

# Data-Driven Dynamic Optimization with Auxiliary Covariates

by

Christopher George McCord

B.S.E., Princeton University (2015)

Submitted to the Sloan School of Management  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author .....  
Sloan School of Management  
May 17, 2019

Certified by.....  
Dimitris Bertsimas  
Boeing Professor of Operations Research  
Co-director, Operations Research Center  
Thesis Supervisor

Accepted by .....  
Patrick Jaillet  
Dugald C. Jackson Professor, Department of Electrical Engineering and  
Computer Science  
Co-director, Operations Research Center



# Data-Driven Dynamic Optimization with Auxiliary Covariates

by

Christopher George McCord

Submitted to the Sloan School of Management  
on May 17, 2019, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy in Operations Research

## Abstract

Optimization under uncertainty forms the foundation for many of the fundamental problems the operations research community seeks to solve. In this thesis, we develop and analyze algorithms that incorporate ideas from machine learning to optimize uncertain objectives directly from data.

In the first chapter, we consider problems in which the decision affects the observed outcome, such as in personalized medicine and pricing. We present a framework for using observational data to learn to optimize an uncertain objective over a continuous and multi-dimensional decision space. Our approach accounts for the uncertainty in predictions, and we provide theoretical results that show this adds value. In addition, we test our approach on a Warfarin dosing example, and it outperforms the leading alternative methods.

In the second chapter, we develop an approach for solving dynamic optimization problems with covariates that uses machine learning to approximate the unknown stochastic process of the uncertainty. We provide theoretical guarantees on the effectiveness of our method and validate the guarantees with computational experiments.

In the third chapter, we introduce a distributionally robust approach for incorporating covariates in large-scale, data-driven dynamic optimization. We prove that it is asymptotically optimal and provide a tractable general-purpose approximation scheme that scales to problems with many temporal stages. Across examples in shipment planning, inventory management, and finance, our method achieves improvements of up to 15% over alternatives.

In the final chapter, we apply the techniques developed in previous chapters to the problem of optimizing the operating room schedule at a major US hospital. Our partner institution faces significant census variability throughout the week, which limits the amount of patients it can accept due to resource constraints at peak times. We introduce a data-driven approach for this problem that combines machine learning with mixed integer optimization and demonstrate that it can reliably reduce the maximal weekly census.

Thesis Supervisor: Dimitris Bertsimas

Title: Boeing Professor of Operations Research  
Co-director, Operations Research Center

## Acknowledgments

As my time at grad school comes to a close, I cannot help but feel incredibly fortunate to have had the opportunity to study at MIT. The past four years have certainly helped shape the trajectory of my life, both professionally and personally.

I owe a special thanks to my advisor, Dimitris, who has been an important mentor and has helped me mature as a researcher. He has often accused me of being overly reserved in my opinions, so let me explicitly express that I admire his passion for impacting the world through research and respect his genuine concern for the well-being of all of his students.

I am lucky that Vivek Farias and Colin Fogarty were both willing to serve on my thesis committee. They provided very helpful feedback and challenged me in ways that forced me to better understand the problems I was working on. In addition, I am grateful that Colin served on my general exam committee, along with John Tsitsiklis. Their comments helped me strengthen Chapter 3. I am also grateful for Rahul Mazumder, who has answered all of my questions about computational statistics over the past few years. I developed a much stronger understanding thanks to his patience. In addition, I owe thanks to Roy Welsch, who advised me and provided funding for me during my first year.

My thesis would not have been possible without the ORC administrative staff, Laura Rose, Andrew Carvalho, and Nikki Fortes, who have all been friendly and helpful in addressing my numerous questions and requests.

I am grateful for my collaborators, both internal and external. Brad Sturt, who worked with me on Chapter 4, constantly pushed me to explain my thought process clearly, and my understanding of the problem, as well as my writing, improved as a result. Martin Copenhaver worked with me on Chapter 5, and I benefited significantly from his deep understanding of the healthcare industry, as well as his impressive work ethic. I also collaborated with Nishanth Mundru on several projects, and I am grateful for his extensive knowledge of the literature and his attention to detail. In addition, Chapter 5 would not have been possible if not for my external collaborators at Beth

Israel Deaconess Medical Center, especially Manu and Ryan.

I was very fortunate to serve as a teaching assistant for classes taught by Rob Freund and Caroline Uhler. Both are outstanding lecturers, and I learned a lot by observing them teach. In addition, both provided me with important mentorship during my early years of grad school. I am also blessed to have had the mentorship and spiritual guidance of Father Dan Moloney, the TCC chaplain, throughout the past four years. Father Dan consistently provided insightful analysis of Scripture and highlighted its relevance to issues in the lives of students.

When I first started at MIT, I did not expect to develop many lasting friendships, but I am extremely lucky to have been surrounded by some incredible people in the ORC. Yeesian Ng has been a great friend throughout my four years in grad school. In addition to teaching me a significant amount about robust optimization, he also gave me very helpful feedback on my defense presentation. I'm glad I got to know him through both the ORC and the TCC. Jourdain Lamperski has also been a great friend, and I admire his creativity and passion for research. He has outstanding geometric intuition about problems, and I've learned a lot from him. He's also really entertaining to be around, and I've enjoyed getting to know him and his wife, Lauren. I am very grateful for my friendship with Scott Hunter. Scott has an incredible work ethic and the humility to constantly push himself to be better in all aspects of life. I'm lucky to call him a friend because he pushes me to be better as well.

My work over these past four years would never have been possible without the support of my friends and family. I was very fortunate to have a few amazing friends in my life during my time as an undergrad, and I am even more fortunate to still have them in my life today. Finally, I cannot possibly express fully my gratitude for the support and sacrifices of my family. My brother, my mom, and my dad have always taken interest in my work and have consistently encouraged me in all areas of my life, and, for that, I am eternally grateful.

# Contents

<b>1</b>	<b>Introduction</b>	<b>15</b>
1.1	Background . . . . .	16
1.2	Contributions . . . . .	18
<b>2</b>	<b>Optimization over Continuous and Multi-Dimensional Decisions with Observational Data</b>	<b>23</b>
2.1	Introduction . . . . .	23
2.1.1	Notation . . . . .	25
2.1.2	A Motivating Example for Uncertainty Penalization . . . . .	26
2.1.3	Related Work . . . . .	28
2.1.4	Contributions . . . . .	30
2.2	Approach . . . . .	30
2.2.1	Parameter Tuning . . . . .	32
2.3	Theory . . . . .	33
2.3.1	Tractability . . . . .	36
2.4	Results . . . . .	37
2.4.1	Pricing . . . . .	37
2.4.2	Warfarin Dosing . . . . .	38
2.5	Conclusions . . . . .	41
<b>3</b>	<b>From Predictions to Prescriptions in Multistage Optimization Prob- lems</b>	<b>43</b>
3.1	Introduction . . . . .	43

3.1.1	Multistage Optimization and Sample Average Approximation .	45
3.1.2	Related Work . . . . .	46
3.1.3	Contributions and Structure . . . . .	48
3.2	Approach . . . . .	49
3.2.1	Notation . . . . .	51
3.3	Asymptotic Optimality . . . . .	52
3.3.1	$k$ -Nearest Neighbor Weight Functions . . . . .	53
3.3.2	CART Weight Functions . . . . .	60
3.3.3	Random Forest Weight Functions . . . . .	68
3.4	Finite Sample Guarantees . . . . .	70
3.5	Computational Examples . . . . .	78
3.5.1	Multistage Inventory Control . . . . .	78
3.5.2	Multistage Lot Sizing . . . . .	81
3.6	Conclusion . . . . .	83
<b>4</b>	<b>Sample Robust Optimization with Covariates</b>	<b>85</b>
4.1	Introduction . . . . .	86
4.1.1	Contributions . . . . .	87
4.1.2	Comparison to Related Work . . . . .	89
4.2	Problem Setting . . . . .	91
4.2.1	Notation . . . . .	92
4.3	Sample Robust Optimization with Covariates . . . . .	93
4.3.1	Preliminary: Sample Robust Optimization . . . . .	93
4.3.2	Incorporating covariates into sample robust optimization . . .	94
4.4	Asymptotic Optimality . . . . .	96
4.4.1	Main result . . . . .	96
4.4.2	Review of the Wasserstein metric . . . . .	98
4.4.3	Concentration of the weighted empirical measure . . . . .	99
4.4.4	Proof of main result . . . . .	106
4.5	Tractable Approximations . . . . .	106



4.6	Computational Experiments . . . . .	109
4.6.1	Shipment planning . . . . .	110
4.6.2	Dynamic inventory management . . . . .	113
4.6.3	Portfolio optimization . . . . .	116
4.7	Conclusion . . . . .	118
<b>5</b>	<b>Data-Driven Optimization of Operating Room Blocks</b>	<b>119</b>
5.1	Introduction . . . . .	119
5.1.1	Related work . . . . .	122
5.1.2	Contributions and Structure . . . . .	124
5.2	Problem Formulation and Deterministic Approach . . . . .	125
5.3	Data-driven Approach . . . . .	129
5.3.1	Review of predictive-to-prescriptive analytics framework . . . . .	129
5.3.2	Data-driven OR scheduling formulation . . . . .	132
5.4	Application and Impact . . . . .	136
5.4.1	Results . . . . .	136
5.4.2	Extension for Targeted Growth . . . . .	138
5.5	Conclusion . . . . .	139
<b>6</b>	<b>Conclusion</b>	<b>141</b>
<b>A</b>	<b>Appendix for Chapter 2</b>	<b>143</b>
A.1	Proofs . . . . .	143
A.2	Optimization with Linear Predictive Models . . . . .	152
A.3	Data Generation . . . . .	153
A.3.1	Pricing . . . . .	153
A.3.2	Warfarin Dosing . . . . .	154
A.4	Sensitivity to Selection of Tuning Parameters . . . . .	155
<b>B</b>	<b>Appendix for Chapter 3</b>	<b>157</b>
B.1	Additional Results on Random Forest Weight Functions . . . . .	157
B.2	Decomposition Algorithm . . . . .	160

<b>C</b>	<b>Appendix for Chapter 4</b>	<b>167</b>
C.1	Properties of Weight Functions . . . . .	167
C.2	Proofs from Section 4.4.4 . . . . .	171
C.3	Technical Details for Section 4.5 . . . . .	177
C.3.1	Proof of Theorem 4.3 . . . . .	177
C.3.2	Compact representation of multi-policy approximation . . . . .	178
C.3.3	Proof of Theorem 4.4 . . . . .	180

# List of Figures

2-1	Regret of two methods as a function of the number of training samples, $n$ . PCM=Predicted cost minimization, UP-PCM=Uncertainty penalized predicted cost minimization. . . . .	27
2-2	Comparison of relevant methods on a pricing example and a Warfarin dosing example. . . . .	38
3-1	Out of sample results with various weight functions for a twelve stage inventory control problem. Vertical axis represents expected cost of policy (smaller is better). . . . .	80
3-2	Average out-of-sample cost of policies computed using various weight functions for twelve stage lot sizing problem. Horizontal axis shows number of training examples. . . . .	82
4-1	Out-of-sample profit for the shipment planning example. . . . .	112
4-2	Out-of-sample objective for the portfolio optimization example. . . . .	118
5-1	Average overnight census by day of the week. . . . .	121
5-2	Out-of-sample reduction in average maximum weekly census for each of the various models. . . . .	138
A-1	Effect of $\lambda$ tuning on the random forest method for Warfarin example	156



# List of Tables

3.1	Summary of notation. . . . .	52
4.1	Relationship of four methods. . . . .	109
4.2	Statistical significance for shipment planning problem. . . . .	112
4.3	Average out-of-sample cost for dynamic procurement problem. . . . .	114
4.4	Statistical significance for dynamic procurement problem. . . . .	115
4.5	Average computation time (seconds) for dynamic procurement problem.	115



# Chapter 1

## Introduction

Optimization under uncertainty forms the foundation for many of the the fundamental applications the operations research community seeks to solve. A retailer managing inventory levels, an investor allocating money, a hospital scheduling resources, and a physician prescribing treatments all face problems that share the same characteristics. Each aims to make a decision to optimize an objective function that depends on an uncertain quantity. The retailer faces uncertain future demand when deciding how much inventory to order to maximize profit. The investor faces uncertain future returns when trying to construct a portfolio with an optimal risk-reward tradeoff. The hospital faces unknown future demand for services when creating its operating room schedule. The physician faces uncertainty in how the patient will respond when prescribing a treatment to minimize adverse effects.

Formally, we model the uncertainty in each of these problems as a random variable with an unknown distribution. In practice, decision makers have access to data consisting of past realizations of the uncertainty, so we aim to learn to make near-optimal decisions from this data. In addition, decision makers often have access to *auxiliary covariates* that can help predict future uncertainty. For example, the retailer has data on market trends and social media interest that can help predict future demand, and the physician knows demographic and genetic factors that can help predict how a patient will respond to specific treatments. The covariates provide valuable information for the decision-maker, but it is not always clear how to efficiently use them. Many

effective machine learning methods exist for using covariates to make point predictions. However, learning to make good decisions presents a greater challenge because it often requires an understanding of the distribution of the uncertainty rather than just a point prediction. In this thesis, we develop and analyze data-driven algorithms that incorporate ideas from machine learning in order to solve optimization problems with auxiliary covariates.

## 1.1 Background

We now review some existing data-driven approaches to optimization problems with covariates. These basic problems are characterized by four components:

- Decision:  $z \in \mathcal{Z} \subset \mathbb{R}^{d_z}$ ,
- Uncertainty:  $Y \in \mathcal{Y} \subset \mathbb{R}^{d_y}$ ,
- Cost function:  $c : \mathcal{Z} \times \mathcal{Y} \rightarrow \mathbb{R}$ ,
- Auxiliary covariates:  $X \in \mathcal{X} \subset \mathbb{R}^{d_x}$ .

For a newly observed vector of auxiliary covariates,  $x \in \mathcal{X}$ , we want to find a decision that minimizes the conditional expected cost,

$$\min_{z \in \mathcal{Z}} \mathbb{E}[c(z; Y) \mid X = x]. \tag{1.1}$$

While we do not know the conditional distribution of  $Y$  given  $X$ , we do have access to data consisting of  $N$  observations,

$$(x^1, y^1), \dots, (x^N, y^N).$$

One approach to solving (1.1) is *sample average approximation* (SAA). In this approach, we replace the unknown distribution of  $Y$  with the empirical distribution of



the training data, resulting in the following problem:

$$\min_{z \in \mathcal{Z}} \frac{1}{N} \sum_{i=1}^N c(z; y^i).$$

Despite its simplicity SAA produces asymptotically optimal decisions for (1.1) if  $Y$  is independent of the covariates  $X$  (under mild technical conditions on  $\mathcal{Z}$  and  $c$ ). In other words, if the covariates provide no information on the uncertainty, then the optimal decisions for the SAA problem converge to optimal decisions of (1.1) as the number of training samples goes to infinity. For more details, see Shapiro et al. [89].

A key limitation of SAA is that it ignores the covariates, so when  $X$  provides information about  $Y$ , it may produce decisions that are asymptotically suboptimal for (1.1). Bertsimas and Kallus [15] propose a modified approach in which they replace the unknown distribution of  $Y$  by a weighted empirical distribution,

$$\min_{z \in \mathcal{Z}} \sum_{i=1}^N w_i^N(x) c(z; y^i), \tag{1.2}$$

where  $w^N : \mathcal{X} \rightarrow \mathbb{R}^N$  is a *weight function*. Many choices of weight functions exist, but, intuitively, for each  $i \in \{1, \dots, N\}$ ,  $w_i^N(x)$  should measure the similarity between the new covariate vector,  $x$ , and the covariate vector of the  $i$ th training example,  $x^i$ . Bertsimas and Kallus [15] list several examples, including the  $k$ -nearest neighbor weight function,

$$w_i^{N, \text{kNN}}(x) := \begin{cases} \frac{1}{k}, & x^i \text{ is a } k\text{-nearest neighbor of } x \text{ out of } \{x^1, \dots, x^N\}, \\ 0, & \text{otherwise,} \end{cases}$$

where  $k \in \mathbb{N}$ , and the kernel regression weight function,

$$w_i^{N, KR}(x) := K\left(\frac{\|x^i - x\|}{h}\right),$$

where  $h > 0$  is a bandwidth parameter and  $K(\cdot)$  is a kernel function (such as the Gaussian kernel,  $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ ). In addition, they introduce weight functions

inspired by classification and regression trees (CART) and random forests. We will discuss these further in later chapters.

All of the above-mentioned weight functions are inspired by nonparametric machine learning methods. These nonparametric methods have proven effective at prediction because they can learn complex relationships between the covariates and the response variable without requiring the practitioner to know an explicit parametric form. Similarly, Bertsimas and Kallus [15] show that, under appropriate conditions, optimal decisions for (1.2) with these weight functions are asymptotically optimal for (1.1), without any parametric assumptions on the relationship between  $X$  and  $Y$ . In other words, incorporating covariates via a nonparametric weight function in (1.2) can lead to asymptotically better decisions than SAA, even without specific knowledge of how the covariates affect the uncertainty.

## 1.2 Contributions

This thesis extends the approach in (1.2) along several dimensions. In particular, we consider *observational* optimization problems in which the decision,  $z$ , affects the observation of the uncertainty,  $Y$ . This occurs in several important applications, such as personalized medicine and pricing. We also consider *dynamic* optimization problems, which consist of a series of temporal stages. In each stage,  $t \in \{1, \dots, T\}$ , the decision maker chooses a decision  $z_t$  and then observes uncertainty  $Y_t$ . For example, consider the retailer managing its inventory of a product. Each week, the retailer observes the current inventory level and decides how much to order for the following week. Crucially, the retailer bases its decision on the remaining inventory level at that time, which depends on the realizations of demand in past weeks. Therefore, in dynamic problems, the decision in each stage is not a fixed quantity, but rather a function of the uncertainty revealed by that point in time. We develop and analyze approaches to these problems in the following chapters.

## Optimization over Continuous and Multi-dimensional Decisions with Observational Data

In Chapter 2, we consider observational optimization problems. The data in this case consists of tuples of covariates, decision, and uncertainty,  $(x^1, z^1, y^1), \dots, (x^N, z^N, y^N)$ . This setting introduces two challenges not present in the basic setting of (1.1). First, the data is incomplete. We do not observe what the realization of the uncertainty would have been under a different decision. Second, there may be confounding between  $z$  and  $y$  in the training data. We address these challenges with a method that not only predicts the cost of each potential decision, but also estimates the variance of those predictions. Accounting for the variance of the predicted cost adds significant value when the decision space,  $\mathcal{Z}$ , is continuous or multidimensional and there are many potential decisions. We establish this through both theoretical and computational results. As an example, we consider the problem of personalized Warfarin dosing, in which the decision,  $z$ , is the dose assigned to the patient, the uncertainty,  $Y$ , is the response of the patient, as measured by the international normalized ratio, and the covariates for each patient,  $X$ , include demographic information, medical history, and the genotype variant at certain sites. We compare our method against alternatives on real data and find that our approach obtains the best out-of-sample performance across all sizes of training sets.

## From Predictions to Prescriptions in Multistage Optimization Problems

In Chapter 3, we consider  $T$ -stage dynamic optimization problems with covariates. We assume that, in addition to the uncertainty,  $Y$ , evolving over time, the covariates,  $X$ , also evolve over time. The decision maker must take into account this additional available information when making decisions. We introduce a data-driven approach to these dynamic problems that uses machine learning to define a stochastic process with finite support that approximates the true stochastic process of the uncertainty,  $(Y_1, \dots, Y_T)$ . We can solve the resulting problem with techniques from dynamic pro-

gramming and approximate dynamic programming (see Bertsekas [11] for reference). We prove that the decisions produced by our approach are asymptotically optimal under mild, nonparametric conditions, and we establish finite sample guarantees on the optimality gap. We validate this theory with computational results on two inventory management examples.

## Sample Robust Optimization with Covariates

In Chapter 4, we again consider  $T$ -stage dynamic optimization problems with covariates. However, instead of approximating the stochastic process of the uncertainty, we introduce robustness to (1.2) and optimize over *decision rules*. A decision rule is a collection of functions that specify what decision to make in each stage based on the information revealed up until that point. For the example of the retailer, a decision rule for week  $t$  specifies an ordering quantity as a function of the observed demands from weeks  $1, \dots, t - 1$ . We establish that our approach produces asymptotically optimal decision rules. The proof of this relies on a novel result regarding the convergence in Wasserstein distance of the weighted empirical distribution introduced in (1.2) to the conditional distribution of the uncertainty given the covariates. We also demonstrate the tractability of the approach by applying it to several computational examples with up to twelve stages.

## Data-Driven Optimization of Operating Room Blocks

In Chapter 5, we use the techniques developed in this thesis to optimize the operating room block schedule of a major US hospital. Our partner institution, like many other hospitals, faces significant census variability throughout the week, which limits the amount of patients it can accept due to resource constraints during peak times. To optimize the block schedule to minimize census peaks, we develop a mixed integer optimization approach that employs predictive machine learning to estimate the distributions of surgical patients' lengths of stay. In addition, by using ideas from Chapter 2, our approach accounts for the impact that moving the date of surgery has

on patients' lengths of stay. We demonstrate the effectiveness of the approach with simulation results using real data and find that we can reliably reduce the maximal weekly census with only a few changes to the schedule.



# Chapter 2

## Optimization over Continuous and Multi-Dimensional Decisions with Observational Data

In this chapter, we consider the optimization of an uncertain objective over continuous and multi-dimensional decision spaces in problems in which we are only provided with observational data. We propose a novel algorithmic framework that is tractable, asymptotically consistent, and superior to comparable methods on example problems. Our approach leverages predictive machine learning methods and incorporates information on the uncertainty of the predicted outcomes for the purpose of prescribing decisions. We demonstrate the efficacy of our method on examples involving both synthetic and real data sets.

### 2.1 Introduction

We study the general problem in which a decision maker seeks to optimize a known objective function that depends on an uncertain quantity. The uncertain quantity has an unknown distribution, which may be affected by the action chosen by the decision maker. Many important problems across a variety of fields fit into this framework. In healthcare, for example, a doctor aims to prescribe drugs in specific dosages to

regulate a patient’s vital signs. In revenue management, a store owner must decide how to price various products in order to maximize profit. In online retail, companies decide which products to display for a user to maximize sales. The general problem we study is characterized by the following components:

- Decision variable:  $z \in \mathcal{Z} \subset \mathbb{R}^p$ ,
- Outcome:  $Y(z) \in \mathcal{Y}$  (We adopt the potential outcomes framework [83], in which  $Y(z)$  denotes the (random) quantity that would have been observed had decision  $z$  been chosen.),
- Auxiliary covariates (also called side-information or context):  $x \in \mathcal{X} \subset \mathbb{R}^d$ ,
- Cost function:  $c(z; y) : \mathcal{Z} \times \mathcal{Y} \rightarrow \mathbb{R}$ . (This function is known *a priori*.)

We allow the auxiliary covariates, decision variable, and outcome to take values on multi-dimensional, continuous sets. A decision-maker seeks to determine the action that minimizes the conditional expected cost:

$$\min_{z \in \mathcal{Z}} \mathbb{E}[c(z; Y(z)) | X = x]. \quad (2.1)$$

Of course, the distribution of  $Y(z)$  is unknown, so it is not possible to solve this problem exactly. However, we assume that we have access to *observational data*, consisting of  $n$  independent and identically distributed observations,  $(X_i, Z_i, Y_i)$  for  $i = 1, \dots, n$ . Each of these observations consists of an auxiliary covariate vector, a decision, and an observed outcome. This type of data presents two challenges that differentiate our problem from a predictive machine learning problem. First, it is incomplete. We only observe  $Y_i := Y_i(Z_i)$ , the outcome associated with the applied decision. We do not observe what the outcome would have been under a different decision. Second, the decisions were not necessarily chosen independently of the outcomes, as they would have been in a randomized experiment, and we do not know how the decisions were assigned. Following common practice in the causal inference literature, we make the ignorability assumption of Hirano and Imbens [60].



**Assumption 2.1** (Ignorability).

$$Y(z) \perp\!\!\!\perp Z \mid X \quad \forall z \in \mathcal{Z}$$

In other words, we assume that historically the decision  $Z$  has been chosen as a function of the auxiliary covariates  $X$ . There were no unmeasured confounding variables that affected both the choice of decision and the outcome.

Under this assumption, we are able to rewrite the objective of (2.1) as

$$\mathbb{E}[c(z; Y) \mid X = x, Z = z].$$

This form of the objective is easier to learn because it depends only on the observed outcome, not on the counterfactual outcomes. A direct approach to solve this problem is to use a regression method to predict the cost as a function of  $x$  and  $z$  and then choose  $z$  to minimize this predicted cost. If the selected regression method is uniformly consistent in  $z$ , then the action chosen by this method will be asymptotically optimal under certain conditions. (We will formalize this later.) However, this requires choosing a regression method that ensures the optimization problem is tractable. For this work, we restrict our attention to linear and tree-based methods, such as CART [30] and random forests [29], as they are both effective and tractable for many practical problems.

A key issue with the direct approach is that it tries to learn too much. It tries to learn the expected outcome under every possible decision, and the level of uncertainty associated with the predicted expected cost can vary between different decisions. This method can lead us to select a decision which has a small point estimate of the cost, but a large uncertainty interval.

### 2.1.1 Notation

Throughout this chapter, we use capital letters to refer to random quantities and lower case letters to refer to deterministic quantities. Thus, we use  $Z$  to refer to the

decision randomly assigned by the (unknown) historical policy and  $z$  to refer to a specific action. For a given, auxiliary covariate vector,  $x$ , and a proposed decision,  $z$ , the conditional expectation  $\mathbb{E}[c(z; Y)|X = x, Z = z]$  means the expectation of the cost function  $c(z; Y)$  under the conditional measure in which  $X$  is fixed as  $x$  and  $Z$  is fixed as  $z$ . We ignore details of measurability throughout and assume this conditional expectation is well defined. Throughout, all norms are  $\ell_2$  norms unless otherwise specified. We use  $(X, Z)$  to denote vector concatenation.

### 2.1.2 A Motivating Example for Uncertainty Penalization

We start with a simple example in which there are  $m$  possible decisions. We have data on  $n$  i.i.d. observations of past decisions and outcomes,  $(Z_i, Y_i)$  for  $i = 1, \dots, n$ . For simplicity, there are no auxiliary covariates for this problem, and the decision,  $Z_i$ , has been chosen independently of the potential outcomes,  $(Y_i(1), \dots, Y_i(m))$  (so Assumption 2.1 holds).

The goal is to choose  $z$  to minimize the expected outcome. If  $\mu_z = \mathbb{E}[Y(z)]$  is known for all  $z$ , the optimal decision is given by:  $z^* \in \arg \min_z \mu_z$ . However, this mean is unknown, so it must be estimated. We define  $\hat{\mu}_z = \frac{1}{\sum_i 1\{Z_i = z\}} \sum_i 1\{Z_i = z\} Y_i$ , the empirical expectation of the outcome under  $z$ . Predicted cost minimization (PCM) minimizes  $\hat{\mu}_z$ .

To motivate uncertainty penalization, we note that if we assume the rewards are almost surely bounded,  $|Y| \leq 1$ , an application of Bernstein's inequality shows that with probability at least  $1 - \delta$ ,

$$\mu_z \leq \hat{\mu}_z + \sqrt{\frac{2\sigma_z^2 \ln(m/\delta)}{\sum_i 1\{Z_i = z\}}} + \frac{4 \ln(m/\delta)}{3 \sum_i 1\{Z_i = z\}}, \quad \forall z,$$

where  $\sigma_z^2 = \text{Var}(Y_i(z))$ . This bound motivates the addition of a term to the objective that penalizes the uncertainty of the estimate  $\hat{\mu}_z$ . In this simple example,  $\hat{\mu}_z$  is an unbiased estimator for  $\mu_z$ , with  $\text{Var}(\hat{\mu}_z) = \mathbb{E}[\sigma_z^2 / \sum_i 1\{Z_i = z\}]$ . The penalty term above has exactly the form of the square root of this variance. We define the

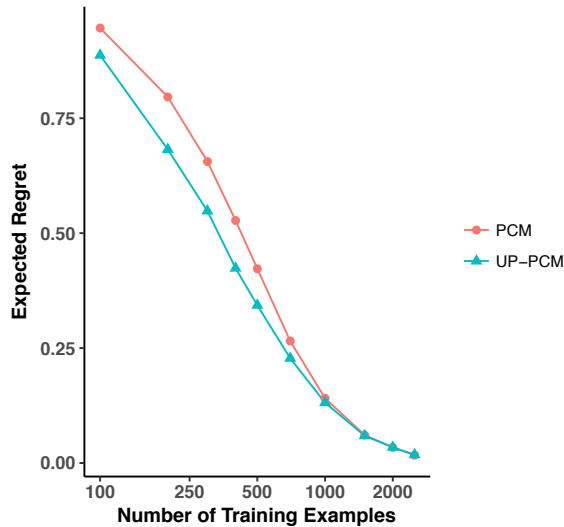


Figure 2-1: Regret of two methods as a function of the number of training samples,  $n$ . PCM=Predicted cost minimization, UP-PCM=Uncertainty penalized predicted cost minimization.

uncertainty penalized objective:

$$\hat{\mu}_z^\lambda = \hat{\mu}_z + \lambda \sqrt{\frac{1}{\sum_i 1\{Z_i = z\}}},$$

where  $\lambda$  is a tuning parameter. We note that we could also include the action-dependent variance  $\sigma_z^2$  in the penalty term. However, this is typically unknown, and we have observed that estimating it without a homoscedasticity assumption can introduce too much noise for the objective to be effective.

For this experiment, we fix  $m = 100$  possible actions, each with a fixed mean response, which was drawn from a standard Gaussian distribution. To construct the training data, at each time step, a decision was chosen uniformly at random from the set of possible decisions and the outcome is the mean response plus noise sampled from a standard Gaussian distribution. The regret of each method was computed and averaged over ten thousand trials for a range of values of  $n$ . We held  $\lambda$  fixed at 1. Our results are shown in Figure 2-1. Uncertainty penalized empirical risk minimization clearly outperforms predicted cost minimization for small training

sample sizes. However, the level of outperformance decreases as  $n$  increases.

This example serves to motivate the need for uncertainty penalization, especially when training data is limited. The direct approach of choosing the action with the smallest predicted cost is inefficient because the predicted costs can have different levels of uncertainty associated with them. Figure 2-1 demonstrates that we are better off choosing a decision whose cost we can predict with high confidence, rather than decisions whose costs we know little about.

### 2.1.3 Related Work

Recent years have seen tremendous interest in the area of data-driven optimization. Much of this work combines ideas from the statistics and machine learning literature with techniques from mathematical optimization. Bertsimas and Kallus [15] developed a framework that uses nonparametric machine learning methods to solve data-driven optimization problems in the presence of auxiliary covariates. They take advantage of the fact that for many machine learning algorithms, the predictions are given by a linear combination of the training samples' target variables. Kao et al. [67] and Elmachtoub and Grigas [47] developed algorithms that make predictions tailored for use in specific optimization problems. However, they all deal with the setting in which the decision does not affect the outcome. This is insufficient for many applications, such as pricing, in which the demand for a product is clearly affected by the price. Bertsimas and Kallus [16] later studied the limitations of predictive approaches to pricing problems. In particular, they demonstrated that confounding in the data between the decision and outcome can lead to large optimality gaps if ignored. They proposed a kernel-based method for data-driven optimization in this setting, but it does not scale well with the dimension of the decision space. Misic [75] developed an efficient mixed integer optimization formulation for problems in which the predicted cost is given by a tree ensemble model. This approach scales fairly well with the dimension of the decision space but does not consider the need for uncertainty penalization.

Another relevant area of research is causal inference (see Rosenbaum [83] for an

overview), which concerns the study of causal effects from observational data. Much of the work in this area has focused on determining whether a treatment has a significant effect on the population as a whole. However, a growing body of work has focused on learning optimal, personalized treatments from observational data. Athey and Wager [2] proposed an algorithm that achieves optimal (up to a constant factor) regret bounds in learning a treatment policy when there are two potential treatments. Kallus [63] proposed an algorithm to efficiently learn a treatment policy when there is a finite set of potential treatments. Bertsimas et al. [21] developed a tree-based algorithm that learns to personalize treatment assignments from observational data. It is based on the optimal trees machine learning method [13] and has performed well in experiments. Considerably less attention has been paid to problems with a continuous decision space. Hirano and Imbens [60] introduced the problem of inference with a continuous treatment, and Flores [50] studied the problem of learning an optimal policy in this setting. Recently, Kallus and Zhou [65] developed an approach to policy learning with a continuous decision variable that generalizes the idea of inverse propensity score weighting. Our approach differs in that we focus on regression-based methods, which we believe scale better with the dimension of the decision space and avoid the need for density estimation.

The idea of uncertainty penalization has been explored as an alternative to empirical risk minimization in statistical learning, starting with Maurer and Pontil [73]. Swaminathan and Joachims [91] applied uncertainty penalization to the offline bandit setting. Their setting is similar to the one we study. An agent seeks to minimize the prediction error of his/her decision, but only observes the loss associated with the selected decision. They assumed that the policy used in the training data is known, which allowed them to use inverse propensity weighting methods. In contrast, we assume ignorability, but not knowledge of the historical policy, and we allow for more complex decision spaces.

We note that our approach bears a superficial resemblance to the upper confidence bound (UCB) algorithms for multi-armed bandits (cf. Bubeck et al. [31]). These algorithms choose the action with the highest upper confidence bound on its predicted

expected reward. Our approach, in contrast, chooses the action with the highest lower confidence bound on its predicted expected reward (or lowest upper confidence bound on predicted expected cost). The difference is that UCB algorithms choose actions with high upside to balance exploration and exploitation in the online bandit setting, whereas we work in the offline setting with a focus on solely exploitation.

### 2.1.4 Contributions

Our primary contribution is an algorithmic framework for observational data driven optimization that allows the decision variable to take values on continuous and multidimensional sets. We consider applications in personalized medicine, in which the decision is the dose of Warfarin to prescribe to a patient, and in pricing, in which the action is the list of prices for several products in a store.

## 2.2 Approach

In this section, we introduce the uncertainty penalization approach for optimization with observational data. Recall that the observational data consists of  $n$  i.i.d. observations,  $(X_1, Z_1, Y_1), \dots, (X_n, Z_n, Y_n)$ . For observation  $i$ ,  $X_i$  represents the pertinent auxiliary covariates,  $Z_i$  is the decision that was applied, and  $Y_i$  is the observed response. The first step of the approach is to train a predictive machine learning model to estimate  $\mathbb{E}[c(z; Y)|X = x, Z = z]$ . When training the predictive model, the feature space is the cartesian product of the auxiliary covariate space and the decision space,  $\mathcal{X} \times \mathcal{Z}$ . We have several options for how to train the predictive model. We can train the model to predict  $Y$ , the cost  $c(Z, Y)$ , or a combination of these two responses. In general, we denote the prediction of the ML algorithm as a linear combination of the cost function evaluated at the training examples,

$$\hat{\mu}(x, z) := \sum_{i=1}^n w_i(x, z)c(z; Y_i).$$

We require the predictive model to satisfy a generalization of the honesty property of Wager and Athey [97].

**Assumption 2.2** (Honesty). *The model trained on  $(X_1, Z_1, Y_1), \dots, (X_n, Z_n, Y_n)$  is honest, i.e., the weights,  $w_i(x, z)$ , are determined independently of the outcomes,  $Y_1, \dots, Y_n$ .*

This honesty assumption reduces the bias of the predictions of the cost. We also enforce several restrictions on the weight functions.

**Assumption 2.3** (Weights). *For all  $(x, z) \in \mathcal{X} \times \mathcal{Z}$ ,  $\sum_{i=1}^n w_i(x, z) = 1$  and for all  $i$ ,  $w_i(x, z) \in [0, 1/\gamma_n]$ . In addition,  $\mathcal{X} \times \mathcal{Z}$  can be partitioned into  $\Gamma_n$  regions such that if  $(x, z)$  and  $(x, z')$  are in the same region,  $\|w(x, z) - w(x, z')\|_1 \leq \alpha \|z - z'\|_2$ .*

The direct approach to solving (2.1) amounts to choosing  $z \in \mathcal{Z}$  that minimizes  $\hat{\mu}(x, z)$ , for each new instance of auxiliary covariates,  $x$ . However, the variance of the predicted cost,  $\hat{\mu}(x, z)$ , can vary with the decision variable,  $z$ . Especially with a small training sample size, the direct approach, minimizing  $\hat{\mu}(x, z)$ , can give a decision with a small, but highly uncertain, predicted cost. We can reduce the expected regret of our action by adding a penalty term for the variance of the selected decision. If Assumption 2.2 holds, the conditional variance of  $\hat{\mu}(x, z)$  given  $(X_1, Z_1), \dots, (X_n, Z_n)$  is given by

$$V(x, z) := \sum_i w_i^2(x, z) \text{Var}(c(z; Y_i) | X_i, Z_i).$$

In addition,  $\hat{\mu}(x, z)$  may not be an unbiased predictor, so we also introduce a term that penalizes the conditional bias of the predicted cost given  $(X_1, Z_1), \dots, (X_n, Z_n)$ . Since the true cost is unknown, it is not possible to exactly compute this bias. Instead, we compute an upper bound under a Lipschitz assumption (details in Section 2.3).

$$B(x, z) := \sum_i w_i(x, z) \|(X_i, Z_i) - (x, z)\|_2.$$

Overall, given a new vector of auxiliary covariates,  $x \in \mathcal{X}$ , our approach makes a

decision by solving

$$\min_{z \in \mathcal{Z}} \hat{\mu}(x, z) + \lambda_1 \sqrt{V(x, z)} + \lambda_2 B(x, z), \quad (2.2)$$

where  $\lambda_1$  and  $\lambda_2$  are tuning parameters.

As a concrete example, we can use the CART algorithm of Breiman et al. [30] or the optimal regression tree algorithm of Bertsimas and Dunn [13] as the predictive method. These algorithms work by partitioning the training examples into clusters, i.e., the leaves of the tree. For a new observation, a prediction of the response variable is made by averaging the responses of the training examples that are contained in the same leaf.

$$w_i(x, z) = \begin{cases} \frac{1}{N(x, z)}, & (x, z) \in l(x, z), \\ 0, & \text{otherwise,} \end{cases}$$

where  $l(x, z)$  denotes the set of training examples that are contained in the same leaf of the tree as  $(x, z)$ , and  $N(x, z) = |l(x, z)|$ . The variance term will be small when the leaf has a large number of training examples, and the bias term will be small when the diameter of the leaf is small. Assumption 2.2 can be satisfied by ignoring the outcomes when selecting the splits or by dividing the training data into two sets, one for making splits and one for making predictions. Assumption 2.3 is satisfied with  $\alpha = 0$  if the minimum number of training samples in each leaf is  $\gamma_n$  and the maximum number of leaves in the tree is  $\Gamma_n$ .

### 2.2.1 Parameter Tuning

Before proceeding, we note that the variance terms,  $\text{Var}(c(z; Y_i) \mid X_i, Z_i)$ , are often unknown in practice. In the absence of further knowledge, we assume homoscedasticity, i.e.,  $\text{Var}(Y_i \mid X_i, Z_i)$  is constant. It is possible to estimate this value by training a machine learning model to predict  $Y_i$  as a function of  $(X_i, Z_i)$  and computing the mean squared error on the training set. However, it may be advantageous to include this value with the tuning parameter  $\lambda_1$ .

We have several options for tuning parameters  $\lambda_1$  and  $\lambda_2$  (and whatever other



parameters are associated with the predictive model). Because the counterfactual outcomes are unknown, it is not possible to use the standard approach of holding out a validation set during training and evaluating the error of the model on that validation set for each combination of possible parameters. One option is to tune the predictive model’s parameters using cross validation to maximize predictive accuracy and then select  $\lambda_1$  and  $\lambda_2$  using the theory we present in Section 2.3. Another option is to split the data into a training and validation set and train a predictive model on the validation data to impute the counterfactual outcomes. We then select the model that minimizes the predicted cost on the validation set. For the examples in Section 2.4, we use a combination of these two ideas. We train a random forest model on the validation set (in order to impute counterfactual outcomes), and we then select the model that minimizes the sum of the mean squared error and the predicted cost on the validation data. In Appendix A.4, we include computations that demonstrate, for the Warfarin example of Section 2.4.2, the method is not too sensitive to the choice of  $\lambda_1$  and  $\lambda_2$ .

## 2.3 Theory

In this section, we describe the theoretical motivation for our approach and provide finite-sample generalization and regret bounds. For notational convenience, we define

$$\mu(x, z) := \mathbb{E}[c(z; Y(z)) | X = x] = \mathbb{E}[c(z; Y) | X = x, Z = z],$$

where the second equality follows from the ignorability assumption. Before presenting the results, we first present a few additional assumptions.

**Assumption 2.4** (Regularity). *The set  $\mathcal{X} \times \mathcal{Z}$  is nonempty, closed, and bounded with diameter  $D$ .*

**Assumption 2.5** (Objective Conditions). *The objective function satisfies the following properties:*

1.  $|c(z; y)| \leq 1 \quad \forall z, y.$

2. For all  $y \in \mathcal{Y}$ ,  $c(\cdot; y)$  is  $L$ -Lipschitz.

3. For any  $x, x' \in \mathcal{X}$  and any  $z, z' \in \mathcal{Z}$ ,  $|\mu(x, z) - \mu(x', z')| \leq L\|(x, z) - (x', z')\|$ .

These assumptions provide some conditions under which the generalization and regret bounds hold, but similar results hold under alternative sets of assumptions (e.g. if  $c(z; Y)|Z$  is subexponential instead of bounded). With these additional assumptions, we have the following generalization bound. All proofs are contained in Appendix A.1.

**Theorem 2.1.** *Suppose assumptions 2.1-2.5 hold. Then, with probability at least  $1 - \delta$ ,*

$$\mu(x, z) - \hat{\mu}(x, z) \leq \frac{4}{3\gamma_n} \ln(K_n/\delta) + 2\sqrt{V(x, z) \ln(K_n/\delta)} + L \cdot B(x, z) \quad \forall z \in \mathcal{Z},$$

where  $K_n = \Gamma_n (9D\gamma_n (\alpha(LD + 1 + \sqrt{2}) + L(\sqrt{2} + 3)))^p$ .

This result uniformly bounds, with high probability, the true cost of action  $z$  by the predicted cost,  $\hat{\mu}(x, z)$ , a term depending on the uncertainty of that predicted cost,  $V(x, z)$ , and a term proportional to the bias associated with that predicted cost,  $B(x, z)$ . It is easy to see how this result motivates the approach described in (2.2). One can also verify that the generalization bound still holds if  $(X_1, Z_1), \dots, (X_n, Z_n)$  are chosen deterministically, as long as  $Y_1, \dots, Y_n$  are still independent. Using Theorem 2.1, we are able to derive a finite-sample regret bound.

**Theorem 2.2.** *Suppose assumptions 2.1-2.5 hold. Define*

$$z^* \in \arg \min_z \mu(x, z),$$

$$\hat{z} \in \arg \min_z \hat{\mu}(x, z) + \lambda_1 \sqrt{V(x, z)} + \lambda_2 B(x, z).$$

If  $\lambda_1 = 2\sqrt{\ln(2K_n/\delta)}$  and  $\lambda_2 = L$ , then with probability at least  $1 - \delta$ ,

$$\mu(x, \hat{z}) - \mu(x, z^*) \leq \frac{2}{\gamma_n} \ln(2K_n/\delta) + 4\sqrt{V(x, z^*) \ln(2K_n/\delta)} + 2L \cdot B(x, z^*),$$

where  $K_n = \Gamma_n (9D\gamma_n (\alpha(LD + 1 + \sqrt{2}) + L(\sqrt{2} + 3)))^p$ .

By this result, the regret of the approach defined in (2.2) depends only on the variance and bias terms of the optimal action,  $z^*$ . Because the predicted cost is penalized by  $V(x, z)$  and  $B(x, z)$ , it does not matter how poor the prediction of cost is at suboptimal actions. Theorem 2.2 immediately implies the following asymptotic result, assuming the auxiliary feature space and decision space are fixed as the training sample size grows to infinity.

**Corollary 2.1.** *In the setting of Theorem 2.2, if  $\gamma_n = \Omega(n^\beta)$  for some  $\beta > 0$ ,  $\Gamma_n = O(n)$ , and  $B(x, z^*) \rightarrow_p 0$  as  $n \rightarrow \infty$ , then*

$$\mu(x, \hat{z}) \rightarrow_p \mu(x, z^*)$$

as  $n \rightarrow \infty$ .

The assumptions can be satisfied, for example, with CART or random forest as the learning algorithm with parameters set in accordance with Lemma 2 of Wager and Athey [97]. This next example demonstrates that there exist problems for which the regret of the uncertainty penalized method is strictly better, asymptotically, than the regret of predicted cost minimization.

**Example 2.1.** *Suppose there are  $m + 1$  different actions and two possible, equally probable states of the world. In one state, action 0 has a cost that is deterministically 1, and all other actions have a random cost that is drawn from  $\mathcal{N}(0, 1)$  distribution. In the other state, action 0 has a cost that is deterministically 0, and all other actions have a random cost, drawn from a  $\mathcal{N}(1, 1)$  distribution. Suppose the training data consists of  $m$  trials of each action. If  $\hat{\mu}(j)$  is the empirical average cost of action  $j$ , then the predicted cost minimization algorithm selects the action that minimizes  $\hat{\mu}(j)$ . The uncertainty penalization algorithm adds a penalty of the form suggested by Theorem 2.2,  $\lambda \sqrt{\frac{\sigma_j^2 \ln m}{m}}$ . If  $\lambda \geq \sqrt{2}$ , the (Bayesian) expected regret of the uncertainty penalization algorithm is asymptotically strictly less than the expected regret of the*

predicted cost minimization algorithm,  $\mathbb{E}R^{UP} = o(\mathbb{E}R^{PCM})$ , where the expectations are taken over both the training data and the unknown state of the world.

This example is simple but demonstrates that there exist settings in which predicted cost minimization is asymptotically suboptimal to the method we have described. In addition, the proof illustrates how one can construct tighter regret bounds than the one in Theorem 2.2 for problems with specific structure.

### 2.3.1 Tractability

The tractability of (2.2) depends on the algorithm that is used as the predictive model. For many kernel-based methods, the resulting optimization problems are highly nonlinear and do not scale well when the dimension of the decision space is more than 2 or 3. For this reason, we advocate using tree-based and linear models as the predictive model. Tree based models partition the space  $\mathcal{X} \times \mathcal{Z}$  into  $\Gamma_n$  leaves, so there are only  $\Gamma_n$  possible values of  $w(x, z)$ . Therefore, we can solve (2.2) separately for each leaf. For  $j = 1, \dots, \Gamma_n$ , we solve

$$\begin{aligned} \min \quad & \hat{\mu}(x, z) + \lambda_1 \sqrt{V(x, z)} + \lambda_2 B(x, z) \\ \text{s.t.} \quad & z \in \mathcal{Z} \\ & (x, z) \in L_j, \end{aligned} \tag{2.3}$$

where  $L_j$  denotes the subset of  $\mathcal{X} \times \mathcal{Z}$  that makes up leaf  $j$  of the tree. Because each split in the tree is a hyperplane,  $L_j$  is defined by an intersection of hyperplanes and thus is a polyhedral set. Clearly,  $B(x, z)$  is a convex function in  $z$ , as it is a nonnegative linear combination of convex functions. If we assume homoscedasticity, then  $V(x, z)$  is constant for all  $(x, z) \in L_j$ . If  $c(z; y)$  is convex in  $z$  and  $\mathcal{Z}$  is a convex set, (2.3) is a convex optimization problem and can be solved by convex optimization techniques. Furthermore, since the  $\Gamma_n$  instances of (2.3) are all independent, we can solve them in parallel. Once (2.3) has been solved for all leaves, we select the solution from the leaf with the overall minimal objective value.

For tree ensemble methods, such as random forests [29] or xgboost [34], opti-

mization is more difficult. We compute optimal decisions using a coordinate descent heuristic. From a random starting action, we cycle through holding all decision variables fixed except for one and optimize that decision using discretization. We repeat this until convergence from several different random starting decisions. For linear predictive models, the resulting problem is often a second order conic optimization problem, which can be handled by off-the-shelf solvers (details given in Appendix A.2).

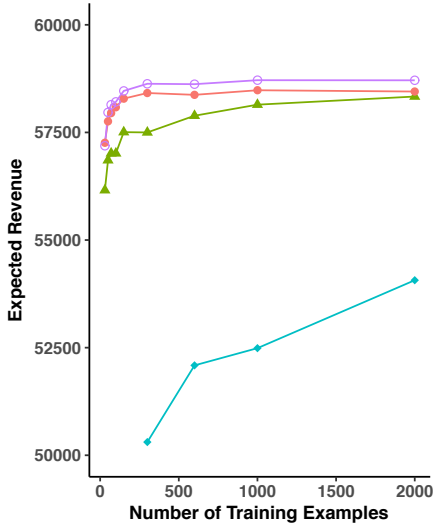
## 2.4 Results

In this section, we demonstrate the effectiveness of our approach with two examples. In the first, we consider pricing problem with synthetic data, while in the second, we use real patient data for personalized Warfarin dosing.

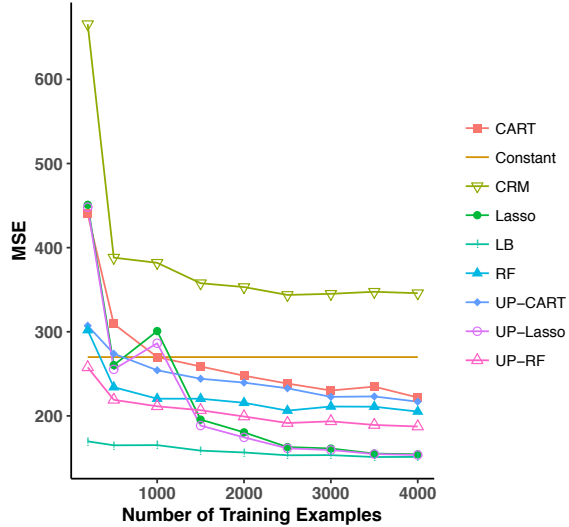
### 2.4.1 Pricing

In this example, the decision variable,  $z \in \mathbb{R}^5$ , is a vector of prices for a collection of products. The outcome,  $Y$ , is a vector of demands for those products. The auxiliary covariates may contain data on the weather and other exogenous factors that may affect demand. The objective is to select prices to maximize revenue for a given vector of auxiliary covariates. The demand for a single product is affected by the auxiliary covariates, the price of that product, and the price of one or more of the other products, but the mapping is unknown to the algorithm. The details on the data generation process can be found in Appendix A.3.

In Figure 2-2a, we compare the expected revenues of the strategies produced by several algorithms. CART, RF, and Lasso refer to the direct methods of training, respectively, a decision tree, a random forest, and a lasso regression [93] to predict revenue, as a function of the auxiliary covariates and prices, and choosing prices, for each vector of auxiliary covariates in the test set, that maximize predicted revenue. (Note that the revenues for CART and Lasso were too small to be displayed on the plot. Unsurprisingly, the linear model performs poorly because revenue does not vary linearly with price. We restrict all prices to be at most 50 to ensure the optimization



(a) Pricing example.



(b) Warfarin example.

Figure 2-2: Comparison of relevant methods on a pricing example and a Warfarin dosing example.

problems are bounded.) UP-CART, UP-RF, and UP-Lasso refer to the uncertainty penalized analogues in which the variance and bias terms are included in the objective. For each training sample size,  $n$ , we average our results over one hundred separate training sets of size  $n$ . At a training size of 2000, the uncertainty penalized random forest method improves expected revenue by an average of \$270 compared to the direct RF method. This improvement is statistically significant at the 0.05 significance level by the Wilcoxon signed-rank test ( $p$ -value  $4.4 \times 10^{-18}$ , testing the null hypothesis that mean improvement is 0 across 100 different training sets).

## 2.4.2 Warfarin Dosing

Warfarin is a commonly prescribed anticoagulant that is used to treat patients who have had blood clots or who have a high risk of stroke. Determining the optimal maintenance dose of Warfarin presents a challenge as the appropriate dose varies sig-

nificantly from patient to patient and is potentially affected by many factors including age, gender, weight, health history, and genetics. However, this is a crucial task because a dose that is too low or too high can put the patient at risk for clotting or bleeding. The effect of a Warfarin dose on a patient is measured by the International Normalized Ratio (INR). Physicians typically aim for patients to have an INR in a target range of 2-3.

In this example, we test the efficacy of our approach in learning optimal Warfarin dosing with data from Consortium et al. [41]. This publicly available data set contains the optimal stable dose, found by experimentation, for a diverse set of 5410 patients. In addition, the data set contains a variety of covariates for each patient, including demographic information, reason for treatment, medical history, current medications, and the genotype variant at CYP2C9 and VKORC1. It is unique because it contains the optimal dose for each patient, permitting the use of off-the-shelf machine learning methods to predict this optimal dose as a function of patient covariates. We instead use this data to construct a problem with observational data, which resembles the common problem practitioners face. Our access to the true optimal dose for each patient allows us to evaluate the performance of our method out-of-sample. This is a commonly used technique, and the resulting data set is sometimes called *semi-synthetic*. Several researchers have used the Warfarin data for developing personalized approaches to medical treatments. In particular, Kallus [64] and Bertsimas et al. [21] tested algorithms that learned to treat patients from semi-synthetic observational data. However, they both discretized the dosage into three categories, whereas we treat the dosage as a continuous decision variable.

To begin, we split the data into a training set of 4000 patients and a test set of 1410 patients. We keep this split fixed throughout all of our experiments to prevent cheating by using insights gained by visualization and exploration on the training set. Similar to Kallus [64], we assume physicians prescribe Warfarin as a function of BMI. We assume the response that the physicians observe is related to the difference between the dose a patient was given and the true optimal dose for that patient. It is a noisy observation, but it, on average, gives directional information (whether

the dose was too high or too low) and information on the magnitude of the distance from the optimal dose. The precise details of how we generate the data are given in Appendix A.3. For all methods, we repeat our work across 100 randomizations of assigned training doses and responses. To measure the performance of our methods, we compute, on the test set, the mean squared error (MSE) of the prescribed doses relative to the true optimal doses. Using the notation described in Section 2.1,  $X_i \in \mathbb{R}^{99}$  represents the auxiliary covariates for patient  $i$ . We work in normalized units so the covariates all contribute equally to the bias penalty term.  $Z_i \in \mathbb{R}$  represents the assigned dose for patient  $i$ , and  $Y_i \in \mathbb{R}$  represents the observed response for patient  $i$ . The objective in this problem is to minimize  $(\mathbb{E}[Y(z)|X = x])^2$  with respect to the dose,  $z$ .<sup>1</sup>

Figure 2-2b displays the results of several algorithms as a function of the number of training examples. We compare CART, without any penalization, to CART with uncertainty penalization (UP-CART), and we see that uncertainty penalization offers a consistent improvement. This improvement is greatest when the training sample size is smallest. (Note: for CART with no penalization, when multiple doses give the same optimal predicted response, we select the mean.) Similarly, when we compare the random forest and Lasso methods with their uncertainty-penalizing analogues, we again see consistent improvements in MSE. The “Constant” line in the plot measures the performance of a baseline heuristic that assigns a fixed dose of 35 mg/week to all patients. The “LB” line provides an unattainable lower bound on the performance of all methods that use the observational data. For this method, we train a random forest to predict the optimal dose as a function of the patient covariates. We also compare our methods with the Counterfactual Risk Minimization (CRM) method of Swaminathan and Joachims [91]. We allow their method access to the true propensity scores that generated the data and optimize over all regularized linear policies for which the proposed dose is a linear function of the auxiliary covariates. We tried

---

<sup>1</sup>This objective differs slightly from the setting described in Section 2.3 in which the objective was to minimize the conditional expectation of a cost function. However, it is straightforward to modify the results to obtain the same regret bound (save a few constant factors) when minimizing  $g(\mathbb{E}[c(z; Y(z))|X = x])$  for a Lipschitz function,  $g$ .



multiple combinations of tuning parameters, but the method always performed poorly out-of-sample. We suspect this is due to the size of the policy space. Our lasso based method works best on this data set when the number of training samples is large, but the random forest based method is best for smaller sample sizes. With the maximal training set size of 4000, the improvements of the CART, random forest, and lasso uncertainty penalized methods over their unpenalized analogues (2.2%, 8.6%, 0.5% respectively) are all statistically significant at the 0.05 family-wise error rate level by the Wilcoxon signed-rank test with Bonferroni correction (adjusted  $p$ -values  $2.1 \times 10^{-4}$ ,  $4.3 \times 10^{-16}$ ,  $1.2 \times 10^{-6}$  respectively).

## 2.5 Conclusions

In this chapter, we introduced a data-driven framework that combines ideas from predictive machine learning and causal inference to optimize an uncertain objective using observational data. Unlike most existing algorithms, our approach handles continuous and multi-dimensional decision variables by introducing terms that penalize the uncertainty associated with the predicted costs. We proved finite sample generalization and regret bounds and provided a sufficient set of conditions under which the resulting decisions are asymptotically optimal. We demonstrated, both theoretically and with real-world examples, the tractability of the approach and the benefit of the approach over unpenalized predicted cost minimization.



# Chapter 3

## From Predictions to Prescriptions in Multistage Optimization Problems

In this chapter, we introduce a framework for solving finite-horizon multistage optimization problems under uncertainty in the presence of auxiliary data. We assume the joint distribution of the uncertain quantities is unknown, but noisy observations, along with observations of auxiliary covariates, are available. We utilize effective predictive methods from machine learning (ML), including  $k$ -nearest neighbors regression ( $k$ NN), classification and regression trees (CART), and random forests (RF), to develop specific methods that are applicable to a wide variety of problems. We demonstrate that our solution methods are asymptotically optimal under mild conditions. Additionally, we establish finite sample guarantees for the optimality of our method with  $k$ NN weight functions. Finally, we demonstrate the practicality of our approach with computational examples. We see a significant decrease in cost by taking into account the auxiliary data in the multistage setting.

### 3.1 Introduction

Many fundamental problems in operations research (OR) involve making decisions, dynamically, subject to uncertainty. A decision maker seeks a sequence of actions that minimize the cost of operating a system. Each action is followed by a stochastic event,

and future actions are functions of the outcomes of these stochastic events. This type of problem has garnered much attention and has been studied extensively by different communities and under various names (dynamic programming, multistage stochastic optimization, Markov decision process, etc.). Much of this work, dating back to Bellman [6], has focused on the setting in which the distribution of the uncertain quantities is known a priori.

In practice, it is rare to know the joint distribution of the uncertain quantities. However, in today's data-rich world, we often have historical observations of the uncertain quantities of interest. Some existing methods work with independent and identically distributed (i.i.d.) observations of the uncertainties (cf. Swamy and Shmoys [92], Shapiro [87]). However, in general, auxiliary data has been ignored in modeling multistage problems, and this can lead to inadequate solutions.

In practice, we often have data,  $\{y^1, \dots, y^N\}$ , on uncertain quantities of interest,  $Y \in \mathcal{Y} \subset \mathbb{R}^{d_y}$ . In addition, we also have data,  $\{x^1, \dots, x^N\}$ , on auxiliary covariates,  $X \in \mathcal{X} \subset \mathbb{R}^d$ , which can be used to predict the uncertainty,  $Y$ . For example,  $Y$  may be the unknown demand for a product in the coming weeks, and  $X$  may include data about the characteristics of the particular product and data about the volume of Google searches for the product.

The machine learning (ML) community has developed many methods (for reference, see [26]) that enable the prediction of an uncertain quantity ( $Y$ ) given covariates ( $X$ ). These methods have been quite effective in generating predictions of quantities of interest in OR applications [54, 56]. However, turning good predictions into good decisions can be challenging. One naive approach is to solve the multistage optimization problem of interest as a deterministic problem, using the predicted values of the uncertainties. However, this ignores the uncertainty by using point predictions and can lead to inadequate decisions.

In this chapter we combine ideas from the OR and ML communities to develop a data-driven decision-making framework that incorporates auxiliary data into multistage stochastic optimization problems.

### 3.1.1 Multistage Optimization and Sample Average Approximation

Before proceeding, we first review the formulation of a multistage optimization problem with uncertainty. The problem is characterized by five components:

- The state at time  $t$ ,  $s_t \in S_t$ , contains all relevant information about the system at the start of time period  $t$ .
- The uncertainty,  $y_t \in \mathcal{Y}_t$ , is a stochastic quantity that is revealed prior to the decision at time  $t$ . Throughout this chapter, we assume the distribution of the uncertainty at time  $t$  does not depend on the current state or past decisions.
- The decision at time  $t$ ,  $z_t \in Z_t(s_t, y_t) \subset \mathbb{R}^{p_t}$ , which is chosen after the uncertainty,  $y_t$ , is revealed.
- The immediate cost incurred at time  $t$ ,  $g_t(z_t)$ , which is a function of the decision at time  $t$ .
- The dynamics of the system, which are captured by a known transition function that specifies how the state evolves,  $s_{t+1} = f_t(z_t)$ .

We note that it is without loss of generality that the cost and transition functions only depend on the decision variable because the feasible set  $Z_t$  is allowed to depend on  $s_t$  and  $y_t$ . To summarize, the system evolves in the following manner: at time  $t$ , the system is known to be in state  $s_t$ , when the previously unknown value,  $y_t$ , is observed. Then the decision  $z_t$  is determined, resulting in immediate cost,  $g_t(z_t)$ , and the system evolves to state  $s_{t+1} = f_t(z_t)$ .

Consider a finite horizon,  $T + 1$  stage problem, in which the initial state,  $s_0$ , is known. We formulate the problem as follows:

$$\min_{z_0 \in Z_0(s_0)} g_0(z_0) + \mathbb{E}[\tilde{Q}_1(f_0(z_0); Y_1)], \quad (3.1)$$

where

$$\tilde{Q}_t(s_t; y_t) = \min_{z_t \in Z_t(s_t, y_t)} g_t(z_t) + \mathbb{E}[\tilde{Q}_{t+1}(f_t(z_t); Y_{t+1})]$$

for  $t = 1, \dots, T - 1$ , and  $\tilde{Q}_T(s_T; y_T) = \min_{z_T \in Z_T(s_T, y_T)} g_T(z_T)$ . The function  $\tilde{Q}_t(s_t; y_t)$  is often called the value function or cost-to-go function. It represents the expected future cost that an optimal policy will incur, starting with a system in state  $s_t$  with realized uncertainty  $y_t$ . Of course, in practice it is impossible to solve this problem because the distributions of  $Y_t$  are unknown. All that we know about the distribution of  $Y_t$  comes from the available data.

A popular data-driven method for solving this problem is *sample average approximation* (SAA) [89]. In SAA, it is assumed that we have access to independent, identically distributed (i.i.d.) training samples of  $Y$ ,  $(y_1^i, \dots, y_T^i)$  for  $i = 1, \dots, N$ . The key idea of SAA is to replace the expectations over the unknown distributions of  $Y$  with empirical expectations. That is, we replace  $\mathbb{E}[\tilde{Q}_T(s_T; y_T)]$  with  $\frac{1}{N} \sum_{i=1}^N \tilde{Q}_T(s_T; y_T^i)$ . With these known, finite distributions of the uncertain quantities, the problem can be solved exactly or approximately by various dynamic programming techniques. Additionally, under certain conditions, the decisions obtained by solving the SAA problem are asymptotically optimal for (3.1) [86]. The basic SAA method does not incorporate auxiliary data. In practice, this can be accounted for by training a generative, parametric model and applying SAA with samples from this model conditioned on the observed auxiliary data. However, this approach does not necessarily lead to asymptotically optimal decisions, so we instead focus on a variant of SAA that starts directly with the data.

### 3.1.2 Related Work

Multistage optimization under uncertainty has attracted significant interest from various research communities. Bellman [6] studied these problems under the name dynamic programming. For reference, see Bertsekas [11]. These problems quickly become intractable as the state and action space grow, with a few exceptions that admit closed form solutions, like linear quadratic control [46]. However, there exists a large body of literature on approximate solution methods (see, e.g., Powell [80]).

When the distribution of the uncertainties is unknown, but data is available, SAA is a common approach [87]. Alternative approaches include robust dynamic

programming [62] and distributionally robust multistage optimization [55]. Another alternative approach is adaptive, or adjustable, robust optimization (cf. Ben-Tal et al. [7], Bertsimas et al. [20]). In this approach, the later stage decisions are typically constrained to be affine or piecewise constant functions of past uncertainties, usually resulting in highly tractable formulations.

In the artificial intelligence community, reinforcement learning (RL) studies a similar problem in which an agent tries to learn an optimal policy by intelligently trying different actions (cf. Sutton and Barto [90]). RL methods typically work very well when the exact dynamics of the system are unknown. However, they struggle to incorporate complex constraints that are common in OR problems. A vast literature also exists on bandit problems, which seek to find a series of decisions that balance exploration and exploitation (cf. Berry and Fristedt [10]). Of particular relevance is the contextual bandit problem (cf. Chapelle and Li [33], Chu et al. [38]), in which the agent has access to auxiliary data on the particular context in which it is operating. These methods have been very effective in online advertising and recommender systems [70].

Recently, the single stage optimization problem with auxiliary data has attracted interest in the OR community. Ban and Rudin [3] studied a news-vendor problem in the presence of auxiliary data. Cohen et al. [40] used a contextual bandit approach in a dynamic pricing problem with auxiliary data. Ferreira et al. [49] used data on the sales of past products, along with auxiliary data about the products, to solve a price optimization problem for never before sold products. Bertsimas and Kallus [15] developed a framework for integrating predictive machine learning methods in a single-stage stochastic optimization problem. Recently, Ban et al. [4] developed a method to solve a multistage dynamic procurement problem with auxiliary data. They used linear or sparse linear regression to build a different scenario tree for each realization of auxiliary covariates. Their approach assumes the uncertainty is a linear function of the auxiliary covariates with additive noise. Our approach is more general because we do not assume a parametric form of the uncertainty.

Statistical decision theory and ML have been more interested in the problems of

estimation and prediction than the problem of prescription (cf. Berger [9]). However, of integral importance to our work are several highly effective, yet simple, nonparametric regression methods. These include  $k$ -nearest neighbor regression [1], CART [30], and random forests [29].

### 3.1.3 Contributions and Structure

In this chapter, we consider the analogue of (3.1) in the presence of auxiliary data. For each  $t = 0, \dots, T-1$ , before the decision  $z_t$  is made, we observe auxiliary covariates  $x_t \in \mathcal{X}_t \subset \mathbb{R}^{d_t}$ . Our training data consists of  $(x_0^1, \dots, x_{T-1}^1), \dots, (x_0^N, \dots, x_{T-1}^N)$  and  $(y_1^1, \dots, y_T^1), \dots, (y_1^N, \dots, y_T^N)$ . By saying training data is i.i.d., we mean observation  $i$ ,  $(x_0^i, \dots, x_{T-1}^i, y_1^i, \dots, y_T^i)$ , was sampled independently of all other observations from the same joint distribution on  $\mathcal{X}_1 \times \dots \times \mathcal{X}_{T-1} \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_T$ .

We assume throughout that the auxiliary covariates evolve according to a Markov process, i.e.,  $X_t$  is conditionally independent of  $X_0, \dots, X_{t-2}, Y_1, \dots, Y_{t-2}$  given  $X_{t-1}$ . This framework can model more complex dependencies because we are able to choose the auxiliary covariate space. We could for example, append  $X_0, \dots, X_{t-1}$  to the auxiliary covariates observed at time  $t$ . In addition, we assume  $Y_t$  is conditionally independent of all past observations given  $X_{t-1}$ .

The problem we seek to solve is defined:

$$v^*(x_0) = \min_{z_0 \in \mathcal{Z}_0(s_0)} g_0(z_0) + \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0], \quad (3.2)$$

where

$$Q_t(s_t; y_t, x_t) = \min_{z \in \mathcal{Z}_t(s_t, y_t)} g_t(z) + \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | X_t = x_t]$$

for  $t = 1, \dots, T-2$ , with

$$\begin{aligned} & Q_{T-1}(s_{T-1}; y_{T-1}, x_{T-1}) \\ &= \min_{z \in \mathcal{Z}_{T-1}(s_{T-1}, y_{T-1})} g_{T-1}(z) + \mathbb{E}[Q_T(f_{T-1}(z); Y_T) | X_{T-1} = x_{T-1}], \end{aligned}$$



and  $Q_T(s_T; y_T) = \min_{z \in Z_T(s_T, y_T)} g_T(z)$ . We summarize our key contributions here.

1. In Section 3.2, we extend the framework introduced by Bertsimas and Kallus [15] to the multistage setting. Similarly to SAA, we replace the expectations in (3.2) with sums over the value functions, evaluated at observations of  $Y$ . However, unlike SAA, we weight the observations according to their relevance to the current problem’s auxiliary data, using weight functions inspired by popular machine learning methods.
2. In Section 3.3, we prove the asymptotic optimality and consistency of our method with  $k$ -nearest neighbor, CART, and random forest weight functions, under fairly mild conditions, for the multistage problem. (We formalize these definitions in that section.) The result for the  $k$ -nearest neighbor weight function is new for the multistage setting, and the results for the CART and random forest weight functions are new for both the single-stage and multistage settings.
3. In Section 3.4, we establish finite sample guarantees for our method with  $k$ -nearest neighbor weight functions. These guarantees are new for both the single-stage and multistage problems.
4. In Section 3.5, we demonstrate the practical tractability of our method with several computational examples using synthetic data. In addition, our results show that accounting for auxiliary data can have significant value.

## 3.2 Approach

In this section, we introduce our framework for solving multistage optimization problems under uncertainty in the presence of auxiliary covariates (3.2). Motivated by the framework developed by Bertsimas and Kallus [15], and analogous to SAA, we replace the expectations over an unknown distribution with finite weighted sums of the value functions evaluated at the observations in the data. The weights we use are obtained from ML methods.

First, we use our training data to learn weight functions,  $w_{N,i}^t(x_{t-1})$ , which quantify the similarity of a new observation,  $x_{t-1}$ , to each of the training examples,  $x_{t-1}^1, \dots, x_{t-1}^N$ . We then replace the conditional expectations in (3.2) with weighted sums. In particular,

$$\hat{v}_N(x_0) = \min_{z \in Z_0(s_0)} g_0(z) + \sum_{i=1}^N w_{N,i}^1(x_0) \hat{Q}_1(f_0(z); y_1^i, x_1^i), \quad (3.3)$$

where

$$\hat{Q}_t(s_t; y_t, x_t) = \min_{z \in Z_t(s_t, y_t)} g_t(z) + \sum_{i=1}^N w_{N,i}^{t+1}(x_t) \hat{Q}_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i)$$

for  $t = 1, \dots, T-2$ , with

$$\begin{aligned} & \hat{Q}_{T-1}(s_{T-1}; y_{T-1}, x_{T-1}) \\ &= \min_{z \in Z_{T-1}(s_{T-1}, y_{T-1})} g_{T-1}(z) + \sum_{i=1}^N w_{N,i}^T(x_{T-1}) \hat{Q}_T(f_{T-1}(z); y_T^i), \end{aligned}$$

$$\text{and } \hat{Q}_T(s_T; y_T) = \min_{z \in Z_T(s_T, y_T)} g_T(z).$$

We note that this is analogous to the sample average approximation method, which can be represented in this framework with the weight functions equal to  $\frac{1}{N}$ . The weight functions can be computed from various predictive machine learning methods. We list here a few examples that we find effective in practice.

**Definition 3.1.** *Motivated by  $k$ -nearest neighbor regression [1], the  $k$ NN weight function is given by*

$$w_{N,i}(x) = \begin{cases} 1/k, & x^i \text{ is a } k\text{NN of } x, \\ 0, & \text{otherwise.} \end{cases}$$

**Definition 3.2.** *Motivated by classification and regression trees [30], the CART*

weight function is given by

$$w_{N,i}(x) = \begin{cases} 1/|R(x)|, & x^i \in R(x), \\ 0, & \text{otherwise.} \end{cases},$$

where  $R(x)$  is the set of training points in the same partition as  $x$  in the CART model.

**Definition 3.3.** Motivated by random forests [29], the random forests weight function is given by

$$w_{N,i}(x) = \frac{1}{B} \sum_{b=1}^B \frac{1}{|R^b(x)|} \mathbb{1}\{x^i \in R^b(x)\},$$

where  $R^b(x)$  is the set of training points in the same partition as  $x$  in tree  $b$  of the random forest.

We offer two observations on the formulation in (3.3) before proceeding.

1. In SAA, we are justified in replacing  $\mathbb{E}[\theta]$  by  $\frac{1}{N} \sum_{i=1}^N \theta^i$  by the strong law of large numbers. We will see in Section 3.3 that a conditional strong law of large numbers holds under certain conditions, and this justifies replacing  $\mathbb{E}[h(Y_t) | X = x]$  with  $\sum_{i=1}^N w_{N,i}(x)h(y_t^i)$ .
2. If the weight functions are nonnegative and sum to one (as is the case with those presented here), then we can think of this formulation as defining a dynamic programming problem in which the uncertain quantities have a known, finite distribution. This means we can readily apply exact and approximate dynamic programming algorithms to solve (3.3). For the reader's convenience, we provide a decomposition algorithm, tailored to our approach, in Appendix B.2, which can be used to exactly solve small to moderately sized problems.

### 3.2.1 Notation

We summarize the relevant notation we use in Table 3.1. We use lower case letters for  $x_t^i$  and  $y_t^i$  to denote observed quantities in the data and capital letters,  $X_t^i$  and  $Y_t^i$ , to denote random quantities. When we discuss the asymptotic optimality of solutions

$x_t$	Auxiliary data observed at time $t$ , $x_t \in \mathcal{X}_t \subset \mathbb{R}^{d_t}$
$s_t$	State of the system at the beginning of time period $t$ , $s_t \in S_t$
$z_t$	Decision made at time $t$ , $z_t \in Z_t(s_t, y_t) \subset \mathbb{R}^{p_t}$
$y_t$	Uncertain quantity observed after the decision at time $t - 1$ , $y_t \in \mathcal{Y}_t$
$f_t(z_t)$	Transition function that gives the evolution of the state to $s_{t+1}$
$g_t(z_t)$	Cost of decision $z_t$ at time $t$
$w_{N,i}^t(x_{t-1})$	Weight function for stage $t$ , gives weighting for training sample $i$
$Q_t(s_t; y_t, x_t)$	Value function in full information problem (3.2)
$\hat{Q}_t(s_t; y_t, x_t)$	Value function in approximate problem (3.3)
$v^*(x_0)$	Optimal objective value of full information problem (3.2)
$\hat{v}_N(x_0)$	Optimal objective value of approximate problem (3.3)

Table 3.1: Summary of notation.

in Definitions 3.4 and 3.5, the data are random quantities, and thus solutions to (3.3) are also random. For notational convenience, we sometimes write  $Q_T(s_T; y_T, x_T)$  even though  $Q_T$  does not depend on  $x_T$  (because the auxiliary data observed after the last decision is made is irrelevant to the problem).

### 3.3 Asymptotic Optimality

In the setting without auxiliary covariates, under certain conditions, the SAA estimator is strongly asymptotically optimal [86]. Here, we provide similar results for the multistage setting with auxiliary data.

**Definition 3.4.** *We say  $\hat{z}_0^N(x_0)$ , a sequence of optimal solutions to (3.3), is strongly asymptotically optimal if, for  $x_0$  almost everywhere (a.e.),*

1. *The estimated cost of  $\hat{z}_0^N(x_0)$  converges to the true optimal cost:*

$$\min_{z \in Z_0(s_0)} g_0(z) + \sum_{i=1}^N w_{N,i}^1(x_0) \hat{Q}_1(f_0(z); Y_1^i, X_1^i) \rightarrow v^*(x), \quad (3.4)$$

*almost surely.*

2. *The true cost of  $\hat{z}_0^N(x_0)$  converges to the true optimal cost:*

$$\mathbb{E}[g_0(\hat{z}_0^N) + Q_1(f_0(\hat{z}_0^N); Y_1, X_1) | X_0 = x_0, \hat{z}_0^N] \rightarrow v^*(x_0),$$

almost surely.

3. The limit points of  $\{\hat{z}_0^N(x_0)\}$  are contained in the set of true optimal solutions:

$$L(\{\hat{z}_0^N(x_0) : N \in \mathbb{N}\}) \subset \arg \min_{z \in Z_0(s_0)} \mathbb{E}[g_0(z) + Q_0(f_0(z); Y_1, X_1) | X_0 = x_0],$$

almost surely, where  $L(S)$  denotes the limit points of the set  $S$ .

**Definition 3.5.** We say  $\hat{z}_0^N(x_0)$ , a sequence of optimal solutions to (3.3), is weakly asymptotically optimal if, for  $x_0$  almost everywhere (a.e.),

1. The estimated cost of  $\hat{z}_0^N(x_0)$  converges to the true optimal cost in probability:

$$\min_{z \in Z_0(s_0)} g_0(z) + \sum_{i=1}^N w_{N,i}^1(x_0) \hat{Q}_1(f_0(z); Y_1^i, X_1^i) \rightarrow_P v^*(x). \quad (3.5)$$

2. The true cost of  $\hat{z}_0^N(x_0)$  converges to the true optimal cost in probability:

$$\mathbb{E}[g_0(\hat{z}_0^N) + Q_1(f_0(\hat{z}_0^N); Y_1, X_1) | X_0 = x_0, \hat{z}_0^N] \rightarrow_P v^*(x_0).$$

### 3.3.1 $k$ -Nearest Neighbor Weight Functions

We begin by defining some assumptions under which asymptotic optimality will hold for (3.3) with the  $k$ -nearest neighbor weight functions.

**Assumption 3.1** (Regularity). For each  $t = 0, \dots, T$ , there exists a closed and bounded set  $W_t$  such that for any  $s_t$  and  $y_t$ , the feasible region,  $Z_t(s_t, y_t)$ , is contained in  $W_t$ .

**Assumption 3.2** (Existence). The full information problem (3.2) is well defined for all stages,  $t = 0, \dots, T - 1$ :  $\mathbb{E}|g_t(z_t) + Q_{t+1}(f_t(z_t); Y_{t+1}, X_{t+1})| < \infty$  for all  $z_t \in W_t$  and  $Z_t(s_t, y_t)$  is nonempty for all  $s_t$  and  $y_t$ .

**Assumption 3.3** (Continuity). The function  $g_{t-1}(z_{t-1}) + Q_t(f_{t-1}(z_{t-1}); y_t, x_t)$  is equicontinuous in  $z_{t-1}$  at all stages. That is, for  $t = 1, \dots, T$ ,  $\forall z \in W_{t-1}, \epsilon > 0$ ,

$\exists \delta > 0$  such that

$$\sup_{y_t \in \mathcal{Y}_t} \sup_{x_t \in \mathcal{X}_t} |g_{t-1}(z) + Q_t(f_{t-1}(z); y_t, x_t) - g_{t-1}(z') - Q_t(f_{t-1}(z'); y_t, x_t)| \leq \epsilon$$

for all  $z' \in \{z' : \|z - z'\| \leq \delta\} \cap W_t$ . Additionally, the final cost function,  $g_T(z_T)$  is continuous.

**Assumption 3.4** (Distribution of Uncertainties). *The following hold:*

1. The stochastic process  $X_0, X_1, \dots, X_{T-1}$  satisfies the Markov property.
2. For each  $t = 1, \dots, T$ ,  $Y_t$  is conditionally independent of  $Y_1, \dots, Y_{t-1}$  and  $X_0, \dots, X_{t-2}$  given  $X_{t-1}$ .
3. For each  $t = 0, \dots, T-1$ , the support of  $X_t$ ,  $\mathcal{X}_t$ , is compact.
4. The noise terms have uniformly bounded tails, i.e., defining  $N_t(s_t; x_{t-1}) = Q_t(s_t; Y_t, X_t) - \mathbb{E}[Q_t(s_t; Y_t, X_t) | X_{t-1} = x_{t-1}]$ , there exists  $\lambda > 0$  such that

$$\max_{t=1, \dots, T} \sup_{x_{t-1} \in \mathcal{X}_{t-1}} \sup_{s_t \in \mathcal{S}_t} \mathbb{E}[e^{\lambda |N_t(s_t; x_{t-1})|} | X_{t-1} = x_{t-1}] < \infty.$$

5. For all  $t$  and  $s_t$ ,  $\mathbb{E}[Q_t(s_t; Y_t, X_t) | X_{t-1} = x]$  is a continuous function of  $x$ .

We remark that Assumption 3.3 does not preclude the possibility of integral constraints on  $z_t$ . In fact, if  $Z_t$  is a finite discrete set, the assumption is automatically satisfied (choose  $\delta < \min_{z, z' \in Z_t} \|z - z'\|$ ). Condition 4 of Assumption 3.4 can be satisfied if  $N_t(s_t; x_{t-1})$  are uniformly bounded random variables, are subgaussian random variables with uniformly bounded subgaussian norms, or are subexponential random variables with uniformly bounded subexponential norms. With these assumptions, we have the following result regarding the asymptotic optimality of (3.3) with the  $k$ -nearest neighbor weight functions.

**Theorem 3.1.** *Suppose Assumptions 3.1-3.4 hold, and the training data is i.i.d. Let  $w_{N,i}^t(x_{t-1})$  be the  $k$ -nearest neighbor weight functions for  $t = 1, \dots, T$  with  $k =$*

$\min\{\lceil CN^\delta \rceil, N - 1\}$  for  $C > 0, \delta \in (0, 1)$ . Then  $\{\hat{z}_0^N(x)\}$ , a sequence of optimal solutions to (3.3), is strongly asymptotically optimal.

Comparing the result with that of Bertsimas and Kallus [15] for the single-stage problem, we see it is quite similar. Both theorems assume regularity, existence, and continuity. The difference in the multistage setting is that these assumptions must hold for the value function at each stage. We also note that Bertsimas and Kallus [15] listed several sets of regularity assumptions that could hold, whereas we only list one for clarity. However, the extension of the other sets of regularity assumptions to the multistage setting is straightforward.

To prove this result, we rely on several technical lemmas. Lemmas 3.1, 3.3, and 3.4 are refined versions of results originally stated in Bertsimas and Kallus [15].

**Lemma 3.1.** *Suppose for each  $N \in \mathbb{N}$ ,  $\hat{C}_N(z|x)$  and  $C(z|x)$  are equicontinuous functions, i.e.,  $\forall \epsilon > 0$  and  $z \in \mathcal{Z}$ ,  $\exists \delta > 0$  s.t.  $\sup_{z' \in B_\delta(z) \cap \mathcal{Z}} \sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - \hat{C}_N(z'|x)| \leq \epsilon$ , and likewise for  $C$ . If for every  $z \in \mathcal{Z}$ ,*

$$\sup_{x \in \mathcal{X}} \left| \hat{C}_N(z|x) - C(z|x) \right| \rightarrow 0,$$

*then the convergence is uniform over any compact subset of  $\mathcal{Z}$ .*

*Proof.* Let  $z_N \in \mathcal{Z}$  be a sequence that converges to  $z \in \mathcal{Z}$ . For any  $\epsilon > 0$ , by assumption,  $\exists \delta > 0$  such that  $\sup_{z' \in B_\delta(z) \cap \mathcal{Z}} |\hat{C}_N(z|x) - C(z'|x)| \leq \epsilon/2$  for all  $x$ . By the convergence of  $z_N$ , there exists an  $N_1$  such that  $\forall N \geq N_1$ ,  $z_N \in B_\delta(z) \cap \mathcal{Z}$ . This implies,  $\forall N \geq N_1$ ,  $|\hat{C}_N(z_N|x) - \hat{C}_N(z|x)| \leq \epsilon/2$  for all  $x$ . Additionally, by assumption,  $\exists N_2$  s.t.  $\forall N \geq N_2$ ,  $\sup_{x \in \mathcal{X}} \left| \hat{C}_N(z|x) - C(z|x) \right| \leq \epsilon/2$ . Therefore,  $\forall N \geq \max(N_1, N_2)$

$$\sup_{x \in \mathcal{X}} \left| \hat{C}_N(z_N|x) - C(z|x) \right| \leq \epsilon.$$

Therefore, for any convergent sequence in  $\mathcal{Z}$ ,  $z_N \rightarrow z$ ,  $\sup_{x \in \mathcal{X}} |\hat{C}_N(z_N|x) - C(z|x)| \rightarrow 0$ .

Given a compact subset  $E \subset \mathcal{Z}$ , suppose that  $\sup_{z \in E} \sup_{x \in \mathcal{X}} \hat{C}_N(z|x) - C(z|x) \not\rightarrow 0$ . This implies  $\exists \epsilon > 0$  and a sequence  $z_N \in E$  such that  $\sup_{x \in \mathcal{X}} |\hat{C}_N(z_N|x) - C(z_N|x)| > \epsilon$  occurs

infinitely often. Define  $z_{N_k}$  to be the subsequence for which this event occurs. Since  $E$  is compact, by the Bolzano-Weirestrass theorem,  $z_{N_k}$  has a convergent subsequence in  $E$ . If we define  $\{s_N\}$  to be this subsubsequence, we have that  $s_N \rightarrow s \in E$  and  $\sup_{x \in \mathcal{X}} |\hat{C}_N(s_N|x) - C(s_N|x)| > \epsilon$  for all  $N$ . We have  $\sup_{x \in \mathcal{X}} |\hat{C}_N(s_N|x) - C(s_N|x)| \leq \sup_{x \in \mathcal{X}} |\hat{C}_N(s_N|x) - C(s|x)| + \sup_{x \in \mathcal{X}} |C(s|x) - C(s_N|x)|$ . The first term converges to 0 because of what we showed above. The second term converges to 0 by the equicontinuity assumption. This is a contradiction, so it must be that  $\sup_{z \in E} \sup_{x \in \mathcal{X}} |\hat{C}(z|x) - C(z|x)| \rightarrow 0$  for any compact set  $E \subset \mathcal{Z}$ .  $\square$

This lemma shows that, given an equicontinuity assumption, pointwise convergence of a function implies uniform convergence over a compact set. We will apply this with  $C = Q_t$ . The following two lemmas establish that strong asymptotic optimality follows from the uniform convergence of the objective of (3.3) to that of (3.2).

**Lemma 3.2.** *Suppose  $\mathcal{Z}$  is a compact set and  $h(z)$  and  $g(z)$  are two continuous functions. If  $z^* \in \arg \min_{z \in \mathcal{Z}} h(z)$  and  $z' \in \arg \min_{z \in \mathcal{Z}} g(z)$ , then*

$$|h(z^*) - g(z')| \leq \sup_{z \in \mathcal{Z}} |h(z) - g(z)|,$$

and

$$|h(z^*) - h(z')| \leq 2 \sup_{z \in \mathcal{Z}} |h(z) - g(z)|.$$

*Proof.* First, because of the optimality of  $z^*$  and  $z'$ ,

$$\begin{aligned} & |h(z^*) - g(z')| \\ & \leq \begin{cases} |h(z^*) - g(z^*)|, & h(z^*) \leq g(z') \\ |h(z') - g(z')|, & h(z^*) > g(z') \end{cases} \\ & \leq \sup_{z \in \mathcal{Z}} |h(z) - g(z)|. \end{aligned}$$



Second,

$$\begin{aligned} |h(z^*) - h(z')| &\leq |h(z') - g(z')| + |g(z') - h(z^*)| \\ &\leq 2 \sup_{z \in \mathcal{Z}} |h(z) - g(z)|. \end{aligned}$$

□

**Lemma 3.3.** *Fix  $x \in \mathcal{X}$  and suppose  $\sup_{z \in \mathcal{Z}} |\hat{C}_N(z|x) - C(z|x)| \rightarrow 0$  as  $N \rightarrow \infty$  and  $C(z|x)$  is a continuous function of  $z$ . In addition, suppose constraint set  $\mathcal{Z}$  is nonempty, closed, and bounded. Any sequence  $z_N \in \arg \min_{z \in \mathcal{Z}} \hat{C}_N(z|x)$  for  $N \in \mathbb{N}$  has all of its limit points contained in  $\arg \min_{z \in \mathcal{Z}} C(z|x)$ .*

*Proof.* Suppose there is a subsequence  $z_{N_k}$ , converging to  $z \notin \arg \min_{z \in \mathcal{Z}} C(z|x)$ . (We must still have that  $z \in \mathcal{Z}$  because  $\mathcal{Z}$  is compact.) Let  $\epsilon = C(z|x) - \min_{z \in \mathcal{Z}} C(z|x) > 0$ . By the continuity assumption,  $\exists k_1$  such that for all  $k \geq k_1$   $|C(z_{N_k}|x) - C(z|x)| \leq \epsilon/4$ . Additionally, by assumption, we can find a  $k_2$  such that  $\forall k \geq k_2$ ,  $|\hat{C}_{N_k}(z_{N_k}|x) - C(z_{N_k}|x)| \leq \sup_{z \in \mathcal{Z}} |\hat{C}_{N_k}(z|x) - C(z|x)| \leq \epsilon/4$ . This implies that for any  $k \geq \max(k_1, k_2)$ ,

$$\min_{z \in \mathcal{Z}} \hat{C}_{N_k}(z|x) = \hat{C}_{N_k}(z_{N_k}|x) \geq C(z_{N_k}|x) - \epsilon/4 \geq C(z|x) - \epsilon/2 = \min_{z \in \mathcal{Z}} C(z|x) + \epsilon/2.$$

From lemma 3.2, we know that

$$|\min_{z \in \mathcal{Z}} \hat{C}_{N_k}(z|x) - \min_{z \in \mathcal{Z}} C(z|x)| \leq \sup_{z \in \mathcal{Z}} |\hat{C}_{N_k}(z|x) - C(z|x)|,$$

which goes to 0 as  $k \rightarrow \infty$ , and is thus a contradiction. Therefore, all limit points of  $\arg \min_z \hat{C}_N(z|x)$  must be contained in  $\arg \min_z C(z|x)$ . □

Lemma 3.4 shows that, given an equicontinuity assumption, if a function indexed by  $z$  converges almost surely for each  $z$  in a compact set, then the convergence holds almost surely for all  $z$ .

**Lemma 3.4.** *Suppose for each  $N \in \mathbb{N}$ ,  $\hat{C}_N(z|x)$  and  $C(z|x)$  are equicontinuous functions: i.e.,  $\forall \epsilon > 0$  and  $z \in \mathcal{Z}$ ,  $\exists \delta > 0$  s.t.  $\sup_{z' \in B_\delta(z) \cap \mathcal{Z}} \sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - \hat{C}_N(z'|x)| \leq \epsilon$ , and*

likewise for  $C$ . In addition, suppose that for each  $z \in \mathcal{Z}$ ,  $\sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - C(z|x)| \rightarrow 0$  almost surely ( $\hat{C}_N(z|x)$  is a random quantity). Furthermore, assume  $\mathcal{Z}$  is compact. Then, almost surely,  $\sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - C(z|x)|$  for all  $z \in \mathcal{Z}$ .

*Proof.* Let  $\mathcal{Z}' = \mathcal{Z} \cap \mathbb{Q}^d \cup \{\text{isolated points of } \mathcal{Z}\}$ . Because  $\mathcal{Z}'$  is countable, for  $x$  almost everywhere,  $P\left(\bigcap_{z' \in \mathcal{Z}'} \{\sup_{x \in \mathcal{X}} |\hat{C}_N(z'|x) - C(z'|x)| \rightarrow 0\}\right) = 1$  by the continuity of probability measures. For any sample path for which this occurs, consider any  $z \in \mathcal{Z}$ . We have, for any  $z' \in \mathcal{Z}'$ ,

$$\begin{aligned} \sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - C(z|x)| &\leq \sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - \hat{C}_N(z'|x)| + \sup_{x \in \mathcal{X}} |C(z'|x) - C(z|x)| \\ &\quad + \sup_{x \in \mathcal{X}} |\hat{C}_N(z'|x) - C(z'|x)|. \end{aligned}$$

By equicontinuity and the density of  $\mathcal{Z}'$ , we can pick  $z' \in \mathcal{Z}'$  such that each of the first two terms is less  $\epsilon/3$  for any  $\epsilon > 0$ . By assumption, we can also find an  $N_1$  such that the third term is bounded by  $\epsilon/3$  for all  $N \geq N_1$ , so we have  $\sup_{x \in \mathcal{X}} |\hat{C}_N(z|x) - C(z|x)| \leq \epsilon$  for all  $N \geq N_1$ . This is true for any  $z \in \mathcal{Z}$  for this particular sample path. Since the set of sample paths for which this is true constitutes a measure 1 event, we have the desired result.  $\square$

We also restate a result from Biau and Devroye [25] (Theorem 12.1) regarding the uniform consistency of the  $k$ -nearest neighbor regression estimator.

**Lemma 3.5.** *Let  $(X^1, Y^1), \dots, (X^N, Y^N) \in \mathbb{R}^d \times \mathbb{R}$  be i.i.d. observations of random variables  $(X, Y)$ . Assume  $X$  has support on a compact set  $\mathcal{X} \subset \mathbb{R}^d$  and there exists  $\lambda > 0$  such that*

$$\sup_{x \in \mathbb{R}^d} \mathbb{E}[\exp(\lambda|Y - \mathbb{E}[Y|X = x]|)|X = x] < \infty.$$

*In addition, assume  $\mathbb{E}[Y|X = x]$  is a continuous function. For some  $C > 0, \delta \in (0, 1)$ , let  $k_N = \min\{\lceil CN^\delta \rceil, N-1\}$ . If  $m_N(x)$  is the  $k_N$  nearest neighbor regression estimator*

for  $Y$  and  $m(x) = \mathbb{E}[Y|X = x]$ , then

$$\sup_{x \in \mathcal{X}} |m_N(x) - m(x)| \rightarrow 0$$

almost surely.

From these lemmas, the proof of Theorem 3.1 follows.

*Proof of Theorem 3.1.* We need to show that

$$\sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \rightarrow 0$$

a.s. for  $x_0$  a.e. The desired result then follows from lemmas 3.2 and 3.3. To begin, we have:

$$\begin{aligned} & \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \\ & \leq \left| \sum_i w_{N,i}^1(x_0) Q_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \\ & \quad + \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \sum_i w_{N,i}^1(x_0) Q_1(f_0(z_0); y_1^i, x_1^i) \right|. \end{aligned}$$

Expanding the second term on the right hand side, and using the fact that  $\sum_i w_{N,i}^1(x_0) = 1$  and  $w_{N,i}^1(x_0) \geq 0$ , we have:

$$\begin{aligned} & \left| \sum_i w_{N,i}^1(x_0) \left( \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - Q_1(f_0(z_0); y_1^i, x_1^i) \right) \right| \\ & \leq \sum_i w_{N,i}^1(x_0) \left| \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - Q_1(f_0(z_0); y_1^i, x_1^i) \right| \\ & \leq \sup_{x_1 \in \mathcal{X}_1} \left| \min_{z_1 \in W_1} \left( g_1(z_1) + \sum_i w_{N,i}^2(x_1) \hat{Q}_2(f_1(z_1); y_2^i, x_2^i) \right) \right. \\ & \quad \left. - \min_{z_1 \in W_1} (g_1(z_1) + \mathbb{E}[Q_2(f_1(z_1); Y_2, X_2) | X_1 = x_1]) \right| \\ & \leq \sup_{x_1 \in \mathcal{X}_1} \sup_{z \in W_1} \left| \sum_{i=1}^N w_{N,i}^2(x_1) \hat{Q}_2(f_1(z); y_2^i, x_2^i) - \mathbb{E}[Q_2(f_1(z); Y_2) | x_1] \right|, \end{aligned}$$

where we have used lemma 3.2. Therefore, we have:

$$\begin{aligned}
& \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1)|x_0] \right| \\
& \leq \left| \sum_i w_{N,i}^1(x_0) Q_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1)|x_0] \right| \\
& \quad + \sup_{x_1 \in \mathcal{X}_1} \sup_{z_1 \in W_1} \left| \sum_{i=1}^N w_{N,i}^2(x_1) \hat{Q}_2(f_1(z_1); y_2^i, x_2^i) - \mathbb{E}[Q_2(f_1(z_1); Y_2, X_2)|x_1] \right|.
\end{aligned}$$

Repeating the above argument for  $t = 2, \dots, T - 1$ , we have:

$$\begin{aligned}
& \sup_{z_0 \in \mathcal{Z}_0} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1)|X_0 = x_0] \right| \\
& \leq \sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) Q_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1)|X_0 = x_0] \right| \\
& \quad + \sum_{t=1}^{T-1} \sup_{x_t \in \mathcal{X}_t} \sup_{z_t \in W_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z_t); y_{t+1}^i, x_{t+1}^i) \right. \\
& \quad \left. - \mathbb{E}[Q_{t+1}(f_t(z_t); Y_{t+1}, X_{t+1})|x_t] \right|.
\end{aligned}$$

To see that each term on the right hand side goes to 0 a.s., we first apply lemma 3.5 to each term. This shows that each term (without the supremums over  $z_t$ ) goes to 0 a.s. for each  $z_t$ . Next we apply lemma 3.4 to each term to show that the convergence holds simultaneously for all  $z_t$  with probability 1. Finally, we apply lemma 3.1 to show the convergence of each term is uniform over  $z_t$  a.s. To do so, we let  $\hat{C}_N(z|x) = \sum_i w_{N,i}^{t+1}(x) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i)$  and  $C(z|x) = \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1})|x]$ . We can verify the equicontinuity assumption holds for each of these functions because of Assumption 3.3 and Jensen's inequality (because  $w_{N,i}^{t+1}(x)$  define a probability distribution). This completes the proof.  $\square$

### 3.3.2 CART Weight Functions

In order to study the asymptotic properties of (3.3) with the CART and random forest weight functions, we need to consider modified versions of the original algorithms of

Breiman et al. [30] and Breiman [29]. Since greedy decision trees have proven difficult to analyze theoretically, we instead consider a modified tree learner introduced by Wager and Athey [97]. Formally, a regression tree is defined as

$$T(x; \xi, X_1, Y_1, \dots, X_n, Y_n) = \frac{1}{|\{i : X_i \in R(x)\}|} \sum_{\{i: X_i \in R(x)\}} Y_i,$$

where  $R(x)$  identifies the region of the tree containing  $x$ , and  $\xi$  is an auxiliary source of randomness. Trees are built by recursively partitioning the feature space. At each step of the training process, for each region, a feature is selected and a cutoff is chosen to define an axis-aligned hyperplane to partition the region into two smaller regions. This is repeated until every region contains some minimum number of training points. In order to guarantee consistency, we place several restrictions on how the trees are built. We use the following definitions from Wager and Athey [97].

**Definition 3.6** (Random-split, regular, and honest trees). *Let the regression tree  $T(x; \xi, X_1, Y_1, \dots, X_n, Y_n)$  be the type defined above.*

1.  *$T$  is a random-split tree if at each step in the training procedure, the probability that the next split occurs in the  $j$ th feature is at least  $\pi/d$  for all  $j = 1, \dots, d$ , with some  $\pi > 0$ . This source of this randomness is  $\xi$ .*
2.  *$T$  is a regular tree if at each split leaves at least a fraction  $\lambda > 0$  of the available training examples on each side of the split. Additionally, the tree is grown to full depth  $k \in \mathbb{N}$ , meaning there are between  $k$  and  $2k - 1$  training examples in each region of the feature space.*
3.  *$T$  is an honest tree if the splits are made independently of the response variables  $\{Y_1, \dots, Y_n\}$ . This can be achieved by ignoring the response variable entirely when making splits or by splitting the training data into two halves, one for making splits and one for making predictions. (If the latter is used, the tree is regular if at least a fraction  $\lambda$  of the available prediction examples are on each side of the split.)*

The standard implementation of the CART algorithm does not satisfy these definitions, but it is straightforward to modify the original algorithm so that it does. It involves modifying how splits are chosen. If we learn weight functions using trees that do satisfy these definitions, we can guarantee the solutions to (3.3) are weakly asymptotically optimal. We first introduce two additional assumptions. We note that these assumptions are strengthened versions of Assumptions 3.3 and 3.4.

**Assumption 3.5** (Distribution of Auxiliary Covariates). *The distribution of the auxiliary data,  $X$ , is uniform on  $[0, 1]^d$  (independent in each feature)<sup>1</sup>.*

**Assumption 3.6** (Continuity). *For each  $t = 1, \dots, T$ , there exists an  $L_t < \infty$  such that  $\forall y_t \in \mathcal{Y}_t, z_t \in W_t$  and  $\forall z, z' \in W_{t-1}$ ,*

$$|Q_t(f_{t-1}(z); y_t, x_t) - Q_t(f_{t-1}(z'); y_t, x_t)| \leq L_t \|z - z'\|.$$

*Furthermore, for each  $t = 1, \dots, T$ ,  $\mathbb{E}[Q_t(s_t; Y_t, X_t) | X = x]$  is  $M_t$ -Lipschitz continuous in  $x$ , for all  $s_t$ .*

**Theorem 3.2.** *Suppose Assumptions 3.1-3.6 hold, and the training data is i.i.d. Let  $w_{N,i}^t(x_{t-1})$  be the CART weight functions for  $t = 1, \dots, T$ , and assume the trees are honest, random split, and regular with  $k$ , the minimum number of training examples in each leaf, equal to  $\min\{\lceil CN^\delta \rceil, N - 1\}$  for  $C > 0, \delta \in (0, 1)$ . Then  $\{\hat{z}_0^N(x)\}$ , a sequence of optimal solutions to (3.3), is weakly asymptotically optimal.*

The proof of this result follows closely the proof of Theorem 3.1. To begin, we prove a result regarding the bias of tree based predictors. It relies on the same argument Wager and Athey [97] used in proving their Lemma 2.

**Lemma 3.6.** *Suppose  $T$  is a regular, random-split tree as in Definition 3.6, and the training covariates  $X_1, \dots, X_N$  are i.i.d. uniform( $[0, 1]^d$ ) random variables. If  $R(x)$*

---

<sup>1</sup>The result holds under more general distributional assumptions, but uniformity is assumed for simplicity. For example, we could assume  $X$  has a continuous density function on  $[0, 1]^d$ , bounded away from 0 and  $\infty$ . See Wager and Athey [97] for a further discussion.

denotes the partition of  $[0, 1]^d$  containing  $x \in \mathbb{R}^d$ , and  $\frac{N}{k} \rightarrow \infty$  as  $N \rightarrow \infty$ , then

$$\sup_x \text{diam}(R(x)) \rightarrow_P 0.$$

*Proof.* As in the proof of Lemma 2 in Wager and Athey [97], we define  $c(x)$  to be the number of splits leading to the leaf  $R(x)$  and  $c_j(x)$  to be the number of these splits that are on the  $j$ th coordinate. Following the same arguments as Wager and Athey [97], we have, conditional on  $X_1, \dots, X_N$ ,

$$c_j(x) \geq B_j(x),$$

where  $B_j(x) \sim \text{Binom} \left( \left\lfloor \frac{\log(N/(2k-1))}{\log \lambda^{-1}} \right\rfloor, \frac{\pi}{d} \right)$ . In addition, for  $N$  sufficiently large, with probability at least  $1 - k/N$  (over the training data),

$$\text{diam}_j(R(x)) \leq (1 - \lambda)^{0.99c_j(x)}.$$

We call this event  $A_N$ . From here, we have, for any  $\epsilon > 0$ ,

$$\begin{aligned} & P \left( \sup_x \text{diam}(R(x)) > \epsilon \mid X_1, \dots, X_N, A_N \right) \\ & \leq P \left( 0.99 \inf_j \inf_x c_j(x) < \frac{\log \epsilon^{-1}}{\log(1 - \lambda)^{-1}} \mid X_1, \dots, X_N, A_N \right) \\ & \leq d \left( \frac{N}{k} \right) P \left( c_j(x) \leq \frac{\log \epsilon^{-1}}{0.99 \log(1 - \lambda)^{-1}} \mid X_1, \dots, X_N, A_N \right) \\ & \leq d \left( \frac{N}{k} \right) P \left( B_j(x) \leq \frac{\log \epsilon^{-1}}{0.99 \log(1 - \lambda)^{-1}} \mid X_1, \dots, X_N, A_N \right) \\ & \leq d \left( \frac{N}{k} \right) \exp \left( -2 \left[ \frac{\log \frac{N}{2k-1}}{\log \lambda^{-1}} \right] \left( \frac{\pi}{d} \left[ \frac{\log \frac{N}{2k-1}}{\log \lambda^{-1}} \right] - \frac{\log \epsilon^{-1}}{0.99 \log(1 - \lambda)^{-1}} \right)^2 \right) \\ & \leq d \left( \frac{N}{k} \right) \exp(-C_1 \log^3(N/(2k-1))), \end{aligned}$$

for  $N/k$  sufficiently large, where  $C_1 > 0$  is a constant that does not depend on  $N$  or  $k$ . The second inequality follows from the union bound since there are a maximum of  $N/k$  total partitions in the tree, and the fourth inequality follows from Hoeffding's

inequality. Putting everything together, we have:

$$\begin{aligned}
& P \left( \sup_x \text{diam}(R(x)) > \epsilon \right) \\
& \leq P \left( \sup_x \text{diam}(R(x)) > \epsilon \middle| A_N \right) + \frac{k}{N} \\
& \leq d \left( \frac{N}{k} \right) \exp \left( -C_1 \log^3(N/(2k-1)) \right) + \frac{k}{N}.
\end{aligned}$$

It is easy to verify that the final expression goes to 0 as  $N \rightarrow \infty$ , so the proof is complete.  $\square$

Next, we establish the uniform consistency of the CART regression estimator.

**Lemma 3.7.** *Suppose  $T$  is a regular, random-split, honest tree as in Definition 3.6, the training data  $(X_1, Y_1), \dots, (X_N, Y_N)$  are i.i.d. with  $X_i$  uniform on  $[0, 1]^d$ , and  $\mathbb{E}[Y|X = x]$  is  $L$ -Lipschitz continuous. In addition, assume there exists  $\lambda > 0$  such that the uniform noise condition is satisfied:  $\sup_x \mathbb{E}[\exp(\lambda|Y_i - \mathbb{E}[Y_i|X_i = x]|)|X_i = x] < \infty$ . Finally, suppose  $\log N/k \rightarrow 0$  and  $N/k \rightarrow \infty$  as  $N \rightarrow \infty$ . If  $\hat{\mu}_N(x)$  denotes the prediction of  $T$  at  $X \in \mathbb{R}^d$  and  $\mu(x) = \mathbb{E}[Y|X = x]$ , then*

$$\sup_x |\hat{\mu}_N(x) - \mu(x)| \rightarrow_P 0.$$

*Proof.* To begin, we have

$$\begin{aligned}
& \sup_x |\hat{\mu}_N(x) - \mu(x)| \\
& \leq \sup_x |\hat{\mu}_N(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]| + \sup_x |\mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N] - \mu(x)| \\
& \leq \sup_x \left| \sum_i w_{N,i}(x)(Y_i - \mathbb{E}[Y_i|X_i]) \right| + \sup_x \left| \sum_i w_{N,i}(x)(\mu(X_i) - \mu(x)) \right| \\
& \leq \sup_x \left| \sum_i w_{N,i}(x)(Y_i - \mathbb{E}[Y_i|X_i]) \right| + \sup_x \sum_i w_{N,i}(x) |\mu(X_i) - \mu(x)|
\end{aligned}$$



$$\begin{aligned}
&\leq \sup_x \left| \sum_i w_{N,i}(x)(Y_i - \mathbb{E}[Y_i|X_i]) \right| + L \sup_x \sum_i w_{N,i}(x) \|X_i - x\| \\
&\leq \sup_x \left| \sum_i w_{N,i}(x)(Y_i - \mathbb{E}[Y_i|X_i]) \right| + L \sup_x \text{diam}(R(x)),
\end{aligned}$$

where  $w_{N,i}(x)$  is the CART weight function corresponding to tree  $T$ . In the third and fourth inequalities we used Jensen's inequality and the Lipschitz continuity assumption. By lemma 3.6, the latter term goes to 0 in probability. For the former, we define  $M(x) = |\{i : X_i \in R(x)\}|$  to be the number of training examples in the leaf containing  $x$ . For fixed  $x$ , if  $c = \sup_x \mathbb{E}[\exp(\lambda|Y_i - \mathbb{E}[Y_i|X_i = x]|)|X_i = x]$ , then by Lemma 12.1 of Biau and Devroye [25], we have, for any  $\epsilon > 0$ ,

$$\begin{aligned}
&P \left( \left| \sum_i w_{N,i}(x)(Y_i - \mathbb{E}[Y_i|X_i]) \right| > \epsilon \middle| X_1, \dots, X_N \right) \\
&\leq 2 \exp \left( -M(x) \frac{\min(\epsilon, \min(1, 2c)/\lambda)^2 \lambda^2}{8c} \right).
\end{aligned}$$

Because there are a maximum of  $N/k$  leaves in the tree, there are a maximum of  $N/k$  values of the weight function  $w_{N,i}(x)$ . Therefore, we can use the union bound to show:

$$\begin{aligned}
&P \left( \sup_x \left| \sum_i w_{N,i}(x)(Y_i - \mathbb{E}[Y_i|X_i]) \right| > \epsilon \middle| X_1, \dots, X_N \right) \\
&\leq \frac{2N}{k} \exp \left( -M(x) \frac{\min(\epsilon, \min(1, 2c)/\lambda)^2 \lambda^2}{8c} \right) \\
&\leq \frac{2N}{k} \exp \left( -k \frac{\min(\epsilon, \min(1, 2c)/\lambda)^2 \lambda^2}{8c} \right) \\
&= \exp \left( \log(2N) - \log k - k \frac{\min(\epsilon, \min(1, 2c)/\lambda)^2 \lambda^2}{8c} \right)
\end{aligned}$$

where the second inequality follows because  $M(x) \geq k$  by assumption. Taking the expectation of both sides and the limit as  $N \rightarrow \infty$  completes the proof.  $\square$

We prove one more intermediate result, and the proof of Theorem 3.2 will follow.

**Lemma 3.8.** *Suppose  $h(z; y)$  is an  $L$  Lipschitz continuous function for all  $y$  (with*

respect to  $\|\cdot\|_p$ ), and  $\mathcal{Z} \subset \mathbb{R}^d$  is nonempty, closed, and bounded with diameter  $D$ . It follows that

$$P\left(\sup_{z \in \mathcal{Z}} h(z; Y) > \epsilon\right) \leq \left(\frac{2\rho DL}{\epsilon}\right)^d \sup_{z \in \mathcal{Z}} P\left(h(z; Y) > \frac{\epsilon}{2}\right),$$

where  $\rho > 0$  is a constant that depends only on  $p$ .

*Proof.* This result follows from a standard covering number argument. We can construct a  $\nu$ -net of  $\mathcal{Z}$ ,  $z_1, \dots, z_K$  with  $K \leq \left(\frac{\rho D}{\nu}\right)^d$ . That is,  $\forall z \in \mathcal{Z}$ , there exists  $i$  such that  $\|z - z_i\| \leq \nu$ . If we define  $i(z)$  to be the function that returns this index, then, for all  $y$ ,

$$\begin{aligned} h(z; y) &\leq h(z_{i(z)}; y) + |h(z; y) - h(z_{i(z)}; y)| \leq h(z_{i(z)}; y) + L\|z - z_{i(z)}\| \\ &\leq h(z_{i(z)}; y) + L\nu. \end{aligned}$$

Taking the supremum of both sides over  $z \in \mathcal{Z}$ , we have

$$\sup_{z \in \mathcal{Z}} h(z; y) \leq \max_{i=1, \dots, K} h(z_i; y) + L\nu.$$

Next, we select  $\nu = \frac{\epsilon}{2L}$ , and we have:

$$\begin{aligned} P\left(\sup_{z \in \mathcal{Z}} h(z; Y) > \epsilon\right) &\leq P\left(\max_{i=1, \dots, K} h(z_i; Y) > \frac{\epsilon}{2}\right) \\ &\leq \left(\frac{2\rho DL}{\epsilon}\right)^d \sup_{z \in \mathcal{Z}} P\left(h(z; Y) > \frac{\epsilon}{2}\right), \end{aligned}$$

where the final inequality follows from the union bound.  $\square$

*Proof of Theorem 3.2.* The proof follows the same outline as the proof of Theorem 3.1. We need to show

$$\sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \rightarrow_P 0,$$

for  $x_0$  a.e. The desired result then follows from lemma 3.2. Following the same steps

as in the proof of Theorem 3.1, we have:

$$\begin{aligned}
& \sup_{z_0 \in \mathcal{Z}_0} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \\
& \leq \sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) Q_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \\
& \quad + \sum_{t=1}^{T-1} \sup_{x_t \in \mathcal{X}_t} \sup_{z_t \in W_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z_t); y_{t+1}^i, x_{t+1}^i) \right. \\
& \quad \left. - \mathbb{E}[Q_{t+1}(f_t(z_t); Y_{t+1}, X_{t+1}) | x_t] \right|.
\end{aligned}$$

Next, we have, for all  $x_t$ ,

$$\begin{aligned}
& \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right| \\
& - \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z'); y_{t+1}^i, x_{t+1}^i) - \mathbb{E}[Q_{t+1}(f_t(z'); Y_{t+1}, X_{t+1}) | x_t] \right| \\
& \leq \left| \sum_i w_{N,i}^{t+1}(x_t) (Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) - Q_{t+1}(f_t(z'); y_{t+1}^i, x_{t+1}^i)) \right| \\
& \quad + |\mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) - Q_{t+1}(f_t(z'); Y_{t+1}, X_{t+1}) | x_t]| \\
& \leq 2 \sup_{x_{t+1}, y_{t+1}} |(Q_{t+1}(f_t(z); y_{t+1}, x_{t+1}) - Q_{t+1}(f_t(z'); y_{t+1}, x_{t+1}))| \\
& \leq 2L_{t+1} \|z - z'\|.
\end{aligned}$$

Rearranging, and taking the supremum over both sides, we have

$$\begin{aligned}
& \sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right| \\
& \leq \sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z'); y_{t+1}^i, x_{t+1}^i) \right. \\
& \quad \left. - \mathbb{E}[Q_{t+1}(f_t(z'); Y_{t+1}, X_{t+1}) | x_t] \right| + 2L_{t+1} \|z - z'\|,
\end{aligned}$$

which demonstrates

$$\sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right|$$

is  $2L_{t+1}$  Lipschitz. We now apply lemma 3.8 to get:

$$\begin{aligned} & P \left( \sup_{z_t \in W_t} \sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) \right. \right. \\ & \quad \left. \left. - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right| > \epsilon \right) \\ & \leq \left( \frac{2\rho D_t L_{t+1}}{\epsilon} \right)^{p_t} P \left( \sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) \right. \right. \\ & \quad \left. \left. - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right| > \frac{\epsilon}{2} \right). \end{aligned}$$

By lemma 3.7, the right hand side goes to 0 in probability. Repeating this argument for all  $t$  completes the result.  $\square$

### 3.3.3 Random Forest Weight Functions

A random forest is an ensemble method that aggregates regression trees as base learners in order to make predictions. To aggregate the trees into a random forest, Breiman suggested training each tree on a bootstrapped sample of the training examples. In order to facilitate the theoretical analysis, we instead build a forest by training trees on subsamples of size  $s_N$  of the training data. The random forest estimator is given by:

$$R(x; \xi, X_1, Y_1, \dots, X_N, Y_N) = \frac{1}{B} \sum_{b=1}^B T(x; \xi_b, X_{S_b}, Y_{S_b}), \quad (3.6)$$

where  $(X_{S_b}, Y_{S_b})$  denotes a random subset of training examples of size  $s_N$ . Given this definition of a random forest, we have the following asymptotic optimality result for random forest weight functions.

**Theorem 3.3.** *Suppose Assumptions 3.1-3.6 hold, and the training data is i.i.d. Let  $w_{N,i}^t(x_{t-1})$  be the random forest weight functions for  $t = 1, \dots, T$ . Assume the trees*

that make up the forest are honest, random split, and regular and  $k$ , the minimum number of training examples in each leaf, equals  $\min\{\lceil C_1 N^\delta \rceil, N - 1\}$  for  $C_1 > 0, \delta \in (0, 1)$ . Furthermore, assume  $s_N$ , the subsample size, equals  $\min\{\lceil C_2 N^\alpha \rceil, N - 1\}$  for  $C_2 > 0, \alpha \in (\delta, 1)$ . Then  $\{\hat{z}_0^N(x)\}$ , a sequence of optimal solutions to (3.3), is weakly asymptotically optimal.

*Proof.* The proof exactly mirrors the proof of Theorem 3.2, except we use lemma 3.9 in place of lemma 3.7.  $\square$

This theorem shows it is possible to obtain asymptotically optimal solutions to (3.3) with random forest weight functions. The random forest model we use is slightly different than Breiman's original algorithm, which is implemented in common machine learning libraries. For example, we require that  $k$ , the minimum number of training examples in each leaf, grows with  $N$ , whereas the original algorithm has  $k$  fixed at 1. We include an additional theorem in the appendix which proves the strong asymptotic optimality of random forest weight functions with  $k$  fixed for the single stage version of the problem. However, the proof does not extend to the multistage problem we consider here. The proof of Theorem 3.3 uses the following lemma.

**Lemma 3.9.** *Suppose  $R$  is a random forest consisting of  $B$  regular, random-split, honest trees, each trained on a random subset of the training data of size  $s_N$ . Assume the training data  $(X_1, Y_1), \dots, (X_N, Y_N)$  are i.i.d. with  $X_i$  uniform on  $[0, 1]^d$ ,  $\mathbb{E}[Y|X = x]$  is  $L$ -Lipschitz continuous, and there exists  $\lambda > 0$  such that the uniform noise condition is satisfied:  $\sup \mathbb{E}[\exp(\lambda|Y_i - \mathbb{E}[Y_i|X_i = x]|)|X_i = x] < \infty$ . Finally, suppose  $\log s_N/k_N \rightarrow 0$  and  $s_N/k_N \rightarrow \infty$  as  $N \rightarrow \infty$ . If  $\hat{\mu}_N(x)$  denotes the prediction of  $R$  at  $X \in \mathbb{R}^d$  and  $\mu(x) = \mathbb{E}[Y|X = x]$ , then*

$$\sup_x |\hat{\mu}_N(x) - \mu(x)| \rightarrow_P 0.$$

*Proof.* We define the prediction of the  $b$ th tree in the ensemble to equal  $\hat{\mu}_N^b(x)$ , so we

have, by Jensen's inequality,

$$\sup_x |\hat{\mu}_N(x) - \mu(x)| \leq \frac{1}{B} \sum_{b=1}^B \sup_x |\hat{\mu}_N^b(x) - \mu(x)|.$$

By lemma 3.7, we immediately have that each term on the right hand side goes to 0 in probability. Because  $B$  does not depend on  $N$ , this completes the result.  $\square$

### 3.4 Finite Sample Guarantees

In this section, we establish finite sample, probabilistic guarantees for the difference between the cost of a solution to (3.3) and the true optimal cost. We focus on  $k$ -nearest neighbor weight functions. To the best of our knowledge, this is the first finite sample bound for either the single-stage or multistage setting with auxiliary data.

To facilitate the presentation of our result, we begin by discussing convergence rate results for the single stage setting. Without auxiliary data, the problem we want to solve is given by

$$\min_{z \in \mathcal{Z}} \mathbb{E}[c(z; Y)].$$

If  $\hat{z}_N$  represents the SAA approach applied to this problem, then, under appropriate conditions, the regret,  $\mathbb{E}[c(\hat{z}_N, Y)] - \min_{z \in \mathcal{Z}} \mathbb{E}[c(z; Y)]$ , is  $\tilde{O}_p\left(\frac{1}{\sqrt{N}}\right)$ , where the  $\tilde{O}_p$  notation suppresses logarithmic dependencies (see, for example, [89]). This implies that for any confidence level,  $\alpha$ , we know that the regret is bounded by a term of order  $1/\sqrt{N}$  with probability at least  $1 - \alpha$ . We contrast this with the setting in which we have auxiliary data,

$$\min_{z \in \mathcal{Z}} \mathbb{E}[c(z; Y)|X = x].$$

If  $\hat{z}_N(x)$  represents a solution to this problem using the approach of Bertsimas and Kallus [15] with the  $k$ -nearest neighbor weight functions, then the regret,

$$\mathbb{E}[c(\hat{z}_N(x), Y)|X = x] - \min_{x \in \mathcal{X}} \mathbb{E}[c(z; Y)|X = x] = \tilde{O}_p\left(\frac{1}{N^{1/2d}}\right),$$

for  $d \geq 2$ , where  $d$  is the dimension of the auxiliary covariate space. The problem

with auxiliary data is clearly harder. The baseline with respect to which we compute the regret is smaller because it takes into account the value of the auxiliary covariates. Furthermore, many of the  $x^i$ 's in the training data will be very different from the  $x$  we are concerned with. In fact, with the  $k$ -nearest neighbor weight functions, we effectively throw out all but the  $k$  most relevant training examples. Because of this, we pay a penalty that depends on the dimension of the auxiliary covariate space.

To formalize the above discussion for the multistage setting, we add two additional assumptions.

**Assumption 3.7** (Distribution of auxiliary covariates). *For each  $t$ ,  $X_t$  has its support contained in  $[0, 1]^{d_t}$  and  $\forall x_t \in \text{support}(X_t)$  and  $\forall \epsilon > 0$ ,  $P(X \in B_\epsilon(x_t)) > g\epsilon^d$  with  $g > 0$ , where  $B_\epsilon(x) = \{x' : \|x - x'\| \leq \epsilon\}$ .*

This assumption is satisfied, for example, if  $X_t$  is uniformly distributed or has finite support and, thus, is more general than Assumption 3.5.

**Assumption 3.8** (Subgaussian noise terms). *The noise terms are uniformly subgaussian, i.e., defining  $N_t(s_t; x_{t-1}) = Q_t(s_t; Y_t, X_t) - \mathbb{E}[Q_t(s_t; Y_t, X_t) | X_{t-1} = x_{t-1}]$ , there exists  $\sigma^2 > 0$  such that*

$$\max_{t=1, \dots, T} \sup_{x_{t-1} \in \mathcal{X}_{t-1}} \sup_{s_t \in S_t} \mathbb{E}[e^{\lambda N_t(s_t; x_{t-1})} | X_{t-1} = x_{t-1}] \leq \exp(\lambda \sigma^2 / 2),$$

for all  $\lambda > 0$ .

This assumption implies condition 4 of Assumption 3.4. With these additional assumptions, we have the following theorem.

**Theorem 3.4.** *Suppose Assumptions 3.1-3.4 and Assumptions 3.6-3.8 hold, the training data,  $(X_1, Y_1), \dots, (X_N, Y_N)$ , is i.i.d., and  $w_{N,i}^t(x_t)$  are the  $k_N$  nearest neighbor weight functions with  $k_N = \min\{\lceil CN^\delta \rceil, N - 1\}$  for  $C > 0$  and  $\delta \in (0, 1)$ . We define*

$$R_N(x_0) = g_0(\hat{z}_0^N) + \mathbb{E}[g_0(\hat{z}_0^N) + Q_1(f_0(\hat{z}_0^N); Y_1, X_1) | X_0 = x_0] - v^*(x_0),$$

where  $\hat{z}_0^N$  is an optimal solution to (3.3). For any  $\alpha \in (0, 1)$ , with probability at least

$1 - \alpha$ ,

$$R_N(x_0) \leq \frac{TC_1}{N^{\delta/2}} \left( \sqrt{\log \frac{1}{\alpha}} + \sqrt{2d \log N} + C_2 \right) + TC_3 \left( \left( \frac{1}{N} \log \frac{1}{\alpha} \right)^{1/2d} + C_4 N^{(\delta-1)/d} + \frac{C_5}{N^{1/2d}} \right),$$

for  $N \geq 25d$ , for  $x_0$  almost everywhere. (Here,  $d = \max_t d_t$ .)  $C_1, C_2, C_3, C_4$ , and  $C_5$  are constants that may depend only logarithmically on  $T$  and  $\log \frac{1}{\alpha}$  and are defined in (3.7) in the proof.

$R_N(x_0)$  represents the regret of the solution to (3.3), i.e., the difference between the cost of the solution and the true optimal cost,  $v^*(x_0)$ . We can optimize the bound by choosing  $\delta = \frac{2}{2+d}$  and restate the result more prosaically:

$$R_N(x_0) = \begin{cases} \tilde{O}_p \left( \frac{T}{N^{1/3}} \right), & d = 1, \\ \tilde{O}_p \left( \frac{T}{N^{1/2d}} \right), & d \geq 2. \end{cases}$$

As before, this result is best understood in comparison with the multistage SAA problem without any auxiliary covariates. For this problem, regret is  $\tilde{O}_p \left( \frac{T}{\sqrt{N}} \right)$  [89, ch. 5]. We pay a penalty that depends on  $d$ , the maximum dimension of the auxiliary covariate spaces,  $\mathcal{X}_t$ .

To prove this result, we rely on the following lemma, which provides a finite sample guarantee on the error of the  $k$ NN regression estimator.

**Lemma 3.10.** *Suppose  $X$  has support  $\mathcal{X} \subset [0, 1]^d$ ,  $P(X \in B_\epsilon(x)) > g\epsilon^d$  for all  $x \in \mathcal{X}$ , and  $Y - \mathbb{E}[Y|X = x]$  is conditionally subgaussian given  $X = x$  with variance proxy  $\sigma^2$ , uniformly for all  $x \in \mathcal{X}$ . Assume the training data,  $(X_1, Y_1), \dots, (X_N, Y_N)$  is i.i.d. and that  $\mathbb{E}[Y|X = x]$  is  $L$ -Lipschitz. If  $\hat{\mu}_N(x)$  denotes the  $k$ NN regression*



estimator at  $x$  and  $\mu(x) = \mathbb{E}[Y|X = x]$ , then

$$\begin{aligned} & P\left(\sup_{x \in \mathcal{X}} |\hat{\mu}_N(x) - \mu(x)| > \epsilon\right) \\ & \leq \left(\frac{4\sqrt{d}\rho L}{\epsilon}\right)^d \exp\left(-\frac{2}{N} \left(Ng\left(\frac{\epsilon}{4L}\right)^d + 1 - k\right)^2\right) \\ & \quad + 2\left(\frac{25}{d}\right)^d N^{2d} \exp\left(-\frac{k\epsilon^2}{8\sigma^2}\right), \end{aligned}$$

for any  $\epsilon \geq 2L\left(\frac{k-1}{Ng}\right)^{1/d}$  and  $N \geq 2d$ .

*Proof.* We decompose  $\sup_{x \in \mathcal{X}} |\hat{\mu}_N(x) - \mu(x)|$  into a sum of two terms:  $\sup_{x \in \mathcal{X}} |\hat{\mu}_N(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]|$  and  $\sup_{x \in \mathcal{X}} |\mu(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]|$ . For the latter term, we utilize the Lipschitz assumption to show

$$\sup_{x \in \mathcal{X}} |\mu(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]| \leq L \sup_{x \in \mathcal{X}} \frac{1}{k} \sum_{i=1}^k \|X_{(i)}(x) - x\|,$$

where  $X_{(i)}(x)$  denotes the  $i$ th nearest neighbor of  $x$  out of  $X_1, \dots, X_N$ , as measured by Euclidean distance. Next, we note that  $\frac{1}{k} \sum_{i=1}^k \|X_{(i)}(x) - x\| \leq \|X_{(k)}(x) - x\|$ . This gives:

$$\begin{aligned} & P\left(\sup_{x \in \mathcal{X}} |\mu(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]| > L\epsilon\right) \\ & \leq P\left(\sup_{x \in \mathcal{X}} \|X_{(k)}(x) - x\| > \epsilon\right) \\ & = P\left(\inf_{x \in \mathcal{X}} |\{i : X_i \in B_\epsilon(x)\}| \leq k-1\right). \end{aligned}$$

Next, we construct an  $\epsilon/2$ -net for  $\mathcal{X}$ ,  $\hat{x}_1, \dots, \hat{x}_m$ , with  $m \leq \left(\frac{2\sqrt{d}\rho}{\epsilon}\right)^d$ . For any  $x \in \mathcal{X}$ , there exists a  $j$  such that  $B_{\epsilon/2}(\hat{x}_j) \subset B_\epsilon(x)$ . Therefore, we can upper bound

the above expression by

$$\begin{aligned} & P\left(\min_{j=1,\dots,m} |\{i : X_i \in B_{\epsilon/2}(\hat{x}_j)\}| \leq k-1\right) \\ & \leq \left(\frac{2\sqrt{d}\rho}{\epsilon}\right)^d P(B \leq k-1), \end{aligned}$$

where  $B \sim \text{Binom}(N, g(\epsilon/2)^d)$ . Applying Hoeffding's bound, we have:

$$\begin{aligned} & P\left(\sup_{x \in \mathcal{X}} |\mu(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]| > \epsilon\right) \\ & \leq \left(\frac{2\sqrt{d}\rho L}{\epsilon}\right)^d \exp\left(-\frac{2}{N} (Ng(\epsilon/2L)^d + 1 - k)^2\right), \end{aligned}$$

for any  $\epsilon \geq 2L \left(\frac{k-1}{Ng}\right)^{1/d}$ .

For the second part, we use Theorem 12.2 of Biau and Devroye [25], which says the number of possible distinct orderings of neighbors of  $X_1, \dots, X_N \in \mathbb{R}^d$  is less than or equal to  $\left(\frac{25}{d}\right)^d N^{2d}$  for all  $N \geq 2d$ . Therefore,

$$\begin{aligned} & P\left(\sup_{x \in \mathcal{X}} |\hat{\mu}_N(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]| > \epsilon \mid X_1, \dots, X_N\right) \\ & \leq \left(\frac{25}{d}\right)^d N^{2d} \sup_{x \in \mathcal{X}} P\left(|\hat{\mu}_N(x) - \mathbb{E}[\hat{\mu}_N(x)|X_1, \dots, X_N]| > \epsilon \mid X_1, \dots, X_N\right) \\ & = \left(\frac{25}{d}\right)^d N^{2d} \sup_{x \in \mathcal{X}} P\left(\left|\frac{1}{k} \sum_{i=1}^k (Y_{(i)}(x) - \mu(X_{(i)}(x)))\right| > \epsilon \mid X_1, \dots, X_N\right), \end{aligned}$$

where  $Y_{(i)}$  denotes the observation  $Y$  corresponding to  $X_{(i)}(x)$ . It is easy to verify, see Proposition 8.1 of Biau and Devroye [25] for example, that  $Y_{(1)}(x) - \mu(X_{(1)}(x)), \dots, Y_{(k)}(x) - \mu(X_{(k)}(x))$  are conditionally independent given  $X_1, \dots, X_N$ . Therefore, we apply Hoeffding's bound for sums of subgaussian random variables to get, for any

$\epsilon > 0$ ,

$$\begin{aligned} & P \left( \sup_{x \in \mathcal{X}} |\hat{\mu}_N(x) - \mathbb{E}[\hat{\mu}_N(x) | X_1, \dots, X_N]| > \epsilon \middle| X_1, \dots, X_N \right) \\ & \leq 2 \left( \frac{25}{d} \right)^d N^{2d} \exp \left( -\frac{k\epsilon^2}{2\sigma^2} \right). \end{aligned}$$

Taking the expectation of both sides and combining with the previous part completes the result.  $\square$

Now, we can prove the main result.

*Proof of Theorem 3.4.* By lemma 3.2, the regret is bounded by

$$2 \sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right|.$$

Following the same steps as in the proof of Theorem 3.1, we have:

$$\begin{aligned} & 2 \sup_{z_0 \in \mathcal{Z}_0} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \\ & \leq 2 \sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) Q_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| \\ & \quad + 2 \sum_{t=1}^{T-1} \sup_{x_t \in \mathcal{X}_t} \sup_{z_t \in W_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z_t); y_{t+1}^i, x_{t+1}^i) \right. \\ & \quad \left. - \mathbb{E}[Q_{t+1}(f_t(z_t); Y_{t+1}, X_{t+1}) | x_t] \right|. \end{aligned}$$

We next apply lemma 3.8, as in the proof of Theorem 3.2, to see, for each  $t$ ,

$$\begin{aligned} & P \left( 2 \sup_{z_t \in W_t} \sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) \right. \right. \\ & \quad \left. \left. - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right| > \frac{\epsilon}{T} \right) \\ & \leq \left( \frac{4T\rho D_t L_t}{\epsilon} \right)^{p_t} P \left( \sup_{x_t \in \mathcal{X}_t} \left| \sum_i w_{N,i}^{t+1}(x_t) Q_{t+1}(f_t(z); y_{t+1}^i, x_{t+1}^i) \right. \right. \\ & \quad \left. \left. - \mathbb{E}[Q_{t+1}(f_t(z); Y_{t+1}, X_{t+1}) | x_t] \right| > \frac{\epsilon}{4T} \right). \end{aligned}$$

Next, we use lemma 3.10 to upper bound this expression by

$$\begin{aligned} & \left( \frac{4T\rho D_t L_t}{\epsilon} \right)^{p_t} \left[ \left( \frac{16T\sqrt{d_t}\rho M_t}{\epsilon} \right)^{d_t} \exp \left( -\frac{2}{N} \left( Ng \left( \frac{\epsilon}{16TM_t} \right)^{d_t} + 1 - k_N \right)^2 \right) \right. \\ & \left. + 2 \left( \frac{25}{d_t} \right)^{d_t} N^{2d_t} \exp \left( -\frac{k_N \epsilon^2}{64T^2 \sigma^2} \right) \right], \end{aligned}$$

for any  $\epsilon \geq 8TM_t \left( \frac{k_N - 1}{Ng} \right)^{1/d_t}$  and  $N \geq 2d_t$ . Combining the results for  $t = 0, \dots, T-1$  with the union bound, and plugging in for the definitions of  $k_N$ ,  $d$ ,  $p$ ,  $L$ ,  $M$ , and  $D$ , we have:

$$\begin{aligned} & P \left( 2 \sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| > \epsilon \right) \\ & \leq T \left( \frac{4T\rho DL}{\epsilon} \right)^p \left[ \left( \frac{16T\sqrt{d}\rho M}{\epsilon} \right)^d \exp \left( -\frac{2}{N} \left( Ng \left( \frac{\epsilon}{16TM} \right)^d + 1 - CN^\delta \right)^2 \right) \right. \\ & \left. + 2 \left( \frac{25}{d} \right)^d N^{2d} \exp \left( -\frac{CN^\delta \epsilon^2}{64T^2 \sigma^2} \right) \right], \end{aligned}$$

for all  $\epsilon \geq 8TM \left( \frac{CN^\delta - 1}{Ng} \right)^{1/d}$  and  $N \geq 2d$ . From this we deduce that for any  $\alpha \in (0, 1)$ ,

$$\begin{aligned} & P \left( 2 \sup_{z_0 \in \mathcal{Z}} \left| \sum_i w_{N,i}^1(x_0) \hat{Q}_1(f_0(z_0); y_1^i, x_1^i) - \mathbb{E}[Q_1(f_0(z_0); Y_1, X_1) | X_0 = x_0] \right| > \epsilon \right) \\ & \leq \alpha \end{aligned}$$

is implied by the following system of inequalities:

$$\begin{aligned}
\frac{CN^\delta \epsilon^2}{64T^2 \sigma^2} &\geq \log \frac{1}{\alpha} + 2d \log N + p \log \frac{1}{\epsilon} + \log \left( 4T \left( \frac{25}{d} \right)^d (4T\rho DL)^p \right), \\
2 \left( \sqrt{N}g \left( \frac{\epsilon}{16TM} \right)^d - CN^{\delta-1/2} \right)^2 &\geq \log \frac{1}{\alpha} + (p+d) \log \frac{1}{\epsilon} \\
&\quad + \log(2T(4T\rho DL)^p(16T\sqrt{d}\rho M)^d), \\
\epsilon &\geq 8TM \left( \frac{CN^{\delta-1}}{g} \right)^{1/d}.
\end{aligned}$$

Following some algebraic manipulations, we see the above system of inequalities is implied by:

$$\begin{aligned}
\epsilon &\geq \frac{TC_1}{N^{\delta/2}} \left( \sqrt{\log \frac{1}{\alpha}} + \sqrt{2d \log N} + C_2 \right) \\
&\quad + TC_3 \left( \left( \frac{1}{N} \log \frac{1}{\alpha} \right)^{1/2d} + C_4 N^{(\delta-1)/d} + \frac{C_5}{N^{1/2d}} \right),
\end{aligned}$$

where

$$\begin{aligned}
C_1 &= \frac{8\sigma}{\sqrt{C}}, \tag{3.7} \\
C_2 &= \sqrt{\log \left( 4T \left( \frac{25}{d} \right)^d (4T\rho DL)^p \right)} + \sqrt{\frac{p\delta}{2} \log \frac{2}{TC_1 \sqrt{\log \frac{1}{\alpha}}}}, \\
C_3 &= \frac{16M}{(\sqrt{2}g)^{1/d}}, \\
C_4 &= \left( \sqrt{2}C \right)^{1/d} + \frac{8M}{C_3} \left( \frac{C}{g} \right)^{1/d}, \\
C_5 &= \left( \log(2T(4T\rho DL)^p(16T\sqrt{d}\rho M)^d) \right)^{1/2d} + \left( (p+d)\delta \log \frac{2}{TC_1 \sqrt{\log \frac{1}{\alpha}}} \right)^{1/2d}.
\end{aligned}$$

□

## 3.5 Computational Examples

In this section, we illustrate the practical applicability of our approach with two examples using synthetic data. These examples also serve to demonstrate the value of accounting for auxiliary data.

### 3.5.1 Multistage Inventory Control

First, we consider a multistage inventory control problem [14], in which we manage the inventory level of a single product subject to  $T$  periods of uncertain demand. At each time step, we observe auxiliary data, which may include data about the product as well as data on time-varying external factors that may be used to predict demand, such as the season of the year, the price of the S&P 500 index, or the demand for a similar product during the previous time period. We also have historical data of the demand for products we've sold in the past as well as the corresponding auxiliary data for each of these products.

At the beginning of time period  $t$ , we observe the demand  $y_t$  and the new auxiliary covariates  $x_t$ . Demand can be served by ordering  $z_t^2$  units at price  $c_2$  for immediate delivery or by ordering  $z_t^1$  units at price  $c_1 < c_2$  for delivery at the beginning of the next time period. If there is a shortfall in the inventory, orders can be backlogged, incurring a cost of  $-c_b$  per unit. If there is excess inventory at the end of a time period, we pay a holding cost of  $c_h$  per unit. In addition, there is an ordering budget, so the cumulative advance orders ( $\sum_{s \leq t} z_s^1$ ) must not exceed  $\bar{z}_{\text{tot},t}$  at any time  $t$ . We assume all ordering and inventory quantities are continuous, and we represent the amount of inventory at the end of time period  $t$  by  $I_t$  (with  $I_{-1} = 0$  and  $z_{-1}^1 = 0$ ). If demand is known, we have the following deterministic formulation of the problem.

$$\begin{aligned}
\min \quad & \sum_{t=0}^T c_1 z_t^1 + c_2 z_t^2 + \max\{c_b I_t, c_h I_t\} \\
\text{s.t.} \quad & I_{t+1} = I_t + z_t^1 + z_{t+1}^2 - y_{t+1} \quad \forall t = -1, \dots, T-1 \\
& \sum_{s=0}^t z_s^1 \leq \bar{z}_{\text{tot},t} \quad \forall t = 0, \dots, T \\
& z_t^1, z_t^2 \geq 0 \quad \forall t = 0, \dots, T.
\end{aligned}$$

We used the parameters  $c_1 = 5$ ,  $c_2 = 10$ ,  $c_h = 5$ , and  $c_b = -10$ . We assumed the initial inventory to be 0 and set  $\bar{z}_{\text{tot},t} = 50(t+1)$ . To generate training data, we sampled  $x_t^i$  independently from a 3 dimensional AR(1) process such that  $x_t^i = 0.7x_{t-1}^i + w_t^i$ , where  $w_t^i$  is a sample of a  $\mathcal{N}(0, I_3)$  random variable. We used a factor model for the demand.

$$y_t^i = \max\{0, 50 + 12a_t^T(x_{t-1}^i + 0.25\phi_t) + 5b_t^T x_{t-1}^i \theta_t\} \quad \forall t = 1, \dots, T$$

where  $\{\phi_t\}$  are drawn independently from a 3 dimensional standard Gaussian and  $\{\theta_t\}$  are drawn independently from a 1 dimensional standard Gaussian. At each time step, the factor loadings,  $a_t$  and  $b_t$ , are permutations of  $\begin{pmatrix} 0.8 & 1 & 1 \end{pmatrix}^T$  and  $\begin{pmatrix} -1 & 1 & 0 \end{pmatrix}^T$ , respectively (held constant for all samples). The results we present here show the average cost of policies based on out-of-sample testing as a function of the amount of training observations,  $N$ . All results are averaged over one hundred realizations of training sets.

Figure 3-1 shows the expected cost of policies learned using our method versus the number of training samples. We see that the SAA approach, which ignores the auxiliary data, is suboptimal. Our method, using the  $k$ -nearest neighbor and random forest weight functions, is asymptotically optimal. We obtain a reduction in cost of over 15% by accounting for auxiliary data.

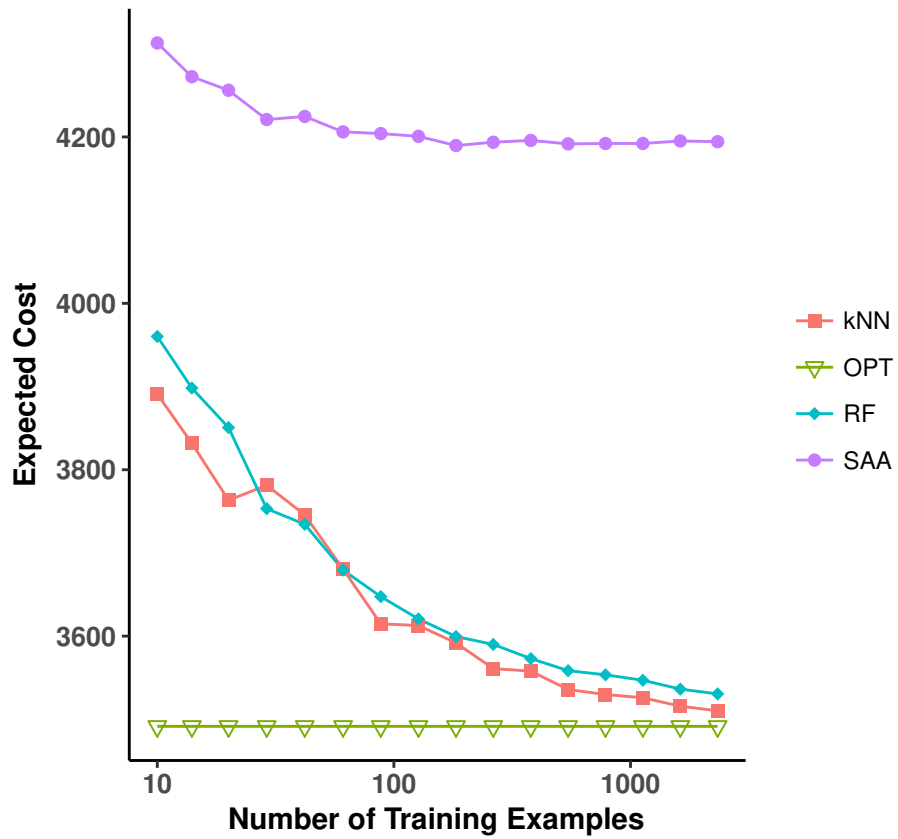


Figure 3-1: Out of sample results with various weight functions for a twelve stage inventory control problem. Vertical axis represents expected cost of policy (smaller is better).



### 3.5.2 Multistage Lot Sizing

For our second computational example, we consider a multistage lot sizing problem [14]. This problem is similar to the multistage inventory control problem, but it includes binary decision variables. The continuous ordering decision for immediate delivery,  $z_t^2$ , is replaced with  $M$  binary ordering decisions,  $z_{tj}^2$ ,  $j = 1, \dots, M$ . Each of these decisions corresponds to a quantity,  $q_j$ , which is delivered immediately for cost  $c_{2j} > c_1$  per unit. Additionally, there is no longer the option to backorder demand. All demand must be satisfied immediately. These restrictions make the problem more realistic because it is often not feasible to produce an arbitrary amount of a product immediately, and it is difficult to estimate the cost of lost customer goodwill due to backordering. Instead, in order to meet demand, the decision maker must buy from another supplier a fixed quantity of product at a higher price.

If demand is known, we have the following deterministic formulation (where we assume  $I_{-1} = 0$ ,  $z_{-1}^1 = 0$ , and  $y_0 = 0$ ).

$$\begin{aligned}
 \min \quad & \sum_{t=0}^T c_1 z_t^1 + \sum_{j=1}^M c_{2j} q_j z_{tj}^2 + c_h I_t \\
 \text{s.t.} \quad & I_{t+1} = I_t + z_t^1 + \sum_{j=1}^M c_{2j} q_j z_{tj}^2 + c_h I_t \\
 & \sum_{s=0}^t z_s^1 \leq \bar{z}_{\text{tot},t} \quad \forall t = 0, \dots, T \\
 & z_t^1, I_t \geq 0 \quad \forall t = 0, \dots, T \\
 & z_{tj}^2 \in \{0, 1\} \quad \forall t = 0, \dots, T \quad \forall j = 1, \dots, M.
 \end{aligned}$$

We used the same parameters and data generating procedure as in the multistage inventory control example. The only differences were that we capped  $y_t$  at 200 (to ensure feasibility) and we drew  $c_{2j}$  independently from a uniform distribution on  $(5, 10)$  for each  $j$ .

To solve the problem, we use an approximate DP algorithm. To develop the algorithm, we recall that basestock policies are optimal for a wide variety of inventory

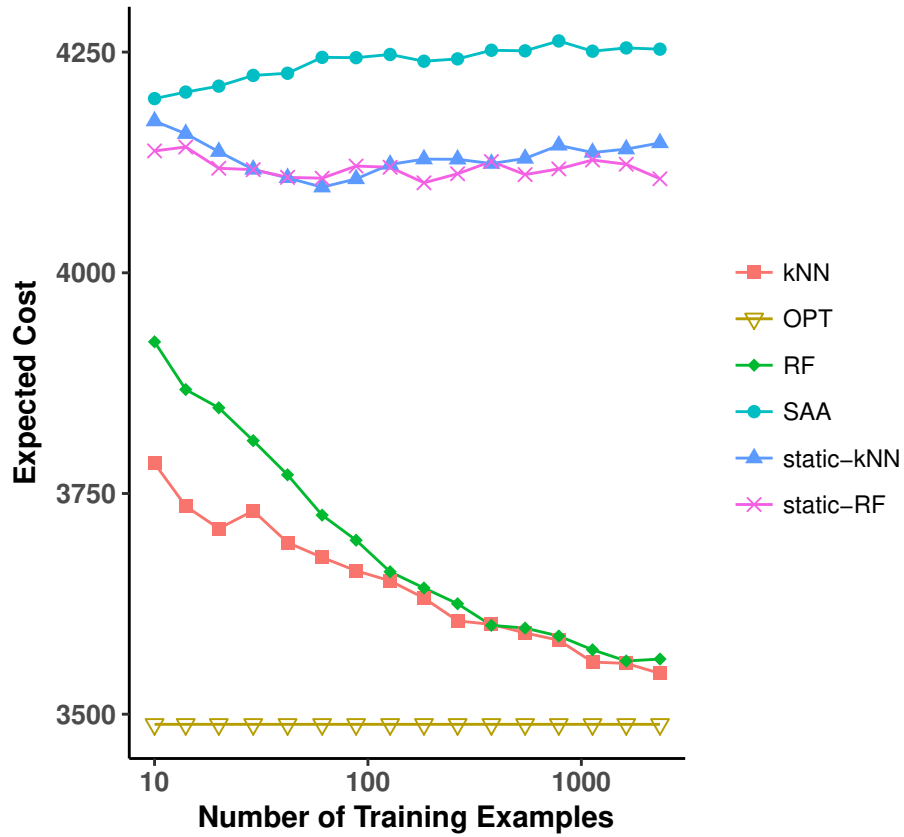


Figure 3-2: Average out-of-sample cost of policies computed using various weight functions for twelve stage lot sizing problem. Horizontal axis shows number of training examples.

control problems. A basestock policy is one in which there is some ideal amount of inventory,  $r_t$ , which we desire at the start of time period  $t$ . If we have less than  $r_t$ , we place advanced orders (if available) to have this amount. If we have more than  $r_t$ , we order nothing in advance. We then serve the remaining demand with immediate orders. A basestock policy will not be optimal for the lot sizing problem because of the nonconvexity of the value function, but it does provide a reasonable approximation. To account for auxiliary data, we have  $N$  different basestock amounts for each time period. Therefore,  $r_t^i$  denotes the target basestock at time  $t$  when the observed auxiliary data is  $x_t^i$ . Parametrizing the policy space with  $NT$  parameters greatly reduces the amount of computation required to solve the problem and allows us to solve much larger problem instances than we could otherwise.

Figure 3-2 shows the results of our method for the SAA,  $k$ -nearest neighbor, and random forest weight functions on a twelve stage lot sizing problem. The SAA method is again clearly suboptimal. We see that the static  $k$ NN and static RF methods, which only use the auxiliary covariates at time 0, offer a significant improvement over the SAA method. However, the  $k$ NN and RF methods which take into account the auxiliary data that arrives over time outperforms all of the above methods, again illustrating the value of auxiliary data. With very little additional computational cost, we are able to obtain an improvement in cost of nearly 15%.

## 3.6 Conclusion

In this chapter, we introduced a data-driven framework for solving multistage optimization problems under uncertainty with auxiliary covariates. We demonstrated how to develop specific methods by integrating predictive machine learning methods such as  $k$ NN, CART, and random forests. Our approach is well suited for multistage optimization problems in which the distribution of the uncertainties is unknown, but samples of the uncertainty and auxiliary data are available.

We demonstrated that our method with the  $k$ -nearest neighbor, CART, and random forest weight functions are asymptotically optimal. We also provided finite sam-

ple guarantees for the method with  $k$ NN weight functions. Additionally, we showed how to apply the framework with two computational examples. Because we can think of (3.3) as a dynamic programming problem, we have at our disposal a variety of exact and approximate solution techniques. The problem is often tractable in practice and can lead to significant improvements over methods that ignore auxiliary data.

We leave for future work the extension in which the decision affects the distribution of the uncertainty. This type of problem appears in applications such as pricing where the choice of price affects the distribution of the demand. We also leave for future research the development of efficient variants of our methods for specific applications. In a world in which the availability of data continues to grow, our proposed approach utilizes this data efficiently and has the potential to make a significant impact in OR applications.

# Chapter 4

## Sample Robust Optimization with Covariates

In this chapter, we present a data-driven framework for incorporating side information in dynamic optimization under uncertainty. Specifically, our approach uses predictive machine learning methods (such as  $k$ -nearest neighbors, kernel regression, and random forests) to weight the relative importance of various data-driven uncertainty sets in a robust optimization formulation. Through a novel measure concentration result for local machine learning methods, we prove that the proposed *sample robust optimization with covariates* framework is asymptotically optimal for stochastic dynamic optimization with covariates. We also describe a general-purpose approximation for these optimization problems, based on overlapping linear decision rules, which is computationally tractable and produces high-quality solutions for dynamic problems with many stages. Across a variety of examples in shipment planning, inventory management, and finance, our method achieves improvements of up to 15% over alternatives and requires less than one minute of computation time on problems with twelve stages.

## 4.1 Introduction

Dynamic decision making under uncertainty forms the foundation for numerous fundamental problems in operations research and management science. In these problems, a decision maker attempts to minimize an uncertain objective over time, as information incrementally becomes available. For example, consider a retailer with the goal of managing the inventory of a new short life cycle product. Each week, the retailer must decide an ordering quantity to replenish its inventory. Future demand for the product is unknown, but the retailer can base its ordering decisions on the remaining inventory level, which depends on the realized demands in previous weeks. A risk-averse investor faces a similar problem when constructing and adjusting a portfolio of assets in order to achieve a desirable risk-return tradeoff over a horizon of many months. Additional examples abound in energy planning, airline routing, and ride sharing, as well as in many other areas.

To make high quality decisions in dynamic environments, the decision maker must accurately model future uncertainty. Often, practitioners have access to *side information* or *auxiliary covariates*, which can help predict that uncertainty. For a retailer, although the future demand for a newly introduced clothing item is unknown, data on the brand, style, and color of the item, as well as data on market trends and social media, can help predict it. For the risk-averse investor, while the returns of the assets in future stages are uncertain, recent asset returns and prices of relevant options can provide crucial insight into upcoming volatility. Consequently, organizations across many industries are continuing to prioritize the use of predictive analytics in order to leverage vast quantities of data to understand future uncertainty and make better operational decisions.

A recent body of work has aimed to leverage predictive analytics in decision making under uncertainty. For example, Ban and Rudin [3], Bertsimas and Kallus [15], Hannah, Powell, and Blei [59], Ho and Hanasusanto [61] investigate prescriptive approaches, based on sample average approximation, that use machine learning to assign weights to the historical data based on covariates. Bertsimas and Van Parys

[19] propose adding robustness to those weights to achieve optimal asymptotic budget guarantees. Elmachtoub and Grigas [47] develop an approach for linear optimization problems in which a machine learning model is trained to minimize the decision cost. All of these approaches are specialized for single-stage or certain two-stage optimization problems, and do not immediately generalize to problems with many stages. For a class of dynamic inventory problems, Ban, Gallien, and Mersereau [4] propose a data-driven approach by fitting the stochastic process and covariates to a parametric regression model, which is asymptotically optimal when the model is correctly specified. Bertsimas and McCord [18] propose a different approach based on dynamic programming to use nonparametric machine learning methods to handle auxiliary covariates. However, these dynamic approaches require scenario tree enumeration and suffer from the curse of dimensionality. To the best of our knowledge, no previous work leverages machine learning in a computationally tractable, data-driven framework for decision making in dynamic environments with covariates.

Recently, Bertsimas, Shtern, and Sturt [22] developed a data-driven approach for dynamic optimization under uncertainty that they call *sample robust optimization* (SRO). Their SRO framework constructs a robust optimization problem in which an uncertainty set is constructed around each historical sample path. They show this data-driven framework enjoys nonparametric out-of-sample performance guarantees for a class of dynamic linear optimization problems without covariates and show that this framework can be approximated using decision rule techniques from robust optimization. Furthermore, Bertsimas, Shtern, and Sturt [24] show that a class of *two-stage* sample robust optimization problems can be tractably approximated to near optimality using overlapping linear decision rules.

### 4.1.1 Contributions

In this chapter, we present a new framework for leveraging machine learning in sample robust optimization. Specifically, we propose combining the predictive analytics approach of Bertsimas and Kallus [15] with the SRO framework to incorporate auxiliary covariates into dynamic optimization. Through a new measure concentration

result for local machine learning methods, we show that the proposed *sample robust optimization with covariates* framework is asymptotically optimal, providing the assurance that the resulting decisions are nearly optimal in the presence of big data. We also demonstrate the tractability of the approach by introducing a multi-policy approximation for dynamic optimization problems with many stages. To the best of our knowledge, our method is the first nonparametric approach for tractably solving dynamic optimization problems with covariates, offering practitioners a general-purpose tool for better decision making with predictive analytics. We summarize our main contributions as follows:

- We present a general-purpose framework for leveraging machine learning in data-driven dynamic optimization with covariates. Our approach extends the sample robust optimization framework by assigning weights to the uncertainty sets based on covariates. The weights are computed using machine learning methods such as  $k$ -nearest neighbor regression, kernel regression, and random forest regression. By using all available data, we show that our method produces decisions that achieve improved out-of-sample performance.
- We provide theoretical justification for the proposed framework in the big data setting. First, we develop a new measure concentration result for local machine learning methods (Theorem 4.2), which shows that the weighted empirical distribution produced by local predictors converges quickly to the true conditional distribution. To the best of our knowledge, such a result for local machine learning is the first of its kind. We use this result to establish that the proposed framework is asymptotically optimal for dynamic optimization with covariates (Theorem 4.1) without any parametric assumptions.
- To find high quality solutions in practical computation times, we introduce a multi-policy approximation scheme for dynamic optimization problems with many stages. Specifically, we propose using separate linear decision rules for each uncertainty set to approximate the costs incurred in each stage. We show that the approximation is computationally tractable, both with respect to the



number of stages and size of the historical dataset, and results in high-quality solutions.

- In a variety of examples (shipment planning, inventory management, and finance), across a variety of time horizons, our proposed method outperforms alternatives, in a statistically significant manner, achieving up to 15% improvement in average out-of-sample cost. Moreover, our algorithm is practical and scalable, requiring less than one minute of computation time on examples with up to twelve stages.

The chapter is organized as follows. Section 4.2 introduces the problem setting and notation. Section 4.3 proposes the new framework for incorporating machine learning into dynamic optimization. Section 4.4 develops theoretical guarantees on the proposed framework. Section 4.5 presents the multi-policy approximation scheme for the proposed framework. Section 4.6 presents a detailed investigation and computational simulations of the proposed methodology in finance, shipment planning, and inventory management. We conclude in Section 4.7.

### 4.1.2 Comparison to Related Work

This chapter follows a recent body of literature on data-driven optimization under uncertainty in operations research and management science. Much of this work has focused on the paradigm of distributionally robust optimization, in which the optimal solution is that which performs best in expectation over a worst-case probability distribution from an ambiguity set. Motivated by probabilistic guarantees, distributionally robust optimization has found particular applicability in data-driven settings in which the ambiguity set is constructed using historical data, such as Delage and Ye [42], Esfahani and Kuhn [48], Van Parys, Esfahani, and Kuhn [94], Xu, Caramanis, and Mannor [99]. In particular, the final steps in our convergence result (Section 4.4.4) draw heavily from similar techniques from Esfahani and Kuhn [48] and Bertsimas, Shtern, and Sturt [22]. In contrast to previous work, this chapter develops a new measure concentration result for the weighted empirical distribution (Section 4.4.3)

which enables machine learning and covariates to be incorporated into sample robust optimization and Wasserstein-based distributionally robust optimization for the first time.

Several recent papers have focused on tractable approximations of two- and multi-stage *distributionally* and *sample* robust optimization. Many approaches are based around policy approximation schemes, including lifted linear decision rules [23],  $K$ -adaptivity [58], and finite adaptability [22]. Alternative approaches include tractable approximations of copositive formulations [57]. Closest related to the approximation scheme in this chapter are Chen, Sim, and Xiong [36] and Bertsimas, Shtern, and Sturt [24], which address two-stage problems via overlapping decision rules. Chen, Sim, and Xiong [36] propose a *scenario-wise* modeling approach that leads to novel approximations of various distributionally robust applications, including two-stage distributionally robust optimization using Wasserstein ambiguity sets and expectations of piecewise convex objective functions in single-stage problems. Independently, Bertsimas, Shtern, and Sturt [24] investigate a multi-policy approximation of two-stage sample robust optimization that introduces a separate linear decision rule for each uncertainty set and prove that the approximation gap converges to zero as the amount of data goes to infinity. In Section 4.5 of this chapter, we show how to extend similar techniques to dynamic problems with many stages.

As discussed previously, the methodology in this chapter also follows recent work on incorporating covariates in optimization under uncertainty using local predictive methods (such as  $k$ -nearest neighbor regression, kernel regression, and random forests). In particular, the asymptotic optimality justification of Bertsimas and Kallus [15] in single-stage settings relies on the strong universal consistency for local predictive models (*e.g.*, Walk [98]). Our proof of asymptotic optimality instead relies on convergence guarantees rooted in distributionally robust optimization. The reason we use a different approach is that the arguments for the convergence for local predictive models from Bertsimas and Kallus [15] require finite dimensional decision variables. In contrast, the convergence guarantees in this chapter apply for dynamic optimization over general spaces of policies.

## 4.2 Problem Setting

We consider finite-horizon discrete-time stochastic dynamic optimization problems. The uncertain quantities observed in each stage are denoted by random variables  $\boldsymbol{\xi}_1 \in \Xi_1 \subseteq \mathbb{R}^{d_\xi^1}, \dots, \boldsymbol{\xi}_T \in \Xi_T \subseteq \mathbb{R}^{d_\xi^T}$ . The decisions made in each stage are denoted by  $\mathbf{x}_1 \in \mathcal{X}_1 \subseteq \mathbb{R}^{d_x^1}, \dots, \mathbf{x}_T \in \mathcal{X}_T \subseteq \mathbb{R}^{d_x^T}$ . Given realizations of the uncertain quantities and decisions, we incur a cost of

$$c(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T, \mathbf{x}_1, \dots, \mathbf{x}_T) \in \mathbb{R}.$$

A decision rule  $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_T)$  is a collection of measurable functions  $\boldsymbol{\pi}_t : \Xi_1 \times \dots \times \Xi_{t-1} \rightarrow \mathcal{X}_t$  which specify what decision to make in stage  $t$  based of the information observed up to that point. Given realizations of the uncertain quantities and choice of decision rules, the resulting cost is

$$c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T, \cdot) := c(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T, \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_T(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{T-1})).$$

Before selecting the decision rules, we observe auxiliary covariates  $\boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^{d_\gamma}$ . For example, in the aforementioned fashion setting, the auxiliary covariates contain information on the brand, style, and color of a new clothing item and the uncertainties represent the demand for the product in each week of the item's lifecycle. Given a realization of the covariates  $\boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}}$ , our goal is to find decision rules which minimize the conditional expected cost:

$$v^*(\bar{\boldsymbol{\gamma}}) := \min_{\boldsymbol{\pi} \in \Pi} \mathbb{E} \left[ c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \middle| \boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}} \right]. \quad (4.1)$$

We refer to (4.1) as *dynamic optimization with covariates*. The optimization takes place over a collection  $\Pi$  which is any subset of the space of all non-anticipative decision rules.

In this chapter, we assume that the joint distribution of the covariates and uncertain quantities  $(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)$  is unknown, and our knowledge consists of historical

data of the form

$$(\boldsymbol{\gamma}^1, \boldsymbol{\xi}_1^1, \dots, \boldsymbol{\xi}_T^1), \dots, (\boldsymbol{\gamma}^N, \boldsymbol{\xi}_1^N, \dots, \boldsymbol{\xi}_T^N),$$

where each of these tuples consists of a realization of the auxiliary covariates and the ensuing realizations of the uncertainties over the stages. For example, in the fashion setting, each tuple corresponds to the covariates of a past fashion item as well as its demand over its lifecycle. We do not assume any parametric structure on the relationship between the covariates and future uncertainty.

The goal of this chapter is a general-purpose, computationally tractable, data-driven approach for approximately solving dynamic optimization with covariates. In the following sections, we propose and analyze a new framework which leverages non-parametric machine learning, trained on historical data, to predict future uncertainty from covariates in a way that leads to near-optimal decision rules to (4.1).

### 4.2.1 Notation

The joint probability distribution of the covariates  $\boldsymbol{\gamma}$  and uncertain quantities  $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)$  is denoted by  $\mathbb{P}$ . For the purpose of proving theorems, we assume throughout this chapter that the historical data are independent and identically distributed (i.i.d.) samples from this distribution  $\mathbb{P}$ . In other words, we assume that the historical data satisfies

$$((\boldsymbol{\gamma}^1, \boldsymbol{\xi}^1), \dots, (\boldsymbol{\gamma}^N, \boldsymbol{\xi}^N)) \sim \mathbb{P}^N,$$

where  $\mathbb{P}^N := \mathbb{P} \times \dots \times \mathbb{P}$  is the product measure. The set of all probability distributions supported on  $\Xi := \Xi_1 \times \dots \times \Xi_T \subseteq \mathbb{R}^{d_\xi}$  is denoted by  $\mathcal{P}(\Xi)$ . For each of the covariates  $\bar{\boldsymbol{\gamma}} \in \Gamma$ , we assume that its conditional probability distribution satisfies  $\mathbb{P}_{\bar{\boldsymbol{\gamma}}} \in \mathcal{P}(\Xi)$ , where  $\mathbb{P}_{\bar{\boldsymbol{\gamma}}}(\cdot)$  is shorthand for  $\mathbb{P}(\cdot \mid \boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}})$ . We sometimes use subscript notation for expectations to specify the underlying probability distribution; for example, the

following two expressions are equivalent:

$$\mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}} [f(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)] \equiv \mathbb{E} [f(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mid \gamma = \bar{\gamma}].$$

### 4.3 Sample Robust Optimization with Covariates

In this section, we present our approach for incorporating machine learning in dynamic optimization. We first review the sample robust optimization framework of Bertsimas, Shtern, and Sturt [22], and then we introduce our new *sample robust optimization with covariates* framework.

#### 4.3.1 Preliminary: Sample Robust Optimization

Consider a stochastic dynamic optimization problem of the form (4.1) in which there are no auxiliary covariates. The underlying joint distribution of the random variables  $\boldsymbol{\xi} \equiv (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)$  is unknown, but we have data consisting of sample paths,  $\boldsymbol{\xi}^1 \equiv (\boldsymbol{\xi}_1^1, \dots, \boldsymbol{\xi}_T^1), \dots, \boldsymbol{\xi}^N \equiv (\boldsymbol{\xi}_1^N, \dots, \boldsymbol{\xi}_T^N)$ . For this setting, Bertsimas, Shtern, and Sturt [22] propose using sample robust optimization to find approximate solutions in stochastic dynamic optimization. To apply their framework, one constructs an uncertainty set around each sample path in the training data and then chooses decision rules that optimize the average of the worst-case realizations of the cost. Formally, this framework results in the following robust optimization problem:

$$\min_{\boldsymbol{\pi} \in \Pi} \sum_{i=1}^N \frac{1}{N} \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T), \quad (4.2)$$

where  $\mathcal{U}_N^i \subseteq \Xi$  is an uncertainty set around  $\boldsymbol{\xi}^i$ . Intuitively speaking, (4.2) chooses the decision rules by averaging over the historical sample paths which are adversarially perturbed. Under mild probabilistic assumptions on the underlying joint distribution and appropriately constructed uncertainty sets, the authors prove that sample robust optimization converges asymptotically to the underlying stochastic problem. They also show that (4.2) is amenable to approximations similar to dynamic robust

optimization.

### 4.3.2 Incorporating covariates into sample robust optimization

We now present our new framework, based on sample robust optimization, for solving dynamic optimization with covariates. In the proposed framework, we first train a machine learning algorithm on the historical data to predict future uncertainty  $(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)$  as a function of the covariates. From the trained learner, we obtain weight functions  $w_N^i(\bar{\boldsymbol{\gamma}})$ , for  $i = 1, \dots, N$ , each of which captures the relevance of the  $i$ th training sample to the new covariates,  $\bar{\boldsymbol{\gamma}}$ . We incorporate the weights into sample robust optimization by multiplying the cost associated with each training example by the corresponding weight function. The resulting *sample robust optimization with covariates* framework is as follows:

$$\hat{v}^N(\bar{\boldsymbol{\gamma}}) := \min_{\boldsymbol{\pi} \in \Pi} \sum_{i=1}^N w_N^i(\bar{\boldsymbol{\gamma}}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T), \quad (4.3)$$

where the uncertainty sets are defined

$$\mathcal{U}_N^i := \{ \boldsymbol{\zeta} \in \Xi : \|\boldsymbol{\zeta} - \boldsymbol{\xi}^i\| \leq \epsilon_N \},$$

and  $\|\cdot\|$  is some  $\ell_p$  norm with  $p \geq 1$ .

The above framework provides the flexibility for the practitioner to construct weights from a variety of machine learning algorithms. We focus on weight functions which come from nonparametric machine learning methods. Examples of viable predictive models include  $k$ -nearest neighbors (kNN), kernel regression, classification and regression trees (CART), or random forests (RF). We describe these four classes of weight functions.

**Definition 4.1.** *The  $k$ -nearest neighbor weight functions are given by:*

$$w_{N,kNN}^i(\bar{\gamma}) := \begin{cases} \frac{1}{k_N}, & \text{if } \gamma^i \text{ is a } k_N\text{-nearest neighbor of } \bar{\gamma} \text{ out of } \{\gamma^1, \dots, \gamma^N\}, \\ 0, & \text{otherwise.} \end{cases}$$

For technical details, we refer the reader to Biau and Devroye [25].

**Definition 4.2.** *The kernel regression weight functions are given by:*

$$w_{N,KR}^i(\bar{\gamma}) := \frac{K(\|\gamma^i - \bar{\gamma}\|/h_N)}{\sum_{j=1}^N K(\|\gamma^j - \bar{\gamma}\|/h_N)},$$

where  $K(\cdot)$  is the kernel function and  $h_N$  is the bandwidth parameter. Examples of kernel functions include the Gaussian kernel,  $K(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$ , the triangular kernel,  $K(u) = (1 - u)\mathbb{1}\{u \leq 1\}$ , and the Epanechnikov kernel,  $K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}\{u \leq 1\}$ . For more information on kernel regression, see Friedman, Hastie, and Tibshirani [52, Chapter 6].

The next two types of weight functions we present are based on classification and regression trees [30] and random forests [29]. We refer the reader to Bertsimas and Kallus [15] for technical implementation details.

**Definition 4.3.** *The classification and regression tree weight functions are given by:*

$$w_{N,CART}^i(\bar{\gamma}) := \begin{cases} \frac{1}{|l^N(\bar{\gamma})|}, & i \in l^N(\bar{\gamma}), \\ 0, & \text{otherwise,} \end{cases}$$

where  $l^N(\bar{\gamma})$  is the set of indices  $i$  such that  $\gamma^i$  is contained in the same leaf of the tree as  $\bar{\gamma}$ .

**Definition 4.4.** *The random forest weight functions are given by:*

$$w_{N,RF}^i(\bar{\gamma}) := \frac{1}{B} \sum_{b=1}^B w_{N,CART}^{i,b}(\bar{\gamma}),$$

where  $B$  is the number of trees in the ensemble, and  $w_{N,CART}^{i,b}(\bar{\gamma})$  refers to the weight function of the  $b$ th tree in the ensemble.

All of the above weight functions come from nonparametric machine learning methods. They are highly effective as predictive methods because they can learn complex relationships between the covariates and the response variable without requiring the practitioner to state an explicit parametric form. Similarly, as we prove in Section 4.4, solutions to (4.3) with these weight functions are asymptotically optimal for (4.1) without any parametric restrictions on the relationship between  $\gamma$  and  $\xi$ . In other words, incorporating covariates into sample robust optimization via (4.3) leads to better decisions asymptotically, even without specific knowledge of how the covariates affect the uncertainty.

## 4.4 Asymptotic Optimality

In this section, we establish asymptotic optimality guarantees for sample robust optimization with auxiliary covariates. We prove that, under mild conditions, (4.3) converges to (4.1) as the number of training samples goes to infinity. Thus, as the amount of data grows, sample robust optimization with covariates becomes an optimal approximation of the underlying stochastic dynamic optimization problem. Crucially, our convergence guarantees do not require parametric restrictions on the space of decision rules (*e.g.*, linearity) or parametric restrictions on the joint distribution of the covariates and uncertain quantities. These theoretical results are consistent with empirical experiments in Section 4.6.

### 4.4.1 Main result

We begin by presenting our main result. The proof of the result depends on some technical assumptions and concepts from distributionally robust optimization. For simplicity, we defer the statement and discussion of technical assumptions regarding the underlying probability distribution and cost until Sections 4.4.3 and 4.4.4, and



first discuss what is needed to apply the method in practice. The practitioner needs to select a weight function, parameters associated with that weight function, and the radius,  $\epsilon_N$ , of the uncertainty sets. While these may be selected by cross validation, we show that the method will in general converge if the parameters are selected to satisfy the following:

**Assumption 4.1.** *The weight functions and uncertainty set radius satisfy one of the following:*

1.  $\{w_N^i(\cdot)\}$  are  $k$ -nearest neighbor weight functions with  $k_N = \min(\lceil k_3 N^\delta \rceil, N - 1)$  for constants  $k_3 > 0$  and  $\delta \in (\frac{1}{2}, 1)$ , and  $\epsilon_N = \frac{k_1}{N^p}$  for constants  $k_1 > 0$  and  $0 < p < \min\left(\frac{1-\delta}{d_\gamma}, \frac{2\delta-1}{d_\xi+2}\right)$ .
2.  $\{w_N^i(\cdot)\}$  are kernel regression weight functions with the Gaussian, triangular, or Epanechnikov kernel function and  $h_N = k_4 N^{-\delta}$  for constants  $k_4 > 0$  and  $\delta \in \left(0, \frac{1}{2d_\gamma}\right)$ , and  $\epsilon_N = \frac{k_1}{N^p}$  for constants  $k_1 > 0$  and  $0 < p < \min\left(\delta, \frac{1-\delta d_\gamma}{2+d_\xi}\right)$ .

Given Assumption 4.1, our main result is the following.

**Theorem 4.1.** *Suppose the weight function and uncertainty sets satisfy Assumption 4.1, the joint probability distribution of  $(\gamma, \xi)$  satisfies Assumptions 4.2-4.4 from Section 4.4.3, and the cost function satisfies Assumption 4.5 from Section 4.4.4. Then, for every  $\bar{\gamma} \in \Gamma$ ,*

$$\lim_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) = v^*(\bar{\gamma}), \quad \mathbb{P}^\infty\text{-almost surely.}$$

The theorem says that objective value of (4.3) and the out-of-sample cost of optimal decisions to (4.3) converge almost surely to the optimal value of the full-information problem, (4.1), as  $N$  goes to infinity. The assumptions of the theorem require that the joint distribution of the auxiliary covariates and uncertain quantities and the feasible decision rules are well behaved. We will discuss these technical assumptions in more detail in the following sections.

In order to prove the asymptotic optimality of sample robust optimization with covariates, we view (4.3) through the more general lens of Wasserstein-based distri-

butionally robust optimization. We first review some properties of the Wasserstein metric and then prove a key intermediary result from which our main result follows.

#### 4.4.2 Review of the Wasserstein metric

The Wasserstein metric provides a distance function between probability distributions. In particular, given two probability distributions  $\mathbb{Q}, \mathbb{Q}' \in \mathcal{P}(\Xi)$ , the type-1 Wasserstein distance is defined as the optimal objective value of a minimization problem:

$$d_1(\mathbb{Q}, \mathbb{Q}') := \inf \left\{ \mathbb{E}_{(\boldsymbol{\xi}, \boldsymbol{\xi}') \sim \Pi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\| : \begin{array}{l} \Pi \text{ is a joint distribution of } \boldsymbol{\xi} \text{ and } \boldsymbol{\xi}' \\ \text{with marginals } \mathbb{Q} \text{ and } \mathbb{Q}', \text{ respectively} \end{array} \right\}.$$

The Wasserstein metric is particularly appealing because a distribution with finite support can have a finite distance to a continuous distribution. This allows us to construct a Wasserstein ball around an empirical distribution that includes continuous distributions, which cannot be done with other popular measures such as the Kullback-Leibler divergence [69]. We remark that the 1-Wasserstein metric satisfies the axioms of a metric, including the triangle inequality [39]:

$$d_1(\mathbb{Q}_1, \mathbb{Q}_2) \leq d_1(\mathbb{Q}_1, \mathbb{Q}_3) + d_1(\mathbb{Q}_3, \mathbb{Q}_2), \quad \forall \mathbb{Q}_1, \mathbb{Q}_2, \mathbb{Q}_3 \in \mathcal{P}(\Xi).$$

Important to our analysis, the 1-Wasserstein metric admits a dual form, as shown by Kantorovich and Rubinstein [66],

$$d_1(\mathbb{Q}, \mathbb{Q}') = \sup_{\text{Lip}(h) \leq 1} |\mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}}[h(\boldsymbol{\xi})] - \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}'}[h(\boldsymbol{\xi})]|,$$

where the supremum is taken over all 1-Lipschitz functions. Note that the absolute value is optional in the dual form of the metric, and the space of Lipschitz functions can be restricted to those which satisfy  $h(0) = 0$ , without loss of generality. Finally, we remark that Fournier and Guillin [51] prove under a light-tailed assumption that the 1-Wasserstein distance between the empirical distribution and its underlying dis-

tribution concentrates around zero with high probability. Theorem 4.2 in the following section extends this concentration result to the setting with auxiliary covariates.

### 4.4.3 Concentration of the weighted empirical measure

Given a local predictive method, let the corresponding weighted empirical measure be defined as

$$\hat{\mathbb{P}}_{\bar{\gamma}}^N := \sum_{i=1}^N w_N^i(\bar{\gamma}) \delta_{\xi^i},$$

where  $\delta_{\xi}$  denotes the Dirac probability distribution which places point mass at  $\xi$ . In this section, we prove under mild assumptions that the weighted empirical measure  $\hat{\mathbb{P}}_{\bar{\gamma}}^N$  concentrates quickly to  $\mathbb{P}_{\bar{\gamma}}$  with respect to the 1-Wasserstein metric. We introduce the following assumptions on the underlying joint probability distribution:

**Assumption 4.2** (Conditional Subgaussianity). *There exists a parameter  $\sigma > 0$  such that*

$$\mathbb{P}(\|\xi\| - \mathbb{E}[\|\xi\| \mid \gamma = \bar{\gamma}] > t \mid \gamma = \bar{\gamma}) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \forall t > 0, \bar{\gamma} \in \Gamma.$$

**Assumption 4.3** (Lipschitz Continuity). *There exists  $0 < L < \infty$  such that*

$$d_1(\mathbb{P}_{\bar{\gamma}}, \mathbb{P}_{\bar{\gamma}'} ) \leq L \|\bar{\gamma} - \bar{\gamma}'\|, \quad \forall \bar{\gamma}, \bar{\gamma}' \in \Gamma.$$

**Assumption 4.4** (Smoothness of Auxiliary Covariates). *The set  $\Gamma$  is compact, and there exists  $g > 0$  such that*

$$\mathbb{P}(\|\gamma - \bar{\gamma}\| \leq \epsilon) \geq g\epsilon^{d_\gamma}, \quad \forall \epsilon > 0, \bar{\gamma} \in \Gamma.$$

With these assumptions, we are ready to prove the concentration result, which relies on the dual form of the Wasserstein metric and a discrete approximation of the space of 1-Lipschitz functions.

**Theorem 4.2.** *Suppose the weight function and uncertainty sets satisfy Assumption 4.1 and the joint probability distribution of  $(\gamma, \xi)$  satisfies Assumptions 4.2-4.4.*

Then, for every  $\bar{\gamma} \in \Gamma$ ,

$$\mathbb{P}^\infty \left( \{d_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) > \epsilon_N\} \text{ i.o.} \right) = 0.$$

*Proof.* Without loss of generality, we assume throughout the proof that all norms  $\|\cdot\|$  refer to the  $\ell_\infty$  norm.<sup>1</sup> Fix any  $\bar{\gamma} \in \Gamma$ . It follows from Assumption 4.1 that

$$\{w_N^i(\bar{\gamma})\} \text{ are not functions of } \boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^N; \quad (4.4)$$

$$\sum_{i=1}^N w_N^i(\bar{\gamma}) = 1 \text{ and } w_N^1(\bar{\gamma}), \dots, w_N^N(\bar{\gamma}) \geq 0, \quad \forall N \in \mathbb{N}; \quad (4.5)$$

$$\epsilon_N = \frac{k_1}{N^p}, \quad \forall N \in \mathbb{N}, \quad (4.6)$$

for constants  $k_1, p > 0$ . Moreover, Assumption 4.1 also implies that there exists constants  $k_2 > 0$  and  $\eta > p(2 + d_\xi)$  such that

$$\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \sum_{i=1}^N w_N^i(\bar{\gamma}) \|\boldsymbol{\gamma}^i - \bar{\gamma}\| = 0, \quad \mathbb{P}^\infty\text{-almost surely}; \quad (4.7)$$

$$\mathbb{E}_{\mathbb{P}^N} \left[ \exp \left( \frac{-\theta}{\sum_{i=1}^N w_N^i(\bar{\gamma})^2} \right) \right] \leq \exp(-k_2 \theta N^\eta), \quad \forall \theta \in (0, 1), N \in \mathbb{N}. \quad (4.8)$$

The proof of the the above statements under Assumption 4.1 is found in Appendix C.1.

Now, choose any fixed  $q \in (0, \eta/(2 + d_\xi) - p)$ , and let

$$b_N := N^q, \quad B_N := \{\boldsymbol{\zeta} \in \mathbb{R}^{d_\xi} : \|\boldsymbol{\zeta}\| \leq b_N\}, \quad I_N := \mathbb{1} \{\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^N \in B_N\}.$$

---

<sup>1</sup>To see why this is without loss of generality, consider any other  $\ell_p$  norm where  $p \geq 1$ . In this case,

$$\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_p \leq d_\xi^{1/p} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_\infty.$$

By the definition of the 1-Wasserstein metric, this implies

$$d_1^p(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq d_\xi^{1/p} d_1^\infty(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N),$$

where  $d_1^p$  refers to the 1-Wasserstein metric with the  $\ell_p$  norm. If  $\epsilon_N$  satisfies Assumption 4.1,  $\epsilon_N/d_\xi^{1/p}$  also satisfies Assumption 4.1, so the result for all other choices of  $\ell_p$  norms follows from the result with the  $\ell_\infty$  norm.

Finally, we define the following intermediary probability distributions:

$$\hat{\mathbb{Q}}_{\bar{\gamma}}^N := \sum_{i=1}^N w_N^i(\bar{\gamma}) \mathbb{P}_{\gamma^i}, \quad \hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N := \sum_{i=1}^N w_N^i(\bar{\gamma}) \mathbb{P}_{\gamma^i|B_N},$$

where  $\mathbb{P}_{\gamma^i|B_N}(\cdot)$  is shorthand for  $\mathbb{P}(\cdot \mid \gamma = \gamma^i, \boldsymbol{\xi} \in B_N)$ .

Applying the triangle inequality for the 1-Wasserstein metric and the union bound,

$$\begin{aligned} \mathbb{P}^\infty \left( \left\{ \mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) > \epsilon_N \right\} \text{ i.o.} \right) &\leq \mathbb{P}^\infty \left( \left\{ \mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{Q}}_{\bar{\gamma}}^N) > \frac{\epsilon_N}{3} \right\} \text{ i.o.} \right) \\ &\quad + \mathbb{P}^\infty \left( \left\{ \mathbf{d}_1(\hat{\mathbb{Q}}_{\bar{\gamma}}^N, \hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N) > \frac{\epsilon_N}{3} \right\} \text{ i.o.} \right) \\ &\quad + \mathbb{P}^\infty \left( \left\{ \mathbf{d}_1(\hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N, \hat{\mathbb{P}}_{\bar{\gamma}}^N) > \frac{\epsilon_N}{3} \right\} \text{ i.o.} \right). \end{aligned}$$

We now proceed to bound each of the above terms.

**Term 1:**  $\mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{Q}}_{\bar{\gamma}}^N)$ :

By the dual form of the 1-Wasserstein metric,

$$\mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{Q}}_{\bar{\gamma}}^N) = \sup_{\text{Lip}(h) \leq 1} \left| \mathbb{E}[h(\boldsymbol{\xi}) \mid \gamma = \bar{\gamma}] - \sum_{i=1}^N w_N^i(\bar{\gamma}) \mathbb{E}[h(\boldsymbol{\xi}) \mid \gamma = \gamma^i] \right|,$$

where the supremum is taken over all 1-Lipschitz functions. By (4.5) and Jensen's inequality, we can upper bound this by

$$\begin{aligned} \mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{Q}}_{\bar{\gamma}}^N) &\leq \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \sup_{\text{Lip}(h) \leq 1} \left| \mathbb{E}[h(\boldsymbol{\xi}) \mid \gamma = \bar{\gamma}] - \mathbb{E}[h(\boldsymbol{\xi}) \mid \gamma = \gamma^i] \right| \right) \\ &= \sum_{i=1}^N w_N^i(\bar{\gamma}) \mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \mathbb{P}_{\gamma^i}) \\ &\leq L \sum_{i=1}^N w_N^i(\bar{\gamma}) \|\bar{\gamma} - \gamma^i\|, \end{aligned}$$

where the final inequality follows from Assumption 4.3. Therefore, it follows from (4.7) that

$$\mathbb{P}^\infty \left( \left\{ \mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{Q}}_{\bar{\gamma}}^N) > \frac{\epsilon_N}{3} \right\} \text{ i.o.} \right) = 0. \quad (4.9)$$

**Term 2:**  $\mathbf{d}_1(\hat{\mathbb{Q}}_{\bar{\gamma}}^N, \hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N)$ :

Consider any Lipschitz function  $\text{Lip}(h) \leq 1$  for which  $h(0) = 0$ , and let  $\bar{N} \in \mathbb{N}$  satisfy  $b_{\bar{N}} \geq \sigma + \sup_{\bar{\gamma} \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| | \gamma = \bar{\gamma}]$  (which is finite because of Assumption 4.4). Then, for all  $N \geq \bar{N}$ , and all  $\bar{\gamma}' \in \Gamma$ ,

$$\begin{aligned} & \mathbb{E}[h(\boldsymbol{\xi}) | \gamma = \bar{\gamma}'] - \mathbb{E}[h(\boldsymbol{\xi}) | \gamma = \bar{\gamma}', \boldsymbol{\xi} \in B_N] \\ &= \mathbb{E}[h(\boldsymbol{\xi}) \mathbb{1}\{\boldsymbol{\xi} \notin B_N\} | \gamma = \bar{\gamma}'] + \mathbb{E}[h(\boldsymbol{\xi}) \mathbb{1}\{\boldsymbol{\xi} \in B_N\} | \gamma = \bar{\gamma}'] \\ &\quad - \mathbb{E}[h(\boldsymbol{\xi}) | \gamma = \bar{\gamma}', \boldsymbol{\xi} \in B_N] \\ &= \mathbb{E}[h(\boldsymbol{\xi}) \mathbb{1}\{\boldsymbol{\xi} \notin B_N\} | \gamma = \bar{\gamma}'] + \mathbb{E}[h(\boldsymbol{\xi}) | \gamma = \bar{\gamma}', \boldsymbol{\xi} \in B_N] \mathbb{P}(\boldsymbol{\xi} \in B_N | \gamma = \bar{\gamma}') \\ &\quad - \mathbb{E}[h(\boldsymbol{\xi}) | \gamma = \bar{\gamma}', \boldsymbol{\xi} \in B_N] \\ &= \mathbb{E}[h(\boldsymbol{\xi}) \mathbb{1}\{\boldsymbol{\xi} \notin B_N\} | \gamma = \bar{\gamma}'] - \mathbb{E}[h(\boldsymbol{\xi}) | \gamma = \bar{\gamma}', \boldsymbol{\xi} \in B_N] \mathbb{P}(\boldsymbol{\xi} \notin B_N | \gamma = \bar{\gamma}') \\ &\leq \mathbb{E}[\|\boldsymbol{\xi}\| \mathbb{1}\{\boldsymbol{\xi} \notin B_N\} | \gamma = \bar{\gamma}'] + b_N \mathbb{P}(\boldsymbol{\xi} \notin B_N | \gamma = \bar{\gamma}') \\ &= \int_{b_N}^{\infty} \mathbb{P}(\|\boldsymbol{\xi}\| > t | \gamma = \bar{\gamma}') dt + b_N \mathbb{P}(\|\boldsymbol{\xi}\| \geq b_N | \gamma = \bar{\gamma}') \\ &\leq (\sigma + b_N) \exp \left( -\frac{1}{2\sigma^2} \left( b_N - \sup_{\bar{\gamma}' \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| | \gamma = \bar{\gamma}'] \right)^2 \right). \end{aligned}$$

The first inequality follows because  $|h(\boldsymbol{\xi})| \leq b_N$  for all  $\boldsymbol{\xi} \in B_N$  and  $|h(\boldsymbol{\xi})| \leq \|\boldsymbol{\xi}\|$  otherwise. For the second inequality, we used the Gaussian tail inequality  $\int_x^\infty e^{-t^2/2} dt \leq e^{-x^2/2}$  for  $x \geq 1$  [96, Proposition 2.1.2] along with Assumption 4.2. Because this

bound holds uniformly over all  $h$ , and all  $\bar{\gamma}' \in \Gamma$ , it follows that

$$\begin{aligned}
d_1(\hat{\mathbb{Q}}_{\bar{\gamma}}^N, \hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N) &= \sup_{\text{Lip}(h) \leq 1, h(0)=0} \left| \sum_{i=1}^N w_N^i(\bar{\gamma}) (\mathbb{E}[h(\boldsymbol{\xi})|\gamma = \gamma^i] - \mathbb{E}[h(\boldsymbol{\xi})|\gamma = \gamma^i, \boldsymbol{\xi} \in B_N]) \right| \\
&\leq \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\text{Lip}(h) \leq 1, h(0)=0} |\mathbb{E}[h(\boldsymbol{\xi})|\gamma = \gamma^i] - \mathbb{E}[h(\boldsymbol{\xi})|\gamma = \gamma^i, \boldsymbol{\xi} \in B_N]| \\
&\leq \sup_{\bar{\gamma}' \in \Gamma} \sup_{\text{Lip}(h) \leq 1, h(0)=0} |\mathbb{E}[h(\boldsymbol{\xi})|\gamma = \bar{\gamma}'] - \mathbb{E}[h(\boldsymbol{\xi})|\gamma = \bar{\gamma}', \boldsymbol{\xi} \in B_N]| \\
&\leq (\sigma + b_N) \exp \left( -\frac{1}{2\sigma^2} \left( b_N - \sup_{\bar{\gamma}' \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| | \gamma = \bar{\gamma}'] \right)^2 \right),
\end{aligned}$$

for all  $N \geq \bar{N}$ . It is easy to see that the right hand side above divided by  $\epsilon_N/3$  goes to 0 as  $N$  goes to infinity, so

$$\mathbb{P}^\infty \left( \left\{ d_1(\hat{\mathbb{Q}}_{\bar{\gamma}}^N, \hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N) > \frac{\epsilon_N}{3} \right\} \text{ i.o.} \right) = 0.$$

**Term 3:**  $d_1(\hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N, \hat{\mathbb{P}}_{\bar{\gamma}}^N)$ :

By the law of total probability,

$$\mathbb{P}^N \left( d_1(\hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N, \hat{\mathbb{P}}_{\bar{\gamma}}^N) > \frac{\epsilon_N}{3} \right) \leq \mathbb{P}^N(I_N = 0) + \mathbb{P}^N \left( d_1(\hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N, \hat{\mathbb{P}}_{\bar{\gamma}}^N) > \frac{\epsilon_N}{3} \middle| I_N = 1 \right).$$

We now show that each of the above terms have finite summations. First,

$$\begin{aligned}
\sum_{N=1}^{\infty} \mathbb{P}^N(I_N = 0) &\leq \sum_{N=1}^{\infty} N \sup_{\bar{\gamma}' \in \Gamma} \mathbb{P}(\boldsymbol{\xi} \notin B_N | \gamma = \bar{\gamma}') \\
&\leq \sum_{N=1}^{\infty} N \sup_{\bar{\gamma}' \in \Gamma} \exp \left( -\frac{(b_N - \mathbb{E}[\|\boldsymbol{\xi}\| | \gamma = \bar{\gamma}'])^2}{2\sigma^2} \right) \\
&< \infty.
\end{aligned}$$

The first inequality follows from the union bound, the second inequality follows from Assumption 4.2, and the final inequality follows because  $\sup_{\bar{\gamma}' \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| | \gamma = \bar{\gamma}'] < \infty$  and the definition of  $b_N$ .

Second, for each  $l \in \mathbb{N}$ , we define several quantities. Let  $\mathcal{P}_l$  be the partitioning

of  $B_N = [-b_N, b_N]^{d_\xi}$  into  $2^{ld_\xi}$  translations of  $(-b_N 2^{-l}, b_N 2^{-l})^{d_\xi}$ . Let  $\mathcal{H}_l$  be the set of piecewise constant functions which are constant on each region of the partition  $\mathcal{P}_l$ , taking values on  $\{kb_N 2^{-l} : k \in \{0, \pm 1, \pm 2, \pm 3, \dots, \pm 2^l\}\}$ . Note that  $|\mathcal{H}_l| = (2^{l+1} + 1)^{2^{ld_\xi}}$ . Then, we observe that for all Lipschitz functions  $\text{Lip}(h) \leq 1$  which satisfy  $h(0) = 0$ , there exists a  $\hat{h} \in \mathcal{H}_l$  such that

$$\sup_{\zeta \in B_N} |h(\zeta) - \hat{h}(\zeta)| \leq b_N 2^{-l+1}.$$

Indeed, within each region of the partition,  $h$  can vary by no more than  $b_N 2^{-l+1}$ . The possible function values for  $\hat{h}$  are separated by  $b_N 2^{-l}$ . Because  $h$  is bounded by  $\pm b_N$ , this implies the existence of  $\hat{h} \in \mathcal{H}_l$  such that  $\hat{h}$  has a value within  $b_N 2^{-l+1}$  of  $h$  everywhere within that region. The identical reasoning holds for all other regions of the partition.

Therefore, for every  $l \in \mathbb{N}$ ,

$$\begin{aligned} & \mathbb{P}^N \left( \mathbf{d}_1(\hat{\mathbb{Q}}_{\tilde{\gamma}|B_N}^N, \hat{\mathbb{P}}_{\tilde{\gamma}}^N) > \frac{\epsilon_N}{3} \middle| I_N = 1 \right) \\ &= \mathbb{P}^N \left( \sup_{\substack{\text{Lip}(h) \leq 1 \\ h(0)=0}} \sum_{i=1}^N w_N^i(\tilde{\gamma}) (h(\xi^i) - \mathbb{E}[h(\xi)|\gamma = \gamma^i, \xi \in B_N]) > \frac{\epsilon_N}{3} \middle| I_N = 1 \right) \\ &\leq \mathbb{P}^N \left( \sup_{\hat{h} \in \mathcal{H}_l} \sum_{i=1}^N w_N^i(\tilde{\gamma}) (\hat{h}(\xi^i) - \mathbb{E}[\hat{h}(\xi)|\gamma = \gamma^i, \xi \in B_N]) > \frac{\epsilon_N}{3} - 2 \cdot b_N 2^{-l+1} \middle| I_N = 1 \right) \\ &\leq |\mathcal{H}_l| \sup_{\hat{h} \in \mathcal{H}_l} \mathbb{P}^N \left( \sum_{i=1}^N w_N^i(\tilde{\gamma}) (\hat{h}(\xi^i) - \mathbb{E}[\hat{h}(\xi)|\gamma = \gamma^i, \xi \in B_N]) > \frac{\epsilon_N}{3} - b_N 2^{-l+2} \middle| I_N = 1 \right), \end{aligned}$$

where the final inequality follows from the union bound. We choose  $l = \left\lceil 2 + \log_2 \frac{6b_N}{\epsilon_N} \right\rceil$ , in which case

$$\frac{\epsilon_N}{3} - b_N 2^{-l+2} \geq \frac{\epsilon_N}{6}.$$



Furthermore, for all sufficiently large  $N$ ,

$$|\mathcal{H}_l| = (2^{l+1} + 1)^{2^{ld_\xi}} \leq \left(96 \frac{b_N}{\epsilon_N}\right)^{24^{d_\xi} (b_N/\epsilon_N)^{d_\xi}} = \exp\left(24^{d_\xi} \left(\frac{b_N}{\epsilon_N}\right)^{d_\xi} \log \frac{96b_N}{\epsilon_N}\right).$$

Applying Hoeffding's inequality, and noting  $|\hat{h}(\boldsymbol{\xi}^i)|$  is bounded by  $b_N$  when  $\boldsymbol{\xi}^i \in B_N$ , we have the following for all  $\hat{h} \in \mathcal{H}_l$ :

$$\begin{aligned} & \mathbb{P}^N \left( \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \hat{h}(\boldsymbol{\xi}^i) - \mathbb{E}[\hat{h}(\boldsymbol{\xi}) | \boldsymbol{\xi} \in B_N, \gamma = \gamma^i] \right) > \frac{\epsilon_N}{6} \middle| I_N = 1 \right) \\ &= \mathbb{E} \left[ \mathbb{P}^N \left( \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \hat{h}(\boldsymbol{\xi}^i) - \mathbb{E}[\hat{h}(\boldsymbol{\xi}) | \boldsymbol{\xi} \in B_N, \gamma = \gamma^i] \right) > \frac{\epsilon_N}{6} \middle| I_N = 1, \gamma^1, \dots, \gamma^N \right) \middle| I_N = 1 \right] \\ &\leq \mathbb{E} \left[ \exp \left( -\frac{\epsilon_N^2}{72 \sum_{i=1}^N (w_N^i(\bar{\gamma}))^2 b_N^2} \right) \middle| I_N = 1 \right] \\ &= \mathbb{E} \left[ \exp \left( -\frac{\epsilon_N^2}{72 \sum_{i=1}^N (w_N^i(\bar{\gamma}))^2 b_N^2} \right) I_N \right] \left( \frac{1}{\mathbb{P}^N(I_N = 1)} \right) \\ &\leq 2\mathbb{E} \left[ \exp \left( -\frac{\epsilon_N^2}{72 \sum_{i=1}^N (w_N^i(\bar{\gamma}))^2 b_N^2} \right) \right] \\ &\leq 2 \exp \left( -\frac{k_2 N^\eta \epsilon_N^2}{72 b_N^2} \right), \end{aligned}$$

for  $N$  sufficiently large that  $\mathbb{P}(I_N = 1) \geq 1/2$  and  $\epsilon_N^2/72b_N^2 < 1$ . Note that (4.8) was used for the final inequality. Combining these results, we have

$$\mathbb{P}^N \left( \mathbf{d}_1(\hat{\mathbb{P}}_{\bar{\gamma}}^N, \hat{\mathbb{Q}}_{\bar{\gamma}|B_N}^N) > \epsilon_N/3 \middle| I_N = 1 \right) \leq 2 \exp \left( 24^{d_\xi} \left( \frac{b_N}{\epsilon_N} \right)^{d_\xi} \log \frac{96b_N}{\epsilon_N} - \frac{k_2 \epsilon_N^2 N^\eta}{72 N b_N^2} \right),$$

for  $N$  sufficiently large. For some constants  $c_1, c_2 > 0$ , and sufficiently large  $N$ , this is upper bounded by

$$2 \exp \left( -c_1 N^{\eta-2(p+q)} + c_2 N^{d_\xi(q+p)} \log N \right).$$

Since  $0 < d_\xi(p+q) < \eta - 2(p+q)$ , we can conduct a limit comparison test with  $1/N^2$  to see that this term has a finite sum over  $N$ , which completes the proof.  $\square$

#### 4.4.4 Proof of main result

Theorem 4.2 provides the key ingredient for the proof of the main consistency result. We state one final assumption, which requires that the objective function of (4.1) is upper semicontinuous and bounded by linear functions of the uncertainty.

**Assumption 4.5.** *For all  $\pi \in \Pi$ ,  $c^\pi(\zeta_1, \dots, \zeta_T)$  is upper semicontinuous in  $\zeta$  and  $|c(\zeta, \mathbf{x})| \leq C(1 + \|\zeta\|)$  for all  $\zeta \in \Xi$  and some  $C > 0$ .*

Under this assumption, the proof of Theorem 4.1 follows from Theorem 4.2 via arguments similar to those used by Esfahani and Kuhn [48] and Bertsimas, Shtern, and Sturt [22]. We state it fully in Appendix C.2.

### 4.5 Tractable Approximations

In the previous sections, we presented the new framework of sample robust optimization with covariates and established its asymptotic optimality without any significant structural restrictions on the space of decision rules. In this section, we focus on tractable methods for approximately solving the robust optimization problems that result from this proposed framework. Specifically, we show how to extend the multi-policy approximation scheme from Bertsimas, Shtern, and Sturt [24] to find high-quality solutions for these problems. In combination with linear decision rules, this approach enables us to approximately solve real-world problems with more than ten stages in less than one minute, as we demonstrate in Section 4.6.

We focus on dynamic optimization problems with cost functions of the form

$$\begin{aligned}
 & c(\xi_1, \dots, \xi_T, \mathbf{x}_1, \dots, \mathbf{x}_T) \\
 &= \sum_{t=1}^T \left( \mathbf{f}_t^\top \mathbf{x}_t + \mathbf{g}_t^\top \xi_t + \min_{\mathbf{y}_t \in \mathbb{R}^{d_y^t}} \left\{ \mathbf{h}_t^\top \mathbf{y}_t : \sum_{s=1}^t \mathbf{A}_{t,s} \mathbf{x}_s + \sum_{s=1}^t \mathbf{B}_{t,s} \xi_s + \mathbf{C}_t \mathbf{y}_t \leq \mathbf{d}_t \right\} \right). \tag{4.10}
 \end{aligned}$$

Such cost functions appear frequently in applications such as inventory management and supply chain networks. Unfortunately, it is well known that these cost functions are convex in the uncertainty  $\xi_1, \dots, \xi_T$ . Thus, even evaluating the worst-case cost

over a convex uncertainty set is computationally demanding in general, as it requires the maximization of a convex function.

As an intermediary step towards developing an approximation scheme for (4.3) with the above cost function, we consider the following optimization problem:

$$\begin{aligned}
\tilde{v}^N(\bar{\gamma}) := & \min_{\substack{\boldsymbol{\pi} \in \Pi \\ \mathbf{y}_t^i \in \mathcal{R}_t \forall i,t}} \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \sum_{t=1}^T (\mathbf{f}_t^\top \boldsymbol{\pi}_t(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{t-1}) + \mathbf{g}_t^\top \boldsymbol{\zeta}_t + \mathbf{h}_t^\top \mathbf{y}_t^i(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_t)) \\
\text{s.t.} & \sum_{s=1}^t \mathbf{A}_{t,s} \boldsymbol{\pi}_s(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{s-1}) + \sum_{s=1}^t \mathbf{B}_{t,s} \boldsymbol{\zeta}_s + \mathbf{C}_t \mathbf{y}_t^i(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_t) \leq \mathbf{d}_t \\
& \forall \boldsymbol{\zeta} \in \mathcal{U}_N^i, i \in \{1, \dots, N\}, t \in \{1, \dots, T\},
\end{aligned} \tag{4.11}$$

where  $\mathcal{R}_t$  is the set of all functions  $\mathbf{y} : \Xi_1 \times \dots \times \Xi_t \rightarrow \mathbb{R}^{d_y^t}$ . In this problem, we have introduced auxiliary decision rules which capture the optimal value of the minimization portion of the cost function in (4.10) in each stage. We refer to (4.11) as a *multi-policy* approach, as it involves different auxiliary decision rules for each uncertainty set. The following theorem shows that (4.11) is equivalent to (4.3).

**Theorem 4.3.** *For cost functions of the form (4.10),  $\tilde{v}^N(\bar{\gamma}) = \hat{v}^N(\bar{\gamma})$ .*

*Proof.* See Appendix C.3. □

We observe that (4.11) involves optimizing over decision rules, and thus is computationally challenging to solve in general. Nonetheless, we can obtain a tractable approximation of (4.11) by further restricting the space of primary and auxiliary decision rules. For instance, we can restrict all primary and auxiliary decision rules as linear decision rules of the form

$$\boldsymbol{\pi}_t(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{t-1}) = \mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s} \boldsymbol{\zeta}_s, \quad \mathbf{y}_t^i(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_t) = \mathbf{y}_{t,0}^i + \sum_{s=1}^t \mathbf{Y}_{t,s}^i \boldsymbol{\zeta}_s.$$

One can alternatively elect to use a richer class of decision rules, such as lifted linear decision rules [35, 53] or finite adaptability [12]. In all cases, feasible approximations that restrict the space of decision rules of (4.11) provide an upper bound on the cost  $\tilde{v}^N(\bar{\gamma})$  and produce decision rules that are feasible for (4.11).

The key benefit of the multi-policy approximation scheme is that it offers many degrees of freedom in approximating the nonlinear cost function. Specifically, in (4.11), a separate auxiliary decision rule  $\mathbf{y}_t^i$  captures the value of the cost function for each uncertainty set in each stage. We approximate each  $\mathbf{y}_t^i$  with a linear decision rule, which only needs to be locally accurate, *i.e.*, accurate for realizations in the corresponding uncertainty set. As a result, (4.11) with linear decision rules results in significantly tighter approximations of (4.3) compared to using a single linear decision rule,  $\mathbf{y}_t$ , for all uncertainty sets in each stage. Moreover, these additional degrees of freedom come with only a mild increase in computation cost, and we substantiate these claims via computational experiments in Section 4.6.2.

For completeness, we now show how to reformulate the multi-policy approximation scheme with linear decision rules into a deterministic optimization problem using standard techniques from robust optimization. First, it is readily observed that any instance of (4.11) using linear decision rules can be rewritten in the following compact form:

$$\begin{aligned}
& \min_{\substack{\mathbf{x}_0 \in \mathbb{R}^{d_x}, \mathbf{X} \in \mathbb{R}^{d_x \times d_\xi} \\ \mathbf{y}_0^i \in \mathbb{R}^{d_y}, \mathbf{Y}^i \in \mathbb{R}^{d_y \times d_\xi}}} \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\zeta \in \mathcal{U}_N^i} \{ \mathbf{f}^\top(\mathbf{x}_0 + \mathbf{X}\zeta) + \mathbf{g}^\top\zeta + \mathbf{h}^\top(\mathbf{y}_0^i + \mathbf{Y}^i\zeta) \} \\
& \text{s.t.} \quad \mathbf{A}(\mathbf{x}_0 + \mathbf{X}\zeta) + \mathbf{B}\zeta + \mathbf{C}(\mathbf{y}_0^i + \mathbf{Y}^i\zeta) \leq \mathbf{d} \\
& \quad \mathbf{x}_0 + \mathbf{X}\zeta \in \mathcal{X} \\
& \quad \forall \zeta \in \mathcal{U}_N^i, i \in \{1, \dots, N\},
\end{aligned} \tag{4.12}$$

where  $\mathcal{X} := \mathcal{X}_1 \times \dots \times \mathcal{X}_T$  and the matrices  $\mathbf{X}$  and  $\mathbf{Y}$  are non-anticipative; for technical details, see Appendix C.3. Note that the linear decision rules in the above optimization problem are represented using  $O(d_\xi \max\{d_x, Nd_y\})$  decision variables, where  $d_x := d_x^1 + \dots + d_x^T$  and  $d_y := d_y^1 + \dots + d_y^T$ . Thus, the complexity of representing the primary and auxiliary linear decision rules scales efficiently both in the size of the dataset and the number of stages. For simplicity, we present the reformulation for the case in which there are no constraints on the decision variables and no support constraints on the random variables (*i.e.*,  $\mathcal{X} = \mathbb{R}^{d_x}$  and  $\Xi = \mathbb{R}^{d_\xi}$ ), but a more general

Table 4.1: Relationship of four methods.

$\epsilon_N$	$w_N^i(\bar{\gamma}) = \frac{1}{N}$ for all $i$	$w_N^i(\bar{\gamma})$ from machine learning
$= 0$	Sample average approximation	Bertsimas and Kallus [15]
$> 0$	Bertsimas, Shtern, and Sturt [22]	This chapter

reformulation for when  $\Xi = \mathbb{R}_+^{d_\xi}$  is found in Appendix C.3.

**Theorem 4.4.** *Suppose  $\Xi = \mathbb{R}^{d_\xi}$  and  $\mathcal{X} = \mathbb{R}^{d_x}$ . Then, (4.12) is equivalent to*

$$\begin{aligned}
 & \min_{\substack{\mathbf{x}_0 \in \mathbb{R}^{d_x} \\ \mathbf{X} \in \mathbb{R}^{d_x \times d_\xi} \\ \mathbf{y}_0^i \in \mathbb{R}^{d_y} \\ \mathbf{Y}^i \in \mathbb{R}^{d_y \times d_\xi}}} \sum_{i=1}^N w_N^i(\bar{\gamma}) (\mathbf{f}^\top (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\xi}^i) + \mathbf{g}^\top \boldsymbol{\xi}^i + \mathbf{h}^\top (\mathbf{y}_0^i + \mathbf{Y}^i \boldsymbol{\xi}^i) + \epsilon_N \|\mathbf{X}^\top \mathbf{f} + \mathbf{g} + (\mathbf{Y}^i)^\top \mathbf{h}\|_*) \\
 \text{s.t.} \quad & \mathbf{A} (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\xi}^i) + \mathbf{B}\boldsymbol{\xi}^i + \mathbf{C} (\mathbf{y}_0^i + \mathbf{Y}^i \boldsymbol{\xi}^i) + \epsilon_N \|\mathbf{A}\mathbf{X} + \mathbf{B} + \mathbf{C}\mathbf{Y}^i\|_* \leq \mathbf{d} \\
 & \forall i \in \{1, \dots, N\}.
 \end{aligned}$$

where  $\|\cdot\|_*$  is the dual norm and  $\|\mathbf{Z}\|_* := (\|\mathbf{z}_1\|_*, \dots, \|\mathbf{z}_r\|_*) \in \mathbb{R}^r$  for any matrix  $\mathbf{Z} \in \mathbb{R}^{r \times n}$ .

*Proof.* The proof of Theorem 4.4, as well as generalizations, are found in Appendix C.3. □

## 4.6 Computational Experiments

We perform computational experiments to assess the out-of-sample performance and computational tractability of the proposed methodologies across several applications. These examples are two-stage shipment planning (Section 4.6.1), dynamic inventory management (Section 4.6.2), and portfolio optimization (Section 4.6.3).

We compare several methods using different machine learning models. These methods include the proposed sample robust optimization with covariates, sample average approximation (SAA), the predictions to prescriptions (PtP) approach of Bertsimas and Kallus [15], and sample robust optimization without covariates (SRO). In Table 4.1, we show that each of the above methods are particular instances of

(4.3) from Section 4.3. The methods in the left column ignore covariates by assigning equal weights to each uncertainty set, and the methods in the right column incorporate covariates by choosing the weights based on predictive machine learning. The methods in the top row do not incorporate any robustness, and the methods in the bottom row incorporate robustness via a positive  $\epsilon_N$  in the uncertainty sets. In addition, for the dynamic inventory management example, we also implement and compare to the residual tree algorithm described in Ban, Gallien, and Mersereau [4].

In each experiment, the relevant methods are applied to the same training datasets, and their solutions are evaluated against a common testing dataset. Further details are provided in each of the following sections.

### 4.6.1 Shipment planning

We first consider a two-stage shipment planning problem in which a decision maker seeks to satisfy demand in several locations from several production facilities while minimizing production and transportation costs. Our problem setting closely follows Bertsimas and Kallus [15], in which the decision maker has access to auxiliary covariates (promotions, social media, market trends), which may be predictive of future sales in each retail location.

**Problem Description.** The decision maker first decides the quantity of inventory  $x_f \geq 0$  to produce in each of the production facilities  $f \in \mathcal{F} := \{1, \dots, |\mathcal{F}|\}$ , at a cost of  $p_1$  per unit. The demands  $\xi_\ell \geq 0$  in each location  $\ell \in \mathcal{L} := \{1, \dots, |\mathcal{L}|\}$  are then observed. The decision maker fulfills these demand by shipping  $s_{f\ell} \geq 0$  units from facility  $f \in \mathcal{F}$  to location  $\ell \in \mathcal{L}$  at a per-unit cost of  $c_{f\ell} > 0$ . Additionally, after observing demand, the decision maker has the opportunity to produce additional units  $y_f \geq 0$  in each facility at a cost of  $p_2 > p_1$  per unit. The fulfillment of each unit of demand generates  $r > 0$  in revenue. Given the above notation and dynamics, the

cost incurred by the decision maker is

$$\begin{aligned}
c(\boldsymbol{\xi}, \mathbf{x}) = & \sum_{f \in \mathcal{F}} p_1 x_f - \sum_{\ell \in \mathcal{L}} r \xi_\ell + \min_{\mathbf{s} \in \mathbb{R}_+^{\mathcal{L} \times \mathcal{F}}, \mathbf{y} \in \mathbb{R}_+^{\mathcal{F}}} \sum_{f \in \mathcal{F}} p_2 y_f + \sum_{f \in \mathcal{F}} \sum_{\ell \in \mathcal{L}} c_{f\ell} s_{f\ell} \\
\text{s.t.} & \sum_{f \in \mathcal{F}} s_{f\ell} \geq \xi_\ell & \forall \ell \in \mathcal{L} \\
& \sum_{\ell \in \mathcal{L}} s_{f\ell} \leq x_f + y_f & \forall f \in \mathcal{F}.
\end{aligned}$$

**Experiments.** We perform computational experiments using the same parameters and data generation procedure as Bertsimas and Kallus [15]. Specifically, we consider an instance with  $|\mathcal{F}| = 4$ ,  $|\mathcal{L}| = 12$ ,  $p_1 = 5$ ,  $p_2 = 100$ , and  $r = 90$ . The network topology, transportation costs, and the joint distribution of the covariates  $\boldsymbol{\gamma} \in \mathbb{R}^3$  and demands  $\boldsymbol{\xi} \in \mathbb{R}^{12}$  are the same as Bertsimas and Kallus [15], with the exception that we generate the covariates as i.i.d. samples as opposed to an ARMA process (but with the same marginal distribution).

In our experiments, we compare sample robust optimization with covariates, sample average approximation, sample robust optimization without covariates, and predictions to prescriptions. For the robust approaches (bottom row of Table 4.1), we construct the uncertainty sets from Section 4.3 using the  $\ell_1$  norm and  $\Xi = \mathbb{R}_+^{12}$ , solve these problems using the multi-policy approximation with linear decision rules described in Section 4.5, and consider uncertainty sets with radius  $\epsilon \in \{100, 500\}$ . For the approaches using covariates (right column of Table 4.1), we used the  $k_N$ -nearest neighbors with parameter  $k_N = \frac{2N}{5}$ . All solutions were evaluated on a test set of size 100 and the results were averaged over 100 independent training sets.

**Results.** In Figure 4-1, we present the average out-of-sample profits of the various methods. The results show that the best out-of-sample average profit is attained when using the proposed sample robust optimization with covariates. Interestingly, we observe no discernible differences between sample average approximation and sample robust optimization in Figure 4-1, suggesting the value gained by incorporating covariates in this example. Compared to the approach of Bertsimas and Kallus [15],

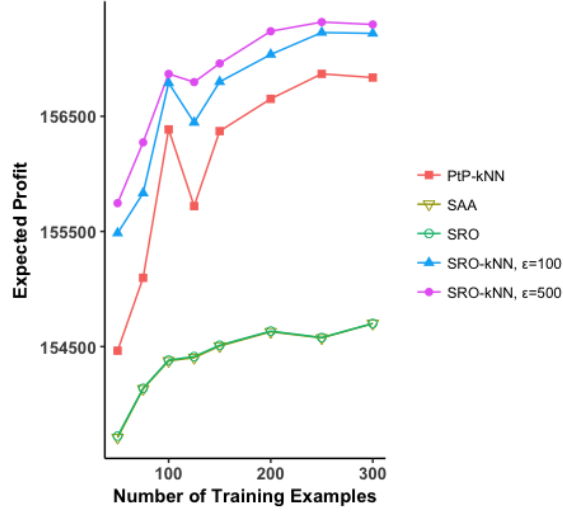


Figure 4-1: Out-of-sample profit for the shipment planning example.

Table 4.2: Statistical significance for shipment planning problem.

$N$	$\epsilon$	
	100	500
50	$4.6 \times 10^{-13}$	$5.3 \times 10^{-16}$
75	$1.3 \times 10^{-14}$	$6.4 \times 10^{-12}$
100	$1.2 \times 10^{-13}$	$1.1 \times 10^{-7}$
125	$2.6 \times 10^{-15}$	$1.5 \times 10^{-11}$
150	$3.4 \times 10^{-12}$	$1.2 \times 10^{-6}$
200	$1.4 \times 10^{-12}$	$1.0 \times 10^{-8}$
250	$3.4 \times 10^{-10}$	$1.0 \times 10^{-4}$
300	$1.8 \times 10^{-6}$	$5.2 \times 10^{-4}$

The  $p$ -values from the Wilcoxon signed rank test for comparison with the predictive to prescriptive analytics method (PtP- $k$ NN) and sample robust optimization with covariates (SRO- $k$ NN). After adjusting for multiple hypothesis testing, all results are significant at the  $\alpha = 0.05$  significance level because all  $p$ -values are less than  $\frac{0.05}{16} \approx 3.1 \times 10^{-3}$ .

sample robust optimization with covariates achieves a better out-of-sample average performance for each choice of  $\epsilon$ . Table 4.2 shows that these differences are statistically significant. This example demonstrates that, in addition to enjoying asymptotic optimality guarantees, sample robust optimization with covariates provides meaningful value across various values of  $N$ .



## 4.6.2 Dynamic inventory management

We next consider a dynamic inventory control problem over the first  $T = 12$  weeks of a new product. In each week, a retailer observes demand for the product and can replenish inventory via procurement orders to different suppliers with lead times. Our problem setting closely follows Ban, Gallien, and Mersereau [4], motivated by the fashion industry in which retailers have access to auxiliary covariates on the new product (color, brand, Google trends, promotions) which can be predictive of how demand unfolds over time.

**Problem Description.** In each stage  $t \in \{1, \dots, T\}$ , the retailer procures inventory from multiple suppliers to satisfy demand for a single product. The demands for the product across stages are denoted by  $\xi_1, \dots, \xi_T \geq 0$ . In each stage  $t$ , and before the demand  $\xi_t$  is observed, the retailer places procurement orders with various suppliers indexed by  $\mathcal{J} = \{1, \dots, |\mathcal{J}|\}$ . Each supplier  $j \in \mathcal{J}$  has per-unit order cost of  $c_{tj} \geq 0$  and a lead time of  $\ell_j$  stages. At the end of each stage, the firm incurs a per-unit holding cost of  $h_t$  and a backorder cost of  $b_t$ . If inventory is fully backlogging and the firm starts with zero initial inventory, the cost incurred by the firm over the time horizon is captured by

$$c(\xi_1, \dots, \xi_T, \mathbf{x}_1, \dots, \mathbf{x}_T) = \sum_{t=1}^T \sum_{j \in \mathcal{J}} c_{tj} x_{tj} + \min_{y_t \in \mathbb{R}} y_t$$

$$\text{s.t. } y_t \geq h_t \left( \sum_{j \in \mathcal{J}} \sum_{s=1}^{t-\ell_j} x_{sj} - \sum_{s=1}^t \xi_s \right)$$

$$y_t \geq -b_t \left( \sum_{j \in \mathcal{J}} \sum_{s=1}^{t-\ell_j} x_{sj} - \sum_{s=1}^t \xi_s \right).$$

**Experiments.** The parameters of the procurement problem were chosen based on Ban, Gallien, and Mersereau [4]. Specifically, we consider the case of two suppliers where  $c_{t1} = 1.0$ ,  $c_{t2} = 0.5$ ,  $h_t = 0.25$ , and  $b_t = 11$  for each stage. The first supplier has no lead time and the second supplier has a lead time of one stage. We generate

Table 4.3: Average out-of-sample cost for dynamic procurement problem.

Method	$k$	$\epsilon$							
		0	100	200	300	400	500	600	700
Sample robust optimization									
Linear decision rules									
no covariates		9669	8783	8590	8789	9150	9604	10102	10614
k-nearest neighbors	26	9600	8566	8411	8642	9030	9494	10001	10528
	20	9640	8544	8375	8603	8996	9464	9974	10505
	13	9862	8561	8365	8573	8960	9433	9943	10473
Linear decision rules, multi-policy									
no covariates		8967	7759	7360	7320	7460	7716	8038	8412
k-nearest neighbors	26	11346	8728	7651	7269	7241	7381	7636	7966
	20	13012	9460	7925	7328	<b>7195</b>	7289	7519	7835
	13	16288	10975	8576	7585	7243	7236	7412	7697

Average out-of-sample cost for the dynamic procurement problem using sample robust optimization with  $N = 40$ . For each uncertainty set radius  $\epsilon$  and parameter  $k$ , average was taken over 100 training sets and 100 test points. Optimal is indicated in bold. The residual tree algorithm with a binning of  $B = 2$  in each stage gave an average out-of-sample cost of 27142.

training and test data from the same distribution as the shipment planning problem in Section 4.6.1. In this case, the demands produced by this process are interpreted as the demands over the  $T = 12$  stages. We perform computational experiments comparing sample robust optimization with and without covariates, as well as the residual tree algorithm proposed by Ban, Gallien, and Mersereau [4]. In particular, we compare sample robust optimization with covariates *with* the multi-policy approximation as well as *without* the multi-policy approximation (in which we use a single auxiliary linear decision rule for  $y_t$  for all uncertainty sets in each stage). The uncertainty sets are defined with the  $\ell_2$  norm and  $\Xi = \mathbb{R}_+^{12}$ . The out-of-sample cost resulting from the decision rules were averaged over 100 training sets of size  $N = 40$  and 100 testing points, and sample robust optimization with covariates used  $k$ -nearest neighbors with varying choices of  $k$  and radius  $\epsilon \geq 0$  of the uncertainty sets.

**Results.** In Table 4.3, we show the average out-of-sample cost resulting from sample robust optimization with covariates using linear decision rules, with and without the multi-policy approximation from Section 4.5. In both settings, we used  $k$ -nearest neighbors as the machine learning method and evaluated the out-of-sample performance by applying the linear decision rules for the ordering quantities. The results of these computational experiments in Table 4.3 demonstrate that significant improvements in average out-of-sample performance are found when combining the multi-

Table 4.4: Statistical significance for dynamic procurement problem.

Method	$k$	$\epsilon$							
		0	100	200	300	400	500	600	700
Sample robust optimization									
Linear decision rules									
no covariates		*	*	*	*	*	*	*	*
k-nearest neighbors	26	*	*	*	*	*	*	*	*
	20	*	*	*	*	*	*	*	*
	13	*	*	*	*	*	*	*	*
Linear decision rules, multi-policy									
no covariates		*	*	*	*	*	*	*	*
k-nearest neighbors	26	*	*	*	*	$1.4 \times 10^{-5}$	*	*	*
	20	*	*	*	*	-	*	*	*
	13	*	*	*	*	$5.8 \times 10^{-3}$	$1 \times 10^{-3}$	*	*

The  $p$ -values of the Wilcoxon signed rank test for comparison with sample robust optimization using linear decision rules with multi-policy,  $k = 20$ , and  $\epsilon = 400$ . An asterisk denotes that the  $p$ -value was less than  $10^{-8}$ . After adjusting for multiple hypothesis testing, each result is significant at the  $\alpha = 0.05$  significance level if its  $p$ -value is less than  $\frac{0.05}{63} \approx 7.9 \times 10^{-4}$ .

Table 4.5: Average computation time (seconds) for dynamic procurement problem.

Method	$k$	$\epsilon$							
		0	100	200	300	400	500	600	700
Sample robust optimization									
Linear decision rules									
no covariates		3.86	25.04	24.75	25.82	28.70	35.37	31.13	31.95
k-nearest neighbors	26	4.02	25.43	23.39	25.15	27.88	33.42	30.87	31.60
	20	3.99	25.98	23.56	24.93	27.41	32.67	30.69	31.50
	13	4.19	26.53	24.89	24.99	26.79	31.64	30.23	31.32
Linear decision rules, multi-policy									
no covariates		0.16	28.31	30.01	29.05	31.13	36.03	35.57	36.09
k-nearest neighbors	26	0.15	27.74	28.69	27.78	30.54	34.44	35.50	36.15
	20	0.15	27.87	28.51	27.74	30.60	34.36	35.65	36.99
	13	0.14	27.78	28.30	27.27	30.00	33.67	35.91	37.76

Average computation time (seconds) for the dynamic procurement problem using sample robust optimization with  $N = 40$ . For each choice of uncertainty set radius  $\epsilon$  and parameter  $k$ , average was taken over 100 training sets. The residual tree algorithm of Ban, Gallien, and Mersereau [4] with a binning of  $B = 2$  in each stage had an average computation time of 23.20 seconds. We were unable to run this algorithm with binning of  $B = 3$  in each stage.

policy approximation with covariates via  $k$ -nearest neighbors. We show in Table 4.4 that these results are statistically significant. For comparison, we also implemented the residual tree algorithm from Ban, Gallien, and Mersereau [4]. When using their algorithm with a binning of  $B = 2$  in each stage, their approach resulted in an average out-of-sample cost of 27142. We were unable to run with a binning of  $B = 3$  in each stage due to time limitations of  $10^3$  seconds, as the size of the resulting linear optimization problem scales on the order  $O(B^T)$ . Such results are consistent with the estimations of computation times presented in [4, Section 6.3]. The running times of the various methods are displayed in Table 4.5.

### 4.6.3 Portfolio optimization

Finally, we consider a single-stage portfolio optimization problem in which we wish to find an allocation of a fixed budget to  $n$  assets. Our goal is to simultaneously maximize the expected return while minimizing the conditional value at risk (cVaR) of the portfolio. Before selecting our portfolio, we observe auxiliary covariates which include general market indicators such as index performance as well as macroeconomic numbers released by the US Bureau of Labor Statistics.

**Problem Description.** We denote the portfolio allocation among the assets by  $\mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}_+^n : \sum_{j=1}^n x_j = 1\}$ , and the returns of the assets by the random variables  $\boldsymbol{\xi} \in \mathbb{R}^n$ . The conditional value at risk at the  $\alpha \in (0, 1)$  level measures the expected loss of the portfolio, conditional on losses being above the  $1 - \alpha$  quantile of the loss distribution. Rockafellar and Uryasev [82] showed that the cVaR of a portfolio can be computed as the optimal objective value of a convex minimization problem. Therefore, our portfolio optimization problem can be expressed as a convex optimization problem with an auxiliary decision variable,  $\beta \in \mathbb{R}$ . Thus, given an observation  $\bar{\gamma}$  of the auxiliary covariates, our goal is to solve

$$\min_{\mathbf{x} \in \mathcal{X}, \beta \in \mathbb{R}} \mathbb{E} \left[ \beta + \frac{1}{\alpha} \max(0, -\mathbf{x}^\top \boldsymbol{\xi} - \beta) - \lambda \mathbf{x}^\top \boldsymbol{\xi} \mid \boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}} \right], \quad (4.13)$$

where  $\lambda \in \mathbb{R}_+$  is a trade-off parameter that balances the risk and return objectives.

**Experiment.** Our experiments are based on a similar setting from Bertsimas and Van Parys [19, Section 5.2]. Specifically, we perform computational experiments on an instance with parameters  $\alpha = 0.05$  and  $\lambda = 1$ , and the joint distribution of the covariates and asset returns are chosen the same as Bertsimas and Van Parys [19, Section 5.2].

In our experiments, we compare sample robust optimization with covariates, sample average approximation, sample robust optimization without covariates, and predictions to prescriptions. For the robust approaches (bottom row of Table 4.1), we

construct the uncertainty sets from Section 4.3 using the  $\ell_1$  norm. For each training sample size, we compute the out-of-sample objective on a test set of size 1000, and we average the results over 100 instances of training data.

In order to select  $\epsilon_N$  and other tuning parameters associated with the machine learning weight functions, we first split the data into a training and validation set. We then train the weight functions using the training set, compute decisions for each of the instances in the validation set, and compute the out-of-sample cost on the validation set. We repeat this for a variety of parameter values and select the combination that achieves the best cost on the validation set.

Following a similar reformulation approach as Esfahani and Kuhn [48], we solve the robust approaches *exactly* by observing that

$$\begin{aligned}
& \min_{\mathbf{x} \in \mathcal{X}, \beta \in \mathbb{R}} \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \left\{ \beta + \frac{1}{\alpha} \max\{0, -\mathbf{x}^\top \boldsymbol{\zeta} - \beta\} - \lambda \mathbf{x}^\top \boldsymbol{\zeta} \right\} \\
= & \min_{\mathbf{x} \in \mathcal{X}, \beta \in \mathbb{R}} \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \left\{ \max \left\{ \beta - \lambda \mathbf{x}^\top \boldsymbol{\zeta}, \left( \frac{1}{\alpha} + \lambda \right) \mathbf{x}^\top \boldsymbol{\zeta} \right\} \right\} \\
= & \min_{\mathbf{x} \in \mathcal{X}, \beta \in \mathbb{R}} \sum_{i=1}^N w_N^i(\bar{\gamma}) \max \left\{ \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \{ \beta - \lambda \mathbf{x}^\top \boldsymbol{\zeta} \}, \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \left( \frac{1}{\alpha} + \lambda \right) \mathbf{x}^\top \boldsymbol{\zeta} \right\}, \\
= & \min_{\mathbf{x} \in \mathcal{X}, \beta \in \mathbb{R}, \mathbf{v} \in \mathbb{R}^N} \sum_{i=1}^N w_N^i(\bar{\gamma}) v_i \\
\text{s.t.} & \quad v_i \geq \beta - \lambda \mathbf{x}^\top \boldsymbol{\zeta} \\
& \quad v_i \geq \left( \frac{1}{\alpha} + \lambda \right) \mathbf{x}^\top \boldsymbol{\zeta} \\
& \quad \forall \boldsymbol{\zeta} \in \mathcal{U}_N^i, i \in \{1, \dots, N\}.
\end{aligned}$$

The final expression can be reformulated as a deterministic optimization problem by reformulating the robust constraints.

**Results.** In Figure 4-2, we show the average out-of-sample objective values using the various methods. Consistent with the computational results of Esfahani and Kuhn [48] and Bertsimas and Van Parys [19], the results underscore the importance of robustness in preventing overfitting and achieving good out-of-sample performance

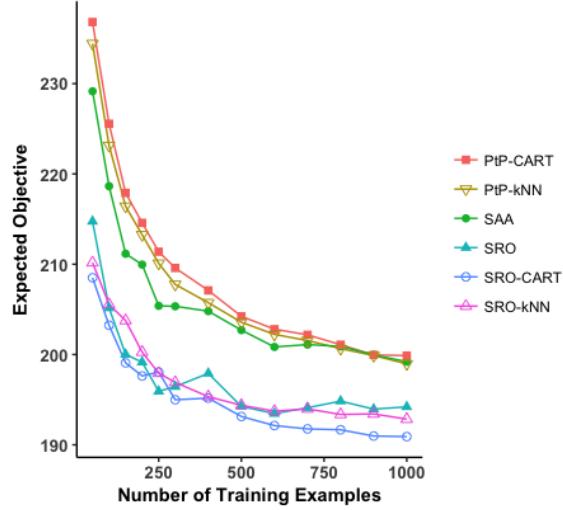


Figure 4-2: Out-of-sample objective for the portfolio optimization example.

in the small data regime. Indeed, we observe that the sample average approximation, which ignores the auxiliary data, outperforms PtP- $k$ NN and PtP-CART when the amount of training data is limited. We believe this is due to the fact the latter methods both throw out training examples, so the methods overfit when the training data is limited, leading to poor out-of-sample performance. In contrast, our methods (SRO- $k$ NN and SRO-CART) typically achieve the strongest out-of-sample performance, even though the amount of training data is limited.

## 4.7 Conclusion

In this chapter, we introduced *sample robust optimization with covariates*, a new framework for solving dynamic optimization problems with side information. Through three computational examples, we demonstrated that our method achieves significantly better out-of-sample performance than scenario-based alternatives. We complemented these empirical observations with theoretical analysis, showing our non-parametric method is asymptotically optimal. Finally, we showed our approach inherits the tractability of robust optimization, scaling to problems with many stages via the multi-policy approximation scheme.

# Chapter 5

## Data-Driven Optimization of Operating Room Blocks

Many hospitals face significant census variability throughout the week, which limits the amount of patients they can accept due to resource constraints during the peak times. Partnering with a major US hospital, we develop a data-driven algorithm for optimizing the operating room block schedule to alleviate census peaks. Our approach adapts the predictive-to-prescriptive analytics framework of Bertsimas and Kallus [15] to efficiently use all available data. It employs predictive machine learning to estimate surgical patients' length of stay distributions and then incorporates the estimates into an integer optimization formulation. Our approach accommodates many types of constraints and also accounts for the impact that the day of the week on which a surgery occurs has on the patient's ensuing length of stay. We demonstrate the effectiveness of the approach with simulation results using real hospital data and find that we can reliably reduce the maximal weekly census with only a few changes to the block schedule.

### 5.1 Introduction

Interest in data-driven optimization has grown significantly in recent years. As the amount of data collected by governments, private companies, and hospitals has

surged, so too has the need for efficient algorithms to turn that data into decisions that result in lower costs and higher revenues. In this chapter, we perform a case study in which we apply modern techniques from data-driven optimization to schedule the operating rooms at a major US hospital.

Many hospitals, including our partner institution, operate near capacity. The challenge facing them is that the hospital can reach 100% occupancy during peak census periods, creating a situation in which additional patients cannot be admitted, despite demand being present. In particular, in the case of our partner institution, the peak census at times has been found to prevent high-priority transfer patients from being admitted at the time of the request. While some of these transfer requests can be accommodated after a waiting period, other patients can not wait, so the transferring institutions end up sending the patients elsewhere. Therefore, we observe a negative societal impact from census peaks when patients are unable to gain access to their preferred institution for care. In addition, census peaks cause organizational challenges due to the increased amount of communication and collaboration required to accommodate admission requests during periods of limited capacity. Finally, census peaks have a drastic financial impact on healthcare institutions, as the lost transfer requests encountered during these peaks represent millions of dollars in missed revenue opportunity.

Our approach aims to address these challenges by scheduling the operating rooms to minimize the peak weekly hospital census, thus maximizing access and revenue potential. Prior to this project, we found that total hospital census tended to be higher between Tuesday and Thursday than between Friday and Monday. As shown in Figure 5-1, although same-day admission (SDA) patients account for less than 20% of the total census, they provide the primary source of census variability. At a high level, this phenomenon makes sense, due to the fact that same-day surgical admits are only scheduled between Monday and Friday, and these patients generally tend to have shorter lengths of stay than medical patients who are admitted throughout the seven-day week. This presents an opportunity for improvement because same-day admits are non-emergency cases, so modifying the operating room (OR) schedule



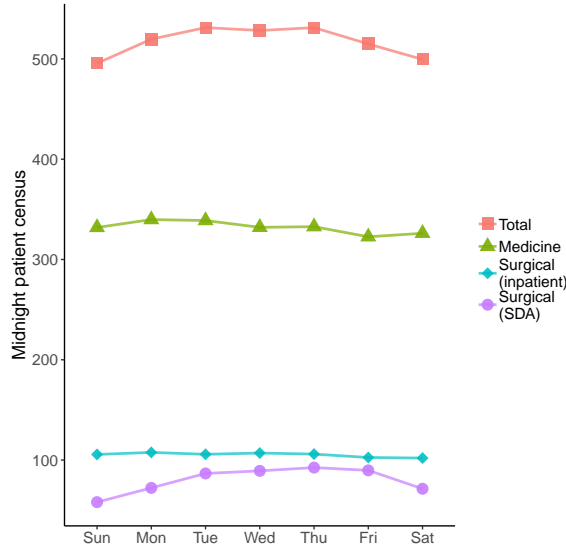


Figure 5-1: Average overnight census by day of the week.

affects the subset of patients that creates census peaks.

The operating room schedule at our partner institution, as well as at many other hospitals, consists of *blocks*, each belonging to a particular surgeon. Each block occurs in the same time slot each week. Some surgeons own multiple blocks, which they may use for different procedure types. We can affect the operating room schedule by moving the blocks to different time slots in the week.

The hospital does not behave as a deterministic system. There are several key sources of uncertainty. One is the patients' lengths of stay following their procedure. We have access to data on the historical arrivals and stays of patients, which are able to use to construct predictive models for these uncertain quantities. In particular, the data includes:

- Covariates for each surgical patient, i.e., age, gender, diagnosis, etc.
- The name, department, and speciality of the surgeon performing each operation.
- The length of stay of each surgical patient.
- The total census of all other patients in the hospital on each day, i.e., emergency patients and transfer patients.

From this historical data, we can build models to predict the distributions of future patients' lengths of stay. Another source of uncertainty is the number and types of patients that will undergo procedures during each block. While SDA patients are scheduled ahead of time, the hospital sets the operating room schedule many months in advance, so patient arrivals are uncertain at that time. In Section 5.3, we explain how we incorporate predictive models into our data-driven optimization framework, allowing us to account for uncertainty while optimizing the operating room block schedule.

### 5.1.1 Related work

There has been significant interest in developing efficient scheduling procedures for hospital operating rooms. McManus et al. [74] found that at a large children's hospital, variability in demand for services inhibits the efficient distribution of hospital resources, and a significant proportion of that variability is due to scheduled surgeries. Similarly, Ragavan et al. [81], in a different hospital, found that variation in scheduled surgery admissions was more than three times higher than variation in emergency surgery admissions. Both support our observations at our partner institution that there is significant potential for improving hospital efficiency by optimizing the operating room schedule.

Researchers have considered various objectives to try to improve. Denton et al. [43] attempted to minimize operating room idle time. Similarly, VanBerkel and Blake [95] aimed to minimize patient waiting time (and thus maximize the throughput of the operating room). Dexter and Macario [44] consider various scheduling systems and their effect on the efficiency of the operating room, which they define as a combination of underutilization and overutilization. Underutilization refers to the situation in which operating rooms are available, but unused, and overutilization refers to the situation in which the operating room time exceeds that which was scheduled. Overutilization can occur because the durations of surgeries are nondeterministic. This type of objective is reminiscent of a newsvendor approximation to the scheduling problem (see, for example, Petruzzi and Dada [79]).

More related to the objective we seek to optimize is work by Marcon and Dexter [72] and Marcon and Dexter [71]. They consider the downstream impact of the operating room schedule on the post-anesthesia care unit (PACU). Specifically, they find that reducing variability in the rate of flow of patients from the operating room to the PACU improves efficiency and reduces delays. Kim and Horowitz [68] considered the effect of operating room scheduling on the intensive care unit (ICU). We believe these types of downstream effects are crucial and must be incorporated into the objective for effective operating room scheduling.

Financial objectives might be the most general and most appropriate type of objective to consider for our purpose. However, as Dexter et al. [45] note, accounting data for hospitals is often lacking, so it is not always possible to accurately compute the financial impact of scheduling decisions. For that reason, we instead use an objective that is correlated with hospital profit, but is easily computable using admission, discharge, and transfer data.

When considering the decision to optimize, the most general setting is one in which the algorithm has the ability to schedule every patient (either in an online fashion or offline for a fixed horizon) individually. However, this is impractical because staffing considerations impose considerable constraints on the space of feasible schedules. Blake et al. [27] develop an approach that creates a master schedule that remains the same from week to week for the surgeons, and then makes limited modifications each week to account for the particular demand. Due to the needs of our partner institution, we focus on optimizing a schedule that will remain fixed for many months, but it would be straightforward to extend our approach to the setting in which a new schedule is optimized each week. In addition, we take the blocks in the schedule as fixed entities. Each block typically corresponds to a single surgeon, but surgeons may have multiple blocks which they use for different types of procedures. It is also possible to have blocks which are shared by multiple surgeons. Previous work by Beliën et al. [5] and Chow et al. [37] has focused on optimizing the groupings of surgeons/procedures into blocks. We do not discuss that problem in this chapter, but it is possible to extend our approach to optimize for that as well.

Uncertainty plays a key role in the census impact of an operating room schedule. First, patients’ lengths of stay are unknown. Second, although SDA procedures are scheduled in advance, the number of types of patients that will receive procedures each day are unknown at the time the operating room schedule is made. Little previous work has accounted for these uncertainty sources (see Cardoen et al. [32] for a catalogue). Persson and Persson [78] developed a simulation model for operating room scheduling that accounts for uncertainty in patient arrivals and the durations of the procedures, but, to our knowledge, no previous work has incorporated patient length of stay uncertainty into an optimization framework.

In our work, we focus on uncertainty in patient arrivals and in lengths of stay. As our objective concerns the down-stream effects of the operating room, we believe these are the most important sources of uncertainty for our problem. The approach most closely related to ours is developed by Zenteno et al. [100]. Their method finds the OR block schedule that would have minimized peak hospital census for the training data. It assumes each patient’s length of stay does not depend on the day of the week the procedure occurred. Our approach differs in that we use patient and surgeon covariate data to build a machine learning model from which we impute predicted length of stay distributions for each patient. This de-noising results in improved out-of-sample performance in simulations.

### 5.1.2 Contributions and Structure

The primary purpose of this case study is to demonstrate how to apply the data-driven optimization techniques developed by Bertsimas and Kallus [15] to a real-world hospital scheduling problem. While the formulations we describe are specific to the operating room scheduling problem at our partner institution, the same techniques can apply to more general operational problems facing hospitals. In Section 5.2, we formulate the problem as an integer optimization problem, assuming all uncertain quantities are known deterministically from historical data. Solving this formulation gives the schedule that would have worked best on the training data, although it will not necessarily perform well on test data. In Section 5.3, we develop a data-

driven formulation by incorporating a machine learning model on patient and surgeon covariates. Finally, in Section 5.4, we present the results of our method on simulations, based on real hospital data, and describe the real-world impact of the model.

## 5.2 Problem Formulation and Deterministic Approach

To begin, we introduce a deterministic formulation for the OR scheduling problem. It takes as fixed the patient arrivals and finds the schedule that would have been best historically if all patients' lengths of stays remained the same.

**Data and problem parameters** For concreteness, we define the data and problem parameters as follows.

- $\mathcal{B}$ : set of blocks in the schedule. Each block consists of a surgeon and a procedure type. (Some surgeons perform different types of procedures in different blocks.)
- $\mathcal{J}$ : the set of potential time slots in which blocks can be scheduled. For our particular problem, this consists of each week day, but  $\mathcal{J}$  can also include specific times of the day.
- $\hat{d}(j)$  for  $j \in \mathcal{J}$ : the day of the week (of  $\{1, \dots, 7\}$ ) on which time slot  $j$  occurs.
- $j(b)$  for  $b \in \mathcal{B}$ : the time slot in which block  $b$  is currently scheduled.
- $T$ : the number of weeks in the training data.<sup>1</sup>
- $\mathcal{P}$ : the set of all SDA patients in the training data.
- $\mathcal{P}_b$  for  $b \in \mathcal{B}$ : the subset of  $\mathcal{B}$  consisting of all block  $b$  patients in the training data.

---

<sup>1</sup>We do not include the first and last weeks of the training data in this count because we consider the effect of shifting each block by up to one week. The training data actually consists of weeks 0 through  $T + 1$ .

- $x_i$  for  $i \in \mathcal{P}$ : covariates for patient  $i$ . This includes demographic data on the patient, the type of procedure the patient receives, and the name of the primary surgeon performing the patient's procedure.
- $N_{b,d}^t$  for  $t \in \{0, \dots, T+1\}$ ,  $b \in \mathcal{B}$ , and  $d \in \{1, \dots, 7\}$ : number of patients from block  $b$  who are in the hospital at 11:59 pm on day  $d$  of week  $t$ .
- $N_d^{t,\text{other}}$  for  $t \in \{1, \dots, T\}$  and  $d \in \{1, \dots, 7\}$ : total number of patients, other than same-day surgical admits, that are in the hospital at 11:59 pm on day  $d$  of week  $t$ .
- $M_b$  for  $b \in \mathcal{B}$ : average number of patients who undergo procedures each week in block  $b$ .
- $C_j$  for  $j \in \mathcal{J}$ : capacity of operating rooms during time slot  $j$ . This is computed as the maximum historical utilization of the operating rooms during any time slot.
- $C_{j,\mathcal{A}}$  for  $j \in \mathcal{J}$  and  $\mathcal{A} \subset \mathcal{B}$ : capacity of operating rooms during time slot  $j$  for blocks in the set  $\mathcal{A}$ . This consists of blocks corresponding to particular procedure types that require specific surgical equipment.
- $\mathcal{S}$ : set of surgeons that hold block time.
- $\mathcal{B}_s$  for  $s \in \mathcal{S}$ : subset of blocks which are used by surgeon  $s$ .

These quantities are all readily computable from the data. For convenience, we also define the quantity

$$N_{b,d}^{t,j} := N_{b,(d+\hat{d}(j)-\hat{d}(j(b))) \bmod 7}^{t+\lfloor (d+\hat{d}(j)-\hat{d}(j(b)))/7 \rfloor} \quad (5.1)$$

This simply computes the number of patients from block  $b$  that would have been in the hospital on day  $d$  of week  $t$  if block  $b$  had been scheduled in time slot  $j$ . For the deterministic model, we assume patient lengths of stay remain the same regardless of when their procedures occur, so (5.1) just shifts the census for each block by the appropriate amount of days. Next, we describe the integer optimization formulation that we construct to optimize the block schedule.

**Decision variables** We optimize the schedule by moving around the time slots corresponding to the blocks. We define the binary variables:

$$z_{b,j} := \begin{cases} 1, & \text{if block } b \text{ is scheduled for time slot } j, \\ 0, & \text{otherwise,} \end{cases}$$

for  $b \in \mathcal{B}$  and  $j \in \mathcal{J}$ . For convenience, we also define the binary variables:

$$x_s := \begin{cases} 1, & \text{if any of the blocks belonging to surgeon } s \text{ are modified,} \\ 0, & \text{otherwise,} \end{cases}$$

for  $s \in \mathcal{S}$ . These variables allow us to count how many surgeons' schedules are modified.

**Objective** The objective we seek to minimize is the average maximum weekly census of the hospital. Formally, this is given by:

$$\frac{1}{T} \sum_{t=1}^T \max_{d \in \{1, \dots, 7\}} \left[ \sum_{b \in \mathcal{B}} \sum_{j \in \mathcal{J}} N_{b,d}^{t,j} z_{b,j} + N_d^{t,\text{other}} \right].$$

This objective accounts for same-day surgical admit patients, whose admission dates are affected by the OR schedule, and all other inpatients, whose admission dates do not depend on the decision variables. We note that, as written, the objective is not a linear function of the decision variables. However, since each term in the summation is the maximum of a finite number of linear functions, we can easily linearize it in a mixed integer optimization formulation.

**Constraints** We include several types of constraints. The first constraints ensure each block is scheduled exactly once.

$$\sum_{j \in \mathcal{J}} z_{b,j} = 1, \quad \forall b \in \mathcal{B}.$$

The next set of constraints force the schedule to respect the capacities of the operating rooms.

$$\begin{aligned} \sum_{b \in \mathcal{B}} M_b z_{b,j} &\leq C_j, \quad \forall j \in \mathcal{J}, \\ \sum_{b \in \mathcal{A}} M_b z_{b,j} &\leq C_{j,\mathcal{A}}, \quad \forall j \in \mathcal{J}, \forall \mathcal{A} \subset \mathcal{B}. \end{aligned}$$

We also constrain the problem so each surgeon can only have one block in time slot  $j$ .

$$\sum_{b \in \mathcal{B}_s} z_{b,j} \leq 1, \quad \forall j \in \mathcal{J}, \forall s \in \mathcal{S}.$$

Finally, we add constraints that limit the number of surgeons whose schedules can be modified, which we limit by the parameter  $\lambda$ . We have found that almost 90% of the potential improvement can be captured by modifying less than 10% of the surgeons' schedules. To constrain  $x_s$  to be the indicator of whether surgeon  $s$ 's schedule is modified, we have

$$\sum_{b \in \mathcal{B}_s} (1 - z_{b,j(b)}) \leq |\mathcal{B}_s| x_s, \quad \forall s \in \mathcal{S}.$$

To limit the number of changes, we have

$$\sum_{s \in \mathcal{S}} x_s \leq \lambda.$$

These constraints are general and apply to most hospitals. There may also be specific constraints needed for any particular hospital. For example, there may be certain blocks that cannot be scheduled for Fridays because of a lack of resources for the post-op patients over the weekends. MIO permits the flexibility to incorporate these types of constraints into the model. We denote the projection of the set of the above constraints onto the  $z$  decision variables by  $\mathcal{Z}$ .



## 5.3 Data-driven Approach

Several shortcomings of the deterministic approach described in Section 5.2 make it overly optimistic in its prediction of the maximum weekly census. First, it assumes that all patients' lengths of stay remain the same, regardless of which day of the week their procedure occurs. In practice, weekends often delay a patient's discharge due to lower availability of resources, so a patient's length of stay depends on when the procedure occurs. For example, if a patient had a procedure on Wednesday and was discharged on Friday, it is unlikely that patient would be discharged before Monday if he or she instead had that procedure on Thursday. Second, the deterministic formulation overfits to the training data. By building a model to predict length of stay distributions, we can significantly improve the estimate of schedule changes on the census. In Section 5.3.1, we review the predictive-to-prescriptive analytics framework, which we then apply to the operating room scheduling problem in Section 5.3.2.

### 5.3.1 Review of predictive-to-prescriptive analytics framework

The predictive-to-prescriptive analytics framework, introduced by Bertsimas and Kallus [15], combines optimization with machine learning in order to incorporate auxiliary covariates in data-driven optimization problems. In a traditional stochastic optimization problem, the objective is to minimize a function that depends on an uncertain quantity  $Y$  in expectation,

$$\min_{z \in \mathcal{C}} \mathbb{E}[c(z; Y)], \quad (5.2)$$

where  $c$  is a cost function and  $z$  is the decision variable which lies in constraint set  $\mathcal{C}$ . In practice, the distribution of  $Y$  is unknown, but instead independent, identically distributed observations  $Y_1, \dots, Y_N$  are available. The method of sample average approximation replaces the expectation over an unknown distribution with an empirical expectation,

$$\min_{z \in \mathcal{C}} \frac{1}{N} \sum_{i=1}^N c(z; Y_i). \quad (5.3)$$

Under certain technical conditions on the cost function, the constraint set, and the distribution of  $Y$ , the sample average approximation method produces asymptotically optimal solutions. That is, as  $N$  grows to infinity, the set of optimal solutions of (5.3) converges to the set of optimal solutions of (5.2). (For more information, see Shapiro et al. [89].)

By contrast, the predictive-to-prescriptive analytics framework instead aims to minimize the conditional expectation of an objective function, given some observed auxiliary covariates,  $x \in \mathcal{X}$ :

$$\min_{z \in \mathcal{C}} \mathbb{E}[c(z; Y) | X = x]. \quad (5.4)$$

In this setting, the training data consists of observations of the covariates paired with the uncertain quantity,  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Instead of weighting all of the training samples equally, as in sample average approximation, the predictive-to-prescriptive analytics framework weights each training example according to its similarity to the new auxiliary covariate vector:

$$\min_{z \in \mathcal{C}} \sum_{i=1}^n w_i^N(x) c(z; Y_i), \quad (5.5)$$

where  $\{w_i^N(x)\}$  are weight functions derived from a machine learning method. For example, CART weight functions are derived from the classification and regression tree algorithm of Breiman et al. [30]. This tree-based method partitions the space of covariates,  $\mathcal{X}$ , into leaves, and the corresponding weight functions are given by:

$$w_i^{N, \text{CART}}(x) = \frac{1}{|\{i : X_i \in l(x)\}|} \mathbf{1}\{X_i \in l(x)\},$$

where  $l(x)$  is the leaf node of the trained regression tree that contains  $x$ . Similarly, random forest weight functions are based on the random forest regression algorithm of Breiman [29]. They average the weights of an ensemble of  $B$  CART learners, each

trained on a random subset of the data.

$$w_i^{N,\text{RF}}(x) = \frac{1}{B} \sum_{b=1}^B w_{i,b}^{N,\text{CART}},$$

where  $w_{i,b}^{N,\text{CART}}$  refers to the weight function corresponding to the  $b$ th tree in the ensemble. Bertsimas and Kallus [15] detail several further examples of weight functions. They also prove that under an appropriate choice of weight function, and mild technical assumptions, the solutions to (5.5) are asymptotically optimal for (5.4). The value of considering (5.4) instead of (5.2) is that it takes into account all available information.

An extension, developed by Bertsimas et al. [21] and Bertsimas and McCord [17], deals with the setting in which the decision variable affects the distribution of the uncertain quantity,  $Y$ . Under the Neyman-Rubin potential outcomes model [84], there exist multiple potential states of the world, one for each possible decision. That is, there exist random variables  $\{Y(z)\}$  for all  $z \in \mathcal{C}$ . However, the only state that is observed in the training data is the one that corresponds to the selected decision. Therefore, the training data consists of tuples of covariates, decision, and outcome,  $(X_i, Z_i, Y_i(Z_i))$  for  $i = 1, \dots, N$ .

Bertsimas and McCord [17] proved that, under technical assumptions, applying the predictive-to-prescriptive analytics framework with  $(X_i, Z_i)$  as the auxiliary covariates gives an asymptotically consistent algorithm in this setting. That is, instead of solving (5.5), one should solve

$$\min_{z \in \mathcal{C}} \sum_{i=1}^n w_i^N(x, z) c(z; Y_i),$$

where  $\{w_i^N(x, z)\}$  are weight functions trained on the covariate/decision tuples.

To model the OR block scheduling problem with the predictive-to-prescriptive analytics framework, we use the same decision variables,  $z_{b,j}$ , and constraints,  $\mathcal{Z}$ , as in Section 5.2. The auxiliary covariates consist of the patient covariates for all patients. The uncertain quantity,  $Y$ , consists of individual patients' lengths of stay. In this

setting, the decision does affect the uncertainty because the day of the week on which a procedure occurs affects the ensuing length of stay for the patient. In the next section, we develop a data-driven formulation of the OR block scheduling problem. Although the problem differs slightly from the idealized predictive-to-prescriptive analytics setting, the insights gained from the setting discussed here provide significant value.

### 5.3.2 Data-driven OR scheduling formulation

For the data-driven formulation, we use the same decision variables and constraints as in the deterministic formulation of Section 5.2. However, we modify the objective to account for uncertainty. We begin by defining the following random variables:

- $\tilde{N}_{b,d}^j$ : the number of block  $b$  patients in the hospital at 11:59 pm of day  $d$  of the week, assuming block  $b$  is scheduled in time slot  $j$ .
- $\tilde{N}_d^{\text{other}}$ : the number of non-SDA patients in the hospital at 11:59 pm of day  $d$  of the week.

We seek to minimize, in expectation, the maximum weekly census. That is, we aim to solve

$$\min_{z \in \mathcal{Z}} \mathbb{E} \left[ \max_{d \in \{1, \dots, 7\}} \sum_{b \in \mathcal{B}} \sum_{j \in \mathcal{J}} \tilde{N}_{b,d}^j z_{b,j} + \tilde{N}_d^{\text{other}} \right]. \quad (5.6)$$

We note that two sources of uncertainty contribute to the objective. First is the uncertainty in patient arrivals within each block. As we discussed previously, SDA procedures are typically scheduled several weeks in advance. However, the schedule remains fixed for many months at a time, so the patient arrivals in each block are unknown when optimizing the schedule. The second source of uncertainty is the uncertainty in patients' lengths of stay. When patients arrive for procedures, we do not know how long they will remain in the hospital.

In order to handle the first type of uncertainty, we fix the patient arrival pattern from the historical data for each block. We assume that, for each week in the historical data, the patients in each block remain the same regardless of when in the

schedule that block is moved. We average over many weeks of historical data to create a reasonable estimate of the distribution of arrivals. To express this formally, we introduce the following random variables, which are defined on the probability space corresponding to the measure on which the patient arrival pattern is fixed.

- $\tilde{N}_{b,d}^{t,j}$ : the number of block  $b$  patients in the hospital at 11:59 pm of day  $d$  of week  $t$ , assuming block  $b$  is scheduled in time slot  $j$ .
- $\tilde{N}_d^{t,\text{other}}$ : the number of non-SDA patients in the hospital at 11:59 pm of day  $d$  of week  $t$ .

Our first approximation to (5.6) is given by

$$\min_{z \in \mathcal{Z}} \mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^T \left[ \max_{d \in \{1, \dots, 7\}} \sum_{b \in \mathcal{B}} \sum_{j \in \mathcal{J}} \tilde{N}_{b,d}^{t,j} z_{b,j} + \tilde{N}_d^{t,\text{other}} \right] \middle| \{x_i : i \in \mathcal{P}\} \right], \quad (5.7)$$

where we condition on the patient covariates. In (5.7), the only remaining source of uncertainty is the uncertainty in patients' lengths of stay.

Because  $\tilde{N}_d^{t,\text{other}}$  counts a large number of patients across many different departments, we estimate it with  $N_d^{t,\text{other}}$ , the observed number of non-SDA patients in the hospital on day  $d$  of week  $t$ . This estimate is justified by concentration of measure results which show that functions of many independent random variables that do not depend too much on any single random variable are approximately constant.  $\tilde{N}_d^{t,\text{other}}$  is a function of many individual patients' lengths of stay, which we assume to be mutually independent, but is not highly sensitive to any single patient's length of stay, so we can reasonably assume  $N_d^{t,\text{other}}$  is a good estimate of it. (See Boucheron et al. [28] for a reference on measure concentration.)

The deterministic formulation of Section 5.2 also estimates  $\tilde{N}_{b,d}^{t,j}$  with a constant,  $N_{b,d}^{t,j}$ . However, most blocks in the operating room schedule only contribute a couple of patients towards the census at a time, so the measure concentration justification for this approximation does not hold. We instead use machine learning to estimate the distributions of patients' lengths of stay.

To do so, we rewrite  $\tilde{N}_{b,d}^{t,j}$  as a function of the individual patients' lengths of stay.

$$\tilde{N}_{b,d}^{t,j} = \sum_{i \in \mathcal{P}_b} \mathbb{1}\{\tilde{L}_i^j > l_{i,j,d,t}^*\}, \quad (5.8)$$

where  $\tilde{L}_i^j$  is the random variable for the length of stay of patient  $i$  if the procedure is scheduled in time slot  $j$ , and  $l_{i,j,d,t}^*$  is the minimum length of stay necessary for that patient to be in the hospital on day  $d$  of week  $t$ . For example, if patient  $i$  has a procedure during week 1 in a time slot on Wednesday, then  $l_{i,j,d,t}^* = 2$  for  $d = \text{Friday}$  and  $t = 1$ . If patient  $i$ 's procedure is scheduled after day  $d$  of week  $t$ , then  $l_{i,j,d,t}^*$  is defined to be infinity. The right hand side of (5.8) simply counts the number of block  $b$  SDA patients in the hospital on day  $d$  of week  $t$ . We note that  $\tilde{L}_i^j$  exists in the potential outcomes model, but in the historical data we only observe the length of stay for each patient corresponding to the block time slot in which the procedure actually occurred, which we denote  $L_i$ .

In order to estimate the distribution of the objective in (5.7), we independently estimate the distribution of  $\tilde{L}_i^j$  for each patient  $i$  by constructing a length of stay prediction model. As features we include the patient covariates,  $x_i$ . This includes demographic information on patient  $i$ , a categorical variable indicating the type of procedure, and a categorical variable indicating the name of the primary surgeon. We also include as a feature the day of the week on which the procedure occurred. We do so because we know the day of the week on which the procedure occurs affects the patient's ensuing length of stay. Including this feature allows the model to learn the effect of admission day on length of stay from the historical data.

As in the predictive-to-prescriptive analytics framework, we train a machine learning model on these features and denote the resulting weight functions by  $\{w_k(x, j)\}$ . That is,  $w_k(x, j)$  denotes the weight assigned to patient  $k$  from the training data when trying to predict the length of stay for a patient with covariates  $x$  and procedure scheduled in block  $j$ . The model's estimate of the distribution of  $\tilde{L}_i^j$  is given

by

$$\mathbb{P}\left(\tilde{L}_i^j > l\right) \approx \sum_{k \in \mathcal{P}} w_k(x_i, j) \mathbb{1}\{L_k > l\}. \quad (5.9)$$

Using these distributional estimates for  $\tilde{L}_i^j$ , and the assumption that patients' lengths of stay are independent, we can compute estimated distributions for  $\tilde{N}_{b,d}^{t,j}$  and for the objective of (5.7).

However, this estimated distribution for the objective of (5.7) relies on the product of  $|\mathcal{P}|$  independent distributions, one for each SDA patient in the training data. Solving the resulting problem requires an exponential number of decision variables and is intractable for the number of patients in our problem. We instead approximately solve (5.7) for the estimated distribution of the objective implied by (5.9). To do so, we independently sample  $N$  lengths of stay for each patient from the distribution (5.9). We let  $\hat{L}_i^{j,n}$  represent the  $n$ th sample of  $\tilde{L}_i^j$  and

$$\hat{N}_{b,d}^{t,j,n} := \sum_{i \in \mathcal{P}_b} \mathbb{1}\{\hat{L}_i^{j,n} > l_{i,j,d,t}^*\},$$

the number of block  $b$  patients in the hospital on day  $d$  of week  $t$  if  $b$  is scheduled in time slot  $j$ , under scenario  $n$ . We then solve an approximate version of (5.7),

$$\min_{z \in \mathcal{Z}} \frac{1}{N} \sum_{n=1}^N \left[ \frac{1}{T} \sum_{t=1}^T \left[ \max_{d \in \{1, \dots, 7\}} \sum_{b \in \mathcal{B}} \sum_{j \in \mathcal{J}} \hat{N}_{b,d}^{t,j,n} z_{b,j} + N_d^{t,\text{other}} \right] \right]. \quad (5.10)$$

This type of sampling approach is justified by consistency results for sample average approximation (see, for example, Shapiro et al. [89]). We can solve (5.10) by linearizing the objective and applying an off-the-shelf MIO solver such as Gurobi [77].

We conclude by highlighting the two major differences between the deterministic formulation of Section 5.2 and the data-driven formulation of this section. First, the deterministic formulation estimates each patient's length of stay as a deterministic quantity while the data-driven formulation estimates a length of stay distribution for each patient using covariate data. The improved estimates of patients' length of stay distributions help curb the model's overfitting and lead to better performance

out-of-sample. Second, the deterministic formulation assumes each patient’s length of stay remains the same regardless of when the procedure occurs. The data-driven formulation learns from the data how the day of surgery affects the patient’s ensuing length of stay.

We note that we did not directly apply the predictive-to-prescriptive analytics framework to the OR block scheduling problem. The objective of (5.7) depends on all patients lengths of stay, so there is only a single observation of the uncertain quantity. However, because we assume patients’ lengths of stays are independent, the data consists of  $|\mathcal{P}|$  i.i.d. observations of patient covariates and their lengths of stay,  $\{(x_i, L_i) : i \in \mathcal{P}\}$ . Therefore, we were able use ideas from the predictive-to-prescriptive analytics framework to construct improved estimates of patient length of stay distributions. With these improved, data-driven estimates, the out-of-sample performance of the optimized schedule improves tangibly.

## 5.4 Application and Impact

Incorporating the results of the models described above has the potential to provide significant operational impact in the hospital setting. We first describe the results of a simulation study, based on real hospital data, and then discuss the real-world impact of the model.

### 5.4.1 Results

In order to demonstrate the positive impact our approach to operating room scheduling, we conduct a study using 12 months of real data from our partner institution. We use the first 6 months to train the models and then the last 5 months of data to evaluate the performance of the proposed schedules. We leave out a month in the middle to minimize the overlap of patients between the training and testing sets.

In order to evaluate proposed schedule changes on the test data, we need to estimate how patients’ lengths of stay are affected when their admission dates change. Using domain knowledge, we developed the following heuristic. We assume each



patient’s length of stay remains the same under the new schedule, unless the patient is scheduled to be discharged on a weekend, in which case we extended the length of stay until the following Monday. This is conservative because it assumes lengths of stay only increase due to schedule changes, but it accounts for the lesser availability of hospital services on weekends, which often delays discharge. In addition, in order to augment the limited test data we have available, we extend the test horizon by adding 100 additional weeks via bootstrapping. That is, we sample, with replacement, 100 weeks of patient arrivals from the set of weeks in the test data. We add each of these weeks of patient arrivals to extend the test horizon. Although these patients and their lengths of stay have already been seen in the test data, the hospital census in each of these additional test weeks is new because the order of the weeks matters. We believe this provides a realistic setting to test the impact of schedule changes over a long horizon.

We compare the impacts of proposed schedules from the deterministic formulation and the data-driven formulation. For the data-driven formulation we use both the CART and the random forest weight functions. We set the number of scenarios we sample for the data-driven formulation,  $N$ , equal to 100. We solve all optimization problems using Gurobi v8.0.1 [77].

Our results are shown in Figure 5-2. The horizontal axis plots the number of surgeons whose schedules the model is allowed to change (the  $\lambda$  parameter from Section 5.2), and the vertical axis shows the out-of-sample improvement in average maximum weekly census (compared to the current OR schedule). With all approaches, we see diminishing returns in the number of changes. For most numbers of allowed changes, the data-driven approaches outperform the deterministic approach. There are a few instances in which the deterministic performance is superior, but in general its performance is more erratic because it overfits more to the training data. In particular, for between 12 and 17 allowed changes, the deterministic formulation produces schedules that perform very poorly compared to the schedules produced by the data-driven approaches. There are many very different feasible schedules that have similar objective values, so small errors in estimation of the objective can lead to

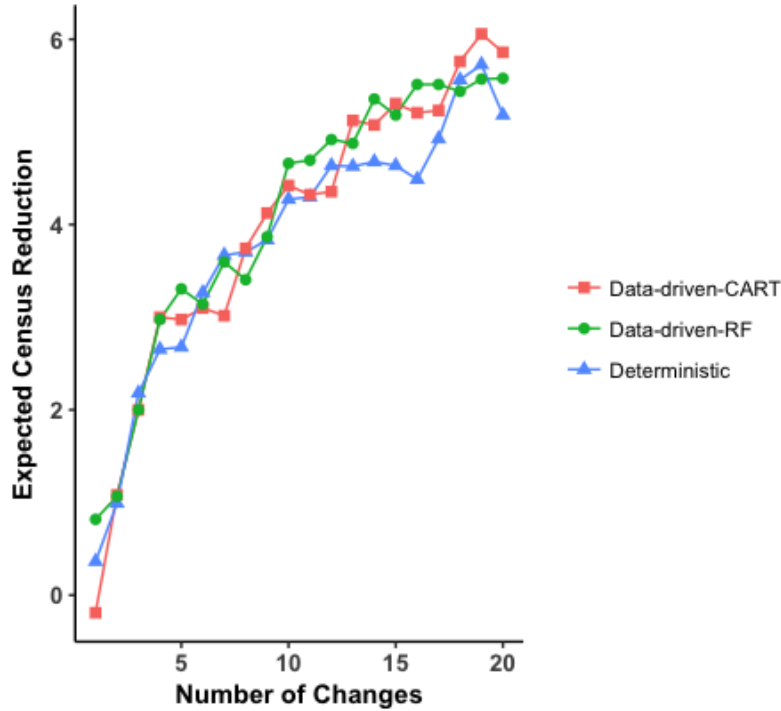


Figure 5-2: Out-of-sample reduction in average maximum weekly census for each of the various models.

big changes in out-of-sample performance. Therefore, the de-noising of the objective function, provided by the data-driven approaches, gives more stable out-of-sample performance, in much the same way that regularization benefits out-of-sample performance in linear regression.

### 5.4.2 Extension for Targeted Growth

As of early 2019, our partner institution has modified the schedules of three surgeons in response to the suggestions of the data-driven model. In addition to suggesting changes to existing surgeons' schedules, our model has also allowed institutions to examine the expected census impact of new surgeons being added to the OR schedule to inform the days on which they should operate. As part of our partner institution's strategic growth plan, the Department of Orthopedics planned to increase its annual total joint replacement volume by bringing on a new surgeon. Without access to the model, Orthopedics likely would have placed the surgeon's OR day on whichever day

had block time availability. However, by using the model, the department worked with the institution's OR block governance committee to place the new surgeon on the day that would have the smallest impact on peak census. As a result, Orthopedics expects to achieve its targeted volume growth (representing approximately 350 inpatient days) without increasing peak census above current state, thereby allowing the inpatient units to handle incremental demand without any adverse effects on capacity.

## 5.5 Conclusion

In this chapter, we developed a new, data-driven approach to the operating room block scheduling problem. We adapted the predictive-to-prescriptive analytics framework of Bertsimas and Kallus [15] to this particular problem in order to minimize the average maximum weekly census. With machine learning methods such as CART and random forest, we estimated the distributions of patients' lengths of stay and incorporated these estimates into a mixed integer optimization problem formulation. In addition, our model used the training data to learn the effects of a patient's admission day on the ensuing length of stay. We showed via simulation results that our approach out-performs the deterministic approach to OR scheduling, which does not take into account all available data. Our data-driven optimization approach has the potential to significantly benefit our partner institution by reducing census variability and, thus, increasing the effective capacity of the hospital.



# Chapter 6

## Conclusion

This thesis developed and analyzed data-driven algorithms for optimization under uncertainty with auxiliary covariates. In Chapter 2, we introduced a method to solve observational optimization problems. We demonstrated, both theoretically and empirically, that accounting for the uncertainty in the predicted cost of each decision improves out-of-sample performance. In Chapter 3, we developed an approach for dynamic problems. We showed how to use machine learning to construct a stochastic process with finite support that approximates the true, unknown stochastic process of the uncertainty, and we showed that solving the dynamic programming problem corresponding to the approximate stochastic process gives asymptotically optimal solutions to the true problem. In Chapter 4, we introduced an alternative, distributionally robust approach to dynamic optimization with covariates. By introducing robustness, this approach avoids the need to consider an exponential number of scenarios and proves tractable and effective both in theory and in computational examples. In the process of developing this approach, we showed how to use nonparametric machine learning methods to learn the conditional distribution of a random variable given covariates, instead of just the conditional expectation. Finally, in Chapter 5, we applied these methods to schedule a hospital's operating rooms, demonstrating how data-driven optimization can provide significant societal and financial value in real-world applications.



# Appendix A

## Appendix for Chapter 2

### A.1 Proofs

To begin, we prove the following lemma.

**Lemma A.1.** *Suppose Assumptions 2.1-2.5 hold. If  $(x, z)$  and  $(x, z')$  are in the same partition of  $\mathcal{X} \times \mathcal{Z}$ , as specified by Assumption 2.3, then*

$$|\Psi(z, \delta) - \Psi(z', \delta)| \leq \left( \alpha(LD + 1 + \sqrt{2\lambda_{\max} \ln 1/\delta}) + L(\sqrt{2 \ln 1/\delta} + 3) \right) \|z - z'\|,$$

where  $\Psi(z, \delta) = \mu(x, z) - \hat{\mu}(x, z) - \frac{2}{3\gamma_n} \ln(1/\delta) - \sqrt{2V(x, z) \ln(1/\delta)} - L \cdot B(x, z)$ .

*Proof.* We first note  $|\mu(x, z) - \mu(x, z')| \leq L\|z - z'\|$  by the Lipschitz assumption on  $c(z; y)$ .

Next, since  $(x, z)$  and  $(x, z')$  are contained in the same partition,

$$\begin{aligned}
|\hat{\mu}(x, z) - \hat{\mu}(x, z')| &= \left| \sum_i w_i(x, z)c(z; Y_i) - w_i(x, z')c(z'; Y_i) \right| \\
&\leq \left| \sum_i w_i(x, z)c(z; Y_i) - w_i(x, z)c(z'; Y_i) \right| \\
&\quad + \left| \sum_i w_i(x, z)c(z'; Y_i) - w_i(x, z')c(z'; Y_i) \right| \\
&\leq L\|z - z'\| + \|w(x, z) - w(x, z')\|_1 \\
&\leq (L + \alpha)\|z - z'\|,
\end{aligned}$$

where we have used Holder's inequality, the uniform bound on  $c$ , and Assumption 2.3.

Similarly, for the bias term,

$$\begin{aligned}
|LB(x, z) - LB(x, z')| &= L \left| \sum_i w_i(x, z)\|(X_i, Z_i) - (x, z)\| - w_i(x, z')\|(X_i, Z_i) - (x, z')\| \right| \\
&\leq L \left| \sum_i w_i(x, z)\|(X_i, Z_i) - (x, z)\| - w_i(x, z)\|(X_i, Z_i) - (x, z')\| \right| \\
&\quad + L \left| \sum_i w_i(x, z)\|(X_i, Z_i) - (x, z')\| - w_i(x, z')\|(X_i, Z_i) - (x, z')\| \right| \\
&\leq L \sum_i w_i(x, z) \left| \|(X_i, Z_i) - (x, z)\| - \|(X_i, Z_i) - (x, z')\| \right| \\
&\quad + L\|w(x, z) - w(x, z')\|_1 \sup_i \|(X_i, Z_i) - (x, z)\| \\
&\leq (L + L\alpha D)\|z - z'\|.
\end{aligned}$$

Next, we consider variance term. We let  $\Sigma(z)$  denote the diagonal matrix with



$\text{Var}(c(z; Y_i)|X_i, Z_i)$  for  $i = 1, \dots, n$  as entries. As before,

$$\begin{aligned}
\left| \sqrt{V(x, z)} - \sqrt{V(x, z')} \right| &= \left| \sqrt{\sum_i w_i^2(x, z) \text{Var}(c(z; Y_i)|X_i, Z_i)} \right. \\
&\quad \left. - \sqrt{\sum_i w_i^2(x, z') \text{Var}(c(z'; Y_i)|X_i, Z_i)} \right| \\
&\leq \left| \sqrt{\sum_i w_i^2(x, z) \text{Var}(c(z; Y_i)|X_i, Z_i)} \right. \\
&\quad \left. - \sqrt{\sum_i w_i^2(x, z) \text{Var}(c(z'; Y_i)|X_i, Z_i)} \right| \\
&\quad + \left| \sqrt{\sum_i w_i^2(x, z) \text{Var}(c(z'; Y_i)|X_i, Z_i)} \right. \\
&\quad \left. - \sqrt{\sum_i w_i^2(x, z') \text{Var}(c(z'; Y_i)|X_i, Z_i)} \right| \\
&= \left| \sqrt{w(x, z)^T \Sigma(z) w(x, z)} - \sqrt{w(x, z)^T \Sigma(z') w(x, z)} \right| \\
&\quad + \left| \|w(x, z)\|_{\Sigma(z')} - \|w(x, z')\|_{\Sigma(z')} \right|,
\end{aligned}$$

where  $\|v\|_{\Sigma} = \sqrt{v^T \Sigma v}$ . One can verify that, because  $\Sigma$  is positive semidefinite,  $\|\cdot\|_{\Sigma}$  is seminorm that satisfies the triangle inequality. Therefore, we can upper bound the latter term by

$$\begin{aligned}
\sqrt{(w(x, z) - w(x, z'))^T \Sigma (w(x, z) - w(x, z'))} &\leq \|w(x, z) - w(x, z')\| \\
&\leq \|w(x, z) - w(x, z')\|_1 \\
&\leq \alpha \|z - z'\|,
\end{aligned}$$

where we have used the assumption that  $|c(z; y)| \leq 1$ .

The former term can again be upper bounded by the triangle inequality.

$$\begin{aligned}
& \left| \sqrt{\sum_i w_i^2(x, z) \text{Var}(c(z; Y_i) | X_i, Z_i)} - \sqrt{\sum_i w_i^2(x, z) \text{Var}(c(z'; Y_i) | X_i, Z_i)} \right| \\
& \leq \sqrt{\sum_i w_i^2(x, z) (\sqrt{\text{Var}(c(z; Y_i) | X_i, Z_i)} - \sqrt{\text{Var}(c(z'; Y_i) | X_i, Z_i)})^2} \quad (\text{A.1})
\end{aligned}$$

Noting that  $\sqrt{\text{Var}(c(z; Y_i))} = \|c(z; Y_i) - \mathbb{E}[c(z; Y_i)]\|_{L_2}$  (dropping conditioning for notational convenience), we can apply the triangle inequality to the  $L_2$  norm:

$$\begin{aligned}
& (\|c(z; Y_i) - \mathbb{E}[c(z; Y_i)]\|_{L_2} - \|c(z'; Y_i) - \mathbb{E}[c(z'; Y_i)]\|_{L_2})^2 \\
& \leq \|c(z; Y_i) - c(z'; Y_i) - \mathbb{E}[c(z; Y_i) - c(z'; Y_i)]\|_{L_2}^2 \\
& \leq \mathbb{E}[(c(z; Y_i) - c(z'; Y_i))^2] \\
& \leq L^2 \|z - z'\|^2.
\end{aligned}$$

Therefore, we can upperbound (A.1) by

$$\begin{aligned}
& \sqrt{\sum_i w_i^2(x, z) L^2 \|z - z'\|^2} \\
& \leq \sum_i w_i(x, z) L \|z - z'\| = L \|z - z'\|,
\end{aligned}$$

where we have used the concavity of the square root function. Therefore,

$$|\sqrt{V(x, z)} - \sqrt{V(x, z')}| \leq (\alpha + L) \|z - z'\|.$$

Combining the three results with the triangle inequality yields the desired result.  $\square$

*Proof of Theorem 2.1.* To derive a regret bound, we first restrict our attention to the fixed design setting. Here, we condition on  $X_1, Z_1, \dots, X_n, Z_n$  and bound  $\hat{\mu}(x, z)$  around its expectation. To simplify notation, we write  $X$  to denote  $(X_1, \dots, X_n)$  and  $Z$  to denote  $(Z_1, \dots, Z_n)$ . Note that by the honesty assumption, in this setting,  $\hat{\mu}$  is a simple sum of independent random variables. Applying Bernstein's inequality (see,

for example, Boucheron et al. [28]), we have, for  $\delta \in (0, 1)$ ,

$$P\left(\mathbb{E}[\hat{\mu}(x, z) | X, Z] - \hat{\mu}(x, z) \leq \frac{2}{3\gamma_n} \ln(1/\delta) + \sqrt{2V(x, z) \ln(1/\delta)} \middle| X, Z\right) \geq 1 - \delta.$$

Next, we need to bound the difference between  $\mathbb{E}[\hat{\mu}(x, z) | X, Z]$  and  $\mu(x, z)$ . By the honesty assumption, Jensen's inequality, and the Lipschitz assumption, we have

$$\begin{aligned} |\mathbb{E}[\hat{\mu}(x, z) | X, Z] - \mu(x, z)| &= \left| \sum_i w_i(x, z) (\mu(X_i, Z_i) - \mu(x, z)) \right| \\ &\leq \sum_i w_i(x, z) |\mu(X_i, Z_i) - \mu(x, z)| \\ &\leq L \sum_i w_i(x, z) \|(X_i, Z_i) - (x, z)\| \\ &= L \cdot B(x, z). \end{aligned}$$

Combining this with the previous result, we have, with probability at least  $1 - \delta$  (conditioned on  $X$  and  $Z$ ),

$$\mu(x, z) - \hat{\mu}(x, z) \leq \frac{2}{3\gamma_n} \ln(1/\delta) + \sqrt{2V(x, z) \ln(1/\delta)} + L \cdot B(x, z) \quad (\text{A.2})$$

Next, we extend this result to hold uniformly over all  $z \in \mathcal{Z}$ . To do so, we partition  $\mathcal{X} \times \mathcal{Z}$  into  $\Gamma_n$  regions as in Assumption 2.3. For each region, we construct a  $\nu$ -net. Therefore, we have a set  $\{\hat{z}_1, \dots, \hat{z}_{K_n}\}$  such that for any  $z \in \mathcal{Z}$ , there exists a  $\hat{z}_k$  such that  $(x, z)$  and  $(x, \hat{z}_k)$  are contained in the same region with  $\|z - \hat{z}_k\| \leq \nu$ . For ease of notation, let  $k : \mathcal{Z} \rightarrow \{1, \dots, K_n\}$  return an index that satisfies these criteria. By assumption,  $\mathcal{Z} \subset \mathbb{R}^p$  has finite diameter,  $D$ , so we can construct this set with  $K_n \leq \Gamma_n (3D/\nu)^p$  (e.g., Shalev-Shwartz and Ben-David [85, pg. 337]).

By Lemma A.1 (and using the notation therein), we have

$$\Psi(z, \delta) \leq \Psi(\hat{z}_{k(z)}, \delta) + \nu \left( \alpha(LD + 1 + \sqrt{2 \ln 1/\delta}) + L(\sqrt{2 \ln 1/\delta} + 3) \right).$$

Taking the supremum over  $z$  of both sides, we get

$$\sup_z \Psi(z, \delta) \leq \max_k \Psi(\hat{z}_k, \delta) + \nu \left( \alpha(LD + 1 + \sqrt{2 \ln 1/\delta}) + L(\sqrt{2 \ln 1/\delta} + 3) \right).$$

If we let  $\nu = \frac{1}{3\gamma_n} (\alpha(LD + 1 + \sqrt{2}) + L(\sqrt{2} + 3))^{-1}$ , we have

$$\begin{aligned} & P(\sup_z \Psi(z, \delta) > 0 | X, Z) \\ & \leq P \left( \max_k \Psi(\hat{z}_k, \delta) + \nu \left( \alpha(LD + 1 + \sqrt{2 \ln 1/\delta}) + L(\sqrt{2 \ln 1/\delta} + 3) \right) > 0 \middle| X, Z \right) \\ & \leq P \left( \max_k \Psi(\hat{z}_k, \delta) + \nu \left( \alpha(LD + 1 + \sqrt{2}) + L(\sqrt{2} + 3) \right) \ln 1/\delta > 0 \middle| X, Z \right) \\ & \leq \sum_k P \left( \Psi(\hat{z}_k, \delta) + \frac{\ln 1/\delta}{3\gamma_n} > 0 \middle| X, Z \right) \\ & \leq \sum_k P \left( \Psi(\hat{z}_k, \sqrt{\delta}) > 0 \middle| X, Z \right) \\ & \leq K_n \sqrt{\delta}, \end{aligned}$$

where we have used the union bound and (A.2). Replacing  $\delta$  with  $\delta^2/K_n^2$  and integrating both sides to remove the conditioning completes the proof.  $\square$

*Proof of Theorem 2.2.* By Theorem 2.1, with probability at least  $1 - \delta/2$ ,

$$\begin{aligned} \mu(x, \hat{z}) & \leq \hat{\mu}(x, \hat{z}) + \frac{4}{3\gamma_n} \ln(2K_n/\delta) + \lambda_1 \sqrt{V(x, \hat{z})} + \lambda_2 B(x, \hat{z}) \\ & \leq \hat{\mu}(x, z^*) + \frac{4}{3\gamma_n} \ln(2K_n/\delta) + \lambda_1 \sqrt{V(x, z^*)} + \lambda_2 B(x, z^*), \end{aligned}$$

where the second inequality follows from the definition of  $\hat{z}$ . Using the same argument we used to derive (A.2), since  $z^*$  is not a random quantity, we have, with probability at least  $1 - \delta/2$ ,

$$\begin{aligned} \hat{\mu}(x, z^*) - \mu(x, z^*) & \leq \frac{2}{3\gamma_n} \ln(2/\delta) + \sqrt{2V(x, z^*) \ln(2/\delta)} + L \cdot B(x, z^*) \\ & \leq \frac{2}{3\gamma_n} \ln(2K_n/\delta) + \lambda_1 \sqrt{V(x, z^*)} + \lambda_2 B(x, z^*). \end{aligned}$$

Combining the two inequalities with the union bound yields the desired result.  $\square$

*Proof of Corollary 2.1.* We show  $\mu(x, \hat{z}) - 2L \cdot B(x, z^*) \rightarrow_p \mu(x, z^*)$ . The desired result follows from the assumption regarding  $B(x, z^*)$  and Slutsky's theorem. First, we note, due to the assumption  $|c(z; y)| \leq 1$ ,

$$V(x, z^*) = \sum_i w_i(x, z^*) \text{Var}(c(z^*; Y_i) | X_i, Z_i) \leq \frac{1}{\gamma_n} \sum_i w_i(x, z^*) = \frac{1}{\gamma_n}.$$

We have, for any  $\epsilon > 0$ ,

$$\begin{aligned} & P(|\mu(x, \hat{z}) - 2LB(x, z^*) - \mu(x, z^*)| > \epsilon) \\ & \leq P(\mu(x, \hat{z}) - 2LB(x, z^*) - \mu(x, z^*) > \epsilon/2) \\ & \quad + P(\mu(x, z^*) - \mu(x, \hat{z}) + 2LB(x, z^*) > \epsilon/2). \end{aligned}$$

By Theorem 2.2, for large enough  $n$ , the first term is upper bounded by

$$\begin{aligned} & 2K_n \exp\left(-\frac{\epsilon^2}{4(2/\gamma_n + 4\sqrt{V(x, z^*)})^2}\right) \\ & \leq 2K_n \exp\left(-\frac{\epsilon^2}{4(2/\sqrt{\gamma_n} + 4/\sqrt{\gamma_n})^2}\right) \\ & = 2\Gamma_n \left(9D\gamma_n \left(\alpha(LD + 1 + \sqrt{2}) + L(\sqrt{2} + 3)\right)\right)^p \exp\left(-\frac{\gamma_n \epsilon^2}{144}\right) \\ & \leq C_1 n^{1+\beta} \exp(-C_2 n^\beta) \rightarrow 0. \end{aligned}$$

Because  $\mu(x, z^*) \leq \mu(x, \hat{z})$ , the latter term is upper bounded by

$$P(B(x, z^*) > \epsilon/4L) \rightarrow 0.$$

$\square$

*Proof of Example 2.1.* First we consider the case that the zero variance action has cost 0, and the other actions have cost 1 (call this event  $A$ ). Because the cost of the optimal action is 0 and the cost of a suboptimal action is 1, the expected regret in

this problem equals the probability of the algorithm selecting a suboptimal action. Noting that  $\hat{\mu}(j) \sim \mathcal{N}(1, 1/m)$  for  $j = 1, \dots, m$ , we can express the expected regret of the predicted cost minimization algorithm as

$$\mathbb{E}[R^{PCM}|A] = P(\hat{\mu}(j) < 0 \text{ for some } j \in \{1, \dots, m\}|A) = P\left(\max_j W_j > \sqrt{m}\right),$$

where  $W_1, \dots, W_m$  are i.i.d. standard normal random variables. Similarly, the expected regret of the uncertainty penalized algorithm can be expressed as

$$\begin{aligned} \mathbb{E}[R^{UP}|A] &= P\left(\hat{\mu}(j) < -\frac{\lambda\sqrt{\ln m}}{\sqrt{m}} \text{ for some } j \in \{1, \dots, m\} \middle| A\right) \\ &= P\left(\max_j W_j > \sqrt{m} + \lambda\sqrt{\ln m}\right) \end{aligned}$$

We can construct an upper bound on  $\mathbb{E}R^{UP}$  with the union bound and a concentration inequality (as in the proof of Theorem 2.1). Applying the Gaussian tail inequality (see, for example, Vershynin [96, Proposition 2.1.2]), we have

$$\begin{aligned} \mathbb{E}[R^{UP}|A] &\leq mP(W_1 > \sqrt{m} + \lambda\sqrt{\ln m}) \\ &\leq \frac{\sqrt{m}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{m} + \lambda\sqrt{\ln m})^2\right) \\ &= \frac{\sqrt{m}}{m^{\lambda^2/2}\sqrt{2\pi}} \exp(-m/2) \exp(-\lambda\sqrt{m \ln m}) \\ &\leq \frac{1}{\sqrt{m}\sqrt{2\pi}} e^{-m/2}, \end{aligned}$$

where we have used the assumption  $\lambda \geq \sqrt{2}$ .

To lower bound the expected regret of the predicted cost minimization algorithm, we can use a similar Gaussian tail inequality.

$$\begin{aligned}
\mathbb{E}[R^{PCM}|A] &= 1 - [1 - P(W_1 > \sqrt{m})]^m \\
&\geq 1 - \left[1 - \left(1 - \frac{1}{m}\right) \frac{1}{\sqrt{m}\sqrt{2\pi}} e^{-m/2}\right]^m \\
&\geq 1 - \left[1 - \frac{1}{2\sqrt{m}\sqrt{2\pi}} e^{-m/2}\right]^m \\
&\geq 1 - \left[\left[1 - \frac{1}{2\sqrt{m}\sqrt{2\pi}} e^{-m/2}\right]^{2\sqrt{2\pi m} \exp(m/2)}\right]^{\sqrt{m} \exp(-m/2)/2\sqrt{2\pi}},
\end{aligned}$$

where the second inequality is valid for all  $m \geq 2$ . One can verify that  $(1 - 1/n)^n$  is a monotonically increasing function that converges to  $e^{-1}$ . Therefore, for all  $m \geq 2$ ,

$$\mathbb{E}[R^{PCM}|A] \geq 1 - \exp\left(-\frac{\sqrt{m}}{2\sqrt{2\pi}} \exp(-m/2)\right).$$

Next, we use these bounds to compute the ratio  $\mathbb{E}[R^{UP}|A]/\mathbb{E}[R^{PCM}|A]$  in the limit as  $m \rightarrow \infty$ .

$$\frac{\mathbb{E}[R^{UP}|A]}{\mathbb{E}[R^{PCM}|A]} \leq \frac{\frac{1}{\sqrt{m}\sqrt{2\pi}} e^{-m/2}}{1 - \exp\left(-\frac{\sqrt{m}}{2\sqrt{2\pi}} \exp(-m/2)\right)}.$$

Applying L'Hopital's rule, the limit of the right hand side is equal to the limit of

$$\begin{aligned}
&\frac{2(2\pi)^{-1/2} (-m^{-3/2} e^{-m/2} - m^{-1/2} e^{-m/2})}{(2\pi)^{-1/2} [m^{-1/2} e^{-m/2} - m^{1/2} e^{-m/2}] \exp\left(-\frac{\sqrt{m}}{2\sqrt{2\pi}} e^{-m/2}\right)} \\
&= 2 \frac{-1 - m}{m - m^2} \cdot \exp\left(\frac{\sqrt{m}}{2\sqrt{2\pi}} e^{-m/2}\right) \rightarrow 0.
\end{aligned}$$

Next, we consider the case that the zero variance action has cost 1, and the other actions have cost 0. The expected regret equals the probability that the zero variance

action is selected. For sufficiently large  $m$ ,

$$\begin{aligned}
\mathbb{E}[R^{UP}|A^c] &= P\left(\hat{\mu}(j) > 1 - \frac{\lambda\sqrt{\ln m}}{\sqrt{m}} \quad \forall j \in \{1, \dots, m\} \middle| A^c\right) \\
&\leq P(W_1 > \sqrt{m} - \lambda\sqrt{\ln m})^m \\
&\leq P(W_1 > \sqrt{m}/2)^m \\
&\leq \left(\frac{2}{\sqrt{2\pi}}e^{-m/8}\right)^m \\
&\leq e^{-m^2/8} = o(\mathbb{E}[R^{UP}|A]).
\end{aligned}$$

Therefore, for sufficiently large  $m$  and some constant  $C$ ,

$$\begin{aligned}
\frac{\mathbb{E}[R^{UP}]}{\mathbb{E}[R^{PCM}]} &= \frac{\mathbb{E}[R^{UP}|A] + \mathbb{E}[R^{UP}|A^c]}{\mathbb{E}[R^{PCM}|A] + \mathbb{E}[R^{PCM}|A^c]} \\
&\leq \frac{\mathbb{E}[R^{UP}|A] + \mathbb{E}[R^{UP}|A^c]}{\mathbb{E}[R^{PCM}|A]} \\
&\leq (1 + C)\frac{\mathbb{E}[R^{UP}|A]}{\mathbb{E}[R^{PCM}|A]} \rightarrow 0.
\end{aligned}$$

□

## A.2 Optimization with Linear Predictive Models

Here, we detail the optimization of (2) with linear predictive models. We focus on the case that  $c(z; Y) = Y$  for simplicity. For these models, we posit the outcome is a linear function of the auxiliary covariates and decision. That is there exists a  $\beta$  such that, given  $X = x$ ,  $Y(z) = (x, z)^T \beta + \epsilon$ , where  $\epsilon$  is a mean 0 subgaussian noise term with variance  $\sigma^2$ . If we let  $A$  denote the design matrix for the problem, a matrix with rows consisting of  $(X_i, Z_i)$  for  $i = 1, \dots, n$ , then the ordinary least squares (OLS) estimator for  $\beta$  is given by

$$\hat{\beta}^{OLS} = (A^T A)^{-1} A^T Y.$$

The ordinary least squares estimator is unbiased, so when solving (2), we set  $\lambda_2 = 0$ . The variance of  $(x, z)^T \hat{\beta}^{OLS}$  is given by  $\sigma^2 (x, z)^T (A^T A)^{-1} (x, z)$ .  $(A^T A)^{-1}$  is a positive



semidefinite matrix, so  $\sqrt{V(x, z)}$  is convex. Therefore, (2) becomes

$$\min_{z \in \mathcal{Z}} (x, z)^T \hat{\beta}^{OLS} + \lambda_1 \sigma \sqrt{(x, z)^T (A^T A)^{-1} (x, z)},$$

which is a second order conic optimization problem if  $\mathcal{Z}$  is polyhedral and can be solved efficiently by commercial solvers. Even if  $\mathcal{Z}$  is a mixed integer set, commercial solvers such as Gurobi [77] can still solve the problem for sizes of practical interest.

For regularized linear models such as ridge and lasso regression, we use a similar approach. Although these estimators are biased, we set  $\lambda_2 = 0$  for computational reasons. The ridge estimator for  $\beta$  has a similar form to the OLS estimator:

$$\hat{\beta}^{Ridge} = (A^T A + \alpha I)^{-1} A^T Y,$$

for some  $\alpha \geq 0$ . The resulting optimization problem is essentially the same as with the OLS estimator. The lasso estimator does not have a closed form solution, but we can approximate it as in Tibshirani [93]:

$$P \hat{\beta}^{Lasso} \approx (P A^T A P^T + \alpha P W)^{-1} P A^T Y,$$

where  $W = \text{diag}(1/|\beta_1^*|, \dots, 1/|\beta_{d+p}^*|)$ ,  $\beta^*$  is the true lasso solution, and  $P$  is a projection matrix that projects to the nonzero components of  $\beta^*$ . (The zero components of  $\beta^*$  are still 0 in the approximation.) With this approximation, the resulting optimization takes the same form as those for the OLS and ridge estimators.

## A.3 Data Generation

### A.3.1 Pricing

For our synthetic pricing example, we consider a store offering 5 products. We generate auxiliary covariates,  $X_i$ , from a  $\mathcal{N}(10, 1)$  distribution. We generate historical

prices,  $Z_i$ , from a Gaussian distribution,

$$\mathcal{N} \left( X_i^T \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}, 100I \right).$$

We compute the expected demand for each product as:

$$\mu = \begin{pmatrix} 500 - (Z_i^1)^2/10 - X_i^1 \cdot Z_i^1/10 - (X_i^1)^2/10 - Z_i^2 \\ 500 - (Z_i^2)^2/10 - X_i^1 \cdot Z_i^2/10 - (X_i^1)^2/10 - Z_i^1 \\ 500 - (Z_i^3)^2/10 - X_i^2 \cdot Z_i^3/10 - (X_i^2)^2/10 + Z_i^1 + Z_i^2 \\ 500 - (Z_i^4)^2/10 - X_i^2 \cdot Z_i^4/10 - (X_i^2)^2/10 + Z_i^1 + Z_i^2 \\ 500 - (Z_i^5)^2/10 - X_i^2 \cdot Z_i^5/20 - X_i^1 \cdot Z_i^5/20 - (X_i^2)^2/10 \end{pmatrix},$$

and generate  $Y_i$  from a  $\mathcal{N}(\mu, 2500I)$  distribution. This example serves to simulate the situation in which some products are complements and some are substitutes.

### A.3.2 Warfarin Dosing

To simulate how physicians might assign Warfarin doses to patients, we compute a normalized BMI for each patient (i.e. body mass divided by height squared, normalized by the population standard deviation of BMI). For each patient, we then sample a dose (in mg/week),  $Z_i$ , from

$$Z_i \sim \mathcal{N}(30 + 15 \cdot \text{BMI}_i, 64).$$

If  $Z_i$  is negative, we assign a dose drawn uniformly from  $[0, 20]$ . If the data dose not contain the patients height and/or weight, we assign a dose drawn uniformly from  $[10, 50]$ , a standard range for Warfarin doses.

To simulate the response that a physician observes for a particular patient, we

compute the difference between the the assigned dose and the true optimal dose for that patient,  $Z_i^*$ , and add noise. We then cap the response so it is less than or equal to 40 in absolute value. The reasoning behind this construction is that the INR measurement gives the physician some idea of whether the assigned dose is too high or too low and whether it is close to the optimal dose. However, if the dose is very far from optimal, then the information INR provides is not very useful in determining the optimal dose (it is purely directional). The response of patient  $i$  is given by

$$Y_i = \begin{cases} -40, & R_i < -40 \\ R_i, & -40 \leq R_i \leq 40, \\ 40, & R_i > 40 \end{cases}$$

where  $R_i \sim \mathcal{N}(Z_i - Z_i^*, 400)$ .

## A.4 Sensitivity to Selection of Tuning Parameters

To test the sensitivity of the method to the selection of tuning parameters, we conduct an experiment on the Warfarin example with the random forest as the base learner. We compute the out-of-sample error for many combinations of  $\lambda_1$  and  $\lambda_2$ . From Figure A-1, we see that the out-of-sample performance is not too sensitive to the selection of parameters. All of the selected parameter combinations out-perform the unpenalized method with the exception of  $(\lambda_1 = 100, \lambda_2 = 0)$ , which is an extreme choice. This demonstrates that the tuning parameter selection does not have to be extremely precise to improve performance.

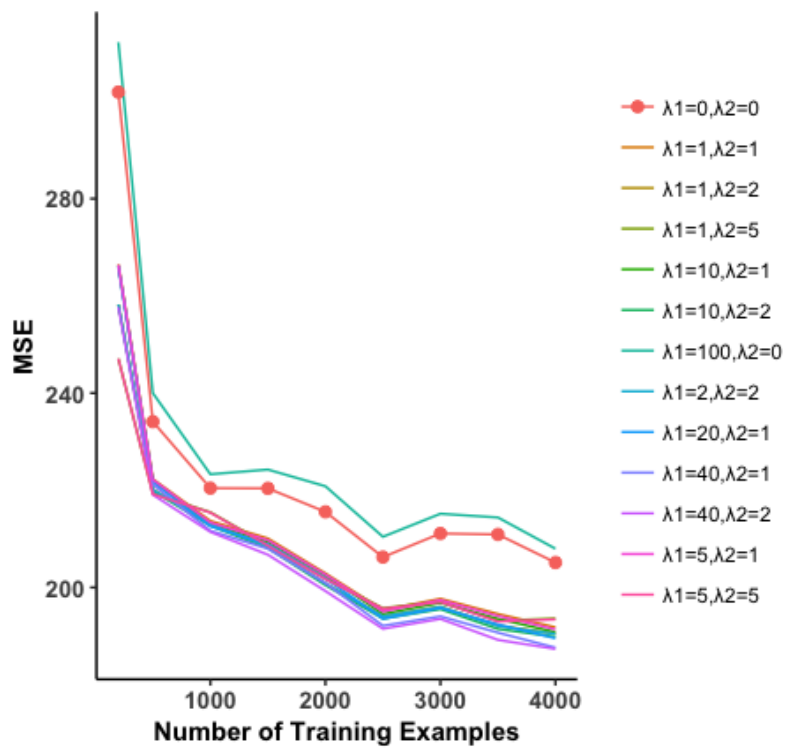


Figure A-1: Effect of  $\lambda$  tuning on the random forest method for Warfarin example

# Appendix B

## Appendix for Chapter 3

### B.1 Additional Results on Random Forest Weight Functions

Here we provide an additional result on the strong asymptotic optimality of the method with random forest weight functions in the single stage setting. Here, we consider random forests as defined in Wager and Athey [97]. The random forest consists of an ensemble of trees, each trained on a subsample of the data of size  $s$ . Each of the trees in the forest is a regular, random-split, and honest regression tree as in Definition 3.6. The prediction of the random forest at  $x$  is given by

$$R(x, X_1, Y_1, \dots, X_N, Y_N) = \binom{N}{s}^{-1} \sum_{i_1, \dots, i_s} \mathbb{E}_\xi [T(x, \xi, X_{i_1}, Y_{i_1}, \dots, X_{i_s}, Y_{i_s})]. \quad (\text{B.1})$$

In practice, this is estimated by training trees on random subsamples of the data and random draws of  $\xi$ .

**Assumption B.1** (Random Forest Specification). *The random forest, as defined in (B.1), has random-split, regular, and honest regression trees as its base learners. In addition,  $\lambda \leq 0.2$ , and the subsample size,  $s_N$ , scales with  $N^\beta$ . That is,  $s_N = \min(\lfloor CN^\beta \rfloor, N - 1)$  with  $C > 0$  and  $\beta \in \left[ \left( 2 + \frac{\pi \log(1-\lambda)}{d \log \lambda} \right)^{-1}, \frac{1}{2} \right)$ .*

This assumption ensures the forest consists of diverse trees, each with low bias,

so they can be aggregated into a consistent regressor.

**Theorem B.1.** *Suppose  $\mathcal{Z} \subset \mathbb{R}^p$  is nonempty and compact, and the training data is i.i.d. In addition, suppose  $|c(z; y)| \leq 1$ , for all  $z \in \mathcal{Z}$ ,  $y \in \mathcal{Y}$ , and that  $c(z; y)$  is a well-defined  $L$ -Lipschitz continuous function of  $z$  for all  $y$ . Finally, suppose  $X^i$  is uniformly distributed on  $[0, 1]^d$  and  $\mathbb{E}[c(z; Y)|X = x]$  is an  $M$ -Lipschitz continuous function of  $x$  for all  $z$ . Let  $w_{N,i}(x)$  be the random forest weight function, satisfying assumption B.1. Then  $\{\hat{z}_0^N(x)\}$ , a sequence of optimal solutions to*

$$\min_{z \in \mathcal{Z}} \sum_{i=1}^N w_{N,i}(x) c(z; y^i),$$

*is strongly asymptotically optimal with respect to the true problem,  $\min_{z \in \mathcal{Z}} \mathbb{E}[c(z; Y)|X = x]$ .*

This result shows that our method can be strongly asymptotically optimal for the single stage problem with random forest weight functions. Some of the assumptions are slightly stronger than in Theorem 3.3. For example, we require that the value functions are bounded and the random forests are slightly different. However, we do not require that the minimum number of training samples per leaf,  $k$ , grows with  $N$  as we do for Theorem 3.3. This is consistent with Breiman's original random forest algorithm in which  $k$  is fixed at 1. To prove this theorem, we first prove a result on the strong consistency of the random forest estimator.

**Lemma B.1.** *Suppose  $(X_1, Y_1), \dots, (X_N, Y_N) \in \mathbb{R}^{d+1}$  are i.i.d. samples, the distribution of  $X_1$  uniform on  $[0, 1]^d$ , and  $|Y| \leq 1$ , a.s. Let  $\mu(x) = \mathbb{E}[Y_1|X_1 = x]$  be a Lipschitz continuous function. Define  $\hat{\mu}_N(x)$  to be the prediction of a random forest that satisfies Assumption B.1. Then, almost surely,*

$$\hat{\mu}_N(x) \rightarrow \mu(x)$$

*for  $x$  a.e.*

*Proof.* We first note that the bias of the predictions goes to 0,  $|\mathbb{E}\hat{\mu}_N(x) - \mu(x)| \rightarrow 0$ ,

for  $x$  a.e. by Theorem 3 from [97]. (The assumptions are satisfied by Assumption B.1.) Next, we define  $h_x(Z_1, \dots, Z_N) = \hat{\mu}_N(x)$ , where  $Z_i = (X_i, Y_i)$ . We note that for any  $Z_k, Z'_k \in \mathcal{X} \times \mathcal{Y}$ ,

$$|h_x(Z_1, \dots, Z_k, \dots, Z_N) - h_x(Z_1, \dots, Z'_k, \dots, Z_N)| \leq \frac{2s_N}{N}.$$

This is due to the assumption that  $|Y_i| \leq 1$ . Since each tree predicts at  $x$  by averaging the  $Y_i$ s corresponding to training samples in the same partition of the feature space as  $x$ , the prediction of any tree is bounded between -1 and 1. Changing a single training sample only affects the trees trained on subsets of the data including that example. This only affects a fraction  $s_N/N$  of the trees in the forest. Since the prediction of the random forest is the average of the predictions of all the trees, the most the prediction can be changed by altering a single training sample is  $2s_N/N$ . Applying McDiarmid's inequality, we have, for  $\epsilon > 0$ ,

$$P(|\hat{\mu}_N(x) - \mathbb{E}\hat{\mu}_N(x)| > \epsilon) \leq 2 \exp\left(-\frac{N\epsilon^2}{2s_N^2}\right).$$

By our assumption, this can be rewritten

$$P(|\hat{\mu}_N(x) - \mathbb{E}\hat{\mu}_N(x)| > \epsilon) \leq 2 \exp\left(-\frac{N^{1-2\beta}\epsilon^2}{C^2 2}\right),$$

where  $\beta < 1/2$ . From this we see  $\sum_{N=1}^{\infty} P(|\hat{\mu}_N(x) - \mathbb{E}\hat{\mu}_N(x)| > \epsilon) < \infty$ , so, by the Borel Cantelli lemma,  $|\hat{\mu}_N(x) - \mathbb{E}\hat{\mu}_N(x)| \rightarrow 0$  a.s. Combining the two results with the triangle inequality completes the proof.  $\square$

*Theorem B.1.* We need to show

$$\sup_{z \in \mathcal{Z}} \left| \sum_{i=1}^N w_{N,i}(x) c(z; y^i) - \mathbb{E}[c(z; Y) | X = x] \right| \rightarrow 0$$

a.s. for  $x$  a.e. The desired result then follows from lemmas 3.2 and 3.3. If we ignore the supremum, we can apply lemma B.1 to see it goes to 0 a.s. Next, we apply lemma 3.4 to see that the convergence holds simultaneously for all  $z$  with probability

1. Finally, we apply lemma 3.1 to show the convergence is uniform over  $z$  a.s.  $\square$

## B.2 Decomposition Algorithm

For the reader's convenience, we present a decomposition algorithm, similar to that of Shapiro [88], tailored for use with our methods. Given weight functions, which we compute from an appropriate machine learning algorithm, our goal is to solve (3.3). If the decisions spaces and state spaces are finite sets with small cardinality, it may be possible to solve the problem exactly using the classical dynamic programming algorithm (see, for example, Bertsekas [11]). However, in many OR problems, this is not the case, and we need to deal with continuous decision and state spaces.

We note that it is possible to formulate (3.3) as a large, but finite sized, single stage optimization problem. To do so, we create  $N$  copies of  $z_1$ ,  $N^2$  copies of  $z_2$ , etc. These copies of the decision variables represent our contingency plan. For each potential realization of  $Y_1, \dots, Y_t$ , we have a distinct copy of  $z_t$ . If the  $f_t$  and  $g_t$  functions are linear and the  $Z_t$  sets are polyhedral, the resulting problem will be a linear optimization problem, which can be solved in time that is polynomial in the size of the formulation. However, the number of variables in the formulation is  $O(N^T)$ , and this formulation becomes impractical for moderately sized  $N$  and  $T$ . In order to solve larger problems, we resort to a Benders-like decomposition approach. (For a review of Benders decomposition methods, see, for example, Murphy [76].)

Algorithm 1 describes the approach. The main idea is that we maintain a piecewise linear, convex lower bound,  $\psi_t(s_t, x_{t-1})$ , on  $\sum_{i=1}^N w_{N,i}^t(x_{t-1}) \hat{Q}_t(s_t; y_t^i, x_t^i)$  for each  $t = 1, \dots, T$ . We have the relaxed problems:

$$P_t(s_t, x_t, y_t, \psi_{t+1}) := \min_{z_t \in Z_t(s_t, y_t)} g_t(z_t) + \psi_{t+1}(f_t(z_t), x_t) \quad (\text{B.2})$$

(For  $t = 0$ , there is no dependence on  $y_0$ .) If  $g_t$  and  $f_t$  are linear functions and  $Z_t$  is a polyhedral set, then the relaxed problem can be reformulated as a linear optimization problem because  $\psi_{t+1}$  is the maximum of a finite set of linear functions.



The algorithm consists of two main steps, which are repeated until convergence. First,  $M$  sample paths for  $X$  and  $Y$  are sampled (with replacement) from the training data. Then, in the forward step, for each of these sample paths, trial states are computed by solving the relaxed problems from  $t = 0$  to  $t = T$ , assuming the system evolves according to the corresponding sample path. Using the costs of each of these sample paths, we can compute a statistical upper bound on the optimal value of (3.3) (assuming  $M$  and  $N$  are such that a central limit theorem is a reasonable approximation).

In the backward step, we update the  $\psi_t$  functions. We proceed backwards, starting with  $t = T$ , and solve  $P_t$  for each of the trial states. We compute the cut coefficients (which we will discuss in more detail next), and then average them across all possible realizations  $y_t^j$  and  $x_t^j$ , according to the distribution  $\{w_{N,i}^t(x_{t-1}^j)\}_{i=1}^N$ . We then update  $\psi_t$  with this new cut. Finally, we update the lower bound on the optimal value of the problem by solving  $P_0(s_0, x_0, \psi_1)$ .

There are several possible stopping criteria we can use. One is to stop when the statistical upper bound (line 15 in Algorithm 1) is within a specified  $\epsilon > 0$  of the lower bound (line 26). This will give us an  $\epsilon$ -optimal solution with probability at least  $1 - \alpha$  (assuming  $M$  is sufficiently large and  $N$  is much larger than  $M$  so that the central limit theorem is a reasonable approximation). Alternatively, we can stop when the lower bound stabilizes or after a fixed number of iterations. All three of these can allow the algorithm to construct lower bounds,  $\{\psi_t\}$ , that are reasonable approximations to the value functions.

The crucial component required for this algorithm to work is the cuts. We begin with a definition from [101].

**Definition B.1** (Valid, tight, and finite cut). *Let  $(\beta_t, \pi_t)$  be the stage  $t$  cut coefficients computed in the backward step of Algorithm 1. We say that the cut is:*

1. *Valid if*

$$\sum_{i=1}^N w_{N,i}^t(\hat{x}_{t-1}) \hat{Q}_t(s_t; y_t^i, x_t^i) \geq \beta_t + \pi_t^T s_t \quad \forall s_t.$$

---

**Algorithm 1:** SDDP algorithm for multistage optimization.

---

**Input:**  $s_0, x_0$ , weight functions  $\{w_{N,i}^t(x_{t-1})\}_{i=1,\dots,N;t=1,\dots,T}$

- 1 Initialize:  $LB \leftarrow -\infty, UB \leftarrow \infty$ , and initial lower bounds  $\{\psi_t\}_{t=1}^T$
- 2 **while** *stopping criterion not satisfied* **do**
- 3     Sample  $M$  scenarios  $\{y_1^j, x_1^j, y_2^j, \dots, x_{T-1}^j, y_T^j\}_{j=1}^M$  with replacement from the training set, such that  $y_1^j$  and  $x_1^j$  are sampled from the probability distribution defined by  $w_{N,i}^1(x_0)$ ,  $y_2^j$  and  $x_2^j$  are sampled from the probability distribution defined by  $w_{N,i}^2(x_1^j)$ , etc.
- 4     /\* Forward step \*/
- 5     **for**  $j = 1, \dots, M$  **do**
- 6         Initialize  $s_0^j \leftarrow s_0$
- 7         **for**  $t = 0, \dots, T$  **do**
- 8             Solve problem  $P_t(s_t^j, x_t^j, y_t^j, \psi_{t+1})$  for optimal solution  $z_t^j$
- 9             Update state  $s_{t+1}^j \leftarrow f_t(z_t^j)$
- 10         **end**
- 11          $v^j \leftarrow \sum_{t=0}^T g_t(z_t^j)$
- 12     **end**
- 13     /\* Statistical upper bound \*/
- 14      $\hat{\mu} \leftarrow \frac{1}{M} \sum_{j=1}^M v^j$  and  $\hat{\sigma}^2 \leftarrow \frac{1}{M-1} \sum_{j=1}^M (v^j - \hat{\mu})^2$
- 15      $UB \leftarrow \hat{\mu} + z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{M}}$
- 16     /\* Backward Step \*/
- 17     **for**  $t = T, \dots, 1$  **do**
- 18         **for**  $j = 1, \dots, M$  **do**
- 19             **for**  $i = 1, \dots, N$  **do**
- 20                 Solve  $P_t(s_t^j, x_t^i, y_t^i, \psi_{t+1})$  to compute cut coefficients  $(\beta_t^{ji}, \pi_t^{ji})$
- 21                 **end**
- 22                 Update
- 23                 
$$\psi_t(s_t, x_{t-1}^j) \leftarrow \max \left\{ \psi_t(s_t, x_{t-1}^j), \sum_{i=1}^N w_{N,i}^t(x_{t-1}^j) (\beta_t^{ji} + (\pi_t^{ji})^T s_t) \right\}$$
- 24             **end**
- 25         **end**
- 26     /\* Lower bound update \*/
- 27     Solve  $P_0(s_0, x_0, \psi_1)$  and set  $LB$  to its optimal value
- 28 **end**

---

2. *Tight if*

$$\sum_{i=1}^N w_{N,i}^t(\hat{x}_{t-1}) P_t^*(\hat{s}_t, x_t^i, y_t^i, \psi_{t+1}) = \beta_t + \pi_t^T \hat{s}_t,$$

where  $P_t^*$  represents the optimal value of (B.2),  $\hat{s}_t$  is the trial state computed during the corresponding forward pass of the algorithm, and  $\hat{x}_{t-1}$  is the auxiliary covariate from the forward pass of the algorithm.

3. *Finite if solving (B.2) for fixed  $\psi_{t+1}$  can only generate finitely many possible cuts.*

Under the conditions that  $f_t$  and  $g_t$  are linear functions,  $Z_t$  are polyhedral sets, and at every stage, (B.2) is feasible with finite optimal value, it is shown in [88] that the SDDP algorithm will converge to an optimal solution in a finite number of iterations with probability 1, provided the cuts used are valid, tight, and finite. [101] showed that this result also holds if the state variables are purely binary, instead of continuous.

The validity of the cuts ensures that the  $\{\psi_t\}$  functions maintain lower bounds on the value functions at each stage. Next, we describe several classes of valid cuts that can be used within the SDDP algorithm.

## Benders' Cut

A well known class of cuts is the Bender's cut [8]. These cuts are valid for linear problems, even with integer constraints, and are tight for linear optimization problems (or more generally convex optimization problems) in which strong duality holds. To compute the cuts for stage  $t$  in the SDDP algorithm we solve the following form of  $P_t(s_t^j, x_t^i, y_t^i, \psi_{t+1})$ :

$$\begin{aligned} \min_{z_t, s_t} \quad & g_t(z_t) + \psi_{t+1}(f_t(z_t), x_t^i) \\ \text{s.t.} \quad & z_t \in Z_t(s_t, y_t^i) \\ & s_t = s_t^j. \end{aligned}$$

We then let  $\pi_t^{ji}$  be the optimal dual solution (of the LO relaxation if there are integer variables) corresponding to the indicated constraint and set  $\beta_t^{ji} =$

$P_t^*(s_t^j, x_t^i, y_t^i, \psi_{t+1}) - (\pi_t^{ji})^T s_t^j$ , where  $P_t^*(s_t^j, x_t^i, y_t^i, \psi_{t+1})$  is the optimal value of the above problem. In order for these cuts to be finite, we should always use basic solutions for  $\pi_t^{ji}$ .

### Integer Optimality Cut

If the state space is binary, the SDDP algorithm with Benders' cuts is not guaranteed to produce an optimal solution. This is because they are not guaranteed to be tight. Instead, we can solve the above integer optimization problem to optimality and choose cuts defined by the linear expression:

$$(P_t^*(s_t^j, x_t^i, y_t^i, \psi_{t+1}) - L_t) \left( \sum_k (s_{tk}(1 - s_{tk}^j) + (1 - s_{tk})s_{tk}^j) \right) + P_t^*(s_t^j, x_t^i, y_t^i, \psi_{t+1}).$$

These cuts are valid, tight, and finite when the state space is binary. However, they tend to be very ineffective in practice.

### Lagrangian Cut

The third class of cut we describe was introduced by [101] and shown to be valid and tight when the state space is binary. They are much more effective than the integer optimality cuts in practice. These cuts are computed by solving the Lagrangian dual problem:

$$\max_{\pi_t^{ji}} \mathcal{L}_t(\pi_t^{ji}) + (\pi_t^{ji})^T s_t^j,$$

where

$$\begin{aligned} \mathcal{L}_t(\pi_t^{ji}) &= \min_{z_t, s_t} g_t(z_t) + \psi_{t+1}(f_t(z_t), x_t^i) - (\pi_t^{ji})^T s_t \\ \text{s.t.} \quad & z_t \in Z_t(s_t, y_t^i) \\ & s_t \in [0, 1]^p. \end{aligned}$$

We then use the cut with  $\pi_t^{ji}$  equal to the optimal solution of the Lagrangian dual problem and  $\beta_t^{ji}$  equal to the optimal value of  $\mathcal{L}_t(\pi_t^{ji})$ .

These three classes of cuts allow us to solve problems where the state space is continuous or pure binary with the SDDP algorithm. When the state space is a

mixed integer set, we can perform a binary expansion to desired accuracy on the continuous variable to convert the problem to the pure binary case. Of course, we can also combine different classes of cuts, and this can speed convergence.



# Appendix C

## Appendix for Chapter 4

### C.1 Properties of Weight Functions

In this section, we show that the  $k$ -nearest neighbor and kernel regression weight functions satisfy several guarantees. These results are used in the proof of Theorem 4.2 in Section 4.4.3. The main result of this section is the following. For convenience, the equations below are numbered the same as in the proof of Theorem 4.2 in Section 4.4.3.

**Theorem C.1.** *If Assumptions 4.1 and 4.4 hold, then*

$$\{w_N^i(\bar{\gamma})\} \text{ are not functions of } \xi^1, \dots, \xi^N; \quad (4.4)$$

$$\sum_{i=1}^N w_N^i(\bar{\gamma}) = 1 \text{ and } w_N^1(\bar{\gamma}), \dots, w_N^N(\bar{\gamma}) \geq 0, \quad \forall N \in \mathbb{N}. \quad (4.5)$$

Moreover, there exists constants  $k_2 > 0$  and  $\eta > p(2 + d_\xi)$  such that

$$\lim_{N \rightarrow \infty} \frac{1}{\epsilon_N} \sum_{i=1}^N w_N^i(\bar{\gamma}) \|\gamma^i - \bar{\gamma}\| = 0, \quad \mathbb{P}^\infty\text{-almost surely}; \quad (4.7)$$

$$\mathbb{E}_{\mathbb{P}^N} \left[ \exp \left( \frac{-\theta}{\sum_{i=1}^N w_N^i(\bar{\gamma})^2} \right) \right] \leq \exp(-k_2 \theta N^\eta), \quad \forall \theta \in (0, 1), N \in \mathbb{N}. \quad (4.8)$$

*Proof.* We observe that (4.4) and (4.5) follow directly from the definitions of the weight functions. The proofs of (4.7) and (4.8) are split into two parts, one for the

$k$ -nearest neighbor weights and one for kernel regression weights.

### **k-Nearest Neighbors:**

For the proof of (4.7), we note

$$\sum_{i=1}^N w_N^i(\bar{\gamma}) \|\gamma^i - \bar{\gamma}\| \leq \|\gamma^{(k_N)}(\bar{\gamma}) - \bar{\gamma}\|,$$

where  $\gamma^{(k_N)}(\bar{\gamma})$  denotes the  $k_N$ th nearest neighbor of  $\bar{\gamma}$  out of  $\gamma^1, \dots, \gamma^N$ . Therefore, for any  $\lambda > 0$ ,

$$\begin{aligned} \mathbb{P}^N \left( \sum_{i=1}^N w_N^i(\bar{\gamma}) \|\gamma^i - \bar{\gamma}\| > \lambda \epsilon_N \right) &\leq \mathbb{P}^N (\|\gamma^{(k_N)}(\bar{\gamma}) - \bar{\gamma}\| > \lambda \epsilon_N) \\ &\leq \mathbb{P}^N (|\{i : \|\gamma^i - \bar{\gamma}\| \leq \lambda \epsilon_N\}| \leq k_N - 1). \end{aligned}$$

By Assumption 4.4, this probability is upper bounded by  $\mathbb{P}(\beta \leq k - 1)$ , where  $\beta \sim \text{Binom}(N, g(\lambda \epsilon_N)^{d_\gamma})$ . By Hoeffding's inequality,

$$\mathbb{P}^N \left( \sum_{i=1}^N w_N^i(\bar{\gamma}) \|\gamma^i - \bar{\gamma}\| > \lambda \epsilon_N \right) \leq \exp \left( \frac{-2(Ng(\lambda k_1/N^p)^{d_\gamma} - k_N + 1)^2}{N} \right),$$

for  $k_N \leq Ng(\lambda k_1/N^p)^{d_\gamma} + 1$ . We note that this condition on  $k_N$  is satisfied for  $N$  sufficiently large because  $\delta + pd_\gamma < 1$  by Assumption 4.1. Because the right hand side in the above inequality has a finite sum over  $N$ , (4.7) follows by the Borel Cantelli lemma.

For the proof of (4.8), it follows from Assumption 4.1 that

$$\sum_{i=1}^N w_N^i(\bar{\gamma})^2 \leq k_3 N^{1-2\delta}$$

deterministically (for all sufficiently large  $N$  such that  $\lceil k_3 N^\delta \rceil \leq N - 1$ ) and  $p(2 + d_\xi) > 2\delta - 1$ . Thus, (4.8) follows with  $\eta = 2\delta - 1$ .



**Kernel regression:**

Assumption 4.1 stipulates that the kernel function  $K(\cdot)$  is Gaussian, triangular, or Epanechnikov, which are defined in Section 4.3. It is easy to verify that these kernel functions satisfy the following:

1.  $K$  is nonnegative, finite valued, and monotonically decreasing (for nonnegative inputs).
2.  $u^\alpha K(u) \rightarrow 0$  as  $u \rightarrow \infty$  for any  $\alpha \in \mathbb{R}$ .
3.  $\exists u^* > 0$  such that  $K(u^*) > 0$ .

For the proof of (4.7), define  $q > 0$  such that  $p < q < \delta$ . Letting  $D$  be the diameter of  $\Gamma$  and  $g_N(\bar{\gamma}) = \sum_{i=1}^N K(\|\gamma^i - \bar{\gamma}\|/h_N)$ , we have

$$\begin{aligned}
& \sum_{i=1}^N w_N^i(\bar{\gamma}) \|\gamma^i - \bar{\gamma}\| \\
&= \sum_{i=1}^N w_N^i(\bar{\gamma}) \mathbf{1}\{\|\gamma^i - \bar{\gamma}\| \leq N^{-q}\} \|\gamma^i - \bar{\gamma}\| \\
&\quad + \frac{1}{g_N(\bar{\gamma})} \sum_{i=1}^N K\left(\frac{\|\gamma^i - \bar{\gamma}\|}{h_N}\right) \mathbf{1}\{\|\gamma^i - \bar{\gamma}\| > N^{-q}\} \|\gamma^i - \bar{\gamma}\| \\
&\leq N^{-q} + \frac{NDK(N^{-q}/h_N)}{g_N(\bar{\gamma})},
\end{aligned}$$

where the inequality follows from the monotonicity of  $K$ . By construction,  $N^{-q}/\epsilon_N \rightarrow 0$ , so we just need to handle the second term. We note, for any  $\lambda > 0$ ,

$$\mathbb{P}^N \left( \frac{NDK(N^{-q}/h_N)}{g_N(\bar{\gamma})} > \lambda \epsilon_N \right) \leq \mathbb{P}^N \left( \sum_{i=1}^N Z_i^N K(u^*) < \frac{NDK(N^{-q}/h_N)}{\lambda \epsilon_N} \right),$$

where  $Z_i^N = \mathbf{1}\{\|\gamma^i - \bar{\gamma}\| \leq u^* h_N\}$ . To achieve this inequality, we lower bounded each term in  $g_N(\bar{\gamma})$  by  $K(u^*)$  or 0, because of the monotonicity of  $K$ . By Hoeffding's

inequality,

$$\begin{aligned}
& \mathbb{P}^N \left( \sum_{i=1}^N Z_i^N K(u^*) < \frac{NDK(N^{-q}/h_N)}{\lambda\epsilon_N} \right) \\
& \leq \exp \left( - \frac{2 \left( N\mathbb{E}Z_i^N - \frac{ND}{\lambda\epsilon_N K(u^*)} K(N^{-q}/h_N) \right)_+^2}{N} \right) \\
& \leq \exp \left( - \frac{2 \left( Ng(u^*h_N)^{d_\gamma} - \frac{ND}{\lambda\epsilon_N K(u^*)} K(N^{-q}/h_N) \right)_+^2}{N} \right) \\
& = \exp \left( - \left( k_5 N^{1/2-\delta d_\gamma} - k_6 N^{1/2+p} K(k_4 N^{-q+\delta}) \right)_+^2 \right),
\end{aligned}$$

for some constants  $k_5, k_6 > 0$  that do not depend on  $N$ . We used Assumption 4.4 for the second inequality. Because  $\delta > q$ , the second kernel property implies  $N^{1/2+p}K(k_4 N^{-q+\delta})$  goes to 0 as  $N$  goes to infinity, so that term is irrelevant. Because  $1/2 - \delta d_\gamma > 0$  by Assumption 4.1, the right hand side of the inequality has a finite sum over  $N$ , and thus (4.7) follows from the Borel Cantelli lemma.

For the proof of (4.8), define

$$v^N = \begin{pmatrix} K(\|\gamma^1 - \bar{\gamma}\|/h_N) \\ \vdots \\ K(\|\gamma^N - \bar{\gamma}\|/h_N) \end{pmatrix}.$$

We note that

$$\sum_{i=1}^N w_N^i(\bar{\gamma})^2 = \frac{\|v^N\|_2^2}{\|v^N\|_1^2} \leq \frac{\|v^N\|_\infty}{\|v^N\|_1} \leq \frac{K(0)}{K(u^*) \sum_{i=1}^N Z_i^N},$$

where  $Z_i^N$  is defined above. The first inequality follows from Holder's inequality, and the second inequality follows from the monotonicity of  $K$ . Next, we define  $\bar{Z}_i^N$  to be

a Bernoulli random variable with parameter  $g(u^*h_N)^{d_\gamma}$  for each  $i$ . For any  $\theta \in (0, 1)$ ,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^N} \left[ \exp \left( \frac{-\theta}{\sum_{i=1}^N w_N^i(\bar{\gamma})^2} \right) \right] &\leq \mathbb{E}_{\mathbb{P}^N} \left[ \exp \left( \frac{-\theta K(u^*) \sum_{i=1}^N \bar{Z}_i^N}{K(0)} \right) \right] \\
&= (1 - g(u^*h_N)^{d_\gamma} + g(u^*h_N)^{d_\gamma} \exp(-\theta K(u^*)/K(0)))^N \\
&\leq \exp(-Ng(u^*h_N)^{d_\gamma}(1 - \exp(-\theta K(u^*)/K(0)))) \\
&\leq \exp \left( -Ng(u^*h_N)^{d_\gamma} \frac{\theta K(u^*)}{2K(0)} \right) \\
&= \exp \left( -\frac{\theta K(u^*)g(k_4u^*)^{d_\gamma} N^{1-\delta d_\gamma}}{2K(0)} \right).
\end{aligned}$$

The first inequality follows because  $g(u^*h_N)^{d_\gamma}$  is an upper bound on  $\mathbb{P}(\|\gamma^i - \bar{\gamma}\| \leq u^*h_N)$  by Assumption 4.4. The first equality follows from the definition of the moment generating function for a binomial random variable. The next line follows from the inequality  $e^x \geq 1+x$  and the following from the inequality  $1-e^{-x} \geq x/2$  for  $0 \leq x \leq 1$ . Because  $1 - \delta d_\gamma > p(2 + d_\xi)$ , this completes the proof of (4.8) with  $\eta = 1 - \delta d_\gamma$  and  $k_2 = K(u^*)g(k_4u^*)^{d_\gamma}/2K(0)$ . □

## C.2 Proofs from Section 4.4.4

To connect Theorem 4.2, a result regarding concentration in 1-Wasserstein distance, to sample robust optimization, we consider the  $\infty$ -Wasserstein metric, which is given by:

$$d_\infty(\mathbb{Q}, \mathbb{Q}') \equiv \inf \left\{ \Pi\text{-ess sup}_{\Xi \times \Xi} \|\xi - \xi'\| : \begin{array}{l} \Pi \text{ is a joint distribution of } \xi \text{ and } \xi' \\ \text{with marginals } \mathbb{Q} \text{ and } \mathbb{Q}', \text{ respectively} \end{array} \right\},$$

where the essential supremum of the joint distribution is defined as

$$\Pi\text{-ess sup}_{\Xi \times \Xi} \|\xi - \xi'\| = \inf \{M : \Pi(\|\xi - \xi'\| > M) = 0\}.$$

We make use of the following result from Bertsimas, Shtern, and Sturt [22]:

**Lemma C.1.** *For any measurable  $f : \Xi \rightarrow \mathbb{R}$ ,*

$$\sum_{i=1}^N w_i^N(\bar{\gamma}) \sup_{\zeta \in \mathcal{U}_N^i} f(\zeta) = \sup_{\mathbb{Q} \in \mathcal{P}(\Xi): d_\infty(\hat{\mathbb{P}}_N^{\bar{\gamma}}, \mathbb{Q}) \leq \epsilon_N} \mathbb{E}_{\xi \sim \mathbb{Q}}[f(\xi)].$$

The proof of Lemma C.1 follows identical reasoning as in Bertsimas, Shtern, and Sturt [22] and is thus omitted.

Next, we state a result from Bertsimas, Shtern, and Sturt [22] (their Theorem EC.1), which bounds the difference in worst case objective values between 1-Wasserstein and  $\infty$ -Wasserstein distributionally robust optimization problems. We note that Bertsimas, Shtern, and Sturt [22] proved the following result for the case that  $\mathbb{Q}'$  is the unweighted empirical measure, but their proof carries through for the case here in which  $\mathbb{Q}'$  is a weighted empirical measure.

**Lemma C.2.** *Let  $\mathcal{Z} \subseteq \mathbb{R}^d$ ,  $f : \mathcal{Z} \rightarrow \mathbb{R}$  be measurable, and  $\zeta^1, \dots, \zeta^N \in \mathcal{Z}$ . Suppose that*

$$\mathbb{Q}' = \sum_{i=1}^N w^i \delta_{\zeta^i}$$

*for given weights  $w^1, \dots, w^N \geq 0$  that sum to one. If  $\theta_2 \geq 2\theta_1 \geq 0$ , then*

$$\sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z}): d_1(\mathbb{Q}', \mathbb{Q}) \leq \theta_1} \mathbb{E}_{\xi \sim \mathbb{Q}}[f(\xi)] \leq \sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z}): d_\infty(\mathbb{Q}', \mathbb{Q}) \leq \theta_2} \mathbb{E}_{\xi \sim \mathbb{Q}}[f(\xi)] + \frac{4\theta_1}{\theta_2} \sup_{\zeta \in \mathcal{Z}} |f(\zeta)|.$$

We now restate and prove the main result, which combines the new measure concentration result from this paper with similar proof techniques as Bertsimas, Shtern, and Sturt [22] and Esfahani and Kuhn [48].

**Theorem 4.1.** *Suppose the weight function and uncertainty sets satisfy Assumption 4.1, the joint probability distribution of  $(\gamma, \xi)$  satisfies Assumptions 4.2-4.4 from Section 4.4.3, and the cost function satisfies Assumption 4.5 from Section 4.4.4. Then, for every  $\bar{\gamma} \in \Gamma$ ,*

$$\lim_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) = v^*(\bar{\gamma}), \quad \mathbb{P}^\infty\text{-almost surely.}$$

*Proof.* We break the limit into upper and lower parts. The proof of the lower part follows from an argument similar to that used by Bertsimas, Shtern, and Sturt [22]. The proof of the upper part follows from the argument used by Esfahani and Kuhn [48].

### Lower bound.

We first show

$$\liminf_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) \geq v^*(\bar{\gamma}), \quad \mathbb{P}^\infty\text{-almost surely.} \quad (\text{C.1})$$

To begin, we define

$$D_N := \{\boldsymbol{\zeta} : \|\boldsymbol{\zeta}\| \leq \log N\},$$

and let  $\mathbb{P}_{\bar{\gamma}|D_N}(\cdot)$  be shorthand for  $\mathbb{P}(\cdot \mid \gamma = \bar{\gamma}, \boldsymbol{\xi} \in D_N)$ . Then, applying Assumption 4.2,

$$\begin{aligned} & \mathbb{P}^N (\cup_{i=1}^N \mathcal{U}_N^i \not\subseteq D_N) & (\text{C.2}) \\ & \leq \mathbb{P} \left( \max_{i \leq N} \|\boldsymbol{\xi}^i\| + \epsilon_N > \log N \right) \\ & \leq N \mathbb{P}(\|\boldsymbol{\xi}\| > \log N - \epsilon_N) \\ & = N \mathbb{E} [\mathbb{P}(\|\boldsymbol{\xi}\| - \mathbb{E}[\|\boldsymbol{\xi}\| \mid \gamma] > \log N - \epsilon_N - \mathbb{E}[\|\boldsymbol{\xi}\| \mid \gamma] \mid \gamma)] \\ & \leq N \mathbb{E} \left[ \mathbb{P} \left( \|\boldsymbol{\xi}\| - \mathbb{E}[\|\boldsymbol{\xi}\| \mid \gamma] > \log N - \epsilon_N - \sup_{\gamma' \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| \mid \gamma = \gamma'] \mid \gamma \right) \right] \\ & \leq N \mathbb{E} \left[ 2 \exp \left( - \frac{(\log N - \epsilon_N - \sup_{\gamma' \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| \mid \gamma = \gamma'])^2}{2\sigma^2} \right) \right] \\ & = 2 \exp \left( \log N - \frac{(\log N - \epsilon_N - \sup_{\gamma' \in \Gamma} \mathbb{E}[\|\boldsymbol{\xi}\| \mid \gamma = \gamma'])^2}{2\sigma^2} \right), & (\text{C.3}) \end{aligned}$$

which has a finite sum over  $N \in \mathbb{N}$ . Therefore, by the Borel-Cantelli lemma, there exists  $N_0 \in \mathbb{N}$ ,  $\mathbb{P}^\infty$ -almost surely, such that

$$\cup_{i=1}^N \mathcal{U}_N^i \subseteq D_N \quad \forall N \geq N_0.$$

We now choose any  $r > 0$  such that  $\epsilon_N N^{-r}$  satisfies Assumption 4.1, and define  $N_1 := \max\{N_0, 2^{\frac{1}{r}}\}$ . Then, the following holds for all  $N \geq N_1$  and  $\boldsymbol{\pi} \in \Pi$ :

$$\begin{aligned}
& \sup_{\mathbb{Q} \in \mathcal{P}(D_N \cap \Xi): d_1(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \frac{\epsilon_N}{N^r}} \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)] \\
& \leq \sup_{\mathbb{Q} \in \mathcal{P}(D_N \cap \Xi): d_\infty(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \epsilon_N} \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)] + \frac{4}{N^r} \sup_{\boldsymbol{\zeta} \in D_N \cap \Xi} |c^\pi(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T)| \\
& = \sum_{i=1}^N w_i^N(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} c^\pi(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) + \frac{4}{N^r} \sup_{\boldsymbol{\zeta} \in D_N \cap \Xi} |c^\pi(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T)| \\
& \leq \sum_{i=1}^N w_i^N(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} c^\pi(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) + \frac{4C}{N^r} (1 + \log N). \tag{C.4}
\end{aligned}$$

Indeed, the first supremum satisfies the conditions of Lemma C.2 since  $N \geq N_0$  and  $N \geq 2^{\frac{1}{r}}$ , and the equality follows from Lemma C.1 since  $N \geq N_0$ . The final inequality follows from Assumption 4.5 and the construction of  $D_N$ . We observe that the second term on (C.4) converges to zero as  $N \rightarrow \infty$ . Next, we observe that

$$\begin{aligned}
\mathbb{E}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mid \boldsymbol{\gamma} = \bar{\gamma}] & \triangleq \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)] \\
& = \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mathbb{1}\{\boldsymbol{\xi} \notin D_N\}] \\
& \quad + \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mathbb{1}\{\boldsymbol{\xi} \in D_N\}].
\end{aligned}$$

We handle the first term with the Cauchy-Schwartz inequality,

$$\mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mathbb{1}\{\boldsymbol{\xi} \notin D_N\}] \leq \sqrt{\mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}}[c^\pi(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)^2] \mathbb{P}_{\bar{\gamma}}(\boldsymbol{\xi} \notin D_N)}.$$

By Assumptions 4.2 and 4.5, the above bound is finite and converges to zero as  $N \rightarrow \infty$  uniformly over  $\boldsymbol{\pi} \in \Pi$ . The second term is handled by the new concentration measure from this paper. Specifically, it follows from Theorem 4.2 that there exists an  $N_2 \geq N_1$ ,  $\mathbb{P}^\infty$ -almost surely, such that

$$d_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \frac{\epsilon_N}{N^r} \quad \forall N \geq N_2.$$

Therefore, for all  $N \geq N_2$  and decision rules  $\boldsymbol{\pi} \in \Pi$ :

$$\begin{aligned}
& \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}} [c^{\boldsymbol{\pi}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mathbb{1}\{\boldsymbol{\xi} \in D_N\}] \\
&= \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{P}_{\bar{\gamma}}} \left[ \left( c^{\boldsymbol{\pi}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) - \inf_{\boldsymbol{\zeta} \in D_N \cap \Xi} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) \right) \mathbb{1}\{\boldsymbol{\xi} \in D_N\} \right] \\
&\quad + \underbrace{\mathbb{P}_{\bar{\gamma}}(\boldsymbol{\xi} \in D_N) \inf_{\boldsymbol{\zeta} \in D_N \cap \Xi} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T)}_{\alpha_N} \\
&\leq \sup_{\mathbb{Q} \in \mathcal{P}(\Xi): d_1(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \frac{\epsilon_N}{N^r}} \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}} \left[ \left( c^{\boldsymbol{\pi}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) - \inf_{\boldsymbol{\zeta} \in D_N \cap \Xi} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) \right) \mathbb{1}\{\boldsymbol{\xi} \in D_N\} \right] + \alpha_N \\
&= \sup_{\mathbb{Q} \in \mathcal{P}(\Xi \cap D_N): d_1(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \frac{\epsilon_N}{N^r}} \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}} \left[ c^{\boldsymbol{\pi}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) - \inf_{\boldsymbol{\zeta} \in D_N \cap \Xi} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) \right] + \alpha_N \\
&= \sup_{\mathbb{Q} \in \mathcal{P}(\Xi \cap D_N): d_1(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \frac{\epsilon_N}{N^r}} \mathbb{E}_{\boldsymbol{\xi} \sim \mathbb{Q}} [c^{\boldsymbol{\pi}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T)] - \mathbb{P}_{\bar{\gamma}}(\boldsymbol{\xi} \notin D_N) \inf_{\boldsymbol{\zeta} \in D_N \cap \Xi} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T).
\end{aligned}$$

Indeed, the inequality follows because  $N \geq N_2$ . It follows from Assumption 4.5 and (C.3) that the second term in the final equality converges to zero as  $N \rightarrow \infty$  uniformly over  $\boldsymbol{\pi} \in \Pi$ . Combining the above, we conclude that

$$\begin{aligned}
\liminf_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) &= \liminf_{N \rightarrow \infty} \inf_{\boldsymbol{\pi} \in \Pi} \sum_{i=1}^N w_i^N(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} c^{\boldsymbol{\pi}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) \\
&\geq \inf_{\boldsymbol{\pi} \in \Pi} \mathbb{E}[c^{\boldsymbol{\pi}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T) \mid \boldsymbol{\gamma} = \bar{\gamma}] = v^*(\bar{\gamma}),
\end{aligned}$$

where the inequality holds  $\mathbb{P}^\infty$ -almost surely. This completes the proof of (C.1).

## Upper bound.

We now prove that

$$\limsup_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) \leq v^*(\bar{\gamma}), \quad \mathbb{P}^\infty\text{-almost surely.} \tag{C.5}$$

Indeed, for any arbitrary  $\delta > 0$ , let  $\mathbf{x}_\delta \in \mathcal{X}$  be a  $\delta$ -optimal solution for (4.1). By Esfahani and Kuhn [48, Lemma A.1] and Assumption 4.5, there exists a non-increasing

sequence of functions  $f^j(\zeta_1, \dots, \zeta_T)$ ,  $j \in \mathbb{N}$ , such that

$$\lim_{j \rightarrow \infty} f^j(\zeta_1, \dots, \zeta_T) = c^{\mathbf{x}^\delta}(\zeta_1, \dots, \zeta_T), \quad \forall \zeta \in \Xi$$

and  $f^j$  is  $L_j$ -Lipschitz continuous. Furthermore, for each  $N \in \mathbb{N}$ , choose any probability distribution  $\hat{\mathbb{Q}}^N \in \mathcal{P}(\Xi)$  such that  $\mathbf{d}_1(\hat{\mathbb{Q}}^N, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \epsilon_N$  and

$$\sup_{\mathbb{Q} \in \mathcal{P}(\Xi): \mathbf{d}_1(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \epsilon_N} \mathbb{E}_{\xi \sim \mathbb{Q}}[c^{\mathbf{x}^\delta}(\xi_1, \dots, \xi_T)] \leq \mathbb{E}_{\xi \sim \hat{\mathbb{Q}}^N}[c^{\mathbf{x}^\delta}(\xi_1, \dots, \xi_T)] + \delta.$$

For any  $j \in \mathbb{N}$ ,

$$\begin{aligned} \limsup_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) &\leq \limsup_{N \rightarrow \infty} \sup_{\mathbb{Q} \in \mathcal{P}(\Xi): \mathbf{d}_\infty(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \epsilon_N} \mathbb{E}_{\xi \sim \mathbb{Q}}[c^{\mathbf{x}^\delta}(\xi_1, \dots, \xi_T)] \\ &\leq \limsup_{N \rightarrow \infty} \sup_{\mathbb{Q} \in \mathcal{P}(\Xi): \mathbf{d}_1(\mathbb{Q}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) \leq \epsilon_N} \mathbb{E}_{\xi \sim \mathbb{Q}}[c^{\mathbf{x}^\delta}(\xi_1, \dots, \xi_T)] \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_{\xi \sim \hat{\mathbb{Q}}^N}[c^{\mathbf{x}^\delta}(\xi_1, \dots, \xi_T)] + \delta \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_{\xi \sim \hat{\mathbb{Q}}^N}[f^j(\xi_1, \dots, \xi_T)] + \delta \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_{\xi \sim \mathbb{P}_{\bar{\gamma}}}[f^j(\xi_1, \dots, \xi_T)] + L_j \mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{Q}}^N) + \delta \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_{\xi \sim \mathbb{P}_{\bar{\gamma}}}[f^j(\xi_1, \dots, \xi_T)] + L_j (\mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) + \mathbf{d}_1(\hat{\mathbb{Q}}^N, \hat{\mathbb{P}}_{\bar{\gamma}}^N)) + \delta \\ &\leq \limsup_{N \rightarrow \infty} \mathbb{E}_{\xi \sim \mathbb{P}_{\bar{\gamma}}}[f^j(\xi_1, \dots, \xi_T)] + L_j (\mathbf{d}_1(\mathbb{P}_{\bar{\gamma}}, \hat{\mathbb{P}}_{\bar{\gamma}}^N) + \epsilon_N) + \delta \\ &= \mathbb{E}_{\mathbb{P}_{\bar{\gamma}}}[f^j(\xi_1, \dots, \xi_T)] + \delta, \quad \mathbb{P}^\infty\text{-almost surely,} \end{aligned}$$

where we have used the fact  $\mathbf{d}_1(\mathbb{P}, \mathbb{Q}) \leq \mathbf{d}_\infty(\mathbb{P}, \mathbb{Q})$  for the second inequality, the dual form of the 1-Wasserstein metric for the fifth inequality (because  $f^j$  is  $L_j$ -Lipschitz), and Theorem 4.2 for the equality. Taking the limit as  $j \rightarrow \infty$ , and applying the monotone convergence theorem (which is allowed because  $\mathbb{E}_{\xi \sim \mathbb{P}_{\bar{\gamma}}}|f^1(\xi_1, \dots, \xi_T)| \leq L_1 \mathbb{E}_{\xi \sim \mathbb{P}_{\bar{\gamma}}}\|\xi\| + |f^1(0)| < \infty$  by Assumption 4.4), gives

$$\limsup_{N \rightarrow \infty} \hat{v}^N(\bar{\gamma}) \leq \mathbb{E}_{\xi \sim \mathbb{P}_{\bar{\gamma}}}[c^{\mathbf{x}^\delta}(\xi_1, \dots, \xi_T)] + \delta \leq v^*(\bar{\gamma}) + 2\delta, \quad \mathbb{P}^\infty\text{-almost surely.}$$



Since  $\delta > 0$  was chosen arbitrarily, the proof of (C.5) is complete.  $\square$

## C.3 Technical Details for Section 4.5

### C.3.1 Proof of Theorem 4.3

We restate the theorem for convenience.

**Theorem 4.3.** *For cost functions of the form (4.10),  $\tilde{v}^N(\bar{\gamma}) = \hat{v}^N(\bar{\gamma})$ .*

*Proof.* We first show that  $\tilde{v}^N(\bar{\gamma}) \geq \hat{v}^N(\bar{\gamma})$ . Indeed, consider any primary decision rule  $\bar{\pi}$  and auxiliary decision rules  $\bar{\mathbf{y}}_1^i, \dots, \bar{\mathbf{y}}_T^i$  for each  $i \in \{1, \dots, N\}$  which are optimal for (4.11).<sup>1</sup> Then, it follows from feasibility to (4.11) that

$$\mathbf{h}_t^\top \bar{\mathbf{y}}_t^i(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_t) \geq \min_{\mathbf{y}_t \in \mathbb{R}^{d_y^t}} \left\{ \mathbf{h}_t^\top \mathbf{y}_t : \sum_{s=1}^t \mathbf{A}_{t,s} \bar{\boldsymbol{\pi}}_s(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{s-1}) + \sum_{s=1}^t \mathbf{B}_{t,s} \boldsymbol{\zeta}_s + \mathbf{C}_t \mathbf{y}_t \leq \mathbf{d}_t \right\}.$$

for each  $i \in \{1, \dots, N\}$ ,  $\boldsymbol{\zeta} \in \mathcal{U}_N^i$ , and  $t \in \{1, \dots, T\}$ . Thus,

$$\begin{aligned} \hat{v}^N(\bar{\gamma}) &= \min_{\boldsymbol{\pi} \in \Pi} \sum_{i=1}^N w_N^i(\bar{\gamma}) c^\pi(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) \\ &\leq \sum_{i=1}^N w_N^i(\bar{\gamma}) c^{\bar{\boldsymbol{\pi}}}(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T) \\ &\leq \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \sum_{t=1}^T (\mathbf{f}_t^\top \bar{\boldsymbol{\pi}}_t(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{t-1}) + \mathbf{g}_t^\top \boldsymbol{\zeta}_t + \mathbf{h}_t^\top \bar{\mathbf{y}}_t^i(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_t)) = \tilde{v}^N(\bar{\gamma}). \end{aligned}$$

The other side of the inequality follows from similar reasoning. Indeed, let  $\bar{\boldsymbol{\pi}}$  be an optimal solution to (4.3). For each  $i \in \{1, \dots, N\}$  and  $t \in \{1, \dots, T\}$ , define  $\bar{\mathbf{y}}_t^i \in \mathcal{R}_t$  as any decision rule that satisfies

$$\bar{\mathbf{y}}_t^i(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_t) \in \arg \min_{\mathbf{y}_t \in \mathbb{R}^{d_y^t}} \left\{ \mathbf{h}_t^\top \mathbf{y}_t : \sum_{s=1}^t \mathbf{A}_{t,s} \bar{\boldsymbol{\pi}}_s(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_{s-1}) + \sum_{s=1}^t \mathbf{B}_{t,s} \boldsymbol{\zeta}_s + \mathbf{C}_t \mathbf{y}_t \leq \mathbf{d}_t \right\}$$

<sup>1</sup>If no optimal solution exists, then we may choose any  $\eta$ -optimal solution.

for every  $\zeta \in \mathcal{U}_N^i$ . Then,

$$\begin{aligned}\tilde{v}^N(\bar{\gamma}) &\leq \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\zeta \in \mathcal{U}_N^i} \sum_{t=1}^T (\mathbf{f}_t^\top \bar{\pi}_t(\zeta_1, \dots, \zeta_{t-1}) + \mathbf{g}_t^\top \zeta_t + \mathbf{h}_t^\top \bar{\mathbf{y}}_t^i(\zeta_1, \dots, \zeta_t)) \\ &= \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\zeta \in \mathcal{U}_N^i} c^{\bar{\pi}}(\zeta_1, \dots, \zeta_T) = \hat{v}^N(\bar{\gamma}).\end{aligned}$$

Combining the above inequalities, the proof is complete.  $\square$

### C.3.2 Compact representation of multi-policy approximation

Consider the multi-policy approximation scheme

$$\begin{aligned}\min_{\pi, \mathbf{y}} \quad & \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\zeta \in \mathcal{U}_N^i} \sum_{t=1}^T (\mathbf{f}_t^\top \pi_t(\zeta_1, \dots, \zeta_{t-1}) + \mathbf{g}_t^\top \zeta_t + \mathbf{h}_t^\top \mathbf{y}_t^i(\zeta_1, \dots, \zeta_t)) \\ \text{s.t.} \quad & \sum_{s=1}^t \mathbf{A}_{t,s} \pi_s(\zeta_1, \dots, \zeta_{s-1}) + \sum_{s=1}^t \mathbf{B}_{t,s} \zeta_s + \mathbf{C}_t \mathbf{y}_t^i(\zeta_1, \dots, \zeta_t) \leq \mathbf{d}_t \\ & \forall \zeta \in \mathcal{U}_N^i, i \in \{1, \dots, N\}, t \in \{1, \dots, T\}\end{aligned}$$

where we restrict to linear decision rules of the form

$$\pi_t(\zeta_1, \dots, \zeta_{t-1}) = \mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s} \zeta_s, \quad \mathbf{y}_t^i(\zeta_1, \dots, \zeta_t) = \mathbf{y}_{t,0}^i + \sum_{s=1}^t \mathbf{Y}_{t,s}^i \zeta_s.$$

We now show how to transform this problem into a more compact representation.

First, we combine the primary linear decision rules across stages as

$$\mathbf{x}_0 = \begin{bmatrix} \mathbf{x}_{1,0} \\ \vdots \\ \mathbf{x}_{T,0} \end{bmatrix} \in \mathbb{R}^{d_x},$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_{2,1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_{3,1} & \mathbf{X}_{3,2} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{X}_{T-2,1} & \mathbf{X}_{T-2,2} & \mathbf{X}_{T-2,3} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_{T-1,1} & \mathbf{X}_{T-1,2} & \mathbf{X}_{T-1,3} & \cdots & \mathbf{X}_{T-1,T-2} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_{T,1} & \mathbf{X}_{T,2} & \mathbf{X}_{T,3} & \cdots & \mathbf{X}_{T,T-2} & \mathbf{X}_{T,T-1} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{d_x \times d_\xi}.$$

We note that the zero entries in the above matrix are necessary to ensure that the linear decision rules are non-anticipative. Similarly, for each  $i \in \{1, \dots, N\}$  we represent the auxiliary linear decision rules as

$$\mathbf{y}_0^i = \begin{bmatrix} \mathbf{y}_{1,0}^i \\ \vdots \\ \mathbf{y}_{T,0}^i \end{bmatrix} \in \mathbb{R}^{d_y}, \quad \mathbf{Y}^i = \begin{bmatrix} \mathbf{Y}_{1,1}^i & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{Y}_{2,1}^i & \mathbf{Y}_{2,2}^i & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Y}_{T-1,1}^i & \mathbf{Y}_{T-1,2}^i & \cdots & \mathbf{Y}_{T-1,T-1}^i & \mathbf{0} \\ \mathbf{Y}_{T,1}^i & \mathbf{Y}_{T,2}^i & \cdots & \mathbf{Y}_{t,t-1}^i & \mathbf{Y}_{T,T}^i \end{bmatrix} \in \mathbb{R}^{d_y \times d_\xi}.$$

We now combine the problem parameters. Let  $\mathbf{d} = (\mathbf{d}_1, \dots, \mathbf{d}_T) \in \mathbb{R}^m$  and

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_T \end{bmatrix} \in \mathbb{R}^{d_x}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{A}_{T-1,1} & \mathbf{A}_{T-1,2} & \cdots & \mathbf{A}_{T-1,T-1} & \mathbf{0} \\ \mathbf{A}_{T,1} & \mathbf{A}_{T,2} & \cdots & \mathbf{A}_{t,t-1} & \mathbf{A}_{T,T} \end{bmatrix} \in \mathbb{R}^{m \times d_x},$$

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_T \end{bmatrix} \in \mathbb{R}^{d_\xi}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{B}_{T-1,1} & \mathbf{B}_{T-1,2} & \cdots & \mathbf{B}_{T-1,T-1} & \mathbf{0} \\ \mathbf{B}_{T,1} & \mathbf{B}_{T,2} & \cdots & \mathbf{B}_{t,t-1} & \mathbf{B}_{T,T} \end{bmatrix} \in \mathbb{R}^{m \times d_x},$$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_T \end{bmatrix} \in \mathbb{R}^{d_y}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{C}_{1,1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{2,2} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{T-1,T-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{T,T} \end{bmatrix} \in \mathbb{R}^{m \times d_x}.$$

Therefore, using the above compact notation, we can rewrite the multi-policy approximation with linear decision rules as

$$\begin{aligned} & \min_{\substack{\mathbf{x}_0 \in \mathbb{R}^{d_x}, \mathbf{X} \in \mathbb{R}^{d_x \times d_\xi} \\ \mathbf{y}_0^i \in \mathbb{R}^{d_y}, \mathbf{Y}^i \in \mathbb{R}^{d_y \times d_\xi}}} \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\zeta \in \mathcal{U}_N^i} \{ \mathbf{f}^\top (\mathbf{x}_0 + \mathbf{X}\zeta) + \mathbf{g}^\top \zeta + \mathbf{h}^\top (\mathbf{y}_0^i + \mathbf{Y}^i \zeta) \} \\ \text{s.t.} \quad & \mathbf{A}(\mathbf{x}_0 + \mathbf{X}\zeta) + \mathbf{B}\zeta + \mathbf{C}(\mathbf{y}_0^i + \mathbf{Y}^i \zeta) \leq \mathbf{d} \\ & \mathbf{x}_0 + \mathbf{X}\zeta \in \mathcal{X} \\ & \forall \zeta \in \mathcal{U}_N^i, i \in \{1, \dots, N\}. \end{aligned}$$

### C.3.3 Proof of Theorem 4.4

We present here a generalization of Theorem 4.4 in which there are nonnegativity constraints on the support of  $\Xi$ .

**Theorem 4.4.** Suppose  $\Xi = \mathbb{R}_+^{d_\xi}$  and  $\mathcal{X} = \mathbb{R}^{d_x}$ . Then, (4.12) is equivalent to

$$\begin{aligned} \min_{\mathbf{x}_0, \mathbf{X}, \mathbf{y}_0, \mathbf{Y}, \boldsymbol{\Lambda}, \mathbf{s}} \quad & \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \mathbf{f}^\top (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\xi}^i) + \mathbf{g}^\top \boldsymbol{\xi}^i + \mathbf{h}^\top (\mathbf{y}_0 + \mathbf{Y}^i \boldsymbol{\xi}^i) + (\mathbf{s}^i)^\top \boldsymbol{\xi}^i \right. \\ & \left. + \epsilon_N \|\mathbf{X}^\top \mathbf{f} + \mathbf{g} + (\mathbf{Y}^i)^\top \mathbf{h} + \mathbf{s}^i\|_* \right) \\ \text{s.t.} \quad & \mathbf{A} (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\xi}^i) + \mathbf{B}\boldsymbol{\xi}^i + \mathbf{C} (\mathbf{y}_0 + \mathbf{Y}^i \boldsymbol{\xi}^i) + \boldsymbol{\Lambda}^i \boldsymbol{\xi}^i \\ & + \epsilon_N \|\mathbf{A}\mathbf{X} + \mathbf{B} + \mathbf{C}\mathbf{Y}^i + \boldsymbol{\Lambda}^i\|_* \leq \mathbf{d} \quad \forall i \in \{1, \dots, N\} \\ & \mathbf{x}_0 \in \mathbb{R}^{d_x}, \mathbf{X} \in \mathbb{R}^{d_x \times d_\xi}, \mathbf{y}_0 \in \mathbb{R}^{d_y}, \mathbf{Y}^i \in \mathbb{R}^{d_y \times d_\xi}, \boldsymbol{\Lambda}^i \in \mathbb{R}_+^{m \times d_\xi}, \mathbf{s}^i \in \mathbb{R}_+^{d_\xi}, \end{aligned}$$

where  $\|\mathbf{Z}\|_* := (\|\mathbf{z}_1\|_*, \dots, \|\mathbf{z}_r\|_*) \in \mathbb{R}^r$  for any matrix  $\mathbf{Z} \in \mathbb{R}^{r \times n}$ .

It follows from the following reformulation that the case where  $\Xi = \mathbb{R}^{d_\xi}$  easily follows in which each of the variables  $\boldsymbol{\Lambda}^i$  and  $\mathbf{s}^i$  are constrained to zero.

*Proof.* For any  $\mathbf{c} \in \mathbb{R}^{d_\xi}$  and  $\boldsymbol{\xi} \in \Xi$ , it follows directly from strong duality for conic optimization that

$$\max_{\boldsymbol{\zeta} \geq \mathbf{0}} \{\mathbf{c}^\top \boldsymbol{\zeta} : \|\boldsymbol{\zeta} - \boldsymbol{\xi}\| \leq \epsilon\} = \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \{(\mathbf{c} + \boldsymbol{\lambda})^\top \boldsymbol{\xi} + \epsilon \|\mathbf{c} + \boldsymbol{\lambda}\|_*\}.$$

We use this result to reformulate the objective and constraints of (4.12). First, let the  $j$ -th rows of  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  and the  $j$ -th element of  $\mathbf{d}$  be denoted by  $\mathbf{a}_j \in \mathbb{R}^{d_x}$ ,  $\mathbf{b}_j \in \mathbb{R}^{d_\xi}$ ,  $\mathbf{c}_j \in \mathbb{R}^{d_y}$ , and  $d_j \in \mathbb{R}$ . Then, each robust constraint has the form

$$\mathbf{a}_j^\top (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\zeta}) + \mathbf{b}_j^\top \boldsymbol{\zeta} + \mathbf{c}_j^\top (\mathbf{y}_0 + \mathbf{Y}^i \boldsymbol{\zeta}) \leq d_j \quad \forall \boldsymbol{\zeta} \in \mathcal{U}_N^i.$$

Rearranging terms,

$$(\mathbf{a}_j^\top \mathbf{X} + \mathbf{b}_j^\top + \mathbf{c}_j^\top \mathbf{Y}^i) \boldsymbol{\zeta} \leq d_j - \mathbf{a}_j^\top \mathbf{x}_0 - \mathbf{c}_j^\top \mathbf{y}_0 \quad \forall \boldsymbol{\zeta} \in \mathcal{U}_N^i,$$

which applying duality becomes

$$\begin{aligned} \exists \boldsymbol{\lambda}_j^i \geq \mathbf{0} : & (\mathbf{X}^\top \mathbf{a}_j + \mathbf{b}_j + (\mathbf{Y}^i)^\top \mathbf{c}_j + \boldsymbol{\lambda}_j^i)^\top \boldsymbol{\xi}^i + \epsilon_N \|\mathbf{X}^\top \mathbf{a}_j + \mathbf{b}_j + (\mathbf{Y}^i)^\top \mathbf{c}_j + \boldsymbol{\lambda}_j^i\|_* \\ & \leq d_j - \mathbf{a}_j^\top \mathbf{x}_0 - \mathbf{c}_j^\top \mathbf{y}_0^i. \end{aligned}$$

Rearranging terms and applying the definition of  $\|\cdot\|_*$  for matrices, the robust constraints for each  $i \in \{1, \dots, N\}$  are satisfied if and only if

$$\begin{aligned} \exists \boldsymbol{\Lambda}^i \geq \mathbf{0} : & \mathbf{A} (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\xi}^i) + \mathbf{B}\boldsymbol{\xi}^i + \mathbf{C} (\mathbf{y}_0^i + \mathbf{Y}^i\boldsymbol{\xi}^i) + \boldsymbol{\Lambda}^i \boldsymbol{\xi}^i + \epsilon_N \|\mathbf{A}\mathbf{X} + \mathbf{B} + \mathbf{C}\mathbf{Y}^i + \boldsymbol{\Lambda}^i\|_* \\ & \leq \mathbf{d}, \end{aligned}$$

where the dual norm for a matrix is applied separately for each row. Similarly, the objective function takes the form

$$\begin{aligned} & \sum_{i=1}^N w_N^i(\bar{\gamma}) \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} \{ \mathbf{f}^\top (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\zeta}) + \mathbf{g}^\top \boldsymbol{\zeta} + \mathbf{h}^\top (\mathbf{y}_0^i + \mathbf{Y}^i\boldsymbol{\zeta}) \} \\ & = \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \mathbf{f}^\top \mathbf{x}_0 + \mathbf{h}^\top \mathbf{y}_0^i + \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^i} (\mathbf{f}^\top \mathbf{X} + \mathbf{g}^\top + \mathbf{h}^\top \mathbf{Y}^i) \boldsymbol{\zeta} \right) \\ & = \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \mathbf{f}^\top \mathbf{x}_0 + \mathbf{h}^\top \mathbf{y}_0^i \right. \\ & \quad \left. + \inf_{\mathbf{s}^i \geq \mathbf{0}} \{ (\mathbf{X}^\top \mathbf{f} + \mathbf{g} + (\mathbf{Y}^i)^\top \mathbf{h} + \mathbf{s}^i)^\top \boldsymbol{\xi}^i + \epsilon_N \|\mathbf{X}^\top \mathbf{f} + \mathbf{g} + (\mathbf{Y}^i)^\top \mathbf{h} + \mathbf{s}^i\|_* \} \right) \\ & = \sum_{i=1}^N w_N^i(\bar{\gamma}) \left( \mathbf{f}^\top (\mathbf{x}_0 + \mathbf{X}\boldsymbol{\xi}^i) + \mathbf{g}^\top \boldsymbol{\xi}^i + \mathbf{h}^\top (\mathbf{y}_0^i + \mathbf{Y}^i\boldsymbol{\xi}^i) \right. \\ & \quad \left. + \inf_{\mathbf{s}^i \geq \mathbf{0}} \{ (\mathbf{s}^i)^\top \boldsymbol{\xi}^i + \epsilon_N \|\mathbf{X}^\top \mathbf{f} + \mathbf{g} + (\mathbf{Y}^i)^\top \mathbf{h} + \mathbf{s}^i\|_* \} \right). \end{aligned}$$

Combining the reformulations above, we obtain the desired result.  $\square$

# Bibliography

- [1] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- [2] Susan Athey and Stefan Wager. Efficient policy learning. *arXiv preprint arXiv:1702.02896*, 2017.
- [3] Gah-Yi Ban and Cynthia Rudin. The big data newsvendor: practical insights from machine learning. *Operations Research*, 67(1):90–108, 2018.
- [4] Gah-Yi Ban, Jérémie Gallien, and Adam J Mersereau. Dynamic procurement of new products with covariate information: The residual tree method. *Manufacturing & Service Operations Management*, 2018.
- [5] Jeroen Beliën, Erik Demeulemeester, and Brecht Cardoen. A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of scheduling*, 12(2):147, 2009.
- [6] Richard Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences*, 38(8):716–719, 1952.
- [7] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.
- [8] Jacques F Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik*, 4(1):238–252, 1962.
- [9] James O Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media, 2013.
- [10] Donald A Berry and Bert Fristedt. *Bandit Problems: Sequential Allocation of Experiments (Monographs on Statistics and Applied Probability)*. Springer, 1985.
- [11] Dimitri P Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific Belmont, MA, 2017.
- [12] Dimitris Bertsimas and Constantine Caramanis. Finite adaptability in multi-stage linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766, 2010.

- [13] Dimitris Bertsimas and Jack Dunn. Optimal classification trees. *Machine Learning*, pages 1–44, 2017.
- [14] Dimitris Bertsimas and Angelos Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, 63(3):610–627, 2015.
- [15] Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *arXiv preprint arXiv:1402.5481*, 2014.
- [16] Dimitris Bertsimas and Nathan Kallus. The power and limits of predictive approaches to observational-data-driven optimization. *arXiv preprint arXiv:1605.02347*, 2017.
- [17] Dimitris Bertsimas and Christopher McCord. Optimization over continuous and multi-dimensional decisions with observational data. In *Advances in Neural Information Processing Systems*, pages 2966–2974, 2018.
- [18] Dimitris Bertsimas and Christopher McCord. From predictions to prescriptions in multistage optimization problems. *arXiv preprint arXiv:1904.11637*, 2018.
- [19] Dimitris Bertsimas and Bart Van Parys. Bootstrap robust prescriptive analytics. *arXiv preprint arXiv:1711.09974*, 2017.
- [20] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.
- [21] Dimitris Bertsimas, Jack Dunn, and Nishanth Mundru. Optimal prescriptive trees. *Under review*, 2018.
- [22] Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. A data-driven approach for multi-stage linear optimization. 2018. URL [http://www.optimization-online.org/DB\\_FILE/2018/11/6907.pdf](http://www.optimization-online.org/DB_FILE/2018/11/6907.pdf).
- [23] Dimitris Bertsimas, Melvyn Sim, and Meilin Zhang. Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618, 2018.
- [24] Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. Two-stage sample robust optimization. 2019.
- [25] Gérard Biau and Luc Devroye. *Lectures on the Nearest Neighbor Method*. Springer, 2015.
- [26] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [27] John T Blake, Franklin Dexter, and Joan Donald. Operating room managers? use of integer programming for assigning block time to surgical groups: a case study. *Anesthesia & Analgesia*, 94(1):143–148, 2002.



- [28] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford university press, 2013.
- [29] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [30] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
- [31] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.
- [32] Brecht Cardoen, Erik Demeulemeester, and Jeroen Beliën. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932, 2010.
- [33] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In *Advances in Neural Information Processing Systems*, pages 2249–2257, 2011.
- [34] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.
- [35] Xin Chen and Yuhan Zhang. Uncertain linear programs: extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2009.
- [36] Zhi Chen, Melvyn Sim, and Peng Xiong. Robust stochastic optimization: the synergy of robust optimization and stochastic programming. 2019. URL [http://www.optimization-online.org/DB\\_FILE/2017/06/6055.pdf](http://www.optimization-online.org/DB_FILE/2017/06/6055.pdf).
- [37] Vincent S Chow, Martin L Puterman, Neda Salehirad, Wenhai Huang, and Derek Atkins. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management*, 20(3):418–430, 2011.
- [38] Wei Chu, Lihong Li, Lev Reyzin, and Robert E Schapire. Contextual bandits with linear payoff functions. In *AISTATS*, volume 15, pages 208–214, 2011.
- [39] Philippe Clement and Wolfgang Desch. An elementary proof of the triangle inequality for the wasserstein metric. *Proceedings of the American Mathematical Society*, 136(1):333–339, 2008.
- [40] Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. 2016.
- [41] International Warfarin Pharmacogenetics Consortium et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 2009(360):753–764, 2009.

- [42] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
- [43] Brian Denton, James Viapiano, and Andrea Vogl. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science*, 10(1):13–24, 2007.
- [44] Franklin Dexter and Alex Macario. Changing allocations of operating room time from a system based on historical utilization to one where the aim is to schedule as many surgical cases as possible. *Anesthesia & Analgesia*, 94(5):1272–1279, 2002.
- [45] Franklin Dexter, John T Blake, Donald H Penning, Brian Sloan, Patricia Chung, and David A Lubarsky. Use of linear programming to estimate impact of changes in a hospital’s operating room time allocation on perioperative variable costs. *The Journal of the American Society of Anesthesiologists*, 96(3):718–724, 2002.
- [46] Peter Dorato, Vito Cerone, and Chaouki Abdallah. *Linear-quadratic control: an introduction*. Simon & Schuster, 1994.
- [47] Adam N Elmachtoub and Paul Grigas. Smart “predict, then optimize”. *arXiv preprint arXiv:1710.08005*, 2017.
- [48] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018.
- [49] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2015.
- [50] Carlos A Flores. Estimation of dose-response functions and optimal doses with a continuous treatment. 2005.
- [51] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.
- [52] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning*, volume 1. Springer, 2001.
- [53] Angelos Georghiou, Wolfram Wiesemann, and Daniel Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152(1-2):301–338, 2015.

- [54] Sharad Goel, Jake M Hofman, Sébastien Lahaie, David M Pennock, and Duncan J Watts. Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences*, 107(41):17486–17490, 2010.
- [55] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4-part-1):902–917, 2010.
- [56] Daniel Gruhl, Ramanathan Guha, Ravi Kumar, Jasmine Novak, and Andrew Tomkins. The predictive power of online chatter. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pages 78–87. ACM, 2005.
- [57] Grani A Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018.
- [58] Grani A Hanasusanto, Daniel Kuhn, and Wolfram Wiesemann. K-adaptability in two-stage distributionally robust binary programming. *Operations Research Letters*, 44(1):6–11, 2016.
- [59] Lauren Hannah, Warren Powell, and David M Blei. Nonparametric density estimation for stochastic optimization with an observable state variable. In *Advances in Neural Information Processing Systems*, pages 820–828, 2010.
- [60] Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, 226164:73–84, 2004.
- [61] Chin Pang Ho and Grani Hanasusanto. On data-driven prescriptive analytics with side information: a regularized nadaraya-watson approach. 2019. URL [http://www.optimization-online.org/DB\\_FILE/2019/01/7043.pdf](http://www.optimization-online.org/DB_FILE/2019/01/7043.pdf).
- [62] Garud N Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.
- [63] Nathan Kallus. Balanced policy evaluation and learning. *arXiv preprint arXiv:1705.07384*, 2017.
- [64] Nathan Kallus. Recursive partitioning for personalization using observational data. In *International Conference on Machine Learning*, pages 1789–1798, 2017.
- [65] Nathan Kallus and Angela Zhou. Policy evaluation and optimization with continuous treatments. *arXiv preprint arXiv:1802.06037*, 2018.
- [66] L Kantorovich and G Rubinstein. On a space of totally additive functions. *Vestn Lening. Univ*, 13:52–59, 1958.
- [67] Yi-hao Kao, Benjamin V Roy, and Xiang Yan. Directed regression. In *Advances in Neural Information Processing Systems*, pages 889–897, 2009.

- [68] Seung-Chul Kim and Ira Horowitz. Scheduling hospital services: the efficacy of elective-surgery quotas. *Omega*, 30(5):335–346, 2002.
- [69] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [70] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, pages 661–670. ACM, 2010.
- [71] Eric Marcon and Franklin Dexter. Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science*, 9(1):87–98, 2006.
- [72] Eric Marcon and Franklin Dexter. An observational study of surgeons’ sequencing of cases and its impact on postanesthesia care unit and holding area staffing requirements at hospitals. *Anesthesia & Analgesia*, 105(1):119–126, 2007.
- [73] Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.
- [74] Michael L McManus, Michael C Long, Abbot Cooper, James Mandell, Donald M Berwick, Marcello Pagano, and Eugene Litvak. Variability in surgical caseload and access to intensive care services. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 98(6):1491–1496, 2003.
- [75] Velibor V Misić. Optimization of tree ensembles. *arXiv preprint arXiv:1705.10883*, 2017.
- [76] James Murphy. Benders, nested benders and stochastic programming: An intuitive introduction. *arXiv preprint arXiv:1312.3158*, 2013.
- [77] Gurobi Optimization. Gurobi optimizer reference manual, 2016. URL <http://www.gurobi.com>.
- [78] Marie Jeanette Persson and Jan A Persson. Analysing management policies for operating room planning using simulation. *Health Care Management Science*, 13(2):182–191, 2010.
- [79] Nicholas C Petruzzi and Maqbool Dada. Pricing and the newsvendor problem: A review with extensions. *Operations Research*, 47(2):183–194, 1999.
- [80] Warren B Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, volume 703. John Wiley & Sons, 2007.
- [81] Meera V Ragavan, Thane A Blinman, and Evan S Fieldston. Scheduled surgery admissions and occupancy at a children’s hospital: variation we can control to improve efficiency and value in health care delivery. *Annals of Surgery*, 257(3):564–570, 2013.

- [82] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.
- [83] Paul R Rosenbaum. *Observational Studies*. Springer, 2002.
- [84] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.
- [85] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge university press, 2014.
- [86] Alexander Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58(1):57–68, 2003.
- [87] Alexander Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34(1):1–8, 2006.
- [88] Alexander Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72, 2011.
- [89] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.
- [90] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*, volume 1. MIT press Cambridge, 1998.
- [91] Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization. In *Proceedings of the 24th International Conference on World Wide Web*, pages 939–941. ACM, 2015.
- [92] Chaitanya Swamy and David B Shmoys. Sampling-based approximation algorithms for multi-stage stochastic optimization. In *46th Annual IEEE Symposium on Foundations of Computer Science, 2005*, pages 357–366. IEEE, 2005.
- [93] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [94] Bart PG Van Parys, Peyman Mohajerin Esfahani, and Daniel Kuhn. From data to decisions: distributionally robust optimization is optimal. *arXiv preprint arXiv:1704.04118*, 2017.
- [95] Peter T VanBerkel and John T Blake. A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Management Science*, 10(4):373–385, 2007.
- [96] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47. Cambridge University Press, 2018.

- [97] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 2017.
- [98] Harro Walk. Strong laws of large numbers and nonparametric estimation. In *Recent Developments in Applied Probability and Statistics*, pages 183–214. Springer, 2010.
- [99] Huan Xu, Constantine Caramanis, and Shie Mannor. A distributional interpretation of robust optimization. *Mathematics of Operations Research*, 37(1): 95–110, 2012.
- [100] Ana Cecilia Zenteno, Tim Carnes, Retsef Levi, Bethany J Daily, and Peter F Dunn. Systematic or block allocation at a large academic medical center. *Annals of Surgery*, 264(6):973–981, 2016.
- [101] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Nested decomposition of multistage stochastic integer programs with binary state variables. *Optimization Online*, 2016.