an increase in congestion, doctors may take numerous ways to speed up their service rate. In this part, we describe two primary mechanisms that doctors may take at work to adjust their speed.

Multitasking

We first focus on doctors' multitasking behavior. In our ED setting, after seeing the doctor, the doctor may send him or her for another test (out of the doctor's treatment room). Thus, during this period when the patient is out for testing, doctor may select another patient for consultation even though the former patient case has not ended yet. We may interpret this multitasking as doctors being simultaneously responsible for multiple patients, but individual consultations are not necessarily performed simultaneously. Therefore, when patients are out for prescribed tests such as radiology, doctors could make use of those "empty" periods to see other patients, which avoids a wastage of doctors' time. From the patients' point of view, this activity could shorten their waiting time especially for their first wait and accelerate the overall service rate. Literatures also demonstrate that multitasking may incur additional switching "costs" which hinders productivity (Pashler (1994)). However, most P3&P4 ED cases are commonly seen and the procedures are more or less standard and not complex. We could assume this switching "costs" to be negligilble. Another reason we study multitasking is that doctors are not only using this for service acceleration, but also for adjusting their decision strategy – they are not strictly following a FCFS policy.

Shortened consultation time length

The other mechanism we study here is doctors' individual consultation time lengths – a major measure for doctors' service rate. Based on the historical data, it is statistically sufficient to show that the consultation time length would be shortened when system status becomes congested. This "rushing" behavior is also found in other literatures (Kc and Terwiesch (2009), Schultz et al. (1998)) where they show servers simply work faster.

Other mechanisms which we do not list here may also be relevant to service rate adjustment, e.g., staffing for nurses, doctors and equipments.

4.2.2 Data Description & Analytical Results

The data set contains detailed information for each patient visit, such as patient demographics, chief complaint, attending doctor, and time stamps and in charge

Variable	**Mean Value**
Age	39.6
Female	56%
Service Time	3.8 hrs
First Wait	45 mins
No. of tests	3.2

Table 4.2: Summary Statistics of Patients

person of all major events. Table 4.2 provides descriptive statistics of the patient population. We are interested to know about how doctors select the next patient for consultation (either from new patients or from returning patients) and system load's influence on doctors' service rate. The census measure we choose to indicate ED system load is the number of patients in the ED who has registered but their cases not end yet. Other measures could be new patients census, ED in-service census or boarding census (Batt and Terwiesch (2012)). In our setting, we divide the study period into 1-hour intervals and take the average number of patients in the ED as our census variables. The rest of our analysis on system load is solely based on the overall number of patients in the ED including both new patients and in-service patients. The hospital also takes this as a key indicator for their crowd level in ED. Similarly, we use the 1-hour average number of patients in charge by the same doctor as a measure for doctor's multitasking.

For much of the analysis here, we focus on a single chief complaint to study the connection between the system load and service rate. Generally speaking, triage nurse would determine and record patients' chief complaint. According to the ICD code (000 - 999), there are 19 primary chief complaints. The most common chief complaint in ED is *upper respiratory tract infection*, which accounts for 12% of all P3&4 ED cases. In Figure 4.3, we plot the mean consultation time with respect to system status. We find that doctors would spend less time on individual consultation when ED become more congested. Equivalently, the service rate is higher when the system load is high.

Multitasking reveals doctors' another mechanism for service acceleration by

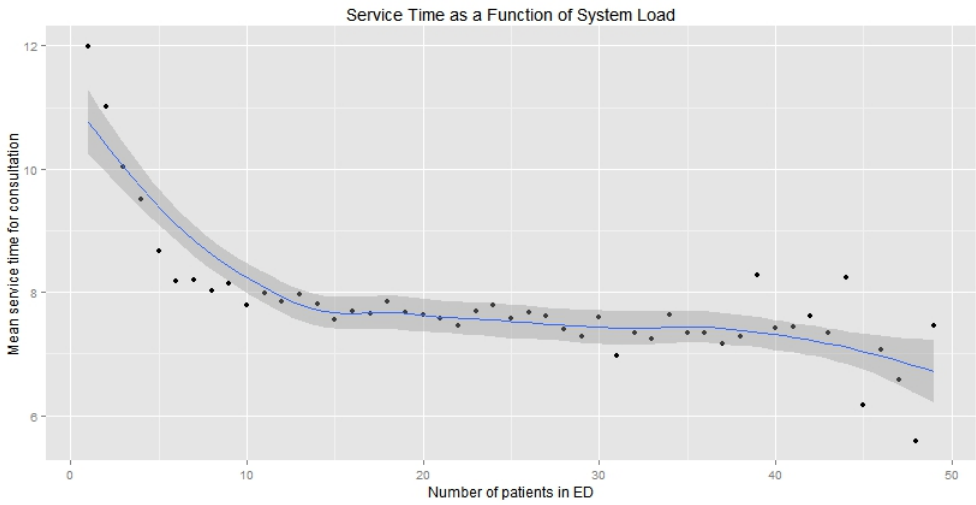

Figure 4.3: Individual consultation time length Vs. system status. -

making efficient use of their time. In fact, from our studied data we observe that doctors tend to take more patients simultaneously as shown in Figure 4.4.

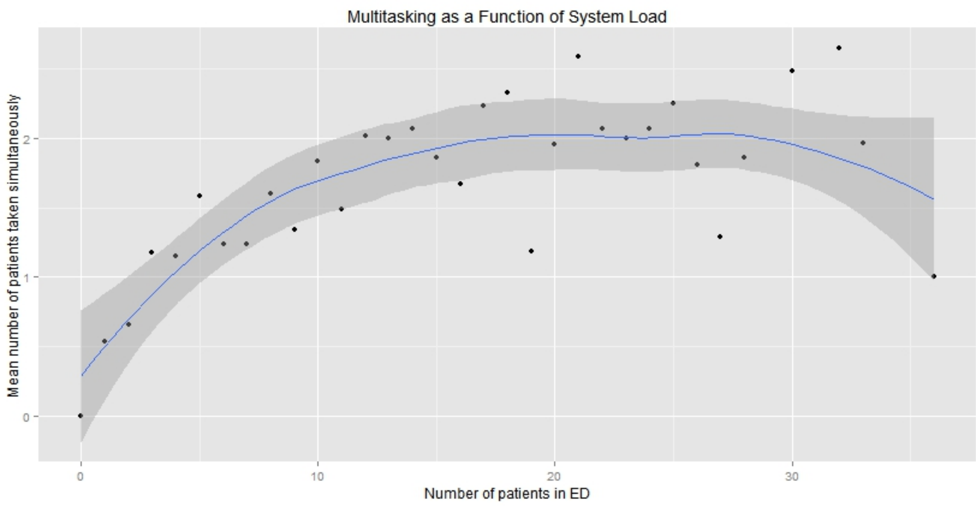

Figure 4.4: Doctor's multitasking Vs. system status. -

4.3 Optimizing Patient Flow Control

For simplicity of model presentation, this section we build the basic ED model with a single doctors (or server) whose service time (exclusively for consultation) is uncertain. Doctor based on system load or status may exhibit different service speeds even for the same type of patients. Patient types could be clustered into multiple groups, for example, certain type of patients may take longer consultation time by the same doctor under the same system load. We assume that patients' treatment processes (including consultations and tests) are independent of doctor, i.e., patients' treatment is a standard process. So we could cluster all patients into distinct clusters based on their disease types which we assume could be identified during the triage stage. In the previous section, we understand that doctors would adjust their service speed according to the system load. Here we introduce the concept of doctor's effort level, denoted by α where $\alpha \in [0, 1]$, and this effort level may decide how long the consultation may take for certain type of patient. Specifically, we allow $\alpha = 1$ (maximum effort) where service time goes to zero while $\alpha = 0$ (minimum effort) service time can be ∞.

4.3.1 Notations

Given parameters:

- I: number of patients, indexed by $i = 1, ..., I$;
- $\mathcal{I}$: set of all patients;
- $\mathcal{I}_0$: set of new patients who are ready for first consultation;
- L_0: number of patients need to be scheduled in new patient pool;
- $\mathcal{I}_1$: set of returning patients who are ready for consultation;
- L_1: number of patients need to be scheduled in doctor's processing pool;
- T_i: the target waiting time for patient i;
- $\bar{s}_i$: upper bound of patient i 's current service time;

- $\underline{s}_i$: lower bound of patient i 's current service time;

Uncertain Variables and Set:

- $\tilde{s}_i$: upcoming service time of patient i, where $\tilde{s}_i \in [\underline{s}_i, \bar{s}_i]$;

Decision Variables:

- x_{ik} : indicating whether patient i has been arranged in position k, where $x_{ik} \in \{0, 1\}$, $\boldsymbol{x}_i = (x_{i1}, ..., x_{i(L_0+L_1)})'$and $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_I)$;
- α: doctor's effort level, and $\alpha \in [0, 1]$, e.g., $\alpha = 1$ means maximum effort while $\alpha = 0$ means minimum effort.

The dynamics of our patient flow work as follows. Once a doctor is available, he or she may decide to pick a patient from either new patient pool $\mathcal{I}_0$ or returning patient pool $\mathcal{I}_1$. So every event is triggered by doctor's status change (e.g., consultation finish). However, the service time for each consultation is uncertain and we are making decisions in a centralized system. Doctor is making decisions out of both the new patient pool $\mathcal{I}_0$ and returning patient pool $\mathcal{I}_1$. Hence the uncertain service time may affect all waiting patients in the design. Meanwhile, the uncertain service time is a function we defined as $(1 - \alpha) -$ quantile, i.e.

Definition 4.3.1. Given doctor's effort level α , uncertain service time $\tilde{s}_i, \forall i \in \mathcal{I}$ are functions defined as follows

$$\tilde{s}_i(\alpha) \quad = \quad \inf \{s : \mathbb{P}(\tilde{s}_i \leq s) \geq 1 - \alpha\} \tag{4.1}$$

where $\mathbb{P} \in \mathbb{F}(\alpha)$ with

$$\mathbb{F}(\alpha) = \left\{ \mathbb{P} : \ \mathbb{P}\left(\ \underline{s}_i \leq \tilde{s}_i(\alpha) \leq \bar{s}_i \ \ \forall i \in \mathcal{I} \ \right) = 1 \ \right\}.$$

As time passes, doctors and patients statuses are updated accordingly. During the process, new patients may join and old patients may be discharged or sent for tests. Regarding the value of $\tilde{s}_i(\alpha)$, $\forall i \in \mathcal{I}$, we could obtain from historical data by deriving the statistical quantiles.

4.3.2 Model Setup

We are subjected to the constraints that the patients' waiting should be within the targeted deadline, but this is not always achievable if system experiences congestion. In most EDs the system load is close the their capacity upper bound, thus a possible key performance indicator (KPI) is the percentage of overall patients which could meet the deadline constraints. However, the disadvantage of this probability measure is its computational intractability which may deter its practical implementation. Instead of solving a complex stochastic problem, we come up with the objective to minimize the doctor's effort level in order to meet the target:

$$\alpha^* = \min_{\boldsymbol{X} \in \mathcal{X}(\alpha)} \alpha$$

where we let $\mathcal{X}(\alpha)$ be the feasible space of scheduling solutions.

If patient i is assigned to the doctor in position $(l+1)$, then this patient's waiting time could be calculated as

$$\sum\nolimits_{k=1}^{l} \sum\nolimits_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{ik} \tilde{s}_i$$

where i' is the index for current patient. In addition, patient i's waiting length should be no more than his or her threshold T_i. So patients' waiting constraints are organized as follows:

$$\begin{aligned} \sum\nolimits_{k=1}^{l} \sum\nolimits_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{ik} \tilde{s}_i \ &\leq \ \sum\nolimits_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{i(l+1)} T_i + (1 - \sum\nolimits_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{i(l+1)}) M, \\ & \qquad l = 1, .., (L_0 + L_1 - 1) \end{aligned} \tag{4.2}$$

where M is a given large number. The other constraints includs:

$$\begin{array}{ll} \sum_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{il} \geq \sum_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{i(l+1)}, & \forall l = 1, .., (L_0 + L_1 - 1) \\ \sum_{k=1}^{L_0+L_j} x_{ik} = 1, & \forall i \in \mathcal{I}_0 \cup \mathcal{I}_1 \\ \sum_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{ik} \leq 1, & \forall k = 1, ..., (L_0 + L_1) \end{array} . \tag{4.3}$$

The first constraint is the sequencing compliance. The second constraint requires every patient to be assigned to at least one doctor. The third constraint ensures each position to be filled with no more than one patient.

We note that this optimization problem α^* is monotonically increasing in $\tilde{s}_i$, $\forall i \in \mathcal{I}$. Since for any $\alpha \leq \alpha^\ddagger$, we have

$$\tilde{s}_i(\alpha) \geq \tilde{s}_i(\alpha^\ddagger) \ .$$

And set $\mathcal{X}(\tilde{\boldsymbol{s}})$ ($\tilde{\boldsymbol{s}} = (\tilde{s}_1, ..., \tilde{s}_I)'$) is defined as

$$\mathcal{X}(\tilde{\boldsymbol{s}}) = \left\{ \boldsymbol{X} \ \middle| \ \begin{array}{l} \boldsymbol{X} \text{ satisfy constraints in 4.2} \\ \boldsymbol{X} \text{ satisfy constraints in 4.3} \\ i \in \mathcal{I}, \forall k = 1, \ldots, (L_0 + L_j) \end{array} \right\}.$$

We could easily derive the following property and thus find the optimal α^* using binary search, since for any given α, we could get the value of $\tilde{s}_i = \tilde{s}_i(\alpha)$ defined in 4.1 from historical data.

Lemma 6. *For any given value sets $(\tilde{\boldsymbol{s}}(\alpha))$ and $(\tilde{\boldsymbol{s}}(\alpha^\ddagger))$ where $\alpha \leq \alpha^\ddagger$, we have*

$$\mathcal{X}(\tilde{\boldsymbol{s}}(\alpha)) \supseteq \mathcal{X}(\tilde{\boldsymbol{s}}(\alpha^\ddagger)).$$

4.4 Simulation Study

In our simulation study, we use a simplified ED setting where only two types of patients are considered. One patient type is with single consultation and after the consultation patient will be discharged. The other type is with two consultations as one radiology test is ordered in between. This setting highlights or emphasizes on returning patients as well as new patients. The two types of patient profile are listed in the Figure 4.5 and Figure 4.6. We evaluate the performance of our proposed policy and other commonly-used policies in terms of both first waiting and length of stay in the ED.

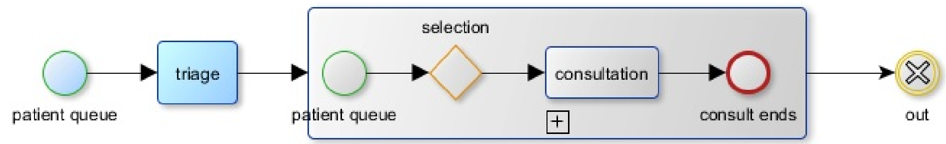

Figure 4.5: Tested patient profiles – single consultation. -

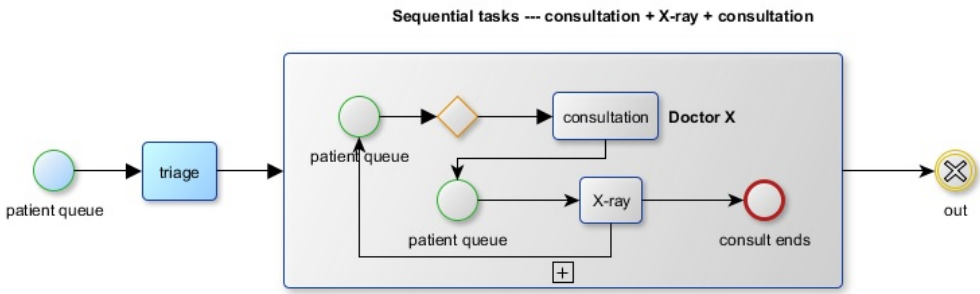

Figure 4.6: Tested patient profile – two consultations with one radiology test. -

4.4.1 Other policies

We run the simulation under four different policies which are listed here. And the detailed description on the system setting for all policies is as follows.

First Come First Servie (FCFS)

We prioritize patient according to their registration time: select patient with earlier registration time stamps among all "available" patients.

Shortest Deadline First (SDF)

We would select the patient whose waiting is closest to the deadline. We use the following way to define the deadline (w^{seq_n} is the waiting time for n^{th} consultation)

$$D^{seq\_1} = 30mins$$
$$D^{seq\_2} = (D - w^{seq\_1})mins$$

where D^{seq_n} is the waiting threshold for n^{th} consultation.

Heuristic Policy from Huang et al.(2014) **(HeuristicPolicy)**

For given $\epsilon \geq 0$, if all new patients satisfy

$$w_i^{seq\_1} \leq D^{seq\_1} - \epsilon, \forall i \in \mathcal{I}_0,$$

then we would select patients from available returning patients first. Otherwise, select the new patient with earliest registration time.

Optimized Target Based Policy(OPT)

The decision is made by solving the optimization problem in the previous section. And the threshold waiting limit is same as the waiting limit in shortest deadline first policy. However, we may not always have a feasible solution if congestion happens and some patient already waits longer than the threshold. In such case, we would suggest another optimization model for that decision making.

In the relaxed model we fixed the service time $s_i = \mathbb{E}_{\mathbb{P}}(\tilde{s}_i)$ (the mean value of $\tilde{s}_i$ from distribution $\mathbb{P}$ and in the simulation we use empirical average instead) and put this given value into the following problem.

$$\begin{aligned} \min \quad & \max_{i \in \mathcal{I}} (w_{ik}^{seq\_n} - D^{seq\_n}) \\ s.t. \quad & w_{i(l+1)}^{seq\_n} = \sum_{k=1}^{l} \sum_{i \in \mathcal{I}_0 \cup \mathcal{I}_1} x_{ik} s_i - (1 - x_{i(l+1)}) * M, \\ & \forall i \in \mathcal{I}_0 \cup \mathcal{I}_j, l = 1, .., (L_0 + L_j - 1) \end{aligned}$$

where $w_{ik}^{seq\_n}$ is the waiting time of patient i if assigned to doctor at position k during the nth consultation.

4.4.2 Input Settings

We get two years ED data from a public Singapore hospital and derive all patient cases's individual service time. Furthermore, based on the characteristics of patient cases and their specified ICD (International Statistical Classification of Diseases and Related Health Problems) code, we clustered those cases into two major groups with differentiated service time distributions. Empirical analysis shows that patients' service time distribution is close to the lognormal distribution. In our patient input files, we use the derived lognormal parameters of service time distribution to generate our patients service time.

We consider a single doctor case with two different patient profiles and two cluster types of patients. The cluster type of patients would determine quantile statistics of service time. Input including two different files: patients arrival and patients profile records, see example in Table 4.3 and 4.4.

Patients ID	**registration time**	**cluster type**
5711104333I	3:06:07 PM	type_1

Table 4.3: Example of input files on patients arrival

Patients ID	task sequence	task type	service time	consult sequence
5711104333I	1	Triage	1	N.A
5711104333I	2	consult	5	1
5711104333I	3	radiology	6	N.A
5711104333I	4	consult_end	10	2

Table 4.4: Example of input files for all patients profile

Other parameters are listed as follows

- Percentage of Type 1 and Type 2 patients : 25% Vs. 75%;
- Percentage of Patients with radiology test : 50%;
- Distribution of consultation time for each cluster type of patients (time unit is minute):

$$\text{Patient Cluster Type I} : \sim lognormal(1.2, 0.85)$$
$$\text{Patient Cluster Type II} : \sim lognormal(1.9, 0.86)$$

Patients Arrival

We assume patients' arrival is Possion process, so the time interval between two adjacent patients follows exponential distribution. Therefore, we set

$$TS_i = TS_{i-1} + exponential(\lambda)$$

where TS_i is the registration time stamp for ith patient (patient $i-1$ arrives earlier than patient i, $i \in \mathbb{Z}^+$) and λ is the parameter for exponential distribution.

4.4.3 Simulation Outcomes

We run the simulation and compared the four different policies listed in Section 5.1. Since we are concerned with the percentage of patients who could meet the targeted deadline in terms of both first waiting and overall length of stay, related terms are defined as follows.

- $D^{seq\text{-}1}$: waiting threshold for the first consultation;
- $w_i^{seq\text{-}1}$: waiting time of patient i for his or her first consultation, $i \in \mathcal{I}$;
- D^{LoS} : threshold of patients' overall length of stay in the ED;
- LoS_i : length of stay of patient i in the ED, $i \in \mathcal{I}$;

Our performance critiriors could be reprensented as:

- $\mathbb{P}(w_i^{seq\text{-}1} \leq D^{seq\text{-}1})$, probability that a new patient would see doctor within targeted deadline (e.g. 30 minutes);
- $\mathbb{E}_{\mathbb{P}}(w_i^{seq\text{-}1})$, the expected first waiting time of patients;
- $\mathbb{E}_{\mathbb{P}}(LoS_i)$, the expected length of stay of patients.

We test three configurations of input parameters with different thresholds and system load. In addition to the listed performance critirions, we plot the density graph and quantile statistics for both first waiting time and overall length of stay.

4.4.3.1 Configuration 1

Configuration parameters are listed below in Table 4.5.

Notation	Description	Value
λ	patient arrival rate which follows possion distribution	0.07
μ	service rate of doctor	0.08
λ/μ	service load	88%
ϵ	threshold part for HeuristicPolicy (minutes)	5
T	simulation duration, in terms of minutes	50000
N	number of patients see by the system	3002
$D^{seq\text{-}1}$	time limit for first waiting (minutes)	45
D	time limit for the accumulated waiting (minutes)	90

Table 4.5: Configuration 1's input parameters

Figure 4.7 shows the density plot for the overall length of stay in ED under configuration 1

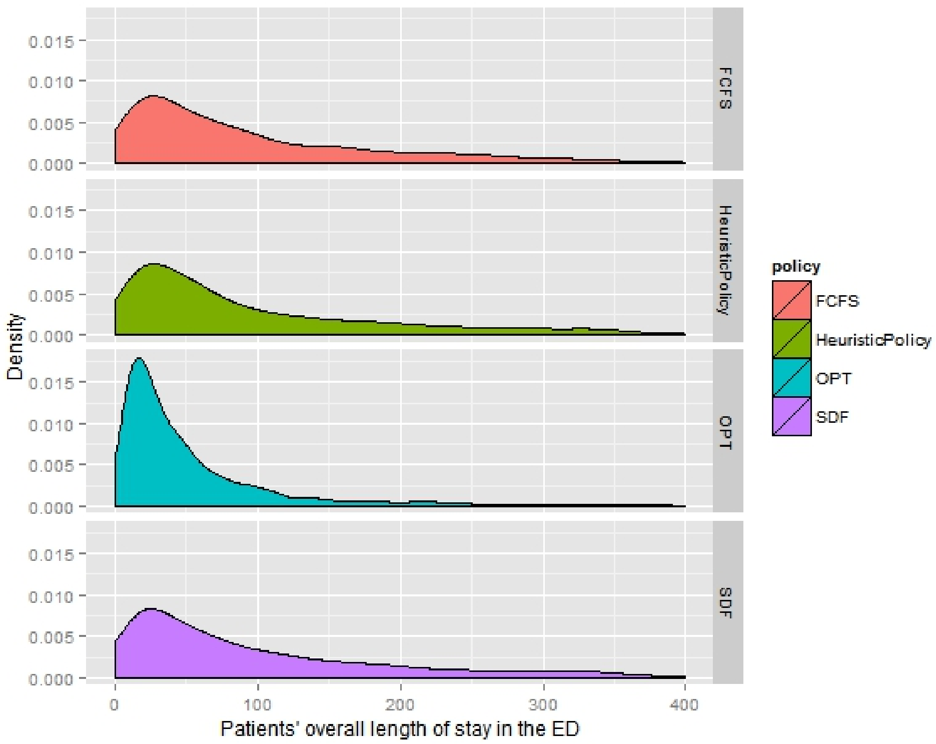

Figure 4.7: Density plot for patients' length of stay under different policies. - (configuration 1)

Figure 4.8 shows the density plot for the first waiting time of patients in ED under configuration 1

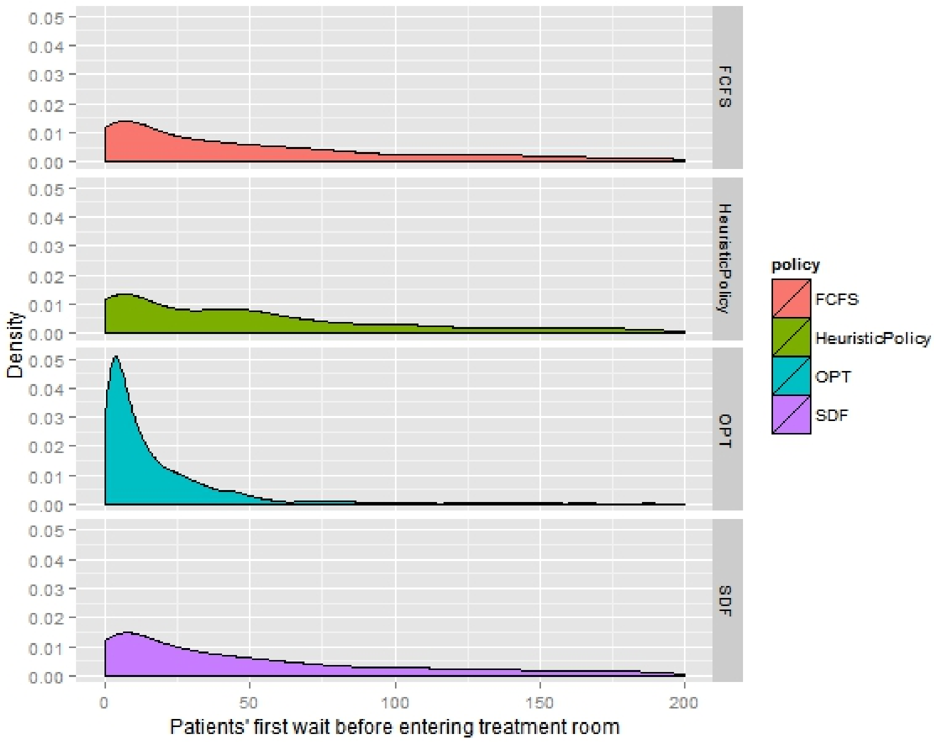

Figure 4.8: Density plot for patients' first waiting time under different policies. - (configuration 1)

Table 4.6 shows the performance results for $\mathbb{P}(w_i^{seq\_1} \leq D^{seq\_1})$, $\mathbb{E}_\mathbb{P}(w_i^{seq\_1})$ and $\mathbb{E}_\mathbb{P}(LoS_i)$ (configuration 1).

Policy	$\mathbb{P}(w_i^{seq\_1} \leq D^{seq\_1})$	$\mathbb{E}_\mathbb{P}(w_i^{seq\_1})$	$\mathbb{E}_\mathbb{P}(LoS_i)$
FCFS	49%	84.33	103.32
SDF	55%	63.69	114.68
HeuristicPolicy	52%	66.85	113.94
OPT	81%	63.43	112.80

Table 4.6: Performance Measure for FCFS, SDF, HeuristicPolicy and OPT (configuration 1).

Table 4.7 shows the performance results for the overall length of stay's quantile information (configuration 1).

Policy	40%	50%	60%	70%	80%	90%
FCFS	49.0	68.0	89.0	123.0	174.0	251.0
SDF	51.0	69.0	96.0	130.0	187.0	289.0
HeuristicPolicy	49.0	65.0	90.0	129.0	187.0	289.0
OPT	29.0	39.0	52.0	77.0	133.0	334.0

Table 4.7: Length of stay's quantile under FCFS, SDF, HeuristicPolicy and OPT (configuration 1).

Table 4.8 shows the performance results for patients' first waiting quantile information (configuration 1).

Policy	40%	50%	60%	70%	80%	90%
FCFS	30.0	47.0	69.0	102.0	153.0	231.0
SDF	24.0	36.0	53.0	76.0	114.0	171.0
HeuristicPolicy	30.6	43.0	56.0	77.0	114.0	172.0
OPT	8.0	12.0	17.0	27.0	44.0	186.0

Table 4.8: First waiting's quantile under FCFS, SDF, HeuristicPolicy and OPT (configuration 1).

4.4.3.2 Configuration 2

We slightly change the thresholds for first waiting and accumulated waiting (D^{seq_1} and D). Configuration parameters under configuration 2 are listed in Table 4.9.

Notation	Description	Value
λ	patient arrival rate which follows possion distribution	0.07
μ	service rate of doctor	0.08
λ/μ	service load	88%
ϵ	threshold part for HeuristicPolicy (minutes)	5
T	simulation duration, in terms of minutes	50000
N	number of patients see by the system	3002
D^{seq_1}	time limit for first waiting (minutes)	60
D	time limit for the accumulated waiting (minutes)	120

Table 4.9: Configuration 2's input parameters

Figure 4.9 shows the density plot for the overall length of stay in ED under configuration 2

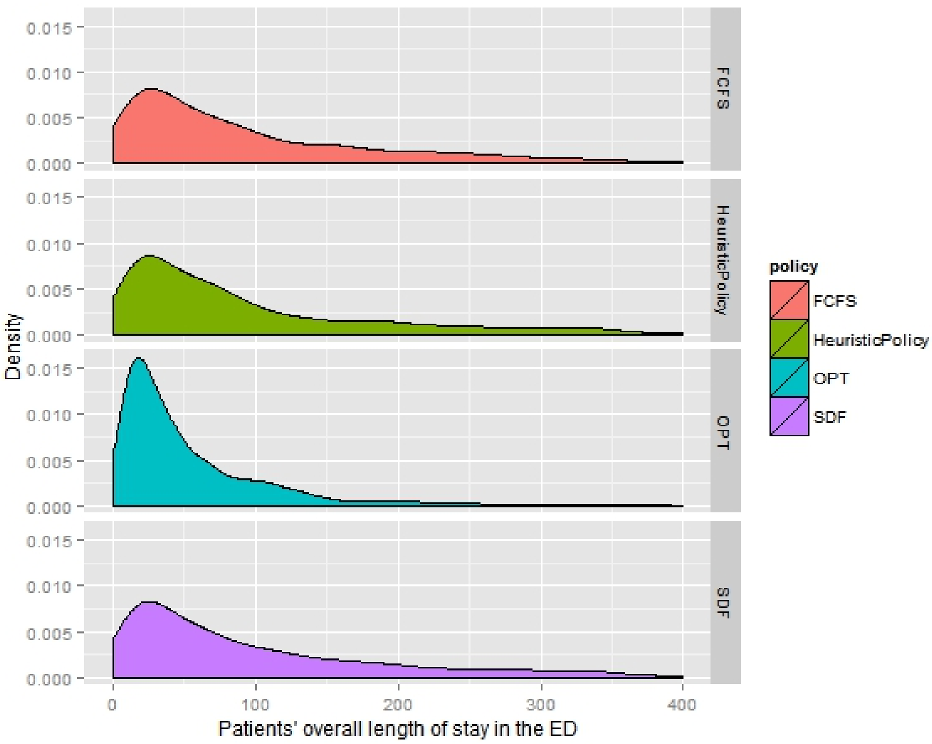

Figure 4.9: Density plot for patients' length of stay under different policies. - (configuration 2)

Figure 4.10 shows the density plot for the first waiting time of patients in ED under configuration 2

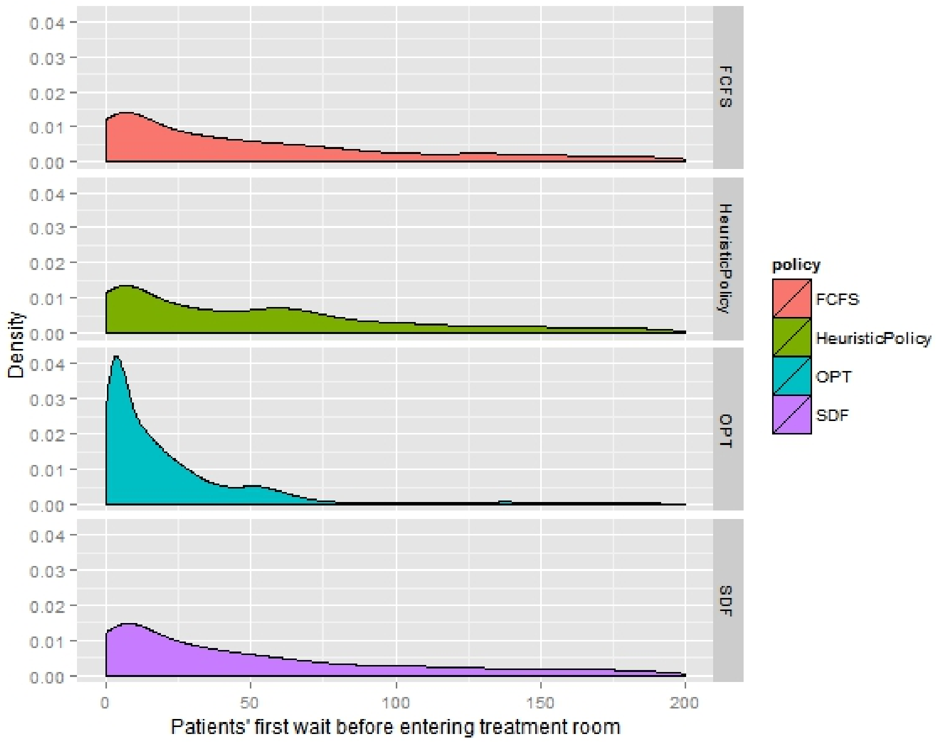

Figure 4.10: Density plot for patients' first waiting time under different policies. - (configuration 2)

Table 4.10 shows the performance results for $\mathbb{P}(w_i^{seq\_1} \leq D^{seq\_1})$, $\mathbb{E}_{\mathbb{P}}(w_i^{seq\_1})$ and $\mathbb{E}_{\mathbb{P}}(LoS_i)$ (configuration 2).

Policy	$\mathbb{P}(w_i^{seq\_1} \leq D^{seq\_1})$	$\mathbb{E}_{\mathbb{P}}(w_i^{seq\_1})$	$\mathbb{E}_{\mathbb{P}}(LoS_i)$
FCFS	57%	84.33	103.32
SDF	64%	63.69	114.68
HeuristicPolicy	59%	67.85	112.94
OPT	84%	64.43	111.80

Table 4.10: Performance Measure for FCFS, SDF, HeuristicPolicy and OPT (configuration 2).

Table 4.11 shows the performance results for the overall length of stay's quantile information (configuration 2).

Policy	40%	50%	60%	70%	80%	90%
FCFS	49.0	68.0	89.0	123.0	174.0	251.0
SDF	51.0	69.0	96.0	130.0	187.0	289.0
HeuristicPolicy	49.0	67.0	87.0	124.0	187.0	289.0
OPT	31.0	42.0	57.0	83.0	127.0	351.0

Table 4.11: Length of stay's quantile under FCFS, SDF, HeuristicPolicy and OPT (configuration 2).

Table 4.12 shows the performance results for patients' first waiting quantile information (configuration 2).

Policy	40%	50%	60%	70%	80%	90%
FCFS	30.0	47.0	69.0	102.0	153.0	231.0
SDF	24.0	36.0	53.0	76.0	114.0	171.0
HeuristicPolicy	30.0	48.0	62.0	79.0	115.0	172.0
OPT	10.0	15.0	21.6	32.0	52.0	161.0

Table 4.12: First waiting's quantile under FCFS, SDF, HeuristicPolicy and OPT (configuration 2).

4.4.3.3 Configuration 3

We increase the service load in the simulation and evaluate their performance. Configuration parameters under configuration 3 are listed in Table 4.13.

Notation	Description	Value
λ	patient arrival rate which follows possion distribution	0.075
μ	service rate of doctor	0.08
λ/μ	service load	93%
ϵ	threshold part for HeuristicPolicy (minutes)	5
T	simulation duration, in terms of minutes	50000
N	number of patients see by the system	3250
$D^{seq\text{-}1}$	time limit for first waiting (minutes)	60
D	time limit for the accumulated waiting (minutes)	120

Table 4.13: Configuration 3's input parameters

Figure 4.11 shows the density plot for the overall length of stay in ED under configuration 3

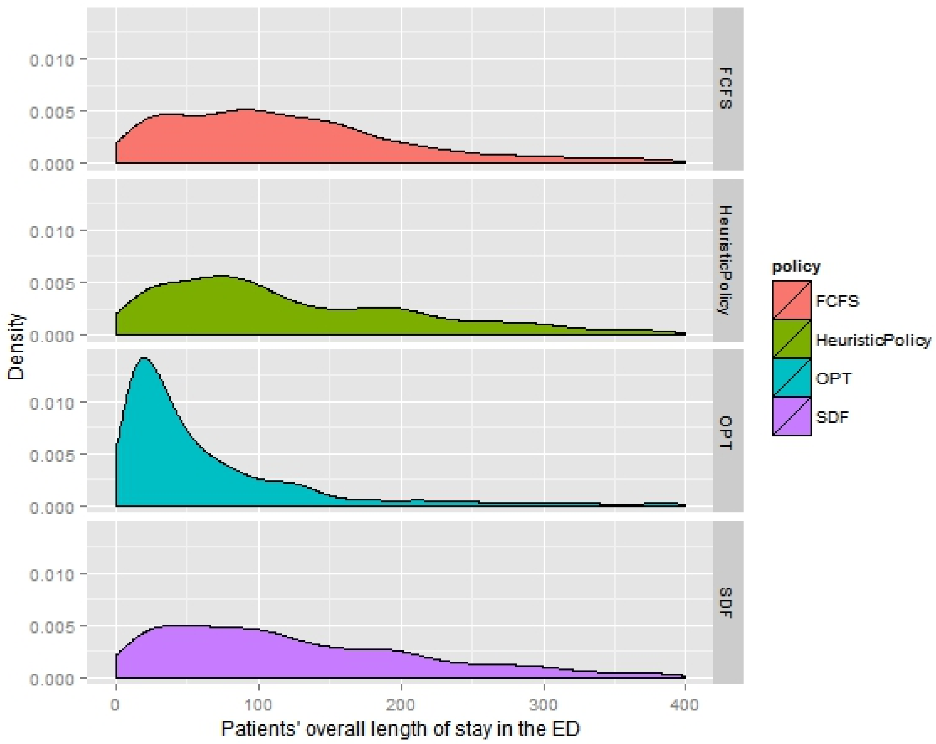

Figure 4.11: Density plot for patients' length of stay under different policies. - (configuration 3)

Figure 4.12 shows the density plot for the first waiting time of patients in ED under configuration 3

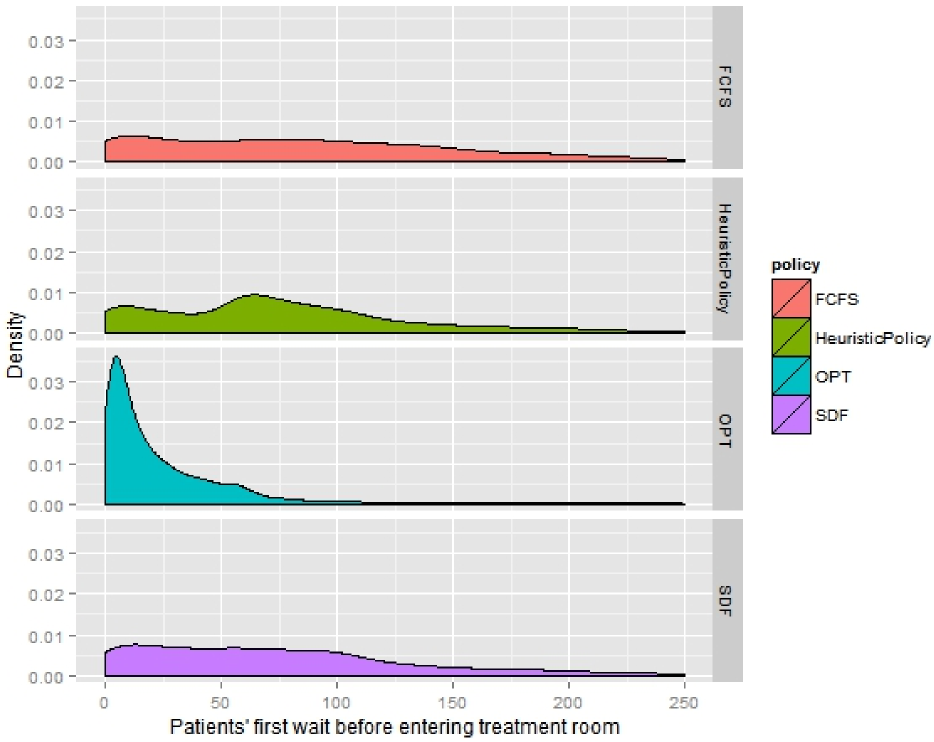

Figure 4.12: Density plot for patients' first waiting time under different policies. - (configuration 3)

Table 4.14 shows the performance results for $\mathbb{P}(w_i^{seq\text{-}1} \leq D^{seq\text{-}1})$, $\mathbb{E}_{\mathbb{P}}(w_i^{seq\text{-}1})$ and $\mathbb{E}_{\mathbb{P}}(LoS_i)$ (configuration 2).

Policy	$\mathbb{P}(w_i^{seq\text{-}1} \leq D^{seq\text{-}1})$	$\mathbb{E}_{\mathbb{P}}(w_i^{seq\text{-}1})$	$\mathbb{E}_{\mathbb{P}}(LoS_i)$
FCFS	35%	112.33	131.32
SDF	45%	85.69	147.68
HeuristicPolicy	40%	88.85	145.94
OPT	77%	84.43	146.30

Table 4.14: Performance Measure for FCFS, SDF, HeuristicPolicy and OPT (configuration 3).

Table 4.15 shows the performance results for the overall length of stay's quantile information (configuration 3).

Policy	40%	50%	60%	70%	80%	90%
FCFS	90.0	110.0	133.0	157.0	194.0	268.5
SDF	89.0	111.0	141.0	179.0	221.0	305.0
HeuristicPolicy	84.0	104.5	137.0	180.0	220.6	304.3
OPT	36.0	50.0	73.0	113.0	233.0	364.0

Table 4.15: Length of stay's quantile under FCFS, SDF, HeuristicPolicy and OPT (configuration 3).

Table 4.16 shows the performance results for patients' first waiting quantile information (configuration 3).

Policy	40%	50%	60%	70%	80%	90%
FCFS	70.0	90.0	113.0	137.0	173.0	245.5
SDF	54.0	69.0	86.0	103.0	130.6	189.0
HeuristicPolicy	61.0	72.0	86.0	103.0	130.0	190.0
OPT	12.0	18.0	28.0	44.0	74.0	294.0

Table 4.16: First waiting's quantile under FCFS, SDF, HeuristicPolicy and OPT (configuration 3).

4.4.4 Performance Discussion

We have shown performance results of different proposed policies above. In terms of performance measure $\mathbb{P}(w_i^{seq\_1} \leq D^{seq\_1})$, our optimized policy significantly outperform the other three policies. This superiority also applies to the percentage of patient cases which are met within given threshold as shown in those quantile information tables (Table 4.7, 4.11 and 4.15). Meanwhile, the expected value of both FW and LoS are not always dominant to the other three approaches. In fact, we observe that the extreme cases (e.g. $\geq$ 90%-quantile) in our optimized polices are waiting or staying longer (about 20% worse off). This could be a tradeoff by letting a majority of patients meet those targets first. In reality, we might assume this situation be within a reasonable range.

Summarizing, the OPT method provides better and more stable performance. And the optimization problem can be solved very efficiently in any modern Mixed Integer Programming (MIP) solvers. A possible limitation of our current methodology is that we only provide the case with a single doctor. Copping with multiple severs (doctors) will be more complex, since it requires some balancing strategy among all available servers. Nevertheless, our research shed light on solving this patient flow control problem in reality both analytically and practically.

5

Conclusion and Discussion

In this dissertation we investigate three topics regarding decision making under uncertainty, ranging from extensible theoretical framework, software tools to practical applications. Our proposed framework under distributionally robust linear optimization could be widely applied, due to its rich expressiveness of uncertainty, extensibility of multi-stage problem and computational advantage. Constructing the uncertainty set could simply be driven by available data, especially in this big data era. And the uncertainty form is specified by only linear and conic quadratic representable expectation constraints. Our generalized decision rule in the recourse functions could easily outperform other existing decision rules such as linear decision rule, extended linear decision rule and deflected linear decision rule. Besides the difficulties it may arise in transforming the original problem into a tractable robust counterpart optimization, we have developed a software modeling package named ROC (written in C++ but could be easily encapsulated) which saves the effort of manual transformation. As the next phase of verification our theoretical foundation, we have explored the area of health-care operations management, and picked two research questions: (1) How to optically assign elective admission bed quotas when facing the challenge of uncertain demand of emergency inpatients? (2) How to optimize the patient flow control in the emergency department with targeted deadlines?

The two applications both make use of the data provided by Singapore hospitals, and we are able to show significant improvements in their respective performance. Our ultimate goal is to let practitioners be able to implement our model or policy in hospital's decision support system easily. In addition to the contributions we emphasized in the thesis, this dissertation sheds light on the discipline of future Business Analytics or data driven decision making. Future research can further exploiting the framework and methodology we proposed here and apply this work to more practical applications, e.g., appointment scheduling, resource allocation, and project management.

For patient flow scheduling problem, we have shown some preliminary results and more work needs to be done in order to make it more practicable and would be eventually be implemented. In the future research, we will continue to work on this part. The study so far only focuses on a single doctor's case and cannot be extended to multiple doctors case, let alone if doctors are of different types. We will add new methodologies to extend our current work within a multiple doctors' context. For example, the service rate will be no longer single doctor based, it could be a time, patient or even system state based variable. And more differentiate would be made between new patients and returning patients.

Other future research can continue exploiting the theoretical framework built for distributionally robust optimization. For example, we may add nonlinear constraints or decision rules to current linear framework. Next, we could incorporate other types of problems such as SDP in conic programming if efficient SDP solver become available.

Bibliography

Alizadeh, F., D. Goldfarb (2003) Second-order cone programming. *Math. Programming* 95(1):2–51.

Anderson, E. J., and P. Nash. 1987. *Linear Programming in Infinite-dimensional Spaces: Theory and Applications.* Wiley, Singapore.

Balintfy, J. L. (1960) A stochastic model for the analysis and prediction of admissions and discharges in hospitals. *Management Science*, **2** 288 – 299.

Batt, R. J., Terwiesch, C. (2012). Doctors under load: An empirical study of state-dependent service times in emergency care. Working paper. 96, 98

Ben-Tal, A., A. Nemirovski (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805. 2, 7, 55

Ben-Tal, A., A. Nemirovski (1999) Robust solutions of uncertain linear programs. *Oper. Res. Lett* 25(1, Auguest):1–13. 2, 7

Ben-Tal, A., A. Nemirovski (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming Ser. A* 88(3):411–424. 7

Ben-Tal, A., A. Nemirovski (2001a) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* SIAM. 19

Ben-Tal, A., A. Nemirovski (2001b) On polyhedral approximations of the second-order cone *Mathematics of Operations Research* 26:193–205 43

Ben-Tal, A., A. Goryashko, E.Guslitzer, A. Nemirovski (2004) Ajustable robust solutions of uncertain linear programs. *Math. Programming* 99:351-376. 2, 8, 12

Bertsimas, D., D. A. Iancu, P. A. Parrilo (2010) Optimality of Affine Policies in Multistage Robust Optimization. *Mathematics of Operations Research* 35(2):363–394. 8

Bertsimas, D., D.B. Brown (2009) Constructing uncertainty sets for robust linear optimization. *Operations Research* 57(6) 1483-1495. 7

Bertsimas, D., M. Sim (2004) The price of robustness. *Oper. Res.* 52(1):35–53. 7, 55

Birge, J. R., F. Louveaux (1997) Introduction to Stochastic Programming. Springer, New York. 7

Breton, M., S. El Hachem (1995) Algorithms for the solution of stochastic dynamic minimax problems. *Comput. Optim. Appl.* 4:317–345. 7

Cardoen, B., E. Demeulemeester, and J. Beliën. 2010. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, **201** 921 – 932.

Chen, W., M. Sim (2009) Goal-driven optimization. *Operations Research* 57(2)(2):342–357. 7

Chen, X., Y. Zhang (2009) Uncertain linear programs: Extended affinely adjustable robust counterparts. *Oper. Res.* 57(6):1469–1482. 8, 10, 26

Chen, X., M. Sim, P. Sun (2007) A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071. 2, 7, 8, 56

Chen, W., M. Sim, J. Sun, and C.-P. Teo. 2010. From CVaR to uncertainty set: implications in joint chance constrained optimization. *Operations Research*, **58** 470 – 485. 56

Chen, X., M. Sim, P. Sun, J. Zhang (2008) A linear decision-based approximation approach to stochastic programming. *Oper. Res.* 56(2):344–357. 2, 8, 10, 26, 32, 33

Cochran, J. K., and K. Roche. 2008. A queuing-based decision support methodology to estimate hospital inpatient bed demand. *The Journal of the Operational Research Society*, **59** 1471 – 1482. 53, 54

Delage, E., Y. Ye (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):596–612. 7, 56

Denton, B., and D. Gupta. 2003. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, **35** 1003 – 1016.

Denton, B. T., A. J. Miller, H. J. Balasubramanian, and T. R. Huschka. 2010. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, **58** 802 – 816.

Dupacova, J.(1987) The minimax approach to stochastic programming and an illustrative application. *Stochastics* 20(1):73–88. 1, 7

Dyer, M., L. Stougie (2006) Computational complexity of stochastic programming problems. *Math. Programming Ser. A* 106:423–432. 8

Ellsberg, D. (1961) Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, 75(4), pp. 643-669. 6

Esogbue, A. O., Singh A. J. (1976) Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, 75(4), pp. 643-669. 54

Gerhak, Y., D. Gupta, and M. Henig. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, **42** 321 – 334.

Garstka, S. J., R. J.-B. Wets (1974) On decision rules in stochastic programming. *Math. Programming* 7(1):117-143. 8

Geer, R., Smith, J. (2004). Strategies to take hospitals off (revenue) diversion. *Healthcare financial management.* 58(3):70-74. 88

Ghaoui, El. L., H. Lebret (1997) Robust solutions to least-squares problems with uncertain data. *SIAM J.Matrix Anal. Appl.*18(4):1035–1064. 7

Ghaoui, El. L., F. Oustry, H. Lebret (1998) Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* 9:33–53. 7, 55

Ghaoui L. El, M. Oks, and F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, **51** 543 – 556.

Gilboa, I., D. Schmeidler (1989) Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2):141–153. 7

Goh, J., M. Sim (2009) Robust optimization made easy with ROME. *Oper. Res.* 59(4):973–985. 3, 52, 56

Goh, J., M. Sim (2010) Distributionally robust optimization and its tractable approximations. *Oper. Res.* 58(4):902–917. 8, 9, 10, 26, 32, 33, 56

Gorunescui F., McClean S. I., Millard P. H. (2004) Using a queueing model to help plan bed allocation in a department of Geriatric medicine. *Health Care Management Science* 5:307–312. 54

Guerriero, F., and R. Guido. 2011. Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, 14:89–114.

Harrison, G. W., A. Shafer, and M. Mackay. 2005. Modelling variability in hospital bed occupancy. *Health Care Management Science*, 8:325–334.

Harper, P. R., and A. K. Shahani. 2002. Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53:11 – 18. 53

Hsu M., M. Bhatt, R. Adolphs, D. Tranel, C.F. Camerer. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310 1680–1683. 6

Huang J., Carmeli B., Mandelbaum A. (2014) Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback. *Working paper.* 5, 91

Isii, K. (1962) On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics* 14(1):185–197. 19

Jebali, A., A. B. Hadjalouane, and P. Ladet. 2005. Operating rooms scheduling. *International Journal of Production Economics*, 99:52 – 62.

Kc, D. S., Terwiesch, C. (2009). Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science.* 55(9):1486-1498. 97

Kao, E. P. C., and G. G. Tung. 1981. Bed allocation in a public health care delivery system. *Management Science*, 27: 507 – 521. 53, 54

Kim, S-C, I. Horowitzb, K. K. Young, and T. A. Buckleye. 2000. Flexible bed allocation and performance in the intensive care unit. *Journal of Operations Management*, 18:427 – 443.

Knight, F. H. (1921) Risk, uncertainty and profit. *Hart, Schaffner and Marx.* 6

Kolesar, P. 1970. A Markovian model for hospital admission scheduling. *Management Science*, 16:384 – 396.

Kokangul, A. 2008. A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit source. *Computer Methods and Programs in Biomedicine*, 90:56 – 65.

Kuhn, D., W. Wiesemann, A. Georghiou (2011) Primal and dual linear decision rules in stochastic and robust optimization. *Math. Programming* 130(1):177–209. 8

Lamiri, M., X. Xie, A. Dolgui, and F. Grimaud. 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185:1026 – 1037. 55

Steve LaValle, Michael S. Hopkins, Eric Lesser, Rebecca Shockley, Nina Kruschwitz (2010) Analytics: The new path to value. *IBM Research Institute.* vi

Lian, J., K. Distefano, S. Shields, C. Heinichen, M. Giampietri, and L. Wang. 2010. Clinical appointment process: improvement through schedule defragmentation. *IEEE Engineering in Medicine and Biology Magazine*, March/April 127 – 134.

Lobo, M., L. Vandenberghe, S. Boyd, H. Lebret (1998) Applications of second-order cone programming. *Linear Algebra and its Applications* 284(1-3):193–228.

Löberg, J. (2012) Automatic robust convex programming. *Optimization methods and software* 27(1): 115–129. 52

Min, D., and Y. Yih. 2010. Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206:642 – 652.

Murty, K. G., and S. N. Kabadi. 1987. Some NP-Complete problems in quadratic and nonlinear programming, *Mathematical Programming*, 39:117 – 129.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin.* 116(2):220. 97

Pitts, S. R., Niska, R. W., Xu, J., Burt, C. W. (2008). National hospital ambulatory medical care survey: 2006 emergency department summary. *Natl Health Stat Report.* 7(7):1-38. 88

Proudlove, N. C., S. Black, and A. Fletcher. 2007. OR and the challenge to improve the NHS: modelling for insight and improvement in in-patients flow. *Journal of the Operational Research Society*, 58:145–158. 54, 55

Popescu, I. (2007) Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* 55(4):98–112. 7

Robb, W. B., M. J. O'Sullivan, A. E. Brannigan, and D. J. Bouchier-Hayes. 2004. Are elective surgical operations cancelled due to increasing medical admissions. *Irish Journal of Medical Sciences*, 173:129 – 132. 53

Rohleder, T., and K. Klassen. 2002. Rolling horizon appointment scheduling: a simulation study. *Health Care Management Science*, 5:201 – 209.

Ruszczynski, A., A. Shapiro (2003) Stochastic Programming. *Handbooks in Operations Research and Management Science* 10. Elsevier Science, Amsterdam. 7

Shapiro A. 2001. On duality theory of conic linear problems. *Semi-Infinite Programming: Recent Advances,* Kluwer Academic Publishers, 135 –165.

Shapiro, A., A. Kleywegt (2002) Minimax analysis of stochastic programs. *Optimization Methods and Software*, 17(3):523–542. 7

Scarf, H. (1958) A min-max solution of an inventory problem. K. Arrow, ed. *Studies in the Mathematical Theory of Inventory and Production.* Stanford University Press, Stanford, CA, 201–209. 1, 7

Schultz, K. L., Juran, D. C., Boudreau, J. W., McClain, J. O., Thomas, L. J. (1998). Modeling and worker motivation in JIT production systems. *Management Science.* 44(12-part-1):1595-1607. 97

See, C.-T., M. Sim (2009) Robust approximation of multiperiod inventory management. *Oper. Res.* 58(3):583–594. 8, 10

Shapiro, A., S. Ahmed (2004) On a class of minimax stochastic programs. *SIAM Journal on Optimization* 14(4):1237–1249. 7

Shapiro, A., A. Nemirovski (2005) On complexity of stochastic programming problems. V. Jeyakumar, A. Rubinov, eds. *Continuous Optimization.* Springer, New York, 111–146. 8

A. L. Soyster (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* 21(5):1154–1157. 2

Teow, K. L., and W. S. Tan. 2008. Allocation of hospital beds in an existing hospital. *Journal of Operations and Logistics*, 2:2–2. 53

Teow, K.L., Heng, BH, Chong Y, Sim J. 2007. Using mathematical programming to optimize bed occupancy by smoothing elective against emergency admissions. *NHG Annual Scientific Congress.* 55

Vandenberghe, L., S. Boyd, and K. Comanor. 2007. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49::52–64.

Vassilacopoulos, G. 1985. A simulation model for bed allocation to hospital inpatient departments. *Simulation*, 45:233–241. 55

Wardrope, J., and P. Driscoll. 2003. Turbulent times. *Emergency Medicine Journal*, 20:116–116. 53

Wiesemann, W., D. Kuhn, M. Sim (2014) Distributionally Robust Convex Optimization. *Under minor revision In Operations Research.* 7, 9, 13, 15, 16, 17, 18

Wolff, R. W. (1989). Stochastic modelling and the theory of queues. Englewood Cliffs, NJ. 96

Xu, H., S. Mannor (2012) Distributionally robust Markov decision processes. *Mathematics of Operations Research* 37(2):288–300. 7

Žáčková, J . (1966) On minimax solution of stochastic linear programming problems. *Časopis pro Pěstování Matematiky*, 91:423–430. 7

Zemach, R. 1970. A model of health service utilization and resource allocation. *Operations Research*, 18:1071–1086.