

Convex Nonlinear and Integer Programming Approaches for Distributionally Robust Optimization of Complex Systems

by

Yiling Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2019

Doctoral Committee:

Assistant Professor Ruiwei Jiang, Co-Chair
Associate Professor Siqian Shen, Co-Chair
Professor Brian Denton
Professor Jon Lee
Assistant Professor Johanna Mathieu

Yiling Zhang

zyiling@umich.edu

ORCID iD: 0000-0002-1077-198X

©Yiling Zhang

2019

ACKNOWLEDGMENTS

First and foremost, I would like to express my most sincere gratitude to my advisor Prof. Siqian Shen. I am so grateful that she took a chance with me when I was a master student and guided me into the field of optimization. It was her tireless passion and inspirations that led me to a career path I believe that will be fun to explore.

I am also grateful to Prof. Ruiwei Jiang, whose invaluable input helped refine and inspire my research. Prof. Johanna Mathieu introduced me through our collaboration to topics in power system operations. Our collaborator, Prof. Mengshi Lu at Purdue University is gratefully appreciated for his enlightening ideas and inspiring discussion on our work. I also appreciate several technical and casual discussions with Prof. Ayca Erdogan at San Jose State University. Thanks to Dr. Jin Dong for his generous mentorship during my internship at the Oak Ridge National Laboratory.

It was also my privilege to take courses with Prof. Marina Epelman, Prof. Jon Lee, and Prof. Brian Denton. They do not only provide me with beautiful mathematical knowledge but also greatly impact my understanding of teaching and learning.

My PhD journey would be less joyful without support and company from my fellow colleagues, Minseok Ryu, Dr. Yan Deng, Dr. Bowen Li, Miao Yu, Hideaki Nakao, Qi Luo, Dr. Wenbo Sun, Qi He, Karmel Shehadeh, Ece Sanci, Sentao Miao, Yuanyuan Guo, Dr. Zhihao Chen, Dr. Yuchen Jiang, Dr. Huanan Zhang, Luze Xu, Xiang Liu, and Zhen Zhang. I would like to extend my thanks to JJ Lin for making fantastic music that keeps me focused during many paper-writing nights.

The last but not the least, I own my special thanks to my parents for their so much love and support.

TABLE OF CONTENTS

Acknowledgments	ii
List of Figures	vi
List of Tables	vii
List of Appendices	ix
List of Abbreviations	x
Abstract	xi
 Chapter	
1 Introduction	1
1.1 Background	1
1.1.1 Distributionally Robust Optimization	3
1.1.2 Ambiguity Sets	3
1.2 Dissertation Overview	4
2 Distributionally Robust Chance-Constrained Optimal Power Flow with Uncertain Renewables and Uncertain Reserves Provided by Loads	7
2.1 Introductory Remarks	7
2.2 Modeling	9
2.2.1 Nomenclature	9
2.2.2 Chance-Constrained OPF	11
2.2.3 Distributionally Robust Reformulation	13
2.2.4 An Alternative Ambiguity Set and DR Reformulation	16
2.3 Benchmark Approaches	17
2.3.1 Reformulation via Gaussian Approximation	17
2.3.2 Reformulation via Scenario Approximation	18
2.4 Studies on the IEEE 9-bus System	19
2.4.1 Test System	19
2.4.2 Flexible Load Modeling	20
2.4.3 Random Sample Generation and Selection	21
2.4.4 Results and Solution Patterns	21
2.5 Studies on the IEEE 39-bus System and IEEE 118-bus System	25
2.5.1 Test System and Sample Generation/Selection	25

2.5.2	Results and Solution Patterns	25
2.6	Concluding Remarks	29
3	Distributionally Robust Appointment Scheduling with Moment-Based Ambiguity Set	30
3.1	Introductory Remarks	30
3.2	DR Appointment Scheduling with Chance-constrained Overtime	32
3.3	Solution Methods for DRCC-S	34
3.3.1	Reformulating the DR objective (3.2a)	34
3.3.2	Reformulating the DR chance constraint (3.2c)	37
3.4	Computational Results	39
3.4.1	Computational Setup	40
3.4.2	Results and Solution Patterns	40
3.5	Concluding Remarks	41
4	Integer Programming for Distributionally Robust Appointment Scheduling with Random No-shows and Service Durations	43
4.1	Introductory Remarks	43
4.1.1	Literature Review	45
4.1.2	Contributions of the Chapter	46
4.1.3	Structure of the Chapter	46
4.2	Formulations of DR Appointment Scheduling	47
4.2.1	Modeling Waiting, Idleness, and Overtime under Uncertainty	47
4.2.2	Supports and Ambiguity Set	49
4.2.3	DR Models with Different Risk Measures	50
4.2.4	Guideline of Selecting Parameter K	51
4.3	Cutting-Plane Approach and Valid Inequalities for DR Expectation Models	53
4.3.1	MINLP Reformulation and a Generic Cutting-Plane Approach	54
4.3.2	MILP Reformulation of the Separation Problem and Valid Inequalities	55
4.4	LP Reformulations of the DR Expectation Model	59
4.5	Computational Results	63
4.5.1	Experiment Setup	65
4.5.2	CPU Time and Computational Details	66
4.5.3	Optimal Objective Values and Scheduling Solution Patterns	68
4.5.4	Results of Out-of-Sample Performance	71
4.5.5	No-Show-Dependent Time Limit	73
4.6	Concluding Remarks	76
5	Solving 0-1 Semidefinite Programs for Distributionally Robust Allocation of Surgery Blocks	77
5.1	Introductory Remarks	77
5.1.1	Literature Review	78
5.1.2	Contributions and Organization of the chapter	81
5.2	DR Chance-Constrained Model	81

5.3	An Equivalent 0-1 SDP and Cutting-plane Algorithm	83
5.4	An Alternative 0-1 SOCP Approximation	86
5.5	Computational Results	89
5.5.1	Experimental Setup	89
5.5.2	Computational Results	90
5.6	Concluding Remarks	92
6	Ambiguous Chance-Constrained Binary Programs under Mean-Covariance Information	93
6.1	Introductory Remarks	93
6.2	Prior Work	95
6.3	DCBP Models and Reformulations	97
6.3.1	Ambiguity Sets	97
6.3.2	SOC Representations of the DRCC	98
6.4	Valid Inequalities for DCBP	105
6.4.1	Submodularity of the 0-1 SOC Constraints	105
6.4.2	Valid Inequalities in a Lifted Space	110
6.5	Numerical Studies	112
6.5.1	Formulation of Ambiguous Chance-Constrained Bin Packing	112
6.5.2	Computational Setup	114
6.5.3	CPU Time Comparison	115
6.6	Concluding Remarks	120
7	Conclusion	121
	Appendices	123
	Bibliography	155

LIST OF FIGURES

Figures

2.1	Optimal generator production and reserve capacities for each generator/load in the IEEE 9-bus system.	24
2.2	Optimal reserve capacities R^* for each generator/load in the IEEE 39-bus system.	27
3.1	Solution patterns with $(\gamma_1, \gamma_2) = (1, 2)$ using Lognormal and Gamma distributed data	41
4.1	An example of $\mathbb{P}(q \in D_q^{(K)})$ for $n = 10$ appointments	52
4.2	Optimal objective values of E- $D_q^{(K)}$ for different settings of parameter R (time limit) and $1 - \nu_i$ (no-show probability)	68
4.3	Appointment schedules produced by E- $D_q^{(2)}$, E- $D_q^{(n+1)}$, and SLP for different settings of parameter R (time limit) and $1 - \nu_i$ (no-show probability)	70
4.4	Appointment schedules produced by E- $D_q^{(2)}$, E- $D_q^{(n+1)}$, and SLP for different settings of parameter R (time limit) and $1 - \nu_i$ (no-show probability) under no-show-dependent time limit	75
6.1	Graphical Solution of Problem (6.8d)	103
6.2	Illustration of the Three SOC Reformulations (6.2), (6.4), and (6.5b) of DRCC (6.3)	104
6.3	CPU time and number of branching nodes under different Ω -values in constraint (6.9)	120

LIST OF TABLES

Tables

2.1	Cost, Reliability, and CPU Time of A1–A4 for the IEEE 9-Bus System with No Congestion	22
2.2	Objective Costs of Various Uncertainty Cases for the IEEE 9-Bus System with No Congestion ($1 - \epsilon_i = 95\%$, $\forall i$)	23
2.3	Cost, Reliability, and CPU Time of A1–A4 for the IEEE 9-Bus System with Congestion ($1 - \epsilon_i = 95\%$, $\forall i$)	23
2.4	Reserves procured by A1–A4 in the IEEE 9-bus system	24
2.5	Cost, Reliability, and CPU Time of A1–A4 for the IEEE 39-Bus System with No congestion ($1 - \epsilon_i = 95\%$)	26
2.6	Cost, Reliability, and CPU Time of A1–A4 for the IEEE 39-Bus System with Congestion ($1 - \epsilon_i = 95\%$)	26
2.7	Objective Costs of Various Uncertainty Cases for the IEEE-39 Bus System with No Congestion ($1 - \epsilon_i = 95\%$, $\forall i$)	26
2.8	Cost, Reliability, and CPU Time of A1 and A4 for the IEEE 118-Bus System ($1 - \epsilon_i = 95\%$)	27
2.9	Reserves procured by A1– A4 in the IEEE 39-bus system	28
2.10	Reserves procured by A1 and A4 in the IEEE 118-bus system	28
3.1	Reliability of optimal solutions to DRCC-S and Stoch-S models	41
3.2	Overtime and waiting time statistics (in minute)	42
4.1	Average CPU time (in second) of solving DR models and SLP with $R = 0$ and $1 - \nu_i = 0.4$	66
4.2	Computational details of solving the MILP models $E-D_q^{(0.3n)}$ and $E-D_q^{(0.7n)}$	67
4.3	Optimal objective value changes according to the value of K and no-show probabilities	69
4.4	Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under perfect information with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$	72
4.5	Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under misspecified distribution with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$	73
4.6	Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under perfect information with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$ and no-show-dependent time limit	74

4.7	Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under misspecified distribution with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$ and no-show-dependent time limit	74
5.1	CPU time (in second) and optimal solutions given by the three benchmark approaches	91
5.2	Average reliability performance of optimal solutions of the three approaches in test samples with only “hMhV” surgeries	91
6.1	CPU time and solution details for solving instances with diagonal matrices . . .	116
6.2	CPU time of DCBP2 solved by different methods with general covariance matrices	117
6.3	CPU time of DCBP2 with general covariance matrices for different problem sizes	118
C.1	Solution reliability in out-of-sample data following a misspecified distribution type	152
C.2	Solution reliability in out-of-sample scenarios with misspecified moments . . .	153
C.3	Optimal open servers and appointment-to-server assignments under general matrices	153
C.4	Solution reliability in out-of-sample scenarios with misspecified moments . . .	154

LIST OF APPENDICES

A Appendix for Chpater 4	123
B Appendix for Chapter 5	149
C Appendix for Chapter 6	150

LIST OF ABBREVIATIONS

- DRO** Distributionally Robust Optimization
- DR** Distributionally Robust
- OPF** Optimal Power Flow
- CC-OPF** Chance-Constrained OPF
- SDP** Semidefinite Programming
- SOCP** Second-order Cone Programming
- SOC** Second-order Cone
- QP** Quadratic Programming
- LP** Linear Programming
- MILP** Mixed-Integer Linear Programming
- MINLP** Mixed-Integer Nonlinear Programming
- DCBP** Distributionally Robust Chance-Constrained Binary Programs
- CVaR** Conditional Value-at-Risk
- pdf** probability density function
- SAA** Sample Average Approximation
- OR** operating rooms

ABSTRACT

The primary focus of the dissertation is to develop Distributionally Robust Optimization (DRO) models and related solution approaches for decision making in energy and healthcare service systems with uncertainties, which often involves nonlinear constraints and discrete decision variables. Without assuming specific distributions, DRO techniques solve for solutions against the worst-case distribution of system uncertainties. In the DRO framework, we consider both risk-neutral (e.g., expectation) and risk-averse (e.g., chance constraint and Conditional Value-at-Risk (CVaR)) measures. The aim is twofold: i) developing efficient solution algorithms for DRO models with integer and/or binary variables, sometimes nonlinear structures and ii) revealing managerial insights of DRO models for specific applications.

We mainly focus on DRO models of power system operations, appointment scheduling, and resource allocation in healthcare. Specifically, we first study stochastic Optimal Power Flow (OPF), where (uncertain) renewable integration and load control are implemented to balance supply and (uncertain) demand in power grids. We propose a Chance-Constrained OPF (CC-OPF) model and investigate its DRO variant which is reformulated as a Semidefinite Programming (SDP) problem. We compare the DRO model with two benchmark models, in the IEEE 9-bus, 39-bus, and 118-bus systems with different flow congestion levels. The DRO approach yields higher a probability of satisfying the chance constraints and shorter solution time. It also better utilizes reserves at both generators and loads when the system has congested flows.

Then we consider appointment scheduling under random service durations with given

(fixed) appointment arrival order. We propose a DRO formulation and derive a conservative SDP reformulation. Furthermore, we study a scheduling variant under random no-shows of appointments and derive tractable reformulations for certain beliefs of no-show patterns.

One preceding problem of appointment scheduling in the healthcare service operations is the surgery block allocation problem that assigns surgeries to operating rooms. We derive an equivalent 0-1 SDP reformulation and a less conservative 0-1 Second-order Cone Programming (SOCP) reformulation for its DRO model.

Finally, we study Distributionally Robust Chance-Constrained Binary Programs (DCBP) for limiting the probability of undesirable events, under mean-covariance information. We reformulate DCBPs as equivalent 0-1 SOCP formulations under two moment-based ambiguity sets. We further exploit the submodularity of the 0-1 SOCP reformulations under diagonal and non-diagonal matrices. We derive extended polymatroid inequalities via submodularity and lifting, which are incorporated into a branch-and-cut algorithm incorporated for efficiently solving DCBPs. We demonstrate the computational efficacy and solution performance with diverse instances of a chance-constrained bin packing problem.

CHAPTER 1

Introduction

1.1 Background

In the era of modern business analytics, one of the biggest challenges in Operations Research concerns the development of optimization problems that can accommodate vast amount of noisy and incomplete data, whilst at the same time truthfully capturing the decision maker's attitude toward risk and ambiguity. DRO is a generalization of the classical robust optimization framework (e.g. Ben-Tal et al., 2009; Ben-Tal and Nemirovski, 2002; Bertsimas and Sim, 2004) complemented with stochastic programming (e.g. Birge and Louveaux, 2011; Shapiro et al., 2009), which addresses distributional ambiguity neither addressed by robust optimization nor by stochastic programming. DRO does not assume a specific probability distribution of system uncertainties but treats uncertain parameters as random variables with ambiguous probability distribution and seeks for solutions against the ambiguous probability distribution in a pre-defined ambiguity set. There are several motivations behind DRO:

1. an accurate estimate of the probability distribution may not be accessible;
2. a probability distribution can evolve over time due to uncertainties' versatile nature;
3. solutions may be sensitive to the ambiguous probability distribution.

In this dissertation, we mainly focus on developing (scalable) solution approaches for DRO within several fields including power system operations, appointment scheduling, and healthcare operations, which have been studied before using effective Operations Research techniques (see, e.g., Bienstock et al., 2014; Erdogan and Denton, 2010; Gupta and Denton, 2008; Margellos et al., 2014).

In power systems, with growing penetrations of renewable energy resources such as wind and solar photovoltaic, their variability increases risks of system operational procedures with integration of renewables (Katiraei and Agüero, 2011; Xie et al., 2011). With the

development of load control algorithms, the increasing penetrations of renewables also lead to additional reserves that can be provided by flexible loads, such as heating and air conditioning. However, the reserve capacity is uncertain because it can be a function of weather and customer usage pattern (Mathieu et al., 2013a). Although even forecasts of load reserves or renewables can be available, forecast errors and also inaccurate information may lead to frequent operational limits (Qiu and Wang, 2015; Roald et al., 2013). Moreover, an accurate probability distribution of renewables and load reserves may be hard to acquire (Carta et al., 2009; Mathieu et al., 2013a; Morgan et al., 2011).

The inaccessibility of accurate probability distributions is also concerned by health-care industry, where appointment scheduling problems arise in numerous settings, such as surgery planning in operating rooms, and scheduling outpatient appointments in primary care and specialty clinics. To plan for appointments' starting time, current literature assumes a wide range of distributions of appointment durations, such as Gamma distribution (Denton and Gupta, 2003; Soriano, 1966), uniform distribution (Denton and Gupta, 2003), exponential distribution (Kaandorp and Koole, 2007), Normal distribution (Denton and Gupta, 2003), and Log-normal distribution (Cayirli and Veral, 2003; Chen and Robinson, 2014). Moreover, the lack of data increases the difficulty of fitting distributions and data is even more limited if broken down by surgery types and surgeons (Denton et al., 2007; Mak et al., 2015).

In the aforementioned application fields, both robust optimization and stochastic programming techniques have been widely employed to solve system operational problems with uncertainty (see, e.g., Begen et al., 2012; Bertsimas et al., 2013; Bienstock et al., 2014; Denton and Gupta, 2003; Jabr, 2013; Jiang et al., 2012; Mittal et al., 2014a; Papavasiliou and Oren, 2013; Roald et al., 2013; Vrakopoulou et al., 2013). However, robust optimization models can lead to overly conservative decisions because they do not exploit distributional knowledge of uncertainty. Contrary to robust optimization, stochastic programming explicitly utilizes distribution information of uncertainty. But as the reasons mentioned above, an accurate estimate of probability distribution can be hard to access.

In the literature, not many have been done to study these problems under ambiguously unknown distribution of uncertain parameters and hence this dissertation work fills in this gap. In the following, we review some basic concepts in DRO and popular choices of ambiguity sets of probability distributions.

1.1.1 Distributionally Robust Optimization

We consider DRO models that impose a min-max objective function and/or distributionally robust constraints to identify an optimal solution by assuming that nature always picks a worst-case probability distribution against the decision maker's choice. Specifically, given $x \in X \subset \mathbb{R}^n$, we consider a risk measure ϱ of a utility function $Q(x, \tilde{\xi}) : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, where $\tilde{\xi} \in \mathbb{R}^m$ consists of all system uncertainties. The risk measure ϱ can be

1. a *risk-neutral* system operator $\varrho(Q(x, \tilde{\xi})) = \mathbb{E}_{\mathbb{P}}[Q(x, \tilde{\xi})]$, the expectation of Q taken with respect to a given fully known probability distribution of \mathbb{P} of $\tilde{\xi}$;
2. *risk-averse* system operators:
 - (a) $\varrho(Q(x, \tilde{\xi})) = \mathbb{P}(Q(x, \tilde{\xi}) \leq Q^{\text{UB}})$, the probability of $Q(x, \tilde{\xi})$ less than its upper bound Q^{UB} with respect to the distribution \mathbb{P} ;
 - (b) $\varrho(Q(x, \tilde{\xi})) = \text{CVaR}_{1-\epsilon}(Q(x, \tilde{\xi}))$, the CVaR (see, e.g., Bertsimas and Sim, 2004; Rockafellar and Uryasev, 2002) of $Q(x, \tilde{\xi})$ with a confidence level $0 < 1 - \epsilon < 1$.

The DRO models impose a generic min-max objective function in the form of

$$\min_{x \in X} \sup_{\mathbb{P} \in \mathcal{D}} \varrho(Q(x, \tilde{\xi})), \quad (1.1)$$

or/and a generic DR constraint in the form of

$$\inf_{\mathbb{P} \in \mathcal{D}} \varrho(Q(x, \tilde{\xi})) \geq \underline{Q}, \quad (1.2)$$

where \underline{Q} is a lower bound if ϱ is the CVaR measure and \underline{Q} can be $0 < 1 - \alpha < 1$ a probability threshold pre-specified according to the decision maker's tolerance towards risk if ϱ is the probability measure. The ambiguity set \mathcal{D} in (1.1) and (1.2) is a pre-defined set of all candidate probability distributions of $\tilde{\xi}$ sharing some common characteristics.

1.1.2 Ambiguity Sets

Two typical ways of constructing ambiguity sets are distance-based and moment-based. The distance-based approaches consider a set of probability distributions within a certain distance from a nominal distribution. Popular choices of statistical distance measures are ϕ -divergence (see, e.g., Ben-Tal et al., 2013; Calafiore, 2007; Jiang and Guan, 2016; Wang

et al., 2013; Yanıkoğlu and den Hertog, 2013) and Wasserstein metrics (see, e.g., Esfahani and Kuhn, 2018; Gao and Kleywegt, 2016; Zhao and Guan, 2018).

Another branch of DRO research considers ambiguity sets constructed based on moment information by requiring candidate probability distributions to satisfy certain moment conditions (see, e.g., Delage and Ye, 2010; Popescu, 2007; Scarf et al., 1958; Zymler et al., 2013). Delage and Ye (2010) present tractable reformulations for the DRO variant of an expectation-based objective function $\min_{x \in X} \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} \left(Q(x, \tilde{\xi}) \right)$, using the conic duality results from Shapiro (2001) and Bertsimas and Popescu (2005). In this dissertation, we derive DRO models mainly based on the moment-based ambiguity set proposed by Delage and Ye (2010) in the form of

$$\mathcal{D} = \mathcal{D}(\Xi, \mu, \Sigma, \gamma_1, \gamma_2) = \left\{ \mathbb{P} : \begin{array}{l} \int_{\tilde{\xi} \in \Xi} \mathbb{P}(d\tilde{\xi}) = 1 \\ (\mathbb{E}_{\mathbb{P}}[\tilde{\xi}] - \mu)^\top \Sigma^{-1} (\mathbb{E}_{\mathbb{P}}[\tilde{\xi}] - \mu) \leq \gamma_1 \\ \mathbb{E}_{\mathbb{P}}[(\tilde{\xi} - \mu)(\tilde{\xi} - \mu)^\top] \preceq \gamma_2 \Sigma \end{array} \right\}, \quad (1.3)$$

where Ξ is the support of $\tilde{\xi}$. The ambiguity set \mathcal{D} is determined by an estimated mean μ (using empirical data) and covariance matrix Σ , and also by parameters γ_1 and γ_2 . The three constraints ensure that (i) the true mean of $\tilde{\xi}$ lies in a μ -centered ellipsoid bounded by γ_1 ; and (ii) the true covariance matrix lies in a positive semi-definite cone bounded by $\gamma_2 \Sigma$. Parameters $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$ are to control the ambiguity set size. The selection of γ_1 and γ_2 depends on the number of samples, support size, and confidence level (see Definition 2 in Delage and Ye (2010)). In general, the two parameters reflect decision makers' risk preference and control the conservatism of solutions. The larger values of the two parameters are, the more moment ambiguity the decision maker tolerates and thus more conservative (robust) solutions are obtained.

1.2 Dissertation Overview

In Chapter 2, we study stochastic OPF, where renewable integration and load control are implemented to balance supply and demand in power grids but often resulting in uncertain balancing reserves. We propose a CC-OPF model and investigate its DRO variant under unknown distribution of random load, renewable productions, and load-control reserve capacities. We consider DRO models with two types of moment-based ambiguity sets. We compare the results of the DRO models with two benchmark models, respectively, derived from Gaussian approximation and scenario approximation approaches, using instances of

the IEEE 9-bus, 39-bus and 118-bus systems with different flow congestion levels, under random renewable input, load, and reserve capacities. We demonstrate that the DRO approach provides an effective distribution-free means for optimizing OPF problems under ambiguous distributional information. It yields higher probability of satisfying the chance constraints and shorter solution time. It also better utilizes reserves at both generators and loads when the system has congested flows.

In Chapter 3, we consider appointment scheduling under random service durations with given (fixed) appointment arrival order. We consider a sequence of appointments arriving at a single server and the system operator has to decide arrival time of each appointment to minimize the total expected total waiting time, while restricting the probability of having server overtime. We propose a DRO formulation with a min-max objective to constrain the total waiting time, and a distributionally robust chance constraint on overtime. We derive a conservative SDP reformulation and conduct computational studies on outpatient treatment scheduling instances benchmarked with a sampling-based stochastic model. Our results show that the DRO solutions demonstrate an approximated “dome” shape suggested as in Denton and Gupta (2003). However, with limited data samples, the “dome” shape pattern does not appear in the optimal scheduling solutions from the stochastic benchmark model.

In Chapter 4, we further consider randomness of no-shows together with random service durations. The probability distribution of the uncertain parameters is assumed to be ambiguous and only the support and first moments are known. We formulate a class of DRO models that incorporate the worst-case expectation/CVaR penalty cost of appointment waiting, server idleness, and overtime into the objective or constraints. Our models flexibly adapt to different prior beliefs of no-show uncertainty. We obtain exact mixed-integer nonlinear programming reformulations, and derive valid inequalities which strengthen the reformulations solved by decomposition algorithms. In particular, we derive convex hulls for special cases of no-show beliefs, yielding polynomial-sized linear programming models for the least and the most conservative supports of no shows. We test various instances to demonstrate the computational efficacy of our approaches and to compare the results of various DRO models given perfect or ambiguous distributional information.

In Chapter 5, we study a preceding problem of appointment scheduling in the healthcare operations field as surgery block allocation, which assigns surgeries to operating rooms. Still, under random surgery durations with unknown distributions, we minimize the cost of opening ORs and surgery assignments while restricting OR overtime risk via a distributionally robust chance constraint in a DRO model. We provide a 0-1 SDP formulation for the DRO model. To tackle the 0-1 SDP, we develop a cutting-plane approach that solves a continuous SDP oracle in each separation procedure to iteratively generate valid

cuts. Also, we derive a less conservative 0-1 SOCP reformulation. Computational tests are conducted on randomly generated outpatient treatment instances to compare solution time, in-sample objective values, and out-of-sample reliability performance. We demonstrate that the cutting-plane method yielded more conservative solutions, and thus having relatively higher cost with better out-of-sample reliability performance. The cutting-plane method also took less CPU time to solve.

One type of constraints included in all the above problems is the set of distributionally robust chance constraints on the utilization of resources, which ensure sufficiently high probabilities of not exceeding the capacity of each resource allocated in service systems. Moreover, these constraints often involve binary variables to select resources, and thus we study generic DCBPs under mean-covariance information in Chapter 6. We reformulate DCBPs as equivalent 0-1 SOCP models under two moment-based ambiguity sets. We further exploit the submodularity in 0-1 second-order cone constraints under diagonal and non-diagonal matrices. We derive extended polymatroid inequalities to strengthen the 0-1 SOCP reformulations via submodularity and lifting. We propose a branch-and-cut algorithm incorporated with the valid inequalities for efficiently solving DCBPs. We also demonstrate the computational efficacy and solution performance with diverse instances of a chance-constrained bin packing problem.

CHAPTER 2

Distributionally Robust Chance-Constrained Optimal Power Flow with Uncertain Renewables and Uncertain Reserves Provided by Loads

2.1 Introductory Remarks

Flexible loads, such as heating and cooling systems, coordinated via load control algorithms are capable of providing reserves to power systems (Bashash and Fathy, 2013; Braslavsky et al., 2013; Callaway, 2009; Mathieu et al., 2013b; Meyn et al., 2015; Zhang et al., 2013). However, the capacity (in MW) of reserves that loads could provide is generally time-varying (Hao et al., 2015; Mathieu et al., 2015). For example, the reserve capacity of an aggregation of air conditioners is a function of weather and load usage patterns (Mathieu et al., 2013a), which are uncertain. Therefore, when the power system operator computes the optimal generation and reserve schedule, the future reserve capacity that will be available from loads is not perfectly known. A conservative option would be to forecast the expected reserve capacity and schedule only a portion of it. Instead, we solve a CC-OPF problem that explicitly considers uncertain reserves from loads, enabling us to more-effectively use this resource.

CC-OPF problems to manage renewable energy and load uncertainty have been posed in Bienstock et al. (2014); Roald et al. (2013, 2015); Vrakopoulou et al. (2013); Zhang and Li (2011), though none of these works considers load reserves or their uncertainty. Vrakopoulou et al. (2014) formulates a multi-period CC-OPF with uncertain load reserves. It is solved with a robust reformulation (Margellos et al., 2014) of the scenario approach (Campi et al., 2009) that makes no assumption on the uncertainty distributions but requires large numbers of uncertainty samples. Li and Mathieu (2015) reformulates the problem assuming Gaussian uncertainty. In practice, it may be difficult to obtain large numbers of uncertainty samples and we would expect the uncertainty to be non-Gaussian.

In this chapter, which is an extension of our preliminary work (Zhang et al., 2015a), we formulate a CC-OPF with uncertain load reserves and reformulate it using Distributionally Robust (DR) optimization (Delage and Ye, 2010), which makes no assumption on uncertainty distributions and does not require large numbers of uncertainty samples. For simplicity of exposition, we use single-period formulation, rather than the multi-period formulation of Li and Mathieu (2015); Vrakopoulou et al. (2014). Using the empirical mean and covariance (i.e., the first two moments) of a small number of uncertainty samples, we construct a convex ambiguity set for the unknown distribution, yielding an SDP reformulation Jiang and Guan (2016), and a tractable SOCP reformulation Wagner (2008). We compare the two DR models with two benchmark approaches, the first of which is an SOCP model obtained by Gaussian approximation, and the second of which is a large-scale Quadratic Programming (QP) model obtained by scenario approximation.

There has been significant recent work on robust OPF, CC-OPF, and DR optimization applied to power system problems. For example, Jabr (2013) formulates a robust OPF problem in which generator participation factors are chosen to ensure a feasible solution for all possible renewable energy injections. The formulation is extended to include chance constraints that manage normally-distributed load forecast error. Ref. Jabr et al. (2015) further extends the method to a multi-period formulation with energy storage. The method assumes the uncertainty set is known. CC-OPFs are posed by Bienstock et al. (2014); Vrakopoulou et al. (2013); Zhang and Li (2011) and solved with scenario-based methods Vrakopoulou et al. (2013) and analytical reformulation assuming normally-distributed uncertainty Bienstock et al. (2014); Roald et al. (2013); Zhang and Li (2011). DR optimization has been applied to the dynamic line rating problem Qiu and Wang (2015) and to quantifying the probability of infeasible dispatch given uncertain wind power injections Wei et al. (2016). The latter uses data-determined uncertainty moments to derive an SDP formulation. Ref. Roald et al. (2015) proposes distributionally robust analytical reformulations of a security constrained CC-OPF. Specifically, the mean, covariance, and structured properties (e.g., symmetry or unimodality) of the uncertainty are used to derive an upper bound for the inverse cumulative distribution function, resulting in a linear programming reformulation. Using upper bounds rather than moments, as in Wei et al. (2016) and our approach, results in conservative solutions. Ref. Lubin et al. (2015) poses a robust CC-OPF problem assuming the uncertainty is normally-distributed but the parameters of the normal distributions are unknown. The paper develops a cutting-plane approach that scales to large systems, and demonstrates the computational efficiency by testing a network with 2209 buses. Note that none of the papers cited in this paragraph consider load reserves or their uncertainty.

The contributions of this work are to i) formulate a single-period CC-OPF with uncertain load reserves; ii) reformulate the CC-OPF, using DR optimization with two types of ambiguity sets, as an SDP and as an SOCP; and iii) compare the cost, reliability, computational time, and optimal decisions of the approach to that of two benchmark approaches. Beyond our preliminary work in Zhang et al. (2015a), in this chapter we have i) corrected the formulation, ii) added a more complete description of the SDP reformulation and derived an SOCP reformulation by using a strengthened ambiguity set, iii) generated results for the IEEE 39-bus system and IEEE 118-bus system (in Zhang et al. (2015a) we considered only the IEEE 9-bus system), iv) computed solutions to a variety of additional cases to put the objective costs in context, and v) included comparisons of the optimal solutions, which give significant additional insights into the performance of the approaches. Our results show that the DR approach provides a good trade-off between cost, reliability, and computational tractability as compared to the other approaches.

The chapter is organized as follows. In Section 2.2, we present the CC-OPF and the DR reformulation. Section 2.3 presents the two benchmark approaches. Sections 2.4 and 2.5 present the computational results of the three approaches for the IEEE 9-bus and 39-, 118-bus bus systems, respectively. Section 2.6 summarizes the chapter. We note that the work in this chapter has been published in Zhang et al. (2016).

2.2 Modeling

2.2.1 Nomenclature

A. Numbers

N_{line}	Number of transmission lines
N_B	Number of buses
N_G	Number of conventional generators
N_W	Number of wind power plants
N_L	Number of loads
m	Number of chance constraints
n	Number of decision variables

B. Parameters

$P_W^f \in \mathbb{R}^{N_W}$	Forecasted wind production
$P_L^f \in \mathbb{R}^{N_L}$	Forecasted load consumption
$\underline{P}_G/\overline{P}_G \in \mathbb{R}^{N_G}$	Min/max generator production
$P_{\text{line}} \in \mathbb{R}^{N_{\text{line}}}$	Line flow limit
c	Vector of energy/reserve costs
$B_{\text{bus}} \in \mathbb{R}^{(N_B-1) \times (N_B-1)}$	Bus susceptance matrix
$B_{\text{flow}} \in \mathbb{R}^{N_{\text{line}} \times N_B}$	Flow susceptance matrix
ϵ_i	Violation probability of chance constraint i

C. Random Variables

$\tilde{P}_W \in \mathbb{R}^{N_W}$	Actual wind production
$\tilde{P}_L \in \mathbb{R}^{N_L}$	Actual load consumption
$\tilde{\underline{P}}_L/\tilde{\overline{P}}_L \in \mathbb{R}^{N_L}$	Actual min/max possible load consumption

D. Auxiliary Random Variables

$P_{\text{mis}} \in \mathbb{R}$	Real-time supply/demand mismatch
$R_G \in \mathbb{R}^{N_G}$	Actual generator reserve dispatch
$R_L \in \mathbb{R}^{N_L}$	Actual load reserve dispatch
$P_{\text{inj}} \in \mathbb{R}^{N_B}$	Net power injections at each bus

E. Decision Variables

$P_G \in \mathbb{R}^{N_G}$	Generator production
$\overline{R}_G, \underline{R}_G \in \mathbb{R}^{N_G}$	Generator up/down-reserve capacities
$\overline{R}_L, \underline{R}_L \in \mathbb{R}^{N_L}$	Load up/down-reserve capacities

$\bar{d}_G, \underline{d}_G \in \mathbb{R}^{N_G}$ Generator up/down distribution vectors

$\bar{d}_L, \underline{d}_L \in \mathbb{R}^{N_L}$ Load up/down distribution vectors

We formulate a CC-OPF which minimizes the costs of producing energy and providing reserves while ensuring that stochastic constraints are satisfied with certain probabilities. We consider a power system in which a fraction of the load at each bus is flexible and can provide reserves by increasing/decreasing its consumption from its baseline consumption. However, we assume that each load's minimum and maximum possible consumption are uncertain. For example, the minimum/maximum possible consumption of an aggregation of air conditioners is a function of the number of air conditioners that are switched on, which is a function of outdoor temperature, which is uncertain. Details on the underlying flexible load model used within our formulation are given in Section 2.4.2.

As in Bienstock et al. (2014); Roald et al. (2013); Vrakopoulou et al. (2013), we use the DC (i.e., linearized) power flow approximation. Also, for ease of exposition and results interpretation, we consider single period (e.g., 1 hour), unlike Vrakopoulou et al. (2014) which considers a multi-period OPF. We first formulate the CC-OPF and then its DR variant.

2.2.2 Chance-Constrained OPF

The linear inequalities that involve random variables are

$$\begin{aligned} \tilde{A}x \geq \tilde{b} = \left\{ \begin{array}{l} \underline{P}_G \leq P_G + R_G \leq \bar{P}_G, \\ \underline{P}_L \leq \tilde{P}_L + R_L \leq \bar{P}_L, \\ -\underline{R}_G \leq R_G \leq \bar{R}_G, \\ -\underline{R}_L \leq R_L \leq \bar{R}_L, \\ -P_{\text{line}} \leq B_{\text{flow}} \begin{bmatrix} 0 \\ B_{\text{bus}}^{-1} \hat{P}_{\text{inj}} \end{bmatrix} \leq P_{\text{line}} \end{array} \right\}, \end{aligned} \quad (2.1)$$

which limit generation, load, generator reserves, load reserves, and line flows, respectively. The notation is defined in Section 2.2.1, with $\tilde{\cdot}$ used to denote random variables. The vector of net power injections at each bus is $P_{\text{inj}} \in \mathbb{R}^{N_B}$ is $C_G(P_G + R_G) + C_W \tilde{P}_W - C_L(\tilde{P}_L + R_L)$, where C_G , C_W , and C_L are matrices mapping generators, wind power plants, and loads to buses; $\hat{P}_{\text{inj}} \in \mathbb{R}^{N_B-1}$ contains the last $N_B - 1$ rows of P_{inj} ; B_{bus} is the bus susceptance matrix (including the susceptances between each bus except the slack bus, which is assumed to be Bus 1) and so the quantity in square brackets is the vector of voltage angles; and B_{flow} is

the flow susceptance matrix (which is used to compute the line flows by multiplying each line susceptance by the difference in voltage angle across the line). The full optimization problem is

[CC-OPF]:

$$\min c^\top \langle 1, P_G, P_G^2, \bar{R}_G, \underline{R}_G, \bar{R}_L, \underline{R}_L \rangle \quad (2.2)$$

$$\text{s.t. } P_{\text{mis}} = \sum_{i=1}^{N_W} (\tilde{P}_{W,i} - P_{W,i}^f) - \sum_{i=1}^{N_L} (\tilde{P}_{L,i} - P_{L,i}^f) \quad (2.3)$$

$$\sum_{i=1}^{N_G} P_{G,i} = \sum_{i=1}^{N_L} P_{L,i}^f - \sum_{i=1}^{N_W} P_{W,i}^f \quad (2.4)$$

$$\sum_{i=1}^{N_G} \underline{d}_{G,i} + \sum_{i=1}^{N_L} \bar{d}_{L,i} = 1 \quad (2.5)$$

$$\sum_{i=1}^{N_G} \bar{d}_{G,i} + \sum_{i=1}^{N_L} \underline{d}_{L,i} = 1 \quad (2.6)$$

$$R_G = \bar{d}_G \max\{-P_{\text{mis}}, 0\} - \underline{d}_G \max\{P_{\text{mis}}, 0\} \quad (2.7)$$

$$R_L = \bar{d}_L \max\{P_{\text{mis}}, 0\} - \underline{d}_L \max\{-P_{\text{mis}}, 0\} \quad (2.8)$$

$$\mathbb{P}(\tilde{A}_i x \geq \tilde{b}_i) \geq 1 - \epsilon_i \quad \forall i = 1, \dots, m \quad (2.9)$$

$$x = \langle P_G, \bar{R}_G, \underline{R}_G, \bar{R}_L, \underline{R}_L, \underline{d}_G, \bar{d}_G, \underline{d}_L, \bar{d}_L \rangle \geq \mathbf{0}, \quad (2.10)$$

where we use $\langle \cdot \rangle$ to denote a stacked column vector. The cost vector

$$c = \langle c_0, c_1, c_2, \bar{c}_G, \underline{c}_G, \bar{c}_L, \underline{c}_L \rangle$$

corresponds to “here-and-now” decisions made before realizing the uncertainty; (2.3) calculates the real-time supply/demand mismatch; (2.4) enforces power balance; (2.5)–(2.6) normalize the distribution vectors, which provide a policy for allocating P_{mis} to generators and loads, as in Bienstock et al. (2014); Vrakopoulou et al. (2013); and (2.7)–(2.8) compute the actual reserves provided by generators and loads, respectively. We assume that there are m constraints in (2.1), and use \tilde{A}_i to represent the i^{th} row of matrix \tilde{A} and \tilde{b}_i to represent the i^{th} entry of vector \tilde{b} . The number of rows in matrix \tilde{A} and entries in vector \tilde{b} is $(2N_{\text{line}} + 4N_G + 4N_L)$. Each chance constraint i in (2.9) should be satisfied with probability $1 - \epsilon_i$. The decision variables, listed in (2.10), are the generator production, generator reserve capacities, load reserve capacities, and distribution vectors. Note that the total reserve requirement (i.e., the sum of all generator and load reserves capacities) is determined endogenously, and is a function of the uncertainty.

We assume symmetric reserve dispatch, i.e., $\bar{d}_G = \underline{d}_G$ and $\bar{d}_L = \underline{d}_L$, since (2.7)–

(2.8) become linear enabling DR reformulation (the Gaussian approximation approach also requires this assumption, while the scenario approximation approach does not). However, we assume unsymmetrical reserve procurement (as in CAISO and ERCOT (MacDonald et al., 2012)), i.e., \underline{R}_L is not enforced to be equal to \overline{R}_L and \underline{R}_G is not enforced to be equal to \overline{R}_G . Modeling symmetrical reserve procurement (as in MISO, PJM, NYISO, and NEISO (MacDonald et al., 2012)) reduces the number of decision variables, does not change the form of the resulting optimization problems, and increases the objective cost.

We could also reformulate the CC-OPF, using Sample Average Approximation (SAA), as a mixed-integer quadratic program. However, in Zhang et al. (2015a), via extensive computational studies, we demonstrated that the formulation is computationally intractable, and leads to solutions that are generally worse than solutions obtained by the approaches in this chapter, especially when the true distribution of the underlying uncertainty is unknown. Therefore, we do not include it here.

2.2.3 Distributionally Robust Reformulation

We first introduce the DR approach for reformulating the CC-OPF, which builds an ambiguity set to bound the probability density function (pdf) of the underlying uncertainty. Our approach follows that of Jiang and Guan (2016), which we summarize here for completeness. To the best of our knowledge, this is the first application of this approach to a CC-OPF problem.

Consider each chance constraint i in (2.9). Suppose that $(\tilde{A}_i, \tilde{b}_i) = (\tilde{A}_i(\xi^i), \tilde{b}_i(\xi^i))$, where ξ^i includes the random variables affecting constraint i . Specifically, for all chance constraints except, $\underline{P}_L \leq \tilde{P}_L + R_L \leq \overline{P}_L$, $\xi^i = \langle \tilde{P}_W, \tilde{P}_L \rangle$. For $\underline{P}_L \leq \tilde{P}_L + R_L$, $\xi^i = \langle \tilde{P}_W, \tilde{P}_L, \underline{P}_L \rangle$ and for $\tilde{P}_L + R_L \leq \overline{P}_L$, $\xi^i = \langle \tilde{P}_W, \tilde{P}_L, \overline{P}_L \rangle$. Then, the DR variants of the chance constraints are

$$\inf_{f(\xi^i) \in \mathcal{D}} \mathbb{P}_{\xi^i}(\tilde{A}_i(\xi^i)x \geq \tilde{b}_i(\xi^i)) \geq 1 - \epsilon_i \quad \forall i = 1, \dots, m. \quad (2.11)$$

Without loss of generality, we can define $\tilde{A}_i(\xi^i)$ and $\tilde{b}_i(\xi^i)$ as affine functions of ξ^i (see Shapiro et al., 2009), i.e.,

$$\tilde{A}_i(\xi^i) = A_{i0} + \sum_{k=1}^{K_i} A_{ik} \xi_k^i, \quad b_i(\xi^i) = b_{i0} + \sum_{k=1}^{K_i} b_{ik} \xi_k^i,$$

where K_i is the dimension of ξ^i . Terms A_{i0} and b_{i0} are the deterministic parts of $\tilde{A}_i(\xi^i)$ and $\tilde{b}_i(\xi^i)$ and terms A_{ik} and b_{ik} are the affine coefficients of ξ_k^i , $\forall k = 1, \dots, K_i$. As a result,

we can reformulate (2.11) as $(\bar{A}_i^x)^\top \xi \leq \bar{b}_i^x$, where $\bar{A}_i^x = \langle \tilde{b}_{i1}^\xi - \tilde{A}_{i1}^\xi x, \dots, \tilde{b}_{iK}^\xi - \tilde{A}_{iK}^\xi x \rangle$ and $\bar{b}_i^x = \tilde{A}_{i0}^\xi x - \tilde{b}_{i0}^\xi$.

For simplicity, we drop the index i of ξ^i in the following. Given data samples $\{\xi^\ell\}_{\ell=1}^N$ of ξ , we calculate the empirical mean vector $\mu_0 = \frac{1}{N} \sum_{\ell=1}^N \xi^\ell$ and covariance matrix $\Sigma_0 = \frac{1}{N} \sum_{\ell=1}^N (\xi^\ell - \mu_0)(\xi^\ell - \mu_0)^\top$, and build an ambiguity set (Delage and Ye, 2010)

$$\mathcal{D} = \left\{ f(\xi) : \begin{array}{l} \int_{\xi \in \mathbb{R}^K} f(\xi) d\xi = 1 \\ (\mathbb{E}[\xi] - \mu_0)^\top \Sigma_0^{-1} (\mathbb{E}[\xi] - \mu_0) \leq \gamma_1 \\ \mathbb{E}[(\xi - \mu_0)(\xi - \mu_0)^\top] \preceq \gamma_2 \Sigma_0 \end{array} \right\},$$

where \mathbb{R}^K is the support of ξ . The ambiguity set \mathcal{D} is determined by μ_0 and Σ_0 , and by parameters γ_1 and γ_2 . The three constraints in \mathcal{D} ensure that (i) the integral of pdf $f(\xi)$ is one; (ii) the true mean of ξ lies in a μ_0 -centered ellipsoid bounded by γ_1 ; and (iii) the true covariance matrix lies in a positive semi-definite cone bounded by $\gamma_2 \Sigma_0$. Delage and Ye (2010) describes how the values of γ_1 and γ_2 can be chosen based on the data sample size, risk parameter, and desired confidence. In practice, the values of γ_1 and γ_2 represent a decision maker's risk preference and can be used to change solution conservatism. In general, larger values of γ_1 and γ_2 will lead to more conservative (robust) solutions. We follow Delage and Ye (2010) and set $\gamma_1 = 0$ and $\gamma_2 = 1$.

We solve a minimization problem over the ambiguity set \mathcal{D} of $f(\xi)$, specifically, Delage and Ye (2010); Jiang and Guan (2016)

$$z_{\mathcal{D}} = \min_{f(\xi)} \int_{\mathbb{R}^K} \mathbb{I}_{\mathcal{A}}(\xi) f(\xi) d\xi \quad (2.12)$$

$$\text{s.t.} \quad \int_{\mathbb{R}^K} f(\xi) d\xi = 1 \quad (2.13)$$

$$\int_{\mathbb{R}^K} \begin{bmatrix} \Sigma_0 & \xi - \mu_0 \\ (\xi - \mu_0)^\top & \gamma_1 \end{bmatrix} f(\xi) d\xi \succeq \mathbf{0} \quad (2.14)$$

$$\int_{\mathbb{R}^K} (\xi - \mu_0)(\xi - \mu_0)^\top f(\xi) d\xi \preceq \gamma_2 \Sigma_0, \quad (2.15)$$

where $\mathbb{I}_{\mathcal{A}}(\xi)$ is an indicator function which equals 1 if $\xi \in \mathcal{A} = \{\xi : (\bar{A}_i^x)^\top \xi \leq \bar{b}_i^x\}$ and 0 otherwise, and (2.13)–(2.15) are the constraints in set \mathcal{D} in integral form. The generalized inequality for symmetric matrices, $X \succeq Y$, means that $X - Y$ is a positive semidefinite matrix; similarly, $X \preceq Y$, means that $Y - X$ is a positive semidefinite matrix. A DR chance constraint i is satisfied when $z_{\mathcal{D}} \geq 1 - \epsilon_i$.

Note that the set \mathcal{D} could be very conservative as the worst-case distributions consist

of few discrete points Vandenberghe et al. (2007). Recent literature investigated inclusion of structural properties (e.g., unimodality) to exclude discrete distributions from the ambiguity set (see, e.g., Van Parys et al. (2016), Hanasusanto et al. (2015)). Others use higher order moments and strengthened supports (e.g., Summers et al. (2015)) to reduce the conservatism of the DR approach. However, the construction of set \mathcal{D} also needs to ensure tractability of the reformulation. In the next subsection, we propose an alternative ambiguity set that matches the exact mean and covariance, which leads to a tractable SOCP reformulation.

The DR chance constraints (2.11) are equivalent to the following SDP model (Jiang and Guan, 2016):

$$\gamma_2 \Sigma_0 \cdot G_i + 1 - r_i + \Sigma_0 \cdot H_i + \gamma_1 q_i \leq \epsilon_i y_i \quad (2.16)$$

$$\begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} \succeq \begin{bmatrix} \mathbf{0} & \frac{1}{2} \bar{A}_i^x \\ \frac{1}{2} (\bar{A}_i^x)^\top & y_i + (\bar{A}_i^x)^\top \mu_0 - \bar{b}_i^x \end{bmatrix} \quad (2.17)$$

$$\begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} \succeq \mathbf{0} \quad (2.18)$$

$$\begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix} \succeq \mathbf{0} \quad (2.19)$$

$$y_i \geq 0, \quad (2.20)$$

for $i = 1, \dots, m$, where we use $X \cdot Y$ to denote the Frobenius inner product of X and Y (i.e., $X \cdot Y = \text{tr}(X^\top Y)$) and $y_i \in \mathbb{R}$, $r_i \in \mathbb{R}$, $q_i \in \mathbb{R}$, $p_i \in \mathbb{R}^K$, $H_i \in \mathbb{R}^{K \times K}$, and $G_i \in \mathbb{R}^{K \times K}$. The proof, given in Jiang and Guan (2016), uses conic duality, and is summarized as follows. Denote \mathbb{S}_+^K as the set of symmetric positive semidefinite $K \times K$ matrices. Define dual variables r_i for (2.13), G_i for (2.14), and $\begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix}$ for (2.15). The conic dual of (2.12)–(2.15) is (Jiang and Guan, 2016)

$$z_{\mathcal{D}} = \max_{G_i, H_i, p_i, q_i, r_i} -\gamma_2 \Sigma_0 \cdot G_i + r_i - \Sigma_0 \cdot H_i - \gamma_1 q_i \quad (2.21)$$

$$\begin{aligned} \text{s.t.} \quad & (\xi - \mu_0)^\top (-G_i) (\xi - \mu_0) + 2p_i^\top (\xi - \mu_0) \\ & + r_i \leq \mathbb{I}_{\mathcal{A}}(\xi), \quad \forall \xi \in \mathbb{R}^K \end{aligned} \quad (2.22)$$

$$G_i \in \mathbb{S}_+^{K \times K} \quad (2.23)$$

$$\begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}. \quad (2.24)$$

As strong duality holds for the primal and dual problems, the existence of feasible solutions

to (2.11) is equivalent to having $z_{\mathcal{D}} \geq 1 - \epsilon_i, \forall i = 1, \dots, m$. After applying Lemma 1 in Jiang and Guan (2016), the SDP formulation (2.16)–(2.20) directly follows from the dual formulation (2.21)–(2.24) by replacing the semi-infinite constraints (2.22) by finite SDP constraints.

Importantly, note that the above approach for bounding the unknown pdf $f(\xi)$ is general and allows the uncertainty ξ to be time-varying, correlated, and endogenous. The complete DR CC-OPF model is

$$\min_{x,y,r,H,G,p,q} \{(2.2) : (2.3)–(2.8), (2.10), (2.16)–(2.20) \forall i = 1, \dots, m\}.$$

2.2.4 An Alternative Ambiguity Set and DR Reformulation

In this section, we provide a DR reformulation based on an alternative ambiguity set

$$\mathcal{D}' = \left\{ f(\xi) : \begin{array}{l} \int_{\xi \in \mathbb{R}^K} f(\xi) d\xi = 1 \\ \mathbb{E}[\xi] - \mu_0 = 0 \\ \mathbb{E}[(\xi - \mu_0)(\xi - \mu_0)^\top] = \Sigma_0 \end{array} \right\},$$

which requires that the true mean and covariance matrix of ξ , given by any distribution in set \mathcal{D}' , be exactly the empirical mean μ_0 and covariance Σ_0 . When using $\gamma_1 = 0$ and $\gamma_2 = 1$ in the set \mathcal{D} , we have $\mathcal{D}' \subset \mathcal{D}$, and thus this set produces less conservative solutions by placing more trust in μ_0 and Σ_0 .

With this ambiguity set, the problem can be reformulated as an SOCP, which is more efficient to compute than the SDP reformation using set \mathcal{D} . Specifically, we rewrite the DR constraints (2.11) as

$$\sup_{f(\xi) \in \mathcal{D}'} \mathbb{P}_\xi \left((\bar{A}_i^x)^\top \xi \geq \bar{b}_i^x \right) \leq \epsilon_i \quad \forall i = 1, \dots, m, \quad (2.25)$$

which are equivalent to

$$\sqrt{(\bar{A}_i^x)^\top \Sigma_0 \bar{A}_i^x} \leq \sqrt{\frac{\epsilon_i}{1 - \epsilon_i}} (\bar{b}_i^x - (\mu_0)^\top \bar{A}_i^x) \quad \forall i = 1, \dots, m. \quad (2.26)$$

The derivation follows a variant of the Chebyshev inequality and was given in Wagner (2008) for general DR individual chance constraints. We demonstrate the procedure and result for the DR CC-OPF as follows. Consider a variant of the Chebyshev inequality for random variable X with mean μ and variance σ^2 as $\mathbb{P}[X \geq (1 + \delta)\mu] \leq \frac{\sigma^2}{\sigma^2 + \mu^2 \delta^2}$ for any constant $0 \leq \delta \leq 1$. According to Bertsimas and Popescu (2005), there exists a distribution

in \mathcal{D}' to make the inequality tight. Then, the equivalence between (2.25) and (2.26) can be established by letting

$$\delta = -1 + \frac{\bar{b}_i^x}{\mathbb{E}[(\bar{A}_i^x)^\top \xi]} = -1 + \frac{\bar{b}_i^x}{(\bar{A}_i^x)^\top \mathbb{E}[\xi]}.$$

Therefore, by using a new ambiguity set \mathcal{D}' , we can solve the alternative DR CC-OPF model

$$\min_{x,y,r,H,G,p,q} \{(2.2) : (2.3)–(2.8), (2.10), (2.26)\}.$$

2.3 Benchmark Approaches

We compare the DR approach to two benchmark approaches used in the literature to solve CC-OPF problems. Both use statistical information and derive convex approximations for the exact CC-OPF.

2.3.1 Reformulation via Gaussian Approximation

Assuming Gaussian uncertainty, the CC-OPF can be reformulated as a convex program (Bienstock et al., 2014; Roald et al., 2013). Prior research has not applied this approach to a CC-OPF with uncertain load reserves, with the exception of Li and Mathieu (2015), which used the multi-period CC-OPF formulation from Vrakopoulou et al. (2014). We briefly describe the derivation of the convex program. First, consider constraints (2.9) in an equivalent form

$$\mathbb{P}(\tilde{A}'_i \bar{x} \leq b'_i) \geq 1 - \epsilon_i \quad i = 1, \dots, m, \quad (2.27)$$

where only the constraint vector \tilde{A}'_i is uncertain and the right-hand side scalar b'_i is deterministic. This is because for each constraint i of the form (2.9) where both \tilde{A}_i and \tilde{b}_i are random, we can always define an artificial variable $x_b \in \mathbb{R}$ and rewrite $\tilde{A}_i x \geq \tilde{b}_i$ as

$$-\tilde{A}_i x + \tilde{b}_i x_b \leq 0 \Leftrightarrow \langle -\tilde{A}_i, \tilde{b}_i \rangle^\top \langle x, x_b \rangle \leq 0,$$

for which we enforce $x_b = 1$. Consequently, we have $\tilde{A}'_i = \langle -\tilde{A}_i, \tilde{b}_i \rangle$, $\bar{x} = \langle x, x_b \rangle$, and $b'_i = 0, \forall i = 1, \dots, m$ in (2.27).

If \tilde{A}'_i follows a multivariate normal distribution, denoted by $\mathcal{N} \sim (\mu_i, \Sigma_i)$ with μ_i and Σ_i being the mean and covariance of \tilde{A}'_i , respectively, then $\tilde{A}'_i \bar{x} - b'_i$ follows a multivariate

normal distribution $\mathcal{N} \sim (\mu_i^\top \bar{x} - b'_i, \bar{x}^\top \Sigma_i \bar{x})$. As a result,

$$\mathbb{P}(\tilde{A}'_i \bar{x} \leq b'_i) = \Phi \left(\frac{b'_i - \mu_i^\top \bar{x}}{\sqrt{\bar{x}^\top \Sigma_i \bar{x}}} \right) \quad i = 1, \dots, m, \quad (2.28)$$

following which, constraints (2.9) are equivalent to

$$b'_i - \mu_i^\top \bar{x} + \Phi^{-1}(\epsilon_i) \sqrt{\bar{x}^\top \Sigma_i \bar{x}} \geq 0 \quad i = 1, \dots, m, \quad (2.29)$$

where $\Phi^{-1}(\epsilon_i)$ denotes the ϵ_i -quantile of the standard normal distribution. We rewrite (2.29) as

$$b'_i - \mu_i^\top \bar{x} \geq \Phi^{-1}(1 - \epsilon_i) \sqrt{\bar{x}^\top \Sigma_i \bar{x}} \quad i = 1, \dots, m, \quad (2.30)$$

which are SOCP constraints if $\Phi^{-1}(1 - \epsilon_i) \geq 0$, i.e., $1 - \epsilon_i \geq 0.5$. This is a mild assumption since the chance constraints must be satisfied with probabilities much higher than 0.5. The first benchmark approach solves an SOCP model:

$$\min_x \{(2.2) : (2.3)–(2.8), (2.10), (2.30)\}.$$

If the uncertainties are not Gaussian, then the above model is a convex approximation for the exact CC-OPF.

2.3.2 Reformulation via Scenario Approximation

Campi et al. (2009) proposes a scenario approximation method for chance-constrained optimization problems by enforcing $A_i^s x \geq b_i^s, \forall i = 1, \dots, m$ in all samples s in an approximate sample set Ω_{ap} . The samples in Ω_{ap} could be (i) data observations or (ii) generated from a known distribution by using Monte Carlo sampling. It usually requires a large sample size $|\Omega_{\text{ap}}|$ to guarantee reliability $1 - \epsilon_i$ with high confidence. Each chance constraint i in (2.9) is replaced with

$$A_i^s x \geq b_i^s \quad \forall s \in \Omega_{\text{ap}}. \quad (2.31)$$

It is shown in Campi et al. (2009) that $|\Omega_{\text{ap}}| \geq \frac{2}{\epsilon} (\ln \frac{1}{\beta} + n)$ where $1 - \beta$ is the confidence level and n is the dimension of the decision vector x . Therefore, the second benchmark approach solves a quadratic programming model:

$$\min_x \{(2.2) : (2.3)–(2.8), (2.10), (2.31)\}.$$

A variant of this approach was applied in Vrakopoulou et al. (2013, 2014), which use a

robust reformulation (Margellos et al., 2014) that reduces the required number of samples and the computational time. Moreover, one could employ scenario reduction procedures from the sample average approximation literature to further improve the solution time of the above model.

2.4 Studies on the IEEE 9-bus System

We first compute solutions for the IEEE 9-bus system using the two benchmark approaches and the DR approach, assuming unknown distributions of the uncertainties. We refer to the approaches as:

- A1: Gaussian approximation approach (SOCP)
- A2: Scenario approximation approach (QP)
- A3: Distributionally robust approach (SDP)
- A4: DR approach with ambiguity set \mathcal{D}' (SOCP)

All computational tests are performed on a Windows 7 machine with Intel(R) Core(TM) i7-2600 CPU 3.40 GHz and 8GB memory. All models are solved by CVX implemented in MATLAB with MOSEK as the optimization solver (Grant and Boyd, 2012).

2.4.1 Test System

We obtained parameters for the IEEE 9-bus system from MATPOWER (Zimmerman et al., 2011), and assume that generator reserves are more expensive than load reserves, specifically, $\bar{c}_G = \underline{c}_G = 10 \times \mathbf{1}$ and $\bar{c}_L = \underline{c}_L = 9.8 \times \mathbf{1}$, where $\mathbf{1}$ is a vector of ones of appropriate size. If generator reserves are less expensive than load reserves, the optimal solution is not to use load reserves as they are uncertain and so less valuable to the system. If the cost of generator and load reserves are set equal, the optimal solution is generally to use a combination of both types of reserves. This is because use of generator reserves can increase generator production costs as generators are dispatched sub-optimally to accommodate reserve provision, and this makes use of some (less valuable) load reserves cost effective.

We add one wind power plant at Bus 6 with rated capacity 75 MW and forecasted output 50 MW. We assume 70% of the load at each bus is perfectly forecastable and not flexible. The remaining load at each bus is flexible but uncertain (in terms of both consumption and min/max possible consumption). The load forecasts are set equal to the test case loads.

2.4.2 Flexible Load Modeling

We assume that the flexible loads are aggregations of electric heat pumps, and each can be modeled as a thermal energy storage unit with power capacity P_c , energy capacity E_c , and baseline power consumption P_L , which all vary with (uncertain) outdoor air temperature $\tilde{\theta}$ (Mathieu et al., 2015). We use Fig. 1 of Vrakopoulou et al. (2014) as a look-up table to determine $P_c(\tilde{\theta})$, $E_c(\tilde{\theta})$, and $\tilde{P}_L = P_L(\tilde{\theta})$.

If the power system operator does not provide a method to manage the energy states of energy-constrained reserve resources, the real-time reserve signal will determine the energy states of the load aggregations at the end of each scheduling period – a zero-mean signal will result in final energy states equivalent to the initial energy states, while all other signals will result in final energy states that are different than initial energy states. At the beginning of each scheduling period, we can compute the capacity (in MW) of reserves that each load aggregation can provide as a function of its forecasted P_c , E_c , P_L , and its current energy state assuming, in the worst case, that the real-time reserve signal will be equivalent to its minimum or maximum over the entire scheduling period (e.g., 1 hour). To mitigate the conservatism that results from computing reserve capacities in this way, several power system operators in the U.S. are now managing the energy states of energy-constrained reserve resources. For example, the California ISO uses the 5-minute energy market to ensure that regulation signals are zero-mean over 15-minute periods California ISO (2011).

Assuming a one hour scheduling period and that a load aggregation providing reserves must be able to operate at full capacity for 15-minutes, which is consistent with California ISO (2011), we compute

$$\tilde{P}_L = \max\left(P_L(\tilde{\theta}) - 4(E_c(\tilde{\theta})/2), 0\right) \quad (2.32)$$

$$\tilde{\bar{P}}_L = \min\left(P_L(\tilde{\theta}) + 4(E_c(\tilde{\theta})/2), P_c(\tilde{\theta})\right) \quad (2.33)$$

where the first values within the min/max operators ensure that the energy constraints are not violated and the second values ensure that the power constraints are not violated. Specifically, we assume that initially the heat pumps are operating normally, and so their initial energy state is half their energy capacity. Then, they can increase or decrease their aggregate power consumption by four times $E_c(\tilde{\theta})/2$ (since there are four 15-minute intervals in one hour) unless the resulting power consumption would violate their power constraints, i.e., the aggregate power must be between 0 and $P_c(\tilde{\theta})$.

2.4.3 Random Sample Generation and Selection

We use forecasted and actual outdoor air temperatures from eleven weather stations in Switzerland to compute temperature errors, and then add these errors to assumed temperature forecasts at each load bus $\theta^f = [13, 10, 14]^T$ °C to create temperature samples, with which we compute samples of \tilde{P}_L , $\tilde{\underline{P}}_L$, and $\tilde{\overline{P}}_L$. We scale the samples to be consistent with our load forecast assumptions. Additionally, we use scaled versions of the wind power samples used in Vrakopoulou et al. (2014), which were computed by applying the Markov chain mechanism developed in Papaefthymiou and Klockl (2008) to forecasted and actual hourly wind power data from Germany from 2006-2011.

We generate 10,000 i.i.d. samples of \tilde{P}_W , \tilde{P}_L , $\tilde{\underline{P}}_L$, $\tilde{\overline{P}}_L$, which comprise the support set Ω . For A1, A3 and A4, we assume a decision maker only has limited knowledge of Ω , and so we randomly select 20 samples from Ω . With these samples, we construct the ambiguity set of the unknown pdf by computing their empirical mean and covariance, which we use to build the SOCP constraints (2.30) in A1, the SDP constraints (2.16)–(2.20) in A3, and the SOCP constraints (2.26) in A4. For A2, we use the bound in Campi et al. (2009) to choose the number of samples for the approximate sample set Ω_{ap} in (2.31). Specifically, for $1 - \epsilon_i = 95\%$ and a confidence parameter $\beta = 0.05$, we select 900 random samples since the bound $|\Omega_{\text{ap}}| \geq \frac{2}{\epsilon}(\ln \frac{1}{\beta} + n) = 932$ where $n = 21$ is the dimension of decision variable vector for the 9-bus system. For $1 - \epsilon_i = 90\%$ and $\beta = 0.05$, we select 500 random samples.

After solving the CC-OPF with A1–A4, we fix the optimal solutions x and test their performance on all 10,000 samples in the support set Ω . For each approach, we re-solve the optimization problem and test the solution ten times, and report the average, maximum, and minimum objective cost, solution reliability (the percent of samples for which the constraints are satisfied when x is fixed), and CPU time.

2.4.4 Results and Solution Patterns

Table 2.1 shows the results for the 9-bus system with no congestion (which occurs when we use the test case line flow limits). We observe that A1 solves the fastest and A2 the slowest. The CPU time taken by A2 also depends on the probability of chance constraint violation as $1 - \epsilon_i = 95\%$ requires many more samples than $1 - \epsilon_i = 90\%$. In contrast, the solution time of A1, A3 and A4 are independent of the probability of chance constraint violation, and only depend upon the problem size. Both A2 and A3 yield higher objective costs than A1. However, A1 results in much lower reliability since the underlying uncertainty is not Gaussian, while A2, A3 and A4 result in reliabilities above the requirements. Note that

Table 2.1: Cost, Reliability, and CPU Time of A1–A4 for the IEEE 9-Bus System with No Congestion

$1 - \epsilon_i =$		A1: Gaussian		A2: Scenario		A3: DR (SDP)		A4: DR (SOCP)	
		95%	90%	95%	90%	95%	90%	95%	90%
Objective cost	avg	4392.63	4330.41	4758.32	4738.73	4875.35	4633.57	4875.41	4633.61
	max	4478.08	4394.57	4895.40	4812.65	5102.61	4789.59	5102.65	4789.62
	min	4308.60	4262.52	4678.17	4649.48	4652.84	4480.48	4652.92	4480.59
Reliability (%)	avg	84.47	75.63	99.65	99.57	99.43	97.45	99.65	97.95
	max	94.07	86.69	99.87	99.83	99.83	99.56	99.83	99.74
	min	65.40	61.98	99.36	99.26	97.60	90.99	98.80	91.94
CPU Time (s)	avg	0.03	0.03	15.21	3.51	0.47	0.46	0.44	0.37
	max	0.05	0.06	15.41	3.85	0.55	0.53	0.36	0.41
	min	0.03	0.02	14.88	3.34	0.34	0.41	0.39	0.34

A3 and A4 perform similarly in terms of cost, reliability, and CPU time. Since, given our choice of γ_1 and γ_2 we have $\mathcal{D}' \subseteq \mathcal{D}$, A4 should result in objective cost values that are less than or equal to those of A3 in each instance. However, in some instances, due to solver limitations for handling SDP models, some results produced with A3 are not fully optimized (i.e., CVX reported “solved/inaccurate”). In contrast, all of the results produced with A4 are fully optimized. Thus, A3 sometimes produces lower costs than A4. When $1 - \epsilon_i = 90\%$, A1 performs substantially worse than when $1 - \epsilon_i = 95\%$, while A2, A3 and A4 yield similar reliability results as they are less sensitive to the change in $1 - \epsilon_i$. In the remainder of the chapter, we use $1 - \epsilon_i = 95\%$, which is more reasonable for power systems.

To put the objective costs in context, we computed the solution to three additional cases. Case 1 assumes no uncertainty, and so we solve a deterministic problem that results in no reserve procurement. Case 2 assumes wind and load uncertainty, but that loads cannot provide reserves, which is consistent with power system operation today. Case 3 assumes that loads can provide reserves, and they are certain (in terms of both consumption and min/max possible consumption) giving us an upper-bound on the cost reductions possible with load reserves. Each is solved with A3. The results for the uncongested system are shown in Table 2.2 along with the comparable results of the complete formulation (referred to as Case 4), i.e., the results of A3 for $1 - \epsilon_i = 95\%$ from Table 2.1. Case 1 has the lowest cost, while Case 2 has the highest. Comparing Case 2 and Case 3 gives the maximum possible value of load reserves, while comparing Case 3 and Case 4 gives the cost of load uncertainty.

We also explore the performance of the approaches under congestion. Specifically, we decrease the line flow limit of the transmission line between Buses 5 and 6 from 150 MW to 40 MW, resulting in congestion on multiple lines. The results are shown in Table 2.3. All

Table 2.2: Objective Costs of Various Uncertainty Cases for the IEEE 9-Bus System with No Congestion ($1 - \epsilon_i = 95\%$, $\forall i$)

		Case 1	Case 2	Case 3	Case 4
Objective cost	avg	4099.97	4891.11	4352.64	4875.35
	max	-	5122.41	4402.08	5102.61
	min	-	4664.12	4292.27	4652.84

Table 2.3: Cost, Reliability, and CPU Time of A1–A4 for the IEEE 9-Bus System with Congestion ($1 - \epsilon_i = 95\%$, $\forall i$)

		A1	A2	A3	A4
Objective cost	avg	4428.72	4884.12	5030.31	5036.35
	max	4538.47	5036.97	5328.71	5330.85
	min	4345.12	4789.28	4767.10	4767.12
Reliability (%)	avg	78.69	99.44	99.27	99.53
	max	91.94	99.79	99.81	99.85
	min	63.52	99.12	97.93	98.43
CPU Time (s)	avg	0.05	48.93	6.40	0.47
	max	0.19	49.44	6.77	0.52
	min	0.02	48.22	6.08	0.42

four approaches yield higher costs as compared to the uncongested cases in Table 2.1. A1’s reliability is poor; its optimal solution can only satisfy the chance constraints in 78.69% of the samples within Ω . In contrast, A2, A3 and A4 achieve higher reliability than required. In each case, the CPU time increases when the system is congested; however, the increase for A4 is small, especially as compare to the increase for A3.

In Fig. 2.1, we plot the optimal generator production and reserve capacities for each generator/load in each approach for $1 - \epsilon_i = 95\%$ for both the uncongested and congested cases. The values plotted are the average over the 10 runs and R^* denotes the sum of the up and down reserve capacities, i.e., $\bar{R} + \underline{R}$. In Table 2.4 we show the optimal reserve procurement given by the three approaches for $1 - \epsilon_i = 95\%$ for both the uncongested (UC) and congested (C) cases. In the column “Generator Reserves” we list the sum of all generator reserves, i.e., $\sum_{j=1}^3 R_{G,j}^*$ and “Load Reserves” are defined similarly. The last column “Total Reserves” is the sum of the generator reserves and load reserves. We can make several observations from Fig. 2.1 and Table 2.4. First, when the system is not congested, the generator production is the same for each approach; however, with congestion the generator production is different for each approach and generators provide more of the total reserves. A1 procures the fewest reserves (corresponding to the lowest cost solution) and

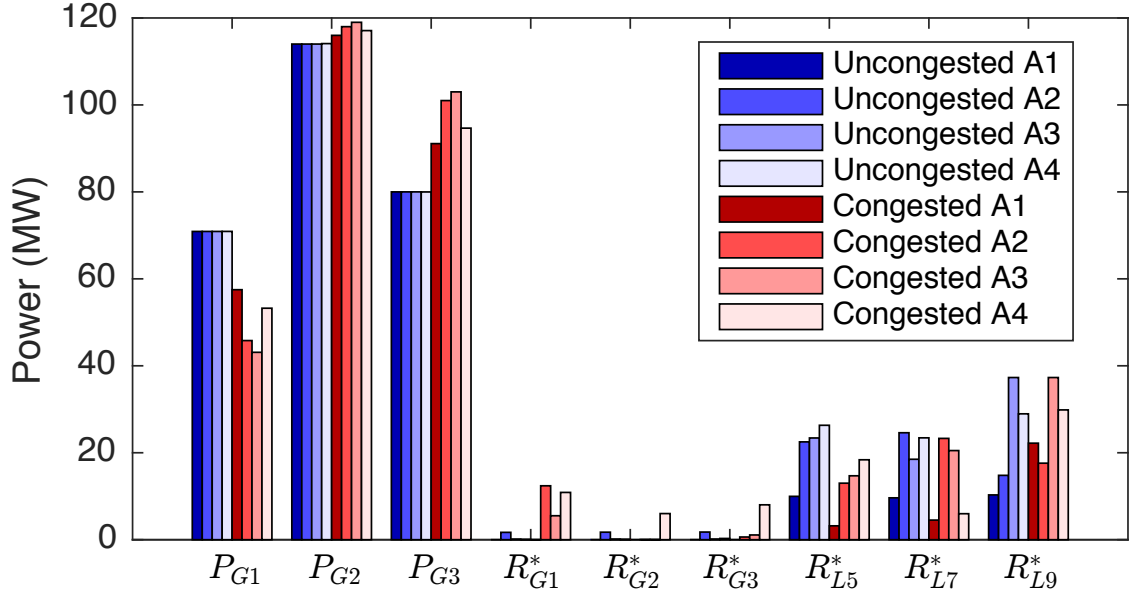


Figure 2.1: Optimal generator production and reserve capacities for each generator/load in the IEEE 9-bus system.

A3/A4 the most (corresponding to the highest cost solution). For each approach, the total reserves procured by the uncongested and congested system is approximately the same.

Table 2.4: Reserves procured by A1–A4 in the IEEE 9-bus system

		Generator Reserves	Load Reserves	Total Reserves
UC	A1	4.7×10^{-5}	29.86	29.86
	A2	5.19	61.89	67.07
	A3	0.34	78.77	79.11
	A4	0.42	78.70	79.12
C	A1	1.2×10^{-4}	29.86	29.86
	A2	13.06	54.01	67.07
	A3	6.63	72.47	79.10
	A4	24.91	54.21	79.12

2.5 Studies on the IEEE 39-bus System and IEEE 118-bus System

2.5.1 Test System and Sample Generation/Selection

We also tested all approaches on the IEEE 39-bus system and both A1 and A4 on the IEEE 118-bus system. We were unable to test A2 or A3 on the 118-bus system as the required CPU times were extremely large. However, it is worth noting that SDP solvers are not yet mature and it is expected that it will become feasible to solve large SDP problems in the near term, which would make A3 more useful.

We obtained test system parameters from MATPOWER (Zimmerman et al., 2011), and again assume $\bar{c}_G = \underline{c}_G = 10 \times \mathbf{1}$, and $\bar{c}_L = \underline{c}_L = 9.8 \times \mathbf{1}$. We add one wind power plant at Bus 6 with rated capacity 300 MW and forecasted output 200 MW. For the 118-bus system, we add three wind power plants at Buses 6, 8, and 15, each with rated capacity 300 MW and forecasted output 200 MW. We treat each wind power injection as an uncorrelated random variable. We assume 95% of the load at each bus is perfectly forecastable and not flexible, and the remaining load is flexible but uncertain, with the load forecasts set equal to the test case loads.

The temperature for each load bus is randomly selected between 10°C and 15°C, and we use the same procedures as used for the 9-bus system to generate 10,000 i.i.d. samples of $\tilde{P}_W, \tilde{P}_L, \underline{\tilde{P}}_L, \bar{\tilde{P}}_L$, which comprise the support set Ω . For A1, A3 and A4, we still use 20 samples randomly picked from the set Ω , but A2 now requires 4000 samples within Ω_{ap} since the number of decision variables increases to $n = 103$. Moreover, the number of chance constraints increases to $m = 216$ as compared to $m = 42$ in the 9-bus system.

Using the test case line flow limits, the system is not congested. For comparison, we also run cases with congestion. Specifically, we decrease the line flow limit of the transmission line between Buses 2 and 3 from 500 MW to 350 MW and between Buses 13 and 14 from 600 MW to 200 MW. To produce congestion in the 118-bus system, we set all line flow limits to 180 MW.

2.5.2 Results and Solution Patterns

We report the objective cost, reliability, and CPU time of approaches A1–A4 for the uncongested system in Table 2.5 and the congested system in Table 2.6. For comparison, the objective costs of Cases 1–4 (as defined in Section 2.4.4) for the uncongested system solved with A3 are in Table 2.7. Figure 2.2 shows the optimal reserve procurement (av-

Table 2.5: Cost, Reliability, and CPU Time of A1–A4 for the IEEE 39-Bus System with No congestion ($1 - \epsilon_i = 95\%$)

		A1	A2	A3	A4
Objective cost	avg	39232.81	40822.56	40214.75	40223.52
	max	39351.18	40952.34	40489.98	40541.92
	min	39060.95	40685.37	39772.05	39764.59
Reliability (%)	avg	85.13	99.89	93.89	99.65
	max	92.28	99.92	97.97	99.82
	min	71.61	99.84	86.34	99.11
CPU Time (s)	avg	0.10	998.11	427.05	0.76
	max	0.12	1036.17	745.54	0.95
	min	0.09	971.84	286.75	0.67

Table 2.6: Cost, Reliability, and CPU Time of A1–A4 for the IEEE 39-Bus System with Congestion ($1 - \epsilon_i = 95\%$)

		A1	A2	A3	A4
Objective cost	avg	42350.49	45489.17	44765.94	44787.62
	max	42788.23	45871.69	45676.41	45744.09
	min	41888.16	45187.54	43996.11	44070.00
Reliability (%)	avg	72.38	99.79	93.25	99.10
	max	86.90	99.87	98.23	99.77
	min	60.47	99.70	85.74	96.60
CPU Time (s)	avg	0.10	1003.45	502.69	0.72
	max	0.12	1040.64	666.80	0.78
	min	0.09	978.80	395.95	0.63

Table 2.7: Objective Costs of Various Uncertainty Cases for the IEEE-39 Bus System with No Congestion ($1 - \epsilon_i = 95\%$, $\forall i$)

		Case 1	Case 2	Case 3	Case 4
Objective cost	avg	38629.05	40259.96	39423.98	40214.75
	max	-	40581.04	39566.85	40489.98
	min	-	39796.61	39217.81	39772.05

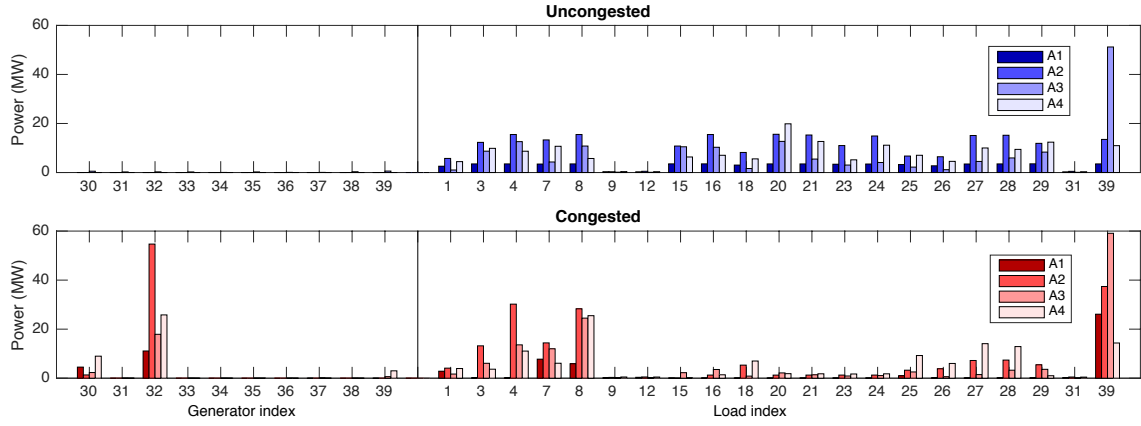


Figure 2.2: Optimal reserve capacities R^* for each generator/load in the IEEE 39-bus system.

Table 2.8: Cost, Reliability, and CPU Time of A1 and A4 for the IEEE 118-Bus System ($1 - \epsilon_i = 95\%$)

		Uncongested		Congested	
		A1	A4	A1	A4
Objective cost	avg	105060	107530	105480	108120
	max	105430	108790	105860	109390
	min	104710	106360	105110	106920
Reliability (%)	avg	29.68	96.57	24.23	95.30
	max	44.62	98.71	40.90	98.63
	min	13.89	92.58	10.11	89.49
CPU Time (s)	avg	7.90	10.71	7.69	11.78
	max	8.69	13.05	8.41	12.76
	min	7.50	9.72	7.36	10.95

Table 2.9: Reserves procured by A1– A4 in the IEEE 39-bus system

		Generator Reserves	Load Reserves	Total Reserves
UC	A1	0.01	61.60	61.61
	A2	0.07	223.76	223.83
	A3	2.64	159.03	161.67
	A4	0.04	162.66	162.70
C	A1	15.59	46.02	61.61
	A2	56.01	167.82	223.83
	A3	21.11	141.04	162.15
	A4	37.83	124.89	162.72

Table 2.10: Reserves procured by A1 and A4 in the IEEE 118-bus system

		Generator Reserves	Load Reserves	Total Reserves
UC	A1	49.62	64.22	113.84
	A4	94.85	206.23	301.07
C	A1	51.65	62.19	113.84
	A4	104.96	196.10	301.06

eraged over the 10 runs) in each approach in the uncongested and congested systems and Table 2.9 summarizes the generator, load, and total reserves procured by each approach in the uncongested and congested systems.

In the 39-bus system, A3 and A4 are *less costly* than A2, since they procure less reserves. Again, A1 is the least expensive, with the best CPU time and the worst reliability. Comparing A3 and A4, A4 achieves higher reliability with significantly lower CPU time and only a slight increase in cost. A2’s reliability is slightly better than A4’s (both are well-above the reliability requirement), but its CPU time is three orders of magnitude larger. It is also worth noting that A2 requires 5583.30 seconds on average to read in data and construct the models (this time is not included in “CPU seconds”) while the average time for A3 and A4 are 298.13 and 21.05 seconds, respectively.

Without congestion, the generator reserves are barely used (they are not clearly visible in the plot, but the totals are listed in Table 2.9). However, with congestion generator reserves, especially from generators 30 and 32 increase. Again, for each approach, the total reserves procured by the system are approximately the same in the uncongested and congested cases.

For the 118-bus system, we report the objective cost, reliability, and CPU time of A1 and A4 for the uncongested and congested system in Table 2.8. Table 2.10 summarizes the generator, load, and total reserves procured by both approaches in the uncongested and

congested systems. In both the uncongested and congested cases, A4 achieves reliabilities above the requirement with modest CPU time (around 10 seconds) demonstrating that the approach should scale to realistically-sized systems. In contrast, A1's reliabilities are well-below the requirement (25-30%, rather than the required 95%). However, the cost of A4's solution is higher, since more reserves are procured, as seen in Table 2.10.

2.6 Concluding Remarks

In this chapter, we proposed a single-period CC-OPF with uncertain reserves from loads. We reformulated the problem using DR optimization and two different ambiguity sets resulting in an SDP problem and an SOCP problem, and compared it to two other reformulations. We conducted a number of computational experiments on the uncongested and congested IEEE 9-bus, 39-bus, and 118-bus systems, and compared the results of the three approaches in terms of objective cost, reliability, CPU, and optimal solution. We find that use of load reserves, even when their reserve capacities are uncertain, decreases system operational costs. We also find that, in contrast to the Gaussian approximation approach, the DR approach yields solutions with reliabilities close to the specified requirements. Additionally, both DR approach require less computation time than the scenario approximation approach, which requires large numbers of uncertainty samples (900 for the 9-bus system and 4000 for the 39-bus system). Furthermore, the DR reformulation that uses SOCP produces solutions with reliabilities above the requirements and requires only modest CPU time (approximately 10 seconds for the IEEE 118-bus system with multiple wind power plants and congestion). In summary, the DR approach, which relies on moments calculated from small uncertainty sample sets (here we use only 20) but makes no assumption on the underlying uncertainty distributions, provides a good trade-off between performance and computational tractability.

CHAPTER 3

Distributionally Robust Appointment Scheduling with Moment-Based Ambiguity Set

3.1 Introductory Remarks

This chapter studies the problem of scheduling a set of appointments with a fixed order of arrivals on a single server. The server does not only refer to a person or a machine but a general service provider which can be an operating room in surgery planning (see, e.g., Denton and Gupta, 2003), an agent in call-center (see, e.g., Gurvich et al., 2010), a computing server in cloud computing data center (see, e.g., Shen and Wang, 2014), or a prototype vehicle in test planning (see, e.g., Shi et al., 2017), depending on specific application settings. The service durations are random and may be correlated. We assign each appointment an arrival time and minimize the expected total waiting time of all the appointments, subject to constrained risk of having server overtime. The traditional stochastic optimization methods require full knowledge of the distribution of uncertain parameters. We employ an ambiguity set of the unknown probability distribution function based on the first and the second moment information. We study a DR formulation that minimizes the worst-case (i.e., maximum) expected waiting time over the ambiguity set, and limits the worst-case overtime risk over the same ambiguity set.

Scheduling under uncertainty has been considered for many applications and the related problems are solved by using simulation, optimization, and approximation algorithms (see, e.g., Begen and Queyranne, 2011; Berg et al., 2014; Epstein et al., 2012; Ge et al., 2013; Klassen and Yoogalingam, 2009; Mittal et al., 2014b). A common goal is to balance appointment waiting, server idleness, and server overtime. Some studies also optimize the sequence of appointments in addition to scheduling their arrival time (Denton and Gupta, 2003; Denton et al., 2007; Mak et al., 2014, 2015; Mittal et al., 2014b). The stochastic optimization literature often assume known distributions of the random service durations (Begen and Queyranne, 2011; Denton and Gupta, 2003; Erdogan and Denton, 2013). On

the other hand, Epstein et al. (2012); Mittal et al. (2014b) focus on robust scheduling model variants and seek “universal” scheduling/sequencing decisions under different uncertain cost structures. We refer to Gabrel et al. (2014) for a thorough review of robust optimization approaches and relevant robust scheduling applications, Pinedo (2016) for a survey of scheduling theories and applications, and Cayirli and Veral (2003); Gupta and Denton (2008) for comprehensive reviews of healthcare scheduling.

Meanwhile, DR approaches have been developed to address the issue of distributional ambiguity in the traditional stochastic programs, by utilizing statistical information of data samples for constructing ambiguity sets of the unknown distributions. We refer to (Delage and Ye, 2010; Zymler et al., 2013) for the representative work that uses moment-based ambiguity sets for optimizing DR expectation-based or chance-constrained models. In particular, by using an ambiguity set based on the first two moments, Jiang and Guan (2016) successfully reformulate DR chance constraints as semidefinite programs (SDPs) based on conic duality.

The issue of distributional ambiguity has received increasing attention in the recent stochastic appointment scheduling literature. For example, both Kong et al. (2013) and Mak et al. (2015) consider DR scheduling problems, and minimize the worst-case expected waiting time of appointments and server overtime. In particular, Kong et al. (2013) consider a cross-moment ambiguity set and derive a copositive programming reformulation. Mak et al. (2015) derive a second-order conic program by only using marginal moments in the ambiguity set. We refer to Deng and Shen (2016); Deng et al. (2016); Qi (2016) for other recent papers that formulate DR models for server planning and/or appointment scheduling by using either moment- or density-based ambiguity sets. Based on the structures of the related reformulations, they discuss continuous or discrete optimization methods for computing the optimal results.

In this problem, we optimize a DR objective that concerns the worst-case expected waiting time, subject to a DR chance constraint for restricting the server overtime. The ambiguity sets in both DR subproblems follow the moment-based form discussed in Jiang and Guan (2016). Different from the previous work, our model incorporates the distributional ambiguity and worst-case analysis in both the objective function of waiting time and the chance constraint of overtime. As the main technical contribution, we utilize special dual structures of the scheduling constraints and also the techniques in Jiang and Guan (2016) to reformulate an SDP approximation of the DR model, which can be solved directly by off-the-shelf solvers. The computational efficacy of our approach depends on the efficiency of solving general continuous SDPs. We demonstrate the effect of distributional ambiguity by comparing the results of our SDP model with the ones of a sampling-based stochas-

tic program on small-scale instances of outpatient treatment scheduling (involving six appointments). One can generalize the study in this chapter to broader service optimization problems with similar structures, e.g., DR inventory control under random demand with unknown distributions (Mak et al., 2014).

The rest of the chapter is organized as follows. Section 3.2 formulates the DR chance-constrained scheduling problem and specifies the ambiguity set. Section 3.3 derives a conservative SDP approximate model of the DR problem. In Section 3.4, we demonstrate the computational results and solution patterns given by the DR approach, and compare them with those of the benchmark stochastic program based on discrete samples of outpatient treatment planning data. We note that the work in this chapter has been published in Zhang et al. (2017).

Assumptions and Notation. We use $|X|$ to denote the cardinality of set X , and use $X \cdot Y$ to denote the Frobenius inner product of X and Y , i.e., $X \cdot Y = \text{tr}(X^T Y)$. We denote \mathbb{S}_+^K as the set of symmetric positive semidefinite $K \times K$ matrices. The generalized inequality for symmetric matrices, $X \succeq Y$, where $X, Y \in \mathbb{S}_+^K$, means that $X - Y \in \mathbb{S}_+^K$. Similarly, $X \preceq Y$ means that $Y - X \in \mathbb{S}_+^K$.

3.2 DR Appointment Scheduling with Chance-constrained Overtime

A set of appointments $1, \dots, m$ sequentially arrive at a single server, for which we plan an arrival time of each appointment. This is equivalent to assigning time intervals x_j between each arrival pair of appointment j and appointment $j + 1$, for all $j = 1, \dots, m - 1$. Let $x = [x_1, \dots, x_{m-1}]^T$. We consider the feasible region:

$$x \in X := \left\{ x : \sum_{j=1}^{m-1} x_j \leq T, x_j \geq 0, \forall j = 1, \dots, m - 1 \right\},$$

where T represents the time limit of operating the server. The service durations of the m appointments are denoted by random vector $s = [s_1, \dots, s_m]^T$ where s_j is the time of serving appointment j , $\forall j = 1, \dots, m$. Instead of knowing the exact distribution of random parameter s , we are given a set of data samples $\{s^k\}_{k=1}^K$, with estimated mean and covariance matrix given by $\mu_0^s = 1/K \sum_{k=1}^K s^k$ and $\Sigma_0^s = 1/K \sum_{k=1}^K (s^k - \mu_0^s)(s^k - \mu_0^s)^T$. Let Ξ^s and f^s be the support and the probability density function of the random parameter s , respectively. We consider the same ambiguity set \mathcal{D}_M^s as the one used in Delage and Ye

(2010); Jiang and Guan (2016), given by

$$\mathcal{D}_M^s = \mathcal{D}_M^s(\Xi^s, \mu_0^s, \Sigma_0^s, \gamma_1, \gamma_2) = \left\{ f^s : \begin{array}{l} \int_{s \in \Xi^s} f^s ds = 1 \\ (\mathbb{E}[s] - \mu_0^s)^\top (\Sigma_0^s)^{-1} (\mathbb{E}[s] - \mu_0^s) \leq \gamma_1 \\ \mathbb{E}[(s - \mu_0^s)(s - \mu_0^s)^\top] \preceq \gamma_2 \Sigma_0^s \end{array} \right\}. \quad (3.1)$$

The constraints in \mathcal{D}_M^s ensures that (i) values of f^s sum to 1 over the support Ξ^s ; (ii) the mean of s lies in an ellipsoid of size proportional to γ_1 centered at the empirical mean μ_0^s ; (iii) the true covariance matrix lies in a positive semidefinite cone bounded by a matrix inequality of $\gamma_2 \Sigma_0^s$. We use parameters γ_1 and γ_2 to control the conservativeness of optimal solutions. In this chapter, we set $\Xi^s = \mathbb{R}^m$.

Define w_j as the waiting time variable of appointment j , for $j = 1, \dots, m$. Without loss of generality, appointment 1 always arrives at time 0, and thus $w_1 = 0$. Let h_j be a unit penalty cost of the waiting time of appointment j , for all $j = 1, \dots, m$. The goal is to minimize the maximum expected waiting penalty, while ensuring that the minimum probability of having no overtime is no less than $1 - \alpha$.

We formulate the DR chance-constrained scheduling problem as

$$\text{[DRCC-S]:} \quad \min_{x \in X} \quad \max_{f^s \in \mathcal{D}_M^s} \mathbb{E}_{f^s} \left[\min_w \sum_{j=2}^m h_j w_j \right] \quad (3.2a)$$

$$\text{s.t.} \quad w_j + x_{j-1} \geq s_{j-1} + w_{j-1}, \quad \forall j = 2, \dots, m \quad (3.2b)$$

$$\inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} \left(\sum_{j=1}^{m-1} x_j + w_m + s_m \leq T \right) \geq 1 - \alpha \quad (3.2c)$$

$$w_1 = 0, \quad w_j \geq 0, \quad \forall j = 2, \dots, m, \quad (3.2d)$$

where the objective function (3.2a) minimizes the maximum expected cost of total waiting time for any distribution $f^s \in \mathcal{D}_M^s$. The arrival times of appointments $j - 1$ and j are $\sum_{i=1}^{j-2} x_i$ and $\sum_{i=1}^{j-1} x_i$, respectively, for all $j = 2, \dots, m$. The starting time of appointment j is calculated as its arrival time plus possible waiting time of appointment j , i.e., $\sum_{i=1}^{j-1} x_i + w_j$, which will be no earlier than completing appointment $j - 1$, at time $\sum_{i=1}^{j-2} x_i + w_{j-1} + s_{j-1}$. Canceling out $\sum_{i=1}^{j-2} x_i$ in the two terms, we have constraints (3.2b) hold for all $j = 2, \dots, m$. The DR chance constraint (3.2c) enforces the worst-case (minimum) probability of finishing all the appointments before the time limit T being no less than $1 - \alpha$ for any $f^s \in \mathcal{D}_M^s$, in which the value of $\sum_{j=1}^{m-1} x_j + w_m + s_m$ represents the time of completing the last appointment m . The waiting time of appointment 1 is set to be zero and all the waiting times are nonnegative according to constraints (3.2d).

3.3 Solution Methods for DRCC-S

In this section, we develop reformulations of the DR objective (3.2a) and the DR chance constraint (3.2c), to solve DRCC-S as a monolithic model. The DRCC-S in model (3.2) involves recourse variables w_j , $\forall j = 1, \dots, m$, which we eliminate during the reformulation process and derive an approximated SDP that only involves continuous decision vector x .

3.3.1 Reformulating the DR objective (3.2a)

To reformulate the DR expectation-based objective (3.2a), we specify the inner problem of DRCC-S, for any given $x \in X$, as follows.

$$\max_{f^s} \int_{s \in \Xi^s} \left(\min_{w \in \Gamma_w(x, s)} \sum_{j=2}^m h_j w_j \right) f^s ds \quad (3.3a)$$

$$\text{s.t.} \quad \int_{s \in \Xi^s} f^s ds = 1 \quad (3.3b)$$

$$\int_{s \in \Xi^s} (s - \mu_0^s)(s - \mu_0^s)^\top f^s ds \preceq \gamma_2 \Sigma_0^s \quad (3.3c)$$

$$\int_{s \in \Xi^s} \begin{bmatrix} \Sigma_0^s & s - \mu_0^s \\ (s - \mu_0^s)^\top & \gamma_1 \end{bmatrix} f^s ds \succeq 0, \quad (3.3d)$$

where in the objective (3.3a), $\Gamma_w(x, s) := \{w \in \mathbb{R}_+^m : (3.2b), (3.2d)\}$ represents the set of feasible w -solutions that satisfy constraints (3.2b) and (3.2d). Constraints (3.3b), (3.3c) and (3.3d) are related to the three constraints in set \mathcal{D}_M^s specified in (3.1). Model (3.3) computes the worst-case (maximum) expected total waiting time for any f^s realization in the ambiguity set \mathcal{D}_M^s .

Denote $\chi_{ij} = \begin{cases} 0 & \text{if } 1 \leq i = j \leq m - 1 \\ \sum_{l=i+1}^j h_l & \text{if } 1 \leq i < j \leq m - 1 \end{cases}$. For each k, j such that $1 \leq k \leq j \leq m - 1$, let vector $\bar{a}^{kj} = \underbrace{[0, \dots, 0]_{k-1}}_{k-1}, \chi_{kj}, \chi_{(k+1)j}, \dots, \chi_{jj}, \underbrace{[0, \dots, 0]_{m-1-j}}_{m-1-j}$. We utilize the one-

to-one correspondence (see Mak et al., 2014, 2015) between partitions of the set $\{1, \dots, m\}$ of appointments and extreme points of the feasible region of appointment scheduling, defined by (3.2b) and (3.2d), to derive equivalent convex reformulation of the inner problem (3.3).

Proposition 1. The inner problem (3.3) has the same optimal objective value as the one of

the following SDP:

$$\min_{\varepsilon, G', H', p', q'} \sum_{i=1}^{m-1} \varepsilon_i + \gamma_2(\Sigma_0^s \cdot G') - (\mu_0^s)^\top G' \mu_0^s + \Sigma_0^s \cdot H' - 2(\mu_0^s)^\top p' + \gamma_1 q' \quad (3.4a)$$

$$\text{s.t.} \quad \begin{bmatrix} G' & -(\frac{1}{2}\bar{a}^{kj} + p' + G' \mu_0^s) \\ -(\frac{1}{2}\bar{a}^{kj} + p' + G' \mu_0^s)^\top & \sum_{i=k}^j \varepsilon_i + \sum_{i=k}^j \chi_{ij} x_i \end{bmatrix} \succeq 0, \forall 1 \leq k \leq j \leq m-1 \quad (3.4b)$$

$$G' \succeq 0, G' \in \mathbb{R}^{m \times m}, \begin{bmatrix} H' & p' \\ (p')^\top & q' \end{bmatrix} \succeq 0, \begin{bmatrix} H' & p' \\ (p')^\top & q' \end{bmatrix} \in \mathbb{R}^{(m+1) \times (m+1)}. \quad (3.4c)$$

Proof. We employ conic duality (see Shapiro, 2001) and a special structure in the dual problem of Model (3.3) to show the above result. First, we associate dual variables $r' \in \mathbb{R}$, $G' \in \mathbb{R}^{m \times m}$, and $\begin{bmatrix} H' & p' \\ (p')^\top & q' \end{bmatrix} \in \mathbb{R}^{(m+1) \times (m+1)}$ with constraints (3.3b), (3.3c) and (3.3d), respectively. We formulate the dual of Model (3.3) as

$$\min_{G', H', p', q', r'} r' + \gamma_2(\Sigma_0^s \cdot G') - (\mu_0^s)^\top G' \mu_0^s + \Sigma_0^s \cdot H' - 2(\mu_0^s)^\top p' + \gamma_1 q' \quad (3.5a)$$

$$\text{s.t.} \quad r' \geq \min_{w \in \Gamma_w(x, s)} \sum_{j=2}^m h_j w_j + 2s^\top (p' + G' \mu_0^s) - s^\top G' s, \forall s \in \mathbb{R}^m \quad (3.5b)$$

$$r' \in \mathbb{R}, G' \succeq 0, \begin{bmatrix} H' & p' \\ (p')^\top & q' \end{bmatrix} \succeq 0. \quad (3.5c)$$

We consider the minimization problem on the right-hand side of the constraint (3.5b):

$$\min_w \left\{ \sum_{j=2}^m h_j w_j : (3.2b), (3.2d) \right\}. \quad (3.6)$$

We associate dual variables δ_i , $i = 1, \dots, m-1$ with constraints (3.2b), and formulate the dual of model (3.6) as

$$\max_{\delta} \sum_{j=1}^{m-1} \delta_j (s_j - x_j) \quad (3.7a)$$

$$\text{s.t.} \quad \delta_{j-1} - \delta_j \leq h_j, \forall j = 2, \dots, m-1 \quad (3.7b)$$

$$\delta_{m-1} \leq h_m \quad (3.7c)$$

$$\delta_j \geq 0, \forall j = 1, \dots, m-1. \quad (3.7d)$$

We use $\Gamma_\delta := \{\delta : (3.7b)-(3.7d)\}$ to denote the dual feasible region, which is independent on the decision x and uncertainty s . According to Mak et al. (2015), the values of extreme points of set Γ_δ have either $\delta_{m-1} = h_m$ or $\delta_{m-1} = 0$, depending on whether $w_m = 0$ or

$w_m > 0$, respectively, which result from the complementary slackness formed between the primal (3.6) and the dual (3.7). We use the value of δ_{m-1} as an incumbent solution, and rearrange inequalities in (3.7b) as $\delta_{j-1} \leq \delta_j + h_j$, $\forall j = 2, \dots, m-1$. Recursively, following the complementary slackness, the optimal value of w_j determines that whether δ_{j-1} takes the upper bound value $\delta_j + h_j$ or the lower bound value 0.

Then, there is a one-to-one correspondence between an extreme point of set Γ_δ and a partition of the set $\{1, \dots, m-1\}$ into intervals. For each interval $\{k, \dots, j\} \subset \{1, \dots, m-1\}$ in the partition, the corresponding extreme-point solution has $\delta_j = 0$, and $\delta_i = \sum_{l=i+1}^j h_l$ for all $i = k, \dots, j-1$. That is, for each interval $\{k, \dots, j\}$ in the partition and $i \in \{k, \dots, j\}$, the value of δ_i is given by χ_{ij} , whose value was specified in Proposition 1.

Thus, define indicating binary variables t_{kj} , $\forall 1 \leq k \leq j \leq m-1$, such that $t_{kj} = 1$ if interval $\{k, \dots, j\}$ belongs to the partition (i.e., $t_{kj} = 1$ if $\delta_i = \chi_{ij}$), and $t_{kj} = 0$ otherwise. The variables t_{kj} represent a valid partition if and only if each element i only belongs to one interval, stated as

$$\sum_{k=1}^i \sum_{j=i}^{m-1} t_{kj} = 1, \quad \forall i = 1, \dots, m-1. \quad (3.8)$$

Denote set $\Gamma_t = \{t_{kj} : (3.8), t_{kj} \geq 0, \forall 1 \leq k \leq j \leq m-1\}$, where we relax the binary variables t_{kj} as continuous variables. Still, any extreme point of the set Γ_t has binary-valued t_{kj} , because the coefficient matrix in (3.8) has a consecutive-ones property and is thus totally unimodular.

Recall constraint (3.5b). We reformulate its right-hand side by the equivalent maximization model (3.7) following the strong duality, yielding:

$$r' \geq \max_{s \in \mathbb{R}^m} \max_{\delta \in \Gamma_\delta} \sum_{i=1}^{m-1} \delta_i (s_i - x_i) + 2s^\top (p' + G' \mu_0^s) - s^\top G' s \quad (3.9a)$$

$$= \max_{\delta \in \Gamma_\delta} \max_{s \in \mathbb{R}^m} \sum_{i=1}^{m-1} \delta_i (s_i - x_i) + 2s^\top (p' + G' \mu_0^s) - s^\top G' s. \quad (3.9b)$$

We can interchange the order of $\max_{\delta \in \Gamma_\delta}$ and $\max_{s \in \mathbb{R}^m}$, because the objective function is bounded with respect to δ and s . We form the relationship between the partitions and extreme points in Γ_δ with the above dual formulation. Moreover, constraint (3.9b) is equivalent to

$$\max_{t \in \Gamma_t} \sum_{k=1}^{m-1} \sum_{j=k}^{m-1} \max_{s \in \mathbb{R}^m} \left(\sum_{i=k}^j \chi_{ij} (s_i - x_i) + 2s^\top (p' + G' \mu_0^s) - s^\top G' s \right) t_{kj}. \quad (3.10)$$

To convert problem (3.10) as a minimization problem, we associate dual variables

ε_i , $i = 1, \dots, m - 1$ with the constraints in Γ_t , and formulate the dual problem as

$$\begin{aligned} \min_{\varepsilon} \quad & \sum_{i=1}^{m-1} \varepsilon_i & (3.11a) \\ \text{s.t.} \quad & \sum_{i=k}^j \varepsilon_i \geq \sum_{i=k}^j \chi_{ij}(s_i - x_i) + 2s^\top(p' + G'\mu_0^s) - s^\top G' s \quad \forall s \in \mathbb{R}^m, \forall 1 \leq k \leq j \leq m-1 & (3.11b) \end{aligned}$$

Note that constraint (3.11b) is equivalent to $\sum_{i=k}^j \varepsilon_i \geq \max_s \sum_{i=k}^j \chi_{ij}(s_i - x_i) + 2s^\top(p' + G'\mu_0^s) - s^\top G' s$, $\forall s \in \mathbb{R}^m$, which can be written as an SDP

$$\begin{bmatrix} G' & -(\frac{1}{2}\bar{a}^{kj} + p' + G'\mu_0^s) \\ -(\frac{1}{2}\bar{a}^{kj} + p' + G'\mu_0^s)^\top & \sum_{i=k}^j \varepsilon_i + \sum_{i=k}^j \chi_{ij}x_i \end{bmatrix} \succeq 0, \quad \forall 1 \leq k \leq j \leq m-1, \quad (3.12)$$

with the values of \bar{a}^{kj} specified in Proposition 1, for each k, j such that $1 \leq k \leq j \leq m-1$. Finally, combining (3.5a), (3.5c), (3.9b), (3.11a), and (3.12), we show that the SDP model (3.4) is equivalent to the inner problem (3.3), and thus complete the proof. \square

Next we combine the reformulation (3.4) with the outer minimization problem in the objective function (3.2a). According to Propositions 2 and 1, we can conclude the following result.

Theorem 3.1. The optimal solution to DRCC-S can be solved via a conservative SDP model

$$\text{[R-DRCC-S]: } \min \{(3.4a) : (3.14a)–(3.14d), (3.4b)–(3.4c), x \in X\}.$$

3.3.2 Reformulating the DR chance constraint (3.2c)

A w -solution in constraint (3.2c) needs to satisfy constraints (3.2b) and (3.2d), which are equivalent to $w_1 = 0$ and $w_j = \max\{0, s_{j-1} - x_{j-1} + w_{j-1}\}$, for $j = 2, \dots, m$. Recursively substituting w_{j-1} in w_j , $j = 2, \dots, m$, we have

$$w_j = \max_{k=1, \dots, j-1} \left\{ \sum_{i=k}^{j-1} (s_i - x_i), 0 \right\} = \max_{k=1, \dots, j} \left\{ \sum_{i=k}^{j-1} (s_i - x_i) \right\},$$

where the last equality follows the convention that $\sum_{i=j}^{j-1} (s_i - x_i) = 0$.

We replace $w_m = \max_{k=1, \dots, m} \left\{ \sum_{i=k}^{m-1} (s_i - x_i) \right\}$ in the DR chance constraint (3.2c),

yielding an equivalent DR joint chance constraint as follows.

$$\begin{aligned}
& \inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} \left(\sum_{j=1}^{m-1} x_j + \max_{k=1, \dots, m} \left\{ \sum_{i=k}^{m-1} (s_i - x_i) \right\} + s_m - T \leq 0 \right) \geq 1 - \alpha \\
\Leftrightarrow & \inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} \left(\max_{k=1, \dots, m} \left\{ \sum_{i=k}^m s_i + \sum_{j=1}^{k-1} x_j - T \right\} \leq 0 \right) \geq 1 - \alpha \\
\Leftrightarrow & \inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} \left(\sum_{i=k}^m s_i + \sum_{j=1}^{k-1} x_j - T \leq 0, \forall k = 1, \dots, m \right) \geq 1 - \alpha. \tag{3.13}
\end{aligned}$$

We derive a reformulation of (3.13) by using a result studied in (Jiang and Guan, 2016) for reformulating and solving general DR joint chance constraints.

Proposition 2. (Jiang and Guan, 2016) The DR variant of a joint chance constraint in (3.13) can be approximated by a conservative SDP as:

$$\cdot G + 1 - r + \Sigma_0^s \cdot H + \gamma_1 q \leq \alpha \lambda \tag{3.14a}$$

$$\begin{bmatrix} G & -p \\ -(p)^\top & 1 - r \end{bmatrix} \succeq \begin{bmatrix} M & \frac{1}{2}(c + 2M\mu_0^s) \\ \frac{1}{2}(c + 2M\mu_0^s)^\top & \lambda + (\mu_0^s)^\top M \mu_0^s + c^\top \mu_0^s + d \end{bmatrix} \tag{3.14b}$$

$$\begin{bmatrix} M & \frac{1}{2}(c - \bar{a}_k) \\ \frac{1}{2}(c - \bar{a}_k)^\top & d + T - \sum_{j=1}^{k-1} x_j \end{bmatrix} \succeq 0 \quad \forall k = 1, \dots, m \tag{3.14c}$$

$$\lambda \geq 0, r \in \mathbb{R}, M \in \mathbb{S}_+^{m \times m},$$

$$\begin{bmatrix} G & -p \\ -(p)^\top & 1 - r \end{bmatrix} \in \mathbb{S}_+^{(m+1) \times (m+1)}, \begin{bmatrix} H & p \\ (p)^\top & q \end{bmatrix} \in \mathbb{S}_+^{(m+1) \times (m+1)}, \tag{3.14d}$$

where $\bar{a}_k = \underbrace{[0, \dots, 0]_{k-1}}_{k-1}, \underbrace{[1, \dots, 1]_{m-k+1}}_{m-k+1}]^\top \in \mathbb{R}^m$, for $k = 1, \dots, m$. (The elements in each \bar{a}_k , $k = 1, \dots, m$ are coefficients of s_1, \dots, s_m in the DR chance constraint (3.13).)

We describe the key idea in (Jiang and Guan, 2016) for proving Proposition 2 as follows. Denote a variable matrix $M \in \mathbb{S}_+^{m \times m}$, a variable vector $c \in \mathbb{R}^m$, and a variable scalar $d \in \mathbb{R}$. We can bound the piecewise linear function $\max_{k=1, \dots, m} \left\{ \sum_{i=k}^m s_i + \sum_{j=1}^{k-1} x_j - T \right\}$ from above by a quadratic function $s^\top M s + c^\top s + d$ for some appropriately assigned values of M , c , and d , such that

$$\inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} (s^\top M s + c^\top s + d \leq 0) \leq \inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} \left(\max_{k=1, \dots, m} \left\{ \sum_{i=k}^m s_i + \sum_{j=1}^{k-1} x_j - T \right\} \leq 0 \right).$$

Therefore, the DR chance constraint $\inf_{f^s \in \mathcal{D}_M^s} \mathbb{P} (s^\top M s + c^\top s + d \leq 0) \geq 1 - \alpha$ implies

the DR joint chance constraint (3.13) (or equivalently (3.2c)). Following conic duality, (Jiang and Guan, 2016) show that (3.14a), (3.14b), and (3.14d) are equivalent to the above DR chance constraint with the quadratic function; (3.14c) ensures that the quadratic function is always above the piecewise linear function. As a result, the SDP model in Proposition 2 serves as a conservative approximation to (3.13) (and to (3.2c)), which indicates that its optimal solution can satisfy (3.2c), but could be too conservative such that the left-hand side of (3.2c) is satisfied at a probability level higher than $1 - \alpha$. The details of the proof of Proposition 2 are given in Jiang and Guan (2016).

3.4 Computational Results

We show the computational results for DRCC-S by solving R-DRCC-S, and compare them with the results of a sampling-based stochastic linear program that minimizes the expected penalty of overtime and waiting time (see, e.g., Denton and Gupta, 2003). The benchmark model considers discrete samples of the random variable s , given by s^1, \dots, s^K , where K is the number of samples, and let p_k be the probability that $s = s^k$, for all $k = 1, \dots, K$, such that $\sum_{k=1}^K p_k = 1$. Given a scheduling solution x , both of the overtime and waiting time depend on the realization of the random service durations s , and therefore, we define $W^k \geq 0$ and w_j^k , $j = 1, \dots, m$ as auxiliary decision variables representing overtime of the server and waiting time of appointments, respectively, for each $k = 1, \dots, K$. We penalize the overtime by a unit cost h_W and penalize each appointment j 's waiting time by a unit cost h_j , $\forall j = 1, \dots, m$. We minimize the expected total penalty cost of waiting and overtime in the following benchmark model.

[Stoch-S]:

$$\min_{\substack{x \in X, w^1, \dots, w^K \\ W^1, \dots, W^K}} \sum_{k=1}^K p_k \left(\sum_{j=2}^m h_j w_j^k + h_W W^k \right) \quad (3.15a)$$

$$\text{s.t.} \quad w_j^k + x_{j-1} \geq s_{j-1}^k + w_{j-1}^k, \quad \forall j = 2, \dots, m, \quad \forall k = 1, \dots, K \quad (3.15b)$$

$$\sum_{j=1}^{m-1} x_j + w_m^k + s_m^k - W^k \leq T, \quad \forall k = 1, \dots, K \quad (3.15c)$$

$$W^k \geq 0, \quad w_1^k = 0, \quad w_j^k \geq 0, \quad \forall j = 2, \dots, m, \quad \forall k = 1, \dots, K \quad (3.15d)$$

The constraints (3.15b), (3.15c), (3.15d) are generalized from the previous constraints (3.2b), (3.2c), (3.2d), respectively. They are specified for each sample k , for all $k = 1, \dots, K$, and compute the values of w_j^k and W^k dependent on the values of x and s^k .

3.4.1 Computational Setup

We test $m = 6$ appointments and set the operating time limit $T = 8$ hours. We use $h_j = 1$ for each appointment j as the unit waiting time cost, and set $h_W = 5$ as the unit overtime cost, used in both the DR model and the benchmark model. The computation consists of two phases: First, we compute optimal solutions of DRCC-S and Stoch-S by using a limited number of scenarios (i.e., training data); second, we fix the solutions in test data which involves a large number of scenarios sampled from the underlying true distribution. We consider an application of outpatient treatment planning, where the true distribution f^s of the random service time s often follows Lognormal (see Gul et al., 2011), and we generate the training data based on two distributional types: Lognormal and Gamma, where the latter indicates a misspecification of the true distribution. In this chapter, we use 20 scenarios for the training data and 10,000 scenarios for the test data. We generate the realized service duration of each appointment in the training/test data by following a given distribution type with both the mean value and standard deviation being 12.5 minutes in all the scenarios. For DRCC-S, we test two overtime risk levels: $\alpha = 5\%$ and 10% . All computational tests are performed on a Windows 7 machine with Intel(R) Core(TM) i7-2600 CPU 3.40 GHz and 8GB memory. We solve all the models by implementing CVX in MATLAB with MOSEK as the optimization solver.

3.4.2 Results and Solution Patterns

We run the first phase of computation under Lognormal and Gamma distributed training data, and depict optimal solution patterns to DRCC-S with $\alpha = 5\%$, 10% , and to Stoch-S in Figure 3.1. For R-DRCC-S, we show the results for the case when $(\gamma_1, \gamma_2) = (1, 2)$.

We summarize the observations as follows. First, under limited data samples and distributional ambiguity, optimal scheduling solutions from the stochastic benchmark model do not demonstrate a “dome” shape as shown in the literature (Denton and Gupta, 2003) when large scale data samples are available. (In a dome shape, the inter-arrival time is shorter for earlier and later appointments, and is longer for the middle ones.) Second, we are able to demonstrate an approximate “dome” shape of the DRCC-S solutions. As $1 - \alpha$ increases in the DR constraint (3.2c), we observe decreased inter-arrival time to prevent overtime.

Next, we conduct the second phase of computation to evaluate the performance of solutions to DRCC-S and to Stoch-S in the test data with 10,000 scenarios. “DR(1,2)” and “DR(0,1)” in Table 3.1 indicate the results for R-DRCC-S with $(\gamma_1, \gamma_2) = (1, 2)$ and $= (0, 1)$, respectively. We compute the solution reliability as the percentage of “overtime-free” scenarios in the test data, and present the results for the cases when training data

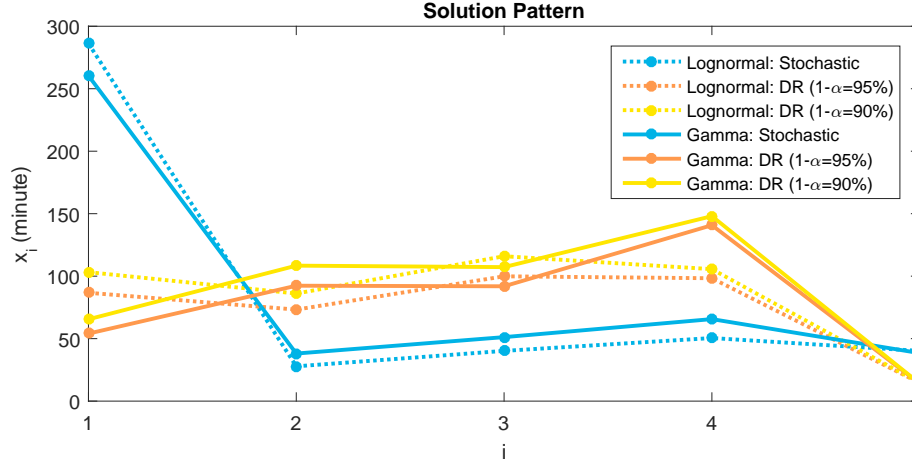


Figure 3.1: Solution patterns with $(\gamma_1, \gamma_2) = (1, 2)$ using Lognormal and Gamma distributed data

follows “Lognormal” or “Gamma”. The reliability results associated with DR(1,2) are all better than those given by Stoch-S. Specifically, when we solve the models with Gamma distributed training data, the reliability results become much better; meanwhile, DR(0,1) could perform worse than the desirable reliability $1 - \alpha = 90\%$, due to that we use smaller γ_1, γ_2 , and thus a smaller, more risk-seeking ambiguity set.

Table 3.1: Reliability of optimal solutions to DRCC-S and Stoch-S models

Approach	$1 - \alpha$	Lognormal	Gamma
Stochastic	N/A	94.44%	89.29%
DR(1,2)	95%	99.67%	99.30%
	90%	97.00%	92.13%
DR(0,1)	95%	96.75%	91.63%
	90%	73.84%	24.80%

In Table 3.2, we also report the overtime and waiting time statistics based on the test data. The average waiting time given by Stoch-S solutions is shorter than that given by DRCC-S solutions, while the standard deviation of waiting time in the latter is smaller. Also, DRCC-S solutions provide shorter overtime, in terms of expectation, standard deviation, and $\geq 50\%$ higher quantiles. This reflects that while trading off between appointment waiting and server overtime, the DR approach emphasizes more on ensuring shorter overtime with sufficiently high probability subject to unknown distributions.

3.5 Concluding Remarks

Table 3.2: Overtime and waiting time statistics (in minute)

		Avg.	Std.	Quantiles			
				50%	75%	90%	99%
Stochastic	Waiting time	2.94	22.34	0.00	0.00	0.00	85.97
	Overtime	1.14	7.43	0.00	0.00	0.00	32.36
DR(1,2) with Lognormal training data with $1 - \alpha = 95\%$	Waiting time	3.66	18.84	0.00	0.63	10.38	49.53
	Overtime	0.12	3.09	0.00	0.00	0.00	0.00

In this chapter, we study the problem of scheduling a set of appointments' arrival time with a preset fixed order of arrivals on a single server. Our DRO problem has distributional ambiguity and worst-case analysis in both the objective function of appointment waiting time and the chance constraint of server overtime. We derived a conservative approximation as SDP which can be directly solved by off-the-shelf solvers. Our computational tests on outpatient treatment scheduling demonstrated that the DRO solutions have an approximate "dome" shape while the solutions by solving the stochastic benchmark model do not have, when only limited data samples are available.

CHAPTER 4

Integer Programming for Distributionally Robust Appointment Scheduling with Random No-shows and Service Durations

4.1 Introductory Remarks

We consider an appointment scheduling problem that involves a single server and a set of appointments following a fixed order of arrivals. A system operator needs to schedule an arrival time for each appointment with random no-shows and service duration. This problem is fundamental for establishing service quality and operational efficiency in many service systems, and has been studied in the context of surgery planning in hospitals (see, e.g., Denton and Gupta, 2003), call-center staffing (see, e.g., Gurvich et al., 2010), and cloud computing server operations (see, e.g., Shen and Wang, 2014). Random no-shows are often observed in outpatient appointment scheduling (e.g., Berg et al., 2014; Lee et al., 2005), which may cause equipment and personnel idleness and losses of opportunities of serving other appointments. As observed by Ho and Lau (1992), random no-shows affect the performance of an appointment system more than random service durations. Additionally, even though many administrative mechanisms exist to reduce the likelihood of no-shows, it is not entirely possible to eliminate no-shows and their negative impacts (see, e.g., Barron, 1980; Moore et al., 2001). In view of these challenges, Cayirli and Veral (2003) point out that a better approach is to adapt the appointment schedule to the anticipated no-shows.

A common goal is to minimize the expected cost associated with appointment waiting time, server idle time, and overtime if the distributional information is fully accessible. In Section 4.1.1, we provide an extensive review of the literature on variants of stochastic appointment scheduling under specific objectives, metrics, and applications. In reality, it is challenging to accurately estimate the probability distribution of no-shows and service durations. The data of no-shows could be limited because of low probability of occurrence

and the heterogeneity of appointments. In view of a wide range of plausible substitutes (e.g., Log-Normal, Normal, and Uniform) for modeling the service-time uncertainty, one could easily misspecify its distribution. Then with ambiguous estimates of no-show and service-duration distributions, we could schedule unnecessarily long (respectively, short) time in between appointments, resulting in significant server idleness (respectively, appointment waiting or server overtime). To address the distributional ambiguity issue, Kong et al. (2013) propose a DR model using a cross-moment ambiguity set that consists of all distributions with common mean and covariance of the random service durations. They obtain a copositive cone programming reformulation and solve a semidefinite program to approximate the optimal results. The most relevant to this chapter, Mak et al. (2015) study a DR model using a marginal-moment ambiguity set of the random service durations. They obtain tractable reformulations by successfully solving a nonconvex optimization problem based on a binary encoding of its feasible region. More importantly, both Kong et al. (2013) and Mak et al. (2015) point out that DR models can yield appointment schedules that (1) perform stably under various probability distributions of service durations and (2) improve the appointment system performance under extreme scenarios. These observations motivate us to study DR models when faced with both no-show and service duration uncertainty.

In this chapter, we generalize the DR appointment scheduling model in Mak et al. (2015) by incorporating heterogeneous no-shows and their distributional ambiguity. We aim to produce appointment schedules with good out-of-sample performance, even only given a few historical data. To the best of our knowledge, this chapter is the first to consider both discrete (no-shows) and continuous (service durations) randomness for DR appointment scheduling. This generalization results in a challenging Mixed-Integer Nonlinear Programming (MINLP) problem, to which the approach by Mak et al. (2015) fails to solve. The main contribution of the chapter is to derive effective integer programming approaches for solving the generalized DR model, including valid inequalities that effectively accelerate the computation of the MINLP (see our computational studies in Section 4.5). We also show that these valid inequalities recover the convex hulls for two important special cases, leading to polynomial-sized Linear Programming (LP) reformulations that are computationally tractable and can be implemented in desktop solvers to benefit practitioners. We note that the work in this chapter has been published in Jiang et al. (2015).

4.1.1 Literature Review

The studies of stochastic appointment scheduling (see, e.g., Berg et al., 2014; Erdogan and Denton, 2013; Gupta and Denton, 2008) often assume uncertain service durations following fully known distributions. Denton and Gupta (2003) formulate a two-stage stochastic LP model for appointment scheduling and demonstrate that the optimal time intervals allocated in between appointments form a “dome shape” if the unit idleness costs are high relative to the unit waiting costs. Mittal et al. (2014b), Begen and Queyranne (2011), Begen et al. (2012), and Ge et al. (2013) develop approximation algorithms for deriving near-optimal solutions to various stochastic or robust appointment scheduling problems. Pinedo (2016) conducts a comprehensive survey of various scheduling problems including their models, theories, and applications.

Ho and Lau (1992) are among the first to take into account no-show uncertainty in scheduling problems. They propose a heuristic approach to double book the first two arrivals and subsequently schedule the remaining appointments. Erdogan and Denton (2013) incorporate no-shows into a stochastic LP model by Denton and Gupta (2003), and also discuss a stochastic dynamic programming variant of the problem. Cayirli and Veral (2003); Hassin and Mendel (2008); Liu et al. (2010); Robinson and Chen (2010) demonstrate the impact of no-shows on static and dynamic appointment scheduling, and discuss general policies to mitigate negative effects such as system idleness. A number of heuristic policies and approximation algorithms have been proposed to schedule appointments under uncertain no-shows (see, e.g., Cayirli et al., 2012; Kong et al., 2018; LaGanga and Lawrence, 2012; Lin et al., 2011; Luo et al., 2012; Muthuraman and Lawley, 2008; Parizi and Ghate, 2015; Zacharias and Pinedo, 2014; Zeng et al., 2010). To our best knowledge, no papers have handled no-shows in a DR framework, which could lead to intractable binary integer programming models due to the discrete nature of 0-1 no-shows (see our computational studies later in Section 4.5.)

In this chapter, we assume a fixed sequence of appointment arrivals. We refer to Denton et al. (2007); Gupta and Denton (2008); He et al. (2015); Mak et al. (2014, 2015); Mancilla (2009) for studies that also involve sequencing decisions, and Denton et al. (2010); Gurvich et al. (2010); Shylo et al. (2012) for studies that optimize server allocation under random service durations. Deng et al. (2016) and Deng and Shen (2016) analyze integrated models for optimizing server allocation, appointment sequencing, scheduling decisions under service time uncertainty, and formulate chance constraints for restricting server overtime use. For DR appointment scheduling, we refer to Kong et al. (2013, 2018); Mak et al. (2015); Zhang et al. (2017). For generic DR optimization using moment-based ambiguity sets, we refer to Bertsimas et al. (2010); Bertsimas and Popescu (2005); Delage and Ye (2010);

Scarf et al. (1958).

4.1.2 Contributions of the Chapter

We summarize the main contributions of this chapter as follows.

1. Depending on system operators' risk preferences, we formulate DR models that incorporate the worst-case expected/CVaR of waiting, idleness, and overtime costs as objective or constraints. Meanwhile, the DR models can flexibly adapt to different prior beliefs of the maximum number of consecutive no-shows, covering from the least conservative case (i.e., no consecutive no-shows) to the most conservative case (i.e., arbitrary no-shows).
2. We develop effective solution approaches for each DR model. The exact reformulations of the DR models result in mixed-integer trilinear programs. We linearize and derive valid inequalities to strengthen the reformulations, which can significantly reduce computational time of solving various instances by using decomposition algorithms. For special no-show beliefs, our derivation leads to polynomial-sized LP reformulations that can readily be implemented in LP solvers.
3. We test diverse instances to show the computational efficacy and demonstrate the performance of DR models under various uncertainties and levels of conservativeness. We provide guidelines for choosing appropriate DR models and no-show beliefs, depending on historical no-show rates, computation budget, and targeted tradeoffs between quality of service and operational cost.

4.1.3 Structure of the Chapter

The remainder of the chapter is organized as follows. In Section 4.2, we formulate the DR expectation/CVaR models and their variants based on different risk preferences. In Section 4.3, we derive an MINLP of the DR expectation model, as well as valid inequalities for accelerating a generic cutting-plane algorithm. In Section 4.4, we derive polynomial-sized LP reformulations for special cases of no-show beliefs. In Section 4.5, we test various instances to demonstrate the computational efficacy and solution performance of different DR models. In Section 4.6, we summarize the chapter and provide future research directions. In the e-companion (EC), we describe models and approaches for problems under a general waiting-time cost and a DR CVaR setting, respectively. We also present all the proofs there.

Notation: The convex hull of a set X is denoted by $\text{conv}(X)$. The abbreviation “w.l.o.g.” represents “without loss of generality.” We follow the convention that $\sum_{k=i}^j a_k = 0$ if $i > j$.

4.2 Formulations of DR Appointment Scheduling

We consider n appointments arriving at a single server following a fixed order of arrivals given as $1, \dots, n$. Each appointment i has a random service duration s_i . We interpret the possibility of random no-show for appointment i by a 0-1 Bernoulli random variable q_i such that $q_i = 1$ if appointee i shows up, and $q_i = 0$ otherwise. The goal is to optimize the decision of scheduling an arrival time for each appointment, or equivalently, assigning time intervals between appointments i and $i + 1$ for all $i = 1, \dots, n - 1$.

4.2.1 Modeling Waiting, Idleness, and Overtime under Uncertainty

Let variable x_i represent the scheduled time interval between appointments i and $i + 1$, $\forall i = 1, \dots, n - 1$. Under random no-shows and service durations, one or multiple of the following three scenarios can happen: (i) an appointment cannot start on time due to overtime operations of previous appointments, (ii) the server is idle and waiting for the next appointment due to an early finish or no-shows of previous appointments, and (iii) the server cannot finish serving all appointments within a given time limit, denoted by T . For all $i = 1, \dots, n$, let variable w_i represent the waiting time of appointment i , and variable u_i represent the server idle time after finishing appointment i . Also, let variable W represent the server’s overtime beyond the fixed time limit T to finish all n appointments. The feasible region of decision x is defined as

$$X = \left\{ x : x_i \geq 0, \forall i = 1, \dots, n, \sum_{i=1}^n x_i = T \right\}, \quad (4.1)$$

to ensure that we assign nonnegative time in between all consecutive appointments, and appointment n is scheduled to arrive before the end of the service horizon T . The dummy variable $x_n \geq 0$ represents $T - \sum_{i=1}^{n-1} x_i$, i.e., the time from the scheduled start of the last appointment to the server time limit.

Given decision $x \in X$ and a realization of uncertain parameters (q, s) , the appointment waiting time $w = [w_1, \dots, w_n]^\top$ and server idleness $u = [u_1, \dots, u_n]^\top$ are given by

$$w_i = \max\{0, q_{i-1}s_{i-1} + w_{i-1} - x_{i-1}\}, \text{ and } u_{i-1} = \max\{0, x_{i-1} - q_{i-1}s_{i-1} - w_{i-1}\}, \forall i = 2, \dots, n. \quad (4.2)$$

We denote nonnegative parameters c_i^w , c_i^u , and c^o as unit penalty costs of waiting, idleness of appointment i , and server overtime, respectively, which satisfy $c_{i+1}^u - c_i^u \leq c_{i+1}^w$ for all $i = 1, \dots, n-1$. This assumption is standard (see Denton and Gupta (2003); Ge et al. (2013); Kong et al. (2013); Mak et al. (2015)). In fact, if this assumption does not hold and $c_{i+1}^u > c_i^u + c_{i+1}^w$ for some i , then the system operator would rather enforce idleness even if appointment $i+1$ has arrived and keep it waiting, which is not realistic due to practical concerns. Under this assumption, we can formulate a linear program to compute the total cost of waiting, idleness, and overtime for given x, q, s :

$$Q(x, q, s) := \min_{w, u, W} \sum_{i=1}^n (c_i^w w_i + c_i^u u_i) + c^o W \quad (4.3a)$$

$$\text{s.t.} \quad w_i - u_{i-1} = q_{i-1} s_{i-1} + w_{i-1} - x_{i-1} \quad \forall i = 2, \dots, n \quad (4.3b)$$

$$W - u_n = q_n s_n + w_n - x_n \quad (4.3c)$$

$$w_i \geq 0, w_1 = 0, u_i \geq 0, W \geq 0, \quad \forall i = 1, \dots, n. \quad (4.3d)$$

The objective function (4.3a) minimizes a linear cost function of the waiting, idleness, and overtime. Constraints (4.3b) yield either the waiting time of appointment i or the server's idle time after finishing appointment $i-1$, both of which will have the same solutions values as given by (4.2) (see Proposition 1 in Ge et al. (2013)). Similarly, constraint (4.3c) yields either the over time W or the idle time u_n . Since appointment 1 always arrives at time 0, we have $w_1 = 0$ and all the waiting, idleness, and overtime variables are nonnegative according to constraints (4.3d).

In (4.3), note that the waiting time costs $c_i^w w_i$ are modeled from the perspective of servers (e.g., operating rooms). In particular, we assume that appointment no-shows take place after the server has been set up for serving the appointments. Hence, the waiting time costs stem from equipment and personnel idleness, as well as from the losses of opportunities of serving other appointments, and they are incurred regardless whether the appointments show up. From the perspective of appointments, the waiting time costs should be modeled as $c_i^w w_i q_i$, i.e., the waiting time costs are waived if an appointment does not show up. In this chapter, we focus on the DR models and solution methods for the former case, i.e., server-based waiting time costs. In A.1, we elaborate how our DR approaches can adapt for a more general setting that incorporates both server-based and appointment-based waiting time costs.

4.2.2 Supports and Ambiguity Set

The classical stochastic appointment scheduling approaches seek an optimal $x \in X$ to minimize the expectation of random cost $Q(x, q, s)$ subject to uncertainty (q, s) with a known joint probability distribution denoted as $\mathbb{P}_{q,s}$. We assume that $\mathbb{P}_{q,s}$ is only known belonging to an ambiguity set $\mathcal{F}(D, \mu, \nu)$ that is determined by the support D of (q, s) and the mean values $\mu = [\mu_1, \dots, \mu_n]^\top$ and $\nu = [\nu_1, \dots, \nu_n]^\top$, where μ_i represents the mean $\mathbb{E}[s_i]$, and ν_i represents the average show-up probability $\mathbb{E}[q_i]$ of appointment i for each $i = 1, \dots, n$. We consider support $D = D_q \times D_s$ where D_q is the support of random no-show parameter q and D_s is the support of random service duration parameter s . We assume upper and lower bounds of the duration of each appointment i and accordingly

$$D_s := \{s \geq 0 : s_i^l \leq s_i \leq s_i^u, \forall i = 1, \dots, n\}.$$

The full support $D_q = \{q : q \in \{0, 1\}^n\}$ contains all no-show scenarios. However, it often leads to over-conservative schedules. In this chapter, we parameterize the no-show support by an integer $K \in \{2, \dots, n+1\}$ such that $D_q = D_q^{(K)}$ rules out consecutive no-shows in any K consecutive appointments. Accordingly,

$$D_q^{(K)} := \left\{ q \in \{0, 1\}^n : \sum_{j=i}^{i+K-1} q_j \geq 1, \forall i = 1, \dots, n - K + 1 \right\}.$$

Note that (i) when $K = 2$, $D_q^{(2)}$ rules out all consecutive no-shows, and (ii) when $K = n+1$, we have $D_q^{(n+1)} = \{0, 1\}^n = D_q$ as the full support. Also, the parameterized supports $D_q^{(2)} \subset D_q^{(3)} \subset \dots \subset D_q^{(n+1)}$ form a spectrum of conservativeness levels, with $D_q^{(2)}$ being the least conservative and $D_q^{(n+1)}$ being the most general/conservative. In practice, the system operator has the flexibility to select parameter K according to her targeted conservativeness, regardless whether the ruled-out realizations may still occur. The conservativeness refers to the trade-off between optimality and robustness (see, e.g., Ben-Tal and Nemirovski, 2000; Bertsimas and Sim, 2004). If we select $K = n+1$, then support $D_q^{(n+1)} \equiv \{0, 1\}^n$ contains all possible no-show scenarios and so is most robust. Meanwhile, $D_q = D_q^{(n+1)}$ leads to the largest value of $\sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x, q, s)]$ among all $K \in \{2, \dots, n+1\}$. In this sense, $D_q^{(n+1)}$ is the most conservative. On the contrary, $D_q^{(2)}$ is the least conservative because it leads to the smallest values of $\mathbb{P}\{q \in D_q^{(K)}\}$ and $\sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x, q, s)]$ among all $K \in \{2, \dots, n+1\}$.

Although $D_q^{(K)}$ with $K \neq n+1$ does not contain all possible no-shows, we can select a value of K such that $\mathbb{P}\{q \in D_q^{(K)}\}$ exceeds a sufficiently high probability. In Section

4.2.4, we provide a practical and rigorous guideline for how to select the value of K . More importantly, $D_q^{(K)}$ can lead to better out-of-sample performance. For example, the DR schedules given by $D_q^{(2)}$ outperform those using $D_q^{(n+1)}$ in the out-of-sample simulations, in which arbitrary consecutive no-shows may still happen (see Section 4.5.4).

We specify the ambiguity set $\mathcal{F}(D, \mu, \nu)$ as

$$\mathcal{F}(D, \mu, \nu) := \left\{ \mathbb{P}_{q,s} \geq 0 : \begin{array}{l} \int_{D_q \times D_s} d\mathbb{P}_{q,s} = 1 \\ \int_{D_q \times D_s} s_i d\mathbb{P}_{q,s} = \mu_i \quad \forall i = 1, \dots, n \\ \int_{D_q \times D_s} q_i d\mathbb{P}_{q,s} = \nu_i \quad \forall i = 1, \dots, n \end{array} \right\}, \quad (4.4)$$

where $\mathbb{P}_{q,s}$ matches the mean values of service durations and no-shows. The ambiguity set $\mathcal{F}(D, \mu, \nu)$ does not incorporate higher moments (e.g., variance and correlations) of service time and no-shows for several reasons. First, with a small amount of data, it is often unclear whether/how the service time and no-shows are correlated. Second, the introduction of higher moments undermines the computational tractability of the DR models, which can be achieved by using $\mathcal{F}(D, \mu, \nu)$ and the solution algorithm derived later. Finally, as we find in the computational study (see Section 4.5), the DR models based on $\mathcal{F}(D, \mu, \nu)$ already provide near-optimal results, and the benefit of incorporating higher moments is not significant in our case.

4.2.3 DR Models with Different Risk Measures

We consider DR appointment scheduling models that impose a min-max DR objective and/or DR constraints. Specifically, given $x \in X$, we consider a risk measure ϱ of $Q(x, q, s)$ where

1. a *risk-neutral* system operator sets $\varrho(Q(x, q, s)) = \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s)]$, i.e., the expected total cost of waiting, idleness, and overtime;
2. a *risk-averse* system operator sets $\varrho(Q(x, q, s)) = \text{CVaR}_{1-\epsilon}(Q(x, q, s))$, i.e., the CVaR of the total cost with $1 - \epsilon \in (0, 1)$ confidence.

Then, the DR models impose a generic min-max DR objective in the form

$$\min_{x \in X} \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \varrho(Q(x, q, s)), \quad (4.5a)$$

and/or generic DR constraints in the form

$$\sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \varrho(Q(x, q, s)) \leq \bar{Q}. \quad (4.5b)$$

where $\bar{Q} \in \mathbb{R}$ represents a bounding threshold for the risk measure from above. Both DR objective (4.5a) and constraints (4.5b) protect the risk measure by hedging against all probability distributions in $\mathcal{F}(D, \mu, \nu)$. A DR model can impose either or both of DR objective (4.5a) and constraints (4.5b), and can use either expectation or CVaR as risk measures in (4.5a)–(4.5b), i.e., $\varrho(Q(x, q, s)) = \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x, q, s)]$ or $\varrho(Q(x, q, s)) = \text{CVaR}_{1-\epsilon}(Q(x, q, s))$. Furthermore, the system operator can tune the cost parameters c_i^w , c_i^u , and c^o to let $Q(x, q, s)$ represent different consequences (e.g., performance metric, quality of service, resource opportunity cost, etc.) associated with waiting, idleness, and overtime in (4.5a)–(4.5b). For example, by setting $c^o = 1$ and $c_i^u = c_i^w = 0$ for all i , we use

$$\sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \text{CVaR}_{1-\epsilon}(Q(x, q, s)) \leq \bar{Q} \quad (4.6)$$

to constrain the CVaR of overtime W below threshold \bar{Q} . The CVaR constraints provide a safe guarantee on the performance metrics with high probabilities. For this particular cost parameter setting, constraint (4.6) guarantees that $\inf_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{P}_{q,s} \{W \leq \bar{Q}\} \geq 1 - \epsilon$, i.e., the overtime W is controlled under threshold \bar{Q} with the smallest possible probability being no less than $1 - \epsilon$. This provides an appropriate “end-of-the-day” guarantee (see, e.g., Shylo et al., 2012; Wang et al., 2014; Zhang et al., 2015b). For presentation brevity, we have relegated the discussions on the CVaR-based model to A.2.

4.2.4 Guideline of Selecting Parameter K

In practice, a system operator may evaluate the probability of the random variables $q = (q_1, \dots, q_n)$ belonging to $D_q^{(K)}$, i.e., $\mathbb{P}(q \in D_q^{(K)})$. Then, she can select a value of K such that $\mathbb{P}(q \in D_q^{(K)})$ exceeds a given threshold such as 90%. To this end, she can gradually increase K from 2 until $\mathbb{P}(q \in D_q^{(K)})$ exceeds the threshold for the first time.

Observation 4.1. $\mathbb{P}(q \in D_q^{(K)}) = 1$ if $K > n$. If $2 \leq K \leq n$ and the components of q are jointly independent, then $\mathbb{P}(q \in D_q^{(K)}) = 1 - [Q^n]_{1, m(K)+1}$, where $m(i) := \frac{1}{2}i(2n - i + 3)$ for $i = 0, \dots, K$ and Q represents a $(m(K) + 1) \times (m(K) + 1)$ matrix such that (1) $Q_{m(i)+j, i+j+1} = \nu_{i+j}$ for all $i = 0, \dots, K - 1$ and $j = 1, \dots, n - i$, (2) $Q_{m(i)+j, m(i+1)+j} = 1 - \nu_{i+j}$ for all $i = 0, \dots, K - 2$ and $j = 1, \dots, n - i$, (3) $Q_{m(K-1)+j, m(K)+1} = 1 - \nu_{K-1+j}$ for all $j = 1, \dots, n - K + 1$, and (4) $Q_{m(K)+1, m(K)+1} = Q_{m(i), m(i)} = 1$ for all $i = 1, \dots, K$.

Proof. Proof of Observation 4.1 If $K > n$, it is clear that we cannot have K consecutive no-shows and so $\mathbb{P}(q \in D_q^{(K)}) = 1$. If $2 \leq K \leq n$, we construct a Markov chain with $m(K) + 1$ states, where state $m(i) + j$ represents i consecutive no-shows in the first $i + j - 1$ appointments for all $i = 0, \dots, K - 1$ and $j = 1, \dots, n - i + 1$, and state $m(K) + 1$ represents that K consecutive no-shows happen. By construction, matrix Q is the one-step transition matrix of this Markov chain, where states $m(K) + 1$ and $m(i)$ for all $i = 1, \dots, K$ are absorbing. Thus, $[Q^n]_{1, m(K) + 1}$, representing component $(1, m(K) + 1)$ of matrix Q^n , is equal to the probability of having K consecutive no-shows in the n appointments. \square

Using Observation 4.1, the selection of K can be conveniently done in a spreadsheet. In Figure 4.1, we display an example of $\mathbb{P}(q \in D_q^{(K)})$ with $n = 10$, $K = 1, \dots, 11$, and $\nu_i = 0.1, \dots$, or 0.9 for all $i = 1, \dots, 10$. We observe that $K = 2$ is sufficient for $\mathbb{P}(q \in D_q^{(K)}) \geq 90\%$ when $\nu_i \leq 0.1$, i.e., when the no-show probability for each appointment is no greater than 0.1 . This observation motivates us to select $D_q = D_q^{(2)}$ for scheduling appointments with low no-show probabilities.

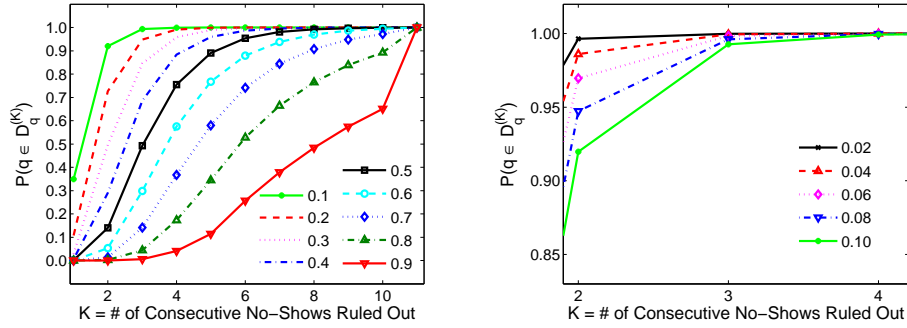


Figure 4.1: An example of $\mathbb{P}(q \in D_q^{(K)})$ for $n = 10$ appointments

Next, we develop reformulations and solution methods. For presentation brevity, we only analyze DR expectation models in Sections 4.3 and 4.4. We present the results of DR CVaR models in A.2. All the proofs are organized in A.3 (for DR expectation models) and A.4 (for DR CVaR models).

4.3 Cutting-Plane Approach and Valid Inequalities for DR Expectation Models

We analyze the DR expectation models by specifying a generic objective form (4.5a) as

$$\min_{x \in X} \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s)], \quad (4.7)$$

which minimizes the worst-case expected cost of waiting, idleness, and overtime. We first consider the inner maximization problem $\sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s)]$ for a fixed $x \in X$, where $\mathbb{P}_{q,s}$ is the decision variable. It can be detailed as a linear functional optimization problem

$$\max_{\mathbb{P}_{q,s} \geq 0} \int_{D_q \times D_s} Q(x, q, s) d\mathbb{P}_{q,s} \quad (4.8a)$$

$$\text{s.t.} \quad \int_{D_q \times D_s} s_i d\mathbb{P}_{q,s} = \mu_i \quad \forall i = 1, \dots, n \quad (4.8b)$$

$$\int_{D_q \times D_s} q_i d\mathbb{P}_{q,s} = \nu_i \quad \forall i = 1, \dots, n \quad (4.8c)$$

$$\int_{D_q \times D_s} d\mathbb{P}_{q,s} = 1, \quad (4.8d)$$

where $D_q = D_q^{(K)}$ for some $K \in \{2, \dots, n+1\}$. Letting ρ_i , γ_i , and θ be dual variables associated with constraints (4.8b), (4.8c), and (4.8d), respectively, we present problem (4.8) in its dual form as

$$\min_{\rho \in \mathbb{R}^n, \gamma \in \mathbb{R}^n, \theta \in \mathbb{R}} \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \quad (4.9a)$$

$$\text{s.t.} \quad \sum_{i=1}^n s_i \rho_i + \sum_{i=1}^n q_i \gamma_i + \theta \geq Q(x, q, s) \quad \forall (q, s) \in D_q \times D_s. \quad (4.9b)$$

Here $\rho = [\rho_1, \dots, \rho_n]^\top$, $\gamma = [\gamma_1, \dots, \gamma_n]^\top$, and θ are unrestricted, and (4.9b) are associated with primal variables $\mathbb{P}_{q,s}$, $\forall (q, s) \in D_q \times D_s$. Under the standard assumptions that μ_i belongs to the interior of set $\{\int_{D_q \times D_s} s_i d\mathbb{Q}_{q,s} : \mathbb{Q}_{q,s} \text{ is a probability distribution over } D_q \times D_s\}$, and that ν_i belongs to the interior of set $\{\int_{D_q \times D_s} q_i d\mathbb{Q}_{q,s} : \mathbb{Q}_{q,s} \text{ is a probability distribution over } D_q \times D_s\}$ for each appointment i , strong duality holds between (4.8) and (4.9). Furthermore, for

a fixed (ρ, γ, θ) , constraints (4.9b) are equivalent to

$$\theta \geq \max_{(q,s) \in D_q \times D_s} \left\{ Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma q_i) \right\}.$$

Thus, due to the objective of minimizing θ , the dual formulation (4.9) is equivalent to

$$\min_{\rho \in \mathbb{R}^n, \gamma \in \mathbb{R}^n} \left\{ \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \max_{(q,s) \in D_q \times D_s} \left\{ Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \right\}. \quad (4.10)$$

4.3.1 MINLP Reformulation and a Generic Cutting-Plane Approach

Note that $Q(x, q, s)$ is a minimization problem and thus in (4.10) we have an inner max-min problem. We next analyze the structure of $Q(x, q, s)$ for given solution x and realized value of (q, s) . We formulate $Q(x, q, s)$ in (4.3) in its dual form as

$$Q(x, q, s) = \max_y \sum_{i=1}^n (q_i s_i - x_i) y_i \quad (4.11a)$$

$$\text{s.t. } y_{i-1} - y_i \leq c_i^w \quad \forall i = 2, \dots, n \quad (4.11b)$$

$$-y_i \leq c_i^u \quad \forall i = 1, \dots, n \quad (4.11c)$$

$$y_n \leq c^o, \quad (4.11d)$$

where variable y_{i-1} represents the dual associated with each constraint i in (4.3b) for all $i = 2, \dots, n$, and variable y_n represents the dual of constraint (4.3c). Constraints (4.11b), (4.11c), (4.11d) are related to primal variables w_i , $i = 2, \dots, n$, u_i , $i = 1, \dots, n$, and W in (4.3), respectively. Therefore, formulation (4.10) is equivalent to

$$\min_{\rho, \gamma} \left\{ \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \max_{(q,s) \in D_q \times D_s} \left\{ Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \right\} \quad (4.12a)$$

$$= \min_{\rho, \gamma} \left\{ \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \max_{y \in Y} h(x, y, \rho, \gamma) \right\}, \quad (4.12b)$$

where Y represents the feasible region of variable y in (4.11) given by (4.11b)–(4.11d), and

$$h(x, y, \rho, \gamma) := \max_{(q,s) \in D_q \times D_s} \left\{ \sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\}. \quad (4.12c)$$

The derivation of $h(x, y, \rho, \gamma)$ follows that we can interchange the order of $\max_{(q,s) \in D_q \times D_s}$ and $\max_{y \in Y}$ in (4.12a). Combining the inner problem in the form of (4.12b) with the outer

minimization problem in (4.7), we derive a reformulation of the DR expectation model (4.7) as:

$$\min_{x \in X, \rho, \gamma, \delta} \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \delta \quad (4.13a)$$

$$\text{s.t. } \delta \geq \max_{y \in Y} h(x, y, \rho, \gamma) \equiv \max_{y \in Y, (q, s) \in D_q \times D_s} \left\{ \sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\}. \quad (4.13b)$$

Next, we analyze structural properties of $\max_{y \in Y} h(x, y, \rho, \gamma)$ as a function of variables x , ρ , and γ .

Lemma 4.1. For any fixed variables x , ρ , and γ , $\max_{y \in Y} h(x, y, \rho, \gamma) < +\infty$. Furthermore, function $\max_{y \in Y} h(x, y, \rho, \gamma)$ is convex and piecewise linear in x , ρ , and γ with a finite number of pieces.

We refer to A.3.1 for a detailed proof. Lemma 4.1 indicates that constraint (4.13b) essentially describes the epigraph of a convex and piecewise linear function of decision variables in model (4.13). This observation facilitates us applying a separation-based decomposition algorithm to solve model (4.13) (or equivalently, the DR expectation model (4.7)), presented in Algorithm 4.1. This algorithm is finite because we identify a new piece of the function $\max_{y \in Y} h(x, y, \rho, \gamma)$ each time when the set $\{L(x, \rho, \gamma, \delta) \geq 0\}$ is augmented in Step 7, and the function has a finite number of pieces according to Lemma 4.1.

The main difficulty of the above decomposition algorithm lies in solving the separation problem (4.14). In general, this problem is a mixed-integer trilinear program because of the integrality restrictions on variables q_i and the trilinear terms $q_i s_i y_i$ in the objective function. This creates obstacles for optimally solving the separation problem if presented in its current form. In Section 4.3.2, we linearize and reformulate the separation problem (4.14) as a Mixed-Integer Linear Programming (MILP) that can readily be solved by optimization solvers. Moreover, we will derive valid inequalities to strengthen this MILP, and test their computational efficiency later.

4.3.2 MILP Reformulation of the Separation Problem and Valid Inequalities

Our approach is inspired by Mak et al. (2015), where the authors point out that an optimal solution y^* to a similar separation problem but not involving no-shows, exists at an

Algorithm 4.1: A decomposition algorithm for solving DR expectation model (4.7).

Input: *feasible regions* X, Y , and $D_q \times D_s$; *set of cuts* $\{L(x, \rho, \gamma, \delta) \geq 0\} = \emptyset$
1 *Solve the master problem*

$$\begin{aligned} \min_{x \in X, \rho, \gamma, \delta} \quad & \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \delta \\ \text{s.t.} \quad & L(x, \rho, \gamma, \delta) \geq 0 \end{aligned}$$

and record an optimal solution $(x^*, \rho^*, \gamma^*, \delta^*)$.
2 *With* (x, ρ, γ) *fixed to be* (x^*, ρ^*, γ^*) , *solve the separation problem*

$$\max_{y \in Y} h(x, y, \rho, \gamma) \equiv \max_{y \in Y, (q, s) \in D_q \times D_s} \left\{ \sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \quad (4.14)$$

and record an optimal solution (y^*, q^*, s^*) .
3 **if** $\delta^* \geq \sum_{i=1}^n (q_i^* s_i^* - x_i^*) y_i^* - \sum_{i=1}^n (\rho_i^* s_i^* + \gamma_i^* q_i^*)$ **then**
4 | *stop and return* x^* *as an optimal solution to formulation* (4.7).
5 **else**
6 | *add the cut* $\delta \geq \sum_{i=1}^n (q_i^* s_i^* - x_i) y_i^* - \sum_{i=1}^n (s_i^* \rho_i + q_i^* \gamma_i)$ *to the set of cuts*
 $\{L(x, \rho, \gamma, \delta) \geq 0\}$ *and go to Step 2.*
7 **end**

extreme point of polyhedron Y . They then successfully decompose the separation problem by appointment for each $i = 1, \dots, n$ and reformulate it by using the extreme points of Y . Different in this chapter, for fixed x, ρ , and γ , the separation problem is a mixed-integer trilinear program involving binary variables $q_i, i = 1, \dots, n$. Moreover, except for the case $D_q = D_q^{(n+1)} = \{0, 1\}^n$, $h(x, y, \rho, \gamma)$ is not decomposable by appointment in view of the cross-appointment nature of D_q . Therefore, the approach in Mak et al. (2015) is no longer applicable, and $\max_{y \in Y} h(x, y, \rho, \gamma)$ becomes much more challenging.

Our analysis consists of the following steps. We start by showing the convexity of $h(x, y, \rho, \gamma)$ in variable y . Then, it follows from fundamental convex analysis that maximizing convex function $h(x, y, \rho, \gamma)$ on polyhedron Y will yield an optimal solution at one of the extreme points of Y . Also considering the cost of idleness, we extend the result of extreme-point representation in Mak et al. (2015) and reformulate the separation problem (4.14) using a polynomial number of binary variables to replace the continuous variables $y_i, i = 1, \dots, n$.

Lemma 4.2. For fixed x, ρ , and γ , function $h(x, y, \rho, \gamma)$ is convex in variable y .

We refer to A.3.2 for a proof. According to Lemma 4.2, an optimal solution y^* to the separation problem (4.14) exists at one of the extreme points of Y having linear constraints

(4.11b)–(4.11d). Consider

$$Y = \left\{ c^o \geq y_n \geq -c_n^u, y_n + c_n^w \geq y_{n-1} \geq -c_{n-1}^u, \dots, y_2 + c_2^w \geq y_1 \geq -c_1^u \right\}. \quad (4.15)$$

It can be observed that any extreme point \hat{y} of Y satisfy (i) either $\hat{y}_n = -c_n^u$ or $\hat{y}_n = c^o$, and (ii) for all $i = 1, \dots, n-1$, dual constraint $\hat{y}_{i+1} + c_{i+1}^w \geq \hat{y}_i \geq -c_i^u$ is binding at either the lower bound or the upper bound.

This observation motivates us to establish an alternative formulation of (4.14) using new binary variables. For notation convenience, we define a dummy variable y_{n+1} , which always takes the lower-bound value $-c_{n+1}^u := 0$. There is a one-to-one correspondence between an extreme point of Y and a partition of the integers $1, \dots, n+1$ into intervals. For each interval $\{k, \dots, j\} \subseteq \{1, \dots, n+1\}$ in the partition, y_j takes on the lower bound value $-c_j^u$ and other y_i equal to their upper bounds, i.e., $y_i = y_{i+1} + c_{i+1}^w, \forall i = k, \dots, j-1$. As a result, for each interval $\{k, \dots, j\}$ in the partition and $i \in \{k, \dots, j\}$, the value of y_i is given by:

$$y_i = \pi_{ij} := \begin{cases} -c_j^u + \sum_{\ell=i+1}^j c_\ell^w & 1 \leq i \leq j \leq n, \\ c^o + \sum_{\ell=i+1}^n c_\ell^w & 1 \leq i \leq n, j = n+1, \end{cases} \quad (4.16)$$

and $y_{n+1} = \pi_{n+1, n+1} := 0$. Define binary variables t_{kj} for all $1 \leq k \leq j \leq n+1$, such that $t_{kj} = 1$ if interval $\{k, \dots, j\}$ belongs to the partition (i.e., $t_{kj} = 1$ if $y_i = \pi_{ij}$) and $t_{kj} = 0$ otherwise. For a valid partition, we require each index i belonging to exactly one interval, and thus $\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1, \forall i = 1, \dots, n+1$. For notation convenience, we define $x_{n+1} = q_{n+1} = s_{n+1} := 0$. Using binary variables t_{kj} , we reformulate the separation problem (4.14) as

$$\max_t \max_{(q,s) \in D_q \times D_s} \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(\sum_{i=k}^j (q_i s_i - x_i) \pi_{ij} \right) t_{kj} - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \quad (4.17a)$$

$$\text{s.t.} \quad \sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1 \quad \forall i = 1, \dots, n+1 \quad (4.17b)$$

$$t_{kj} \in \{0, 1\}, \quad \forall 1 \leq k \leq j \leq n+1. \quad (4.17c)$$

Note that the objective function (4.17a) contains trilinear terms $q_i s_i t_{kj}$ with binary variables q_i, t_{kj} , and continuous variables s_i . To linearize formulation (4.17), we define $p_{ikj} \equiv q_i t_{kj}$ and $o_{ikj} \equiv q_i s_i t_{kj}$ for all $1 \leq k \leq j \leq n+1$ and $k \leq i \leq j$. Also, we introduce the following McCormick inequalities (4.18a)–(4.18b) and (4.18c)–(4.18d) for variables p_{ikj}

and o_{ikj} , respectively.

$$p_{ikj} - t_{kj} \leq 0, \quad (4.18a)$$

$$p_{ikj} - q_i \leq 0, \quad p_{ikj} - q_i - t_{kj} \geq -1, \quad p_{ikj} \geq 0, \quad (4.18b)$$

$$o_{ikj} - s_i^L p_{ikj} \geq 0, \quad o_{ikj} - s_i^U p_{ikj} \leq 0, \quad (4.18c)$$

$$o_{ikj} - s_i + s_i^L(1 - p_{ikj}) \leq 0, \quad o_{ikj} - s_i + s_i^U(1 - p_{ikj}) \geq 0. \quad (4.18d)$$

Thus, the separation problem (4.14) is equivalent to an MILP as:

$$\max_{t,q,s,p,o} \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^j (\pi_{ij} o_{ikj} - x_i \pi_{ij} t_{kj}) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \quad (4.19a)$$

$$\text{s.t.} \quad (4.17b)–(4.17c), \quad (4.18a)–(4.18d), \quad (4.19b)$$

$$s_i \in [s_i^L, s_i^U], \quad q \in D_q \subseteq \{0, 1\}^n. \quad (4.19c)$$

We can replace Steps 3–8 of Algorithm 4.1 proposed in Section 4.3 based on this MILP reformulation:

- 3 *With (x, ρ, γ) fixed to be (x^*, ρ^*, γ^*) , solve formulation (4.19) and record an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ **if** $\delta^* \geq \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^j (\pi_{ij} o_{ikj}^* - x_i^* \pi_{ij} t_{kj}^*) - \sum_{i=1}^n (\rho_i^* s_i^* + \gamma_i^* q_i^*)$ **then***
- 4 *stop and return x^* as an optimal solution to formulation (4.7);*
- 5 **else**
- 6 *add the cut $\delta \geq \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^j (\pi_{ij} o_{ikj}^* - \pi_{ij} t_{kj}^* x_i) - \sum_{i=1}^n (s_i^* \rho_i + q_i^* \gamma_i)$ to the set of cuts $\{L(x, \rho, \gamma, \delta) \geq 0\}$ and go to Step 2;*
- 7 **end**

Remark 4.1. We note that Algorithm 4.1 applies to various types of no-show support D_q . For example, we can specify $D_q = \{q \in \{0, 1\}^n : \sum_{i=1}^n (1 - q_i) \leq Q_{\max}\}$, where Q_{\max} represents the maximum number of no-shows. In this case, we only need to replace the definition of D_q in (4.19c) when applying Algorithm 4.1. In fact, Algorithm 4.1 is general regardless of the specific form of set D_q , to select which we take into account the operator's beliefs and/or preferences, and the computational tractability. In this chapter, we specify $D_q = D_q^{(K)}$ due to its flexibility (see Section 4.2.4) and computational tractability (see Proposition 3 and Theorem 4.2).

We further identify a set of valid inequalities to strengthen formulation (4.19). We summarize the valid inequalities in the following proposition and delegate its proof in A.3.3. The inequalities (4.20a)–(4.20f) can be added to the MILP (4.19) solved in Step

3, to strengthen the reformulation.

Proposition 3. The following inequalities are valid for set $F = \{(t, q, s, p, o) : (4.19b)–(4.19c)\}$:

$$\sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} = q_i \quad \forall i = 1, \dots, n+1, \quad (4.20a)$$

$$s_i - \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^L p_{ikj}) \geq s_i^L \quad \forall 1 \leq i \leq n+1, \quad (4.20b)$$

$$s_i - \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^U p_{ikj}) \leq s_i^U \quad \forall 1 \leq i \leq n+1, \quad (4.20c)$$

$$\sum_{\ell=i}^{i+K-1} p_{\ell kj} \geq t_{kj} \quad \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j - K + 1, \quad (4.20d)$$

$$\sum_{k=1}^{i-K+2} \sum_{\ell=i-K+2}^i p_{\ell ki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \geq \sum_{k=1}^{i-K+2} t_{ki} \quad \forall i = K-1, \dots, n, \quad (4.20e)$$

$$\sum_{k=1}^i p_{iki} + \sum_{\ell=i+1}^{i+K-1} \sum_{j=i+K-1}^{n+1} p_{\ell(i+1)j} \geq \sum_{j=i+K-1}^{n+1} t_{(i+1)j} \quad \forall i = 1, \dots, n - K + 2. \quad (4.20f)$$

Remark 4.2. Note that the above inequalities hold valid for all $K = 2, \dots, n+1$. We also note two features that (i) valid inequalities (4.20a)–(4.20f) are *polynomially* many, and (ii) all coefficients of these inequalities are in *closed-form*. Features (i) and (ii) can significantly accelerate Algorithm 4.1, because Feature (i) ensures that model (4.14) (i.e., (4.19) after reformulation) remains small by incorporating these inequalities, and Feature (ii) implies that we do not need to separate these inequalities.

4.4 LP Reformulations of the DR Expectation Model

In this section, we present tractable reformulations of the DR expectation model (4.7) as we derive the convex hull of separation problem (4.14) for $D_q = D_q^{(2)}$ (i.e., no conservative no-shows) and $D_q = D_q^{(n+1)}$ (i.e., arbitrary no-shows). This leads to polynomial-sized LP reformulations of model (4.7). We note that these LP reformulations are derived for these the special cases and do not simplify the general model. For the general cases (i.e., $D_q = D_q^{(K)}$ with $3 \leq K \leq n$), we can apply Algorithm 4.1 to solve model (4.7) and obtain

a globally optimal appointment schedule.

Case 1. (No Consecutive No-Shows) Recall that F represents the mixed-integer feasible region of formulation (4.19), i.e., $F = \{(t, q, s, p, o) : (4.19b)–(4.19c)\}$. We show that the valid inequalities identified in Proposition 3 are sufficient to describe $\text{conv}(F)$. We first notice that when $K = 2$: (i) inequalities (4.20d) are equivalent to $p_{ikj} + p_{(i+1)kj} \geq t_{kj}$ for all $1 \leq k < j \leq n + 1$ and $k \leq i \leq j - 1$, and (ii) inequalities (4.20e) and (4.20f) are identical and equivalent to

$$\sum_{k=1}^i p_{iki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \geq \sum_{k=1}^i t_{ki} \quad \forall i = 1, \dots, n. \quad (4.21)$$

This leads to the following theorem, of which a proof is relegated to A.3.4.

Theorem 4.1. Polyhedron

$$CF := \{(t, q, s, p, o) : (4.17b), (4.18a), (4.18c), (4.20a)–(4.20d), (4.21)\}$$

is the convex hull of set F , i.e., $CF = \text{conv}(F)$.

Therefore, we can reformulate the separation problem (4.14) as an LP model:

$$\begin{aligned} \max_{t, q, s, p, o} \quad & \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^j (\pi_{ij} o_{ikj} - x_i \pi_{ij} t_{kj}) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \\ \text{s.t.} \quad & (t, q, s, p, o) \in CF. \end{aligned}$$

To combine the separation problem with the outer minimization problem in (4.13), we

present the above reformulation in its dual form:

$$\min \sum_{i=1}^{n+1} (\alpha_i + s_i^U \tau_i^U - s_i^L \tau_i^L) \quad (4.22a)$$

$$\text{s.t.} \quad \sum_{i=k}^j (\alpha_i - \sigma_{ikj}) + \sum_{i=k}^{j-1} \lambda_{ikj} + \sum_{i=j}^{\min\{j,n\}} \phi_i \geq - \sum_{i=k}^j \pi_{ij} x_i \quad \forall 1 \leq k \leq j \leq n+1, \quad (4.22b)$$

$$\zeta_i \leq \gamma_i \quad \forall 1 \leq i \leq n, \quad (4.22c)$$

$$\tau_i^L - \tau_i^U \leq \rho_i \quad \forall 1 \leq i \leq n, \quad (4.22d)$$

$$\begin{aligned} \sigma_{ikj} + s_i^L \varphi_{ikj}^L - s_i^U \varphi_{ikj}^U + \zeta_i - s_i^L \tau_i^L + s_i^U \tau_i^U - \sum_{\ell=\max\{k,i-1\}}^{\min\{j-1,i\}} \lambda_{\ell kj} \\ - \sum_{\ell=\min\{2i-k-1,i\} \vee 1}^{\max\{2i-j,i-1\} \wedge n} \phi_\ell \geq 0 \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \end{aligned} \quad (4.22e)$$

$$- \varphi_{ikj}^L + \varphi_{ikj}^U + \tau_i^L - \tau_i^U \geq \pi_{ij} \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \quad (4.22f)$$

$$\varphi_{ikj}^L, \varphi_{ikj}^U, \tau_i^L, \tau_i^U, \lambda_{ikj}, \phi_i, \sigma_{ikj} \geq 0 \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \quad (4.22g)$$

where we denote $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$ for $a, b \in \mathbb{R}$ for notation convenience. Here the dual variables α_i , σ_{ikj} , $\varphi_{ikj}^{L/U}$, ζ_i , $\tau_i^{L/U}$, λ_{ikj} , and ϕ_i are associated with constraints (4.17b), (4.18a), (4.18c), (4.20a), (4.20b)–(4.20c), (4.20d), and (4.21) respectively (after transforming all “ \geq ” inequalities into the “ \leq ” form), and constraints (4.22b)–(4.22f) are associated with primal variables t_{kj} , q_i , s_i , p_{ikj} , and o_{ikj} respectively. In (4.22b), the term $\sum_{i=j}^{\min\{j,n\}} \phi_i$ becomes ϕ_j for all $1 \leq j \leq n$, and will disappear for $j = n+1$. In (4.22e), when $k \leq i < j$, the term $-\sum_{\ell=\max\{k,i-1\}}^{\min\{j-1,i\}} \lambda_{\ell kj}$ becomes $-\lambda_{ikj} - \lambda_{(i-1)kj}$; when $k < i = j$, it becomes a singleton $-\lambda_{(i-1)kj}$; and when $k = i = j$, it does not appear. Similarly, when $2 \leq k = i = j \leq n$, the term $-\sum_{\ell=\min\{2i-k-1,i\} \vee 1}^{\max\{2i-j,i-1\} \wedge n} \phi_\ell$ becomes $-\phi_i - \phi_{i-1}$; when $j > i = k$ or $k = i = j = n+1$, the term only contains $-\phi_{i-1}$; when $k < i = j$ or $1 = k = i = j$, the term only contains $-\phi_i$; and in all other cases, i.e., when $1 \leq k < i < j \leq n+1$, the term does not appear. We can then reformulate the DR expectation model in an LP form as follows.

Theorem 4.2. Under no-consecutive no-show assumption, i.e., $D_q = D_q^{(2)}$, the DR expectation model (4.7) is equivalent to the following linear program:

$$\begin{aligned} \min \quad & \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \sum_{i=1}^{n+1} (\alpha_i + s_i^U \tau_i^U - s_i^L \tau_i^L) \\ \text{s.t.} \quad & (4.22\text{b})\text{--}(4.22\text{g}), \quad \sum_{i=1}^n x_i = T, \quad x_{n+1} = 0, \quad x_i \geq 0 \quad \forall i = 1, \dots, n. \end{aligned}$$

Case 2. (Arbitrary No-Shows): Given $D_q = \{0, 1\}^n$ and $D_s = \prod_{i=1}^n [s_i^L, s_i^U]$, the optimization problem defining function $h(x, y, \rho, \gamma)$ (see (4.12c)) is separable by each appointment, i.e.,

$$\begin{aligned} h(x, y, \rho, \gamma) &= \max_{(q,s) \in D_q \times D_s} \left\{ \sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \\ &= \sum_{i=1}^n \max_{q_i \in \{0,1\}, s_i \in [s_i^L, s_i^U]} \{ (q_i s_i - x_i) y_i - (\rho_i s_i + \gamma_i q_i) \}. \end{aligned}$$

To reformulate separation problem (4.14), recall the observations on polyhedron Y in Section 4.3.2 and again we represent the extreme points of Y based on variables t_{kj} . It follows that

$$\max_{y \in Y} h(x, y, \rho, \gamma) = \max_{t \geq 0} \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(\sum_{i=k}^j \max_{q_i \in \{0,1\}, s_i \in [s_i^L, s_i^U]} \{ (q_i s_i - x_i) \pi_{ij} - (\rho_i s_i + \gamma_i q_i) \} \right) t_{kj} \quad (4.23\text{a})$$

$$\text{s.t.} \quad \sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1 \quad \forall i = 1, \dots, n+1 \quad (4.23\text{b})$$

$$t_{kj} \in \{0, 1\}, \quad \forall 1 \leq k \leq j \leq n+1. \quad (4.23\text{c})$$

Because the constraint matrix formed by (4.23b)–(4.23c) is totally unimodular (TU), we can relax the integrality constraints (4.23c) without loss of optimality. Hence, formulation

(4.23a)–(4.23c) is an LP model in variables t_{kj} and we can take its dual as:

$$\min_{\alpha, \beta} \sum_{i=1}^{n+1} \alpha_i \quad (4.24a)$$

$$\text{s.t. } \sum_{i=k}^j \alpha_i \geq \sum_{i=k}^j \beta_{ij}, \quad \forall 1 \leq k \leq j \leq n+1 \quad (4.24b)$$

$$\beta_{ij} \geq \max_{q_i \in \{0,1\}, s_i \in [s_i^L, s_i^U]} \{(q_i s_i - x_i) \pi_{ij} - (\rho_i s_i + \gamma_i q_i)\},$$

$$\forall i = 1, \dots, n, \forall j = i, \dots, n+1 \quad (4.24c)$$

$$\beta_{n+1, n+1} = 0, \quad (4.24d)$$

where dual variables α_i , $i = 1, \dots, n+1$ are associated with constraints (4.23b), constraints (4.24b) are associated with variables y_{kj} , each variable β_{ij} represents the value of $\max_{q_i \in \{0,1\}, s_i \in [s_i^L, s_i^U]} \{(q_i s_i - x_i) \pi_{ij} - (\rho_i s_i + \gamma_i q_i)\}$, and $\beta_{n+1, n+1} = 0$ because $q_{n+1} = s_{n+1} = \pi_{n+1, n+1} = 0$. Finally, for each $i = 1, \dots, n$, the related objective function $(q_i s_i - x_i) \pi_{ij} - (\rho_i s_i + \gamma_i q_i)$ is linear in variables q_i and s_i , and thus each of constraints (4.24c) is equivalent to

$$\beta_{ij} \geq -\pi_{ij} x_i - s_i^L \rho_i \quad (4.25a)$$

$$\beta_{ij} \geq -\pi_{ij} x_i - s_i^U \rho_i \quad (4.25b)$$

$$\beta_{ij} \geq -\pi_{ij} x_i - s_i^L \rho_i - \gamma_i + s_i^L \pi_{ij} \quad (4.25c)$$

$$\beta_{ij} \geq -\pi_{ij} x_i - s_i^U \rho_i - \gamma_i + s_i^U \pi_{ij}, \quad (4.25d)$$

because $q_i \in \{0, 1\}$ and $s_i \in \{s_i^L, s_i^U\}$ at optimality. It follows that model (4.13) (i.e., the DR expectation model (4.7)), is equivalent to the following LP model when $D_q = D_q^{(n+1)}$:

$$\min_{x, \rho, \gamma, \alpha, \beta} \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \sum_{i=1}^{n+1} \alpha_i$$

$$\text{s.t. } (4.24b), (4.24d), (4.25a)–(4.25d), \sum_{i=1}^n x_i = T, \quad x_{n+1} = 0, \quad x_i \geq 0, \quad \forall i = 1, \dots, n.$$

4.5 Computational Results

We conduct numerical experiments on the three variants of the DR expectation model (4.7), namely, $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and $E-D_q^{(K)}$ ($K = 3, \dots, n$), yielded by $D_q = D_q^{(2)}$, $D_q^{(n+1)}$, and $D_q^{(K)}$, respectively. For benchmark, we also solve a stochastic linear program (SLP)

that minimizes the expected total cost of waiting, server idleness, and overtime via the SAA approach (Kleywegt et al., 2002). We briefly describe the key computational procedures as follows. First, we follow a distribution belief to generate N i.i.d. samples, of which we randomly pick a small subset of data to compute the empirical mean and support information, and use them to compute the (in-sample) optimal solutions and optimal objective values to the DR models.

For the SLP, we solve an LP model:

$$\begin{aligned}
\text{SLP: } \quad & \min_{x,w,u,W} \quad \frac{1}{N} \sum_{m=1}^N \sum_{i=1}^n (c_i^w w_i^m + c_i^u u_i^m) + c^o W^m & (4.26a) \\
\text{s.t.} \quad & w_i^m - u_{i-1}^m = q_{i-1}^m s_{i-1}^m + w_{i-1}^m - x_{i-1} \quad \forall i = 2, \dots, n, m = 1, \dots, N & (4.26b) \\
& W^m - u_n^m = q_n^m s_n^m + w_n^m + \sum_{i=1}^{n-1} x_i - T \quad \forall m = 1, \dots, N & (4.26c) \\
& \sum_{i=1}^{n-1} x_i \leq T & (4.26d) \\
& x_i \geq 0 \quad \forall i = 1, \dots, n-1 & (4.26e) \\
& w_i^m \geq 0, w_1^m = 0, u_i^m \geq 0, W^m \geq 0, \quad \forall i = 1, \dots, n, m = 1, \dots, N & (4.26f)
\end{aligned}$$

where q_i^m and s_i^m are realizations of parameter q_i and s_i of appointment i in scenario m , respectively, for all $i = 1, \dots, n$ and $m = 1, \dots, N$. Variables w_i^m , u_i^m , and W^m represent recourse waiting time of appointment i , server idle time after serving appointment i , and server overtime in scenario m , respectively, for all $m = 1, \dots, N$. Constraints (4.26b) and (4.26c) obtain the waiting time/idle time/overtime values for each appointment dependent on values of x_i and (q_i^m, s_i^m) .

Section 4.5.1 describes how to set the parameter for the above models; Section 4.5.2 compares the CPU time and details of solving $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, $E-D_q^{(K)}$, and SLP. In Section 4.5.3, we illustrate optimal objective values given by $E-D_q^{(K)}$ with $K = 2, \dots, n$ for different settings of time limit T and no-show probabilities. In Section 4.5.4, we compare the performance of optimal schedules of $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP via out-of-sample simulation tests. Specifically, we follow a certain distribution to generate N' data samples, which represent realizations of random service durations and no-shows. The distributions used for generating the in-sample and out-of-sample data could be different, and when they are the same and N is sufficiently large, the SLP is considered being optimized under the ‘‘perfect information’’ (Birge and Louveaux, 2011). In reality, it is hard to know the exact true distribution, and thus we also test the case where the distribution is ‘‘misspecified.’’

4.5.1 Experiment Setup

We follow procedures in the appointment scheduling literature (e.g., Denton and Gupta, 2003; Mak et al., 2015) to generate random instances as follows. For most instances, we consider $n = 10$ appointments, each having a random service duration s_i with the mean $\mu_i \sim U[36, 44]$ (i.e., uniformly sampled in between the values below and above 10% of 40 minutes) and the standard deviation $\sigma_i = 0.5\mu_i$. We set $T = \sum_{i=1}^n \mu_i + R \cdot \sqrt{\sum_{i=1}^n \sigma_i^2}$ where scalar R adjusts the length of time limit T . Note that in this setting T does not take into consideration the no-show probabilities. In Section 4.5.5, we report the results when T does depend on the no-show probabilities. Each appointment i has a probability ν_i of showing up, and we test $\nu_1 = \dots = \nu_n = 0.8$ or $\nu_1 = \dots = \nu_n = 0.6$. To approximate the upper and lower bounds s_i^U and s_i^L of each service duration s_i , we respectively use the 80%- and 20%-quantile values of the N in-sample data. We set the ratio $c_i^w : c_i^u : c_i^o = 1 : 0.5 : 10$ in all the DR models and in SLP.

We sample $N = 1000$ realizations $(q_1^m, s_1^m), \dots, (q_n^m, s_n^m)$, $m = 1, \dots, N$ by following Log-Normal distributions with the given means and standard deviations of s_i and probability $1 - \nu_i$ of no-shows for each $i = 1, \dots, n$. (The Log-Normal distribution possesses the long-tail property and has been shown accurately describing the shape of service-time distributions in many service systems (see, e.g., Gul et al., 2011, for a study of five-year outpatient surgical data in Mayo Clinic).) We optimize the SLP model by using all the N data points, and only use 20 randomly picked data samples from the N -data set to calculate the first moments of service durations and no-shows, used in all the DR models. Given the optimal schedules produced by different models, we generate $N' = 10,000$ i.i.d. data samples from certain distributions with details given in Section 4.5.4, to evaluate the performance of each schedule.

We increase the size of instances with $n = 10, 15, \dots, 50$ appointments in Section 4.5.2 to compare the CPU time of different models and approaches. For each instance, we generate $N = 1000$ i.i.d. data samples of service durations and no-shows by following the same procedures as above. All LP (i.e., $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP) and MILP (i.e., $E-D_q^{(K)}$ with $K = 3, \dots, n$) models are computed in Python 2.7.10 using Gurobi 5.6.3. The computations are performed on a Windows 7 machine with Intel(R) Core(TM) i7-2600M CPU 3.40 GHz and 8GB memory. The CPU time limit is set as 3 hours for solving each instance.

4.5.2 CPU Time and Computational Details

In this section, we increase the problem size from $n = 10$ to $n = 50$ appointments, and compare CPU time of optimizing different DR models and the SLP. We first vary $R = 0, 0.25, 0.5, 0.75, 1$, and find that the CPU time of all the models are similar for different R values and no-show probabilities. Thus, we fix $R = 0$ and use $T = \sum_{i=1}^n \mu_i = 380.24$ minutes. We consider $\nu_i = 0.6$ for all appointments $i = 1, \dots, n$, and test ten replications for each instance. Table 4.1 reports the average CPU time (in second) of solving models $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, SLP, and $E-D_q^{(K)}$ with $K = 0.3n$ and $K = 0.7n$. Note that the first three are LP models, and specifically, there are $\mathcal{O}(n^3)$ variables and $\mathcal{O}(n^3)$ constraints in the two LP models $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$, but $\mathcal{O}(nN)$ variables and constraints in SLP. The $E-D_q^{(0.3n)}$ and $E-D_q^{(0.7n)}$ models are solved via Algorithm 4.1, and we present the average time for solving the MILP models with (see columns ‘‘Ineq.’’) and without (see columns ‘‘w/o’’) the valid inequalities in Proposition 3. For instances that take longer than 3 hours to solve, we instead report the optimality gap values (in %) achieved at the end of the computation process.

Table 4.1: Average CPU time (in second) of solving DR models and SLP with $R = 0$ and $1 - \nu_i = 0.4$

n	$E-D_q^{(2)}$	$E-D_q^{(n+1)}$	SLP	$E-D_q^{(0.3n)}$		$E-D_q^{(0.7n)}$	
				w/o	Ineq.	w/o	Ineq.
10	0.03	0.00	3.10	6.70	10.68	3.44	1.95
15	0.16	0.01	7.24	52.74	47.56	20.28	4.23
20	0.38	0.01	10.44	158.50	106.67	72.96	15.12
25	2.85	0.02	16.97	409.88	270.60	266.34	47.18
30	5.12	0.05	20.91	1000.38	187.81	823.14	101.68
35	11.50	0.05	28.18	10658.06	401.91	7994.79	340.76
40	28.87	0.15	32.32	(5.76%)	808.38	(10.98%)	614.75
45	26.55	0.20	39.07	(12.49%)	1739.49	(10.24%)	1491.17
50	24.64	0.45	44.49	(31.77%)	3271.83	(43.63%)	3393.62

In Table 4.1, the CPU time of both $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ are shorter than the one of SLP, especially after $n \geq 40$. The time for solving $E-D_q^{(2)}$ is longer than solving $E-D_q^{(n+1)}$, due to the many more constraints involved in the former. Note that the time presented in Table 4.1 is only for solving all the models, but does not include the time spent on reading in data and constructing the constraints, which is negligible for all the DR models, but grows quickly for SLP, ranging from 30 seconds to 60 seconds when $n \geq 35$.

All the DR LP models and SLP are efficiently solved for $n = 10, \dots, 50$, while the MILP models $E-D_q^{(K)}$ with either small or large K -values are computationally intractable,

reflected by the long CPU seconds taken by instances with $n = 35, 40, 45, 50$, especially when no valid inequalities were added. The addition of valid inequalities in Proposition 3 drastically speeds up the decomposition algorithm, and the effect is much more significant when $n \geq 35$. For instance, none of the $n = 40, 45, 50$ cases were solved within 3 hours without the valid inequalities, and the average optimality gaps could be as large as $30\% \sim 45\%$ when $n = 50$. In contrast, after adding the valid inequalities, the decomposition algorithm quickly converges, and on average, it only takes no more than 15, 30, and 60 minutes to optimize the MILPs over instances with $n = 40, 45$, and 50 , respectively.

Next, we present more details of solving the $E-D_q^{(K)}$ MILPs. Table 4.2 illustrates the number of constraints (“# of Cons.”), the total number of branching nodes (“# of Nodes”), the average CPU seconds taken by the master problem and the subproblem in each iteration, and the number of iterations in the decomposition algorithm before it converges to the optimum or reaches the time limit (“# of Cuts”) for solving both $E-D_q^{(0.3n)}$ and $E-D_q^{(0.7n)}$, with or without the valid inequalities.

Table 4.2: Computational details of solving the MILP models $E-D_q^{(0.3n)}$ and $E-D_q^{(0.7n)}$

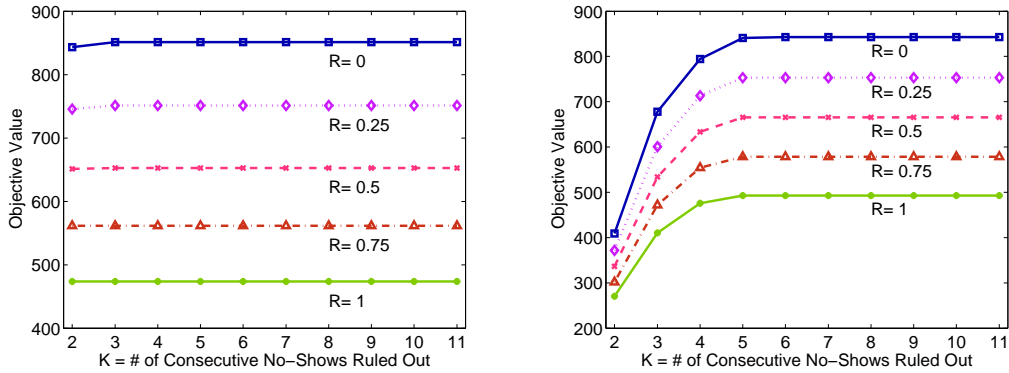
Models	n	with Ineq.					w/o Ineq.				
		# of Cons.	# of Nodes	Avg. Time (s)		# of Cuts	# of Cons.	# of Nodes	Avg. Time (s)		# of Cuts
				Master	Sub				Master	Sub	
$E-D_q^{(0.3n)}$	10	2239	790	0.00	0.22	48	2023	1810	0.00	0.13	49
	15	6271	1363	0.01	0.61	77	5742	8510	0.01	0.62	83
	20	13346	2572	0.03	1.04	99	12435	19112	0.03	1.52	102
	25	24637	2865	0.08	2.38	110	22979	31602	0.11	3.83	104
	30	40686	32	0.23	2.00	84	38247	54697	0.22	9.04	108
	35	62932	0	0.51	3.95	90	59116	1245441	0.53	86.12	123
	40	91602	0	1.05	7.28	97	86459	677713	2.63	632.67	17
	45	128499	0	1.60	16.52	96	121153	387066	7.30	1342.76	8
50	173437	0	3.09	27.48	107	164071	324952	17.57	3582.45	3	
$E-D_q^{(0.7n)}$	10	2097	0	0.01	0.06	29	2019	525	0.00	0.13	25
	15	5882	0	0.01	0.11	35	5736	1975	0.02	0.56	35
	20	12626	0	0.04	0.28	48	12427	6674	0.05	1.41	50
	25	23287	0	0.11	0.71	58	22969	19704	0.13	3.79	68
	30	38636	0	0.23	1.50	59	38235	52000	0.27	10.02	80
	35	59691	0	0.48	3.63	83	59102	848133	0.42	82.86	96
	40	87154	0	0.93	6.06	88	86443	880188	3.90	1538.97	7
	45	122121	0	2.07	15.07	87	121135	385058	7.89	1192.12	9
50	165207	0	3.06	28.36	108	164051	268133	18.05	10786.87	1	

In Table 4.2, we observe that the valid inequalities in Proposition 3 slightly increase the number of constraints, but significantly tighten the MILPs, directly reflected by the significantly reduced branching-and-bound nodes in all the instances. In particular, the decomposition algorithm obtains integer solutions at the root node in each iteration for solving $E-D_q^{(0.3n)}$ when $n \geq 35$, and for $E-D_q^{(0.7n)}$ given any values of n we test. Moreover, the valid inequalities significantly reduce the CPU seconds of computing the separation subproblem in each iteration, especially for instances with $n = 35, 40, 45, 50$. Lastly, by

adding the valid inequalities to the MILP models, the decomposition algorithm takes almost constant number of iterations to converge (i.e., being around 100 iterations when $n \geq 20$). However, if no valid inequalities were added, the number of iterations first increases as we increase n from 10 to 35, and drastically decreases as we continue increasing n to 50.

4.5.3 Optimal Objective Values and Scheduling Solution Patterns

We compare the optimal objective values of the $E-D_q^{(K)}$ models with $K = 2, \dots, n + 1$, and plot their value changes for instances with $R = 0, 0.25, 0.5, 0.75, 1$ and no-show probabilities $1 - \nu_i = 0.2, 0.4$ for all $i = 1, \dots, n$ when $n = 10$. Recall that parameter K represents the minimum number of consecutive appointments in which consecutive no-shows are ruled out. Therefore, as K increases, the support $D_q^{(K)}$ becomes larger, which leads to smaller feasible region for the scheduling decision vector x , and thus the optimal objective value is nondecreasing for $K = 2, \dots, n + 1$. Figure 4.2 illustrates the optimal objective values, in which Figure 4.2(a) corresponds to $1 - \nu_i = 0.2, \forall i = 1, \dots, n$ and Figure 4.2(b) corresponds to larger no-show probabilities $1 - \nu_i = 0.4, \forall i = 1, \dots, n$.



(a) $n = 10, 1 - \nu_i = 0.2, \forall i = 1, \dots, n$ (b) $n = 10, 1 - \nu_i = 0.4, \forall i = 1, \dots, n$

Figure 4.2: Optimal objective values of $E-D_q^{(K)}$ for different settings of parameter R (time limit) and $1 - \nu_i$ (no-show probability)

Table 4.3 presents the detailed optimal objective values in Figure 4.2. Note that the optimal objective values of $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ respectively provide valid lower and upper bounds for the optimal objective value of any $E-D_q^{(K)}$ models with $K = 3, \dots, n$. For each combination of R and $1 - \nu_i$, we mark the first K -value when the optimal objective value of $E-D_q^{(K)}$ equals to the upper bound, i.e., the optimal objective value of $E-D_q^{(n+1)}$.

In Table 4.3, when the no-show probability is smaller (i.e., $1 - \nu_i = 0.2, \forall i = 1, \dots, n$),

Table 4.3: Optimal objective value changes according to the value of K and no-show probabilities

No-show	R	$E-D_q^{(2)}$	$E-D_q^{(K)}$, with $K =$								$E-D_q^{(n+1)}$
			3	4	5	6	7	8	9	10	
$1 - \nu_i = 0.2$	0	843.38	851.08	851.08	851.08	851.08	851.08	851.08	851.08	851.08	851.08
	0.25	745.74	751.32	751.32	751.32	751.32	751.32	751.32	751.32	751.32	751.32
	0.5	651.18	652.54	652.54	652.54	652.54	652.54	652.54	652.54	652.54	652.54
	0.75	561.32	561.32	561.32	561.32	561.32	561.32	561.32	561.32	561.32	561.32
	1	473.69	473.69	473.69	473.69	473.69	473.69	473.69	473.69	473.69	473.69
$1 - \nu_i = 0.4$	0	409.11	677.66	794.41	840.70	842.91	842.91	842.91	842.91	842.91	842.91
	0.25	371.47	600.99	713.55	752.87	752.87	752.87	752.87	752.87	752.87	752.87
	0.5	336.21	534.22	633.12	665.24	665.24	665.24	665.24	665.24	665.24	665.24
	0.75	301.89	472.07	553.93	578.61	578.61	578.61	578.61	578.61	578.61	578.61
	1	270.55	410.02	475.67	492.69	492.69	492.69	492.69	492.69	492.69	492.69

the differences between the upper and lower bounds are very small, indicating that the two LP models $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ can already provide tight approximations for the MILPs of $E-D_q^{(K)}$ with $K = 3, \dots, n$. Considering the long CPU time of solving the MILPs in Table 4.1, one can avoid directly solving $E-D_q^{(K)}$ ($K = 3, \dots, n$), and instead use $K = 2$ or $K = n + 1$. On the other hand, when the no-show probability is larger (i.e., $1 - \nu_i = 0.4, \forall i = 1, \dots, n$), the differences between the results of $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ become larger when R is smaller. In such a case, the choices of different K values will lead to significantly different objective costs. A decision maker can choose either $E-D_q^{(2)}$ or $E-D_q^{(n+1)}$ to optimize the schedule x based on his/her risk preference. Alternatively, he/she can firstly use the two LP models to quickly compute the bounds of the optimal objective value for a general $E-D_q^{(K)}$ for any $K = 3, \dots, n$, and then optimize the $E-D_q^{(K)}$ model for some K by employing the valid inequalities in Proposition 3 and the decomposition algorithm.

We demonstrate in Figure 4.3 the optimal schedules of instances with $n = 10$ appointments, produced by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP for $R = 0, 1$ and $1 - \nu_i = 0.2, 0.4, \forall i = 1, \dots, n$. The points (i, x_i) of every model in each subfigure correspond to the time interval (in minute) assigned in between the arrivals of appointments i and $i + 1$, for all $i = 1, \dots, 9$.

As shown in Figure 4.3, SLP almost equally distributes the time in between each arrival and schedules a long interval after the last appointment, for all combinations of $1 - \nu_i$ and R values. As compared to SLP, both $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ schedule longer inter-arrival time for the early appointments. Intuitively, $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ intend to mitigate the waiting time that may accumulate due to long service durations (also reflected by the shorter waiting time for the DR models in Tables 4.4 and 4.5, reported later in Section 4.5.4). Additionally, to mitigate the random no-shows, $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ intend to double book later appointments (reflected by the small x_8 or x_9 value in all cases). When the no-show

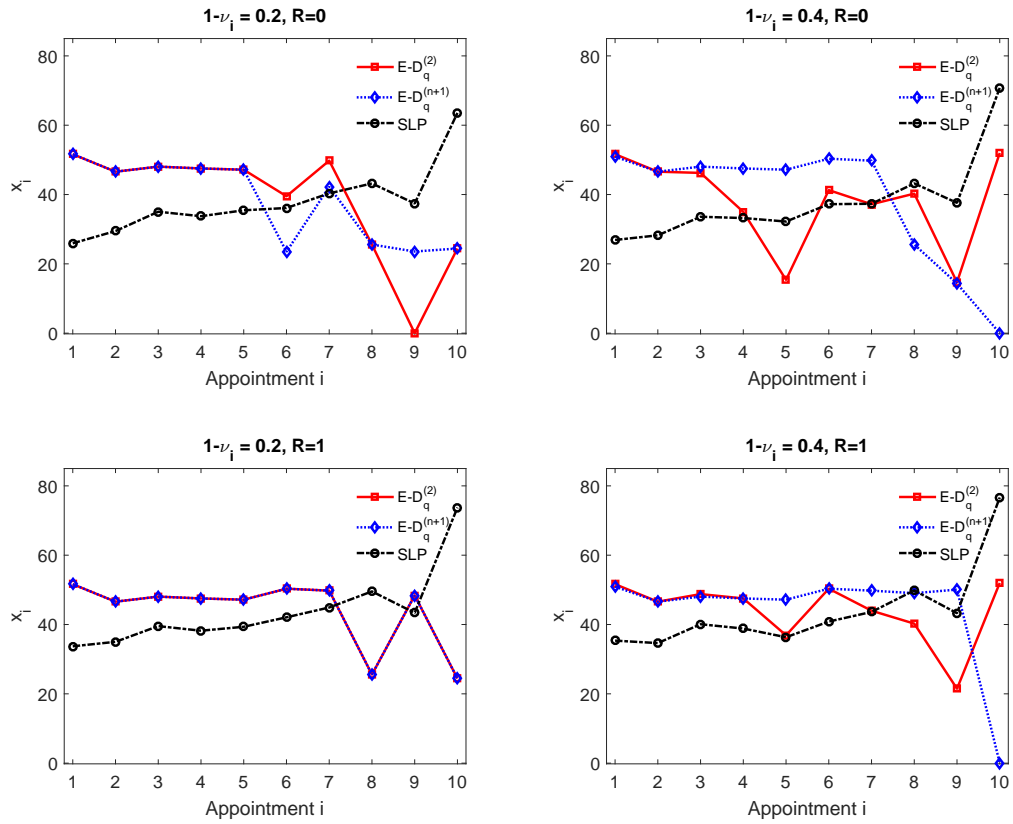


Figure 4.3: Appointment schedules produced by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP for different settings of parameter R (time limit) and $1 - \nu_i$ (no-show probability)

probability is relative small (i.e., $1 - \nu_i = 0.2$, $\forall i = 1, \dots, n$) and the time limit T is sufficiently long (i.e., $R = 1$), both $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ yield the same optimal schedule (also reflected by the same optimal objective value of the two models in Table 4.3).

4.5.4 Results of Out-of-Sample Performance

We compare the out-of-sample simulation performance of the optimal schedules of $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP (see Figure 4.3) under (i) “perfect information” and (ii) misspecified distributional information. Note that $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ models produce solutions that differ the most under large no-show probabilities, and thus we focus on the case when $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$. We examine three cases of the time limit T , by using $R = 0, 0.5, 1$.

We generate two sets of $N' = 10,000$ i.i.d. out-of-sample data $(q_1^m, s_1^m), \dots, (q_n^m, s_n^m)$, $m = 1, \dots, N'$ of the random vector (q, s) following the procedures as follows.

- **Perfect Information:** We use the same distribution (i.e., Log-Normal) and parameter settings as the ones for generating the N in-sample data to sample the N' data points.
- **Misspecified Distribution:** We keep the same mean values μ_i of the random s_i , ν_i of the random q_i , and standard deviation σ_i of the random s_i for each appointment $i = 1, \dots, n$. Therefore, the moment information used in all the DR models and in the SLP remain the same, but we vary the distribution type, as well as correlations among the random service durations and no-shows. Specifically, we follow positively correlated truncated Normal distributions with supports $[0, s_i^U]$, $\forall i = 1, \dots, n$ to generate realizations s_1^m, \dots, s_n^m and follow positively correlated Bernoulli distributions to generate realizations q_1^m, \dots, q_n^m for $m = 1, \dots, N'$. The parameters of the truncated Normal distributions and the Bernoulli distributions are designed by following standard statistical methods¹, in order to yield positive data correlations and also to keep the first two moments of the N' out-of-sample data the same as the ones of the N in-sample data.

To measure the out-of-sample performance of each solution given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP, we fix x as an interested solution in Model (4.26), but use parameters $(q_1^m, s_1^m), \dots, (q_n^m, s_n^m)$, $m = 1, \dots, N'$. We then compute w_i^m , u_i^m , W^m as the waiting time (WaitT), idle time (idleT), and overtime (OverT) in each scenario m , for $m = 1, \dots, N'$. Table 4.4 displays means and quantiles of WaitT, IdleT, and OverT, yielded by the optimal solution of each model under perfect distributional information.

¹See https://en.wikipedia.org/wiki/Truncated_normal_distribution.

Table 4.4: Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under perfect information with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$

Metrics	Model	$R = 0$ (in minute)			$R = 0.5$ (in minute)			$R = 1$ (in minute)		
		WaitT	OverT	IdleT	WaitT	OverT	IdleT	WaitT	OverT	IdleT
Mean	$E-D_q^{(2)}$	10.06	13.97	16.62	7.33	11.26	19.31	5.50	9.18	22.06
	$E-D_q^{(n+1)}$	6.62	49.27	20.15	4.87	33.59	21.54	3.33	28.11	23.96
	SLP	11.20	3.50	15.57	8.70	2.29	18.41	6.70	1.64	21.31
Median	$E-D_q^{(2)}$	0.00	0.00	12.27	0.00	0.00	14.68	0.00	0.00	20.47
	$E-D_q^{(n+1)}$	0.00	44.92	15.47	0.00	29.22	19.29	0.00	24.93	21.79
	SLP	0.00	0.00	8.19	0.00	0.00	12.95	0.00	0.00	17.45
75%-QT	$E-D_q^{(2)}$	8.40	19.94	33.32	1.79	14.95	34.99	0.00	9.12	40.25
	$E-D_q^{(n+1)}$	0.00	72.90	45.40	0.00	52.22	45.40	0.00	45.27	47.16
	SLP	11.94	0.00	28.83	6.49	0.00	34.38	1.33	0.00	37.08
95%-QT	$E-D_q^{(2)}$	50.88	35.42	46.62	40.08	26.06	48.64	33.52	19.27	50.14
	$E-D_q^{(n+1)}$	40.51	117.78	50.33	30.68	90.84	50.33	22.03	76.66	50.33
	SLP	54.03	23.75	43.20	45.89	12.61	46.05	38.41	7.54	49.85

Based on Table 4.4, both $E-D_q^{(2)}$ and $E-D_q^{(n+1)}$ yield slightly better waiting time on average and at different quantiles than the SLP. The $E-D_q^{(2)}$ model results in overtime that is close to the one of SLP (which has perfect distributional information). But the optimal schedule of $E-D_q^{(n+1)}$ performs badly on average and at all quantiles. For example, when $R = 0$, the schedule by $E-D_q^{(2)}$ lasts 10 minutes longer than the 3-minute average overtime given by the SLP optimal schedule, while the optimal schedule of $E-D_q^{(n+1)}$ lasts about 46 minutes longer on average. This is due to the overly conservative no-show support assumption used by $E-D_q^{(n+1)}$. When the distributional information is accurate, $E-D_q^{(n+1)}$ results in over-conservative schedules that perform badly especially in the overtime metric.

Table 4.5 illustrates the means and quantiles of WaitT, IdleT, and OverT, yielded by optimal schedules of the three models under misspecified distributional information.

From Table 4.5, we observe that both DR models yield much better (i.e., 30%–70% shorter) waiting time per appointment than the optimal schedule given by the SLP, when the distribution type becomes different but the first two moments remain unchanged from the assumed case. The time reduction is reflected in all the metrics, including the mean and 50% to 95% quantiles of the random WaitT, for both $R = 0$ and $R = 1$. On the other hand, the three models yield similar IdleT, and the optimal schedule given by $E-D_q^{(n+1)}$ yields slightly longer idle time per appointment, but much longer OverT than both $E-D_q^{(2)}$ and SLP. This observation indicates that the optimal schedules given by SLP can become suboptimal when the probability distributions are misspecified, while $E-D_q^{(2)}$ can produce schedules that are less sensitive to misspecification of distribution types.

Table 4.5: Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under misspecified distribution with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$

Metrics	Model	$R = 0$ (in minute)			$R = 0.5$ (in minute)			$R = 1$ (in minute)		
		WaitT	OverT	IdleT	WaitT	OverT	IdleT	WaitT	OverT	IdleT
Mean	$E-D_q^{(2)}$	19.89	42.30	13.93	15.75	34.24	16.08	11.51	27.39	18.36
	$E-D_q^{(n+1)}$	12.14	69.59	16.65	10.20	52.16	17.87	8.26	44.56	20.07
	SLP	29.32	38.36	13.53	24.30	31.04	15.76	19.61	24.54	18.07
Median	$E-D_q^{(2)}$	0.00	0.00	5.52	0.00	0.00	9.62	0.00	0.00	14.15
	$E-D_q^{(n+1)}$	0.00	46.25	8.61	0.00	30.62	11.83	0.00	30.56	15.64
	SLP	0.00	0.00	5.09	0.00	0.00	9.16	0.00	0.00	13.00
75%-QT	$E-D_q^{(2)}$	15.34	41.98	27.16	10.70	27.98	30.95	4.91	19.47	35.42
	$E-D_q^{(n+1)}$	6.95	93.23	33.03	2.44	65.31	34.86	0.00	53.56	39.49
	SLP	27.64	11.57	26.36	21.18	0.00	30.33	15.06	0.00	34.45
95%-QT	$E-D_q^{(2)}$	109.75	241.69	46.59	89.11	212.08	49.77	65.02	182.47	48.78
	$E-D_q^{(n+1)}$	62.06	241.69	49.77	58.20	212.08	49.77	52.26	182.47	50.08
	SLP	147.84	241.63	43.20	130.09	212.02	46.05	110.73	182.41	49.84

4.5.5 No-Show-Dependent Time Limit

In this section, we make the time limit T depend on the no-show probabilities and report the scheduling solution patterns of SLP, $E-D_q^{(2)}$, and $E-D_q^{(n+1)}$, together with their out-of-sample performance under perfect information and misspecified distributional information. More specifically, we keep all experiment settings the same as in Sections 4.5.3–4.5.4, except for setting $T = \sum_{i=1}^n \nu_i \mu_i + R \cdot \sqrt{\sum_{i=1}^n [(\nu_i - \nu_i^2) \mu_i^2 + \nu_i \sigma_i^2]}$ to take into account the no-show probabilities. Note that $\mathbb{E}_{\mathbb{P}_{q,s}}[q_i s_i] = \nu_i \mu_i$ and $\text{Var}(q_i s_i) = (\nu_i - \nu_i^2) \mu_i^2 + \nu_i \sigma_i^2$ if q_i and s_i are independent. Additionally, $\nu_i < 1$ and so in this setting the time limit is shorter than that in Sections 4.5.3–4.5.4. First, we report the out-of-sample performance under perfect information in Table 4.6 and that under misspecified distributional information in Table 4.7. From these two tables, we make similar observations as in Section 4.5.4. Under perfect information, as compared to SLP, $E-D_q^{(2)}$ yields shorter waiting time, longer overtime, and similar idle time, while $E-D_q^{(n+1)}$ yields a much longer overtime in all settings. Under misspecified distributional information, both DR models result in significant reduction in waiting time and $E-D_q^{(2)}$ yields similar overtime and idle time as SLP does, while $E-D_q^{(n+1)}$ still performs poorly in overtime. This confirms our conclusion that $E-D_q^{(2)}$ can produce near-optimal schedules that are less sensitive to misspecification of distribution types.

Additionally, we demonstrate in Figure 4.4 the optimal schedules produced by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP for $R = 0, 1$ and $1 - \nu_i = 0.2, 0.4$, $\forall i = 1, \dots, n$. By comparing this figure with Figure 4.3, we observe that SLP now intends to double book the first ap-

Table 4.6: Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under perfect information with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$ and no-show-dependent time limit

Metrics	Model	$R = 0$ (in minute)			$R = 0.5$ (in minute)			$R = 1$ (in minute)		
		WaitT	OverT	IdleT	WaitT	OverT	IdleT	WaitT	OverT	IdleT
Mean	$E-D_q^{(2)}$	29.75	52.56	5.06	26.94	28.26	6.37	18.47	21.00	9.37
	$E-D_q^{(n+1)}$	30.74	124.71	12.28	22.62	102.41	13.78	14.77	79.93	15.27
	SLP	34.51	39.40	3.75	28.43	22.56	5.80	21.54	12.76	8.55
Median	$E-D_q^{(2)}$	14.16	42.62	0.00	10.80	8.63	0.00	0.00	0.00	0.00
	$E-D_q^{(n+1)}$	0.00	119.87	0.00	0.00	96.78	0.00	0.00	75.74	4.96
	SLP	19.02	23.80	0.00	11.99	0.00	0.00	4.30	0.00	0.00
75%-QT	$E-D_q^{(2)}$	42.48	77.75	2.48	37.84	44.71	6.73	24.86	33.50	14.68
	$E-D_q^{(n+1)}$	43.24	161.90	24.03	30.07	136.99	25.83	14.59	111.36	25.83
	SLP	48.38	63.39	0.00	39.89	34.70	5.39	29.54	13.68	13.06
95%-QT	$E-D_q^{(2)}$	98.19	145.32	34.88	92.85	109.04	40.23	75.06	90.94	37.75
	$E-D_q^{(n+1)}$	126.95	226.24	48.05	100.50	200.02	48.05	73.78	166.14	48.05
	SLP	108.56	135.82	23.36	95.45	99.25	28.36	81.38	71.52	33.42

Table 4.7: Out-of-sample performance of optimal schedules given by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP under misspecified distribution with no-show probabilities $1 - \nu_i = 0.4$, $\forall i = 1, \dots, n$ and no-show-dependent time limit

Metrics	Model	$R = 0$ (in minute)			$R = 0.5$ (in minute)			$R = 1$ (in minute)		
		WaitT	OverT	IdleT	WaitT	OverT	IdleT	WaitT	OverT	IdleT
Mean	$E-D_q^{(2)}$	52.66	103.65	4.64	49.78	80.99	6.11	37.38	65.68	8.31
	$E-D_q^{(n+1)}$	41.02	158.23	10.10	31.34	132.34	11.25	22.33	106.81	12.43
	SLP	62.29	100.52	4.33	55.23	80.26	6.04	46.12	63.99	8.14
Median	$E-D_q^{(2)}$	14.74	36.78	0.00	11.91	2.20	0.00	2.93	0.00	0.00
	$E-D_q^{(n+1)}$	4.25	138.88	0.00	0.00	110.64	0.00	0.00	80.42	0.00
	SLP	22.36	33.41	0.00	15.86	3.80	0.00	7.58	0.00	0.00
75%-QT	$E-D_q^{(2)}$	62.00	165.04	2.48	56.99	127.71	6.72	37.89	91.01	13.90
	$E-D_q^{(n+1)}$	50.66	211.26	17.45	35.39	185.49	20.42	19.92	148.20	23.49
	SLP	78.78	164.73	3.43	68.43	127.40	9.10	53.96	90.38	13.88
95%-QT	$E-D_q^{(2)}$	217.42	395.80	32.63	211.46	358.47	36.04	177.68	321.14	37.39
	$E-D_q^{(n+1)}$	169.15	395.85	47.52	140.08	358.53	47.52	108.03	321.20	47.52
	SLP	236.10	395.80	23.36	219.89	358.47	28.44	198.36	321.14	31.64

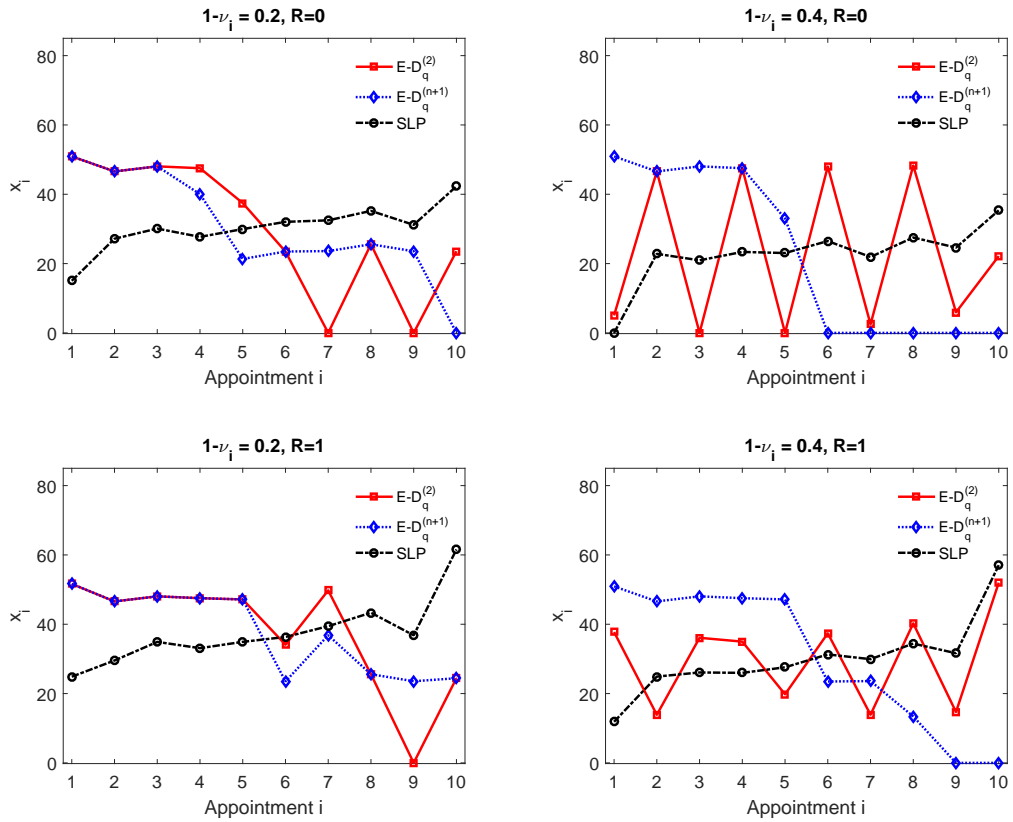


Figure 4.4: Appointment schedules produced by $E-D_q^{(2)}$, $E-D_q^{(n+1)}$, and SLP for different settings of parameter R (time limit) and $1 - \nu_i$ (no-show probability) under no-show-dependent time limit

pointment (reflected by the small x_1 value in three cases) and otherwise behaves similarly. Additionally, $E-D_q^{(n+1)}$ yields similar schedules when the no-show probability is low, but it starts overbooking towards the end when no-shows become more likely (reflected by the small x_i values, $i \geq 6$, when $1 - \nu_i = 0.4$). Similar pattern changes also happen in $E-D_q^{(2)}$ schedules: $E-D_q^{(2)}$ intends to double book later appointments when the no-show probability is low; but when no-shows become more likely, it double books earlier appointments as well. An intuitive explanation is that as the time limit T becomes shorter and no-shows become more likely, double booking can simultaneously mitigate long waiting time and long idle time (also reflected in Tables 4.6–4.7).

4.6 Concluding Remarks

In this chapter, we studied moment-based DR models of the stochastic appointment scheduling problem under uncertainty arising from no-shows and service durations. Our approaches are suitable for an appointment scheduler who has a limited amount of data and considers ambiguous distributions of the two co-existing uncertainties. We derived the following insights on DR appointment scheduling models: (i) one can improve the DR models' ability of utilizing distributional information by using reasonably conservative supports and (ii) the DR model with the least conservative support of no-shows obtains near-optimal schedules under perfect information, and outperforms other DR models and the stochastic program if the distributional type is misspecified.

Based on the computational results, we derived the following recommendations for the practitioners: (i) if the appointment scheduler has accurate information on the no-show probabilities and service duration distributions, then she can double book the first appointment to better reduce the disruptive effects of no-shows, (ii) if the distributional information is ambiguous, then she can double book the later appointments to yield more stable performance, and (iii) under ambiguous distributions, she should double book earlier appointments as well, as the no-show probabilities increase or as the time limit becomes shorter (relative to the number of appointments).

CHAPTER 5

Solving 0-1 Semidefinite Programs for Distributionally Robust Allocation of Surgery Blocks

5.1 Introductory Remarks

In this chapter, we study a server allocation problem in a specific context to allocate surgery blocks to operating rooms (ORs). The goal is to minimize the total cost of opening ORs, while guaranteeing a low risk of having OR overtime. The problem follows a standard setting of OR planning under random surgery durations that has been studied by a variety of recent papers from different aspects (see. e.g., Deng et al., 2016; Denton et al., 2010; Shylo et al., 2012). The majority of the related literature employs stochastic optimization methods that require full knowledge of the exact distribution of the underlying uncertainty. One of the main purposes of this chapter is to provide new paradigms for deriving more reliable solutions for the case when a decision maker only knows limited data from past experiences about the time duration of completing each surgery, but not the full distributional information.

We introduce the detailed problem context as follows. A surgery scheduler opens ORs with limited total operating hours for completing surgeries in a daily base. Each surgery can be operated in a given subset of ORs and has a random duration time. The probabilities of undesirable outcomes such as OR overtime need to be restricted, usually through some quality-of-service requirement specified before making the decisions. The random time for completing different surgeries could be correlated or independent in all our models. We also allow heterogeneous surgery durations, meaning that the random surgery time in different ORs may follow different distribution types and parameters.

In this chapter, we restrict the OR overtime through individual OR-based chance constraints, which yields an alternative formulation to the minimization of the expected arti-

cial penalty cost of OR overtime. The latter is traditionally seen in the stochastic programming literature (e.g., Denton et al., 2010). We refer to Deng et al. (2016); Shylo et al. (2012) for the motivation of applying chance-constrained programming for ensuring the quality of service in OR planning, including (i) the difficulty of accurately estimating the cost associated with unit time delay in each OR and (ii) the drastic change of optimal solutions given small changes of the penalty cost parameter.

In this chapter, we consider a DR chance-constrained model, which minimizes the total cost of operating ORs and assignments, while restricting the worst-case probability of OR overtime under distributional ambiguity. Such an approach is emerging and has shown great power for modeling various optimal decision making problems under limited data or rapidly-changing complex environment. We follow the well-developed DR optimization theories in Delage and Ye (2010), and use the mean vector and covariance matrix to construct the ambiguity set of the unknown distribution function of the jointly distributed surgery durations. Then, following a recent result in Jiang and Guan (2016) for solving generic DR chance constraints, we reformulate the DR model as a SDP problem but with binary variables (used for modeling decisions of opening ORs and surgery block assignment), which cannot be directly optimized in off-the-shelf solvers. To tackle the 0-1 SDP, we develop a cutting-plane approach that solves a continuous SDP oracle in each separation procedure for iteratively generating valid cuts. Also, we propose a 0-1 SOCP reformulation deduced from a tighter ambiguity set, solving which does not require implementing the cutting-plane algorithm. We test instances of an outpatient treatment application, to compare different approaches, specifically in terms of their solution time, in-sample objective values, out-of-sample reliability performance. We note that the work in this chapter has been published in Zhang et al. (2018b).

5.1.1 Literature Review

We categorize the most relevant literature into three categories as follows.

5.1.1.1 Distributionally Robust Optimization and Chance-Constrained Programming

DR approaches handle the issue of unknown distributions in traditional stochastic programming, by utilizing statistical information of historical data samples to prevent making decisions that will perform badly in some future circumstances. To illustrate, consider a stochastic program $\min_{a \in A} \mathbb{E}_{\xi}[c(a, \xi)]$ as an example, where a is a decision vector and ξ is a random parameter vector, both affecting a random cost $c(a, \xi)$. The goal is to minimize the expected cost for a given distribution of ξ with probability function $f(\xi)$. When the

exact form of $f(\xi)$ is not known, we optimize against the worst-case outcome, and build an ambiguity set \mathcal{D} of possible probability distribution functions for ξ . The DR problem, which minimizes the worst-case expected cost over possible $f(\xi) \in \mathcal{D}$, is given by

$$\min_{a \in A} \max_{f(\xi) \in \mathcal{D}} \mathbb{E}_{\xi}[c(a, \xi)].$$

The key challenge in DR optimization lies in constructing an ambiguity set \mathcal{D} of $f(\xi)$ to derive tractable yet not too conservative solutions depending on the amount of available data. Scarf et al. (1958) are the first to consider a DR inventory-control problem, with set \mathcal{D} consisting of linear constraints on moments of random demand. We refer to Bertsimas et al. (2010); Delage and Ye (2010) for the representative work that uses moment-based ambiguity sets for deriving reformulations of generic DR models.

Meanwhile, chance-constrained programs (cf. Shapiro et al., 2009) are commonly used for restricting the probability of undesirable outcomes to a small and controllable value. They provide powerful stochastic optimization tools for guaranteeing the quality of service under uncertainty. A DR variant of a chance constraint ensures that the probability of undesirable outcomes given by the worst-case distribution in an ambiguity set is no more than a given risk tolerance. Jiang and Guan (2016); Zymler et al. (2013) develop reformulations and solution methods for general DR chance constraints by constructing an ambiguity set based on the first two moments of random data. They derive equivalent SDPs for replacing the DR chance constraints by using the conic duality (Shapiro, 2001).

5.1.1.2 Stochastic or Robust Surgery Planning

Next, we review the most related literature in the optimization of healthcare systems, especially for surgery planning under uncertainty. A common objective is to find the best solution that balances conflicting criteria such as operational cost and quality of service (i.e., the chance of having no overtime in the open ORs focused in our chapter). Blake and Donald (2002) formulate an integer programming model for allocating similar surgery types to ORs. Denton et al. (2010) consider allocating surgery blocks of the same specialty with random time duration to ORs, where there exist a fixed cost of opening an OR and a unit cost for over-utilization that might occur in each open OR. The authors minimize the total cost of opening ORs and the expected penalty cost of overtime in all ORs. They formulate the problem as a two-stage stochastic mixed-integer program, solved by an integer L-shaped method using finite samples of surgery durations. Shylo et al. (2012) are among the first to use a chance constraint for restricting the OR overtime, while minimizing the total cost of opening ORs and the expected idle time in ORs. They assume that the surgery

durations follow a multivariate Normal distribution, and reformulate the problem as an equivalent semidefinite program. Deng et al. (2016) investigate an integrated problem of allocating and scheduling surgeries in multiple ORs, with additional binary variables that determine the sequence of performing surgeries in each OR. They consider a DR chance constraint on OR overtime due to unknown distributions of surgery durations. They reformulate the DR chance-constrained problem as a regular chance-constrained program with 0-1 variables based on ϕ -divergence functions used for constructing the ambiguity set.

5.1.1.3 Stochastic Bin Packing and Variants

By viewing each OR as a bin and all the surgeries as items to be packed in the bins, our model is a special case of the stochastic bin packing problem with random item sizes. Zhang et al. (2015b) study a DR chance-constrained model for stochastic bin packing by assuming the distributions of the random item sizes are ambiguous, but the random size of each item in their model does not change in different bins. They discuss a branch-and-price algorithm based on a column generation procedure. Focusing on an application of the stochastic bin packing problem, Deng et al. (2016) study a surgery planning problem that optimizes the allocation of surgeries to ORs and limits the the OR overtime using a joint DR chance-constraint, solved by using an ambiguity set constructed based on statistical ϕ -divergence measures. Shen and Wang (2014) consider the stochastic bin packing structure in the operations of Cloud Computing systems, and formulate a mixed-integer linear programming model for solving the problem based on finite samples of the random processing time of computational jobs.

Stochastic knapsack is another classical resource allocation problem, which packs a subset of items with random sizes to an open bin, and maximizes the total value of packed items. Both the stochastic knapsack and stochastic bin packing problems concern about uncertain item sizes in knapsack constraints. Kleinberg et al. (2000) relate the two models in bandwidth allocation problems in telecommunication, and propose approximation algorithms for solving the chance-constrained variant of the stochastic bin packing model by assuming that each random item size follows a Bernoulli distribution. Kosuch and Lisser (2010) focus on stochastic knapsack problem variants, one with simple recourse and the other with a chance constraint. They provide upper bounds in a branch-and-bound framework for optimizing the two model variants. Cheng et al. (2014) construct SDP relaxations for optimizing DR knapsack constraints. Han et al. (2016) propose a pseudo-polynomial time algorithm for a chance-constrained binary knapsack problem with Normally distributed item sizes.

5.1.2 Contributions and Organization of the chapter

The main contributions of the chapter are as follows. First, we consider the standard OR planning problem from a new angle by taking into account ambiguous distributions, and apply risk-averse measures to evaluate undesirable random outcomes (e.g., the OR overtime risk in our study). Second, we propose a DR chance-constrained model based on moment-based ambiguity sets of the unknown distribution, and formulate an equivalent 0-1 SDP. Third, we develop a cutting-plane algorithm and an alternative 0-1 SOCP for approximating the optimal solution to the 0-1 SDP. Fourth, we demonstrate the data-driven aspect of using the DR approach through cross-sample computational studies.

The remainder of the chapter is organized as follows. Section 5.2 describes the DR chance-constrained model and the moment-based distributional ambiguity set. In Section 5.3, we reformulate the problem as a 0-1 SDP and propose a cutting-plane algorithm for iteratively approximating the results. In Section 5.4, we extend the results in Delage and Ye (2010) to approximate the original ambiguity set, which results in an alternative model as a 0-1 SOCP. In Section 5.5, we compare the computational efficacy of the cutting-plane algorithm and directly solving the 0-1 SOCP. We test instances randomly generated from real data of hospital outpatient treatment time and compare the results of the two methods. Section 5.6 summarizes the chapter and states future research avenues.

Assumptions and Notation. We use $|X|$ to denote the cardinality of set X , and denote $X \cdot Y$ as the Frobenius inner product of X and Y , i.e., $X \cdot Y = \text{tr}(X^T Y)$. We denote \mathbb{S}_+^K as the set of symmetric positive semidefinite $K \times K$ matrices. The generalized inequality for symmetric matrices, $X \succeq Y$, where $X, Y \in \mathbb{S}_+^K$, means that $X - Y \in \mathbb{S}_+^K$. Similarly, $X \preceq Y$, means that $Y - X \in \mathbb{S}_+^K$. We use $\langle \cdot \rangle$ to denote a stacked column vector.

5.2 DR Chance-Constrained Model

Consider a set I of ORs and a set J of surgeries. For each pair of surgery $j \in J$ and OR $i \in I$, binary parameter ρ_{ij} indicates whether or not surgery j can be operated in OR i , such that $\rho_{ij} = 1$ if yes, and $\rho_{ij} = 0$ otherwise. Vector $s = \langle s_{ij}, i \in I, j \in J \rangle$ represents the random service durations, where s_{ij} is the time of completing surgery j in OR i . We assume a fixed cost τ_i^z and time limit T_i of operating OR i , $\forall i \in I$. Let τ_{ij}^y be a fixed cost of assigning surgery j in OR i , $\forall i \in I, j \in J$.

We define variables $z_i \in \{0, 1\}$, $i \in I$ such that $z_i = 1$ if we operate OR i and $z_i = 0$ otherwise. Also, we define variables $y_{ij} \in \{0, 1\}$ for the assignment of surgery j to OR i

for all $j \in J$ and $i \in I$, such that $y_{ij} = 1$ if surgery j is processed in OR i , and $y_{ij} = 0$ otherwise. We minimize the total cost of operating a subset of ORs in I to complete all the surgeries in J , and ensure that the probability of having overtime on each OR i is no more than α_i , where $0 < \alpha_i < 1$ is a given risk level for each $i \in I$. The DR chance-constrained model is given by:

$$\min_{\mathbf{z}, \mathbf{y}} \quad \sum_{i \in I} \tau_i^z z_i + \sum_{i \in I} \sum_{j \in J} \tau_{ij}^y y_{ij} \quad (5.1a)$$

$$\text{s.t.} \quad y_{ij} \leq \rho_{ij} z_i \quad \forall i \in I, j \in J \quad (5.1b)$$

$$\sum_{i \in I} y_{ij} = 1 \quad \forall j \in J \quad (5.1c)$$

$$y_{ij}, z_i \in \{0, 1\} \quad \forall i \in I, j \in J \quad (5.1d)$$

$$\inf_{f(s_i) \in \mathcal{D}_i} \mathbb{P}_{f(s_i)} \left\{ \sum_{j \in J} s_{ij} y_{ij} \leq T_i z_i \right\} \geq 1 - \alpha_i, \quad \forall i \in I \quad (5.1e)$$

where $\mathbb{P}_{f(s_i)} \{\cdot\}$ denotes the probability of event \cdot under a given distribution function $f(s_i)$ of the random vector $s_i = \langle s_{ij}, \forall j \in J \rangle$, for each OR $i \in I$. The objective function (5.1a) minimizes the total cost of operating ORs and assigning surgeries to open ORs. Constraints (5.1b) and (5.1c) ensure that surgeries are only assigned to open ORs, and each surgery can only be allocated to one OR, respectively. Constraints (5.1d) ensure binary values of both variables z_i and y_{ij} . In the DR chance constraints (5.1e), we designate a moment-based ambiguity set, \mathcal{D}_i , of the pdf $f(s_i)$. The sum, $\sum_{j \in J} s_{ij} y_{ij}$, is the total service time for completing all the surgeries on OR i , which needs to be no more than T_i with at least probability $1 - \alpha_i$, for any PDF in the ambiguity set \mathcal{D}_i . That is, we ensure the worst-case (i.e., the minimum) probability of on-time surgery completion at each OR i being no less than $1 - \alpha_i$, for each $i \in I$.

In this chapter, we utilize the moment-based ambiguity set in Delage and Ye (2010) to characterize possible PDF $f(s_i)$ for each $i \in I$ as follows. We consider $\mathcal{D} = \cup_{i \in I} \mathcal{D}_i$, with set \mathcal{D}_i being specified as

$$\mathcal{D}_i^M(\Xi_i, \mu_i^0, \Sigma_i^0, \gamma_1, \gamma_2) = \left\{ f(s_i) : \begin{array}{l} \int_{s_i \in \Xi_i} f(s_i) ds_i = 1 \\ (\mathbb{E}[s_i] - \mu_i^0)^\top (\Sigma_i^0)^{-1} (\mathbb{E}[s_i] - \mu_i^0) \leq \gamma_1 \\ \mathbb{E}[(s_i - \mu_i^0)(s_i - \mu_i^0)^\top] \preceq \gamma_2 \Sigma_i^0 \end{array} \right\}, \quad (5.2)$$

where Ξ_i denotes the support of random parameter s_i , and we let $\Xi_i = \mathbb{R}^{|J|}$, $\forall i \in I$ (rather than $\mathbb{R}_+^{|J|}$) in our later derivation of the solution approaches. Moreover, μ_i^0 and Σ_i^0 represent the empirical mean vector and the covariance matrix of surgery-duration vector s_i , for all

$i \in I$, respectively. Scalars $\gamma_1, \gamma_2 \geq 0$ adjust the size of the ambiguity set \mathcal{D}_i^M . Note that the three constraints in \mathcal{D}_i^M ensure that (i) values of $f(s_i)$ sum to 1 over the support set Ξ_i ; (ii) the mean of s_i lies in an ellipsoid of size proportion to γ_1 centered at the empirical mean μ_i^0 ; (iii) the true covariance matrix lies in a positive semidefinite cone bounded by a matrix inequality of $\gamma_2 \Sigma_i^0$. Parameters γ_1 and γ_2 further control the conservativeness of optimal solutions given by the DR model that uses \mathcal{D}_i^M , for all $i \in I$ (Delage and Ye, 2010). One can also differentiate the values of γ_1 and γ_2 for different OR $i \in I$.

5.3 An Equivalent 0-1 SDP and Cutting-plane Algorithm

In this section, we reformulate the DR chance constraints (5.1e) with $\mathcal{D}_i = \mathcal{D}_i^M$ by using SDP constraints, and develop a cutting-plane algorithm for optimizing the resulting 0-1 SDP.

For each OR i , the worst-case probability in the DR chance constraint (5.1e) is equivalent to the optimal objective value of the following problem:

$$z_{\mathcal{D}_i^M} = \min_{f(s_i)} \int_{\mathbb{R}^{|J|}} \mathbb{I}_{\{s_i: \sum_{j \in J} s_{ij} y_{ij} \leq T_i z_i\}}(s_i) \quad (5.3a)$$

$$\text{s.t.} \quad \int_{\mathbb{R}^{|J|}} f(s_i) ds_i = 1 \quad (5.3b)$$

$$\int_{\mathbb{R}^{|J|}} \begin{bmatrix} \Sigma_i^0 & s_i - \mu_i^0 \\ (s_i - \mu_i^0)^\top & \gamma_1 \end{bmatrix} f(s_i) ds_i \succeq 0 \quad (5.3c)$$

$$\int_{\mathbb{R}^{|J|}} (s_i - \mu_i^0)(s_i - \mu_i^0)^\top f(s_i) ds_i \preceq \gamma_2 \Sigma_i^0, \quad (5.3d)$$

where $\mathbb{I}_A(\xi)$ is an indicator function given set A and uncertainty ξ , such that $\mathbb{I}_A(\xi) = 1$, if $\xi \in A$, and $\mathbb{I}_A(\xi) = 0$ otherwise. Constraints (5.3b)–(5.3d) are equivalent to the moment-matching constraints of the set \mathcal{D}_i^M in (5.2). The optimal objective value $z_{\mathcal{D}_i^M}$ returns the worst-case probability of having no overtime on OR i achieved over the ambiguity set \mathcal{D}_i^M . Therefore, we satisfy the DR chance constraints (5.1e) if $z_{\mathcal{D}_i^M} \geq 1 - \alpha_i$, for each $i \in I$.

Theorem 5.1. (Adopted from Jiang and Guan (2016)) Given a solution $\hat{y}_i = \langle \hat{y}_{ij}, \forall j \in J \rangle$ and \hat{z}_i , for OR i , the optimization problem (5.3) leads to the following SDP constraints that

are equivalent to the DR chance constraints:

$$\mathbf{SDP}(\hat{y}_i, \hat{z}_i): \quad \gamma_2 \Sigma_i^0 \cdot G_i + 1 - r_i + \Sigma_i^0 \cdot H_i + \gamma_1 q_i \leq \alpha_i \lambda_i \quad (5.4a)$$

$$\begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} - \begin{bmatrix} 0 & \frac{1}{2} \hat{y}_i \\ \frac{1}{2} \hat{y}_i^\top & \lambda_i + \hat{y}_i^\top \mu_i^0 - T_i \hat{z}_i \end{bmatrix} \succeq 0 \quad (5.4b)$$

$$\begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} \in \mathbb{S}_+^{(|J|+1) \times (|J|+1)}, \quad \begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix} \in \mathbb{S}_+^{(|J|+1) \times (|J|+1)}, \\ \lambda_i \geq 0. \quad (5.4c)$$

The proof follows Jiang and Guan (2016) that reformulates a generic DR chance constraint as an SDP, and we specify the proof details for our problem in Appendix B.1.

Following Theorem 5.1, we can formulate $|I|$ such SDPs for replacing all the DR chance constraints in (5.1e). As a result, the DR chance-constrained model (5.1) is equivalent to solving a 0-1 SDP as

$$\min_{z, y} \left\{ \sum_{i \in I} \tau_i^z z_i + \sum_{i \in I} \sum_{j \in J} \tau_{ij}^y y_{ij} : (5.1b), (5.1c), (5.1d), (5.4a)–(5.4c) \text{ in } \mathbf{SDP}(y_i) \forall i \in I \right\}. \quad (5.5)$$

However, formulation (5.5) cannot be directly optimized by off-the-shelf solvers and we derive a cutting-plane algorithm as follows. We iteratively solve the following relaxed master problem and subproblems consisting of SDP constraints (5.4a)–(5.4c), to verify whether a given allocation solution satisfies the DR chance constraint (5.1e) for each $i \in I$. The relaxed master problem is given by

$$[\mathbf{MP}]: \quad \min_{z, y} \quad \sum_{i \in I} \tau_i^z z_i + \sum_{i \in I} \sum_{j \in J} \tau_{ij}^y y_{ij} \quad (5.6a)$$

$$\text{s.t.} \quad (5.1b), (5.1c), (5.1d)$$

$$C_i^l y_i \leq c_i^l z_i, \quad i \in I, \quad l = 1, \dots, k_i \quad (5.6b)$$

where in constraints (5.6b), for $i \in I$, $C_i^l y_i \leq c_i^l z_i$, $l = 1, \dots, k_i$ are sets of cuts associated with OR i cumulatively added by solving the subproblems described later. The parameters C_i^l and c_i^l are the coefficients of variable y_i and of variable z_i of each cut i , respectively.

At each iteration, given a tentative binary solution (\hat{y}, \hat{z}) , where $\hat{y} = \langle \hat{y}_i, i \in I \rangle$ and $\hat{z} = \langle \hat{z}_i, i \in I \rangle$, from a relaxation of MP, we check its feasibility for the DR chance constraints (5.1e) by solving subproblems derived from the $\mathbf{SDP}(\hat{y}_i, \hat{z}_i)$, $\forall i \in I$ in Theorem

5.1:

$$\begin{aligned} \text{[SUB}^i(\hat{y}_i, \hat{z}_i)\text{]: } P(\hat{y}_i, \hat{z}_i) &= \min_{G_i, p_i, r_i, H_i, q_i, \lambda_i} \quad \gamma_2 \Sigma_i^0 \cdot G_i + 1 - r_i + \Sigma_i^0 \cdot H_i + \gamma_1 q_i - \alpha_i \quad (5.4a) \\ &\text{s.t.} \quad (5.4b), (5.4c). \end{aligned}$$

Note that the optimal objective value of each subproblem $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$ needs to be no more than zero, and then the inequality (5.4a) in the $\text{SDP}(\hat{y}_i, \hat{z}_i)$ can be ensured. Otherwise, (5.4a) is violated and so is the corresponding DR chance constraint (5.1e) associated with OR i . To derive a valid cut, we examine the dual of $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$ given by

[SUBⁱ(\hat{y}_i, \hat{z}_i)-Dual]:

$$D(\hat{y}_i, \hat{z}_i) = \max_{Q_i, d_i, u_i, v_i} \quad \hat{y}_i^\top d_i + (\hat{y}_i^\top \mu_i^0 - T_i \hat{z}_i) u_i \quad (5.8a)$$

$$\text{s.t.} \quad \begin{bmatrix} \gamma_2 \Sigma_i^0 & v_i \\ v_i^\top & 1 \end{bmatrix} - \begin{bmatrix} Q_i & d_i \\ d_i^\top & u_i \end{bmatrix} \succeq 0 \quad (5.8b)$$

$$u_i - \alpha_i \geq 0 \quad (5.8c)$$

$$\begin{bmatrix} \Sigma_i^0 & -v_i \\ -v_i^\top & \gamma_1 \end{bmatrix} \succeq 0 \quad (5.8d)$$

$$v_i \in \mathbb{R}^{|J|}, \quad \begin{bmatrix} Q_i & d_i \\ d_i^\top & u_i \end{bmatrix} \in \mathbb{S}_+^{(|J|+1) \times (|J|+1)}, \quad (5.8e)$$

where $\begin{bmatrix} Q_i & d_i \\ d_i^\top & u_i \end{bmatrix}$ are dual matrix variables associated with the constraint (5.4b) in $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$, and v_i is the dual vectors corresponding to vector p_i . Furthermore, the strong duality holds, i.e., $P(\hat{y}_i, \hat{z}_i) = D(\hat{y}_i, \hat{z}_i)$ for any given (\hat{y}_i, \hat{z}_i) . Note that $P(\hat{y}_i, \hat{z}_i) \leq 0, \forall i \in I$ implies the satisfaction of every chance constraint in (5.1e). Until every $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$ for each i has their optimal objective value ≤ 0 for a solution (\hat{y}, \hat{z}) given by MP in (5.6), we claim its optimality to the original 0-1 SDP and thus to Formulation (5.1).

When $D(\hat{y}_i, \hat{z}_i) > 0$ for any $i \in I$, due to the strong duality, we have $P(\hat{y}_i, \hat{z}_i) > 0$ for the same $i \in I$. Thus, we can generate a valid inequality (5.6b) specified as

$$(\hat{d}_i + \hat{u}_i \mu_i^0)^\top y_i - \hat{u}_i T_i z_i \leq 0, \quad (5.9)$$

into the MP in (5.6) to cut off (\hat{y}_i, \hat{z}_i) , where \hat{d}_i and \hat{u}_i are optimal solutions to $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$ -Dual. Since there are finite many binary solutions of (y, z) and each cut in the form of (5.9) cuts off at least one solution, the cutting-plane algorithm terminates in finite iterations.

The $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$ -Dual is a small-scale SDP with only continuous variables and can

be solved quickly in the state-of-the-art convex optimization solvers. Moreover, we can compute $\text{SUB}^i(\hat{y}_i, \hat{z}_i)$ -Dual for all $i \in I$ in parallel to further speed up the computation. The implementation details of the cutting-plane approach are given in Algorithm 5.1.

Algorithm 5.1: A cutting-plane algorithm for optimizing the reformulated 0-1 SDP (5.5) of the DR model (5.1) with the ambiguity set $\mathcal{D} = \mathcal{D}_i^M$

```

1 Set  $k_i = t = 0, \forall i \in I, l \leftarrow -\infty, \text{INF}=\text{FALSE}, \text{and } \text{CUTFOUND}=\text{TRUE}$ 
2 while  $\text{INF}=\text{FALSE}$  and  $\text{CUTFOUND}=\text{TRUE}$  do
3   Set  $\text{CUTFOUND}=\text{FALSE}, t \leftarrow t + 1$ . Solve the relaxed master problem MP. Let
    $(\hat{z}^t, \hat{y}^t)$  be an optimal solution and  $\text{obj}^t$  be the optimal objective value. If the
   problem is infeasible, set  $\text{INF}=\text{TRUE}$ 
4   for  $i \in I$  do
5     Solve  $\text{SUB}^i(\hat{y}_i^t, \hat{z}_i^t)$ -Dual (5.8), and let  $(\hat{Q}_i, \hat{d}_i, \hat{u}_i, \hat{v}_i)$  be an optimal solution
6     if  $(\hat{d}_i + \hat{u}_i \mu_i^0)^\top \hat{y}_i^t > \hat{u}_i T_i \hat{z}_i^t$  then
7       Let  $C_i^{k_i+1} = (\hat{d}_i + \hat{u}_i \mu_i^0)^\top$  and  $c_i^{k_i+1} = \hat{u}_i T_i$  as the cut coefficients in
       (5.6b).
8       Add the corresponding cut (5.6b) (or equivalently (5.9)) into the MP.
9       Set  $k_i \leftarrow k_i + 1$  and  $\text{CUTFOUND}=\text{TRUE}$ 
10    end
11  end
12  Set  $l \leftarrow \text{obj}^t$  in MP.
13 end
14 if  $\text{INF}=\text{TRUE}$  then
15   Claim that the DR model (5.1) is infeasible.
16 else
17   Return  $(\hat{z}^t, \hat{y}^t)$  as one optimal solution to the 0-1 SDP (5.5) as well as to the DR
   model (5.1).
18 end

```

5.4 An Alternative 0-1 SOCP Approximation

Alternatively, we propose a method to approximate the optimal results, which requires solving a 0-1 SOCP without implementing of a cutting-plane algorithm. The derivation is based on a general result given in Proposition 4 first shown by Wagner (2008), specified in our notation below.

Proposition 4. Denote an ambiguity set

$$\mathcal{D}_i^C = \mathcal{D}_i^C(\Xi_i, \mu_i, \Sigma_i) = \left\{ f(s_i) : \begin{array}{l} \int_{s_i \in \Xi_i} f(s_i) ds_i = 1, \\ \mathbb{E}[s_i] = \mu_i \\ \mathbb{E}[(s_i - \mu_i)(s_i - \mu_i)^\top] = \Sigma_i \end{array} \right\}. \quad (5.10)$$

Given risk parameter α_i , the mean vector μ_i and covariance matrix Σ_i of the service duration vector s_i , an individual DR chance constraint (i.e., the inner constraint only has one row)

$$\inf_{f(s_i) \in \mathcal{D}_i^C} \mathbb{P}_{f(s_i)} \{s_i^\top y_i \leq T_i z_i\} \geq 1 - \alpha_i \quad (5.11)$$

is equivalent to

$$\sqrt{y_i^\top \Sigma_i y_i} \leq \sqrt{\frac{\alpha_i}{1 - \alpha_i}} (T_i z_i - (\mu_i)^\top y_i), \quad \forall i \in I. \quad (5.12)$$

The proof follows Theorem 2.2 in Wagner (2008).

Theorem 5.2. Given the risk parameter α_i , $\forall i \in I$, the empirical mean vector μ_i^0 and covariance matrix Σ_i^0 of the service duration vector s_i , solutions to the DR chance constraints (5.1e) can be approximated by the following second-order conic constraints with binary variables y_i , $i \in I$:

$$\sqrt{\frac{1}{1 - a - b}} \left(1 + \sqrt{\frac{\alpha_i b}{1 - \alpha_i}}\right) \sqrt{y_i^\top \Sigma_i^0 y_i} \leq \sqrt{\frac{\alpha_i}{1 - \alpha_i}} (T_i z_i - (\mu_i^0)^\top y_i), \quad \forall i \in I. \quad (5.13)$$

Proof. Recall that the equivalence between the DR chance-constrained model (5.1) and the 0-1 SDP (5.5) given ambiguity set $\mathcal{D} = \mathcal{D}_i^M$, following the result in the previous Theorem 5.1. Furthermore, given empirical mean μ_i^0 and covariance Σ_i^0 of the set of samples $\{s_i^n\}_{n=1}^N$ of the random vector s_i , let μ_i and Σ_i be the true mean and covariance of each s_i , $i \in I$ based on the true PDF $f(s_i)$, respectively. Delage and Ye (2010) show the following: Consider the ambiguity set $\mathcal{D}_i^M = \mathcal{D}_i^M(\Xi_i, \mu_i^0, \Sigma_i^0, \gamma_1, \gamma_2)$, with

$$\gamma_1 = \frac{b}{1 - a - b}, \quad \gamma_2 = \frac{1 + b}{1 - a - b}.$$

for some scalars a and b . Now if (μ_i, Σ_i) satisfies

$$(\mu_i^0 - \mu_i)^\top (\Sigma_i)^{-1} (\mu_i^0 - \mu_i) \leq b, \quad (5.14)$$

$$\Sigma_i \preceq \frac{1}{1 - a - b} \Sigma_i^0, \quad (5.15)$$

then any $f(s_i) \in \mathcal{D}_i^C(\Xi_i, \mu_i, \Sigma_i)$ can satisfy the constraints in set $\mathcal{D}_i^M(\Xi_i, \mu_i^0, \Sigma_i^0, \gamma_1, \gamma_2)$ with probability greater than or equal to $1 - \delta$ (defined as the confidence level). Here the sample size N needs to be sufficiently large to obtain reliable empirical μ_i^0 and Σ_i^0 , and the threshold size value depends on the confidence level $1 - \delta$, support Ξ_i , and the dimension of the uncertainty s_i . (We refer the interested readers to Delage and Ye (2010) for detailed discussion about the theoretical sample size and its solution guarantee. The values of a and

b in (5.14)–(5.15) also depend on the confidence level $1 - \delta$, support Ξ_i , dimensionality, and moreover, the sample size N .)

Following the above result, we decompose the reformulation of DR chance constraint (5.1e) into two steps: First, we consider (5.11) with the ambiguity set $\mathcal{D}_i = \mathcal{D}_i^C(\Xi_i, \mu_i, \Sigma_i)$. Second, we require (5.11) holds for any μ_i and Σ_i that satisfy (5.14)–(5.15). According to Proposition 4, (5.11) is equivalent to (5.12) given μ_i and Σ_i , for all $i \in I$. Thus, the above two steps for reformulating the DR chance constraint (5.1e) for a given y_i are equivalent to

$$Z_2(y_i) = \max_{\mu_i, \Sigma_i} \{Z_1(y_i, \mu_i, \Sigma_i) : (5.14) - (5.15)\} \leq 0, \text{ for any feasible } y_i, i \in I, \quad (5.16)$$

where

$$Z_1(y_i, \mu_i, \Sigma_i) = \sqrt{y_i^\top \Sigma_i y_i} - \sqrt{\frac{\alpha_i}{1 - \alpha_i}} (T_i z_i - (\mu_i)^\top y_i). \quad (5.17)$$

Now to optimize model (5.16), we first maximize $Z_1(y_i, \mu_i, \Sigma_i)$ over variable μ_i , i.e., we maximize $(\mu_i)^\top y_i$ with constraint (5.14). Let $t = (\Sigma_i)^{-1/2}(\mu_i - \mu_i^0)$. We consider

$$\max_{\mu_i} (\mu_i)^\top y_i = \max_t ((\Sigma_i^0)^{\frac{1}{2}} y_i)^\top t + (\mu_i^0)^\top y_i \quad (5.18a)$$

$$\text{s.t.} \quad \|t\|_2 \leq \sqrt{b}. \quad (5.18b)$$

One solution is $\hat{t} = \sqrt{b}((\Sigma_i^0)^{\frac{1}{2}} y_i) / \|(\Sigma_i^0)^{\frac{1}{2}} y_i\|_2$, which is along the direction of $(\Sigma_i^0)^{\frac{1}{2}} y_i$ with maximum length of t , because (5.18) maximizes a linear objective function of t with coefficient $(\Sigma_i^0)^{\frac{1}{2}} y_i$ over a ball centered at the origin with radius \sqrt{b} .

Denote $\hat{\mu}_i$ as an optimal solution to the maximization problem (5.18), then

$$(\hat{\mu}_i)^\top y_i = ((\Sigma_i^0)^{\frac{1}{2}} y_i)^\top \hat{t} + (\mu_i^0)^\top y_i = \sqrt{b(y_i)^\top \Sigma_i^0 y_i} + (\mu_i^0)^\top y_i. \quad (5.19)$$

Now,

$$Z_1(y_i, \hat{\mu}_i, \Sigma_i) = \left(1 + \sqrt{\frac{\alpha_i b}{1 - \alpha_i}}\right) \sqrt{y_i^\top \Sigma_i y_i} - \sqrt{\frac{\alpha_i}{1 - \alpha_i}} (T_i z_i - (\mu_i^0)^\top y_i). \quad (5.20)$$

Replacing (5.17) with (5.20) in the overall formulation (5.16), we have an optimal solution to (5.16) as $\hat{\Sigma}^i = 1/(1 - a - b)\Sigma_i^0$. Therefore, the corresponding optimal objective value of (5.16) is

$$\hat{Z}_2(y_i) = \sqrt{\frac{1}{1 - a - b}} \left(1 + \sqrt{\frac{\alpha_i b}{1 - \alpha_i}}\right) \sqrt{y_i^\top \Sigma_i^0 y_i} - \sqrt{\frac{\alpha_i}{1 - \alpha_i}} (T_i z_i - (\mu_i^0)^\top y_i) \leq 0,$$

which is exactly (5.13). This completes the proof. \square

Following Theorem 5.2, we can reformulate the DR chance-constrained model as a 0-1 SOCP:

$$\min_{z,y} \left\{ \sum_{i \in I} \tau_i^z z_i + \sum_{i \in I} \sum_{j \in J} \tau_{ij}^y y_{ij} : (5.1b), (5.1c), (5.1d), (5.13) \right\}. \quad (5.21)$$

Remark 5.1. If the confidence level $1 - \delta$ is closed to 1, then almost all $f(s_i) \in \mathcal{D}_i^C(\Xi_i, \mu_i, \Sigma_i)$, with (μ_i, Σ_i) satisfying (5.14)–(5.15), is in the set \mathcal{D}_i^M . In such a case, the solutions to the 0-1 SOCP are less conservative than those to the DR chance-constrained model (5.1).

5.5 Computational Results

We conduct computational studies to demonstrate the results of the cutting-plane approach (i.e., Algorithm 5.1) for optimizing the 0-1 SDP (5.5) and directly solving the 0-1 SOCP (5.21). We consider the SAA approach (Luedtke and Ahmed, 2008) as a benchmark, which optimizes the MILP reformulation of the chance-constrained model. We test diverse instances, report the CPU time of the three approaches, and compare their solutions in terms of the in-sample results (e.g., the optimal objective value, and whether or not they satisfy the 0-1 SDP constraints (5.4a)–(5.4c)), as well as the out-of-sample performance (e.g., the reliability of having no overtime in each open OR given by the solutions).

5.5.1 Experimental Setup

We consider 32 surgeries and 6 ORs. Each OR has the same opening cost, and thus we set $\tau_i^z = 1$ for all $i \in I$. The cost of assignment y_{ij} , τ_{ij}^y uniformly distributes on the interval $[0, 0.1]$, $\forall i \in I, j \in J$. We assume that the duration time of each surgery follows Log-normal distribution, which has a long-tail property and has been shown in the literature well fitting the shape of the distribution type of service time (Berg et al., 2014). We consider four types of surgeries denoted as “ $\ell M \ell V$ ”, “ $\ell M h V$ ”, “ $h M \ell V$ ”, “ $h M h V$ ”, where “ ℓ ” and “ h ” refer to “low” and “high”, respectively, and “ M ” and “ V ” refer to “Mean” and “Variance”, respectively. Following (Berg et al., 2014), we use 12.5 minutes as the low mean of duration time, and use 25 minutes as the high mean. The low standard deviation is set as 0.3 of the mean value, and the high standard deviation equals to the mean duration. We evenly divide the 32 surgeries into four types and have 8 surgeries for each type, called “mixed” samples.

Using the Monte Carlo sampling approach, we generate training samples, having each surgery following a Log-Normal distribution with its mean and variance specified as above.

We generate 20 i.i.d. sampled scenarios of surgery durations (which can be considered as historical data points) to compute the empirical mean and covariance. We implement the cutting-plane algorithm for optimizing the 0-1 SDP and also solve the 0-1 SOCP approximation directly in solvers. We examine the results of optimal solutions given by the two approaches, including optimal objective values and whether the optimal solution to the 0-1 SOCP can satisfy constraints in the 0-1 SDP. According to Theorem 5.1, the latter verifies whether or not solutions given by 0-1 SOCP satisfy the original DR chance constraints (5.1e). For the MILP reformulation in the SAA approach, we generate 1000 scenarios by following the same distribution and Monte Carlo sampling.

We examine the out-of-sample performance of each solution by testing them in reference samples with 10,000 scenarios, representing realizations in practice. As the surgery scheduler may not know the exact distribution of random surgery durations while using the DR approach, the 20 in-sample data and the 10,000 out-of-sample data do not necessarily follow the same distribution. In specific, we simulate solutions given by the above three approaches in five sets of test samples, each of which has 10,000 i.i.d. scenarios. The five sets of data samples either have one type of surgeries from “ $\ell M \ell V$ ”, “ $\ell M h V$ ”, “ $h M \ell V$ ”, “ $h M h V$,” or we equally mix all the four types of surgeries. The duration time of each surgery follows a Log-Normal distribution with corresponding mean and standard deviation values.

Finally, each OR has 8-hour (i.e., 480-minute) time limit to operate, exceeding which the operation time is considered as overtime. We set parameter $(\gamma_1, \gamma_2) = (0, 1)$ in all the DR chance constraints associated with the 6 ORs. The computations are performed on a Windows 7 machine with Intel(R) Core(TM) i7-2600 CPU 3.40 GHz and 8GB memory. All continuous SDPs and 0-1 SOCPs are solved by CVX implemented in Matlab with MOSEK as the optimization solver.

5.5.2 Computational Results

Table 5.1 presents the optimal solutions and CPU time (in second) given by solving the 0-1 SDP via the cutting-plane algorithm, the 0-1 SOCP approximation, and the SAA-based MILP. We vary the reliability level $1 - \alpha_i$ in each chance constraint related to OR i as 95%, 90%, $\forall i \in I$.

In Table 5.1, the cutting-plane algorithm opens one more OR than the 0-1 SOCP for $1 - \alpha_i = 95\%$ and 90% , $\forall i \in I$. The MILP opens two ORs regardless of the choice of $1 - \alpha_i$. Consequently, the cutting-plane method yields the highest objective cost and more conservative solutions as compared to the other two methods. In terms of CPU time, the

Table 5.1: CPU time (in second) and optimal solutions given by the three benchmark approaches

$1 - \alpha_i$	Approach	CPU (sec)	Obj. Cost	# of open ORs	# of surgeries in each OR
95%	Cutting-plane	10.86	4.50	4	(8, 9, 7, 8)
	0-1 SOCP	124.50	3.66	3	(11, 11, 10)
	MILP	107.47	2.95	2	(16, 16)
90%	Cutting-plane	7.77	3.65	3	(12, 11, 9)
	0-1 SOCP	40.95	3.00	2	(16, 16)
	MILP	109.04	2.95	2	(16, 16)

cutting-plane approach significantly outperforms directly solving the 0-1 SOCP and MILP.

The 0-1 SOCP does not necessarily provide exact solutions to the original DR chance-constrained model (5.1). Thus, we check the feasibility of our solutions in the constraints in $\text{SDP}(\hat{y}_i)$ for each given solution \hat{y}_i . We generate 10 training samples, each with 20 scenarios, and optimize them by solving the corresponding 0-1 SOCPs. Out of the ten optimal solutions, only seven solutions satisfy the SDP constraints.

In the out-of-sample simulation, besides instances with equally mixed surgery types, we also simulate the solutions in test samples that have only one type of surgeries. All out-of-sample tests contain 10,000 data points. We omit the performance details in test samples having “mixed”, “ $\ell M \ell V$ ”, “ $\ell M h V$ ”, and “ $h M \ell V$ ” surgeries, as the solutions given by all the three approaches perform equally well. That is, all the solutions meet the the desired reliability $1 - \alpha_i$ with respect to each OR $i \in I$ in all the four cases.

Table 5.2: Average reliability performance of optimal solutions of the three approaches in test samples with only “ $h M h V$ ” surgeries

$1 - \alpha_i$	Approach	OR #1	OR #2	OR #3	OR #4
95%	Cutting-plane	0.99	0.99	1.00	0.99
	0-1 SOCP	0.98	0.98	N/A*	0.99
	MILP	0.81	N/A	N/A	0.82
90%	Cutting-plane	0.96	0.98	N/A	0.99
	0-1 SOCP	0.81	0.81	N/A	N/A
	MILP	0.81	N/A	N/A	0.82

* N/A: the corresponding OR is not open and thus no reliability result.

In Table 5.2, we report the average reliability of the ten solutions given by each approach in the test sample with all surgeries being “ $h M h V$.” The out-of-sample reliability of having no overtime in each open OR is shown under the last four columns in Table 5.2. The solutions of the cutting-plane method can achieve very high reliability in all the open ORs (usually greater than 95%), while the 0-1 SOCP leads to solutions that achieve lower reli-

bility values than the desired ones, especially when $1 - \alpha_i = 90\%$. In particular, when the desired reliability values of having no overtime in each open OR are 90%, the 0-1 SOCP can only provide solutions that achieve 0.81 probability of having no overtime in ORs #1 and #2. Under $1 - \alpha_i = 95\%$ and 90%, the reliability levels achieved by MILP solutions are quite low in all open ORs. This result again demonstrates that the exact 0-1 SDP (solved by the cutting-plane algorithm) is more conservative than the 0-1 SOCP approximation and the SAA-based MILP.

5.6 Concluding Remarks

In this chapter, we considered OR planning under random surgery durations, and incorporated chance constraints to restrict undesirable OR overtime. We studied a DR model variant by assuming unknown distributions of the uncertainty, and built an ambiguity set based on the first and the second moments of random service durations. The DR chance constraints can be replaced by SDP constraints, but the replacement resulted in a 0-1 SDP with binary variables indicating the decisions of opening ORs and assigning surgeries. To tackle the 0-1 SDP, we developed a cutting-plane algorithm, and also a 0-1 SOCP approximation. Both approaches can be implemented in off-the-shelf solvers to provide exact or relaxed solutions to the original problem. Via computational studies, we demonstrated that the cutting-plane method yielded more conservative solutions, having relatively higher cost, and better out-of-sample reliability performance. Moreover, the cutting-plane method took less CPU time than the other approaches.

Our formulation and solution approaches can be applied to a wide variety of applications that involve server allocation and bin packing procedures (e.g., surgery planning, cloud computing server planning, etc.), under the circumstances that the service time is random and has unknown distributions. In our future research, we plan to analyze more related applications, and analyze special problem structures.

CHAPTER 6

Ambiguous Chance-Constrained Binary Programs under Mean-Covariance Information

6.1 Introductory Remarks

We consider chance-constrained binary programs that involve a set of individual chance constraints. More specifically, we consider I individual chance constraints and, for each $i \in [I] := \{1, \dots, I\}$, let $y_i \in \{0, 1\}^J$ be a binary decision vector such that $y_i := [y_{i1}, \dots, y_{iJ}]^\top$ and \tilde{t}_i be the corresponding random coefficients such that $\tilde{t}_i := [\tilde{t}_{i1}, \dots, \tilde{t}_{iJ}]^\top$. Then, we consider the following individual chance constraints:

$$\mathbb{P}\left\{\tilde{t}_i^\top y_i \leq T_i\right\} \geq 1 - \alpha_i \quad \forall i \in [I], \quad (6.1)$$

where $T_i \in \mathbb{R}$, \mathbb{P} represents the joint probability distribution of $\{\tilde{t}_{ij} : i \in [I], j \in [J]\}$, and each α_i represents an allowed risk tolerance of constraint violation that often takes a small value (e.g., $\alpha_i = 0.05$).

The individual chance constraints have wide applications in service and operations management, providing an effective and convenient way of controlling capacity violation and ensuring high quality of service. For example, in surgery allocation, y_i represents yes-no decisions of allocating J surgeries in operating rooms (OR) i , for all $i \in [I]$. The operational time limit of each OR (i.e., T_i) is usually deterministic, but the processing time of each surgery (i.e., \tilde{t}_{ij}) is usually random due to the variety of patients, surgical teams, and surgery characteristics. Then, chance constraints (6.1) make sure that each OR will not go overtime with a large probability, offering an appropriate “end-of-the-day” guarantee.

Chance-constrained programs are difficult to solve, mainly because the feasible region described by constraints (6.1) is non-convex in general (Prékopa, 2003). Nonetheless, promising special cases have been identified to recapture the convexity of chance-

constrained models. In particular, if $\{\tilde{t}_{ij} : j \in [J]\}$ are assumed to follow a Gaussian distribution with a known mean μ_i and covariance matrix Σ_i , then the chance constraints (6.1) are equivalent to the Second-order Cone (SOC) constraints

$$\mu_i^\top y_i + \Phi^{-1}(1 - \alpha_i) \sqrt{y_i^\top \Sigma_i y_i} \leq T_i \quad \forall i \in [I], \quad (6.2)$$

where $\Phi(\cdot)$ represents the cumulative distribution function of the standard Gaussian distribution. In this case, feasible binary solutions y_i to constraints (6.2) can be quickly found by off-the-shelf optimization solvers. In another promising research stream, the probability distribution \mathbb{P} of \tilde{t}_{ij} is replaced by a finite-sample approximation, leading to a SAA of the chance-constrained model (Luedtke and Ahmed, 2008; Pagnoncelli et al., 2009). The SAA model is then recast as a mixed-integer linear program (MILP), on which many strong valid inequalities can be derived to accelerate the branch-and-cut algorithm (see, e.g., Küçükyavuz, 2012; Liu et al., 2016; Luedtke, 2014; Luedtke et al., 2010; Song et al., 2014).

However, a basic challenge of the chance-constrained approach is that the perfect knowledge of probability distribution \mathbb{P} may not be accessible. Under many circumstances, we only have a series of historical data that can be considered as samples taken from the true (while ambiguous) distribution. As a consequence, the solution obtained from a chance-constrained model can be sensitive to the choice of distribution \mathbb{P} we employ in (6.1) and hence perform poorly in out-of-sample tests. This phenomenon is often observed when solving stochastic programs and is called the *optimizer's curse* (Smith and Winkler, 2006). A natural way of addressing this curse is that, instead of a single estimate of \mathbb{P} , we employ a set of plausible probability distributions, termed the ambiguity set and denoted \mathcal{D} . Then, from a robust perspective, we ensure that chance constraints (6.1) hold valid with regard to all probability distributions belonging to \mathcal{D} , i.e.,

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left\{ \tilde{t}_i^\top y_i \leq T_i \right\} \geq 1 - \alpha_i \quad \forall i \in [I], \quad (6.3)$$

and accordingly we call (6.3) *distributionally robust chance constraints* (DRCCs).

In this chapter, we consider DCBP that involve binary y_i -variables and DRCCs (6.3). Without making the Gaussian assumption on \mathbb{P} , we show that a DCBP is equivalent to a 0-1 SOC program when \mathcal{D} is characterized by the first two moments of \tilde{t}_i . Furthermore, building upon existing work on valid inequalities for submodular/supermodular functions, we exploit the submodularity of the 0-1 SOC program to derive valid inequalities. As demonstrated in extensive computational experiments of bin packing instances, these valid

inequalities significantly accelerate the branch-and-cut algorithm for solving the related DCBPs. Notably, the proposed submodular approximations and the resulting valid inequalities apply to general 0-1 SOC programs than DCBPs.

The remainder of the chapter is organized as follows. Section 6.2 reviews the prior work related to optimization techniques used in this chapter and stochastic bin packing problems. Section 6.3 presents two 0-1 SOC representations, respectively, for DRCCs under two moment-based ambiguity sets. Section 6.4 utilizes submodularity and lifting to derive valid inequalities to strengthen the 0-1 SOC formulations. Section 6.5 demonstrates the computational efficacy of our approaches for solving different 0-1 SOC reformulations of a DCBP for chance-constrained bin packing, with diverse problem sizes and parameter settings. Section 6.6 summarizes the chapter and discusses future research directions. We note that the work in this chapter has been published in Zhang et al. (2018a).

6.2 Prior Work

Chance-constrained binary programs with uncertain technology matrix and/or right-hand side are computationally challenging, largely because (i) the non-convexity of chance constraints and (ii) the discrete variables. The majority of existing literature requires full distributional knowledge of the random coefficients and applies the SAA approach to approximate the models as MILPs. For example, Song et al. (2014) considered a generic chance-constrained binary packing problem using finite samples of the random item weights, and derived lifted cover inequalities to accelerate the computation. For generic chance-constrained programs, the SAA approach and valid inequalities for the related MILPs have been well studied in the literature (see, e.g., Küçükyavuz, 2012; Luedtke, 2014; Luedtke and Ahmed, 2008; Luedtke et al., 2010). We also refer to Deng et al. (2016); Shen and Wang (2014); Shylo et al. (2012) for wide applications of chance-constrained binary programs, mainly in service systems and operations. As compared to the existing work, this chapter waives the assumption of full distributional information and only relies on the first two moments of the uncertainty.

DRO has received growing attention, mainly because it provides effective modeling and computational approaches for handling ambiguous distributions of random variables in stochastic programming by using available distributional information. Moment information has been widely used for building ambiguity sets in various DRO models (see, e.g., Bertsimas et al. (2010); Delage and Ye (2010); Wiesemann et al. (2014)). Using moment-based ambiguity sets, Calafiore and El Ghaoui (2006); Chen et al. (2010); Cheng et al. (2014); El Ghaoui et al. (2003); Jiang and Guan (2016); Wagner (2008); Zymler et al.

(2013) derived exact reformulations and/or approximations for DRCCs, often in the form of SDPs. In special cases, e.g., when the first two moments are *exactly* matched in the ambiguity set, the SDPs can further be simplified as SOC programs. While many existing ambiguity sets exactly match the first two moments of uncertainty (see, e.g., El Ghaoui et al., 2003; Wagner, 2008; Zymmler et al., 2013), Delage and Ye (2010) proposed a data-driven approach to construct an ambiguity set that can model moment estimation errors. In this chapter, we consider both types of moment-based ambiguity sets. To the best of our knowledge, for the first time, we provide an SOC representation of DRCCs using the general ambiguity set proposed by Delage and Ye (2010).

Meanwhile, DRO has received much less attention in discrete optimization problems, possibly due to the difficulty of solving 0-1 nonlinear programs. For example, most off-the-shelf solvers cannot directly handle 0-1 SDPs, which often arise from discrete optimization problems with DRCCs. To the best of our knowledge, our results on chance constraints are most related to Cheng et al. (2014) that studied DRCCs in the binary knapsack problem and derived 0-1 SDP reformulations. As compared to Cheng et al. (2014), we investigate a different ambiguity set and derive a 0-1 SOC representation. Additionally, we solve the 0-1 SOC reformulation to *global* optimality instead of considering an SDP relaxation as in Cheng et al. (2014).

In the seminal work Nemhauser et al. (1978), the authors identified submodularity in combinatorial and discrete optimization problems and proved a sufficient and necessary condition for 0-1 quadratic functions being submodular. We use this condition to exploit the submodularity of our 0-1 SOC reformulations. Indeed, submodular and supermodular knapsack sets (the discrete lower level set of a submodular function and discrete upper level set of a supermodular function, respectively) often arise when modeling utility, risk, and chance constraints on discrete variables. Extended polymatroid inequalities (see Edmonds (1970); Nemhauser and Wolsey (1999)) can be efficiently obtained through greedy-based separation procedures to optimize submodular/supermodular functions. Recently, Atamtürk and Bhardwaj (2015); Atamtürk and Narayanan (2009); Bhardwaj (2015) proposed cover and packing inequalities for efficiently solving submodular and supermodular knapsack sets with 0-1 variables. Our results on valid inequalities are most related to Atamtürk and Bhardwaj (2018); Bhardwaj (2015) that identified a sufficient condition for the submodularity of 0-1 SOC constraints and strengthened their formulations by using the corresponding extended polymatroid inequalities (see Section 2.2 of Atamtürk and Bhardwaj (2018)). In contrast, we derive a different way to exploit the submodularity of general 0-1 SOC constraints. In particular, we apply the sufficient and necessary condition derived by Nemhauser et al. (1978) to search for “optimal” submodular approximations of the 0-1

SOC constraints (see Section 6.4.1).

The main contributions of the chapter are three-fold. First, using the general moment-based ambiguity set proposed by Delage and Ye (2010), we equivalently reformulate DRCCs as 0-1 SOC constraints that can readily be solved by solvers. Second, we exploit the (hidden) submodularity of the 0-1 SOC constraints and derive extended polymatroid valid inequalities to accelerate solving DCBP. In particular, we provide an efficient way of finding “optimal” submodular approximations of the 0-1 SOC constraints in the original variable space, and furthermore show that any 0-1 SOC constraint possesses submodularity in a lifted space. The valid inequalities in original and lifted spaces can both be efficiently separated via the well-known greedy algorithm (see, e.g., Atamtürk and Gómez (2017); Atamtürk and Narayanan (2008); Edmonds (1970)). Third, we conduct extensive numerical studies to demonstrate the computational efficacy of our solution approaches.

6.3 DCBP Models and Reformulations

We study DRCCs under two alternatives of ambiguity set \mathcal{D} based on the first two moments of \tilde{t}_i , $i \in [I]$. The first ambiguity set, denoted \mathcal{D}_1 , exactly matches the mean and covariance matrix of each \tilde{t}_i . In contrast, the second ambiguity set, denoted \mathcal{D}_2 , considers the estimation errors of sample mean and sample covariance matrix (see Delage and Ye (2010)). In Section 6.3.1, we introduce these two ambiguity sets and their calibration based on historical data. In Section 6.3.2, we derive SOC representations of DRCC (6.3) under \mathcal{D}_1 and \mathcal{D}_2 , respectively. While the former case (i.e., (6.3) under \mathcal{D}_1) has been well studied (see, e.g., Calafiore and El Ghaoui (2006); El Ghaoui et al. (2003); Zymler et al. (2013)), to the best of our knowledge, this is the first work to show the SOC representation of the latter case based on general covariance matrices (i.e., (6.3) under \mathcal{D}_2).

6.3.1 Ambiguity Sets

Suppose that a series of independent historical data samples $\{\tilde{t}_i^n\}_{n=1}^N$ are drawn from the true probability distribution \mathbb{P} of \tilde{t}_{ij} . Then, the first two moments of \tilde{t}_i can be estimated by the sample mean and sample covariance matrix

$$\mu_i = \frac{1}{N} \sum_{n=1}^N \tilde{t}_i^n, \quad \Sigma_i = \frac{1}{N} \sum_{n=1}^N (\tilde{t}_i^n - \mu_i)(\tilde{t}_i^n - \mu_i)^\top.$$

Throughout this chapter, we assume that both Σ_i and the true covariance matrix of \tilde{t}_i are symmetric and positive definite. As promised by the law of large numbers, as the data size

N grows, μ_i and Σ_i converge to the true mean and true covariance matrix of \tilde{t}_i , respectively. Hence, when N takes a large value, a natural choice of the ambiguity set consists of all probability distributions that match the sample moments μ_i and Σ_i , i.e.,

$$\mathcal{D}_1 = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^J) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \mu_i, \\ \mathbb{E}_{\mathbb{P}}[(\tilde{t}_i - \mu_i)(\tilde{t}_i - \mu_i)^\top] = \Sigma_i, \forall i \in [I] \end{array} \right\},$$

where $\mathcal{P}(\mathbb{R}^J)$ represents the set of all probability distributions on \mathbb{R}^J .

Under many circumstances, however, the historical data can be inadequate. With a small N , there may exist considerable estimation errors in μ_i and Σ_i , which brings a layer of “moment ambiguity” into \mathcal{D}_1 and adds to the existing distributional ambiguity of \mathbb{P} . To address the moment ambiguity and take into account the estimation errors, Delage and Ye (2010) proposed an alternative ambiguity set

$$\mathcal{D}_2 = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^J) : \begin{array}{l} (\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] - \mu_i)^\top \Sigma_i^{-1} (\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] - \mu_i) \leq \gamma_1, \\ \mathbb{E}_{\mathbb{P}}[(\tilde{t}_i - \mu_i)(\tilde{t}_i - \mu_i)^\top] \preceq \gamma_2 \Sigma_i, \forall i \in [I] \end{array} \right\},$$

where $\gamma_1 > 0$ and $\gamma_2 > \max\{\gamma_1, 1\}$ represent two given parameters. Set \mathcal{D}_2 designates that (i) the true mean of \tilde{t}_i is within an ellipsoid centered at μ_i , and (ii) the true covariance matrix of \tilde{t}_i is bounded from above by $\gamma_2 \Sigma_i - (\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] - \mu_i)(\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] - \mu_i)^\top$ (note that $\mathbb{E}_{\mathbb{P}}[(\tilde{t}_i - \mu_i)(\tilde{t}_i - \mu_i)^\top] = \mathbb{E}_{\mathbb{P}}[(\tilde{t}_i - \mathbb{E}_{\mathbb{P}}[\tilde{t}_i])(\tilde{t}_i - \mathbb{E}_{\mathbb{P}}[\tilde{t}_i])^\top] + (\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] - \mu_i)(\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] - \mu_i)^\top$).

Delage and Ye (2010) offered a rigorous guideline for selecting γ_1 and γ_2 values (see Theorem 2 in Delage and Ye (2010)) so that \mathcal{D}_2 includes the true distribution of \tilde{t}_i with a high confidence level. In practice, we can select the values of γ_1 and γ_2 via cross validation. For example, we can divide the N data points into two halves. We estimate sample moments (μ_i^1, Σ_i^1) based on the first half of the data and (μ_i^2, Σ_i^2) based on the second half. Then, we characterize \mathcal{D}_2 based on (μ_i^1, Σ_i^1) , and select γ_1 and γ_2 such that probability distributions with moments (μ_i^2, Σ_i^2) belong to \mathcal{D}_2 .

6.3.2 SOC Representations of the DRCC

Now we derive SOC representations of DRCC (6.3) for all $i \in [I]$. For notation brevity, we define vector $y := [y_{i1}, \dots, y_{iJ}]$ and omit the subscript i throughout this section. First, we review the celebrated SOC representation of DRCC (6.3) under ambiguity set \mathcal{D}_1 in the following theorem.

Theorem 6.1. (Adapted from El Ghaoui et al. (2003), also see Wagner (2008)) The DRCC

(6.3) with $\mathcal{D} = \mathcal{D}_1$ is equivalent to the following SOC constraint:

$$\mu^\top y + \sqrt{\frac{1-\alpha}{\alpha}} \sqrt{y^\top \Sigma y} \leq T. \quad (6.4)$$

Theorem 6.1 shows that we can recapture the convexity of DRCC (6.3) by employing ambiguity set \mathcal{D}_1 to model the \tilde{t} uncertainty. Perhaps more surprisingly, in this case, the convex feasible region characterized by DRCC (6.3) is SOC representable. It follows that the continuous relaxation of the DCBP model is an SOC program, which can be solved very efficiently by standard nonlinear optimization solvers.

Next, we show that DRCC (6.3) under the ambiguity set \mathcal{D}_2 is also SOC representable. This implies that the computational complexity of the DCBP remains the same even if we take the moment ambiguity into account. We present the main result of this section in the following theorem.

Theorem 6.2. DRCC (6.3) with $\mathcal{D} = \mathcal{D}_2$ is equivalent to

$$\mu^\top y + \left(\sqrt{\gamma_1} + \sqrt{\left(\frac{1-\alpha}{\alpha}\right)(\gamma_2 - \gamma_1)} \right) \sqrt{y^\top \Sigma y} \leq T \quad (6.5a)$$

if $\gamma_1/\gamma_2 \leq \alpha$, and is equivalent to

$$\mu^\top y + \sqrt{\frac{\gamma_2}{\alpha}} \sqrt{y^\top \Sigma y} \leq T \quad (6.5b)$$

if $\gamma_1/\gamma_2 > \alpha$.

Rujeerapaiboon et al. (2015) considered an ambiguity set similar to \mathcal{D}_2 and derived an SOC representation of DRCCs in portfolio optimization under an assumption of weak sense white noise, i.e., the uncertainty is stationary and mutually uncorrelated over time (see Definition 4 and Theorem 5 in Rujeerapaiboon et al. (2015)). In contrast, the SOC representation in Theorem 6.2 holds for general covariance matrices. We prove Theorem 6.2 in two steps. In the first step, we project the random vector \tilde{t} and its ambiguity set \mathcal{D}_2 from \mathbb{R}^J to the real line, i.e., \mathbb{R} . This simplifies DRCC (6.3) as involving a one-dimensional random variable. In the second step, we derive optimal (i.e., worst-case) mean and covariance matrix in \mathcal{D}_2 that attain the worst-case probability bound in (6.3). We then apply Cantelli's inequality to finish the representation. We present the first step of the proof in the following lemma.

Lemma 6.1. Let \tilde{s} be a random vector in \mathbb{R}^J and $\tilde{\xi}$ be a random variable in \mathbb{R} . For a given

$y \in \mathbb{R}^J$, define ambiguity sets $\mathcal{D}_{\tilde{s}}$ and $\mathcal{D}_{\tilde{\xi}}$ as

$$\mathcal{D}_{\tilde{s}} = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^J) : \mathbb{E}_{\mathbb{P}}[\tilde{s}]^{\top} \Sigma^{-1} \mathbb{E}_{\mathbb{P}}[\tilde{s}] \leq \gamma_1, \quad \mathbb{E}_{\mathbb{P}}[\tilde{s} \tilde{s}^{\top}] \preceq \gamma_2 \Sigma \right\} \quad (6.6a)$$

and

$$\mathcal{D}_{\tilde{\xi}} = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}) : |\mathbb{E}_{\mathbb{P}}[\tilde{\xi}]| \leq \sqrt{\gamma_1} \sqrt{y^{\top} \Sigma y}, \quad \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] \leq \gamma_2 (y^{\top} \Sigma y) \right\}. \quad (6.6b)$$

Then, for any Borel measurable function $f : \mathbb{R}^J \rightarrow \mathbb{R}$, we have

$$\inf_{\mathbb{P} \in \mathcal{D}_{\tilde{s}}} \mathbb{P}\{f(y^{\top} \tilde{s}) \leq 0\} = \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}} \mathbb{P}\{f(\tilde{\xi}) \leq 0\}.$$

Proof: We first show that $\inf_{\mathbb{P} \in \mathcal{D}_{\tilde{s}}} \mathbb{P}\{f(y^{\top} \tilde{s}) \leq 0\} \geq \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}} \mathbb{P}\{f(\tilde{\xi}) \leq 0\}$. Pick a $\mathbb{P} \in \mathcal{D}_{\tilde{s}}$, and let \tilde{s} denote the corresponding random vector and $\tilde{\xi} = y^{\top} \tilde{s}$. It follows that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] &= y^{\top} \mathbb{E}_{\mathbb{P}}[\tilde{s}] \\ &\leq \max_{s: s^{\top} \Sigma^{-1} s \leq \gamma_1} y^{\top} s \\ &= \max_{z: \|z\|_2 \leq \sqrt{\gamma_1}} (\Sigma^{1/2} y)^{\top} z = \sqrt{\gamma_1} \sqrt{y^{\top} \Sigma y}, \end{aligned} \quad (6.7a)$$

where inequality (6.7a) is because $\mathbb{E}_{\mathbb{P}}[\tilde{s}]^{\top} \Sigma^{-1} \mathbb{E}_{\mathbb{P}}[\tilde{s}] \leq \gamma_1$. Similarly, we have $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}] \geq -\sqrt{\gamma_1} \sqrt{y^{\top} \Sigma y}$. Meanwhile, note that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] &= y^{\top} \mathbb{E}_{\mathbb{P}}[\tilde{s} \tilde{s}^{\top}] y \\ &\leq y^{\top} (\gamma_2 \Sigma) y = \gamma_2 (y^{\top} \Sigma y), \end{aligned} \quad (6.7b)$$

where inequality (6.7b) is because $\mathbb{E}_{\mathbb{P}}[\tilde{s} \tilde{s}^{\top}] \preceq \gamma_2 \Sigma$. Hence, the probability distribution of $\tilde{\xi}$ belongs to $\mathcal{D}_{\tilde{\xi}}$. It follows that $\inf_{\mathbb{P} \in \mathcal{D}_{\tilde{s}}} \mathbb{P}\{f(y^{\top} \tilde{s}) \leq 0\} \geq \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}} \mathbb{P}\{f(\tilde{\xi}) \leq 0\}$.

Second, we show that $\inf_{\mathbb{P} \in \mathcal{D}_{\tilde{s}}} \mathbb{P}\{f(y^{\top} \tilde{s}) \leq 0\} \leq \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}} \mathbb{P}\{f(\tilde{\xi}) \leq 0\}$. Pick a $\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}$, and let $\tilde{\xi}$ denote the corresponding random variable and $\tilde{s} = [\tilde{\xi} / (y^{\top} \Sigma y)] \Sigma y$. It follows that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{s}]^{\top} \Sigma^{-1} \mathbb{E}_{\mathbb{P}}[\tilde{s}] &= \mathbb{E}_{\mathbb{P}}[\tilde{\xi}]^2 \frac{y^{\top} \Sigma}{y^{\top} \Sigma y} \Sigma^{-1} \frac{\Sigma y}{y^{\top} \Sigma y} = \frac{\mathbb{E}_{\mathbb{P}}[\tilde{\xi}]^2}{y^{\top} \Sigma y} \\ &\leq \frac{\gamma_1 y^{\top} \Sigma y}{y^{\top} \Sigma y} = \gamma_1, \end{aligned} \quad (6.7c)$$

where inequality (6.7c) is because $|\mathbb{E}_{\mathbb{P}}[\tilde{\xi}]| \leq \sqrt{\gamma_1} \sqrt{y^\top \Sigma y}$. Meanwhile, note that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{s}\tilde{s}^\top] &= \mathbb{E}_{\mathbb{P}} \left[\tilde{\xi}^2 \frac{\Sigma y}{y^\top \Sigma y} \frac{y^\top \Sigma}{y^\top \Sigma y} \right] \\ &= \mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] \frac{(\Sigma y)(\Sigma y)^\top}{(y^\top \Sigma y)^2} \\ &\preceq \gamma_2 (y^\top \Sigma y) \frac{(\Sigma y)(\Sigma y)^\top}{(y^\top \Sigma y)^2} \end{aligned} \quad (6.7d)$$

$$\preceq \gamma_2 (y^\top \Sigma y) \frac{(y^\top \Sigma y) \Sigma}{(y^\top \Sigma y)^2} = \gamma_2 \Sigma, \quad (6.7e)$$

where inequality (6.7d) is because $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}^2] \leq \gamma_2 (y^\top \Sigma y)$ and inequality (6.7e) is because $(\Sigma y)(\Sigma y)^\top \preceq (y^\top \Sigma y) \Sigma$, which holds because for all $z \in \mathbb{R}^J$,

$$\begin{aligned} z^\top (\Sigma y)(\Sigma y)^\top z &= [(\Sigma^{1/2} z)^\top (\Sigma^{1/2} y)]^2 \\ &\leq \|\Sigma^{1/2} z\|^2 \|\Sigma^{1/2} y\|^2 \quad (\text{Cauchy-Schwarz inequality}) \\ &= (y^\top \Sigma y)(z^\top \Sigma z) \\ &= z^\top [(y^\top \Sigma y) \Sigma] z. \end{aligned}$$

Hence, the probability distribution of \tilde{s} belongs to $\mathcal{D}_{\tilde{s}}$. It follows that $\inf_{\mathbb{P} \in \mathcal{D}_{\tilde{s}}} \mathbb{P}\{f(y^\top \tilde{s}) \leq 0\} \leq \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}} \mathbb{P}\{f(\tilde{\xi}) \leq 0\}$ because $\tilde{\xi} = y^\top \tilde{s}$, and the proof is completed. \square

Remark 6.1. Popescu (2007) and Yu et al. (2009) showed a similar projection property for \mathcal{D}_1 , i.e., when the first two moments of \tilde{s} are exactly known. Lemma 6.1 employs a different transformation approach to show the projection property for \mathcal{D}_2 when these moments are ambiguous in the sense of Delage and Ye (2010).

We are now ready to finish the proof of Theorem 6.2.

Proof of Theorem 6.2: First, we define random vector $\tilde{s} = \tilde{t} - \mu$, random variable $\tilde{\xi} = y^\top \tilde{s}$, constant $b = T - \mu^\top y$, and set S such that

$$S = \{(\mu_1, \sigma_1) \in \mathbb{R} \times \mathbb{R}_+ : |\mu_1| \leq \sqrt{\gamma_1} \sqrt{y^\top \Sigma y}, \mu_1^2 + \sigma_1^2 \leq \gamma_2 y^\top \Sigma y\}.$$

It follows that

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{D}_2} \mathbb{P}\{\tilde{t}^\top y \leq T\} &= \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{s}}} \mathbb{P}\{y^\top \tilde{s} \leq b\} \\ &= \inf_{\mathbb{P} \in \mathcal{D}_{\tilde{\xi}}} \mathbb{P}\{\tilde{\xi} \leq b\} \end{aligned} \quad (6.8a)$$

$$= \inf_{(\mu_1, \sigma_1) \in S} \inf_{\mathbb{P} \in \mathcal{D}_1(\mu_1, \sigma_1^2)} \mathbb{P}\{\tilde{\xi} \leq b\}, \quad (6.8b)$$

where $\mathcal{D}_{\tilde{s}}$ and $\mathcal{D}_{\tilde{\xi}}$ are defined in (6.6a) and (6.6b), respectively, equality (6.8a) follows from Lemma 6.1, and equality (6.8b) decomposes the optimization problem in (6.8a) into two layers: the outer layer searches for the optimal (i.e., worst-case) mean and covariance, while the inner layer computes the worst-case probability bound under the given mean and covariance. For the inner layer, based on Cantelli's inequality, we have

$$\inf_{\mathbb{P} \in \mathcal{D}_1(\mu_1, \sigma_1^2)} \mathbb{P}\{\tilde{\xi} \leq b\} = \begin{cases} \frac{(b-\mu_1)^2}{\sigma_1^2 + (b-\mu_1)^2}, & \text{if } b \geq \mu_1, \\ 0, & \text{o.w.} \end{cases}$$

As DRCC (6.3) states that $\inf_{\mathbb{P} \in \mathcal{D}_2} \mathbb{P}\{\tilde{t}^\top y \leq T\} \geq 1 - \alpha > 0$, we can assume $b \geq \mu_1$ for all $(\mu_1, \sigma_1) \in S$ without loss of generality. That is,

$$b \geq \max_{(\mu_1, \sigma_1) \in S} \mu_1 = \sqrt{\gamma_1} \sqrt{y^\top \Sigma y}.$$

It follows that

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{D}_2} \mathbb{P}\{\tilde{t}^\top y \leq T\} &= \inf_{(\mu_1, \sigma_1) \in S} \frac{(b - \mu_1)^2}{\sigma_1^2 + (b - \mu_1)^2} \\ &= \inf_{(\mu_1, \sigma_1) \in S} \frac{1}{\left(\frac{\sigma_1}{b - \mu_1}\right)^2 + 1}. \end{aligned} \quad (6.8c)$$

Note that the objective function value in (6.8c) decreases as $\sigma_1/(b - \mu_1)$ increases. Hence, (6.8c) shares optimal solutions with the following optimization problem:

$$\inf_{(\mu_1, \sigma_1) \in S} -\left(\frac{\sigma_1}{b - \mu_1}\right). \quad (6.8d)$$

The feasible region of problem (6.8d) is depicted in the shaded area of Figure 6.1. Furthermore, we note that the objective function of (6.8d) equals to the slope of the straight line connecting points $(b, 0)$ and (μ_1, σ_1) (see Figure 6.1 for an example). It follows that an optimal solution (μ_1^*, σ_1^*) to problem (6.8d), and so to problem (6.8c), lies in one of the

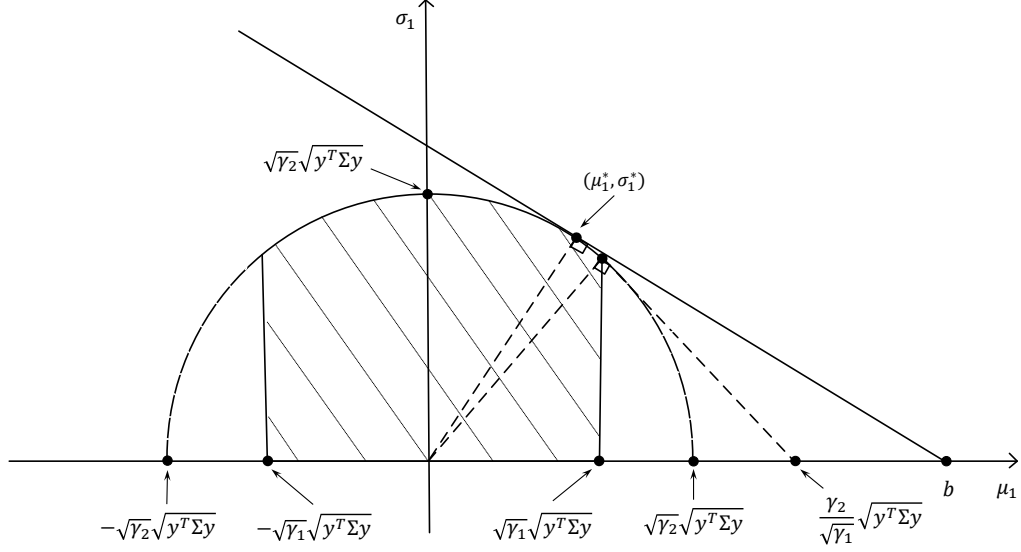


Figure 6.1: Graphical Solution of Problem (6.8d)

following two cases:

Case 1. If $\sqrt{\gamma_1} \sqrt{y^\top \Sigma y} \leq b \leq (\gamma_2 / \sqrt{\gamma_1}) \sqrt{y^\top \Sigma y}$, then $\mu_1^* = \sqrt{\gamma_1} \sqrt{y^\top \Sigma y}$ and $\sigma_1^* = \sqrt{\gamma_2 - \gamma_1} \sqrt{y^\top \Sigma y}$.

Case 2. If $b > (\gamma_2 / \sqrt{\gamma_1}) \sqrt{y^\top \Sigma y}$, then $\mu_1^* = (\gamma_2 y^\top \Sigma y) / b$ and $\sigma_1^* = \sqrt{\gamma_2 y^\top \Sigma y - (\gamma_2 y^\top \Sigma y)^2 / b^2}$.

Denoting $\kappa(b, y) = \frac{b}{\sqrt{y^\top \Sigma y}}$, we have

$$\inf_{\mathbb{P} \in \mathcal{D}_2} \mathbb{P}\{\tilde{t}^\top y \leq T\} = \begin{cases} \frac{1}{\left(\frac{\sqrt{\gamma_2 - \gamma_1}}{\kappa(b, y) - \sqrt{\gamma_1}}\right)^2 + 1}, & \text{if } \sqrt{\gamma_1} \leq \kappa(b, y) \leq \frac{\gamma_2}{\sqrt{\gamma_1}}, \\ \frac{\kappa(b, y)^2 - \gamma_2}{\kappa(b, y)^2}, & \text{if } \kappa(b, y) > \frac{\gamma_2}{\sqrt{\gamma_1}}. \end{cases} \quad (6.8e)$$

Second, based on (6.8e), the DRCC $\inf_{\mathbb{P} \in \mathcal{D}_2} \mathbb{P}\{\tilde{t}^\top y \leq T\} \geq 1 - \alpha$ has the following representations:

$$\text{DRCC} \Leftrightarrow \begin{cases} \frac{b}{\sqrt{y^\top \Sigma y}} \geq \sqrt{\gamma_1} + \sqrt{\left(\frac{1-\alpha}{\alpha}\right)(\gamma_2 - \gamma_1)}, & \text{if } \sqrt{\gamma_1} \leq \frac{b}{\sqrt{y^\top \Sigma y}} \leq \frac{\gamma_2}{\sqrt{\gamma_1}}, \\ \frac{b}{\sqrt{y^\top \Sigma y}} \geq \sqrt{\frac{\gamma_2}{\alpha}}, & \text{if } \frac{b}{\sqrt{y^\top \Sigma y}} > \frac{\gamma_2}{\sqrt{\gamma_1}}. \end{cases}$$

It follows that DRCC (6.3) is equivalent to an SOC constraint by discussing the following two cases:

Case 1. If $\gamma_1/\gamma_2 \leq \alpha$, then $\gamma_2/\sqrt{\gamma_1} \geq \sqrt{\gamma_1} + \sqrt{[(1-\alpha)/\alpha](\gamma_2 - \gamma_1)}$ and $\gamma_2/\sqrt{\gamma_1} \geq \sqrt{\gamma_2/\alpha}$. It follows that (i) if $\sqrt{\gamma_1} \leq b/\sqrt{y^\top \Sigma y} \leq \gamma_2/\sqrt{\gamma_1}$, then DRCC is equivalent to $b/\sqrt{y^\top \Sigma y} \geq \sqrt{\gamma_1} + \sqrt{[(1-\alpha)/\alpha](\gamma_2 - \gamma_1)}$ and (ii) if $b/\sqrt{y^\top \Sigma y} > \gamma_2/\sqrt{\gamma_1}$, then DRCC always holds. Combining sub-cases (i) and (ii) yields that DRCC (6.3) is equivalent to $b/\sqrt{y^\top \Sigma y} \geq \sqrt{\gamma_1} + \sqrt{[(1-\alpha)/\alpha](\gamma_2 - \gamma_1)}$.

Case 2. If $\gamma_1/\gamma_2 > \alpha$, then $\gamma_2/\sqrt{\gamma_1} < \sqrt{\gamma_1} + \sqrt{[(1-\alpha)/\alpha](\gamma_2 - \gamma_1)}$ and $\gamma_2/\sqrt{\gamma_1} < \sqrt{\gamma_2/\alpha}$. It follows that (i) if $\sqrt{\gamma_1} \leq b/\sqrt{y^\top \Sigma y} \leq \gamma_2/\sqrt{\gamma_1}$, then DRCC always fails and (ii) if $b/\sqrt{y^\top \Sigma y} > \gamma_2/\sqrt{\gamma_1}$, then DRCC is equivalent to $b/\sqrt{y^\top \Sigma y} \geq \sqrt{\gamma_2/\alpha}$. Combining sub-cases (i) and (ii) yields that DRCC (6.3) is equivalent to $b/\sqrt{y^\top \Sigma y} \geq \sqrt{\gamma_2/\alpha}$.

The proofs of the above two cases are both completed by the definition of b . □

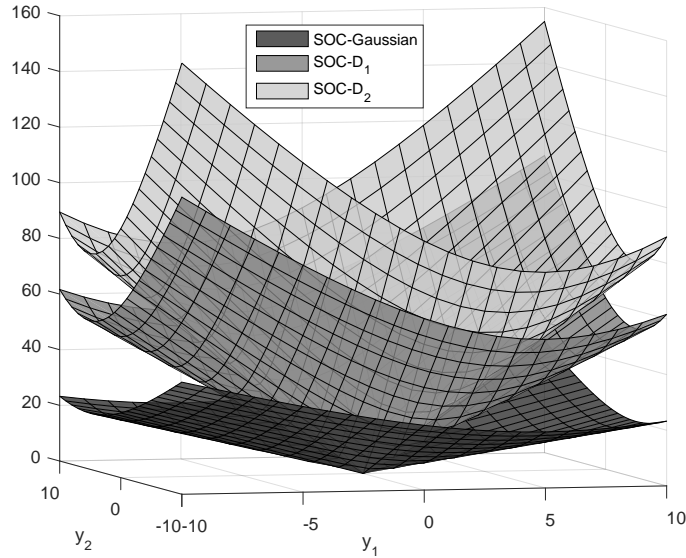


Figure 6.2: Illustration of the Three SOC Reformulations (6.2), (6.4), and (6.5b) of DRCC (6.3)

To sum up, we have two exact 0-1 SOC constraint reformulations of DRCC (6.3) under ambiguity sets \mathcal{D}_1 and \mathcal{D}_2 , being constraints (6.4), (6.5a)/(6.5b), respectively.

Example 6.1. We consider a DRCC (6.3) with $I = 1$ (the subscript i is hence omitted), $J = 2$, $1 - \alpha = 95\%$, mean vector $\mu = [0 \ 0]^\top$ and covariance matrix $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. For the ambiguity set \mathcal{D}_2 , we set $\gamma_1 = 1$ and $\gamma_2 = 2$. We note that $\Phi^{-1}(1 - \alpha) = 1.6449$ in the SOC reformulation (6.2), $\sqrt{(1 - \alpha)/\alpha} = 4.3589$ in the SOC reformulation (6.4), and

$\sqrt{\gamma_2/\alpha} = 6.3246$ in the SOC reformulation (6.5b) (since $\gamma_1/\gamma_2 > \alpha$). Without specifying the value of T , we depict the boundaries of the second-order cones associated with the three SOC reformulations in Figure 6.2. From this figure, we observe that the Gaussian approximation leads to the widest cone and so the largest SOC feasible region corresponding to DRCC (6.3). \square

6.4 Valid Inequalities for DCBP

Although 0-1 SOC constraint reformulations can be directly handled by the off-the-shelf solvers, as we report in Section 6.5, the resultant 0-1 SOC programs are often time-consuming to solve, primarily because of the binary restrictions on variables. In this section, we derive valid inequalities for DRCC (6.3), with the objective of accelerating the branch-and-cut algorithm for solving DCBP with individual DRCCs and also general 0-1 SOC programs in commercial solvers. Specifically, we exploit the submodularity of 0-1 SOC constraints. In Section 6.4.1, we derive a sufficient condition for submodularity and two approximations of the 0-1 SOC constraints that satisfy this condition. Using the submodular approximations, we derive extended polymatroid inequalities. In Section 6.4.2, we show the submodularity of the 0-1 SOC constraints in a lifted (i.e., higher-dimensional) space. While the extended polymatroid inequalities are well-known (see, e.g., Atamtürk and Gómez (2017); Atamtürk and Narayanan (2008); Edmonds (1970)), to the best of our knowledge, the two submodular approximations and the submodularity of 0-1 SOC constraints in the lifted space appear to be new and have not been studied in any literature.

6.4.1 Submodularity of the 0-1 SOC Constraints

We consider SOC constraints of the form

$$\mu^\top y + \sqrt{y^\top \Lambda y} \leq T, \quad (6.9)$$

where Λ represents a $J \times J$ positive definite matrix. Note that all SOC reformulations (6.4), (6.5a), and (6.5b) derived in Section 6.3, as well as the SOC reformulation (6.2) of a chance constraint with Gaussian uncertainty, possess the form of (6.9). Before investigating the submodularity of (6.9), we review the definitions of submodular functions and extended polymatroids.

Definition 6.1. Define the collection of set $[J]$'s subsets $\mathcal{C} := \{R : R \subseteq [J]\}$. A set function $h: \mathcal{C} \rightarrow \mathbb{R}$ is called submodular if and only if $h(R \cup \{j\}) - h(R) \geq h(S \cup \{j\}) - h(S)$ for all $R \subseteq S \subseteq [J]$ and all $j \in [J] \setminus S$.

Throughout this section, we refer to a set function as $h(R)$ and $h(y)$ interchangeably, where $y \in \{0, 1\}^J$ represents the indicating vector for subset $R \subseteq [J]$, i.e., $y_j = 1$ if $j \in R$ and $y_j = 0$ otherwise.

Definition 6.2. For a submodular function $h(y)$, the polyhedron

$$\text{EP}_h = \{\pi \in \mathbb{R}^J : \pi(R) \leq h(R), \quad \forall R \subseteq [J]\}$$

is called an extended polymatroid associated with $h(y)$, where $\pi(R) = \sum_{j \in R} \pi_j$.

For a submodular function $h(y)$ with $h(\emptyset) = 0$, inequality

$$\pi^\top y \leq t, \tag{6.10}$$

termed an *extended polymatroid inequality*, is valid for the epigraph of h , i.e.,

$$\{(y, t) \in \{0, 1\}^J \times \mathbb{R} : t \geq h(y)\}, \text{ if and only if } \pi \in \text{EP}_h \text{ (see Edmonds (1970))}.$$

Furthermore, the separation of (6.10) can be efficiently done by a greedy algorithm (Edmonds, 1970), which we briefly describe in Algorithm 6.1.

Algorithm 6.1: A greedy algorithm for separating extended polymatroid inequalities (6.10).

Input: A point (\hat{y}, \hat{t}) with $\hat{y} \in [0, 1]^J$, $\hat{t} \in \mathbb{R}$, and a function h that is submodular on $\{0, 1\}^J$.

1 Sort the entries in \hat{y} such that $y_{(1)} \geq \dots \geq y_{(J)}$. Obtain the permutation

$\{(1), \dots, (J)\}$ of $[J]$;

2 Letting $R_{(j)} := \{(1), \dots, (j)\}$, $\forall j \in \mathcal{J}$, compute $\hat{\pi}_{(1)} = h(R_{(1)})$, and

$\hat{\pi}_{(j)} = h(R_{(j)}) - h(R_{(j-1)})$ for $j = 2, \dots, J$;

3 **if** $\hat{\pi}^\top \hat{y} > \hat{t}$ **then**

4 | We generate a valid extended polymatroid inequality $\hat{\pi}^\top y \leq t$;

5 **else**

6 | The current solution (\hat{y}, \hat{t}) satisfies $h(\hat{y}) \leq \hat{t}$;

7 **end**

8 **return** either (\hat{y}, \hat{t}) is feasible, or a violated extended polymatroid inequality $\hat{\pi}^\top y \leq t$

The strength and efficient separation of the extended polymatroid inequality motivate us to explore the submodularity of function $g(y) := \mu^\top y + \sqrt{y^\top \Lambda y}$. In a special case, Λ is assumed to be a diagonal matrix and so the random coefficients $\tilde{t}_{ij}, j \in [J]$ for the same i are uncorrelated. In this case, Atamtürk and Narayanan (2008) successfully show that $g(y)$ is submodular. As a result, we can strengthen a DCBP by incorporating extended polymatroid inequalities in the form $\pi^\top y \leq T$, where $\pi \in \text{EP}_g$. Unfortunately, the submodularity of $g(y)$ quickly fades when the off-diagonal entries of Λ become non-zero, e.g., when Λ is associated with a general covariance matrix. We present an example as follows.

Example 6.2. Suppose that $[J] = \{1, 2, 3\}$, $\mu = [0, 0, 0]^\top$, and

$$\Lambda = \begin{bmatrix} 0.6 & -0.2 & 0.2 \\ -0.2 & 0.7 & 0.1 \\ 0.2 & 0.1 & 0.6 \end{bmatrix}.$$

The three eigenvalues of Λ are 0.2881, 0.7432, and 0.8687, and so $\Lambda \succ 0$. However, function $g(y) = \mu^\top y + \sqrt{y^\top \Lambda y}$ is not submodular because $g(R \cup \{j\}) - g(R) < g(S \cup \{j\}) - g(S)$, where $R = \{1\}$, $S = \{1, 2\}$, and $j = 3$. \square

In this section, we provide a sufficient condition for function $g(y)$ being submodular for general Λ . To this end, we apply a necessary and sufficient condition derived in Nemhauser et al. (1978) for quadratic function $y^\top \Lambda y$ being submodular (see the second paragraph on Page 276 of Nemhauser et al. (1978), following Proposition 3.5). We summarize this condition in the following theorem.

Theorem 6.3. (Nemhauser et al. (1978)) Define function $h : \{0, 1\}^J \rightarrow \mathbb{R}$ such that $h(y) := y^\top \Lambda y$, where $\Lambda \in \mathbb{R}^{J \times J}$ represents a symmetric matrix. Then, $h(y)$ is submodular if and only if $\Lambda_{rs} \leq 0$ for all $r, s \in [J]$ and $r \neq s$.

Note that Theorem 6.3 does not assume $\Lambda \succeq 0$ and so can be applied to general (convex or non-convex) quadratic functions. Theorem 6.3 leads to a sufficient condition for function $g(y)$ being submodular. We summarize the main result of this section in the following theorem and present a proof.

Theorem 6.4. Let $\Lambda \in \mathbb{R}^{J \times J}$ represent a symmetric and positive semidefinite matrix that satisfies (i) $2 \sum_{s=1}^J \Lambda_{rs} \geq \Lambda_{rr}$ for all $r \in [J]$ and (ii) $\Lambda_{rs} \leq 0$ for all $r, s \in [J]$ and $r \neq s$. Then, function $g(y) = \mu^\top y + \sqrt{y^\top \Lambda y}$ is submodular.

Proof: As $\mu^\top y$ is submodular in y , it suffices to prove that $\sqrt{y^\top \Lambda y}$ is submodular. Hence, we can assume $\mu = 0$ without loss of generality. We let $f(x) = \sqrt{x}$ and $h(y) = y^\top \Lambda y$. Then, $g(y) = f(h(y))$.

First, we note that $h(y \vee e_r) \geq h(y)$ for all $y \in \{0, 1\}^J$ and $r \in [J]$, where $a \vee b = [\max\{a_1, b_1\}, \dots, \max\{a_J, b_J\}]^\top$ for $a, b \in \mathbb{R}^J$. Indeed, if $y_r = 1$ then $y \vee e_r = y$ and so $h(y \vee e_r) = h(y)$. Otherwise, if $y_r = 0$, then $y \vee e_r = y + e_r$ and so

$$\begin{aligned} h(y \vee e_r) &= y^\top \Lambda y + 2e_r^\top \Lambda y + e_r^\top \Lambda e_r \\ &= y^\top \Lambda y + 2 \sum_{s: y_s=1} \Lambda_{rs} + \Lambda_{rr} \\ &\geq y^\top \Lambda y + 2 \sum_{\substack{s=1 \\ s \neq r}}^J \Lambda_{rs} + \Lambda_{rr} \\ &\geq y^\top \Lambda y = h(y), \end{aligned}$$

where the first inequality is due to $y_r = 0$ and condition (ii), and the second inequality is due to condition (i). It follows that $h(y') \geq h(y)$ for all $y, y' \in \{0, 1\}^J$ such that $y' \geq y$. Hence, $h(y)$ is increasing.

Second, based on Theorem 6.3, $h(y)$ is submodular due to condition (ii). It follows that $g(y) = f(h(y))$ is submodular because function f is concave and nondecreasing and function h is submodular and increasing (see, e.g., Proposition 2.2.5 in Simchi-Levi et al. (2014))¹. \square

6.4 generalizes the sufficient condition in Atamtürk and Narayanan (2008) because conditions (i)–(ii) are automatically satisfied if Λ is diagonal and positive definite.

For general $\Lambda \succ 0$ that does not satisfy sufficient conditions (i)–(ii), we can approximate SOC constraint (6.9) by replacing Λ with a matrix Δ that satisfies these conditions. We derive relaxed and conservative submodular approximations of constraint (6.9) in the following theorem.

Theorem 6.5. Constraint (6.9) implies the SOC constraint

$$\mu^\top y + \sqrt{y^\top \Delta^L y} \leq T, \quad (6.11)$$

where function $g^L(y) := \mu^\top y + \sqrt{y^\top \Delta^L y}$ is submodular and Δ^L is an optimal solution of

¹Proposition 2.2.5 in Simchi-Levi et al. (2014) assumes that $y \in \mathbb{R}^n$ and provides a sufficient condition for g being supermodular. It can be shown that a similar proof of this proposition applies to our case.

SDP

$$\min_{\Delta} \|\Delta - \Lambda\|_2 \quad (6.12a)$$

$$\text{s.t. } 0 \preceq \Delta \preceq \Lambda, \quad (6.12b)$$

$$2 \sum_{s=1}^J \Delta_{rs} \geq \Delta_{rr}, \quad \forall r \in [J], \quad (6.12c)$$

$$\Delta_{rs} \leq 0, \quad \forall r, s \in [J] \text{ and } r \neq s. \quad (6.12d)$$

Additionally, constraint (6.9) is implied by the SOC constraint

$$\mu^\top y + \sqrt{y^\top \Delta^u y} \leq T, \quad (6.13)$$

where function $g^u(y) := \mu^\top y + \sqrt{y^\top \Delta^u y}$ is submodular and Δ^u is an optimal solution of SDP

$$\min_{\Delta} \|\Delta - \Lambda\|_2 \quad (6.14a)$$

$$\text{s.t. } \Delta \succeq \Lambda, \quad (6.12c)\text{--}(6.12d). \quad (6.14b)$$

Proof: By construction, $g^l(y)$ is submodular because Δ^l satisfies constraints (6.12c)–(6.12d) and so conditions (i)–(ii). Additionally, constraint (6.9) implies (6.11) because Δ^l satisfies constraint (6.12b) and so $\Delta^l \preceq \Lambda$. Similarly, we obtain that $g^u(y)$ is submodular and constraint (6.9) is implied by (6.13). \square

Note that there always exist matrices Δ^l and Δ^u that are feasible to SDPs (6.12a)–(6.12d) and (6.14a)–(6.14b), respectively. For example, $\text{diag}(\lambda_{\min}, \dots, \lambda_{\min}) \in \mathbb{R}^{J \times J}$ satisfy constraints (6.12b)–(6.12d), where λ_{\min} represents the smallest eigenvalue of matrix Λ . Additionally, $\text{diag}(\lambda_{\max}, \dots, \lambda_{\max}) \in \mathbb{R}^{J \times J}$ satisfy constraints (6.14b), where λ_{\max} represents the largest eigenvalue of matrix Λ .

By minimizing the ℓ^2 distance between Δ and Λ in objective functions (6.12a) and (6.14a), we find “optimal” approximations of Λ that satisfies sufficient conditions (i)–(ii) in 6.4. Accordingly, we obtain “optimal” submodular approximations of the 0-1 SOC constraint (6.9). There are many possible alternatives of the ℓ^2 norm in (6.12a) and (6.14a). For example, formulations (6.12a)–(6.12d) and (6.14a)–(6.14b) remain SDPs if the ℓ^2 norm is replaced by the ℓ^1 norm or the ℓ^∞ norm. We have empirically tested ℓ^1 , ℓ^2 , and ℓ^∞ norms based on a server allocation problem (see Section 6.5.2 for a brief description) and the ℓ^2 norm leads to the largest improvement on CPU time. In computation, we only need to

solve these two SDPs once to obtain Δ^L and Δ^U . Then, extended polymatroid inequalities can be obtained from the relaxed approximation (6.11). Additionally, the conservative approximation (6.13) leads to an upper bound of the optimal objective value of the related DCBP.

Example 6.3. Recall the 3×3 matrix Λ in Example 6.2 and the corresponding function $g(y) = \mu^\top y + \sqrt{0.6y_1^2 + 0.7y_2^2 + 0.6y_3^2 - 0.4y_1y_2 + 0.4y_1y_3 + 0.2y_2y_3}$ being not submodular. We set $\mu = [0, 0, 0]^\top$ and apply 6.5 to optimize the two SDPs (6.12) and (6.14), yielding the following two positive semidefinite matrices:

$$\Delta^L = \begin{bmatrix} 0.35 & -0.15 & 0 \\ -0.15 & 0.37 & 0 \\ 0 & 0 & 0.38 \end{bmatrix} \text{ and}$$

$$\Delta^U = \begin{bmatrix} 0.83 & -0.22 & 0 \\ -0.22 & 0.95 & 0 \\ 0 & 0 & 0.82 \end{bmatrix}.$$

Due to 6.4, $g^L(y) := \mu^\top y + \sqrt{y^\top \Delta^L y} = \sqrt{0.35y_1^2 + 0.37y_2^2 + 0.38y_3^2 - 0.3y_1y_2}$ and $g^U(y) := \mu^\top y + \sqrt{y^\top \Delta^U y} = \sqrt{0.83y_1^2 + 0.95y_2^2 + 0.82y_3^2 - 0.44y_1y_2}$ are submodular.

Now suppose that $T = 0.8$ and we are given $\hat{y} = [\hat{y}_1, \hat{y}_2, \hat{y}_3]^\top = [1, 0.5, 0.9]^\top$ with $g(\hat{y}) = \mu^\top \hat{y} + \sqrt{\hat{y}^\top \Lambda \hat{y}} = 1.23 > 0.8 = T$. First, with respect to constraint $g^L(y) \leq T$, we follow 6.1 to find an extended polymatroid inequality and note that this inequality is also valid for constraint $g(y) \leq T$. Specifically, we sort the entries of \hat{y} to obtain $\hat{y}_1 \geq \hat{y}_3 \geq \hat{y}_2$ and $\{1, 3, 2\}$, the corresponding permutation. It follows that $R_{(1)} = \{1\}$, $R_{(2)} = \{1, 3\}$, and $R_{(3)} = \{1, 3, 2\}$. Hence, $\hat{\pi}_{(1)} = g^L([1, 0, 0]^\top) = 0.59$, $\hat{\pi}_{(2)} = g^L([1, 0, 1]^\top) - g^L([1, 0, 0]^\top) = 0.26$, and $\hat{\pi}_{(3)} = g^L([1, 1, 1]^\top) - g^L([1, 0, 1]^\top) = 0.04$. This generates the extended polymatroid inequality $0.59y_1 + 0.26y_3 + 0.04y_2 \leq 0.8$ that cuts off \hat{y} . Second, we can replace constraint $g(y) \leq T$ with $g^U(y) \leq T$ in a DCBP to obtain a conservative approximation. \square

6.4.2 Valid Inequalities in a Lifted Space

In Section 6.4.1, we derive extended polymatroid inequalities based on a relaxed approximation of SOC constraint (6.9). These valid inequalities might not be facet-defining if matrix Λ does not satisfy sufficient conditions (i)–(ii) in 6.4. In this section, we show that

the submodularity of (6.9) holds for general Λ in a lifted (i.e., higher-dimensional) space. Accordingly, we derive extended polymatroid inequalities in the lifted space.

To this end, we reformulate constraint (6.9) as $\mu^\top y \leq T$ and $y^\top \Lambda y \leq (T - \mu^\top y)^2$, i.e., $y^\top (\Lambda - \mu\mu^\top) y + 2T\mu^\top y \leq T^2$. Note that $\mu^\top y \leq T$ is because $T - \mu^\top y \geq \sqrt{y^\top \Lambda y} \geq 0$ by (6.9). Then, we define $w_{jk} = y_j y_k$ for all $j, k \in [J]$ and augment vector y to vector $v = [y_1, \dots, y_J, w_{11}, \dots, w_{1J}, w_{21}, \dots, w_{JJ}]^\top$. We can incorporate the following McCormick inequalities to define each w_{ij} :

$$w_{jk} \leq y_j, \quad w_{jk} \leq y_k, \quad w_{jk} \geq y_j + y_k - 1, \quad w_{ij} \geq 0. \quad (6.15)$$

Accordingly, we rewrite (6.9) as $[\mu^\top, 0^\top]v \leq T$ and

$$a^\top v + v^\top B_N v \leq T^2, \quad (6.16)$$

where we decompose $(\Lambda - \mu\mu^\top)$ to be the sum of two matrices, one containing all positive entries and the other containing all nonpositive entries. Accordingly, we define vector $a := [2T\mu; B_p]^\top$ with $B_p \in \mathbb{R}_+^{J^2}$ representing all the positive entries after vectorization, and matrix $B_N \in \mathbb{R}_-^{(J+J^2) \times (J+J^2)}$ collects all nonpositive entries, i.e., $B_N = \begin{bmatrix} -(\Lambda - \mu\mu^\top)_- & 0 \\ 0 & 0 \end{bmatrix}$, where $(x)_- = -\min\{0, x\}$ for $x \in \mathbb{R}$.

As $a^\top v + v^\top B_N v$ is a submodular function of v by Theorem 6.3, we can incorporate extended polymatroid inequalities to strengthen the lifted SOC constraints (6.9). We summarize this result in the following theorem.

Theorem 6.6. Define function $h : \{0, 1\}^{J+J^2} \rightarrow \mathbb{R}$ such that $h(v) := a^\top v + v^\top B_N v$. Then, h is submodular. Furthermore, inequality $\pi^\top v \leq T^2$ is valid for set $\{v \in \{0, 1\}^{J+J^2} : h(v) \leq T^2\}$ for all $\pi \in \text{EP}_h$ and the separation of this inequality can be done by Algorithm 6.1.

Note that this lifting procedure introduces J^2 additional variables w_{ijk} for each i . However, w_{ijk} can be treated as continuous variables when solving DCBP in view of the McCormick inequalities (6.15), and the number of w_{ijk} variables can be reduced by half because $w_{ijk} = w_{ikj}$. In our numerical studies later, we derive more valid inequalities to strengthen the formulation in the lifted space for distributionally robust chance-constrained bin packing problems that involve DRCCs, using the bin packing structure.

Example 6.4. Recall Example 6.2 and the corresponding function $g(y)$ being not submodular. We set $\mu = [0, 0, 0]^\top$ and $T = 0.8$, and rewrite the constraint $g(y) \leq T$ in the form of

(6.16) as

$$0.6v_4 + 0.2v_6 + 0.7v_8 + 0.1v_9 + 0.2v_{10} + 0.1v_{11} + 0.6v_{12} - 0.4v_1v_2 \leq 0.64,$$

where $v := [y_1, y_2, y_3, w_{11}, w_{12}, \dots, w_{33}]^\top = [v_1, \dots, v_{12}]^\top$. Now suppose that we are given $\hat{y} = [1, 0.5, 0.9]^\top$. The corresponding $\hat{v} = [1, 0.5, 0.9, 1, 0.5, 0.9, 0.5, 0.25, 0.45, 0.9, 0.45, 0.81]^\top$ violates the above inequality. We follow 6.1 to find an extended polymatroid inequality in the lifted space of v . Specifically, we sort the entries of \hat{v} to obtain $\hat{v}_1 \geq \hat{v}_4 \geq \hat{v}_3 \geq \hat{v}_6 \geq \hat{v}_{10} \geq \hat{v}_{12} \geq \hat{v}_2 \geq \hat{v}_5 \geq \hat{v}_7 \geq \hat{v}_9 \geq \hat{v}_{11} \geq \hat{v}_8$ and the corresponding permutation $\{1, 4, 3, 6, 10, 12, 2, 5, 7, 9, 11, 8\}$. It follows that $\hat{\pi}_{(2)} = 0.6$, $\hat{\pi}_{(4)} = 0.2$, $\hat{\pi}_{(5)} = 0.2$, $\hat{\pi}_{(6)} = 0.6$, $\hat{\pi}_{(7)} = -0.4$, $\hat{\pi}_{(10)} = 0.1$, $\hat{\pi}_{(11)} = 0.1$, $\hat{\pi}_{(12)} = 0.7$, and all other $\hat{\pi}_{(i)}$'s equal zero. This generates the following extended polymatroid inequality

$$0.6v_4 + 0.2v_6 + 0.2v_{10} + 0.6v_{12} - 0.4v_2 + 0.1v_9 + 0.1v_{11} + 0.7v_8 \leq 0.64$$

that cuts off \hat{v} . □

6.5 Numerical Studies

We numerically evaluate the performance of our proposed models and solution approaches. In Section 6.5.1, we present the formulation of a chance-constrained bin-packing problem with DRCCs and the related 0-1 SOC reformulations. We describe the solution methods and more valid inequalities based on the bin packing structure. In Section 6.5.2, we describe the experimental setup of the stochastic bin packing instances. Our results consist of two parts, which report the CPU time (Section 6.5.3) and the out-of-sample performance of solutions given by different models (Appendix C.3), respectively. More specifically, Section 6.5.3 demonstrates the computational efficacy of the valid inequalities we derived in Section 6.4 for the original or lifted SOC constraints. Appendix C.3 demonstrates that DCBP solutions can well protect against the distributional ambiguity as opposed to the solutions obtained by following the Gaussian distribution assumption or by the SAA method.

6.5.1 Formulation of Ambiguous Chance-Constrained Bin Packing

For the classical bin packing problem, $[I]$ is the set of bins and $[J]$ is the set of items, where each bin i has a weight capacity T_i and each item j , if assigned to bin i , has a random

weight \tilde{t}_{ij} . The deterministic bin packing problem aims to assign all J items to a minimum number of bins, while respecting the capacity of each bin. If we consider a slightly more general setting by introducing a cost for each assignment, then the DCBP of bin packing with DRCCs is presented as:

$$\min_{\mathbf{z}, \mathbf{y}} \quad \sum_{i=1}^I c_i^z z_i + \sum_{i=1}^I \sum_{j=1}^J c_{ij}^y y_{ij} \quad (6.17a)$$

$$\text{s.t.} \quad y_{ij} \leq \rho_{ij} z_i \quad \forall i \in [I], j \in [J] \quad (6.17b)$$

$$\sum_{i=1}^I y_{ij} = 1 \quad \forall j \in [J] \quad (6.17c)$$

$$y_{ij}, z_i \in \{0, 1\} \quad \forall i \in [I], j \in [J], \quad (6.17d)$$

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left\{ \sum_{j=1}^J \tilde{t}_{ij} y_{ij} \leq T_i \right\} \geq 1 - \alpha_i \quad \forall i \in [I], \quad (6.17e)$$

where c_i^z represents the cost of opening bin i and c_{ij}^y represents the cost of assigning item j to bin i . For each $i \in [I]$ and $j \in [J]$, we let binary variables z_i represent if bin i is open (i.e., $z_i = 1$ if open and $z_i = 0$ otherwise), binary variables y_{ij} represent if item j is assigned to bin i (i.e., $y_{ij} = 1$ if assigned and $y_{ij} = 0$ otherwise), and parameters ρ_{ij} represent if we can assign item j to bin i (i.e., $\rho_{ij} = 1$ if we can and $\rho_{ij} = 0$ otherwise). The objective function (6.17a) minimizes the total cost of opening bins and assigning items to bins. Constraints (6.17b) ensure that all items are assigned to open bins, constraints (6.17c) ensure that each item is assigned to one and only one bin, and constraints (6.17e) are the DRCCs.

In our computational studies, we follow Section 6.3 to derive 0-1 SOC reformulations of model (6.17) and then follow Section 6.4 to derive valid inequalities in the original and lifted space for the 0-1 SOC reformulations. We strengthen the extended polymatroid inequalities, as well as derive valid inequalities in the lifted space containing variables z_i , y_{ij} , and w_{ijk} to further strengthen the formulation as follows. We refer to Appendices C.1 and C.2 for the detailed proofs of the valid inequalities below, and will test their effectiveness later.

Theorem 6.7. For all extended polymatroid inequalities $\pi^\top y_i \leq T$ with regard to bin i , $\forall i \in [I]$, inequality

$$\pi^\top y_i \leq T z_i \quad (6.18a)$$

is valid for the DCBP formulation. Similarly, for all extended polymatroid inequalities

$\pi^\top v_i \leq T^2$ with regard to bin i , $\forall i \in [I]$, inequality

$$\pi^\top v_i \leq T^2 z_i \quad (6.18b)$$

is valid for the DCBP formulation.

Theorem 6.8. Consider set

$$L = \left\{ (z, y, w) \in \{0, 1\}^{I \times (IJ) \times (IJ^2)} : (6.17b)–(6.17d), w_{ijk} = y_{ij}y_{ik}, \forall j, k \in [J] \right\}.$$

Without loss of optimality, the following inequalities are valid for L :

$$w_{ijk} \geq y_{ij} + y_{ik} + \sum_{\substack{\ell=1 \\ \ell \neq i}}^I w_{\ell jk} - 1 \quad \forall j, k \in [J] \quad (6.19a)$$

$$w_{ijk} \geq y_{ij} + y_{ik} - z_i \quad \forall i \in [I], \forall j, k \in [J] \quad (6.19b)$$

$$\sum_{\substack{j=1 \\ j \neq k}}^J w_{ijk} \leq \sum_{j=1}^J y_{ij} - z_i \quad \forall i \in [I], \forall k \in [J] \quad (6.19c)$$

$$\sum_{j=1}^J \sum_{k=j+1}^J w_{ijk} \geq \sum_{j=1}^J y_{ij} - z_i \quad \forall i \in [I]. \quad (6.19d)$$

Remark 6.2. We note that valid inequalities (6.19a)–(6.19d) are *polynomially many* and all coefficients are in *closed-form*. Hence, we do not need any separation processes for these inequalities, and we can incorporate them in the DCBP formulation without dramatically increasing its size.

6.5.2 Computational Setup

We first consider $I = 6$ servers (i.e., bins) and $J = 32$ appointments (i.e., items) to test the DCBP model under various distributional assumptions and ambiguity sets. The daily operating time limit (i.e., capacity) T_i of each server i varies in between $[420, 540]$ minutes (i.e., 7–9 hours). We let the opening cost c_i^z of each server i be an increasing function of T_i such that $c_i^z = T_i^2/3600 + 3T_i/60$, and let all assignment costs c_{ij}^y , $\forall i \in [I], j \in [J]$ vary in between $[0, 18]$, so that the total opening cost and the total assignment cost have similar magnitudes. The above problem size and parameter settings follow the literature of surgery block allocation (see, e.g., Deng et al. (2016); Denton et al. (2010); Shylo et al. (2012)).

To generate samples of random service time (i.e., random item weight), we consider

“high mean (hM)” and “low mean (ℓ M)” being 25 minutes and 12.5 minutes, respectively. We set the coefficient of variation (i.e., the ratio of the standard deviation to the mean) as 1.0 for the “high variance (hV)” case and as 0.3 for the “low variance (ℓ V)” case. We equally mix all four types of appointments with “hMhV”, “hM ℓ V”, “ ℓ MhV”, “ ℓ M ℓ V”, and thus have eight appointments of each type. We sample 10,000 data points as the random service time of each appointment on each server, following a Gaussian distribution with the above settings of mean and standard deviation. We will hereafter call them the in-sample data. To formulate the 0-1 SOC models with diagonal covariance matrices, we use the empirical mean and standard deviation of each \tilde{t}_{ij} obtained from the in-sample data and set $\alpha_i = 0.05, \forall i \in [I]$. Using the same α_i -values, we formulate the 0-1 SOC models under general covariance matrices, for which we use the empirical mean and covariance matrix obtained from the in-sample data. The empirical covariance matrices we obtain have most of their off-diagonal entries being non-zero, and some being quite significant.

All the computation is performed on a Windows 7 machine with Intel(R) Core(TM) i7-2600 CPU 3.40 GHz and 8GB memory. We implement the optimization models and the branch-and-cut algorithm using commercial solver GUROBI 5.6.3 via Python 2.7.10. The GUROBI default settings are used for optimizing all the 0-1 programs, and we set the number of threads as one. When implementing the branch-and-cut algorithm, we add the violated extended polymatroid inequalities using GUROBI `callback` class by `Model.cbCut()` for both integer and fractional temporary solutions. For all the nodes in the branch-and-bound tree, we generate violated cuts at each node as long as any exists. The optimality gap tolerance is set as 0.01%. We also set the threshold for identifying violated cuts as 10^{-4} , and set the time limit for computing each instance as 3600 seconds.

6.5.3 CPU Time Comparison

We solve 0-1 SOC reformulations or approximations of DCBP, and use either a diagonal or a general covariance matrix in each model. Our valid inequalities significantly reduce the solution time of directly solving the 0-1 SOC models in GUROBI, while the extended polymatroid inequalities generated based on the approximate and lifted SOC constraints perform differently depending on the problem size. The details are presented as follows.

6.5.3.1 Computing 0-1 SOC models with diagonal matrices

We first optimize 0-1 SOC models with a diagonal matrix in constraint (6.9), of which the left-hand side function $g(y)$ is submodular, and thus we use extended polymatroid inequalities (6.18a) with $\pi \in EP_g$ in a branch-and-cut algorithm. Table 6.1 presents the CPU time

(in seconds), optimal objective values, and other solution details (including “**Server**” as the number of open servers, “**Node**” as the total number of branching nodes, and “**Cut**” as the total number of cuts added) for solving the three 0-1 SOC models **DCBP1** (i.e., using ambiguity set \mathcal{D}_1), **DCBP2** (i.e., using ambiguity set \mathcal{D}_2 with $\gamma_1 = 1$, $\gamma_2 = 2$), and **Gaussian** (assuming Gaussian distributed service time). We also implement the SAA approach (i.e., row “**SAA**”) by optimizing the MILP reformulation of the chance-constrained bin packing model based on the 10,000 in-sample data points. We compare the branch-and-cut algorithm using our extended polymatroid inequalities (in rows “**B&C**”) with directly solving the 0-1 SOC models in GUROBI (in rows “**w/o Cuts**”).

Table 6.1: CPU time and solution details for solving instances with diagonal matrices

Approach	Model	Time (s)	Opt. Obj.	Server	Opt. Gap	Node	Cut
B&C	DCBP1	0.73	328.99	3	0.00%	83	82
	DCBP2	27.50	366.54	3	0.00%	2146	2624
	Gaussian	0.13	297.94	2	0.00%	0	0
w/o Cuts	DCBP1	95.73	328.99	3	0.01%	76237	N/A
	DCBP2	LIMIT	380.09	2	9.15%	409422	N/A
	Gaussian	0.02	297.94	2	0.00%	16	N/A
SAA	MILP	21.20	297.94	2	0.00%	89	N/A

In Table 6.1, the branch-and-cut algorithm quickly optimizes DCBP1 and DCBP2. Especially, if being directly solved by GUROBI, DCBP2 cannot be solved within the 3600-second time limit and ends with a 9.15% optimality gap. Solving DCBP1 by using the branch-and-cut algorithm is much faster than solving the large-scale SAA-based MILP model, while the solution time of DCBP2 is similar to the latter. The two DCBP models also yield higher objective values, since they both provide more conservative solutions that open one more server than either the Gaussian or the SAA-based approach.

6.5.3.2 Computing 0-1 SOC models with general covariance matrices

In this section, we focus on testing DCBP2 yielded by the ambiguity set \mathcal{D}_2 with parameters $\gamma_1 = 1$, $\gamma_2 = 2$, and $\alpha_i = 0.05$, $\forall i \in [I]$. We use empirical covariance matrices of the in-sample data. Note that these covariance matrices are general and non-diagonal. We compare the time of solving the 0-1 SOC reformulations of DCBP2 on ten independently generated instances. We examine two implementations of the branch-and-cut algorithm: one uses extended polymatroid inequalities (6.18a) with $\pi \in EP_{g,L}$ based on the relaxed 0-1 SOC constraint (6.11), and the other uses the extended polymatroid inequalities (6.18b) based on the lifted SOC constraint.

Table 6.2 reports the CPU time (in seconds) and number of branching nodes (in column “**Node**”) for various methods. First, we directly solve the 0-1 SOC models of DCPB2 in GUROBI, without or with the linear valid inequalities (6.19a)–(6.19d), and report their results in columns “**w/o Cuts**” and “**Ineq.**”, respectively. We then implement the branch-and-cut algorithm, and examine the results of using extended polymatroid inequalities (6.18a) with $\pi \in EP_{gL}$ (reported in columns “**B&C-Relax**”) and cuts (6.18b) based on lifted SOC constraints (reported in columns “**B&C-Lifted**”). (The latter does not involve valid inequalities (6.18a) used in the former.) The time reported under **B&C-Relax** also involves the time of solving SDPs for obtaining Δ^L and the relaxed 0-1 SOC constraint (6.11). The time of solving related SDPs are quite small (varying from 1 to 2 seconds for instances of different sizes) and negligible as compared to the total B&C time. For both B&C methods, we also present the number of extended polymatroid inequalities (see column “**Cut**”) added.

Table 6.2: CPU time of DCPB2 solved by different methods with general covariance matrices

Instance	w/o Cuts		Ineq.		B&C-Relax			B&C-Lifted		
	Time (s)	Node	Time (s)	Node	Time (s)	Node	Cut	Time (s)	Node	Cut
1	286.29	10409	156.50	795	51.99	9095	702	35.03	618	823
2	433.32	10336	167.91	687	26.63	6524	698	12.34	405	235
3	284.17	10434	206.82	971	70.43	17420	621	29.84	595	729
4	310.11	10302	139.06	656	15.37	2467	723	25.31	419	617
5	329.32	10453	181.83	777	56.53	12349	737	35.09	678	921
6	365.28	10300	168.26	652	23.89	4807	695	26.73	555	595
7	296.55	10759	198.87	873	45.21	11585	738	21.08	440	626
8	278.62	10490	211.05	900	53.84	14540	721	47.78	1064	1686
9	139.24	7771	177.41	632	19.90	3918	645	19.37	216	360
10	297.72	10330	159.52	822	30.36	6877	649	29.43	400	727

In Table 6.2, we highlight the solution time of the method that runs the fastest for each instance. Note that without the extended polymatroid inequalities or the valid inequalities, the GUROBI solver takes the longest time for solving all the instances except instance #9. Adding the valid inequalities (6.19a)–(6.19d) to the solver reduces the solution time by 40% or more in almost all the instances, and drastically reduces the number of branching nodes. The extended polymatroid inequalities (6.18a)–(6.18b) further reduce the CPU time significantly (see columns “**B&C-Relax**” and “**B&C-Lifted**”). Moreover, for all the instances having 6 servers and 32 appointments, the algorithm using the extended polymatroid inequalities (6.18b) runs faster in eight out of ten instances than the algorithm using cuts (6.18a) based on relaxed SOC constraints without lifting. It indicates that the extended polymatroid inequalities generated by the lifted SOC constraints are more effective than

those generated by the relaxed SOC constraints. This observation is overturned when we later increase the problem size.

In the following, we continue reporting the CPU time of solving DCBP2 with general covariance matrix. We vary the problem sizes (i.e., values of I and J) in Section 6.5.3.3, and vary the values of Λ in the SOC constraint (6.9) in Section 6.5.3.4.

6.5.3.3 Solving 0-1 SOC models under different problem sizes

We use the same problem settings as in Section 6.5.2, and vary $I = 6, 8, 10$ and $J = 32, 40$ to test DCBP2 instances with different sizes. We still keep an equal mixture of all the four appointment types in each instance. Table 6.3 presents the computational time (in seconds), the total number of branching nodes (“**Node**”), and the total number of extended polymatroid inequalities generated (“**Cuts**”; if applicable) for solving the 0-1 SOC reformulation of DCBP2 by directly using GUROBI (“**w/o Cuts**”) and by using the two implementations of the extended polymatroid inequalities (“**B&C-Relax**” and “**B&C-Lifted**”).

Table 6.3: CPU time of DCBP2 with general covariance matrices for different problem sizes

Method	Inst.	$J = 32$					$J = 40$					
		1	2	3	4	5	6	7	8	9	10	
$I = 6$	B&C-Relax	Time (s)	51.99	26.63	70.43	15.37	56.53	6.87	12.76	1.59	2.36	12.73
		Node	9095	6524	17420	2467	12349	1009	1322	176	285	1270
		Cut	702	698	621	723	737	174	604	171	179	602
	B&C-Lifted	Time (s)	35.03	12.34	29.84	25.31	35.09	64.58	98.18	91.12	60.11	59.50
		Node	618	405	595	419	678	274	484	447	289	234
		Cut	823	235	729	617	921	470	690	688	462	394
	w/o Cuts	Time (s)	286.29	433.32	284.17	310.11	329.32	1654.31	208.12	1182.46	1580.41	1266.27
		Node	10409	10336	10434	10302	10453	10525	1272	10732	10658	10642
	$I = 8$	B&C-Relax	Time (s)	41.57	139.41	55.22	261.24	305.72	23.91	9.73	17.76	27.16
Node			8342	29042	12267	49820	61334	2130	1240	1561	2607	1024
Cut			737	770	742	803	790	714	199	728	702	690
B&C-Lifted		Time (s)	106.03	28.55	84.64	97.05	13.56	331.29	273.14	307.06	178.41	161.39
		Node	678	502	647	634	125	1177	836	1397	457	529
		Cut	114	691	128	143	216	1781	1175	2066	703	719
w/o Cuts		Time (s)	866.12	597.43	649.72	683.18	497.15	2265.53	2428.60	2294.62	1781.95	851.99
		Node	10338	10305	10309	10306	14386	11441	11219	11708	11128	5241
$I = 10$		B&C-Relax	Time (s)	3.75	9.28	6.56	3.23	16.71	29.94	80.34	22.58	24.48
	Node		637	972	659	549	2274	2336	7315	1870	1959	34306
	Cut		241	552	390	230	741	767	714	736	729	715
	B&C-Lifted	Time (s)	108.43	117.44	120.60	22.10	111.37	186.72	714.45	197.42	549.90	661.13
		Node	668	785	828	291	779	766	1108	811	896	1209
		Cut	108	191	314	281	188	1196	850	1106	568	808
	w/o Cuts	Time (s)	987.92	1140.23	183.06	1113.09	1425.83	2382.97	2917.03	LIMIT	2052.42	2451.62
		Node	10353	10357	4992	10307	10401	11015	11197	12101	10812	11001

In Table 6.3, we again highlight the solution time of the method that runs the fastest in each instance. We keep the first five instances we reported in Table 6.2 for instances with $I = 6$, $J = 32$, and report five instances for other (I, J) combinations. From Table 6.3, we observe that both implementations of the extended polymatroid inequalities run

significantly faster than directly using GUROBI, especially when we increase the problem sizes (i.e., I increased from 6 to 10, and J increased from 32 to 40). In particular, the CPU time of directly using GUROBI is consistently 1 or 2 orders of magnitude larger than that of our approaches. For smaller (I, J) -values (e.g., $(I, J) = (6, 32)$ or $(I, J) = (8, 32)$), we see that **B&C-Lifted** sometimes runs faster than **B&C-Relax**, but for all the other (I, J) combinations, the latter completely dominates the former. This is expected because the cuts (6.18b) are generated in a lifted space with J^2 additional variables for each server $i \in [I]$. Therefore, it makes sense that the scalability of **B&C-Lifted** is worse than that of **B&C-Relax**, which uses cuts (6.18a) without lifting.

6.5.3.4 Solving 0-1 SOC models with different Λ -values in (6.9)

We again focus on instances with $I = 6$ and $J = 32$ under the same general covariance matrix Σ obtained from the in-sample data points. We let $\Lambda := \Omega^2 \Sigma$ in the 0-1 SOC constraint (6.9) and adjust the scalar Ω to obtain different Λ . We want to show how the computational time of directly using GUROBI increases as we increase Ω , as compared to using the branch-and-cut algorithm with extended polymatroid inequalities. (The results of **B&C-Lifted** are used here and similar observations can be made if the results of **B&C-Relax** are used.)

Considering specific cases of the SOC constraint (6.9) for modeling DCBP, we have $\Omega = \Phi^{-1}(1 - \alpha) = 1.64$ for the Gaussian approximation model when $\alpha = 0.05$, and $\Omega = \sqrt{\gamma_2/\alpha} = 6.32$ for DCBP2 when $\gamma_2 = 2$ and $\alpha = 0.05$. We test four values of Ω equally distributed in between $[1.64, 6.32]$ including the two end points.

Figure 6.3 depicts the average CPU time and the average number of branching nodes of solving five independent DCBP2 instances for each Ω -setting. Specifically, the four values 1.64, 3.20, 4.76, and 6.32 of Ω lead to 0.98, 97.68, 299.29, and 363.56 CPU seconds when directly using GUROBI, respectively, together with the significantly growing number of branching nodes 0, 909.8, 6662, and 10418, respectively. On the other hand, the branch-and-cut algorithm with the extended polymatroid inequalities respectively takes 1.03, 35.19, 16.09, and 26.63 CPU seconds on average for solving the same instances, and branches on average 0, 188.6, 247.2, and 374.6 nodes, respectively. This indicates that our approach is more scalable than directly using the off-the-shelf solvers.

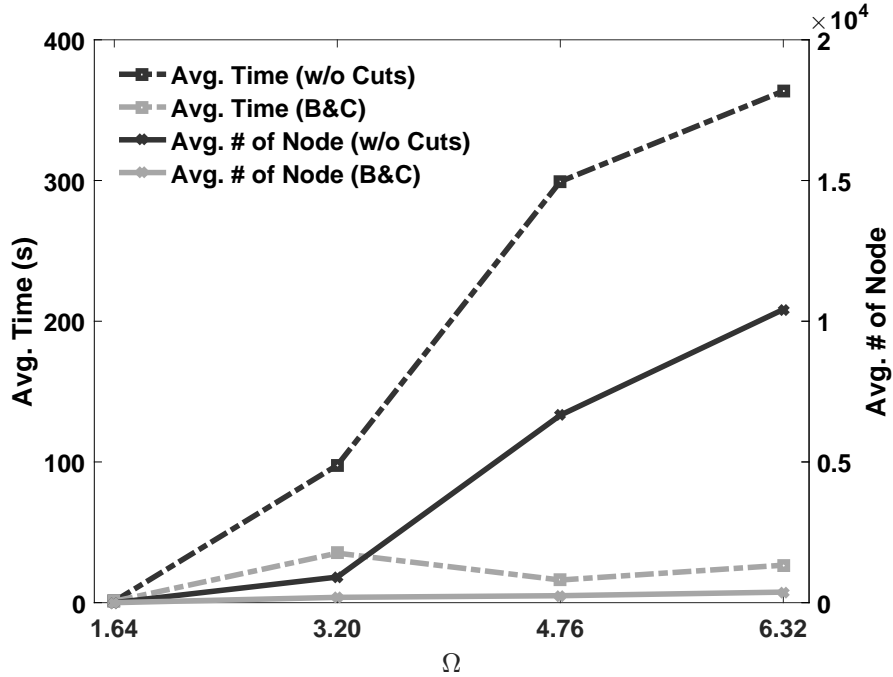


Figure 6.3: CPU time and number of branching nodes under different Ω -values in constraint (6.9)

6.6 Concluding Remarks

In this chapter, we considered distributionally robust individual chance constraints, where the true distributional information of the constraint coefficients is ambiguous and only the empirical first and second moments are given. The goal is to restrict the worst-case probability of violating a linear constraint under a given threshold. We provided 0-1 SOC representations of DRCCs under two types of ambiguity sets. In addition, we derived an efficient way of obtaining extended polymatroid inequalities for the 0-1 SOC constraints in both original and lifted spaces. Via extensive numerical studies, we demonstrated that our solution approaches significantly accelerate solving the DCBP model as compared to the state-of-the-art commercial solvers. In particular, a branch-and-cut algorithm with extended polymatroid inequalities in the original space scales very well as the problem size grows.

CHAPTER 7

Conclusion

In this dissertation work, we focus on developing scalable solution approaches for DRO models (especially with integer decisions and nonlinear structures) in various applications including energy system operations, appointment scheduling, and surgery block allocation in healthcare industry. Although the solution approaches are developed for specific applications, the appointment scheduling and bin packing (of which the surgery block allocation is a special problem setting in healthcare) itself can be adapted to a wide range of applications. Therefore, the solution approaches are also applicable to other problems (e.g., especially to chance-constrained binary programs). Moreover, through extensive computational experiments, we study the impact of ambiguous probability distribution on the system operator's decisions by employing different moment-based ambiguity sets, as an accurate estimate of system uncertainty's probability distribution is hard to obtain due to insufficiency of historical data. In particular, in Chapter 2, we observed that the DRO approach yields better reliabilities compared to the Gaussian approximation approach by locating more reserves on generator and load buses and also achieves a good trade-off between performance and computational efficiency. In Chapter 3, our results demonstrated that an approximate "dome" shape shown in the solutions of DRO models but not in those of a sampling-based stochastic model when only limited data samples are available. In Chapter 4, we showed that the DRO models with the least conservative support of no-shows obtain near-optimal schedules under perfect information and outperform DR models with other support of no-shows and stochastic program if the distributional type is misspecified. In Chapter 5, the DRO models yield better out-of-sample reliability performance of not exceeding server's capacity compared to the stochastic program. While, the conservatism of DRO solutions depends on the choice of ambiguity set's parameters. In Chapter 6, we derived efficient branch-and-cut algorithms with valid cuts obtained from extended polymatroid inequalities in both the original and lifted spaces. We showed that the algorithm with cuts from the original spaces scales very well as the problem size grows.

For the future research, we plan to apply the DRO techniques and solution approaches developed in this dissertation to other power system operation problems, such as transmission expansion planning and building load control. We also plan to investigate DRO models under other types of ambiguity sets, which could take into account not only the moment information but also density, support or structural information of their probability distribution. Moreover, instead of considering one optimization problem in a system, we plan to investigate systems using DRO models with two or more players, each of which has optimization problems of their own interests while there are some linking constraints among them.

APPENDIX A

Appendix for Chpater 4

A.1 General Waiting Time Costs

In this section, we discuss a general setting that incorporates both server-based and appointment-based cases to model the waiting costs. We show that the models in Section 4.2 and solution methods in Section 4.3 can adapt for this general setting. For presentation brevity, we focus on DR expectation models and the adaptation of methods for DR CVaR models in A.2 is similar.

To be general, we model the waiting time cost for each appointment i as $c_i^s w_i + c_i^a w_i q_i = (c_i^s + c_i^a q_i) w_i$, where c_i^s represents the server-based waiting time cost and c_i^a represents the appointment-based waiting time cost. Note that this setting applies for server-based and appointment-based cases, because we can set $c_i^s := 0$ or $c_i^a := 0$ if the corresponding cost does not apply. Accordingly, the general DR expectation model can be formulated as

$$\min_{x \in X} \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{E}_{\mathbb{P}_{q,s}} [Q^G(x, q, s)], \quad (\text{A.1})$$

where $Q^G(x, q, s)$ represents the cost function of the waiting, idleness, and overtime under the general setting. Replacing $c_i^w w_i$ with $(c_i^s + c_i^a q_i) w_i$ in (4.11), we formulate the dual of $Q^G(x, q, s)$ as

$$Q^G(x, q, s) = \max_y \sum_{i=1}^n (q_i s_i - x_i) y_i \quad (\text{A.2a})$$

$$\text{s.t.} \quad y_{i-1} - y_i \leq c_i^s + c_i^a q_i \quad \forall i = 2, \dots, n \quad (\text{A.2b})$$

$$-y_i \leq c_i^u \quad \forall i = 1, \dots, n \quad (\text{A.2c})$$

$$y_n \leq c^o, \quad (\text{A.2d})$$

and we let polyhedron $Y^G := \{y : (\text{A.2b})\text{--}(\text{A.2d})\}$ represent the feasible region of variable

y . As model (A.2) is a linear program in variables y , there exists an optimal solution to (A.2) that resides at an extreme point of Y^G . It can be observed (see, e.g., Mak et al. (2015); Zangwill (1966, 1969)) that any extreme point \hat{y} of Y^G satisfy (i) either $\hat{y}_n = -c_n^u$ or $\hat{y}_n = c^o$, and (ii) for all $i = 1, \dots, n-1$, dual constraint $\hat{y}_{i+1} + c_{i+1}^s + c_{i+1}^a q_{i+1} \geq \hat{y}_i \geq -c_i^u$ is binding at either the lower bound or the upper bound. Similar to the analysis in Section 4.3.2, we define binary variables t_{kj} for all $1 \leq k \leq j \leq n+1$ to represent extreme points \hat{y} , such that $t_{kj} = 1$ if $\hat{y}_j = -c_j^u$ and $\hat{y}_i = \hat{y}_{i+1} + c_{i+1}^s + c_{i+1}^a q_{i+1}$, $\forall i = k, \dots, j-1$. It follows that

$$\hat{y}_i = \pi_{ij}^G := \begin{cases} -c_j^u + \sum_{\ell=i+1}^j (c_\ell^s + c_\ell^a q_\ell) & 1 \leq i \leq j \leq n, \\ c^o + \sum_{\ell=i+1}^n (c_\ell^s + c_\ell^a q_\ell) & 1 \leq i \leq n, j = n+1, \end{cases} \quad (\text{A.3})$$

where $\hat{y}_{n+1} = \pi_{n+1, n+1}^G := 0$. Based on this binary representation, we can rewrite $Q^G(x, q, s)$ as

$$\begin{aligned} Q^G(x, q, s) &= \max_t \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(\sum_{i=k}^j (q_i s_i - x_i) \pi_{ij}^G \right) t_{kj} \\ &\equiv \sum_{k=1}^n \sum_{j=k}^n \sum_{i=k}^j \left[(q_i s_i - x_i) \left(-c_j^u + \sum_{\ell=i+1}^j (c_\ell^s + c_\ell^a q_\ell) \right) \right] t_{kj} + \\ &\quad \sum_{k=1}^n \sum_{i=k}^{n+1} \left[(q_i s_i - x_i) \left(c^o + \sum_{\ell=i+1}^n (c_\ell^s + c_\ell^a q_\ell) \right) \right] t_{k(n+1)} \quad (\text{A.4a}) \end{aligned}$$

$$\text{s.t.} \quad \sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1 \quad \forall i = 1, \dots, n+1 \quad (\text{A.4b})$$

$$t_{kj} \in \{0, 1\}, \quad \forall 1 \leq k \leq j \leq n+1. \quad (\text{A.4c})$$

Note that the objective function (A.4a) contains multilinear terms $q_i s_i t_{kj}$ and $q_i q_\ell s_i t_{kj}$ with binary variables q_i , q_ℓ , and t_{kj} , and continuous variables s_i . To linearize formulation (A.4), we define $p_{ikj} \equiv q_i t_{kj}$, $o_{ikj} \equiv q_i s_i t_{kj}$, and $r_{ilkj} = q_i q_\ell s_i t_{kj}$ for all $1 \leq k \leq i \leq j \leq n+1$ and $i+1 \leq \ell \leq j$. We then linearize the multilinear terms by applying McCormick inequalities (4.18a)–(4.18b) for variables p_{ikj} , (4.18c)–(4.18d) for variables o_{ikj} , and (A.5a)–(A.5b) as follows for variables r_{ilkj} .

$$r_{ilkj} \geq 0, \quad r_{ilkj} - q_\ell s_i \leq 0, \quad (\text{A.5a})$$

$$r_{ilkj} - o_{ikj} \leq 0, \quad r_{ilkj} - o_{ikj} + s_i^u (1 - q_\ell) \geq 0. \quad (\text{A.5b})$$

If follows that $Q^G(x, q, s)$ equals to the optimal objective value of the following MILP:

$$\begin{aligned}
\max_{t,p,o,r} \quad & \sum_{k=1}^n \sum_{j=k}^n \sum_{i=k}^j \left[\left(-c_j^u + \sum_{\ell=i+1}^j c_\ell^s \right) o_{ikj} + \left(c_j^u - \sum_{\ell=i+1}^j c_\ell^s \right) x_i t_{kj} + \sum_{\ell=i+1}^j c_\ell^a r_{ilkj} \right. \\
& \left. - \sum_{\ell=i+1}^j c_\ell^a x_i p_{lkj} \right] + \sum_{k=1}^n \sum_{i=k}^{n+1} \left[\left(c^o + \sum_{\ell=i+1}^n c_\ell^s \right) o_{ik(n+1)} - \left(c^o + \sum_{\ell=i+1}^n c_\ell^s \right) x_i t_{k(n+1)} \right. \\
& \left. + \sum_{\ell=i+1}^n c_\ell^a r_{ilk(n+1)} - \sum_{\ell=i+1}^n c_\ell^a x_i p_{lk(n+1)} \right] \\
\text{s.t.} \quad & (4.18a)-(4.18d), (A.4b), (A.5a)-(A.5b), \tag{A.6a} \\
& t_{kj}, p_{ikj} \in \{0, 1\}, \quad \forall 1 \leq k \leq i \leq j \leq n+1. \tag{A.6b}
\end{aligned}$$

To finish reformulating the general DR expectation model (A.1), we follow a similar dualization process described in Section 4.3 and rewrite formulation (A.1) as

$$\begin{aligned}
\min_{x \in X, \rho \in \mathbb{R}^n, \gamma \in \mathbb{R}^n, \theta \in \mathbb{R}} \quad & \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \tag{A.7a} \\
\text{s.t.} \quad & \theta \geq H^G(x, \rho, \gamma) \equiv \max_{(q,s) \in D_q \times D_s} \left\{ Q^G(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \tag{A.7b}
\end{aligned}$$

Similar to Lemma 4.1, we observe that for any fixed variables x , ρ , and γ , $H^G(x, \rho, \gamma) < +\infty$. Furthermore, function $H^G(x, \rho, \gamma)$ is convex and piecewise linear in x , ρ , and γ with a finite number of pieces. Hence, we can adapt Algorithm 4.1 to solve model (A.1) in a decomposition framework. We present this adaptation in Algorithm A.1. Similar to Algorithm 4.1, we observe that Algorithm A.1 is finite. Finally, for the separation problem in Step 3, we remark that (i) feasible region D_q can be set as $D_q^{(K)}$ for $K = 2, \dots, n+1$ based on the scheduler's targeted conservativeness, (ii) the separation problem is an MILP and can be solved by off-the-shelf software, and (iii) we can incorporate the same valid inequalities identified in Proposition 3 to accelerate solving the separation problem and

hence the decomposition algorithm.

Algorithm A.1: A decomposition algorithm for solving general DR expectation model (A.1).

Input: Feasible regions X and $D_q \times D_s$; set of cuts $\{L(x, \rho, \gamma, \theta) \geq 0\} = \emptyset$.

1 Solve the master problem

$$\begin{aligned} \min_{x \in X, \rho, \gamma, \theta} \quad & \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \\ \text{s.t.} \quad & L(x, \rho, \gamma, \theta) \geq 0 \end{aligned}$$

and record an optimal solution $(x^*, \rho^*, \gamma^*, \theta^*)$;

2 With (x, ρ, γ) fixed to be (x^*, ρ^*, γ^*) , solve the separation problem

$$\begin{aligned} \max_{t, p, q, s, o, r} \quad & \sum_{k=1}^n \sum_{j=k}^n \sum_{i=k}^j \left[\left(-c_j^u + \sum_{\ell=i+1}^j c_\ell^s \right) o_{ikj} + \left(c_j^u - \sum_{\ell=i+1}^j c_\ell^s \right) x_i t_{kj} + \sum_{\ell=i+1}^j c_\ell^a r_{ilkj} - \right. \\ & \left. \sum_{\ell=i+1}^j c_\ell^a x_i p_{lkj} \right] + \sum_{k=1}^n \sum_{i=k}^{n+1} \left[\left(c^o + \sum_{\ell=i+1}^n c_\ell^s \right) o_{ik(n+1)} - \left(c^o + \sum_{\ell=i+1}^n c_\ell^s \right) x_i t_{k(n+1)} + \right. \\ & \left. \sum_{\ell=i+1}^n c_\ell^a r_{ilk(n+1)} - \sum_{\ell=i+1}^n c_\ell^a x_i p_{lk(n+1)} \right] - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \end{aligned}$$

s.t. (4.18a)–(4.18d), (A.4b), (A.5a)–(A.5b),

$$t_{kj}, p_{ikj} \in \{0, 1\}, \forall 1 \leq k \leq i \leq j \leq n+1, (q, s) \in D_q \times D_s$$

and record an optimal solution $(t^*, p^*, q^*, s^*, o^*, r^*)$;

3 if θ^* is greater than or equal to the optimal objective value of the separation problem, then

4 | stop and return x^* as an optimal solution to formulation (A.1).

5 else

6 | add the cut

$$\begin{aligned} \theta \geq \quad & \sum_{k=1}^n \sum_{j=k}^n \sum_{i=k}^j \left[\left(-c_j^u + \sum_{\ell=i+1}^j c_\ell^s \right) o_{ikj}^* + \left(c_j^u - \sum_{\ell=i+1}^j c_\ell^s \right) t_{kj}^* x_i + \sum_{\ell=i+1}^j c_\ell^a r_{ilkj}^* - \right. \\ & \left. \sum_{\ell=i+1}^j c_\ell^a p_{lkj}^* x_i \right] + \sum_{k=1}^n \sum_{i=k}^{n+1} \left[\left(c^o + \sum_{\ell=i+1}^n c_\ell^s \right) o_{ik(n+1)}^* - \left(c^o + \sum_{\ell=i+1}^n c_\ell^s \right) t_{k(n+1)}^* x_i + \right. \\ & \left. \sum_{\ell=i+1}^n c_\ell^a r_{ilk(n+1)}^* - \sum_{\ell=i+1}^n c_\ell^a p_{lk(n+1)}^* x_i \right] - \sum_{i=1}^n (s_i^* \rho_i + q_i^* \gamma_i) \end{aligned}$$

7

to the set of cuts $\{L(x, \rho, \gamma, \theta) \geq 0\}$ and go to Step 2.

8 end

A.2 Solution Approaches for DR CVaR Models

In this section, we reformulate DR CVaR constraints in (4.6) with cost parameters c^o , c_i^u , and c_i^w . We first represent CVaR by an alternative definition (Rockafellar and Uryasev, 2000, 2002):

$$\text{CVaR}_{1-\epsilon}(Q(x, q, s)) = \inf_{z \in \mathbb{R}} \left\{ z + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s) - z]^+ \right\},$$

where $[a]^+ := \max\{a, 0\}$ for $a \in \mathbb{R}$. It follows that

$$\begin{aligned} \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \text{CVaR}_{1-\epsilon}(Q(x, q, s)) &= \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \inf_{z \in \mathbb{R}} \left\{ z + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s) - z]^+ \right\} \\ &= \inf_{z \in \mathbb{R}} \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \left\{ z + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s) - z]^+ \right\} \quad (\text{A.8a}) \\ &= \inf_{z \in \mathbb{R}} \left\{ z + \frac{1}{\epsilon} \sup_{\mathbb{P}_{q,s} \in \mathcal{F}(D, \mu, \nu)} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s) - z]^+ \right\}, \quad (\text{A.8b}) \end{aligned}$$

where (A.8a) follows the Sion's minimax theorem (Sion, 1958) because $z + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}_{q,s}} [Q(x, q, s) - z]^+$ is convex in z , concave (in particular, linear) in variables $\mathbb{P}_{q,s}$, and $\mathcal{F}(D, \mu, \nu)$ is weakly compact.

A.2.1 MILP Reformulation and Decomposition Algorithm

Based on a similar dualization process in Section 4.3 (see the primal and dual formulations (4.8) and (4.9)), we reformulate the inner maximization problem in (A.8b) as a minimiza-

tion problem, and combine it with the outer minimization problem to obtain

$$\begin{aligned}
& \inf_{z, \rho, \gamma, \theta} z + \frac{1}{\epsilon} \left(\sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \right) \\
& \text{s.t. } \sum_{i=1}^n s_i \rho_i + \sum_{i=1}^n \gamma_i q_i + \theta \geq [Q(x, q, s) - z]^+ \quad \forall (q, s) \in D_q \times D_s \\
= & \inf_{z, \rho, \gamma, \theta} z + \frac{1}{\epsilon} \left(\sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \right) \\
& \text{s.t. } \sum_{i=1}^n s_i \rho_i + \sum_{i=1}^n \gamma_i q_i + \theta \geq 0 \quad \forall (q, s) \in D_q \times D_s \tag{A.9a} \\
& \sum_{i=1}^n s_i \rho_i + \sum_{i=1}^n \gamma_i q_i + \theta \geq Q(x, q, s) - z \quad \forall (q, s) \in D_q \times D_s, \tag{A.9b}
\end{aligned}$$

where constraints (A.9a) and (A.9b) are derived based on the definition of $[\cdot]^+$. Thus, the DR CVaR constraint (4.6) is equivalent to

$$\bar{Q} \geq z + \frac{1}{\epsilon} \left(\sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \right) \tag{A.10a}$$

$$\min_{(q, s) \in D_q \times D_s} \left\{ \sum_{i=1}^n \rho_i s_i + \sum_{i=1}^n \gamma_i q_i \right\} + \theta \geq 0 \tag{A.10b}$$

$$\theta + z \geq \max_{(q, s) \in D_q \times D_s} \left\{ Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\}, \tag{A.10c}$$

where constraint (A.10a) is linear, but (A.10b) and (A.10c) need further analysis. First, we replace constraint (A.10b) by equivalent linear constraints in the following proposition, whose proof is relegated to A.4.1.

Proposition 5. For fixed ρ and γ , and $D_q = D_q^{(K)}$ with $K \in \{2, \dots, n+1\}$, (A.10b) is equivalent to linear constraints:

$$\theta + \sum_{i=1}^{n-K+1} \beta_i + \sum_{i=1}^n (s_i^L \chi_i^L - s_i^U \chi_i^U - \eta_i) \geq 0, \tag{A.11a}$$

$$-\eta_i + \sum_{j=\max\{i-K+1, 1\}}^{\min\{i, n-K+1\}} \beta_j \leq \gamma_i \quad \forall 1 \leq i \leq n, \tag{A.11b}$$

$$\chi_i^L - \chi_i^U \leq \rho_i \quad \forall 1 \leq i \leq n, \tag{A.11c}$$

$$\beta_i, \chi_i^L, \chi_i^U, \eta_i \geq 0 \quad \forall 1 \leq i \leq n. \tag{A.11d}$$

Second, note that the right-hand side of constraint (A.10c) is equivalent to that of constraint (4.13b) in the reformulated DR expectation model, and so the reformulated separation problem (4.19) and Algorithm 4.1 described in Section 4.3 can be easily adapted to handle constraint (A.10c). Furthermore, the valid inequalities (4.20a)–(4.20f) can be incorporated to accelerate solving the adapted separation problem and implementing the decomposition algorithm.

A.2.2 LP Reformulations of the DR CVaR Model

We derive LP reformulations for the DR CVaR constraint (4.6) when $D_q = D_q^{(2)}$ (i.e., no consecutive no-shows) and when $D_q = D_q^{(n+1)}$ (i.e., arbitrary no-shows).

Case 1. (No Consecutive No-Shows) Recall that DR CVaR constraint (4.6) is equivalent to constraints (A.10a), (A.11a)–(A.11d) with $K = 2$, and (A.10c). When $D_q = D_q^{(2)}$, we apply Theorem 4.1 to further reformulate (A.10c) as linear constraints $\theta + z \geq \sum_{i=1}^{n+1} (\alpha_i + s_i^u \tau_i^u - s_i^l \tau_i^l)$ and (4.22b)–(4.22g), resulting in the following proposition.

Proposition 6. When $D_q = D_q^{(2)}$, the DR CVaR constraint (4.6) is equivalent to linear constraints (A.10a), (A.11a)–(A.11d) with $K = 2$, $\sum_{i=1}^{n+1} (\alpha_i + s_i^u \tau_i^u - s_i^l \tau_i^l) \leq \theta + z$, and (4.22b)–(4.22g).

We remark that the LP reformulation in Proposition 6 is of the size $\mathcal{O}(n^3)$ because constraints (4.22b)–(4.22g) incorporate $\mathcal{O}(n^3)$ decision variables and linear constraints. In this section, we focus on a specific DR CVaR constraint (4.6) that restricts overtime only. That is, $c_i^u = c_i^w = 0$ for all $1 \leq i \leq n$ and $c^o = 1$, and $Q(x, q, s) = Q^W(x, q, s) := \min_{w, u, W} W$ subject to constraints (4.3b)–(4.3d). Next, we derive a more compact LP reformulation of this DR CVaR constraint with $\mathcal{O}(n^2)$ variables and constraints. To that end, we derive an $\mathcal{O}(n^2)$ LP reformulation for constraint (A.10c). We begin by specializing the extreme point representation of polyhedron Y for $Q(x, q, s) = Q^W(x, q, s)$.

Lemma A.1. When $c_i^u = c_i^w = 0$ for all $1 \leq i \leq n$ and $c^o = 1$, the set of extreme points of polyhedron Y defined in (4.15) is $\{\sum_{\ell=k}^n e_\ell : k = 1, \dots, n\} \cup \{\mathbf{0}_n\}$, where e_ℓ represents an n -dimensional unit vector with component ℓ equaling to one and any other component equaling to zero; $\mathbf{0}_n$ is an n -dimensional zero vector.

Recall the observation in Section 4.3.2 that each extreme point (y_1, \dots, y_{n+1}) of Y is associated with a partition of set $\{1, \dots, n+1\}$ into intervals. The result in Lemma A.1 follows from (4.16) when the cost parameters take the above specified values. Define binary variables t_k for all $1 \leq k \leq n$ to represent the set of extreme points of Y , such that $t_k = 1$ if the extreme point is $\sum_{\ell=k}^n e_\ell$ and $t_k = 0$ otherwise. Note that extreme

point $\mathbf{0}_n$ is represented by $t_k = 0$ for all $1 \leq k \leq n$. For a valid representation, we require $\sum_{k=1}^n t_k \leq 1$. It follows that the right-hand side of (A.10c) (with $Q(x, q, s) = Q^W(x, q, s)$) is equivalent to

$$\begin{aligned} \max_{t,q,s} \quad & \sum_{k=1}^n \left(\sum_{i=k}^n (q_i s_i - x_i) \right) t_k - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \\ \text{s.t.} \quad & \sum_{k=1}^n t_k \leq 1, \quad q \in D_q, \quad s \in D_s, \quad t \in \{0, 1\}^n \end{aligned}$$

as a mixed-integer bilinear program with binary vectors q and t , and continuous vector s . We linearize the bilinear terms by defining $p_{ki} \equiv t_k q_i$ and $o_{ki} \equiv t_k q_i s_i$ for all $1 \leq k \leq i \leq n$. Also, we introduce McCormick inequalities (A.12b)–(A.12c) and (A.12d)–(A.12e) for variables p_{ki} and o_{ki} , respectively to further reformulate the separation problem as a mixed-integer linear program:

$$\max_{t,q,s,p,o} \quad \sum_{k=1}^n \sum_{i=k}^n (o_{ki} - x_i t_k) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \quad (\text{A.12a})$$

$$\text{s.t.} \quad p_{ki} - t_k \leq 0 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.12b})$$

$$p_{ki} - q_i \leq 0, \quad p_{ki} - q_i - t_k \geq -1, \quad p_{ki} \geq 0 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.12c})$$

$$o_{ki} - s_i^l p_{ki} \geq 0, \quad o_{ki} - s_i^u p_{ki} \leq 0 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.12d})$$

$$o_{ki} - s_i + s_i^l (1 - p_{ki}) \leq 0, \quad o_{ki} - s_i + s_i^u (1 - p_{ki}) \geq 0 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.12e})$$

$$\sum_{k=1}^n t_k \leq 1, \quad (\text{A.12f})$$

$$q \in D_q, \quad s \in D_s, \quad t \in \{0, 1\}^n. \quad (\text{A.12g})$$

Similar as before, we aim to derive the convex hull of the feasible region of problem (A.12), i.e., the mixed-integer feasible region described by constraints (A.12b)–(A.12g). We denote the feasible region as set G and derive $\text{conv}(G)$ in the following theorem, whose proof is in A.4.2.

Theorem A.1. When $D_q = D_q^{(2)}$, the following inequalities are valid for set $G = \{(t, q, s, p, o) :$

(A.12b)–(A.12g)}:

$$\sum_{k=1}^n p_{kn} \leq q_n, \quad (\text{A.13a})$$

$$p_{ki} + p_{k(i+1)} \geq t_k \quad \forall 1 \leq k \leq i \leq n-1, \quad (\text{A.13b})$$

$$\sum_{k=1}^i (p_{ki} - t_k) \geq q_i - 1 \quad \forall 1 \leq i \leq n, \quad (\text{A.13c})$$

$$\sum_{k=1}^i (p_{ki} + p_{k(i+1)}) \leq \sum_{k=1}^i t_k + q_i + q_{i+1} - 1 \quad \forall 1 \leq i \leq n-1, \quad (\text{A.13d})$$

$$s_i - \sum_{k=1}^i (o_{ki} - s_i^l p_{ki}) \geq s_i^l \quad \forall 1 \leq i \leq n, \quad (\text{A.13e})$$

$$s_i - \sum_{k=1}^i (o_{ki} - s_i^u p_{ki}) \leq s_i^u \quad \forall 1 \leq i \leq n. \quad (\text{A.13f})$$

Furthermore, polyhedron $CG := \{(t, q, s, p, o) : (\text{A.12b}), (\text{A.12d}), (\text{A.13a})\text{--}(\text{A.13f})\}$ is the convex hull of set G , i.e., $CG = \text{conv}(G)$.

Theorem A.1 provides us a compact LP reformulation of the right-hand side of constraint (A.10c) with $\mathcal{O}(n^2)$ variables and constraints:

$$\max_{t, q, s, p, o} \sum_{k=1}^n \sum_{i=k}^n (o_{ki} - x_i t_k) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \quad (\text{A.14a})$$

$$\text{s.t. } (t, q, s, p, o) \in CG. \quad (\text{A.14b})$$

Finally, by resorting to the dual formulation of (A.14), we represent constraint (A.10c)

as

$$\sum_{i=1}^n (\alpha_i - s_i^L \tau_i^L + s_i^U \tau_i^U) - \sum_{i=1}^{n-1} \phi_i \leq \theta + z \quad (\text{A.15a})$$

$$\sum_{i=k}^n (\alpha_i - \sigma_{ki}) + \sum_{i=k}^{n-1} (\lambda_{ki} - \phi_i) \geq - \sum_{i=k}^n x_i \quad \forall 1 \leq k \leq n, \quad (\text{A.15b})$$

$$\alpha_i - \sum_{\ell=\max\{i-1,1\}}^{\min\{i,n-1\}} \phi_\ell - \sum_{\ell=n}^{\max\{i,n-1\}} \zeta \geq -\gamma_i \quad \forall 1 \leq i \leq n, \quad (\text{A.15c})$$

$$\tau_i^U - \tau_i^L \geq -\rho_i \quad \forall 1 \leq i \leq n, \quad (\text{A.15d})$$

$$\begin{aligned} & \sigma_{ki} + s_i^L \varphi_{ki}^L - s_i^U \varphi_{ki}^U - \alpha_i - s_i^L \tau_i^L + s_i^U \tau_i^U \\ & + \sum_{\ell=n}^{\max\{i,n-1\}} \zeta + \sum_{\ell=\max\{i-1,k\}}^{\min\{i,n-1\}} (\phi_\ell - \lambda_{k\ell}) \geq 0 \quad \forall 1 \leq k \leq i \leq n, \end{aligned} \quad (\text{A.15e})$$

$$-\varphi_{ki}^L + \varphi_{ki}^U + \tau_i^L - \tau_i^U \geq 1 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.15f})$$

$$\sigma_{ki}, \varphi_{ki}^L, \varphi_{ki}^U, \zeta, \lambda_{ki}, \alpha_i, \phi_i, \tau_i^L, \tau_i^U \geq 0 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.15g})$$

where dual variables σ_{ki} , $\varphi_{ki}^{L/U}$, ζ , λ_{ki} , α_i , ϕ_i , and $\tau_i^{L/U}$ are associated with constraints (A.12b), (A.12d), (A.13a), (A.13b), (A.13c), (A.13d), and (A.13e)–(A.13f), respectively (after transforming all “ \geq ” inequalities into the “ \leq ” form), and dual constraints (A.15b)–(A.15f) are associated with primal variables t_k , q_i , s_i , p_{ki} , and o_{ki} respectively. This results in an $\mathcal{O}(n^2)$ LP reformulation of the DR CVaR constraint on overtime.

Proposition 7. When $D_q = D_q^{(2)}$, $c_i^u = c_i^w = 0$ for all $1 \leq i \leq n$ and $c^o = 1$, the DR CVaR constraint (4.6) on overtime is equivalent to linear constraints (A.10a), (A.11a)–(A.11d) with $K = 2$, and (A.15a)–(A.15g).

Case 2. (Arbitrary No-Shows) Recall that DR CVaR constraint (4.6) is equivalent to constraints (A.10a), (A.11a)–(A.11d) with $K = n + 1$, and (A.10c). As $D_q = D_q^{(n+1)}$, we can apply the results in Section 4.4 (see Case 2) to further reformulate (A.10c) as linear constraints $\theta + z \geq \sum_{i=1}^{n+1} \alpha_i$, (4.24b), (4.24d), and (4.25a)–(4.25d). This results in the following proposition.

Proposition 8. When $D_q = D_q^{(n+1)}$, the DR CVaR constraint (4.6) is equivalent to constraints (A.10a), (A.11a)–(A.11d) with $K = n + 1$, $\sum_{i=1}^{n+1} \alpha_i \leq \theta + z$, (4.24b), (4.24d), and (4.25a)–(4.25d).

A.3 Proofs for the DR Expectation Model

A.3.1 Proof of Lemma 4.1

Proof. Proof of Lemma 4.1 First, feasible regions Y and $D_q \times D_s$ are both independent of x , ρ , and γ , and bounded. Hence, $\max_{y \in Y} h(x, y, \rho, \gamma) \equiv \max_{y \in Y, (q, s) \in D_q \times D_s} \{ \sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \} < +\infty$. Second, for any fixed y , q , and s , $\sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i)$ is a linear function of x , ρ , and γ . It follows that $\max_{y \in Y} h(x, y, \rho, \gamma)$ is the maximum of a set of linear functions of x , ρ , and γ , and hence convex and piecewise linear. Third, it is clear that each linear piece of function $\max_{y \in Y} h(x, y, \rho, \gamma)$ is associated with one distinct extreme point of polyhedra Y , D_q , and D_s respectively. Therefore, the number of pieces of function $\max_{y \in Y} h(x, y, \rho, \gamma)$ is finite because each of these polyhedra has a finite number of extreme points. This completes the proof. \square

A.3.2 Proof of Lemma 4.2

Proof. Proof of Lemma 4.2 For fixed x , ρ , and γ , in view of the definition of function $h(x, y, \rho, \gamma)$ in (4.12c), we have $h(x, y, \rho, \gamma) = \max_{(q, s) \in D_q \times D_s} H(q, s, y)$, where $H(q, s, y)$ is a linear function of variable y . It follows that $h(x, y, \rho, \gamma)$ is the supremum of a set of convex functions of y , and hence itself convex in variable y . \square

A.3.3 Proof of Proposition 3

Proof. Proof of Proposition 3 First, because $p_{ikj} \equiv q_i t_{kj}$, equality (4.20a) can be obtained via multiplying equalities $\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1$ by q_i on both sides.

Second, because $o_{ikj} \equiv q_i s_i t_{kj} \equiv s_i p_{ikj}$, and by equalities (4.20a) and $s_i \in [s_i^L, s_i^U]$, we have

$$\begin{aligned} \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^L p_{ikj}) &= (s_i - s_i^L) \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} = (s_i - s_i^L) q_i \leq (s_i - s_i^L), \\ \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^U p_{ikj}) &= (s_i - s_i^U) \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} = (s_i - s_i^U) q_i \geq (s_i - s_i^U), \end{aligned}$$

which shows the validity of inequalities (4.20b) and (4.20c).

Third, for $1 \leq k < j \leq n + 1$ and $k \leq i \leq j - K + 1$, because $\sum_{\ell=i}^{i+K-1} q_\ell \geq 1$ by the

definition of $D_q^{(K)}$, we have

$$\sum_{\ell=i}^{i+K-1} p_{\ell k j} = \sum_{\ell=i}^{i+K-1} q_{\ell} t_{k j} = \left(\sum_{\ell=i}^{i+K-1} q_{\ell} \right) t_{k j} \geq t_{k j},$$

which shows the validity of inequalities (4.20d).

Fourth, for $i = 1, \dots, n$, because $\sum_{k=1}^i \sum_{j=i}^{n+1} t_{k j} = 1$ and $\sum_{k=1}^{i+1} \sum_{j=i+1}^{n+1} t_{k j} = 1$ by constraints (4.17b), we have

$$0 = \sum_{k=1}^i \sum_{j=i}^{n+1} t_{k j} - \sum_{k=1}^{i+1} \sum_{j=i+1}^{n+1} t_{k j} = \sum_{k=1}^i t_{k i} - \sum_{j=i+1}^{n+1} t_{(i+1)j}. \quad (\text{A.16})$$

We show the validity of inequalities (4.20e) for all $i = K - 1, \dots, n$. If $\sum_{k=1}^{i-K+2} t_{k i} = 0$, then the conclusion holds because each $p_{i k j} \geq 0$. Now suppose that $\sum_{k=1}^{i-K+2} t_{k i} = 1$, then $\sum_{j=i+1}^{n+1} t_{(i+1)j} = 1$ in view of (A.16). It follows that

$$\begin{aligned} \sum_{k=1}^{i-K+2} \sum_{\ell=i-K+2}^i p_{\ell k i} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} &= \left(\sum_{\ell=i-K+2}^i q_{\ell} \right) \left(\sum_{k=1}^{i-K+2} t_{k i} \right) + q_{i+1} \sum_{j=i+1}^{n+1} t_{(i+1)j} \\ &= \sum_{\ell=i-K+2}^{i+1} q_{\ell} \geq 1, \end{aligned}$$

where the last inequality is due to the definition of $D_q^{(K)}$.

Finally, we show the validity of inequalities (4.20f) for all $i = 1, \dots, n - K + 2$. If $\sum_{j=i+K-1}^{n+1} t_{(i+1)j} = 0$, then the conclusion holds because each $p_{i k j} \geq 0$. Now suppose that $\sum_{j=i+K-1}^{n+1} t_{(i+1)j} = 1$, then $\sum_{k=1}^i t_{k i} = 1$ in view of (A.16). It follows that

$$\begin{aligned} \sum_{k=1}^i p_{i k i} + \sum_{\ell=i+1}^{i+K-1} \sum_{j=i+K-1}^{n+1} p_{\ell(i+1)j} &= q_i \left(\sum_{k=1}^i t_{k i} \right) + \left(\sum_{\ell=i+1}^{i+K-1} q_{\ell} \right) \left(\sum_{j=i+K-1}^{n+1} t_{(i+1)j} \right) \\ &= \sum_{\ell=i}^{i+K-1} q_{\ell} \geq 1, \end{aligned}$$

where the last inequality is due to the definition of $D_q^{(K)}$.

□

A.3.4 Proof of Theorem 4.1

Recall that polyhedron $CF = \{(t, q, s, p, o) : (4.17b), (4.18a), (4.18c), (4.20a)–(4.20d), (4.21)\}$ in Theorem 4.1. We first study the extreme points of polyhedron CF and show their properties as follows.

Proposition 9. Every extreme point (t, q, s, p, o) of CF satisfies the following:

1. $t_{kj}, p_{ikj} \in \{0, 1\}$ for all $1 \leq k \leq j \leq n + 1$ and $k \leq i \leq j$;
2. $q_i \in \{0, 1\}$ for all $1 \leq i \leq n + 1$;
3. $p_{ikj} = q_i t_{kj}$ and $o_{ikj} = q_i s_i t_{kj}$ for all $1 \leq k \leq j \leq n + 1$ and $k \leq i \leq j$.

Proof. Proof of Proposition 9 Consider arbitrary cost coefficients c_i^q and c_i^s for all $1 \leq i \leq n + 1$, c_{kj}^t for all $1 \leq k \leq j \leq n + 1$, and c_{ikj}^p and c_{ikj}^o for all $1 \leq k \leq j \leq n + 1$ and $k \leq i \leq j$. We construct a related linear program

$$\begin{aligned}
 \text{(LP-CF)} \quad & \min_{t, q, s, p, o} \sum_{i=1}^{n+1} (c_i^q q_i + c_i^s s_i) + \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(c_{kj}^t t_{kj} + \sum_{i=k}^j (c_{ikj}^p p_{ikj} + c_{ikj}^o o_{ikj}) \right) \\
 & \text{s.t. } (t, q, s, p, o) \in CF.
 \end{aligned}$$

To prove that each extreme point of CF satisfies properties 1, 2, and 3, we show for any values of c_{kj}^t , c_i^q , c_i^s , c_{ikj}^p , and c_{ikj}^o , there exists an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ to (LP-CF) that satisfies properties 1, 2, and 3 (cf. Nemhauser and Wolsey, 1999; Wolsey, 1998).

First, in view of equalities (4.20a), we can assume that $c_i^q = 0$ for all $1 \leq i \leq n + 1$ w.l.o.g., because we can always replace each c_{ikj}^p with $c_{ikj}^p + c_i^q$ so that variables p_{ikj} will carry the cost of decisions q_i . It follows that we can ignore variables q_i in (LP-CF) because they do not contribute to the objective function and their values entirely depend on p_{ikj} by constraints (4.20a). Also, we note that (i) $s_i^l p_{ikj} \leq o_{ikj} \leq s_i^u p_{ikj}$ by (4.18c), and so $p_{ikj} \geq 0$ for all $1 \leq k \leq i \leq j \leq n + 1$, and (ii) for all $1 \leq i \leq n + 1$, $\sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} \leq \sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1$ by (4.18a) and (4.17b).

Second, we rewrite (LP-CF) as a two-stage formulation as follows:

$$\begin{aligned}
 \min_{t, p} \quad & \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(c_{kj}^t t_{kj} + \sum_{i=k}^j c_{ikj}^p p_{ikj} \right) + V(p) \\
 \text{s.t.} \quad & (t, p) \in CF_{t, p},
 \end{aligned}$$

where polyhedron $CF_{t,p} := \{(t, p) : (4.17b), (4.18a), (4.20d), (4.21)\}$ and $V(p)$ represents a value function of p defined as

$$\begin{aligned} \text{(LP-CF}(p)) \quad V(p) &:= \min_{s,o} \sum_{i=1}^{n+1} c_i^s s_i + \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^j c_{ikj}^o o_{ikj} \\ \text{s.t.} \quad (s, o) &\in CF_{s,o}(p), \end{aligned}$$

where

$$CF_{s,o}(p) = \left\{ (s, o) : o_{ikj} \geq s_i^l p_{ikj} \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \quad (\text{A.17a}) \right.$$

$$o_{ikj} \leq s_i^u p_{ikj} \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \quad (\text{A.17b})$$

$$s_i - \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^l p_{ikj}) \geq s_i^l \quad \forall 1 \leq i \leq n+1, \quad (\text{A.17c})$$

$$s_i - \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^u p_{ikj}) \leq s_i^u \quad \forall 1 \leq i \leq n+1 \left. \right\} \quad (\text{A.17d})$$

represents a parametric polyhedron depending on the values of p_{ikj} . We solve (LP-CF(p)) by considering its dual formulation

$$\begin{aligned} V(p) = \max_{\psi, \omega} & \sum_{i=1}^{n+1} \sum_{k=1}^i \sum_{j=i}^{n+1} \left(s_i^l p_{ikj} \psi_{ikj}^l - s_i^u p_{ikj} \psi_{ikj}^u \right) + \sum_{i=1}^{n+1} \left[s_i^l \left(1 - \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} \right) \omega_i^l - \right. \\ & \left. s_i^u \left(1 - \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} \right) \omega_i^u \right] \end{aligned}$$

$$\text{s.t.} \quad \psi_{ikj}^l - \psi_{ikj}^u - \omega_i^l + \omega_i^u = c_{ikj}^o \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \quad (\text{A.18a})$$

$$\omega_i^l - \omega_i^u = c_i^s \quad \forall 1 \leq i \leq n+1, \quad (\text{A.18b})$$

where dual variables $\psi_{ikj}^{l/u}$ and $\omega_i^{l/u}$ are associated with primal constraints (A.17a)–(A.17b) and (A.17c)–(A.17d), respectively (after transforming all “ \leq ” inequalities into the “ \geq ” form), and dual constraints (A.18a) and (A.18b) are associated with primal variables o_{ikj} and s_i , respectively. Because $p_{ikj} \geq 0$ for all $1 \leq k \leq i \leq j \leq n+1$ and $1 - \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} \geq 0$ for all $1 \leq i \leq n+1$, a dual optimal solution to problem (LP-CF(p)) is $\psi_{ikj}^{l*} = (c_{ikj}^o + c_i^s)^+$, $\psi_{ikj}^{u*} = (-c_{ikj}^o - c_i^s)^+$, $\omega_i^{l*} = (c_i^s)^+$, and $\omega_i^{u*} = (-c_i^s)^+$. It follows that

$$V(p) = \sum_{i=1}^{n+1} [s_i^l (c_i^s)^+ - s_i^u (-c_i^s)^+] + \sum_{i=1}^{n+1} \sum_{k=1}^i \sum_{j=i}^{n+1} \left[s_i^l (c_{ikj}^o + c_i^s)^+ - s_i^u (-c_{ikj}^o - c_i^s)^+ - \right.$$

$s_i^l(c_i^s)^+ + s_i^u(-c_i^s)^+$ is a linear function of p . Therefore, (LP-CF) is equivalent to optimizing a linear function of (t, p) on polyhedron $CF_{t,p}$. It follows that there exists an optimal solution (t^*, p^*, s^*, o^*) to (LP-CF) where (t^*, p^*) is an extreme point of polyhedron $CF_{t,p}$.

Third, we show that all extreme points of $CF_{t,p}$ are integral. To this end, we show that the constraint matrix describing $CF_{t,p}$ is totally unimodular (TU). For presentation convenience, we rewrite the constraints defining $CF_{t,p}$ in inequalities as follows:

$$\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} \geq 1 \quad \forall i = 1, \dots, n+1, \quad (\text{A.19a})$$

$$-\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} \geq -1 \quad \forall i = 1, \dots, n+1, \quad (\text{A.19b})$$

$$-t_{kj} + p_{ikj} + p_{(i+1)kj} \geq 0 \quad \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1, \quad (\text{A.19c})$$

$$-\sum_{k=1}^i t_{ki} + \sum_{k=1}^i p_{iki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \geq 0 \quad \forall i = 1, \dots, n, \quad (\text{A.19d})$$

$$t_{kj} - p_{ikj} \geq 0 \quad \forall 1 \leq k \leq j \leq n+1, \forall k \leq i \leq j, \quad (\text{A.19e})$$

and we denote the constraint matrix as

$$\mathcal{CF}_{t,p}^0 := \begin{bmatrix} (\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj}), & \forall 1 \leq i \leq n+1 \\ (-\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj}), & \forall 1 \leq i \leq n+1 \\ (-t_{kj} + p_{ikj} + p_{(i+1)kj}), & \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1 \\ (-\sum_{k=1}^i t_{ki} + \sum_{k=1}^i p_{iki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j}), & \forall 1 \leq i \leq n \\ (t_{kj} - p_{ikj}), & \forall 1 \leq k \leq i \leq j \leq n+1 \end{bmatrix},$$

where the five row sub-matrices in matrix $\mathcal{CF}_{t,p}^0$ are associated with the left-hand side of constraints (A.19a)–(A.19e), respectively. To show that matrix $\mathcal{CF}_{t,p}^0$ is TU, we conduct pivot operations on the matrix with variables p_{ikj} and t_{kj} . Note that a matrix is TU if and only if it remains TU after pivot operations (Nemhauser and Wolsey, 1999). We conduct the following pivot operations in order.

1. For all $1 \leq k \leq j \leq n+1$ and $k \leq i \leq j$, pivot with variable p_{ikj} based on the component -1 in sub-matrix $(t_{kj} - p_{ikj})$ (corresponding to constraints (A.19e)). This pivot operation is equivalent to (a) adding $t_{kj} - p_{ikj}$, for all $1 \leq k \leq j \leq n+1$ and

$k \leq i \leq j$, to the left-hand side of every constraint (A.19c)–(A.19d) in which variable p_{ikj} has coefficient 1, and (b) multiplying the left-hand side of each constraint (A.19e) by -1 . As a result, the matrix after pivoting becomes

$$\mathcal{CF}_{t,p}^1 := \begin{bmatrix} (\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj}), & \forall 1 \leq i \leq n+1 \\ (-\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj}), & \forall 1 \leq i \leq n+1 \\ (t_{kj}), & \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1 \\ (\sum_{j=i+1}^{n+1} t_{(i+1)j}), & \forall 1 \leq i \leq n \\ (-t_{kj} + p_{ikj}), & \forall 1 \leq k \leq i \leq j \leq n+1 \end{bmatrix}.$$

Note that sub-matrix $(-t_{kj} + p_{ikj} + p_{(i+1)kj})$ becomes (t_{kj}) because each $-t_{kj} + p_{ikj} + p_{(i+1)kj}$ on the left-hand side of (A.19c) is summed with $t_{kj} - p_{ikj}$ and $t_{kj} - p_{(i+1)kj}$, and so the coefficient of each t_{kj} changes from -1 to 1 after pivoting.

2. For all $1 \leq k < j \leq n+1$, pivot with variable t_{kj} based on any component 1 in sub-matrix (t_{kj}) (note that there are multiple components 1 corresponding to each variable t_{kj} in sub-matrix (t_{kj}) and we can pick any one of them). Since all components in each row of sub-matrix (t_{kj}) are zeros except one equaling 1, these pivot operations (a) make all coefficients of all variables t_{kj} zeros in matrix $\mathcal{CF}_{t,p}^1$ as long as $1 \leq k < j \leq n+1$, and (b) keep all coefficients of all variables p_{ikj} unchanged. As a result, the matrix after pivoting becomes

$$\mathcal{CF}_{t,p}^2 := \begin{bmatrix} (t_{ii}), & \forall 1 \leq i \leq n+1 \\ (-t_{ii}), & \forall 1 \leq i \leq n+1 \\ (t_{kj}), & \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1 \\ (t_{(i+1)(i+1)}), & \forall 1 \leq i \leq n \\ \left\{ \begin{array}{l} (-t_{ii} + p_{iii}), & \forall 1 \leq i \leq n+1 \\ (p_{ikj}), & 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1 \end{array} \right. \end{bmatrix}.$$

It follows that matrix $\mathcal{CF}_{t,p}^2$ contains only $\{-1, 0, 1\}$ entries, has no more than two nonzero entries in each row, and the sum of the entries is zero for each row containing two nonzero entries. Hence, matrix $\mathcal{CF}_{t,p}^2$ is TU, and so is matrix $\mathcal{CF}_{t,p}^0$.

Therefore, the extreme points of polyhedron $CF_{t,p}$ are integral and so property 1 is proved.

Fourth, to show property 2, we consider any extreme point (t, q, s, p, o) of polyhedron

CF. By constraints (4.18a) and (4.20a), we have $q_i = \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj} \leq \sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1$, and so $q_i \in \{0, 1\}$ because each $p_{ikj} \in \{0, 1\}$ by property 1. This shows property 2.

Finally, to show property 3, we consider any extreme point (t, q, s, p, o) of polyhedron *CF*. We show $p_{ikj} = q_i t_{kj}$ by discussing the following cases.

1. If $q_i = 0$, then $p_{ikj} = 0$ for all $1 \leq k \leq i$ and $i \leq j \leq n + 1$ because $q_i = \sum_{k=1}^i \sum_{j=i}^{n+1} p_{ikj}$. It follows that $p_{ikj} = q_i t_{kj}$.
2. If $q_i = 1$, then there exist $1 \leq k^* \leq i$ and $i \leq j^* \leq n + 1$ such that $p_{ik^*j^*} = 1$ and any other $p_{ikj} = 0$. It follows that $t_{k^*j^*} = 1$ because $p_{ikj} - t_{kj} \leq 0$ by constraint (4.18a), and any other $t_{kj} = 0$ because $\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1$ given by (4.17b). Therefore, we have $p_{ik^*j^*} = q_i t_{k^*j^*} = 1$ and $p_{ikj} = q_i t_{kj} = 0$ for all other $1 \leq k \leq i$ and $i \leq j \leq n + 1$.

For all $1 \leq i \leq n + 1$, since $\sum_{k=1}^i \sum_{j=i}^{n+1} t_{kj} = 1$, there exist $1 \leq k^* \leq i$ and $i \leq j^* \leq n + 1$ such that $t_{k^*j^*} = 1$ and any other $t_{kj} = 0$. Since (t, q, s, p, o) is an extreme point of polyhedron *CF*, each o_{ikj} satisfies either inequality (A.17a) or (A.17b) at equality, and each s_i satisfies either inequality (A.17c) or (A.17d) at equality. We discuss the following two cases to show $o_{ikj} = q_i s_i t_{kj}$.

1. If $q_i = 0$, then $p_{ikj} = q_i t_{kj} = 0$ for all $1 \leq k \leq i$ and $i \leq j \leq n + 1$. It follows from inequalities (A.17a)–(A.17b) that each corresponding $o_{ikj} = 0$. Therefore, we have $o_{ikj} = s_i p_{ikj} = 0$, or equivalently $o_{ikj} = q_i s_i t_{kj} = 0$, for all $1 \leq k \leq j \leq n + 1$ and $k \leq i \leq j$.
2. If $q_i = 1$, then $p_{ik^*j^*} = q_i t_{k^*j^*} = 1$ and $p_{ikj} = 0$ for all other $1 \leq k \leq i$ and $i \leq j \leq n + 1$. Then, inequalities (A.17a)–(A.17b) yield $o_{ikj} = s_i p_{ikj} = 0$ for all $1 \leq k \leq i$ and $i \leq j \leq n + 1$ such that $(k, j) \neq (k^*, j^*)$. Furthermore, inequalities (A.17c)–(A.17d) yield

$$s_i - \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^l p_{ikj}) = s_i - o_{ik^*j^*} + s_i^l p_{ik^*j^*} \geq s_i^l,$$

$$s_i - \sum_{k=1}^i \sum_{j=i}^{n+1} (o_{ikj} - s_i^u p_{ikj}) = s_i - o_{ik^*j^*} + s_i^u p_{ik^*j^*} \leq s_i^u.$$

It follows that $s_i = o_{ik^*j^*}$. Therefore, we have $o_{ik^*j^*} = s_i p_{ik^*j^*}$.

□

We are now ready to show Theorem 4.1.

Proof. Proof of Theorem 4.1 ($CF \supseteq \text{conv}(F)$) By Proposition 3, since polyhedron CF consists of either trivial equalities/inequalities or valid inequalities of set F , we have $(t, q, s, p, o) \in CF$ if $(t, q, s, p, o) \in F$. It follows that $CF \supseteq \text{conv}(F)$.

($CF \subseteq \text{conv}(F)$) By Proposition 9, since each extreme point (t, q, s, p, o) of CF satisfies properties 1, 2, and 3, $(t, q, s, p, o) \in F$. By the Minkowski's Theorem on polyhedron, we have $x \in \text{conv}(F)$ if $x \in CF$. It follows that $CF \subseteq \text{conv}(F)$. This completes the proof. \square

A.4 Proofs for the DR CVaR Model

A.4.1 Proof of Proposition 5

Proof. Proof of Proposition 5 We analyze the following two cases based on the value of K .

When $K \in \{2, \dots, n\}$: For the embedded minimization problem in constraint (A.10b), we observe that the constraint matrix of D_q , described by constraints $\sum_{j=i}^{i+K-1} q_j \geq 1$ for all $1 \leq i \leq n - K + 1$, is an interval matrix and thus TU. It follows that $\text{conv}(D_q) = \{q \in [0, 1]^n : \sum_{j=i}^{i+K-1} q_j \geq 1, \forall 1 \leq i \leq n - K + 1\}$. Because the feasible regions of variables q and s (i.e., D_q and D_s) are disjoint in (A.10b), we can replace D_q with $\text{conv}(D_q)$ and obtain

$$\theta + \min_{q,s} \left\{ \sum_{i=1}^n \rho_i s_i + \sum_{i=1}^n \gamma_i q_i \right\} \geq 0 \quad (\text{A.20a})$$

$$\text{s.t. } s_i^l \leq s_i \leq s_i^u \quad \forall 1 \leq i \leq n, \quad (\text{A.20b})$$

$$\sum_{j=i}^{i+K-1} q_j \geq 1 \quad \forall 1 \leq i \leq n - K + 1, \quad (\text{A.20c})$$

$$q_i \leq 1 \quad \forall 1 \leq i \leq n, \quad (\text{A.20d})$$

$$q_i, s_i \geq 0 \quad \forall 1 \leq i \leq n. \quad (\text{A.20e})$$

Presenting linear program (A.20) in its dual form yields (A.11a)–(A.11d), where dual variables $\chi_i^{l/u}$, β_i , and η_i are associated with constraints (A.20b), (A.20c), and (A.20d) respectively, and dual constraints (A.11b) and (A.11c) are associated with primal variables q_i and s_i , respectively.

When $K = n + 1$: In this case, $D_q = \{0, 1\}^n$ and so $\text{conv}(D_q) = [0, 1]^n$. It follows that

constraint (A.10b) is equivalent to

$$\begin{aligned} \theta + \min_{q,s} \left\{ \sum_{i=1}^n \rho_i s_i + \sum_{i=1}^n \gamma_i q_i \right\} &\geq 0 \\ \text{s.t. } s_i^L &\leq s_i \leq s_i^U \quad \forall 1 \leq i \leq n, \\ q_i &\leq 1 \quad \forall 1 \leq i \leq n, \\ q_i, s_i &\geq 0 \quad \forall 1 \leq i \leq n, \end{aligned}$$

Similar to the case when $K \in \{2, \dots, n\}$, we can present the embedded LP in its dual form to obtain the following linear constraints:

$$\begin{aligned} \theta + \sum_{i=1}^n (s_i^L \chi_i^L - s_i^U \chi_i^U - \eta_i) &\geq 0, \\ -\eta_i &\leq \gamma_i \quad \forall 1 \leq i \leq n, \\ \chi_i^L - \chi_i^U &\leq \rho_i \quad \forall 1 \leq i \leq n, \\ \chi_i^L, \chi_i^U, \eta_i &\geq 0 \quad \forall 1 \leq i \leq n. \end{aligned}$$

We note that these linear constraints are equivalent to (A.11a)–(A.11d) because $\sum_{i=1}^{n-K+1} \beta_i = \sum_{i=1}^0 \beta_i = 0$ and $\sum_{j=\max\{i-K+1, 1\}}^{\min\{i, n-K+1\}} \beta_j = \sum_{j=1}^0 \beta_j = 0$. The proof is completed. \square

A.4.2 Proof of Theorem A.1

We take the following three steps to prove Theorem A.1.

Step 1: We prove the validity of inequalities (A.13a)–(A.13f) in the following proposition.

Proposition 10. When $D_q = D_q^{(2)}$, inequalities (A.13a)–(A.13f) are valid for set $G = \{(t, q, s, p, o) : (A.12b)–(A.12g)\}$.

Proof. Proof of Proposition 10 First, since $\sum_{k=1}^n t_k \leq 1$ by constraint (A.12f) and $q_n \geq 0$, we have

$$\sum_{k=1}^n p_{kn} = q_n \sum_{k=1}^n t_k \leq q_n,$$

which shows inequality (A.13a).

Second, for all $1 \leq k \leq i \leq n-1$, since $q_i + q_{i+1} \geq 1$ by the definition of D_q and $t_k \geq 0$, we have

$$p_{ki} + p_{k(i+1)} = t_k(q_i + q_{i+1}) \geq t_k,$$

which shows inequalities (A.13b).

Third, for all $1 \leq i \leq n$, since $t_i \geq 0$, $\forall i$ and thus $\sum_{k=1}^i t_k \leq \sum_{k=1}^n t_k \leq 1$ by constraint (A.12f) and since $q_i \leq 1 \Rightarrow q_i - 1 \leq 0$, we have

$$\sum_{k=1}^i (p_{ki} - t_k) = (q_i - 1) \sum_{k=1}^i t_k \geq q_i - 1,$$

which shows inequality (A.13c).

Fourth, for all $1 \leq i \leq n-1$ because (a) $\sum_{k=1}^i t_k \leq \sum_{k=1}^n t_k \leq 1$ by constraint (A.12f) and $t_k \geq 0$, $\forall k$, and (b) $q_i + q_{i+1} \geq 1$ by the definition of D_q , we have

$$\begin{aligned} \sum_{k=1}^i (p_{ki} + p_{k(i+1)}) - (q_i + q_{i+1}) &= (q_i + q_{i+1}) \sum_{k=1}^i t_k - (q_i + q_{i+1}) \\ &= (q_i + q_{i+1}) \left(\sum_{k=1}^i t_k - 1 \right) \leq \sum_{k=1}^i t_k - 1, \end{aligned}$$

which shows inequality (A.13d).

Finally, for each $1 \leq i \leq n$, since $\sum_{k=1}^n t_k \leq 1$ by constraint (A.12f), and $t_k, q_i \in [0, 1]$, we have $\sum_{k=1}^i p_{ki} = q_i (\sum_{k=1}^i t_k) \leq q_i (\sum_{k=1}^n t_k) \leq 1$. Also, because $s_i \in [s_i^l, s_i^u]$, it follows that

$$\begin{aligned} \sum_{k=1}^i (o_{ki} - s_i^l p_{ki}) &= (s_i - s_i^l) \sum_{k=1}^i p_{ki} \leq s_i - s_i^l, \\ \sum_{k=1}^i (o_{ki} - s_i^u p_{ki}) &= (s_i - s_i^u) \sum_{k=1}^i p_{ki} \geq s_i - s_i^u, \end{aligned}$$

which shows inequalities (A.13e)–(A.13f). □

Step 2: We show the properties of the extreme points of polyhedron CG in the following proposition. Recall that $CG = \{(t, q, s, p, o) : (A.12b), (A.12d), (A.13a)–(A.13f)\}$.

Proposition 11. Each extreme point (t, q, s, p, o) of CG has the following properties:

1. $t_k, q_i, p_{ki} \in \{0, 1\}$ for all $1 \leq k \leq i \leq n$;
2. $p_{ki} = t_k q_i$ and $o_{ki} = t_k q_i s_i$ for all $1 \leq k \leq i \leq n$.

Proof. Proof of Proposition 11 For any $c_k^t, c_i^q, c_i^s, c_{ki}^p$, and c_{ki}^o for all $1 \leq k \leq i \leq n$, we

consider linear program

$$\begin{aligned}
(\text{LP-CG}) \quad & \min_{t,q,s,p,o} \sum_{i=1}^n (c_i^q q_i + c_i^s s_i) + \sum_{k=1}^n \left(c_k^t t_k + \sum_{i=k}^n (c_{ki}^p p_{ki} + c_{ki}^o o_{ki}) \right) \\
& \text{s.t. } (t, q, s, p, o) \in CG.
\end{aligned}$$

To prove that each extreme point of CG satisfies properties 1 and 2, we show that for any c_k^t , c_i^q , c_i^s , c_{ki}^p , and c_{ki}^o , there exists an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ to (LP-CG) that satisfies properties 1 and 2. First, we rewrite (LP-CG) as a two-stage formulation as follows:

$$\begin{aligned}
\min_{t,q,p} \quad & \sum_{i=1}^n c_i^q q_i + \sum_{k=1}^n \left(c_k^t t_k + \sum_{i=k}^n c_{ki}^p p_{ki} \right) + V(p) \\
& \text{s.t. } (t, q, p) \in CG_{t,q,p},
\end{aligned}$$

where polyhedron $CG_{t,q,p} := \{(t, q, p) : (\text{A.12b}), (\text{A.13a})\text{--}(\text{A.13d})\}$ and $V(p)$ represents a value function of p defined as

$$(\text{LP-CG}(p)) \quad V(p) := \min_{s,o} \left\{ \sum_{i=1}^n c_i^s s_i + \sum_{k=1}^n \sum_{i=k}^n c_{ki}^o o_{ki} : (s, o) \in CG_{s,o}(p) \right\},$$

$$\begin{aligned}
\text{where } CG_{s,o}(p) &= \left\{ (s, o) : (\text{A.12d}), (\text{A.13e}), (\text{A.13f}) \right\} \\
&= \left\{ (s, o) : o_{ki} \geq s_i^L p_{ki} \quad \forall 1 \leq k \leq i \leq n, \right. \tag{A.21a}
\end{aligned}$$

$$o_{ki} \leq s_i^U p_{ki} \quad \forall 1 \leq k \leq i \leq n, \tag{A.21b}$$

$$s_i - \sum_{k=1}^i (o_{ki} - s_i^L p_{ki}) \geq s_i^L \quad \forall 1 \leq i \leq n, \tag{A.21c}$$

$$s_i - \sum_{k=1}^i (o_{ki} - s_i^U p_{ki}) \leq s_i^U \quad \forall 1 \leq i \leq n \left. \right\} \tag{A.21d}$$

represents a parametric polyhedron depending on the values of p_{ki} . We solve (LP-CG(p))

by considering its dual formulation

$$V(p) = \max_{\psi, \omega} \sum_{i=1}^n \sum_{k=1}^i (s_i^L p_{ki} \psi_{ki}^L - s_i^U p_{ki} \psi_{ki}^U) + \sum_{i=1}^n \left[s_i^L \left(1 - \sum_{k=1}^i p_{ki} \right) \omega_i^L - s_i^U \left(1 - \sum_{k=1}^i p_{ki} \right) \omega_i^U \right]$$

$$\text{s.t. } \psi_{ki}^L - \psi_{ki}^U - \omega_i^L + \omega_i^U = c_{ki}^o \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.22a})$$

$$\omega_i^L - \omega_i^U = c_i^s \quad \forall 1 \leq i \leq n, \quad (\text{A.22b})$$

where dual variables $\psi_{ki}^{L/U}$ and $\omega_i^{L/U}$ are associated with primal constraints (A.21a)–(A.21b) and (A.21c)–(A.21d), respectively (after transforming all “ \leq ” inequalities into the “ \geq ” form), and dual constraints (A.22a) and (A.22b) are associated with primal variables o_{ki} and s_i , respectively. Because (i) $s_i^L p_{ki} \leq o_{ki} \leq s_i^U p_{ki}$ by (A.12d), and so $p_{ki} \geq 0$ for all $1 \leq k \leq i \leq n$, and (ii) $s_i^L (1 - \sum_{k=1}^i p_{ki}) \leq s_i - \sum_{k=1}^i o_{ki} \leq s_i^U (1 - \sum_{k=1}^i p_{ki})$ by (A.13e)–(A.13f), and so $1 - \sum_{k=1}^i p_{ki} \geq 0$ for all $1 \leq i \leq n$, a dual optimal solution to problem (LP-CG(p)) is $\psi_{ki}^{L*} = (c_{ki}^o + c_i^s)^+$, $\psi_{ki}^{U*} = (-c_{ki}^o - c_i^s)^+$, $\omega_i^{L*} = (c_i^s)^+$, and $\omega_i^{U*} = (-c_i^s)^+$. It follows that $V(p) = \sum_{i=1}^n [s_i^L (c_i^s)^+ - s_i^U (-c_i^s)^+] + \sum_{i=1}^n \sum_{k=1}^i \left[s_i^L (c_{ki}^o + c_i^s)^+ - s_i^U (-c_{ki}^o - c_i^s)^+ - s_i^L (c_i^s)^+ + s_i^U (-c_i^s)^+ \right] p_{ki}$ is a linear function of p . Therefore, (LP-CG) is equivalent to optimizing a linear function of (t, q, p) on polyhedron $CG_{t,q,p}$. It follows that there exists an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ to (LP-CG) where (t^*, q^*, p^*) is an extreme point of polyhedron $CG_{t,q,p}$.

Second, we show that all extreme points of $CG_{t,q,p}$ are integral. To this end, we show that the constraint matrix describing $CG_{t,q,p}$ is TU. For presentation convenience, we rewrite the constraints defining $CG_{t,q,p}$ as follows:

$$q_i + q_{i+1} + \sum_{k=1}^i t_k - \sum_{k=1}^i (p_{ki} + p_{k(i+1)}) \geq 1 \quad \forall 1 \leq i \leq n-1 \quad (\text{A.23a})$$

$$q_n - \sum_{k=1}^n p_{kn} \geq 0, \quad (\text{A.23b})$$

$$-q_i - \sum_{k=1}^i t_k + \sum_{k=1}^i p_{ki} \geq -1 \quad \forall 1 \leq i \leq n, \quad (\text{A.23c})$$

$$-t_k + p_{ki} + p_{k(i+1)} \geq 0 \quad \forall 1 \leq k \leq i \leq n-1, \quad (\text{A.23d})$$

$$t_k - p_{ki} \geq 0 \quad \forall 1 \leq k \leq i \leq n, \quad (\text{A.23e})$$

and we denote the constraint matrix as

$$\mathcal{CG}_{t,q,p}^0 := \begin{bmatrix} (q_i + q_{i+1} + \sum_{k=1}^i t_k - \sum_{k=1}^i (p_{ki} + p_{k(i+1)})), & \forall 1 \leq i \leq n-1 \\ (q_n - \sum_{k=1}^n p_{kn}) \\ (-q_i - \sum_{k=1}^i t_k + \sum_{k=1}^i p_{ki}), & \forall 1 \leq i \leq n \\ (-t_k + p_{ki} + p_{k(i+1)}), & \forall 1 \leq k \leq i \leq n-1 \\ (t_k - p_{ki}), & \forall 1 \leq k \leq i \leq n \end{bmatrix},$$

where the five rows of sub-matrices are associated with constraints (A.23a)–(A.23e), respectively. To show that matrix $\mathcal{CG}_{t,q,p}^0$ is TU, we conduct pivot operations on the matrix with variables p_{ki} , t_k , and q_i . Note that a matrix is TU if and only if it remains TU after pivot operations (cf. Nemhauser and Wolsey, 1999). We conduct the following pivot operations in order.

1. For all $1 \leq k \leq i \leq n$, pivot with variable p_{ki} based on the component -1 in sub-matrix $(t_k - p_{ki})$ (corresponding to constraints (A.23e)). This pivot operation is equivalent to (a) adding $t_k - p_{ki}$, for all $1 \leq k \leq i \leq n$, to the left-hand side of every constraint (A.23c)–(A.23d) in which variable p_{ki} has coefficient 1, (b) adding $p_{ki} - t_k$, for all $1 \leq k \leq i \leq n$, to the left-hand side of every constraint (A.23a)–(A.23b) in which variable p_{ki} has coefficient -1 and (c) multiplying each left-hand side of constraint (A.23e) by -1 . As a result, the matrix after pivoting becomes

$$\mathcal{CG}_{t,q,p}^1 := \begin{bmatrix} (q_i + q_{i+1} - \sum_{k=1}^i t_k), & \forall 1 \leq i \leq n-1 \\ (q_n - \sum_{k=1}^n t_k) \\ (-q_i), & \forall 1 \leq i \leq n \\ (t_k), & \forall 1 \leq k \leq i \leq n-1 \\ (-t_k + p_{ki}), & \forall 1 \leq k \leq i \leq n \end{bmatrix},$$

Note that sub-matrix $(-t_k + p_{ki} + p_{k(i+1)})$ becomes (t_k) because the left-hand side of each constraint (A.23d), $-t_k + p_{ki} + p_{k(i+1)}$, is summed with $t_k - p_{ki}$ and $t_k - p_{k(i+1)}$ and so the coefficient of each t_k changes from -1 to 1 after pivoting. Meanwhile, sub-matrix $(-q_i - \sum_{k=1}^i t_k + \sum_{k=1}^i p_{ki})$ becomes $(-q_i)$ after pivoting because, for each $1 \leq i \leq n$, $-q_i - \sum_{k=1}^i t_k + \sum_{k=1}^i p_{ki}$ is summed with $t_k - p_{ki}$ for all $1 \leq k \leq i$.

2. For all $1 \leq k \leq n-1$, pivot with variable t_k based on any component 1 in sub-matrix (t_k) (note that there are multiple components 1 associated with each variable

t_k in (t_k) and we can pick any one of them). Since all components in each row of sub-matrix (t_k) are zeros except one equaling 1, these pivot operations (a) make all coefficients of all variables t_k zeros in matrix $\mathcal{CG}_{t,q,p}^1$ as long as $1 \leq k \leq n-1$, and (b) keep all coefficients of all variables q_i and p_{ki} unchanged. As a result, the matrix after pivoting becomes

$$\mathcal{CG}_{t,q,p}^2 := \begin{bmatrix} (q_i + q_{i+1}), & \forall 1 \leq i \leq n-1 \\ (q_n - t_n) \\ (-q_i), & \forall 1 \leq i \leq n \\ (t_k), & \forall 1 \leq k \leq i \leq n-1 \\ \left\{ \begin{array}{l} (p_{ki}), & \forall 1 \leq k \leq i \leq n-1 \\ (-t_n + p_{nn}), \end{array} \right. \end{bmatrix}.$$

3. For all $1 \leq i \leq n$, pivot with variable q_i based on any component -1 in sub-matrix $(-q_i)$ in $\mathcal{CG}_{t,q,p}^2$. Since $(-q_i)$ is an identity matrix, these pivot operations eliminate all coefficients of variables q_i in all other sub-matrices. As a result, the matrix after pivoting becomes

$$\mathcal{CG}_{t,q,p}^3 := \begin{bmatrix} (0), & \forall 1 \leq i \leq n-1 \\ (-t_n) \\ (q_i), & \forall 1 \leq i \leq n \\ (t_k), & \forall 1 \leq k \leq i \leq n-1 \\ \left\{ \begin{array}{l} (p_{ki}), & \forall 1 \leq k \leq i \leq n-1 \\ (-t_n + p_{nn}), \end{array} \right. \end{bmatrix}.$$

It follows that matrix $\mathcal{CG}_{t,q,p}^3$ contains only $\{-1, 0, 1\}$ entries, has no more than two nonzero entries in each row, and the sum of the entries is zero for each row containing two nonzero entries. Hence, matrix $\mathcal{CG}_{t,q,p}^3$ is TU, and so is matrix $\mathcal{CG}_{t,q,p}^0$.

Therefore, the extreme points of polyhedron $CG_{t,q,p}$ are integral and so property 1 is proved.

Third, to show property 2, we consider any extreme point (t, q, s, p, o) of polyhedron CG . Because $\sum_{k=1}^n t_k \leq 1$ and each $t_k \in \{0, 1\}$ by property 1, we show $p_{ki} = t_k q_i$ by discussing the following two cases on values of t_k .

1. If $t_k = 0$ for all $1 \leq k \leq n$, then $p_{ki} = 0$ for all $1 \leq k \leq i \leq n$ because $p_{ki} \leq t_k$. It follows that $p_{ki} = t_k q_i = 0$.
2. If there exists $1 \leq k^* \leq n$ such that $t_{k^*} = 1$, then any other $t_k = 0$. It follows that $p_{ki} = 0$, and so $p_{ki} = t_k q_i = 0$ for all $1 \leq k \leq i \leq n$ and $k \neq k^*$. For all $k^* \leq i \leq n$, constraints (A.13c) yield

$$-q_i - \sum_{k=1}^i t_k + \sum_{k=1}^i p_{ki} = -q_i - 1 + p_{k^*i} \geq -1 \quad \Rightarrow \quad p_{k^*i} \geq q_i.$$

Also, for all $k^* \leq i \leq n-1$, constraints (A.13d) yield

$$q_i + q_{i+1} + \sum_{k=1}^i t_k - \sum_{k=1}^i (p_{ki} + p_{k(i+1)}) = q_i + q_{i+1} + 1 - p_{k^*i} - p_{k^*(i+1)} \geq 1$$

$$\Rightarrow p_{k^*i} + p_{k^*(i+1)} \leq q_i + q_{i+1}.$$

It follows that $p_{k^*i} + p_{k^*(i+1)} = q_i + q_{i+1}$ for all $k^* \leq i \leq n-1$. Furthermore, constraint (A.13a) implies $q_n - \sum_{k=1}^n p_{kn} = q_n - p_{k^*n} \geq 0$, and so $q_n = p_{k^*n}$. Therefore, $q_i = p_{k^*i}$, or equivalently $q_i = t_{k^*} p_{k^*i}$ since $t_{k^*} = 1$, for all $k^* \leq i \leq n$.

Since (t, q, s, p, o) is an extreme point of polyhedron CG , each o_{ki} satisfies either inequality (A.21a) or (A.21b) at equality, and each s_i satisfies either inequality (A.21c) or (A.21d) at equality. We discuss the following cases to show $o_{ki} = s_i p_{ki} = t_k q_i s_i$.

1. If $t_k = 0$ for all $1 \leq k \leq n$, then $p_{ki} = 0$ for all $1 \leq k \leq i \leq n$ because $p_{ki} \leq t_k$. It follows that $o_{ki} = 0$ by constraints (A.12d). Therefore, $o_{ki} = s_i p_{ki} = 0$.
2. If there exists $1 \leq k^* \leq n$ such that $t_{k^*} = 1$, then any other $t_k = 0$. It follows that $p_{ki} = 0$, and so $o_{ki} = s_i p_{ki} = 0$ for all $1 \leq k \leq i \leq n$ and $k \neq k^*$. Then, for all $k^* \leq i \leq n$, inequalities (A.21c)–(A.21d) yield

$$s_i - \sum_{k=1}^i (o_{ki} - s_i^l p_{ki}) = s_i - o_{k^*i} + s_i^l p_{k^*i} \geq s_i^l, \quad (\text{A.24a})$$

$$s_i - \sum_{k=1}^i (o_{ki} - s_i^u p_{ki}) = s_i - o_{k^*i} + s_i^u p_{k^*i} \leq s_i^u. \quad (\text{A.24b})$$

Hence, each s_i satisfies either inequality (A.24a) or (A.24b) at equality. We discuss the following two sub-cases to finish the proof.

Sub-case 1. If $p_{k^*i} = 0$, then $o_{k^*i} = 0$ by constraints (A.12d). Therefore, $o_{k^*i} = s_i p_{k^*i} = 0$.

Sub-case 2. If $p_{k^*i} = 1$, then inequalities (A.24a)–(A.24b) imply $s_i = o_{k^*i}$. Therefore, $o_{k^*i} = s_i p_{k^*i}$.

□

Step 3: Finally, we prove Theorem A.1 based on the previous two propositions.

Proof. Proof of Theorem A.1 ($CG \supseteq \text{conv}(G)$) By Proposition 10, since polyhedron CG consists of either trivial equalities/inequalities or valid inequalities of set G , we have $(t, q, s, p, o) \in CG$ if $(t, q, s, p, o) \in G$. It follows that $CG \supseteq \text{conv}(G)$.

($CG \subseteq \text{conv}(G)$) By Proposition 11, since each extreme point (t, q, s, p, o) of CG satisfies properties 1, 2, and 3, $(t, q, s, p, o) \in G$. By the Minkowski's Theorem on polyhedron, we have $x \in \text{conv}(G)$ if $x \in CG$. It follows that $CG \subseteq \text{conv}(G)$. This completes the proof.

□

APPENDIX B

Appendix for Chapter 5

B.1 Proof of Theorem 5.1

Proof. Define dual variables r_i for (5.3b), $\begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix}$ for (5.3c), and G_i for (5.3d). The conic dual problem of (5.3) is

$$z_{\mathcal{D}_i^M} = \max_{G_i, p_i, r_i, H_i, q_i} \gamma_2 \Sigma_i^0 \cdot G_i + 1 - r_i + \Sigma_i^0 \cdot H_i + \gamma_1 q_i \quad (\text{B.1a})$$

$$\text{s.t.} \quad (s_i - \mu_i^0)^\top (-G_i)(s_i - \mu_i^0) + 2(p_i)^\top (s_i - \mu_i^0) + r_i \leq \mathbb{I}_{\{s_i: \sum_{j \in J(i)} s_{ij} \hat{y}_{ij} \leq T_i \hat{z}_i\}}(s_i), \quad \forall s_i \in \mathbb{R}^{|J|} \quad (\text{B.1b})$$

$$G_i \in \mathbb{S}_+^{|J| \times |J|}, \quad \begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix} \in \mathbb{S}_+^{(|J|+1) \times (|J|+1)}. \quad (\text{B.1c})$$

As strong duality holds for the primal and dual problems (Shapiro, 2001), satisfying the DR chance constraints (5.1e) with the ambiguity set \mathcal{D}_i^M , is equivalent to having solutions with objective value $z_{\mathcal{D}_i^M} \geq 1 - \alpha_i$, $i \in I$. After applying Lemma 1 in Jiang and Guan (2016), the dual problem (B.1) is equivalent to $\text{SDP}(\hat{y}_i, \hat{z}_i)$, after replacing the semi-infinite constraints (B.1b) with finite number of SDP constraints. This completes the proof. \square

APPENDIX C

Appendix for Chapter 6

C.1 Proof of Theorem 6.7

Proof: When $z_i = 1$, inequality (6.18a) reduces to the extended polymatroid inequality. When $z_i = 0$, we have $y_{ij} = 0$ for all $j \in [J]$ due to constraints (6.17b). It follows that inequality (6.18a) holds valid.

When $z_i = 1$, inequality (6.18b) reduces to the extended polymatroid inequality. When $z_i = 0$, we have $y_{ij} = 0$ for all $j \in [J]$ due to constraints (6.17b) and so $w_{ijk} = 0$ for all $j, k \in [J]$. It follows that $v_i = 0$ by definition. Hence, inequality (6.18b) holds valid. \square

C.2 Proof of Theorem 6.8

Proof: (**Validity of inequality (6.19a)**) If $j = k$, then $w_{ijk} = y_{ij}^2 = y_{ij}$. In this case, inequality (6.19a) reduces to $y_{ij} \geq 2y_{ij} + \sum_{\substack{\ell=1 \\ \ell \neq i}}^I y_{\ell j} - 1$, which clearly holds because $y_{ij} + \sum_{\substack{\ell=1 \\ \ell \neq i}}^I y_{\ell j} = \sum_{i=1}^I y_{ij} \leq 1$. If $j \neq k$, then we discuss the following two cases:

[1] If $\max\{y_{ij}, y_{ik}\} = 1$, then we assume $y_{ij} = 1$ without loss of generality. It follows that $y_{\ell j} = 0$ due to constraints (6.17c) and so $w_{\ell jk} = y_{\ell j}y_{\ell k} = 0$ for all $\ell \neq i$. Hence, $\sum_{\substack{\ell=1 \\ \ell \neq i}}^I w_{\ell jk} = 0$ and inequality (6.19a) reduces to $w_{ijk} \geq y_{ij} + y_{ik} - 1$, which holds valid.

[2] If $\max\{y_{ij}, y_{ik}\} = 0$, then $y_{ij} = y_{ik} = 0$ and $w_{ijk} = 0$. It remains to show $\sum_{\substack{\ell=1 \\ \ell \neq i}}^I w_{\ell jk} \leq 1$. Indeed, since $w_{\ell jk} \leq y_{\ell j}$, we have $\sum_{\substack{\ell=1 \\ \ell \neq i}}^I w_{\ell jk} \leq \sum_{\substack{\ell=1 \\ \ell \neq i}}^I y_{\ell j} \leq \sum_{\ell=1}^I y_{\ell j} = 1$, where the last equality is due to constraints (6.17c).

(**Validity of inequality (6.19b)**) This inequality clearly holds valid when $z_i = 1$. When $z_i = 0$, we have $y_{ij} = y_{ik} = 0$ due to constraints (6.17b). It follows that $w_{ijk} = y_{ij}y_{ik} = 0$

and so the inequality holds valid.

(Validity of inequality (6.19c)) This inequality holds valid when $z_i = 0$. When $z_i = 1$, this inequality is equivalent to $y_{ik} \sum_{j=1, j \neq k}^J y_{ij} \leq \sum_{j=1}^J y_{ij} - 1$. We discuss the following two cases:

[1] If $y_{ik} = 0$, then $\sum_{j=1}^J y_{ij} \geq 1$ without loss of optimality because $z_i = 1$. Inequality (6.19c) holds valid.

[2] If $y_{ik} = 1$, then $y_{ik} \sum_{j=1, j \neq k}^J y_{ij} = \sum_{j=1, j \neq k}^J y_{ij}$. Meanwhile, $\sum_{j=1}^J y_{ij} - 1 = \sum_{j=1, j \neq k}^J y_{ij} + y_{ik} - 1 = \sum_{j=1, j \neq k}^J y_{ij}$. Inequality (6.19c) holds valid.

(Validity of inequality (6.19d)) This inequality holds valid when $z_i = 0$. When $z_i = 1$, we have $\sum_{j=1}^J y_{ij} \geq 1$ without loss of optimality. It follows that $\sum_{j=1}^J y_{ij} = 1$ or $\sum_{j=1}^J y_{ij} \geq 2$, and so $(\sum_{j=1}^J y_{ij} - 1)(\sum_{j=1}^J y_{ij} - 2) \geq 0$. Hence,

$$\begin{aligned} \left(\sum_{j=1}^J y_{ij} - 1 \right) \left(\sum_{j=1}^J y_{ij} - 2 \right) &= \sum_{j=1}^J y_{ij}^2 + 2 \sum_{j=1}^J \sum_{k=j+1}^J y_{ij} y_{ik} - 3 \left(\sum_{j=1}^J y_{ij} \right) + 2 \\ &= \sum_{j=1}^J y_{ij} + 2 \sum_{j=1}^J \sum_{k=j+1}^J w_{ijk} - 3 \left(\sum_{j=1}^J y_{ij} \right) + 2 \\ &= 2 \sum_{j=1}^J \sum_{k=j+1}^J w_{ijk} - 2 \left[\left(\sum_{j=1}^J y_{ij} \right) - 1 \right] \geq 0. \end{aligned}$$

Inequality (6.19d) follows. □

C.3 Out-of-Sample Performance of DCBP

Through testing instances of chance-constrained bin packing, we show that DCBP solutions have very low probabilities of violating capacities in all the out-of-sample tests, even when the distributional information is misspecified. Specifically, we evaluate the out-of-sample performance of the optimal solutions to the DCBP1, DCBP2, Gaussian-based 0-1 SOC models, and the SAA-based MILP model. To generate the out-of-sample reference scenarios, we consider either misspecified distribution type or misspecified moment information as follows.

- **Misspecified distribution type:** We sample 10,000 out-of-sample data points from a two-point distribution having the same mean and standard deviation of each random

variable \tilde{t}_{ij} for $i \in [I]$ and $j \in [J]$ as the in-sample data. The service time is realized as $\mu_{ij} + \frac{(1-p)}{\sqrt{p(1-p)}}\sigma_{ij}$ with probability p ($0 < p < 1$) and as $\mu_{ij} - \frac{\sqrt{p(1-p)}}{(1-p)}\sigma_{ij}$ with probability $1 - p$, where μ_{ij} and σ_{ij} are the sample mean and standard deviation of \tilde{t}_{ij} obtained from the in-sample data. We set $p = 0.3$ so that we have smaller probability of having larger service time realizations.

- **Misspecified moments:** Alternatively, we sample 10,000 data points from the Gaussian distribution, but only consider the hMℓV type of appointments, instead of an equal mixture of all the four types. In each sample and for each $i \in [I]$, we draw a standard-Gaussian random number ρ_i , and for each $j \in [J]$, generate a service time realization as $\mu_{ij} + \rho_i\sigma_{ij}$.

Performance of solutions under diagonal matrices Under diagonal matrices, both of the two DCBP models open three servers (i.e., Servers 4, 5, 6 by DCBP1 and Servers 2, 4, 5 by DCBP2), while the Gaussian and SAA approaches only open Servers 4 and 6. We first use the 10,000 out-of-sample data points given by misspecified distribution type, namely, the two-point distribution. Table C.1 reports each solution’s probability of having the total time of assigned appointments not exceeding the capacity of the server to which they are assigned.

Table C.1: Solution reliability in out-of-sample data following a misspecified distribution type

Model	Server 2	Server 4	Server 5	Server 6
DCBP1	N/A	1.00	1.00	1.00
DCBP2	1.00	1.00	1.00	N/A
Gaussian	N/A	0.69	N/A	0.91
SAA	N/A	0.69	N/A	0.91

“N/A”: the server is not opened by using the corresponding method.

Recall that $\alpha_i = 0.05$ for all i used in all four approaches. The reliability results of the Gaussian and SAA approaches are significantly lower than the desired probability threshold $1 - \alpha_i = 0.95$ on Server 4, and slightly lower than 0.95 on Server 6. On the other hand, the optimal solutions of DCBP1 and DCBP2 do not exceed the capacity of any open servers.

Next, we use the 10,000 out-of-sample data points given by misspecified moments. Table C.2 reports the reliability performance of each optimal solution. The DCBP2 solution still outperforms solutions given by all the other approaches and achieves the desired reliability in all the three open servers. The Gaussian and SAA solutions perform poorly when the moment information is different from the empirical inputs. The DCBP1 solution

Table C.2: Solution reliability in out-of-sample scenarios with misspecified moments

Model	Server 2	Server 4	Server 5	Server 6
DCBP1	N/A	0.94	1.00	1.00
DCBP2	0.98	1.00	0.99	N/A
Gaussian	N/A	0.59	N/A	0.89
SAA	N/A	0.59	N/A	0.89

“N/A”: the server is not opened by using the corresponding method.

respects the capacities of Servers 5 and 6 with sufficiently high probability (i.e., > 0.95), but yields a slightly lower reliability (0.94) than the threshold on Server 4.

Performance of solutions under general matrices We optimize all the models under general matrices by using the empirical covariance matrices of the in-sample data, and report their corresponding solutions in Table C.3. Each entry illustrates the number of appointments assigned to an open server. Note that the Gaussian and SAA approaches yield the same solution of opening servers and assigning appointments.

Table C.3: Optimal open servers and appointment-to-server assignments under general matrices

Model	Server 3	Server 4	Server 5	Server 6
DCBP1	12	N/A	13	7
DCBP2	12	11	9	N/A
Gaussian	15	N/A	17	N/A
SAA	15	N/A	17	N/A

“N/A”: the server is not opened by using the corresponding method.

We test the solutions shown in Table C.3 in the out-of-sample scenarios under misspecified distribution type, and present their reliability performance in Table C.4. We again show that under general covariance matrices, the DCBP2 model yields the most conservative solution that does not exceed any open server’s capacity, while DCBP1 only ensures the desired reliability on Servers 3 and 6, but not on Server 5. The Gaussian and SAA approaches cannot produce solutions that can achieve the desired reliability threshold on any of their open servers.

Table C.4: Solution reliability in out-of-sample scenarios with misspecified moments

Model	Server 3	Server 4	Server 5	Server 6
DCBP1	1.00	N/A	0.91	1.00
DCBP2	1.00	1.00	1.00	N/A
Gaussian	0.91	N/A	0.91	N/A
SAA	0.91	N/A	0.91	N/A

“N/A”: the server is not opened by using the corresponding method.

BIBLIOGRAPHY

- Atamtürk, A. and Bhardwaj, A. (2015). Supermodular covering knapsack polytope. *Discrete Optimization*, 18:74–86.
- Atamtürk, A. and Bhardwaj, A. (2018). Network design with probabilistic capacities. *Networks*, 71(1):16–30.
- Atamtürk, A. and Gómez, A. (2017). Polymatroid inequalities for p-order conic mixed 0-1 optimization. Available at arXiv e-prints <https://arxiv.org/abs/1705.05918>.
- Atamtürk, A. and Narayanan, V. (2008). Polymatroids and mean-risk minimization in discrete optimization. *Operations Research Letters*, 36(5):618–622.
- Atamtürk, A. and Narayanan, V. (2009). The submodular knapsack polytope. *Discrete Optimization*, 6(4):333–344.
- Barron, W. M. (1980). Failed appointments. Who misses them, why they are missed, and what can be done. *Primary care*, 7(4):563–574.
- Bashash, S. and Fathy, H. (2013). Modeling and control of aggregate air conditioning loads for robust renewable power management. *IEEE Transactions on Control Systems Technology*, 21(4):1318–1327.
- Begen, M. A., Levi, R., and Queyranne, M. (2012). Technical note – a sampling-based approach to appointment scheduling. *Operations Research*, 60(3):675–681.
- Begen, M. A. and Queyranne, M. (2011). Appointment scheduling with discrete random durations. *Mathematics of Operations Research*, 36(2):240–257.
- Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.
- Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ.
- Ben-Tal, A. and Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424.
- Ben-Tal, A. and Nemirovski, A. (2002). Robust optimization – methodology and applications. *Mathematical Programming*, 92(3):453–480.
- Berg, B. P., Denton, B. T., Erdogan, S. A., Rohleder, T., and Huschka, Todd, R. (2014). Optimal booking and scheduling in outpatient procedure centers. *Computers & Operations Research*, 50:24–37.
- Bertsimas, D., Doan, X. V., Natarajan, K., and Teo, C.-P. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602.

- Bertsimas, D., Litvinov, E., Sun, X. A., Zhao, J., and Zheng, T. (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63.
- Bertsimas, D. and Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804.
- Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52(1):35–53.
- Bhardwaj, A. (2015). *Binary conic quadratic knapsacks*. PhD thesis, University of California, Berkeley.
- Bienstock, D., Chertkov, M., and Harnett, S. (2014). Chance-constrained optimal power flow: Risk-aware network control under uncertainty. *SIAM Review*, 56(3):461–495.
- Birge, J. R. and Louveaux, F. V. (2011). *Introduction to Stochastic Programming*. Springer, New York, NY.
- Blake, J. T. and Donald, J. (2002). Mount sinai hospital uses integer programming to allocate operating room time. *Interfaces*, 32(2):63–73.
- Braslavsky, J., Perfumo, C., and Ward, J. (2013). Model-based feedback control of distributed air-conditioning loads for fast demand-side ancillary services. In *Proceedings of IEEE Conference on Decision and Control*, pages 6274–6279.
- Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877.
- Calafiore, G. C. and El Ghaoui, L. (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22.
- California ISO (2011). Regulation energy management draft final proposal.
- Callaway, D. S. (2009). Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. *Energy Conversion and Management*, 50(5):1389–1400.
- Campi, M., Calafiore, G., and Prandini, M. (2009). The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149–157.
- Carta, J. A., Ramirez, P., and Velazquez, S. (2009). A review of wind speed probability distributions used in wind energy analysis: Case studies in the canary islands. *Renewable and sustainable energy reviews*, 13(5):933–955.
- Cayirli, T. and Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *Production and operations management*, 12(4):519–549.
- Cayirli, T., Yang, K. K., and Quek, S. A. (2012). A universal appointment rule in the presence of no-shows and walk-ins. *Production and Operations Management*, 21(4):682–697.
- Chen, R. R. and Robinson, L. W. (2014). Sequencing and scheduling appointments with potential call-in patients. *Production and Operations Management*, 23(9):1522–1538.
- Chen, W., Sim, M., Sun, J., and Teo, C. P. (2010). From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Operations Research*, 58(2):470–485.
- Cheng, J., Delage, E., and Lissner, A. (2014). Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization*, 24(3):1485–1506.
- Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.
- Deng, Y. and Shen, S. (2016). Decomposition algorithms for optimizing multi-server appointment scheduling with chance constraints. *Mathematical Programming*, 157(1):245–276.

- Deng, Y., Shen, S., and Denton, B. (2016). Chance-constrained surgery planning under conditions of limited and ambiguous data.
- Denton, B. and Gupta, D. (2003). A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35(11):1003–1016.
- Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science*, 10(1):13–24.
- Denton, B. T., Miller, A. J., Balasubramanian, H. J., and Huschka, T. R. (2010). Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, 58(4):802–816.
- Edmonds, J. (1970). Submodular functions, matroids, and certain polyhedra. *Combinatorial structures and their applications*, pages 69–87.
- El Ghaoui, L., Oks, M., and Oustry, F. (2003). Worst-case Value-at-Risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556.
- Epstein, L., Levin, A., Marchetti-Spaccamela, A., Megow, N., Mestre, J., Skutella, M., and Stougie, L. (2012). Universal sequencing on an unreliable machine. *SIAM Journal on Computing*, 41(3):565–586.
- Erdogan, S. A. and Denton, B. T. (2010). Surgery planning and scheduling. *Wiley encyclopedia of operations research and management science*.
- Erdogan, S. A. and Denton, B. T. (2013). Dynamic appointment scheduling of a stochastic server with uncertain demand. *INFORMS Journal on Computing*, 25(1):116–132.
- Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166.
- Gabrel, V., Murat, C., and Thiele, A. (2014). Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3):471–483.
- Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.
- Ge, D., Wan, G., Wang, Z., and Zhang, J. (2013). A note on appointment scheduling with piecewise linear cost functions. *Mathematics of Operations Research*, 39(4):1244–1251.
- Grant, M. and Boyd, S. (2012). CVX: Matlab software for disciplined convex programming, version 2.0 beta. <http://cvxr.com/cvx>.
- Gul, S., Denton, B. T., Fowler, J. W., and Huschka, T. (2011). Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations management*, 20(3):406–417.
- Gupta, D. and Denton, B. T. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9):800–819.
- Gurvich, I., Luedtke, J., and Tezcan, T. (2010). Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science*, 56(7):1093–1115.
- Han, J., Lee, K., Lee, C., Choi, K.-S., and Park, S. (2016). Robust optimization approach for a chance-constrained binary knapsack problem. *Mathematical Programming*, 157(1):277–296.
- Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2015). A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62.
- Hao, H., Sanandaji, B., Poolla, K., and Vincent, T. (2015). Aggregate flexibility of thermostatically controlled loads. *IEEE Transactions on Power Systems*, 30(1):189–198.

- Hassin, R. and Mendel, S. (2008). Scheduling arrivals to queues: A single-server model with no-shows. *Management Science*, 54(3):565–572.
- He, S., Sim, M., and Zhang, M. (2015). Data-driven patient scheduling in emergency departments: A hybrid robust-stochastic approach. Available at *Optimization-Online* http://www.optimization-online.org/DB_HTML/2015/11/5213.html.
- Ho, C.-J. and Lau, H.-S. (1992). Minimizing total cost in scheduling outpatient appointments. *Management science*, 38(12):1750–1764.
- Jabr, R. A. (2013). Adjustable robust OPF with renewable energy sources. *IEEE Transactions on Power Systems*, 28(4):4742–4751.
- Jabr, R. A., Karaki, S., and Korbane, J. A. (2015). Robust multi-period OPF with storage and renewables. *IEEE Transactions on Power Systems*, 30(5):2790–2799.
- Jiang, R. and Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming, Series A*, 158(1):291–327.
- Jiang, R., Shen, S., and Zhang, Y. (2015). Distributionally robust appointment scheduling with random no-shows and service durations. Available at *Optimization-Online*: http://www.optimization-online.org/DB_FILE/2015/08/5080.pdf.
- Jiang, R., Wang, J., and Guan, Y. (2012). Robust unit commitment with wind power and pumped storage hydro. Forthcoming in *IEEE Transactions on Power Systems*.
- Kaandorp, G. C. and Koole, G. (2007). Optimal outpatient appointment scheduling. *Health Care Management Science*, 10(3):217–229.
- Katiraei, F. and Aguerro, J. R. (2011). Solar PV integration challenges. *IEEE Power and Energy Magazine*, 9(3):62–71.
- Klassen, K. J. and Yoogalingam, R. (2009). Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management*, 18(4):447–458.
- Kleinberg, J., Rabani, Y., and Tardos, É. (2000). Allocating bandwidth for bursty connections. *SIAM Journal on Computing*, 30(1):191–217.
- Kleywegt, A. J., Shapiro, A., and Homem-de-Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502.
- Kong, Q., Lee, C.-Y., Teo, C.-P., and Zheng, Z. (2013). Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3):711–726.
- Kong, Q., Li, S., Liu, N., Teo, C.-P., and Yan, Z. (2018). Appointment scheduling under schedule-dependent patient no-show behavior. Available at RePub: <https://repub.eur.nl/pub/106439/Noshow-MS-1030c.pdf>.
- Kosuch, S. and Lissner, A. (2010). Upper bounds for the 0-1 stochastic knapsack problem and a b&b algorithm. *Annals of Operations Research*, 176(1):77–93.
- Küçükyavuz, S. (2012). On mixing sets arising in chance-constrained programming. *Mathematical Programming*, 132(1–2):31–56.
- LaGanga, L. R. and Lawrence, S. R. (2012). Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management*, 21(5):874–888.
- Lee, V. J., Earnest, A., Chen, M. I., and Krishnan, B. (2005). Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases. *BMC Health Services Research*, 5(1):51.

- Li, B. and Mathieu, J. (2015). Analytical reformulation of chance-constrained optimal power flow with uncertain load control. In *Proceedings of IEEE PowerTech*, pages 1–6, Eindhoven, Netherlands.
- Lin, J., Muthuraman, K., and Lawley, M. (2011). Optimal and approximate algorithms for sequential clinical scheduling with no-shows. *IIE Transactions on Healthcare Systems Engineering*, 1(1):20–36.
- Liu, N., Ziya, S., and Kulkarni, V. G. (2010). Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management*, 12(2):347–364.
- Liu, X., Küçükyavuz, S., and Luedtke, J. (2016). Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, 157(1):219–243.
- Lubin, M., Dvorkin, Y., and Backhaus, S. (2015). A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Transactions on Power Systems*, 31:3840–3849.
- Luedtke, J. (2014). A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming*, 146(1-2):219–244.
- Luedtke, J. and Ahmed, S. (2008). A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699.
- Luedtke, J., Ahmed, S., and Nemhauser, G. (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2):247–272.
- Luo, J., Kulkarni, V. G., and Ziya, S. (2012). Appointment scheduling under patient no-shows and service interruptions. *Manufacturing & Service Operations Management*, 14(4):670–684.
- MacDonald, J., Cappers, P., Callaway, D., and Kiliccote, S. (2012). Demand response providing ancillary services. In *Proceedings of the Grid Interop*.
- Mak, H.-Y., Rong, Y., and Zhang, J. (2014). Sequencing appointments for service systems using inventory approximations. *Manufacturing & Service Operations Management*, 16(2):251–262.
- Mak, H.-Y., Rong, Y., and Zhang, J. (2015). Appointment scheduling with limited distributional information. *Management Science*, 61(2):316–334.
- Mancilla, C. (2009). *Stochastic sequencing and scheduling of an operating room*. PhD thesis, Lehigh University, Department of Industrial and Systems Engineering.
- Margellos, K., Goulart, P., and Lygeros, J. (2014). On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Transactions on Automatic Control*, 59(8).
- Mathieu, J., González Vayá, M., and Andersson, G. (2013a). Uncertainty in the flexibility of aggregations of demand response resources. In *Proceedings of the IEEE Industrial Electronics Society Annual Conference*, pages 8052–8057, Vienna, Austria.
- Mathieu, J., Kamgarpour, M., Lygeros, J., Andersson, G., and Callaway, D. (2015). Arbitraging intraday wholesale energy market prices with aggregations of thermostatic loads. *IEEE Transactions on Power Systems*, 30(2):763–772.
- Mathieu, J., Koch, S., and Callaway, D. (2013b). State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Transactions on Power Systems*, 28(1):430–440.
- Meyn, S., Barooah, P., Busic, A., Chen, Y., and Ehren, J. (2015). Ancillary service to the grid using intelligent deferrable loads. *IEEE Transactions on Automatic Control*, 60(11):2847–2862.

- Mittal, S., Schulz, A. S., and Stiller, S. (2014a). Robust appointment scheduling. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- Mittal, S., Schulz, A. S., and Stiller, S. (2014b). Robust appointment scheduling. In *LIPICs-Leibniz International Proceedings in Informatics*, volume 28. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- Moore, C. G., Wilson-Witherspoon, P., and Probst, J. C. (2001). Time and money: Effects of no-shows at a family practice residency clinic. *Family Medicine-Kansas City-*, 33(7):522–527.
- Morgan, E. C., Lackner, M., Vogel, R. M., and Baise, L. G. (2011). Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52(1):15–26.
- Muthuraman, K. and Lawley, M. (2008). A stochastic overbooking model for outpatient clinical scheduling with no-shows. *IIE Transactions*, 40(9):820–837.
- Nemhauser, G. and Wolsey, L. (1999). *Integer and Combinatorial Optimization*. Wiley.
- Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294.
- Pagnoncelli, B., Ahmed, S., and Shapiro, A. (2009). Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416.
- Papaefthymiou, G. and Klockl, B. (2008). MCMC for wind power simulation. *IEEE Transactions on Energy Conversion*, 23(1):234–240.
- Papavasiliou, A. and Oren, S. S. (2013). Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research*, 61(3):578–592.
- Parizi, M. S. and Ghate, A. (2015). Multi-class, multi-resource advance scheduling with no-shows, cancellations and overbooking. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2550560>.
- Pinedo, M. L. (2016). *Scheduling: Theory, Algorithms, and Systems*. Springer.
- Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112.
- Prékopa, A. (2003). Probabilistic programming. In *Stochastic Programming: Handbooks in Operations Research and Management Science*, volume 10, chapter 5, pages 267–351. Elsevier.
- Qi, J. (2016). Mitigating delays and unfairness in appointment systems. *Management Science*, 63(2):566–583.
- Qiu, F. and Wang, J. (2015). Distributionally robust congestion management with dynamic line ratings. *IEEE Transactions Power Systems*.
- Roald, L., Oldewurtel, F., Krause, T., and Andersson, G. (2013). Analytical reformulation of security constrained optimal power flow with probabilistic constraints. In *Proceedings of IEEE PowerTech*.
- Roald, L., Oldewurtel, F., Van Parys, B., and Andersson, G. (2015). Security constrained optimal power flow with distributionally robust chance constraints. *arXiv preprint arXiv:1508.06061*.
- Robinson, L. W. and Chen, R. R. (2010). A comparison of traditional and open-access policies for appointment scheduling. *Manufacturing & Service Operations Management*, 12(2):330–346.
- Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional Value-at-Risk. *Journal of Risk*, 2(3):21–42.

- Rockafellar, R. T. and Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471.
- Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. (2015). Robust growth-optimal portfolios. *Management Science*, 62(7):2090–2109.
- Scarf, H., Arrow, K., and Karlin, S. (1958). A min-max solution of an inventory problem. In *Studies in the Mathematical Theory of Inventory and Production*, volume 10, pages 201–209. Stanford University Press, Stanford, CA.
- Shapiro, A. (2001). On duality theory of conic linear problems. In *Semi-Infinite Programming*, pages 135–165. Springer.
- Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on Stochastic Programming: Modeling and Theory*, volume 9. SIAM, Philadelphia, PA.
- Shen, S. and Wang, J. (2014). Stochastic modeling and approaches for managing energy footprints in cloud computing services. *INFORMS Service Science*, 6(1):15–33.
- Shi, Y., Reich, D., Epelman, M., Klampfl, E., and Cohn, A. (2017). An analytical approach to prototype vehicle test scheduling. *Omega*, 67:168–176.
- Shylo, O. V., Prokopyev, O. A., and Schaefer, A. J. (2012). Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing*, 25(4):682–692.
- Simchi-Levi, D., Chen, X., and Bramel, J. (2014). *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management*. Springer Verlag.
- Sion, M. (1958). On general minimax theorems. *Pacific J. Math*, 8(1):171–176.
- Smith, J. E. and Winkler, R. L. (2006). The optimizer’s curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322.
- Song, Y., Luedtke, J. R., and Küçükyavuz, S. (2014). Chance-constrained binary packing problems. *INFORMS Journal on Computing*, 26(4):735–747.
- Soriano, A. (1966). Comparison of two scheduling systems. *Operations Research*, 14(3):388–397.
- Summers, T., Warrington, J., Morari, M., and Lygeros, J. (2015). Stochastic optimal power flow based on conditional value at risk and distributional robustness. *International Journal of Electrical Power & Energy Systems*, 72:116–125.
- Van Parys, B. P., Goulart, P. J., and Kuhn, D. (2016). Generalized Gauss inequalities via semidefinite programming. *Mathematical Programming*, 156(1-2):271–302.
- Vandenbergh, L., Boyd, S., and Comanor, K. (2007). Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52.
- Vrakopoulou, M., Margellos, K., Lygeros, J., and Andersson, G. (2013). A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration. *IEEE Transactions on Power Systems*, 28(4).
- Vrakopoulou, M., Mathieu, J., and Andersson, G. (2014). Stochastic optimal power flow with uncertain reserves from demand response. In *Proceedings of Hawaii International Conference on System Science (HICSS)*, pages 2353–2362.
- Wagner, M. R. (2008). Stochastic 0–1 linear programming under limited distributional information. *Operations Research Letters*, 36(2):150–156.
- Wang, Y., Tang, J., and Fung, R. Y. (2014). A column-generation-based heuristic algorithm for solving operating theater planning problem under stochastic demand and surgery cancellation risk. *International Journal of Production Economics*, 158:28–36.

- Wang, Z., Glynn, P., and Ye, Y. (2013). Likelihood robust optimization for data-driven problems. *arXiv preprint arXiv:1307.6279*.
- Wei, W., Li, N., Wang, J., and Mei, S. (2016). Estimating the probability of infeasible real-time dispatch without exact distributions of stochastic wind generations. *IEEE Transactions on Power Systems*, 31(6):5022–5032.
- Wiesemann, W., Kuhn, D., and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376.
- Wolsey, L. A. (1998). *Integer Programming*. Wiley, New York, NY.
- Xie, L., Carvalho, P. M., Ferreira, L. A., Liu, J., Krogh, B. H., Popli, N., and Ilic, M. D. (2011). Wind integration in power systems: Operational challenges and possible solutions. *Proceedings of the IEEE*, 99(1):214–232.
- Yanikoğlu, İ. and den Hertog, D. (2013). Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4):666–681.
- Yu, Y.-L., Li, Y., Schuurmans, D., and Szepesvári, C. (2009). A general projection property for distribution families. In *Advances in Neural Information Processing Systems*, pages 2232–2240.
- Zacharias, C. and Pinedo, M. (2014). Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, 23(5):788–801.
- Zangwill, W. I. (1966). A deterministic multi-period production scheduling model with backlogging. *Management Science*, 13(1):105–119.
- Zangwill, W. I. (1969). A backlogging model and a multi-echelon model of a dynamic economic lot size production system – a network approach. *Management Science*, 15(9):506–527.
- Zeng, B., Turkcan, A., Lin, J., and Lawley, M. (2010). Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities. *Annals of Operations Research*, 178(1):121–144.
- Zhang, H. and Li, P. (2011). Chance constrained programming for optimal power flow under uncertainty. *IEEE Transactions on Power Systems*, 26(4):2417–2424.
- Zhang, W., Lian, J., Chang, C., and Kalsi, K. (2013). Aggregated modeling and control of air conditioning loads for demand response. *IEEE Transactions on Power Systems*, 28(4):4655–4664.
- Zhang, Y., Jiang, R., and Shen, S. (2018a). Ambiguous chance-constrained binary programs under mean-covariance information. *SIAM Journal on Optimization*, 28(4):2922–2944.
- Zhang, Y., Shen, S., and Erdogan, S. A. (2017). Distributionally robust appointment scheduling with moment-based ambiguity set. *Operations Research Letters*, 45(2):139–144.
- Zhang, Y., Shen, S., and Erdogan, S. A. (2018b). Solving 0–1 semidefinite programs for distributionally robust allocation of surgery blocks. *Optimization Letters*, 12(7):1503–1521.
- Zhang, Y., Shen, S., and Mathieu, J. (2016). Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads. *IEEE Transactions on Power Systems*.
- Zhang, Y., Shen, S., and Mathieu, J. L. (2015a). Data-driven optimization approaches for optimal power flow with uncertain reserves from load control. In *2015 American Control Conference (ACC)*, pages 3013–3018. IEEE.
- Zhang, Z., Denton, B., and Xie, X. (2015b). Branch and price for chance constrained bin packing. Available at Optimization-Online: http://www.optimization-online.org/DB_HTML/2015/11/5217.html.

- Zhao, C. and Guan, Y. (2018). Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267.
- Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. (2011). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19.
- Zymler, S., Kuhn, D., and Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198.