Title
Modeling Consumer Behavior in Electricity Markets - Theory and Applications

Author
Zhou, Datong Paul

Publication Date
2018

Peer reviewed|Thesis/dissertation

Modeling Consumer Behavior in Electricity Markets – Theory and Applications

by

Datong Paul Zhou

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Claire J. Tomlin, Chair
Professor Kameshwar Poolla
Professor Francesco Borrelli
Assistant Professor Anil J. Aswani

Fall 2018

Modeling Consumer Behavior in Electricity Markets – Theory and Applications

Abstract

Modeling Consumer Behavior in Electricity Markets – Theory and Applications

by

Datong Paul Zhou

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Claire J. Tomlin, Chair

This thesis models consumer behavior in electricity markets from both a theoretical and practical perspective. The main underlying theme of this thesis is residential demand response, which is enabled by the development of smart grid technologies and the subsequent collection of large amounts of data. The work in this thesis touches on themes at the intersection of machine learning, economics, and game theory.

In the first part of this thesis, we seek to analyze the interaction between different players in the modern smart grid. Topics discussed include the incentivization of residential users to elicit their private information, how peer effects affect residential energy consumption, as well as hedging strategies against quantity and price risks in the electric market. To support these expositions with even more fundamental analyses, a framework for budget-constrained and combinatorial multi-armed bandits is introduced. The motivation behind including this rather generic topic lies in the sequential and repetitive nature of interactions between different players in the smart grid, which - under certain assumptions - could be captured with this methodology.

The second part of this thesis is concerned with the estimation of a residential Demand Response program carried out by OhmConnect, Inc. during a 14-month period. To evaluate the effects of monetary (and non-monetary) incentives on the reduction in electricity consumption, we first develop a short-term load forecasting method and compare various estimators. The estimation accuracy is further improved by incorporating mixtures of Gaussians and Hidden Markov Models into the estimators under consideration. Next, we develop a two-stage estimation framework to estimate individual treatment effects of Demand Response and compare the aggregated effect, namely the Average Treatment Effect, to the outcome of a randomized controlled trial. The ability to estimate individual treatment effects allows us to design an adaptive targeting framework, which seeks to maximize cost efficiency of this program. Lastly, the effect of moral suasion (non-monetary incentives) that only appeal to the environmental consciousness of users is explored.

Contents

Contents **i**

List of Figures **iv**

List of Tables **vii**

1 Introduction and Background **1**

I Theory **5**

2 Eliciting Private User Information in DR **6**
2.1 Introduction 6
2.2 Market Participants and Interactions 9
2.3 Demand Response Mechanism 11
2.4 Effect of Baseline "Gaming" 16
2.5 Simulations 17
2.6 Conclusion 21

3 Peer Effects in Electricity Consumption **22**
3.1 Introduction 22
3.2 Game-Theoretic Model 24
3.3 Theoretical Statements 27
3.4 Comparison of Pricing Schemes 30
3.5 Profit Maximization with User Selection 31
3.6 Conclusion 34

4 Hedging Strategies in Electricity Markets **37**
4.1 Introduction 37
4.2 Market Participants 39
4.3 Optimal Hedging Strategies 40
4.4 The Effect of Uncertainty 43
4.5 Choosing the Best Option 45

4.6 Simulations 46
4.7 Conclusion 51

5 Budget-Constrained Combinatorial Multi-Armed Bandits 52
5.1 Introduction 52
5.2 Main Results 54
5.3 Stochastic Setting 55
5.4 Adversarial Setting 56
5.5 Proofs 62
5.6 Discussion and Conclusion 66

II Applications 68

6 HVAC Temperature Control in Commercial Buildings 69
6.1 Introduction 69
6.2 Preliminaries 71
6.3 Data-Driven Model 73
6.4 Physics-Based Model 79
6.5 Quantitative Comparison of Both Models 83
6.6 Discussion and Conclusion 85

7 Short-Term Load Forecasting on Smart Meter Data 87
7.1 Introduction 87
7.2 Forecasting Methods 88
7.3 Mixture Models 90
7.4 Hidden Markov Models 92
7.5 Short-Term Load Forecasting 96
7.6 Non-Experimental Estimates of DR Reduction 98
7.7 Experiments on Data 99
7.8 Conclusion 104

8 Evaluation of a Residential DR Program 105
8.1 Introduction 105
8.2 Demand Response Mechanism 109
8.3 Experimental Setup and Data Characteristics 110
8.4 Nonexperimental Treatment Effect Estimation 120
8.5 Nonexperimental Estimation Results 130
8.6 ATE Estimation with Fixed Effects Models 135
8.7 Comparison of Estimation Methods 138
8.8 Effect of Adaptive Targeting 138
8.9 Effect of Moral Suasion 146

8.10 Conclusion 147

9 Conclusion and Further Work **149**

III Appendices 150

A Proofs **151**
A.1 Proofs for Chapter 2 151
A.2 Proofs for Chapter 3 152
A.3 Proofs for Chapter 4 157
A.4 Proofs for Chapter 5 161

B Supplementary Material **183**
B.1 Supplementary Material for Chapter 8 183

Bibliography **191**

List of Figures

2.1 Energy Market Participants for DR and their Interactions 9
2.2 DR Mechanism Timeline 13
2.3 Lognormal Consumption Distribution Fit for Selected User, 5-6 pm 17
2.4 Compound Statistics for Lognormal Consumption Distribution. Left: Shape, Middle: Location, Right: Scale 18
2.5 Number of Targeted Users and Total Payment to Users for DR Mechanism (blue) vs. Omniscient Allocation (green), $n = 500,\ q = 5.0,\ \alpha_i \sim \text{unif}[0.05, 0.06]$. 19
2.6 Composition of Target Aggregate Reduction M for varying Baselines. Red: $\sum_{i\in\mathcal{T}} \delta_i^{\text{BL}}$. Blue: $\sum_{i\in\mathcal{T}} \delta_i^{r}$. Parameters: $n = 500, q = 5.0, \alpha_i \sim \text{unif}\ [0.05, 0.06]$ . . 20
2.7 Payments to Users to Elicit M for varying Baseline Accuracies. 20

3.1 $\|W - \tilde{W}\|_2$ for 12 fully connected users embedded in a network of $n = 24$ customers, $\gamma = 0.05$. 30
3.2 Basic network architectures for $n = 4$: Fully connected, star, ring 30
3.3 Profit of monopolist, single prices (3.10) and (3.13a), average user consumption, and maximum consumption under perfect price discrimination (green), single pricing under complete information (3.10) (yellow), and single pricing under incomplete information (3.13a) (red). 10,000 iterations, $a \sim \text{unif}[8, 12],\ b \sim \text{unif}[0.75, 1.25],\ c_i = 2\ \forall\ i \in \mathcal{I}$. 32
3.4 Average profit, percentage of optimal choice of heuristic, regret, and average infinity norm of consumption for the utility's profit maximization problem under the single price (3.10). 10,000 iterations, $a \sim \text{unif}[8, 12],\ b \sim \text{unif}[0.75, 1.25],\ c_i = 2$. 35

4.1 Energy Market Participants and their Interactions 39
4.2 Timeline of Hedging 40
4.3 Distribution of Aggregate Hourly Consumption for Varying Aggregation Sizes, 4-6 pm. Top: 250 Users, Middle: 150 Users, Bottom: 50 Users. 47
4.4 Distributions of CAISO LMPs conditional on previous prices exceeding threshold ξ for $\xi \in \{80\frac{\text{USD}}{\text{MWh}}, 90\frac{\text{USD}}{\text{MWh}}, 100\frac{\text{USD}}{\text{MWh}}\}$. 48
4.5 Boundaries and contours of equal expected profit for forward option and DR, 250 users, $\lambda_f = 0.05\frac{\text{USD}}{\text{kWh}}$. 49

4.6 Boundaries and contours of equal expected profit for DR and call option, 250 users, $\lambda_f = 0.05\frac{\text{USD}}{\text{kWh}}$. 50
4.7 Boundaries and contours of equal expected profit for forward and call option, 250 users, $\lambda_f = 0.05\frac{\text{USD}}{\text{kWh}}$. 50

6.1 Zones for the 4th Floor of Sutardja Dai Hall (SDH) 72
6.2 Estimated Internal Gain q_{IG} from the Data-Driven Model by Season, Lumped Case 76
6.3 Estimated Internal Gain q_{IG} from the Data-Driven Model by Zone and Season, Individual Case 78
6.4 Estimated Internal Gain f_{IG} from the Physics-Based Model by Zone and Season 80
6.5 Simulated Temperatures from the Data-Driven Model (blue), Physics-Based Model (orange) and Actual Temperatures (green) 82
6.6 Optimal Temperature for MPC with Data-Driven Model (blue), MPC with Physics-Based Model (orange) and Actual Temperature (green) 84
6.7 Optimal Control Strategy for MPC with Data-Driven Model (blue), MPC with Physics-Based Model (orange) and Actual Input (green) 85

7.1 Hidden Markov Model. Hidden States q, Observations y 92
7.2 Markov State Transition Diagram, 24 Hour Periodicity. For Example, "5" Signifies Time Between 5 a.m. - 6 a.m. 93
7.3 Prediction Accuracy by Forecasting Method. "+" Signifies Model with HMM Latent Variable, "Mix." Denotes CGMM. Blue Boxes Span 25-75th Percentile, Whiskers 10-90th. 101
7.4 Pointwise Prediction Error of DR Treatment Effect on the User Level; Bias $\hat{\mu}$ and Variance $\hat{\sigma}^2$ of Model with Latent Variable from HMM. 102
7.5 Estimated Reduction Across Users by Hour of Day (Yellow) vs. Estimated Reduction for Placebo Events (Gray) for Automated and Non-Automated Users Conditional on Estimated Latent Variable. Red: Median, Green: Mean. Blue Boxes Span 25-75th Percentile, Whiskers 10-90th. 103

8.1 Interactions of Agents in Residential Demand Response 110
8.2 Setup of Experiment 111
8.3 Distribution of Enrollment over Time 113
8.4 Distribution of Utility Account Connections over Time 113
8.5 Availability of Smart Meter Data Across Experimental Users 114
8.6 Geographic Distribution of Enrolled Users 114
8.7 Geographic Distribution of Users 116
8.8 Number of Messages Sent to Users 116
8.9 Number of Phase 2 DR Events per Individual Over Time 117
8.10 Distribution of Phase 1 DR Events by Hour of the Day 118
8.11 Distribution of Phase 1 DR Events by Day of the Week 118
8.12 Mean Consumption for Targeted vs. Non-Targeted Users by Hour of the Day . . 119

8.13 Mean Temperature for Targeted vs. Non-Targeted Users by Hour of the Day . . 120
8.14 Estimation of the Counterfactual $\hat{y}^0_{it}$ using Treatment Covariates $\mathbf{x}_{it}$ and Predicted Reduction $\hat{y}^0_{it} - y^1_{it}$ 123
8.15 Separation of consumption time series into training set (green), DR Events (grey), and Periods of Spillover Effects (blue) 124
8.16 Mean Prediction Bias on $\mathcal{D}_{i,p}$ and Sample Standard Deviation Across Users for Selected Estimators 126
8.17 Distribution of User-Level Prediction Biases on Placebo Events 127
8.18 Distribution of Event-wise Residuals for Selected Estimators 130
8.19 Deciles of Non-Targeted Users Ranked by ITE estimated in both Phase 1 and Phase 2 of the Experiment 131
8.20 95% Confidence Intervals for Estimators and Varying Group Sizes (5, 10, 25, 50 users) 132
8.21 CATEs by Incentive Level with Bootstrapped Confidence Intervals 133
8.22 Distribution of ITEs with Bootstrapped Confidence Intervals 135
8.23 CATEs by Incentive Level with Confidence Intervals, Comparison Fixed Effects Estimators and Non-Experimental Estimators 138
8.24 Distribution of Payouts to Users 141
8.25 Differences in Estimated Reduction between Non-Targeted and Targeted Users, computed with RF 142
8.26 Estimated Reductions of Non-Targeted and Targeted Users, computed with RF 143
8.27 Estimated Reductions in Phase 3 147

B.1 Distribution of Number of Phase 1 DR Events Across Users with Completed Phase 1183
B.2 Distribution of DR Events by Hour of the Day Across Users with Completed Phase 1 184
B.3 Distribution of DR Events by Day of the Week Across Users with Completed Phase 1 184
B.4 Distribution of DR Events by Month of the Year Across Users with Completed Phase 1 184
B.5 CATEs by Month of Year with Confidence Intervals, Comparison Fixed Effects Estimators and Non-Experimental Estimators 186
B.6 Correlation between Average Ambient Air Temperature and ITEs. 189
B.7 Correlation between Average Ambient Air Temperature and CATEs. 190

List of Tables

2.1 Example User Characteristics 15

5.1 Regret Bounds in Adversarial and Stochastic Bandit Settings 54

6.1 VAV Boxes by Zone 72
6.2 RMS by Zone and Season for Data-Driven and Physics-Based Models 79

8.1 Number of Total Users Enrolled by Group, Data Availability, and Users with DR Events as of September 28, 2017. 115
8.2 Number of Phase 1 DR Events for Encouraged and Non-Encouraged Users by Incentive Level 115
8.3 Number of Phase 2 DR Events for Low Targeted, High Targeted, and Non Targeted Users by Incentive Level 117
8.4 Standard Deviation of Event-wise Residuals ($y_{it} - \hat{y}_{it}$) 131
8.5 Pearson Correlation Coefficient for Non-Targeted User Ranks Across Phase 1 and Phase 2, Minimum 20 Events per User 132
8.6 ATE Estimates and Demand Curve by Estimator, all 4791 Users 133
8.7 Width of 95 % Confidence Intervals around ATE Point Estimate Conditional on Incentive Level, All Estimators 134
8.8 Estimated CATEs by Automation Status, RF Estimator (E6) 134
8.9 Estimated Percentage of Significant Reducers according to Permutation Test, RF Estimator (E6) 136
8.10 Width of 95 % Confidence Intervals around ATE Point Estimate by Incentive Level, All Estimators 139
8.11 Targeting Results for 2,733 users between June 29, 2017 - December 31, 2017. $[H_1] = \frac{\text{kWh}}{\text{USD}}$, $[H_2] = \frac{\text{kWh}^2}{\text{USD}^2}$, [CATE] = kWh. 146
8.12 Aggregated Phase 3 Reductions 147

B.1 Balance Checks for Users in Control and Treatment Group 185
B.2 Fixed Effect Regression Results by Incentive Level 186
B.3 Fixed Effect Regression Results by Month of Year 187
B.4 Fixed Effect Regression Results by Automation Status 188
B.5 Fixed Effect Regression Results by Automation Uptake Encouragement 188

Acknowledgments

First and foremost, I would like to thank my PhD adviser, Prof. Claire Tomlin, for supporting me throughout my academic journey. Your ability to point me to relevant resources, experts in specific research domains, as well as the independence you have granted me allowed me to get a glimpse into the academic world.

I would also like to thank Maximilian Balandat, James Gillan, and Qie Hu for fruitful collaborations that have resulted in numerous publications. Your support especially in the earlier stages of my PhD program has proven to be invaluable.

Moreover, I am grateful for the supervision I received from Prof. Munther Dahleh and Mardavij Roozbehani during my exchange at MIT.

Lastly, I appreciate the support of my PhD committee. You deserve the utmost recognition and praise.

Chapter 1

Introduction and Background

Information technology has arguably been the most fascinating, yet disruptive development of the last few decades. Data from all possible sources is collected at an ever-increasing rate, and while it appears that many people are growing more concerned about privacy issues, it is undeniable that even the most skeptical person is likely to benefit from digital products in at least some ways. What makes information technology such a dominant force as opposed to other classical industries that have existed for centuries? While we certainly will not be able to answer this question exhaustively, one possible answer is the innate capability of information technology to use granular data to make predictions, infer causal relationships, and to design policies that allow us make decisions quickly and often with a provable guarantee on performance.

While data itself has no inherent value (it does not pay any dividend) and does not improve in quality over time like a good bottle of wine, it is the academic community and entrepreneurs who have developed new fields of study that harness the availability of "big data". Artificial Intelligence and Machine Learning have given rise to entirely new academic areas of research. The "Internet of things" connects individual users in fundamentally new ways, allowing fast and flexible communication between agents. Examples for such industries are transportation, utilities, retail, healthcare, financial services, and manufacturing.

This thesis seeks to shed light on the role of large volumes of data on electricity markets, which is a prime example for how a traditional industry has been disrupted and modernized by the ability to collect, store, and transmit data between individual agents. The electricity market has a few peculiarities that are worth mentioning, such as the difficulty of forecasting both demand and supply, an issue that is exacerbated by the increasing penetration of intermittent and variable renewable energy sources, the insensitivity of demand to price fluctuations, and its prohibitive cost of storage. These issues result in highly volatile short-term prices of electricity, which has spurred areas of research at the intersection of machine learning and economics to alleviate the previously mentioned issues.

Demand-side management (DSM) is an umbrella term that encompasses a wide range of interventions aiming to alleviate supply and demand imbalances of electricity, often by integrating end-users of electricity as virtual power plants into the control loop that exists be-

tween generators and load-serving entities (LSEs). The idea is to communicate the marginal price of electricity in an approximately real-time fashion to end-users in order to incentivize users to behave according to a certain objective, which often corresponds to temporarily reducing or increasing their consumption to help align supply and demand of electricity. Such end-users of electricity span a wide range of entities as small as residential customers and as large as commercial buildings or factories. Timescales of such interventions range from seconds to days ahead. While there exists a plethora of work from engineers and economists on day-ahead interventions, the role of short-term interventions remains largely unexplored, a gap that ultimately motivates the creation of this thesis.

Outline

This thesis consists of two main parts. The first part is a theoretical section in which we develop a few tools that allow us to quantify the role of private information and peer effects in electricity consumption as well as strategies to hedge against price and quantity risks in dynamic electricity markets. The purpose of this section is two-fold: On one hand, it contributes to the extant literature in mechanism design, game theory, and online learning, and therefore are interesting topics in their own right. On the other hand, the theoretical section lays the foundation for the second part of this thesis, which is concerned with the investigation of demand-side management for commercial buildings and residential households. The tools and theorems introduced in the theoretical section thus naturally motivate a more practical perspective, which is covered by case studies in California.

Chapter 2 explores the role of private information in Demand Response. The principal agent model has long existed in literature and is a classical framework in economics to model information asymmetries between a principal and its agents. To incentivize users to behave according to a desired objective, the principal needs to design a suitable mechanism in order to elicit private user information. We develop such a mechanism in the context of Residential Demand Response, in which a principal requests a set of households to temporarily reduce their electricity consumption in exchange for a monetary incentive. Each household, however, has a certain willingness to reduce, which is parameterized by variables known only to the household itself. The challenge for the principal is to design an incentive compatible and individually rational mechanism that deals with the heterogeneity of users in their price elasticity of demand, elicits aggregate reductions that are large enough, and does so in a cost-efficient fashion. The success of this mechanism hinges on a reliable estimator that predicts electricity consumption on the individual user level, as the amount of prediction inaccuracy correlates negatively with cost efficiency, a fact we support with numerical simulations.

Chapter 2 is succeeded by Chapter 3, which shifts our attention to the interaction between end-users of electricity, rather than the interactions between a principal and a pool of end-users. This chapter is motivated by case studies that investigate the effect of peer pressure on electricity consumption. To answer this question, we develop a two-stage game theoretic model that reflects the behavior of both the principal and each individual agent. This two-stage game can then be solved − under certain assumptions − by finding a subgame

perfect equilibrium. We derive optimal pricing mechanisms for various information levels and simulate the optimal profit of the principal. Further, we investigate the role of network uncertainty on the pricing strategy.

Next, Chapter 4 aims to provide a more holistic model of price and quantity risks that ensue from the interplay between generators of electricity, the electric wholesale market, electric utilities, and end users of electricity. These risks are a logical consequence of the peculiarities of electricity as a good, namely the inelasticity of user demand, prohibitive cost of storage, and its steep supply curve. These factors imply that small changes in demand or supply of electricity result in disproportionately large changes in prices. The purpose of this chapter is to investigate to what extent such price risks can be hedged against with bilateral contracts between electric utilities and generators. Such mechanisms could be forward contracts, call options, as well as Demand Response, all of which we investigate in more detail. Further, we quantify how the amount of uncertainty affects the optimal expected profit of the load serving entity. Lastly, we use numerical simulations to do pairwise comparisons between the options to illustrate decision boundaries of equal expected profit.

The last chapter of the theoretical section of this thesis (Chapter 5) analyzes the multi-armed bandit problem with a budget. We extend the basic framework, which was first introduced by Auer et al. 2002, to a setting in which the player has to play exactly K out of N possible arms. Furthermore, playing each arm comes with a reward and a cost, both of which are uniformly distributed (the original setting does not involve a cost component). The idea of this chapter is to derive an abstract model that quantifies repeated interactions between a principal and a pool of possible agents. Indeed, all previous chapters involve repeated interactions between such a principal (load serving entity) and a set of possible arms (end users of electricity), and taking an action at each round typically comes at a cost. The principal has a budget $B > 0$ to spend on interventions and seeks to maximize the reward. We investigate this problem for both the stochastic and adversarial setting.

The applied section of this thesis begins with a discussion about the controllability of commercial buildings for frequency control (Chapter 6). We derive two distinct models for the evolution of temperature as a function of the control input, which is the airflow by the HVAC system. The first model is a purely data-driven model that makes no structural assumptions other than its linearity between input and output. The second model, however, is a physics-based model that is inspired by modeling individual components of the buildings as a resistance-capacitance model. While the first model is a higher fidelity model, the second one is capable of modeling individual zones of the building with higher accuracy, as it naturally partitions the building into more states. The fact that the data-driven model is leaner allows us to perform energy efficient control in a relatively straightforward fashion. Respecting comfort constraints on the temperature, a model predictive control scheme can noticeably reduce the cost of the HVAC system.

The final two chapters of this thesis are devoted to the investigation of a case study on Residential Demand Response in California. Chapter 7 presents an overview of short-term load forecasting methods that seek to predict the consumption of residential households for the next hour. We contribute to this thoroughly researched field by proposing Hidden

Markov Models and Conditional Mixtures of Gaussians that are specifically tailored to residential smart meter data. By assuming exactly two latent states, namely "low" and "high" consumption, we can derive predictive models for each of these states, a mixture of which noticeably improves prediction accuracy compared to a baseline estimator that does not make use of such latent states. The Hidden Markov Model further assumes transition probabilities between these latent states, as a user who is in the "low" ("high") state is most likely to remain in the "low" ("high") state in the next hour, achieving the highest estimation accuracy.

Chapter 8 presents the findings of a randomized controlled trial on approximately $10,000$ residential customers that investigates the causal effect of short-term monetary incentives on the reduction in electricity consumption. This study is carried out in collaboration with OhmConnect, Inc., a start-up company based in the San Francisco Bay Area. This experiment includes three phases. Phase 1 aims to estimate the Average Treatment Effect across the entire subject population. Phase 2 seeks to adaptively target residential customers by sending customized messages with varying incentive levels in order to improve cost efficiency from the perspective of the DR provider. Phase 3 is dedicated to the effect of moral suasion, namely whether or not users respond to incentives that solely appeal to the environmental consciousness of users. This experimental approach is replicated by using a non-experimental estimation framework, which is able to provide more granular estimates of treatment effects, namely on the *individual* user level. As these user-specific models are not inherently causal − estimation bias plays an important role − the randomized controlled trial serves as a ground truth benchmark for non-experimental estimates.

Finally, Chapter 9 concludes this thesis.

Part I

Theory

Chapter 2

Eliciting Private User Information in Demand Response

2.1 Introduction

With the restructuring of the traditional, vertically integrated energy market towards a competitive market, Demand-Side Management (DSM) has become a viable tool for alleviating supply and demand imbalances of electricity. Facilitated by advancements in information and communications technology, smart metering infrastructure allows end-users of electricity to "participate" in the electric market as virtual power plants through properly designed incentive mechanisms. DSM is motivated by the inelasticity of energy supply, which causes small variations in demand to result in a price boom or bust, respectively. These price fluctuations are aggravated by the inherent volatility of renewable generation resources, their increasing levels of penetration, and the prohibitively high capital cost of energy storage. Since a load-serving entity (LSE) is required to procure electricity at fluctuating prices to cover the electricity demand of its residential households under contract instantaneously and at quasi-fixed tariffs, price risks are almost entirely borne by the LSE. Incentivizing users to temporarily reduce their consumption (and charging a fee if users do not reduce) during periods of high prices therefore partially passes such price risks on to customers.

While the area of DSM has attracted a vast array of research across different domains (see Palensky and Dietrich (2011) for a summary), we in this chapter focus on the area of Demand Response (DR), where end-users of electricity are incentivized to reduce their demand temporarily during designated hours, precisely when there is a shortage of electricity supply. Users receive a reward for each unit of reduction, but incur a penalty for increasing their consumption. Demand Response providers (DRPs) bundle these reductions and can offer these reductions as a bid directly into the competitive wholesale electricity market, or enter bilateral contracts with load-serving utilities. While DR is traditionally carried out on a commercial level, residential customers are targeted for load reduction programs, as well. For instance, in California, the Public Utilities Commission (CPUC) launched a "Demand

Response Auction Mechanism" (DRAM) in July 2015 (*Resolution E-4728. Joint Utility Proposal for a DRAM Pilot*) to allow DRPs to offer reduction capacity from residential customers directly into the day-ahead electricity market, where they are subject to regular market clearing prices and shortfall penalties. Utilities are required to purchase a fixed minimum monthly amount of this reduction capacity.

To make an informed capacity bid into the market, the DRP must take various factors into account, such as the expected Locational Marginal Price (LMP) which determines its market clearing price, the elasticity of users' demand given an incentive, and the number of users under contract. If the DRP bids too much capacity, the aggregate reduction among its user base will likely fail to reach the capacity volume, thereby incurring a shortfall penalty; similarly, a suboptimal revenue arises from too small a bid. The DRP can improve its bidding strategy by learning users' behavior in response to incentives. However, users' preferences are typically private information and hence unknown to the utility. The challenge thus becomes to elicit this private information. We cast this problem as a mechanism design problem, where the DRP as the auctioneer solicits bids from each of its residential customers through an *incentive compatible* and *individually rational* mechanism. The motivation behind this approach is to increase allocative efficiency, that is, the utility would like to solicit reductions only from the highest reducers, who are most willing to reduce their consumption in exchange for the lowest possible reward. In this chapter, we design such a mechanism that fulfills these criteria and benchmark its performance against the omniscient case, where user characteristics are common knowledge.

A crucial question that arises from this setting is how to measure the reduction of any individual user during a DR event, given that only the consumption outcome under a treatment can be observed, but not its counterfactual (the consumption had there been no DR event). This is the fundamental problem of causal inference (Holland 1986). To estimate the reduction during any particular DR event, it is thus essential to estimate the counterfactual, which we refer to in this context as "baseline". Estimating this baseline in the absence of a Randomized Controlled Trial is a modern area of research at the intersection of economics and machine learning. Examples for such baseline estimates can be found in Athey and Imbens (2016), Abadie, Diamond, and Hainmueller (2012), Zhou, Balandat, and Tomlin (2016b), and Zhou, Balandat, and Tomlin (2016a). In this chapter, we employ the "10-in-10" baseline employed by the California Independent System Operator (CAISO) (*CAISO Fifth Replacement FERC Electric Tariff*), which estimates the counterfactual for a particular DR event as the mean consumption of the 10 most recent days during the same hour as the DR event. Using this baseline, the measured reduction for any selected user can be formulated as the sum of a virtual reduction, which reflects the estimation error in the baseline prediction, and the actual reduction due to price elasticity of user demand. We observe that the DR provider can achieve a virtual reduction from those users for which the baseline is high. That is, the DR provider receives payments for virtual, non-existent reductions which are indirectly paid for by utilities. However, we show that a more accurate baseline diminishes the impact of such virtual reductions.

2.1.1 Related Work

Modeling consumer behavior in response to monetary incentives in DR and their heterogeneity is a growing area of research. In Kwac and Rajagopal (2013), the authors formulate the problem of targeting the "right" customers for DR as a stochastic knapsack problem in order to achieve a target reduction with high probability. However, users' responses are modeled as a linear model without private user information.

Other works have incorporated a contractual formulation between consumers and suppliers in DR settings. For example, Li, Chen, and Dahleh (2015) design a DR market where suppliers bid supply curves in the presence of a supply shortage to the load-serving entity and analyze the ensuing market equilibria. In Balandat et al. (2014), the authors formulate a contract between an aggregator of buildings, individual buildings, and the wholesale electricity market to exploit flexibility of commercial buildings' HVAC consumption. In a similar fashion, Han, Han, and Sezaki (2010) formulate a contract design problem between an aggregator and individual electric vehicle owners to maximize its revenue by providing power capacity to the grid operator.

To quantify the impact of DR signals on the reduction of consumption, Zhou, Balandat, and Tomlin (2016b) and Zhou, Balandat, and Tomlin (2016a) estimate individual treatment effects in response to hourly DR events by comparing the estimated counterfactual consumption to the actual, observed consumption. Li and Zhang (2016) formulates an optimal treatment assignment strategy to precisely measure the treatment effect of DR.

The application of Mechanism Design on DR is covered in Samadi et al. (2012), where the authors maximize the social welfare of consumers and the energy provider by designing a consumption controller with a Vickrey-Clarke-Groves auction. In Ma et al. (2016) and Li and Li (2016), the authors incorporate uncertainty into consumers' reduction behavior and introduce the notion of reliability for achieving a designated amount of aggregate reduction.

2.1.2 Contributions

Unlike previous works, which modeled reductions as multiples of unit reductions, we account for the *Fundamental Problem of Causal Inference* (Holland 1986) into the mechanism design formulation between DRP and users, which is our main contribution. Specifically, we estimate reductions using the CAISO "10-in-10" baseline as the counterfactual estimate. As a consequence of uncertain baseline predictions, *virtual reductions* arise. Using observational data from residential customers in California, we quantify the extent to which these virtual reductions counteract DR, and how these reductions diminish as baseline estimates become more precise.

2.1.3 Notation

Let $[\cdot]_+ = \max(0, \cdot)$. Vectors are printed in boldface. Let $\mathbf{a}_{-i}$ denote the vector of all components in $\mathbf{a}$ excluding i. $\mathbf{1}_{(\cdot)}$ denotes the indicator function.

2.1.4 Outline

The remainder of this chapter is organized as follows: Section 2.2 characterizes DR market participants and their interactions, based on which Section 2.3 presents a mechanism for the DR provider to elicit private user information and to achieve an aggregate reduction among its users under contract. Section 2.4 elucidates the difference between virtual and actual reductions as an artifact of an uncertain baseline estimate. The mechanism is simulated on residential smart meter data in California in Section 2.5, where we experimentally show how more accurate baselines reduce the amount of virtual reductions. Section 2.6 concludes. All proofs are relegated to Appendix A.1.

2.2 Market Participants and Interactions

2.2.1 Residential Demand Response

Figure 2.1 describes the interaction between the DRP, end-users, the electric utility, and the wholesale electricity market.

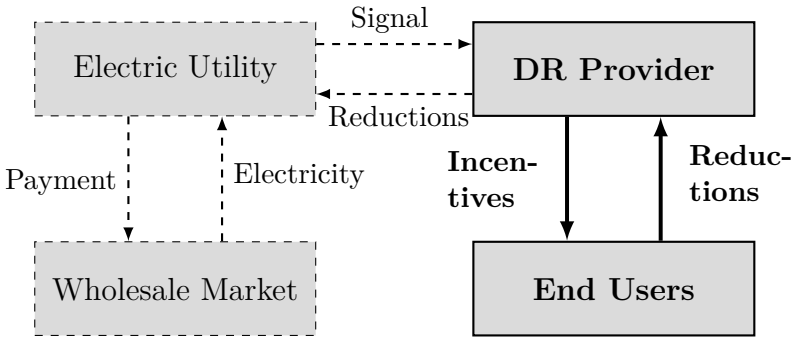

Figure 2.1: Energy Market Participants for DR and their Interactions

The DRAM requires electric utilities to acquire demand flexibility from DRPs, which they submit as part of their supply curves as a bid into the real-time wholesale electricity market. If these bids are cleared, the utility sends the DRP a signal to ask for a specified aggregate reduction among its users. The DRP elicits reductions by incentivizing a subset of its customers $\mathcal{T} \subseteq \mathcal{I}$ with user-specific per-unit rewards $\{r_i \in \mathbb{R}_+ \mid i \in \mathcal{T}\}$, where $\mathcal{I} = \{1, \ldots, n\}$ denotes the set of users. In exchange for the monetary incentive, users reduce consumption by $\{\delta_i \in \mathbb{R} \mid i \in \mathcal{T}\}$. A per-unit penalty $q \in \mathbb{R}_+$, which is assumed to be identical for all users and common knowledge, is enforced for an increase in consumption beyond the baseline. Users in the non-targeted group $\mathcal{I} \setminus \mathcal{T}$ are excluded from the incentive program. In this thesis, we only focus on the interaction between the DRP and the end-users from the perspective of the DRP. To maximize its profit, the goal of the DRP is to achieve an a-priori defined aggregate reduction with minimal payments to its users.

2.2.2 Residential Customers

Each rational, profit-maximizing user $i \in \mathcal{I}$ is endowed with the "10-in-10" baseline $\hat{x}_i \in \mathbb{R}_+$ employed by the California Independent System Operator (CAISO) (*CAISO Fifth Replacement FERC Electric Tariff*), which is an estimate of her counterfactual consumption (*PJM Empirical Analysis of Demand Response Baseline Methods*) for a particular hour. For notational ease, we drop time indices, but we emphasize the need to re-calculate $\hat{x}_i$ for any individual hour. The baseline for a particular hour on a weekday is calculated as the mean of the hourly consumptions on the 10 most recent business days during the hour of interest. For weekend days and holidays, the mean of the 4 most recent observations is calculated. User i's measured load reduction δ_i, provided she is given incentive r_i to reduce during a particular hour, is simply the difference between the baseline $\hat{x}_i$ and the actual, materialized consumption x_i:

$$\delta_i = \begin{cases} 0 & , \text{ if } \quad i \notin \mathcal{T} \\ \hat{x}_i - x_i & , \text{ if } \quad i \in \mathcal{T} \end{cases} \tag{2.1}$$

Due to the widespread existence of advanced metering infrastructure, the baseline $\hat{x}_i$ is assumed to be common knowledge among the DRP and user i. The utility of user i is defined as follows:

$$u_i = \begin{cases} 0 & , \text{ if } \quad i \notin \mathcal{T} \\ r_i \cdot [\hat{x}_i - x_i]_+ - q \cdot [x_i - \hat{x}_i]_+ & , \text{ if } \quad i \in \mathcal{T} \end{cases} \tag{2.2}$$

which equals the payment from the DRP to user i. That is, if the user is under a DR contract with the DRP, she is rewarded with $r_i \in \mathbb{R}_+$ for each unit of reduction, and charged q for each unit of consumption above the baseline $\hat{x}_i$.

We model users' consumption in response to r_i, denoted with $x_i(r_i)$, with a semi-logarithmic demand curve, an assumption frequently made in economics:

$$\begin{aligned} x_i(r_i) &= \bar{x}_i \cdot \exp(-\alpha_i r_i) \\ \log x_i(r_i) &= \log \bar{x}_i - \alpha_i r_i \qquad \forall\ i \in \mathcal{I} \end{aligned} \tag{2.3}$$

In (2.3), $\bar{x}_i \in \mathbb{R}_+$ and $\alpha_i \in \mathbb{R}_+$ are random variables signifying the base demand (the intercept or the consumption with $r_i = 0$) and the slope of the demand curve in log-linear coordinates, respectively. This semi-logarithmic demand curve captures the fact that the amount of reduction is marginally decreasing in the reward r_i and saturates. User i's type $\boldsymbol{\theta}_i$ is information correlated with $(\bar{x}_i, \alpha_i)$ (not necessarily $(\bar{x}_i, \alpha_i)$ itself) and user i's private information.

2.2.3 Demand Response Provider

The DRP aims to maximize its profit Π in expectation:

$$\Pi = \bar{r} \cdot \min(\Delta, M) - \bar{q} \cdot [M - \Delta]_+ - \sum_{i \in \mathcal{I}} \delta_i \left(r_i \cdot \mathbf{1}_{\delta_i > 0} + q_i \cdot \mathbf{1}_{\delta_i \leq 0} \right). \tag{2.4}$$

Π is random in $\delta_1, \ldots, \delta_n$. $\Delta = \sum_{i \in \mathcal{I}} \delta_i$ is the total sum of reductions and $M \in \mathbb{R}_+$ the target capacity the DRP has to provide to the utility. $\bar{r}$ and $\bar{q} \in \mathbb{R}_+$ denote the per-unit reward and shortfall penalty the DRP is subject to in the wholesale electricity market. Note that $\bar{q} \neq q$ and $\bar{r} \neq r_i$. The first term of (2.4) represents the profit the DRP earns for materialized reductions, the second term captures the shortfall penalty for unfulfilled reductions, and the last term is the sum of payments disbursed to individual customers.

Assumption 1. *The DRP is risk-neutral and profit-maximizing.*

Assumption 2. *The per-unit penalty $\bar{q}$ in the wholesale electricity market and the per-unit reward $\bar{r}$ are greater than the maximum per-unit reward disbursed to any customer, i.e.* $\min(\bar{q}, \bar{r}) > \max_{1 \leq i \leq n}(r_i)$.

With Assumptions 1 and 2, (2.4) can be rewritten as follows:

$$\begin{aligned} \underset{r_1,\ldots,r_n}{\text{minimize}} \quad & \mathbb{E}_{\delta_1,\ldots,\delta_n}\left[\sum\nolimits_{i \in \mathcal{I}} \delta_i \left(r_i \mathbf{1}_{\delta_i > 0} + q_i \mathbf{1}_{\delta_i \leq 0}\right)\right] \\ \text{subject to} \quad & \mathbb{E}_{\delta_1,\ldots,\delta_n}\left[\sum\nolimits_{i \in \mathcal{I}} \delta_i\right] \geq M. \end{aligned} \tag{2.5}$$

That is, the DRP aims to find an optimal vector of per-unit rewards $\mathbf{r}^*$ that minimizes the expected total amount of payments disbursed to the users while satisfying the constraint that the expected sum of reductions exceeds M.

2.3 Demand Response Mechanism

To find an approximation to the solution of (2.5), the utility needs to elicit user i's private type $\boldsymbol{\theta}_i$ with an incentive compatible (IC) and individually rational (IR) mechanism. IR guarantees that participation in the mechanism, provided users act rationally, results in an expected payoff that is at least as large as in the case of non-participation (outside option), which is zero in our case (2.2). IC is required to ensure that users report their types truthfully to the DRP.

2.3.1 Mechanism Design Basics

We first introduce basic notation relevant to our problem. Let $\boldsymbol{\theta}$ denote the collection of types $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$, where each $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i \; \forall \; i \in \mathcal{I}$ is drawn from its type space $\boldsymbol{\Theta}_i$. It is assumed that $\boldsymbol{\theta}$ is drawn from a commonly known joint distribution F defined on the product space $\boldsymbol{\Theta} = \times_{i=1}^{n} \boldsymbol{\Theta}_i$. Each agent is assumed to seek expected utility maximization of her utility function $u_i(\mathbf{y}, \boldsymbol{\theta}_i) : \mathcal{Y} \times \boldsymbol{\Theta}_i \mapsto \mathbb{R}$, where $\mathbf{y} = (\mathbf{d}, \mathbf{r}) \in \mathcal{Y} = \{0, 1\}^n \times \mathbb{R}_+^n$ is the collective choice consisting of the vector of allocation decisions $\mathbf{d}$ and the vector of rewards $\mathbf{r}$. The social choice function $f(\boldsymbol{\theta}) : \boldsymbol{\Theta} \mapsto \mathcal{Y}$ maps a particular collection of types $\boldsymbol{\theta}$ to $\mathbf{y}$.

Let $\mathcal{S}_i, \ldots, \mathcal{S}_n$ denote the strategy spaces of users $i \in \mathcal{I}$. A realized strategy vector $\mathbf{s} \in \times_{i=1}^{n} \mathcal{S}_i$ defines an outcome function $g(s_1, \ldots, s_n) : \times_{i=1}^{n} \mathcal{S}_i \mapsto \mathcal{Y}$. Together they define a mechanism $\Gamma = (\mathcal{S}_1, \ldots, \mathcal{S}_n, g(\cdot))$, which transforms users' strategies into a social choice function through the outcome function $g(\cdot)$. $(\Gamma, F, \{u_i\}_{i=1}^{n})$ defines a Bayesian Game with payoffs $u_i(g(s_1, \ldots, s_n), \boldsymbol{\theta}_i)$ and strategies $\mathbf{s}_i : \boldsymbol{\Theta}_i \mapsto \mathcal{S}_i$.

The *revelation principle* (Osborne and Rubinstein 1994) allows us to focus on direct mechanisms, i.e. $\mathcal{S}_i = \boldsymbol{\Theta}_i$ and $g(s_1, \ldots, s_N) \equiv g(\boldsymbol{\theta}) = f(\boldsymbol{\theta})$, which is the well-known fact that any equilibrium of any mechanism is identical to an equilibrium of a direct mechanism, provided truthful reporting. We focus on the dominant strategy equilibrium:

Definition 1 (Dominant Strategy Equilibrium (DSE))**.** *A Dominant Strategy Equilibrium is given by*

$$\boldsymbol{\theta}_i = \arg\max_{\mathbf{z}_i \in \boldsymbol{\Theta}_i} \mathbb{E}_{\mathbf{z}_i} \left[u_i(f(\mathbf{z}_i, \mathbf{z}_{-i}), \boldsymbol{\theta}_i) \right] \quad \forall i \in \mathcal{I},\ \mathbf{z} \in \boldsymbol{\Theta} \tag{2.6}$$

That is, if the supremum of user i's expected utility u_i is achieved with truthful reporting $s_i^*(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_i$, regardless of other users reports $\mathbf{z}_{-i} \in \boldsymbol{\Theta}_{-i}$, then the social choice function $f(\cdot)$ is dominant strategy incentive compatible.

2.3.2 Timing, User Types, and Reward Calculation

The DR mechanism unfolds as follows:

- The users $i \in \mathcal{I}$ discover their types $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$. The baselines $\hat{x}_1, \ldots, \hat{x}_n$ become common knowledge.
- The users reveal their types $\{\mathbf{z}_i\}_{i=1}^{n}$ to the DRP, where $\mathbf{z}_i$ not necessarily corresponds to the true type $\boldsymbol{\theta}_i$.
- The DRP implements the collective choice $f(\mathbf{z}) = \mathbf{y} = (\mathbf{d}, \mathbf{r})$ through mechanism Γ.
- Users observe $f(\mathbf{z})$ and adjust their consumption according to (2.3) and d_i, r_i.

For better visualization, Figure 2.2 depicts these steps.

An important observation is that, after the implementation of $f(\mathbf{z})$ at $t = 2$, the DRP calculates its *expected* profit $\mathbb{E}[\Pi]$ and the *expected* payments disbursed to each user i. Due to the Myerson-Satterthwaite Theorem (Myerson and Satterthwaite 1983), we do not perform any ex-post analysis on the *realized* consumptions $\mathbf{x}(\mathbf{r})$ at $t = 3$.

To model the fact that users' base electricity consumption is often driven by habits rather than rational profit-maximization (Maréchal 2010), we assume the user-specific intercept $\bar{x}_i$ to be drawn from an a-priori defined distribution G with characteristic parameters $\boldsymbol{\xi}_i$ encoded in user i's private type. $\boldsymbol{\xi}_i$ itself is distributed according to the joint distribution $F_{\boldsymbol{\xi}}$, and so $\bar{x}_i$ is a compound random variable. The slope, however, is assumed to be explicitly known for each user and drawn from distribution F_α. Thus $\boldsymbol{\theta}_i = (\alpha_i \sim F_\alpha, \boldsymbol{\xi}_i \sim F_{\boldsymbol{\xi}})$,

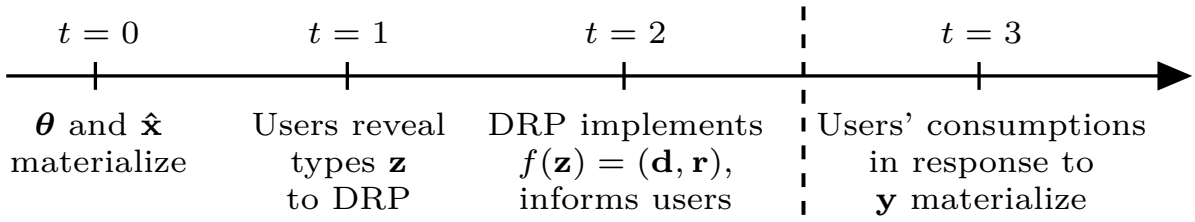

Figure 2.2: DR Mechanism Timeline

where $\bar{x}_i \sim G_{\boldsymbol{\xi}_i} \sim G_{\boldsymbol{\xi}_i \sim F_{\boldsymbol{\xi}}}$. All distributions have support on $\mathbb{R}_+$. We make the following assumption:

Assumption 3. *The types $(\alpha_i, \boldsymbol{\xi}_i)$ are drawn from independent, absolutely continuous distributions F_α and $F_{\boldsymbol{\xi}}$. Each component k in $\boldsymbol{\xi}_i$ is independently drawn from the marginal distribution $F_{\boldsymbol{\xi}_k}$ s.t. $F_{\boldsymbol{\xi}} = F_{\boldsymbol{\xi}_1} \cdot \ldots \cdot F_{\boldsymbol{\xi}_m}$, where m is the dimension of $\boldsymbol{\xi}_i$. G is pairwise independent of F_α and $F_{\boldsymbol{\xi}}$.*

User i's expected utility μ_i, given the realized types α_i and $\boldsymbol{\xi}_i$, allocation $d_i = 1$, and reward r_i, is obtained by taking the expectation of (2.2) with respect to the random variable $\bar{x}_i \sim G_{\boldsymbol{\xi}_i}$:

$$\mu_i(d_i = 1, r_i) = \int_{\mathbb{R}_+} u_i(\alpha_i, r_i, x)\ dG_{\boldsymbol{\xi}_i}(x), \tag{2.7}$$

which is strictly monotonically increasing in reward r_i, cf. (2.2). Letting $\mathcal{G}$ denote the CDF of G, (2.7) for $r_i = 0$ becomes

$$\mu_i(d_i = 1, r_i = 0) = q_i \left[\hat{x}_i(1 - \mathcal{G}(\hat{x}_i)) - \int_{\hat{x}_i}^{\infty} x\ dG_{\boldsymbol{\xi}_i}(x)\right]$$

which is negative. Hence, there is a unique $\tilde{r}_i$ such that $\mu_i(d_i = 1, \tilde{r}_i) = 0$, i.e. the unique threshold reward level for which user i's expected utility is zero. We approximate $\tilde{r}_i$ with Newton's method, exploiting the fact μ_i is monotonically increasing in r_i. Due to the same property, any reward $r_i \geq \tilde{r}_i$ fulfills the IR constraint as $\mu_i(d_i = 0) = 0$ (cf. Eq. 2.2).

2.3.3 Mechanism for Demand Response

We now present the Demand Response Mechanism:

1. Each user announces her private type $\mathbf{z}_i \in \boldsymbol{\Theta}_i$ to the DRP. We will later show that this mechanism is incentive compatible, so that users report their types truthfully. In the following, we thus let $\mathbf{z}_i = \boldsymbol{\theta}_i$.

2. The DRP calculates the unique $\tilde{r}_i$ for each user based on the reports $\boldsymbol{\theta}_i$ with Newton's method on (2.7).

3. The DRP sorts $\{\tilde{r}_i \mid i \in \mathcal{I}\}$ in ascending order. Call this sorted set $\mathcal{R}$.

4. The DRP implements the social choice $\mathbf{y}$ as follows:

$$j_{\max} = \min_j \left\{ j \in \mathbb{N}_+ \;\middle|\; \sum_{i=1}^{j} \delta_i(\tilde{r}_j|\boldsymbol{\theta}_i) \geq M \right\} \tag{2.8a}$$

$$j(i) = \min_k \left\{ k \in \mathbb{N}_+ \;\middle|\; \sum_{s=1, s\neq i}^{k} \delta_s(\tilde{r}_k|\boldsymbol{\theta}_s) \geq M \right\} \quad \forall\, i \in \{1, \ldots, j_{\max}\} =: \mathcal{T} \tag{2.8b}$$

$$r_i \leftarrow \tilde{r}_{j(i)} \geq \tilde{r}_i \quad \forall\, i \in \mathcal{T} \tag{2.8c}$$

The allocation decision and the reward vector are

$$\mathbf{d} = (1, \ldots, 1, \mathbf{0}_{n-j_{\max}}), \tag{2.9a}$$

$$\mathbf{r} = (\tilde{r}_{j(1)}, \ldots, \tilde{r}_{j(j_{\max})}, \mathbf{0}_{n-j_{\max}}). \tag{2.9b}$$

In the above mechanism, $\delta_i(\tilde{r}_j|\boldsymbol{\theta}_i)$ denotes the expected reduction of user i, given the reward level $\tilde{r}_j$ conditional on truthful reporting $\mathbf{z}_i = \boldsymbol{\theta}_i$, which is computed by taking the expectation on (2.1) and (2.3) with respect to $\boldsymbol{\xi}_i$.

The mechanism first determines the set of targeted users $\mathcal{T}$ by selecting the smallest index $j_{\max} \in \{1, \ldots, n\}$, such that the sum of expected reductions of users 1 through $j_{\max}$, if each user were given the reward $\tilde{r}_{j_{max}}$, exceeds the desired aggregate amount M (2.8a). Notice that since the set $\mathcal{R}$ is sorted in ascending order, $\tilde{r}_{j_{\max}} \geq \tilde{r}_i \;\forall\; i \leq j_{\max}$. Because $\mu_i(d_i = 1, r_i)$ is strictly monotonically increasing in r_i, all targeted users will respond to incentive level $\tilde{r}_{j_{\max}}$.

Next, the reward for each user $i \in \mathcal{T}$ is determined by running the same exact mechanism (2.8a) on $\mathcal{I} \setminus i$, i.e. the set of all users excluding i (2.8b). Denote the user with the largest threshold reward $\tilde{r}_{j(i)}$ in this new set with $j(i)$. This reward level is then assigned to user i (2.8c).

In summary, the first $j_{\max}$ users (2.8a) with the smallest threshold rewards $\tilde{r}_i$ are offered user-specific unit-rewards ((2.8b), (2.8c)). The remaining $n - j_{\max}$ users are not targeted.

Lastly, to ensure that the mechanism returns a valid index $j_{\max}$, we restrict M to the range $\left[0,\ \sum_{i=2}^{n-1} \delta_i(\tilde{r}_{n-1}|\boldsymbol{\theta}_i)\right]$. If M exceeds this range, there are not enough users to achieve expected aggregate reduction M on the given n users.

Theorem 1. *If* $M \in \left[0,\ \sum_{i=2}^{n-1} \delta_i(\tilde{r}_{n-1}|\boldsymbol{\theta}_i)\right]$, *the DR Mechanism terminates. The mechanism fulfills the IR constraint. Truthful reporting, i.e.* $s_i^*(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_i$, *establishes a DSE.*

Since truthful reporting establishes a DSE (Theorem 1), Mechanism I is also IC, due to the revelation principle (Milgrom 2004).

Remark 1. *Due to the fact that* $\{(\alpha_i, \boldsymbol{\xi}_i)\}_{i=1}^n$ *are realizations of continuous random variables, no ties need to be broken in* (2.8a), (2.8b) *and the sorting of the users into* $\mathcal{R}$*, because identical threshold rewards* $\tilde{r}_i = \tilde{r}_j,\ i, j \in \mathcal{I},\ i \neq j$*, only occur with probability zero.*

The presented mechanism runs in $\mathcal{O}(n \log n)$ time, as it takes $\mathcal{O}(n \log n)$ time to create the sorted list $\tilde{R}$ and $\log n$ time to determine the correct index $j_{\max}$ (2.8a) with a binary search on all possible values of $j = 1, \ldots, n$. Once $j_{\max}$ has been found, we have to determine the reward level for each user by running the same mechanism again, which amounts to $\mathcal{O}(n \log n)$. This yields a runtime of $\mathcal{O}(n \log n)$.

Remark 2. *This mechanism is motivated by the classic Vickrey-Clarke-Groves Mechanism (Milgrom 2004), as it allocates an "item" (in our case reward) to the "highest" bidders (in our case lowest threshold reward levels).*

2.3.4 Numerical Example

Table 2.1 lists threshold rewards $\tilde{r}_i$ and reduction functions of 6 hypothetical users in a synthetic user pool. The linearity of $\{\delta_i\}_{i=1}^6$ is assumed for ease of exposition. Let $M = 4.3$.

Pool of Users						
User#	1	2	3	4	5	6
$\tilde{r}_i$	0.5	1.0	1.5	1.8	2.0	2.1
$\delta_i(r_i)$	$1 + r_1$	$2 + \frac{r_2}{2}$	$1 + \frac{r_3}{3}$	$2 + \frac{r_4}{4}$	$1 + \frac{r_5}{2}$	$1 + \frac{r_6}{5}$

Table 2.1: Example User Characteristics

(2.8a) selects $j_{\max} = 2$ such that $\delta_1(\tilde{r}_2) + \delta_2(\tilde{r}_2) = (1 + 1) + (2 + \frac{1}{2} \cdot 1) = 4.5 \geq M$. Thus $\mathcal{T} = \{1, 2\}$. (2.8b) then determines $j(1)$ and $j(2)$ by solving (2.8a) on $\mathcal{T} \setminus 1$ and $\mathcal{T} \setminus 2$, respectively:

- For $i = 1$, $j(1) = 4$ because $\delta_2(\tilde{r}_4) + \delta_3(\tilde{r}_4) + \delta_4(\tilde{r}_4) = (2 + 1.8/2) + (1 + 1.8/3) + (2 + 1.8/4) = 6.95 \geq M$. Indeed, $j(1) \neq 3$ because $\delta_2(\tilde{r}_3) + \delta_3(\tilde{r}_3) = (2 + 1.5/2) + (1 + 1.5/3) = 4.25 < M$.
- For $i = 2$, $j(2) = 4$ because $\delta_1(\tilde{r}_4) + \delta_3(\tilde{r}_4) + \delta_4(\tilde{r}_4) = (1 + 1.8) + (1 + 1.8/3) + (2 + 1.8/4) = 6.85 \geq M$. Indeed, $j(2) \neq 3$ because $\delta_1(\tilde{r}_3) + \delta_3(\tilde{r}_3) = (1 + 1.5) + (1 + 1.5/3) = 4 < M$

User 1 and 2's rewards therefore are $\tilde{r}_4$, see (2.8c).

2.4 Effect of Baseline "Gaming"

By expanding user i's reduction of consumption (2.1),

$$\delta_i = (\hat{x}_i - \bar{x}_i) + \bar{x}_i(1 - e^{-\alpha_i r_i}) =: \delta_i^{\mathrm{BL}} + \delta_i^r, \tag{2.10}$$

it becomes clear that the measured reduction δ_i of user i is comprised of two components: δ_i^{BL}, which captures the difference between the baseline $\hat{x}_i$ and the base consumption (i.e. the consumption with no reward), and the actual reduction δ_i^r due to the elasticity of user i in response to the reward level r_i. δ_i^{BL} is a "virtual reduction", which, if positive (negative), represents the amount of falsely measured reduction (increase). From an economic perspective, $\delta_i^{\mathrm{BL}} > 0$ results in falsely allocated credit from the utility to the DRP as well as from the DRP to users i. On the contrary, $\delta_i^{\mathrm{BL}} < 0$ is synonymous with a misallocated monetary transfer from user i to the utility as well as from the utility to the DRP proportional to the amount of $|\delta_i^{\mathrm{BL}}|$. To diminish the effect of virtual reduction, the baseline estimates should become as precise as possible. We make the following assumption:

Assumption 4. *The random variables α_i and $\boldsymbol{\xi}_i$ for different points in time are independent.*

Assumption 4 excludes the possibility of baseline manipulation (Campaigne, Balandat, and Ratliff 2016), which captures the fact that users can inflate or deflate their baseline, given the knowledge of future DR events, in order to increase their calculated reduction δ_i (2.1). For example, a user can increase her expected utility (2.2) for a DR event by consciously over-consuming prior to the DR event so as to increase the baseline $\hat{x}_i$, which results in a higher payment $r_i \cdot [\hat{x}_i - x_i]_+$, despite having a zero actual reduction δ_i^r. However, as DR events are difficult to forecast, the mild assumption that users do not consciously manipulate their baseline justifies Assumption 4, that is, users consume independently of the past and the future.

As a result, averaging 10 recent observations for weekdays (or 4 for weekends and holidays), excluding hours of past DR events, results in an unbiased estimate of the mean consumption x_i, but with considerable variance around x_i. From a theoretical perspective, the baseline estimate approaches zero variance as the number of previous observations to estimate $\hat{x}_i$ goes to infinity, due to the Central Limit Theorem and Assumption 4. In the next Section, we simulate the effect of more precise baseline estimates on the quantity of virtual reductions δ_i^{BL}.

As the analysis of the economic implications of this virtual baseline reduction component is outside the scope of this thesis, the reader is referred to (Borenstein, Jaske, and Rosenfeld 2002a), which explicitly characterizes the magnitude of marginal competitive rents in California's wholesale electricity market, and (Zhou, Balandat, and Tomlin 2016b; Zhou, Balandat, and Tomlin 2016a), where the authors suggest alternative baselining methodologies based on Machine Learning, which weaken the effect of such virtual reductions.

2.5 Simulations

In this section, we simulate the presented mechanism and the effect of virtual reductions stemming from imperfect baseline predictions. We utilize hourly smart meter data from 1,000 residential customers serviced by the three largest utilities in California (Pacific Gas & Electric, San Diego Gas & Energy, and Southern California & Edison).

2.5.1 Approximation of Base Consumption

Figure 2.3 shows the distribution of the hourly base consumptions between 5-6 pm in the absence of DR events of a selected user. The restriction to 5-6 pm is arbitrarily chosen. For a more thorough analysis, we would have to analyze all 24 hours of the day separately.

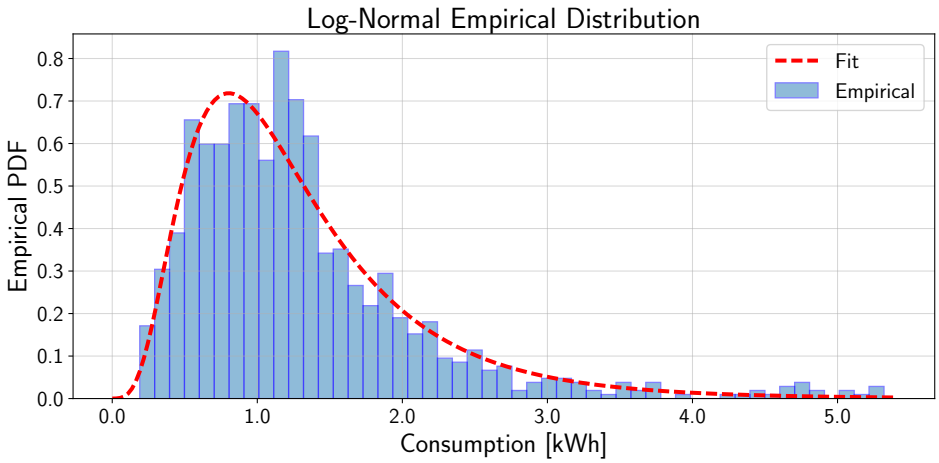

Figure 2.3: Lognormal Consumption Distribution Fit for Selected User, 5-6 pm

It is found that the base consumption $\bar{x}_i$ can be approximated with a log-normal distribution, whose density

$$\mathcal{N}(\log x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x-\ell)-\mu)^2}{2\sigma^2}\right) \tag{2.11}$$

is fully parameterized by the shape $\sigma > 0$, scale $s = e^{\mu} > 0$, and location parameter ℓ. As (2.11) has support on (ℓ, ∞), the location ℓ denotes the lower bound on the support of the base consumption distribution.

Fitting a log-normal distribution to the hourly consumptions between 5-6 pm across all users yields a distribution of the compound statistics $\boldsymbol{\xi}_i = (\sigma,\ s,\ \ell)$, given below:

$$\bar{x}_i \sim \text{Lognormal}(\sigma, s, \ell) \qquad \sigma \sim \mathcal{N}(\mu_n, \sigma_n)$$
$$s \sim \text{Cauchy}(\ell_c, s_c) \qquad \ell \sim \text{Exponential}(\lambda_e)$$

That is, the shape parameter σ is best approximated with a Gaussian distribution $\mathcal{N}(\mu_n, \sigma_n)$, the location ℓ by a Cauchy distribution parameterized by location ℓ_c and scale parameter s_c, and the scale parameter s by an exponential distribution with parameter λ_e. Figure 2.4 shows the distribution of these compound statistics across all 1,000 users.

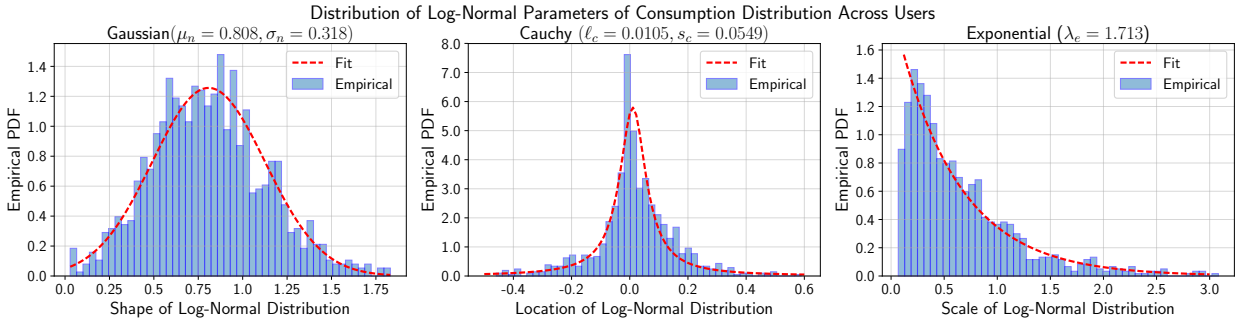

Figure 2.4: Compound Statistics for Lognormal Consumption Distribution. Left: Shape, Middle: Location, Right: Scale

2.5.2 Performance of DR Mechanism

We compare the DR Mechanism (2.8a)-(2.8c) to the hypothetical case of an omniscient DRP, which knows $\{(\alpha_i, \boldsymbol{\xi}_i)\}_{i=1}^n$. Despite this being an unrealistic scenario, it provides a near-optimal approximation of the minimum payment disbursed to the users necessary to elicit a target reduction of M. Given the sorted list $\mathcal{R}$ of user-specific threshold rewards, the omniscient DRP implements the social choice $\mathbf{y}^o = (\mathbf{d}^o, \mathbf{r}^o)$ as follows:

$$j^o = \min_j \left\{ j \in \mathbb{N}_+ \middle| \sum_{i=1}^{j} \delta_i(\tilde{r}_i) \geq M \right\} \tag{2.12a}$$

$$\mathcal{T}^o = \{1, \ldots, j^o\} \tag{2.12b}$$

$$r_i^o = \tilde{r}_i \quad \forall\, i \in \mathcal{T}^o \tag{2.12c}$$

$$\mathbf{d}^o = (1, \ldots, 1, \mathbf{0}_{n-j^o}) \tag{2.12d}$$

That is, the DRP determines the smallest index j^o to obtain the desired expected aggregate reduction M (2.12a) where each user $\{1, \ldots, j^o\}$ is given their individual threshold reward $\tilde{r}_i$ (2.12c). These are the targeted users (2.12b), (2.12d).

Due to $\{(\alpha_i, \boldsymbol{\xi}_i)\}_{i=1}^n$ being publicly known, users are unable to extract information rent from the DRP, which is the payment to the users required to elicit their private information (Laffont and Martimort 2002). Hence, the DRP can offer targeted users their threshold reward $\tilde{r}_i$, which keeps users at an expected utility (2.7) of zero. To guarantee user participation, the DRP has to offer the reward level $\tilde{r}_i + \varepsilon$ to each user $i \in \mathcal{T}^o$, where ε is an arbitrarily small positive number.

Figure 2.5 compares the DR Mechanism (2.8a)-(2.8c) to the omniscient allocation with respect to the number of targeted users (left) and the total amount of rewards disbursed (right) on $n = 500$ users whose parameters $\boldsymbol{\xi}_i = (\sigma_i, s_i, \ell_i)$ are sampled from the fitted distributions in Figure 2.4. As expected, the omniscient allocation is more economical at eliciting a particular aggregate reduction target M due to the lack of private user information, namely about 45% better than the DR mechanism. However, it needs to target more customers as each customer in the omniscient case receives a smaller reward level than in the DR mechanism.

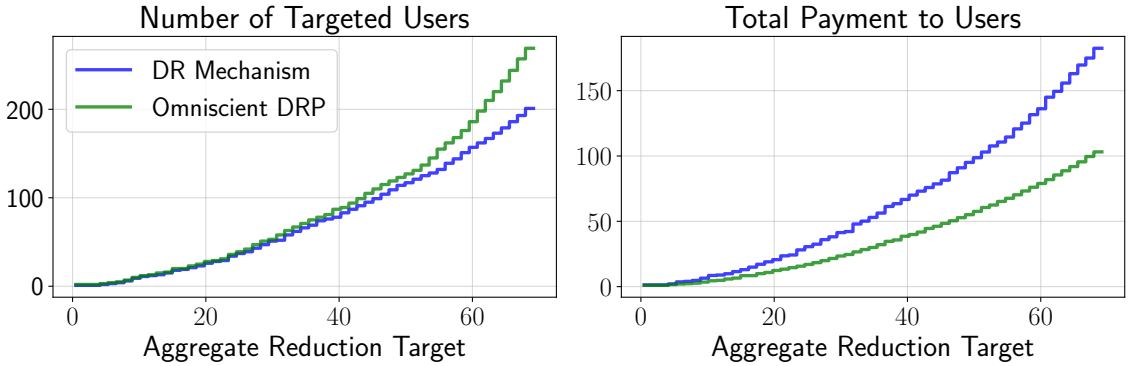

Figure 2.5: Number of Targeted Users and Total Payment to Users for DR Mechanism (blue) vs. Omniscient Allocation (green), $n = 500$, $q = 5.0$, $\alpha_i \sim \text{unif}[0.05, 0.06]$.

2.5.3 Virtual Reductions

Figure 2.6 shows the total reduction $\sum_{i\in\mathcal{T}} \delta_i$ of all targeted users and its components $\sum_{i\in\mathcal{T}} \delta_i^{\text{BL}}$ and $\sum_{i\in\mathcal{T}} \delta_i^{r}$ as a function of M for $n = 500$ users, $q = 5$, and elasticities $\{\alpha_i\}_{i=1}^n$ drawn from a uniform distribution with support $[0.05, 0.06]$. The baseline computed with a particular number x of previous days taken into consideration is calculated as the mean of x randomly drawn samples from the empirical consumption distribution (2.11).

As can be seen from Figure 2.6, almost the entire reduction is attributed to the baseline component $\sum_{i\in\mathcal{T}} \delta_i^{\text{BL}}$ for small M. With larger values of M, the contribution of $\sum_{i\in\mathcal{T}} \delta_i^{\text{BL}}$ decreases marginally and finally starts decreasing. This can be explained by the fact that sorting users in $\mathcal{R}$ tends to put users with the highest δ_i^{BL} towards the start of the array, while those with the lowest (and negative) δ_i^{BL} bunch up at the end of $\mathcal{R}$. Consequently, as more users are assigned to $\mathcal{T}$, the sum of baseline reductions decreases. The actual reduction $\sum_{i\in\mathcal{T}} \delta_i^{r}$ increases exponentially with the number of users targeted, because as more users are assigned to $\mathcal{T}$, the per-unit reward levels also increase, which results in a superlinear growth of $\sum_{i\in\mathcal{T}} \delta_i^{r}$.

For increasing numbers of baseline averaging components, that is, the number of previous days to calculate the baseline, the variance of the baseline estimate $\bar{x}_i - \hat{x}_i$ decreases, and

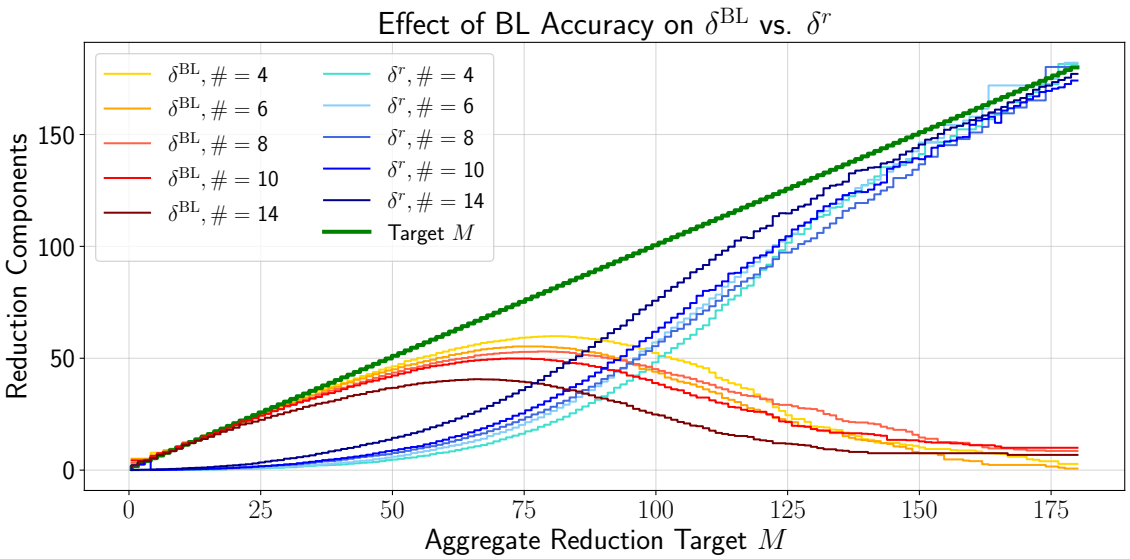

Figure 2.6: Composition of Target Aggregate Reduction M for varying Baselines. Red: $\sum_{i\in\mathcal{T}} \delta_i^{\mathrm{BL}}$. Blue: $\sum_{i\in\mathcal{T}} \delta_i^r$. Parameters: $n = 500, q = 5.0, \alpha_i \sim \text{unif } [0.05, 0.06]$

so the virtual reductions decrease. For the limiting case of a perfect baseline, the virtual reductions are zero.

Finally, Figure 2.7 depicts the total amount of payments the DRP has to make to the users for varying baseline accuracies in the range $M \in [0, 100]$, where virtual payments have the largest effect (see Figure 2.6). For more inaccurate baselines (fewer number of averaging days), the DRP has to pay users less as it can exploit the virtual reduction component.

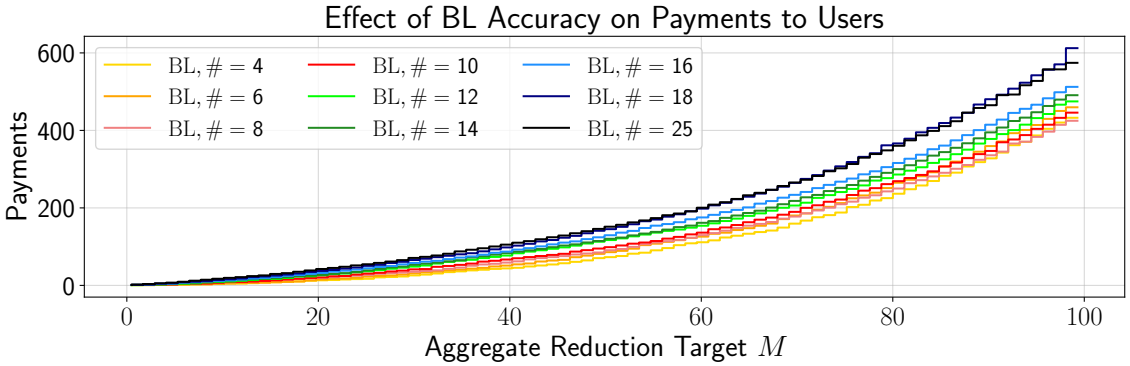

Figure 2.7: Payments to Users to Elicit M for varying Baseline Accuracies.

2.6 Conclusion

We modeled Residential Demand Response with a Mechanism Design framework, in which a Demand Response Provider asks a subset of its customers under contract to reduce their electricity consumption temporarily in exchange for a monetary reward. Each user's consumption in response to a per-unit reduction incentive is modeled as a logarithmic demand curve where the intercept and the slope are private information of users. While each user has a fixed slope, the user-specific intercept, which corresponds to the consumption given no incentive, is modeled as a realization of a compound random variable, capturing the fact that users often do not consume electricity in a profit-maximizing fashion, but rather are following habits, and hence have no explicit utility function. To make an informed choice about the magnitude of reductions in response to incentives to achieve an a-priori defined aggregate reduction target M, the Demand Response Provider asks for residential customers' bids to elicit their private information. Reductions are measured against a counterfactual estimate of the consumption in the hypothetical case of no DR event, which in this chapter is the "10-in-10"-baseline employed by the California Independent System Operator. Since this baseline is plagued by high variance, the Demand Response Provider can exploit "virtual reductions" emanating from high baseline estimates, which are false-positive reductions despite the users not having reduced, but whose role diminishes as the baseline becomes more precise. The Demand Response mechanism is validated on hourly smart meter data of residential customers in California.

Our analysis is an initial step towards quantifying economic implications of Demand Response on a residential level. While we approximated users' base demand (i.e. in the absence of incentives) reasonably well with existing smart meter data, the price elasticity of users in response to incentives is unknown, a fact that is complicated by the fundamental problem of causal inference. Thus, to further validate our analysis on real data, credible parameters for users' slope of the demand curve would be necessary.

Lastly, extending the single period analysis employed in this chapter towards a dynamical problem, which allows for baseline manipulation of users, is a logical next step. Comparing the "10-in-10"-baseline to improved baseline estimates obtained with Machine Learning techniques, which exploit serial correlation of consumption time series and are able to achieve higher predictive accuracy, would shed further light on the economics of Residential Demand Response, as the accuracy of prediction correlates negatively with the amount of virtual reductions.

Chapter 3

Analysis of Peer Effects on Electricity Consumption

3.1 Introduction

While the previous chapter was concerned with the interactions between a principal (Demand Response Provider), this chapter seeks to shift the focus towards the interaction between users.

Energy efficiency programs have emerged as a viable resource to yield economic benefits to utility systems and to reduce the amount of greenhouse gas emissions. Demand-side management aims to modify consumer demand through financial incentive schemes and to induce behavioral changes through education. Specifically, users are offered rewards to conserve energy during peak hours or to shift usage to off-peak times. With communications and information technology constantly improving, which are characteristic elements of today's smart grid, demand-side management technologies are becoming increasingly feasible.

Previous academic work by psychologists, political scientists, and behavioral economists has found that social comparisons can have a significant impact on people's behavior, exploiting the willingness of individuals to conform to a standard, receive social acclaim, or simply the belief that other people's choices are informative in the presence of limited or imperfect information (Akerlof and Kranton 2000; Mani, Rahwan, and Pentland 2013).

Motivated by this line of academic work and the pressing need to improve energy efficiency, various companies and groups, for instance OPOWER, have conducted randomized control trials to investigate the impact of peer effects on energy consumption of residential households by sending out quarterly energy reports (so called *Home Energy Reports*) to users with a comparison of their usage to their closest neighbors (Ayres, Raseman, and Shih 2009). While all experiments unanimously found an average reduction among the highest consuming users of around 1-2% (Allcott 2011), ambiguous results were found among low consumers, with one study reporting a "boomerang effect", that is, an increase of energy demand among the most efficient households.

Network effects in social networks and platforms often exhibit positive externalities, capturing the intuitive fact that an increased amount of platform activity promotes a local increase in platform activity. From a game-theoretic perspective, it is known that an analysis of games under such strategic complements admits well-behaved solutions if utility functions are supermodular with parameters drawn from a lattice (Topkis 1998; Zhou 1994). Examples for such games can be found in modeling technology adoption, human capital decisions, and criminal and social networks (Calvó-Armengol, Patacchini, and Zenou 2009). The opposite effect, that is, in games of strategic substitutes where an increased amount of activity leads to local reductions of activity, is observed in information sharing and the provision of public goods (Bramoullé and Kranton 2007). However, since utility functions in this setting tend to lose the feature of supermodularity, finding equilibria is an inherently hard problem (Jackson and Zenou 2014), and so these settings have been significantly less studied.

In an attempt to characterize the most influential players in a network, (Ballester, Calvó-Armengol, and Zenou 2006) develops a quadratic model with continuous action spaces, a parameterization which we employ in this thesis. Other research directions aiming at understanding the impact of network effects on social phenomena include diffusion models for the spread of information with the goal of influence maximization (Kempe, Kleinberg, and Tardos 2003), repeated games to learn user interactions over time (Acemoglu et al. 2011), or the analysis of systemic risk and stability (Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015) in financial networks. The problem of profit maximization of a monopolist selling a divisible good, which is closely related with our work, has been investigated in (Candogan, Bimpikis, and Ozdaglar 2012), where the authors assume a constant marginal cost of production. However, to the best of our knowledge, a modeling approach for the impact of peer effects on energy consumption, whose generation typically has quadratic marginal cost, has yet to be formulated.

In this thesis, we propose a two-stage game-theoretic model for the energy consumption of a network of users, serviced by the load-serving entity that is obligated to cover the households' energy demand at all times. We analytically solve for the equilibria of this game under full information of the network structure and users' parameters to characterize the influence of peer effects on aggregate consumption and utility profit, for both the case of perfect price discrimination and a single price valid for all users. For the case of incomplete information, we obtain approximations of the utility's profit, user consumptions, and the optimal pricing scheme. Further, we analyze the profit-maximization problem by selecting the best subset of users to be exposed to peer effects, and present a heuristic solution to this NP-hard selection problem. Lastly, we provide theoretical statements on the properties of users which ensure that the consumption under peer effects is reduced.

The remainder of this chapter is organized as follows: Section 3.2 presents the two-stage game-theoretic model between the utility and the network of consumers and derives consumption and price equilibria. Based on this model, Section 3.3 presents various theorems on the reduction of consumption in response to the peer effect as well as on the effect of uncertainty of the network structure on the optimal profit. Section 3.4 compares the utility's profit under the pricing schemes derived in Section 3.2. Next, the challenge of maximizing the

utility's profit by imposing a binary constraint on the number of users exposed to peer effects is formulated and solved with a heuristic approach in Section 3.5. Section 3.6 concludes this chapter. All proofs are relegated to Appendix A.2.

3.2 Game-Theoretic Model

3.2.1 Players

Define the set of consumers as $\mathcal{I} = \{1, \dots, n\}$. Let $W \in \mathbb{R}^{n \times n}$ define the interaction matrix which describes the network links and strengths between users. More precisely, let $w_{ij} \in [0, 1]$ denote the strength of influence of user j on i. We assume $w_{ii} = 0 \; \forall \; i \in \mathcal{I}$ and normalize the row sums, $\sum_{j \in \mathcal{I}} w_{ij} = 1 \; \forall \; i \in \mathcal{I}$. Each element $w_{ij} > 0$ in W corresponds to a directed edge from agent j to agent i, that is, the adjacency matrix G of the resulting directed graph is the transpose of W. Each user i derives a utility $u_i \in \mathbb{R}$ from consuming x_i units of electricity as follows:

$$u_i = a_i x_i - b_i x_i^2 - p_i x_i + \gamma_i x_i \left(\sum_{j \in \mathcal{I}} w_{ij} x_j - x_i \right). \tag{3.1}$$

In (3.1), a_i and b_i denote user-specific parameters to describe the concave and increasing direct utility from consuming x_i units of electricity, and p_i denotes the unit price set by the utility. The last term captures the strategic complementarity between user i and its neighbors. It is positive if user i consumes less than the average of its neighbors, and vice versa. The difference between the average consumption and the user consumption is scaled by a proportionality constant γ_i and the consumption level x_i.

Since each user consumes x_i units of electricity at unit price p_i, the utility's profit reads as follows:

$$\Pi = \sum_{i \in \mathcal{I}} p_i x_i - c_i x_i^2, \tag{3.2}$$

where the marginal cost of production $2c_i x_i$ is assumed to be linear in the production quantity x_i, which is a standard and often made assumption. For expositional ease, we further assume that the utility generates electricity itself and does not procure it from the wholesale electricity market. Relaxing this assumption would introduce uncertainty in wholesale prices, a problem which is outside the scope of this thesis.

3.2.2 Two-Stage Game

To model the hierarchy between the utility, which acts as a monopolist that has the power to set prices, and the users, we formulate a two-stage game as follows:

1. The utility determines the optimal price $\mathbf{p}^*$ so as to maximize its profit by taking into account users' consumption decisions as a function of any particular price vector $\mathbf{p}$, that is,
$$\mathbf{p}^* = \arg\max_{\mathbf{p}\geq\mathbf{0}} \sum_{i\in\mathcal{I}} p_i x_i(p_i) - c_i x_i^2(p_i) \tag{3.3}$$

2. Each agent observes the price p_i^* and $\mathbf{x}_{-i}$ and consumes x_i^* units of electricity so as to maximize her utility, that is, $x_i^* = \arg\max_{x_i\geq 0} u_i(x_i, \mathbf{x}_{-i}, \gamma_i, W)$.

We will solve this two-stage game by finding a subgame perfect equilibrium for the cases of perfect price discrimination and a single price for all users. We also differentiate between the full-information case where the utility has knowledge about all $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$, and the case in which only their expectations $\mathbb{E}[a]$ and $\mathbb{E}[b]$ are known.

3.2.3 Subgame-Perfect Equilibrium

Assumption 5. *$a_i > p_i$ and $b_i > \gamma_i \;\forall\; i \in \mathcal{I}$.*

Theorem 2. *Given the price vector $\mathbf{p}$ and consumption vector $\mathbf{x}_{-i}$, the utility maximizing response of user i is*
$$x_i^* = \frac{a_i - p_i + \gamma_i \sum_{j\in\mathcal{I}} w_{ij} x_j}{2(b_i + \gamma_i)}. \tag{3.4}$$
Further, $\{x_1^*, \ldots, x_n^*\}$ constitute a unique Nash Equilibrium of the second stage game.

Recall that $w_{ii} = 0 \;\forall\; i \in \mathcal{I}$, which allows the right hand side of (3.4) to depend on $\mathbf{x}_{-i}$ only. Assumption 5 is necessary to ensure that (3.4) is indeed a maximum attained at a non-negative value. With the definitions $B := \operatorname{diag}(2b_1, \ldots, 2b_n)$ and $\Gamma := \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$, (3.4) can be rewritten as
$$\mathbf{x}^* = (B + 2\Gamma - \Gamma W)^{-1}(\mathbf{a} - \mathbf{p}). \tag{3.5}$$

Definition 2 (Katz-Bonacich Centrality (Katz 1953; Bonacich 1987))**.** *Given the adjacency matrix G, the weight vector $\mathbf{w}$, and the scalar $0 \leq \alpha < 1/\rho(G)$, where $\rho(G)$ denotes the spectral radius of G, the weighted Katz-Bonacich Centrality is defined as*
$$\mathcal{K}_{\mathbf{w}}(G, \alpha) = (I - \alpha G)^{-1}\mathbf{w} = \sum_{k=0}^{\infty} (\alpha G)^k \mathbf{w}. \tag{3.6}$$
The centrality of a particular node i can be interpreted as the sum of total number of walks from i to its neighbors discounted exponentially by α and weighted by w_i.

For the special case $\gamma_1 = \ldots = \gamma_n = \gamma$, and noting that $G = W^\top$, (3.5) can be rewritten in terms of the weighted Katz-Bonacich Centrality:
$$\begin{aligned}\mathbf{x}^* &= (B + 2\gamma I)^{-1}\left(I - \gamma W^\top (B + 2\gamma I)^{-1}\right)^{-1}(\mathbf{a} - \mathbf{p}) \\ &= (B + 2\gamma I)^{-1}\mathcal{K}_{\mathbf{a}-\mathbf{p}}(W^\top (B + 2\gamma I)^{-1}, \gamma)\end{aligned}$$

We note that $(B + 2\Gamma - \Gamma W)$ is strictly diagonally dominant for all $\gamma \geq 0$, with positive diagonal entries. The Gershgorin Circle Theorem then states that all its eigenvalues are strictly positive, from which invertibility follows.

We first focus on the full information case and present the equilibria in Theorems 3 and 4. Let $C = \text{diag}(c_1, \ldots, c_n)$.

Theorem 3. *Under perfect price discrimination, the profit-maximizing solution* $\mathbf{p}^*$ *to the first stage game is*

$$\mathbf{p}^* = \underbrace{\frac{\mathbf{a}}{2}}_{(1)} + \underbrace{CZ\frac{\mathbf{a}}{2}}_{(2)} - \underbrace{W^\top \Gamma Z\frac{\mathbf{a}}{4}}_{(3)} + \underbrace{\Gamma W Z\frac{\mathbf{a}}{4}}_{(4)}, \tag{3.7}$$

$$Z = \left[2\Gamma + B + C - \left(\frac{W^\top\Gamma}{2} + \frac{\Gamma W}{2}\right)\right]^{-1}.$$

The four components are interpreted as follows:

1. *A constant term* $a_i/2$*, c.f.* a_i *in* (3.1)*,*
2. *An additional cost that correlates with cost* c_i*,*
3. *An incentive for strongly influential users* $W^\top\Gamma$*,*
4. *An additional cost for strongly influenced users* ΓW*.*

The optimal consumption under this policy is

$$\mathbf{x}^* = \left(C + B + 2\Gamma - \frac{W^\top\Gamma}{2} - \frac{\Gamma W}{2}\right)^{-1}\frac{\mathbf{a}}{2}. \tag{3.8}$$

For the special case of symmetric networks, i.e. $W = W^\top$*, the optimal profit* Π^* *becomes*

$$\Pi^* = \frac{1}{4}\mathbf{a}^\top(C + B + 2\Gamma - \Gamma W)^{-1}\mathbf{a}. \tag{3.9}$$

Theorem 4. *Under complete information, i.e. the utility knows* a_i *and* b_i $\forall$ $i \in \mathcal{I}$*, the profit-maximizing single price* p_u^* *is*

$$p_u^* = \left[1 - \frac{\mathbf{1}^\top A^{-1}\mathbf{1}}{2\cdot\mathbf{1}^\top\left(A^{-1} + A^{-1}CA^{-1}\right)\mathbf{1}}\right]\bar{a} \tag{3.10}$$

and the consumption equilibrium writes

$$\mathbf{x}^* = A^{-1}\left[\mathbf{a} - \left(1 - \frac{\mathbf{1}^\top A^{-1}\mathbf{1}}{2\cdot\mathbf{1}^\top\left(A^{-1} + A^{-1}CA^{-1}\right)\mathbf{1}}\right)\bar{a}\mathbf{1}\right], \tag{3.11}$$

where $A = B + 2\Gamma - \Gamma W$ *and* $\bar{a} = \sum_{i=1}^n a_i/n$*.*

Lemma 1. *For symmetric networks, i.e.* $W = W^\top$*, the single profit-maximizing price* (3.10) *and its corresponding consumption* (3.11) *simplify to*

$$p_u^* = \frac{1}{n}\sum_{i=1}^{n} p_i^* \tag{3.12a}$$

$$\mathbf{x}_u^* = (B + 2\Gamma - \Gamma W)^{-1}(\mathbf{a} - \bar{a}\mathbf{1}) + (C + B + 2\Gamma - \Gamma W)^{-1}\frac{\bar{a}}{2}\mathbf{1}. \tag{3.12b}$$

By construction of the optimal prices and consumptions, the optimal profit under a single price is less than under perfect price discrimination, that is, $\Pi_u^* \geq \Pi^*$.

Next, for the incomplete information scenario and additional assumptions $W = W^\top$ and $C = cI$, the utility can approximate the profit-maximizing price as in Theorem 5.

Theorem 5. *In the case of incomplete information, that is, only the expectations of* $\{a_i\}_{i=1}^n$ *and* $\{b_i\}_{i=1}^n$ *are known and denoted with* $\mathbb{E}[a]$ *and* $\mathbb{E}[b]$*, the optimal single profit-maximizing price* $\tilde{p}_u^*$ *and the expected corresponding consumption equilibrium* $\mathbb{E}[\tilde{x}_i]$ *are bounded below by*

$$\tilde{p}_u^* \geq \frac{\mathbb{E}[a]}{2}\left[1 + \frac{c}{n}\mathbf{1}^\top\left[2\Gamma + (2\mathbb{E}[b] + c)I - \Gamma W\right]^{-1}\mathbf{1}\right], \tag{3.13a}$$

$$\mathbb{E}[\tilde{x}_i] \geq \frac{\mathbb{E}[a] - \tilde{p}_{u,LB}^*}{n} \cdot \mathbf{1}^\top(2\Gamma + 2\mathbb{E}[b]I - \Gamma W)^{-1}\mathbf{1}. \tag{3.13b}$$

where $\tilde{p}_{u,LB}^*$ *denotes the lower bound on the single profit-maximizing price* $\tilde{p}_u^*$ (3.13a)*.*

Theorem 6 (Profit Maximizing Price without Peer Effects)**.** *In the case of incomplete information and in the absence of any peer effects, the single profit-maximizing price* $\hat{p}^*$ *and the expected user consumption* $\mathbb{E}[\hat{x}_i]$ *are*

$$\hat{p}^* = \frac{\mathbb{E}[b] + c}{2\mathbb{E}[b] + c}\mathbb{E}[a], \tag{3.14a}$$

$$\mathbb{E}[\hat{x}_i^*] = \frac{\mathbb{E}[a]}{2(2\mathbb{E}[b] + c)} \quad \forall\ i \in \mathcal{I}. \tag{3.14b}$$

3.3 Theoretical Statements

We next seek to analyze under what conditions the aggregate consumption across all users is less than in the absence of peer effects, which is a desirable goal from the energy efficiency perspective.

Theorem 7. *If* $a_i =: a$*,* $b_i =: b$*, and* $\gamma_i =: \gamma\ \forall\ i \in \mathcal{I}$*, and Assumption 5 holds, then* x_i^* (3.4) *is strictly monotonically decreasing in* γ*, independent of the network topology* W*.*

Theorem 7 is interesting because identical consumers will reduce their optimal consumption compared to the case of no peer effects, even though $x_i^* = x_j^*\ \forall\ i, j \in \mathcal{I}$ and hence the peer effect term $\gamma_i x_i\left(\sum_{j\in\mathcal{I}} w_{ij}x_j - x_i\right)$ is zero.

Theorem 8 (**Influence of High Consumer**). *Given that* $w_{ij} = \left(\sum_{j\in\mathcal{I}} 1_{w_{ij}>0}\right)^{-1} \ \forall\ i \in \mathcal{I}$, *that is, all connections are of equal weight, and* $b_i =: b$ *and* $\gamma_i =: \gamma\ \forall\ i \in \mathcal{I}$. *Define the set of users* $\mathcal{N} := \{i \in \mathcal{I} \setminus j\}$ *with the characteristic* $a_i - p_i =: \alpha\ \forall\ i \in \mathcal{N}$. *Further, let* j *be a "high consumer", that is,* $a_j - p_j =: \bar{\alpha} > n\alpha$. *Denote the set of all neighbors of* j *as* $\mathcal{C}_j := \{i \in \mathcal{N} \mid w_{ij} > 0\}$. *Then, independent of the network topology, for all users* $i \in \mathcal{C}_j$, x_i^* *is increasing for small enough values of* γ *whereas* x_j^* *is strictly monotonically decreasing in* γ.

Let m_i denote the number of neighbors of consumer i. Theorem 8 can be restated as in Lemma 2.

Lemma 2. $x_i^*,\ i \in \mathcal{C}_j$ *is increasing for small enough values of* γ *if* $\bar{\alpha} \geq m_j + 1$. *Equivalently, if* $\bar{\alpha} = k\alpha,\ k \in \mathbb{N}$, *only the subset* $\{i \in \mathcal{C}_j \mid m_i \leq k-1\}$, *i.e. the set of users with fewer than* $k-1$ *neighbors, shows an initial increase in consumption as a function of* γ.

Theorem 8 and Lemma 2 describe conditions on the average consumption of any particular user's neighbors to observe a "boomerang effect", given there is a unique "high" consumer among a pool of users of identical characteristics.

Theorem 9 (**Targeted Peer Effects**). *For a general setting of* $n \geq 2$ *users with non-identical parameters* a_i, b_i *and a fixed price* p *among all users, exposing exactly two connected users to the peer effect, w.l.o.g. referred to as users "1" and "2", reduces the sum of their consumptions under the following conditions:*

$$b_1 \leq \frac{(a_1 - p)\left[4(b_2 + \gamma) - \gamma w_{12} w_{21}\right]}{4(b_2 + \gamma)\sum_{j=3}^{n} w_{1j}x_j + 2w_{12}\left(a_2 - p + \gamma \sum_{j=3}^{n} w_{2j}x_j\right)} \tag{3.15a}$$

$$b_2 \leq \frac{(a_2 - p)\left[4(b_1 + \gamma) - \gamma w_{12} w_{21}\right]}{4(b_1 + \gamma)\sum_{j=3}^{n} w_{2j}x_j + 2w_{21}\left(a_1 - p + \gamma \sum_{j=3}^{n} w_{1j}x_j\right)} \tag{3.15b}$$

where $x_j,\ j \in \{3, \ldots, n\}$ *is given by* $x_j = (a_j - p)/(2b_j)$. *For the special case of* $n = 2$, *this condition reads*

$$b_1 \leq \frac{(a_1 - p)\,(4b_2 + 3\gamma)}{2(a_2 - p)} \quad \textit{and} \quad b_2 \leq \frac{(a_2 - p)\,(4b_1 + 3\gamma)}{2(a_1 - p)}$$

Theorem 9 states that if two connected users both receive notifications of their neighbors' consumption, the sum of their consumptions decreases as long as they are not "too different" from each other and their neighbors. Thus, the total consumption of a network of users correlates negatively with the number of users given the treatment. Analogous bounds can be found for exposing more than two users to the peer effect at the expense of notational ease.

Finally, we investigate the case of incomplete information about the network structure for the case of symmetric networks, i.e. $W = W^\top$. It is assumed that the monopolist only knows an approximation of W, denoted with $\tilde{W}$, where $\tilde{W} = \tilde{W}^\top$. Under perfect price discrimination, the utility can set profit-maximizing prices in the first stage of the game, assuming that users' consumption $\tilde{\mathbf{x}}$ in the first stage is determined according to $\tilde{W}$. The real consumption $\mathbf{x}^*$, however, follows the actual W (which is unknown to the utility). Theorem 10 provides a lower bound on the ratio of the optimal expected profit under network uncertainty to the profit obtainable under perfect network information.

Theorem 10 (**Uncertainty in W**). *Assume that $W = W^\top$ and $\Gamma = \gamma I, \gamma \geq 0$. If the monopolist has access only to the estimate $\tilde{W}$ with $\tilde{W} = \tilde{W}^\top$, then, under perfect price discrimination, the ratio of optimal expected profit $\tilde{\Pi}^*$ to profit Π^* under perfect knowledge of W is bounded below:*

$$\frac{\tilde{\Pi}^*}{\Pi^*} \geq \frac{\lambda_{\min}(C + B + 2\Gamma - \Gamma W)}{\lambda_{\max}(C + B + 2\Gamma - \Gamma W) + \gamma \|W - \tilde{W}\|_2}, \tag{3.16}$$

where $\|\cdot\|_2$ is the Euclidian matrix norm.

For the edge case $\tilde{W} = 0$, we have $\|W\|_2 = 1$ due to the well-known fact that the maximal eigenvalue of an adjacency matrix is the degree of the graph. Due to row normalizations of W, the degree is 1, which corresponds to the eigenvector $\mathbf{1}$ associated with eigenvalue 1. To qualitatively show that the bound (3.16) becomes tighter as $\tilde{W}$ approaches W, observe that $\|W - \tilde{W}\|_2$ corresponds to the largest singular value of $W - \tilde{W}$, which is identical to its spectral radius because $W - \tilde{W}$ is Hermitian. Finally, the Gershgorin Circle Theorem states that every eigenvalue of $W - \tilde{W}$ lies within at least one of the disks that is centered at the origin, each of which has radius $R_i = \sum_{j \neq i} |w_{ij} - \tilde{w}_{ij}|$. As $w_{ij} \to \tilde{w}_{ij}$, $R_i \to 0$.

To illustrate the bound (3.16), let $n = 24$ and $W \in \mathbb{R}^{24 \times 24}$ be the ground truth interaction matrix of 12 randomly chosen, fully connected users, whose parameters a_i, b_i, and $c_i \; \forall \; i \in \mathcal{I}$ are randomly drawn from appropriate distributions. Assuming that the monopolist knows that 12 out of 24 users are fully connected, we iterate through all $\binom{24}{12}$ combinations and calculate $\|\tilde{W} - W\|_2$ and the profit bound (3.16) as a function of the number of correct user assignments, where we take the mean across any particular number of correct assignments. As the number of correct assignments increases, the metric for the mismatch between W and $\tilde{W}$, namely $\|\tilde{W} - W\|_2$ decreases, whereas the profit bound increases, see Figure 3.1.

Theorem 11 (Efficiency). *The consumption equilibrium $\mathbf{x}^*$ (3.8) is inefficient as the social welfare $\mathcal{S}$ attained at (3.8) is suboptimal. Specifically, $x_i^* < x_i^o \; \forall \; i \in \mathcal{I}$, where $\mathbf{x}^o$ denotes the consumption that maximizes social welfare, which reads*

$$\mathbf{x}^o = \left(C + \frac{B}{2} + \Gamma - \frac{W^\top \Gamma}{2} - \frac{\Gamma W}{2} \right)^{-1} \frac{\mathbf{a}}{2}. \tag{3.17}$$

Allocating users per-unit subsidies $s_i = (b_i + \gamma_i) x_i^2 / 2$ (Pigouvian Subsidy) can restore the social optimum.

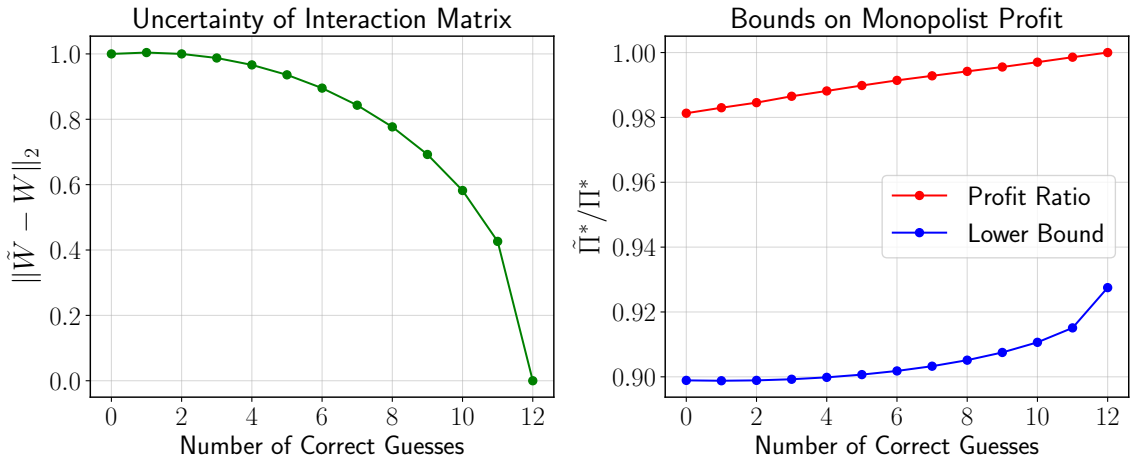

Figure 3.1: $\|W - \tilde{W}\|_2$ for 12 fully connected users embedded in a network of $n = 24$ customers, $\gamma = 0.05$.

3.4 Comparison of Pricing Schemes

3.4.1 Network Topologies

In the remainder of this chapter, we assume users to be connected to each other through one of the basic network topologies displayed in Figure 3.2.

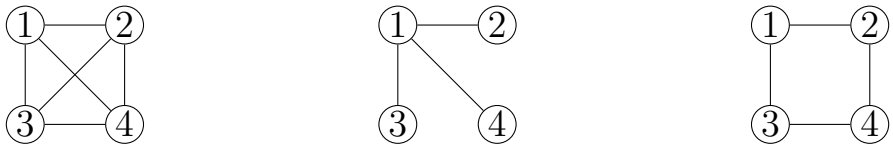

Figure 3.2: Basic network architectures for $n = 4$: Fully connected, star, ring

3.4.2 Simulation

We now simulate the consumption and price equilibria as well as the profit of the monopolist as a function of the network strength parameter γ under the following three pricing scenarios:

- **Case 1**: Monopolist has complete information of **a** and **b** and sets prices with perfect price discrimination (3.7);
- **Case 2**: Monopolist has complete information of **a** and **b** and sets the profit-maximizing single price (3.10);
- **Case 3**: Monopolist has access only to $\mathbb{E}[a]$ and $\mathbb{E}[b]$ and sets the lower bound on the single price (3.13a).

We simulate a network of $n = 10$ fully connected users with a_i and b_i randomly drawn from uniform distributions with support $[8, 12]$ and $[0.75, 1.25]$, respectively. The cost is set to $c_i = 2$ for all users. As the results for the star and ring network are qualitatively similar to the fully connected network, we omit discussions of these cases. The optimal prices for each of the cases (1)-(3) are then calculated, which fixes the users' consumptions and the monopolist's profit. Repeating this process 10,000 times and taking the mean across all iterations yields the characteristics in Figure 3.3.

As expected, the profit under perfect price discrimination (3.7) exceeds the profit obtained with cases (2) and (3), where, somewhat surprisingly, setting the lower bound on the prices (case (3)) does not give up too much profit, compared to case (2). This indicates that the lower bound on the optimal price (3.13a) is "close" to the actual optimum, which is proven by the second subplot, from which it follows that (3.13a) falls short of (3.10) by less than $< 1\%$.

Consequently, the lower price bound (3.13a) results in a higher average user consumption than in case (2), which directly follows from the consumption equilibrium (3.4). The average user consumption under perfect price discrimination is sandwiched between cases (2) and (3).

Lastly, the maximum user consumption for perfect price discrimination is about 30% lower than in cases (2) and (3), which has beneficial side-effects on grid operation. This observation also motivates the heuristic user-selection algorithm presented in the next section.

3.5 Profit Maximization with User Selection

3.5.1 Problem Formulation

We now seek to answer the following question: Given the single, exogenous price p and the parameters $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ sampled from distributions with means $\mathbb{E}[a]$ and $\mathbb{E}[b]$, respectively and are known to the monopolist, which users should be targeted to maximize profit? This situation can arise if the utility is obligated to charge customers at a rate p per unit of electricity and only wants to spend a limited budget on informing users about their peers' behavior. In other words, which best subset of all users should be exposed to the peer effect such that the utility achieves maximum profit under exogenous price p? The profit maximizing problem of the utility thus writes

$$\begin{aligned} \underset{\delta_1,\dots,\delta_n}{\text{maximize}} \quad & \sum_{i=1}^{n} p x_i - c_i x_i^2 \\ \text{subject to} \quad & \mathbf{x} = (B + 2\Delta\Gamma - \Delta\Gamma W)^{-1} (\mathbf{a} - p\mathbf{1}) \\ & \sum_{i=1}^{n} \delta_i = m, \quad \delta_i \in \{0, 1\} \end{aligned} \tag{3.18}$$

where $\Delta = \operatorname{diag}(\delta_1, \dots, \delta_n)$ and $\delta_i = 1$ and $\delta_i = 0$ denote that user i is targeted or non-targeted, respectively. This is an NP-hard Mixed Integer Quadratically Constrained Program

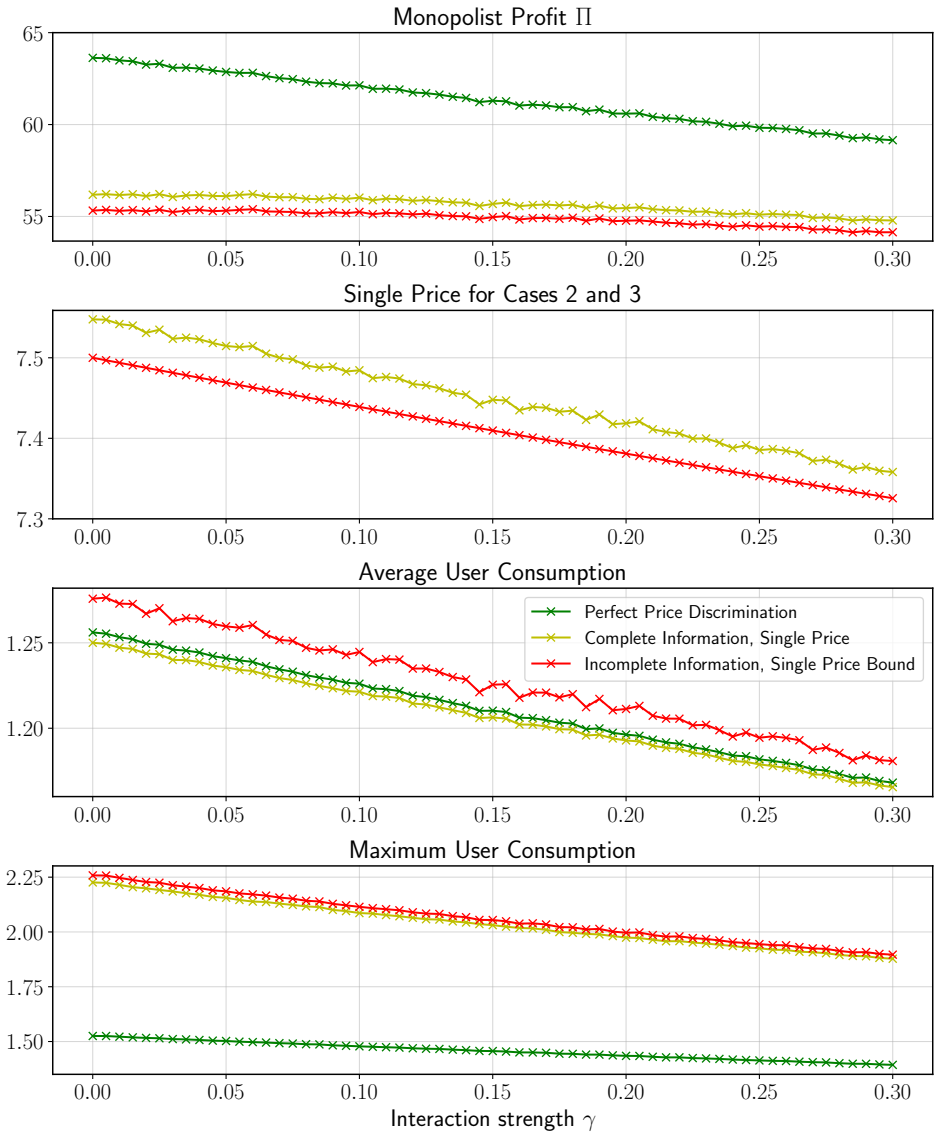

Figure 3.3: Profit of monopolist, single prices (3.10) and (3.13a), average user consumption, and maximum consumption under perfect price discrimination (green), single pricing under complete information (3.10) (yellow), and single pricing under incomplete information (3.13a) (red). 10,000 iterations, $a \sim \text{unif}[8, 12]$, $b \sim \text{unif}[0.75, 1.25]$, $c_i = 2 \; \forall \; i \in \mathcal{I}$.

(MIQCP) due to the binary constraint to expose exactly m, $0 \leq m \leq n$ users to the network effect and the quadratic objective, and so (3.18) does not admit a closed form solution. An analytical solution requires exhaustive search, which is computationally infeasible for any real network of users. Therefore, we resort to the following heuristic which was hinted at at the end of Section 3.4: Given the user parameters $\mathbf{a}$ and $\mathbf{b}$ and the single price p, we first compute the consumptions in the absence of any network effects, denoted with $\tilde{\mathbf{x}} = B^{-1}(\mathbf{a} - p\mathbf{1})$. Next, we calculate the optimal consumptions with the expectations of $\mathbb{E}[a]$ and $\mathbb{E}[b]$, which we denote with $\mathbb{E}[x]$. Lastly, the pairwise differences $|\mathbb{E}[x] - \tilde{x}_i|$ are put into a sorted list, and the heuristic selection algorithm returns the indices of the m largest values in this list. That is, $\Delta_h = \operatorname{diag}(\delta_{h,1}, \ldots, \delta_{h,n})$, where $\delta_{h,i} = 1$ if consumer i belongs to the set of the m largest $|\mathbb{E}[x] - \tilde{x}_i|$, and $\delta_{h,i} = 0$ otherwise.

The idea of this heuristic is motivated by Theorem 8, according to which a high consumer in a network of low consumers can result in a consumption increase of low consumers. Since the user parameters are sampled from a finite distribution, a single price on non-identical users always results in suboptimal profit, but approaches optimality as users become more similar. Exposing the highest and lowest consumers (measured against $\mathbb{E}[x]$) to the network effect nudges high users (low users) to consume less (more), thereby making the users more similar in their consumption, which in turn increases the utility's profit.

Further, the fact that the maximum user consumption under perfect price discrimination (which achieves notably better profit than single pricing, see Figure 3.3) is about 30 % lower than under single pricing corroborates the notion of exposing high consumers to the peer effects. According to Theorem 8, such users reduce their consumption in response to the peer effect, which reduces the maximum user consumption to increase profit.

The utility needs to find the sweet spot between the following two extremes: Targeting too few users results in a suboptimal increase in profit. On the other hand, according to Theorem 9, targeting too many users leads to an overall consumption decrease because targeting a customer whose neighbors are already exposed to the network causes the neighbors to reduce their consumption further.

Note that this heuristic neither takes into account the interaction matrix W nor the fact whether the deviation of the actual consumption from the expected one is positive or negative, and so it could be improved by running a classification algorithm on the features $|\mathbb{E}[x] - \tilde{x}_i|_+, |\mathbb{E}[x] - \tilde{x}_i|_-$, and $\gamma_i \tilde{x}_i \left(\sum_{j \in \mathcal{I}} w_{ij} \tilde{x}_j - \tilde{x}_i\right)$.

3.5.2 Simulation

We let $c_i = 2$, $n = 10$ as in Section 3.4 and analyze all three network topologies depicted in Figure 3.2. a_i and b_i are sampled from the same uniform distributions. We set the exogenous price as the profit-maximizing price in the absence of peer effects (3.14a), from which the expected consumption $\mathbb{E}[x]$ is determined with (3.14b). The analytical solution to the MIQCP (3.18) is determined with `Gurobi` (*Gurobi Optimization* 2016). We repeat this calculation 10,000 times and take the mean across all iterations. To describe the performance

of the heuristic, we define the performance metric S as follows:

$$S_m = \frac{\Pi_m^h - \Pi^E}{\Pi_m^* - \Pi^E} \cdot 100\%, \tag{3.19}$$

where Π_m^* and Π_m^h denote the profit under the analytical solution of (3.18) and the heuristic with m targeted users, respectively. Π^E denotes the profit in the absence of any peer effects ($m = 0$) achieved with exogenous price p where the users consume according to $\tilde{\mathbf{x}} = B^{-1}(\mathbf{a} - p\mathbf{1})$. S_m captures the fraction of the heuristic's achieved profit improvement of the total possible improvement.

Figure 3.4 shows the objective for the heuristic Π^h (solid lines) and analytical solution Π^* (colored dashed line) for all network topologies as a function of m. The expected profit with $m = 0$ follows by taking the expectation of the profit

$$\mathbb{E}[\Pi]_{m=0} = n \cdot \mathbb{E}_{a\sim U[8,12]}\mathbb{E}_{b\sim U[0.75,1.25]} \left[px - cx^2\right] \Big|_{x=\frac{a-p}{2b}},$$

which is depicted as the black dashed line. Further, the percentage of cases where the heuristic selects the identical subset of users as the analytical solution is depicted in the second subplot. S_m and the maximum user consumption as a function of m are provided in the third and fourth subplot, respectively.

For all network topologies, it can be seen that the optimal solution to (3.18) achieves an increase in profit by $\approx 1\%$ for $m \in \{2, 3, 4\}$ compared to the case of no targeting, while at the same time reducing the peak consumption by $\approx 4\%$. The performance of the heuristic decreases in the number of consumers targeted and reaches its minimum at $\approx 75\%$, $\approx 82\%$, and $\approx 90\%$ for the ring, star, and fully connected network, respectively. The percentage of optimal choices across all 10,000 iterations is always $> 22\%$. These results suggest that the presented heuristic achieves a good approximation of the optimal solution, which is NP-hard and computationally intractable for larger, real-world networks.

3.6 Conclusion

Motivated by home energy reports that benchmark the consumption of individual users against their neighbors, we proposed a two-stage game-theoretic model for a network of electricity consumers, in which each consumer seeks to optimize her individual utility function that includes a peer effect term. Specifically, users derive positive utility from consuming less energy than the average of their neighbors, and vice versa. We investigated profit-maximizing pricing schemes for the complete and incomplete information scenario as well as for the single price and perfect price discrimination case. We provided theoretical statements with regard to overall consumption, efficiency, and profit under network uncertainty. For the case of targeting only a subset of all available consumers under an exogenous single price, we formulated the monopolist's profit maximization problem. The resulting NP-hard

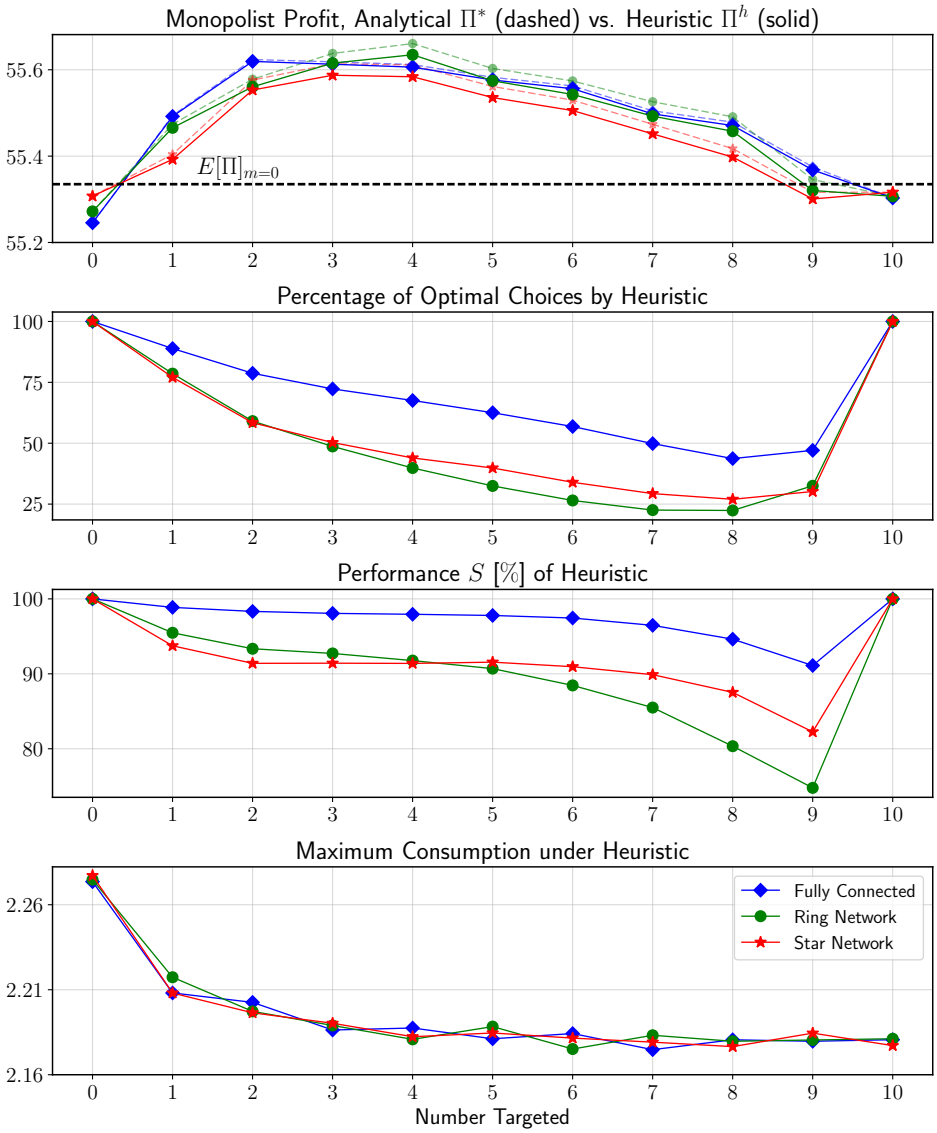

Figure 3.4: Average profit, percentage of optimal choice of heuristic, regret, and average infinity norm of consumption for the utility's profit maximization problem under the single price (3.10). 10,000 iterations, $a \sim \text{unif}[8, 12]$, $b \sim \text{unif}[0.75, 1.25]$, $c_i = 2$.

optimization problem was solved with a heuristic approach, which simply targets those users who deviate most from the expected consumption in the hypothetical absence of peer effects. Compared to the analytical solution, this heuristic was shown to achieve acceptable accuracy.

This work could be extended by incorporating time. In particular, if we allow the monopolist to also procure electricity from the wholesale market whose prices are fluctuating, an algorithmic and online treatment of this problem becomes necessary. The goal then becomes to learn user preferences and the network structure over time. Further, the selection problem to target the most valuable users for the objective of profit maximization calls for modeling peer effects in auction settings, where the desired goal is to design a truthful and incentive compatible mechanism to elicit user preferences.

Chapter 4

Hedging Strategies in Electricity Markets

4.1 Introduction

Historically, electricity was supplied by vertically integrated entities which maintained full functional control over the entire supply chain, including generation, transmission, and distribution assets. This static structure constituted an impediment for new energy providers on both the supply and retail end to participate in the energy market. In the United States, the Federal Energy Regulatory Commission issued Orders 888 and 889 in April 1996 to remove such barriers of entry in an attempt to promote competition and market efficiency (*FERC Order No. 888. Transmission Open Access. Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities (Final Rule)*; *FERC Order No. 889: OASIS: Open Access Same-Time Information System and Standards of Conduct (Final Rule)*). The result of this market design process was a combination of a central electricity pool operating day-ahead, overseen by Independent System Operators (ISOs), and bilateral trading between generating companies and electric utilities, which supplanted the traditional, vertically integrated entities.

As a consequence of the restructuring process, generators and utilities in the electricity market started facing price and quantity risks ensuing from the inelasticity of user demand, the steep supply curve due to the slowly changing nature of power plants' output adjustment, and prohibitive cost of energy storage. These factors allow small increases or decreases of demand to result in a price boom or bust, respectively. Furthermore, despite the fact that the economic consensus calls for passing along varying electricity prices to end-users in order to increase economic efficiency (Borenstein, Jaske, and Rosenfeld 2002b; Borenstein 2005; Borenstein and Holland 2005), policymakers have retained quasi-fixed electricity tariffs, e.g. Time-of-Use pricing. In conjunction with the obligation of utilities to service end-users with electricity at all times, risks associated with sudden price spikes are borne by the utility. This

market situation has resulted in several crises. For instance, unseasonably warm climate in the summer of 2000 resulted in California's wholesale electricity prices rising to average prices of more than 140 USD/MWh, leading to the bankruptcy of Pacific Gas & Electric, California's largest utility, and high profits of electricity generators (Borenstein, Jaske, and Rosenfeld 2002a). Similar crises occurred in Texas (2004) and in the Midwestern United States (1998).

These crises resulted in the following notable developments. Firstly, electric utilities and generating companies started to hedge against price fluctuations through contracts on different scales of time, ranging from short-term forward contracts to long-term contracts, thereby locking in a fixed price and quantity to be delivered over a contractually specified period of time. Secondly, Demand-Side Management (DSM), which aims to affect consumer behavior during periods of peak demand, emerged as a viable tool to partially relay price risks to end-users. For instance, companies like `OPOWER` provide Demand Response (DR) services to utilities, allowing them to offer monetary rewards to end-users in exchange for a reduction in electricity consumption during hours of peak demand (*OPOWER*).

Motivated by these shortcomings, a large body of research, particularly in operations research, has studied optimal hedging contracts, most often from the utility perspective, including (Oum, Oren, and Deng 2006; Oum and Oren 2009), where the authors construct an optimal one-step hedging portfolio with standard power options, or (Hatami, Seifi, and Skeikh-El-Eslami 2009), which finds an optimal energy procurement policy with stochastic programming over a specified period. Borenstein 2006 analyzes hedging instruments against price volatility for industrial customers. Wolak 2001 investigates hedging strategies for electricity generators.

While there exists a large body of literature on operational and algorithmic aspects of DR (e.g. load scheduling and shifting (Mohsenian-Rad et al. 2010; Li, Chen, and Low 2011; Aalami, Moghaddam, and Yousefi 2010)), significantly less research has focused on the role of DR programs as an alternative way of hedging. Notable examples are Deng and Xu (2009), where the authors discuss interruptible service contracts, and Sezgen, Goldman, and Krishnarao (2007), which estimates the economic value of DR programs for commerical customers by adapting models used to value energy options. To the best of our knowledge, no significant research has investigated the option value of residential DR programs. To close this gap, we derive a stylized model for the utility's profit under such DR schemes and determine its optimal profit. The methodology we use is closest in spirit with Bitar et al. (2012), where the authors determine the optimal bidding volume of wind generators in a conventional energy market. We compare the profit under Demand Response to the case of forward contracts and call options by incorporating the conditional value at risk (Rockafellar and Uryasev 2002) measure. Using smart meter data of residential customers in California, we find that DR can yield higher expected profits than under forward contracts and call options, especially in the presence of high wholesale electricity prices.

The remainder of this chapter is organized as follows: In Section 4.2, we describe the interactions between the participants in energy markets. Section 4.3 introduces forward contracts, options, and Demand Response as hedging instruments for the Demand Response

Provider. The effect of uncertainty in the user demand on the expected profit of the Demand Response Provider is investigated in Section 4.4. We compute optimal, profit-maximizing portfolios for load-serving entities in Section 4.5 and simulate decision boundaries between them in Section 4.6. Section 4.7 concludes. All proofs are relegated to Appendix A.3.

Notation

Let $\mathbb{E}[\cdot]$ denote the expectation of a random variable. Let $[\cdot]_+$ denote the hinge function, i.e. $[x]_+ = \max(0, x)$.

4.2 Market Participants

Figure 4.1 illustrates the interaction between generating companies, load-serving entities (utilities), the wholesale electricity market, and end-users of electricity.

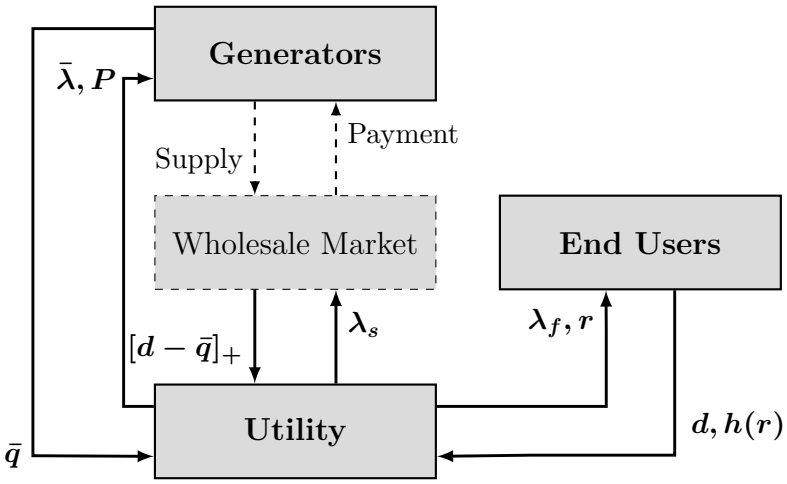

Figure 4.1: Energy Market Participants and their Interactions

The electric utility can strike one-to-one contracts with generating companies to purchase a fixed amount of electricity $\bar{q}$ at a locked-in price $\bar{\lambda}$ to be delivered at some *a-priori* specified time in the future. P denotes the premium for each reserved unit of electricity. The utility provides end-users with electricity at a fixed unit rate λ_f and is obligated to cover the random demand d at all times. The rate λ_f is exogenously set by the Public Utilities Commission. However, the utility can use DR to incentivize users to temporarily reduce their demand. This is achieved by offering the reward r to end-users, which elicits a demand reduction $h(r)$. If the demand d exceeds $\bar{q}$, i.e. the purchased amount of electricity through one-to-one contracts with generators, the utility has to procure $[d - \bar{q}]_+$ units of electricity from the wholesale market at uncertain wholesale price λ_s per unit. The market clearing price λ_s, reflected by Locational Marginal Prices (LMPs), is a random variable and depends on the

ratio of energy supply by generators, the total demand $[d - \bar{q}]_+$, operational constraints, as well as congestion of the grid.

The interactions between generators and the utility as well as between end-users and the utility are instruments to hedge utilities against high prices λ_s. If the utility expects high wholesale prices λ_s, then it has an incentive to reduce customer demand d by engaging in DR, or to procure cheaper electricity through contracts with generating companies. We make the following assumptions:

Assumption 6. *The utility is risk-neutral.*

Assumption 7. *The utility is price-taking.*

Assumption 7 is a natural assumption, stating that the utility cannot influence prices by exerting market power. Together with Assumption 6, the question we seek to answer in the remainder of this chapter is how the utility maximizes its expected profit in the presence of the random variables d and λ_s and hedging instruments. For simplicity, we focus on a single load zone to avoid spatial heterogeneity of LMPs.

4.3 Optimal Hedging Strategies

Let λ_s and d be random variables with cumulative distribution functions (CDF) $G(\cdot)$ and $F(\cdot)$, respectively. G and F are assumed to have support $[0, \infty)$ and $[d_{\min}, d_{\max}]$, respectively, where $0 \leq d_{\min} \leq d_{\max}$. We assume the absence of any energy storage capabilities and focus on a single-period setting, where the LSE can purchase hedging instruments at time 0, possessing only an estimate of consumer demand d and real-time spot price λ_s at time 1. At time 1, the random variables d and λ_s materialize and the LSE's profit Π as a function of the hedging instruments purchased at time 0 is determined. Figure 4.2 illustrates the hedging process. The LSE aims to maximize its *expected* profit $\mathbb{E}[\Pi]$ by deciding on its portfolio of hedging instruments at $t = 0$.

Figure 4.2: Timeline of Hedging

In the following, we analyze the cases for (a) no hedging instruments, (b) forward contract, (c) call option, and (d) DR and derive explicit expressions for the optimal contracts and corresponding profits for cases (b)-(d).

4.3.1 Base Case (No Hedging)

If the LSE does not buy any options at stage 0, its expected profit at time 1 is simply

$$\mathbb{E}[\Pi] = (\lambda_f - \mathbb{E}[\lambda_s]) \cdot \mathbb{E}[d]. \tag{4.1}$$

We will compare the profit of this base case to the forward contract, call option, and DR in the following.

4.3.2 Forward Contract

A forward contract is a one-on-one agreement between the LSE and an electricity generator, which obligates the generator (at time 0) to deliver a fixed amount of electricity $\bar{q}$ at a locked-in price $\bar{\lambda}_F$ to the LSE at some point in the future (time 1). Forward contracts possess high flexibility and are traded as over-the-counter products. The LSE seeks to sign such a contract if it has reason to believe the expected wholesale price at the time of delivery to exceed $\bar{\lambda}_F$, and the generator will do so in the opposite case. If, at time 1, $\bar{q} > d$, the LSE has purchased too much volume at time 0, and so $\bar{q} - d$ are wasted. Conversely, if $\bar{q} < d$ at time 1, $d - \bar{q}$ units of electricity have to be bought at real-time spot price λ_s.

The profit Π_F under a forward contract of volume $\bar{q}$ at unit price $\bar{\lambda}$ is therefore expressed as

$$\Pi_F = \lambda_f d - \bar{\lambda}_F \bar{q} - \lambda_s [d - \bar{q}]_+. \tag{4.2}$$

Theorem 12 (Optimal Forward Contract)**.** *With $\mathbb{E}[\lambda_s] > \bar{\lambda}_F$, the optimal contract volume $\bar{q}^*$ and the optimal expected profit $\mathbb{E}[\Pi_F^*]$ become*

$$\bar{q}^* = F^{-1}\left(1 - \frac{\bar{\lambda}_F}{\mathbb{E}[\lambda_s]}\right), \tag{4.3a}$$

$$\mathbb{E}[\Pi_F^*] = \lambda_f \mathbb{E}[d] - \mathbb{E}[\lambda_s] \int_{F^{-1}\left(1 - \frac{\bar{\lambda}_F}{\mathbb{E}[\lambda_s]}\right)}^{\infty} x f(x)\, dx. \tag{4.3b}$$

4.3.3 Call Option

Similar to fixed forward contracts, the LSE can strike one-on-one deals with a counterparty over an agreed volume $\bar{q}$ at strike price $\bar{\lambda}_C$. The key difference is that the LSE can, but is not obliged to, exercise the call option if $\bar{\lambda}_C < \lambda_s$ at time 1. Typically the buyer of the call option pays a premium P for each unit of the call option.

The profit Π_C under a call option with volume $\bar{q}$ at strike price $\bar{\lambda}_C$ at the premium P per unit can thus be written as

$$\Pi_C = \lambda_f d - \lambda_s \left[d - \bar{q}\right]_+ - P\bar{q} - \min(\bar{\lambda}_C, \lambda_s) \cdot \min(d, \bar{q}). \tag{4.4}$$

The last term of (4.4) encodes the fact that the LSE can cover up to $\bar{q}$ units at the cheaper of the strike price $\bar{\lambda}_C$ or the wholesale price λ_s. The remainder $[d - \bar{q}]_+$ has to be purchased from the spot market at price λ_s.

Theorem 13 (Optimal Call)**.** *With* $\mathbb{E}[\lambda_s] > P + \bar{\lambda}_C - \int_0^{\bar{\lambda}_C} G(y)dy$*, the profit-maximizing call volume* $\bar{q}^*$ *and the corresponding optimal expected profit* $\mathbb{E}[\Pi_C^*]$ *are*

$$\bar{q}^* = F^{-1}\left(1 - \frac{P}{\mathbb{E}[\lambda_S] - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy}\right), \tag{4.5a}$$

$$\mathbb{E}[\Pi_C^*] = \left(\lambda_f - \bar{\lambda} + \int_0^{\bar{\lambda}_C} G(y)dy\right)\mathbb{E}[d] - \left(\mathbb{E}[\lambda_s] - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy\right)\int_{\bar{q}^*}^{\infty} xf(x)dx. \tag{4.5b}$$

4.3.4 Demand Response

We model the effect of demand response as a shift in the distribution of the consumer towards zero, induced by the monetary reward $r \in \mathbb{R}_+$ transferred from the LSE to the consumer as a lump sum. Note that the real reduction of the consumer in response to the DR signal has to be estimated by constructing the counterfactual consumption in the absence of the DR signal, whose estimation is analyzed in the second part of this thesis (cf. Chapters 7 and 8). The interested reader is referred to (Zhou, Balandat, and Tomlin 2016b; Zhou, Balandat, and Tomlin 2016a) before jumping to these chapters.

Let $f(d)$ denote the probability density function of d in the absence of any reward with support $[d_{\min}, d_{\max}]$. Let $F(d|r)$ denote the cumulative distribution function of the random variable d, given the reward level r. Then the distribution shift is modeled as

$$F(d|r) = \begin{cases} 0, & \text{if } d < d_{\min} \\ F(d + h(r)), & \text{if } d \geq d_{\min} \end{cases} \tag{4.6}$$

where $h(r)$ is a concave, increasing function representing the elasticity of the user in response to reward r, i.e. the relative reduction of consumption as a function of r. $h(r)$ is equivalent to the shift of the location parameter of distribution $f(\cdot)$. We make the following assumption:

Assumption 8. *The reward* $r \geq 0$ *induces a linear shift, i.e.*

$$h(r) = \alpha r, \quad \alpha > 0. \tag{4.7}$$

With Assumption 8 and the definition of the distribution shift, it becomes clear that the distribution $f(\cdot|r)$, given a reward $r > 0$, has support $[d_{\min}, d_{\max} - h(r)]$ with discrete mass $\int_{d_{\max}-h(r)}^{d_{\max}} f(x)\ dx$ at $d_{\min}$.

Assumption 8 is necessary for analytical tractability of the DR hedging case. We note that the linearity of $h(r)$ is unrealistic, since it implies that for large enough reward levels r, the user consumes zero with probability 1. However, for small reward levels, a linear price elasticity of demand $h(r)$ can be justified.

The LSE's profit Π_{DR} with Demand Response is

$$\Pi_{\text{DR}} = (\lambda_f - \lambda_s)d(r) - r. \tag{4.8}$$

From (4.8), it immediately follows that DR only makes sense in the presence of large expected spot prices $\mathbb{E}[\lambda_s]$ at time 1 which exceed the fixed contractual price λ_f. Then the optimal profit Π^*_{DR} is the minimal expected loss of the LSE.

Theorem 14 (Optimal Demand Response)**.** *With $\mathbb{E}[\lambda_s] > \lambda_f$, the profit-maximizing reward r^* and the optimal expected profit $\mathbb{E}\Pi^*_{\text{DR}}$ are*

$$r^* = \begin{cases} \frac{1}{\alpha} F^{-1}\left(1 - \frac{1}{\alpha \cdot \left(\mathbb{E}[\lambda_s] - \lambda_f\right)}\right), & \text{if } \frac{1}{\alpha} < \mathbb{E}[\lambda_s] - \lambda_f \\ 0, & \text{otherwise} \end{cases} \tag{4.9a}$$

$$\mathbb{E}\Pi^*_{\text{DR}} = \begin{cases} (\lambda_f - \mathbb{E}[\lambda_s]) \int_{\alpha r^*}^{\infty} x f(x) dx, & \text{if } \frac{1}{\alpha} < \mathbb{E}[\lambda_s] - \lambda_f \\ (\lambda_f - \mathbb{E}[\lambda_s])\ \mathbb{E}[d], & \text{otherwise} \end{cases} \tag{4.9b}$$

The condition $\alpha > (\mathbb{E}[\lambda_s] - \lambda_f)^{-1}$ for the optimal reward means that the ability to shift, $1/\alpha$, must be greater than the inverse of the expected price difference $(\mathbb{E}[\lambda_s] - \lambda_f)^{-1}$ to make DR profitable. The higher the expected price difference $\mathbb{E}[\lambda_s] - \lambda_f$, the less stringent the requirement on α, which agrees with intuition.

Theorem 15 (Diversified Portfolios)**.** *For general demand distributions, the optimal portfolio can either consist of a unique option or a combination of call and forward contract options, but never of a combination of DR and either call or forward contract options. For the special case of a uniform demand distribution, the optimal portfolio always consists of a unique option, i.e. diversified portfolios consisting of more than one option are always suboptimal.*

Depending on the properties of the demand distribution $F(\cdot)$, a mixed portfolio of call and forward contract options can exist, but is impossible to obtain in closed form for general distributions. This is consistent with the approach in (Oum, Oren, and Deng 2006) where the authors replicate the optimal portfolio (which would be continuous) with a finite set of options. Due to Theorem 15, we restrict our attention to optimal portfolios consisting of a unique option in the remainder of this chapter.

4.4 The Effect of Uncertainty

For a better understanding of the optimal profits under the different contracts $\Pi^*_F, \Pi^*_C, \Pi^*_{\text{DR}}$ introduced in the previous section, we relate these quantities to properties of the consumption distribution $F(\cdot)$.

4.4.1 Influence of Distribution Tail

By incorporating the *Conditional Value-at-Risk* (CVaR) measure (Rockafellar and Uryasev 2002), we can relate the optimal profits to the tail properties of the consumption density $f(\cdot)$.

The CVaR at confidence level $\alpha \in (0,1)$ of a random variable X with CDF $F(\cdot)$ representing loss is formally defined as

$$\mathrm{CVaR}_\alpha(X) = \mathbb{E}[X \mid X \geq F^{-1}(\alpha)] \tag{4.10}$$

and can be interpreted as the expected loss attained in the worst $(1-\alpha)\cdot 100\%$ of cases or the expectation of the $(1-\alpha)$ probability tail of X. With this definition, the optimal expected profits under the different options Π_F^*, Π_C^*, and Π_{DR}^* are reformulated in Proposition 1.

Proposition 1. *With $\alpha > (\mathbb{E}[\lambda_s] - \lambda_f)^{-1}$ and the definition of CVaR, the optimal expected profits under the forward contract $\mathbb{E}[\Pi_F^*]$, the call option $\mathbb{E}[\Pi_C^*]$, and Demand Response $\mathbb{E}[\Pi_{\mathrm{DR}}^*]$ can be expressed as follows:*

$$\begin{aligned}
\mathbb{E}[\Pi_F^*] &= \lambda_f \mathbb{E}[d] - \bar{\lambda}_F \mathbb{E}[d \mid d \geq F^{-1}(1 - \bar{\lambda}_F/\mathbb{E}[\lambda_s])] \\
&= \lambda_f \mathbb{E}[d] - \bar{\lambda}_F \cdot \mathrm{CVaR}_{\alpha_F}(d) && \text{(4.11a)} \\
\mathbb{E}[\Pi_C^*] &= \left(\lambda_f - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy\right) \mathbb{E}[d] - P \cdot \mathrm{CVaR}_{\alpha_C}(d) && \text{(4.11b)} \\
\mathbb{E}[\Pi_{\mathrm{DR}}^*] &= -\frac{1}{\alpha} \cdot \mathrm{CVaR}_{\alpha_{\mathrm{DR}}}(d) && \text{(4.11c)}
\end{aligned}$$

where we used the definitions

$$\begin{aligned}
\alpha_F &= 1 - \frac{\bar{\lambda}_F}{\mathbb{E}[\lambda_s]} && \text{(4.12a)} \\
\alpha_C &= 1 - \frac{P}{\mathbb{E}[\lambda_s] - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy} && \text{(4.12b)} \\
\alpha_{\mathrm{DR}} &= 1 - \frac{1}{\alpha \cdot (\mathbb{E}[\lambda_s] - \lambda_f)} && \text{(4.12c)}
\end{aligned}$$

From Proposition 1, it follows that the optimal profit decreases as the conditional expectation of the tail increases, that is, the more heavy-tailed the consumption distribution $f(\cdot)$ becomes. It is illustrative to analyze the optimal decisions and corresponding optimal expected profits for perfect information of d, which are given in the following:

$$\begin{aligned}
\bar{q}_F^*|d = d, \qquad \bar{q}_C^*|d = d, \qquad r^*|d = d/\alpha \\
\mathbb{E}[\Pi_F^*|d] &= (\lambda_f - \bar{\lambda}_F) \cdot d && \text{(4.13a)} \\
\mathbb{E}[\Pi_C^*|d] &= \left(\lambda_f - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy - P\right) d && \text{(4.13b)} \\
\mathbb{E}[\Pi_{\mathrm{DR}}|d] &= -d/\alpha && \text{(4.13c)}
\end{aligned}$$

$\bar{q}_F^*|d$ and $\bar{q}_C^*|d$ denote the optimal forward contract and call volume, respectively. $r^*|d$ signifies the optimal DR reward.

4.4.2 Influence of Statistical Dispersion

In this section, we attempt to construct a relationship between the statistical dispersion of the consumption distribution $F(\cdot)$ and the optimal expected profit. Intuitively, the more spread out the distribution $F(\cdot)$, the lower the expected profit. While many measures for statistical dispersion exist in the literature, such as interquartile ranges, absolute deviation, variance-to-mean-ratio, etc., we express the optimal expected profits $\mathbb{E}[\Pi_F^*]$, $\mathbb{E}[\Pi_C^*]$, and $\mathbb{E}[\Pi_{\mathrm{DR}}^*]$ in terms of the standard deviation σ for the special case of a uniform distribution with support $[d_{\min}, d_{\max}]$ for expositional ease and analytical tractability.

Proposition 2. *For the uniform distribution $F(\cdot)$ with support $[d_{\min}, d_{\max}]$, the optimal expected profits under the conditions $\mathbb{E}[\lambda_s] > \max\left(\bar{\lambda}_F, P + \bar{\lambda}_C - \int_0^{\bar{\lambda}_C} G(y)\, dy\right)$ and $\alpha > (\mathbb{E}[\lambda_s] - \lambda_f)^{-1}$ are expressed as follows:*

$$\mathbb{E}[\Pi_F^*] = \lambda_f \mathbb{E}[d] - \bar{\lambda}_F d_{\min} - \sqrt{3}\mathbb{E}[\lambda_s](1-\alpha_F^2)\sigma \tag{4.14a}$$

$$\begin{aligned}\mathbb{E}\Pi_C^* = &\left(\lambda_f - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy\right)\mathbb{E}[d] - P d_{\min} \\ &- \sqrt{3}\left(\mathbb{E}[\lambda_s] - \bar{\lambda}_C + \int_0^{\bar{\lambda}_C} G(y)dy\right)(1-\alpha_C^2)\sigma\end{aligned} \tag{4.14b}$$

$$\mathbb{E}[\Pi_{\mathrm{DR}}^*] = -d_{\min}/\alpha - \sqrt{3}(\mathbb{E}[\lambda_s] - \lambda_f)(1-\alpha_{\mathrm{DR}}^2)\sigma \tag{4.14c}$$

For the case of perfect information, i.e. $\sigma = 0$ and $d_{\min} = d_{\max} = d$, the equations for the optimal expected profit under perfect information (4.13a)-(4.13c) are recovered. Equations (4.14a)-(4.14c) explain that the optimal expected profit for each case decreases linearly in σ, giving rise to the notion that more "spread out" distributions diminish the expected profit. The rate of decrease depends on case-specific parameters, whose relation to each other determines which hedging option is profit-maximizing for a particular case. As consumption distributions typically are plagued by a large amount of uncertainty (large σ), improved load predictions to decrease σ have a direct economic benefit to the utility.

4.5 Choosing the Best Option

We now derive conditions on the random variables λ_s and d with distributions $G(\cdot)$ and $F(\cdot)$ and the option parameters $\bar{\lambda}_F$, $\bar{\lambda}_C$, P, and α announced at time 0 to determine the best hedging strategy consisting of a unique option. For analytical tractability, we make the following assumptions:

Assumption 9. *The real-time spot price λ_s is uniformly distributed with support $[0, s_{\max}]$, that is, $G(y) = \frac{1}{s_{\max}}\mathbf{1}_{0\le y\le s_{\max}}$.*

Assumption 10. *The consumption is uniformly distributed in $[0, d_{\max}]$, that is, $F(x) = \frac{1}{d_{\max}}\mathbf{1}_{0\le\ x\le\ d_{\max}}$.*

Theorem 16. *Under Assumptions 9 and 10 and* $\mathbb{E}[\lambda_s] > \lambda_f$*, the forward contract is preferred over the call option, if*

$$\bar{\lambda}_F \leq \mathbb{E}[\lambda_s] - \frac{\mathbb{E}[\lambda_s] - \bar{\lambda}_C + \bar{\lambda}_C^2/(4\mathbb{E}[\lambda_s]) - P}{\sqrt{1 - \frac{\bar{\lambda}_C - \bar{\lambda}_C^2/(4\mathbb{E}[\lambda_s])}{\mathbb{E}[\lambda_s]}}}. \tag{4.15}$$

DR is preferred over the forward contract, if

$$\frac{1}{\alpha} \leq (\mathbb{E}[\lambda_s] - \lambda_f) \left[1 - \sqrt{\frac{\mathbb{E}[\lambda_s]}{\mathbb{E}[\lambda_s] - \lambda_f}} \left(1 - \frac{\bar{\lambda}_F}{\mathbb{E}[\lambda_s]}\right)\right]. \tag{4.16}$$

Finally, DR is preferred over the call option, if

$$\frac{1}{\alpha} \leq (\mathbb{E}[\lambda_s] - \lambda_f) \left[1 - \sqrt{\frac{L}{(\mathbb{E}[\lambda_s] - \lambda_f)}} \left(1 - \frac{P}{L}\right)\right]. \tag{4.17}$$

with $L = (\mathbb{E}[\lambda_s] - \bar{\lambda}_C + \bar{\lambda}_C^2)/(4\mathbb{E}[\lambda_s])$ *and where* $\bar{\lambda}_F$ *and* $\bar{\lambda}_C$ *denote the unit price for each reserved unit of electricity under the forward contract and the call option, respectively.*

4.6 Simulations

Assumptions 9 and 10 admitted a closed form solution to the best hedging instrument, stated in (4.15)-(4.17). For a more elaborate analysis, we now repeat this exercise by approximating the demand distribution $F(\cdot)$ as well as the distribution of spot prices $G(\cdot)$ with real data from California to approximate decision boundaries for which the expected profits under different hedging instruments are identical. Since closed-form solution under this more realistic scenario do not exist, we plot these optimal decision boundaries as a function of the hedging parameters $P, \bar{\lambda}_F, \bar{\lambda}_C$, and α.

4.6.1 Empirical Distribution of Demand

We use hourly smart meter data from residential customers in California from the utilities Pacific Gas & Electric, San Diego Gas & Electric, and Southern California Edison to create a demand distribution for different sizes of user aggregations. The observations are restricted to hourly consumptions between 4-5 pm and 5-6 pm. Figure 4.3 shows the empirical PDFs and CDFs for different sizes of user aggregations. We approximate both functions as follows:

$$\hat{f}(x) = a(x - d_{\min})e^{-cx}, \quad a, c \in \mathbb{R}_+, x \in [d_{\min}, d_{\max}] \tag{4.18a}$$

$$\hat{F}(x) = \frac{a}{c^2}\left(cd_{\min} - cx - 1\right)e^{-cx} + \gamma, \quad \gamma \in \mathbb{R} \tag{4.18b}$$

With the constraints $\hat{F}(d_{\min}) = 0$ and $\hat{F}(d_{\max}) = 1$, the parameters a and γ can be found as a function of the decay parameter c. It can be seen that the approximations (4.18a) and (4.18b) fit the observed data reasonably well.

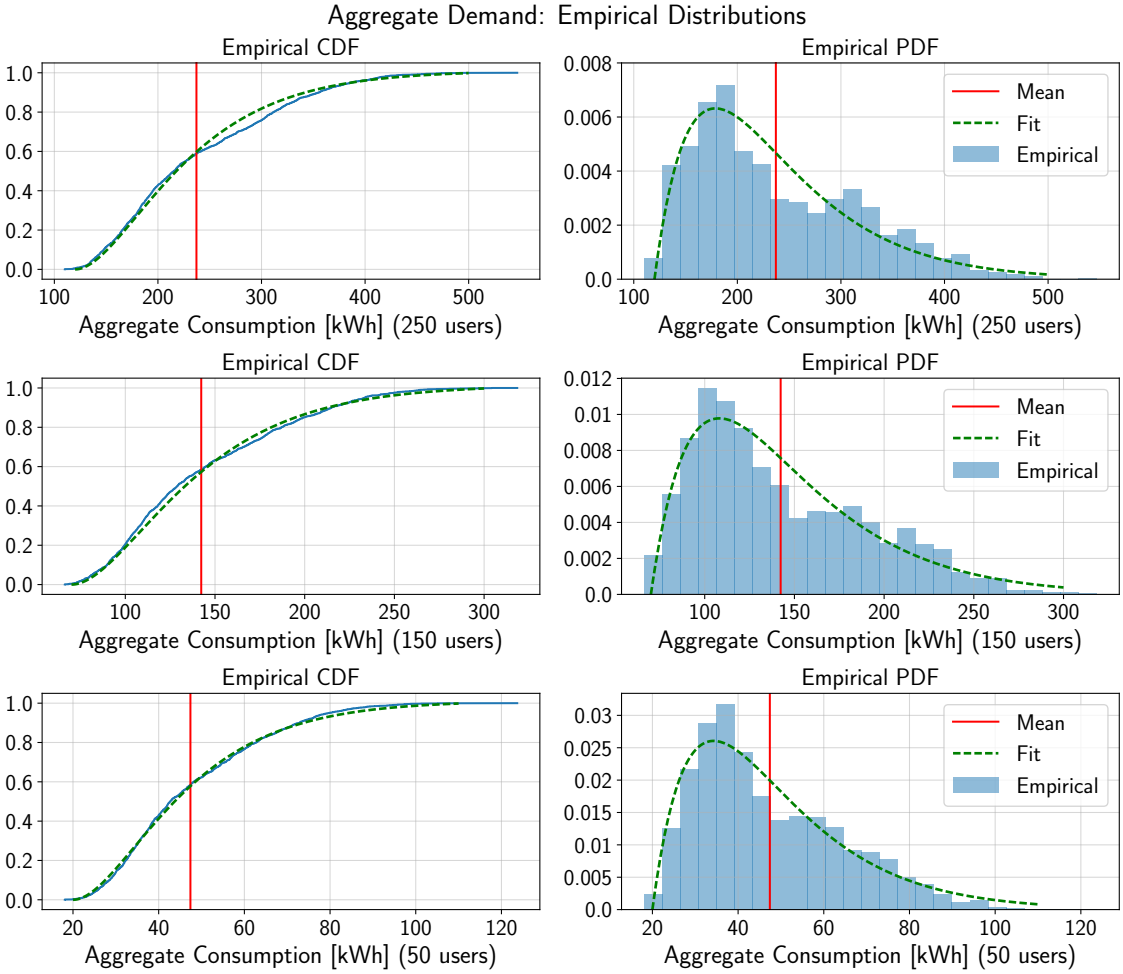

Figure 4.3: Distribution of Aggregate Hourly Consumption for Varying Aggregation Sizes, 4-6 pm. Top: 250 Users, Middle: 150 Users, Bottom: 50 Users.

4.6.2 Empirical Distribution of Wholesale Prices

To obtain the price distribution $G(\cdot)$, we convert 5-minute locational marginal prices (LMPs) λ_s set by the California Independent System Operator into an hourly format. The distribution $G(\cdot)$ of "high" LMPs is obtained by fitting a density function to the normalized histogram of those LMPs for which the two previous LMPs exceed the threshold $\xi > 0$, i.e. we consider all $\{\lambda_s | \lambda_{s,t-1} \geq \xi, \lambda_{s,t-2} \geq \xi\}$ for different thresholds ξ. We approximate the density function with a log-normal distribution:

$$\mathcal{N}(\ln x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \tag{4.19}$$

which has support $[0, \infty)$, that is, we disregard negative LMPs. Figure 4.4 shows the observed data and the approximations for thresholds $\xi = 80, 90, 100 \frac{\text{USD}}{\text{MWh}}$.

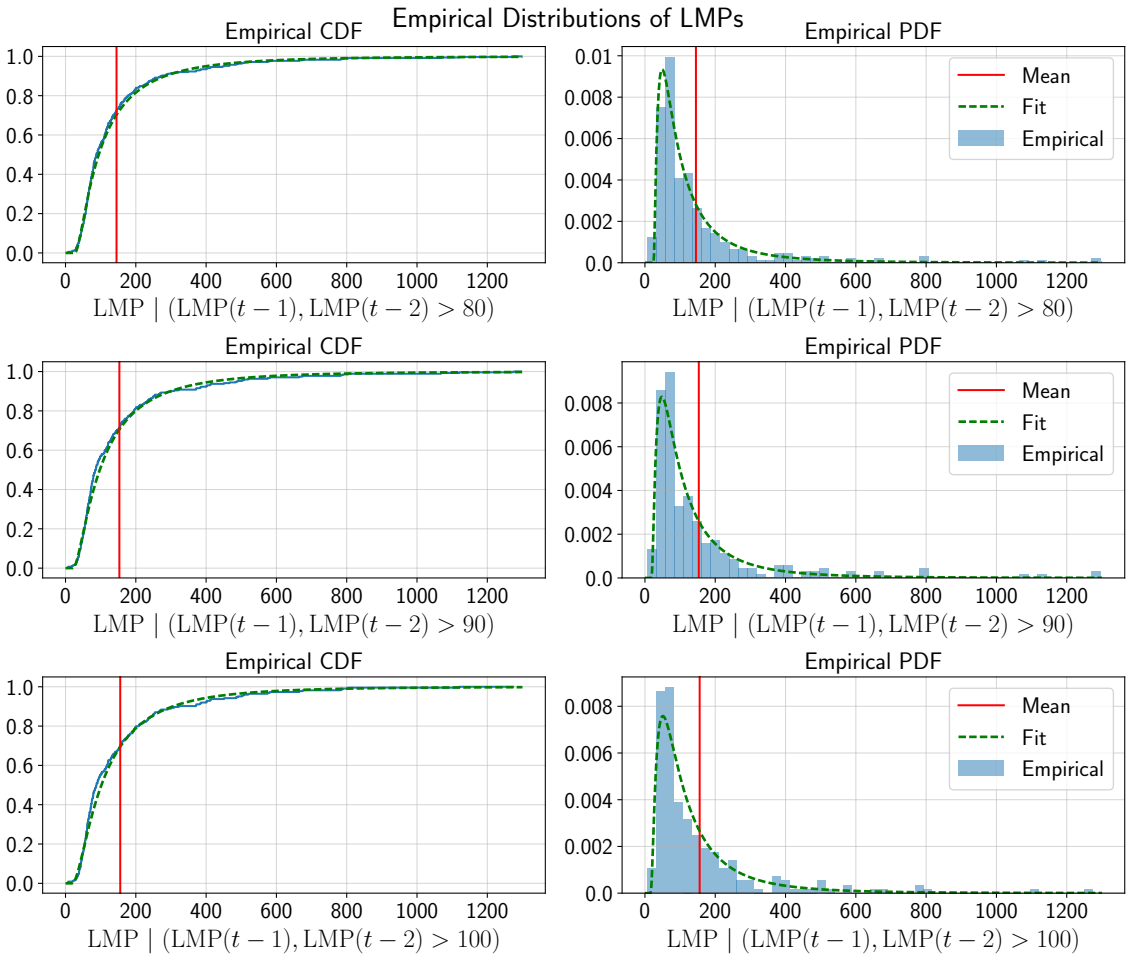

Figure 4.4: Distributions of CAISO LMPs conditional on previous prices exceeding threshold ξ for $\xi \in \{80\frac{\text{USD}}{\text{MWh}}, 90\frac{\text{USD}}{\text{MWh}}, 100\frac{\text{USD}}{\text{MWh}}\}$.

4.6.3 Pairwise Comparison of Hedging Instruments

We now compute decision boundaries of equal expected profit for all 3 pairs of hedging instruments with Newton's method, using the demand and price distributions derived in (4.18a), (4.18b), and (4.19).

4.6.3.1 DR vs. Forward Contract

Figure 4.5 shows the decision boundary of elasticity α above which the optimal expected profits under DR is greater than under the forward contract, that is, $\mathbb{E}[\Pi_{\text{DR}}] \geq \mathbb{E}[\Pi_F]$, for different expected spot prices $\mathbb{E}[\lambda_s]$ and forward contract prices $\bar{\lambda}_F$, assuming $\lambda_f \leq \mathbb{E}[\lambda_s]$. It is observed that α decreases as $\bar{\lambda}_F$ or the expected wholesale price $\mathbb{E}[\lambda_s]$ increase. The negative correlation of α with $\bar{\lambda}_F$ is consistent with expectations as a higher $\bar{\lambda}_F$ makes

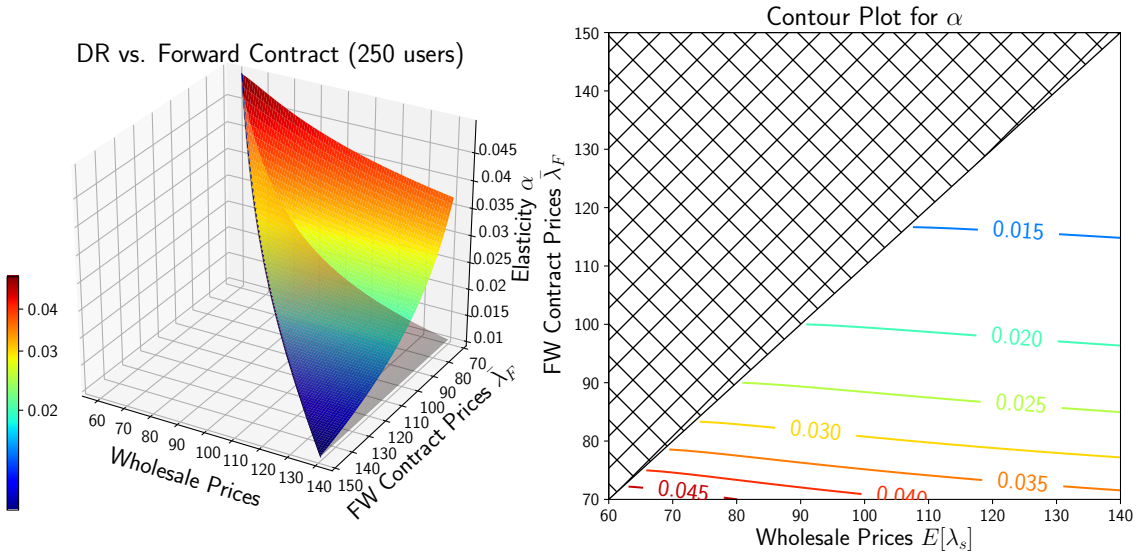

Figure 4.5: Boundaries and contours of equal expected profit for forward option and DR, 250 users, $\lambda_f = 0.05 \frac{\text{USD}}{\text{kWh}}$.

forward contracts more expensive. The fact that decreasing wholesale prices $\mathbb{E}[\lambda_s]$ make DR more competitive than forward contracts can be explained by comparing (4.3b) to (4.9b), which states that the entire demand d has to be covered at price λ_s in the DR case, compared to only $[d - \bar{q}]_+$ in the forward contract case. Also shown in Figure 4.5 is the lower bound on α (gray transparent surface) below which DR is non-profitable, i.e. $\{(\mathbb{E}[\lambda_s] - \lambda_f)^{-1} \mid 70 \leq \mathbb{E}[\lambda_s] \leq 150\}$, where we set the residential tariff to $\lambda_f = 0.05$ USD/kWh.

4.6.3.2 DR vs. Call

Figure 4.6 shows the decision boundary of α for different call strike prices λ_C and premium levels P above which $\mathbb{E}[\Pi_{\text{DR}}] \geq \mathbb{E}[\Pi_C]$ with $\xi = 80$. As the premium and strike price for the call option increase (and hence the call option becomes less attractive), DR becomes more profitable because α decreases.

4.6.3.3 Forward Contract vs. Call

Lastly, Figure 4.7 shows the decision surface for $\bar{\lambda}_F$ as a function of the call option parameters P and $\bar{\lambda}_C$ above which the forward contract is more profitable in expectation, i.e. $\mathbb{E}[\Pi_F] \geq \mathbb{E}[\Pi_C]$. As expected, the forward contract becomes more attractive as either the premium P or the call strike price $\bar{\lambda}_C$ increase.

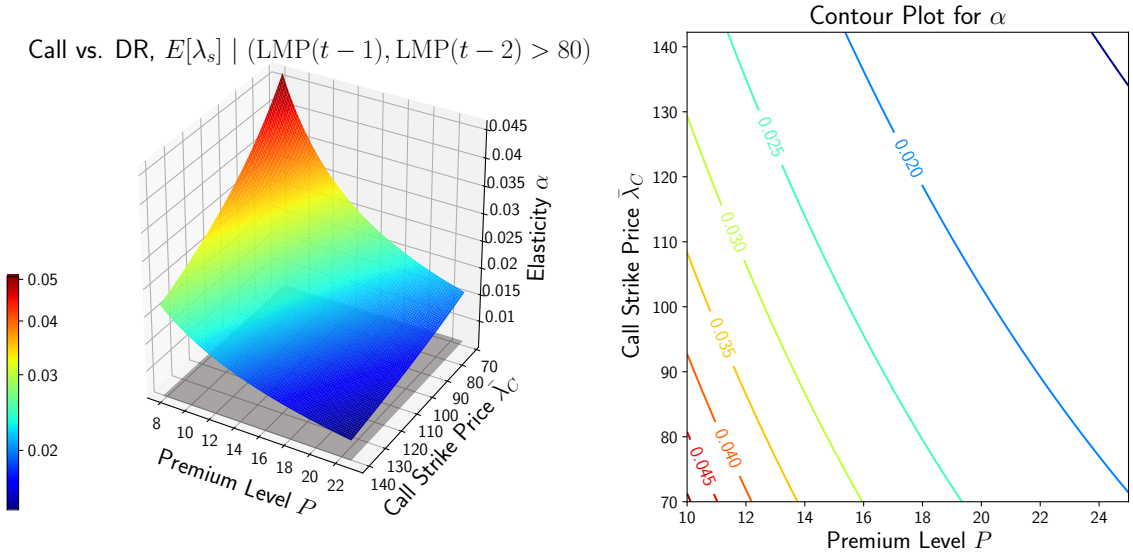

Figure 4.6: Boundaries and contours of equal expected profit for DR and call option, 250 users, $\lambda_f = 0.05 \frac{\mathrm{USD}}{\mathrm{kWh}}$.

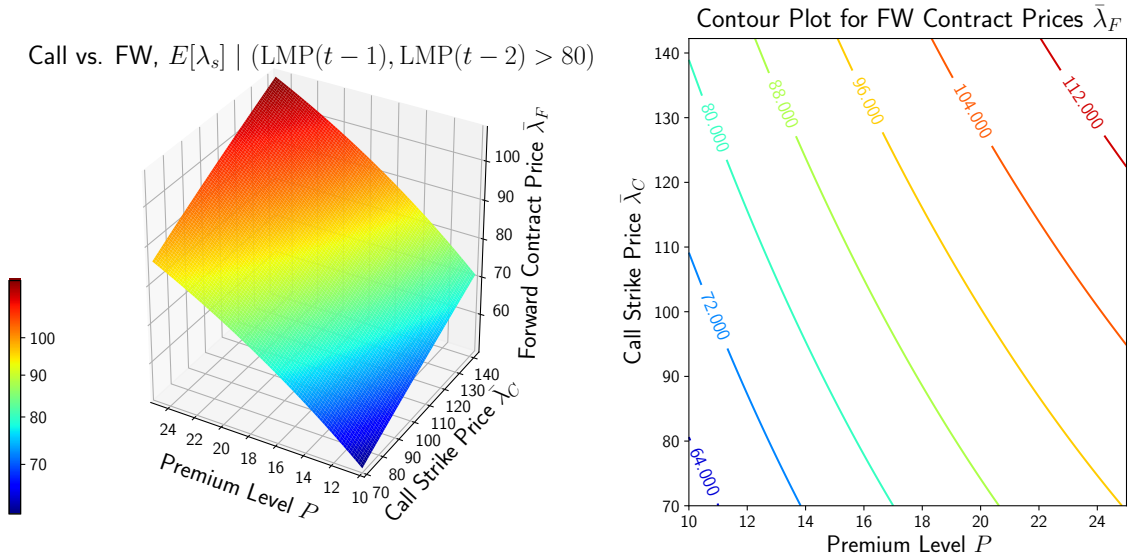

Figure 4.7: Boundaries and contours of equal expected profit for forward and call option, 250 users, $\lambda_f = 0.05 \frac{\mathrm{USD}}{\mathrm{kWh}}$.

4.6.4 Evaluation

Assuming a residential tariff of $0.05\frac{\text{USD}}{\text{kWh}}$, a lower bound on the elasticity α of approximately $0.02\frac{\text{MWh}}{\text{USD}} = 20\frac{\text{kWh}}{\text{USD}}$ at first glance seems to be an unachievable goal. However, note that wholesale prices can spike at up to $1000\frac{\text{USD}}{\text{MWh}}$, which is far outside the range of our calculations. Further, we disregarded transmission losses and capacity costs inherent to generators and utilities, which make the delivery of electricity under the forward contract and the call option more expensive, thereby lowering the bound on α.

4.7 Conclusion

We analyzed hedging instruments for load-serving entities to mitigate price risks associated with volatile energy supply and demand. Hedging against such risks is motivated by the fact that load-serving entities are obligated to meet energy demand of customers under contract instantaneously, which, in the absence of any hedging instruments, has to be procured in its entirety from the wholesale electricity market (at potentially high prices). Forward contracts and call options between load-serving entities and generating companies as well as Demand Response programs for end-users are methods to share this risk with other market participants. We formulated the optimal hedging strategy as a profit maximization problem which is random in the aggregate demand and wholesale electricity price. The optimal expected profit under each hedging instrument was found to be monotonically decreasing in the statistical dispersion of the demand distribution, and linearly decreasing for the special case of a uniform distribution. Using smart meter consumption data and locational marginal prices in California, we compared the optimal expected profits between the hedging methods in a pairwise fashion to generate decision boundaries of equal profit.

Our results can be extended in several regards. Firstly, a more involved analysis that takes into account operational constraints of the smart grid, e.g. transmission capacities and grid congestion, would add credibility to the suggestions of this chapter. Secondly, analyzing how the optimal expected profit increases as a function of diminished uncertainty in electric wholesale prices and aggregate consumer demand due to forecasting is interesting from the perspective of profit maximization. Lastly, forgoing Assumptions 6 and 7 allows utilities and generating companies to exercise market power. This calls for a game-theoretic formulation of the profit-maximization problem from the perspective of both generating companies and utilities, where each player seeks bids from the other in a mechanism design framework.

Chapter 5

Budget-Constrained Combinatorial Multi-Armed Bandits

5.1 Introduction

Chapters 2, 3, and 4 modeled the effect of *repeated* interactions between various actors in the electricity market. The purpose of this chapter is to formalize these sequential actions into a more general framework that "learns" which actions are the "best" over time.

The multi-armed bandit (MAB) problem has been extensively studied in machine learning and statistics as a means to model online sequential decision making. In the classic setting popularized by Auer, Cesa-Bianchi, and Fischer 2002 and Auer et al. 2002, the decision-maker selects exactly one arm at a given round t, given the observations of realized rewards from arms played in previous rounds $1, \ldots, t-1$. The goal is to maximize the cumulative reward over a fixed horizon T, or equivalently, to minimize regret, which is defined as the difference between the cumulative gain achieved, had the decision-maker always played the best arm, and the realized cumulative gain. The analysis of this setting reflects the fundamental tradeoff between the desire to learn better arms (exploration) and the possibility to play arms believed to have high payoff (exploitation).

A variety of practical applications of the MAB problem include placement of online advertising to maximize the click-through rate, in particular online sponsored search auctions (Rusmevichientong and Williamson 2005) and ad-exchange platforms (Chakraborty et al. 2010), channel selection in radio networks (Huang, Liu, and Ding 2008), or learning to rank web documents (Radlinski, Kleinberg, and Joachims 2008). As acknowledged by Ding et al. 2013, taking an action (playing an arm) in practice is inherently costly, yet the vast majority of existing bandit-related work used to analyze such examples forgoes any notion of cost. Furthermore, the above-mentioned applications rarely proceed in a strictly sequential way. A more realistic scenario is a setting in which, at each round, *multiple* actions are taken among the set of all possible choices.

These two shortcomings motivate the theme of this chapter, as we investigate the MAB

problem under a budget constraint in a setting with time-varying rewards and costs and multiple plays. More precisely, given an a-priori defined budget B, at each round the decision maker selects a combination of K distinct arms from N available arms and observes the individual costs and rewards, which corresponds to the *semi-bandit* setting. The player pays for the materialized costs until the remaining budget is exhausted, at which point the algorithm terminates and the cumulative reward is compared to the theoretical optimum and defines the weak regret, which is the expected difference between the payout under the best fixed choice of arms for all rounds and the actual gain. In this chapter, we investigate both the stochastic and the adversarial case. For the stochastic case, we derive an upper bound on the expected regret of order $O(NK^4 \log B)$, utilizing Algorithm `UCB-MB` inspired by the upper confidence bound algorithm `UCB1` first introduced by Auer, Cesa-Bianchi, and Fischer 2002. For the adversarial case, Algorithm `Exp3.M.B` upper and lower-bounds the regret with $O(\sqrt{NB \log(N/K)})$ and $\Omega((1 - K/N)^2 \sqrt{NB/K})$, respectively. These findings extend existing results from Uchiya, Nakamura, and Kudo 2010 and Auer et al. 2002, as we also provide an upper bound that holds with high probability. To the best of our knowledge, this is the first case that addresses the *adversarial* budget-constrained case, which we therefore consider to be the main contribution of this chapter.

5.1.1 Related Work

In the extant literature, attempts to make sense of a cost component in MAB problems occur in Tran-Thanh et al. 2010 and Tran-Thanh et al. 2012, who assume *time-invariant* costs and cast the setting as a knapsack problem with only the rewards being stochastic. In contrast, Ding et al. 2013 proposed algorithm `UCB-BV`, where per-round costs and rewards are sampled in an IID fashion from unknown distributions to derive an upper bound on the regret of order $O(\log B)$. The papers that are closest to our setting are Badanidiyuru, Kleinberg, and Slivkins 2013 and Xia et al. 2016. The former investigates the stochastic case with a resource consumption. Unlike our case, however, the authors allow for the existence of a "null arm", which is tantamount to skipping rounds, and obtain an upper bound of order $O(\sqrt{B})$ rather than $O(\log B)$ compared to our case. The latter paper focuses on the stochastic case, but does not address the adversarial setting at all.

Slightly less relevant settings to the setup investigates in this thesis are found in the best arm identification problem (Audibert, Bubeck, and Munos 2010), (Gabillon, Ghavamzadeh, and Lazaric 2012), where, given a budget, the goal is to identify the best arm rather than maximizing the cumulative reward. A different, rather experimental line of work is concerned with budgeted learning (Deng et al. 2007). Lastly, Seldin et al. 2014 discusses MABs with paid observations, where the player can query the rewards of any number of arms for an additional payment, even for the arm that has not been played at any given round.

The extension of the single play to the multiple plays case, where at each round $K \geq 1$ arms have to be played, was introduced in Anantharam, Varaiya, and Walrand 1986 and Agrawal, Hegde, and Teneketzis 1990. However, their analysis is based on the original

bandit formulation introduced by Lai and Robbins 1985, where the regret bounds only hold asymptotically (in particular not for a finite time), rely on hard-to-compute index policies, and are distribution-dependent. Influenced by the works of Auer, Cesa-Bianchi, and Fischer 2002 and Agrawal 2002, who popularized the usage of easy-to-compute upper confidence bounds (UCB), a recent line of work has further investigated the combinatorial bandit setting. For example, Gai, Krishnamachari, and Jain 2012 derived an $O(NK^4 \log T)$ regret bound in the stochastic semi-bandit setting, utilizing a policy they termed "Learning with Linear Rewards" (LLR). Similarly, Chen, Wang, and Yuan 2013 utilize a framework where the decision-maker queries an oracle that returns a fraction of the optimal reward. Other, less relevant settings to this chapter are found in Cesa-Bianchi and Lugosi 2009 and later Combes et al. 2015, who consider the adversarial bandit setting, where only the sum of losses for the selected arms can be observed. Furthermore, Kale, Reyzin, and Schapire 2010 investigate bandit slate problems to take into account the ordering of the arms selected at each round. Lastly, Komiyama, Honda, and Nakagawa 2015 utilize Thompson Sampling to model the stochastic MAB problem.

5.2 Main Results

In this section, we formally define the budgeted, multiple play multi-armed bandit setup and present the main theorems, whose results are provided in Table 5.1 together with a comparison to existing results in the literature. We first describe the stochastic setting (Section 5.3) and then proceed to the adversarial one (Section 5.4). Illuminating proofs for the theorems in this chapter are presented in Section 5.5. Technical proofs are relegated to Appendix A.4 of this thesis.

Algorithm	Upper Bound	Lower Bound	Authors
`Exp3`	$O\left(\sqrt{NT \log N}\right)$	$\Omega\left(\sqrt{NT}\right)$	Auer et al. 2002
`Exp3.M`	$O\left(\sqrt{NTK \log \frac{N}{K}}\right)$	$\Omega\left(\left(1-\frac{K}{N}\right)^2 \sqrt{NT}\right)$	Uchiya, Nakamura, and Kudo 2010
`Exp3.M.B`	$O\left(\sqrt{NB \log \frac{N}{K}}\right)$	$\Omega\left(\left(1-\frac{K}{N}\right)^2 \sqrt{NB/K}\right)$	This thesis
`Exp3.P`	$O\left(\sqrt{NT \log\left(NT/\delta\right)} + \log(NT/\delta)\right)$		Auer et al. 2002
`Exp3.P.M`	$O\left(K^2 \sqrt{NT \frac{N-K}{N-1} \log\left(NT/\delta\right)} + \frac{N-K}{N-1} \log(NT/\delta)\right)$		This thesis
`Exp3.P.M.B`	$O\left(K^2 \sqrt{\frac{NB}{K} \frac{N-K}{N-1} \log\left(\frac{NB}{K\delta}\right)} + \frac{N-K}{N-1} \log\left(\frac{NB}{K\delta}\right)\right)$		This thesis
`UCB1`	$O(N \log T)$		Auer, Cesa-Bianchi, and Fischer 2002
`LLR`	$O(NK^4 \log T)$		Gai, Krishnamachari, and Jain 2012
`UCB-BV`	$O(N \log B)$		Ding et al. 2013
`UCB-MB`	$O(NK^4 \log B)$		This thesis

Table 5.1: Regret Bounds in Adversarial and Stochastic Bandit Settings

5.3 Stochastic Setting

The definition of the stochastic setting is based on the classic setup introduced in Auer, Cesa-Bianchi, and Fischer 2002, but is enriched by a cost component and a multiple play constraint. Specifically, given a bandit with N distinct arms, each arm indexed by $i \in [N]$ is associated with an unknown reward and cost distribution with unknown means $0 < \mu_r^i \leq 1$ and $0 < c_{\min} \leq \mu_c^i \leq 1$, respectively. Realizations of costs $c_{i,t} \in [c_{\min}, 1]$ and rewards $r_{i,t} \in [0, 1]$ are independently and identically distributed. At each round t, the decision maker plays exactly K arms ($1 \leq K \leq N$) and subsequently observes the individual costs and rewards only for the played arms, which corresponds to the *semi-bandit* setting. Before the game starts, the player is given a budget $0 < B \in \mathbb{R}_+$ to pay for the materialized costs $\{c_{i,t} \mid i \in a_t\}$, where a_t denotes the indexes of the K arms played at time t. The game terminates as soon as the sum of costs at round t, namely $\sum_{j \in a_t} c_{j,t}$ exceeds the remaining budget.

Notice the minimum $c_{\min}$ on the support of the cost distributions. This assumption is not only made for practical reasons, as many applications of bandits come with a minimum cost, but also to guarantee well-defined "bang-per-buck" ratios $\mu^i = \mu_r^i / \mu_c^i$, which our analysis in this thesis relies on.

The goal is to design a deterministic algorithm $\mathcal{A}$ such that the expected payout $\mathbb{E}\left[G_{\mathcal{A}}(B)\right]$ is maximized, given the budget and multiple play constraints. Formally:

$$\begin{aligned} \underset{a_1,\ldots,a_{\tau_{\mathcal{A}}(B)}}{\text{maximize}} \quad & \mathbb{E}\left[\sum_{t=1}^{\tau_{\mathcal{A}}(B)} \sum_{i \in a_t} r_{i,t}\right] \\ \text{subject to} \quad & \mathbb{E}\left[\sum_{t=1}^{\tau_{\mathcal{A}}(B)} \sum_{i \in a_t} c_{i,t} \leq B\right] \\ & |a_t| = K,\ 1 \leq K \leq N\ \forall\ t \in [\tau_{\mathcal{A}}(B)] \end{aligned} \tag{5.1}$$

In (5.1), $\tau_{\mathcal{A}}(B)$ is the stopping time of algorithm $\mathcal{A}$ and indicates after how many steps the algorithm terminates, namely when the budget is exhausted. The expectation is taken over the randomness of the reward and cost distributions.

The performance of algorithm $\mathcal{A}$ is evaluated on its expected regret $\mathcal{R}_{\mathcal{A}}(B)$, which is defined as the difference between the expected payout (gain) $\mathbb{E}[G_{\mathcal{A}^*}]$ under the optimal strategy $\mathcal{A}^*$ (which in each round plays a^*, namely the set of K arms with the largest bang-per-buck ratios) and the expected payout $\mathbb{E}[G_{\mathcal{A}}]$ under algorithm $\mathcal{A}$:

$$\mathcal{R}_{\mathcal{A}}(B) = \mathbb{E}[G_{\mathcal{A}^*}(B)] - \mathbb{E}[G_{\mathcal{A}}(B)]. \tag{5.2}$$

Our main result in Theorem 17 upper bounds the regret achieved with Algorithm 17. Similar to Auer, Cesa-Bianchi, and Fischer 2002 and Ding et al. 2013, we maintain time-varying upper confidence bounds $U_{i,t}$ for each arm i

$$U_{i,t} = \bar{\mu}_t^i + e_{i,t}, \tag{5.3}$$

where $\bar{\mu}_t^i$ denotes the sample mean of the observed bang-per-buck ratios up to time t and $e_{i,t}$ the exploration term defined in Algorithm 17. At each round, the K arms associated with the K largest confidence bounds are played. For initialization purposes, we allow all N arms to be played exactly once prior to the while-loop.

Theorem 17. *There exist constants c_1, c_2, and c_3, which are functions of $N, K, c_{\min}, \Delta_{\min}, \mu_i, \mu_c$ only, such that Algorithm 1 (`UCB-MB`) achieves expected regret*

$$\mathcal{R}_{\mathcal{A}}(B) \leq c_1 + c_2 \log(B + c_3) = O(NK^4 \log B). \tag{5.4}$$

In Theorem 17, $\Delta_{\min}$ denotes the smallest possible difference of bang-per-buck ratios among non-optimal selections $a \neq a^*$, i.e. the second best choice of arms:

$$\Delta_{\min} = \sum_{j \in a^*} \mu^j - \max_{a, a \neq a^*} \sum_{j \in a} \mu^j. \tag{5.5}$$

Similarly, the proof of Theorem 17 also relies on the largest such difference $\Delta_{\max}$, which corresponds to the worst possible choice of arms:

$$\Delta_{\max} = \sum_{j \in a^*} \mu^j - \min_{a, a \neq a^*} \sum_{j \in a} \mu^j. \tag{5.6}$$

Comparing the bound given in Theorem 17 to the results in Table 5.1, we recover the $O(N \log B)$ bound from Ding et al. 2013 for the single-play case.

Algorithm 1 `UCB-MB` for Stochastic MAB

Initialize: $t = 1$. Play all arms together exactly once. Let $\bar{\mu}_{r,1}^i = r_{i,1}$, $\bar{\mu}_{c,1}^i = c_{i,1}$, $\bar{\mu}_1^i = \frac{\bar{\mu}_{r,1}^i}{\bar{\mu}_{c,1}^i}$ $\forall\ i \in [N]$, $n_{i,1} = 1$, $e_{i,1} = 0\ \forall\ i \in [N]$, $G_{\mathcal{A}} = 0$.

while true **do**
  $a_t \leftarrow$ Indexes of K arms with K largest $U_{i,t}$.
  if $\sum_{j \in a_t} c_{j,t} > B$ **then**
    return Gain $G_{\mathcal{A}}$, stopping time $\tau_{\mathcal{A}}(B) = t$
  end if
  $G_{\mathcal{A}} \leftarrow G_{\mathcal{A}} + \sum_{i \in a_t} r_{i,t}, \quad B \leftarrow B - \sum_{i \in a_t} c_{i,t}$
  $n_{i,t} \leftarrow n_{i,t} + 1 \quad \forall\ i \in a_t$
  $t \leftarrow t + 1$
  $e_{i,t} \leftarrow \frac{\sqrt{(K+1)\log t / n_{i,t}}(1 + 1/c_{\min})}{c_{\min} - \sqrt{(K+1)\log t / n_{i,t}}}$
end while

5.4 Adversarial Setting

We now consider the adversarial case that makes no assumptions on the reward and cost distributions whatsoever. The setup for this case was first proposed and analyzed by Auer

et al. 2002 for the single play case (i.e. $K = 1$), a fixed horizon T, and an oblivious adversary. That is, the entire seqence of rewards for all arms is fixed in advance and in particular cannot be adaptively changed during runtime. The proposed randomized algorithm `Exp3` enjoys $O(\sqrt{NT \log N})$ regret. Under *semi-bandit* feedback, where the rewards for a given round are observed for each arm played, Uchiya, Nakamura, and Kudo 2010 derived a variation of the single-play `Exp3` algorithm, which they called `Exp3.M` and enjoys regret $O\left(\sqrt{NTK \log(N/K)}\right)$, where K is the number of plays per round.

We consider the extension of the classic setting as in Uchiya, Nakamura, and Kudo 2010, where the decision maker has to play exactly $1 \leq K \leq N$ arms. For each arm i played at round t, the player observes the reward $r_i(t) \in [0, 1]$ and, unlike in previous settings, additionally the cost $0 < c_{\min} < c_i(t) < 1$. As in the stochastic setting (Section 5.3), the player is given a budget $B > 0$ to pay for the costs incurred, and the algorithm terminates after $\tau_{\mathcal{A}}(B)$ rounds when the sum of materialized costs in round $\tau_{\mathcal{A}}(B)$ exceeds the remaining budget. The gain $G_{\mathcal{A}}(B)$ of algorithm $\mathcal{A}$ is the sum of observed rewards up to and including round $\tau_{\mathcal{A}}(B) - 1$. The expected regret $\mathcal{R}_{\mathcal{A}}(B)$ is defined as in (5.2), where the gain of algorithm $\mathcal{A}$ is compared against the best set of arms that an omniscient algorithm $\mathcal{A}^*$, which knows the reward and cost sequences in advance, would select, given the budget B. In contrast to the stochastic case, the expectation is now taken with respect to algorithm $\mathcal{A}$'s internal randomness.

5.4.1 Upper Bounds on the Regret

We begin with upper bounds on the regret for the budget constrained MAB with multiple plays and later transition towards lower bounds and upper bounds that hold with high probability. Algorithm 2, which we call `Exp3.M.B`, provides a randomized algorithm to achieve sublinear regret. Similar to the original `Exp3` algorithm developed by Auer et al. 2002, Algorithm `Exp3.M.B` maintains a set of time-varying weights $\{w_i(t)\}_{i=1}^N$ for all arms, from which the probabilities for each arm being played at time t are calculated (line 10). As noted in Uchiya, Nakamura, and Kudo 2010, the probabilities $\{p_i(t)\}_{i=1}^N$ sum to K (because exactly K arms need to be played), which requires the weights to be capped at a value $v_t > 0$ (line 3) such that the probabilities $\{p_i(t)\}_{i=1}^N$ are kept in the range $[0, 1]$. In each round, the player draws a set of distinct arms a_t of cardinality $|a_t| = K$, where each arm has probability $p_i(t)$ of being included in a_t (line 11). This is done by employing algorithm `DependentRounding` introduced by Gandhi, Khuller, and Parthasarathy 2006, which runs in $O(K)$ time and $O(N)$ space. At the end of each round, the observed rewards and costs for the played arms are turned into estimates $\hat{r}_i(t)$ and $\hat{c}_i(t)$ such that $\mathbb{E}[\hat{r}_i(t) \mid a_t, \ldots, a_1] = r_i(t)$ and $\mathbb{E}[\hat{c}_i(t) \mid a_t, \ldots, a_1] = c_i(t)$ for $i \in a_t$ (line 16). Arms with $w_i(t) < v_t$ are updated according to $(\hat{r}_i(t) - \hat{c}_i(t))$, which assigns larger weights as $\hat{r}_i(t)$ increases and $\hat{c}_i(t)$ decreases, as one might expect.

Algorithm 2 `Exp3.M.B`: Budget Constrained Multi-Armed Bandit, Multiple Play, Adversarial

Initialize: $w_i = 1$ for $i \in [N]$, gain $G_{\mathcal{A}} = 0$.

while $B > 0$ **do**
  if $\arg\max_{i\in[N]} w_i(t) \geq \left(\frac{1}{K} - \frac{\gamma}{N}\right) \sum_{j=1}^{N} \frac{w_i(t)}{1-\gamma}$ **then**
    Determine v_t as follows: $1/K - \gamma/N =$

$$\frac{v_t(1-\gamma)}{\sum_{i=1}^{N} v_t \cdot \mathbb{1}(w_i(t) \geq v_t) + w_i(t) \cdot \mathbb{1}(w_i(t) < v_t)}$$

    Define set $\tilde{S}(t) = \{i \in [N] \mid w_i(t) \geq v_t\}$.
    Define weights $\tilde{w}_i(t) = v_t$ for $i \in \tilde{S}(t)$.
  else
    Define set $\tilde{S}(t) = \{\}$.
  end if
  Define weights $\tilde{w}_i(t) = w_i(t)$ for $i \in [N] \setminus \tilde{S}(t)$.
  Calculate probabilities for each $i \in [N]$:

$$p_i(t) = K\left((1-\gamma)\frac{\tilde{w}_i(t)}{\sum_{j=1}^{N} \tilde{w}_j(t)} + \frac{\gamma}{N}\right).$$

  Play arms $a_t \sim p_1, \ldots, p_N$.
  if $\sum_{i\in a_t} c_i(t) > B$ **then**
    return Gain $G_{\texttt{Exp3.M.B}}$, stopping time $\tau_{\mathcal{A}}(B) = t$
  end if
  $B \leftarrow B - \sum_{i\in a_t} c_i(t)$, $G_{\mathcal{A}} \leftarrow G_{\mathcal{A}} + \sum_{i\in a_t} r_i(t)$.
  Calculate estimated rewards and costs to update weights for each $i \in [N]$:

$$\hat{r}_i(t) = r_i(t)/p_i(t) \cdot \mathbb{1}(i \in a_t)$$
$$\hat{c}_i(t) = c_i(t)/p_i(t) \cdot \mathbb{1}(i \in a_t)$$
$$w_i(t+1) = w_i(t) \exp\left[\frac{K\gamma}{N}\left[\hat{r}_i(t) - \hat{c}_i(t)\right] \mathbb{1}_{i\in\tilde{S}(t)}\right]$$

end while

Theorem 18. *Algorithm* `Exp3.M.B` *achieves regret*

$$\mathcal{R} \leq 2.63\sqrt{1 + \frac{B}{gc_{\min}}}\sqrt{gN\log(N/K)} + K, \tag{5.7}$$

where g is an upper bound on $G_{\max}$, the maximal gain of the optimal algorithm. This bound

is of order $O(\sqrt{BN\log(N/K)})$.

The runtime of Algorithm `Exp3.M.B` and its space complexity is linear in the number of arms, i.e. $O(N)$. If no bound g on $G_{\max}$ exists, we have to modify Algorithm 2. Specifically, the weights are now updated as follows:

$$w_i(t+1) = w_i(t)\exp\left[\frac{K\gamma}{N}\left[\hat{r}_i(t) - \hat{c}_i(t)\right]\cdot \mathbb{1}_{i\in a_t}\right]. \tag{5.8}$$

This replaces the original update step in line 16 of Algorithm 2. As in Algorithm `Exp3.1` in Auer et al. 2002, we use an adaptation of Algorithm 2, which we call `Exp3.1.M.B`, see Algorithm 3. In Algorithm 3, we define cumulative expected gains and losses

$$\hat{G}_i(t) = \sum_{s=1}^{t}\hat{r}_i(s), \tag{5.9a}$$

$$\hat{L}_i(t) = \sum_{s=1}^{t}\hat{c}_i(s). \tag{5.9b}$$

and make the following, necessary assumption:

Assumption 11. $\sum_{i\in a} r_i(t) \geq \sum_{i\in a} c_i(t)$ *for all* $a \in \mathcal{S}$ *possible* K*-combinations and* $t \geq 1$.

Assumption 11 is a natural assumption, which is motivated by "individual rationality" reasons. In other words, a user will only play the bandit algorithm if the reward at any given round, for any possible choice of arms, is at least as large as the cost that incurs for playing. Under the caveat of this assumption, Algorithm `Exp3.1.M.B` utilizes Algorithm `Exp3.1.M` as a subroutine in each epoch until termination.

Algorithm 3 Algorithm `Exp3.1.M.B` with Budget B

Initialize: $t = 1$, $w_i = 1$ for $i \in [N]$, $r = 0$.

1: **while** $\sum_{t=1}^{T}\sum_{i\in a_t} c_i(t) \leq B$ **do**
2: Define $g_r = \frac{N\log(N/K)}{(e-1)-(e-2)c_{\min}}4^r$
3: Restart `Exp3.M.B` with $\gamma_r = \min(1, 2^{-r})$
4: **while** $\max_{a\in\mathcal{S}}\sum_{i\in a}(\hat{G}_i(t) - \hat{L}_i(t)) \leq g_r - \frac{N(1-c_{\min})}{K\gamma_r}$ **do**
5: Draw $a_t \sim p_1, \ldots, p_N$, observe $r_i(t)$ and $c_i(t)$ for $i \in a_t$, calculate $\hat{r}_i(t)$ and $\hat{c}_i(t)$.
6: $\hat{G}_i(t+1) \leftarrow \hat{G}_i(t) + \hat{r}_i(t)$ for $i \in [N]$
7: $\hat{L}_i(t+1) \leftarrow \hat{L}_i(t) + \hat{c}_i(t)$ for $i \in [N]$
8: $t \leftarrow t+1$
9: **end while**
10: **end while**
11: **return** Gain $G_{\texttt{Exp3.1.M.B}}$

Proposition 3. *For the multiple plays case with budget, the regret of Algorithm* ***Exp3.1.M.B*** *is upper bounded by*

$$\begin{aligned}\mathcal{R} \leq\ & 8\left[(e-1)-(e-2)c_{\min}\right]\frac{N}{K}+2N\log\frac{N}{K}+K+\\ & 8\sqrt{\left[(e-1)-(e-2)c_{\min}\right](G_{\max}-B+K)N\log(N/K)}.\end{aligned} \tag{5.10}$$

This bound is of order $O((G_{\max}-B)N\log(N/K))$ and, due to Assumption 11, not directly comparable to the bound in Theorem 18. One case in which (5.10) outperforms (5.7) occurs whenever only a loose upper bound of g on $G_{\max}$ exists or whenever $G_{\max}$, the return of the best selection of arms, is "small".

5.4.2 Lower Bound on the Regret

Theorem 19 provides a lower bound of order $\Omega((1-K/N)^2\sqrt{NB/K})$ on the weak regret of algorithm `Exp3.M.B`.

Theorem 19. *For $1 \leq K \leq N$, the weak regret $\mathcal{R}$ of Algorithm* ***Exp3.M.B*** *is lower bounded as follows:*

$$\mathcal{R} \geq \varepsilon\left(B-\frac{BK}{N}-2Bc_{\min}^{-3/2}\varepsilon\sqrt{\frac{BK\log(4/3)}{N}}\right), \tag{5.11}$$

where $\varepsilon \in (0, 1/4]$. Choosing ε as

$$\varepsilon = \min\left(\frac{1}{4},\ \frac{(1-K/N)c_{\min}^{3/2}}{4\sqrt{\log(4/3)}}\sqrt{\frac{N}{BK}}\right)$$

yields the bound

$$\mathcal{R} \geq \min\left(\frac{c_{\min}^{3/2}(1-K/N)^2}{8\sqrt{\log(4/3)}}\sqrt{\frac{NB}{K}},\ \frac{B(1-K/N)}{8}\right). \tag{5.12}$$

This lower bound differs from the upper bound given in Theorem 17 by a factor of $\sqrt{K\log(N/K)}(N/(N-K))^2$. For the single-play case $K = 1$, this factor is $\sqrt{\log N}$, which recovers the gap from Auer et al. 2002.

5.4.3 High Probability Upper Bounds on the Regret

For a fixed number of rounds (no budget considerations) and single play per round ($K = 1$), Auer et al. 2002 proposed Algorithm `Exp3.P` to derive the following upper bound on the

regret that holds with probability at least $1-\delta$:

$$G_{\max} - G_{\texttt{Exp3.P}} \leq 4\sqrt{NT\log\left(NT/\delta\right)} + 4\sqrt{\frac{5}{3}NT\log N} + 8\log\left(\frac{NT}{\delta}\right). \tag{5.13}$$

Theorem 20 extends the non-budgeted case to the multiple play case.

Theorem 20. *For the multiple play algorithm ($1 \leq K \leq N$) and a fixed number of rounds T, the following bound on the regret holds with probability at least $1-\delta$:*

$$\begin{aligned}\mathcal{R} &= G_{\max} - G_{\textit{Exp3.P.M}} \\ &\leq 2\sqrt{5}\sqrt{NKT\log(N/K)} + 8\frac{N-K}{N-1}\log\left(\frac{NT}{\delta}\right) \\ &\quad + 2(1+K^2)\sqrt{NT\frac{N-K}{N-1}\log\left(\frac{NT}{\delta}\right)}.\end{aligned} \tag{5.14}$$

For $K=1$, (5.14) recovers (5.13) save for the constants, which is due to a better ε-tuning in this thesis compared to Auer et al. 2002. Agreeing with intuition, this upper bound becomes zero for the edge case $K \equiv N$.

Theorem 20 can be derived by using a modified version of Algorithm 2, which we name `Exp3.P.M`. The necessary modifications to `Exp3.M.B` are motivated by Algorithm `Exp3.P` in Auer et al. 2002 and are provided in the following:

- Replace the outer while loop with **for** $t = 1, \ldots, T$ **do**
- Initialize parameter α:

$$\alpha = 2\sqrt{(N-K)/(N-1)\log\left(NT/\delta\right)}.$$

- Initialize weights w_i for $i \in [N]$:

$$w_i(1) = \exp\left(\alpha\gamma K^2\sqrt{T/N}/3\right).$$

- Update weights for $i \in [N]$ as follows:

$$w_i(t+1) = w_i(t)\exp\left[\mathbb{1}_{i\notin\tilde{S}(t)}\frac{\gamma K}{3N}\left(\hat{r}_i(t) + \frac{\alpha}{p_i(t)\sqrt{NT}}\right)\right]. \tag{5.15}$$

Since there is no notion of cost in Theorem 20, we do not need to update any cost terms.

Lastly, Theorem 21 extends Theorem 20 to the budget constrained setting using algorithm `Exp3.P.M.B`.

Theorem 21. *For the multiple play algorithm ($1 \leq K \leq N$) and the budget $B > 0$, the following bound on the regret holds with probability at least $1 - \delta$:*

$$\begin{aligned}\mathcal{R} &= G_{\max} - G_{\texttt{Exp3.P.M.B}} \\ &\leq 2\sqrt{3}\sqrt{\frac{NB(1-c_{\min})}{c_{\min}}\log\frac{N}{K}} + 4\sqrt{6}\frac{N-K}{N-1}\log\left(\frac{NB}{Kc_{\min}\delta}\right) \\ &\quad + 2\sqrt{6}(1+K^2)\sqrt{\frac{N-K}{N-1}\frac{NB}{Kc_{\min}}\log\left(\frac{NB}{Kc_{\min}\delta}\right)}.\end{aligned} \tag{5.16}$$

To derive bound (5.16), we again modify the following update rules in Algorithm 2 to obtain Algorithm `Exp3.P.M.B`:

- Initialize parameter α:

$$\alpha = 2\sqrt{6}\sqrt{(N-K)/(N-1)\log\left(NB/(Kc_{\min}\delta)\right)}.$$

- Initialize weights w_i for $i \in [N]$:

$$w_i(1) = \exp\left(\alpha\gamma K^2\sqrt{B/(NKc_{\min})}/3\right).$$

- Update weights for $i \in [N]$ as follows:

$$w_i(t+1) = w_i(t)\exp\left[\mathbb{1}_{i\notin\tilde{S}(t)}\frac{\gamma K}{3N}\left(\hat{r}_i(t) - \hat{c}_i(t) + \frac{\alpha\sqrt{Kc_{\min}}}{p_i(t)\sqrt{NB}}\right)\right].$$

The estimated costs $\hat{c}_i(t)$ are computed as $\hat{c}_i(t) = c_i(t)/p_i(t)$ whenever arm i is played at time t, as is done in Algorithm 2.

5.5 Proofs

5.5.1 Proof of Theorem 17

The proof of Theorem 17 is divided into two technical lemmas introduced in the following. Due to space constraints, the proofs are relegated to Appendix A.4.

First, we bound the number of times a non-optimal selection of arms is made up to stopping time $\tau_{\mathcal{A}}(B)$. For this purpose, let us define a counter $C_{i,t}$ for each arm i, initialized to zero for $t = 1$. Each time a non-optimal vector of arms is played, that is, $a_t \neq a^*$, we increment the smallest counter in the set a_t:

$$C_{j,t} \leftarrow C_{j,t} + 1, \quad j = \arg\min_{i\in a_t} C_{i,t}. \tag{5.17}$$

Ties are broken randomly. By definition, the number of times arm i has been played until time t is greater than or equal to its counter $C_{i,t}$. Further, the sum of all counters is exactly the number of suboptimal choices made so far:

$$n_{i,t} \geq C_{i,t} \quad \forall\, i \in [N],\ t \in [\tau_{\mathcal{A}}(B)].$$
$$\sum_{i=1}^{N} C_{i,t} = \sum_{\tau=1}^{t} \mathbb{1}(a_\tau \neq a^*) \quad \forall\, t \in [\tau_{\mathcal{A}}(B)].$$

Lemma 3 bounds the value of $C_{i,t}$ from above.

Lemma 3. *Upon termination of algorithm $\mathcal{A}$, there have been at most $O\left(NK^3 \log \tau_{\mathcal{A}}(B)\right)$ suboptimal actions. Specifically, for each $i \in [N]$:*

$$\mathbb{E}\left[C_{i,\tau_{\mathcal{A}}(B)}\right] \leq 1 + K\frac{\pi^2}{3} + (K+1)\left(\frac{\Delta_{\min} + 2K(1+1/c_{\min})}{c_{\min}\Delta_{\min}}\right)^2 \log \tau_{\mathcal{A}}(B).$$

Secondly, we relate the stopping time of algorithm $\mathcal{A}$ to the optimal action a^*:

Lemma 4. *The stopping time $\tau_{\mathcal{A}}$ is bounded as follows:*

$$\frac{B}{\sum_{i\in a^*} \mu_c^i} - c_2 - c_3 \log\left(c_1 + \frac{2B}{\sum_{i\in a^*} \mu_c^i}\right) \leq \tau_{\mathcal{A}} \leq \frac{2B}{\sum_{i\in a^*} \mu_c^i}\mu_c^i + c_1,$$

where c_1, c_2, and c_3 are the same positive constants as in Theorem 17 that depend only on $N, K, c_{\min}, \Delta_{\min}, \mu_c^i, \mu_r^i$.

Utilizing Lemmas 3 and 4 in conjunction with the definition of the weak regret (5.2) yields Theorem 17. See Appendix A.4 for further technicalities.

5.5.2 Proof of Theorem 18

The proof of Theorem 18 in influenced by the proof methods for Algorithms `Exp3` by Auer et al. 2002 and `Exp3.M` by Uchiya, Nakamura, and Kudo 2010. The main challenge is the absence of a well-defined time horizon T due to the time-varying costs. To remedy this problem, we define $T = \max\left(\tau_{\mathcal{A}}(B), \tau_{\mathcal{A}^*}(B)\right)$, which allows us to first express the regret as a function of T. In a second step, we relate T to the budget B.

5.5.3 Proof of Proposition 3

The proof of Proposition 3 is divided into the following two lemmas:

Lemma 5. *For any subset $a \in \mathcal{S}$ of K unique elements from $[N]$, $1 \leq K \leq N$:*

$$\sum_{t=S_r}^{T_r} \sum_{i\in a_t} (r_i(t) - c_i(t)) \geq \sum_{i\in a} \sum_{t=S_r}^{T_r} (\hat{r}_j(t) - \hat{c}_j(t)) - 2\sqrt{(e-1) - (e-2)c_{\min}}\sqrt{g_r N \log(N/K)}, \tag{5.18}$$

where S_r and T_r denote the first and last time step at epoch r, respectively.

Lemma 6. *The total number of epochs R is bounded by*

$$2^{R-1} \leq \frac{N(1-c_{\min})}{Kc} + \sqrt{\frac{\hat{G}_{\max} - \hat{L}_{\max}}{c}} + \frac{1}{2}, \tag{5.19}$$

where $c = \frac{N \log(N/K)}{(e-1)-(e-2)c_{\min}}$.

To derive Proposition 3, we combine Lemmas 5 and 6 and utilize the fact that algorithm `Exp3.1.M.B` terminates after $\tau_{\mathcal{A}}(B)$ rounds. See Appendix A.4 for details.

5.5.4 Proof of Theorem 19

The proof follows existing procedures for deriving lower bounds in adversarial bandit settings, see Auer et al. 2002, Cesa-Bianchi and Lugosi 2006. The main challenges are found in generalizing the single play setting to the multiple play setting ($K > 1$) as well as incorporating a notion of cost associated with bandits.

Select exactly K out of N arms at random to be the arms in the "good" subset a^*. For these arms, let $r_i(t)$ at each round t be Bernoulli distributed with bias $\frac{1}{2} + \varepsilon$, and the cost $c_i(t)$ attain $c_{\min}$ and 1 with probability $\frac{1}{2} + \varepsilon$ and $\frac{1}{2} - \varepsilon$, respectively, for some $0 < \varepsilon < 1/2$ to be specified later. All other $N - K$ arms are assigned rewards 0 and 1 and costs $c_{\min}$ and 1 independently at random. Let $\mathbb{E}_{a^*}[\cdot]$ denote the expectation of a random variable conditional on a^* as the set of good arms. Let $\mathbb{E}_u[\cdot]$ denote the expectation with respect to a uniform assignment of costs $\{c_{\min}, 1\}$ and rewards $\{0, 1\}$ to all arms. Lemma 7 is an extension of Lemma A.1 in Auer et al. 2002 to the multiple-play case with cost considerations:

Lemma 7. *Let $f : \{\{0,1\}, \{c_{\min}, 1\}\}^{\tau_{\max}} \to [0, M]$ be any function defined on reward and cost sequences $\{\mathbf{r}, \mathbf{c}\}$ of length less than or equal $\tau_{\max} = \frac{B}{Kc_{\min}}$. Then, for the best action set a^*:*

$$\mathbb{E}_{a^*}\left[f(\mathbf{r}, \mathbf{c})\right] \leq \mathbb{E}_u[f(\mathbf{r}, \mathbf{c})] + \frac{Bc_{\min}^{-3/2}}{2}\sqrt{-\mathbb{E}_u[N_{a^*}]\log(1 - 4\varepsilon^2)},$$

where N_{a^*} denotes the total number of plays of arms in a^* during rounds $t = 1$ through $t = \tau_{\mathcal{A}}(B)$, that is:

$$N_{a^*} = \sum_{t=1}^{\tau_{\mathcal{A}}(B)} \sum_{i \in a^*} \mathbb{1}\left(i \in a_t\right).$$

Lemma 7, whose proof is relegated to Appendix A.4, allows us to bound the gain under the existence of K optimal arms by treating the problem as a uniform assignment of costs and rewards to arms. The technical parts of the proof can also be found in Appendix A.4.

5.5.5 Proof of Theorem 20

The proof strategy is to acknowledge that Algorithm `Exp3.P.M` uses upper confidence bounds $\hat{r}_i(t) + \frac{\alpha}{p_i(t)\sqrt{NT}}$ to update weights (5.15). Lemma 8 asserts that these confidence bounds are valid, namely that they upper bound the actual gain with probability at least $1-\delta$, where $0 < \delta \ll 1$.

Lemma 8. *For* $2\sqrt{\frac{N-K}{N-1}\log\left(\frac{NT}{\delta}\right)} \leq \alpha \leq 2\sqrt{NT}$,

$$\mathbb{P}\left(\hat{U}^* > G_{\max}\right) \geq \mathbb{P}\left(\bigcap_{a\subset\mathcal{S}} \sum_{i\in a} \hat{G}_i + \alpha\hat{\sigma}_i > \sum_{i\in a} G_i\right) \geq 1-\delta,$$

where $a \subset \mathcal{S}$ *denotes an arbitrary subset of* K *unique elements from* $[N]$. $\hat{U}^*$ *denotes the upper confidence bound for the optimal gain.*

Next, Lemma 9 provides a lower bound on the gain of Algorithm `Exp3.P.M` as a function of the maximal upper confidence bound.

Lemma 9. *For* $\alpha \leq 2\sqrt{NT}$, *the gain of Algorithm* ***`Exp3.P.M`*** *is bounded below as follows:*

$$G_{\texttt{Exp3.P.M}} \geq \left(1 - \frac{5}{3}\gamma\right)\hat{U}^* - \frac{3N}{\gamma}\log(N/K) - 2\alpha^2 - \alpha(1+K^2)\sqrt{NT}, \tag{5.20}$$

where $\hat{U}^* = \sum_{j\in a^*} \hat{G}_j + \alpha\hat{\sigma}_j$ *denotes the upper confidence bound of the optimal gain achieved with optimal set* a^*.

Therefore, combining Lemmas 8 and 9 upper bounds the actual gain of Algorithm `Exp3.P.M` with high probability. See Appendix A.4 for technical details.

5.5.6 Proof of Theorem 21

The proof of Theorem 21 proceeds in the same fashion as in Theorem 20. Importantly, the upper confidence bounds now include a cost term. Lemma 10 is the equivalent to Lemma 8 for the budget constrained case:

Lemma 10. *For* $2\sqrt{6}\sqrt{\frac{N-K}{N-1}\log\frac{NB}{Kc_{\min}\delta}} \leq \alpha \leq 12\sqrt{\frac{NB}{Kc_{\min}}}$,

$$\mathbb{P}\left(\hat{U}^* > G_{\max} - B\right) \geq \mathbb{P}\left(\bigcap_{a\subset\mathcal{S}} \sum_{i\in a} \hat{G}_i - \hat{L}_i + \alpha\hat{\sigma}_i > \sum_{i\in a} G_i - L_i\right) \geq 1-\delta,$$

where $a \subset \mathcal{S}$ *denotes an arbitrary time-invariant subset of* K *unique elements from* $[N]$. $\hat{U}^*$ *denotes the upper confidence bound for the cumulative optimal gain minus the cumulative*

cost incurred after $\tau_a(B)$ rounds (the stopping time when the budget is exhausted):

$$a^* = \max_{a \in \mathcal{S}} \sum_{t=1}^{\tau_a(B)} (r_i(t) - c_i(t)),$$
$$\hat{U}^* = \sum_{i \in a^*} \left(\alpha \hat{\sigma}_i + \sum_{t=1}^{\tau_{a^*}(B)} (\hat{r}_i(t) - \hat{c}_i(t)) \right). \tag{5.21}$$

In Lemma 10, $G_{\max}$ denotes the optimal cumulative reward under the optimal set a^* chosen in (5.21). $\hat{G}_i$ and $\hat{L}_i$ denote the cumulative expected reward and cost of arm i after exhaustion of the budget (that is, after $\tau_a(B)$ rounds), respectively.

Lastly, Lemma 11 lower bounds the actual gain of Algorithm `Exp3.P.M.B` as a function of the upper confidence bound (5.21).

Lemma 11. *For $\alpha \leq 2\sqrt{\frac{NB}{Kc_{\min}}}$, the gain of Algorithm `Exp3.P.M.B` is bounded below as follows:*

$$G_{\texttt{Exp3.P.M.B}} \geq \left(1 - \gamma - \frac{2\gamma}{3} \frac{1 - c_{\min}}{c_{\min}}\right) \hat{U}^* - \frac{3N}{\gamma} \log \frac{N}{K} - 2\alpha^2 - \alpha(1 + K^2) \frac{BN}{Kc_{\min}}.$$

Combining Lemmas 10 and 11 completes the proof, see the appendix.

5.6 Discussion and Conclusion

We discussed the budget-constrained multi-armed bandit problem with N arms, K multiple plays, and an a-priori defined budget B. We explored the stochastic as well as the adversarial case and provided algorithms to derive regret bounds in the budget B. For the stochastic setting, our algorithm `UCB-MB` enjoys regret $O(NK^4 \log B)$. In the adversarial case, we showed that algorithm `Exp3.M.B` enjoys an upper bound on the regret of order $O(\sqrt{NB \log(N/K)})$ and a lower bound $\Omega((1 - K/N)^2 \sqrt{NB/K})$. Lastly, we derived upper bounds that hold with high probability.

Our work can be extended in several dimensions in future research. For example, the incorporation of a budget constraint in this thesis leads us to believe that a logical extension is to integrate ideas from economics, in particular mechanism design, into the multiple plays setting (one might think about auctioning off multiple items simultaneously), c.f. Babaioff, Sharma, and Slivkins (2009). A possible idea is to investigate to what extent the regret varies as the number of plays K increases. Further, it would not be a stretch to imagine that in such settings, repeated interactions with customers (playing arms) give rise to strategic considerations, in which customers can misreport their preferences in the first few rounds to maximize their long-run surplus. While the works of Amin, Rostamizadeh, and Syed 2013 and Mohri and Munoz 2014 investigate repeated interactions with a single player only, we believe an extension to a pool of buyers is worth exploring. In this setting, we would

expect that the extent of strategic behavior decreases as the number of plays K in each round increases, since the decision-maker could simply ignore users in future rounds who previously declined offers.

Part II

Applications

Chapter 6

HVAC Temperature Control in Commercial Buildings

6.1 Introduction

Part I of this thesis was concerned with developing a theoretical framework for the interaction of agents in the smart grid. Part II, however, focuses on practical applications, namely the investigation of dynamic loads, forecasting smater meter data, Demand Response, and temperature control in commercial buildings. This chapter intends to shed light on controlling the interior temperature of commercial buildings in a way that maximizes energy efficiency while respecting comfort constraints of occupants.

According to Pérez-Lombard, Ortiz, and Pout 2008, residential and commercial buildings account for up to 40% of the total electricity consumption in developed countries, with an upward trend. Heating, ventilation and air-conditioning (HVAC) systems are a major source of this consumption (*U.S. Department of Energy Buildings Energy Data Book*; McQuade 2009). Nevertheless, their power consumption can be flexibly scheduled without compromising occupant comfort, due to the thermal capacity of buildings. As a result, HVAC systems have become the focal point of research, with the goal of utilizing this source of consumption flexibility. From the point of view of energy efficiency, researchers have studied optimization of building control in order to minimize power consumption (Ma, Anderson, and Borrelli 2011; Široky et al. 2011; Parisio et al. 2014). Further, by participating in the regulation of electricity's frequency, buildings can assist in supporting the supply quality of electricity and the grid stability (Balandat et al. 2014; Vrettos et al. 2014; Lin et al. 2015; Baccino et al. 2014).

In frequency regulation, there are several focal points of research based on the formulation and solution of optimization problems. First, Baccino et al. 2014 and Lin et al. 2015 have implemented the optimization scheme into the operation system of a building and experimentally tested it. Second, the robustness of frequency reserve provision has been studied by Vrettos et al. 2014, using Model Predictive Control (MPC). Third, the contract

design problem for the scenario when an aggregation of commercial buildings provide their total regulation capacity to the wholesale electricity market through an aggregator has been investigated with a game-theoretical approach by Balandat et al. 2014.

All of the above research activities are based on a valid mathematical model describing the thermal behavior of buildings, which is an essential part of controller design as well as *in silico* assessment of the performance of any proposed controller. Traditionally, buildings have been modeled with high-dimensional physics-based models such as resistance-capacitance (RC) models (Maasoumy et al. 2014; Hao et al. 2013), TRNSYS (Duffy et al. 2009), and EnergyPlus (Zhao, Lam, and Ydstie 2013). These models are motivated by the thermodynamics of the building and explicitly model the heat transfer between building components. The advantage of such models is their high granularity of temperature modeling, but a drawback is their high dimensionality, rendering them computationally expensive. Although there has been extensive work on model reduction, this remains to be a non-trivial task.

A large body of this work focuses on linear models, whereas physics-based models for commercial buildings with a variable air volume (VAV) HVAC system are bilinear in nature. Furthermore, existing model reduction techniques often result in a loss of interpretability of states (Dobbs 2012) and disproportionate increase in the model's prediction error (Goyal and Barooah 2012). This justifies the identification procedure of a purely data-driven, low-dimensional model, as it cannot be obtained from the reduction of physics-based models with acceptable prediction accuracy.

Motivated by these shortcomings, a new direction of research attempts to identify low-dimensional, data-driven models with Input-Output models (Lin et al. 2015) and semiparametric regression (Aswani et al. 2012). The intention is to alleviate the computational complexity in expense for coarser and less accurate temperature predictions.

A crucial question that arises within these two extremes is the extent to which the estimated temperature model is compatible with controller design. Take Model Predictive Control (MPC) for example, where the classical physics-based models require an MPC strategy to be solved online with high computational demand. Even then, the inherent bilinearity ensuing from the physics of the HVAC system often requires robustification, in the form of stochastic MPC formulations with chance constraints (Ma et al. 2012). In contrast, regression-based models provide convenient difference equations that are easy and fast to use for MPC. A logical question to ask is how lean a model can be for a reasonable control application, without trading off too much accuracy and granularity of the temperature predictions.

The contribution of this chapter is two-fold. First, we aim to improve existing data-driven model identification techniques. Unlike Radecki and Hencey 2012; Radecki and Hencey 2013, who model the evolution of the building's energy consumption without a specific control input, we identify a model for temperature evolution in multiple building zones amenable to control design, i.e. with airflows as inputs. Our model also differs from Aswani et al. 2012, which uses HVAC supply air temperature as the single control input, resulting in a simpler identification problem, but, on the other hand, offers less flexibility for control.

Second, and more importantly, we perform a *quantitative comparison* of data-driven

and physics-based models in terms of open-loop prediction accuracy and closed-loop control strategies, based on the *same testbed* (the entire floor of an office building) using *experimental data* collected from the building, as opposed to simulated data.

We conclude that a low-dimensional data-driven model is suitable for building control applications, such as frequency regulation, due to its minor loss of prediction accuracy compared to high-dimensional physics-based models, but significant gain in computational ease.

To the best of our knowledge, the extant body of literature has analyzed data-driven and physical models for the identification of temperature evolution in commercial buildings only in an isolated fashion (in particular not on the same testbed), (Široky et al. 2011; Hu et al. 2016). In addition, some of these models were identified for fictitious buildings with synthetic data (Cole, Hale, and Edgar 2013; Goyal, Ingley, and Barooah 2013), while others used experimental data collected under environments with little or no disturbance, e.g. without occupants (Lin et al. 2015). Our work differs from these existing works in two aspects. First, we use experimental data to identify models for a multi-zone commercial building under regular operation, which is subject to significant disturbances such as occupancy. Second, although the existing literature mentions the differences between data-driven and physics-based models, the prevailing isolationist approach does not provide any quantitative comparison with respect to building control applications - a gap we aim to fill by juxtaposing a data-driven with a physics-based model.

The remainder of this chapter is organized as follows: In Section 6.2, we describe the testbed and the experimental data. Section 6.3 presents the identification process for a purely data-driven model with semiparametric regression, followed by Section 6.4, which details the procedure for identifying a physics-based model. Section 6.5 then compares the performance of both models in terms of open-loop prediction accuracy and closed-loop energy efficient optimal control. We show that, despite the higher accuracy of the complex physics-based model, the optimal control strategies with respect to HVAC operation cost while maintaining the thermal comfort of occupants is almost identical for both systems. Section 6.6 concludes.

6.2 Preliminaries

6.2.1 Testbed for System Identification

We model the temperature evolution of the fourth floor of Sutardja Dai Hall (SDH), a building on the University of California, Berkeley campus. This floor contains offices for research staff and open workspaces for students, and is divided into six zones for modeling purposes (Figure 6.1).

SDH is equipped with a variable air volume (VAV) HVAC system, which consists of large supply fans driving air through heat exchangers, cooling it down to a desired supply air temperature (SAT), and then distribute air to VAV boxes located throughout the building. There are 21 VAV boxes located on the fourth floor that govern the airflow to each room. In

addition, the supply air may be reheated at the VAV box before entering the room. Table 6.1 provides information about the mapping of the 21 VAV boxes to the zones they serve.

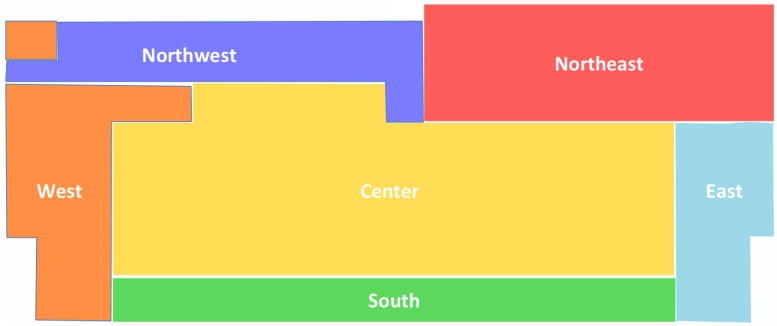

Figure 6.1: Zones for the 4th Floor of Sutardja Dai Hall (SDH)

Zone	VAV Boxes
Northwest (NW)	6, 8, 11
West (W)	1, 2, 3
South (S)	5, 9, 15, 16
East (E)	20, 21
Northeast (NE)	10, 14, 17, 19
Center (C)	4, 7, 12, 13, 18

Table 6.1: VAV Boxes by Zone

6.2.2 Collection of Experimental Data

We collected 51 weeks of one-minute resolution temperature data for the six zones along with the airflow rates of the 21 VAV boxes, SAT, and the outside air temperature. The hourly global horizontal solar radiation data was obtained from a nearby weather station (*CIMIS Station Reports* 2015), from which the incidence solar radiation of the four geographic directions was calculated with the `PV_LIB` toolbox (*PV Performance Modeling Collaborative*). All collected data were down-sampled or interpolated to 15 minute intervals.

These 51 weeks of data span periods when the building was under normal operation as well as periods with excitation experiments. For accurate parameter identification, temperatures of neighboring zones should not have strong correlation (Lin, Middelkoop, and Barooah 2012). For buildings in regular operation, this is generally possible through forced response experiments. Because of commercial buildings' large thermal inertia, each forced excitation must last sufficiently long before temperature changes are observed. With these points in mind, we conducted our experiment as follows: Starting at 8 am, every 2 hours, the air

inflow rate to one zone is set to its maximum value, minimum air flow rates are set for each of its neighboring zones and a random air flow rate is chosen for each remaining zone. This is repeated for each of the 6 zones. This experiment is performed during Saturdays to (a) minimize effects due to occupancy on our data, and thus facilitate subsequent parameter identification; (b) minimize disturbance to building operation (Hu et al. 2016).

For accurate parameter identification, temperatures of neighboring zones should not have strong correlation (Lin, Middelkoop, and Barooah 2012). Our testbed is a regular office building in operation, thus forced response experiments were performed during Saturdays to (a) increase identifiability of the building model; (b) minimize effects due to occupancy on our data, and thus facilitate subsequent parameter identification; (c) minimize disturbance to building operation (Hu et al. 2016).

6.2.3 Data Splitting

Next, we defined the seasons "fall" (early September - mid December), "winter" (mid December - late January), and "spring" (late January - mid May) to account for different occupancy levels during the fall and spring semesters and the winter break. A 90%-10% random split into training and test data was used for fitting and testing the models, respectively.

6.2.4 Notation

Let $\hat{(\cdot)}$ denote either the conditional expectation or the estimated value of a variable. Let $\tilde{(\cdot)}$ the predicted value of a variable, respectively.

6.3 Data-Driven Model

Using semiparametric regression, we identify a difference equation for the temperature evolution, for a lumped zone model and a multi-zone model of the fourth floor of SDH. Semiparametric regression in buildings has been proposed by Aswani et al. 2012, where the authors used only one week of historic data to model the temperature evolution and used the HVAC's supply air temperature as the single control input including an exogenous heating load that captures the effect of occupancy, electric devices, outside air temperature, and solar radiation.

We extend this approach by taking into account multiple weeks, which we separate into three seasons (fall, winter, spring) so as to characterize the different levels of the exogenous heating load for different temporal seasons. In addition, we model the room temperatures as a function of airflow rates from multiple VAVs to obtain a model which can be used for more sophisticated control strategies. We make use of cross-validation across all weeks to find the optimal model, therefore allowing for a more general analysis of the thermal behavior rather than restricting ourselves to identifying a model tailored to a single week, as is done in Aswani et al. 2012.

6.3.1 Lumped Zone

6.3.1.1 Model Setup

In order to facilitate analysis, the entire 4th floor of SDH is treated as a single zone, with the scalar temperature x corresponding to the average temperature on the entire floor and the input u as the sum of the inflow of all 21 VAV boxes. This lumped model assumes a uniform temperature on the entire floor, x, and has been commonly used in literature (Ma et al. 2012; Oldewurtel et al. 2010). Then, the temperature evolution is assumed to have the following form:

$$x(k+1) = ax(k) + bu(k) + c^\top v(k) + q_{\text{IG}}(k) + \epsilon(k), \tag{6.1}$$

where u denotes the total air inflow to the entire floor and $v := [v_{\text{Ta}}, v_{\text{Ts}}, v_{\text{solE}}, v_{\text{solN}}, v_{\text{solS}}, v_{\text{solW}}]^\top$ is a vector of known disturbances that describes ambient air temperature, the HVAC system's supply air temperature, and solar radiation from each of the four geographic directions. In addition, q_{IG} represents the internal gains due to occupancy and electric devices, and ϵ denotes independent and identically distributed zero mean noise with constant and finite variance which is conditionally independent of x, u, v, and q_{IG}. Finally, a, $b \in \mathbb{R}$ and $c \in \mathbb{R}^6$ are unknown coefficients to be estimated using semiparametric regression (Ruppert, Wand, and Carroll 2003).

6.3.1.2 Smoothing of Time Series

The q_{IG} term of Equation (6.1) is treated as a nonparametric term, so that (6.1) becomes a partially linear model (Härdle, Liang, and Gao 2000). By taking conditional expectations on both sides of (6.1), we obtain

$$\hat{x}(k+1) = a\hat{x}(k) + b\hat{u}(k) + c^\top \hat{v}(k) + \mathbb{E}\left[q_{\text{IG}}(k)|k\right] + \mathbb{E}\left[\epsilon(k)|k\right], \tag{6.2}$$

where the conditional expectations $\hat{x}(\cdot) = \mathbb{E}\left[x(\cdot)|\cdot\right]$, $\hat{u}(\cdot) = \mathbb{E}\left[u(\cdot)|\cdot\right]$, and $\hat{v}(\cdot) = \mathbb{E}\left[v(\cdot)|\cdot\right]$ are used. Noting that $\mathbb{E}\left[\epsilon(\cdot)|\cdot\right] = 0$ and assuming $\mathbb{E}\left[q_{\text{IG}}(\cdot)|\cdot\right] = q_{\text{IG}}(\cdot)$, subtracting (6.2) from (6.1) gives

$$x(k+1) - \hat{x}(k+1) = a\left(x(k) - \hat{x}(k)\right) + b\left(u(k) - \hat{u}(k)\right) + c^\top\left(v(k) - \hat{v}(k)\right) + \epsilon(k). \tag{6.3}$$

The unknown internal gains term has been eliminated, and thus the coefficients a, b, c in (6.3) can be estimated with any regression method. The conditional expectations $\hat{x}(\cdot), \hat{u}(\cdot)$ and $\hat{v}(\cdot)$ are obtained by smoothing the respective time series (Aswani et al. 2012). We made use of locally weighted linear regression with a tricube weight function, where we use k-fold cross-validation to determine the optimal kernel width that assigns weights ψ_i

$$\psi_i = \left(1 - \left|\frac{z - z_i}{d(z)}\right|^3\right)^3, \tag{6.4}$$

that belong to $z_i \in \mathcal{Z}$, which is a neighbor of the data point z to be smoothed along the abscissa within the span $\mathcal{Z}$ (all data points around z taken into account to smooth the time

series at z), and $d(z)$ the distance from z to the farthest predictor within $\mathcal{Z}$. The width $d(z)$ of the span $\mathcal{Z}$ is determined with k-fold cross-validation.

The error measure used for in-sample estimates is the *Root Mean Squared* (RMS) *Error* between the measured temperatures $\bar{x}(k)$ and the model's predicted temperatures $x(k)$ over a time horizon of N steps (e.g. we chose 24 hours, $N = 96$):

$$\text{RMS error} = \left(\frac{1}{N} \sum_{k=1}^{N} \left[\bar{x}(k) - x(k) \right]^2 \right)^{1/2}. \tag{6.5}$$

6.3.1.3 Bayesian Constrained Least Squares

A major challenge in identifying the model is that commercial buildings are often insufficiently excited. For example, SDH's room temperatures under regular operation only vary within a range of 2°C whereas inflow of the VAV boxes hardly varies at all. To overcome this, forced response experiments (Section 6.2.2) were conducted to compensate for the lack of excitation. Further, we use Bayesian regression, which allows prior knowledge of the building physics to be incorporated in the identification of coefficients. Specifically, Gaussian prior distributions are used for the coefficients a and b, i.e., $a \sim \mathcal{N}(\mu_a, \Sigma_a)$ and $b \sim \mathcal{N}(\mu_b, \Sigma_b)$, where $\mathcal{N}(\mu, \Sigma)$ denotes a jointly Gaussian distribution with mean μ and covariance matrix Σ. In addition, a, b and c are constrained to be identical for the different seasons, since they model the underlying physics of the building which are assumed to be invariant throughout the year.

Let $\mathcal{T} = \{1, 2, \cdots, N\}$ denote N weeks of training data. Let the set of training weeks from the fall season be represented by $\mathcal{F} = \{i \in \mathcal{T} \text{ such that } i \text{ is a week in fall}\}$. Similarly, define the sets of training weeks from the winter and spring as $\mathcal{W}$ and $\mathcal{S}$. The coefficient identification problem reads as follows:

$$\begin{aligned}
(\hat{a}, \hat{b}, \hat{c}) = & \arg\min_{a,b,c} \ (J_{\mathcal{F}} + J_{\mathcal{W}} + J_{\mathcal{S}}) + \|\Sigma_a^{-1/2}(a - \mu_a)\|^2 + \|\Sigma_b^{-1/2}(b - \mu_b)\|^2 \\
\text{s.t. } J_{\mathcal{X}} = & \sum_{i \in \mathcal{X}} \| x_i(k+1) - \hat{x}_i(k+1) - a\left(x_i(k) - \hat{x}_i(k)\right) \\
& - b\left(u_i(k) - \hat{u}_i(k)\right) - c^\top \left(v_i(k) - \hat{v}_i(k)\right) \|^2 \\
& \text{for } \mathcal{X} \in \{\mathcal{F}, \mathcal{W}, \mathcal{S}\}, \\
& 0 < a < 1, \ b \leq 0, \ c \geq 0.
\end{aligned} \tag{6.6}$$

where subscripts f, w, and s represent fall, winter and spring, respectively. $J_{\mathcal{F}}$, $J_{\mathcal{W}}$ and $J_{\mathcal{S}}$ denote the sum of squared errors between actual and predicted temperatures for fall, winter, and spring, respectively. The sign constraints on the parameters b and c capture the fact that temperature correlates positively with all components in v and negatively with VAV airflow. The range of a is a consequence of Newton's Law of Cooling.

To find the effect of the VAV inflow on the 15-minute temperature evolution, we computed the 15-minute incremental reductions in temperature Δx recorded during the excitation experiments. It is assumed that the large inflow u dominates all other effects such that we can assume $\Delta x = x(k+1) - x(k) = b \cdot u(k)$ for all k during the excitation period. The

estimated prior μ_b can then be isolated. The prior μ_a was set as the optimal $\hat{a}$ identified by (6.6) without the prior terms. The covariance matrices Σ_a and Σ_b were chosen subjectively.

6.3.1.4 Estimation of Internal Gains

With the estimated coefficients $\hat{a}, \hat{b}, \hat{c}$ in hand, the internal gains q_{IG} can be estimated by manipulating (6.2):

$$\hat{q}_{\text{IG}}(k) = \hat{x}(k+1) - \left(\hat{a}\hat{x}(k) + \hat{b}\hat{u}(k) + \hat{c}^\top \hat{v}(k)\right). \tag{6.7}$$

(6.7) is used to estimate an instance of the internal gains function for each week i in the training set $\mathcal{T}$. The q_{IG} for each season is defined as the average of estimated weekly gains for all weeks $i \in \mathcal{X}$ and $\mathcal{X} \in \{\mathcal{F}, \mathcal{W}, \mathcal{S}\}$.

6.3.1.5 Results

The estimated internal gains for each season are shown in Figure 6.2. Observe that, for

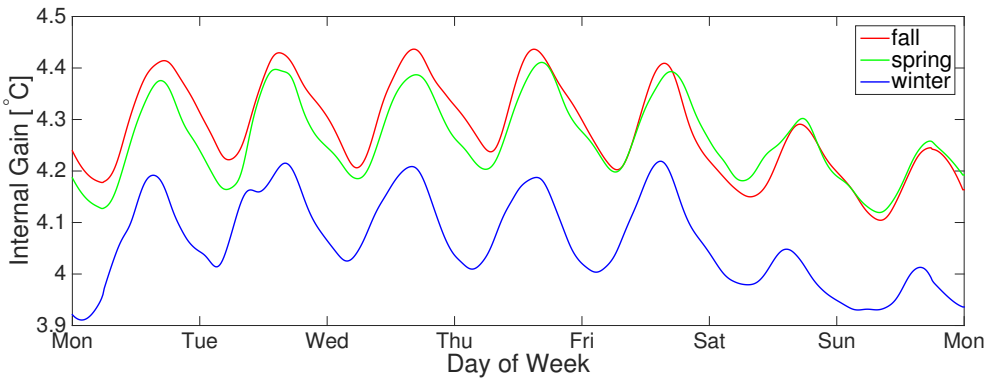

Figure 6.2: Estimated Internal Gain q_{IG} from the Data-Driven Model by Season, Lumped Case

all three seasons, the internal gains exhibit a daily trend with local peaks around the late afternoon and local minima at night. Moreover, the amplitudes of the internal gains are considerably smaller during weekends, suggesting a lighter occupancy. It can further be seen that the magnitude of the internal gains is smallest for the winter season, which is consistent with intuition as many building occupants are absent.

Lastly, since the Bayesian Constrained Least Squares algorithm (6.6) has identified a set of parameter estimates $\hat{a}, \hat{b}, \hat{c}$ valid for all three seasons to account for the time-invariant physics of the building, the temperature predictions are of the same nature for all three seasons. We thus conclude that the inherent differences between the seasonal temperature data are captured by the internal gains and can be compared between the seasons on a relative level.

The identified models for the seasons $\mathcal{X} \in \{\mathcal{F}, \mathcal{W}, \mathcal{S}\}$ found with (6.6) are

$$\begin{aligned} x(k+1) &= 0.80 \cdot x(k) - 0.18 \cdot u(k) + [0.0019, 0.028, \mathbf{0}]\, v(k) + q_{\mathrm{IG},\mathcal{X}}(k) \\ &= 0.80 \cdot x(k) - 0.18 \cdot u(k) + 0.0019 \cdot v_{\mathrm{Ta}}(k) + 0.028 \cdot v_{\mathrm{Ts}}(k) + q_{\mathrm{IG},\mathcal{X}}(k) \end{aligned} \tag{6.8}$$

The estimated effect of solar radiation on the room temperature is orders of magnitude less than that of other factors and hence can be neglected. This is in agreement with our testbed having no windows in the South and Northeast zones, and most of the windows are covered with blinds.

The average RMS prediction errors are 0.22°C, 0.17°C and 0.23°C for fall, winter and spring, respectively, showing that our model predicts the temperature reasonably well.

6.3.2 Individual Zones

6.3.2.1 Model Setup

Rather than approximating the entire 4th floor of SDH as a single zone, we now identify a multivariate thermodynamic model for each of the six individual zones:

$$\begin{aligned} x(k+1) = Ax(k) + Bu(k) + Cv(k) + q_{\mathrm{IG},\mathcal{X}}(k) \\ \text{for } \mathcal{X} \in \{\mathcal{F}, \mathcal{W}, \mathcal{S}\}, \end{aligned} \tag{6.9}$$

where x, $q_{\mathrm{IG},\mathcal{X}} \in \mathbb{R}^6$, and the control input $u \in \mathbb{R}^6$ represent the temperatures, the internal gains of each zone, and the total air flow to each zone, respectively. In the lumped case, it was observed that solar radiation only had a negligible effect on the building's thermodynamics compared to the input and other disturbances, and thus we omit the solar radiation in the subsequent analysis: $v := [v_{\mathrm{Ta}}, v_{\mathrm{Ts}}]^\top \in \mathbb{R}^2$.

Inspired by Newton's Law of Cooling, only adjacent zones influence each other's temperature, which defines the sparsity pattern of the coefficient matrices that are to be estimated:

$$A_{ij} = \begin{cases} \neq 0, & \text{if } i = j \text{ or } (i,j) \text{ adjacent} \\ 0, & \text{otherwise.} \end{cases} \tag{6.10}$$

The diagonal elements of A denote autoregressive terms for zone temperatures, whereas non-diagonal elements describe the heat exchange between adjacent rooms. The matrix B is diagonal by definition of u. The sparsity pattern of C is found by physical adjacency of a respective zone to an exterior wall of a given geographic direction.

6.3.2.2 Model Identification

The procedure for the estimation of the parameter matrices $\hat{A}$, $\hat{B}$, $\hat{C}$, and the internal gains follows (6.6), but with a modified choice of the (now matrix-valued) priors μ_a and μ_b: μ_b and the diagonal entries of μ_a are obtained by scaling the corresponding priors from the

lumped zone case in order to account for the thermal masses of the individual zones, which are smaller than in the lumped case. The off-diagonal elements of μ_a, which represent the heat transfer between adjacent zones, were set to a value close to zero, according to our calculations with the heat transfer equation $\dot{q} = U \cdot A \cdot \Delta x$ and Koehler and Borrelli 2013.

6.3.2.3 Results

Figure 6.3 shows the estimated internal gains for the three seasons fall, winter, and spring for the six single zones, computed with the smoothed time series (6.7). It can be seen that the different zones exhibit different magnitudes of internal gains, with average values of the internal gains ranging between 1.0°C and 3.6°C for different zones and seasons. Similar to the lumped zone case (Figure 6.2), daily peaks of the internal gains profiles can be recognized, with a slight decrease in magnitude on weekend days. Table 6.2 reports the average prediction RMS error by zone and season.

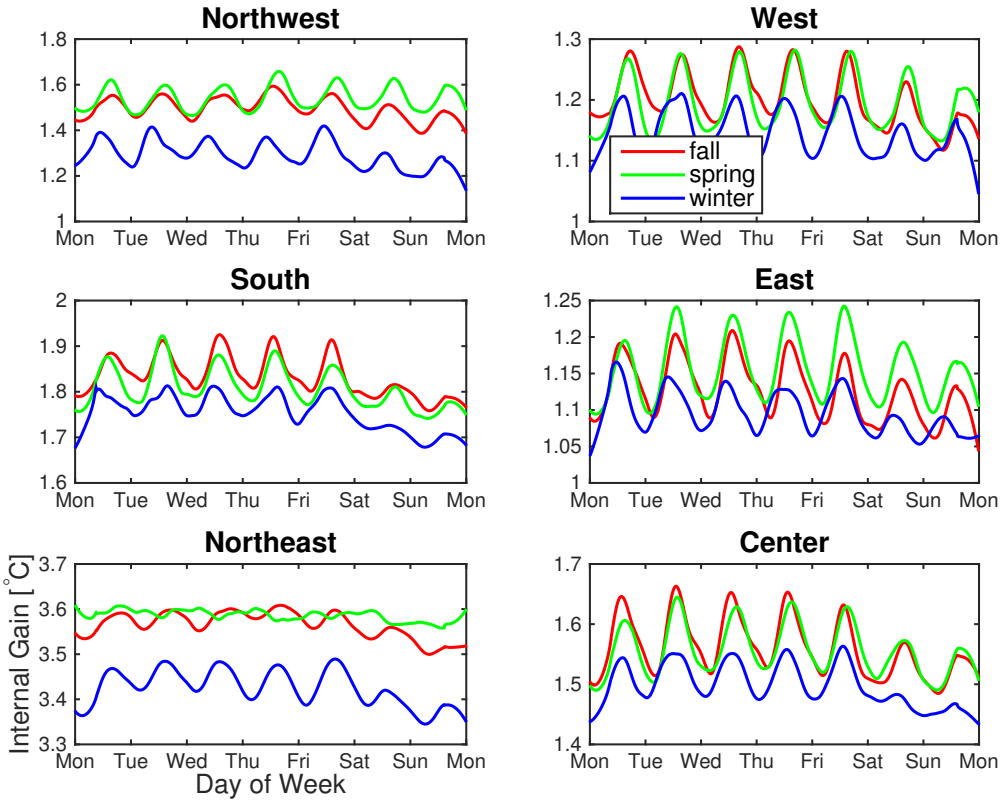

Figure 6.3: Estimated Internal Gain q_{IG} from the Data-Driven Model by Zone and Season, Individual Case

Data-Driven Model							
Season	NW	W	S	E	NE	C	Mean
Fall	0.98	0.61	0.28	0.42	0.28	0.36	0.488
Winter	1.41	0.34	0.29	0.26	0.25	0.21	0.460
Spring	0.56	0.25	0.31	0.71	0.17	0.34	0.390
Physics-Based Model							
Season	NW	W	S	E	NE	C	Mean
Fall	0.61	0.46	0.39	0.39	0.20	0.32	0.396
Winter	0.55	0.39	0.34	0.32	0.18	0.24	0.338
Spring	0.45	0.28	0.24	0.33	0.09	0.19	0.263

Table 6.2: RMS by Zone and Season for Data-Driven and Physics-Based Models

6.4 Physics-Based Model

We describe the physics-based modeling approach proposed in Hu et al. 2016, which obtains an RC-model using the Building Resistance-Capacitance Modeling (BRCM) MATLAB toolbox (Sturzenegger et al. 2012). A main advantage of this approach is that the resulting model has a small number of parameters, even for a complex multi-zone building; furthermore, these parameters have strong physical meaning, which aids in their identification. We re-identify the building model using the same training dataset as used in Section 6.3, and estimate distinct internal gains functions for different seasons.

6.4.1 Model Setup

The physics-based model has the following form Hu et al. 2016:

$$\begin{aligned} x(k+1) &= Ax(k) + B_v v(k) + B_{\mathrm{IG}} f_{\mathrm{IG}}(k) \\ &\quad + \textstyle\sum_{i=1}^{21} \Big(B_{xu_i} x(k) + B_{vu_i} v(k) \Big) u_i(k) \end{aligned} \tag{6.11a}$$

$$y(k) = Cx(k), \tag{6.11b}$$

where state vector $x \in \mathbb{R}^{289}$ and $y \in \mathbb{R}^6$ represent temperatures of all building elements (walls, ceilings, floors, etc.) on the 4th floor and the average temperatures of the six zones shown in Figure 6.1, respectively. $u \in \mathbb{R}^{21}$ denotes the airflow rate from the 21 VAV boxes and $v := [v_{\mathrm{Ta}}, v_{\mathrm{Ts}}]^\top$ the vector of known disturbances. As in the data-driven model, heat gains due to solar radiation are omitted from the analysis. Finally, $f_{\mathrm{IG}}(k) : \mathbb{N} \to \mathbb{R}^6$ captures internal gains in each of the six zones on the 4th floor. For week m in the training set $\mathcal{T}$:

$$f_{\mathrm{IG}}(k) = f^c_{\mathrm{IG}} + \begin{cases} f^v_{\mathrm{IG},\mathcal{F}}(k), & \text{if } m \in \mathcal{F}, \\ f^v_{\mathrm{IG},\mathcal{W}}(k), & \text{if } m \in \mathcal{W}, \\ f^v_{\mathrm{IG},\mathcal{S}}(k), & \text{if } m \in \mathcal{S}, \end{cases} \tag{6.12}$$

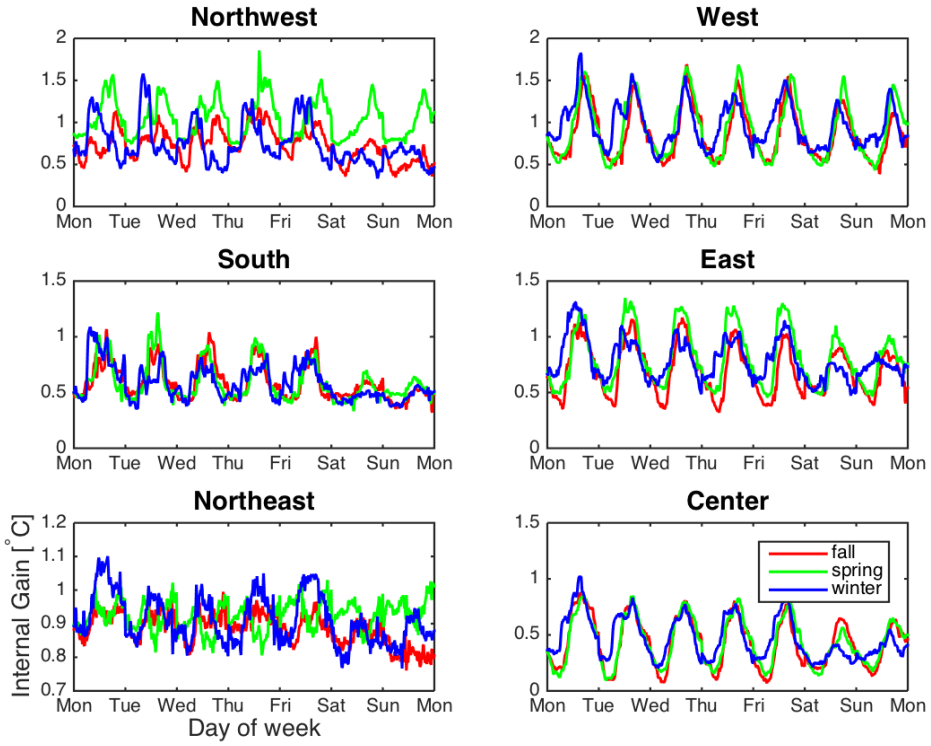

Figure 6.4: Estimated Internal Gain f_{IG} from the Physics-Based Model by Zone and Season

where f_{IG}^c is an unknown constant vector representing background heat gains due to idle electric appliances. Functions $f_{\text{IG},\mathcal{F}}^v(\cdot)$, $f_{\text{IG},\mathcal{W}}^v(\cdot)$ and $f_{\text{IG},\mathcal{S}}^v(\cdot)$ are unknown nonparametric functions that capture the time-varying heat gain due to occupancy and equipments in fall, winter and spring, respectively. The system matrices A, B_v, B_{IG}, B_{xu_i} and B_{vu_i} are functions of the window heat transmission coefficient U_{win} and convection coefficients of the interior wall γ_{IW}, exterior wall γ_{EW}, floor γ_{floor}, and ceiling γ_{ceil}. Define $\gamma := \left[U_{\text{win}}, \gamma_{\text{IW}}, \gamma_{\text{EW}}, \gamma_{\text{floor}}, \gamma_{\text{ceil}}, f_{\text{IG}}^{c\top}\right]^\top \in \mathbb{R}^{11}$, then to identify the physics-based model, we need to estimate the parameter vector γ as well as the functions $f_{\text{IG},\mathcal{X}}^v(\cdot)$, $\mathcal{X} \in \{\mathcal{F}, \mathcal{W}, \mathcal{S}\}$.

6.4.2 Model Identification

For a fair comparison, the same data used to train and test the data-driven model is used to train and validate the physics-based model. The model identification process is performed in two steps: First, the subset of the training data collected during weekends is used to estimate the parameters, γ. Second, the nonparametric functions $f_{\text{IG},\mathcal{X}}^v(\cdot)$ are estimated from the complete training dataset.

6.4.2.1 Parameter Estimation

We first set $f^v_{\mathrm{IG},\mathcal{X}}(\cdot) = 0$ during the weekend days to evaluate them at a later point. With $f^v_{\mathrm{IG},\mathcal{X}}(\cdot) = 0$, (6.11) reduces to a purely parametric model:

$$\begin{aligned} x(k+1) &= Ax(k) + B_v v(k) + B_{\mathrm{IG}} f^c_{\mathrm{IG}} \\ &\quad + \textstyle\sum_{i=1}^{21} \Big(B_{xu_i} x(k) + B_{vu_i} v(k)\Big) u_i(k), \\ y(k) &= Cx(k). \end{aligned} \tag{6.13}$$

The optimal model parameters are estimated by solving the following optimization problem:

$$\begin{aligned} \hat{\gamma} = \ & \arg\min_{\gamma > 0} \ \textstyle\sum_{m \in \mathcal{T}} \sum_k \|y_m(k,\gamma) - \bar{y}_m(k)\|^2 \\ \text{s.t.} \ & \ y_m(k,\gamma) \text{ and } x_m(k,\gamma) \text{ satisfy (6.13) with} \\ & \ x_m(0) = x_{\mathrm{KF},m}(0) \\ & \ u_m(k) = \bar{u}_m(k), v_m(k) = \bar{v}_m(k) \ \forall \ k, \end{aligned} \tag{6.14}$$

where $\bar{u}$, $\bar{v}$ and $\bar{y}$ denote the measured inputs, disturbances, and zone temperatures, respectively. In other words, we choose γ such that, when the model is simulated with this set of parameter values and the measured inputs and disturbances, the sum of squared errors between the measured zone temperatures and the simulated temperatures is minimized. The initial state $x_m(0)$ is required to simulate the model, however, not all states are measurable (e.g. wall temperature). Thus, we estimate the initial states using a Kalman Filter $x_{\mathrm{KF},m}(0)$ and set $x_m(0) = x_{\mathrm{KF},m}(0)$. Furthermore, to compensate for the lack of sufficient excitation of the building, physically plausible initial guesses for γ are chosen. The optimal parameter values are similar to those reported in Hu et al. 2016 and hence omitted.

6.4.2.2 Estimation of $f^v_{\mathrm{IG}}(\cdot)$

Let $f^v_{\mathrm{IG},m}(\cdot)$ be an instance of the internal gains function $f^v_{\mathrm{IG}}(\cdot)$ estimated for week m in the training set. The optimal estimate for a given season, is then defined as the the average of all estimates for that season:

$$\hat{f}^v_{\mathrm{IG},\mathcal{F}}(k) = \textstyle\sum_{m \in \mathcal{F}} f^v_{\mathrm{IG},m}(k) / \|\mathcal{F}\| \quad \forall \ k, \tag{6.15}$$

where $\|\mathcal{F}\|$ represents the cardinality of set $\mathcal{F}$. To estimate $f^v_{\mathrm{IG},m}(\cdot)$ for a given week m, let $\tilde{x}(k)$ and $\tilde{y}(k)$ denote the predicted states and zone temperatures at time k, with $f^v_{\mathrm{IG},w}(k-1) = 0$. That is,

$$\begin{aligned} \tilde{x}(k) &= Ax(k-1) + B_v v(k-1) + B_{\mathrm{IG}} f^c_{\mathrm{IG}} \\ &\quad + \textstyle\sum_{i=1}^{21} \Big(B_{xu_i} x(k-1) + B_{vu_i} v(k-1)\Big) \\ &\quad \cdot u_i(k-1), \\ \tilde{y} &= C\tilde{x}(k). \end{aligned} \tag{6.16}$$

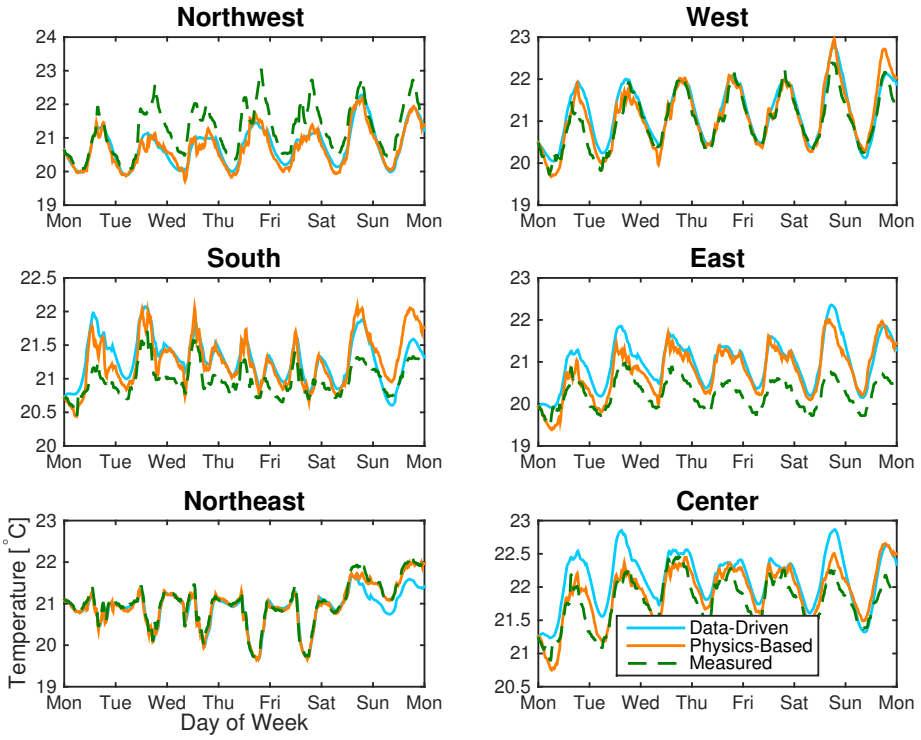

Figure 6.5: Simulated Temperatures from the Data-Driven Model (blue), Physics-Based Model (orange) and Actual Temperatures (green)

By noting $x(k) = \tilde{x}(k) + B_{\text{IG}} f^v_{\text{IG},m}(k-1)$, $f^v_{\text{IG},m}(k-1)$ can be estimated by solving $(CB_{\text{IG}}) \cdot f^v_{\text{IG},m}(k-1) = \bar{y}(k) - \tilde{y}(k)$, a set of linear equations, using Ordinary Least Squares. $\bar{y}(k)$ denotes the measured zone temperatures at time k.

6.4.3 Results

The average daily prediction RMS errors by zone and season are reported in Table 6.2. Figure 6.4 shows the estimated temperature contribution of the internal gains for fall, winter and spring. The zones that correspond to open workspaces and conference rooms ("West", "South", "East", "Center") show discernible daily peaks in their internal gains profiles with a slight decrease during weekends. Lastly, there is little variation in the internal gains across different seasons.

6.5 Quantitative Comparison of Both Models

6.5.1 Prediction Accuracy

The physics-based model (Model B) is found to have a higher prediction accuracy compared to the data-driven model for the individual zones (Model A) presented in Section 6.3.2: According to Table 6.2, the mean RMS error for Model B across zones is 0.11°C lower than for Model A. This is also illustrated in Figure 6.5, which shows 7-day open-loop predictions of the temperature of a randomly selected holdout test week in the spring period, simulated with both models initialized with the measured temperature. To the best of our knowledge, a quantitative comparison at this level is non-existent, as previous building models were developed for different testbeds, fictitious buildings or from simulated data. This chapter attempts to close this gap by providing a quantitative comparison between the low-dimensional data-driven model and the high-dimensional physics-based model's prediction accuracy for the *same* multi-zone commercial building under *regular* operation.

6.5.2 Energy Efficient Control

Next, we explore the extent to which Model A's slightly lower prediction accuracy affects its resulting controller's closed-loop performance in energy efficient control. We formulate an MPC problem to find the optimal control strategy that minimizes the cost of HVAC operation over the same week used in Figure 6.5, while guaranteeing the temperature to stay within a comfort zone $[T_{\min}, T_{\max}]$, which we chose as $[20°\text{C}, 22°\text{C}]$ (Hansen and Burroughs 2013), and confining the control input to the minimum and maximum airflow settings of the HVAC system $[u_{\min}, u_{\max}]$:

$$
\begin{aligned}
\min_{u,\varepsilon} \quad & \sum_{k=1}^{N} u(k)^2 + \rho\|\varepsilon\|_2 \\
\text{s.t. } & x(0) = \bar{x}(0) \\
& x(k+1) = \begin{cases} (6.9), & \text{Model A} \\ (6.11\text{a}), & \text{Model B} \end{cases} \\
& u_{\min} - \varepsilon \le u(k) \le u_{\max} + \varepsilon \qquad \forall k \in [0, N-1] \\
& \begin{cases} T_{\min} \le x(k) \le T_{\max}, & \text{Model A} \\ T_{\min} \le Cx(k) \le T_{\max}, & \text{Model B} \quad (6.11\text{b}) \end{cases} \qquad \forall k \in [1, N]
\end{aligned}
\tag{6.17}
$$

The temperature is initialized with the measured temperature $\bar{x}(0)$ at the beginning of the week-long simulation. Soft constraints on the control input with a penalty parameter ρ ensure feasibility of the problem. The penalty represents the cost of decreasing airflow below the set minimum value. This is physically feasible as the set minimum airflow rate for our testbed is significantly higher than the standard minimum required by building standards. To find the

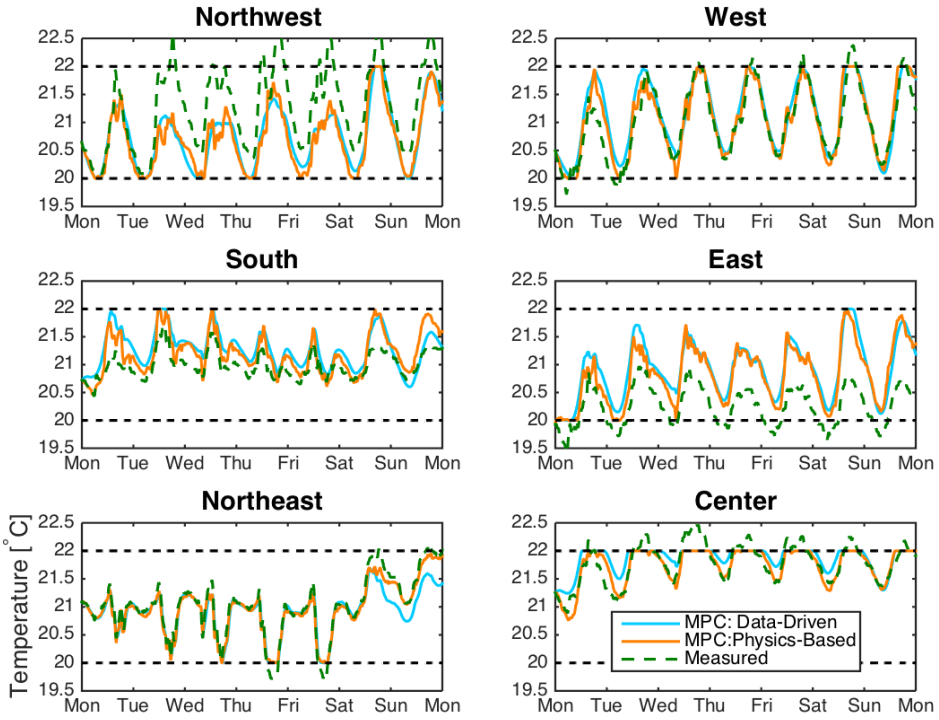

Figure 6.6: Optimal Temperature for MPC with Data-Driven Model (blue), MPC with Physics-Based Model (orange) and Actual Temperature (green)

optimal control strategy, we make use of receding horizon control with a prediction horizon of three 15-minute time steps.

Figure 6.6 shows the temperature trajectory computed by the energy efficient controller (6.17) computed with both models A and B, together with the measured temperature as a reference. It can be seen that both control schemes are capable of maintaining the temperature within $[20°\text{C}, 22°\text{C}]$, with a control strategy that is of comparable cost (1,006 and 1,731 for Model A and Model B, respectively, where $\rho = 100$), shown in Figure 6.7. Furthermore, it is interesting to observe that variations in the control input do not impact the periodicity of the temperature qualitatively, which can be explained by the regularity of the identified internal gains.

These findings suggest that both models perform equally well in designing an energy efficient control strategy. However, computing this strategy for Model A was cheap (< 5 minutes) compared to Model B (≈ 20 hours) on a 2 GHz Intel Core i7, 16 GB 1600 MHz DDR3 machine. Further, we note that in real-world applications, the MPC would use state feedback to initialize the temperature with sensor measurements at every time step, whereas in our simulation, it operates in an "open loop" fashion and hence propagates the estimation

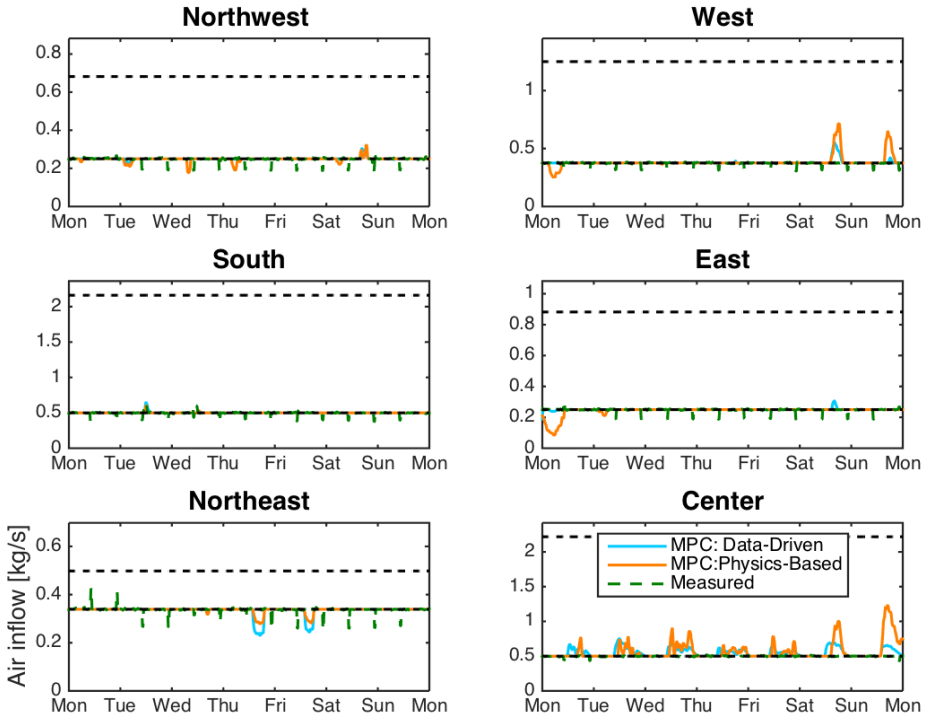

Figure 6.7: Optimal Control Strategy for MPC with Data-Driven Model (blue), MPC with Physics-Based Model (orange) and Actual Input (green)

error with time. This will reduce the difference in the prediction quality by both controllers even further, since the RMS error is now to be evaluated on a much shorter prediction horizon, thereby further corroborating the finding of almost identical control schemes.

Observing that Model A only suffers a negligible loss of accuracy compared to Model B for an open loop optimal control scheme, our findings suggest the suitability of Model A to applications with temperature-critical zones in which even more precise temperature estimates are needed, e.g. long-term planning of reserve provision for frequency regulation.

6.6 Discussion and Conclusion

We identified a low-dimensional data-driven model, using semiparametric regression, and a high-dimensional physics-based resistance-capacitance model for the thermal behavior of the same multi-zone commercial building. Both state-space models were fitted on experimental data collected during regular building operation and capture the effect of disturbances such as occupancy and electrical appliances that commercial buildings are subjected to, without

installation of any additional hardware such as occupancy sensors.

The identification of both models on the *same building* enabled us to quantitatively compare the performance of these types of models when applied to a real building, which has not been investigated before. Our results showed that the RMS error of the open-loop temperature prediction of the physics-based model across different thermal zones and temporal seasons is 0.11°C lower than in the data-driven model, a 25% reduction. However, simulating energy efficient MPC schemes under both models suggested both models perform equally well in terms of cost function minimization and constraint satisfaction despite the significantly higher complexity of the physics-based model.

It is widely known in this field that low-dimensional data-driven models have lower prediction accuracy than high-dimensional physics-based models, and thus have been only proposed for control of less temperature-critical buildings or zones. However, our work investigated an identification method for data-driven models for multi-zone commercial buildings in regular operation and demonstrated that the lower open-loop prediction accuracy of such data-driven models is insignificant in closed-loop control schemes compared to a high-dimensional physics-based model. Based on these findings, we suggest that such data-driven models may be suitable for applications that were previously considered inappropriate, e.g. frequency regulation.

To verify our hypotheses, one would have to design and implement a control scheme suitable for frequency regulation, using the data-driven model, into the building operation system of SDH. This is a challenge left for future work.

Chapter 7

Short-Term Load Forecasting on Smart Meter Data

7.1 Introduction

While Chapter 6 analyzed the electric load associated with temperature control of a commercial building, we now pivot to residential users and their loads. The central purpose of the remainder of this thesis is to quantify the extent to which residential households are willing to reduce their electricity consumption temporarily in exchange for a monetary incentive. In contrast to Chapter 6, in which we were capable of deriving explicit equations for temperature models that were amenable to HVAC control, we have to avail ourselves of more complex machinery to be able to build suitable models for the behavior of residential customers. This chapter is a precursor to the actual analysis of a residential Demand Response (cf. Chapter 8), as it provides a few fundamental tools for predicting electricity consumption on an individual user level. Specifically, this chapter discusses how to forecast smart meter data with advanced models, namely Hidden-Markov Models (HMMs) and Conditional Mixtures of Gaussians (CGMMs).

In the extant literature, short-term load forecasting (STLF) has been extensively studied with different approaches and on different levels of aggregations of users, ranging from the individual level to city-wide predictions (Sevlian and Rajagopal 2014; Mirowski et al. 2014). Statistical time series models (Arora and Taylor 2014; Taylor and McSharry 2007), standard parametric regression models such as Ordinary Least Squares, Lasso- and Ridge-Regression (Mirowski et al. 2014), and non-parametric methods including k-Nearest Neighbors, Support Vector Regression (Elattar, Goulermas, and Wu 2010), and Neural Networks (Edwards, New, and Parker 2012) have been evaluated with respect to different metrics for accuracy. Widely explored Bayesian Methods for STLF are Gaussian Processes (Lauret, David, and Caloigne 2012), Bayesian Neural Network approaches, e.g. for input selection problems (Hippert and Taylor 2010), and Kalman-Filtering methods (Al-Hamadi and Soliman 2004) with Hybrid Neural Network extensions (Guan et al. 2013). HMMs for STLF have been applied primarily

for the purpose of occupancy detection (Kleiminger et al. 2014) and Nonintrusive Load Monitoring (Parson et al. 2011). (Han, Gao, and Fan 2012) and (Ai, Fan, and Gao 2014) utilize occupancy information to increase the energy efficiency of building operation. To the best of our knowledge, CGMMs have not been investigated for STLF.

This chapter is structured as follows: In Section 7.2, we briefly outline classical Machine Learning (ML) methods used for STLF. Sections 7.3 and 7.4 describe technical details of CGMMs and HMMs tailored to the specific needs of STLF, followed by Section 7.5, which outlines the procedure of incorporating the estimated latent variables into STLF. Section 7.8 concludes this chapter.

7.2 Forecasting Methods

The following well-established forecasting methods, which regress the outcomes Y on the covariates X, are discussed in the remainder of this chapter:

- Ordinary Least Squares Regression (OLS)
- k Nearest Neighbors Regression (KNN)
- Support Vector Regression (SVR)
- Decision Tree Regression (DT)

Notation: Let $Y \in \mathbb{R}^N$ denote a column vector of N scalar outcomes $\{y_1, \dots, y_N\}$, e.g. in our case electricity consumption, and $X \in \mathbb{R}^{N \times d}$ the design matrix whose k-th row represents the covariates $x_k \in \mathbb{R}^d$ associated with outcome y_k. Let y and x denote a generic outcome and its associated covariate vector, respectively.

7.2.1 Ordinary Least Squares Regression

Assuming a linear relationship between covariate-outcome pairs (X, Y),

$$Y = Xw, \tag{7.1}$$

the regression coefficients $w \in \mathbb{R}^d$ are estimated using Ordinary Least Squares Regression (OLS).

7.2.2 K-Nearest Neighbors-Regression (KNN)

Given a point in feature space x, the goal is to find the k training points $x_1, \dots, x_k$ that are closest in distance to x. We choose the commonly used Euclidian norm (though other

choices can be justified) as a measure for distance in feature space. The prediction of the outcome $\hat{y}$ is the average of the outcomes of the k nearest neighbors

$$\hat{y} = \frac{1}{k}(y_1 + \ldots + y_k). \tag{7.2}$$

The number of neighbors k for an optimal fit is found using common cross-validation techniques.

7.2.3 Support Vector Regression

Support Vector Regression (SVR) solves the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \; & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ \text{s.t. } & y_i - w^\top \phi(x_i) - b \leq \epsilon + \xi_i, \\ & w^\top \phi(x_i) + b - y_i \leq \epsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, \quad i \in [1, \ldots, N]\,. \end{aligned} \tag{7.3}$$

In (7.3), ϵ defines an error tube within which no penalty is associated, ξ and ξ^* denote slack variables that guarantee the existence of a solution for all ϵ, b is a real constant, C is the regularization constant, w are the regression coefficients to be estimated, and $\phi(\cdot)$ a map between the input space and a higher dimensional feature space. (7.3) is typically solved by transforming it into dual form, thereby avoiding the explicit calculation of $\phi(\cdot)$ with the so-called Kernel trick. We choose the commonly used Gaussian Kernel function.

7.2.4 Decision Tree Regression (DT)

This non-parametric learning method finds decision rules that partition the feature space into up to 2^n pieces, where n denotes the maximal depth of the tree. For a given iteration step, enumeration of all nodes and possible splitting scenarios (exhaustive search) yields a tuple $\theta^* = (j, t_m)$ that minimizes the sum of the ensuing child node impurities $G(\theta^*, m)$, where j denotes the j-th feature and m the m-th node of the tree. This is written as

$$\theta^* = \arg\min_{\theta} G(\theta, m), \tag{7.4a}$$

$$G(\theta, m) = \frac{n_{\text{left}}^m}{N_m} H(Q_{\text{left}}(\theta)) + \frac{n_{\text{right}}^m}{N_m} H(Q_{\text{right}}(\theta)). \tag{7.4b}$$

where Q_{left} and Q_{right} denote the set of covariate-outcome pairs belonging to the left and right child node of parent node m, respectively; and n_{left}^m and n_{right}^m denote their respective

count. The impurity measure $H(\cdot)$ at a node minimizes the mean squared error

$$c(\cdot) = \frac{1}{N(\cdot)} \sum_{i \in N(\cdot)} y_i, \tag{7.5a}$$

$$H(\cdot) = \frac{1}{N(\cdot)} \sum_{i \in N(\cdot)} \left[y_i - c(\cdot) \right]^2, \tag{7.5b}$$

with $N(\cdot)$ representing the number of covariate-outcome pairs at the node of interest.

DTs are readily fitted using exhaustive search for each split. Cross-validation, usually on the maximal depth of the tree or the minimal number of samples per node, avoids overfitting of the tree. The optimized tree is then used for forecasting the outcome by taking the average of all outcomes belonging to a given node m. This yields a decision tree with piecewise constant predictions.

7.3 Mixture Models

In this section, we describe the fitting procedure of CGMMs on data that combine multiple linear regression models to act as an ensemble learner. Given a set of covariate-outcome pairs (in our case y_i denotes energy consumption),

$$\mathcal{D} = \{(x_i, y_i) : i = 1, \dots, N\}, \tag{7.6}$$

the idea is to model the probability distribution of any observation y with corresponding covariates x as the output of an ensemble of linear regressions

$$\mathbb{P}(y|x, \underbrace{\mathbf{w}, \sigma^2, \pi}_{=:\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y|w_k \cdot x, \sigma^2), \tag{7.7}$$

where $\pi = \{\pi_1, \dots, \pi_K\}$ and $\mathbf{w} = \{w_1, \dots, w_K\}$ denote K mixing proportions with $\sum_{i=1}^{K} \pi_k = 1$ and the regression coefficients for each learner, respectively. σ^2 signifies the noise variance, where, according to Bishop 2006, we make the following

Assumption 12. *σ^2 is equal across all mixture components $k = 1, \dots, K$.*

Assumption 12 can be relaxed by using mixture-specific noise covariances $\{\sigma_1^2, \dots, \sigma_K^2\}$, in which case we would have to modify (7.10a)−(7.10d).

The rationale behind fitting a CGMM on a residential load forecasting setting is to describe electricity consumption as a result of different user behaviors $\{\pi_k\}_{k=1}^{K}$, i.e. to interpret the mixing proportions as archetypes of human consumption behavior, each of which has an observable output distribution governed by an OLS model. The convex combination of these archetypes allows for an interpolation between these archetypes.

7.3.1 Parameter Estimation

Given the training data $\mathcal{D}$, the Expectation-Maximization Algorithm (EM-Algorithm) (Rabiner 1989; Bishop 2006) allows us to derive an iterative procedure to learn the parameters $\theta = \{\{\pi_k\}_{k=1}^K, \{w_k\}_{k=1}^K, \sigma^2\}$. We first define the expected complete log likelihood $\ell(\theta|\mathcal{D}_c)$, where

$$\mathcal{D}_c = \{(x_i, y_i, z_i) : i = 1, \ldots, N\} \tag{7.8}$$

denotes the fully observed dataset whose latent variables $\{z_1, \ldots, z_N\}$ are assumed to be known. The latent variable belonging to x_i is a vector $z_i = [z_{i1}, \ldots, z_{iK}]^\top$, where z_{ik} denotes the probability that x_i was generated by mixture component k. The complete log-likelihood is

$$\ell(\theta|\mathcal{D}_c) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \left(\pi_k \mathcal{N}(y_i|w_k \cdot x_i, \sigma^2)\right) \tag{7.9}$$

under the assumption of known z_{ik}. The EM-Algorithm alternates between the E-Step, whose task is to determine the expected value of the latent variables $z_{ik}, 1 \leq i \leq N, 1 \leq k \leq K$ with respect to the conditional probability distribution (7.7), and the M-Step, which updates the parameters θ with the results from the E-Step by taking the derivative of the expected value of (7.9) with respect to the desired parameters θ. This is carried out iteratively until some convergence criterion is reached, i.e. the incremental increase of the expected complete log likelihood (7.9) falls below a threshold. The update steps for one iteration are as follows:

$$\hat{z}_{ik} = \frac{\hat{\pi}_k \mathcal{N}(y_i|\hat{w}_k \cdot x_i, \hat{\sigma}^2)}{\sum_{j=1}^K \hat{\pi}_j \mathcal{N}(y_i|\hat{w}_j \cdot x_i, \hat{\sigma}^2)}, \tag{7.10a}$$

$$\hat{\pi}_k = \frac{1}{N} \sum_{i=1}^{N} \hat{z}_{ik}, \tag{7.10b}$$

$$\hat{w}_k = \left[X^\top D X\right]^{-1} X^\top D Y, \; D = \operatorname{diag}(\hat{z}_{1k}, \ldots, \hat{z}_{Nk}), \tag{7.10c}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \hat{z}_{ik} (y_i - \hat{w}_k \cdot x_i)^2, \tag{7.10d}$$

where we have to incorporate the constraint $\sum_{k=1}^K \hat{\pi}_k = 1$ as a Lagrange Multiplier in the derivation.

7.3.2 Predicting New Data

To predict the outcome $\hat{y}$ of an out-of-sample data point x, we suggest a different approach than is employed by Bishop 2006: Instead of using the estimated mixing proportions $\{\hat{\pi}_k\}_{k=1}^K$ as the weights for a convex combination of the estimated regression coefficients $\{\hat{w}_k\}_{k=1}^K$, we

choose the weights as the estimated latent variables $\{\hat{z}_{jk}\}_{k=1}^{K}$ of x's nearest neighbor x_j:

$$j = \arg \min_{1 \leq i \leq N} \|x_i - x\|_2 \tag{7.11a}$$

$$\hat{y} = \sum_{k=1}^{K} \hat{z}_{jk} \hat{w}_k \cdot x \tag{7.11b}$$

The rationale behind this approach is to exploit potential spatial separation in the set of training data, i.e. the fact that different regions of the covariate space are best fit by a specific learner. By locating the nearest neighbor of x, the same set of weights that proved to be most accurate for the training of the data points in the region around x are to be used for the prediction of $\hat{y}$.

7.4 Hidden Markov Models

In this section, we briefly outline the training procedure of HMMs. Figure 7.1 shows the graphical model of a standard HMM with a hidden layer (transparent nodes), representing latent variables, and observations (shaded nodes).

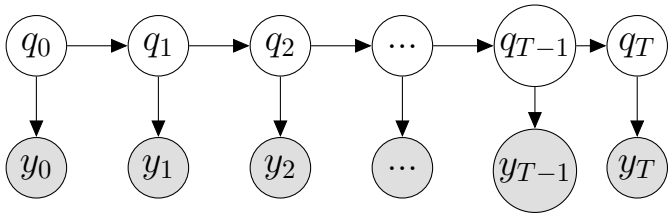

Figure 7.1: Hidden Markov Model. Hidden States q, Observations y

7.4.1 Hidden Layer

We model the latent variables in the hidden layer (see Figure 7.1) as a first order, time-invariant, Discrete Time Markov Chain (DTMC) with a set of transition probabilities

$$a_{ij} = \mathbb{P}(q_t = j | q_{t-1} = i), \quad 1 \leq i, j \leq M, \tag{7.12}$$

where $t = 0, 1, 2, \ldots, T$ denote time instants associated with state changes and q_t the hidden state at time t. Due to the Markov Property, we have that, conditional on q_t, q_{t+1} is independent of q_{t-1}. The state transition coefficients a_{ij} have the properties

$$0 \leq a_{ij} \leq 1, \quad \sum_{j=1}^{M} a_{ij} = 1, \quad i, j \in \{1, \ldots, M\}, \tag{7.13}$$

where M denotes the number of states (=latent variables).

We postulate the existence of two different latent states for each hour of the day (HoD) between 6 a.m. - 8 p.m., and a single state for the remaining hours, hence $M = 38$. For the former hours, binary states describing each hour shall encode information about "high" ("H") or "low" ("L") consumption, which might be an indicator for occupancy ("H" = at home, "L" = not at home). For the remaining HoDs, we note that first, no DR events in our data set were recorded outside this window, and second, little variation in the smart meter recordings was observed, which is consistent with Zhou, Balandat, and Tomlin 2016b, where the authors find little variation in clustered load shapes during the night. Due to the Markov Property, state transitions are restricted to states belonging to the next hour only, which renders the Markov transition matrix $A \in \mathbb{R}^{38\times 38}$ sparse. Figure 7.2 shows the state transition diagram (without probabilities on the edges, which are to be estimated from data, see Section 7.4.3).

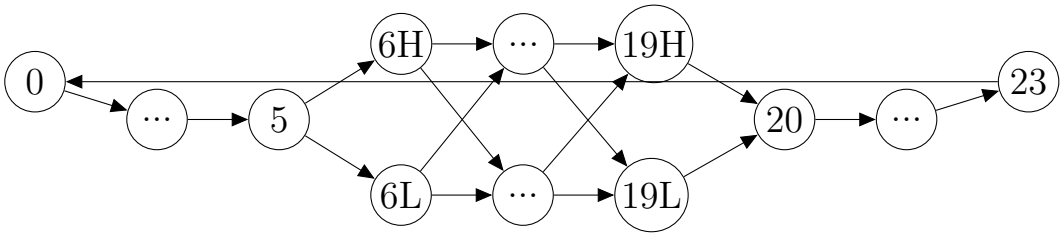

Figure 7.2: Markov State Transition Diagram, 24 Hour Periodicity. For Example, "5" Signifies Time Between 5 a.m. - 6 a.m.

A logical extension is to allow for multi-step dependencies, which can be achieved by enlarging the state space of the DTMC such that the previous $n > 1$ states jointly determine the next transition. A more granular description of the state transitions, however, would come at the cost of a higher computational complexity, a tradeoff whose analysis is outside the scope of this chapter.

A consequence of this modeling approach is that, if the consumption is high at time $t-1$, it is likely that the hidden state $q_{t-1} = \mathrm{H}$ and $q_t = \mathrm{H}$, and so we expect a high consumption at time t, as well. Conversely, if the consumption at time $t-1$ is low (i.e. due to an absent user), the most likely hidden state $q_{t-1} = \mathrm{L}$ and $q_t = \mathrm{L}$, and thus we would expect a low consumption at time t. It turns out that the parameter estimation on actual smart meter data set automatically assigns higher probabilities to transitions to the next hour of the same type than to the opposite type, indicating that switches between "H" and "L" do not occur frequently. This is consistent with our intuition: If the latent variable represents periods of expected presence or absence at home, users are more likely to remain either at home or absent, rather than switching every hour.

7.4.2 Observations

Assumption 13. *Conditional on the current hidden state q_t, the observable energy consumption y_t (=observation/emission) is assumed to be normally distributed with parameters*

$(\mu_{q_t}, \sigma^2_{q_t})$:

$$\mathbb{P}(y_t|q_t) = \frac{1}{\sqrt{2\pi\sigma^2_{q_t}}} \exp\left(-\frac{(y_t - \mu_{q_t})^2}{2\sigma^2_{q_t}}\right). \tag{7.14}$$

An obvious extension is to choose alternative distributions, an idea we do not investigate further in this thesis.

7.4.3 Parameter Estimation and Inference

Given an observed sequence of emissions $Y := \{y_0, y_1, \ldots, y_T\}$ with known initial state distribution π_{q_0}, the parameters of the HMM $\theta := \{\{a_{ij}\}, \{\mu_{q_t}\}, \{\sigma^2_{q_t}\}\}$, i.e. the transition probabilities and emission parameters, can be estimated with the EM-Algorithm. Starting from the complete log-likelihood

$$\ell(\theta|\mathcal{D}_c) = \log\left(\pi_{q_0} \prod_{t=0}^{T-1} a_{q_t,q_{t+1}} \prod_{t=0}^{T} \mathcal{N}(y_t|\mu_{q_t}, \sigma^2_{q_t})\right), \tag{7.15}$$

with the fully observed data set

$$\mathcal{D}_c = \{(y_n, q_n, a_{q_n,q_{n+1}}) : n \in [0, T-1]\} \cup \{\pi_{q_0}, y_T, q_T\}, \tag{7.16}$$

minimizing the expected value of (7.15) with respect to the desired variables θ to be estimated yields the update equations for the M-Step of the EM-algorithm (also called Baum-Welch Updates):

$$\hat{\pi}_i = \mathbb{P}(q_0 = i|Y) \tag{7.17a}$$

$$\hat{a}_{ij} = \frac{\sum_{t=0}^{T-1} \mathbb{P}(q_t = i, q_{t+1} = j|Y)}{\sum_{t=0}^{T-1} \sum_{j=1}^{M} \mathbb{P}(q_t = i, q_{t+1} = j|Y)} \tag{7.17b}$$

$$\hat{\mu}_i = \frac{\sum_{t=0}^{T} y_t \cdot \mathbb{P}(q_t = i|Y)}{\sum_{t=0}^{T} \mathbb{P}(q_t = i|Y)} \tag{7.17c}$$

$$\hat{\sigma}^2_i = \frac{\sum_{t=0}^{T} \mathbb{P}(q_t = i|Y)(y_t - \hat{\mu}_i)^2}{\sum_{t=0}^{T} \mathbb{P}(q_t = i|Y)} \tag{7.17d}$$

To arrive at Equations (7.17a) and (7.17b), the stochastic constraints described in (7.13) and sparsity patterns of the transition matrix A as well as $\sum_{i=1}^{M} \pi_i = 1$ are used as Lagrange multipliers during the minimization of (7.15).

Using Bayes Rule, the E-Step of the EM-algorithm computes the sufficient statistics $\mathbb{P}(q_t = i, q_{t+1} = j|Y)$ and $\mathbb{P}(q_t = i|Y)$ with the well-known *Alpha-Beta-Recursion*:

$$\begin{aligned}\mathbb{P}(q_t|Y) &= \frac{\mathbb{P}(Y|q_t)\mathbb{P}(q_t)}{\mathbb{P}(Y)} \\ &= \frac{\mathbb{P}(y_0, \dots, y_{t-1}, q_t)\mathbb{P}(y_t|q_t)\mathbb{P}(y_{t+1}, \dots, y_T|q_t)}{\mathbb{P}(Y)} \\ &=: \frac{\alpha(q_t)\mathbb{P}(y_t|q_t)\beta(q_t)}{\mathbb{P}(Y)}. \end{aligned} \tag{7.18}$$

We note that $\alpha(q_t)$ is defined as $\mathbb{P}(y_0, \dots, y_{t-1}, q_t)$ rather than $\mathbb{P}(y_0, \dots, y_t, q_t)$ as is done in Rabiner 1989 and Jordan 2007. This is done for a simplified treatment of its update step (7.19) and the prediction problem (7.23).

Using Bayes Rule, $\alpha(q_t)$ and $\beta(q_t)$ can be updated recursively:

$$\begin{aligned}\alpha(q_{t+1}) &= \mathbb{P}(y_0, \dots, y_t, q_{t+1}) \\ &= \sum_{q_t} \mathbb{P}(y_0, \dots, y_t, q_t, q_{t+1}) \\ &= \sum_{q_t} \alpha(q_t)\mathbb{P}(y_t|q_t)a_{q_t,q_{t+1}}. \end{aligned} \tag{7.19}$$

$$\begin{aligned}\beta(q_t) &= \mathbb{P}(y_{t+1}, \dots, y_T|q_t) \\ &= \sum_{q_{t+1}} \mathbb{P}(y_{t+1}, \dots, y_T, q_{t+1}|q_t) \\ &= \sum_{q_{t+1}} \beta(q_{t+1})\mathbb{P}(y_{t+1}|q_{t+1})a_{q_t,q_{t+1}}. \end{aligned} \tag{7.20}$$

Note that $\mathbb{P}(y_t|q_t),\ 0 \leq t \leq T$ can be computed with (7.14).

$\alpha(q_1)$ is initialized as π_{q_0} and $\beta(q_T)$ as a vector of ones.

With the definition of $\alpha(q_t)$ and $\beta(q_t)$, $\mathbb{P}(q_t, q_{t+1}|Y)$ is computed as follows:

$$\begin{aligned}\mathbb{P}(q_t, q_{t+1}|Y) &= \frac{\mathbb{P}(Y|q_t, q_{t+1})\mathbb{P}(q_t, q_{t+1})}{\mathbb{P}(Y)} \\ &= \frac{\alpha(q_t)\beta(q_{t+1})a_{q_t,q_{t+1}}\mathbb{P}(y_t|q_t)\mathbb{P}(y_{t+1}|q_{t+1})}{\mathbb{P}(Y)}. \end{aligned} \tag{7.21}$$

In summary, the EM-algorithm iterates between the E-Step to compute the sufficient statistics $\mathbb{P}(q_t = i, q_{t+1} = j|Y)$ and $\mathbb{P}(q_t = i|Y)$ with Equations (7.18), (7.19), (7.20), and (7.21) while fixing the parameters in (7.17a)−(7.17d), and the M-Step to update the parameters in (7.17a)−(7.17d) while fixing the sufficient statistics until some convergence criterion on the expected value of (7.15) is reached.

7.4.4 Filtering, Smoothing, and Predicting the Latent Variable

After the parameters of the HMM have been estimated, we turn to the problem of estimating the probabilities of the most likely hidden state. Given the observation sequence $Y := \{y_0, y_1, \ldots, y_T\}$, the *filtering problem* calculates $\mathbb{P}(q_T|Y)$:

$$\begin{aligned}\mathbb{P}(q_T|y_0,\ldots,y_T) &= \frac{\mathbb{P}(y_0,\ldots,y_T|q_T)\mathbb{P}(q_T)}{\mathbb{P}(y_0,\ldots,y_T)} \\ &= \frac{\alpha(q_T)\mathbb{P}(y_T|q_T)}{\mathbb{P}(y_0,\ldots,y_T)}.\end{aligned} \tag{7.22}$$

Alternatively, the *prediction problem* can be used to predict the probability of the next hidden state at time $T+1$, i.e.

$$\begin{aligned}\mathbb{P}(q_{T+1}|y_0,\ldots,y_T) &= \frac{\mathbb{P}(y_0,\ldots,y_T|q_{T+1})\mathbb{P}(q_{T+1})}{\mathbb{P}(y_0,\ldots,y_T)} \\ &= \frac{\alpha(q_{T+1})}{\mathbb{P}(y_0,\ldots,y_T)}.\end{aligned} \tag{7.23}$$

Lastly, the *smoothing problem* can be solved to ex-post predict the probability of the latent variable at a past time $1 \leq p < T$:

$$\begin{aligned}\mathbb{P}(q_p|y_0,\ldots,y_T) &= \frac{\mathbb{P}(y_0,\ldots,y_T|q_p)\mathbb{P}(q_p)}{\mathbb{P}(y_0,\ldots,y_T)} \\ &= \frac{\alpha(q_p)\mathbb{P}(y_p|q_p)\beta(q_p)}{\mathbb{P}(y_0,\ldots,y_T)}.\end{aligned} \tag{7.24}$$

7.5 Short-Term Load Forecasting

In the following, we describe *online* forecasting algorithms that allow for including knowledge about the estimated latent variables obtained from HMMs and CGMMs into the ML methods introduced in Section 7.2. We make the following

Assumption 14. *The consumption time series Y is stationary, i.e. there are no structural changes in consumption behavior over time.*

7.5.1 Covariates for Prediction

The following observable covariates are used for all forecasting methods:

- Five previous hourly consumptions
- Five previous hourly ambient air temperatures

- A categorical variable for the hour of day for ML methods without latent variable and the CGMM
- A categorical variable interacting the hour of day with the estimated latent variable obtained from HMM for ML methods with HMM

7.5.2 Prediction with Hidden Markov Model

Algorithm 4 Algorithm for Online Prediction with HMM

Input: Training Data $\mathcal{D}_{\text{tr}} := \{(x_t, y_t) : t = 0, \dots, T\}$, Test Data $\mathcal{D}_{\text{te}} := \{(x_t, y_t) : t = T+1, \dots, \tau\}$, ML Method

Initialize all $\mu_1, \dots, \mu_{38}, \sigma_1^2, \dots, \sigma_{38}^2$ suitably
Initialize all a_{ij}, observing (7.13) and Figure 7.2
while $\Delta \mathbb{E}\left[\ell(\theta|\mathcal{D}_c)\right] < \epsilon$ **do**
    Do E-Step: Calculate (7.14) and (7.21) for $t = [0, \dots, T-1]$, $q_t, q_{t+1} = [1, \dots, 38]$ with (7.18)−(7.20)
    Do M-Step: Update HMM parameters with (7.17a)−(7.17d)
end while
Solve smoothing problem (7.24) for $t = 0, \dots, T-1$
Solve filtering problem (7.22) for $t = T$
Round $\mathbb{P}(\hat{q}_0|\mathcal{D}_{\text{tr}}), \dots, \mathbb{P}(\hat{q}_T|\mathcal{D}_{\text{tr}})$ to 0 / 1
Fit ML Method on $\{((x_t, \mathbb{P}(\hat{q}_t)), y_t) : t \in 0, \dots, T\}$
for s in $[T+1, \tau]$ **do**
    Solve prediction problem (7.23) at time s
    Round $\mathbb{P}(\hat{q}_s)$ to 0 / 1
    Predict $\hat{y}_s$ with ML method on covariates $(x_s, \mathbb{P}(\hat{q}_s))$
end for
return $\hat{y}_{T+1}, \dots, \hat{y}_\tau$

Algorithm 4 describes the procedure of fitting an HMM on training data $\mathcal{D}_{\text{tr}}$, which yields estimated latent variables to be used as additional covariates for the ML methods presented in Section 7.2 to perform stepwise prediction on the covariates of the test data $\mathcal{D}_{\text{te}}$. The prediction accuracy of these outcomes is then compared to those outcomes predicted by ML methods that are trained on the training data $\mathcal{D}_{\text{tr}}$ without estimated latent variables in the covariates.

7.5.3 Prediction with Conditional Gaussian Mixture Model

Algorithm 5 describes the online prediction method for a CGMM with $k = 2$ on a given set of training and test data. $\hat{w}$ obtained by OLS is perturbed with zero mean Gaussian Noise ϵ

to obtain the initializations w_1, w_2. Note that this step is necessary to break the symmetry of the update steps (7.10a)−(7.10d), which would keep $w_1 = w_2 = \hat{w}$ unchanged.

Algorithm 5 Algorithm for Online Prediction with CGMM

Input: Training Data $\mathcal{D}_{\text{tr}} := \{(x_t, y_t) : t = 0, \ldots, T\}$, Test Data $\mathcal{D}_{\text{te}} := \{(x_t, y_t) : t = T+1, \ldots, \tau\}$

```
Fit OLS model on D_tr to obtain ŵ
Initialize w_1 ← ŵ + ε
Initialize w_2 ← ŵ + ε
while ΔE[ℓ(θ|D_c)] < ε do
    Update CGMM parameters (7.10a)−(7.10d)
end while
for s in [T+1, τ] do
    Predict ŷ_s with (7.11a) and (7.11b)
end for
return ŷ_{T+1}, ..., ŷ_τ
```

Note that in both Algorithms 4 and 5, the model-specific parameters could be updated after each prediction as more data from the test sequence is observed and hence enters $\mathcal{D}_{\text{tr}}$.

7.5.4 Metric for Forecasting Accuracy

The Mean Absolute Percentage Error (MAPE) of predictions of a set of discrete values $v_i \in \mathcal{V}$ is used to evaluate the accuracy of the predictor:

$$\text{MAPE} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \left| \frac{\hat{v}_i - v_i}{v_i} \right| \cdot 100\%, \tag{7.25}$$

where $\hat{v}_i$ denotes the estimate of v_i.

7.6 Non-Experimental Estimates of DR Reduction

To estimate individual treatment effects, we adopt the potential outcomes framework (Rubin 1974) with binary treatments $T_t \in \{0, 1\}$, where $T_t = 1$ corresponds to a DR intervention at time t, and $T_t = 0$ denotes its absence, hence "control". Let y_t^0 and y_t^1 denote the response (i.e. the electricity consumption) that would be observed if an individual received treatment 0 and 1 at time t, respectively. The goal is to estimate the conditional treatment effect, i.e.

$$\Delta(x) = \mathbb{E}\left[y^1 | x \in \mathcal{X}\right] - \mathbb{E}\left[y^0 | x \in \mathcal{X}\right], \tag{7.26}$$

where x denotes a vector of observable covariates in the covariate space $\mathcal{X}$. Assuming an unconfounded assignment mechanism of treatments to individuals and independency of the

potential outcomes of time, then, conditional on the covariates (see Rubin 1974 for details), the true causal effect of DR, namely $(y_t^0 - y_t^1)$, cannot be found because only one of y_t^0 and y_t^1 can be observed (cf. Fundamental Problem of Causal Inference (Holland 1986)).

Thus, Causal Inference is a "Missing Data Problem". Given the observed treatment outcomes $y_{t_1}^1, \ldots, y_{t_n}^1$, to estimate the true causal effect of treatment, one would require a credible estimate of the *counterfactuals* $\hat{y}_{t_1}^0, \ldots, \hat{y}_{t_n}^0$, i.e. the outcomes in the hypothetical absence of treatment.

To compute such estimates in a non-experimental way, we split the available consumption data into a pretreatment training set with time indices $t \in \mathcal{P}$ consisting of "regular" electricity consumption, i.e. all smart meter readings before the customers' signup date with the DR provider, and a test set with corresponding times $t \in \mathcal{S}$ thereafter which itself consists of smart meter readings during DR hours $\mathcal{T}$ (treatment) and outside DR hours $\mathcal{C}$ (control), hence $\mathcal{S} = \mathcal{T} \cup \mathcal{C}$. Let

$$\mathcal{D}_P = \{\left(x_{i,t}^0, y_{i,t}^0\right) : t \in \mathcal{P}\} \tag{7.27a}$$

$$\mathcal{D}_C = \{\left(x_{i,t}^0, y_{i,t}^0\right) : t \in \mathcal{C}\} \tag{7.27b}$$

$$\mathcal{D}_T = \{\left(x_{i,t}^1, y_{i,t}^1\right) : t \in \mathcal{T}\} \tag{7.27c}$$

denote covariate/outcome pairs for the pretreatment period, the control observations, and the treatment observations of user i, respectively. By fitting any regression model presented in Section 7.2 on the pretreatment training data $\mathcal{D}_P$ of a given user i, and under Assumption 14, applying this model on the treatment covariates $\{x_{i,t}^1 : t \in \mathcal{T}\}$ yields user i's estimated counterfactuals $\{\hat{y}_{i,t}^0 : t \in \mathcal{T}\}$. In particular, Assumption 14 states that DR treatments are interpreted as transitory shocks that do not result in a change in the consumption behavior for $t \in \mathcal{C}$. An elementwise comparison of $\{\hat{y}_{i,t}^0 : t \in \mathcal{T}\}$ and $\{y_{i,t}^1 : t \in \mathcal{T}\}$ yields the pointwise estimated reduction of user i's consumption $\{\hat{y}_{i,t}^\Delta : t \in \mathcal{T}\}$ during each DR event:

$$\{\hat{y}_{i,t}^\Delta : t \in \mathcal{T}\} = \{(\hat{y}_{i,t}^0 - y_{i,t}^1) : t \in \mathcal{T}\}. \tag{7.28}$$

$\hat{y}_{i,t}^\Delta > 0$ corresponds to an estimated reduction of $\hat{y}_{i,t}^\Delta$, and conversely, $\hat{y}_{i,t}^\Delta < 0$ implies an estimated increase by $|\hat{y}_{i,t}^\Delta|$.

7.7 Experiments on Data

We conduct a case study on a data set of a residential DR program including residential customers in the western United States, collected between 2012 and 2014. Aligned with those readings are timestamps of notifications sent by the DR provider to the users that prompt them to reduce their consumption for a short period, typically until the next full hour. A subset of the users have smart home devices that can be remotely shut off by the DR provider with the users' consent. Ambient air temperature measurements were logged from publicly available data sources to capture the correlation between temperature and electricity consumption.

7.7.1 Characteristics of Data and Data Preprocessing

Users with the following characteristics are excluded from the analysis:

- Users with residential solar photovoltaics (PV)
- Users with corrupt smart meter readings, i.e. unrealistically high recordings

The consumption series of the remaining users are then aligned with available temperature readings and mapped to the range $[0, 1]$ to be able to compare users on a relative level. The temperature data is standardized to zero mean and unit variance. Lastly, the pretreatment data is tested for stationarity with the augmented Dickey-Fuller Test (Fuller 1995) to assert, with a significance level of more than 99%, the absence of a unit root, which motivates Assumption 14.

7.7.2 Experiments on Semi-Synthetic Data

As only one of $\{y^0_{i,t}, y^1_{i,t}\}$ for a given user i at time t can be observed, we construct semisynthetic data for which both values and hence the true causal effect $(y^0_{i,t} - y^1_{i,t})$ are known. This allows us to evaluate the accuracy of predicted counterfactual consumptions and the ensuing non-experimental estimates of DR reduction (7.28). For this purpose, we take actual pretreatment training data $\mathcal{D}_P$ (7.27a) for each user i, which is free of any DR messages. Next, we split this training set into two pieces by introducing an artificial signup date $\tilde{t}$ valid across all users. We thus obtain a synthetic training set $\tilde{\mathcal{D}}_P = \{\left(x^0_{i,t}, y^0_{i,t}\right) : t \in \mathcal{P}, t < \tilde{t}\}$ and a synthetic test set $\tilde{\mathcal{D}}_S = \{\left(x^0_{i,t}, y^0_{i,t}\right) : t \in \mathcal{P}, t \geq \tilde{t}\}$ for user i. Next, a random subset $\tilde{\mathcal{T}}$ of all available time indices in the synthetic test set $\tilde{\mathcal{D}}_S$ between 6 a.m. - 8 p.m. is assigned a synthetic treatment, for which the respective consumption is decreased by a random value $\in [0, \bar{c}]$. By doing so, both the treatment and control outcomes for $t \in \tilde{\mathcal{T}}$ become available, and so we obtain the semisynthetic data set

$$\tilde{\mathcal{D}}_{\tilde{T}} := \{\left(x^0_{i,t}, y^0_{i,t}, y^1_{i,t}\right) : t \in \tilde{\mathcal{T}}\}. \tag{7.29}$$

Thus, any non-experimental estimate of the DR treatment effect for $t \in \tilde{\mathcal{T}}$ can be benchmarked on the known (synthetic) counterfactual $\{y^0_{i,t} : t \in \tilde{\mathcal{T}}\}$.

This semisynthetic data set is used for two purposes. First, we evaluate the MAPE (7.25) of the estimators from Section 7.2, with and without latent variables. This is done by training them on $\mathcal{D}_{\text{tr}} = \tilde{D}_P$ and testing on $\mathcal{D}_{\text{te}} = \tilde{D}_S$, which yields out-of-sample counterfactual consumptions $\{\hat{y}^0_{i,t} : t \in \tilde{\mathcal{T}}\}$ across all users i, see Algorithms 4 and 5. Second, we conduct a comparison of the eventwise errors of estimated DR reductions for all ML methods with the HMM latent variable (CGMM is not considered further), which, for a given user i at time t, are obtained as follows:

$$\hat{y}^{\Delta}_{i,t} - y^{\Delta}_{i,t} = \left(\hat{y}^0_{i,t} - y^1_{i,t}\right) - \left(y^0_{i,t} - y^1_{i,t}\right) = \hat{y}^0_{i,t} - y^0_{i,t}. \tag{7.30}$$

The ground truth counterfactual $y^0_{i,t}$ is available for the semisynthetic data (7.29) by construction, but would be unavailable for real-world data.

Figure 7.3 shows a boxplot of the distribution of average MAPEs across users for the prediction methods introduced in Section 7.2 with and without the latent variable from HMM, and for the CGMM (Section 7.3). It can be seen that the information about the latent

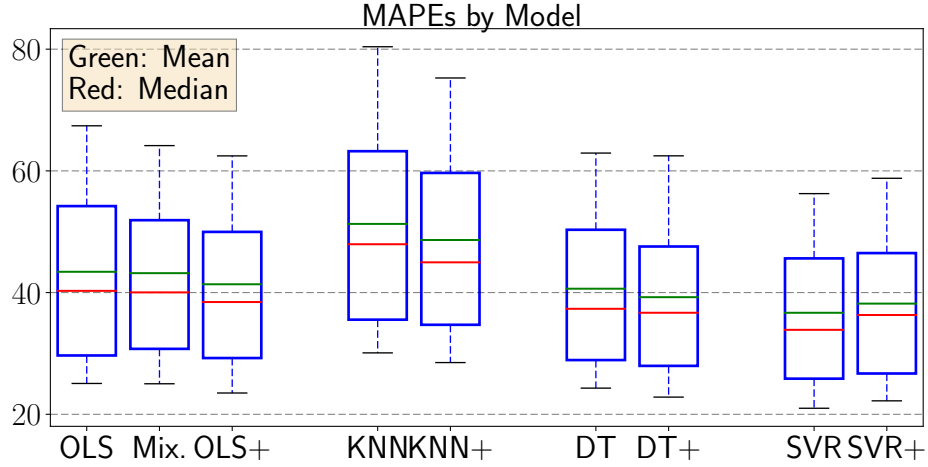

Figure 7.3: Prediction Accuracy by Forecasting Method. "+" Signifies Model with HMM Latent Variable, "Mix." Denotes CGMM. Blue Boxes Span 25-75th Percentile, Whiskers 10-90th.

variable improves the prediction accuracy in all cases but SVR. Further, the lower MAPE obtained with DT and SVR is consistent with the findings in Mirowski et al. 2014; Taylor and McSharry 2007. The higher MAPE for KNN compared to OLS can be explained by the different magnitudes of the covariates introduced in Section 7.5.1, which gives categorical variables disproportionate weight. The CGMM performs better than OLS, but worse than OLS with the latent variable. Note that other more sophisticated predictors (e.g. Neural Networks) have lower MAPEs at the cost of longer computation times and potential loss of interpretability, but are likely to show a similar improvement in terms of MAPE by incorporating information about the estimated latent variable as the amount of training data increases. For a comparison between the prediction accuracy of state-of-the-art estimators, the reader is referred to Mirowski et al. 2014; Taylor and McSharry 2007 for further information.

Figure 7.4 shows histograms of eventwise prediction errors (7.30) in the estimated DR reduction for single events and across all users i. Green bars and red bars signify prediction errors of forecasting methods that do and do not make use of the estimated latent variable from HMM, respectively. Aligned with these plots are the sample mean and covariance of the errors for the models that take the latent variable into account. The bias-variance decomposition

$$\mathbb{E}\left[\left(\hat{y}^{\Delta}_{i,t} - y^{\Delta}_{i,t}\right)^2\right] = \mathrm{Bias}(\hat{y}^{\Delta}_{i,t})^2 + \mathrm{Var}(\hat{y}^{\Delta}_{i,t}) + \epsilon, \tag{7.31}$$

where ϵ denotes the irreducible error, is invoked in the following. Noting that $\hat{\mu}$ and $\hat{\sigma}^2$ in Figure 7.4 correspond to the bias and variance in (7.31) from the model with latent variable from HMM, the tradeoff becomes clear when comparing OLS, DT, and SVR. A lower variance of DT and SVR comes at the cost of a higher bias. For KNN, both bias and variance are larger than in OLS, which is explained by the poor predictive performance of KNN (see Figure 7.3). For a subsequent analysis of individual treatment effects (ITEs), we choose the least biased estimator that uses latent variables, in our case OLS, despite its higher overall prediction error compared to SVR and DT. This follows the consensus of econometrics concerned with estimating ITEs (Bound, Jaeger, and Baker 1995; Angrist and Pischke 2009).

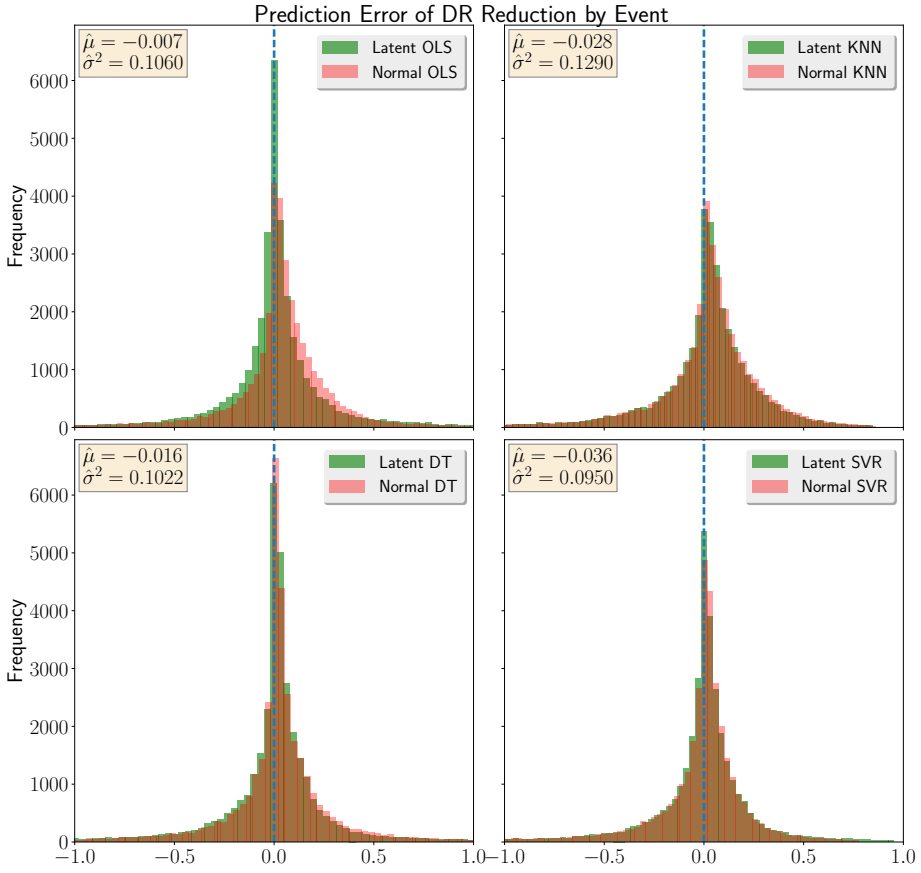

Figure 7.4: Pointwise Prediction Error of DR Treatment Effect on the User Level; Bias $\hat{\mu}$ and Variance $\hat{\sigma}^2$ of Model with Latent Variable from HMM.

7.7.3 Experiments on Actual Data

In the following, we analyze ITEs for users with and without smart home devices. The analysis of reduction is carried out with OLS that utilizes an estimate of the HMM latent

state because it is found that this method has the lowest bias on semisynthetic data (see Figure 7.4).

In the real data case, only the treatment outcomes $\{y^1_{i,t} : t \in \mathcal{T}\}$ for user i are observed during DR events, and so the counterfactuals $\{\hat{y}^0_{i,t} : t \in \mathcal{T}\}$ are predicted to calculate a non-experimental estimate of the DR reduction (7.28). Using Algorithm 4 on the pre-signup data $\mathcal{D}_P$ (7.27a) as training data $\mathcal{D}_{\text{tr}}$ for each user and $\mathcal{D}_{\text{te}} = \mathcal{D}_C \cup \mathcal{D}_T$ (7.27b), (7.27c), the pointwise reductions across all users and each treatment $t \in \mathcal{T}$ are calculated. Figure 7.5 shows boxplots of estimated DR reductions conditional on (a) the hour of day, (b) users with and without smart home devices, and (c) the predicted latent states. The gray bars represent "placebo" events (i.e. a subset of hours $t \in \mathcal{C}$ outside DR treatments hours, but after the signup date) estimated by the same model.

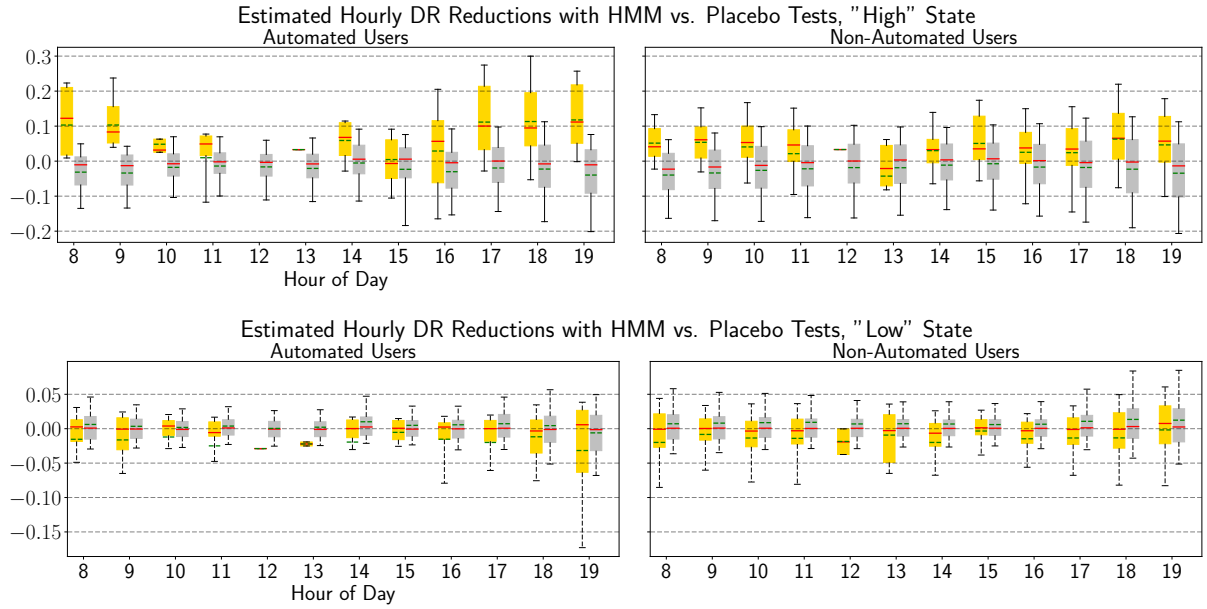

Figure 7.5: Estimated Reduction Across Users by Hour of Day (Yellow) vs. Estimated Reduction for Placebo Events (Gray) for Automated and Non-Automated Users Conditional on Estimated Latent Variable. Red: Median, Green: Mean. Blue Boxes Span 25-75th Percentile, Whiskers 10-90th.

Figure 7.5 gives rise to two observations: First, the estimated reduction conditional on the "high" latent state is greater in magnitude for users with smart home devices, following the intuition that the "high" state describes the operation of smart home devices which can be conveniently shut off during DR hours. In contrast, the lower estimated reductions of regular users during "high" latent states might reflect the additional hassle cost that incurs for users to manually reduce their consumption. Second, the estimated reductions for both users with and without smart home devices and conditional on the "low" latent state show

mean reductions around zero, contrary to the expectation of a small positive reduction. This might indicate the existence of a threshold representing the standby consumption of users, below which it is hard or impossible to reduce consumption further.

This finding could be particularly meaningful to the DR provider, as it presents a recommendation as to when to call DR events and for which users, which could improve the allocative efficiency of DR targeting and be a stepping stone towards calculating optimal bids.

7.8 Conclusion

We developed non-experimental estimators from Machine Learning for estimating ITEs of Residential Demand Response and showed that incorporating a latent variable, either with a Conditional Gaussian Mixture Model or a Hidden Markov Model, allows for an improvement in prediction accuracy. This Bayesian approach is motivated by the need to obtain interpretable and physically meaningful results capturing the users' electricity consumption behavior. We then tested the forecasting algorithms on semi-synthetic data to find that Ordinary Least Squares in conjunction with a latent variable produces the least biased estimator for DR reduction. Lastly, this estimator was applied on a residential DR data set to determine hourly reductions of electricity consumption for both users with and without automated electric devices. The highest reductions were found to be among users with home automation devices during "high" estimated latent states, which in turn provides a recommendation for DR providers for targeting purposes, i.e. to focus on automated users for the highest yield in reduction.

This chapter provides only a foundation for more profound analyses in the area of Residential Demand Response. In particular, latent variables can be added as an additional covariate to more computationally demanding estimators, for instance Neural Networks or Random Forests, in order to assess the gain in forecasting precision with latent variables. This is an area to be explored by the established area of STLF, which has traditionally been focusing on maximizing the precision of forecasting algorithms. Further, various extensions to modeling the HMM are worth exploring, such as enlarging the state space of the Markov Chain to enforce a dependency on more than just the previous hour, or increasing the number of hidden states for a given hour (i.e "low", "medium", and "high" consumption). Lastly, the estimated latent variable could be related to a measure of occupancy in residential dwellings, and so a validation of the estimated latent states on ground truth data on occupancy would be interesting if privacy concerns could be overcome.

Chapter 8

Evaluation of a Residential Demand Response Program

8.1 Introduction

The final chapter of this thesis is concerned with the evaluation of a residential Demand Response program in California. The topics discussed in Chapter 7 have already foreshadowed the central theme of this chapter, namely the estimation of individual loads for causal inference.

Cyber-Physical Systems (CPS) utilize intelligent mechanisms to improve the efficiency and usability of the interacting elements and are often augmented by human capabilities to achieve a desired goal. However, recent advances in CPS emerging from societal-scale infrastructure have been limited by the lack of good models for human interaction with systems. Significant efforts to introduce such models exist in transportation or energy, which is the case studied here. We study the effect of incentivizing people to participate in Demand Response to reduce electricity consumption with a Randomized Controlled Trial (RCT). While our application is specific, our techniques employed in this thesis are generalizable to any societal-scale CPS which seeks to incentivize users to achieve a desired human behavior.

Following the 1970s energy crisis, programs for demand-side management (DSM) (Palensky and Dietrich 2011) were introduced on a global scale. Such programs seek to temporarily reduce consumers' electricity usage through financial incentive schemes during periods of electricity supply shortage. These programs are enabled by the integration of information and communications technology in the electric grid, which has inspired a large body of research aiming at better understanding the interaction between the consumption habits of consumers, load-serving entities and the electric wholesale market while taking into account constraints of the electric grid operation. Energy supply is highly inelastic due to the slowness of power plants' output adjustment, which causes small increases and decreases in demand to result in a price boom or bust, respectively. This issue is exacerbated by the variable nature of electricity demand (mainly influenced by ambient temperatures (Pardo, Meneu,

and Valor 2002)), prohibitively costly energy storage, and steady growth of renewable, yet volatile electricity generation.

Despite the fact that electric utilities and generating companies hedge against such price fluctuations through long-term contracts, a large portion of electricity remains to be procured through the wholesale electricity market. Since utilities are obligated to provide end-users with electricity at a quasi-fixed tariff at all times (Federal Energy Regulatory Commission 2016), e.g. Time-of-Use pricing, they have to bear price risks. Therefore, DSM is an attempt to protect utilities against such price risks by partially relaying them to end-users, which increases market efficiency according to the economic consensus (Borenstein, Jaske, and Rosenfeld 2002a; Borenstein 2005; Borenstein and Holland 2005).

DSM describes a set of interventions aiming at affecting customer behavior on different scales of application and time (Palensky and Dietrich 2011). Examples include programs on a utility-wide scale, community scale, or individual household scale, for the purpose of long-term energy efficiency or short-term demand interventions. Previous studies in residential DSM have mainly focused on critical-peak pricing and real-time pricing, using *day-ahead* notifications that were sent often only during selected seasonal periods of the year (Jessoe and Rapson 2014; Wolak 2010; Allcott 2011). In contrast, this thesis focuses on extremely short-term, *hour-ahead* behavioral interventions during hours of peak demand or shortages of electricity supply when demand reductions can counteract high electricity prices reflected by Locational Marginal Prices (LMPs) (Hong and Hsiao 2002). Since electric wholesale markets tend to fluctuate more on an hour-ahead rather than day-ahead level, such *hour-ahead* notifications could have more leverage to deliver welfare effects compared to *day-ahead* interventions.

The California Public Utilities Commission (CPUC) has launched a Demand Response Auction Mechanism (DRAM) in July 2015 (*Public Utilities Commission of the State of California: Resolution E-4728. Approval with Modifications to the Joint Utility Proposal for a Demand Response Auction Mechanism Pilot* 2015) which requires utilities to procure a minimum monthly amount of reduction capacity from Demand Response (DR) aggregators. The real-time market determines electricity prices by matching demand and utilities' supply curves subject to the procured capacity. A utility whose bid is cleared then asks the DR provider to incentivize its customers to temporarily reduce their consumption relative to their projected consumption without intervention. This is the counterfactual, referred to in this context as *baseline*, based on which compensations for (non-)fulfilled reductions are settled: If the consumer uses less (more) energy than the baseline, she receives a reward (incurs a penalty). In a similar fashion, if the aggregator falls short of delivering the promised load reduction, it incurs a penalty. For a profit-maximizing bid, the DR provider needs to estimate the counterfactual consumption as precisely as possible, among other aspects such as the size of its customer base, the LMP, and the elasticity of users' demand in response to incentives.

The estimation of the actually delivered reduction both on the household and aggregation level arguably is the most critical component of the DR bidding process. If the reductions are estimated with a biased counterfactual, either the DR provider or the utility clearing the

bids is systematically discriminated against. If the baseline is unbiased but plagued by high variance, the profit settlement is highly volatile. Existing baselines employed by major power grid operators in the United States (e.g. California Independent System Operator (CAISO), New York ISO) are calculated with simple arithmetic averages of previous observations (*California Independent System Operator Corporation (CAISO): Fifth Replacement FERC Electric Tariff* 2014) and therefore are inaccurate. Estimating more accurate baselines with non-experimental, unbiased estimators on the one hand and an experimental control group within a RCT on the other hand, is a significant contribution of this thesis.

8.1.1 Contributions

We estimate the *average treatment effect* (ATE) of hour-ahead notifications on the reduction of electricity consumption by evaluating a Randomized Controlled Trial (RCT) on ≈ 5000 residential households in California serviced by the three main electric utilities (PG&E, SDG&E, SCE). This experiment is funded by the California Energy Commission, which − to the best of our knowledge − is the first one to experiment with hour-ahead notifications on a residential household level. We estimate an ATE of -0.13 kWh per DR Event and user and further discover notable geographic and temporal heterogeneity among users, as the largest estimated reductions occur in summer months as well as in regions with warmer climate, suggesting that air conditioning units play a decisive role in DR programs.

In addition to this experimental approach, we also develop a *non-experimental method* for estimating this causal effect on an *individual* user level, which is easily aggregated into an ATE, thereby making sense of the underlying heterogeneity of residential customers, an approach that crucially does not require an experiment. Importantly, we utilize experimental observations and the existence of a control group as a benchmark for the non-experimental ATE estimate and find that the results in both cases are close to each other. Interestingly, the non-experimental approach even achieves tighter confidence intervals of the estimated causal effect. Motivated by these results, we claim that our method is applicable to estimating treatment effects in any setting with high-frequency time-series data whenever an RCT is hard to conduct, for example due to budget or ethical constraints.

Furthermore, we design an *adaptive targeting* method, which exploits heterogeneity in users' responses to incentive signals to assign differing price levels to different subsets of the treatment population. Specifically, we separate users based on their previous responses into two distinct groups, each of which either only receives low or high incentives. Using this partitioning method, we observe an increase of the per-dollar yield of $\approx$43%. Taken together, we bring together ideas at the intersection of Machine Learning (ML), economics, and energy markets. The observational data is provided by the company `OhmConnect, Inc.` (*OhmConnect, Inc. 2015* 2017), headquartered in the San Francisco Bay Area, and is being held under a confidentiality agreement.

This chapter unfolds as follows: In Section 8.2, we describe the market setting for Residential Demand Response. Section 8.3 explains the experimental setup and provides summary statistics on the RCT data. We then develop the non-experimental estimation framework in

Section 8.4, where we pay particular attention to estimation bias and empirical de-biasing methods (Section 8.4.3).

The non-experimental estimation results are provided in Section 8.5, both on an individual and aggregate level. We discover notable geographic and temporal heterogeneity among users. That is, the largest estimated reductions occur during early afternoon hours and early evenings, as well as in regions with warmer climate, suggesting that air conditioning units play a decisive role in DR programs. Next, in Section 8.6 we estimate the ATE using a classical Fixed-Effects Estimator from econometrics (Diggle et al. 2013). By using varying regression specifications, we estimate the demand curve of Demand Response (Section 8.6.1) and conditional ATEs by hour of the day (Section 8.6.2), month of the year (Section 8.6.3), or smart home automation status of users (Section 8.6.4). Section 8.7 compares the estimates obtained by both approaches. The effect of adaptive targeting on cost efficiency is discussed in Section 8.8. Lastly, the role of moral suasion and non-monetary incentives is explored in Section 8.9. Section 8.10 concludes. Additional results including summary statistics of the data set, estimation results, and supporting numeric data are relegated to Appendix B.1.

8.1.2 Related Work

LaLonde 1986 provides the first (unsuccessful) benchmarking comparison between non-experimental and experimental estimates using data from the National Supported Work Demonstration. Later studies (Dehejia and Wahba 1999; Smith and Todd 2005) found that model misspecifications in (LaLonde 1986) were responsible for the mismatch, and that propensity score estimates were able to correct them. The authors argue that there exists no general framework to ensure validity of non-experimental estimates, which therefore requires each non-experimental estimate to be benchmarked against an experiment. This motivates our thesis.

Given the rapid growth of collected user data, non-experimental estimates become more and more valuable. Moreover, there are situations where RCTs, the experimental standard, are infeasible to conduct, e.g. due to budget or ethical constraints. These facts have spurred research at the intersection of machine learning and economics, whose general idea is to partition observations under treatment and control in order to fit a nominal model on the latter set, which, when applied on the treatment set, yields counterfactual estimates, from which the treatment effect is computed by subtracting out actual observed treatment outcomes.

Examples for such nominal models are found in Bollinger and Hartmann 2015, who evaluates welfare effects of home automation by calculating the Kolmogorov-Smirnov Statistic between users, which are then used as weights for kernel-based non-parametric regression. In Abadie, Diamond, and Hainmueller 2012, a convex combination of US states is computed as the counterfactual estimate for tobacco consumption to estimate the effect of a tobacco control program in California on tobacco consumption. In Athey and Imbens 2016; Wager and Athey 2016, the estimators are random forests trained by recursive partitioning of the feature space and novel cross-validation criteria. Brodersen et al. 2015 develops Bayesian structural time series models combined with a Monte-Carlo sampling method for treatment

effect inference of market interventions. In Zhou, Balandat, and Tomlin 2016b and Zhou, Balandat, and Tomlin 2016a, the authors estimated reductions of a DR program on a small dataset of 500 users in California with classical ML regression methods.

Fitting an estimator on smart meter time-series is essentially a short-term load forecasting (STLF) problem, whose goal is to fit estimators on observed data to predict future consumption with the highest possible accuracy. The two main directions are based on time-series modeling and classical regression analysis. Within STLF, tools employed are ARIMA models with a seasonal component (Taylor and McSharry 2007; Soares and Medeiros 2008) and classic regression models where support vector regression (SVR) (Pai and Hong 2005; Hippert, Pedreira, and Souza 2002) and neural networks yield the highest accuracy (Senjyu et al. 2002; Elattar, Goulermas, and Wu 2010). A comprehensive comparison between ML techniques for forecasting and differing levels of load aggregation is provided in Mirowski et al. 2014. Other methods include Kernel Density Estimation (Arora and Taylor 2016) and fuzzy methods (Song et al. 2005).

In the context of smart meter data mining, much of the existing work focuses on disaggregation of energy consumption to identify contributions of discrete appliances from the total observed consumption (Chen et al. 2011; Fei et al. 2013) and to learn consumption patterns (Molina-Markham et al. 2010; Zhou, Balandat, and Tomlin 2016a). Studies in applied economics typically emphasize the estimation of ATEs of experimental interventions. To increase precision of the estimates, the employed regression models often employ unit-level fixed effects (Allcott 2011; Jessoe, Miller, and Rapson 2015), which is an implicit way of training models for the consumption of individual consumers. In this work, we make these user-level models explicit, allowing for more general ML techniques. Importantly, our approach is original as it permits to perform causal inference on the level of *individual* treatment effects in a straightforward fashion by employing estimators from STLF. To the best of our knowledge, this thesis is the first of its kind to analyze the potential of Demand Response interventions on a residential level, combining ideas at the intersection of causal inference from econometrics and Machine Learning for estimation.

8.2 Demand Response Mechanism

According to DRAM (*Public Utilities Commission of the State of California: Resolution E-4728. Approval with Modifications to the Joint Utility Proposal for a Demand Response Auction Mechanism Pilot* 2015), electric utilities are obligated to offer "demand flexibility" through Demand Response Providers (DRPs). Utilities solicit bids from DRPs and accept the highest ones up to a monthly target capacity. In the real-time wholesale electricity market, the utility submits supply bids including these acquired capacities, which, when cleared, have to be delivered by the DRP under contract over a contractually specified period of time. The DRP does so by eliciting reductions among a suitably chosen subset of its residential end-use customers by offering them a monetary incentive. Such an aggregation of users is also known as a *Proxy Demand Resource* product (PDR) (*California Independent System*

Operator Corporation (CAISO): Fifth Replacement FERC Electric Tariff 2014). The DRP receives a payment from the wholesale market for each unit of reduction up to its original capacity bid, but incurs a shortfall penalty for each unit of unfulfilled obligation. Figure 8.1 illustrates the interaction between all agents.

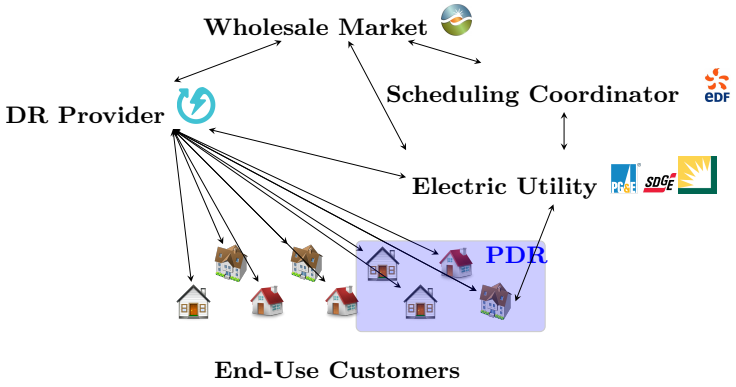

Figure 8.1: Interactions of Agents in Residential Demand Response

We focus on the DRP-User interaction and in particular answer the question of how to measure and quantify reductions of end-users' electricity consumption in response to monetary incentives. The regulatory standard in California measures reductions with the CAISO 10-in-10 baseline (*California Independent System Operator Corporation (CAISO): Fifth Replacement FERC Electric Tariff* 2014), which computes the estimated reduction for a particular user at a particular time. During a DR event, the materialized consumption is compared to the estimated baseline. The user is rewarded (penalized) the difference between those two multiplied with a reward level. Due to the inherent noisiness of the CAISO baseline, we present alternative approaches to estimate reductions (namely the individual level non-experimental estimates and the fixed effects regression, which implicitly uses the control group as a counterfactual).

Our data set consists of DR events of length one hour. Specifically, users receive notifications of a DR event up to 20 minutes into an hour, which lasts until the end of the hour. The notifications sent during the RCT communicate the incentive level to the user, which was chosen from the set $\mathcal{R} = \{0.05, 0.10, 0.25, 1.00, 3.00\} \frac{\text{USD}}{\text{kWh}}$.

8.3 Experimental Setup and Data Characteristics

8.3.1 Setup of the Experiment

The experiment is carried out by OhmConnect, Inc., using funds provided by the California Energy Commission. Figure 8.2 draws a flowchart of the experimental setup.

Over the course of the experimental time period (Nov. 2016 - Dec. 2017), each consumer that signs up for the study is *randomly* assigned to one of the following groups:

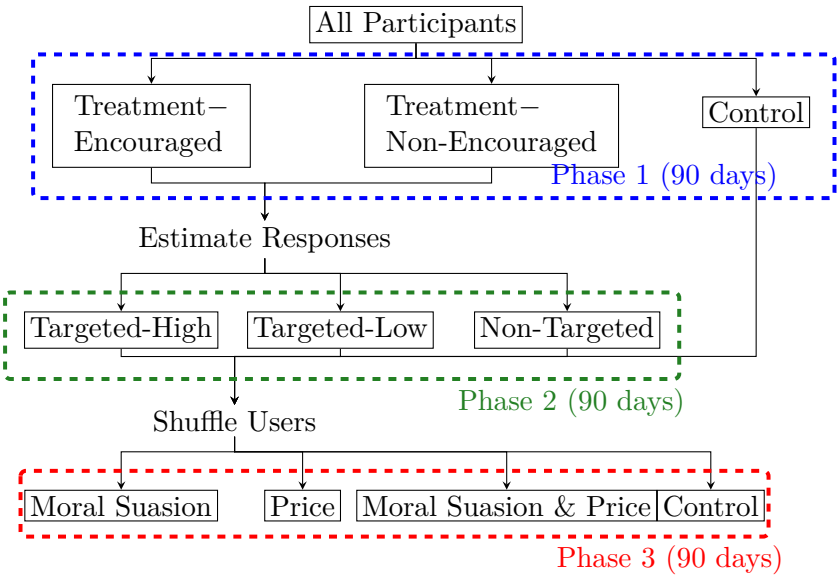

Figure 8.2: Setup of Experiment

- Treatment−Encouraged: The user receives an average number of 25 DR events in the 90 days following the signup, with incentive levels being randomly chosen from the set $\mathcal{R} = \{0.05, 0.25, 0.50, 1.00, 3.00\}\frac{\text{USD}}{\text{kWh}}$. Additionally, the user is given a rebate for purchasing a smart home automation device.
- Treatment−Non-Encouraged: Same as in Treatment-Encouraged, but without smart home automation rebate.
- Control: Users do not receive any notifications for DR events for a period of 90 days after sigup.

Taken together, these three groups form Phase 1 of the experiment. Users in the control group that have reached 90 days of age are removed from the study. Users in either the Treatment−Encouraged or Treatment−Non-Encouraged groups that have reached 90 days of age are pooled and systematically assigned to one of the following groups for Phase 2 interventions:

- Targeted-High: The user receives an average number of 25 DR events for a period of 90 days after being rolled over into Phase 2. Each reward level is randomly drawn from the set $\{1.00, 3.00\}\frac{\text{USD}}{\text{kWh}}$.
- Targeted-Low: Same as in Targeted-High, but rewards are randomly drawn from $\{0.05, 0.25, 0.50\}\frac{\text{USD}}{\text{kWh}}$.
- Non-Targeted: Same as in targeted groups, with rewards drawn from the complete set, i.e. $\{0.05, 0.25, 0.50, 1.00, 3.00\}\frac{\text{USD}}{\text{kWh}}$.

Lastly, users with completed Phase 2 as well as Phase 1 control users after 90 days of age undergo Phase 3 (moral suasion), which occurs on an event-by-event level. Specifically, users receive one of four treatment types selected uniformly at random:

- Moral Suasion Only: Message with environmental content
- Price Only: Message with monetary incentive of 1 USD/kWh
- Moral Suasion and Price: Combination of above
- Control: User receives no notification

Users with completed Phase 3 (or synonymously all users in Treatment−Encouraged and Treatment−Non-Encouraged after 180 days of treatment) are removed from the study. In Sections 8.4-8.6, we evaluate Phase 1 of the experiment whereas Sections 8.8 and 8.9 are dedicated to adaptive targeting (Phase 2) and moral suasion (Phase 3), respectively. In the remainder of this chapter, we use the term "treatment users" to refer to users in the "Treatment-Encouraged" and "Treatment-Non-Encouraged" group.

Recruitment began on 11/14/2016 with the launch of the pilot and concluded on 8/14/17. The pilot period recruitment ran from 11/14/2016 to 12/31/16 and the study period recruitment from 1/1/2017 to 8/14/2017. Due to a technical implementation problem, the recruitment period was cut short two weeks and ended prior to the originally planned 9/1/2017. While this represents an unfortunate loss in data, it did not seriously affect the statistical power of the study.

Figure 8.3 illustrates the number of study participants that were recruited for the RCT broken out by time of enrollment. This is done separately for users assigned to the three different experimental groups of Phase 1 (Control, Encouraged, Non-Encouraged). Recruitment began on November 15, 2016 and ended on August 15, 2017. We observe lower enrollment figures from April 2017 – June 2016 with a noticeable peak towards the end of the recruitment period. As can be seen from the figure, the height of the red and green bars for a particular vertical slice appear to have approximately the same height, indicating that encouraged and non-encouraged users are balanced in size. In contrast, the blue bar is about half as large as the green or red bar, which is consistent with the 40/40/20 assignment of users into encouraged, non-encouraged, and control groups.

In a similar fashion, Figure 8.4 plots the number of users that were recruited into the study and successfully connected their electric utility accounts. About half of all recruited users connected their utility accounts. We were unable to use the recruited users who did not connect their utility accounts because we have no energy data for them. We observe that the shape of the boxplot looks similar to the one in Figure 8.3, suggesting that users across the three different experimental groups were equally likely to connect their electric utility accounts. The average number of recruits per day was 58 with a standard deviation of 48, a minimum of 5, and a maximum of 295.

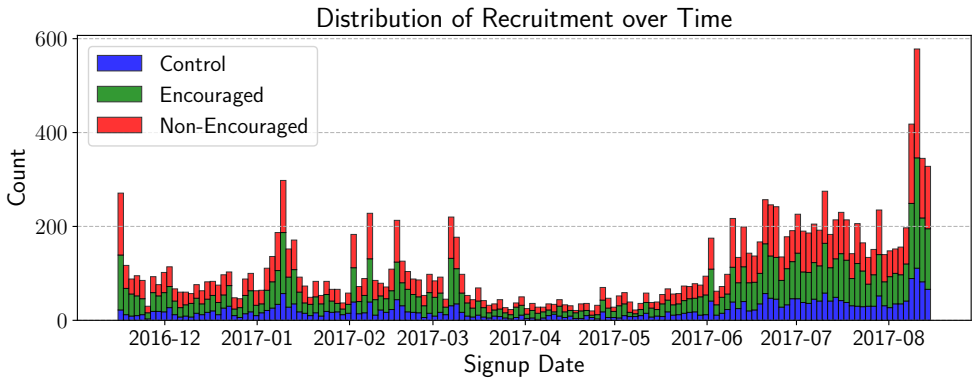

Figure 8.3: Distribution of Enrollment over Time

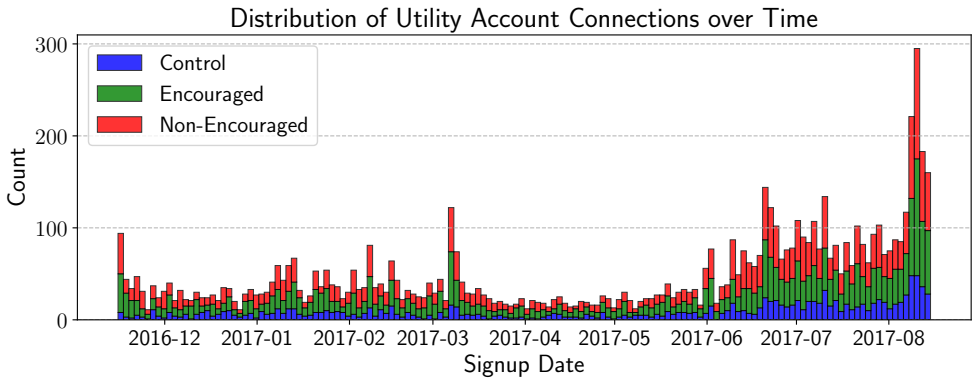

Figure 8.4: Distribution of Utility Account Connections over Time

8.3.2 Summary Statistics

Figure 8.5 illustrates the distribution of the lengths of available historical smart meter data among all users that have successfully connected their utility accounts. Users from Southern California Edison (SCE) have the shortest availability and those serviced by San Diego Gas & Electric (SDG&E) have the longest. We observe peaks at 365 days and 730 days, which correspond to 1 or 2 years of data availability. The black dashed lines reflect the median availability of historical smart meter data, which is 374 days for PG&E, 273 days for SCE, and 403 days for SDG&E.

Figure 8.6 provides a scatter plot of the geographic distribution of control, encouraged, and non-encouraged users broken out by electric utility. As expected, most users are concentrated in the urban areas of the San Francisco Bay Area, San Diego, and Los Angeles. It is visually striking that there appear to exist no structural differences in the distribution of users across either treatment group or electric utility, which is an intuition to be confirmed in the balance checks provided in Section 8.3.4.

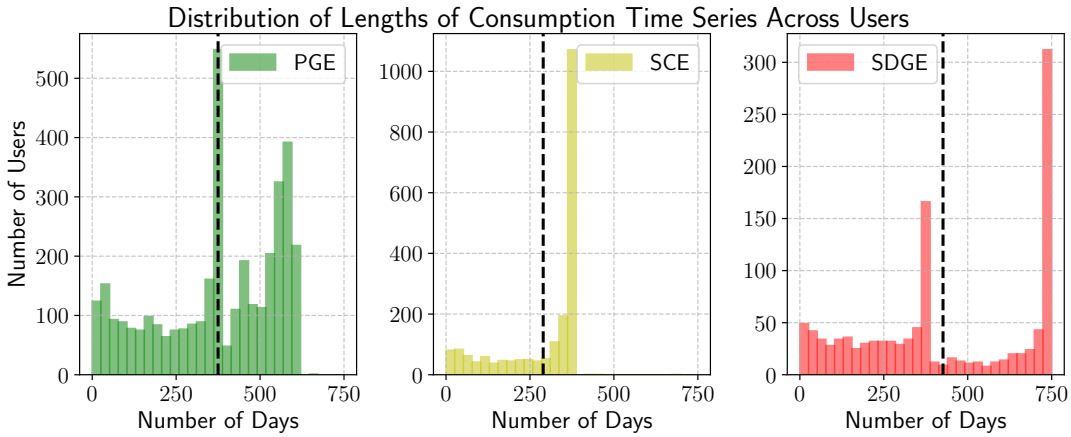

Figure 8.5: Availability of Smart Meter Data Across Experimental Users

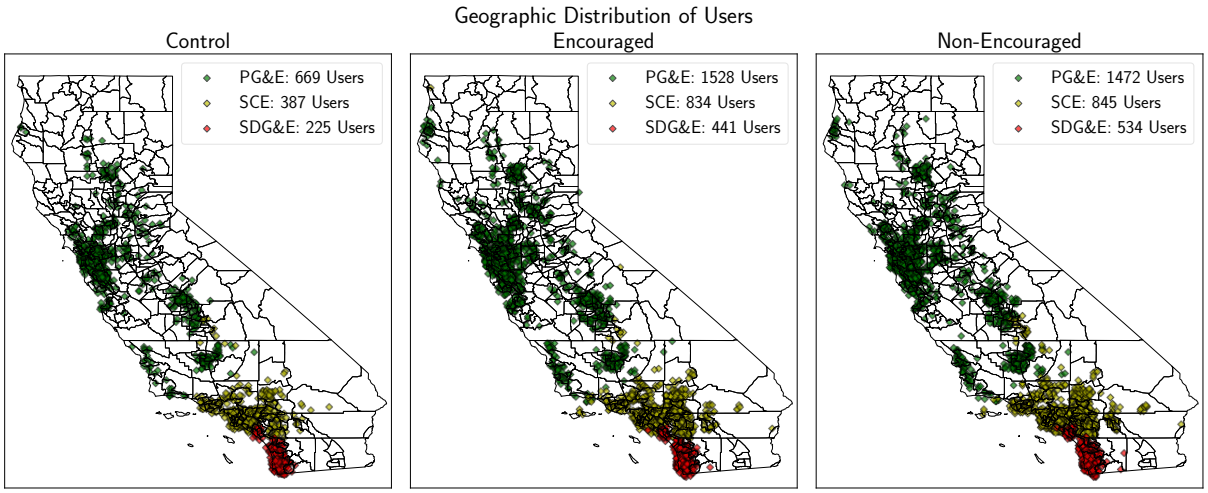

Figure 8.6: Geographic Distribution of Enrolled Users

Table 8.1 reports the number of users by experiment group and proportion of users for which we were able to scrape historical smart meter reading data. The table shows that the randomized assignment of users to groups roughly follows a 1:2:2 ratio (Control vs. Treatment−Encouraged vs. Treatment−Non-Encouraged).

Historical Smart Meter Data Availability by Group			
Group	# Enrolled	# With Data	# With DR
Control	3266	1311	–
Treatment−Encouraged	6735	2873	2389
Treatment−Non-Enc.	6689	2910	2402

Table 8.1: Number of Total Users Enrolled by Group, Data Availability, and Users with DR Events as of September 28, 2017.

Users without DR events or for which we were unable to scrape historical data are omitted from the study. Since the assignment of users into the different experimental groups was randomized (see Section 8.3.4), dropping such users does not affect the evaluation of the experiment. Hence, in the remainder of this chapter, we omit the attribute "with historical data" when referring to users in the experiment for the sake of brevity, unless otherwise stated.

Figure 8.7 shows the geographic distribution of the remaining users across California. More than half of all study participants are serviced by Pacific Gas & Electric. The remaining users are covered by Southern California Edison (Los Angeles area) and San Diego Gas & Electric.

Figures 8.8 and 8.9 illustrate the number of DR messages received per individual over time.

Table 8.2 shows the distribution of Phase 1 event incentive levels for DR events among the 5906 remaining users with historical data and DR events. There is a total of $145,838$ messages, which equates to ≈ 24.7 Phase 1 messages per user.

DR Events for Encouraged and Non-Encouraged Users in Phase 1			
Event Incentive Level	Encouraged	Non-Enc.	Control
0.05 USD/kWh	14422 (19.9%)	14666 (20.0%)	–
0.25 USD/kWh	14575 (20.1%)	14637 (20.0%)	–
0.50 USD/kWh	14426 (19.9%)	14663 (20.0%)	–
1.00 USD/kWh	14572 (20.1%)	14611 (19.9%)	–
3.00 USD/kWh	14562 (20.0%)	14704 (20.1%)	–
Total	72557 (100%)	73281 (100%)	–

Table 8.2: Number of Phase 1 DR Events for Encouraged and Non-Encouraged Users by Incentive Level

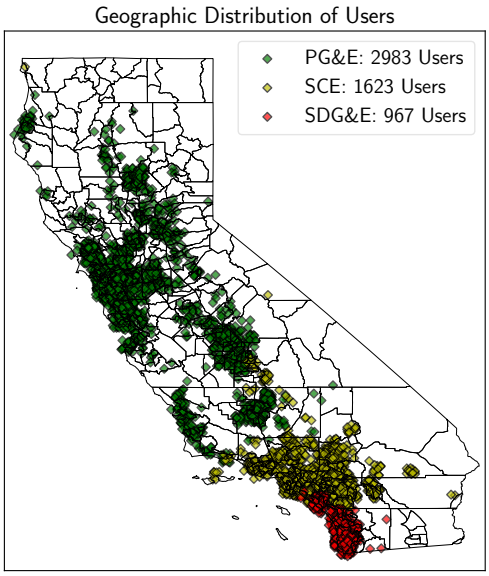

Figure 8.7: Geographic Distribution of Users

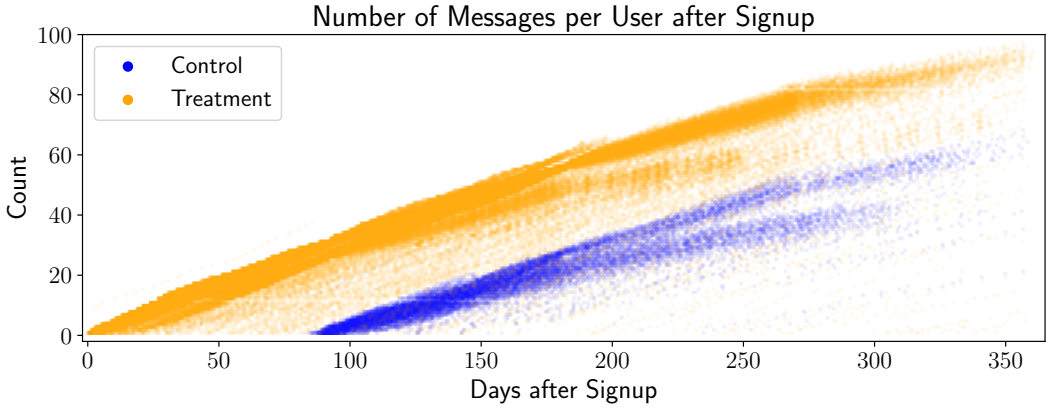

Figure 8.8: Number of Messages Sent to Users

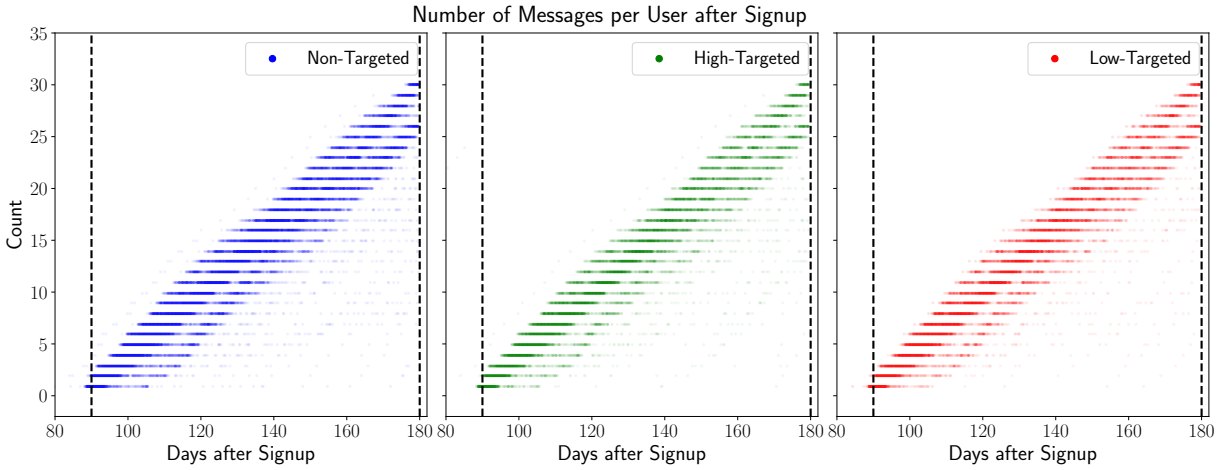

Figure 8.9: Number of Phase 2 DR Events per Individual Over Time

In a similar fashion, Table 8.3 shows the breakdown of Phase 2 events across the three experimental groups, which confirms the random assignment of rewards to users.

DR Events for Targeted (Low, High) and Non-Targeted Users, Phase 2			
Incentive Level	High-Targeted	Low-Targeted	Non-Targeted
0.05 USD/kWh	–	8745 (33.0%)	10421 (19.7%)
0.25 USD/kWh	–	8608 (32.5%)	10516 (19.8%)
0.50 USD/kWh	–	9109 (34.5%)	10675 (20.1%)
1.00 USD/kWh	12874 (49.9%)	–	10871 (20.5%)
3.00 USD/kWh	12927 (50.1%)	–	10508 (19.9%)
Total	25801 (100%)	26462 (100%)	52991 (100%)
	52263 (49.7%)		52991 (50.3%)

Table 8.3: Number of Phase 2 DR Events for Low Targeted, High Targeted, and Non Targeted Users by Incentive Level

Figure 8.10 shows the distribution of DR events in Phase 1 across all users by hour of the day. As can be seen from the figure, most events occurred in the late afternoon and early evening.

In a similar fashion, Figure 8.11 shows the distribution of Phase 1 messages by day of the week.

8.3.3 Weather Data

Hourly measurements of ambient air temperature are scraped from the publicly accessible California Irrigation Management Information System "California Irrigation Management Information System" 2017. As there are fewer weather stations than distinct user ZIP codes,

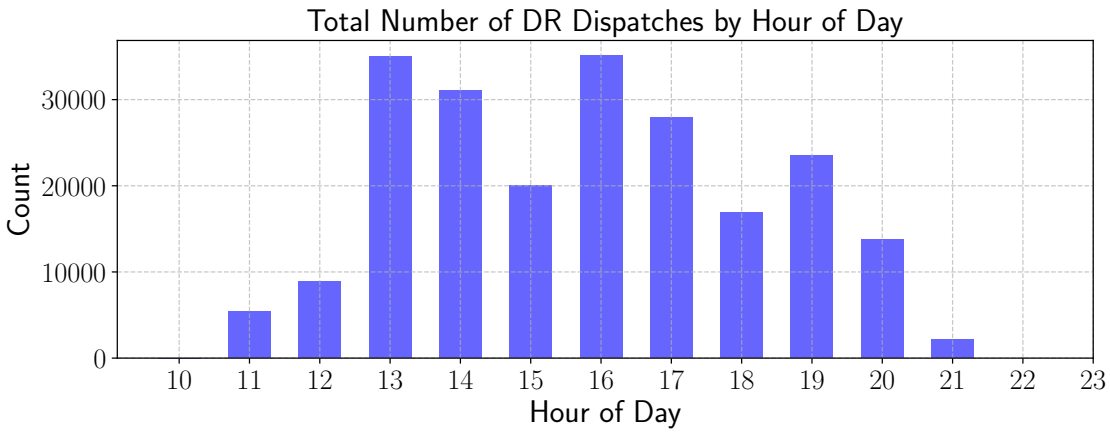

Figure 8.10: Distribution of Phase 1 DR Events by Hour of the Day

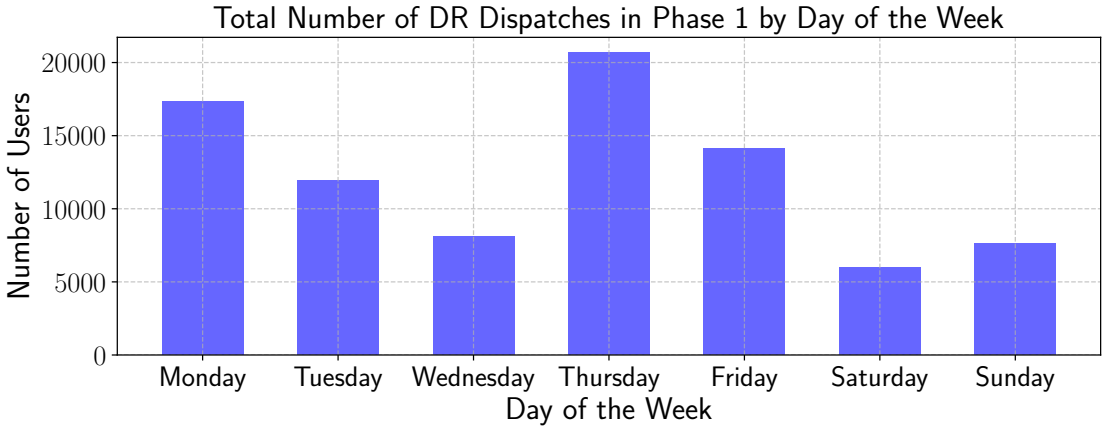

Figure 8.11: Distribution of Phase 1 DR Events by Day of the Week

we linearly interpolate user-specific temperatures at their ZIP codes from the two closest weather stations in latitude and longitude by calculating geodesic distances with Vincenty's formulae (Vincenty 1975).

8.3.4 Balance Checks

To verify that users were randomly assigned to control and treatment groups, we perform a balance check on the distribution of observed air temperatures and electricity consumptions across both groups. Notice that the relatively large sample size renders a classical differences-in-means t-test inappropriate. Therefore, we utilize Cohen's d to estimate the effect size based on the differences between means, which is insensitive to the large sample size. Given two discrete distributions P and Q with sample sizes n_1/n_2 and means $\bar{x}_1/\bar{x}_2$, Cohen's d is

defined as

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}, \quad s = \left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right)^{1/2}, \tag{8.1}$$

where s_1 and s_2 are the sample standard deviations for distributions P and Q, respectively. In addition, we use the Hellinger Distance H as a nonparametric comparison to quantify the similarity between the distributions *Hellinger Distance*:

$$H(P, Q) = \frac{1}{\sqrt{2}} \left(\sum_{i=1}^{k} (\sqrt{p_i} - \sqrt{q_i})^2 \right)^{1/2} \tag{8.2}$$

where $P = \{p_1, \ldots, p_k\}$ and $Q = \{q_1, \ldots, q_k\}$. To compute (8.1) and (8.2), we discretize the temperature and consumption distributions appropriately. Table B.1 in the Appendix provides these metrics together with the differences in means for a selected subset of hours of the day, which was chosen to coincide with those hours of the day for which DR events were observed (see Figure B.2). We omit the metrics for the remaining hours of the day as they are very similar to the listed ones. As the Hellinger Distance $H \in [0, 1]$, with 0 corresponding to a perfect similarity and 1 to total dissimilarity, we can assume that the assignment of users into treatment and control group is as good as random.

Figures 8.12 and 8.13 plot the mean consumption and temperature for targeted and non-targeted users, indicating that both groups are balanced in terms of these two metrics.

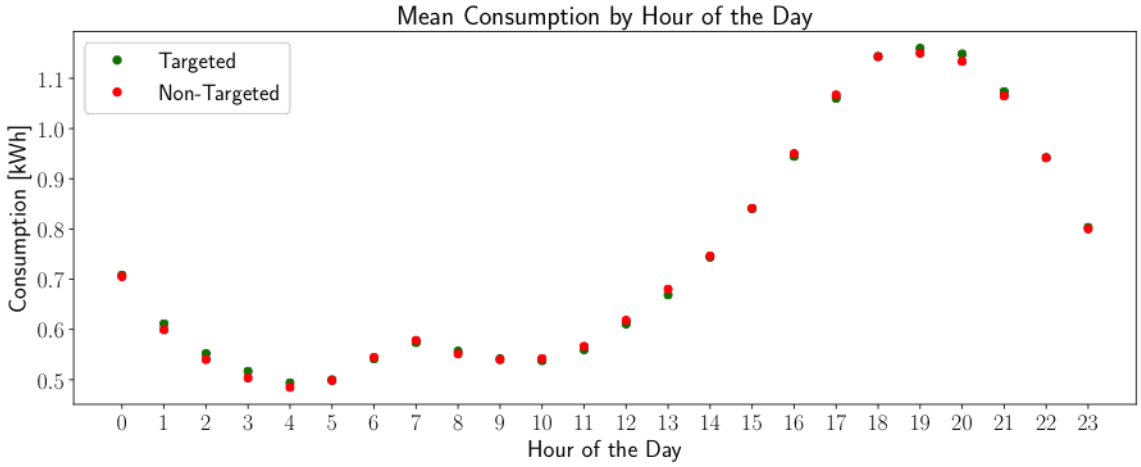

Figure 8.12: Mean Consumption for Targeted vs. Non-Targeted Users by Hour of the Day

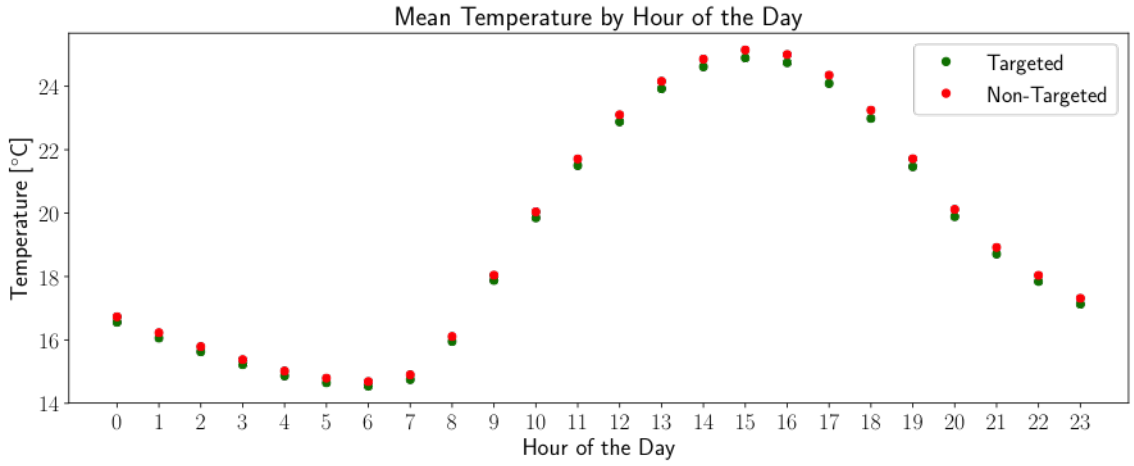

Figure 8.13: Mean Temperature for Targeted vs. Non-Targeted Users by Hour of the Day

8.4 Nonexperimental Treatment Effect Estimation

8.4.1 Potential Outcomes Framework

To estimate the effect of the DR intervention program, we adopt the *potential outcomes* framework introduced by Rubin (1974) (Rubin 1974). Let $\mathcal{I} = \{1, \ldots, n\}$ denote the set of users. The indicator $D_{it} \in \{0, 1\}$ encodes the fact whether or not user i received DR treatment at time t. Each user is endowed with a consumption time series $\mathbf{y}_i = \{y_{i1}, \ldots, y_{i\tau}\}$ and associated covariates $X_i = \{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i\tau}\} \in \times_{i=1}^{\tau} \mathcal{X}_i$, $\mathcal{X}_i \subset \mathbb{R}^{n_x}$, where time is indexed by $t \in \mathbb{T} = \{1, \ldots, \tau\}$ and n_x is the dimension of the covariate space $\mathcal{X}_i$. Let y_{it}^0 and y_{it}^1 denote user i's electricity consumption at time t for $D_{it} = 0$ and $D_{it} = 1$, respectively. Let $\mathcal{C}_i$ and $\mathcal{T}_i$ denote the set of control and treatment times for user i. That is,

$$\mathcal{C}_i = \{t \in \mathbb{T} \mid D_{it} = 0\}, \quad \mathcal{T}_i = \{t \in \mathbb{T} \mid D_{it} = 1\}. \tag{8.3}$$

The number of treatment hours is much smaller than the number of non-treatment hours. Thus $0 < |\mathcal{T}_i|/|\mathcal{C}_i| \ll 1$.

Further, let $\mathcal{D}_{i,t}$ and $\mathcal{D}_{i,c}$ denote user i's covariate-outcome pairs of treatment and control times, respectively. That is,

$$\mathcal{D}_{i,t} = \{(\mathbf{x}_{it}, y_{it}) \mid t \in \mathcal{T}_i\}, \ \mathcal{D}_{i,c} = \{(\mathbf{x}_{it}, y_{it}) \mid t \in \mathcal{C}_i\}. \tag{8.4}$$

The one-sample estimate of the treatment effect on user i at time t, given the covariates $\mathbf{x}_{it} \in \mathbb{R}^{n_x}$, is

$$\beta_{it}(\mathbf{x}_{it}) := y_{it}^1(\mathbf{x}_{it}) - y_{it}^0(\mathbf{x}_{it}) \quad \forall\ i \in \mathcal{I},\ t \in \mathbb{T}, \tag{8.5}$$

which varies across time, the covariate space, and the user population. Marginalizing this one-sample estimate over the set of treatment times $\mathcal{T}_i$ and the covariate space $\mathcal{X}_i$ yields the

user-specific Individual Treatment Effect (ITE) β_i

$$\beta_i := \mathbb{E}_{\mathcal{X}_i}\mathbb{E}_{t\in\mathcal{T}_i}\left[y_{it}^1 - y_{it}^0 \;\middle|\; \mathbf{x}_{it}\right] = \frac{1}{|\mathcal{T}_i|}\sum_{t\in\mathcal{T}_i} y_{it}^1 - y_{it}^0. \tag{8.6}$$

The average treatment effect on the treated (ATT) follows from (8.6):

$$\text{ATT} := \mathbb{E}_{i\in\mathcal{I}}[\beta_i] = \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{1}{|\mathcal{T}_i|}\sum_{t\in\mathcal{T}_i}(y_{it}^1 - y_{it}^0). \tag{8.7}$$

Since users were put into different experimental groups in a *randomized* fashion, the ATT and the average treatment effect (ATE) are identical (Pischke and Angrist 2009). Lastly, the conditional average treatment effect (CATE) on $\tilde{\mathbf{x}}$ is obtained by marginalizing the conditional distribution of one-sample estimates (8.5) on $\tilde{\mathbf{x}}$ over all users and treatment times, where $\tilde{\mathbf{x}} \in \mathbb{R}^{\tilde{n}_x}$ is a subvector of $\mathbf{x} \in \mathbb{R}^{n_x},\ 0 < \tilde{n}_x < n_x$:

$$\text{CATE}(\tilde{\mathbf{x}}) := \mathbb{E}_{i\in\mathcal{I}}\mathbb{E}_{t\in\mathcal{T}_i}\left[(y_{it}^1 - y_{it}^0) \;\middle|\; \tilde{\mathbf{x}}_{it} = \tilde{\mathbf{x}}\right]. \tag{8.8}$$

The CATE captures heterogeneity among users, e.g. with respect to specific hours of the day, the geographic distribution of users, the extent to which a user possesses "smart home" appliances, group or peer effects, etc. with the important requirement that the set of all treatment times $\mathcal{T}_i$ is drawn from a fixed, prior distribution G over the hours of the day. This differentiation is crucial since the one-sample estimate of the ITE, namely $\hat{y}_{it}^0 - y_{it}^1$ for any $t \in \mathcal{T}_i$, depends on its consumption level and willingness to reduce, both of which can vary considerably for different hours of the day. In our case, $\mathcal{T}_i$ is drawn from the empirical distribution of DR events as depicted in Figure 8.10. To rule out the existence of unobserved factors that could influence the assignment mechanism generating the complete observed data set $\{(\mathbf{x}_{it}, y_{it}, D_{it}) \mid i \in \mathcal{I}, t \in \mathbb{T}\}$, we make the following standard assumptions:

Assumption 15 (Unconfoundedness of Treatment Assignment)**.** *Given the covariates* $\{\mathbf{x}_{it}\}_{t\in\mathbb{T}}$*, the potential outcomes are independent of treatment assignment:*

$$(y_{it}^0, y_{it}^1) \perp D_{it} \mid \mathbf{x}_{it} \quad \forall i \in \mathcal{I}, t \in \mathbb{T}. \tag{8.9}$$

Assumption 16 (Stationarity of Potential Outcomes)**.** *Given the covariates* $\{\mathbf{x}_{it}\}_{t\in\mathbb{T}}$*, the potential outcomes are independent of time, that is,*

$$(y_{it}^0, y_{it}^1) \perp t \mid \mathbf{x}_{it} \quad \forall i \in \mathcal{I}, t \in \mathbb{T}. \tag{8.10}$$

Assumption 15 is the "ignorable treatment assignment" assumption introduced by Rosenbaum and Rubin (Rosenbaum and Rubin 1983). Under this assumption, the assignment of DR treatment to users is implemented in a *randomized* fashion, which allows the calculation of unbiased ATEs (8.7) and CATEs (8.8). Assumption 16, motivated by the time-series nature of the observational data, ensures that the set of observable covariates $\{\mathbf{x}_{it} \mid t \in \mathbb{T}\}$

can capture seasonality effects in the estimation of the potential outcomes. That is, the conditional distribution of the potential outcomes, given covariates, remains constant.

Normalizing (8.6) with the estimated counterfactuals yields the *elasticity* of user i, which, with a slight abuse of notation, we also denote with β_i:

$$\beta_i := \mathbb{E}\left[1 - y_{it}^0/y_{it}^1\right] = \frac{1}{|\mathcal{T}_i|}\sum_{j\in\mathcal{T}_i}\left(1 - y_{ij}^0/y_{ij}^1\right) \tag{8.11}$$

where the same distinction between a conditional and general elasticity as ITE / CITE can be made.

The *fundamental problem of causal inference* (Holland 1986) refers to the fact that either the treatment or the control outcome can be observed, but never both (granted there are no missing observations). That is,

$$y_{it} = y_{it}^0 + D_{it}\cdot(y_{it}^1 - y_{it}^0) \quad \forall\, t \in \mathbb{T}. \tag{8.12}$$

Thus, the ITE (8.6) is not identified, because one and only one of both potential outcomes is observed, namely $\{y_{it}^1 \mid t \in \mathcal{T}_i\}$ for the treatment times and $\{y_{it}^0 \mid t \in \mathcal{C}_i\}$ for the control times. It therefore becomes necessary to estimate counterfactuals.

8.4.2 Non-Experimental Estimation of Counterfactuals

Consider the following model for the estimation of such counterfactuals:

$$y_{it} = f_i(\mathbf{x}_{it}) + D_{it}\cdot\beta_{it}(\mathbf{x}_{it}) + \varepsilon_{it}, \tag{8.13}$$

where ε_{it} denotes noise uncorrelated with covariates and treatment assignment. $f_i(\cdot) : \mathbb{R}^{n_x} \mapsto \mathbb{R}$ is the conditional mean function and pertains to $D_{it} = 0$. To obtain an estimate for $f_i(\cdot)$, denoted with $\hat{f}_i(\cdot)$, control outcomes $\{y_{it}^0 \mid t \in \mathcal{C}_i\}$ are first regressed on $\{\mathbf{x}_{it} \mid t \in \mathcal{C}_i\}$, namely their observable covariates. In a second step, the counterfactual $\hat{y}_{it}^0$ for any $t \in \mathcal{T}_i$ can be estimated by evaluating $\hat{f}_i(\cdot)$ on its associated covariate vector $\mathbf{x}_{it}$. Finally, subtracting $\hat{y}_{it}^0$ from y_{it}^1 isolates the one-sample estimate $\beta_{it}(\mathbf{x}_{it})$, from which the user-specific ITE (8.6) can be estimated. Figure 8.14 illustrates this process of estimating the reduction during a DR event by subtracting the actual consumption y_{it}^1 from the predicted counterfactual $\hat{y}_{it}^0 = \hat{f}_i(\mathbf{x}_{it})$. Despite the fact that consumption can be predicted for horizons longer than a single hour, we restrict our estimators $f_i(\cdot)$ to a single hour prediction horizon as DR events are at most one hour long.

To estimate $f_i(\cdot)$, we use the following classical regression methods (Hastie, Tibshirani, and Friedman 2009), referred to as *estimators*:

E1: Ordinary Least Squares Regression (OLS)

E2: L1 Regularized (LASSO) Linear Regression (L1)

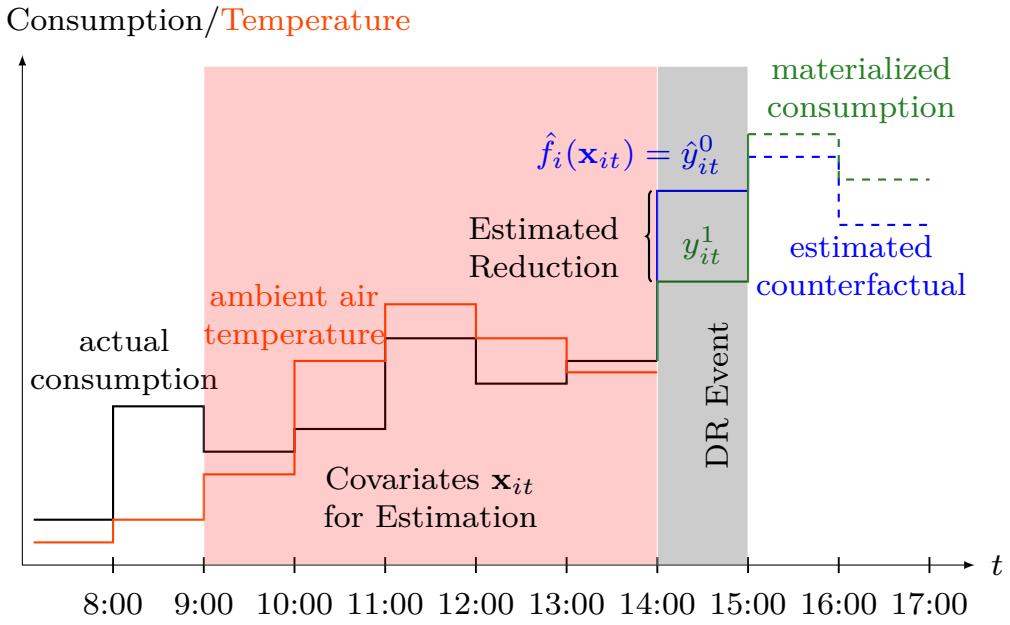

Figure 8.14: Estimation of the Counterfactual $\hat{y}^0_{it}$ using Treatment Covariates $\mathbf{x}_{it}$ and Predicted Reduction $\hat{y}^0_{it} - y^1_{it}$

E3: L2 Regularized (Ridge) Linear Regression (L2)

E4: k-Nearest Neighbors Regression (KNN)

E5: Decision Tree Regression (DT)

E6: Random Forest Regression (RF)

DT (E5) and RF (E6) follow the procedure of Classification and Regression Trees (Breiman et al. 1984). We compare estimators (E1)-(E6) to the CAISO 10-in-10 Baseline (BL) (*California Independent System Operator Corporation (CAISO): Fifth Replacement FERC Electric Tariff* 2014) (the regulatory standard for DR settlements in California), which, for any given hour on a weekday, is calculated as the mean of the hourly consumptions on the 10 most recent business days during the selected hour. For weekend days and holidays, the mean of the 4 most recent observations is calculated. This BL is further adjusted with a *Load Point Adjustment*, which corrects the BL by a factor proportional to the consumption three hours prior to a DR event (*California Independent System Operator Corporation (CAISO): Fifth Replacement FERC Electric Tariff* 2014) excluding the hour immediately prior to the event.

Since users tend to exhibit a temporary increase in consumption in the hours following the DR intervention (Palensky and Dietrich 2011), we remove $n_r = 8$ hourly observations following each DR event in order to prevent estimators (E1)-(E6) from learning from such spillover effects. This process is illustrated in Figure 8.15.

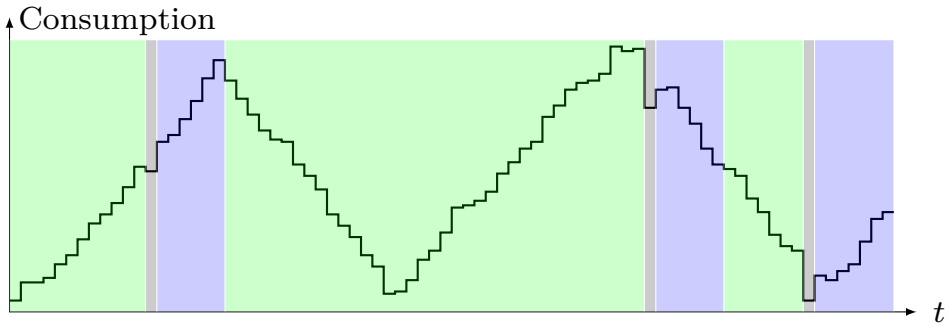

Figure 8.15: Separation of consumption time series into training set (green), DR Events (grey), and Periods of Spillover Effects (blue)

Hence the training data $\mathcal{D}_{i,tr} \subset \mathcal{D}_{i,t}$ used to estimate the conditional mean function $\hat{f}_i(\cdot)$ (8.13) consists of all observations leading up to a DR event, excluding those that are within 8 hours of any DR event. To estimate user i's counterfactual outcome $\hat{y}^0_{it}$ during a DR event $t \in \mathcal{T}_i$, we use the following covariates:

- 5 hourly consumption values preceding time t
- Air temperature at time t and 4 preceding measurements
- Hour of the day, an indicator variable for (non-)business days, and month of the year as categorical variables

Thus, the covariate vector writes

$$\begin{aligned}\mathbf{x}_{it} = [&y^0_{it-1} \quad \cdots \quad y^0_{it-5} \quad T_{it} \quad \cdots \quad T_{it-4} \\ &\mathrm{C}(\mathrm{HoD}_{it}) : \mathrm{C}(\mathrm{is\_Bday}_{it}) \quad \mathrm{C}(\mathrm{MoY}_{it})].\end{aligned} \tag{8.14}$$

In (8.14), T_{it} denotes temperature, HoD_{it} hour of day, $\mathrm{is\_Bday}_{it}$ an indicator variable for business days, and MoY_{it} the month of year (all for user i at time t). "C" denotes dummy variables and ":" their interaction.

8.4.3 Placebo Treatments and De-biasing of Estimators

As previously mentioned, a crucial element of an estimator is unbiasedness. If an estimator systematically predicts counterfactuals that are too large (small), users receive an excess reward (are paid less) proportional to the amount of prediction bias. For a fair economic settlement, it is thus desirable to minimize the amount of bias. In our application, such prediction bias is caused by the following two factors:

- Inherent bias of estimators: With the exception of OLS (E1), (E2)-(E6) are inherently biased, which is justified due to the well-known bias-variance tradeoff.

- Seasonal and temporal bias: Due to the experimental design, DR events for a particular user are concentrated within a period of 180 days after signing up. Further, DR events are called only in the afternoon and early evening (see Figure B.2). Thus, fitting an estimator on all available historical data is likely to introduce bias during these time periods of interest, namely in the 90 days after signing up and in particular during DR events.

To deal with these challenges, we use the de-biasing procedure presented in Algorithm 6, which was first introduced in Balandat 2016. We first separate a subset of non-DR

Algorithm 6 Unbiased Estimation of Counterfactuals

Input: Treatment data $\mathcal{D}_{i,t}$, control data $\mathcal{D}_{i,c}$, Estimator, η **Output:** Counterfactuals $\{\hat{y}_{it} \mid t \in \mathcal{T}_i\}$

1: Split $\mathcal{D}_{i,c}$ into training data $\mathcal{D}_{i,tr}$ and placebo data $\mathcal{D}_{i,pl}$ according to empirical distribution of $\mathcal{T}_i$. Split control times $\mathcal{C}_i$ into training times $\mathcal{C}_{i,tr}$ and placebo times $\mathcal{C}_{i,pl}$
2: Compute weights for $\mathcal{D}_{i,tr} = \{(x_{it}, y_{it}) \mid t \in \mathcal{C}_{i,tr}\}$ according to (8.16a)-(8.16c), using η
3: Fit conditional mean function $\hat{f}_i$ on $\mathcal{D}_{i,tr}$ with weights
4: Estimate placebo counterfactuals $\{\hat{y}^0_{it} \mid t \in \mathcal{C}_{i,pl}\}$
5: Compute bias on placebo treatment set
6: Estimate treatment counterfactuals $\{\hat{y}^0_{it} \mid t \in \mathcal{T}_i\}$
7: Subtract placebo treatment bias from estimated treatment counterfactuals:

$$\hat{y}^0_{it} \leftarrow \hat{y}^0_{it} - \frac{1}{|\mathcal{C}_{i,pl}|} \sum_{\tau \in \mathcal{C}_{i,pl}} (\hat{y}^0_{i\tau} - y^0_{i\tau}) \quad \forall\, t \in \mathcal{T}_i \tag{8.15}$$

events from user i's control data $\mathcal{D}_{i,c}$, which we call the placebo set $\mathcal{D}_{i,pl}$ with associated placebo treatment times $\mathcal{C}_{i,pl}$ (we chose $\mathcal{D}_{i,pl}$ to be of size 25). This placebo set is drawn according to user i's empirical distribution of Phase 1 DR events by hour of day and month of year. Next, the non-experimental estimator of choice is fitted (using cross-validation to find hyperparameters to minimize the mean squared prediction error) on the training set $\mathcal{D}_{i,tr}$. Importantly, to account for temporal bias, we assign weights to the training samples, ensuring that samples in "similar" hours or seasons as actual DR events are assigned larger weights. Specifically, the weights w_{it} are determined as follows:

$$w_{it} \propto w^{\text{HoD}}_{it} w^{\text{MoY}}_{it}, \tag{8.16a}$$

$$w^{\text{HoD}}_{it} = \eta + \sum_{\tau \in \mathcal{C}_{i,tr}} \mathbb{1}(\text{HoD}_{it} = \text{HoD}_{i\tau}), \tag{8.16b}$$

$$w^{\text{MoY}}_{it} = \eta + \sum_{\tau \in \mathcal{C}_{i,tr}} \mathbb{1}(\text{MoY}_{it} = \text{MoY}_{i\tau}), \tag{8.16c}$$

where $\eta > 0$ is a constant to be chosen a-priori.

Then, the fitted model is used to predict counterfactuals associated with placebo events. This yields a set of $|\mathcal{C}_{i,pl}|$ paired samples from which we can obtain a proxy of the estimation

bias that remains even after assigning sample weights according to the previous step. Finally, to obtain an empirically de-biased estimate of actual Phase 1 DR events, we simply subtract this proxy of the estimation bias from predicted Phase 1 DR event outcomes.

The attentive reader might wonder why the computation of weights is part of the algorithm, given that the subtraction of placebo treatment bias from the estimated treatment counterfactuals is already performed. This question is answered by the desire to reduce both the estimation bias and the variance pertinent to the set $\{\hat{y}^0_{it} \mid t \in \mathcal{C}_{i,pl}\}$. This is illustrated in Figure 8.16, which plots both the mean prediction bias on the placebo sets across users and their sample standard deviation for a selected subset of non-experimental estimators.

It can be seen that for finite weight parameters, the prediction bias is closer to zero than for an infinite value of η (which corresponds to no sample weighting, see (8.16b), (8.16c)). Further, the sample standard deviation of user-level biases decreases as η decreases. It appears that $\eta = 0.2$ is a good choice that minimizes the variance of estimated biases, which is the parameter of η we use throughout this chapter.

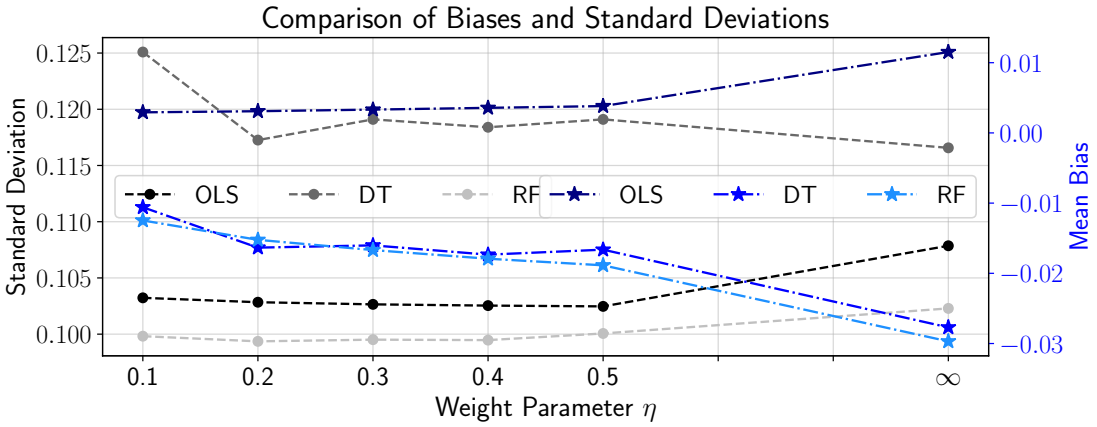

Figure 8.16: Mean Prediction Bias on $\mathcal{D}_{i,p}$ and Sample Standard Deviation Across Users for Selected Estimators

For the remaining estimators, Figure 8.17 plots the distribution of user-level biases on the placebo test without sample weighting. This figure confirms the need for dealing with seasonal and temporal bias.

8.4.4 Estimation of Individual Treatment Effects

To obtain point estimates for user i's ITE β_i, we can either average all one-sample estimates (8.5) according to (8.6), or use the Hodges-Lehmann Estimator (8.17), which appears to be advantageous in our application as it is robust against estimation inaccuracies. In both cases, we would like to compute confidence intervals (CIs) around the ITEs, which we can do with either the Wilcoxon Signed Rank Test (Section 8.4.4.1) or with a simple permutation test (Section 8.4.4.2).

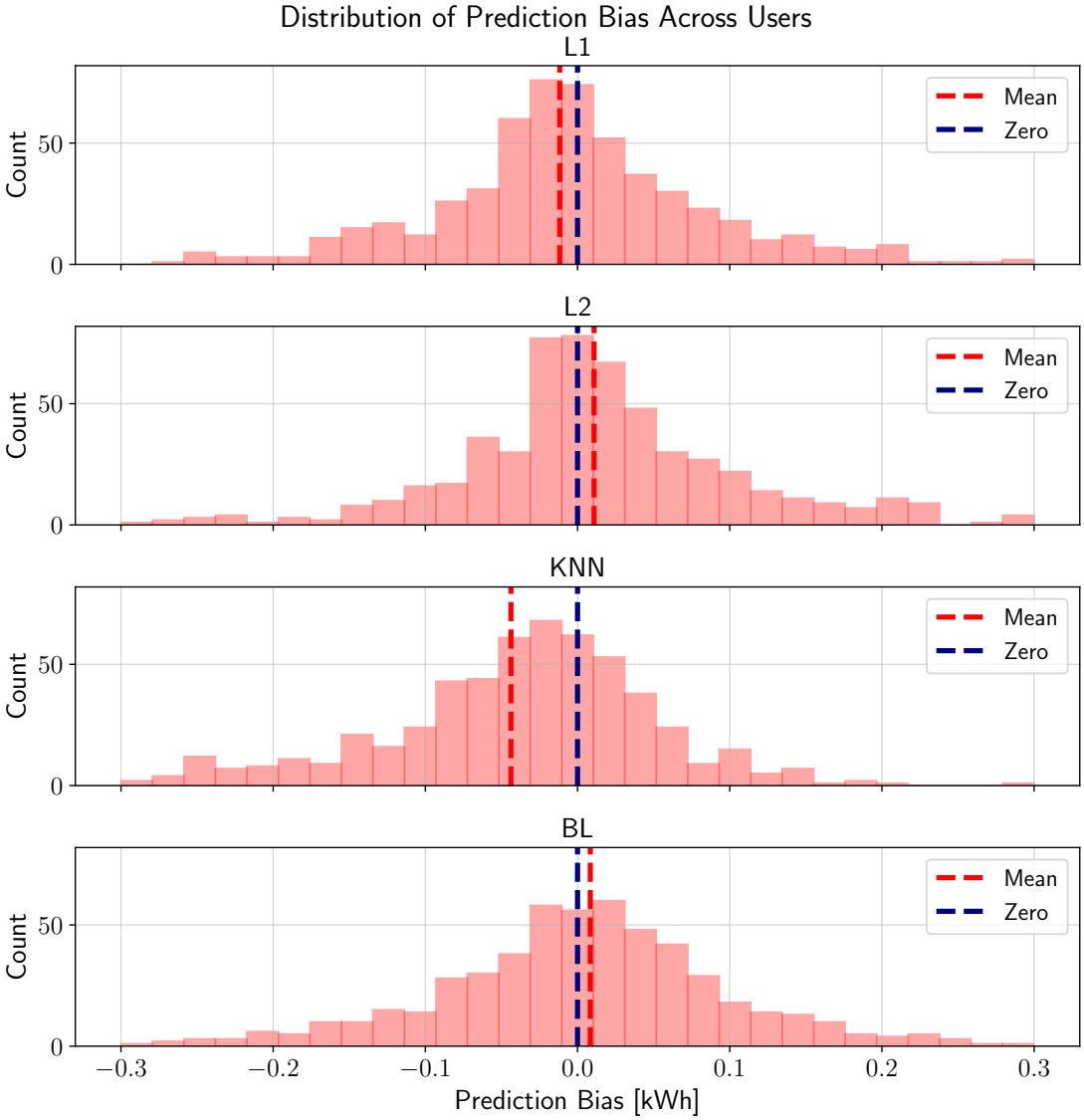

Figure 8.17: Distribution of User-Level Prediction Biases on Placebo Events

8.4.4.1 Nonparametric Signed Rank Test

A nonparametric comparison of the treatment observations $\{y_{it}^1 \mid t \in \mathcal{T}_i\}$ and their estimated counterfactuals $\{\hat{y}_{it}^0 \mid t \in \mathcal{T}_i\}$ admits p-values and coverage probabilities for confidence intervals without requiring assumptions on the underlying data generating process, while being robust against outliers. The paired replicate nature of these sets as well as their relatively small size (which precludes the use of the Central Limit Theorem) calls for an analysis using signed ranks (Hollander, Wolfe, and Chicken 2013). We estimate ITEs with the Hodges-Lehmann Estimator (HLM) that is associated with the Wilcoxon Signed Rank Statistic. Let $z_{it} := \hat{y}_{it}^0 - y_{it}^1 \ \forall \ t \in \mathcal{T}_i$ denote the pairwise sample difference at treatment time t. The following assumption has to be made (Hollander, Wolfe, and Chicken 2013):

Assumption 17. *The elements of* $\{z_{it} \mid t \in \mathcal{T}_i\}$ *are mutually independent.*

Assumption 18. *The elements of* $\{z_{it} \mid t \in \mathcal{T}_i\}$ *are mutually independent. Each* z_{it} *comes from an absolutely continuous distribution* F_i *symmetric about median treatment effect* θ_i*.*

From Assumption 18, it follows immediately that $\theta_i \equiv \beta_i$. To estimate $\hat{\theta}_i$, first the set of pairwise averages

$$\mathcal{Z}_i := \{(z_{i\mathcal{T}[t]} + z_{i\mathcal{T}[u]})/2 \mid 1 \leq t \leq u \leq |\mathcal{T}_i|\} \tag{8.17}$$

of cardinality $|\mathcal{T}_i|(|\mathcal{T}_i|+1)/2$ is constructed. Then $\hat{\theta}_i$ is calculated as the median of $\mathcal{Z}_i$. The Hodges-Lehmann Estimator is intrinsically related to the Wilcoxon Signed Rank Test with null hypothesis and its corresponding alternative

$$H_0 : \theta_i = 0, \tag{8.18a}$$
$$H_1 : \theta_i \neq 0, \tag{8.18b}$$

which constitutes a two-sided test at significance level α. The test statistic W_i is calculated by sorting $\{|z_{it}| \mid t \in \mathcal{T}_i\}$ in ascending order, assigning ordinal ranks $\{R_{it} \mid t \in \mathcal{T}_i\}$ to the ordered items, and finally adding up the ranks as follows:

$$W_i = \sum_{t=1}^{|\mathcal{T}_i|} R_{i\mathcal{T}_i[t]} \cdot \mathbb{1}(z_{i\mathcal{T}_i[t]} > 0), \tag{8.19}$$

where $\mathcal{T}_i[t]$ denotes the element at the t-th position of $\mathcal{T}_i[t]$. It can be shown that $W_i \sim G_{W_i}$ is distributed symmetrically about the mean $|\mathcal{T}_i|(|\mathcal{T}_i|+1)/4$ with discrete support on $[0, |\mathcal{T}_i|(|\mathcal{T}_i|+1)/2]$. The more $\{z_{it} \mid t \in \mathcal{T}_i\}$ alternate in signs, the closer W_i to the mean. The estimated treatment effect $\hat{\theta}_i$ is identified to be the scalar that has to be subtracted from each element of $\mathcal{Z}_i$ such that the new set

$$\tilde{\mathcal{Z}}_i := \{(z_{i\mathcal{T}[t]} + z_{i\mathcal{T}[u]})/2 - \hat{\theta}_i \mid 1 \leq t \leq u \leq |\mathcal{T}_i|\} \tag{8.20}$$

puts statistic W_i closest to its mean. Equivalently, the distance $||\mathcal{T}_i|(|\mathcal{T}_i|+1)/4 - W_i|$ is minimized for $\{(z_{it} - \hat{\theta}_i) \mid t \in \mathcal{T}_i\}$, which maximizes the resulting p-value associated with H_0 (8.18a).

Remark 3. *Because* $\{y_{it}^1 \mid t \in \mathcal{T}_i\}$ *and* $\{\hat{y}_{it}^0 \mid t \in \mathcal{T}_i\}$ *are generated by an absolutely continuous distribution (Assumption 18), the pairwise differences* $\{z_{it} \mid t \in \mathcal{T}_i\}$ *consist of unique values with probability 1. Therefore no ties need to be broken.*

The test statistic W_i admits a p-value and $(1-\alpha)$ confidence interval $\left[\underline{\theta}_i, \overline{\theta}_i\right]$ (Lehmann and D'Abrera 2006) as follows:

$$p = \left|\Phi_i(W_i) + 1 - \Phi_i\left(\frac{\mathcal{T}_i(|\mathcal{T}_i|+1)}{2} - W_i\right)\right|, \tag{8.21a}$$

$$\left[\underline{\theta}_i, \overline{\theta}_i\right] = \left[\mathcal{Z}[\Phi_i^{-1}(\alpha/2)],\ \mathcal{Z}[\Phi_i^{-1}(1-\alpha/2)]\right], \tag{8.21b}$$

where $\Phi_i(\cdot)$ and $\Phi_i^{-1}(\cdot)$ are the CDF and percent point function of distribution G_{W_i}, respectively. The coverage probability of θ_i is therefore given by

$$\mathbb{P}(\underline{\theta}_i \leq \theta_i \leq \overline{\theta}_i) = 1 - \alpha. \tag{8.22}$$

The confidence interval corresponds to the range of $\tilde{\theta}_i$ for which the Wilcoxon Signed Rank Test does not reject the modified null hypothesis $H_0 : \theta = \tilde{\theta}_i$ with corresponding alternative $H_1 : \theta \neq \tilde{\theta}_i$ at significance level α. Thus, (8.18a) is rejected at confidence level $1-\alpha$ if and only if either

- $\underline{\theta}_i \leq \overline{\theta}_i < 0$: (8.18a) is rejected at the lower tail (that is, user i *reduces* consumption significantly), or
- $0 < \underline{\theta}_i \leq \overline{\theta}_i$: (8.18a) is rejected at the upper tail (that is, user i *increases* consumption significantly).

8.4.4.2 Nonparametric Permutation Test

To obtain an estimate of whether or not a given user i has actually reduced consumption, we utilize a nonparametric permutation test with the null hypothesis of a zero ITE:

$$H_0 : \beta_i = 0, \qquad H_1 : \beta_i \neq 0. \tag{8.23}$$

Given user i's paired samples $\{z_{it} = \hat{y}_{it}^0 - y_{it}^1 \mid t \in \mathcal{T}_i\}$ during DR periods, the p-value associated with H_0 (8.23) is

$$p = \frac{\sum_{D \in \mathcal{P}_i} \mathbb{1}(\bar{D} \leq \hat{\beta}_i)}{2^{|\mathcal{T}_i|}}. \tag{8.24}$$

In (8.24), $\bar{D}$ denotes the mean of D. $\mathcal{P}_i$ denotes the set of all possible assignments of signs to the pairwise differences in the set $\{z_{it} = y_{it}^1 - \hat{y}_{it}^0 \mid t \in \mathcal{T}_i\}$. That is,

$$\mathcal{P}_i = \{s_1 z_{i1}, \ldots, s_{|\mathcal{T}_i|} z_{i|\mathcal{T}_i|} \mid s_j \in \{\pm 1\}, 1 \leq j \leq |\mathcal{T}_i|\} \tag{8.25}$$

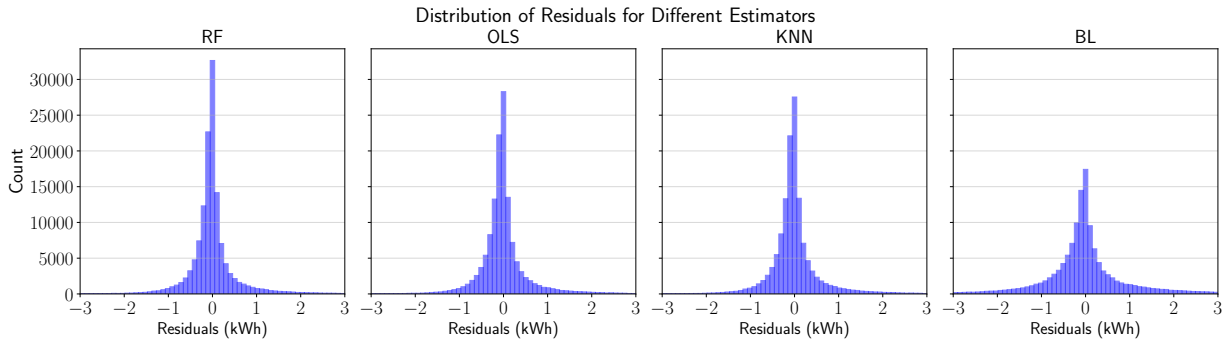

Figure 8.18: Distribution of Event-wise Residuals for Selected Estimators

which is of size $2^{|\mathcal{T}_i|}$. Finally, the p-value from (8.23) is calculated as the fraction of all possible assignments whose means are less than or equal the estimated ITE $\hat{\beta}_i$. In practice, as the number of DR events per user in Phase 1 is about 25 (see Figure 8.2), the number of total possible assignments becomes computationally infeasible. Thus, we randomly generate a subset of 10^5 assignments from $\mathcal{P}_i$ to compute the p-value in (8.24).

Moreover, we use the percentile bootstrap method (Efron and Tibshirani 1994) to compute a confidence interval of the estimated ITE for user i around the point estimate $\hat{\beta}_i$.

8.5 Nonexperimental Estimation Results

8.5.1 Validation of Individual Treatment Effects

In this section we analyze the prediction accuracy of individual treatment effect estimators by analyzing raw residuals. Figure 8.18 plots a histogram of the distribution of such residuals for four selected estimators. Table 8.4 provides the standard deviation of this distribution for all estimators, and we observe that RF is the "best" estimator in terms of tightness around zero. As expected, the CAISO 10-in-10 baseline performs worst.

Moreover, we carry out a ranking exercise of non-targeted users across Phases 1 and 2 of the experiment. The intuition is that non-targeted users do not experience a change in their average reward level across both phases (unlike targeted users) and hence are expected to have the same reduction behavior in both phases of the experiment. Under the assumption of no user attrition and adversarial behavior, a perfect estimator should estimate the same ITE in Phase 1 as well as in Phase 2 for a particular user. Equivalently, if we rank users according to their Phase 1 ITEs predicted by a perfect estimator, we would expect the same ordering as in Phase 2. Thus, if we scatter plot the ranks of user ITEs estimated in Phase 1 and Phase 2, the accuracy of a particular estimator correlates with how tightly the dots are put around the diagonal. Figure 8.19 shows such a plot with deciles instead of ranks, i.e. a point (x, y) represents a user whose rank is in the x-th (y-th) decile in Phase 1 (Phase 2) rolled over on September 21, 2017. For illustrative purposes, the dots are transparent

Distribution of Residuals for Estimators	
Estimator	Standard Deviation
RF	**0.5710**
OLS	0.5936
L2	0.5997
L1	0.6009
KNN	0.6050
DT	0.6866
BL	1.6741

Table 8.4: Standard Deviation of Event-wise Residuals ($y_{it} - \hat{y}_{it}$)

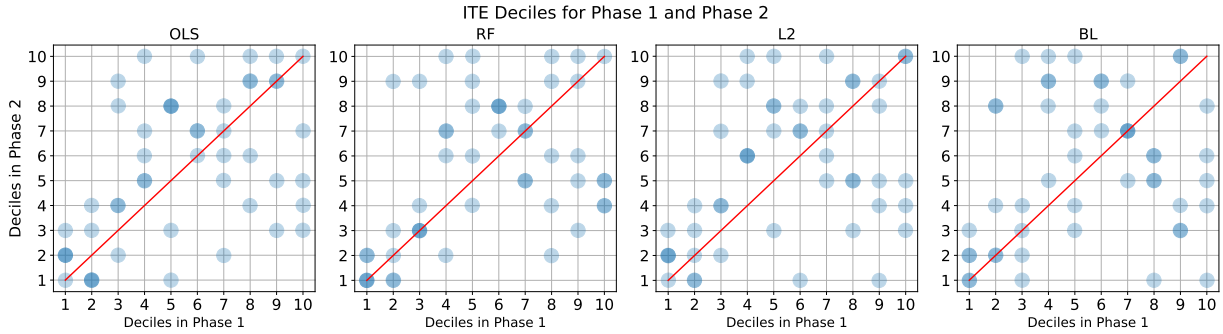

Figure 8.19: Deciles of Non-Targeted Users Ranked by ITE estimated in both Phase 1 and Phase 2 of the Experiment

such that darker dots indicate multiple counts. It can be seen that OLS and RF achieve the tightest fit around the diagonal, whereas BL performs worst.

Table 8.5 provides the R^2 scores for selected rollover dates, namely dates for which there exists at least one non-targeted user with both at least 20 events in each of Phase 1 and 2. Averaging all R^2 scores across all such rollover dates is evidence for RF being the most accurate estimator on an individual user level.

8.5.2 Validation of Average Treatment Effects

In a similar fashion as Section 8.5.1, we now aggregate the simulated user-level responses into a population-wide ATE. Figure 8.20 plots the means and 95% confidence intervals for varying estimators and group sizes, where the means and upper/lower confidence bounds themselves have distributions, as we repeat the simulations many times. Consistent with previous findings, larger group sizes result in tighter confidence intervals.

R^2 for Non-Targeted Users for Phase 1 and Phase 2 Ranks				
Rollover Date	RF	OLS	L2	BL
2017/07/27	**0.7820**	0.3856	0.7126	0.3554
2017/08/03	0.3415	0.1789	0.2195	**0.8117**
2017/08/10	0.7348	0.6464	**0.9042**	0.6390
2017/08/31	**0.3131**	0.0772	0.2888	0.1015
2017/09/07	0.4275	**0.4857**	0.3936	0.2213
2017/09/14	0.3234	0.2276	0.2931	**0.4665**
2017/09/21	**0.3477**	0.3007	0.3138	0.1937
2017/09/28	0.2780	0.2406	0.1937	**0.2889**
2017/10/03	0.0274	0.0138	−0.0674	**0.0446**
Mean	**0.3973**	0.2841	0.3613	0.3470

Table 8.5: Pearson Correlation Coefficient for Non-Targeted User Ranks Across Phase 1 and Phase 2, Minimum 20 Events per User

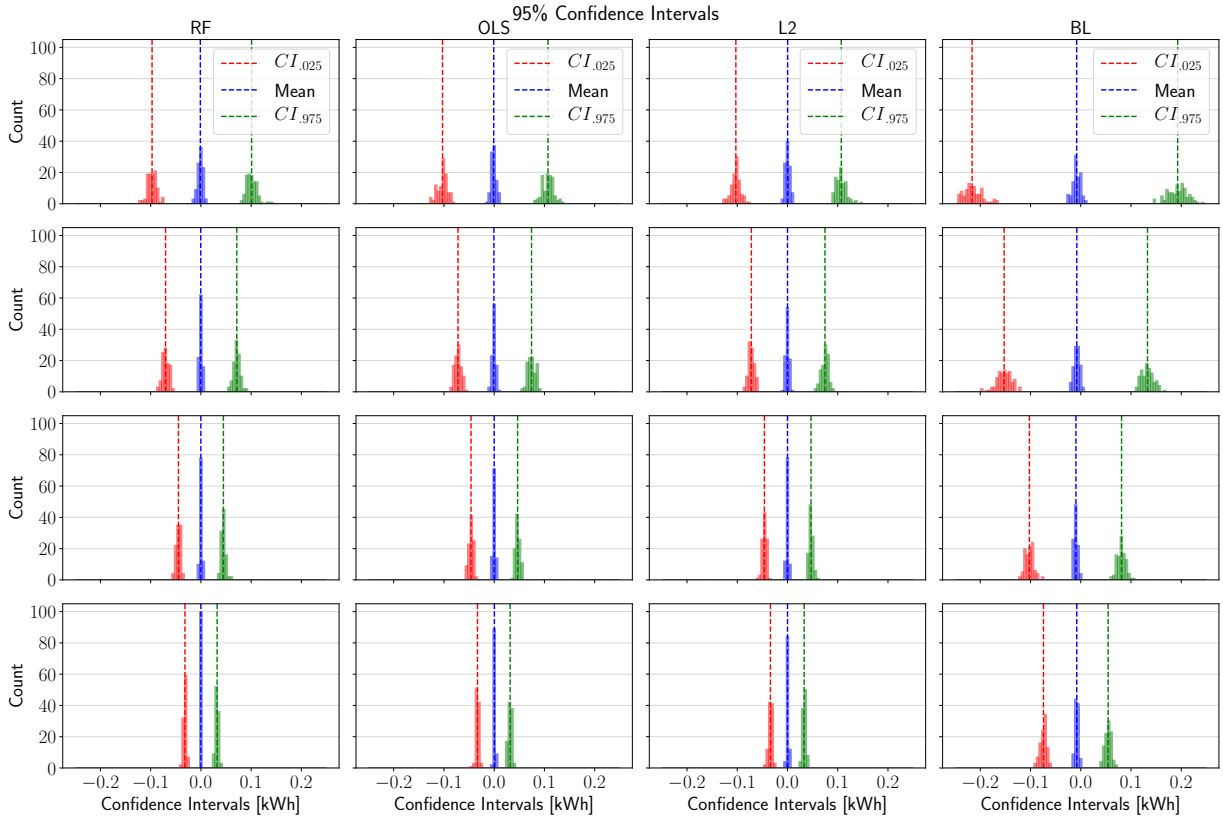

Figure 8.20: 95% Confidence Intervals for Estimators and Varying Group Sizes (5, 10, 25, 50 users)

8.5.3 Average Treatment Effects

Figure 8.21 shows ATE point estimates and their 99% bootstrapped confidence intervals conditional on differing reward levels for all estimators as well as the CAISO BL. Due to empirical de-biasing with Algorithm 6, the point estimates for estimators E1-E6 are close to each other. BL appears to be biased in favor of the DRP, as it systematically predicts smaller reductions than E1-E6.

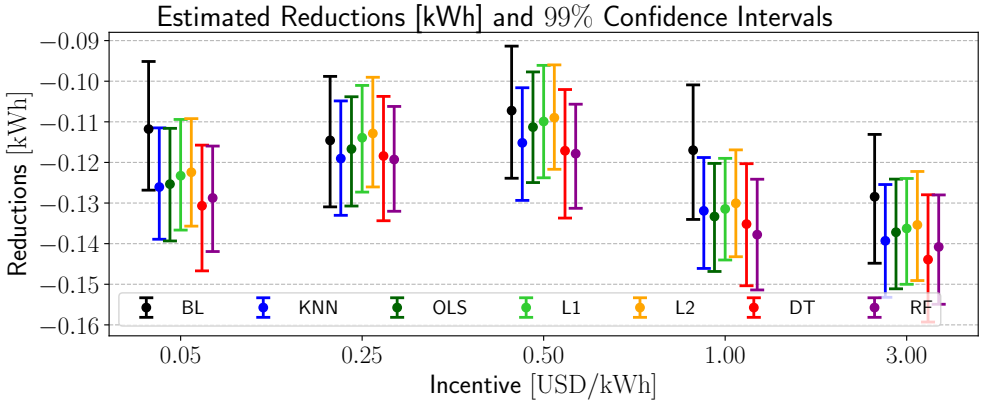

Figure 8.21: CATEs by Incentive Level with Bootstrapped Confidence Intervals

Table 8.6 reports the estimated ATE (in kWh and percent reduction) and the estimated intercept and slope of the demand curve, aggregated across all incentive levels. Due to the idiosyncratic nature of the CATE for $r = 0.5\frac{\text{USD}}{\text{kWh}}$, the slope and intercept have to be interpreted with caution. However, the results give rise to a notable correlation between larger incentives levels and larger reductions.

ATEs, Intercepts (kWh), and Slopes (kWh) for all Methods				
Estimator	ATE (kWh)	ATE (%)	Intercept	Slope · 3USD
BL	−0.116	−12.9	−0.110	−0.018
KNN	−0.126	−13.9	−0.120	−0.020
OLS	−0.125	−13.8	−0.119	−0.019
L1	−0.123	−13.6	−0.116	−0.021
L2	−0.122	−13.5	−0.115	−0.021
DT	−0.129	−14.2	−0.122	−0.022
RF	−0.129	−14.2	−0.123	−0.019

Table 8.6: ATE Estimates and Demand Curve by Estimator, all 4791 Users

To compare the prediction accuracy of the estimators, Table 8.10 reports the width of the confidence intervals for each method and incentive level. The inferiority of the CAISO

baseline compared to the non-experimental estimators, among which RF achieves the tightest confidence intervals, becomes apparent. Therefore, in the remainder of this chapter, we restrict all results achieved with non-experimental estimators to those obtained with RF.

Width of CATE Confidence Intervals (kWh) by Incentive Level					
	0.05 $\frac{\text{USD}}{\text{kWh}}$	0.25 $\frac{\text{USD}}{\text{kWh}}$	0.50 $\frac{\text{USD}}{\text{kWh}}$	1.00 $\frac{\text{USD}}{\text{kWh}}$	3.00 $\frac{\text{USD}}{\text{kWh}}$
BL	0.0317	0.0321	0.0325	0.0331	0.0317
KNN	0.0274	0.0282	0.0277	0.0273	0.0278
OLS	0.0277	0.0269	0.0272	0.0266	0.0270
L1	0.0272	0.0263	0.0277	**0.0250**	0.0261
L2	0.0265	0.0270	0.0257	0.0263	0.0269
DT	0.0310	0.0306	0.0317	0.0301	0.0313
RF	**0.0260**	**0.0258**	**0.0256**	0.0273	**0.0260**

Table 8.7: Width of 95 % Confidence Intervals around ATE Point Estimate Conditional on Incentive Level, All Estimators

8.5.4 Individual Treatment Effects

Figure 8.22 plots ITEs for a randomly selected subset of 800 users who received at least 10 DR events in Phase 1, estimated with RF. Users are sorted by their point estimates (blue), whose 95% bootstrapped confidence intervals are drawn in black. Yellow lines represent users with at least one active smart home automation device. By marginalizing the point estimates over all users with at least 10 events, we obtain an ATE of -0.139 kWh (-14.2%), which is close to -0.129 kWh as reported in Table 8.6. The difference ensues from only considering users with at least 10 DR events. The 99% ATE confidence interval is $[-0.154, -0.125]$ kWh.

Table 8.8 reports estimated ATEs for users with or without active smart home automation devices, which are obtained by aggregating the relevant estimated ITEs from Figure 8.22. We notice larger responses as well as a larger percentage of estimated reducers among automated users.

ATEs Conditional on Automation Status for Users with $\geq$ 10 DR Events				
	# Users	% Reducers	ATE (kWh)	ATE (%)
Automated	372	80.4	-0.332	-38.2
Non-Automated	3853	66.7	-0.121	-12.0
All	4225	67.9	-0.139	-14.2

Table 8.8: Estimated CATEs by Automation Status, RF Estimator (E6)

Table 8.9 reports the percentage of significant reducers for different confidence levels, obtained with the permutation test under the null (8.23). From Tables 8.8 and 8.9, it

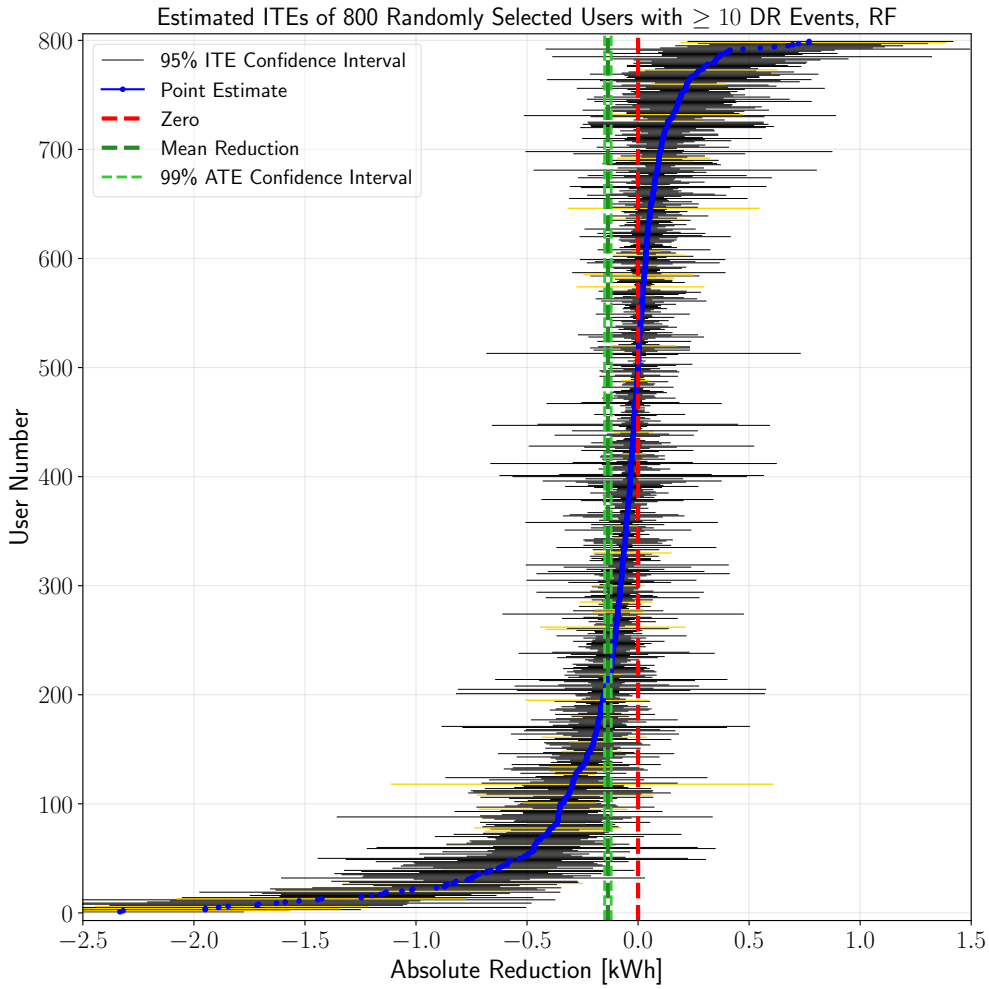

Figure 8.22: Distribution of ITEs with Bootstrapped Confidence Intervals

becomes clear that automated users show larger reductions than non-automated ones, which agrees with expectations. Lastly, Figure B.6 in the Appendix gives rise to a noticeable positive correlation between ambient air temperature and the ITE.

8.6 ATE Estimation with Fixed Effects Models

To estimate the ATE of DR interventions on electricity consumption, we consider the following fixed-effects model with raw consumption (kWh) as the dependent variable:

$$\text{kWh}_{it} = X_{it} \cdot \beta + \alpha_{it} + u_{it}. \tag{8.26}$$

In (8.26), subscripts i and t refer to user i at time t, respectively. X_{it} is a row vector of observable covariates, α_{it} are unobserved fixed effects, and u_{it} is the noise term which is

Fraction of Significant Reducers (among sample of size 4225)			
	$1-\alpha = 0.9$	$1-\alpha = 0.95$	$1-\alpha = 0.99$
# Automated	182	158	131
% of Total	48.9	42.5	35.2
# Non-Automated	1240	1041	695
% of Total	32.2	27.0	18.0
# All	1422	1199	826
% of Total	33.7	28.4	19.6

Table 8.9: Estimated Percentage of Significant Reducers according to Permutation Test, RF Estimator (E6)

assumed to be uncorrelated with the regressors and Gaussian distributed with zero mean and finite variance. The fixed effects term α_{it} removes persistent differences across users in their hourly and monthly consumption interacted with a business day indicator variable:

$$\alpha_{it} \sim \mathrm{C}(\mathrm{HoD}_{it}) : \mathrm{C}(\mathrm{is\_Bday}_{it}) + \mathrm{C}(\mathrm{MoY}_{it}). \tag{8.27}$$

Recall Assumption 15, which states that for an unbiased estimate of the ATE, we require a *randomized* assignment of reward levels to users. However, we observe a notable correlation between the reward level and the CAISO BL, suggesting that the DRP might systematically assign larger reward levels to users with a higher projected baseline. To test this conjecture with a nonparametric hypothesis test, we again make use of a permutation test with the following null and alternative hypothesis:

$$H_0 : \max_{i,j \in \mathcal{R}} \mathrm{KS}_{ij} = 0, \qquad H_1 : \max_{i,j \in \mathcal{R}} \mathrm{KS}_{ij} > 0, \tag{8.28}$$

where KS_{ij} denotes the KS distance between the empirical CAISO BL distributions for all DR events observed for reward levels i and j. The idea of this formulation is to exploit the idea that a randomized assignment of reward levels to users is reflected in a small KS distance and vice versa. To perform a permutation test, we pool all CAISO BL observations during DR events and randomly assign them to differing reward levels, where each group has the same sample size as the original group. Repeating this process many times yields a permutation distribution of the metric $\max_{i,j \in \mathcal{R}} \mathrm{KS}_{ij}$, from which the p-value associated with (8.28) follows. After 10,000 such iterations, we obtain a p-value of 0.0975, such that we can reject the null at the $1-\alpha = 0.9$ confidence level. This is indicative of a non-random assignment of reward levels to users. To account for this fact, we include the CAISO BL in the regression specifications described more closely in the following subsections.

Due to space constraints, the regression tables, which include point estimates and their 95% confidence intervals, t-values of the regression, and clustered standard errors, are relegated to the appendix.

8.6.1 Estimation by Incentive Level

To estimate the CATE by incentive level, the covariate matrix X_{it} in (8.26) is specified as follows:

$$X_{it} = [\text{is\_treat}_{it} \quad \text{BL}_{it} \quad T_{it} \quad R_{it}]\,, \tag{8.29a}$$
$$R_{it} = [\mathbb{1}(r_{it} = 0.05) \quad \cdots \quad \mathbb{1}(r_{it} = 3.00)]\,. \tag{8.29b}$$

In (8.29a), is_treat_{it} is an indicator set to one for all treatment users (and zero for all control users). BL_{it} is the CAISO baseline for user i at time t, which is necessary to control for the non-random assignment of reward levels to users, T_{it} is the ambient air temperature, and R_{it} is the reward level.

8.6.2 Estimation by Hour of the Day

To estimate the CATE by hour of the day, we pool all reward levels into the indicator variable is_DR_{it}, which is one if user i received treatment at time t and zero otherwise:

$$X_{it} = [\text{is\_treat}_{it} \quad \text{BL}_{it} \quad T_{it} \quad \text{C(HoD)} : \text{is\_DR}_{it}]\,. \tag{8.30}$$

8.6.3 Estimation by Month of the Year

The CATE by month of the year is found in a similar fashion to the CATE by hour of the day:

$$X_{it} = [\text{is\_treat}_{it} \quad \text{BL}_{it} \quad T_{it} \quad \text{C(MoY)} : \text{is\_DR}_{it}]\,. \tag{8.31}$$

8.6.4 Role of Smart Home Automation

The CATE by automation status is determined by introducing the indicator is_auto_{it}:

$$X_{it} = [\text{is\_treat}_{it} \quad \text{BL}_{it} \quad T_{it} \quad \text{C}(\text{is\_auto}_{it}) : \text{is\_DR}_{it}]\,. \tag{8.32}$$

8.6.5 Effect of Automation Uptake Encouragement

Lastly, the effect of incentivizing users to purchase a smart home automation device on energy consumption during DR events is determined as follows:

$$\begin{aligned} X_{it} = [&\text{is\_enc}_{it} \quad \text{is\_nonenc}_{it} \quad \text{BL}_{it} \quad T_{it} \\ &\text{is\_enc}_{it} \cdot \text{is\_DR}_{it} \quad \text{is\_nonenc}_{it} \cdot \text{is\_DR}_{it}]. \end{aligned} \tag{8.33}$$

In (8.33), the indicators is_enc and is_nonenc are 1 for all users in the "Treatment-Encouraged" and in "Treatment-Non-Encouraged", respectively, and zero otherwise.

8.7 Comparison of Estimation Methods

In this section, we benchmark the results obtained from the best non-experimental estimator (RF) to those from the fixed effects model with specification (8.29a).

Figure 8.23 compares the point CATEs by reward levels and their 95% confidence intervals. It can be seen that the point estimates are close to each other (-0.123 kWh aggregated for fixed effects vs. -0.129 for non-experimental approach with RF, a less than 5% difference), a finding that suggests that our non-experimental estimation technique produces reliable estimates comparable to the experimental gold standard. The fact that the confidence intervals are notably tighter for RF corroborates this notion. The remaining comparisons by

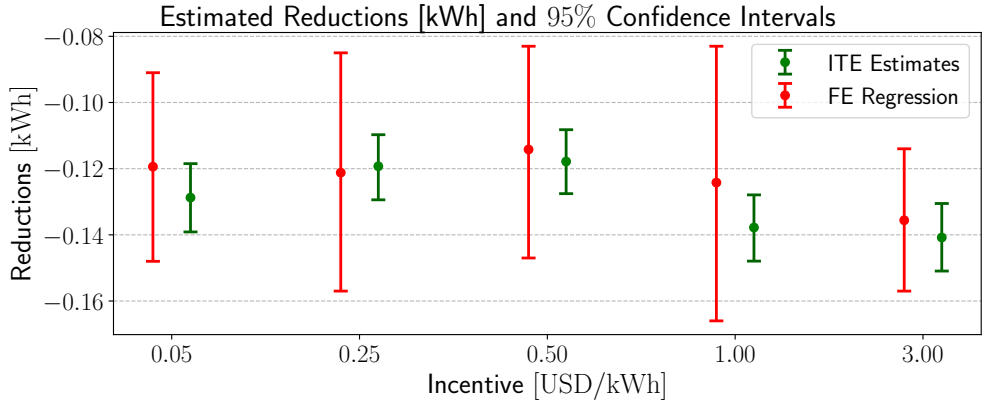

Figure 8.23: CATEs by Incentive Level with Confidence Intervals, Comparison Fixed Effects Estimators and Non-Experimental Estimators

month of the year, automation status, and automation takeup encouragement, accompanied by the Fixed Effects regression results, are relegated to the appendix.

8.8 Effect of Adaptive Targeting

The goal of adaptive targeting is to maximize the reduction per dollar paid to the users, which is achieved by either minimizing the payout and/or maximizing users' reductions. Since we have no a-priori understanding about the distribution of *individual* users' elasticities in response to incentives, we focus on minimizing the payouts to users. Assuming that users are relatively price inelastic (indeed Figure 8.23 only shows a weak negative slope of the demand curve), we explore the potential of assigning large incentives to low reducers (and small incentives to high reducers) to minimize the total payout from DRP to users.

Width of CATE Confidence Intervals (kWh) by Incentive Level					
	0.05 $\frac{\text{USD}}{\text{kWh}}$	0.25 $\frac{\text{USD}}{\text{kWh}}$	0.50 $\frac{\text{USD}}{\text{kWh}}$	1.00 $\frac{\text{USD}}{\text{kWh}}$	3.00 $\frac{\text{USD}}{\text{kWh}}$
BL	0.0508	0.0459	0.0503	0.0474	0.0481
KNN	0.0422	0.0409	0.0437	0.0396	0.0420
OLS	0.0407	0.0387	0.0428	0.0401	0.0419
L1	0.0395	0.0402	0.0430	**0.0391**	0.0414
L2	0.0391	0.0400	0.0429	0.0419	0.0419
DT	0.0452	0.0437	0.0452	0.0449	0.0468
RF	**0.0389**	**0.0383**	**0.0419**	0.0392	**0.0402**

Table 8.10: Width of 95 % Confidence Intervals around ATE Point Estimate by Incentive Level, All Estimators

8.8.1 Simulations for Adaptive Targeting

First and foremost, we carry out a simple targeting exercise to determine which criterion is most suitable for maximizing the desired metric in Phase 2 of the experiment. The idea is to assign users into one of two targeted groups, based on one of the following criteria estimated from Phase 1 responses:

- ITE (kWh or % reduction)
- ITE normalized by average reward level received (kWh or % reduction)
- Intercept of estimated individual demand curve (kWh or % reduction)
- Slope of estimated individual demand curve (kWh or % reduction)
- Random assignment (benchmark)

Each of these four criteria are computed in kWh and % values. Algorithm 7 describes the steps of the simulation in detail.

We repeat Algorithm 7 2,000 times to obtain a distribution of the metrics, which we report in Figures 8.24 - 8.26.

Figure 8.24 plots the distribution of payouts from the DRP to targeted and non-targeted simulated users. The top panel verifies that random assignment of targeted users to $\mathcal{H}$ and $\mathcal{L}$ achieves an average payout that matches the payout of the non-targeted group, thereby verifying the validity of random assignment. Panels 2-5 include this random benchmark, which serves as a comparison for the payout to the targeted users under varying targeting criteria $\{\theta_i\}_{i\in\mathcal{I}}$. We observe that the criteria ITE, ITE normalized, as well as intercept all reduce the payout significantly, however, ITE performs best. According to the bottom panel, the slope criterion underperforms.

Algorithm 7 Simulation of Adaptive Targeting of Users

Input: Set of non-targeted Phase 2 users $\mathcal{N}$

1: Randomly split $\mathcal{N} = \mathcal{T} \cup \overline{\mathcal{T}}$ of equal size, i.e. simulated-targeted $\mathcal{T}$ and simulated non-targeted $\overline{\mathcal{T}}$

2: Sort users in $\mathcal{T}$ on targeting criteria θ_i using actual variation from Phase 1 and assign to simulated high/low groups $\mathcal{H}$, $\mathcal{L}$, i.e. $\mathcal{T} = \mathcal{H} \cup \mathcal{L}$. Use targeting strategy g:

- For the first three criteria: $g(\hat{\theta}_i) = \begin{cases} H & \text{if } \hat{\theta}_i > \text{median}(\{\theta_i\}_{i\in\mathcal{I}}) \\ L & \text{if } \hat{\theta}_i \leq \text{median}(\{\theta_i\}_{i\in\mathcal{I}}) \end{cases}$,
- For the slope criterion: $g(\hat{\theta}_i) = \begin{cases} H & \text{if } \hat{\theta}_i \leq \text{median}(\{\theta_i\}_{i\in\mathcal{I}}) \\ L & \text{if } \hat{\theta}_i > \text{median}(\{\theta_i\}_{i\in\mathcal{I}}) \end{cases}$,
- For random assignment: $\mathcal{T} = \mathcal{H} \cup \mathcal{L}$ randomly. Serves as benchmark to gauge other four targeting methods against.

3: Generate synthetic reward levels:

- Users in $\mathcal{L}$ receive low incentives uniformly drawn from $\{0.05\frac{\text{USD}}{\text{kWh}}, 0.25\frac{\text{USD}}{\text{kWh}}, 0.50\frac{\text{USD}}{\text{kWh}}\}$
- Users in $\mathcal{H}$ receive high incentives uniformly drawn from $\{1\frac{\text{USD}}{\text{kWh}}, 3\frac{\text{USD}}{\text{kWh}}\}$

4: Using actual DR events $\{\mathcal{T}_i\}_{i\in\mathcal{I}}$ and experimental variation from Phase 2, calculate the following metrics:

- The difference in average consumption between low targeted ($\mathcal{L}$) and high targeted ($\mathcal{H}$) groups (ATE of targeting):

$$\Delta y = \sum_{i\in\mathcal{H}} \sum_{t\in\mathcal{T}_i} y_{it} - \sum_{i\in\mathcal{L}} \sum_{t\in\mathcal{T}_i} y_{it} \tag{8.34}$$

- Estimated ITEs (reductions) for high and low-targeted groups:

$$\hat{\theta}_{\mathcal{H}} := \sum_{i\in\mathcal{H}} \sum_{t\in\mathcal{T}_i} (\hat{y}_{it} - y_{it}), \tag{8.35a}$$

$$\hat{\theta}_{\mathcal{L}} := \sum_{i\in\mathcal{L}} \sum_{t\in\mathcal{T}_i} (\hat{y}_{it} - y_{it}) \tag{8.35b}$$

- Difference of estimated reductions for high and low-targeted groups:

$$\Delta\hat{\theta} := \hat{\theta}_{\mathcal{H}} - \hat{\theta}_{\mathcal{L}} = \sum_{i\in\mathcal{H}} \sum_{t\in\mathcal{T}_i} (\hat{y}_{it} - y_{it}) - \sum_{i\in\mathcal{L}} \sum_{t\in\mathcal{T}_i} (\hat{y}_{it} - y_{it}) \tag{8.36}$$

- Total payouts for $\mathcal{T} = \mathcal{H} \cup \mathcal{L}$ and non-targeted $\overline{\mathcal{T}}$ based on actual responses during Phase 2 DR events. Use actual incentive levels from the actual experiment for $\overline{\mathcal{T}}$, but synthetic levels based on targeting assignment for $\mathcal{L}$ and $\mathcal{H}$:

$$\pi_{\mathcal{T}} = \sum_{i\in\mathcal{T}} \sum_{t\in\mathcal{T}_i} r_{it}(\hat{y}_{it} - y_{it}), \tag{8.37a}$$

$$\pi_{\overline{\mathcal{T}}} = \sum_{i\in\overline{\mathcal{T}}} \sum_{t\in\mathcal{T}_i} r_{it}(\hat{y}_{it} - y_{it}), \tag{8.37b}$$

where we compute counterfactuals $\hat{y}_{it}$ with either the CAISO 10-in-10 BL or an ML estimator of choice.

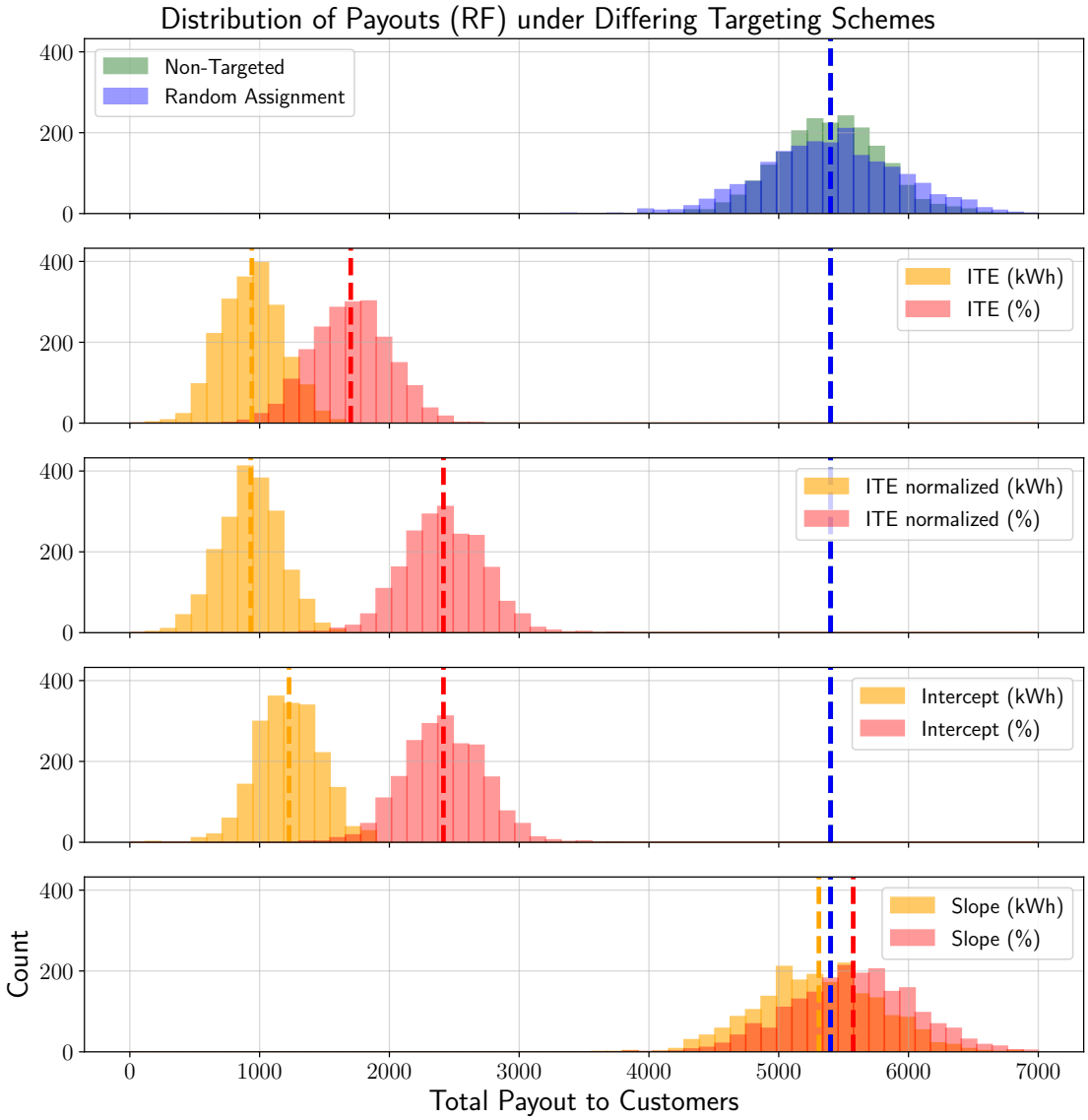

Figure 8.24: Distribution of Payouts to Users

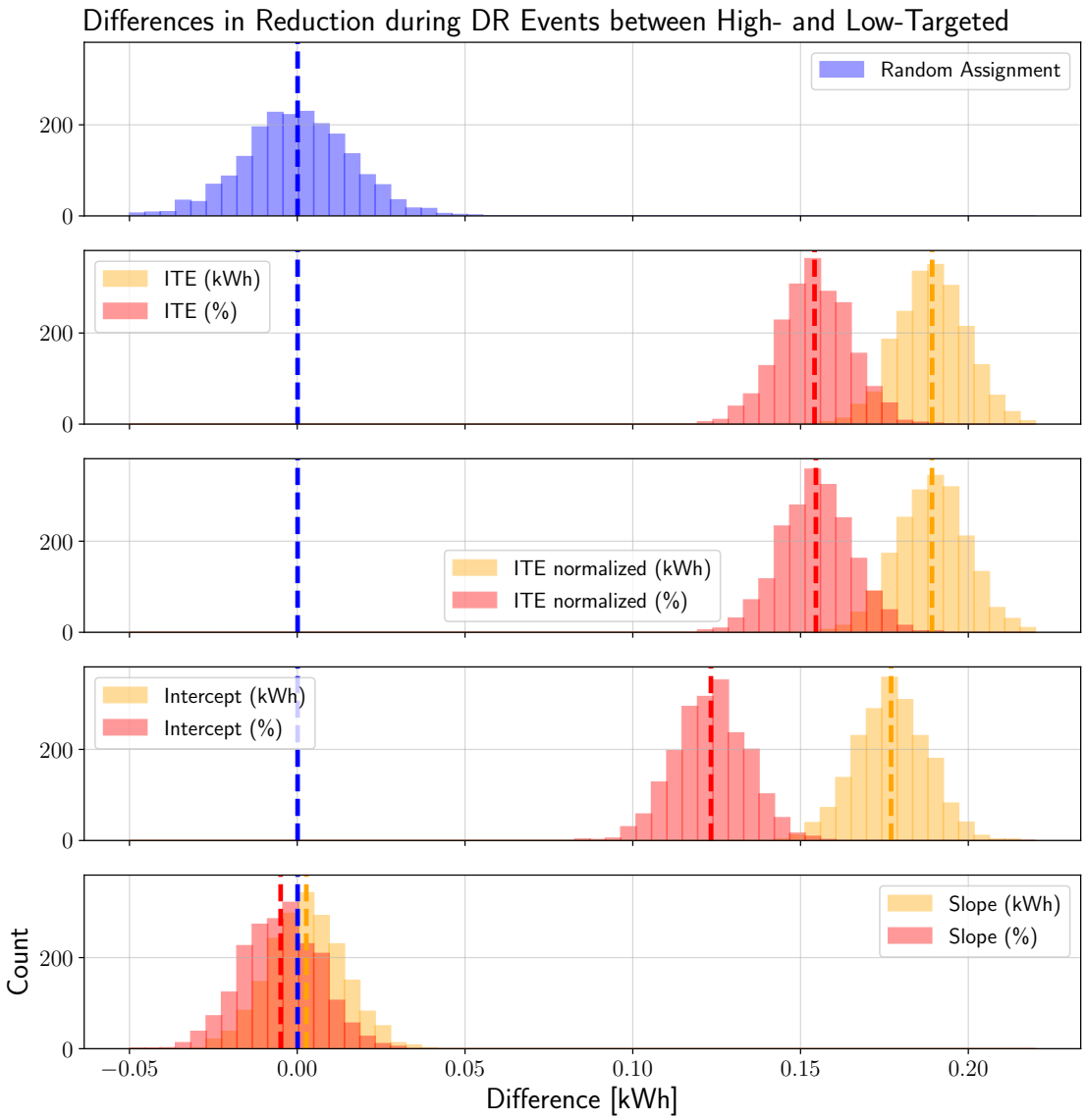

Figure 8.25: Differences in Estimated Reduction between Non-Targeted and Targeted Users, computed with RF

Figure 8.25 plots the distribution of differences in means between estimated Phase 2 reductions of simulated low and high targeted users Δy. Random assignment (top panel) shows a distribution centered around zero. Panels 2-4 graph distributions centered around a positive value, which is consistent with the objective of assigning high (low) rewards to low (large) Phase 1 reducers. ITE (kWh) achieves the largest mean difference whereas the slope criterion appears to be ineffective.

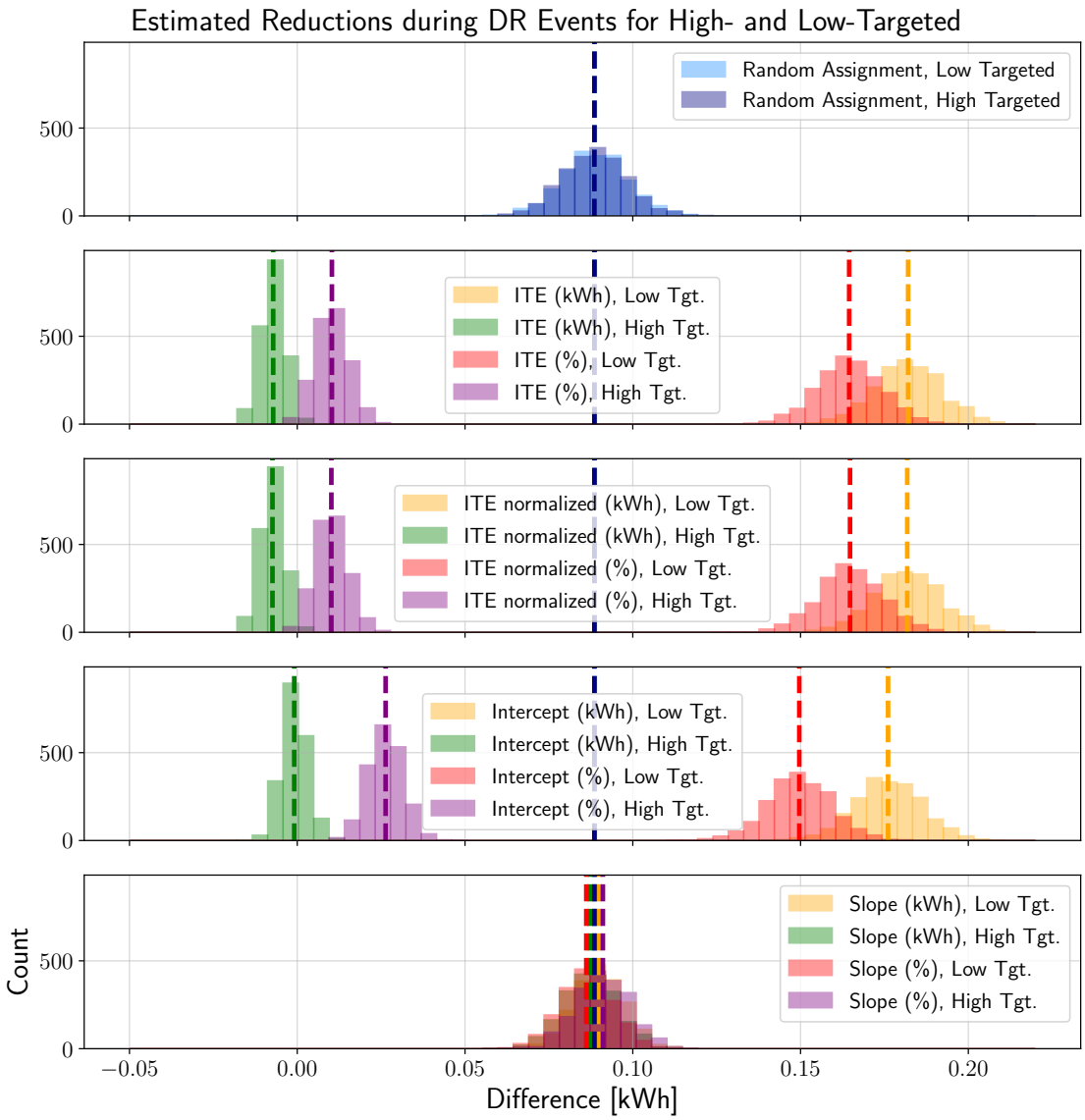

Figure 8.26: Estimated Reductions of Non-Targeted and Targeted Users, computed with RF

Figure 8.26 supports the results in Figure 8.25, as it breaks down the reductions of simulated low and high targeted users. The differences in means between the reductions estimated for low and high targeted users (for a particular targeting criterion) is exactly the mean given in Figure 8.25.

Algorithm 8 describes the targeting assignment algorithm on a given set of users, which we denote with $\mathcal{S}$. Users are transitioned into Phase 2 on a weekly basis. That is, for a particular week, all users who have reached 90 days of age in Phase 1 form the current weekly cohort, which is *randomly* split into a non-targeted group $\mathcal{S}_n$ and targeted group $\mathcal{S}_t$ of equal

Algorithm 8 Adaptive Targeting of Users

Input: Set of users $\mathcal{S}$ with completed Phase 1
Output: Groups $\mathcal{S}_n$, $\mathcal{S}_l$, and $\mathcal{S}_h$ of proportion $2:1:1$

1: Randomly split $\mathcal{S} = \mathcal{S}_n \cup \mathcal{S}_t$, $-1 \leq |\mathcal{S}_n| - |\mathcal{S}_t| \leq 1$
2: Estimate ITE $\hat{\beta}_i$ with E6 for each user $i \in \mathcal{S}_t$.
3: Sort $\{\beta_i\}_{i \in \mathcal{S}}$ in ascending order
4: Split $\mathcal{S}_t = \mathcal{S}_l \cup \mathcal{S}_h$, such that $-1 \leq |\mathcal{S}_l| - |\mathcal{S}_h| \leq 1$ and $\hat{\beta}_i < \hat{\beta}_j \; \forall \; i \in \mathcal{S}_l, j \in \mathcal{S}_h$
5: Assign users in $\mathcal{S}_l$ and $\mathcal{S}_h$ to low- and high-targeted, respectively

size (ties are broken randomly). For each user in $\mathcal{S}_t$, we calculate the ITE based on Phase 1 events. These ITEs are then sorted in ascending order. The 50% of the largest reducers (with the most negative ITEs) are defined to be the low-targeted group $\mathcal{S}_l$, whereas the other half is assigned to high-targeted group $\mathcal{S}_h$. This targeting scheme appears to be a double-edged sword: On the one hand, the DRP pays less money to large reducers and also achieves larger reductions for previously small reducers, increasing the desired ratio. On the other hand, previously large reducers now reduce less (in response to smaller rewards) and previously small reducers are paid more money for increased reductions, thereby counteracting the desired goal. However, the latter factors are dominated by the gains from the former ones, as we show in Section 8.8.5.

8.8.2 Targeting Assignment Algorithm

Recall the model for the estimation of user-level counterfactuals (8.13), here repeated for convenience:

$$y_{it} = f_i(\mathbf{x}_{it}) + D_{it} \cdot \beta_{it}(\mathbf{x}_{it}) + \varepsilon_{it}. \tag{8.38}$$

Also, recall that an estimate of user i's ITE $\hat{\beta}_i$ is obtained by marginalizing $\{\hat{\beta}_{it}\}_{t \in \mathcal{T}_i}$. We now want to define a targeting strategy $g(\beta_i, \boldsymbol{\beta}_{-i}) = g(\boldsymbol{\beta})$ that maps ITE estimates $\boldsymbol{\beta}$ to treatment assignments (we abuse notation and write $\boldsymbol{\beta}$ in lieu of $\hat{\boldsymbol{\beta}}$).

Defining $\boldsymbol{\beta}_i := \{\beta_{it}\}_{t \in \mathcal{T}_i} \in \mathbb{R}^{|\mathcal{T}_i|}$ and $\boldsymbol{r}_i := \{r_{it}\}_{t \in \mathcal{T}_i} \in \mathbb{R}^{|\mathcal{T}_i|}$, i.e. the collection of all one-sample treatment effects and incentive levels, we define targeting to be efficient if

$$\mathbb{E}_{i \in \mathcal{I}}[h(\boldsymbol{\beta}_i, \boldsymbol{r}_i) \mid g(\beta_i)] > \mathbb{E}_{i \in \mathcal{I}}[h(\boldsymbol{\beta}_i, \boldsymbol{r}_i) \mid D_i], \tag{8.39}$$

where $h(\cdot, \cdot) = \mathbb{R}^{|\mathcal{T}_i|} \times \mathbb{R}^{|\mathcal{T}_i|} \to \mathbb{R}$ maps the estimated one-sample treatment effects $\boldsymbol{\beta}_i$ and incentive levels $\boldsymbol{r}_i$ to a desired (scalar) targeting metric. D_i is random assignment to reward levels. In other words, the marginalization of user-level metrics shall outperform random assignment in terms of a scalar metric. The idea for Phase 2 targeting is simple:

- Right after Phase 1, estimate a set of ITEs $\{\hat{\beta}_i\}_{i \in \mathcal{I}}$.

- Right before Phase 2 begins, analyze the distribution of estimated ITEs, define a targeting metric h, and then design the targeting strategy (or assignment strategy) g for subsequent Phase 2 interventions.

8.8.3 Targeting Assignment Strategy and Metrics

We use a simple partitioning methodology that assigns users from Phase 1 to two different targeted groups in Phase 2. More specifically, let $g(\beta_i, \boldsymbol{\beta}_{-i})$ be defined as follows:

$$g(\beta_i, \boldsymbol{\beta}_{-i}) = \begin{cases} H & \text{if } \beta_i > \text{median}(\{\beta_i\}_{i\in\mathcal{I}}) \\ L & \text{if } \beta_i \leq \text{median}(\{\beta_i\}_{i\in\mathcal{I}}) \end{cases}, \tag{8.40}$$

where $H = \text{unif}\{1, 3\}$ and $L = \text{unif}\{0.05, 0.10, 0.25\}$, that is, users in H (L) receive high (low) incentive levels during Phase 2. More concretely, the 50% of the largest (smallest) reducers from Phase 1 receive low (high) incentives in Phase 2. This targeting scheme appears to be a double-edged sword: On the one hand, the DRP pays less money to large reducers and also achieves larger reductions for previously small reducers, increasing the desired ratio. On the other hand, previously large reducers now reduce less (in response to smaller rewards) and previously small reducers are paid more money for increased reductions, thereby counteracting the desired goal. However, the latter factors are dominated by the gains from the former ones, as we show in Section 8.8.5.

From the perspective of the DR Provider, the goal is to maximize the reduction per dollar paid to the users, which is achieved by either minimizing the payout and/or maximizing users' reductions. Accordingly, we define the targeting metric $h_1(\boldsymbol{\beta}_i, \boldsymbol{r}_i)$ to be

$$h_1(\boldsymbol{\beta}_i, \boldsymbol{r}_i) = \boldsymbol{\beta}_i^\top \mathbf{1} / (\boldsymbol{\beta}_i^\top \boldsymbol{r}_i), \tag{8.41}$$

that is, the sum of event-wise reductions $\boldsymbol{\beta}_i^\top \mathbf{1}$ divided by the payments made to user i, $\boldsymbol{\beta}_i^\top \boldsymbol{r}_i$. Alternatively, to take varying incentive levels offered to the user into account, we can also extend (8.41) as follows:

$$h_2(\boldsymbol{\beta}_i, \boldsymbol{r}_i) = \frac{\boldsymbol{\beta}_i^\top \mathbf{1}}{\boldsymbol{\beta}_i^\top \boldsymbol{r}_i} \Big/ \frac{\boldsymbol{r}_i^\top \mathbf{1}}{|\boldsymbol{r}_i|} = \frac{h_2 |\boldsymbol{r}_i|}{\boldsymbol{r}_i^\top \mathbf{1}}. \tag{8.42}$$

That is, we normalize h_1 defined in (8.41) by the average incentive level offered to user i, as higher incentive levels offered to users are likely to elicit larger treatment effects.

8.8.4 Practical Considerations

Users are transitioned into Phase 2 on a weekly basis. That is, for a particular week, all users who have reached 90 days of age in Phase 1 form the current weekly cohort, which is *randomly* split into a non-targeted group $\overline{\mathcal{T}}$, whose users continue to receive all incentive levels, and a targeted group $\mathcal{T}$ of equal size (ties are broken randomly). $\overline{\mathcal{T}}$ acts as a control group against which we benchmark the metrics (8.41) and (8.42). Users in $\mathcal{T}$ are then assigned to H and L according to (8.40).

8.8.5 Results of Adaptive Targeting

To obtain targeting metrics averaged across users in the targeted and non-targeted groups, we marginalize (8.41) and (8.42) across targeted and non-targeted users to obtain

- $H_1^{\mathcal{T}}$ and $H_1^{\overline{\mathcal{T}}}$, i.e. (8.41) for all targeted/non-targeted users
- $H_2^{\mathcal{T}}$ and $H_2^{\overline{\mathcal{T}}}$, i.e. (8.42) for all targeted/non-targeted users

Table 8.11 shows the results. RF predicts a difference of $H_1^{\mathcal{T}} - H_1^{\overline{\mathcal{T}}} \approx 1.464 \frac{\text{kWh}}{\text{USD}}$, or an

Targeting Metrics for Phase 2						
Estimator	$H_1^{\mathcal{T}}$	$H_1^{\overline{\mathcal{T}}}$	$H_2^{\mathcal{T}}$	$H_2^{\overline{\mathcal{T}}}$	$\text{CATE}_{\mathcal{T}}$	$\text{CATE}_{\overline{\mathcal{T}}}$
BL	1.411	1.044	1.273	1.070	−0.0743	−0.0861
L2	2.934	1.070	2.655	1.096	−0.0441	−0.0580
OLS	2.684	1.069	2.429	1.096	−0.0484	−0.0628
RF	2.549	1.085	2.306	1.112	−0.0464	−0.0609

Table 8.11: Targeting Results for 2,733 users between June 29, 2017 - December 31, 2017. $[H_1] = \frac{\text{kWh}}{\text{USD}}$, $[H_2] = \frac{\text{kWh}^2}{\text{USD}^2}$, [CATE] = kWh.

increase of $\approx$ 135%. That is, the targeting mechanism is capable of achieving the same reduction for the targeted as for the non-targeted group with just $\approx$ 43% of the payout. If we normalize H_1 with the average reward offered, RF predicts a difference of $H_2^{\mathcal{T}} - H_2^{\overline{\mathcal{T}}} \approx 1.194 \frac{\text{kWh}}{\text{USD}}$, which is an increase of $\approx$ 107%.

Additionally, observe that the absolute value of the CATE among targeted ($|\text{CATE}_{\mathcal{T}}|$) users is smaller than for non-targeted users ($|\text{CATE}_{\overline{\mathcal{T}}}|$), which indicates that the extent to which low-targeted users (large reducers in Phase 1) reduce their reductions dominates the increase in reduction among high-targeted users (small reducers in Phase 2). However, this net reduction of willingness to temporarily reduce electricity consumption is greatly dominated by the assignment of high (low) rewards to low (large) reducers in Phase 1, which explains the 135% and 107% increase of the H_1 and H_2 metrics, respectively.

8.9 Effect of Moral Suasion

Phase 3 of the experiment explores the opportunity of eliciting reductions in electricity consumption by offering non-monetary incentives to users. Specifically, text messages attempt to appeal to the environmental consciousness of participants. Example text messages for the experimental groups are found below:

- **Moral Suasion Only**: “Saving energy now could keep a dirty power plant off!”

- **Price Only**: "You will receive 100 points for every kWh you reduce below your forecast"
- **Moral Suasion and Price**: "You will receive 100 points for every kWh you reduce below your forecast. Saving energy now could keep a dirty power plant off!"

Figure 8.27 provides point estimates and 95% bootstrapped confidence intervals of estimated reductions in Phase 3 broken out by month of the year. Table 8.12 reports the reductions aggregated over all months. The results show that environmental language is able to achieve more than half of the reduction ($\approx$ 55%) that is achievable with a pure monetary incentive at 1 USD/kWh. Interestingly, combining both price and moral suasion only achieves an increase of $\approx$ 7% compared to the price only case, suggesting that people are a lot more price conscious than environmentally conscious.

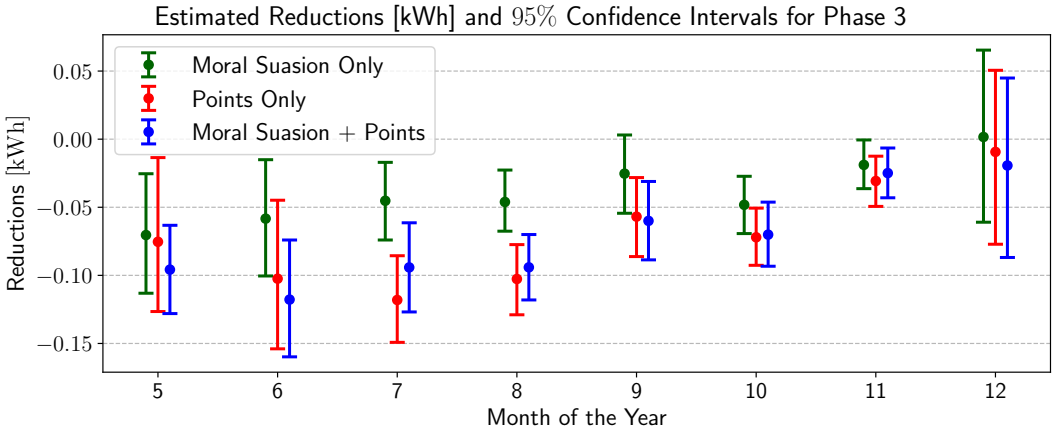

Figure 8.27: Estimated Reductions in Phase 3

Estimated Reductions [kWh] by Message Type for Phase 3			
	Moral Suasion Only	Price Only	Both
# Messages	44290 (33.8%)	43578 (33.3%)	43166 (32.9%)
Reduction	−0.040 kWh	−0.073 kWh	−0.078 kWh

Table 8.12: Aggregated Phase 3 Reductions

8.10 Conclusion

We analyzed Residential Demand Response as a human-in-the-loop cyber-physical system that incentivizes users to curtail electricity consumption during designated hours. Utilizing data collected from a Randomized Controlled Trial funded by the CEC and conducted by a

Demand Response provider in the San Francisco Bay Area, we estimated the causal effect of hour-ahead price interventions on electricity reduction. To the best of our knowledge, this is the first major study to investigate DR on such short time scales.

We developed a *non-experimental* estimation framework and benchmarked its estimates against those obtained from an *experimental* Fixed-Effects Linear Regression Model. Importantly, the former does not depend on the existence of an experimental control group to construct counterfactuals that are necessary to estimate the treatment effect. Instead, we employ off-the-shelf regression models to learn a consumption model on non-DR periods, which can then be used to predict counterfactuals during DR hours of interest. We find that the estimated treatment effects from both approaches are close to each other. The estimated ATE is -0.13 kWh (14%) per Demand Response event and household. Further, we observe a weak positive correlation between the incentive level and the estimated reductions, suggesting that users are only weakly elastic in response to incentives.

The fact that the estimates obtained from both approaches are close to each other is encouraging, as our non-experimental framework permits to go a step further compared to the experimental method in that it allows for an estimation of *individual* treatment effects. From an economic perspective, being able to differentiate low from high responders allows for an adaptive targeting scheme, whose goal is to minimize the total payout to users while maximizing total reductions. We utilize this fact to achieve an increase of the reduction-per-reward ratio of 43%.

Moreover, we discovered notable heterogeneity of users in time and by automation status, since the largest reductions were observed in summer months as well as among users with at least one connected smart home automation device. Further, the ambient air temperature was found to positively correlate with the amount of reductions, suggesting that air conditioning units are a major contributor to reductions.

Lastly, we emphasize that our non-experimental estimation framework presented in this chapter is generalizable to any human-in-the-loop cyber-physical system that requires the incentivization of users to achieve a desired objective. This is because our non-experimental framework admits results on an individual user level, which could be of particular interest in the incentivization of users in transportation systems.

Future work includes the analysis of adversarial user behavior (baseline gaming) and advanced effects including peer and network effects influencing Demand Response.

Chapter 9

Conclusion and Further Work

We investigated interactions of market participants in electricity markets by deriving results from the domains of game theory, mechanism design, and online learning. In addition to these theoretical findings, we reported results from a Residential Demand Response study on $10,000$ users in California and conducted experiments on a commercial building on the University of California, Berkeley campus to explore energy efficient control mechanisms.

While this thesis provides a solid representation of a few of the topics surrounding electricity markets in California, much more work remains to be done. Indeed, the theoretical section of this thesis makes a series of simplifying assumptions for the ease of exposition; dropping those would lead to a more involved analysis and further technicalities. Examples for such assumptions are

- The risk-neutrality of the Demand Response Provider, the structure of the Demand Response penalty and reward, and the distribution of private user types that determine the shape of the demand curve of individual users in Chapter 2,
- Marginal costs of generating companies that increase linearly as well as an assumption on the user parameters that is necessary to find a subgame perfect equilibrium in Chapter 3,
- The fact that the utility is assumed to be risk-neutral and price-taking in Chapter 4.

Lastly, the case studies presented in the second part of this thesis, while interesting in their own right, lead us to believe that further analyses could shed more light on the transferability of our findings. Sutardja Dai Hall is a large commercial building on the campus and is not structurally different from other office buildings across the nation; time-series in the form of hourly smart meter data for the Demand Response study can be similarly found in numerous other applications. Thus, utilizing the machinery presented in this thesis on different, yet structurally similar applications would be a direction worth exploring for future students.

Part III

Appendices

Appendix A

Proofs

A.1 Proofs for Chapter 2

A.1.1 Proof of Theorem 1

A.1.1.1 Individual Rationality

Notice first that each user $i \in \mathcal{T}$ is given the reward $\tilde{r}_{j(i)}$, where $j(i) \geq j_{\max} \geq i$. The first inequality is a consequence of (2.8b), which for each $i \in \mathcal{T}$ re-runs (2.8a) on the subset of users $\mathcal{I} \setminus i$. Thus, to achieve the aggregate reduction M on users $\mathcal{I} \setminus i$, where each user would be given the highest threshold reward of the targeted group, requires more users to be targeted than running the same mechanism on $\mathcal{I}$. Hence $j(i) \geq j_{\max}$. The second inequality is due to the fact that $\tilde{R}$ is sorted in ascending order, which also implies

$$\mathbb{E}[u_i(\tilde{r}_{j(i)}|\boldsymbol{\theta}_i)] \geq \mathbb{E}[u_i(\tilde{r}_i|\boldsymbol{\theta}_i)] = 0.$$

due to the monotonically increasing property of the expected utility in the reward. Thus, participation in the mechanism and being assigned to $\mathcal{T}$ results in a non-negative expected utility, compared to a zero utility for non-participation. On the other hand, users $i \notin \mathcal{T}$ receive a zero payment and so the expected utility is zero.

A.1.1.2 Incentive Compatibility

To show that the DR mechanism is incentive compatible, first note that the reward level $r(i)$ for each $i \in \mathcal{T}$ is determined *independently* of user i's bid $\mathbf{z}_i$. For each $i \notin \mathcal{T}$, user i is not given a reward. To show IC, we must therefore iterate through the following two cases:

1. $i \in \mathcal{T}$ for $\mathbf{z}_i = \boldsymbol{\theta}_i$, i.e. user i is assigned treatment with truthful reporting. This implies user i is given reward $\tilde{r}_{j(i)}$, which results in a positive expected utility. Now suppose user i had reported $\mathbf{z}_i \neq \boldsymbol{\theta}_i$. Then either the user is still assigned treatment, in which case her reward remains the same, or the user is not assigned treatment, in which case

her reward reduces to zero. Thus, misreporting could lead to a zero utility when the user could have had a positive expected utility.

2. $i \notin \mathcal{T}$ for $\mathbf{z}_i = \boldsymbol{\theta}_i$, i.e. user i is outside the treatment group with truthful reporting. If user i had reported $\mathbf{z}_i \neq \boldsymbol{\theta}_i$, then either the user is still outside the treatment group, which results in a zero utility, or the user is now in the treatment group. In the latter case, note that user i is assigned reward $\tilde{r}_{j_{\max}}$, as $j_{\max}$ is exactly the solution to (2.8b). Finally, because $\tilde{r}_{j_{\max}} < \tilde{r}_i$ (due to $i \notin \mathcal{T}$), the expected utility turns negative. Thus, misreporting does not improve the expected utility, but could lead to a negative expected utility when the user could have had a zero utility.

Combining the two cases above, it follows that misreporting the true type either yields a utility that is identical to or less than the utility in case of truthful reporting. Therefore the maximum expected utility is attained with truthful reporting, $\mathbf{z}_i = \boldsymbol{\theta}_i$, and so the DR mechanism is incentive compatible.

Lastly, to show that the mechanism terminates if $0 \leq M \leq \sum_{i=2}^{n-1} \delta_i(\tilde{r}_{n-1}|\boldsymbol{\theta}_i)$, simply notice that $j(i) \leq n \;\forall\; i \in \mathcal{T}$ (2.8b) because $\delta_1(\tilde{r}_k|\boldsymbol{\theta}_1) \geq \delta_i(\tilde{r}_k|\boldsymbol{\theta}_i),\; i \leq k \leq j(i)$ due to the monotonically increasing property of (2.1) and (2.7). Hence running the mechanism on $\mathcal{T} \setminus i$ always satisfies M.

A.2 Proofs for Chapter 3

A.2.1 Proof of Theorem 2

With Assumption 5, (3.4) follows by evaluating the first order optimality condition of (3.1) and acknowledging that its second derivative is strictly negative. Uniqueness of the Nash Equilibrium follows from Topkis' Theorem on supermodular games (Topkis 1998), which holds due to the continuity of the payoff functions (3.1) on the compact set $\mathbb{R}_+$ and increasing differences in $(x_i, \mathbf{x}_{-i})$ as $\frac{\partial^2 u_i}{\partial x_i \partial x_j} \geq 0 \;\forall\; i, j \in \mathcal{I}$.

A.2.2 Proof of Theorem 3

(3.7) is obtained by solving

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \mathbf{p}^\top \mathbf{x} - \mathbf{x}^\top C \mathbf{x} \\ \text{subject to} \quad & \mathbf{x} = (B + 2\Gamma - \Gamma W)^{-1} (\mathbf{a} - \mathbf{p}) \end{aligned} \tag{A.1}$$

and applying the Matrix Inversion Lemma for general matrices A, U, C, V of appropriate dimensions:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U \left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}.$$

The optimal profit Π^* is obtained by plugging $\mathbf{p}^*$ and $\mathbf{x}^*$ into the utility function of the monopolist.

A.2.3 Proof of Theorem 4

(3.10) is obtained in the same fashion as (3.7):

$$\begin{aligned} \underset{p}{\text{maximize}} \quad & p\mathbf{1}^\top \mathbf{x} - \mathbf{x}^\top C\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} = (B + 2\Gamma - \Gamma W)^{-1}(\mathbf{a} - p\mathbf{1}) \end{aligned}$$

Eliminating $\mathbf{x}$ from both equations and evaluating the first order optimality condition with respect to p yields (3.10).

A.2.4 Proof of Theorem 5

To derive (3.13a), we first note that since $W = W^\top$ and $C = cI$, the profit maximizing solution under complete information (3.7) simplifies to

$$p_u^* = \frac{\mathbf{1}^\top \mathbf{a}}{2n} + \mathbf{1}^\top \left[2\Gamma + B + cI - \Gamma W\right]^{-1} \frac{\mathbf{a}c}{2n}.$$

After taking the expectation with respect to the random variables $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ to obtain

$$\tilde{p}_u^* = \frac{\mathbb{E}[a]}{2} + \frac{\mathbb{E}[a]c}{2n} \mathbb{E}_b \left[\mathbf{1}^\top \left[2\Gamma + B + cI - \Gamma W\right]^{-1} \mathbf{1}^\top\right],$$

we first show convexity of the last term in $\operatorname{diag}(2b_1, \ldots, 2b_n)$. Define the matrices

$$\begin{aligned} D &= \Gamma + \frac{c}{2} I - \frac{\Gamma W}{2} + \alpha \operatorname{diag}(2b_1, \ldots, 2b_n), \\ E &= \Gamma + \frac{c}{2} I - \frac{\Gamma W}{2} + (1-\alpha) \operatorname{diag}\left(2\bar{b}_1, \ldots, 2\bar{b}_n\right), \end{aligned}$$

where $\alpha \in (0, 1)$. D and E are clearly positive definite due to the Levy-Desplanques Theorem (Horn and Johnson 2012). It is then to be shown that

$$\begin{aligned} g(X) := \mathbf{1}^\top X^{-1} \mathbf{1},\ X := \left(\alpha D + (1-\alpha)E\right)^{-1}, \\ X := 2\Gamma + B + cI - \Gamma W \end{aligned}$$

is a convex function on the domain of all positive definite matrices. Using the Schur Decomposition, which states

$$\begin{bmatrix} S & T^\top \\ T & U \end{bmatrix} \succeq 0 \Leftrightarrow S \succeq T^\top U^{-1} T,$$

and since positive definite matrices are convex,

$$\alpha \begin{bmatrix} \mathbf{1}^\top D^{-1} \mathbf{1} & \mathbf{1}^\top \\ \mathbf{1} & D \end{bmatrix} + (1-\alpha) \begin{bmatrix} \mathbf{1}^\top E^{-1} \mathbf{1} & \mathbf{1}^\top \\ \mathbf{1} & E \end{bmatrix}$$

$$= \begin{bmatrix} \alpha \mathbf{1}^\top D^{-1}\mathbf{1} + (1-\alpha)\mathbf{1}^\top E^{-1}\mathbf{1} & \mathbf{1}^\top \\ \mathbf{1} & \alpha D + (1-\alpha)E \end{bmatrix} \succeq 0,$$

This immediately shows convexity of $g(X)$:

$$\begin{aligned} \alpha g(D) + (1-\alpha)g(E) &= \alpha \mathbf{1}^\top D^{-1}\mathbf{1} + (1-\alpha)\mathbf{1}^\top E^{-1}\mathbf{1} \\ &\geq \mathbf{1}^\top (\alpha D + (1-\alpha)E)^{-1}\mathbf{1} \\ &= g(\alpha D + (1-\alpha)E). \end{aligned}$$

Finally, applying Jensen's inequality in the multivariate case on the multivariate random variable $Y := \text{diag}(2b_1, \ldots, 2b_n)$, we obtain

$$\mathbb{E}_Y \left[g(X) \right] \geq g\left(\mathbb{E}_Y [X] \right),$$

from which (3.13a) follows directly.

A.2.5 Proof of Theorem 6

Under the given conditions, it follows immediately that $\mathbb{E}[x_1^*] = \ldots = \mathbb{E}[x_n^*]$. With this constraint, taking the expectation of (3.4) yields $\mathbb{E}[x^*(p)]$ as a function of p. Plugging $\mathbb{E}[x^*(p)]$ into the utility's profit function (3.2) and taking the expectation with respect to a and b allows to compute the optimal uniform price p^* (3.14b). Next, setting $p = p^*$ in $\mathbb{E}[x^*(p)]$ yields (3.14b).

A.2.6 Proof of Theorem 7

Taking the derivative of (3.5) with respect to γ yields:

$$\frac{d\mathbf{x}}{d\gamma} = -\frac{1}{4\gamma(b+\gamma)} K^{-1}F^{-1}(\mathbf{a} - \mathbf{p}), \quad \gamma > 0$$

where we used the abbreviations

$$K := \left(I - \frac{\gamma}{(2b+2\gamma)} W \right), \; F := \left(I + \frac{b}{\gamma}(2I - W)^{-1} \right).$$

K is a strictly diagonally dominant *M-Matrix* because it can be expressed in the form $sI - B$ with $s = 1$ and has negative off-diagonal elements (Berman and Plemmons 1994). This special property guarantees that its inverse exists and is strictly diagonally dominant and entrywise positive. F is strictly diagonally dominant with positive off-diagonal entries, because $(2I - W)^{-1}$ is an *M-Matrix*. The Levy-Desplanques Theorem (Horn and Johnson 2012) then implies that F^{-1} exists, is diagonally dominant, and possesses nonnegative diagonal elements. Despite the possible negativity of its off-diagonal elements, we show that the row sums of $K^{-1}F^{-1}$ are positive. Take, for example the i-th row sum:

$$\sum_{j=1}^{n} (K^{-1}F^{-1})_{ij} = \sum_{j=1}^{n} \sum_{s=1}^{n} K_{is}^{-1} F_{sj}^{-1}$$

$$
\begin{aligned}
&= \sum_{s=1}^{n} K_{is}^{-1} F_{ss}^{-1} + \sum_{s=1}^{n} K_{is}^{-1} \sum_{j=1, j\neq s}^{n} F_{sj}^{-1} \\
&> \sum_{s=1}^{n} K_{is}^{-1} F_{ss}^{-1} - \sum_{s=1}^{n} K_{is}^{-1} \; F_{ss}^{-1} \\
&= 0.
\end{aligned}
$$

Together with $a_i > p_i \; \forall \; i \in \mathcal{I}$ (see Assumption 5), this shows that $\frac{d\mathbf{x}}{d\gamma} < 0$ for $\gamma > 0$.

A.2.7 Proof of Theorem 8

Define $L := (2I - W)$, which is a diagonally dominant matrix. Evaluating $\frac{d\mathbf{x}}{d\gamma}$ at $\gamma = 0$ yields

$$
\left.\frac{d\mathbf{x}}{d\gamma}\right|_{\gamma=0} = -\frac{1}{4}(2I - W)(\mathbf{a} - \mathbf{p}) = -\frac{1}{4} L\boldsymbol{\alpha}, \tag{A.2}
$$

where $\boldsymbol{\alpha}$ is the column vector of all $\{\alpha_i \mid i \in \mathcal{I}\}$. Evaluating this derivative for user $i \neq j,\ i \in \mathcal{C}_j$ yields

$$
\begin{aligned}
-4\left.\frac{dx_i}{d\gamma}\right|_{\gamma=0} &= L_{ii}\alpha + L_{ij}\bar{\alpha} + \sum_{k \in \mathcal{I}\setminus\{i,j\}} L_{ik}\alpha \\
&= 2\alpha - \frac{\bar{\alpha}}{n-1} - \frac{(n-2)\alpha}{n-1} \\
&< 2\alpha - \frac{n\alpha}{n-1} - \frac{(n-2)\alpha}{n-1} = 0.
\end{aligned}
$$

Hence we have $\left.\frac{dx_i}{d\gamma}\right|_{\gamma=0} > 0$. On the other hand, for the "high" consumer j, the derivative reads

$$
-4\left.\frac{dx_j}{d\gamma}\right|_{\gamma=0} = L_{jj}\alpha + \sum_{k \in \mathcal{I}\setminus j} L_{jk}\alpha = 2\alpha - \frac{n-2}{n-1}\alpha > 0,
$$

which completes the proof.

A.2.8 Proof of Lemma 2

The proof is similar to the one used for Theorem 8. For each user i, $i \in \mathcal{I} \setminus j$, the derivative reads

$$
\begin{aligned}
-4\left.\frac{dx_i}{d\gamma}\right|_{\gamma=0} &= 2\alpha - \frac{k\alpha}{m_i} - \frac{m_i - 2}{m_i - 1}\alpha = \frac{\alpha(m_i - k)}{m_i - 1} \\
&\leq \frac{(k-1) - k}{m_i - 1}\alpha < 0.
\end{aligned}
$$

For user j, we have

$$-4\frac{dx_j}{d\gamma}\bigg|_{\gamma=0} = 2\alpha - \frac{\alpha}{m_j - 1} = \frac{2m_j - 1}{m_j - 1}\alpha > 0.$$

A.2.9 Proof of Theorem 9

From Theorem 6, any user j with index $3, \ldots, n$, given the price p, consumes $(a_j - p)/(2b_j)$. To find x_1^* and x_2^*, we solve (3.4) for users 1 and 2:

$$\begin{bmatrix} 2(b_1+\gamma) & -\gamma w_{12} \\ -\gamma w_{21} & 2(b_2+\gamma) \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} a_1 - p + \gamma \sum_{j=3}^n w_{1j} x_j \\ a_2 - p + \gamma \sum_{j=3}^n w_{2j} x_j \end{bmatrix}$$

Comparing $x_1^* + x_2^*$ to the consumptions without peer effect, that is, $(a_1 - p)/(2b_1) + (a_2 - p)/(2b_2)$ yields the desired inequalities. For the special case $n = 2$, note that $w_{12} = w_{21} = 1$ and $w_{2j},\ j \geq 3$ as well as $w_{1j},\ j \geq 3$, are zero.

A.2.10 Proof of Theorem 10

The optimal pricing vector $\tilde{\mathbf{p}}^*$ under network uncertainty and its corresponding consumption vector $\tilde{\mathbf{x}}^*$ can be determined by solving (A.1) (with $W = \tilde{W}$) with respect to $\mathbf{p}$. $\tilde{\mathbf{x}}^*$ is then determined by plugging $\tilde{\mathbf{p}}^*$ back into (3.4). Let $F := B + 2\Gamma - \Gamma W$, $\tilde{F} := \lambda + 2\Gamma - \Gamma\tilde{W}$. Then $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{x}}$ are

$$\begin{aligned} \tilde{\mathbf{p}}^* &= \mathbf{a} - \tilde{F}(C + \tilde{F})^{-1}\mathbf{a}/2, \\ \tilde{\mathbf{x}}^* &= F^{-1}\tilde{F}(C + \tilde{F})^{-1}\mathbf{a}/2. \end{aligned}$$

The optimal profit $\tilde{\Pi}^* = \tilde{\mathbf{p}}^{*\top}\tilde{\mathbf{x}}^* - \tilde{\mathbf{x}}^{*\top}C\tilde{\mathbf{x}}^*$ can then be expressed as follows:

$$\tilde{\Pi}^* = \frac{1}{4}\mathbf{a}^\top(C + \tilde{F})^{-1}\mathbf{a} + \mathcal{O}(\gamma^2) \geq \frac{1}{4}\mathbf{a}^\top(C + \tilde{F})^{-1}\mathbf{a}.$$

Using the definition of Rayleigh quotients (Horn and Johnson 2012), we thus obtain the following ratio on the profit under uncertainty:

$$\frac{\tilde{\Pi}^*}{\Pi^*} \geq \frac{\mathbf{a}^\top(C + \tilde{F})^{-1}\mathbf{a}}{\mathbf{a}^\top(C + F)^{-1}\mathbf{a}} \geq \frac{\lambda_{\min}((C + \tilde{F})^{-1})}{\lambda_{\max}((C + F)^{-1})}.$$

$(C + \tilde{F})$ as well as $(C + F)$ are symmetric positive definite matrices due to their diagonal dominance with nonpositive off-diagonal elements. Hence the eigenvalues of their inverses are strictly positive. Utilizing the identity $\lambda_{\min}(A)^{-1} = 1/\lambda_{\max}(A)$ for any nonsingular matrix A, and $\|A+B\| \leq \|A\|+\|B\|$ (a fundamental property of matrix norms), further simplifications yield

$$\frac{\tilde{\Pi}^*}{\Pi^*} \geq \frac{\lambda_{\min}(C + F)}{\lambda_{\max}(C + \tilde{F})} = \frac{\lambda_{\min}(C + F)}{\|C + F + \gamma(W - \tilde{W})\|_2}$$

$$\geq \frac{\lambda_{\min}(C+F)}{\lambda_{\max}(C+F)+\gamma\|(W-\tilde{W})\|_2},$$

where we used the fact that for a symmetric positive definite matrix A, we have $\|A\|_2 \equiv \sqrt{\lambda_{\max}(A^\top A)} = \sqrt{\lambda_{\max}(A^2)} = \lambda_{\max}(A)$.

A.2.11 Proof of Theorem 11

The social welfare $\mathcal{S}$ is the sum of all users' and the monopolist's utility:

$$\mathcal{S} = \sum_{i\in\mathcal{I}} \left(a_i x_i - b_i x_i^2 - c_i x_i^2 + \gamma_i x_i \left(\sum_{j\in\mathcal{I}} w_{ij} x_j - x_i \right) \right).$$

For each $i \in \mathcal{I}$, minimizing $\mathcal{S}$ with respect to x_i yields

$$\frac{d\mathcal{S}}{dx_i} = a_i - 2(b_i + c_i + \gamma_i)x_i + \gamma_i \sum_{j\in\mathcal{I}} w_{ij}x_j + \gamma_i \sum_{j\in\mathcal{I}} w_{ji}x_i,$$

where the last term on the right hand side signifies the externalities user i imposes on its neighbors, but which are unaccounted for in the individual users' utility maximization. Solving for x_i and vectorizing the equation yields (3.17).

To show that $\mathbf{x}_i^o > \mathbf{x}_i^*$ for $\gamma > 0$, it suffices to show that $A = (C + B/2 + \Gamma - W^\top\Gamma/2 - \Gamma W/2)^{-1}$ is entrywise greater that $B = (C + B + 2\Gamma - W^\top\Gamma/2 - \Gamma W/2)^{-1}$. By performing Gauss-Jordan Elimination on A and B and exploiting the fact that A and B are diagonally dominant matrices with positive values on the diagonal and negative off-diagonal entries, this claim follows.

To show that a Pigouvian Subsidy of $s_i = \frac{1}{2}(b_i + \gamma_i)x_i^2$ restores social welfare, note that the user's utility function u_i^o now reads

$$u_i^o = a_i x_i - \frac{1}{2} b_i x_i^2 - p_i x_i + \gamma_i x_i \left(\sum_{j\in\mathcal{I}} g_{ij} x_j - \frac{1}{2} x_i \right).$$

The solution to the subgame-perfect equilibrium under the new user utility u_i^o yields x_i^o.

A.3 Proofs for Chapter 4

Throughout this section, we utilize Leibniz' Integral Rule, see Lemma 12.

Lemma 12 (Leibniz Integral Rule)**.** *For a function $f(x,t)$ with both $f(x,t)$ and $\frac{\partial f}{\partial x}$ continuous in $t \in [a(x), b(x)]$ and $x \in [x_0, x_1]$, where $a(x)$ and $b(x)$ are continuous in $x \in [x_0, x_1]$, for $x \in [x_0, x_1]$:*

$$\frac{d}{dt}\left(\int_{a(t)}^{b(t)} f(x,t)\ dx \right) = \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t} dx + f(b(t),t)\cdot b'(t) - f(a(t),t)\cdot a'(t)$$

A.3.1 Proof of Theorem 12

Taking the expectation of (4.2) with respect to the random variables λ_s and d yields:

$$\begin{aligned}\mathbb{E}[\Pi_F] = & -\bar{q}\cdot\bar{\lambda}_F + \lambda_f\int_0^{\bar{q}} xf(x)\,dx + \lambda_f\bar{q}(1-F(\bar{q})) \\ & + (\lambda_f - \mathbb{E}[\lambda_s])\int_{\bar{q}}^{\infty}(x-\bar{q})f(x)\,dx \end{aligned} \tag{A.3}$$

Using Lemma 12, its derivatives with respect to $\bar{q}$ are

$$\begin{aligned}\frac{d\mathbb{E}[\Pi_F]}{d\bar{q}} &= -\bar{\lambda}_F + \lambda_f(1-F(\bar{q})) + (\lambda_f - \mathbb{E}[\lambda_s])(F(\bar{q})-1), \\ \frac{d^2\mathbb{E}[\Pi_F]}{d\bar{q}^2} &= -\lambda_f f(\bar{q}) + f(\bar{q})(\lambda_f - \mathbb{E}[\lambda_s]) < 0,\end{aligned}$$

from which the optimal contract volume $\bar{q}^*$ (4.3a) follows. Plugging $\bar{q}^*$ back into (A.3) yields

$$\mathbb{E}[\Pi_F] = -\bar{\lambda}_F F^{-1}\left(1-\frac{\bar{\lambda}_F}{\mathbb{E}[\lambda_s]}\right) + \lambda_f\int_0^{\bar{q}} xf(x)\,dx + \frac{\lambda_f\bar{\lambda}_F}{\mathbb{E}[\lambda_s]}\,F^{-1}\left(1-\frac{\bar{\lambda}_F}{\mathbb{E}[\lambda_s]}\right),$$

from which the optimal profit (4.3b) follows.

A.3.2 Proof of Theorem 13

Similar to the previous proof, we take the expectation of (4.4) with respect to λ_s and d:

$$\begin{aligned}\mathbb{E}[\Pi_C] = & \lambda_f\mathbb{E}[d] - \int_0^{\bar{q}} xf(x)dx\int_0^{\bar{\lambda}_C} yg(y)dy - P\bar{q} - r \\ & - \bar{\lambda}_C(1-G(\bar{\lambda}_C))\int_0^{\bar{q}} xf(x)dx - \bar{q}(1-F(\bar{q}))\int_0^{\bar{\lambda}_C} yg(y)dy \\ & - \bar{q}(1-F(\bar{q}))(1-G(\bar{\lambda}_C))\bar{\lambda}_C - \mathbb{E}[\lambda_S]\int_{\bar{q}}^{\infty}(x-\bar{q})f(x)dx.\end{aligned}$$

The first order optimality condition reads

$$\begin{aligned}\frac{d\mathbb{E}\Pi_C}{d\bar{q}} = & -P + \mathbb{E}[\lambda_s](1-F(\bar{q})) \\ & - (1-F(\bar{q}))\left[\int_0^{\bar{\lambda}_C} yg(y)dy + \bar{\lambda}_C(1-G(\bar{\lambda}_C))\right],\end{aligned}$$

which yields (4.5a) at the optimum. To show that this is a maximum, we compute the second derivative:

$$\frac{d^2\mathbb{E}\Pi_C}{d\bar{q}^2} = f(\bar{q})\left[\int_0^{\bar{\lambda}_C} yg(y)dy + \bar{\lambda}_C(1-G(\bar{\lambda}_C)) - \mathbb{E}[\lambda_s]\right],$$

which is negative as we show below:

$$\begin{aligned}\int_0^{\bar{\lambda}_C} y g(y)dy + \bar{\lambda}_C(1-G(\bar{\lambda}_C)) &\overset{?}{<} \mathbb{E}[\lambda_s]\\ \bar{\lambda}_C G(\bar{\lambda}_C) - \int_0^{\bar{\lambda}_C} G(y)dy + \bar{\lambda}_C - \bar{\lambda}_C G(\bar{\lambda}_C) &\overset{?}{<} \mathbb{E}[\lambda_s]\\ 0 \leq \bar{\lambda}_C - \int_0^{\bar{\lambda}_C} G(y)dy < \bar{\lambda}_C &< \mathbb{E}[\lambda_s]\end{aligned}$$

Finally, the optimal expected profit $\mathbb{E}[\Pi_C^*]$ (4.5b) follows from plugging (4.5a) back into the expectation of (4.4).

A.3.3 Proof of Theorem 14

Taking the expectation of (4.8) with respect to λ_s and r by performing Lebesgue-Stieltjes Integration gives

$$\begin{aligned}\mathbb{E}[\Pi_{\mathrm{DR}}] &= (\lambda_f - \mathbb{E}[\lambda_s]) \int_{d_{\min}}^{d_{\max}-h(r)} x f(x+h(r))dx - r\\ &\quad + (\lambda_f - \mathbb{E}[\lambda_s])\ d_{\min} \int_{d_{\max}-h(r)}^{d_{\max}} f(x)dx\\ &= (\lambda_f - \mathbb{E}[\lambda_s]) \int_{d_{\min}+h(r)}^{d_{\max}} (x-h(r)) f(x)dx - r\end{aligned} \tag{A.4}$$

where we used the change of variables $x + h(r) \to x$ and the fact that $F(d_{\max}) = F(d_{\max} - h(r)) = 1$. With the Leibniz Integral Rule, its derivatives with respect to r read

$$\begin{aligned}\frac{d\mathbb{E}[\Pi_{\mathrm{DR}}]}{dr} &= (\lambda_f - \mathbb{E}[\lambda_s])[1-F(h(r))](-h'(r)) - 1\\ \frac{d^2\mathbb{E}[\Pi_{\mathrm{DR}}]}{dr^2} &= \underbrace{(\lambda_f - \mathbb{E}[\lambda_s])}_{\leq 0} \underbrace{[f(h)h' + (F(h)-1)h'']}_{\geq 0}\Big|_{h=h(r)}\end{aligned}$$

For the linear shift, i.e. $h(r) = \alpha r$, first order optimality yields (4.9a), which is valid only under the condition that $\alpha > (\mathbb{E}[\lambda_s] - \lambda_f)^{-1}$. The second derivative is negative due to the concavity of $h(r)$, which results in $h''(r) \leq 0$. The optimal profit Π_{DR}^* follows from plugging r^* back into (A.4):

$$\begin{aligned}\mathbb{E}[\Pi_{\mathrm{DR}}^*] &= (\lambda_f - \mathbb{E}[\lambda_s]) \int_{\alpha r^*}^{\alpha r^* + d_{\max}} (x - \alpha r^*) f(x)dx - r^*\\ &\quad + (\lambda_f - \mathbb{E}[\lambda_s])\ d_{\min}\left[F(d_{\max}) - F(d_{\max} - h(r))\right]\\ &= (\lambda_f - \mathbb{E}[\lambda_s]) \int_{F^{-1}(1-\frac{1}{\alpha(\mathbb{E}[\lambda_s]-\lambda_f)})}^{\infty} x f(x)dx\end{aligned}$$

A.3.4 Proof of Theorem 15

This theorem can be proved by showing that the determinant of the Hessian of the two-dimensional optimization problem is negative, and hence yields a saddle at each joint minimum of portfolios ($(r^*, \bar{q}_C^*$ for DR + call, $(r^*, \bar{q}_F^*)$ for DR + forward contract, $(\bar{q}_F^*, \bar{q}_C^*)$ for call + forward contract). The objectives for each of these pairwise portfolios are

$$\Pi_{\text{FC}} = \lambda_f d - \bar{\lambda}_F \bar{q}_F - P\bar{q}_C - (d - \bar{q}_F - \bar{q}_C)\lambda_s \mathbf{1}_{d>\bar{q}_F+\bar{q}_C} - (d - \bar{q}_F)\min(\lambda_s, \bar{\lambda}_C)\mathbf{1}_{\bar{q}_F \le d \le \bar{q}_F+\bar{q}_C},$$
$$\Pi_{\text{FD}} = (\lambda_f - \lambda_s)[d(r) - \bar{q}_F]_+ - \bar{\lambda}_F \bar{q}_F - r + \lambda_F d(r)\mathbf{1}_{d(r)\le\bar{q}_F} + \lambda_f \bar{q}_F \mathbf{1}_{d(r)>\bar{q}_F},$$
$$\Pi_{\text{CD}} = \lambda_f d - \lambda_s[d(r) - \bar{q}_C]_+ - P\bar{q}_C - r + \min(\bar{\lambda}_C, \lambda_s)\left[-d(r)\mathbf{1}_{d(r)\le\bar{q}_C} - \bar{q}_C \mathbf{1}_{d(r)>\bar{q}_C}\right],$$

where $\Pi_{\text{FC}}, \Pi_{\text{FD}}$, and Π_{CD} denote the profit under the pairwise portfolios (forward contract, call), (forward contract, DR), (call, DR), respectively. Taking the expectation w.r.t to the random variables d and λ_s and the derivatives w.r.t. the decision variables yields the Hessian matrix, from which further analysis proves the claim.

A.3.5 Proof of Proposition 1

Using the definition of the conditional expectation for continuous random variables X, Y

$$\mathbb{E}[X|Y] = \int_{x\in\mathbb{R}} p_{X|Y}(x|y)dx,$$

it follows that

$$\mathbb{E}[d \mid d \ge \tau] = \frac{\int_\tau^{d_{\max}} x f(x)dx}{\int_\tau^{d_{\max}} f(x)dx}, \quad d_{\min} < \tau < d_{\max}. \tag{A.5}$$

Applying (A.5) on (4.3b), (4.5b), and (4.9b) with $\tau = \alpha_F$ (4.12a), $\tau = \alpha_C$ (4.12b), and $\tau = \alpha_{\text{DR}}$ (4.12c), respectively, yields the desired expressions.

A.3.6 Proof of Proposition 2

For a uniform distribution with support $[d_{\min}, d_{\max}]$, the PDF is $f(x) = 1/(d_{\max} - d_{\min})\mathbf{1}(d_{\min} \le x \le d_{\max})$, and the inverse CDF is $F^{-1}(z) = d_{\min} + (d_{\max} - d_{\min})z$, $z \in [0, 1]$. Straightforward manipulation of the optimal expected profits (4.3b), (4.5b), and (4.9b) and using the formula for the standard deviation

$$\sigma = \frac{d_{\max} - d_{\min}}{2\sqrt{3}}$$

yields (4.14a), (4.14b), and (4.14c).

A.3.7 Proof of Theorem 16

Straightforward by pairwise comparison of equations (4.11a)-(4.11c).

A.4 Proofs for Chapter 5

A.4.1 Proofs for Stochastic Setting

A.4.1.1 Proof of Lemma 3

Recall the definition of counters $C_{i,t}$. Each time a non-optimal vector of arms is played, that is, $a_t \neq a^*$, we increment the smallest counter in the set a_t:

$$C_{j,t} \leftarrow C_{j,t} + 1, \quad j = \arg\min_{i \in a_t} C_{i,t} \tag{A.6}$$

Proof. Let $I_i(t)$ denote the indicator that $C_{i,t}$ is incremented at time t. Then for any time τ, we have:

$$\begin{aligned}
C_{i,\tau} &= \sum_{t=2}^{\tau} \mathbb{1}\left[I_i(t) = 1\right] = m + \sum_{t=2}^{\tau} \mathbb{1}\left[I_i(t) = 1,\ C_{i,t} \geq m\right] \\
&= m + \sum_{t=2}^{\tau} \mathbb{1}\left[\sum_{j \in a_t} \bar{\mu}^j_{t-1} + e_{j,t-1} \geq \sum_{j \in a^*} \bar{\mu}^j_{t-1} + e_{j,t-1}\ ,\ C_{i,t} \geq m\right] \\
&\leq m + \sum_{t=1}^{\tau} \mathbb{1}\left[\sum_{j \in a_{t+1}} U_{j,t} \geq \sum_{j \in a^*} U_{j,t}\ ,\ C_{i,t} \geq m\right] \\
&\leq m + \sum_{t=1}^{\tau} \mathbb{1}\left[\max_{m \leq n_{s(1)}, \ldots, n_{s(K)} \leq t} \sum_{j=1}^{K} U_{n_{s(j),t}} \geq \min_{1 \leq n_{s^*(1)}, \ldots, n_{s^*(K)} \leq t} \sum_{j=1}^{K} U_{n_{s^*(j),t}}\right] \\
&\leq m + \sum_{t=1}^{\infty} \sum_{n_{s(1)}=m}^{t} \cdots \sum_{n_{s(K)}=m}^{t} \sum_{n_{s^*(1)}=1}^{t} \cdots \sum_{n_{s^*(K)}=1}^{t} \mathbb{1}\left[\sum_{j=1}^{K} U_{n_{s(j),t}} \geq \sum_{j=1}^{K} U_{n_{s^*(j),t}}\right].
\end{aligned}$$

$s(j)$ and $s^*(j)$ denote the j-th nonzero element in a_{t+1} and a^*, respectively. $U_{n_{s(j),t}} = \bar{\mu}_t^{s(j)} + e_{s(j),t}$ is the upper confidence bound of arm $s(j)$ at time t after it has been played $n_{s(j),t}$ times.

Using the choice of m in Lemma A13, we obtain the lower bound on the expectation of $C_{i,\tau(B)}$ as stated in Lemma 3:

$$\begin{aligned}
&\mathbb{E}[C_{i,\tau(B)}] \\
&\leq m + \sum_{t=1}^{\infty} \sum_{n_{s(1)}=m}^{t} \cdots \sum_{n_{s(K)}=m}^{t} \sum_{n_{s^*(1)}=1}^{t} \cdots \sum_{n_{s^*(K)}=1}^{t} 2Kt^{-2(K+1)} \\
&\leq \left\lceil (K+1)\log\tau(B)\left(\frac{\Delta_{\min} + 2K(1+1/c_{\min})}{c_{\min}\Delta_{\min}}\right)^2\right\rceil + \sum_{t=1}^{\infty}(t-m+1)^K t^K 2Kt^{-2(K+1)} \\
&\leq (K+1)\left(\frac{\Delta_{\min} + 2K(1+1/c_{\min})}{c_{\min}\Delta_{\min}}\right)^2 \log\tau(B) + 1 + 2K\sum_{t=1}^{\infty} t^{-2}
\end{aligned}$$

$$\leq \underbrace{(K+1)\left(\frac{\Delta_{\min}+2K(1+1/c_{\min})}{c_{\min}\Delta_{\min}}\right)^2}_{=:\gamma}\log\tau(B)+\underbrace{1+K\frac{\pi^2}{3}}_{=:\delta}. \tag{A.7}$$

□

Lemma 13. *For the choice*

$$m \geq (K+1)\log\tau(B)\left[\frac{\Delta_{\min}+2K(1+1/c_{\min})}{c_{\min}\Delta_{\min}}\right]^2$$

we obtain the following bound:

$$\mathbb{P}\left(\sum_{j=1}^{K}\bar{\mu}_t^{s(j)}+e_{s(j),t}\geq\sum_{j=1}^{K}\bar{\mu}_t^{s^*(j)}+e_{s^*(j),t}\right)\leq 2Kt^{-2(K+1)}.$$

Proof. The proof follows ideas employed in Auer, Cesa-Bianchi, and Fischer 2002. Assuming that the event

$$\sum_{j=1}^{K}\bar{\mu}_t^{s(j)}+e_{s(j),t}\geq\sum_{j=1}^{K}\bar{\mu}_t^{s^*(j)}+e_{s^*(j),t} \tag{A.8}$$

is true, at least one of the following events must also be true:

$$\sum_{j=1}^{K}\bar{\mu}_t^{s^*(j)}\leq\sum_{j=1}^{K}\mu_{s^*(j)}-e_{s^*(j),t} \tag{A.9a}$$

$$\sum_{j=1}^{K}\bar{\mu}_t^{s(j)}\geq\sum_{j=1}^{K}\mu_{s(j)}+e_{s(j),t} \tag{A.9b}$$

$$\sum_{j=1}^{K}\mu_{s^*(j)}<\sum_{j=1}^{K}\mu_{s(j)}+2e_{s(j),t} \tag{A.9c}$$

To show this claim, assume the probabilities of events (A.9a) or (A.9b) occurring is zero. Then it follows that

$$\begin{aligned}\sum_{j=1}^{K}\bar{\mu}_t^{s(j)}+e_{s(j),t}&\geq\sum_{j=1}^{K}\bar{\mu}_t^{s^*(j)}+e_{s^*(j),t}\\&\overset{\text{(A.9a)}}{>}\mu_{s^*(j)}-e_{s^*(j),t}+e_{s^*(j),t}=\sum_{j=1}^{K}\mu_{s^*(j)}\end{aligned}$$

and

$$\sum_{j=1}^{K}\mu_{s(j)}+2e_{s(j),t}\overset{\text{(A.9b)}}{>}\sum_{j=1}^{K}\bar{\mu}_t^{s(j)}+e_{s(j),t}\geq\sum_{j=1}^{K}\bar{\mu}_t^{s^*(j)}+e_{s^*(j),t}.$$

Hence, it follows that

$$\sum_{j=1}^{K} \mu_{s^*(j)} < \sum_{j=1}^{K} \bar{\mu}_t^{s^*(j)} + e_{s^*(j),t} < \sum_{j=1}^{K} \mu_{s(j)} + 2e_{s(j),t},$$

which is exactly event (A.9c). Thus, it suffices to upper-bound the probability of events (A.9a) and (A.9b), while choosing m such that the third event (A.9c) occurs with probability zero. Using Lemma A14, we have

$$\mathbb{P}\left(A.9a \text{ true}\right) + \mathbb{P}\left(A.9b \text{ true}\right) \leq 2Kt^{-2(K+1)}. \tag{A.10}$$

Now we pick m such that event (A.9c) becomes impossible:

$$\begin{aligned}
&\sum_{j=1}^{K} \mu_{s^*(j)} - \sum_{j=1}^{K} \mu_{s(j)} - \sum_{j=1}^{K} 2e_{s(j),t} \\
&\quad = \sum_{j=1}^{K} \left[\mu_{s^*(j)} - \mu_{s(j)}\right] - 2\sum_{j=1}^{K} \frac{(1+1/c_{\min})\varepsilon_{s(j),t}}{c_{\min} - \varepsilon_{s(j),t}} \\
&=: \ \Delta_{a_{t+1}} - 2\sum_{j=1}^{K} \frac{(1+1/c_{\min})\varepsilon_{s(j),t}}{c_{\min} - \varepsilon_{s(j),t}} \\
&= \Delta_{a_{t+1}} - 2\sum_{j=1}^{K} \frac{(1+1/c_{\min})\sqrt{\frac{(K+1)\log t}{n_{s(j)}}}}{c_{\min} - \sqrt{\frac{(K+1)\log t}{n_{s(j)}}}} \\
&\geq \Delta_{a_{t+1}} - 2K\frac{(1+1/c_{\min})\sqrt{\frac{(K+1)\log \tau(B)}{m}}}{c_{\min} - \sqrt{\frac{(K+1)\log \tau(B)}{m}}} \geq 0,
\end{aligned}$$

where the last inequality is obtained by selecting m as follows:

$$m \geq (K+1)\log\tau(B)\left(\frac{\Delta_{a_{t+1}} + 2K(1+1/c_{\min})}{c_{\min}\Delta_{a_{t+1}}}\right)^2.$$

This choice of m is suitable for the *particular* choice of a_{t+1}. To falsify (A.9c) for all possible choices of a_{t+1}, we let m be defined as follows:

$$m \geq (K+1)\log\tau(B)\left(\frac{\Delta_{\min} + 2K(1+1/c_{\min})}{c_{\min}\Delta_{\min}}\right)^2.$$

□

Lemma 14. *The probabilities of the events* (A.9a) *and* (A.9b) *are upper-bounded as follows:*

$$\mathbb{P}\left(\sum_{j=1}^{K} \bar{\mu}_t^{s^*(j)} \leq \sum_{j=1}^{K} \mu_{s^*(j)} - e_{s^*(j),t}\right) \leq Kt^{-2(K+1)} \tag{A.11a}$$

$$\mathbb{P}\left(\sum_{j=1}^{K}\bar{\mu}_t^{s(j)} \geq \sum_{j=1}^{K}\mu_{s(j)} + e_{s(j),t}\right) \leq Kt^{-2(K+1)} \tag{A.11b}$$

Proof. Using the union bound on (A.11a), we obtain

$$\mathbb{P}\left(\sum_{j=1}^{K}\bar{\mu}_t^{s^*(j)} \leq \sum_{j=1}^{K}\mu_{s^*(j)} - e_{s^*(j),t}\right) \leq \sum_{j=1}^{K}\mathbb{P}\left(\bar{\mu}_t^{s^*(j)} \leq \mu_{s^*(j)} - e_{s^*(j),t}\right)$$

Analyzing

$$\bar{\mu}_t^{s^*(j)} \leq \mu_{s^*(j)} - e_{s^*(j),t} \Leftrightarrow \frac{\bar{\mu}_{r,t}^{s^*(j)}}{\bar{\mu}_{c,t}^{s^*(j)}} \leq \frac{\mu_r^{s^*(j)}}{\mu_c^{s^*(j)}} - e_{s^*(j),t}, \tag{A.12}$$

we claim that at least one of the following two events must be true:

$$\bar{\mu}_{r,t}^{s^*(j)} \leq \mu_r^{s^*(j)} - \varepsilon_{s^*(j),t}, \tag{A.13a}$$
$$\bar{\mu}_{c,t}^{s^*(j)} \geq \mu_c^{s^*(j)} + \varepsilon_{s^*(j),t}, \tag{A.13b}$$

where $\varepsilon_{s^*(j),t}$ is the low-level exploration term for the mean reward and cost defined in (A.14). The claim is true, because if both (A.13a) and (A.13b) were false, then we would have

$$\begin{aligned}
\frac{\mu_r^{s^*(j)}}{\mu_c^{s^*(j)}} - \frac{\bar{\mu}_{r,t}^{s^*(j)}}{\bar{\mu}_{c,t}^{s^*(j)}} &= \frac{\left(\mu_r^{s^*(j)} - \bar{\mu}_{r,t}^{s^*(j)}\right)\mu_c^{s^*(j)} - \left(\mu_c^{s^*(j)} - \bar{\mu}_{c,t}^{s^*(j)}\right)\mu_r^{s^*(j)}}{\bar{\mu}_{c,t}^{s^*(j)}\mu_c^{s^*(j)}} \\
&< \frac{\varepsilon_{s^*(j),t}}{\bar{\mu}_{c,t}^{s^*(j)}} + \frac{\varepsilon_{s^*(j),t}\mu_r^{s^*(j)}}{\bar{\mu}_{c,t}^{s^*(j)}\mu_c^{s^*(j)}} \\
&\leq \frac{\varepsilon_{s^*(j),t}}{c_{\min}} + \frac{\varepsilon_{s^*(j),t}\cdot 1}{c_{\min}^2} = \frac{\varepsilon_{s^*(j),t}(c_{\min}+1)}{c_{\min}^2} \\
&\leq \frac{\varepsilon_{s^*(j),t}(1+1/c_{\min})}{c_{\min} - \varepsilon_{s^*(j),t}} \overset{!}{=} e_{s^*(j),t},
\end{aligned}$$

which contradicts the claim (A.12). Now, choosing $\varepsilon_{s^*(j),t}$ as

$$\varepsilon_{s^*(j),t} = \sqrt{\frac{(K+1)\log t}{n_{s^*(j)}}} \tag{A.14}$$

allows us to bound the probability of (A.13a) and (A.13b) using Hoeffding's Inequality:

$$\begin{aligned}
\mathbb{P}\left(\bar{\mu}_{r,t}^{s^*(j)} \leq \mu_r^{s^*(j)} - \varepsilon_{s^*(j),t}\right) &\leq \exp\left(-2n_{s^*(j)}\varepsilon_{s^*(j),t}^2\right) = t^{-2(K+1)}, \\
\mathbb{P}\left(\bar{\mu}_{c,t}^{s^*(j)} \geq \mu_c^{s^*(j)} + \varepsilon_{s^*(j),t}\right) &\leq \exp\left(-2n_{s^*(j)}\varepsilon_{s^*(j),t}^2\right) = t^{-2(K+1)}.
\end{aligned}$$

From here (A.11a) follows. In a similar fashion, we can bound the probability of event (A.11b) by showing that at least one of

$$\bar{\mu}_{r,t}^{s(j)} \geq \mu_r^{s(j)} + \varepsilon_{s(j),t} \tag{A.15a}$$

$$\bar{\mu}_{c,t}^{s(j)} \leq \mu_c^{s(j)} - \varepsilon_{s(j),t} \tag{A.15b}$$

is true (similar to (A.13a) and (A.13b)), where $\varepsilon_{s(j),t}$ is now defined as

$$\varepsilon_{s(j),t} = \sqrt{\frac{(K+1)\log t}{n_{s(j)}}}. \tag{A.16}$$

More specifically, if both (A.15a) and (A.15b) were false, then (A.9b) would be false, too. Thus, (A.11b) follows. □

A.4.1.2 Proof of Lemma 4

First, notice that the optimal algorithm $\mathcal{A}^*$ knows all bang-per-buck ratios and can simply pull those K arms associated with the K largest ratios, denoted with a^*.

Lemma 15. *The optimal expected payout of $\mathcal{A}^*$, $\mathbb{E}[U_{\mathcal{A}^*}]$, is bounded from above as follows:*

$$\mathbb{E}[U_{\mathcal{A}^*}] \leq \frac{\sum_{i\in a^*}\mu_r^i}{\sum_{i\in a^*}\mu_c^i}(B+1) \tag{A.17}$$

Proof. This can be shown easily by induction. For the base case, consider $-1 \leq B \leq 0$ and so (A.17) holds trivially. Now consider the budget $B' > 0$. Then we have

$$\begin{aligned}
U_{\mathcal{A}^*}(B') &= \sum_{i\in a^*}\mu_r^i + U_{\mathcal{A}^*}\left(B' - \sum_{i\in a^*}\mu_r^i\right) \\
&\overset{\text{(A.17)}}{\leq} \sum_{i\in a^*}\mu_r^i + \frac{\sum_{i\in a^*}\mu_r^i}{\sum_{i\in a^*}\mu_c^i}\left(B' - \sum_{i\in a^*}\mu_c^i + 1\right) \\
&= \sum_{i\in a^*}\mu_r^i - \frac{\sum_{i\in a^*}\mu_r^i}{\sum_{i\in a^*}\mu_c^i}\sum_{i\in a^*}\mu_c^i + \frac{\sum_{i\in a^*}\mu_r^i}{\sum_{i\in a^*}\mu_c^i}(B'+1) = \frac{\sum_{i\in a^*}\mu_r^i}{\sum_{i\in a^*}\mu_c^i}(B'+1).
\end{aligned}$$

□

Now, let us denote the stopping time of the optimal algorithm as $\tau_{\mathcal{A}^*}(B)$. Since we know that $\mathcal{A}^*$ always selects the set of actions a^* in each round, the stopping time is

$$\tau_{\mathcal{A}^*}(B) = \left\lfloor \frac{B}{\sum_{i\in a^*}\mu_c^i} \right\rfloor.$$

Hence, we obtain the following inequality on $\tau_{\mathcal{A}^*}(B)$:

$$\frac{B}{\sum_{i\in a^*}\mu_c^i} - 1 \leq \tau_{\mathcal{A}^*}(B) \leq \frac{B}{\sum_{i\in a^*}\mu_c^i}. \tag{A.18}$$

Lemma A16 bounds stopping time $\tau_{\mathcal{A}}(B)$ of algorithm $\mathcal{A}$:

Lemma 16. *The stopping time of algorithm $\mathcal{A}$ is bounded as follows:*

$$\frac{B - NK\left(\gamma \log \tau_{\mathcal{A}}(B) + \delta\right)}{\sum_{i \in a^*} \mu_c^i} - 1 \leq \tau_{\mathcal{A}}(B) \leq \tau_{\mathcal{A}^*}(B) + \frac{N}{Kc_{\min}}\left(\gamma \log \tau_{\mathcal{A}}(B) + \delta\right).$$

Proof. Let $0 \leq B^* \leq B$ denote the budget spent on pulling optimal arms from the set a^* across all rounds $1, \ldots, \tau_{\mathcal{A}}(B)$. Similarly, let B^- denote the budget spent on pulling non-optimal arms across those rounds. To obtain the upper bound on $\tau_{\mathcal{A}}(B)$, observe the following manipulations:

$$\begin{aligned}
\tau_{\mathcal{A}}(B) &\leq \tau_{\mathcal{A}^*}(B) + \tau_{\mathcal{A}}\left(\sum_{i \notin a^*} n_{i,\tau(B)} c_{\max}\right) \\
&\leq \tau_{\mathcal{A}^*}(B) + \tau_{\mathcal{A}}\left(\sum_{i=1}^{N} C_{i,\tau_{\mathcal{A}}(B)}\right) \\
&\leq \tau_{\mathcal{A}^*}(B) + \frac{\sum_{i=1}^{N} C_{i,\tau(B)}}{Kc_{\min}} \qquad \text{(A.19)} \\
&\overset{\text{(A.7)}}{\leq} \tau_{\mathcal{A}^*}(B) + \frac{N}{Kc_{\min}}\left(\gamma \log \tau_{\mathcal{A}}(B) + \delta\right).
\end{aligned}$$

In (A.19), we used the definition of the counters $C_{i,t}$ and the fact that the minimum payment per round is $Kc_{\min}$, from which $\tau_{\mathcal{A}(B)} \leq B/(Kc_{\min})$ follows for any B. To obtain the lower bound on $\tau_{\mathcal{A}}(B)$, observe the following:

$$\begin{aligned}
\tau_{\mathcal{A}}(B) &= \tau_{\mathcal{A}}(B^* + B^-) \\
&\geq \tau_{\mathcal{A}^*}(B^*) \\
&\geq \tau_{\mathcal{A}^*}\left(B - \sum_{i \notin a^*} n_{i,\tau(B)} c_{\max}\right) \\
&\geq \tau_{\mathcal{A}^*}\left(B - \sum_{i=1}^{N} C_{i,\tau_{\mathcal{A}}(B)}\right) \qquad \text{(A.20)} \\
&\overset{\text{(A.7)}}{\geq} \tau_{\mathcal{A}^*}\left(B - NK\left(\gamma \log \tau(B) + \delta\right)\right) \\
&\overset{\text{(A.18)}}{\geq} \frac{B - NK\left(\gamma \log \tau_{\mathcal{A}}(B) + \delta\right)}{\sum_{i \in a^*} \mu_c^i} - 1.
\end{aligned}$$

(A.20) again uses the definition of the counters $C_{i,t}$. □

Finally, to prove Lemma 4, we need to remove the implicit relation of the bounds on $\tau_{\mathcal{A}}(B)$ presented in Lemma A16. For this purpose, we employ the inequality $\log(\phi) \leq \phi - 1$, which is valid for all $\phi > 0$. Letting $\phi = \frac{Kc_{\min}}{2N\gamma}\tau_{\mathcal{A}}(B)$, we obtain:

$$\log \tau_{\mathcal{A}}(B) \leq \frac{Kc_{\min}}{2N\gamma}\tau_{\mathcal{A}}(B) + \log\left(\frac{2N\gamma}{Kc_{\min}}\right) - 1. \qquad \text{(A.21)}$$

Substituting (A.21) into the upper bound on $\tau_{\mathcal{A}}(B)$ in Lemma A16 yields

$$\begin{aligned}\tau_{\mathcal{A}}(B) &\leq \tau_{\mathcal{A}^*}(B) + \frac{N}{Kc_{\min}}\left[\gamma\left(\frac{Kc_{\min}}{2N\gamma}\tau_{\mathcal{A}}(B) + \log\left(\frac{2N\gamma}{Kc_{\min}}\right) - 1\right) + \delta\right] \\ &\overset{(A.18)}{\leq} \frac{B}{\sum_{i\in a^*}\mu_c^i} + \frac{\tau_{\mathcal{A}}(B)}{2} + \frac{N}{Kc_{\min}}\left[\gamma\left(\log\left(\frac{2N\gamma}{Kc_{\min}}\right) - 1\right) + \delta\right] \\ &\leq \frac{2B}{\sum_{i\in a^*}\mu_c^i} + \underbrace{\frac{2N}{Kc_{\min}}\left[\gamma\left(\log\left(\frac{2N\gamma}{Kc_{\min}}\right) - 1\right) + \delta\right]}_{=:c_1}. \end{aligned} \tag{A.22}$$

Next, taking the logarithm of (A.22) and substituting into the lower bound on $\tau_{\mathcal{A}}(B)$ in Lemma A16 results in the second part of the inequality in Lemma 4, because

$$\tau_{\mathcal{A}}(B) \geq \frac{B}{\sum_{i\in a^*}\mu_c^i} - \underbrace{\left(\frac{NK\delta}{\sum_{i\in a^*}\mu_c^i} + 1\right)}_{=:c_2} - \underbrace{\frac{NK\gamma}{\sum_{i\in a^*}\mu_c^i}}_{=:c_3}\log\left(\frac{2B}{\sum_{i\in a^*}\mu_c^i} + c_1\right),$$

where we again used (A.18). This completes the proof.

A.4.1.3 Proof of Theorem 17

Proof. The constants c_1, c_2, and c_3 were defined in the previous subsection and are repeated here for convenience:

$$\begin{aligned} c_1 &= \frac{2N}{Kc_{\min}}\left[\gamma\left(\log\left(\frac{2N\gamma}{Kc_{\min}}\right) - 1\right) + \delta\right] \\ c_2 &= \left(\frac{NK\delta}{\sum_{i\in a^*}\mu_c^i} + 1\right) \\ c_3 &= \frac{NK\gamma}{\sum_{i\in a^*}\mu_c^i} \end{aligned}$$

Utilizing the definition of weak regret $\mathcal{R}_{\mathcal{A}}$ of a strategy $\mathcal{A}$ as the difference between the expected payout of the best strategy $\mathcal{A}^*$, which has knowledge of all bang-per-buck ratios, and the expected payout of $\mathcal{A}$, we obtain:

$$\begin{aligned}\mathcal{R}_{\mathcal{A}} &= \mathbb{E}[G_{\mathcal{A}^*}] - \mathbb{E}[G_{\mathcal{A}}] \\ &\overset{(A.18)}{\leq} \frac{\sum_{j\in a^*}\mu_r^j}{\sum_{j\in a^*}\mu_c^j}(B+1) - \mathbb{E}\left[\sum_{t=1}^{\tau_{\mathcal{A}}(B)}\sum_{j\in a_t}\mu_r^j\right] \\ &= \left[\frac{\sum_{j\in a^*}\mu_r^j}{\sum_{j\in a^*}\mu_c^j}(B+1) - \tau_{\mathcal{A}}(B)\sum_{j\in a^*}\mu_r^j\right] + \left[\tau_{\mathcal{A}}(B)\sum_{j\in a^*}\mu_r^j - \mathbb{E}\left[\sum_{t=1}^{\tau_{\mathcal{A}}(B)}\sum_{j\in a_t}\mu_r^j\right]\right] \\ &\overset{(5.17)}{\leq} \left[\frac{\sum_{j\in a^*}\mu_r^j}{\sum_{j\in a^*}\mu_c^j}(B+1) - \tau_{\mathcal{A}}(B)\sum_{j\in a^*}\mu_r^j\right] + \sum_{i=1}^{N} C_{i,\tau_{\mathcal{A}}(B)}\Delta_{\max}\end{aligned}$$

$$\leq \left[\frac{\sum_{j\in a^*}\mu_r^j}{\sum_{j\in a^*}\mu_c^j}(B+1) - \sum_{j\in a^*}\mu_r^j\left(\frac{B}{\sum_{j\in a^*}\mu_c^j} - c_2 - c_3\log\left(\frac{2}{\sum_{j\in a^*}\mu_c^j}B + c_1\right)\right)\right] \tag{A.23}$$

$$+ N\Delta_{\max}(\gamma\log\tau_{\mathcal{A}}(B)+\delta)$$

$$\leq \frac{\sum_{j\in a^*}\mu_r^j}{\sum_{j\in a^*}\mu_c^j} + \sum_{j\in a^*}\mu_r^j\left(c_2 + c_3\log\left(\frac{2}{\sum_{j\in a^*}\mu_c^j}B + c_1\right)\right) \tag{A.24}$$

$$+ N\Delta_{\max}\left(\gamma\log\left(\frac{2}{\sum_{j\in a^*}\mu_c^j}B + c_1\right) + \delta\right)$$

$$= O(c_3 + N\gamma\log B) = O(NK^4\log B) \tag{A.25}$$

In (A.23) and (A.24), we used the explicit bounds on $\tau_{\mathcal{A}}(B)$ on B derived in Lemma 4. Lastly, in (A.25), we used the definitions of the constants $c_3 = O(NK\gamma)$ and $\gamma = O(K^3)$. This completes the proof of Theorem 17. □

A.4.2 Proofs for Adversarial Setting

A.4.2.1 Proof of Theorem 18

Proof. Define $W_t = \sum_{i=1}^N w_i(t)$ and $\tilde{W}_t = \sum_{i=1}^N \tilde{w}_i(t)$. Then observe the following manipulations:

$$\frac{W_{t+1}}{W_t} = \sum_{i\in[N]\setminus\tilde{S}(t)}\frac{w_i(t)}{W_t}\exp\left(\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t))\right) + \sum_{i\in\tilde{S}(t)}\frac{w_i(t)}{W_t}$$

$$\leq \sum_{i\in[N]\setminus\tilde{S}(t)}\frac{w_i(t)}{W_t}\left[1 + \frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t)) + (e-2)\left(\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t))\right)^2\right] + \sum_{i\in\tilde{S}(t)}\frac{w_i(t)}{W_t}$$

$$= 1 + \frac{\tilde{W}_t}{W_t}\sum_{i\in[N]\setminus\tilde{S}(t)}\frac{w_i(t)}{\tilde{W}_t}\left[\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t)) + (e-2)\left(\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t))\right)^2\right]$$

$$\leq 1 + \sum_{i\in[N]\setminus\tilde{S}(t)}\frac{p_i(t)/K - \gamma/N}{1-\gamma}\left[\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t)) + (e-2)\left(\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t))\right)^2\right]$$

$$\leq 1 + \frac{\gamma}{(1-\gamma)N}\sum_{i\in[N]\setminus\tilde{S}(t)}p_i(t)(\hat{r}_i(t)-\hat{c}_i(t)) + \frac{(e-2)K\gamma^2}{(1-\gamma)N^2}\sum_{i\in[N]\setminus\tilde{S}(t)}p_i(t)(\hat{r}_i(t)-\hat{c}_i(t))^2$$

$$\leq 1 + \frac{\gamma}{(1-\gamma)N}\sum_{i\in[N]\setminus\tilde{S}(t)}(r_i(t)-c_i(t)) + \frac{(e-2)K\gamma^2}{(1-\gamma)N^2}(1-c_{\min})\sum_{i\in[N]}(\hat{r}_i(t)-\hat{c}_i(t)).$$

In the above manipulations, we used the update rules of the weights and probabilities $p_i(t)$ defined in Algorithm `Exp3.M.B`. Further, we utilized the property $e^x \leq 1 + x + (e-2)x^2$ for $x = K\gamma(\hat{r}_i(t)-\hat{c}_i(t))/N < 1$. In the last line, we exploit the definition of the estimated rewards $\hat{r}_i(t)$ and costs $\hat{c}_i(t)$. Next, since $e^x \geq 1 + x$ for $x \geq 0$, summing over $t = 1, \ldots, T$,

where $T = \max(\tau_{\mathcal{A}}(B), \tau_{\mathcal{A}^*}(B))$ yields

$$\log\left(\frac{W_{T+1}}{W_1}\right) \leq \frac{\gamma}{(1-\gamma)N}\sum_{t=1}^{T}\sum_{i\in a_t\setminus\tilde{S}(t)}(r_i(t)-c_i(t)) + \frac{(e-2)K\gamma^2}{(1-\gamma)N^2}(1-c_{\min})\sum_{t=1}^{T}\sum_{i\in[N]}(\hat{r}_i(t)-\hat{c}_i(t)). \tag{A.26}$$

Let a^* denote the optimal action set for algorithm $\mathcal{A}^*$. Bounding $\log(W_{T+1}/W_1)$ from above yields

$$\begin{aligned}\log\left(\frac{W_{T+1}}{W_1}\right) &\geq \log\left(\frac{\sum_{i\in a^*} w_i(T+1)}{W_1}\right)\\ &\geq \log\left(\frac{K\left(\prod_{i\in a^*} w_i(T+1)\right)^{1/k}}{N}\right)\\ &= \log\left(\frac{K}{N}\right) + \frac{1}{K}\log\left(\prod_{i\in a^*}\prod_{t\in[T]:i\notin\tilde{S}(t)}\exp\left(\frac{K\gamma}{N}(\hat{r}_i(t)-\hat{c}_i(t))\right)\right)\\ &= \log\left(\frac{K}{N}\right) + \frac{1}{K}\sum_{i\in a^*}\sum_{t\in[T]:i\notin\tilde{S}(t)}\frac{K\gamma}{N}\left(\hat{r}_i(t)-\hat{c}_i(t)\right).\end{aligned} \tag{A.27}$$

Combining (A.26) and (A.27) yields

$$\begin{aligned}&\frac{N}{\gamma}\log\left(\frac{K}{N}\right) + \sum_{i\in a^*}\sum_{t:i\notin\tilde{S}}(\hat{r}_i(t)-\hat{c}_i(t))\\ &\leq \frac{1}{1-\gamma}\sum_{t=1}^{T}\sum_{i\in a_t\setminus\tilde{S}(t)}(r_i(t)-c_i(t)) + \frac{(e-2)\gamma K(1-c_{\min})}{N(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[N]}\left(\hat{r}_i(t)-\hat{c}_i(t)\right).\end{aligned} \tag{A.28}$$

Taking the expectations of $\hat{r}_i(t)$ and $\hat{c}_i(t)$ and adding the term $\frac{1}{1-\gamma}\sum_{t=1}^{T}\sum_{i\in\tilde{S}(t)}(r_i(t)-c_i(t))$, which is bounded from below by $\sum_{i\in a^*}\sum_{t:i\in\tilde{S}(t)}(r_i(t)-c_i(t))$, to both sides of (A.28) gives us

$$\begin{aligned}&\frac{N}{\gamma}\log\left(\frac{K}{N}\right) + \sum_{i\in a^*}\sum_{t=1}^{T}(r_i(t)-c_i(t))\\ &\leq \frac{1}{1-\gamma}\sum_{t=1}^{T}\sum_{i\in a_t}(r_i(t)-c_i(t)) + \frac{(e-2)\gamma K(1-c_{\min})}{N(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[N]}\left(r_i(t)-c_i(t)\right).\end{aligned} \tag{A.29}$$

Since $T = \max(\tau_{\mathcal{A}}(B), \tau_{\mathcal{A}^*}(B))$ and due to the fact that algorithm $\mathcal{A}$ terminates after $\tau_{\mathcal{A}}(B)$ rounds, (A.29) becomes

$$\frac{N}{\gamma}\log\left(\frac{K}{N}\right) + \sum_{i\in a^*}\sum_{t=1}^{\tau_{\mathcal{A}^*}(B)}(r_i(t)-c_i(t))$$

$$\leq \frac{1}{1-\gamma}\sum_{t=1}^{\tau_{\mathcal{A}}(B)}\sum_{i\in a_t}(r_i(t)-c_i(t)) + \frac{(e-2)\gamma K(1-c_{\min})}{N(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[N]}(r_i(t)-c_i(t)) \tag{A.30}$$

We now bound the time-dependent terms in (A.30) separately:

$$\sum_{i\in a^*}\sum_{t=1}^{\tau_{\mathcal{A}^*}(B)}(r_i(t)-c_i(t)) \geq G_{\max}-B,$$
$$\sum_{t=1}^{\tau_{\mathcal{A}}(B)}\sum_{i\in a_t}(r_i(t)-c_i(t)) \leq \left(G_{\texttt{Exp3.M.B}}-(B-Kc_{\max})\right),$$

and

$$\sum_{t=1}^{T}\sum_{i\in[N]}c_i(t) \geq \sum_{t=1}^{\tau_{\mathcal{A}}(B)}\sum_{i\in[N]}c_i(t) \geq B-Kc_{\max}$$
$$\sum_{t=1}^{T}\sum_{i\in[N]}r_i(t) = \sum_{i\in[N]}\sum_{t=1}^{\tau_{\mathcal{A}^*}(B)}r_i(t) + \mathbb{1}\left(\tau_{\mathcal{A}}(B)>\tau_{\mathcal{A}^*}(B)\right)\times\sum_{i\in[N]}\sum_{t=\tau_{\mathcal{A}^*}(B)+1}^{\tau_{\mathcal{A}}(B)}r_i(t)$$
$$\leq \frac{N}{K}G_{\max}+\frac{NB(1-c_{\min})}{Kc_{\min}}. \tag{A.31}$$

In (A.31), we used the upper bound

$$\mathbb{1}\left(\tau_{\mathcal{A}}(B)>\tau_{\mathcal{A}^*}(B)\right)\cdot\sum_{i\in[N]}\sum_{t=\tau_{\mathcal{A}^*}(B)+1}^{\tau_{\mathcal{A}}(B)}r_i(t) \leq |\tau_{\mathcal{A}}(B)-\tau_{\mathcal{A}^*}(B)|Nr_{\max}$$
$$\leq \frac{B-\frac{B}{Kc_{\max}}Kc_{\min}}{Kc_{\min}}Nc_{\max} = \frac{NB(1-c_{\min})}{Kc_{\min}}.$$

With these bounds, (A.30) becomes

$$\begin{aligned}
&G_{\max}-G_{\texttt{Exp3.M.B}}\\
&\quad\leq \frac{N}{\gamma}\log\left(\frac{N}{K}\right)+\gamma G_{\max}\left(1+(e-2)\gamma(1-c_{\min})\right)+\gamma B\left(\frac{(e-2)(1-c_{\min})^2}{c_{\min}}-1\right)+K\\
&\quad\leq \frac{N}{\gamma}\log\left(\frac{N}{K}\right)+\gamma G_{\max}(e-1)+\frac{\gamma B(e-1)}{c_{\min}}+K.
\end{aligned}$$

If an upper bound g on $G_{\max}$ exists, i.e. $g \geq G_{\max}$, then γ can be tuned by choosing $\gamma = \min\left(1, \sqrt{\frac{N\log(N/K)}{g(e-1)(1+\frac{B}{gc_{\min}})}}\right)$, which gives us

$$\begin{aligned}
\mathcal{R} &= G_{\max}-G_{\texttt{Exp3.M.B}}\\
&\leq K+2\sqrt{e-1}\sqrt{1+\frac{B}{gc_{\min}}}\sqrt{gN\log(N/K)}
\end{aligned}$$

$$< K + 2.63\sqrt{1 + \frac{B}{gc_{\min}}}\sqrt{gN\log(N/K)},$$

as stated in Theorem 18. □

A.4.2.2 Proof of Proposition 3

A.4.2.2.1 Proof of Lemma 5

Proof. Using the update rule for weights (A.26) in place of the original update rule in Algorithm `Exp3.M.B`, we obtain the following inequality from (A.28) in the proof of Theorem 18:

$$\begin{aligned}\sum_{t=T_r}^{S_r}\sum_{i\in a_t}(r_i(t)-c_i(t)) \geq (1-\gamma_r)\left[\sum_{i\in a}\sum_{t=T_r}^{S_r}(\hat{r}_i(t)-\hat{c}_i(t)) + \frac{N}{\gamma_r}\log\frac{K}{N}\right] \\ - (1-\gamma_r)\frac{(e-2)\gamma_r K}{N(1-\gamma_r)}\sum_{t=T_r}^{S_r}\sum_{i\in[N]}(\hat{r}_i(t)-\hat{c}_i(t)),\end{aligned} \tag{A.32}$$

where a denotes any subset of $[N]$ of size K. According to the termination criterion of Algorithm `Exp3.1.M.B`, for each epoch r we have $\sum_{i\in a}(\hat{G}_i(T_r) - \hat{L}_i(T_r)) \leq g_r - \frac{N(1-c_{\min})}{K\gamma_r}$ for all $a \in \mathcal{S}$ and therefore

$$\sum_{i\in a}\hat{G}_i(T_r+1) \leq g_r - \frac{N(1-c_{\min})}{K\gamma_r} + \frac{N(1-c_{\min})}{K\gamma_r} = g_r.$$

Combining this equation with (A.32) yields (5.18), as stated in Lemma 5. □

A.4.2.2.2 Proof of Lemma 6

Proof. Observe that

$$\begin{aligned}\hat{G}_{\max}(T+1) - \hat{L}_{\max}(T+1) &\geq \hat{G}_{\max}(T_{R-1}+1) - \hat{L}_{\max}(T_{R-1}+1) \\ &\geq g_{R-1} + \frac{N(1-c_{\min})}{K\gamma_{R-1}} \\ &= 4^{R-1}c - 2^{R-1}\frac{N(1-c_{\min})}{K} =: cz^2 - \frac{N(1-c_{\min})}{K}z,\end{aligned} \tag{A.33}$$

where $z = 2^{R-1}$. Clearly, (A.33) is increasing for $z > N(1-c_{\min})/(2Kc)$. Now, if (5.19) were false, then $z > \frac{N(1-c_{\min})}{Kc} + \sqrt{(\hat{G}_{\max} - \hat{L}_{\max})/c} > N(1-c_{\min})/(2Kc)$ would be true, and as a consequence,

$$cz^2 - \frac{N(1-c_{\min})}{K}z > c\left(\frac{N(1-c_{\min})}{Kc} + \sqrt{\frac{\hat{G}_{\max} - \hat{L}_{\max}}{c}}\right)^2$$

$$
\begin{aligned}
&- \frac{N(1-c_{\min})}{K}\left(\frac{N(1-c_{\min})}{Kc} + \sqrt{\hat{G}_{\max}/c}\right) \\
&= \frac{N(1-c_{\min})}{K}\sqrt{\frac{\hat{G}_{\max} - \hat{L}_{\max}}{c}} + \hat{G}_{\max} - \hat{L}_{\max},
\end{aligned}
$$

which contradicts (A.33). □

To prove Proposition 3, we put together the results from Lemmas 5 and 6. Then we obtain

$$
\begin{aligned}
&\sum_{t=1}^{\tau_{\mathcal{A}}(B)} \sum_{i \in a_t} (r_i(t) - c_i(t)) \geq \\
&\qquad \max_{a \in \mathcal{S}} \left(\sum_{t=1}^{\tau_{\mathcal{A}^*}(B)} \sum_{i \in a} (\hat{r}_i(t) - \hat{c}_i(t)) - 2\sqrt{(e-1) - (e-2)c_{\min}} \sum_{r=0}^{R} \sqrt{g_r N \log(N/K)} \right),
\end{aligned}
$$

as we showed in Lemma 5 that this bounds holds for *any* subset of arms. Continuing the above equations, we observe

$$
\begin{aligned}
&\sum_{t=1}^{\tau_{\mathcal{A}}(B)} \sum_{i \in a_t} (r_i(t) - c_i(t)) \\
&\quad \geq \hat{G}_{\max} - \hat{L}_{\max} - 2N \log(N/K) \sum_{r=0}^{R} 2^r \\
&\quad \geq \hat{G}_{\max} - \hat{L}_{\max} + 2N \log(N/K) - 8N \log(N/K) \left(\frac{N(1-c_{\min})}{Kc} + \frac{\hat{G}_{\max} - \hat{L}_{\max}}{c} + \frac{1}{2} \right) \\
&\quad \geq \hat{G}_{\max} - \hat{L}_{\max} - 2N \log \frac{N}{K} - 8((e-1) - (e-2)c_{\min}) \frac{N}{K} \\
&\qquad - 8\sqrt{((e-1) - (e-2)c_{\min}) N \log \frac{N}{K} (\hat{G}_{\max} - \hat{L}_{\max})}.
\end{aligned} \tag{A.34}
$$

On the other hand, we have

$$
\sum_{t=1}^{\tau_{\mathcal{A}}(B)} \sum_{i \in a_t} (r_i(t) - c_i(t)) \leq G_{\texttt{Exp3.1.M.B}} - (B - K). \tag{A.35}
$$

Simply combining (A.34) and (A.35) yields

$$
\begin{aligned}
&G_{\texttt{Exp3.1.M.B}} \\
&\quad \geq B - K + \hat{G}_{\max} - \hat{L}_{\max} - 2N \log \frac{N}{K} - 8((e-1) - (e-2)c_{\min}) \frac{N}{K} \\
&\qquad - 8\sqrt{((e-1) - (e-2)c_{\min}) N \log \frac{N}{K} (\hat{G}_{\max} - \hat{L}_{\max})}
\end{aligned}
$$

$$=: f(\hat{G}_{\max} - \hat{L}_{\max}). \tag{A.36}$$

and it can be shown that $f(\hat{G}_{\max} - \hat{L}_{\max})$ is convex. Thus, taking the expectation of (A.36) and utilizing Jensen's inequality gives

$$\mathbb{E}[G_{\texttt{Exp3.1.M.B}}] \geq \mathbb{E}[f(\hat{G}_{\max} - \hat{L}_{\max})] \geq f(\mathbb{E}[\hat{G}_{\max} - \hat{L}_{\max}]).$$

Further, we notice

$$\begin{aligned}
\mathbb{E}[\hat{G}_{\max} - \hat{L}_{\max}] &= \mathbb{E}\left[\max_{a\in\mathcal{S}} \sum_{i\in a} \hat{G}_i - \hat{L}_i\right] \\
&\geq \max_{a\in\mathcal{S}} \mathbb{E}\left[\sum_{i\in a} \hat{G}_i - \hat{L}_i\right] \\
&= \max_{a\in\mathcal{S}} \sum_{i\in a} \sum_{t=1}^{\tau_{\mathcal{A}}(B)} (r_i(t) - c_i(t)) \\
&\geq G_{\max} - (B - K).
\end{aligned}$$

These results, together with the elementary fact $\mathbb{E}[\hat{L}_{\max}] \leq B$, yield the claim in Proposition 3.

A.4.2.3 Proof of Theorem 19

As mentioned in the main text, we use auxiliary Lemma 7.

A.4.2.3.1 Proof of Lemma 7

Proof. Let $\mathbf{r}_t$ and $\mathbf{c}_t$ denote the vector of rewards and costs observed at time t, respectively. Similarly, let $\mathbf{r}^t$ and $\mathbf{c}^t$ denote all such reward and cost vectors observed up to time t. $\mathbb{P}_u(\cdot)$ or $\mathbb{P}_{a^*}(\cdot)$ are probability measures of a random variable with respect to the uniform assignment of costs $\{c_{\min}, 1\}$ and rewards $\{0, 1\}$ to arms or conditional on a^* being the best subset of arms, respectively. With this notation, we have

$$\begin{aligned}
\mathbb{E}_{a^*}[f(\mathbf{r}, \mathbf{c})] - \mathbb{E}_u[f(\mathbf{r}, \mathbf{c})] &= \sum_{\mathbf{r},\mathbf{c}} f(\mathbf{r}, \mathbf{c})(\mathbb{P}_{a^*}(\mathbf{r}, \mathbf{c}) - \mathbb{P}_u(\mathbf{r}, \mathbf{c})) \\
&\leq \frac{B}{c_{\min}} \sum_{(\mathbf{r},\mathbf{c}):\mathbb{P}_{a^*}(\mathbf{r},\mathbf{c}) \geq \mathbb{P}_u(\mathbf{r},\mathbf{c})} (\mathbb{P}_{a^*}(\mathbf{r}, \mathbf{c}) - \mathbb{P}_u(\mathbf{r}, \mathbf{c})) \\
&\leq \frac{B}{2c_{\min}} \|\mathbb{P}_{a^*} - \mathbb{P}_u\|_1,
\end{aligned} \tag{A.37}$$

where $\|\mathbb{P}_{a^*} - \mathbb{P}_u\|_1 = \sum_{(\mathbf{r},\mathbf{c})} |\mathbb{P}_{a^*}(\mathbf{r}, \mathbf{c}) - \mathbb{P}_u(\mathbf{r}, \mathbf{c})|$. Letting $\text{Bern}(p)$ denote a Bernoulli distribution with parameter p, we obtain, using Pinsker's Inequality

$$\|\mathbb{P}_{a^*} - \mathbb{P}_u\|_1^2 \leq 2 \log 2 \cdot \text{KL}(\mathbb{P}_u \parallel \mathbb{P}_{a^*}), \tag{A.38}$$

the following result:

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_u \parallel \mathbb{P}_{a^*}) &= \sum_{t=1}^{\left\lfloor \frac{B}{Kc_{\min}} \right\rfloor} \mathbb{1}\left(\sum_{\tau=1}^{t-1} \mathbf{1}\cdot\mathbf{c}_\tau \le B - c_{\min}\right) \mathbb{1}\left(\sum_{\tau=1}^{t} \mathbf{1}\cdot\mathbf{c}_\tau \le B\right) \\
&\quad \times \mathrm{KL}\left(\mathbb{P}_u\left(\mathbf{c}_t, \mathbf{r}_t \mid \mathbf{c}^{t-1}, \mathbf{r}^{t-1}\right) \parallel \mathbb{P}_{a^*}\left(\mathbf{c}_t, \mathbf{r}_t \mid \mathbf{c}^{t-1}, \mathbf{r}^{t-1}\right)\right) \\
&\le \sum_{t=1}^{\left\lfloor \frac{B}{Kc_{\min}} \right\rfloor} \mathbb{P}_u(a_t \neq a^*)\mathrm{KL}\left(\mathrm{Bern}(1/2) \parallel \mathrm{Bern}(1/2)\right) \\
&\qquad + \mathbb{P}_u(a_t = a^*)\mathrm{KL}\left(\mathrm{Bern}(1/2) \parallel \mathrm{Bern}(\varepsilon + 1/2)\right) \\
&= \sum_{t=1}^{\left\lfloor \frac{B}{Kc_{\min}} \right\rfloor} \mathbb{P}_u(a_t = a^*)\left(-\frac{1}{2}\log_2(1-4\varepsilon^2)\right) \\
&= \frac{1+c_{\min}}{2c_{\min}}\mathbb{E}_u[N_{a^*}]\left(-\frac{1}{2}\log_2(1-4\varepsilon^2)\right) \qquad \text{(A.39)} \\
&\le \frac{1}{c_{\min}}\mathbb{E}_u[N_{a^*}]\left(-\frac{1}{2}\log_2(1-4\varepsilon^2)\right), \qquad \text{(A.40)}
\end{aligned}$$

where in (A.40) we used $c_{\min} \le 1$. (A.39) uses the expected stopping time under uniform assignment $\mathbb{E}_u[\tau(B)] = \lfloor 2BK^{-1}/(c_{\min}+1)\rfloor$ to obtain

$$\begin{aligned}
\sum_{t=1}^{\left\lfloor \frac{B}{Kc_{\min}} \right\rfloor} \mathbb{P}_u(a_t = a^*) &= \sum_{t=1}^{\left\lfloor \frac{2B}{K(1+c_{\min})} \right\rfloor} \mathbb{P}_u(a_t = a^*) + \sum_{t=\left\lceil \frac{2B}{K(1+c_{\min})} \right\rceil}^{\left\lfloor \frac{B}{Kc_{\min}} \right\rfloor} \mathbb{P}_u(a_t = a^*) \\
&= \frac{B/(Kc_{\min})}{2BK^{-1}/(1+c_{\min})}\mathbb{E}_u(N_{a^*}) \le \frac{1}{c_{\min}}\mathbb{E}_u(N_{a^*}).
\end{aligned}$$

Substituting (A.38) and (A.40) into (A.37) and utilizing $\log_2(x) = \log x/\log 2$ for $x > 0$ yields the statement in the lemma. □

To finalize the proof of Theorem 19, notice that there exist $\binom{N}{K}$ possible combinations of arms of size K. Borrowing notation from Uchiya, Nakamura, and Kudo 2010, let $\mathbf{C}([N], K)$ denote the set of all such subsets. Now, let $\mathbb{E}_*[\cdot]$ denote the expected value with respect to the uniform assignment of "good" arms. With this notation, observe

$$\begin{aligned}
\mathbb{E}_*[G_{\max}] &= \left(\frac{1}{2}+\varepsilon\right)K\mathbb{E}_*[\tau_{\mathcal{A}}(B)] \\
\mathbb{E}_{a^*}[G_{\mathcal{A}}] &= \frac{1}{2}K\mathbb{E}_{a^*}[\tau_{\mathcal{A}}(B)] + \varepsilon\mathbb{E}_{a^*}[N_{a^*}] \\
\mathbb{E}_*[G_{\mathcal{A}}] &= \frac{1}{\binom{N}{K}}\sum_{a^*\in\mathbf{C}([N],K)}\mathbb{E}_{a^*}[G_{\mathcal{A}}] = \frac{1}{2}K\mathbb{E}_*[\tau_{\mathcal{A}}(B)] + \frac{\varepsilon}{\binom{N}{K}}\sum_{a^*\in\mathbf{C}([N],K)}\mathbb{E}_{a^*}[N_{a^*}].
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}_*[G_{\max} - G_{\mathcal{A}}] \\
&\geq \varepsilon K \mathbb{E}_*[\tau_{\mathcal{A}}(B)] - \frac{\varepsilon}{\binom{N}{K}} \sum_{a^* \in \mathbf{C}([N],K)} \left(\mathbb{E}_u[N_{a^*}] + \frac{B}{2c_{\min}^{3/2}} \sqrt{-\mathbb{E}_u[N_{a^*}] \log(1-4\varepsilon^2)} \right) \\
&\geq \varepsilon K \mathbb{E}_u[\tau_{\mathcal{A}}(B)] - \frac{\varepsilon}{\binom{N}{K}} \binom{N}{K} \mathbb{E}_u[\tau_{\mathcal{A}}(B)] \frac{K}{N} K - \frac{\varepsilon}{\binom{N}{K}} \sum_{a^* \in \mathbf{C}([N],K)} \frac{B}{2c_{\min}^{3/2}} \sqrt{-\mathbb{E}_u[N_{a^*}] \log(1-4\varepsilon^2)} \\
&= \varepsilon K \left(1 - \frac{K}{N}\right) \mathbb{E}_u[\tau_{\mathcal{A}}(B)] - \frac{\varepsilon B c_{\min}^{-3/2}}{2\binom{N}{K}} \sqrt{-\binom{N}{K}\binom{N-1}{K-1} \mathbb{E}_u[N_{a^*}] \log(1-4\varepsilon^2)} \qquad \text{(A.41)} \\
&\geq \varepsilon B \left(1 - \frac{K}{N}\right) - \frac{2\varepsilon B}{c_{\min}^{3/2}} \sqrt{\frac{BK}{N} \log(4/3)}. \qquad \text{(A.42)}
\end{aligned}$$

In (A.41), we used Jensen's inequality and the fact that

$$\sum_{a^* \in \mathbf{C}([N],K)} \mathbb{E}_u[N_{a^*}] = \binom{N}{K} \mathbb{E}_u[\tau_{\mathcal{A}}(B)] \frac{K}{N} K.$$

In (A.42), we utilized $B/K \leq \mathbb{E}_u[\tau_{\mathcal{A}}(B)] \leq B/(2K)$ and $-\log(1-4\varepsilon^2) \leq 16\log(4/3)\varepsilon^2$. Finally, to prove (5.12), we tune ε as follows:

$$\varepsilon = \min\left(\frac{1}{4},\ \frac{c_{\min}^{3/2}}{4\log(4/3)} (1 - K/N) \sqrt{\frac{N}{BK}} \right). \qquad \text{(A.43)}$$

Plugging (A.43) back into (A.42) completes the proof.

A.4.2.4 Proof of Theorem 20

A.4.2.4.1 Proof of Lemma 8

Proof. Since

$$\mathbb{P}\left(\bigcap_{a \in \mathcal{S}} \sum_{i \in a} \hat{G}_i + \alpha \hat{\sigma}_i > \sum_{i \in a} G_i \right) \geq \mathbb{P}\left(\bigcap_{i \in [N]} \hat{G}_i + \alpha \hat{\sigma}_i > G_i \right) = 1 - \mathbb{P}\left(\bigcup_{i \in [N]} \hat{G}_i + \alpha \hat{\sigma}_i \leq G_i \right),$$

it suffices to show that (using the union bound)

$$\mathbb{P}\left(\bigcup_{i \in [N]} \hat{G}_i + \alpha \hat{\sigma}_i \leq G_i \right) < \sum_{i=1}^{N} \mathbb{P}\left(\hat{G}_i + \alpha \hat{\sigma}_i \leq G_i \right) < \delta. \qquad \text{(A.44)}$$

To show this, choose an arbitrary $i \in [N]$ and define

$$\hat{\sigma}(t+1) = K\sqrt{NT} + \sum_{\tau=1}^{t} \frac{1}{p_i(\tau)\sqrt{NT}}, \qquad \text{(A.45)}$$

$$s_t = \frac{\alpha K}{2\hat{\sigma}(t+1)} \leq 1. \tag{A.46}$$

Using the shorthand notation $\hat{\sigma}_i := \hat{\sigma}_i(T+1)$, observe

$$\begin{aligned}
\mathbb{P}\left(\hat{G}_i + \alpha\hat{\sigma}_i \leq G_i\right) &= \mathbb{P}\left(\sum_{t=1}^{T}\left(r_i(t) - \hat{r}_i(t) - \alpha\hat{\sigma}_i/2\right) \geq \alpha\hat{\sigma}_i/2\right) \\
&\leq \mathbb{P}\left(s_T\sum_{t=1}^{T}\left(r_i(t) - \hat{r}_i(t) - \frac{\alpha}{2p_i(t)\sqrt{NT}}\right) \geq \frac{\alpha^2 K}{4}\right) \\
&= \mathbb{P}\left(\exp\left[s_T\sum_{t=1}^{T}\left(r_i(t) - \hat{r}_i(t) - \frac{\alpha}{2p_i(t)\sqrt{NT}}\right)\right] \geq \exp\left(\frac{\alpha^2 K}{4}\right)\right) \\
&= \exp\left(-\frac{\alpha^2 K}{4}\right)\mathbb{E}\left[s_T\sum_{t=1}^{T}\left(r_i(t) - \hat{r}_i(t) - \frac{\alpha}{2p_i(t)\sqrt{NT}}\right)\right].
\end{aligned}$$

As in Lemma 6.1 from Auer et al. 2002, define

$$Z_t = \exp\left(s_t\sum_{\tau=1}^{t}\left(r_i(\tau) - \hat{r}_i(\tau) - \frac{\alpha}{2p_i(\tau)\sqrt{NT}}\right)\right),$$

from which it follows for $t = 2, \ldots, T$ that

$$Z_t = \exp\left(s_t\left(r_i(t) - \hat{r}_i(t) - \frac{\alpha}{2p_i(t)\sqrt{NT}}\right)\right) Z_{t-1}^{\frac{s_t}{s_{t-1}}}.$$

Since

$$\frac{\alpha}{2p_i(t)\sqrt{NT}} \geq \frac{\alpha K}{2p_i(t)\hat{\sigma}_i(t+1)} = \frac{s_t}{p_i(t)},$$

we obtain for $t = 2, \ldots, T$:

$$\begin{aligned}
\mathbb{E}_{\hat{r}_i(t)}[Z_t] &\leq \mathbb{E}_{\hat{r}_i(t)}\left[\exp\left[s_t\left(r_i(t) - \hat{r}_i(t) - \frac{s_t}{p_i(t)}\right)\right]\right] Z_{t-1}^{\frac{s_t}{s_{t-1}}} \\
&\leq \mathbb{E}_{\hat{r}_i(t)}\left[1 + s_t\left(r_i(t) - \hat{r}_i(t)\right) + s_t^2\left(r_i(t) - \hat{r}_i(t)\right)^2\right]\exp\left(-\frac{s_t^2}{p_i(t)}\right) Z_{t-1}^{\frac{s_t}{s_{t-1}}} \\
&\leq \left(1 + \frac{s_t^2}{p_i(t)}\right)\exp\left(-\frac{s_t^2}{p_i(t)}\right) Z_{t-1}^{\frac{s_t}{s_{t-1}}} \\
&\leq Z_{t-1}^{\frac{s_t}{s_{t-1}}} \leq 1 + Z_{t-1}.
\end{aligned}$$

Since $\mathbb{E}_{\hat{r}_i(1)}[Z_1] \leq 1$, it follows that $\mathbb{E}_{\hat{r}_i(T)}[Z_T] < T$. Hence, (A.44) writes

$$\sum_{i=1}^{N}\mathbb{P}\left(\hat{G}_i + \alpha\hat{\sigma}_i \leq G_i\right) \leq \sum_{i=1}^{N}\exp\left(-K\frac{N-K}{N-1}\log\left(\frac{NT}{\delta}\right)\right)T$$

$$= NT\left(\frac{\delta}{NT}\right)^{\frac{K(N-K)}{N-1}} \leq NT\frac{\delta}{NT} = \delta. \tag{A.47}$$

In (A.47), we used the fact that the minima of $K(N-K)/(N-1)$ for $1 \leq K < N$ are attained at $K = 1$ and $K = N-1$ and have value 1. Since $\delta/(NT) < 1$, the claim follows. □

A.4.2.4.2 Proof of Lemma 9

Proof. From the definition of the weights in Algorithm `Exp3.P.M`, observe:

$$\begin{aligned}
\frac{W_{t+1}}{W_t} &= \sum_{i\in[N]\setminus\tilde{S}(t)} \frac{w_i(t)}{W_t}\exp\left(\eta\hat{r}_i(t) + \frac{\alpha\eta}{p_i(t)\sqrt{NT}}\right) + \sum_{i\in\tilde{S}(t)}\frac{w_i(t)}{W_t} \\
&\leq \sum_{i\in[N]\setminus\tilde{S}(t)} \frac{w_i(t)}{W_t}\left[1 + \eta\hat{r}_i(t) + \frac{\alpha\eta}{p_i(t)\sqrt{NT}} + 2\eta^2\hat{r}_i(t)^2 + \frac{2\alpha^2\eta^2}{p_i(t)^2 NT}\right] + \sum_{i\in\tilde{S}(t)}\frac{w_i(t)}{W_t} \qquad \text{(A.48)} \\
&= 1 + \frac{W_t'}{W_t}\sum_{i\in[N]\setminus\tilde{S}(t)} \frac{\frac{p_i(t)}{K} - \frac{\gamma}{k}}{1-\gamma}\left[\eta\hat{r}_i(t) + \frac{\alpha\eta}{p_i(t)\sqrt{NT}} + 2\eta^2\hat{r}_i(t)^2 + \frac{2\alpha^2\eta^2}{p_i(t)^2 NT}\right] \qquad \text{(A.49)} \\
&\leq 1 + \frac{\eta}{K(1-\gamma)}\sum_{i\in[N]\setminus\tilde{S}(t)} p_i(t)\hat{r}_i(t) + \frac{\alpha\eta}{K(1-\gamma)}\sqrt{\frac{N}{T}} \\
&\quad + \frac{2\eta^2}{K(1-\gamma)}\sum_{i\in[N]} p_i(t)\hat{r}_i(t)^2 + \frac{2\alpha^2\eta^2}{NTK(1-\gamma)}\sum_{i\in[N]}\frac{1}{p_i(t)} \\
&= 1 + \frac{\eta}{K(1-\gamma)}\sum_{i\in a_t} r_i(t) + \frac{\alpha\eta}{K(1-\gamma)}\sqrt{\frac{N}{T}} + \frac{2\eta^2}{K(1-\gamma)}\sum_{i\in[N]}\hat{r}_i(t) + \frac{2\alpha^2\eta}{K(1-\gamma)}\frac{1}{T}, \qquad \text{(A.50)}
\end{aligned}$$

where we used the properties

$$\begin{aligned}
\hat{r}_i(t) &\leq \frac{1}{p_i(t)} \leq \frac{N}{\gamma K}, \\
\sum_{i\in[N]} p_i(t)\hat{r}_i(t) &= \sum_{i\in[N]} r_i(t), \\
\sum_{i\in[N]} p_i(t)\hat{r}_i(t)^2 &\leq \sum_{i\in[N]}\hat{r}_i(t)
\end{aligned}$$

in (A.50) and the inequality $e^x \leq 1 + x + x^2$ valid for $x \leq 1$ in (A.48). Summing over $t = 1, \ldots, T$ and utilizing the telescoping property of the logarithm yields

$$\log\left(\frac{W_{T+1}}{W_1}\right) \leq \frac{\eta}{K(1-\gamma)}G_{\texttt{Exp3.P.M}} + \frac{2\eta^2}{K(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in[N]}\hat{x}_i(t) + \frac{\alpha\eta\sqrt{NT}}{K(1-\gamma)} + \frac{2\alpha^2\eta}{K(1-\gamma)}$$

$$\leq \frac{\eta}{K(1-\gamma)} G_{\texttt{Exp3.P.M}} + \frac{2\eta^2}{K(1-\gamma)} \frac{N}{K} \hat{U}^* + \frac{\alpha\eta\sqrt{NT}}{K(1-\gamma)} + \frac{2\alpha^2\eta}{K(1-\gamma)}. \tag{A.51}$$

On the other hand, we have

$$\log(W_1) = \log\left[N \exp\left(\frac{\alpha\gamma K^2}{3}\sqrt{\frac{T}{N}}\right)\right] = \log(N) + \alpha K\eta\sqrt{NT}, \tag{A.52}$$

$$\begin{aligned}
\log(W_{T+1}) &\geq \log\left(\sum_{i\in a^*} w_j(T+1)\right) \\
&\geq \log\left[K\left(\prod_{i\in a^*} w_j(T+1)\right)^{1/K}\right] \\
&= \log(K) + \frac{1}{K}\sum_{i\in a^*} \log(w_i(T+1)) \\
&= \log(K) + \frac{1}{K}\sum_{i\in a^*}\left[\frac{\alpha\gamma K^2}{3}\sqrt{\frac{T}{N}} + \sum_{t=1}^{T}\left(\eta\hat{x}_i(t) + \frac{\alpha\eta}{p_i(t)\sqrt{NT}}\right)\right] \\
&= \log(K) + \frac{1}{K}\sum_{i\in a^*}\left(\eta\hat{G}_i + \alpha\eta\hat{\sigma}_i\right),
\end{aligned} \tag{A.53}$$

where (A.52) and (A.53) follow from the definitions of weights in Algorithm `Exp3.P.M` and (A.45), respectively. □

Finally, to show the claim in Theorem 20, simply combine the results from Lemma 8 and Lemma 9. Combining (A.51), (A.52), and (A.53) yields

$$G_{\texttt{Exp3.P.M}} \geq \left(1 - \frac{5\gamma}{3}\right)\hat{U}^* - 2\alpha^2 - \alpha(1+K^2)\sqrt{NT} - \frac{3N}{\gamma}\log(N/K).$$

From Lemma (8), it follows that $\hat{U}^* > G_{\max}$ with probability at least $1-\delta$. Together with the simple fact $G_{\max} \leq KT$, we have that

$$\mathcal{R} = G_{\max} - G_{\texttt{Exp3.P.M}} \leq \frac{5}{3}\gamma KT + 2\alpha^2 + \alpha(1+K^2)\sqrt{NT} + \frac{3N}{\gamma}\log\left(\frac{N}{K}\right).$$

Choosing

$$\gamma = \min\left(\frac{3}{5}, \frac{3}{\sqrt{5}}\sqrt{\frac{N\log(N/K)}{KT}}\right), \tag{A.54}$$

$$\alpha = 2\sqrt{\frac{N-K}{N-1}\log\left(\frac{NT}{\delta}\right)} \tag{A.55}$$

yields (5.14), which is the bound in Theorem 20. If either $T \geq \frac{N\log(N/K)}{5K}$ (to make $\gamma \leq 3/5$ in (A.54)) or $\delta \geq NT\exp\left(-\frac{NT(N-1)}{N-K}\right)$ (to make $\alpha < 2\sqrt{NT}$ in (A.55)) is not fulfilled, then the bound holds trivially.

A.4.2.5 Proof of Theorem 21

A.4.2.5.1 Proof of Lemma 10

Proof. As in the proof for Lemma 8, it suffices to show that

$$\mathbb{P}\left(\bigcup_{i\in[N]} \hat{G}_i - \hat{L}_i + \alpha\hat{\sigma}_i \leq G_i - L_i\right) < \sum_{i=1}^{N} \mathbb{P}\left(\hat{G}_i - \hat{L}_i + \alpha\hat{\sigma}_i \leq G_i - L_i\right) < \delta. \tag{A.56}$$

Let $\hat{\sigma}(t+1)$ and s_t be defined as

$$\hat{\sigma}(t+1) = K\sqrt{\frac{NB}{Kc_{\min}}} + \sum_{\tau=1}^{t} \frac{\sqrt{Kc_{\min}}}{p_i(\tau)\sqrt{NB}}, \tag{A.57}$$

$$s_t = \frac{\alpha K}{12\hat{\sigma}(t+1)} \leq 1. \tag{A.58}$$

Now observe

$$\begin{aligned}
&\mathbb{P}\left(\hat{G}_i - \hat{L}_i + \alpha\hat{\sigma}_i \leq G_i - L_i\right) \\
&= \mathbb{P}\left(\sum_{t=1}^{\tau_a(B)} (r_i(t) - c_i(t) - \hat{r}_i(t) + \hat{c}_i(t) - \alpha\hat{\sigma}_i/2) \geq \alpha\hat{\sigma}_i/2\right) \\
&\leq \mathbb{P}\left(s_{\tau_a(B)} \sum_{t=1}^{\tau_a(B)} \left(r_i(t) - c_i(t) - \hat{r}_i(t) + \hat{c}_i(t) - \frac{\alpha\sqrt{Kc_{\min}}}{2p_i(t)\sqrt{NB}}\right) \geq \frac{\alpha^2 K}{24}\right) \\
&= \exp\left(-\frac{\alpha^2 K}{24}\right) \mathbb{E}\left[s_{\tau_a(B)} \sum_{t=1}^{\tau_a(B)} \left(r_i(t) - c_i(t) - \hat{r}_i(t) + \hat{c}_i(t) - \frac{\alpha\sqrt{Kc_{\min}}}{2p_i(t)\sqrt{NB}}\right)\right].
\end{aligned}$$

Now define Z_t as follows:

$$Z_t = \exp\left(s_t \sum_{\tau=1}^{t} \left(r_i(\tau) - c_i(\tau) - \hat{r}_i(\tau) + \hat{c}_i(\tau) - \frac{\alpha\sqrt{Kc_{\min}}}{2p_i(\tau)\sqrt{NB}}\right)\right)$$

from which it follows that

$$Z_t = \exp\left(s_t \left(r_i(t) - c_i(t) - \hat{r}_i(t) + \hat{c}_i(t) - \frac{\alpha\sqrt{Kc_{\min}}}{2p_i(t)\sqrt{NB}}\right)\right) Z_{t-1}^{\frac{s_t}{s_{t-1}}},\ t = 2, \ldots, \tau_S(B).$$

Since

$$\frac{\alpha\sqrt{Kc_{\min}}}{2p_i(t)\sqrt{NB}} \geq \frac{4\alpha K}{8p_i(t)\hat{\sigma}_i(t+1)} = \frac{4s_t}{p_i(t)},$$

we obtain for $t = 2, \ldots, \tau_S(B)$:

$$\mathbb{E}_t[Z_t] \leq \mathbb{E}_t\left[\exp\left[s_t\left(r_i(t) - \hat{r}_i(t) - \frac{4s_t}{p_i(t)}\right)\right]\right] Z_{t-1}^{\frac{s_t}{s_{t-1}}}$$

$$\leq \mathbb{E}_t\left[1 + s_t\left(r_i(t) - c_i(t) - \hat{r}_i(t) + \hat{c}_i(t)\right) + s_t^2\left(r_i(t) - c_i(t) - \hat{r}_i(t) + \hat{c}_i(t)\right)^2\right]$$
$$\times \exp\left(-\frac{4s_t^2}{p_i(t)}\right) Z_{t-1}^{\frac{s_t}{s_{t-1}}} \tag{A.59}$$
$$\leq \left(1 + \frac{4s_t^2}{p_i(t)}\right)\exp\left(-\frac{4s_t^2}{p_i(t)}\right) Z_{t-1}^{\frac{s_t}{s_{t-1}}}$$
$$\leq Z_{t-1}^{\frac{s_t}{s_{t-1}}} \leq 1 + Z_{t-1} \tag{A.60}$$

In (A.59), we used the following operation:

$$\begin{aligned}
&\mathbb{E}_t\left[\left(\left(r_i(t) - \hat{r}_i(t)\right) - \left(c_i(t) - \hat{c}_i(t)\right)\right)^2\right] \\
&= \mathbb{E}_t\left[\left(r_i(t) - \hat{r}_i(t)\right)^2\right] + \mathbb{E}_t\left[\left(c_i(t) - \hat{c}_i(t)\right)^2\right] - 2\mathbb{E}_t\left[\left(r_i(t) - \hat{r}_i(t)\right)\left(c_i(t) - \hat{c}_i(t)\right)\right] \\
&\leq \mathbb{E}_t[\hat{r}_i(t)^2] + \mathbb{E}_t[\hat{c}_i(t)^2] - 2\mathbb{E}_t[r_i(t)c_i(t) - r_i(t)\hat{c}_i(t) - c_i(t)\hat{r}_i(t) + \hat{r}_i(t)\hat{c}_i(t)] \\
&\leq \frac{2}{p_i(t)} - 2\left[r_i(t)c_i(t) - 2r_i(t)c_i(t) + \frac{r_i(t)c_i(t)}{p_i(t)}\right] \\
&\leq \frac{2}{p_i(t)} + 2\frac{1}{p_i(t)} = \frac{4}{p_i(t)}.
\end{aligned}$$

Since $\mathbb{E}_t[Z_1] \leq 1$, it follows that $\mathbb{E}_{\tau_a(B)}[Z_{\tau_a(B)}] < \tau_S(B)$. Hence, (A.56) writes

$$\begin{aligned}
&\sum_{i=1}^{N} \mathbb{P}\left(\hat{G}_i - \hat{L}_i + \alpha\hat{\sigma}_i \leq G_i - L_i\right) \\
&\quad \leq \sum_{i=1}^{N} \exp\left(-K\frac{N-K}{N-1}\log\left(\frac{NB}{Kc_{\min}\delta}\right)\right)\tau_a(B) \\
&\quad = N\tau_a(B)\left(\frac{Kc_{\min}\delta}{NB}\right)^{\frac{K(N-K)}{N-1}} \leq N\tau_a(B)\frac{Kc_{\min}\delta}{NB} \leq \delta
\end{aligned} \tag{A.61}$$

because $\frac{\tau_a(B)}{B/(Kc_{\min})} \leq 1$. This completes the proof. □

A.4.2.5.2 Proof of Lemma 11

Proof. Using the weight update rule for Algorithm `Exp3.P.M.B`, we obtain (using the same manipulations as in the proof for Lemma 9)

$$\begin{aligned}
\frac{W_{t+1}}{W_t} = 1 &+ \frac{\eta}{K(1-\gamma)}\sum_{t=1}^{T}\sum_{i\in a_t}\left(r_i(t) - c_i(t)\right) \\
&+ \frac{\alpha\eta T}{K(1-\gamma)}\sqrt{\frac{NKc_{\min}}{B}} + \frac{2\alpha^2\eta K c_{\min} T}{BK(1-\gamma)}
\end{aligned}$$

$$+ \frac{2\eta^2(1-c_{\min})}{K(1-\gamma)} \sum_{i\in[N]} \sum_{t=1}^{T} (\hat{r}_i(t) - \hat{c}_i(t)), \tag{A.62}$$

where $T = \max\left(\tau_{a^*}(B), \tau_a(B)\right)$. On the other hand, observe that

$$\log W_1 = \log N + \alpha K \eta \sqrt{\frac{BN}{Kc_{\min}}} \tag{A.63}$$

and

$$\begin{aligned}
\log W_{T+1} &\geq \log K + \frac{1}{K} \sum_{i\in a^*} \log w_j(T+1) && \text{(A.64)}\\
&\geq \log K + \frac{1}{K} \sum_{i\in a^*} \left(\alpha\eta K \sqrt{\frac{BN}{Kc_{\min}}} + \sum_{t=1}^{T} \left(\eta(\hat{r}_i(t) - \hat{c}_i(t)) + \frac{\alpha\eta\sqrt{Kc_{\min}}}{p_i(t)\sqrt{NB}} \right) \right) \\
&= \log K + \frac{1}{K} \sum_{i\in a^*} \left(\alpha\eta\hat{\sigma}_i(T+1) + \eta(\hat{G}_i(T+1) - \hat{L}_i(T+1)) \right) \\
&\geq \log K + \frac{1}{K} \sum_{i\in a^*} \left(\alpha\eta\hat{\sigma}_i(\tau_{a^*}(B)+1) + \eta(\hat{G}_i(\tau_{a^*}(B)+1) - \hat{L}_i(\tau_{a^*}(B)+1)) \right) \\
&= \log K + \frac{\eta}{K}\hat{U}^*. && \text{(A.65)}
\end{aligned}$$

Using the identity $e^x > 1 + x$, the telescoping property of the logarithm in equations (A.62), and (A.63) and (A.65) yield

$$\begin{aligned}
&\log \frac{K}{N} + \frac{\eta}{K}\hat{U}^* - \alpha K \eta \sqrt{\frac{BN}{Kc_{\min}}} \\
&\quad \leq \frac{\eta}{K(1-\gamma)} \sum_{t=1}^{T} \sum_{i\in a_t} (r_i(t) - c_i(t)) \\
&\qquad + \frac{\alpha\eta T}{K(1-\gamma)} \sqrt{\frac{NKc_{\min}}{B}} + \frac{2\alpha^2\eta K c_{\min} T}{BK(1-\gamma)} + \frac{2\eta^2(1-c_{\min})}{K(1-\gamma)} \sum_{i\in[N]} \sum_{t=1}^{T} (\hat{r}_i(t) - \hat{c}_i(t)). \qquad \text{(A.66)}
\end{aligned}$$

From Lemma 10, we have that $\hat{U}^* > G_{\max} - B$ with probability at least $1 - \delta$. Now, manipulating the right hand side of (A.66) and noticing that algorithm `Exp3.P.M.B` terminates after $\tau_{\mathcal{A}}(B)$ rounds yields

$$\begin{aligned}
\text{RHS} &\leq \frac{\eta}{K(1-\gamma)} \left(G_{\texttt{Exp3.P.M.B}} - (B - Kc_{\max}) \right) \\
&\qquad + \frac{\alpha\eta}{K(1-\gamma)} \sqrt{\frac{NB}{Kc_{\min}}} + \frac{2\alpha^2\eta}{K(1-\gamma)} + \frac{2\eta^2(1-c_{\min})}{K(1-\gamma)} \sum_{t=1}^{B/(Kc_{\min})} (\hat{r}_i(t) - \hat{c}_i(t)), \qquad \text{(A.67)}
\end{aligned}$$

where we used the fact that $T = \max\left(\tau_{a^*}(B), \tau_a(B)\right) \leq B/(Kc_{\min})$. Finally, putting LHS and RHS together and utilizing $\sum_{t=1}^{B/(Kc_{\min})}(\hat{r}_i(t) - \hat{c}_i(t)) \leq (N/K)\hat{U}^*$ gives

$$\begin{aligned}
&\frac{K(1-\gamma)}{\eta}\log\frac{K}{N} + (1-\gamma)\hat{U}^* - \alpha K^2(1-\gamma)\sqrt{\frac{BN}{Kc_{\min}}} \\
&\quad \leq G_{\texttt{Exp3.P.M.B}} - B + K + \alpha\sqrt{\frac{BN}{Kc_{\min}}} + 2\eta(1-c_{\min})\frac{N}{K}\hat{U}^* + 2\alpha^2,
\end{aligned}$$

from which (5.20) follows. □

Finally, putting both Lemmas together, we get from Lemma 10 that $\hat{U}^* > G_{\max}$ with probability at least $1-\delta$. Also, note that $G_{\max} = B/c_{\min}$. Combining with Lemma 11 and choosing

$$\begin{aligned}
\gamma = \min\Bigg(&\left(1 + \frac{2}{3}\frac{1-c_{\min}}{c_{\min}}\right)^{-1}, \qquad\qquad (A.68)\\
&\left(\frac{3N\log(N/K)}{(G_{\max}-B)\left(1+2(1-c_{\min})/(3c_{\min})\right)},\right)^{1/2}\Bigg), \\
\alpha = 2\sqrt{6}&\sqrt{\frac{N-K}{N-1}\log\left(\frac{NB}{Kc_{\min}\delta}\right)} \qquad\qquad (A.69)
\end{aligned}$$

yields the desired bound (5.16). If either $B \geq 3N\log(N/K)\left(1 + 2/3 + c_{\min}/(1-c_{\min})\right)$ (to make $\gamma \leq 3/5$ in (A.68)) or $\delta \geq NB/(Kc_{\min})\exp\left(-\frac{6(N-1)NB}{(N-K)Kc_{\min}}\right)$ (in order to make $\alpha < 12\sqrt{NT/(Kc_{\min})}$ in (A.69)) is not fulfilled, then the bound holds trivially.

Appendix B

Supplementary Material

B.1 Supplementary Material for Chapter 8

B.1.1 Summary Statistics

Figures B.1-B.4 illustrate the distribution of the number of DR events received among users with *completed* Phase 1, as well as the total number of DR events broken out by hour of the day, day of the week, and month of the year.

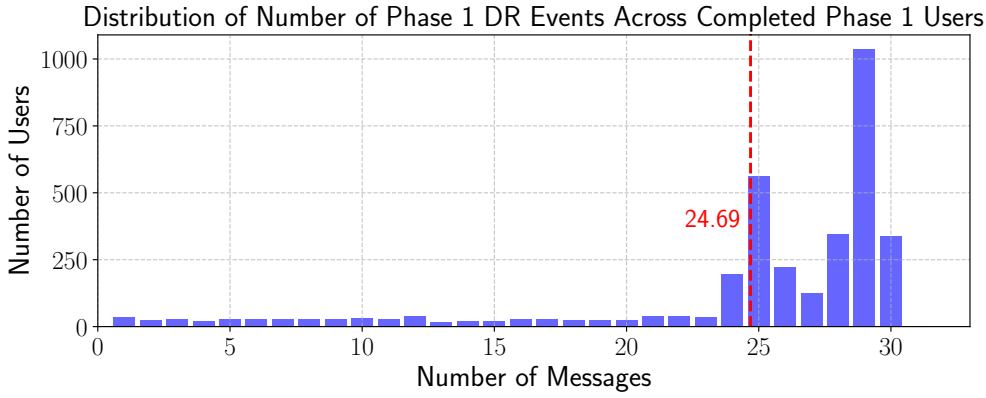

Figure B.1: Distribution of Number of Phase 1 DR Events Across Users with Completed Phase 1

B.1.2 Balance Checks

Table B.1 provides the balance metrics introduced in Section 8.3.4.

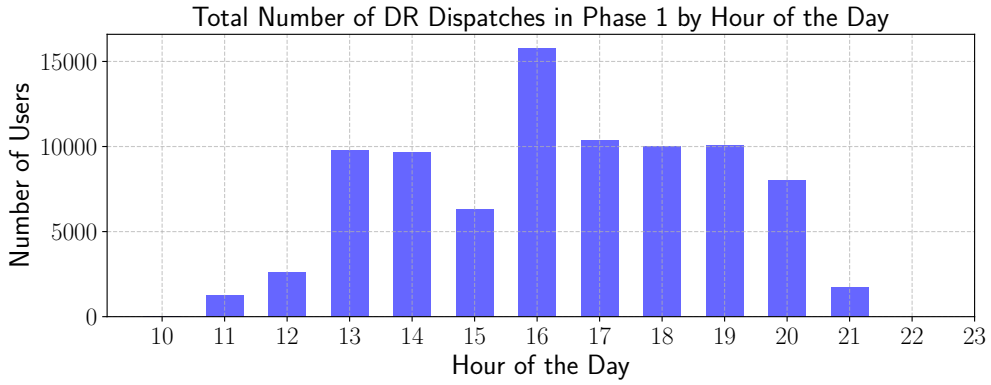

Figure B.2: Distribution of DR Events by Hour of the Day Across Users with Completed Phase 1

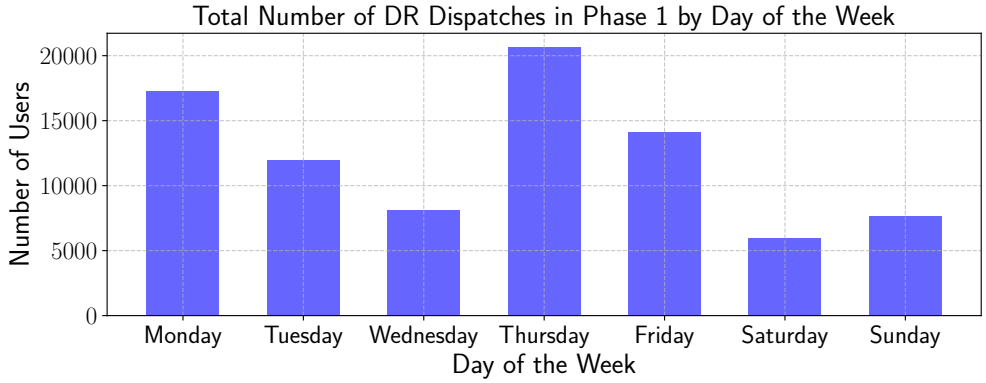

Figure B.3: Distribution of DR Events by Day of the Week Across Users with Completed Phase 1

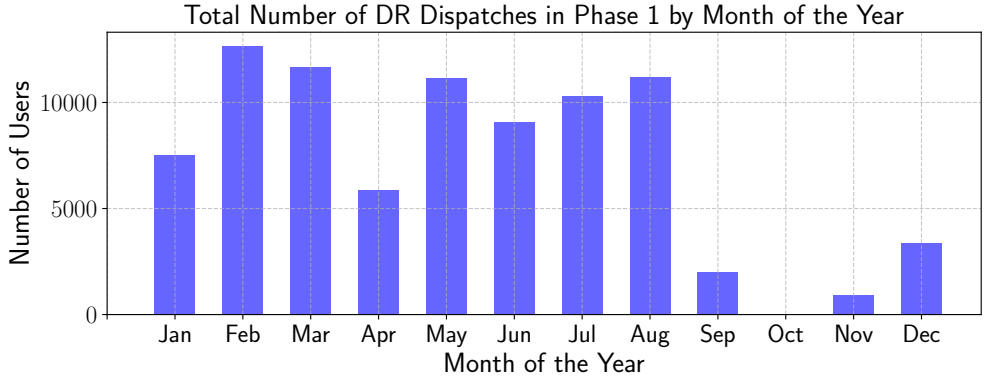

Figure B.4: Distribution of DR Events by Month of the Year Across Users with Completed Phase 1

Balance Metrics for Control and Treatment Group			
	Cohen's D	Hellinger Dist.	Diff. Mean
kWh, HoD = 13	0.023	0.002	0.022
kWh, HoD = 14	0.026	0.002	0.026
kWh, HoD = 15	0.017	0.002	0.017
kWh, HoD = 16	0.006	0.002	0.006
kWh, HoD = 17	0.002	0.002	0.002
kWh, HoD = 18	0.002	0.002	−0.007
kWh, HoD = 19	0.002	0.002	−0.010
kWh, HoD = 20	0.002	0.002	−0.008
air_temp, HoD = 13	0.013	0.005	0.032
air_temp, HoD = 14	0.013	0.005	0.032
air_temp, HoD = 15	0.015	0.005	0.038
air_temp, HoD = 16	0.015	0.005	0.039
air_temp, HoD = 17	0.016	0.004	0.041
air_temp, HoD = 18	0.016	0.003	0.041
air_temp, HoD = 19	0.018	0.003	0.043
air_temp, HoD = 20	0.018	0.002	0.042
# historical obs. (hours)	4.632	0.037	303.7

Table B.1: Balance Checks for Users in Control and Treatment Group

B.1.3 Fixed Effects Regression Tables

Tables B.2-B.4 provide the results of the Fixed Effects Regressions presented in Section 8.6. The point estimates of interest are printed in boldface and are accompanied by the standard errors as well as their 95% confidence intervals. The t-value of the regression gives rise to the p-value, where we use $(*)$, $(**)$, $(***)$ to denote statistical significance at the 90%, 95%, 99% confidence level, respectively.

B.1.4 Comparison of Estimation Methods

Figure 8.23 visually compares the ATEs broken out by incentive level, and it can be seen that both methods produce similar estimates. Figure B.5 does the same for month of the year. Agreeing with intuition, the reductions are notably larger in summer months compared to winter periods. Conditional on the automation status, Table B.4 states that the reductions are -0.331 and -0.103 kWh for automated and non-automated users, respectively, compared to -0.332 and -0.121 kWh calculated by the non-experimental case. These values are close to each other. Lastly, no significant difference in the magnitude of reductions can be found between encouraged and non-encouraged users.

Effect of DR by Incentive Level on Electricity Consumption				
Parameter	Estimate (Std. Err.)	t-Value	95% Conf. Int.	p-value
is_treat_{it}	-0.006 (0.003)	-2.100	[-0.013, 0]	0.047**
BL_{it}	0.879 (0.010)	88.89	[0.859, 0.900]	<0.001***
T_{it}	0.0205 (0.002)	10.79	[0.017, 0.024]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, r_{it} = 0.05)$	**-0.120** (0.014)	-8.532	[-0.148, -0.091]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, r_{it} = 0.25)$	**-0.121** (0.018)	-6.910	[-0.157, -0.085]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, r_{it} = 0.50)$	**-0.115** (0.016)	-7.369	[-0.147, -0.083]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, r_{it} = 1.00)$	**-0.124** (0.020)	-6.219	[-0.166, -0.083]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, r_{it} = 3.00)$	**-0.136** (0.010)	-12.95	[-0.157, -0.114]	<0.001***

Table B.2: Fixed Effect Regression Results by Incentive Level

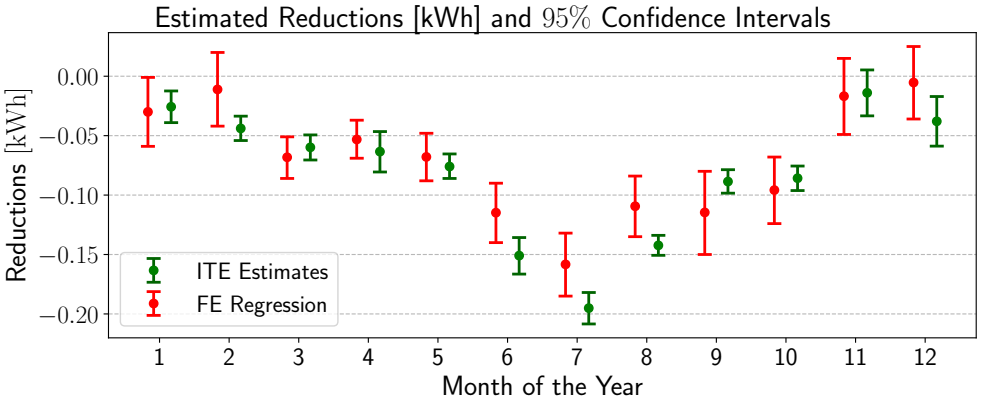

Figure B.5: CATEs by Month of Year with Confidence Intervals, Comparison Fixed Effects Estimators and Non-Experimental Estimators

Effect of DR by Month of Year on Electricity Consumption				
Parameter	Estimate (Std. Err.)	t-Value	95% Conf. Int.	p-value
is_treat_{it}	-0.007 (0.003)	-1.962	[-0.014, 0.001]	0.078*
BL_{it}	0.879 (0.016)	55.52	[0.844, 0.915]	<0.001***
T_{it}	0.021 (0.006)	3.326	[0.007, 0.034]	0.008***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 1)$	**-0.041** (0.010)	-4.298	[-0.063, -0.020]	0.002***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 2)$	**-0.022** (0.009)	-2.571	[-0.041, -0.003]	0.028**
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 3)$	**-0.076** (0.004)	-18.62	[-0.085, -0.067]	0.002***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 4)$	**-0.062** (0.004)	-15.14	[-0.071, -0.053]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 5)$	**-0.094** (0.005)	-20.07	[-0.104, -0.083]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 6)$	**-0.155** (0.008)	-20.25	[-0.172, -0.138]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 7)$	**-0.227** (0.007)	-32.68	[-0.242, -0.211]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 8)$	**-0.159** (0.008)	-19.39	[-0.177, -0.141]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 9)$	**-0.179** (0.014)	-9.071	[-0.218, -0.142]	<0.001***
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 11)$	**-0.029** (0.009)	-3.055	[-0.050, -0.008]	0.012**
$\mathbb{1}(\text{is\_DR}_{it}, \text{MoY}_{it} = 12)$	**-0.022** (0.010)	-2.172	[-0.045, 0.001]	0.055*

Table B.3: Fixed Effect Regression Results by Month of Year

Effect of Home Automation on Electricity Consumption				
Parameter	Estimate (Std. Err.)	t-Value	95% Conf. Int.	p-value
is_treat_{it}	-0.006 (0.003)	-2.101	[-0.013, 0]	0.047^{**}
BL_{it}	0.879 (0.010)	88.94	[0.859, 0.900]	$<0.001^{***}$
T_{it}	0.021 (0.002)	10.79	[0.017, 0.024]	$<0.001^{***}$
$\mathbb{1}(\text{is\_DR}_{it}, \text{is\_auto}_{it})$	**-0.331** (0.042)	-7.800	[-0.418, -0.243]	$<0.001^{***}$
$\mathbb{1}(\text{is\_DR}_{it}, \neg\text{is\_auto}_{it})$	**-0.103** (0.014)	-7.310	[-0.132, -0.074]	$<0.001^{***}$

Table B.4: Fixed Effect Regression Results by Automation Status

Effect of Automation Uptake Incentive on Electricity Consumption				
Parameter	Estimate (Std. Err.)	t-Value	95% Conf. Int.	p-value
is_enc_{it}	-0.005 (0.003)	-1.422	[-0.012, 0.002]	0.168
is_nonenc_{it}	-0.008 (0.003)	-2.485	[-0.015, -0.001]	0.021^{**}
BL_{it}	0.9366 (0.024)	38.38	[0.886, 0.987]	$<0.001^{***}$
T_{it}	0.0206 (0.002)	10.794	[0.017, 0.024]	$<0.001^{***}$
$\mathbb{1}(\text{is\_DR}_{it}, \text{is\_enc}_{it})$	**-0.121** (0.016)	-7.703	[-0.153, -0.088]	$<0.001^{***}$
$\mathbb{1}(\text{is\_DR}_{it}, \text{is\_nonenc}_{it})$	**-0.125** (0.015)	-8.304	[-0.156, -0.094]	$<0.001^{***}$

Table B.5: Fixed Effect Regression Results by Automation Uptake Encouragement

B.1.5 Correlation of Temperature and ITE

As mentioned in the previous subsection, larger reductions are estimated in warm summer months. To test the hypothesis whether or not there exists such a correlation, Figure B.6 scatter plots estimated ITEs as a function of the average ambient air temperature observed during the relevant DR events. We can notice a notable positive correlation of ambient air temperature and the magnitude of reductions. Indeed, a subsequent hypothesis test with the null being a zero slope is rejected with a p-value of less than $1e-9$.

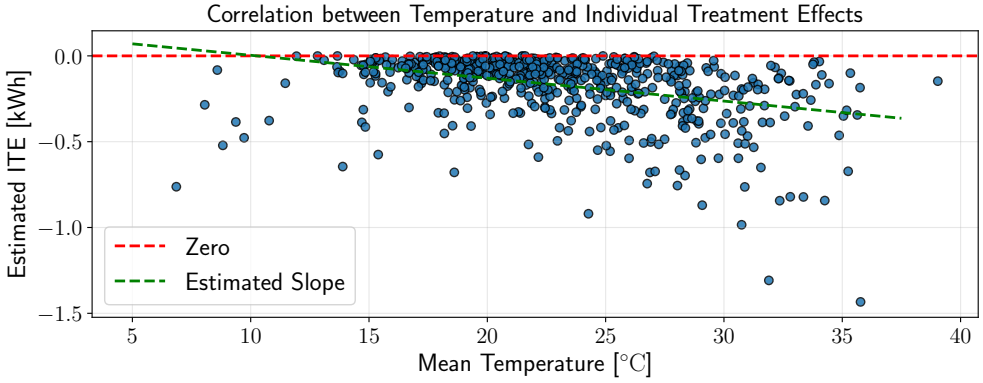

Figure B.6: Correlation between Average Ambient Air Temperature and ITEs.

To support this notion, we marginalize ITEs for each ZIP code to obtain the geographic distribution of CATEs by location, see Figure B.7, and it is visually striking that users in coastal areas in California show smaller reductions than users in the Central Valley, where the climate is hotter.

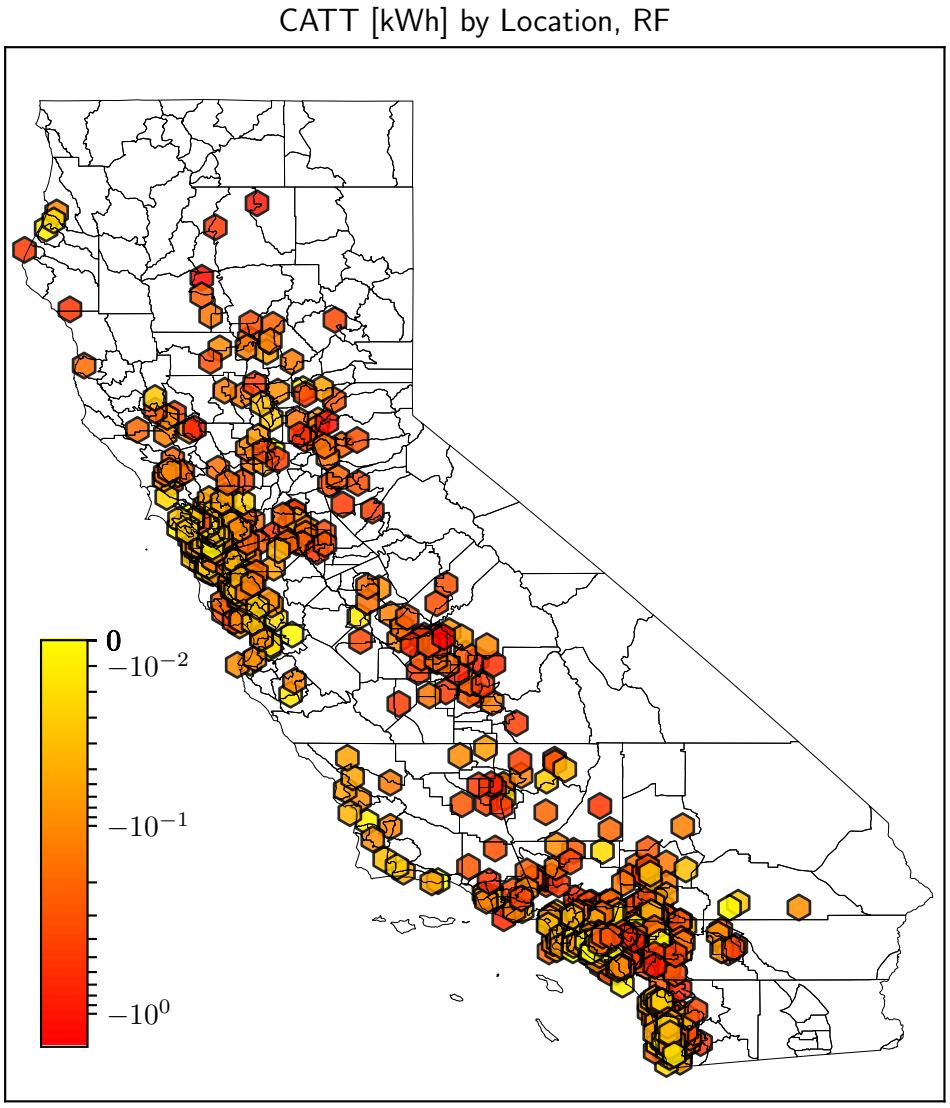

Figure B.7: Correlation between Average Ambient Air Temperature and CATEs.

Bibliography

Aalami, H. A., M. P. Moghaddam, and G. R. Yousefi (2010). "Demand response modeling considering Interruptible Loads and Capacity Market Programs". In: *Applied Energy* 87.1, pp. 243–250.

Abadie, A., A. Diamond, and J. Hainmueller (2012). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

Acemoglu, D., Asuman Ozdaglar, and A. Tahbaz-Salehi (2015). "Systemic Risk and Stability in Financial Networks". In: *The American Economic Review* 105.2, pp. 564–608.

Acemoglu, D. et al. (2011). "Bayesian Learning in Social Networks". In: *Review of Economic Studies* 78, pp. 1201–1236.

Agrawal, R. (2002). "Sample Mean Based Index Policies with O(log n) Regret for the Multi-Armed Bandit Problem". In: *Machine Learning* 47, pp. 235–256.

Agrawal, R., M. V. Hegde, and D. Teneketzis (1990). "Multi-Armed Bandits with Multiple Plays and Switching Cost". In: *Stochastics and Stochastic Reports* 29, pp. 437–459.

Ai, Bing, Zhaoyan Fan, and Robert X. Gao (2014). "Occupancy Estimation for Smart Buildings by an Auto-Regressive Hidden Markov Model". In: *American Control Conference*.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753.

Al-Hamadi, H. M. and S. A. Soliman (2004). "Short-Term Electric Load Forecasting Based on Kalman Filtering Algorithm with Moving Window Weather and Load Model". In: *Electric Power Systems Research* 68.1, pp. 47–59.

Allcott, Hunt (2011). "Social Norms and Energy Conservation". In: *Journal of Public Economics* 95.9, pp. 1082–1095.

Amin, K., A. Rostamizadeh, and U. Syed (2013). "Learning Prices for Repeated Auctions with Strategic Buyers". In: *Advances in Neural Information Processing Systems*, pp. 1169–1177.

Anantharam, V., P. Varaiya, and J. Walrand (1986). "Asymptotically Efficient Allocation Rules for the Multiarmed Bandit Problem - Part I: IID Rewards". In: *IEEE Transactions on Automatic Control* 32, pp. 968–976.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arora, S. and James W. Taylor (2016). "Forecasting Electricity Smart Meter Data Using Conditional Kernel Density Estimation". In: *Omega* 59, Part A, pp. 47–59.

Arora, Siddharth and James W. Taylor (2014). "Forecasting Electricity Smart Meter Data Using Conditional Kernel Density Estimation". In: *Omega*.

Aswani, Anil et al. (2012). "Identifying Models of HVAC Systems Using Semiparametric Regression". In: *American Control Conference*.

Athey, S. and G. Imbens (2016). "Recursive Partitioning for Heterogeneous Causal Effects". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.27, pp. 7353–7360.

Audibert, J.Y., S. Bubeck, and R. Munos (2010). "Best Arm Identification in Multi-Armed Bandits". In: *COLT-23th Conference on Learning Theory*.

Auer, P., N. Cesa-Bianchi, and P. Fischer (2002). "Finite-Time Analysis of the Multiarmed Bandit Problem". In: *Machine Learning* 47, pp. 235–256.

Auer, P. et al. (2002). "The Nonstochastic Multi-Armed Bandit Problem". In: *SIAM Journal on Computing* 32, pp. 48–77.

Ayres, Ian, Sophie Raseman, and Alice Shih (2009). "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage". In: *NBER Working Paper No. 15386*.

Babaioff, M., Y. Sharma, and A. Slivkins (2009). "Characterizing Truthful Multi-Armed Bandit Mechanisms". In: *Proceedings of the 10th ACM Conference on Electronic Commerce*, pp. 79–88.

Baccino, F. et al. (2014). "Frequency Regulation by Management of Building Cooling Systems through Model Predictive Control". In: *Power Systems Computation Conference*, pp. 1–7.

Badanidiyuru, A., R. Kleinberg, and A. Slivkins (2013). "Bandits with Knapsacks". In: *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 207–216.

Balandat, M. (2016). *New Tools for Econometric Analysis of High-Frequency Time Series Data - Application to Demand-Side Management in Electricity Markets*. PhD Dissertation. University of California, Berkeley.

Balandat, M. et al. (2014). "Contract Design for Frequency Regulation by Aggregations of Commercial Buildings". In: *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pp. 38–45.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). "Who's Who in Networks: Wanted The Key Player". In: *Econometrica* 74.5, pp. 1403–1417.

Berman, A. and R. J. Plemmons (1994). *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics.

Bishop, Chris (2006). *Pattern Recognition and Machine Learning*. Springer.

Bitar, E. Y. et al. (2012). "Bringing Wind Energy to Market". In: *IEEE Transactions on Power Systems* 27.3, pp. 1225–1235.

Bollinger, B. and W. R. Hartmann (2015). "Welfare Effects of Home Automation Technology with Dynamic Pricing". In: *Stanford University, Graduate School of Business Research Papers*.

Bonacich, P. (1987). "Power and Centrality: A Family of Measures". In: *American Journal of Psychology* 92, pp. 1170–1182.

Borenstein, S., M. Jaske, and A. Rosenfeld (2002a). "Dynamic Pricing, Advanced Metering, and Demand Response in Electricity Markets". In: *University of California Energy Institute, Center for the Study of Energy Markets*.

Borenstein, Severin (2005). "The Long-Run Efficiency of Real-Time Electricity Pricing". In: *The Energy Journal*.

— (2006). "Customer Risk from Real-Time Retail Electricity Pricing: Bill Volatility and Hedgability". In: *NBER Working Paper No. 12524*.

Borenstein, Severin and S. P. Holland (2005). "On the Efficiency of Competitive Electricity Markets with Time-Invariant Retail Prices". In: *Rand Journal of Economics* 36.3, pp. 469–493.

Borenstein, Severin, M. Jaske, and A. Rosenfeld (2002b). "Dynamic Pricing, Advanced Metering, and Demand Response in Electricity Markets". In: *University of California Energy Institute, Center for the Study of Energy Markets*.

Bound, John, David A. Jaeger, and Regina M. Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak". In: *Journal of the American Statistical Association* 90.430, pp. 443–450.

Bramoullé, Y. and R. Kranton (2007). "Public Goods in Networks". In: *Journal of Economic Theory* 135, pp. 478–494.

Breiman, L. et al. (1984). "Classification and Regression Trees". In: *CRC Press*.

Brodersen, K. et al. (2015). "Inferring Causal Impact Using Bayesian Structural Time-Series Models". In: *The Annals of Applied Statistics* 9.1, pp. 247–274.

CAISO Fifth Replacement FERC Electric Tariff. `http://www.caiso.com/Documents/ConformedTariff_Dec1_2014.pdf`.

CIMIS Station Reports (2015). Tech. rep. California Irrigation Management Information System. URL: `http://www.cimis.water.ca.gov/`.

California Independent System Operator Corporation (CAISO): Fifth Replacement FERC Electric Tariff (2014).

"California Irrigation Management Information System" (2017). In:

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). "Peer Effects and Social Networks in Education". In: *Review of Economic Studies* 76, pp. 1239–1276.

Campaigne, Clay, Maximilian Balandat, and Lillian Ratliff (2016). "Welfare Effects of Dynamic Electricity Pricing". In: *Working Paper*.

Candogan, Ozan, Kostas Bimpikis, and Asuman Ozdaglar (2012). "Optimal Pricing in Networks with Externalities". In: *Operations Research* 60.4, pp. 883–905.

Cesa-Bianchi, N. and G. Lugosi (2006). *Prediction, Learning, and Games*. Cambridge University Press.

— (2009). "Combinatorial Bandits". In: *Proceedings of the 22nd Annual Conference on Learning Theory*.

Chakraborty, T. et al. (2010). "Selective Call Out and Real Time Bidding". In: *WINE* 6484, pp. 145–157.

Chen, F. et al. (2011). "Activity Analysis Based on Low Sample Rate Smart Meters". In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 240–248.

Chen, Wei, Yajun Wang, and Yang Yuan (2013). "Combinatorial Bandits: General Framework, Results and Applications". In: *International Conference on Machine Learning*.

Cole, Wesley J., Elaine T. Hale, and Thomas F. Edgar (2013). "Building Energy Model Reduction for Model Predictive Control Using OpenStudio". In: *American Control Conference*.

Combes, R. et al. (2015). "Combinatorial Bandits Revisited". In: *Advances in Neural Information Processing Systems*, pp. 2116 –2124.

Dehejia, R. H. and Sadek Wahba (1999). "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs". In: *Journal of the American Statistical Association* 94.448, pp. 1053–1062.

Deng, K. et al. (2007). "Bandit-Based Algorithms for Budgeted Learning". In: *IEEE International Conference on Data Mining*, pp. 463–468.

Deng, Shijie and Li Xu (2009). "Mean-Risk Efficient Portfolio Analysis of Demand Response and Supply Resources". In: *Energy* 34.

Diggle, P. J. et al. (2013). *Analysis of Longitudinal Data*. Vol. 2. Oxford University Press.

Ding, W. et al. (2013). "Multi-Armed Bandit with Budget Constraint and Variable Costs". In: *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Dobbs, Justin R. (2012). "A Comparison of Thermal Zone Aggregation Methods". In: *51st Conference on Decision and Control*.

Duffy, M. et al. (2009). "TRNSYS - Features and Functionality for Building Simulation". In: *IBSPA Conference*, pp. 1950 –1954.

Edwards, Richard E., Joshua New, and Lynne E. Parker (2012). "Predicting Future Hourly Residential Electrical Consumption: A Machine Learning Case Study". In: *Energy and Buildings* 49, pp. 591–603.

Efron, Bradley and Robert J. Tibshirani (1994). *An Introduction to the Bootstrap*. CRC Press.

Elattar, Ehab E., John Goulermas, and Q. H. Wu (2010). "Electric Load Forecasting Based on Locally Weighted Support Vector Regression". In: *IEEE Transactions on Systems, Man, and Cybernetics* 40.4.

FERC Order No. 888. Transmission Open Access. Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities (Final Rule).

FERC Order No. 889: OASIS: Open Access Same-Time Information System and Standards of Conduct (Final Rule).

Federal Energy Regulatory Commission (2016). *Assessment of Demand Response and Advanced Metering*. Tech. rep.

Fei, H. et al. (2013). "Heat Pump Detection from Coarse Grained Smart Meter Data with Positive and Unlabeled Learning". In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1330–1338.

Fuller, W. A. (1995). *Introduction to Statistical Time Series*. Wiley-Interscience.

Gabillon, V., M. Ghavamzadeh, and A. Lazaric (2012). "Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence". In: *Advances in Neural Information Processing Systems*, pp. 3212–3220.

Gai, Y., B. Krishnamachari, and R. Jain (2012). "Combinatorial Network Optimization with Unknown Variables: Multi-Armed Bandits with Linear Rewards and Individual Observations". In: *IEEE/ACM Transactions on Networking* 20.5, pp. 1466–1478.

Gandhi, R., S. Khuller, and S. Parthasarathy (2006). "Dependent Rounding and its Applications to Approximation Algorithms". In: *Journal of the ACM (JACM)* 53.3, pp. 324–360.

Goyal, Siddharth and Prabir Barooah (2012). "A Method for Model-Reduction of Nonlinear Thermal Dynamics of Multi-Zone Buildings". In: *Energy and Buildings* 47, pp. 332–340.

Goyal, Siddharth, Herbert A. Ingley, and Prabir Barooah (2013). "Occupancy-Based Zone-Climate Control for Energy-Efficient Buildings: Complexity vs. Performance". In: *Applied Energy* 106, pp. 209–221.

Guan, Che et al. (2013). "Hybrid Kalman Filters for Very STLF and Prediction Interval Estimation". In: *IEEE Transactions on Power Systems* 28.4.

Gurobi Optimization (2016). `http://www.gurobi.com`.

Han, S., S. Han, and K. Sezaki (2010). "Development of an Optimal Vehicle-to-Grid Aggregator for Frequency Regulation". In: *IEEE Transactions on Smart Grid* 1.1, pp. 65–72.

Han, Zhenyu, Robert X. Gao, and Zhaoyan Fan (2012). "Occupancy and Indoor Environment Quality Sensing for Smart Buildings". In: *Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 882–887.

Hansen, Shirley J. and H.E. Burroughs (2013). *Managing Indoor Air Quality*. Lulu Press, Inc.

Hao, H. et al. (2013). "Ancillary Service for the Grid via Control of Commercial Building HVAC Systems". In: *American Control Conference* 467-472.

Härdle, Wolfgang, Hua Liang, and Jiti Gao (2000). *Partially Linear Models*. Springer.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. Springer New York.

Hatami, A. R., H. Seifi, and M.K. Skeikh-El-Eslami (2009). "Optimal Selling Price and Energy Procurement Strategies for a Retailer in an Electricity Market". In: *Electric Power Systems* 79.1, pp. 246–254.

Hippert, H.S., C.E. Pedreira, and R.C. Souza (2002). "Neural Networks for Short-Term Load Forecasting: A Review and Evaluation". In: *IEEE Transactions on Power Systems* 16.1, pp. 44–55.

Hippert, Henrique S. and James W. Taylor (2010). "An Evaluation of Bayesian Techniques for Controlling Model Complexity and Selecting Inputs in a Neural Network for Short-Term Load Forecasting". In: *Neural Networks* 23, pp. 386–395.

Holland, Paul W. (1986). "Statistics and Causal Inference". In: *Journal of the American Statistical Association* 81.396, pp. 945–960.

Hollander, M., D. A. Wolfe, and E. Chicken (2013). *Nonparametric Statistical Methods*. 3rd. John Wiley & Sons.

Hong, Y.-Y. and C.-Y. Hsiao (2002). "Locational Marginal Price Forecasting in Deregulated Electricity Markets Using Artificial Intelligence". In: *IEE Proceedings - Generations, Transmission and Distribution* 149.5, pp. 621–626.

Horn, Roger A. and Charles R. Johnson (2012). *Matrix Analysis*. 2nd ed. Cambridge University Press.

Hu, Q. et al. (2016). "Model Identification of Commercial Building HVAC Systems During Regular Operation - Empirical Results and Challenges". In: *American Control Conference (Accepted)*.

Huang, S., X. Liu, and Zhi Ding (2008). "Opportunistic Spectrum Access in Cognitive Radio Networks". In: *IEEE INFOCOM 2008 Proceedings*, pp. 2101–2109.

Jackson, M. O. and Y. Zenou (2014). "Games on Networks". In: *Handbook of Game Theory*.

Jessoe, K. K., D. L. Miller, and D. S. Rapson (2015). "Can High-Frequency Data and Non-Experimental Research Designs Recover Causal Effects?" In: *Working Paper*.

Jessoe, K. and D. Rapson (2014). "Knowledge is (less) Power: Experimental Evidence from Residential Energy Use". In: *The American Economic Review* 104.4, pp. 1417–1438.

Jordan, Michael I. (2007). *An Introduction to Probabilistic Graphical Models*. In preparation.

Kale, Satyen, Lev Reyzin, and Robert E. Schapire (2010). "Non-Stochastic Bandit Slate Problems". In: *Advances in Neural Information Processing Systems*, pp. 1054–1062.

Katz, L. (1953). "A New Status Index Derived from Socioeconomic Analysis". In: *Psychometrica* 18, pp. 39–43.

Kempe, D., J. Kleinberg, and E. Tardos (2003). "Maximizing the Spread of Influence through a Social Network". In: *9th SIGKDD Intl. Conference on Knowledge Discovery & Data Mining*, pp. 137–146.

Kleiminger, Wilhelm et al. (2014). "Occupancy Detection from Electricity Consumption Data". In: *BuildSys'13*.

Koehler, S. and Francesco Borrelli (2013). "Building Temperature Distributed Control via Explicit MPC and "Trim and Respond" Methods". In: *European Control Conference*.

Komiyama, Junpei, Junya Honda, and Hiroshi Nakagawa (2015). "Optimal Regret Analysis of Thompson Sampling in Stochastic Multi-Armed Bandit Problems with Multiple Plays". In: *International Conference on Machine Learning*, pp. 1152–1161.

Kwac, J. and R. Rajagopal (2013). "Demand Response Targeting Using Big Data Analytics". In: *Big Data, IEEE International Conference on*.

LaLonde, R. J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". In: *The American Economic Review* 76.4, pp. 604–620.

Laffont, Jean-Jacques and David Martimort (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Lai, T. L. and H. Robbins (1985). "Asymptotically Efficient Adaptive Allocation Rules". In: *Advances in Applied Mathematics* 6.1, pp. 4–22.

Lauret, Philippe, Mathieu David, and Didier Caloigne (2012). "Nonlinear Methods for STLF". In: *Energy Procedia* 14, pp. 1404–1409.

Lehmann, E. L. and H. J. M. D'Abrera (2006). *Nonparametrics: Statistical Methods Based on Ranks*. 2nd. Springer New York.

Li, N., L. Chen, and S. H. Low (2011). "Optimal Demand Response Based on Utility Maximization in Power Networks". In: *Power and Energy Society General Meeting*.

Li, Na, Lijun Chen, and M. Dahleh (2015). "Demand Response Using Linear Supply Function Bidding". In: *IEEE Transactions on Smart Grid* 6.4, pp. 1827–1838.

Li, Pan and Baosen Zhang (2016). "An Optimal Treatment Assignment Strategy to Evaluate Demand Response Effect". In: *54th Annual Allerton Conference on Communication, Control, and Computing*.

Li, Y. and N. Li (2016). *Mechanism Design for Reliability in Demand Response with Uncertainty*. `http://scholar.harvard.edu/files/nali/files/accdr2016.pdf`.

Lin, Y., T. Middelkoop, and P. Barooah (2012). "Issues in Identification of Control-Oriented Thermal Models of Zones in Multi-Zone Buildings". In: *51st IEEE Conference on Decision and Control*.

Lin, Yashen et al. (2015). "Experimental Evaluation of Frequency Regulation From Commercial Building HVAC Systems". In: *IEEE Transactions on Smart Grid* 6.2.

Ma, H. et al. (2016). "Incentivizing Reliability in Demand-Side Response". In: *The 25th International Joint Conference on Artificial Intelligence*, pp. 352–358.

Ma, Yudong, Garrett Anderson, and Francesco Borrelli (2011). "A Distributed Predictive Control Approach to Building Temperature Regulation". In: *American Control Conference*, pp. 2089–2094.

Ma, Yudong et al. (2012). "Model Predictive Control for the Operation of Building Cooling Systems". In: *IEEE Transactions on Control Systems Technology* 20, pp. 796–803.

Maasoumy, Mehdi et al. (2014). "Handling Model Uncertainty in Model Predictive Control for Energy Efficient Buildings". In: *Energy and Buildings*.

Mani, A., I. Rahwan, and A. Pentland (2013). "Inducing Peer Pressure to Promote Cooperation". In: *Nature Scientific Reports* 3.

Maréchal, K. (2010). "Not Irrational but Habitual: The Importance of "Behavioural Lock-in" in Energy Consumption"". In: *Ecological Economics* 69.5, pp. 1104–1114.

McQuade, J. M. (2009). *A System Approach to High Performance Buildings*. Tech. rep. United Technologies Corporation.

Milgrom, Paul (2004). *Putting Auction Theory to Work*. Cambridge University Press.

Mirowski, P. et al. (2014). "Demand Forecasting in Smart Grids". In: *Bell Labs Technical Journal* 18.4.

Mohri, M. and A. Munoz (2014). "Optimal Regret Minimization in Posted-Price Auctions with Strategic Buyers". In: *Advances in Neural Information Processing Systems*, pp. 1871–1879.

Mohsenian-Rad, A.-H. et al. (2010). "Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid". In: *IEEE Transactions on Smart Grid* 1.3.

Molina-Markham, A. et al. (2010). "Private Memoirs of a Smart Meter". In: *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pp. 61–66.

Myerson, R. B. and M. A. Satterthwaite (1983). "Efficient Mechanisms for Bilateral Trading". In: *Journal of Economic Theory* 29.2, pp. 265–281.

Nikulin, M. S. *Hellinger Distance*. `"http://www.encyclopediaofmath.org/index.php?title=Hellinger_distance&oldid=16453"`.

OPOWER. `https://opower.com/products/demand-response/`.

OhmConnect, Inc. 2015 (2017). `"https://www.ohmconnect.com"`.

Oldewurtel, Frauke et al. (2010). "Energy Efficient Building Climate Control Using Stochastic Model Predictive Control and Weather Predictions". In: *American Control Conference*.

Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. MIT Press.

Oum, Yumi and Shmuel Oren (2009). "Optimal Static Hedging of Volumetric Risk in a Competitive Wholesale Electricity Market". In: *Decision Analysis* 7.1, pp. 107–122.

Oum, Yumi, Shmuel Oren, and Shijie Deng (2006). "Hedging Quantity Risks with Standard Power Options in a Competitive Wholesale Electricity Market". In: *Naval Research Logistics (NRL)* 53.7, pp. 697–715.

PJM Empirical Analysis of Demand Response Baseline Methods. `https://www.pjm.com/~/media/markets-ops/dsr/pjm-analysis-of-dr-baseline-methods-full-report.ashx`.

PV Performance Modeling Collaborative. URL: `https://pvpmc.sandia.gov/`.

Pai, P.-F. and W.-C. Hong (2005). "Support Vector Machines with Simulated Annealing Algorithms in Electricity Load Forecasting". In: *Energy Conversion and Management* 46.17, pp. 2669–2688.

Palensky, P. and D. Dietrich (2011). "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads". In: *IEEE Transactions on Industrial Informatics* 7.3, pp. 381–388.

Pardo, A., V. Meneu, and E. Valor (2002). "Temperature and Seasonality Influences on Spanish Electricity Load". In: *Energy Economics* 24.1, pp. 55–70.

Parisio, Alessandra et al. (2014). "Control of HVAC Systems via Scenario-Based Explicit MPC". In: *IEEE Conference on Decision and Control*.

Parson, Oliver et al. (2011). "Using Hidden Markov Models for Iterative Non-Intrusive Appliance Modeling". In: *Neural Information Processing Systems Workshop on Machine Learning with Sustainability*.

Pérez-Lombard, Luis, José Ortiz, and Christine Pout (2008). "A Review on Buildings Energy Consumption Information". In: *Energy and Buildings* 40, pp. 394–398.

Pischke, J.-S. and J. D. Angrist (2009). *Mostly Harmless Econometrics*. 1st. Princeton University Press.

Public Utilities Commission of the State of California: Resolution E-4728. Approval with Modifications to the Joint Utility Proposal for a Demand Response Auction Mechanism Pilot (2015).

Public Utilities Commission of the State of California. *Resolution E-4728. Joint Utility Proposal for a DRAM Pilot*.

Rabiner, Lawrence R. (1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". In: *Proceedings of the IEEE* 77.2.

Radecki, Peter and Brandon Hencey (2012). "Online Building Thermal Parameter Estimation via Unscented Kalman Filtering". In: *American Control Conference*.

— (2013). "Online Thermal Estimation, Control, and Self-Excitation of Buildings". In: *52nd Conference on Decision and Control*.

Radlinski, F., R. Kleinberg, and T. Joachims (2008). "Learning Diverse Rankings with Multi-Armed Bandits". In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 784–791.

Rockafellar, R. Tyrrell and Stanislav Uryasev (2002). "Conditional Value-at-Risk for General Loss Distributions". In: *Journal of Banking and Finance* 26.7, pp. 1443–1471.

Rosenbaum, P. R. and D. B. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". In: *Biometrika* 70.1, pp. 41–55.

Rubin, D. B. (1974). "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies". In: *Journal of Educational Psychology* 66.5, pp. 688–701.

Ruppert, David, M.P. Wand, and R.J. Carroll (2003). *Semiparametric Regression*. Cambridge University Press.

Rusmevichientong, P. and D. P. Williamson (2005). "An Adaptive Algorithm for Selecting Profitable Keywords for Search-Based Advertising Services". In: *Proceedings of the 7th ACM Conference on Electronic Commerce*, pp. 260–269.

Samadi, P. et al. (2012). "Advanced Demand Side Management for the Future Smart Grid Using Mechanism Design". In: *IEEE Transactions on Smart Grid* 3.3, pp. 1170–1180.

Seldin, Y. et al. (2014). "Prediction with Limited Advice and Multiarmed Bandits with Paid Observations". In: *International Conference on Machine Learning*, pp. 280–287.

Senjyu, T. et al. (2002). "One-Hour-Ahead Load Forecasting Using Neural Network". In: *IEEE Transactions on Power Systems* 17.1, pp. 113–118.

Sevlian, Raffi Avo and Ram Rajagopal (2014). "A Model For The Effect of Aggregation on Short Term Load Forecasting". In: *IEEE Transactions on Power Systems*.

Sezgen, Osman, C. A. Goldman, and P. Krishnarao (2007). "Option Value of Electricity Demand Response". In: *Energy* 32.2.

Široky, Jan et al. (2011). "Experimental Analysis of Model Predictive Control for an Energy Efficient Building Heating System". In: *Applied Energy* 88, pp. 3079–3087.

Smith, Jeffrey and Petra Todd (2005). "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" In: *Journal of Econometrics* 125, pp. 305–353.

Soares, L. J. and M. C. Medeiros (2008). "Modeling and Forecasting Short-Term Electricity Load: A Comparison of Methods with an Application to Brazilian Data". In: *International Journal of Forecasting* 24.4, pp. 630–644.

Song, K.-B. et al. (2005). "Short-Term Load Forecasting for the Holidays Using Fuzzy Linear Regression Method". In: *IEEE Transactions on Power Systems* 20.1, pp. 96–101.

Sturzenegger, D. et al. (2012). "Semi-Automated Modular Modeling of Buildings for Model Predictive Control". In: *BuildSys 2012 – Workshop of SCM SenSys Conference*.

Taylor, J. W. and P. E. McSharry (2007). "Short-Term Load Forecasting Methods: An Evaluation Based on European Data". In: *IEEE Transactions on Power Systems* 22.4.

Topkis, Donald M. (1998). *Supermodularity and Complementarity*. Princeton University Press.

Tran-Thanh, L. et al. (2010). "Epsilon-First Policies for Budget-Limited Multi-Armed Bandits". In: *Twenty-Fourth AAAI Conference on Artificial Intelligence*, pp. 1211–1216.

Tran-Thanh, L. et al. (2012). "Knapsack Based Optimal Policies for Budget–Limited Multi–Armed Bandits". In: *Twenty-Sixth AAAI Conference on Artificial Intelligence*, pp. 1134–1140.

U.S. Department of Energy Buildings Energy Data Book. URL: `http://buildingsdatabook.eren.doe.gov/`.

Uchiya, T., A. Nakamura, and M. Kudo (2010). "Algorithms for Adversarial Bandit Problems with Multiple Plays". In: *International Conference on Algorithmic Learning Theory*, pp. 375–389.

Vincenty, T. (1975). *Geodetic Inverse Solution Between Antipodal Points*. Tech. rep.

Vrettos, Evangelos et al. (2014). "Robust Provision of Frequency Reserves by Office Building Aggregations". In: *Proceedings of the 19th IFAC World Congress*, pp. 12068 –12073.

Wager, S. and S. Athey (2016). *Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests*. "`https://arxiv.org/pdf/1510.04342v3.pdf`".

Wolak, F. (2010). "An Experimental Comparison of Critical Peak and Hourly Pricing: The PowerCentsDC Program". In: *Department of Economics Stanford University*.

Wolak, Frank A. (2001). "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market". In: *NBER Working Paper 8212*.

Xia, Y. et al. (2016). "Budgeted Multi-Armed Bandits with Multiple Plays". In: *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pp. 2210 –2216.

Zhao, Jie, Khee Poh Lam, and B. Erik Ydstie (2013). "EnergyPlus model-based predictive control (EPMPC) by using MATLAB/SIMULINK and MLE+". In: *Proceedings of 13th Conference of International Building Performance Simulation Association*.

Zhou, D., M. Balandat, and Claire Tomlin (2016a). "A Bayesian Perspective on Residential Demand Response Using Smart Meter Data". In: *54th Annual Allerton Conference on Communication, Control, and Computing*.

Zhou, D., M. Balandat, and Claire Tomlin (2016b). "Residential Demand Response Targeting Using Observational Data". In: *55th IEEE Conference on Decision and Control*.

Zhou, L. (1994). "The Set of Nash Equilibria of a Supermodular Game is a Complete Lattice". In: *Games and Economic Behavior*.