MARKET METHODS FOR SUPPLY AND DEMAND MANAGEMENT IN THE SMART GRID

A Dissertation

by

BAINAN XIA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Srinivas Shakkottai
Committee Members,	P. R. Kumar
	Krishina Narayanan
	Natarajan Gautam
Head of Department,	Miroslav M. Begovic

May 2019

Major Subject: Computer Engineering

ABSTRACT

This study addresses the resource management problem in a large scale networked system with high flexibility. We consider the supply and demand management problem specifically in the context of the future Smart Grid. On the supply side, we design a secondary market to provide stochastic energy service via distributed renewable energy resources. The performance of the proposed market is evaluated in two circumstances, i.e. whether or not the extra energy penetration caused by the market changes the operation point of the power grid. On the demand side, we would like to take the advantages of the residential demand flexibility to relieve consumption peaks and stabilize the system. We conduct certain demand response in a market approach and further build a real experiment system to analyze the performance of such regime.

The study of supply side market is referred to the subheading: *Small-Scale Markets for a Bilateral Energy Sharing Economy* followed by an extension of the corresponding market which brings in the concern that the increased energy penetration may change the operation point of the grid. As for the demand side study, design and analysis of such demand response market is under the subheading: *Mean Field Games in Nudge Systems for Societal Networks* and the real experiment built-up is presented in *Incentive-Based Demand Response: Empirical Assessment and Critical Appraisal*. We model the agent behaviour in both markets via game theoretic approach and analyze the equilibrium performance. We show that a Mean Field Game regime can be applied to accurately approximate these repeated game frameworks and socially desirable equilibria that benefit both system operator and agents exist.

DEDICATION

To my family and friends, for their love and support.

ACKNOWLEDGMENTS

The completion of this dissertation could not have been possible without the expertise of my advisor, Dr. Srinivas Shakkottai. He guided me step by step on the way towards a Ph.D. He continually and convincingly conveyed a spirit of conducting solid research. I feel so lucky to have him as my teacher as well as my friend. I owe my deepest gratitude to him.

I would like to thank Dr. Vijay Subramanian from the University of Michigan, who collaborated in almost every piece of my work. His constancy of enthusiasm on fundamental problems impressed me deeply. I've also learned a lot from his rigorous mathematical thinking. Working with him is one of the greatest experiences in this journey.

I am also very grateful to Dr. Le Xie from Texas A&M University, who opened the door of power system for me. His inputs as a power expert give this dissertation great meaning in real power system.

I would like to acknowledge my committee members, Dr. P. R. Kumar, Dr. Krishina Narayanan and Dr. Natarajan Gautam. I've benefited a lot from their suggestions and comments on my proposal and dissertation. Dr. Kumar encouraged me to follow true interests in choosing research directions, which brought me to my current group. Dr. Narayanan taught me information theory, which turns out to be one of the most important courses that I have taken. The conversation on the EnergyCoupon project with Dr. Gautam gave me great confidence in stepping further toward this direction.

I would like to thank my previous and current group members Jian Li, Rajarshi Bhattacharyya, Vamseedhar Reddyvari, Ki-Yeob Lee, Kartic Bhargav, Desik Rengarajan, Archana Bura and Aria HasanzadeZonuzy, as well as my friends in College Station, for their friendship and support. Thanks for being part of my life.

Last, but not least, I would like to thank my family. Without them none of these would indeed be possible.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of advisor Dr. Srinivas Shakkottai, Dr. P. R. Kumar and Dr. Krishina Narayanan of the Department of Electrical and Computer Engineering and Dr. Natarajan Gautam of the Department of Industrial and Systems Engineering.

The buildup of model presented in Chapter 3 was conducted in collaboration with Jian Li, Xinbo Geng and Hao Ming. The system implementation and the subsequent analysis described in Chapter 4 was conducted in collaboration with Hao Ming, Ki-Yeob Lee, Yuanyuan Li, Yuqi Zhou and Shantanu Bansal. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a research assistantship from Texas A&M University.

NOMENCLATURE

P2P	Peer-to-Peer
A2A	Agent-to-Agent
PV	Photovoltaics
MFG	Mean Field Game
MFE	Mean Field Equilibrium
LMP	Location Marginal Price
CDF	Cumulative Distribution Function
OPF	Optimal Power Flow
LSE	Load Serving Entity
LA	Load Aggregator
EUT	Expected Utility Theory
PT	Prospect Theory
DTMC	Discrete-Time Markov Chain
CTMC	Continuous-Time Markov Chain
DP	Dynamic Program
DR	Demand Response
CIDR	Coupon-Incentive based Demand Response
ERCOT	Electric Reliability Council of Texas
ARIMA	Autoregressive Integrated Moving Average

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Resource management problem in large scale system such as communication network, transportation system and power grid becomes more complex when the uncertainty of the resource is high. In power grid, the resource should be allocated economically so that supply and demand are balanced at any time. With increasing renewable energy penetration, it is more and more difficult for the system operator to fully utilize such time effective resources. Whereas the flexibility of demand could help to handle the uncertainty of supply through intelligent management. Markets facilitate trades taking advantages of such flexibility and potentially allocate resource efficiently. People have proposed many market mediated sharing systems in the forms ranging from bartering to bargaining. Such systems often involve large number of users with infrequent interactions in any random subset of the total population. The users usually make decisions on the time and quantity of possessing the corresponding resource repeatedly. Given these attributes, mean field game framework is a promising approach towards studying such systems.

In Chapter 2, motivated by the ever-increasing installation of photovoltaics in homes and small businesses, we consider a general small-scale market for agent-to-agent (energy) resource sharing, in which each agent could either be a seller (server) or a buyer (client) in each time period. In every time period of the market, a server has a certain amount of resources (e.g., units of energy) that any client could consume, and randomly gets matched with a client. Our target is to maximize the resource utilization in such an agent-to-agent market, where the agents are strategic. During each successful transaction, the server gets money and the client gets resources. Hence, trade ratio maximization implies efficiency maximization in our system. We model the proposed market system through a Mean Field Game approach and prove the existence of Mean Field Equilibria in general, and also ones that can achieve an almost 100% trade ratio. Finally, we carry out a simulation study on a generic problem instance and a case-study of a proposed photovoltaic market, and show the designed market benefits both individuals and the system as a whole. A reasonable extension of the proposed energy sharing market to a larger scale system so that different trade ratios between

different locations induce different operating points of the power grid, i.e. different Locational Marginal Prices (LMP), is discussed at the end of this chapter. This extension essentially dives into the question what if the agents in such market is price anticipating instead of price taking.

In Chapter 3, we consider the general problem of resource sharing in societal networks, consisting of interconnected communication, transportation, energy and other networks important to the functioning of society. Participants in such network need to take decisions daily, both on the quantity of resources to use as well as the periods of usage. With this in mind, we discuss the problem of incentivizing users to behave in such a way that society as a whole benefits. In order to perceive societal level impact, such incentives may take the form of rewarding users with lottery tickets based on good behavior, and periodically conducting a lottery to translate these tickets into real rewards. We will pose the user decision problem as a mean field game (MFG), and the incentives question as one of trying to select a good mean field equilibrium (MFE). In such a framework, each agent (a participant in the societal network) takes a decision based on an assumed distribution of actions of his/her competitors, and the incentives provided by the social planner. The system is said to be at MFE if the agent's action is a sample drawn from the assumed distribution. We will show the existence of such an MFE under general settings, and also illustrate how to choose an attractive equilibrium using as an example demand-response in the (smart) electricity network.

In Chapter 4, we present the system design and results of a real user demand-response experiment in the Smart Grid based on the analytical results discussed in Chapter 3. Demand response (DR) provides both operational and financial benefits to a variety of stakeholders in the power system. As an example, in the deregulated market such as the Electric Reliability Council of Texas (ERCOT), load serving entities (LSEs) usually purchase electricity from the wholesale market and sign fixed retail price contracts with their end-consumers. Therefore, end-consumers' load shift from peak to off-peak hours could benefit the LSE in terms of largely reducing its electricity purchase with extremely high price from the real-time market. As a first-of-its-kind implementation of coupon incentive-based demand response (CIDR), the EnergyCoupon project provides end-consumers with dynamic time-of-the-day DR event announcements, individualized coupon

targets, as well as periodic lottery experiments. This chapter summarizes the design methodology, the critical findings, and potential generalization of such experiment based on the activities in the summer of 2017. Comparison with the conventional time-of-the-day price-based DR program is conducted. It is shown that by combining dynamic coupon with lotteries, the effective cost for demand response providers can be reduced substantially while achieving the same level of demand reduction.

In Chapter 5, we conclude with a summary of the main results of this dissertation.

2. SMALL-SCALE MARKETS FOR A BILATERAL ENERGY SHARING ECONOMY *

2.1 Introduction

The sharing economy is a paradigm shift in the working of the twenty-first century marketplace. Supported by the ease of communication and availability of information provided by the Internet, this marketplace innovation has blurred the line between producers and consumers, turning participants into *prosumers* who can both provide and utilize resources and services. Successful platforms here enable access to resources that are commonplace, but are needed at the right place and at the right time. Typically, these resources are such that "unused value is wasted value," in that idle time cannot be utilized later on. Examples include *peer-to-peer* (P2P) networks such as BitTorrent (bartering of bandwidth), Fon (token-based WiFi sharing), Uber/Lyft (typically, fixed-price car sharing), and Airbnb (marketplace-mediated home sharing).

Prosumers typically provide or consume small amounts of resources, which means that bilateral trade (one-to-one) is the norm. Thus, the sharing platform enables bilateral trading, with options ranging from barter to bargaining with monetary instruments. Prosumers are ephemeral in that they might participate for some duration of time, and then switch to some other platform or stop altogether. The number of participants at any time is large, which is how sharing systems manage to match demand and supply. Also, in most existing sharing systems, prosumers act largely as consumers or producers, but rarely switch roles.

A novel set of applications are now emerging in which prosumers switch roles from being producers to consumers frequently. Here, agents have either demand or resources that are bursty, which results in recurring role changes. Like P2P networks used for content sharing, these applications are associated with easily sharable resources, and provide services that are indistinguishable from traditional sources. In this chapter, we consider *agent-to-agent* (A2A) market design in the context of distributed electricity generation and consumption. We focus on rooftop-photovoltaics

*This chapter has been submitted to IEEE Transactions on Control of Network Systems and is under review. An alternative version is available at https://arxiv.org/abs/1712.04427

(PV) based electricity generation at the level of homes and small businesses, where generation depends on the intensity of sunshine. Here, geographic vagaries mean that one can shift from being a producer to consumer often, and the existing grid can allow incorporation of these resources. A traditional alternative exists here in the form of electricity purchase from a utility provider.

While there are existing platforms using a two-sided market approach for some applications, approaches that focus on prosumers that frequently change roles from provider to consumer are few. Consider a bilateral market in which currency is used as the instrument of trading. A simple mechanism is one in which a consumer (that we term as a *client*) is matched to a random producer (that we term as a *server*), each places a bid, and a trade happens if the client bids higher than the server's demand. The server then receives the currency equal to her bid, and must incur a cost of providing service. Thus, an agent in a client role pays the agent in a server role to obtain resources. Likewise, the agent that is currently a server can use these currency units to obtain resources when it in turn becomes a client. Also, each time an agent in a client role obtains resources, it generates surplus, measured in currency units. This corresponds, for example, to the productivity gains due to obtaining electricity. If the client does not succeed in obtaining service under the sharing economy, it faces a cost, which can be thought of as the negative feeling of having to search for an alternative source or to experience delays. *Would such a market be sustainable, i.e., would there be enough resource trades generating currency (surplus) such that available resource utilization is high?*

In this chapter, we develop a game theoretic framework to model and analyze A2A markets for electric energy under the mechanism described above. The choice of mechanism often depends on the timescale of resource usage, with simple solutions such as bartering being effective at short timescales, and more complex ones like bargaining at long timescales. The timescale of hours for energy sharing suggests that a low complexity solution is desirable, and the value of our proposed solution will be apparent in later sections. In the context of our application, random matching of agents is viable since integration of renewable energy into the electricity grid is already well established in the US (eg. using net-metering in which customers can sell back excess renewable energy generated [4]). The net effect is to simply inject power into the bus to which the producer

is connected, while the consumer draws energy from the grid at the bus to which it is connected. Hence, the electrons produced by the producer are not directly transferred to the consumer, and only monies are transferred between consumer and producer.

Our market model consists of random matching between a large number of ephemeral agents that might leave at any time. We assume that a departing agent is replaced with a new agent, keeping the total number of agents fixed. The state of any agent is the amount of currency that it possesses at that time. A client can be constrained to place a bid only if it has sufficient currency to do so. It is clear that such a budget constraint might restrict entering agents from obtaining resources, and result in low trade volume. Indeed, some P2P networks such as BitTorrent build in a measure of altruism to reduce friction in the system. In this chapter, we consider models in which the client may obtain a loan in order to pay for service. The client must pay back the loan with interest after the trade using the currency (surplus) generated by receiving resources and any budget money in its coffers. We consider two loan options, namely, (i) the client can obtain the loan from an outside lending agency – a bank-loan, or (ii) from the server itself – a peer-loan. The peer-loan model is the most general in that an infinite interest rate will ensure a hard budget constraint, while the bank-loan model is a special case under which the client pays interest, but the server sees a zero interest rate as the interest is transferred out of the system.

2.1.1 Mean Field Games

We investigate the existence of an equilibrium using the framework of Mean Field Games (MFG) [5]. Here, each agent assumes that the matched agent would play an action drawn *independently* from a fixed distribution over its bid space. The agent then chooses an action that is a best response against actions drawn in this manner. The system is said to be at Mean Field Equilibrium (MFE) if this best response action is itself a sample drawn from the assumed bid distribution. This framework considerably reduces computational overhead, and can easily be shown to be an accurate approximation in range of applications [2, 6–8] including our context, when the number of agents is large enough.

The MFG framework offers a relatively simple way of modeling and analyzing large-scale

games when each subset of agents interacts infrequently. In the context of the sharing economy, a particular producer and consumer would rarely be matched together multiple times in their lifetimes, since the number of participants is large and participant lifetime is limited. This implies that little utility is lost due to minimal history retention, and the action choice becomes less complex.

The main related papers in the MFG setting are [6, 7]. In [6], a sytem for auctioning advertisements on a webpage is considered. Here agents are advertisers that bid for these spots, and the main result shows how convergence to the MFE takes place while learning about the value of winning a slot on the webpage. The model is extended in [7] to include hard budget constraints in the sense that agents may only bid an amount less than their existing budget. The budget itself is updated according to an independent arrival process, and the result is a characterization of the reduced bid that would be made in this case. Neither of these considers matching markets of producers and consumers that are interchangeable.

2.1.2 Other Related Work

There has recently been much work in the context of the sharing economy, but little in the way of understanding systems in which agents change their roles often. Most work that deals with this problem considers the special case of data/spectrum sharing in wireless networks. For instance, [9] study pricing models for a system like Fon in which WiFi is shared. In the same manner, [10, 11] study spectrum sharing and mobile data offload in which peers can use each other's resources in the setting of a small number of agents. They consider mechanisms across a small number of agents such as contracts and double auctions. However, they do not consider repeated play with learning of behaviors.

In the context of sharing electricity storage resources, [12] considers a model of charging when prices are low and sharing when prices are high. However, storage is currently very expensive and its penetration is still low as compared to PV installations, particularly in Texas which is the setting of our case study. Hence, we do not assume any storage, and usage by a consumer can only happen with successful trade.

To the best of our knowledge, there is no prior work that considers mechanism design for bi-

lateral (A2A) repeated games with role switching agents between producer (server) and consumer (client) in the mean-field setting. Our initial analysis on this problem [1] only considers the special case of the bank-loan model. Furthermore, it only presents an overview of results and no proofs are included. The current work generalizes the results to the peer-loan model that subsumes hard budget constraints, bank-loan, as well as peer-loan and also emphasizes the methodological contribution by presenting the important proofs.

2.1.3 Main Results

We present a characterization of the mean-field equilibrium bid distribution for the general case of the peer-loan model, and show the existence of a mean-field equilibrium with all the important proofs. We show that there exists a set of equilibria that are simple, and characterized by the server setting a fixed price k, while the client chooses whether or not to bid k based on her budget and estimate of future value. In particular, the client decision turns out to be a set of divisions of the budget into intervals, with k being optimal in some and 0 being optimal in others. In all cases, if the budget is sufficiently large, the client always bids k, while if it is sufficiently small, it always bids 0.

The stable bid k is not unique, and a set of such bids exist (with the minimum being lower bounded by server cost, and the maximum being upper bounded by the client surplus plus cost of not obtaining service), each one of which is an MFE. However, the fraction of time that a trade happens (i.e, the client actually bids k) is not the same for all systems and all values of k. In particular, the bank-loan and peer-loan models both attain higher trade ratios than the hard budget constrained system, particularly in the case when the initial budget of an agent is low. Essentially, a small boost in the form of a loan (which is retuned immediately with interest via the surplus generated by the trade) is successful in reducing friction in the market allowing it to attain high efficiency.

The trade ratio also depends on the value of k itself. Interestingly, maximum trade is not necessarily attained at the lowest possible value of k, but there exists a value between the highest and lowest at which this happens. The reason is that since clients and servers are interchangeable,

extraction of surplus by a server is not always a bad thing from the client's perspective, since it too will gain when the roles are reversed. When the initial budget is low, a client is forced to take a loan in order to bid a high value of k. But when it does so, it transfers a larger sum to the server, which then (subtracting service cost), might be in a position to obtain service without having to take a loan in the future (when it's a client). Thus, aggregation of surplus at servers may not be bad.

For evaluation, we first conduct numerical studies to illustrate the viability of our scheme and compare the performance among the three systems: the hard-budget-constrained model, the bank-loan model and the peer-loan model, in all of which agents alternate probabilistically between being a client or a server. Comparative statics on trade ratios and optimal prices are studied in this setting. In order to make the learnings concrete, we also conduct a case study on a synthetic grid of Texas from the Electric Grid Test Case Repository [13, 14]. We utilize a data trace with per-bus demand at each hour, estimate the available PV energy using weather data traces, inject power at certain buses based on trade achieved in our (secondary) market, and determine the impact both on the feasibility and on location marginal prices (LMPs) in the primary market (as obtained by solving the Optimal Power Flow Problem in each hour). We show that even when injections to the tune of 25% of the peak residential load occur, there is essentially no change to the LMPs and the grid feasibility conditions are not violated. We then estimate the gains on a per agent (household or small business) basis from using our market mechanism to be close to two hundred dollars a year.

2.2 Mean Field Model

We consider a general model of the proposed market with a large number of agents. Each agent maintains a private budget state and can bid any value within her budget plus some affordable loan in the role of a client. The meaning of "affordable" will become clear when the value functions are defined. When a client gets matched to a server, each places a bid. If the server indicates a lower price than what the client proposes, a bilateral trade happens. At the end of a successful trade, the client pays the server's asking price together with the interest on any loan received, receives

service and translates it into a dollar value surplus that directly increases her budget. Meanwhile, the server receives the payment, and pays the cost of the providing service. Thus, the client will bid strategically under some belief about the likely bids of the server, and *vice versa*. Computing

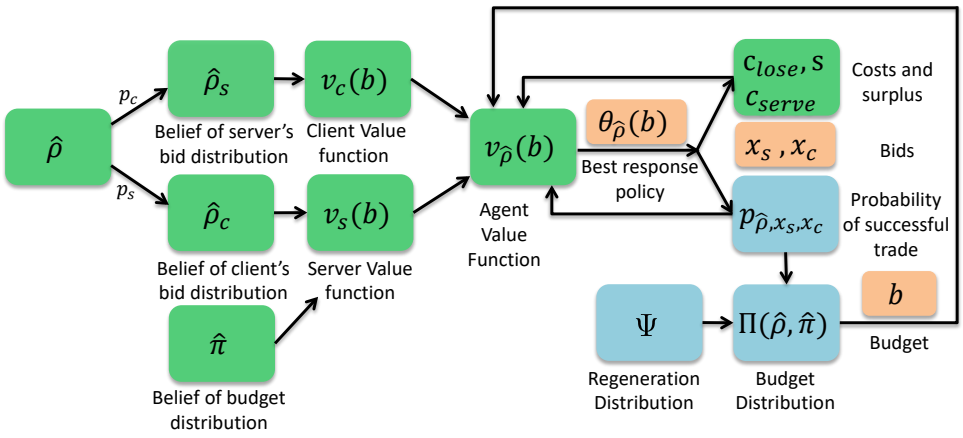

Figure 2.1: Mean Field Game

perfect Bayesian equilibria of such a system is complex, particularly when the number of agents is large. Instead, we use a mean field approximation of the proposed market system, which has proven to be an accurate representation when the number of agents is asymptotically large [6, 15]. Fig. 2.1 illustrates our mean field model from the perspective of a single agent.

At each discrete time step, an agent could either be a client or a server with fixed probabilities, p_c and p_s, respectively. A client places a bid based on her belief of server's bid distribution, and *vice versa*. In addition to the client's bid distribution, a server also maintains a belief of client's budget distribution conditional on the trade happening to evaluate the amount of interest that she might gain through the loan. The conditional budget distribution can be obtained through a re-normalization of the original budget distribution shifted by a constant, given the belief of bid distributions. To simplify notation, we ignore this constant shift which induces no impact on the following analysis, and use the notation of the original budget distribution instead. Note that since clients and servers change roles, this is the same as a belief over a generic agent's budget

distribution, with a density function denoted by $\hat{\pi}$. We assume that $\hat{\pi}$ is continuous. Since the number of the agents is large, which implies that both the number of clients and servers at any instant are large as well, each individual can assume that her opponent's bid is drawn independently from the c.d.f. $\hat{\rho}_c$ or $\hat{\rho}_s$, respectively. Then the complexity of the single agent decision making problem is much reduced. In the rest of this section, we will provide a term-wise description of our mean field model with the accompanying notation.

Time: Time is discrete and indexed by $t \in \{0, 1, ...\}$.

Agent: At each time period t, an agent is either a client or a server, with probability p_c or p_s, respectively. Note that the total number of agents is large and $p_c + p_s = 1$.

Bids: When a client is matched with a server, each places a bid, denoted as x_c and x_s, for the client and server, respectively. When $x_c \geq x_s$, the trade occurs and the client pays x_s to the server.

State: Each agent keeps track of her budget, b, as a private state. At time t, the budget of a agent is updated as following if a trade happens, i.e. $x_c \geq x_s$,

$$b[t+1] - b[t] = \begin{cases} s - x_s - \alpha_c(x_s - b[t])^+, \text{ as a client w.p. } p_c \\ x_s - c_{serve} + \alpha_s(x_s - \hat{B}[t])^+, \text{ as a server w.p. } p_s, \end{cases}$$

where $\hat{B}[t]$ is a random variable representing the budget of a contacted client, and is distributed according to $\hat{\pi}$. If a trade fails, i.e., $x_c < x_s$, then $b[t+1] = b[t]$ for both agents.

Note that s represents the fixed dollar value surplus that a client gains from receiving service, and c_{serve} is the corresponding fixed cost that a server pays for providing service. Parameter α_c is the penalty term when a client overdraws on her budget, i.e., receives a loan from the peer server, which is then paid back in full with interest after service is obtained and surplus is generated, and α_s is the return to the server. Here, both α_c and α_s are at least 1 with $\alpha_s \leq \alpha_c$. For the hard budget model $\alpha_c = +\infty$ and $\alpha_s = 1$, and for the bank-model $\alpha_s = 1$ and $\alpha_c > 1$ with $\alpha_c - 1$ being the bank's interest rate. The support of the budget is $\mathbb{R}_+ := [0, +\infty)$.

Costs: We have already mentioned c_{serve}, which is the server's cost of providing service. In addition, we introduce another cost, c_{lose}, which denotes the cost of failure to obtain service as a

client. This models an instantaneous dissatisfaction suffered by the client, but does not impact her budget.

Regeneration: An agent may quit the system at the start of any time period t with probability $1-\beta$, and may stay with probability $\beta \in (0,1)$. We assume that a new agent enters the system when an old agent leaves, and the budget of the new agent is drawn from a probability distribution Ψ with a density on B_{init} that is a bounded subset of $\mathbb{R}_+$ (for simplicity).

Best Response Policy: As an agent participates in the system, she places bids at each time period. Hence, she needs to solve a repeated decision making problem, given the private budget b, the public belief about the bid distribution $\hat{\rho} = [\hat{\rho}_c, \hat{\rho}_s]$ and the belief about the budget distribution $\hat{\pi}$. The probability of the trade happening can be computed for a given bid using the public belief about her opponent's bid distribution. This probability characterizes the next step transition of an agent's budget. Hence, a dynamic program is defined for an agent to find her best response policy, $\theta_{\hat{\rho}}$, as is shown in green/dark blocks of Fig. 2.1. We will discuss this in detail in section 2.3.

Stationary Distribution of Budget: Given the best response policy, the state transition of an agent is described by equation (2.1) together with the regeneration, which forms the transition kernel of a discrete-time Markov process. The stationary distribution of this Markov process, π, is equivalent to the resulting budget distribution at the end of the lifetime of any given agent with the public belief $\hat{\rho}$ and $\hat{\pi}$.

Mean Field Equilibrium: Given the assumed belief $\hat{\rho}$ and $\hat{\pi}$, solving the dynamic program, the best response policy is obtained, which defines the kernel of the budget Markov process. Thereafter, taking the stationary distribution of budget together with the best response of each state, a new bid distribution $\gamma(\hat{\rho})$ can be calculated. If $\gamma(\hat{\rho})$ turns out to be the same as the public belief $\hat{\rho}$, the system is at an MFE. Detailed discussions of MFE can be found in Section 2.4. Note that there are various ways to spread the public statistics $\hat{\rho}$ and $\hat{\pi}$ in practice, e.g., via mobile apps (see [16]) or a public website.

Discussion: While the above model is kept simple for purposes of exposition, more complex phenomena can easily be added. For instance, the diurnal variation of weather causing client

and server interchange can be included by an additional Markovian state variable indicating if the weather is sunny or overcast. Similarly, variable demand over the course of the day can be included by considering a time-of-day mean field model as in [2]. Availability of storage too can be incorporated with an additional state variable for the agents, whose value could also determine the likelihood of being a client or a server in the current state. While surplus is chosen as deterministic in the above model, we expect similar results to hold with stochastic surplus. Given that one unit of energy is being shared through one trade in the current setup, an opportunistic way to deal with variable surplus is by conducting multiple trades upon request. Behavior with all of these extensions will be more complex, since we would have to include a belief on the additional state variables, and an expectation by the servers of a trade only yielding partial returns. However, the feasibility of the simplest scenario then opens up the possibility of studying these generalizations.

While the technical reason for regeneration is to ensure correlation decay leading to asymptotic independence of client and server budgets [15], it also models real behavior of agents such as moving to a new house, changing utility providers, PV system maintenance, etc. Also, in our current regeneration scheme, we keep the total number of agents being constant over time, which is not necessary in the mean field model. Instead, only the average number of agents needs to be stationary in order to show the stationarity of the budget. Finally, our analysis assumes that mixing happens fast enough that stationary regime analysis is accurate in the current time-block and the resource utilization could be maximized with little regularity. Our experiments in Sections 2.5 and 2.6 show that the required mixing indeed holds, even for large instances. In particular, the case study that is based on a data trace does have hour-by-hour changes in the weather, and accounts for this variation at each step. Beliefs converge even under this setup, which indicates robustness of the approach.

2.3 Best Response Policy

We first introduce a few easily established facts regarding equilibrium behavior of the agents. Here, we consider agents bidding discrete values in $\mathbb{R}_+$. Later, we will show a specific class of highly efficient equilibria exists, in which all servers bid a single price while clients either accept

or reject. We further define random variables $\tilde{X}_s$ and $\tilde{X}_c$ distributed according to $\hat{\rho}_s$ and $\hat{\rho}_c$, and the corresponding p.m.f. are $p_{\tilde{X}_s}$ and $p_{\tilde{X}_c}$. Suppose the bid spaces of clients and servers are upper bounded by $\bar{x}_c$ and $\bar{x}_s$, we have following:

Fact 1. *Given $\hat{\rho}_s$, a client should never bid higher than $\bar{x}_s$, where $\bar{x}_s$ is the upper-end of the support of $\hat{\rho}_s$, i.e., $\hat{\rho}_s(x_s) = 1 \; \forall x_s \geq \bar{x}_s$ and $\hat{\rho}_s(x_s) < 1 \; \forall x_s < \bar{x}_s$.*

Fact 2. *Given $\hat{\rho}_c$, a server should never bid higher than $\bar{x}_c$, where $\bar{x}_c$ is the upper-end of the support of $\hat{\rho}_c$, i.e., $\hat{\rho}_c(x_c) = 1 \; \forall x_c \geq \bar{x}_c$ and $\hat{\rho}_c(x_c) < 1 \; \forall x_c < \bar{x}_c$.*

Fact 3. *Given $p_{\tilde{X}_s}$, a client should never bid x_c for $x_c > 0$ such that $p_{\tilde{X}_s}(x_c) = 0$.*

Fact 4. *Given $p_{\tilde{X}_c}$, a server should never bid x_s such that $p_{\tilde{X}_c}(x_s) = 0$.*

These facts hold since the violation each of them yields a non-positive expected payoff to a generic agent. Thus, we claim that if an equilibrium exists, which we will discuss in section 2.4, then in each equilibrium, by Fact 1 and 2, we have $\bar{x}_s = \bar{x}_c$, meanwhile, by Fact 3 and 4, we have the action space of an agent in each role has the same discrete support, denoted as $\mathcal{D} \subset \mathbb{R}_+$, with the corresponding beliefs. Furthermore, the number of clients and servers are not required to be the same in our setup. The mismatched cases can be absorbed by the extreme values of the corresponding support. If there are more servers than clients, then a server will have certain extra probability to see a client bidding zero. In the opposite case, a client will see an increased probability of servers bidding the upper bound. The existence of such boundaries in the action space is shown later in Lemma 1.

2.3.1 Value Function

As we discussed in section 2.2, the repeated decision making problem for a single agent (with a geometrically distributed lifetime) forms a discounted cost dynamic program. The agent plays two roles (client or server) probabilistically over its lifetime. In each role, the agent encounters a different decision making problem, but based on its private budget that is common to both roles. Throughout, we will track a generic agent just before her role (client or server) is revealed.

However, as an agent takes an action (bids) only after her role is revealed, this is consistent with our set-up mentioned in Section 2.2. Hence, we define the Bellman equation associated with the dynamic program of interest as follows:

$$
\begin{aligned}
v_{\hat{\rho},\hat{\pi}}(b) &= p_s v_s(b) + p_c v_c(b) \\
&= p_s\big(\max_{x_s \in \mathcal{D}} \mathbb{E}_{\hat{\rho}}[\mathbf{1}_{\tilde{x}_c \geq x_s}(\mathbb{E}_{\hat{\pi}}[\beta v_{\hat{\rho},\hat{\pi}}(b + x_s - c_{serve} + \alpha(x_s - \hat{B})^+)] + x_s - c_{serve}) \\
&\quad + \mathbf{1}_{\tilde{x}_c < x_s} \beta v_{\hat{\rho},\hat{\pi}}(b)]\big) \\
&\quad + p_c\big(\max_{x_c \in \mathcal{D} \cap [0, b+s/(1+\alpha)]} \mathbb{E}_{\hat{\rho}}[\mathbf{1}_{x_c \geq \tilde{x}_s}(\beta v_{\hat{\rho},\hat{\pi}}(b + s - \tilde{x}_s - \alpha(\tilde{x}_s - b)^+) + s - \tilde{x}_s) \\
&\quad + \mathbf{1}_{x_c < \tilde{x}_s}(\beta v_{\hat{\rho},\hat{\pi}}(b) - c_{lose})]\big), \qquad (2.1)
\end{aligned}
$$

where $\tilde{x}_s$ and $\tilde{x}_c$ are realizations of random variables, $\tilde{X}_s$ and $\tilde{X}_c$, and $v_{\hat{\rho},\hat{\pi}}(\cdot)$ is the value function of a generic agent, which is in turn composed of an average of $v_s(\cdot)$ and $v_c(\cdot)$ (value functions once her role is revealed). Also $\alpha_s = \alpha_c = \alpha$ for ease of exposition. Since each agent's role is determined exogenously at the beginning of a time period, the evolution of both $v_s(\cdot)$ and $v_c(\cdot)$ depends on $v_{\hat{\rho},\hat{\pi}}(\cdot)$ so that (2.1) remains consistent. We believe that the exogenously driven role choice makes this a natural assumption. Since the role of the agent in our context is usually determined by the external environment, the value of currency should be determined by the underlying market and not the role one's currently playing.

In our model, a client is allowed to overdraw her budget with an upper limit such that the budget does not end up negative after any possible transaction, i.e. a client is allowed to choose up to the maximum value of x_c subject to $(b + s - x_c - \alpha(x_c - b)^+)$ being non-negative. A simple calculation then yields the upper limit of a client's bid as $b + s/(1 + \alpha)$. As mentioned earlier, the interest that a server may gain through the peer loan depends on the realized budget of the peer client, denoted by $\hat{b}$ (drawn according to $\hat{\pi}$). Also, for both clients and servers, when a trade happens, a budget update as well as an instantaneous gain in value is induced, which captures the fact that the trade generates value both in the present and in the future. Further, notice that the expectation of the indicator functions in equation (2.1) can be determined using the probability of

trade happening, which in turn can be calculated directly using $\hat{\rho}$. Then we can further characterize $v_{\hat{\rho},\hat{\pi}}(b)$ as follows:

$$
\begin{aligned}
&v_{\hat{\rho},\hat{\pi}}(b) = p_s\Big(\beta v_{\hat{\rho},\hat{\pi}}(b) + \max_{x_s\in\mathcal{D}}(1-\hat{\rho}_c(x_s))(\beta(\mathbb{E}_{\hat{\pi}}[v_{\hat{\rho},\hat{\pi}}(b+x_s-c_{serve}+\alpha(x_s-\hat{B})^+)] \\
&- v_{\hat{\rho},\hat{\pi}}(b)) + x_s - c_{serve})\Big) + p_c\Big(\max_{x_c\in\mathcal{D}\cap[0,b+s/(1+\alpha)]}\Big[\sum_{\tilde{x}_s=0}^{x_c} p_{\tilde{X}_s}(\tilde{x}_s)\big(\beta v_{\hat{\rho},\hat{\pi}}(b+s \\
&- \tilde{x}_s - \alpha(\tilde{x}_s-b)^+) + s - \tilde{x}_s\big) + (1-\hat{\rho}_s(x_c))(\beta v_{\hat{\rho},\hat{\pi}}(b) - c_{lose})\Big]\Big) \\
&= \beta v_{\hat{\rho},\hat{\pi}}(b) + \max_{(x_s,x_c)\in\mathcal{A}(b)}\Big(\Big[p_s(1-\hat{\rho}_c(x_s))(x_s-c_{serve}) + p_c\big(\hat{\rho}_s(x_c)(s-\mathbb{E}[\tilde{X}_s|\tilde{X}_s\le x_c])- \\
&(1-\hat{\rho}_s(x_c))c_{lose}\big)\Big] + \beta\Big[p_s(1-\hat{\rho}_c(x_s))\Delta v_s(b,x_s,c_{serve},\hat{\pi}) + p_c\sum_{\tilde{x}_s=0}^{x_c} p_{\tilde{X}_s}(\tilde{x}_s)\Delta v_c(b,s,\tilde{x}_s,\alpha)\Big]\Big),
\end{aligned}
\tag{2.2}
$$

where

$\mathcal{A}(b)$ is the two dimensional bid space $\mathcal{D}\times\mathcal{D}\cap[0,b+s/(1+\alpha)]$,

$\Delta v_s(b,x_s,c_{serve},\hat{\pi}) = \int_0^\infty \hat{\pi}(\hat{b})v_{\hat{\rho},\hat{\pi}}(b+x_s-c_{serve}+\alpha(x_s-\hat{b})^+)d\hat{b} - v_{\hat{\rho},\hat{\pi}}(b)$, and $\Delta v_c(b,s,\tilde{x}_s,\alpha) = v_{\hat{\rho},\hat{\pi}}(b+s-\tilde{x}_s-\alpha(\tilde{x}_s-b)^+) - v_{\hat{\rho},\hat{\pi}}(b)$. Where we have used the fact that $\hat{\pi}$ is the density function of (belief) budget of an agent and the latter two functions account for the change in value with a trade for a server and a client, respectively. Then, the space of possible value functions is

$$\mathcal{V} = \{f : (\mathbb{R}_+ \to \mathbb{R}) : \|f\|_\infty < \infty\} = L_\infty.$$

Define the Bellman operator $T_{\hat{\rho},\hat{\pi}}$ on L_∞ as below:

$$
\begin{aligned}
&(T_{\hat{\rho},\hat{\pi}}f)(b) \\
&= \beta f(b) + \max_{(x_s,x_c)\in\mathcal{A}(b)}\Big(\Big(p_s(1-\hat{\rho}_c(x_s))(x_s-c_{serve}) + p_c\big(\hat{\rho}_s(x_c)(s-\mathbb{E}[\tilde{X}_s|\tilde{X}_s\le x_c]) - (1 \\
&- \hat{\rho}_s(x_c))c_{lose}\big)\Big) + \beta\Big[p_s(1-\hat{\rho}_c(x_s))\Delta f_s(b,x_s,c_{serve},\hat{\pi}) + p_c\sum_{\tilde{x}_s=0}^{x_c} p_{\tilde{X}_s}(\tilde{x}_s)\Delta f_c(b,s,\tilde{x}_s,\alpha)\Big]\Big)
\end{aligned}
\tag{2.3}
$$

where

$\Delta f_s(b, x_s, c_{serve}, \hat{\pi}) = \int_0^\infty \hat{\pi}(\hat{b}) f(b + x_s - c_{serve} + \alpha(x_s - \hat{b})^+) d\hat{b} - f(b)$,

$\Delta f_c(b, s, \tilde{x}_s, \alpha) = f(b + s - \tilde{x}_s - \alpha(\tilde{x}_s - b)^+) - f(b)$.

2.3.2 Properties of the Value Function

In order to characterize the best response policy, we need to derive some useful properties of the value function $v_{\hat{\rho},\hat{\pi}}$. We start by proving the convergence of value iteration for the Bellman operator $T_{\hat{\rho},\hat{\pi}}(\cdot)$. This follows immediately from classical results in [17], as long as we can prove the following three lemmas. Define the transition kernel $\mathcal{Q}(B|b, (x_s, x_c))$ for non-empty Borel subset $B \subset \mathbb{R}_+$ by equation (2.1) together with the regeneration. Note that given x_s, x_c, the probability of trade happening can be directly calculated through $\hat{\rho}$.

Lemma 1. *For every state $b \in \mathbb{R}_+$,*

1) There exists an effective bid space $\hat{\mathcal{A}}(b)$, which is compact;

2) The reward-per-stage is lower semi-continuous in (x_s, x_c);

3) The function $\mu(b, x_s, x_c) := \mathbb{E}_{\mathcal{Q}}[u(B)|b, (x_s, x_c)]$ is continuous in $(x_s, x_c) \in \hat{\mathcal{A}}(b)$ for every function $u \in \mathcal{V}$.

Proof. The proof of 1) follows from showing the existence of upper bounds on bids for both client and server yielding $\hat{\mathcal{A}}(b)$. The reward-per-stage in (2.3) is defined as $c(b, (x_s, x_c)) \triangleq p_s(1 - \hat{\rho}_c(x_s))(x_s - c_{serve}) + p_c\big(\hat{\rho}_s(x_c)(s - \mathbb{E}[\tilde{X}_s|\tilde{X}_s \le x_c]) - (1 - \hat{\rho}_s(x_c))c_{lose}\big)$ and given b, x_s, x_c, the kernel $\mathcal{Q}$ is fully determined by $p_s, p_c, \hat{\rho}, \hat{\pi}, \Psi$. The continuity of $\mathcal{Q}$ over the discrete topology of $\hat{\mathcal{A}}(b)$ is natural. Details of this proof are available in Appendix. □

Lemma 2. *There exist constants $\xi \ge 0$ and $\eta \ge 0$ with $1 \le \eta < 1/\beta$, and a function $w \ge 1$ s.t. for every state b*

1) $\sup_{\mathcal{A}(b)} |c(b, (x_s, x_c))| \le \xi w(b)$*; and*

2) $\sup_{\mathcal{A}(b)} \mathbb{E}_{\mathcal{Q}}[w(B)|b, (x_s, x_c)] \le \eta w(b)$.

Lemma 3. *For every state (b), the function $\omega(b, x_s, x_c) := \mathbb{E}_{\mathcal{Q}}[w(B)|b, (x_s, x_c)]$ is continuous in $(x_s, x_c) \in \hat{\mathcal{A}}(x)$.*

Proof. Since c_{serve}, s and c_{lose} are fixed, $c(b,(x_s,x_c))$ is bounded. Then taking a bounded function w, with the continuity of Q, the results in Lemma 2 and 3 are straightforward. □

Theorem 1. *(Hernandez-Lerma [17]) Given the belief $\hat{\rho}_s, \hat{\rho}_c, \hat{\pi}$ and the corresponding p.m.f. $p_{\tilde{X}_s}, p_{\tilde{X}_c}$ we have,*

1) There exists a $j \in \mathbb{N}$ such that $T^j_{\hat{\rho},\hat{\pi}} : \mathcal{V} \to \mathcal{V}$ is a contraction mapping. Hence, there exists a unique $f^*_{\hat{\rho},\hat{\pi}} \in \mathcal{V}$ such that $T_{\hat{\rho},\hat{\pi}} f^*_{\hat{\rho},\hat{\pi}} = f^*_{\hat{\rho},\hat{\pi}}$, and for any $f \in \mathcal{V}$, $T^n_{\hat{\rho},\hat{\pi}} f \to f^*_{\hat{\rho},\hat{\pi}}$ as $n \to \infty$.

2) The fixed point $f^*_{\hat{\rho},\hat{\pi}}$ of operator $T_{\hat{\rho},\hat{\pi}}$ is the unique solution to the Bellman equation, i.e., $f^*_{\hat{\rho},\hat{\pi}} = v^*_{\hat{\rho},\hat{\pi}}$.

Lemma 4. *$v^*_{\hat{\rho},\hat{\pi}}(b)$ is monotonically increasing in b.*

Proof. By Theorem 1, we have proved $v_{\hat{\rho},\hat{\pi}}$ converges to a unique fixed point $v^*_{\hat{\rho},\hat{\pi}}$ over $T_{\hat{\rho},\hat{\pi}}$. Thus, it is sufficient to prove that $T_{\hat{\rho},\hat{\pi}}$ maintains the assumed monotonicity. Full details of the proof are presented in Appendix. □

2.3.3 Best Response Policy Characterization

As discussed in Section 2.2, our goal is to maximize server utilization from the system perspective, which is equivalent to maximizing the expected trade ratio in the market. Furthermore, the budget, which is defined through equation (2.1), increases through successful trade. These observations imply that we should characterize the best response policy not only from the single agent perspective, but also from the perspective of maximizing the expected trade ratio. We will use this goal to motivate a specific family of equilibria for our problem. Given the four facts we discussed at the beginning of this section, we then show that for the best system performance a certain simpler class of bidding functions suffice.

Lemma 5. *All servers bidding the same price within the clients' affordable range maximizes the expected trade ratio.*

Proof. The proof follows from comparing the trade ratios between the scenarios in which the server places multiple bids or a single bid. Using the four facts, the corresponding client bid distributions can be further characterized. Full details are available in Appendix. □

Motivated by Lemma 5, we characterize the best response policy by initializing the belief of server's bid distribution to be $p_{\tilde{X}_s}(k) = 1$ for some fixed k, i.e. all servers bid the same price k. For non-trivial behavior $k \geq c_{serve}$, but k can be higher than s, though not by much, i.e. $s - k \geq c_{lose}$, otherwise, the trade will become worthless; see section 2.5 for the latter.

2.3.3.1 Client's Best Response

Given the belief that all servers bid k, the value function of clients from (2.1) becomes

$$v_c(b) = \max_{x_c \in \mathcal{D} \cap [0, b+s/(1+\alpha)]} \Big(\mathbf{1}_{x_c \geq k} (\beta (v_{\hat{\rho},\hat{\pi}}(b + s - k - \alpha(k-b)^+)) + s - k) + \mathbf{1}_{x_c < k} (\beta v_{\hat{\rho},\hat{\pi}}(b) - c_{lose}) \Big)$$

By Facts 1 and 3, we conclude that the client will bid either 0 or k. If a client bids k, the trade will happen w.p. 1, and will fail otherwise. We define the following useful terms:

$v_{c\_win}(b) = \beta(v^*_{\hat{\rho},\hat{\pi}}(b + s - k - \alpha(k-b)^+)) + s - k$,

$v_{c\_lose}(b) = \beta v^*_{\hat{\rho},\hat{\pi}}(b) - c_{lose}$, $b_{c\_win} = b + s - k - \alpha(k-b)^+$, and $b_{c\_lose} = b$.

Since the budget can never go negative, we have an upper limit on a client's bid of $b + s/(1+\alpha)$. If k lies out of this range, the client will simply bid 0. Now, from Lemma 4, $b_{c\_win} \geq b_{c\_lose}$ i.e. $b \geq ((1+\alpha)k - s)/\alpha = k - \frac{s-k}{\alpha}$ implies that if $v_{c\_win}(b) \geq v_{c\_lose}(b)$, then the client should bid k. Thus, we have a lower bound on the bid as 0, and the upper bound as k. The exact bidding strategy depends on the relationship between $v_{c\_win}(b)$ and $v_{c\_lose}(b)$. A summary of the best responses of a client with budget b is:

$$x^*_c = \begin{cases} 0 & b \in [0, k - \frac{s}{1+\alpha}) \\ 0 \text{ if } v_{c\_win}(b) \leq v_{c\_lose}(b) & b \in [k - \frac{s}{1+\alpha}, k - \frac{s-k}{\alpha}] \\ k \text{ if } v_{c\_win}(b) \geq v_{c\_lose}(b) & b \in [k - \frac{s}{1+\alpha}, k - \frac{s-k}{\alpha}] \\ k & b \in (\frac{(1+\alpha)k - s}{\alpha}, \infty) \end{cases} \tag{2.4}$$

Note that when $k < s/(1+\alpha)$, all clients will bid k, which is an extreme case of a "cheap

resource." It implies the price one needs to pay is too low, as compared to the gain from the trade. We further characterize the best response function $\theta_{c,\hat{\rho}}(b)$ in the following Lemma.

Lemma 6. *$\theta_{c,\hat{\rho}}(b)$ is piecewise constant on $[0, k - \frac{s-k}{\alpha}]$ with a finite number of constant intervals.*

Proof. The proof follows by showing the difference $v_{c\_win}(b) - v_{c\_lose}(b)$ is of bounded total variation. Full details are available in Appendix. □

2.3.3.2 Server's Best Response

Given the client's best response function, we observe that under certain circumstances, the client will bid either 0 or k based on her private state. By Facts 2 and 4, we conclude that the server will again bid either 0 or k. We can refine the server's belief about the client's bid distribution as $p_{\tilde{X}_c} = (\hat{z}, 1-\hat{z})$, where $\hat{z} = \mathbb{P}(\tilde{X}_c = 0)$. Then $v_s(b) = (1-\hat{z})(\mathbb{E}_{\hat{\pi}}[\beta v_{\hat{z},\hat{\pi}}(b + x_s - c_{serve} + \alpha(x_s - \hat{B})^+)] + x_s - c_{serve}) + \hat{z}\beta v_{\hat{z},\hat{\pi}}(b)$. By Lemma 4, for $\forall b \in \mathbb{R}_+$, we have $v_s(b)$ is monotonically increasing in x_s, when $x_s \leq k$. Hence, all servers will bid k.

Given a feasible k (which we refer to as a "unified price" for both clients and servers), the discussion above lends credence to the existence of an equilibrium over the simple set of beliefs given by $\hat{z}, \hat{\pi}$. We will prove that such Mean Field Equilibrium (MFE) indeed exists in section 2.4.

2.4 Mean Field Equilibrium

The main result of this section is to show the existence of an MFE with under simple bidding strategies. Given the unified price k and the (belief) probability of bidding 0 as a client, $\hat{z}$, and the belief of the budget distribution $\hat{\pi}$, the kernel of state transitions in (2.1) is well defined. Denote the fixed point value function as $v^*_{\hat{z},\hat{\pi}}$. Taking the best response of client using (2.4), we have the following budget transitions for a generic agent before she reveals her role:

$$b[t+1] = \begin{cases} b[t] \text{ w.p. } \beta(p_s\hat{z} + p_c\mathbf{1}_{b[t]\in B_0}) \\ b[t] + s - k - \alpha(k - b[t])^+ \text{ w.p. } \beta p_c \mathbf{1}_{b[t]\in\mathbb{R}_+\setminus B_0} \\ b[t] + k - c_{serve} + \alpha(k - \hat{b}[t])^+ \text{ w.p. } \beta\hat{\pi}(\hat{b}[t])p_s(1-\hat{z}) \\ B_{init} \text{ w.p. } (1-\beta)\Psi(B_{init}) \end{cases} \tag{2.5}$$

where, $B_0 \subset \mathbb{R}_+$, in which the agents bid 0 as a client, and Ψ is the probability measure of the agent regeneration process. Set $B_{init} \subseteq \mathbb{R}_+$ is the set of possible budgets with regeneration. When $b[t]$ lies on the boundaries of B_0, by Lemma 6, a client is indifferent to bidding 0 or k. Also, the number of these boundary points in B_0 is finite, which leads to a Borel-null set in B_0. Hence, w.l.o.g. adding these points by assuming the client will bid 0 with some probability p_{tie} and k with the complementary probability will not alter the proofs in the rest of this section.

Observe that from (2.5), given the current state $b[t]$, the next state $b[t+1]$ is independent of the rest of the history. Thus, the transition kernel above defines a Markov process for the budget, and we have following Lemma.

Lemma 7. *The Markov process $\{b[t]\}_{t=0}^{\infty}$ with transition kernel (2.5) is positive recurrent and has a unique stationary distribution, $\pi = \Pi(\hat{z}, \hat{\pi})$, where $\Pi(.)$ is used to denote the mapping between $[\hat{z}, \hat{\pi}]$ and π. Furthermore, given $\hat{z}$ and $\hat{\pi}$, π is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}_+$.*

Proof. The proof of the first statement follows by showing that the one-step transition function satisfies the Doeblin condition, then using the results in [18, Chap. 12]. Then we derive the relationship between $\pi(B)$ and $\pi^{(\tau)}(B|b)$, where τ is the first regeneration time after $t = 0$ and $b(0) = b$ to prove the second statement. Details are in Appendix. □

The best response of each state $\theta_{\hat{z}}$ yields a resultant budget distribution with density π and a new value z. If $\pi = \hat{\pi}$ and $z = \hat{z}$, then we say the system is at an MFE. The main result of this section is to prove the following theorem, where $\theta_{c,\hat{z}}(\cdot)$ is the set-valued (subset of $\{0, k\}$) function

of client bids as a function of its budget. More formally, $\theta_{c,\hat{z}} : \mathbb{R}_+ \to \{0, k\}$ is the client's best response correspondence which maps the budget to a binary choice of bids, either 0 or k.

Theorem 2. *Define* $\gamma(\hat{z}, \hat{\pi}) \triangleq \pi_{\hat{z},\hat{\pi}}(\theta^{-1}_{c,\hat{z}}(0)), \forall \hat{z} \in [0, 1]$*, where* $\theta^{-1}_{c,\hat{z}}(0)$ *is the lower inverse of* $\theta_{c,\hat{z}}(\cdot)$ *at* 0*. There exists an MFE* $(z, k, \theta_{\hat{z}}, \pi)$*, such that* $\pi = \Pi(\hat{z}, \hat{\pi})$ *and* $z = \gamma(\hat{z}, \hat{\pi})$*.*

To prove Theorem 2, we need to show that $\gamma(\cdot)$ and $\Pi(\cdot)$ have a fixed point, i.e., $\gamma(z, \pi) = z$ and $\Pi(z, \pi) = \pi$ for some $z \in [0, 1]$ and some (continuous) probability density π on $\mathbb{R}_+$. We use the Schauder Fixed Point Theorem, which requires that $\gamma(\cdot)$ and $\Pi(\cdot)$ are continuous in $\hat{z}$ and $\hat{\pi}$, and that the closure of their range spaces are compact.

It is straightforward to show that $\Pi(.)$ is a continuous map. Now, from Section 2.3 the best response function $\theta_{c,\hat{z}}(\cdot)$ is a set-valued function with a range containing all non-empty subsets of $\{0, k\}$. Then we show $\theta_{c,\hat{z}}(\cdot)$ is upper hemicontinuous in $\hat{z}$, and combine with the continuity of π. Finally, we prove the single point inverse (lower inverse) of the upper hemicontinuous set-valued function $\theta_{c,\hat{z}}(\cdot)$ is a subset that consists of finite number of continuous pieces, which leads to the continuity of $\gamma(\cdot)$.

Now, $\hat{z}$ lies in the closed interval $[0, 1]$, which is compact and convex. We then consider the range of the mapping Π. Now, $\mathbb{R}_+$ is a $\sigma-$compact metric space, and since Π is a continuous real valued function, it can be easily shown that the closure of the image of the mapping Π is compact in the norm space of $\hat{\pi}$ directly using Theorem 13 in [19]. This completes the requirements of the Schauder Fixed Point Theorem. The steps of the proof are given below.

Lemma 8. $v^*_{\hat{z},\hat{\pi}}$ *is Lipschitz continuous in* $\hat{\pi}$ *and in* $\hat{z}$*.*

Proof. The proof follows using the properties of the contraction mapping $T^j_{\hat{z},\hat{\pi}}$ in Theorem 1. Details are presented in Appendix. □

Theorem 3. $\theta_{c,\hat{z}}(\cdot)$ *is upper hemicontinuous in* $\hat{z}$*.*

Proof. Given $\hat{z}$ and k, we can rewrite $v^*_{\hat{z},\hat{\pi}}$ in a different way such that it can be represented as a increasing piecewise linear convex function. The proof holds by applying Berge's Maximum Theorem. Details are presented in Appendix. □

Theorem 4. $\Pi(\hat{z}, \hat{\pi})$ *is continuous in* $\hat{\pi}$ *and* $\hat{z}$*.*

Proof. The key idea of the proof is using the Portmanteau Theorem to show for any uniform converging sequence $\hat{z}_n \to \hat{z}$, a sequence of density functions $\hat{\pi}_n \to \hat{\pi}$ and any open set B, $\liminf_{n\to\infty} \pi_n(B) \geq \pi(B)$, where $\pi_n = \Pi(\hat{z}_n, \hat{\pi}_n)$ and $\pi = \Pi(\hat{z}, \hat{\pi})$. Details are presented in Appendix. □

Theorem 5. $\gamma(\hat{z}) \triangleq \pi_{\hat{z},\hat{\pi}}(\theta^{-1}_{c,\hat{z}}(0))$ *is continuous in* $\hat{z}$*.*

Proof. We prove this by showing that $\theta^{-1}_{c,\hat{z}}(0)$ is a continuity set for $\pi_{\hat{z},\hat{\pi}}$. See Appendix for details. □

2.5 Simulation

Here we conduct Monte Carlo simulations of all three models of our market system with one million virtual agents.

Before presenting simulation results, we first introduce the parameter settings. We use a regeneration factor $\beta = 0.98$; we assume that agents have an equal probability being clients and servers, i.e., $p_s = p_c = 0.5$; $\alpha = 1.1$, which is the penalty parameter for overdraft; $s = 8$ and $c_{serve} = 6$ make the trade generate reasonable amount of value; $c_{lose} = 0.5$ captures the client disappointment when the trade fails; $\Psi(B_{init}) = U[0, 5]$, new agents come with limited amount of budget, which helps us better analyze the differences among three models; and price $k = 7$ in order to balance the benefits between servers and clients through the trade. Later on in this section, we will show how different prices and initial budgets affect the equilibrium trade ratio in the bank-loan model.

Fig. 2.2 shows the convergence of the MFE bid probability belief, z. In the hard constraint model, clients cannot afford the price $k = 7$, so all clients bid 0 and the system freezes. However, in the bank-loan model and peer-loan model, the system gradually ramps up through different borrowing mechanisms and z dramatically reduces when the system attains more and more wealth through successful trades. Fig. 2.3 presents the CDFs of budget at MFE across the three models, which indicates agents are wealthier in the peer-loan model than in the bank-loan model at MFE (as the bank extracts some of the surplus), while the budget distribution is similar to the initial

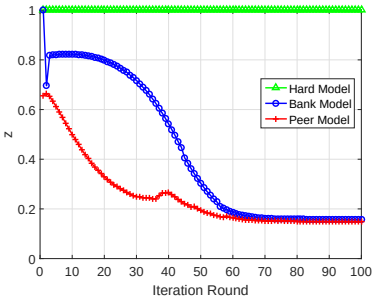

Figure 2.2: Convergence of the belief, z. Reprinted with permission from [1].

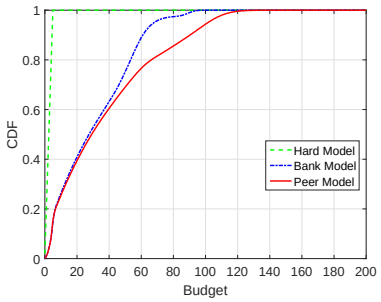

Figure 2.3: CDF of budget at MFE. Reprinted with permission from [1].

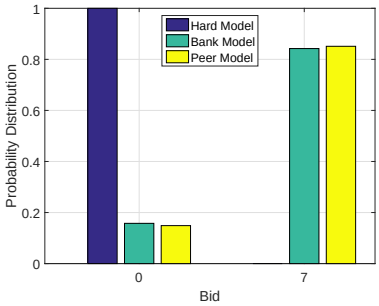

Figure 2.4: Client bid distribution at MFE. Reprinted with permission from [1].

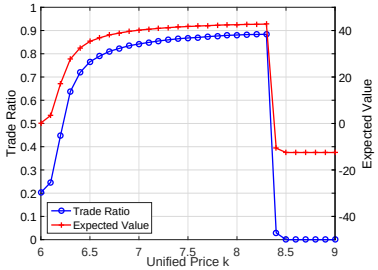

Figure 2.5: Bank-loan model: trade ratio and expected value vs. k, $\Psi(B_{init}) = U[0, 5]$. Reprinted with permission from [1].

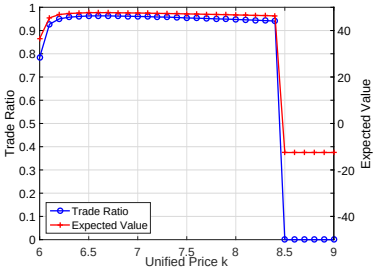

Figure 2.6: Bank-loan model: trade ratio and expected value vs. k, $\Psi(B_{init}) = U[3, 8]$. Reprinted with permission from [1].

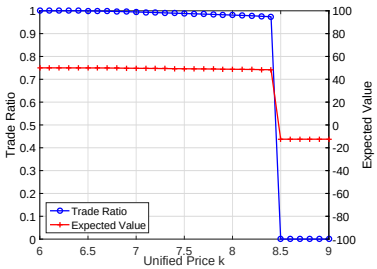

Figure 2.7: Bank-loan model: trade ratio and expected value vs. k, $\Psi(B_{init}) = U[5, 10]$. Reprinted with permission from [1].

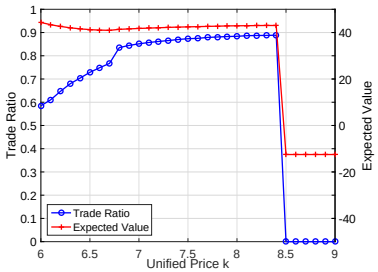

Figure 2.8: Peer-loan model: trade ratio and expected value vs. k, $\Psi(B_{init}) = U[0, 5]$

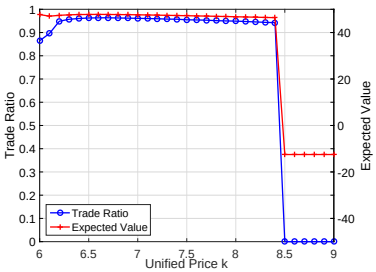

Figure 2.9: Peer-loan model: trade ratio and expected value vs. k, $\Psi(B_{init}) = U[3, 8]$

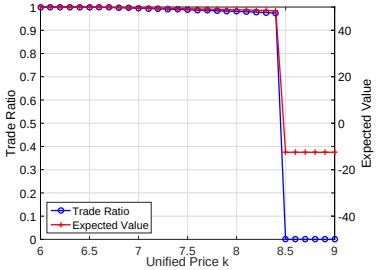

Figure 2.10: Peer-loan model: trade ratio and expected value vs. k, $\Psi(B_{init}) = U[5, 10]$

values in the hard constraint model. Fig. 2.4 shows the binary bid distributions of clients at MFE, which verifies the results in Section 2.3. We further evaluate two important statistics, namely the trade ratio and the expected value in Table 2.1, where the expected value is calculated according to $\mathbb{E}_{\Psi(B_{init})}[v^*_{MFE}]$. As we mentioned earlier, the higher the trade ratio the market system achieves,

Table 2.1: Trade Ratio and Expected Value. Reprinted with permission from [1].

Initial Budget	$\Psi(B_{init}) = U[0,5]$			$\Psi(B_{init}) = U[5,10]$		
Model	Hard	Bank	Peer	Hard	Bank	Peer
Trade Ratio	0%	84.3%	85.2%	97.7%	99.4%	99.5%
Value	-12.49	40.14	41.74	48.53	49.6	49.7

the higher resource utilization it ends up with. Given the unified bid of server and binary bids of client, the expected trade ratio is captured by $1-z$ at MFE. The empirical trade ratio among the one million agents matches this quantity in all cases. Intuitively, a higher trade ratio implies a higher expected value, which is verified in Table 2.1. For the sake of comparison, we also append the results of sufficient initial budget case, i.e. $\Psi(B_{init}) = U[5,10]$, in Table 2.1 as well. We observe that with a sufficient initial budget, the hard constraint model also achieves a reasonably high trade ratio at MFE, and the statistics of bank-loan model catches up with the peer-loan model, and both reach 100% trade ratio.

In the simulations above, we set the unified price $k = 7$ and the initial budget, B_{init}, to be uniform distributed in $[0,5]$. Next, we will show how these two attributes affect the trade ratio of the market system. Figures 2.5, 2.6, 2.7, 2.8, 2.9 and 2.10 illustrate the trade ratio and the corresponding expected value versus different unified prices with different initial budget distributions under the bank-loan and peer-loan models. Recall that $s = 8$ and $c_{serve} = 6$, which means that the reasonable range of k lies in $[6, 8+\epsilon]$, given the client is allowed to overdraw his budget to some extent. We observe that for B_{init} in $[0,5]$, the trade ratio increases in k; for B_{init} in $[3,8]$, the trade ratio first increases then decreases; for B_{init} in $[5,10]$, the trade ratio decreases in k. An intuitive explanation of these results is as follows. When most agents have a sufficient budget, lower price yields higher trade ratio. However, when most agents have a limited budget, it is better to set higher prices to aggregate the wealth at fewer agents, which then can obtain service without taking a loan from the bank or the peer server. It is interesting to note that in the bank-loan model, the expected value follows the corresponding trend in trade ratio, whereas in the peer-loan model, maximum value and maximum trade do not correspond. The reason for this is the interest received

by the server in the peer-model makes up for the lack of a high trade ratio at low prices.

2.6 Case Study: Photovoltaic Market

We consider a PV energy sharing market in which the agents could be householders or small business owners. During the hours with high sunshine, the system can supply the electricity consumption of an agent. Moreover, it generates extra energy that can be fed back to the grid. However, in rainy or overcast weather, solar panels only produce 10%-25% of their rated capacity [20]. This creates the opportunity to share the extra solar energy between locations with good and bad weather.

However, electricity is a product that is hard to differentiate between different producers. In our context, one cannot identify specific electron transactions between matched servers and clients. However, since we build our market on top of the existing grid, servers inject power into the grid and clients demand power from the grid. This gives us the flexibility of conducting trades between any set of locations that are connected on the grid. For our case study, we pick two such cities, Dallas and Houston, since both belong to Texas interconnection and are also associated with retail competition.

To validate the model on a realistic system, we conduct our market on a 2000-bus synthetic Texas grid [14] built using publicly available data of the actual U.S. transmission system in 2016, the topology of which is shown in Fig 2.11. Collecting hourly historical weather data of the two cities in 2016 from [21], we found that approximately 38% of the time, both cities have sunny days, and about 20% of the time, both have cloudy days (hence do not need the market).

We are interested in situations in which the two cities have different weather so that they could share energy. The probabilities of these events are 0.1 (Dallas-bad, Houston-good) and 0.32 (Dallas-good, Houston-bad). Normalizing these values, we obtain two type of agents: Dallas agents that are clients about 23% of the time and are servers 77% of the time, and Houston agents who are exactly the opposite. Given the heterogeneous agent types, we have a slightly different setup from the previous section with homogeneous agents. Also, given the regional weather effects, all Dallas agents are of one type, while all Houston agents are of another type. We will show the

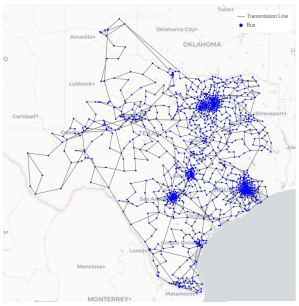

Figure 2.11: Synthetic Texas Grid

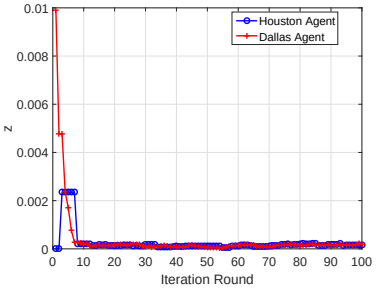

Figure 2.12: Peer-loan model: Convergence of the Coupling Belief

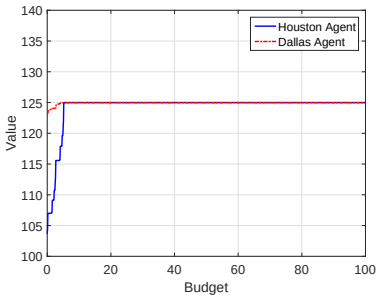

Figure 2.13: Peer-Loan model: MFE Value Function

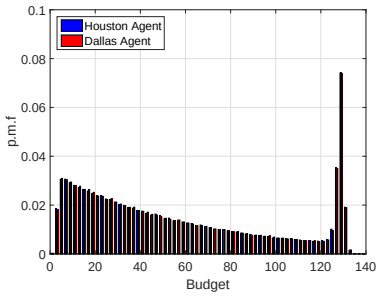

Figure 2.14: Peer-Loan model: MFE Budget Distribution

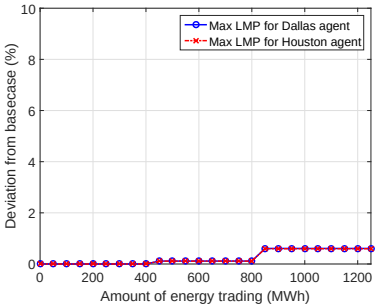

Figure 2.15: Impact on LMPs for Dallas servers

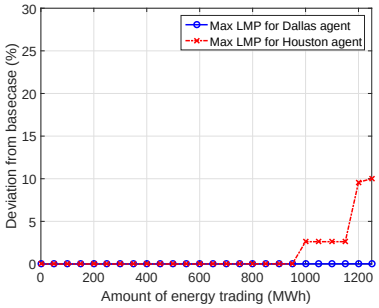

Figure 2.16: Impact on LMPs for Houston servers

convergence of the coupling beliefs and the consistency of budget distributions across agent types under the chosen market price, k. The setting resembles the one in [22], with multiple agent types that undergo exogenous changes.

We choose parameters for our analysis based on readily available data. As we discussed above, p_s and p_c for Dallas (Houston) agents are 77%(23%) and 23%(77%) respectively. Given the average electricity price to be 10 cents/kwh, we set the value of surplus $s = 10$. However, an accurate unit cost for rooftop solar energy is not well established, since it varies by the installation fee, the maintenance fee, the government subsidy, etc., and we use 5 cents/kwh as a conservative estimate [23], which yields $c_{serve} = 5$. We choose the unified price $k = 7.5$ to balance the benefits of the trade between clients and servers. A sufficient budget initialization $\Psi(B_{init}) = U[5, 10]$ is used.

Fig. 2.12 shows the convergence of the coupled beliefs in loan model. We see that the agents

quickly get involved in trades, since the peer-loan ramps up the system fast.Also, the higher probability of being a client for Houston agents makes their value of budget lower, which is shown in Fig. 2.13. Fig. 2.14 illustrates the consistency of budget distributions between agent types. With the balanced price chosen, the consistency of budget distributions implies a high trade ratio. Indeed, the trade ratio turns out to be 99.9%, which matches our discussion of the sufficient budget case in Section 2.5.

We next determine the feasibility of our system operating in conjunction with the existing (primary) market, whose bus-prices are determined by the solution of the DC optimal power flow (OPF) solution with a given demand. Hence, we first determine the baseline location marginal prices (LMPs) when only the primary market exists, and compare with the situation that arises under significant penetration of the our A2S (secondary) market. The physical impact of our secondary market is to inject power at certain buses, and we aim to determine both the feasibility of doing so without violating grid constrains, as well as the impact on the LMPs in the primary market. Thus, we compare the DC optimal power flow (OPF) solution of the original network (baseline) with the case when around one million customers from the two cities (split equally) are involved in our market with different trade ratios.

The average rooftop PV system size in the U.S. is $5kW$ [24]. Given that the average electricity usage is roughly $2.5kWh$ per hour in the daytime [25], a server is able to provide $2.5kWh$ extra energy with one hour full sunshine, which is sufficient to supply a typical client. This implies roughly $1250MWh$ maximum extra energy injection at servers' buses when scaled by one million participants. Fig 2.15 and 2.16 explicitly show the maximum LMP fluctuations are below 10% of the value of the baseline LMP ($\$19.6/MWh$) when different amounts of energy are traded. Further notice that the potential energy penetration through the grid due to the market here is $5000MWh$ ($5kWh*1$ million), which is roughly one quarter of the residential load of the entire Texas system in a typical peak hour during 2016. Thus, the secondary market is both physically possible, and has minimal disruption on existing markets even with significant penetration.

Finally, we evaluate our market system using weather traces in the year 2016 [21]. We con-

servatively define daytime to be the interval between one hour after sunrise to one hour before sunset, and found that in 1379 time periods (hours) a Houston agent is a client, whereas in 402 time periods a Dallas agent is a client. One potentially saves $(1379*(10-7.5)+402*(7.5-5))*2.5/100*99.9\% \approx \$111.3/year$ through our market. Note that the price k is chosen to balance the benefits of the trade, so the savings are consistent between different types of agents. However, for a Net Meter user, the grid usually pays at the rate $5¢$/kwh [26], which gives zero profit as a server and deficits as a client, i.e. $-1379*10/100 = -\$137.9/year$ for an Houston agent and $-402*10/100 = -\$40.2/year$ for an Dallas agent. The effective returns from the market are thus $\$249.2$ and $\$151.5$ for the two agent types, respectively.

2.7 Extension

So far we have discussed the market system between two locations and showed the operation point of the underlying power grid will not deviate too much even with large amount of energy being traded. However, the power grid is a large scale networked system with thousands of buses connected with each other. One geographic location often imply one bus in power system. Even for big cities, several buses could handle the transmission requirement. What if we scale our secondary market across the grid? With large group of agents in multiple locations participating into the market and trading simultaneously with each other based on different weather conditions, the overall extra transmissions due to the market may exceed the threshold of current operation point and cause significant changes in LMPs, which is not observed in Fig 2.15 and 2.16.

In this section, we first show the possibility of changing LMPs dramatically by injecting cheap PV energy to the network through a simple three bus case. Then we briefly describe the results in [27], which describes the polyhedral structure of LMP. Finally, we introduce a possible extension of our proposed market to price anticipating scenarios based on the geometry of LMPs.

2.7.1 A Three-Bus Example

Fig 2.17 shows a three-bus system with one traditional generator on each bus. The maximum capacity of all traditional generators are $450MW$. The cost curves are linear but different among

generators. A load $l = 360MW$ is attached to bus 3. The impedance of all three transmission lines are the same, however, there is a capacity limit on the transmission line between bus 2 and bus 3. We set this capacity limit to be $150MW + \epsilon$ for the purpose of showing the boundary case explicitly. Suppose a PV generator is installed on bus 2, of which the power output is stochastic based on different whether conditions.

If there is no injection for the PV generator, the DC Optimal Power Flow (OPF) is shown in blue text. In this case, generator g_1 takes care of all the demand and $LMP_1 = LMP_2 = LMP_3 = \$10/MWh$, which is the marginal cost of g_1. The system reaches a boundary point when the PV injection equals to $90MW$, the OPF of which is shown in yellow text. When the PV injection is beyond $90MW$, we solve the OPF with $\epsilon = 0.5$ and get $LMP_1 = \$10/MWh, LMP_2 = \$0/MWh, LMP_3 = \$20/MWh$. Hence, the operation point of the system is changed due to the PV injection.

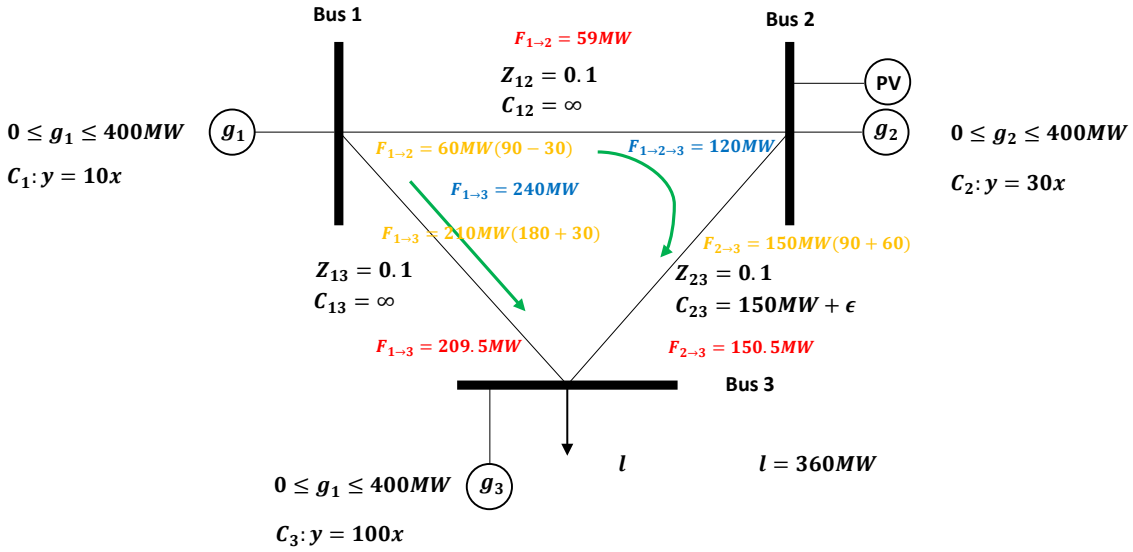

Figure 2.17: A three bus example

2.7.2 The Geometry of Locational Marginal Prices

Although the LMPs may vary due to extra power injection, it is not intractable. It is shown in [27] that the state-space of a power system represented by a DC power flow model can be partitioned into polyhedral price regions in which the LMPs are constant. Here, we use a simple two bus example as shown in Fig 2.18 to illustrate the idea. In this two bus network, we have two

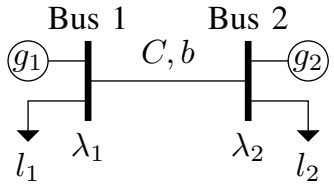

Figure 2.18: A two bus network

generators g_1 and g_2 on bus 1 and bus 2 with maximum generation capacity G_1 and G_2 respectively. Again the cost curves are linear and the marginal cost of g_1 is smaller than g_2. The loads are l_1 and l_2. The line capacity is C, assuming $C < G_i$. Given the setup, the feasible load set is shown in Fig 2.19. According to the DC OPF solutions, the LMPs on the two buses turn out to be constant within the three regions S_1, S_2 and S_3. Specifically, in region S_3, the transmission line is congested. One could further derive the closed form expressions for the LMPs in each region. The technical details regarding this work refer to [27].

2.7.3 Market Model Extension

Given the geometry of LMPs discussed in previous section, it is possible to extend our market model into a more general system, in which agents from different locations share their extra PV generation across the Smart Grid. Our current market model assumes the system operates in a single polyhedral region of LMP, i.e. agents are price taking. However, the transitions between different LMP regions can be fully captured by the trade ratios among different locations, which

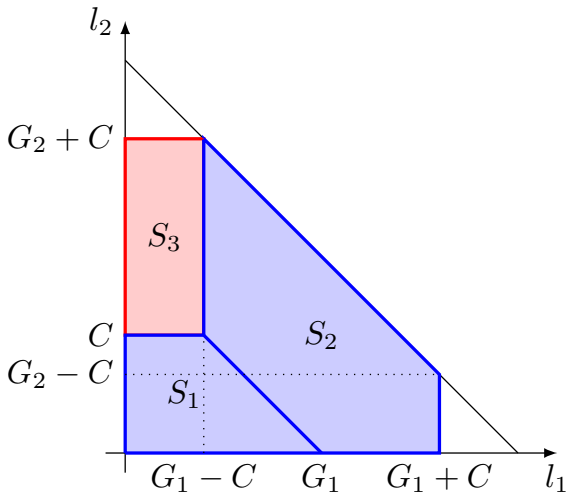

Figure 2.19: Polyhedral price regions for the two bus example

can be treated as a function of $\mathbf{z}$. Here $\mathbf{z}$ is a vector, the entries of which define the trade ratio between each location-wise client-server pair. Each agent in such extended system maintains a belief of the distribution of $\mathbf{z}$. The cost of service as a server will be different in different polyhedral regions, which can be revealed by different ks.

2.8 Conclusion

We considered the problem of market equilibria that arise in an energy sharing economy where agents change roles frequently from provider (server) to consumer (client). We developed a model of bilateral trade under which consumers and providers are matched randomly with each other. Under the MFG setting, we showed that the MFE consists of a single price bid by both client and server. We conducted numerical evaluations to study the effects of different equilibrium prices on trade ratios, and showed in a case study that significant savings are possible in a rooftop PV setting. We proposed a reasonable extension of the market system accounting for the physical constraints in the Smart Grid given the geometry of LMPs.

In next chapter, we will discuss the resource management problem from demand side perspective. Again, we first design a general market mechanism to steer the system toward a desirable

equilibrium and conduct performance evaluation via realistic data in the Smart Grid.

3. MEAN FIELD GAMES IN NUDGE SYSTEMS FOR SOCIETAL NETWORKS*

3.1 Introduction

There has recently been much interest in understanding *societal networks,* consisting of interconnected communication, transportation, energy and other networks that are important to the functioning of human society. These systems usually have a shared resource component, and where the participants have to periodically take decisions on when and how much to utilize such resources, but with indirect knowledge of the aggregate utilization of the shared resource. Research into these networks often takes the form of behavioral studies on decision making by the participants, and whether it is possible to provide incentives to modify their behavior in such a way that the society as a whole benefits [28, 29].

Our candidate application in this chapter is that of a Load Serving Entity (LSE) or a Load Aggregator (LA) (*e.g.,* a utility company) trying to reduce its exposure to daily electricity market volatility by incentivizing demand response in a Smart Grid setting. The reason for our choice is the ready availability of data and reliable models for the cost and payoff structure that enables a realistic study. The data used in this chapter was obtained from the Electric Reliability Council of Texas [30], an organization that manages the wholesale electricity market in the state. The price shows considerable variation, and peaks at about 5 PM each day, which is when maximum demand occurs. A major source of this demand in Texas is air conditioning, which in each home is of the order of 30 kWh per day [31]. Incentivizing customers to move a few kWh of peak-time usage to the sides of the peak each day could lead to much reduced risks of peak price borne by the LSE. Such demand shaping could also have a positive effect on environmental impact of power plant emissions, since supplying peak load is associated with inefficient electricity generation.

As an example, we take the baseline temperature setpoint as 22.5°C, and consider a customer

*Republished with permission of ACM, from Mean Field Games in Nudge Systems for Societal Networks, Jian Li, Bainan Xia, Xinbo Geng, Hao Ming, Srinivas Shakkottai, Vijay Subramanian, Le Xie, Transactions on Modeling and Performance Evaluation of Computing Systems, Volume 3, 2018; permission conveyed through Copyright Clearance Center, Inc.

that every day increases the setpoint by $1°$C in $5-6$ PM and decreases the setpoint by $0.5°$C in the off-peak times. We will see later that even such a small change of the setpoint of the AC, the incentives to the users (a group of fifty homes) can be tuned to achieve different trade-offs: either just a utility increase for the users, just savings to the LSE (of the order of eighty dollars in our case) or any objective that chooses some appropriate mix of the two. This result is under the implicit assumption that the LSE in question is a price-taker so that changes in its demand profile are assumed not to perturb the prices. The shifting of daily energy usage could potentially cause a small increase in the mean and deviation of the internal home temperature, which is a discomfort cost borne by the customer. In our approach, the LSE awards a number of "Energy Coupons" to the customer in proportion to his usage at the non-peak times, and these coupons are used as tickets for a lottery conducted by the LSE. A higher number of coupons would be obtained by choosing an option that potentially entails more discomfort, and would also imply a higher probability of winning at the lottery. Since the customers do not observe the variation of day-ahead prices on a day-to-day basis nor do they see the aggregate demand at the LSE, the lottery scheme serves as a light-weight and easy to implement mechanism to transfer some of this information over to the customers by coupling them. We will explicitly demonstrate the advantage of this coupling over an individual incentive scheme (a fixed reward for peak time reductions) that serves as a benchmark for the comparison.

In our analytical model, each agent has a set of actions that it can take in each play of a repeated game, with each action having a corresponding cost. Higher cost actions yield a higher number of coupons. Agents participate in a lottery in which they are randomly permuted into groups, and one or more prizes are given in each group. The state of each agent is measured using his surplus, which captures the history of plays experienced by the agent, and is a proxy to capture his interest in participating in the incentive system. A win at the lottery increases the surplus, and a loss decreases it. Furthermore, we assume that the agent has a prospect incremental utility function that is increasing and concave for positive surplus and convex for negative surplus. This prospect theory model captures decision making under risk and uncertainty for agents. Any agent could

depart from the system with a fixed probability independent of the others, and a departing agent is replaced by a new entrant with a randomly drawn surplus. The main question we answer in this chapter then is how would agents decide on what action to take at each play? Having answered this we also comment on impact of this on the sum total value of the agents, the return to the system and the trade-off between these two quantities provided by our proposed scheme.

3.1.1 Prospect Theory

Most previous studies account for uncertainty in agent payoffs by means of *expected utility theory* (EUT). Here, the objective of the decision maker is to maximize the probabilistically weighted average utilities under different outcomes, and it is assumed that he/she is capable of making arbitrarily complex deductions. However, EUT does not incorporate observed behavior of human agents, who exhibit bounded rationality and can take decisions deviating from the conventional rational agent norm. For example, empirical studies have shown that agents ascribe high weights to rare, positive events (such as winning a lottery) [32].

Prospect theory (PT) [32–35] is perhaps the most well-known alternative theory to EUT. It was originally developed for binary lotteries [32] and later refined to deal with issues related to multiple outcomes and valuations [33]. This Nobel-prize-in-economics-winning theory has been observed to provide a more accurate description of decision making under risk and uncertainty than EUT. There are three key characteristics of PT. First, the value function is concave for gains, convex for losses, and steeper for losses than for gains. This feature is due to the observation that most (human) decision makers prefer avoiding losses to achieving gains. Thus, the value function is usually S-shaped. Second, a nonlinear transformation of the probability scale is in effect, *i.e.*, (human) decision makers will overweight low probability events and underweight high probability events. The weighting function usually has an inverted S-shape, *i.e.*, it is steepest near endpoints and shallower in the middle of the range, which captures the behaviors related to risk seeking and risk aversion. Finally the third, the framing effect is accounted for, *i.e.*, the (human) decision maker takes into account the relative gains or losses with respect to a reference point rather than the final asset position. As PT fits better in reality than EUT based on many empirical studies, it has been

widely used in many contexts such as social sciences [36, 37], communication networks [38–40] and smart grids [41, 42]. Since we study equilibria that arise through (human) agents' repeated play in lotteries, we use PT as opposed to EUT to account for agent-perceived value while taking decisions.

3.1.2 Mean Field Games

The problem described is an example of a dynamic Bayesian game with incomplete information, wherein each player has to estimate the actions of all his potential opponents in the current lottery (and in the future) without knowing their surpluses, play a best response, and update his beliefs about their states of surplus based on the outcome of the lottery. However, since the set of agents is large and, from the perspective of each agent, each lottery is conducted with a randomly drawn finite set of opponents, an accurate approximation for any agent is to assume that the states of his opponents (and hence actions) are independent of each other. This is the setting of a Mean Field Game (MFG) [5, 43, 44], which we will use as a framework to study equilibria in societal networks. Here, the system is viewed from the perspective of a single agent, who assumes that each opponent's action would be drawn independently from an assumed distribution, and plays a best response action. We say that the system is at a Mean Field Equilibrium (MFE) if this best response action turns out to be a sample drawn from the assumed distribution.

3.1.3 Demand Response in Deregulated Markets

Demand Response is the term used to refer to the idea of customers being incentivized in some manner to change their normal electricity usage patterns in response to peaks in the wholesale price of electric power [45]. Many methods of achieving demand response exist, including an extreme one of turning off power for short intervals to customers a few times a year if the price is very high. Customers expect a subsidy in return, often in terms of a reduced electricity bill.

3.1.4 Main Results

Our objective in this chapter is to design and analyze a system that can provide greater ability for the LSE to realize a desired combination of profit and user value by incentivizing user behavior.

Our main contributions in this chapter are as follows:

(1) We propose a mean field model to capture the dynamics in societal networks. Our model is well suited to large scale systems in which any given subset of agents interact only rarely. This kind of system satisfies a chaos hypothesis that enables us to use the mean field approximation to accurately model agent interactions. The state of the mean field agent is its surplus, and the agent must choose from a finite set of actions based on its surplus and its belief about the action distribution of other agents. The state (surplus) evolves according to a Markov process that increases by winning and decreases by losing at the lottery. Our mean field model of societal networks is quite general, and can be applied to different incentive schemes that are currently being proposed in the field of public transportation and communication network usage.

(2) We conduct a simulation analysis of our scheme under an accurate measurement-based model of the daily usage of electricity in each hour in Texas. We also use the data on wholesale electricity prices during the interval to calculate what times of day would yield the best returns to rewards. We show that under several intuitive coupon allocation options and a $\$15$ weekly reward (lottery prize), customers would change their AC setpoints (as small as $1°$C each day) and each week, the LSE gains a benefit of the order of a $\$80$ over a cluster of 50 homes. While doing so, we also numerically verify that our model satisfies conditions needed for passage to the mean field.

(3) We conduct comparative studies between a benchmark scheme that returns a fixed reward per action (assuming that each customer maximizes his return) versus the lottery scheme, and show that the lottery scheme can outperform the fixed reward scheme by about 100% in terms of total value to the users, and about 20% in terms of profit to the LSE. We also explore the relation between LSE profit and user value for both schemes, and show that as one changes the reward values and coupon allocations, the lottery scheme bounds the achievable region of the benchmark scheme in a Pareto-sense: it is better able to attain a desired combination of user value and LSE profit, and includes combinations unachievable by the individual incentive scheme.

(4) We develop a characterization of a lottery in which multiple rewards can be distributed, but with each participant getting at most one by withdrawing the winner in each round. Each lottery is

played amongst a cluster of M agents drawn from a random permutation of the set of all agents. While the exact form of the lottery is not critical to our results, we present it for completeness.
(5) We characterize the best response policy of the mean field agent, using a dynamic programming formulation. We find that under our assumptions, the value function is continuous in the action distribution, but that multiple actions could turn out to be best responses. Hence, an agent also needs to choose some randomization method across such equal-value actions. If the value function is super-modular, sub-modular or S-shaped (under the prospect-based utility function), the action choices map to surplus intervals, with two actions being of equal value at each interval boundary.
(6) The probability of winning the lottery defines the transition kernel (along with the regeneration distribution) of the Markov process of the surplus, and hence maps an assumed distribution across competitors states to a resultant stationary distribution. We show the existence of a fixed point of this kernel, which is the MFE, by using Kakutani's fixed point theorem. Essentially, the system is a map between the space consisting of the triple of an assumed action distribution, a randomized policy and a surplus distribution back to itself, and our result is to show this map has a fixed point. Our proof of the existence of MFE does not depend on the shape of the utility function, which can be quite general. Since we have a discrete action and state space, showing a fixed point in the space of such triples is quite intricate.

A 2-page conference abstract that includes a high-level overview of our results developed herein was presented to practitioners in [46].

3.1.5 Related Work

In terms of the MFG, our framework is based on work such as [6, 47, 48]. In [6] the setting is that of advertisers bidding for spots on a webpage, and the focus is on learning the value of winning (making a sale though the advertisement) as time proceeds. In [47], apps on smart phones bid for service from a cellular base station, and the goal is to ensure that the service regime that results has low per-packet delays. In both works, the existence of an MFE with desired properties is proved. In [48], the objective is to incentivize truthful revelation of state that would allow for

optimal resource allocation in a device-to-device wireless network. The state space is discrete, and the focus is on the exploration of truthful dynamic mechanisms in the mean field regime. However, unlike that work, we focus on a lottery-based allocation in this chapter. The lottery is simple and well-established, and has been successfully applied in a variety of existing nudge systems. Thus, our goal is to analyze this well-established mechanism, rather than designing new ones. Also, unlike the previous work, all of which focussed on pure strategy equilibria, our current work has a more complex state space and pure strategy equilibria may not exist due to the non-uniqueness of best responses. Hence, we seek a mixed strategy equilibrium, which necessitates a different proof technique.

Nudge systems are typically designed and used to encourage socially beneficial behaviors and individually beneficial behaviors. For instance, lottery schemes are widely used in practice to incentivize good behavior, e.g., to combat (sales) tax evasion in Brazil ([49]), Portugal ([50]), Taiwan ([51]), and for Internet congestion management ([52]). Similarly, [28, 29] provide experimental results on designing lottery-based "nudge engines" to provide incentives to participants to modify their behaviors in the context of evenly distributing load on public transportation. In another scheme, [53] study the impact of nudging on social welfare by sending one-year home energy reports to participants and using multiple price lists to determine participants' willingness to stay in the system for the next year. Our system is a form of nudge engine, but our focus is on analytical characterization of system behavior and attained equilibria with large number of customers with repeated decision-making. We aim to design incentive schemes to modify customer behavior such that the system as the whole benefits from the attained equilibrium.

Our idea of offering coupons for reduced electricity usage at certain times is based on one presented in [54], which suggests offering incentives to coincide with predicted realtime price peaks. An experimental trial based on a similar idea is described in [55], in which the focus is on designing algorithms to coordinate demand flexibility to enable the full utilization of variable renewable generation. In [56], this kind of system is modeled as a Stackelberg game with two stages: setting the coupon values followed by consumer choice. The decision making model in all

the above research is myopic. The authors of [57] study demand-response as trading off the cost of an action (such as modifying energy usage) against the probability of winning at a lottery in terms of a mean field game. However, the game is played in a single step according to their model, and there is no evolution of state or dynamics based on repeated play. Further, their conception of the mean field equilibrium is that the mean value of the action distribution (not the distribution itself) is invariant. Unlike these models, we are interested in characterizing repeated consumer choice with state evolution when the number of customers is large, and identifying the action distribution and benefits (if any) of the resulting equilibrium.

A rich literature studies lottery schemes, and here we can only hope to cover a fraction of them that we see most relevant. In this chapter, we model lotteries as choosing a random permutation of the M agents participating in it, and picking the first K of them as winners, with the distribution on the symmetric group of permutations of $\{1, \cdots, M\}$ being a function of the coupons assigned to the different actions. Assuming that different actions yield different numbers of coupons, we will choose the distribution such that more coupons results in a higher probability of winning. There are various probabilistic models on permutations in the ranking literature [58, 59], Here we use the popular Plackett-Luce model [60] to implement our lotteries. While the Plackett-Luce model is used for concreteness, other probabilistic models on permutations such as the Thurstone model [59] can also be used with the number of coupons as parameters of the distribution as long as more coupons results in a higher probability of winning.

The monotonicity properties in rewards are shared with other literature, such as [61,62]. In particular, [62] focuses on the existence of the mean field equilibrium when playersą́ welfare depends on the distribution of other players actions. However, this previous work studies the existence of pure strategy equilibria, whereas our discrete state and action spaces requires consideration of a mixed strategy equilibrium. The proof of the existence of this equilibrium is one of the major technical contributions of this chapter.

3.1.6 Organization

The chapter is organized as follows. In Section 3.2, we introduce our mean field model. We then conduct simulation-based numerical studies in Section 3.3, on utilizing our framework in the context of demand response in electricity markets. In Section 3.4 we develop a characterization of a lottery in which multiple rewards can be distributed, but with each participant getting at most one by withdrawing the winner in each round. We discuss the basic property of the optimal value function in Section 3.5. The existence of MFE is considered in Section 3.6. We characterize the best response policy of the mean field agent, using a dynamic programming formulation in Section 3.7. We conclude in Section 3.8. To ease exposition of our results, all proofs are relegated to the Appendix.

3.2 Mean Field Model

We consider a general model of a societal network in which the number of agents is large. Agents have a discrete set of actions available to them, and must take one of these actions at each discrete time instant. The actions result in the agents receiving coupons, with higher cost actions resulting in more coupons. The agents are then randomly permuted into clusters of size M, and a nudge is provided via a lottery that is held using the coupons to win real rewards. Thus, agents must take their actions under some belief about the likely actions, and hence the likely coupons held by their competitors in the lottery.

Figure 3.1 illustrates the mean field approximation of our model. We provide justification for the mean field approximation in the discussion at the end of this section. The diagram is drawn from the perspective of a single agent (w.l.o.g, let this be agent 1), who assumes that the actions played by each of his opponents would be drawn independently of each other from the probability mass function ρ. In this section, we will introduce the notation, costs and payoffs of the agent, and provide a brief description of the policy space and equilibrium.

Time: Time is discrete and indexed by $k \in \{0, 1, \cdots\}$.

Agents: The total number of agents is infinite, and we consider a generic agent 1 who in each

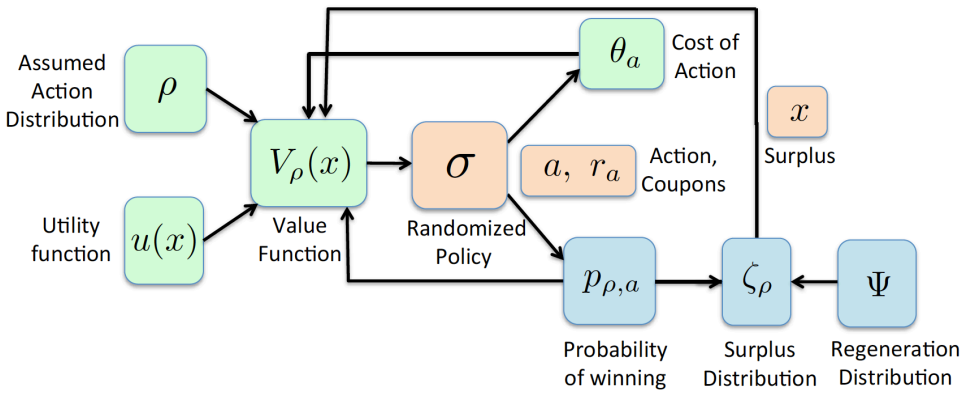

Figure 3.1: Mean Field Game. Reprinted with permission from [2].

lottery will be paired with $M-1$ others drawn randomly.

Actions: We suppose that each agent has the same action space denoted as $\mathcal{A} = \{1, 2, \cdots, |\mathcal{A}|\}$. Hence, the action that this agent takes at time k is $a[k] \in \mathcal{A}$. Under the mean field assumption, the actions of the other agents would be drawn independently from the p.m.f. $\rho = [b_1, b_2, \cdots, b_{|\mathcal{A}|}]$, where b_a is the probability mass associated with action a. We call ρ as the assumed action distribution.

Costs: Each action $a \in \mathcal{A}$ taken at time k has a corresponding cost θ_a. This cost is fixed and represents the discomfort suffered by the agent in having to take that action.

Coupons: When agent takes an action a, it is awarded some fixed number of coupons r_a for playing that action. These coupons are then used by the agents as lottery tickets.

Lottery: We suppose that there are only K rewards for agents in one cluster, where K is a fixed number less than M. The probability of winning is based on the number of coupons that each agent possesses. We model each lottery as choosing a permutation of the M agents participating in it, and picking the first K of them as winners. We denote the winning probability as $p_{\rho,a}$ and derive its explicit form in Section 3.4.

States: The agent keeps track of his history of wins and losses in the lotteries by means of his net surplus at time k, denoted as $x[k]$. The value of surplus is the state of the agent, and is updated in

a Markovian fashion as follows:

$$x[k+1] = \begin{cases} x[k] + w, & \text{if agent 1 wins the lottery,} \\ x[k] - l, & \text{if agent 1 loses the lottery,} \end{cases} \tag{3.1}$$

where w and l is the impact of winning or losing on surplus. Effectively, the assumption is that the agent expects to win at least an amount l at each lottery. Not receiving this amount would decrease his surplus. Similarly, if the prize money at the lottery is $w + l$, the increase in surplus due to winning is w. Surplus values are discrete, and the set of possible values is given by a countable $\mathbb{X}$ that ranges from $(-\infty, +\infty)$.

Value function for prospect: The impact of surplus on the agent's happiness is modeled by an S-shaped incremental utility function $u(x[k])$, which is monotone increasing, concave for a positive surplus and convex for a negative surplus. Moreover, the impact of loss is usually larger than that of gain of the same absolute value. Note that we implicitly assume that the reference for all agents is 0. Then following [33], we use the following value function for prospect

$$u(x) = \begin{cases} u^+(x) = x^\gamma, & x \geq 0, \\ u^-(x) = -\varphi(-x)^\gamma, & x < 0, \end{cases} \tag{3.2}$$

where $\varphi > 1$ is the loss penalty parameter and $0 < \gamma < 1$ is the risk aversion parameter. A larger φ means that the operator is more loss averse, while a smaller γ indicates that the operator is more risk seeking. From empirical studies [33, 35], realistic values are $\varphi = 2.25$ and $\gamma = 0.88$.

Weighting function for prospect: It has been observed empirically that people tend to subjectively weight uncertain outcomes in real-life decision making [63]. In the proposed game, this weighting factors capture the agent's subjective evaluation on the mixed strategy of its opponents. Thus, under PT, instead of objectively observing the probability of winning the lottery $p_{\rho,a}$, each user perceives a weighted version of it, $\phi(p_{\rho,a})$. Here, $\phi(\cdot)$ is a nonlinear transformation that maps the objective probability to a subjective one, which is monotonic increasing in probability. It has

been shown in many PT studies that, people usually overweight low probability outcomes and underweight high probability outcomes. Following [63], we use the weighting function

$$\phi(p) = \exp(-(-\ln p)^{\xi}), \quad \text{for} \quad 0 \leq p \leq 1, \tag{3.3}$$

where $\xi \in (0, 1]$ is the objective weight that characterizes the distortion between subjective and objective probability. Note that under the extreme case of $\xi = 1$, (3.3) reduces to the conventional EUT probability, i.e., $\phi(p) = p$.

Regeneration: An agent may quit the system at any time, independent of others. This event occurs with probability $1 - \beta$, where $\beta \in (0, 1)$. When this happens, a new agent takes the place of the old one, with a state drawn from a probability mass function Ψ.

Best Response Policy: The agent must choose an action at each time, including staying with the status-quo/baseline as an action too. The green/light tiles in Figure 3.1 relate to the problem of the agent determining his best response policy. The agent assumes that the actions taken by each of his $M - 1$ opponents are drawn independently from probability mass function ρ. Given this assumption, the state of his surplus is x and current utility is $u(x)$, the agent must calculate the probability of winning at the lottery $p_{\rho,a}(x)$, if he were to take action $a(x) \in \mathcal{A}$, incurring a cost $\theta_{a(x)}$ and gaining $r_{a(x)}$ coupons. Since the agent must take this decision repeatedly, he must solve a dynamic program to determine his optimal policy. There could be many best response actions, and we assume that the agent chooses a randomized policy $\sigma(x) \triangleq [\sigma_1(x), \sigma_2(x), \cdots, \sigma_a(x), \cdots, \sigma_{|\mathcal{A}|}(x)]$, in which $\sigma_a(x)$ specifies the probability of playing action a when the agent's surplus is x; in other words, we enlarge the space to include mixed strategies as pure strategy equilibria may not exist. The action taken by the agent is a random variable $A \sim \sigma(x)$. The details of the lottery and how to calculate the probability of success are given in Section 3.4. The properties of the best response policy are described in detail in Section 3.7.

Stationary Surplus Distribution: The assumed action distribution ρ, and the best-response randomized policy $\sigma(x)$ yield the state transition kernel of the Markov chain corresponding to the

surplus, via the probability of winning the lottery $p_{\rho,a}(x)$. This is illustrated by means of the blue/dark tiles in Figure 3.1. The transition kernel also is influenced by the regeneration distribution Ψ. The stationary distribution of surplus associated with the transition kernel is denoted as ζ_ρ. This stationary distribution of the single mean field agent is equivalent to the one-step empirical state distribution of infinite agents who all take a (mixed-strategy) action, $\sigma(x)$ when state is x, assuming that the actions of their competitors would be drawn from ρ.

Mean Field Equilibrium: The triple of an assumed action distribution ρ, randomized policy σ and stationary surplus distribution ζ gets mapped via mapping Π^* into a triple of action distribution $\tilde{\rho}$, best-response randomized policy $\tilde{\sigma}$ and a stationary surplus distribution $\tilde{\zeta}$ via the operations described above. A fixed point of the resulting map is called an MFE. For a formal definition and the proof of existence see Section 3.6.

3.2.1 Discussion

Is the MFG a good approximation? Specifically, we need to first show that for any agent, the assumption that the states of any finite subset of agents that it interacts with are independent of it and each other as the number of agents becomes asymptotically large. Second, we need to show that when we repeat the game over time, the empirical distribution of the agents' states converges to a fixed point (mean field limit).

The first result is follows from an argument called *propagation of chaos* via constructing *interaction sets* defined in [15], which characterize the conditions under which any finite subset of the state of the agents are independent of each other. Following a similar argument to [6], we can show that after any finite number of lotteries (finite time), as the total number of agents becomes large enough, the interaction sets of any finite collections of agents become disjoint with high probability. Hence, the states of these agents become independent. Inspired by [6], the proof is divided into two parts: (i) first, we need to show that as the total number of agents JM (J is the number of lotteries) becomes large enough, the probability that agent 1 interacted with the set of agents (that it interacts at the k-th lottery, $k \geq 1$) before the k-th lottery become zero; and (ii) the action distribution ρ_1, the randomized policy σ_1, and the surplus distribution ζ_1 of agent 1 converges to

the assumed distributions ρ, σ, ζ, respectively, as the number of agents JM becomes large enough. We do not present the full argument here due to space limitations and the fact that it follows via identical arguments to [6, 15].

The second result requires the establishment of the so called *Mckean-Vlasov limit*—a differential equation that specifies the evolution of the empirical distribution of state over the transition kernel specified in Figure 3.1. In order to do this, we need to verify three sufficiency conditions presented in [64]) (see Section 2 Assumptions $\boldsymbol{A1}$ - $\boldsymbol{A3}$) built on a continuous-time Markov chain (CTMC). It is easy to move our discrete-time Markov chain (DTMC) setup to their framework by equipping each agent with an independent Poisson clock with rate λ_Q (chosen as 1 w.l.o.g). An agent whose clock ticks is allowed to take an action, and receives a reward with the same probability engendered by a lottery under the same action and with the same belief distribution. The equivalence of the stationary distributions of the CTMC and DTMC versions follows immediately from [65] (Chapter 7), with the Bellman equation of the DTMC system being replaced by the Hamilton-Jacobi-Bellman equation of the CTMC. In our problem, the Q-matrix of the equivalent CTMC is simply $-\lambda_Q I + \lambda_Q P$, where P is the P-matrix of the DTMC version and I is the identity matrix. The most important condition of [64] that needs to be verified is the assumption on the Lipschitz nature of the map between the belief action distribution and the resultant action distribution. We numerically verify in Section 3.3.6.2 that given an action belief ρ, the derivative of this map at each iteration step is bounded, leading to the desired Lipschitz property for both DTMC and CTMC. While this supports the conjecture that the condition holds in our case, the proof is beyond the scope of this work due to the implicit form of the map.

3.3 Numerical Study

We conduct an empirical data-based simulation in the context of electricity usage for home air conditioning to illustrate the likely performance of our nudge system in the context of electricity demand-response. In doing so, we will also numerically study the properties of the mean field approximation. As mentioned in Section 3.1, our context is that of a Load Serving Entity (LSE) trying to incentivize its customers to shape their electricity consumption so as to reduce its cost

of electricity purchase from the wholesale market whose price variation is as shown in Figure 3.2. These incentives could increase the net surplus of the end-users, the profit of the LSE, or the total welfare of these agents as well. Data available for our simulations consist of historical electricity prices from [30], and a data set containing appliance-wise electricity usage for about 1000 homes along with the ambient temperatures over each day in June–August, 2013 [31].

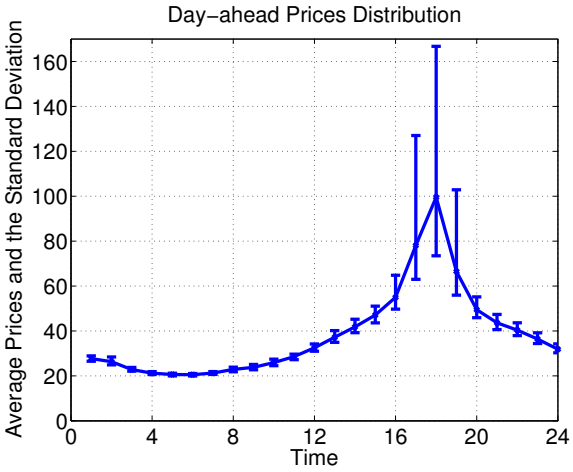

Figure 3.2: Day-ahead electricity market prices in dollars per MWh on an hourly basis between 12 AM to 12 PM, measured between June–August, 2013 in Austin, TX. Standard deviations above and below the mean are indicated separately. Reprinted with permission from [2].

3.3.1 Home Model

A standard continuous time model [66, 67] for describing the evolution of the internal temperature $\tau(t)$ at time t of an air conditioned home is

$$\dot{\tau}(t) = \begin{cases} -\dfrac{1}{RC}(\tau(t) - \tau_a) - \dfrac{\eta}{C}P_m, & \text{if } q(t) = 1, \\ -\dfrac{1}{RC}(\tau(t) - \tau_a), & \text{if } q(t) = 0. \end{cases} \tag{3.4}$$

Here, τ_a is the ambient temperature (of the external environment), R is the thermal resistance of the home, C is the thermal capacitance of the home, η is the efficiency, and P_m is the rated electrical

power of the AC unit. The state of the AC is described by the binary signal $q(t)$, where $q(t) = 1$ means AC is in the ON state at time t and in the OFF state if $q(t) = 0$. The state is determined by the crossings of user specified temperature thresholds as follows:

$$\lim_{\epsilon \to 0} q(t+\epsilon) = \begin{cases} q(t), & |\tau(t) - \tau_r| \leq \Delta, \\ 1, & \tau(t) > \tau_r + \Delta, \\ 0, & \tau(t) < \tau_r - \Delta, \end{cases} \tag{3.5}$$

where τ_r is the temperature setpoint and Δ is the temperature deadband.

Table 3.1: Parameters for a Residential AC Unit. Reprinted with permission from [2].

C(Capacitance)	R(Resistance)	P_m(Power)	η(Coefficient)	τ_r(Setpoint)	Δ(Deadband)
10 kWh/°C	2 °C/kW	6.8 kW	2.5	22.5 °C	0.3 °C

A number of studies investigate the thermal properties of typical homes. We use the parameters shown in Table 3.1 for our simulations. These are based on the derivations presented in [66] for temperature conditioning a 250 m^2 home (about 2700 square feet), which is a common mid-size home in many Texas neighborhoods.

In order to determine the energy usage for AC in our typical home, we need to know how the ambient temperature varies in Texas during the summer months of interest. These values are available in the Pecan Street data set, and we plot the values of 3 days which are arbitrarily chosen over three months for Austin, TX in Figure 3.3.

Next, we calculate the ON-OFF pattern of our typical air conditioner based on the ambient temperature variation over the course of the day. We do this by simulating the controller in (3.5) with the appropriate ambient temperature values taken from Figure 3.3. The pattern is presented in Figure 3.4. We see that there is higher energy usage during the hotter times of the day, as is to be expected. This also corresponds to the peak in wholesale electricity prices shown for the same

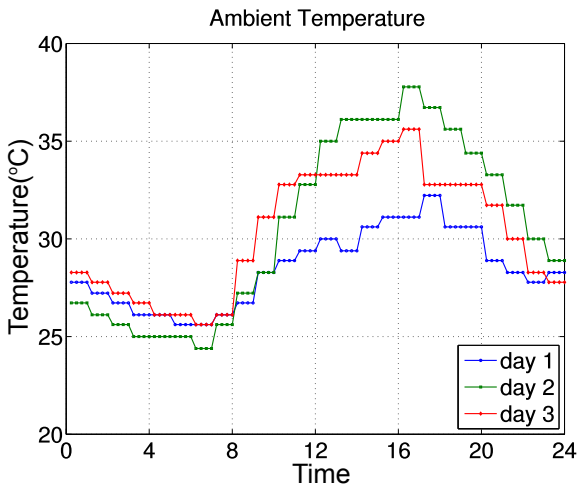

Figure 3.3: Ambient temperature of 3 arbitrary days from June–August, 2013 in Austin, TX. Measurements are taken every 15 minutes from 12 AM to 12 PM. Reprinted with permission from [2].

period in Figure 3.2. The total energy used each day corresponding to our 2500 sq ft home with a 5 ton AC (= 6.8 kW; see Table 3.1) is 32.83 kWh. For comparison, we identified 4 homes in the Pecan Street data set that have parameters in the same ballpark as our typical (simulated) home. The average size of these real homes was 2627 sq feet, with a 4 ton AC on average, and the average electricity consumed for airconditioning was 34.8 kWh per day during time interval corresponding to our simulation. The numbers are quite similar to our simulated home, indicating accuracy of the model.

3.3.2 Actions, Costs and LSE Savings

Since we are interested in peak-period usage, we consider an action set available to the customer that consists of choosing different thermostat setpoints during each hour from $2-8$ PM, i.e, 6 periods (hours) in total. We denote each period by an index j, where $j = 1$ indicates the period $2-3$ PM and so on until $j = 6$, which indicates the period $7-8$ PM. Each action can now be identified with a vector $(y_1, y_2, y_3, y_4, y_5, y_6)$, where y_j indicates the setpoint in the period j. We take the setpoint 22.5°C as the baseline. Hence, the vector $(22.5, 22.5, 22.5, 22.5, 22.5, 22.5)$ indicates a baseline action in which the customer does not change the original setpoint in each period. The

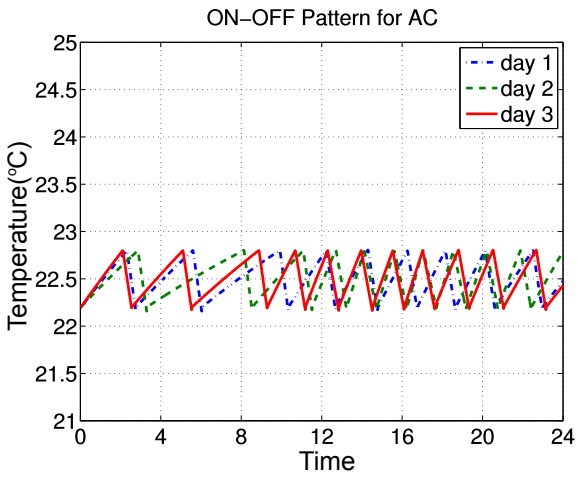

Figure 3.4: Simulated ON/OFF state of AC over a 24 hour period in a home and the corresponding interior temperature. The interior temperature falls when the AC comes on, and rises when it is off. Reprinted with permission from [2].

set of all such setpoint vectors defines an action set $\mathcal{A}$, and we define the action with index $a = 0$ to be the no-change action. Since the setpoints on a thermostat are discrete, the number of actions is finite. We identified 5 other actions that appeared to have the most promise of being used. These actions are shown in the second column of Table 3.2.

We next calculate the cost of taking each action $a \in \mathcal{A}$, which corresponds to the discomfort of having a potentially higher mean and standard deviation in the home temperature, and possibly higher energy consumption. We measure the state of the home under action $a \in \mathcal{A}$ by the tuple consisting of the mean temperature, the standard deviation and energy usage, denoted $[\bar{\tau}_a, \sigma_a, \mathrm{E}_a]$. The baseline state of these parameters is under action 0, denoted by $[\bar{\tau}_0, \sigma_0, \mathrm{E}_0]$ We define the cost of taking any action a as

$$\theta_a = |\bar{\tau}_0 - \bar{\tau}_a| + \lambda|\sigma_0 - \sigma_a| - \varsigma(\mathrm{E}_0 - \mathrm{E}_a), \tag{3.6}$$

where we choose $\lambda = 10$ to make the numerical values of the mean and standard deviation comparable to each other and $\varsigma = 10$ ¢/kWh as the fixed energy price. We note that the map between temperature variation, discomfort suffered, and its measurement in cents is not obvious. However,

given the fact that the customer uses between $1-3$ kWh or about $\text{¢}10-30$ per hour to obtain a temperature differential between the ambient temperature and interior temperature of about $15-20$°C, the discomfort cost of a degree C temperature increase being $\text{¢}1$ seems reasonable in the limited temperature range that we are interested in. Note that the calculation of cost for each action involves simulating the home under that action to determine $[\bar{\tau}_a, \sigma_a, \mathrm{E}_a]$. However, this has to be done only once to create a look-up table, which can be used thereafter. Note also that each action in $\mathcal{A}$ is chosen to be close to energy neutral, i.e., the third term in (3.6) is essentially zero. Thus, we focus on modifying usage time, not the total usage. Table 3.2 shows our selection of actions and their corresponding costs.

When applied over a day, each action could result in some savings to the LSE towards the costs it incurs in purchasing electricity. We measure the day-ahead price of electricity experienced by the LSE in dollars/MWh and denote the price at time period j in day i as $\pi_{i,j}$, where $i = \{1, 2, \cdots, 92\}$ and $j = \{1, \cdots, 6\}$. Each action vector of a customer would impose a net price on the LSE in proportion to the usage. We define the *differential price* measured in dollars imposed by an action $y = (y_1, y_2, y_3, y_4, y_5, y_6)$ versus $z = (z_1, z_2, z_3, z_4, z_5, z_6)$ as

$$H(y, z) = \sum_j \left(k(y_{i,j}) - k(z_{i,j})\right) \pi_{i,j}, \tag{3.7}$$

where k converts the setpoints into electricity usage in each period, which is measured in MWh. Setting y as the baseline action $(22.5, 22.5, 22.5, 22.5, 22.5, 22.5)$ presents a way of measuring the reduction/increase in cost due to the incentive scheme.

We calculated the savings of each action applied over each day of our three month data set and obtained the average savings. These values are shown in Table 3.2 (where the final columns entitled "C_0–C_3" will be discussed in Section 3.3.4). As is clear, the cost of taking each of our selected actions is considerably lower than the savings resulting from that action, and hence it might be possible to create appropriate incentive schemes to encourage their adoption. We will consider two such schemes, namely, (i) a fixed reward scheme used as a benchmark, and (ii) a

lottery based scheme.

Table 3.2: Actions, Costs, LSE Savings and Coupons Awarded. Reprinted with permission from [2].

Index	Action Vector	Cost (¢)	LSE Savings (¢)	C_0	C_1	C_2	C_3
0	$(22.5, 22.5, 22.5, 22.5, 22.5, 22.5)$	0	0	37.4	8.416	8.416	8.416
1	$(21.5, 21.5, 22.25, 23.5, 23.75, 21.25)$	3.68	27.7	715	431.8	521.8	611.8
2	$(21.5, 21.5, 22.25, 23.5, 23.25, 22.25)$	3.15	22.7	693	431.8	511.6	591.5
3	$(21.5, 21.5, 22.25, 24, 23, 22.5)$	2.68	22	577	431.8	466.8	501.7
4	$(22, 22, 22.25, 23, 23, 22.5)$	1.34	19	434	287.4	325.6	363.7
5	$(22, 22, 22.25, 23.25, 22.5, 22.75)$	0.95	16.4	222	287.4	244.2	200.9

3.3.3 Benchmark Incentive Scheme

A simple incentive scheme to get users to adopt cost saving actions (from the LSE's perspective) is to calculate the expected savings of each action, and to deterministically reward each agent with some percentage of the expected savings for taking that action. Such guaranteed savings are similar in spirit to rebate for using public transportation during off-peak hours, and a system of sharing a fraction of the savings by demand-response providers such as OhmConnect [68]. We will use this scheme as a benchmark in order to determine whether shared savings are large enough to encourage meaningful participation. Thus, our benchmark incentive scheme attempts to incentivize each action by returning to the user some fixed fraction of the expected LSE savings for that action presented in Table 3.2. For example, a return of 50% for taking action 1 would imply awarding ¢13.85 each time that action is taken.

3.3.4 Lottery-Based Incentive Scheme

Our second incentive scheme is lottery-based, with Energy Coupons being used as lottery tickets. Now, the baseline action $a = 0$ corresponds to a setpoint of 22.5°C in period 3 at which $\pi_{i,3}$ is highest (Figure 3.2) for any day i. Hence, the LSE should incentivize actions that are likely to reduce the risks of peak day-ahead price by offering Energy Coupons in proportion to the us-

age during the corresponding periods. In the context of our simulation, it is intuitively clear that coupons must be placed at periods of lower price. Our candidate *coupon profiles* are shown in Table 3.3, where coupons are awarded in periods 1 and 6 only if the usage is greater than base usage values of $x_1 = 2.464$ kWh and $x_6 = 2.24$ kWh, respectively. We experimented with a range of coupon profiles to explore their impact on the MFE, and present some examples C_0–C_3.

Table 3.3: Mean Day-ahead Price and Energy Coupon Profiles. Reprinted with permission from [2].

Index	Period	Price/MWh	C_0/kWh	C_1/kWh	C_2/kWh	C_3/kWh
1	$2-3$ PM	\$47	107	100	100	100
2	$3-4$ PM	\$55	5.4	0	0	0
3	$4-5$ PM	\$78	1.8	0	0	0
4	$5-6$ PM	\$99.6	0	0	0	0
5	$6-7$ PM	\$66.5	3.6	0	0	0
6	$7-8$ PM	\$49.5	54	0	20	40

Given the coupon placement by the LSE, the customers need to determine the number of coupons resulting from each action, and use these values to estimate the utility that they would attain. Our six actions are shown in Table 3.2 with their attendant costs and number of coupons received. The LSE conducts an lottery each week across clusters of $M = 50$ homes participating in each lottery. For each cluster, there is $K = 1$ prize for winning the lottery. We assume that the customers choose the same action on each day of the week, and then participate in the lottery.

3.3.5 Utility and Surplus

As described in Section 3.2, the user state consists of his/her surplus. Any rewards result in an increase in surplus by the reward amount w, whereas performing an action but not receiving a reward results in decreasing the surplus by some amount l. Since a reward is assured for each action in the benchmark incentive scheme, there are no surplus decrease events. However, in the lottery scheme, a user that does not win the lottery would see a decrease in surplus. We select l

that results from losing at a lottery to be the average reward obtained form the lottery assuming that every player has an equal probability of winning.

For the customer utility, which maps surplus to utility units, we use the value function of prospect model (defined in (3.2)), $u(x) = x^\gamma$ if $x \geq 0$ or $u(x) = -\varphi(-x)^\gamma$ if $x < 0$, where $\varphi = 2.25$ and $\gamma = 0.88$ according to the empirical studies conducted in [33, 35]. The utility model applies to both the benchmark and the lottery scheme. Under this model, we expect a user who has lost a number of lotteries to stop participating in the system, since his surplus becomes negative and he is not receiving enough of an incentive to stay, given the cost he bears each day. Similarly, a user who has won too many times would have a large surplus, and would also not be keen on participating since the marginal utility he gets may not be high enough for him. The latter observation applies to the benchmark as well, although given the small rewards, we do not expect it to happen frequently.

The participants in the lottery scheme see a distorted probability of winning, parameterized by ξ, as defined in (3.3). This is an important feature of our model, since it captures the attractiveness of lotteries in incentivizing risky actions. Consistent with empirical studies in [63], we choose $\xi = 0.37$.

We assume that a customer remains in the system with probability 0.92, *i.e.,* the average lifetime is 12 time steps, which parallels the fact that the main summer season lasts for about three months. Further, a newly entering customer has zero surplus.

3.3.6 Equilibria Attained by Incentive Schemes

3.3.6.1 Benchmark Scheme

As described in Section 3.3.3, we construct a fixed-reward type of incentive scheme to obtain a benchmark with which to compare the performance of the lottery scheme. Under the benchmark scheme, customers are awarded some percentage of the expected savings that their action is likely to yield to the LSE, shown in Table 3.2. The actual action chosen by the customer will be determined using a dynamic program (DP) similar to the one defined in (3.14). However, since

rewards are deterministic, there is no dependence on the belief over competitors' actions, and the only randomness is from the lifetime of the user. Thus, solving the DP is straightforward, and we can easily obtain a map between surplus and action for a given reward.

3.3.6.2 Lottery Scheme

We next consider the lottery with $M = 50$ competitors. The win probability $p_{\rho,a}$ is the probability that the coupons generated by action a are greater than those generated by $M-1$ independent actions drawn from ρ. We offer a single prize with value $\$15$, which implies that the customer expects to win $¢30$ on average by participating, i.e., the decrease in surplus due to losing at the lottery is $l = 0.3$, while the increase in surplus due to winning is $w = 15 - 0.3 = 14.7$.

We start with a uniform action distribution ρ_0 as the initial condition. In each iteration i, given the belief ρ_i (action distribution of other players), we first determine the value of each state using the Bellman equation for value (3.15), with convergence in roughly 50 steps. We next determine the stationary surplus distribution, and then map it to the resultant stationary action distribution ρ_{i+1}, uniformly choosing all equal value actions. Note that this map $\tilde{\Pi}$, from belief ρ_i to the resultant distribution ρ_{i+1}, is a sequential version of the map Π^* from Section 3.2, and $\tilde{\Pi}$ and Π^* have the same fixed points. As described in Section 3.2.1, the iterative procedure is referred to as the Mckean-Vlasov dynamics. As discussed in Section 3.2.1, we also identify the CTMC version of our system, and simulate it using the same procedure above. Finally, as specified in Section 3.2.1, an important sufficiency condition for convergence is the Lipschitzness of the map $\tilde{\Pi}$. We calculate a numerical derivative (for both the DTMC and CTMC versions)

$$\frac{||\tilde{\Pi}_{i+1}(\rho_{i+1}) - \tilde{\Pi}_i(\rho_i)||_{\text{sup}}}{||\rho_{i+1} - \rho_i||_{\text{sup}}}. \tag{3.8}$$

Figure 3.5 (*Left*) plots the derivative along a simulated trajectory for both DTMC and CTMC. That they are bounded, indicates the Lipschitz property of the maps.

We found that typically convergence occurs rapidly and reaches within 0.1% of the final value within 20 iterations. The eventual values to which each surplus value converge is the mean field

surplus distribution, in which it turns out that customers win at a lottery at most once over an average lifetime of 12 time intervals, as is to be expected with a cluster size of 50 customers at each lottery. The mean field action distribution under the lottery scheme with a $\$15$ reward and the savings attained are shown in Table 3.4. The MFE shifts based on the coupon profile, but the saving is quite robust to profiles that award comparable numbers of coupons in periods 1 and 6.

We observed multiple thresholds at which two actions have identical value. For example, under coupon profile C_0, there are three threshold surplus values -19.2, 1.7 and 190.3 at which we have equal probabilities of choosing between actions 0 and 2, between 2 and 4, and between 4 and 5, respectively.

Table 3.4: Mean Field Equilibria under $15 reward (lottery prize). Reprinted with permission from [2].

Coupon Profile	MFE	Expected Surplus	Expected Value	LSE Saving
C_0	$[0.001, 0, 0.81, 0, 0.19, 0]$	0.3563	$\$189.8$	$\$77$
C_1	$[0.001, 0, 0, 0.584, 0, 0.416]$	0.3704	$\$203.5$	$\$69$
C_2	$[0.001, 0, 0.875, 0.124, 0, 0]$	0.3565	$\$193.1$	$\$79$
C_3	$[0.001, 0, 0.999, 0, 0, 0]$	0.3563	$\$189.4$	$\$79.4$

Example

Figure 3.5 (*Middle*) shows the interior temperature under actions 0, 2, 4, the mean field action distribution and benchmark action when $\$15$ is the total reward amount. We see that the mean field behavior is more aggressive than the benchmark in reducing the interior temperature before the peak period, and shows a marginally higher interior temperature during the peak period. Figure 3.5 (*Right*) shows the comparison of energy consumption between action 0 (doing nothing, with an average energy consumption of 36.5 kWh per day), the mean field action distribution (average energy consumption of 36.7 kWh per day), and the benchmark action (average energy consumption of 36.4 kWh per day). We see that the mean field distribution is more aggressive in moving energy usage away from the peak period as compared to the benchmark, although both have essentially the same energy consumption and an identical reward value of $\$15$ per week. Finally, we compute

that the savings to the LSE over 50 homes each week in this is example is $\$57.4$ in the benchmark scheme and $\$77$ in the lottery scheme. Thus, incentivizing customers by offering a prize of $\$15$ each week is certainly feasible. The MFE illustrates that even as small as $1°$C change of the setpoint of AC each day over several homes can yield significant benefits.

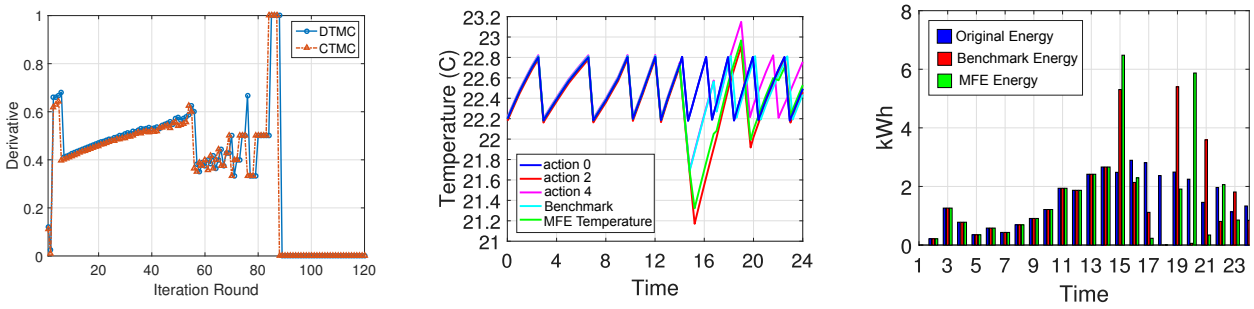

Figure 3.5: *(Left)* Numerical derivative of map; *(Middle)* Simulated ON/OFF state of AC over a 24 hour period under actions 0, 2, 4, the mean field action and the benchmark action on an arbitrary day and the corresponding interior temperature. The temperature graph is slightly offset for actions $2, 4$, the mean field action and the benchmark action for ease of visualization; *(Right)* Average daily energy usage profile. Reprinted with permission from [2].

3.3.7 Performance Analysis of Incentive Schemes

Benchmark Scheme

We consider a range of scenarios wherein the LSE rewards customers for each action with between $1\% - 100\%$ of its expected savings, in steps of 1% increments. The relations between the total weekly reward to customers, savings to the LSE and profit to the LSE, are shown in Figure 3.6 (*Left*). We see that the maximum weekly profit of $\$52$ is achieved when about 9% savings (about $\$10$ in total per week) is the customer reward regardless of the coupon awarding profile, indicating robustness to the exact profile employed. Note that although the reward under the benchmark scheme is indicated by a percentage returned, it corresponds to a dollar value returned based on the actions of the customers, and the total dollar reward values are also shown in green (dashed line) in Figure 3.6 (*Left*).

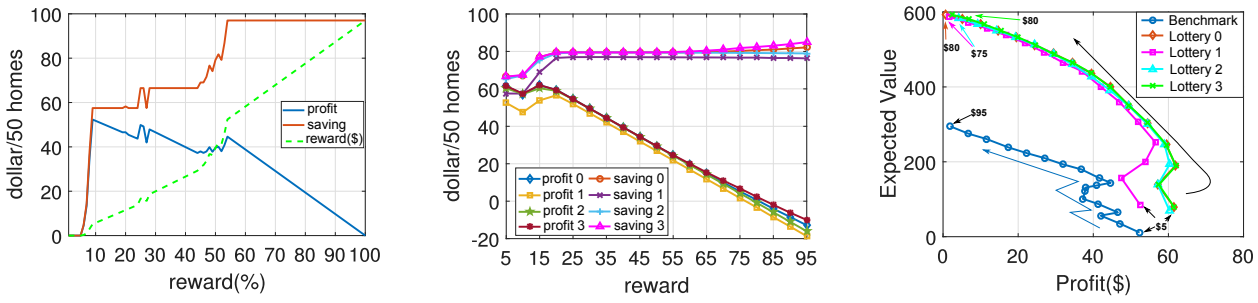

Figure 3.6: The relation between offered reward, LSE savings and LSE profit: *(Left)* Benchmark incentive scheme and *(Middle)* Lottery scheme; *(Right)* The relation between profit to the LSE and the expected value of a generic customer when different rewards are given to the customers. Reprinted with permission from [2].

Lottery Scheme

We next conduct numerical experiments under a range coupon profiles and lottery rewards from $\$5$ to $\$95$ in steps of $\$5$ increments as we did (by using percentage returns) with the benchmark scheme. Hence, we set a reward value, calculate the values of l and w that it implies, and compute the MFE for different coupon awarding profiles. Our results are shown in Figure 3.6 (*Middle*), where we plot the total savings to the LSE as well as its profit (savings minus reward) as a function of the reward offered for winning the lottery. From Figure 3.6 (*Middle*), the maximum profit is achieved by in a robust manner giving a reward of $\$15 - \$20$ for all coupon profiles Also, from observation of the mean field action distribution that results from this reward (Table 3.4), we note that almost all the customers will participate in the system, i.e., the probability of choosing action 0 is close 0.

From Figures 3.6 (*Left*) and 3.6 (*Middle*), we see that the maximum profit using the lottery scheme is a little over $\$62$ per week, while the maximum profit is only about $\$52$ under the benchmark scheme. Given that a typical LSE has several hundred thousand customers, a difference of $\$10$ each week over a cluster of 50 homes is quite significant.

Comparison of Local Social Welfare

Our final step is to characterize the local social welfare under the lottery (under different coupon profiles) and the benchmark. We define the (expected) *local social welfare* (LSW) measured in

dollars/week as

$$\text{LSW} = 50\mathbb{E}_{\text{MFE}}[V_i(X)] + \text{profits of the LSE.} \tag{3.9}$$

For the lottery scheme, the maximum expected value to each user of roughly 600 each is obtained when the LSE is revenue neutral. This is roughly 100% better than what can be achieved with the benchmark individual incentive. This increase is due to both the prospect-based utility as well as coupling across users in the lottery scheme.

Note also that the LSE obtains a maximum profit between $\$62 - \$64.4$ (for different coupon profiles) by giving a $\$15$ reward, under which each customer will take actions according to the MFEs shown in Table 3.4. The corresponding surplus is also shown, which for a generic customer is between $\$189.2 - \$203.5$. Therefore, the local social welfare is about $\$9552$.

For the benchmark, the profit to the LSE is $\$42$ when $\$15$ reward is given as shown in Figure 3.6 (*Left*), under which each customer will only take action 5. The corresponding expected value of a generic customer is about $\$56.2$, and the local social welfare is about $\$2852$. Again, we see that the lottery scheme outperforms the benchmark.

We perform the same analysis to determine the relation between profit to the LSE and the expected value of a generic customer under different rewards and coupon profiles. Our results are shown in Figure 3.6 (*Right*). We explore a range of rewards from $\$5$ to the break-even point ($\$80$ for lottery scheme and $\$95$ for the benchmark, regardless of the coupons awarded). The points on each curve correspond to increasing the reward by $\$5$ in steps in a manner indicated by the arrow marks. From Figure 3.6 (*Right*), we see that the lottery scheme appears to better capture the frontier between LSE profit and customer value than the benchmark scheme, with the lottery-based incentive being better in a Pareto-sense. This is true regardless of the exact maximum coupon choice. Thus, based on a desired level of customer value and LSE profit, the lottery scheme can ensure a better outcome than the benchmark scheme with an appropriate reward.

3.4 Lottery Scheme

We first construct the lottery scheme that will be used in our mean field game. We permute all the agents into clusters, such that there are exactly M agents in each cluster, and conduct a lottery in each such cluster. Suppose there are K rewards for all agents in one cluster, where K is a fixed number less than M. When an agent takes an action, he/she will receive the credit (number of coupons) associated with that action. Then the probability of winning is based on the number of coupons that each agent possesses. We will model the lotteries as choosing a permutation of the M agents participating in it, and picking the first K of them as winners. Then different lottery schemes can be interpreted as choosing different distributions on the symmetric group of permutations on M. In particular, we will use ideas from the Plackett-Luce model to implement our lotteries.

Without loss of generality, we assume that the actions are ordered in decreasing order of the costs so that $\theta_1 \geq \cdots \geq \theta_A$. In order to incentivize agents to take the more costly actions we will insist that the vector of coupons obtained for each action is also in decreasing order of the index, i.e., $r_1 \geq \cdots \geq r_A$.

The specific lottery procedure we consider is the following: for every agent m that takes action $a[m]$ and receives coupons $r_{a[m]} > 0$, we choose an exponential random variable with mean $1/r_{a[m]}$ and then pick the first K agents in increasing order of the realizations of the exponentials. Note the abuse of notation only in this section to use $a[m]$ to refer to the action of agent m. Since we consider only one lottery, we do not consider time k. Let the agent $m = 1, \ldots, M$ receive $r_{a[m]}$ number of coupons. The set of winners is a permutation over the agent indices, and we denote such a permutation by $\mu = [\mu_1, \mu_2, \cdots, \mu_M]$. We then have the probability of the permutation μ given by

$$\mathbb{P}(\mu | r_{a[1]}, \ldots, r_{a[M]}) = \prod_{n=1}^{M-1} \frac{r_{a[\mu_n]}}{\sum_{j=n}^{M} r_{a[\mu_j]}}. \tag{3.10}$$

Essentially, after each agent is chosen as a winner, he is removed and the next lottery is conducted just as before but with fewer agents.

We now analyze the probability of winning in our lottery. For analysis under the mean field assumption, it suffices to consider agent 1 with the coupons it gets by taking action a being denoted as $r_{a[1]}$. Let $\mathcal{M} := \{2, \ldots, M\}$, which is the set of opponents of agent 1. For these agents, suppose there are υ_n agents that choose action n, where $\sum_{n\in\mathcal{A}} \upsilon_n = M - 1$. We denote the vector of these actions by $\vec{\upsilon} = (\upsilon_1, \ldots, \upsilon_A)$.

The conditional probability of agent 1 failing to obtain a reward is given by

$$p^L_{1,\vec{\upsilon}} = \sum_{\kappa_1\in\mathcal{M}_1} \cdots \sum_{\kappa_K\in\mathcal{M}_K} \frac{\prod_{l=1}^K r_{a[\kappa_l]}}{\prod_{l=1}^K (r_{a[1]} + \sum_{m\in\mathcal{M}_l} r_{a[m]})},$$

where L refers to the fact that agent 1 "loses," $\mathcal{M}_1 = \mathcal{M}$, and for $l \geq 2$ we have $\mathcal{M}_l = \mathcal{M}_{l-1} \setminus \{\kappa_{l-1}\}$. Essentially, the above looks at the lottery process round by round, and is a summation of the probabilities of all permutations in which agent 1 does not appear in the first spot in any round.

The above expression considerably simplifies if the summations are instead taken over the actions $\tilde{\kappa}_l$ that the lottery winner κ_l at round $l \in \{1, \ldots, K\}$ can take. Note that we assume that we can distinguish the actions based on the number of coupons given out. If this were not true, then we could further simplify the expression by summing over the coupon space. Given a coupon/action profile $\vec{\upsilon}$, let $\mathcal{J}(\vec{\upsilon})$ denote the actions that have non-zero entries. Additionally, by $\vec{\upsilon} - \vec{1}_{\tilde{\kappa}}$ for $\tilde{\kappa} \in \mathcal{J}(\vec{\upsilon})$ denote the resulting coupon profile obtained by removing one entry at location $\tilde{\kappa}$, and by $r_{\vec{\upsilon}}$ the sum of all the coupons in profile $\vec{\upsilon}$, i.e., $\sum_{\tilde{\kappa}\in\mathcal{J}(\vec{\upsilon})} r_{\tilde{\kappa}}\upsilon_{\tilde{\kappa}}$. Then

$$p^L_{1,\vec{\upsilon}} = \sum_{\tilde{\kappa}_1\in\mathcal{J}(\vec{\upsilon}^1)} \cdots \sum_{\tilde{\kappa}_K\in\mathcal{J}(\vec{\upsilon}^K)} \frac{\prod_{l=1}^K \upsilon^l_{\tilde{\kappa}_l} r_{\tilde{\kappa}_l}}{\prod_{l=1}^K (r_{a[1]} + r_{\vec{\upsilon}^l})}, \tag{3.11}$$

where $\vec{\upsilon}^1 = \vec{\upsilon}$, for $l = 2, \ldots, K$, $\vec{\upsilon}^l = \vec{\upsilon}^{l-1} - \vec{1}_{\tilde{\kappa}_l}$ and $\upsilon^l_{\tilde{\kappa}}$ is the number of entries at location $\tilde{\kappa}$ for coupon profile $\vec{\upsilon}^l$. Note that $p^L_{1,\vec{\upsilon}}$ is a decreasing function of $r_{a[1]}$ for every $\vec{\upsilon}$. Therefore, agent 1 comparing two actions i and j that have $r_{1,i} > r_{1,j}$ will find $p^L_{1,\vec{\upsilon}}(i) < p^L_{1,\vec{\upsilon}}(j)$ for all $\vec{\upsilon}$. Also by taking the limit of $r_{a[1]}$ going to 0, having an action with 0 coupons results in a loss probability of 1 for every $\vec{\upsilon}$.

To determine the probability of winning in the lottery we need to account for the fact that the actions of the opponents are drawn from the distribution ρ (under the mean field assumption). Hence, the probability of obtaining the coupon profile (equivalently action profile) of the opponents $\vec{v} = (v_1, \ldots, v_A)$ is given by the multinomial formula, *i.e.,*

$$\mathbb{P}_\rho(\vec{v}) = \frac{(M-1)! \prod_{i \in \mathcal{A}} b_i^{v_i}}{\prod_{i \in \mathcal{A}} v_i!}. \tag{3.12}$$

Using (3.11) and (3.12), we obtain the winning probability for the mean field agent 1 when taking action a as

$$p_{\rho,a} = 1 - \sum_{\vec{v}:|\mathcal{J}(\vec{v})|=M-1} p_{1,\vec{v}}^L \mathbb{P}_\rho(\vec{v}). \tag{3.13}$$

By lower bounding each term in the conditional probability of not obtaining a reward we get $p_{\rho,a} \leq 1 - \frac{M-K}{M}(\frac{r_A}{r_1})^K =: \overline{p}_W \in (0,1)$. If we ran the lottery without removing the winners (and any of their coupons), we obtain a lower bound on the probability of winning that has a simpler expression. Using this simpler expression we can obtain the lower bound $p_{\rho,1} \geq 1 - (1 - \frac{r_A}{r_A + (M-1)r_1})^K =: \underline{p}_W \in (0,1)$. Note that both bounds are independent of ρ. If we allow an action that yields 0 coupons, then the above bounds become trivial with $\overline{p}_W = 1$ and $\underline{p}_W = 0$.

An important feature of our lottery scheme is that the probability of winning increases with the number of coupons given out. For simplicity we assumed a fixed reward for any win. However, we can extend the lotteries to ones where different rewards are given out at different stages, and also where the rewards are dependent on the number of coupons of the winner. For the latter, we will insist on the rewards being an increasing function of the number of coupons of the winner. Finally, we can also extend to scenarios where we choose the number of stages K in an (exogenous) random fashion in $\{1, \ldots, M-1\}$. Since the analysis carries through unchanged except with more onerous notation, we only discuss the simplest setting.

3.5 Optimal Value Function

As discussed in Section 3.2, the mean field agent must determine the optimal action to take, given his surplus x and the assumed action distribution ρ. We follow the usual quasi-linear combination of prospect function and cost consistent with Von Neumann-Morgenstern utility functions, and under which the impact of winning or losing in the lottery is on the surplus of the agent (and not simply a one-step myopic value change).

The objective of a particular agent i is

$$V_\rho(\boldsymbol{x}[k]) = \max_{\left\{\boldsymbol{a}(\boldsymbol{x}[l])\in\mathcal{A}^{|\mathcal{A}|}\right\}_{l=k}^{\infty}} \mathbb{E}\left\{\sum_{l=k}^{\infty}\beta^{l-k}\left(u_i(x_i[l]) - \theta_{a_i(x_i[l])}\right)\right\},$$

where $\boldsymbol{x}[k] = (x_1[k], \cdots, x_M[k])$, and $\boldsymbol{a}(\boldsymbol{x}[k]) = (a_1(x_1[k]), \cdots, a_M(x_M[k]))$ are the vectors of surplus and actions for each agent in the particular lottery cluster of the agent at time k, respectively. The expectation is over the distribution of competitors actions and the randomness introduced by the lottery.

Under the mean field assumption, the actions of all agents besides i are drawn from a distribution ρ independently of each other. Also, the agent uses a prospect function to estimate the probabilities of winning and losing at the lottery. We can then drop the index of the agent i and the dynamic program that the agent in prospect theory needs to solve is given by the following Bellman equation

$$V_\rho(x) = \max_{a(x)\in\mathcal{A}}\{u(x) - \theta_{a(x)} + \beta[\phi(p_{\rho,a}(x))V_\rho(x+w) + \phi(1-p_{\rho,a}(x))V_\rho(x-l)]\}. \quad (3.14)$$

Note that $p_{\rho,a}(x)$ is a result of a lottery that we described in detail in Section 3.4, and $\phi(\cdot)$ is the weighting function, which overweights small probabilities (of winning the lottery) and underweights moderate and high probabilities (of losing the lottery). Here, we use the weighting function defined in (3.3).

First, we need to define a set of functions as

$$\Phi = \left\{ f : \mathbb{X} \to \mathbb{R} : \sup_{x \in \mathbb{X}} \left| \frac{f(x)}{\Omega(x)} \right| < \infty \right\},$$

where $\Omega(x) = \max\{|u(x)|, 1\}$. Note that Φ is a Banach space with $\Omega-$norm,

$$||f||_\Omega = \sup_{x \in \mathbb{X}} \left| \frac{f(x)}{\Omega(x)} \right| < \infty.$$

Also define the Bellman operator T_ρ as

$$T_\rho f(x) = \max_{a(x) \in \mathcal{A}} \{u(x) - \theta_{a(x)} + \beta[\phi(p_{\rho,a}(x)) f(x+w) + \phi(1 - p_{\rho,a}(x)) f(x-l)]\}, \quad (3.15)$$

where $f \in \Phi$.

We now show that the optimal value function $V_\rho(x)$ exists and it is continuous in ρ.

Lemma 9. *1) There exists a unique $f^* \in \Phi$, such that $T_\rho f^*(x) = f^*(x)$ for every $x \in \mathbb{X}$, and given $x \in \mathbb{X}$, for every $f \in \Phi$, we have $T_\rho^n f(x) \to f^*(x)$, as $n \to \infty$.*
2) The fixed point f^* of operator T_ρ is the unique solution of Equation (3.14)*, i.e. $f^* = V_\rho^*$.*

Lemma 10. *The value function $V_\rho(\cdot)$ is Lipschitz continuous in ρ.*

3.5.1 Stationary distributions

For a generic agent, w.l.o.g., say agent 1, we consider the state process $\{x_1[k]\}_{k=0}^\infty$. It's a Markov chain with countable state-space $\mathbb{X}$, and it has an invariant transition kernel given by a combination of the randomized policy $\sigma(x)$ at each surplus x for any $a(x) \in \mathcal{A}$, and the lottery scheme from Section 3.4. By following this Markov policy, we get a process $\{W[k]\}_{k=0}^\infty$ that takes values in $\{\text{win}, \text{lose}\}$ with probability $p_{\rho,a}(x)$ for the win, drawn conditionally independent of the past (given $x_1[k]$). Then the transition kernel conditioned on $W[k]$ is given by

$$\mathbb{P}(x_1[k] \in B | x_1[k-1] = x, W[k]) = \beta 1_{\{x + w 1_{\{W[k]=\text{win}\}} - l 1_{\{W[k]=\text{lose}\}} \in B\}} + (1-\beta)\Psi(B), \quad (3.16)$$

where $B \subset \mathbb{X}$ and Ψ is the probability measure of the regeneration process for surplus. The unconditioned transition kernel is then

$$\begin{aligned}\mathbb{P}(x_1[k] \in B | x_1[k-1] = x) =& \beta \sum_{a(x)\in\mathcal{A}} \sigma_a(x) p_{\rho,a}(x) 1_{x+w\in B} \\ &+ \beta\big(1 - \sum_{a(x)\in\mathcal{A}} \sigma_a(x) p_{\rho,a}(x)\big) 1_{x-l\in B} + (1-\beta)\Psi(B).\end{aligned} \tag{3.17}$$

Lemma 11. *The Markov chain where the action policy is determined by $\sigma(x)$ based on the states of the users and the transition probabilities in (3.17) is positive recurrent and has a unique stationary surplus distribution. We denote the unique stationary surplus distribution as $\zeta_{\rho\times\sigma}$. Let $\zeta^{(k)}_{\rho\times\sigma}(B|x)$ be the surplus distribution at time k induced by the transition kernel (3.17) conditioned on the event that $X[0] = x$ and there is no regeneration until time k. $\zeta_{\rho\times\sigma}(\cdot)$ and $\zeta^{(k)}_{\rho\times\sigma}(\cdot)$ are related as follows:*

$$\zeta_{\rho\times\sigma}(B) = \sum_{k=0}^{\infty}(1-\beta)\beta^k \mathbb{E}_\Psi\left(\zeta^{(k)}_{\rho\times\sigma}(B|X)\right) = \sum_{k=0}^{\infty}(1-\beta)\beta^k \int \zeta^{(k)}_{\rho\times\sigma}(B|x) d\Psi(x). \tag{3.18}$$

Thus $\zeta_{\rho\times\sigma}(B)$ in terms of $\zeta^{(k)}_{\rho\times\sigma}(B|x)$ is simply based on the properties of the conditional expectation. And note that in $\mathbb{E}_\Psi\left(\zeta^{(k)}_{\rho\times\sigma}(B|X)\right)$, the random variable X is the initial condition of the surplus, distributed as Ψ. For $x \in \mathbb{X}$, the only possible one-step updates are the increase of the surplus to $x + w$ or a decrease to $x - l$, i.e. $B = \{x + w, x - l\}$.

3.6 Mean Field Equilibrium

The action distribution ρ is a probability mass function on the action set $\mathcal{A}$: let b_i be the probability of choosing action i. Note that ρ lives in the probability simplex on $\mathbb{R}^{|\mathcal{A}|}$, which is compact and convex; denote it as Γ_ρ. Let ζ be the stationary surplus distribution and the set of all such possible surplus distributions is denoted as Γ_ζ, which is a compact and convex subset of l_∞: all surplus distributions are dominated by the distribution obtained by allowing the agent to win in every period; all surplus distributions dominate the distribution obtained by allowing the agent to

lose in every period; both these distributions have a finite mean (and convexity follows); and then the compactness result follows using the argument in [69]. For a given surplus x, let $\sigma(x)$ be the action distribution at x. Denote Γ_σ as the set of all possible distributions over the action space for each x, which is compact and convex. We further assume that $\rho \in \Gamma_\rho$, $\zeta \in \Gamma_\zeta$ and $\sigma(x) \in \Gamma_\sigma$ for each $x \in \mathbb{X}$.

Definition 1. *Consider the action distribution ρ, the randomized policy σ and the stationary surplus distribution ζ_ρ: (i) Given the action distribution ρ, determine the success probabilities in the lottery scheme using (3.13) and then compute the value function in (3.14). Taking the best response given by (3.14) results in an action distribution $\tilde{\sigma}$; (ii) Given action distribution ρ, following the randomized policy σ yields transition kernels for the surplus Markov chain and stationary surplus distribution $\tilde{\zeta}_\rho$, (with each transition kernel having a unique stationary distribution); and (iii) Given the stationary surplus distribution ζ_ρ, applying the randomized policy $\sigma(x)$ at each surplus x yields the distribution of actions $\tilde{\rho}$. Define the best response mapping Π^* that maps $\Gamma_\rho \otimes \Gamma_\sigma^{|\mathbb{X}|} \otimes \Gamma_\zeta$ into itself. Then we say that the assumed action distribution ρ, randomized policy σ and stationary surplus distribution ζ_ρ constitute a mean field equilibrium (MFE) if $\Pi^* : \rho \otimes \sigma \otimes \zeta_\rho \mapsto \tilde{\rho} \otimes \tilde{\sigma} \otimes \tilde{\zeta}_\rho$ has $(\rho, \sigma, \zeta_\rho)$ as a fixed point.*

3.6.1 Existence of MFE

Theorem 6. *There exists an MFE of ρ, the randomized policy $\sigma(x)$ at each surplus x and ζ, such that $\rho \in \Gamma_\rho$, $\sigma(x) \in \Gamma_\sigma$ and $\zeta \in \Gamma_\zeta$, $\forall a \in \mathcal{A}$ and $\forall x \in \mathbb{X}$.*

We will be specializing to the spaces $\Gamma_\rho, \Gamma_\sigma, \Gamma_\zeta$ and define the topologies being used in the following proofs first.

1. For the assumed action distribution $\rho \in \Gamma_\rho$ on the finite set $\mathcal{A}$, all norms are equivalent, we will consider the topology of uniform convergence, i.e., using the l_∞ norm given by $||\rho|| = \max_{a \in \mathcal{A}} \rho(a)$.

2. For the randomized policy $\sigma \in \Gamma_\sigma^{|\mathbb{X}|}$, we enumerate the elements in $\mathbb{X}$ as $1, 2, \cdots$, and consider the metric topology generated by norm $||\sigma|| = \sum_{j=1}^{\infty} 2^{-j} |\sigma(x_j)|$, where $|\sigma(x)| =$

$\max_{a\in\mathcal{A}} \sigma(x,a)$. We consider the convergence of any sequence $\{\sigma_n\}_{n=1}^{\infty}$ to σ in this topological space.

3. For the surplus distribution ζ on the countable set $\mathbb{X}$, we consider the topology of pointwise convergence, which can be shown to be equivalent to convergence in l_∞, i.e., uniform convergence, using coupling results presented in [69].

Note that from the definition of Γ_ρ, Γ_σ and Γ_ζ, they are already non-empty, convex and compact. Furthermore, they are jointly convex. Then in order to show that the mapping Π^* satisfies the conditions of Kakutani fixed point theorem, we only need to verify the following three lemmas.

Lemma 12. *Given ρ, by taking the best response given by (3.14), we can obtain the action distribution $\sigma(x)$ for every x, which is upper semicontinuous in ρ.*

Remark 1. *As we have discussed earlier, since our state space and action space are discrete, there might exist multiple best response actions when the agent solves the dynamic program. Thus, a pure equilibrium might not exist. Since the best response can be set-valued, we need to consider mixed strategies. In other words, the agent needs to choose a randomized action policy for each state. Hence, the randomized policy σ is critical in the construction of the MFE given in Definition 1.*

Lemma 13. *Given ρ and $\sigma(x)$, there exists a unique stationary surplus distribution $\zeta(x)$, which is continuous in ρ and $\sigma(x)$.*

Lemma 14. *Given $\zeta(x)$ and $\sigma(x)$, there exists a stationary action distribution ρ, which is continuous in $\zeta(x)$ and $\sigma(x)$.*

3.7 Characteristics of the Best Response Policy

In this section, we characterize the best response policy under the assumption that V_ρ in (3.14) has some properties. Then we discuss the relations between the incremental utility function $u(x)$ and the optimal value function V_ρ.

3.7.1 Existence of Threshold Policy

We make the assumption that given the action distribution ρ, $V_\rho(x)$ is increasing and submodular in x when $x \leq -l$; increasing and linear in x when $-l \leq x \leq w$; and increasing and supermodular in x, when $x \geq w$.

In Section 3.4, our lotteries are constructed such that the probability of winning monotonically increases with the cost of the action. This when combined with the monotonicity, submodularity (decreasing differences) for positive argument and supermodularlity (increasing differences) for negative argument of V_ρ yields the following characterization of the best response policy.

Lemma 15. *For any two action, say actions a_1 and a_2, suppose that $\theta_{a_1} > \theta_{a_2}$, so that $p_{\rho,a_1} > p_{\rho,a_2}$, i.e., $\phi(p_{\rho,a_1}) > \phi(p_{\rho,a_2})$, then there is a threshold value of the surplus queue for user such that preference order for the actions changes from one side of the threshold to the other.*

Using the same argument as Lemma 15, under the assumption that $V_\rho(x)$ is increasing and submodular in $x \in (-\infty, \infty)$, or increasing and supermodular in $x \in (-\infty, \infty)$, we can show the existence of a threshold policy.

3.7.2 Relations between incremental utility function $u(x)$ and the optimal value function V_ρ

3.7.2.1 Concave/Convex incremental utility function

Lemma 16. *Given the action distribution ρ, $V_\rho(x)$ is an increasing and submodular (i.e., decreasing differences) function of x if $u(x)$ is a concave and monotone increasing function of x, supermodular (i.e., increasing differences) function of x if $u(x)$ is a convex and monotone increasing function of x.*

Thus, from Lemma 16 and Lemma 15, the optimal policy takes a threshold form for both concave and convex incremental utility function.

3.7.2.2 Conjecture for S-shaped prospect incremental utility function

We found numerically that with an S-shaped utility function, the value function satisfies the super/sub-modularity conditions on the positive/negative axis respectively. If this holds true in

general, then from Lemma 15, the optimal policy would take a threshold form, and indeed this is what we observed numerically. However, we are not able to prove this result due to the implicit nature of the value function, and we can only conjecture that this condition might hold for some class of S-shaped utility functions.

3.8 Conclusion

In this chapter we developed a general framework for analyzing incentive schemes, referred to as nudge systems, to promote desirable behavior in societal networks by posing the problem in the form of a Mean Field Game (MFG). Our incentive scheme took the form of awarding coupons in such that higher cost actions would correspond to more coupons, and conducting a lottery periodically using these coupons as lottery tickets. Using this framework, we developed results in the characteristics of the optimal policy and showed the existence of the MFE.

We used the candidate setting of an LSE trying to promote demand-response in the form of setting high setpoints in higher price time of the day in order to transfer energy usage from a higher to a lower price time of day for an air conditioning application. We conducted data driven simulations that accurately account for electricity prices, ambient temperature and home air conditioning usage. We showed how the prospect of winning at a lottery could potentially motivate customers to change their AC usage patterns sufficiently that the LSE can more than recoup the reward cost through a likely reduced expenditure in electricity purchase. Further, we showed that a lottery is more effective than a fixed reward at enabling such desirable behavior and can attain a better tradeoff between social value and LSE profits.

Given the desirable analysis so far, we implemented a real system called *EnergyCoupon* and conduct experiment among customers in practice. We will present the experiment results and analysis in the following chapter.

4. INCENTIVE-BASED DEMAND RESPONSE: EMPIRICAL ASSESSMENT AND CRITICAL APPRAISAL*

4.1 Introduction

With the increasing penetration of renewable energy resources such as wind and solar, there is a growing interest on the demand-side management, or demand response (DR). DR has the potential to become a flexible resource to increase the reliability and efficiency to the power system. Reference [70] defines the demand response as the changes of end-consumers' electricity consumption in peak hours from their normal patterns. Many independent system operators in the U.S. such as Electric Reliability Council of Texas (ERCOT), New York ISO (NYISO), California ISO (CAISO) and ISO New England have operated day-ahead and real-time DR programs by means of providing energy, reserve and auxiliary services [71–73].

Demand response in the U.S originated in the 1970s, mainly due to the popularity of usage of central air conditioners [74]; Since then, a huge amount of DR programs are designed and implemented. This chapter categorizes the DR programs in two dimensions: 1) Direct load control or self-controlled (market-based) programs; and 2) the scale of target end-consumers (large industrial/commercial or small residential customers). Direct load control enables the DR operator (such as the utility) to remotely turn off or change the setpoint of the customers' equipments; in such a way, the amount of load shedding can be easily controlled within specified intervals, but it also causes problems in customers comfort and satisfaction (imaging your air conditioner is forced to turn off in hot summer/cold winter). Market-based DR programs tend to use prize signals or other incentives to encourage customers' self-motivated load control behaviors. Such programs usually have less ruin on the customers' comfort and satisfaction, but less precise when a specified amount of reduction needs to be achieved.

*Part of the results reported in this chapter is reprinted with permission of ACM, from EnergyCoupon: A Case Study on Incentive-based Demand Response in Smart Grid, Bainan Xia, Hao Ming, Ki-Yeob Lee, Yuanyuan Li, Yuqi Zhou, Shantanu Bansal, Srinivas Shakkottai, Le Xie, e-Energy, 2017; permission conveyed through Copyright Clearance Center, Inc.

In terms of customer scale, industrial or commercial customers are usually large profit-seeking entities; therefore they have advantages over small residential customers. Energy management systems are developed to help increasing the energy efficiency in data centers, retails, telecoms etc and coordinate with market based signals (such as real-time electricity and gas prices) [75]. On the other hand, residential customers pay more attention on personal habit and comfort; therefore, some price-based mechanisms (such as time-of-usage (TOU), critical peak pricing (CPP) [76] and market-index retail plans offered by the utility) do not have significant impact to the majority of residential end-consumers with fixed-rate retail plans. Given the fact that residential takes the most electricity consumption in the U.S (38%, compared with commercial 37% and industry 25%) [75], the potential of DR in residential is far from fully explored. Table 4.1 summarized some current researches and implementations in different categories of demand response.

Table 4.1: Classification of Demand Response Programs

	Direct load control (centralized)	Market-based (price/incentive-based)
Large customers	Researches [77–79], direct load control programs [80]	Energy-management systems [75]
Small residential	Direct load control programs [80]	Variable rate retail plans [81], CPP [76], EnergyCoupon [3]

Despite the research [76,82] and commercial programs (ENERNOC [83], Ohmconnect [84]) of price-based DR, an alternative market-based solution to residential DR programs named coupon-incentive based demand response (CIDR) aims at providing coupon incentives to reduce the electricity consumption of residential end-consumers during peak hours [85–87]. Compared with the above traditional DR programs, this mechanism has the following advantages: purely voluntary,

penalty-free to customers, and can be implemented to the majority of residential end-consumers who face fixed retail plans. A program named *EnergyCoupon* is the first-of-its-kind implementation of CIDR, with additional inherited innovations: 1) it provides dynamic DR events to end-consumers with *individualized targets*; 2) *periodic lotteries* are designed to convert coupons earned in DR events into dollar-value prizes. A small-scale pilot project was conducted in 2016, and we found clear load profile change of the residential customers [3].

In terms of 2) *periodic lotteries*, a lot of researches and commercial programs shows how the "nudge engine", games and lotteries help to encourage the desired behaviors of human beings. Reference [88] tries to discover the social value of energy saving, [87] models the CIDR system as a two-stage Stackelberg game, and [89–91] use the "mean field games" to describe end-consumers' behaviors in the DR program with lottery-based incentives. On the other hand, the lottery scheme has already been implemented on encouraging the uniform load on public transportation [92] and relieving congested roadways [93]; however, there are few practical works, including some ongoing experiments [94, 95], trying to adopt the lottery scheme on DR programs.

Built upon our previous studies, a larger-scale experiment, which is closer to the real world, was carried out in 2017 with much more comprehensive designs and critical assessments[1]. The improvements in experiment ('17) include but not limited to: 1) an extra comparison group for data analysis and comparison; 2) an improved baseline algorithm ("similar day"); 3) the treatment group divided into two subgroups facing fixed and dynamic DR events. More facts and comparisons between two experiments are listed in Table 4.2. We will show in later sections that these changes help to analyze end-consumers' behaviors in-depth.

The main contributions *EnergyCoupon* are as follows:

1. Providing price and baseline prediction algorithms suitable for DR programs;

2. Systematically documenting the experiment design, collection, and posterior analysis of on residential customers;

[1]Unless otherwise specified, in the remaining part of this chapter, "experiment ('16)" refers to our previous study conducted in the year 2016 and "experiment ('17)" refers to the new one in 2017.

Table 4.2: Overview of EnergyCoupon Experiments in Year 2016 and 2017

Year	2016	2017
Experiment length (weeks)	12	12
Treatment group size	8	29
Comparison group existence	No	Yes
Baseline algorithm	Hybrid	"Similar day"
Active subjects defined by	Energy saving	Lottery participation
Active subjects No.	3	7

3. Experimental result shows the trend of load shedding/shifting effects, different behaviors over fixed/dynamic coupon targets, financial benefits of the LSE and end-consumers, impact of periodic lotteries on human behaviors, as well as the effective cost saving of *EnergyCoupon* over traditional DR programs.

This chapter is organized as follows: Section 4.2 introduces the system architecture and the interface of the EnergyCoupon App. Key algorithms including price prediction, baseline prediction, individualized target settling and lottery are explained in Section 4.3. Experimental design is described in Section 4.4, and data analysis is shown in Section 4.5. We finally conclude our findings in Section 4.6.

4.2 System Overview

The EnergyCoupon system is designed to inform the end-consumers about the upcoming DR event along with individualized targets, measure the demand reduction during the DR event, provide statistics and tips for energy saving, as well as operate periodic lotteries. Fig. 4.1 exhibits the system architecture of EnergyCoupon. As the core component in the architecture, a SQL database is hosted on a server running 24 hours a day, interacting with the data resources (shows in blue

blocks), mathematical algorithms (green blocks) and the lottery scheme (pink blocks). An EnergyCoupon App (both Android/IOS versions available) is developed and installed in the mobile phones of the treatment group subjects. The app (interface shown in Fig. 4.2) receives coupon targets, tips and statistics, and is also used to submit coupons in the lottery. A brief overview of the remaining components is as follows:

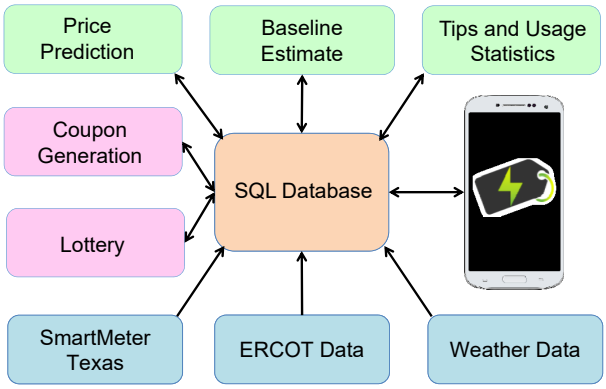

Figure 4.1: System architecture. Reprinted with permission from [3].

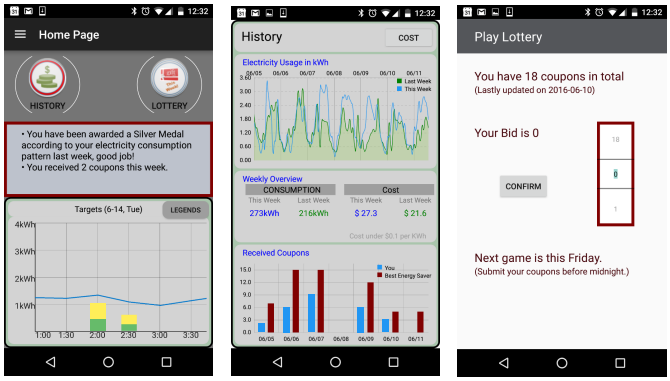

Figure 4.2: EnergyCoupon app interface. (a) Main page, coupon targets and tips (b) Usage statistics (c) Lottery interface. Reprinted with permission from [3].

(1) *SmartMeterTexas*: The source of the electricity consumption of all end-consumers in 15-min resolution [96]. Data is used in baseline prediction and coupon target generation algorithms in Section 4.3.2 and 4.3.3.

(2) *ERCOT Data*: The source of day-ahead and real-time market prices, as well as the system load in ERCOT area [30]. Data is used in the price prediction algorithm described in Section 4.3.1.

(3) *Weather Data*: The source of weather information used in the price (Section 4.3.1) and baseline prediction algorithms (4.3.2).

(4) *Price Prediction*: An algorithm with the purpose of determining whether a dynamic DR event should be announced two hours in advance. This algorithm is introduced in detail in Section 4.3.1.

(5) *Baseline Estimate*: An algorithm with the target of predicting the "normal consumption" of the end-consumer without DR. This algorithm is designed to eliminate the gaming effects described in [97] while keeping a relatively high accuracy. Details are included in Section 4.3.2.

(6) *Tips and Usage Statistics*: Usage statistics are provided including the estimate of the number of coupons the user is likely to win, and his/her behavior compared with neighbors. Personalized tips to save energy are generated based on his/her usage statistics.

(7) *Coupon Generation*: DR events are determined according to price prediction, and personalized target are generated using baseline estimate. See Section 4.3.3 for details.

(8) *Lottery*: Periodic lotteries enable the end-consumers to convert his/her coupons earned into dollar-value gifts. See Section 4.3.4 for more details.

4.3 Algorithms

In this Section, we will elaborate the key analytics behind the experiment ('17). These include the price prediction, baseline estimate, coupon generation, and lottery. These analytics are

important not only to this experiment, but also to any other possible means of demand response mechanisms.

4.3.1 Price Prediction

We would like to incentivize load shift from peak (price) hours to non-peak (price) hours for end-customers. In order to run our system in real-time, we must be capable to predict the high price occurrences ahead of time. Regarding the topic on electricity price prediction, various work has been carried out so far. Time series models are used to predict day-ahead electricity prices in [98, 99]. A combination of wavelet transform and ARIMA model is considered in [100]. A hybrid method mixed with time series and neural network is discussed in [101]. In [102], spot price prediction is conducted with load prediction and wind power generation involved. We have different concerns on price prediction given the specific requirements of EnergyCoupon system, since there is less interest in the exact number of the price. Instead, the label of the price, either high or low, is more valuable. Moreover, time series techniques have good performance in handling data with repeating periods, i.e. 24 hours, and achieve high accuracy in predicting the following successive samples. The high prices we target in our situation do not have such behavior. Also, as an online algorithm, low computing complexity is critical. Together with all the concerns, we design and deploy a specific decision tree to deal with the price prediction in our system.

Decision tree is a well-known classifier with selected features in non-leaf nodes and labels in leaf nodes. Different from traditional ones, we have unbalance error concerns in our EnergyCoupon system, as a false high price alert will not induce much loss since the total budget is controlled by lottery prizes. However, a failure in actual high price catch may cause loss of efficiency. Hence, our decision tree should have higher tolerance in false positive errors than false negative ones. This can be captured by adjusting the penalty ratio between two kinds of errors in training stage, though one must be careful in doing so due to the risk of overfitting the training set.

Considering the DR procedure conducted by our system, we believe a 2-hour time window is reasonable for participants to react. Given the target time slot to be 2-hour in advance, there are enormous features in both spatial and temporal space to choose from. Since air-conditioning

dominates household electricity consumption in Texas and weather has an impact on renewable energy availability as well, we finalize five fundamental feature classes: Price(π), Demand(P), Temperature(T), Humidity(H) and WindSpeed(W). Furthermore, we choose the temporal offsets in each feature class according to the self/cross-correlation between the feature and the price label. A numerical study on high price appearances based on different thresholds is carried out to choose a proper threshold in our study, so as to label data samples. Table 2 in reference [3] shows the prediction accuracy over 90% in validation.

Details on training data preparation, feature selection and performance evaluation are beyond the scope of this chapter. readers may refer to [3] for more information.

4.3.2 Baseline Estimate

As defined by the U.S. Department of Energy, baseline is the "normal consumption pattern" by end-consumers without the impact of DR [70]. Daily baseline prediction algorithm is with crucial importance in our EnergyCoupon program, since it affects the energy reduction measurement, and the number of coupons issued to participants. Energy reduction for an end-consumer i on interval k in a particular day D, $P^D_{DR,i}(k)$ is calculated as the difference between the the subject's predicted baseline $P^D_{base,i}(k)$ and his/her real electricity consumption $P^D_{real,i}(k)$ (as shown in (4.1)); $P^D_{real,i}(k)$ can be measured by the smart meter installed in the his/her household with high reliability.

$$P^D_{DR,i}(k) = P^D_{base,i}(k) - P^D_{real,i}(k). \tag{4.1}$$

Reference [3,97,103] address the limitations of conventional baseline estimate algorithms used by the ISOs [71, 72] considering baseline manipulation and user's dilemma. The "similar day" algorithm was sprouted in our previous work [3].

Deriving from k-nearest neighbors algorithm (k-NN) and kernel regression [104, 105], this proposed "similar day" algorithm 1) predicts the baseline by matching the targeted 6-hour time window with historical windows having the same *time of day* and similar *ambient temperature* (measured by Euclidian distance, see 4.2); 2) efficiently eliminates the gaming effect of participants

by avoiding using end-consumers' consumption during the experiment.

$$T_{MSE}^{D,l,t} = \frac{1}{N_t}\sum_{k=1}^{N_t}(T^{D,t}(k) - T^{l,t}(k))^2, \tag{4.2}$$

where D, l represents the index of the target day and a particular historical day, $t \in \{1, 2, 3, 4\}$ in the index of time window, N_t is the number of samples in each section. Therefore, the day-ahead baseline in this section is calculated as the average consumption of all corresponding N_s "similar days" for the **same** end-consumer,

$$P_{base,i}^{D,t}(k) = \frac{1}{N_s}\sum_{l=1}^{N_s} P_{base,i}^{l,t}(k). \tag{4.3}$$

Previously the "hybrid" method is adopted as the baseline estimate algorithm in experiment ('16) [3]. However, later we discovered that this algorithm neither 1) eliminated gaming effects, nor 2) kept a small prediction error due to large diversity among residential end-consumers. Therefore, the "similar day" algorithm is used in *both baseline estimate and data analysis* in the recent EnergyCoupon experiment in 2017.

4.3.3 Individualized Target Settling and Coupon Generation

There are two types of DR events in the EnergyCoupon program: "fixed" and "dynamic". Both types of event last for 30 min, and only appear during 1-7 pm every day. However, the major difference of two types of event lies in the way to determine the time period. Through price statistics in ERCOT real-time market [30], "fixed" events are pre-determined on the periods that have highest probability to have peak prices. In contrast, "dynamic" events are triggered when the 2 hour-ahead predicted price (calculated using the algorithm in Section 4.3.1) is higher than $\$50/MWh$. "Fixed" event periods are unchanged during the month, with up to no more than 3 times in a day; while there is no restriction for the number of "dynamic" events. These two types of DR events are collectively referred to as "hybrid" event.

After the time period of a DR event is determined, a multi-layer coupon target is generated only

depending on individual predicted baseline, regardless of the specific event type. Fig. 4.3 shows the multi-layer structure of the coupon target. When a particular end-consumer reduces his/her electricity consumption to 70% of the predicted baseline (yellow region), he/she will be awarded 2 coupons; when he/she further reduces the usage to the green region (more than 70% reduction from baseline), he/she will be awarded 5 coupons.

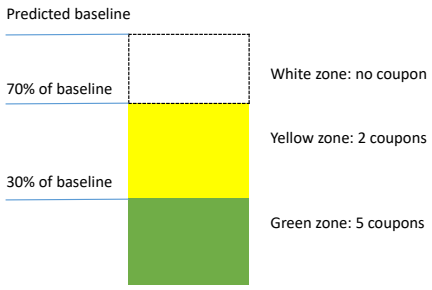

Figure 4.3: Individual target setting. Reprinted with permission from [3].

Fig. 4.4 summarizes the logic flow of a coupon target generated based on algorithms introduced in Section 4.3.1 to 4.3.3.

4.3.4 Lottery Algorithms

Due to the framing effect and the prospect theory [106–109], lottery scheme is considered to provide incentives to desired human behaviors when there is a large group and each person contributes a small impact. In our experiment, weekly lotteries are provided to convert end-consumers' earned coupons into dollar-value prizes. Three amazon gift cards with face value $\$20$, $\$10$ and $\$5$ are issued to the lottery winners every week. A participant is allowed to bid any positive number of coupons (no more than the number he/she has) in each lottery; the more coupons he/she bids, the higher probability he/she will win the prize. End-consumers are also allowed to collect the remaining coupons for future lottery bids.

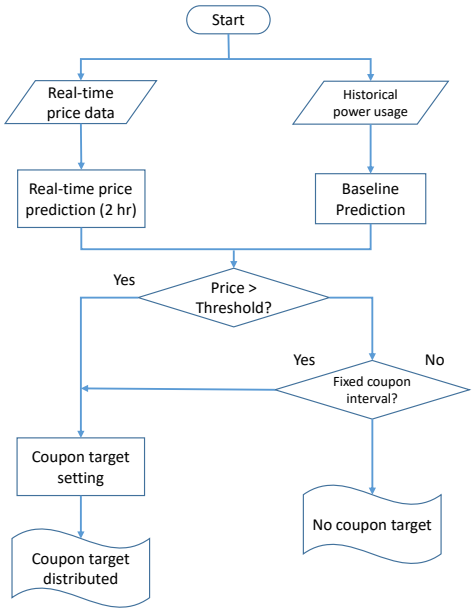

Figure 4.4: EnergyCoupon algorithm flow chart. Reprinted with permission from [3].

4.4 Experiment Design

4.4.1 Brief summary of Experiment ('16)

A small-scale preliminary EnergyCoupon experiment was carried out between June and August in 2016, and 7 end-consumers were recruited in the same residential area in Cypress, Texas. Each subject received a number of 30-min-length DR events along with coupon targets, between 1-7 pm every day, and they are allowed to participate in lotteries with $\$35$ amazon gift card in total every week. Peak time estimate, individualized target settling, coupon generation and lottery scheme followed the algorithms introduced in Section 4.3. "Hybrid" method was used in baseline estimate, and "similar day" was used in posterior data analysis with normalization.

Experimental result shows a load shifting from peak to off-peak hours; it yields substantial savings for the LSE, about $\$0.44/(week \cdot user)$ on average, and $\$1.15/(week \cdot subject)$ for active subjects. Readers can refer to [3] for more details.

4.4.2 Subjects in Experiment ('17)

A larger-scale EnergyCoupon experiment was conducted in the summer of 2017, with 29 anonymous residential end-consumers in Woodland, Texas recruited to form the treatment group. All these end-consumers are employees of a local utility company named *MP2 energy*. The experiment was purely voluntary and end-consumers were free to quit the experiment at any time.

In addition, the electricity consumption data for another 16 anonymous residential end-consumers is also provided by MP2 energy. Those end-consumers form the *comparison group*. The relationship between the treatment and comparison group is exhibited in Fig. 4.5(a).

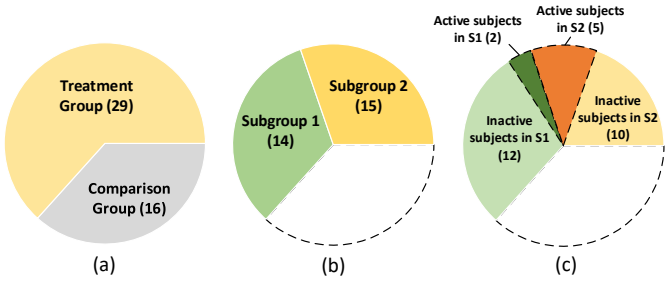

Figure 4.5: Subjects in experiment ('17). (a) Treatment vs. comparison group (b) Subgroup 1 vs. Subgroup 2 (c) Active vs. Inactive subgroups. Numbers in brackets are group sizes.

4.4.3 Procedure in Experiment ('17)

The experiment lasted for 12 weeks from Jun 10, 2017 to Sep 1, 2017. The treatment group subjects were asked to create an account for SmartMeterTexas.com in order for the server to track their energy consumptions during the experiment. In Week 0 (Jun 10-Jun 16, 2017), the treatment group subjects were asked to download and install the EnergyCoupon App, get familiar with interfaces, practice to make energy reduction following individualized coupon targets and participate into the lottery. The electricity consumption data during this period of time is neither considered as the experimental data, nor used as the historical data in baseline estimate.

During the experiment, the treatment group subjects were able to see the coupon targets at

least 2 hours prior to the DR event. A subject who wanted to save energy only needed to turn off/change the setpoints of his/her appliances during the 30-min-length DR event period. The subject's electricity consumption would be recorded by the smart meter installed in the house and data would be available to the server within 36 hours. Each subject would be awarded coupons based on his/her coupon target achievement during the DR events.

In the first three weeks (Jun 17, 2017 to Jul 7, 2017), all the subjects in the treatment group were faced with *"hybrid"* coupon targets; Starting from Jul 8, 2017, and till the end of the experiment, those subjects were randomly assigned to two subgroups (**Subgroup 1 and 2**, or **S1** and **S2** for short) with almost the same scale (14 subjects in S1 and 15 in S2). S1 and S2 subjects only received "fixed" and "dynamic" coupon targets separately (Fig. 4.5(b)).

The "similar day" algorithm is used in baseline estimate, and coupon target generation follows the algorithm in Section 4.3.3. All DR events were generated within 1-7 pm every day.

Weekly lotteries were provided during the experiment, with each lottery cycle beginning at 12:00 am on Saturday and till 11:59 pm on the next Friday. Lotteries are designed following the algorithm explained in Section 4.3.4. For the analysis purpose, at the end the the whole experiment, all subjects in the treatment group are assigned into another two subgroups according to their lottery engagements. Subjects who participated *at least 5 out of totally 11 lotteries* (7 subjects) are called *active* subjects and assigned to the "Active" subgroup, and the remaining treatment group subjects (22 subjects) are regarded as *inactive* subjects and are assigned to the "Inactive" subgroup. As Fig. 4.5(c) shows, 2 active subjects belong to S1 and 5 belong to S2, and 12 of the inactive subjects belong to S1 and 10 belong to S2.

As we have briefly described in Section 4.1, there are major differences between the designs of EnergyCoupon experiment ('16) and ('17). Change of the algorithm and removal of normalization in baseline estimate help to increase the prediction precision and eliminate gaming effect, the availability of comparison group provides an alternative in measuring energy saving for the treatment group, and the assignment of S1 and S2 helps to reveal more behaviors of end-consumers.

4.5 Data Analysis

Through experiment ('17) our server has collected electricity consumption data for all the subjects in the treatment and comparison groups. In this section, the calculation and discussion on the energy saving for the treatment group is from subsection 4.5.1 to 4.5.2; Financial benefit for this experiment is estimated in 4.5.4, and behavioral changes due to lottery is investigated from subsection 4.5.5 to 4.5.7.

4.5.1 Energy Saving for the Treatment Group

There are two ways to measure the electricity reduction for the treatment group during the experiment: compare the real electricity consumption with (i) historical consumption data or (ii) consumers' baseline. Fig. 4.6 exhibits the *energy consumption ratio* of the treatment and comparison groups following method (i). The ratio is defined as the group's weekly consumption between 1-7 pm divided by historical consumption in 2016, and a lower ratio indicates a higher energy reduction level during peak hours.

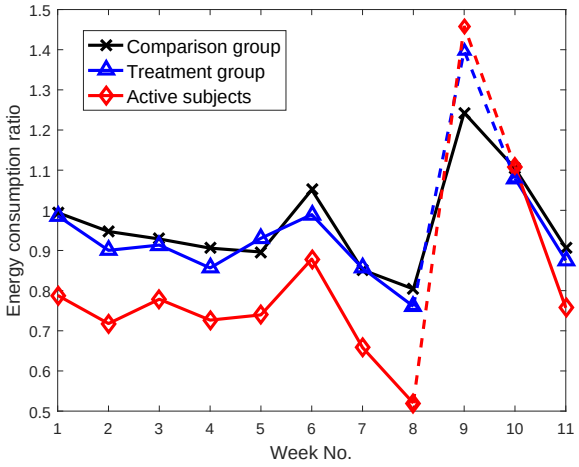

Figure 4.6: Energy saving at 1-7 pm for treatment and comparison group during the experiment (2017), based on the consumption of same days in 2016 [2]

[2]Historical consumption data for some treatment group subjects in Week 9, 2016 is not available; We used dashed lines to show the less reliable trend between Week 8-9 and 9-10.

Fig. 4.6 shows the energy saving for the active subjects during the experiment. While the ratios for inactive and comparison groups are around 1 in most weeks during the experiment, there is clear gap between the active subjects (red curve) and these two groups. Active subjects consume less electricity in most weeks compared with their consumption in 2016, with the maximum saving around 40% in Week 8.

The disadvantage of method (i) is that we cannot tell the exact energy saving, since there are variables that change between the year 2016 and 2017. Energy saving will be measured using method (ii) in the next subsection.

4.5.2 Comparison between Active and Inactive Subjects in Treatment Group

As introduced above, method (i) calculates energy consumption ratio using the estimated baseline as the reference, instead of historical consumption in 2016. Fig. 4.7 shows the energy saving ratios of active, inactive subgroups and the whole treatment group.

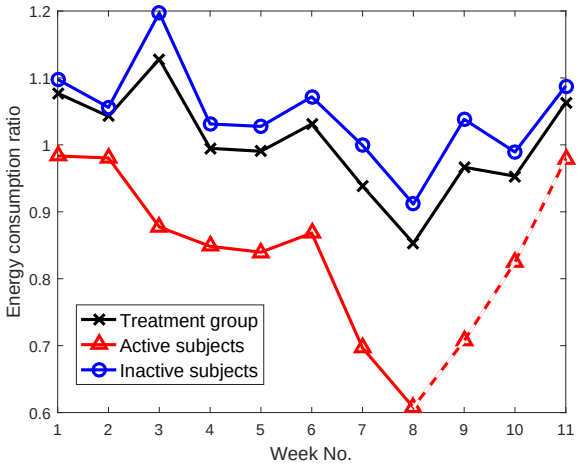

Figure 4.7: Energy saving at 1-7 pm for active/inactive subjects by week, based on their baseline consumptions

The performance of the inactive subgroup is quite constant, with the ratio around 1.0 in most of the weeks, and never falls below 0.9. This is accord with our intuition that less engagement in

lottery is a sign of lack of enthusiasm about energy saving in the EnergyCoupon program. Since inactive subjects take the majority of the treatment group (as shown in Fig. 4.5), the gap between the blue and black curves are minor.

The curve for the active subgroup is far below the other two curves, indicating the energy saving and load pattern change for active subjects during the experiment. Energy saving for the active subgroup gradually increases in the first few weeks and reaches the peak at about 40% in Week 8. After Week 9, the saving begins to decline, until reach only 10% in Week 11. The rebound of the ratio can be explained by the Harvey hurricane arrived in Week 10 and 11, which may distract the subjects from the DR program.

To better visualize the load pattern change for the active subjects, 24-hour average real consumptions (red curves) are illustrated for the active and inactive subgroups *in a particular week* (7/29-8/4/2017), and the corresponding baselines (blue curves) are set as references (Fig. 4.8). Energy saving during 1-7 pm (interval 27-38) for active subjects can be calculated as 28.9%, while that of inactive subjects is only -0.2%. The close-to-zero energy saving for inactive subjects is unsurprising, and also proves the precision of our baseline estimate from another prespective; However, the surprising finding from Fig. 4.8(a) is the load shedding effect in non-peak hours (25.0% energy saving). This observation conflicts with our previous assumption of pure load shifting in [3], and it can be explained by the assumption that there is "inertia" in demand response; incentivized energy reduction in peak hours would influence that of off-peak hours.

4.5.3 Comparison between Subjects in Treatment Group Facing Fixed/dynamic Coupons

Starting from Week 3 and till the end of the experiment, the treatment group subjects are randomly assigned into two subgroups S1 and S2 facing "fixed" and "dynamic" coupon targets seperately. We aim to discover the effectiveness of different types of coupon targets on energy saving of end-consumers. Energy savings for these two subgroups during 1-7 pm are exhibited in Fig. 4.9(a).

As observed from Fig. 4.9(a), the active subjects in S1 and S2 are not homogeneous, as S1 on average saves 35% and S2 saves -5% in Week 1-2. However, a clear "activation" is observed for

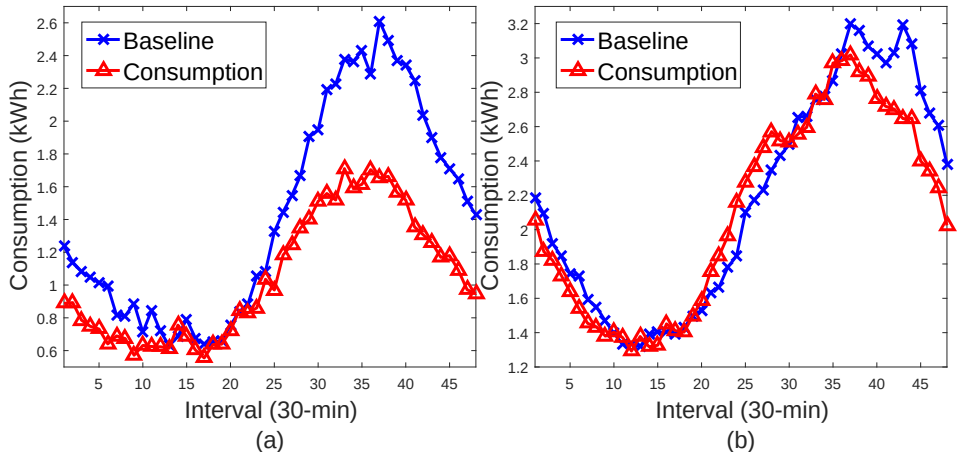

Figure 4.8: Daily consumption vs. baseline for active/inactive subjects (7/29/2017-8/4/2017). (a) Active subjects (b) Inactive subjects

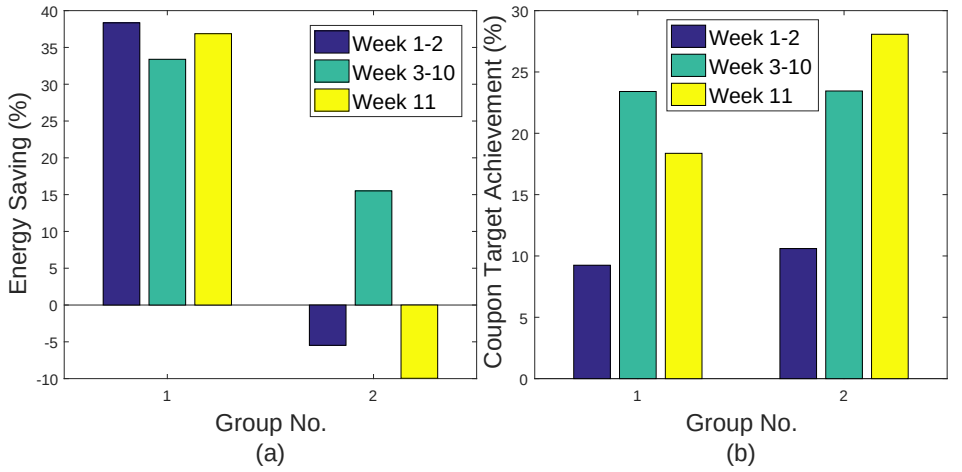

Figure 4.9: Behavior comparisons between subjects facing fixed (Subgroup 1) and dynamic coupons (Subgroup 2). (a) Average energy saving at 1-7 pm (b) Coupon target achievement percentage

S2 subjects, as their energy saving jumps from -5% to -10% in Week 3-10; while no such effect is observed for S1 subjects. In week 11, the energy saving for S2 subjects returns to their initial level, which can be attributed to the hurricane.

Fig. 4.9(b) illustrates the coupon target achievement ratios of active subjects in two subgroups. The ratio is defined as the proportion of DR events that the subjects at least earn one coupon

(reduce at least 30% energy from their baseline). Comparing Fig. 4.9(a)(b), an interesting finding is that although S1 has higher energy saving than S2 in all periods, both subgroups reach a similar level of coupon achievement.

One possible explanation for our observation is that S1 subjects facing "fixed" DR events would prefer to program their home appliances (such as AC) in advance to hit coupon targets, and do not change their setpoints frequently. This results in a decrease of electricity consumption at around 1 pm and the rebound at 7 pm, which clearly is an overreaction to coupon targets. At the same time, S2 subjects facing "dynamic" DR events have to check the app more regularly, since dynamic coupon targets only appear within 2 hours in advance. Therefore, they are more aware of the coupon targets, and can hit the targets and earn coupons more "efficiently" with minimum energy reduction. To support our explanation, Fig. 4.10 provides the load patterns of two active subjects in S1 and S2 for a particular day.

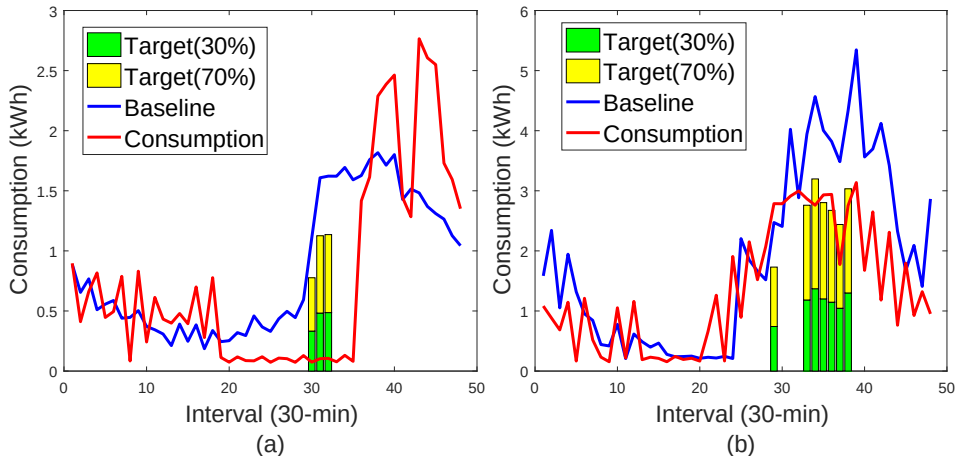

Figure 4.10: Energy consumption curve for two active subjects on 7/19/2017. (a) Subject No.19 (Subgroup 1) (b) Subject No.18 (Subgroup 2)

4.5.4 Financial Benefit Analysis

Previously we assume that all the subjects perform *pure load shifting* from peak to off-peak hours [3]; therefore, the DR program will bring financial benefits to both the LSE and active end-

consumers, since LSE could reduce their purchase in peak price hours, while not losing its retail revenue, and end-consumers win extra rewards from the LSE for their energy saving behaviors during peak hours. However, our finding in Section 4.5.2 conflicts with the *pure load shifting* assumption. Therefore, it's not obvious whether the LSE and end-consumers reach a win-win situation.

The net benefit for the LSE consists for three parts: (a) the saving in electricity purchase in high-price hours (b) the decrease of sales revenue due to the load shedding effect and (c) the cost of rewards issued to lottery winners. Our calculation shows the saving in three parts is $\$2.6$, $\$-2.7$ and $4.0/(week \cdot activesubject)$. Due to the load shedding effect, the loss in (b) offsets the benefit in (a) and the LSE suffers a net loss close to $4.0/(week \cdot activesubject)$.

Table 4.3: Financial Benefit of the LSE and Active Subjects

Subjects	LSE	Active subjects
Savings ($\$/(week \cdot subject)$)	$2.6 - 2.7 - 4.0$	$4.0 + 2.7$

An active subject, in contrast, on average receives $\$4.0$ lottery rewards per week from the LSE; at the same time, the load shedding effect leads to the decrease of his/her electric bill by around $\$2.7$ per week. Therefore, our EnergyCoupon program brings positive financial benefit to active subjects.

Although this experiment did not bring a win-win situation to both the LSE and end-consumers, it still increases the social welfare on the demand side, as the summation of benefits is positive ($\$2.6/(week \cdot subject)$). We can also summarize that the financial benefit of LSE in the DR program is closely related to the level of load shifting in the subjects' behavior; if the load shift is minor, cost saving from its purchase may not cover the loss of retail revenue, which leads to a net financial loss to the LSE.

4.5.5 Influence of Lottery on Human Behaviors

As discussed in Section 4.3.4, the lottery scheme is considered to provide incentive to desirable behaviors of the treatment group. Table 4.4 lists some influences of the lottery on the participants' behaviors.

Table 4.4: Subjects' Behavior Change due to Lottery

Subjects (on average)	Energy saving improvement (1-7 pm) (%)	Next lottery participation prob. (%)	Porb. of ≥ 1 participation in next 3 lotteries (%)
All winners	10.7	56.6	80.5
Prizeless participants	-0.03	40.0	70.0

The first column in Table 4.4 shows that winning a lottery prize has a positive impact on the energy saving, as lottery winners make an energy saving improvement of 10.7% in the next lottery cycle on average (for example, 10% to 20.7%). In contrast, the average energy saving improvement for prizeless participants is close to zero (-0.03%). The second and third columns clearly demonstrate that lottery winners on average tend to have higher engagements than prizeless participants in the next lottery (56.6% to 40.0%) and next three lotteries (80.5% to 70.0%). Therefore, we can summarize that the lottery prize has positive impacts on both energy saving and lottery engagements in future lottery cycles.

4.5.6 Comparison with previous CPP Experiment

In this subsection, we compare our EnergyCoupon experiment with previous price-based DR experiment conducted by Prof. Wolak [76], which was carried out in Anaheim, CA in the year of 2005. Critical peak pricing (CPP) was used in this experiment; CPP days are selected based

on some price prediction algorithm, and on CPP days during noon-6 pm, subjects in the treatment group receives $\$0.35$ for every kWh reduction from baseline. Some comparisons of these two experiments are listed in Table 4.5.

Table 4.5: Comparison between EnergyCoupon and Previous Experiments

Compare items	Critical Peak Pricing	Energy-Coupon
Category	Price-based DR	Incentive-based DR
Treatment group scale	71	29
Experiment length	Jun-Oct	Jun-Aug
Number of DR days	Certain CPP days (12)	Regular (77)
Peak hours	Noon-6 pm	1-7 pm
Energy reduction	12%	10.7%[a]
Effective cost compared with retail price [b]	368%	58.8%

[a] Electricity reductions for active and inactive subjects are 24.8% and 7.36%, respectively.

[b] We choose retail price in Anaheim in 2005 as $\$0.095/kWh$ [110], and average retail price in Woodland, TX in 2017 as $\$0.090/kWh$ [81].

We can observe that our experiment has reached a similar level of energy reduction with the CPP experiment (12% to 10.7%). Since our EnergyCoupon provides DR events every day compared with only 12 CPP days in their experiment, EnergyCoupon project helps to save energy more efficiently than the CPP experiment.

Moreover, effective cost is defined as the cost of reducing $1\ kWh$ of electricity during peak hours. This value in our experiment is calculated by the total value of lottery prizes divided by the energy reduction for all treatment group subjects. Data analysis shows that effective cost in our experiment ($\$0.053/kWh$) is only $1/7$ of that in the CPP experiment.

4.5.7 Cost Saving Decomposition

In this subsection we would like to estimate how the lottery scheme contributes to the effective cost saving in our EnergyCoupon experiment, compared with previous CPP experiment described in Section 4.5.6. We split and model the contributions of (i) the lottery scheme and (ii) other features of EnergyCoupon (CIDR, mobile app) as

$$C_{CPP} = \alpha\beta C_{EnergyCoupon}, \tag{4.4}$$

where $C_{(\cdot)}$ represents the effective cost shown in Table 4.5 ($C_{CPP} = 0.35$, $C_{EnergyCoupon} = 0.053$), and α, β are the contributions of the above two factors ((i), (ii)), modeled as multipliers to the effective cost of EnergyCoupon.

The value α can be estimated using cumulative prospect theory. As a behavioral game theory, this theory describes the individual choice between risky probabilistic alternatives [107]. It models the probability weighting and loss aversion, which leads to the overweighting of small probabilities and underweighting of moderate and high probabilities. A gain prospect $f = (x_i, p_i)$ describes a prospect results in the multiple outcome $x_i, x_i < x_j$ iff $i < j$ with probability p_i, and $\sum_i p_i = 1$. This theory defines a certain equivalent f as

$$V(f) = \sum_{i=0}^{N} \pi_i V(x_i), \tag{4.5}$$

where $V(\cdot)$ is the utility function, π_i are decision weights calculated by

$$\pi_i = w(p_i+, ..., +p_n) - w(p_{i+1}+, ..., +p_n), 0 \leq i < N, \tag{4.6}$$

and $w(\cdot)$ is the probability weighting function.

As a rough estimate, according to the coupons they have earned, on average each active subject has approximately 7.0% chance to win each prize ($\$20$, $\$10$ and $\$5$) in the weekly lottery; the probability for an inactive user to win each prize is around 2.3%. Therefore, the prospect each active/inactive subject faces (f_a and f_b) can be described as

$$\begin{aligned} f_a =&(0.79, \$0, 0.07, \$5, 0.07, \$10, 0.07, \$20), \\ f_b =&(0.931, \$0, 0.023, \$5, 0.023, \$10, 0.023, \$20). \end{aligned} \tag{4.7}$$

Since each lottery prize is relatively small ($< \$200$), the utility function $V(\cdot)$ is linear and can be eliminated from both sides of (4.5) [109]; thus the certain equivalent can be calculated as

$$\begin{aligned} f_a^e =&w(0.07) \times (5 + 10 + 20) = \$5.25, \\ f_b^e =&w(0.023) \times (5 + 10 + 20) = \$2.28. \end{aligned} \tag{4.8}$$

The value of $w(\cdot)$ comes from the estimate in [107]. Equation (4.8) shows the estimate of direct cash needed in our experiment to maintain the same level of incentive to the treatment group, if no lottery scheme is adopted ($f_a^e \times 7 + f_b^e \times 22 = 86.91$). Therefore, multiplier α is estimated as the ratio of equivalent cash divided by total weekly lottery prizes $\alpha = 86.91/35 = 2.48$. According to (4.4), $\beta = 2.66$ and we can conclude that lottery scheme and other EnergyCoupon designs have *equal* levels of contribution to reducing the effective cost.

4.6 Concluding Remarks

This chapter presents the design and critically assesses the empirical experiment of coupon incentive-based demand response for end-consumers over a two-year period in Houston area. Different from traditional price-based DR programs, EnergyCoupon has the following features: (1) Dynamic time-of-the-day DR events and individualized coupon targets; (2) End-consumers receive coupon targets and usage statistics through mobile app; (3) Periodic lottery allows to convert coupons into dollar-value prizes.

Data analysis shows that there is significant load shedding effect for the treatment group, but not much load shifting effect is observed. In addition, there are incentives of lottery prizes on desirable behaviors such as energy saving improvement and lottery participation. Our experiment has much lower DR cost ($¢5.3/kWh$) compared with previous CPP projects ($¢35.0/kWh$); Using prospect theory we estimate that the design of system architecture and lottery scheme have equal contributions to the cost saving.

This chapter is generalizable towards other Internet-of-Things-enabled demand response activities, and could shed light on the overall discussion of incentive-based versus price-based demand response.

Future work would examine the value added from obtaining the consumer behavior data in this experiment. Another possible venue of future work is to further develop a platform that allows for the end users to aggregate and participate in wholesale level ancillary services.

5. CONCLUSION

In this thesis, we explored the market methods to approach resource management problem for both supply and demand in the Smart Grid. In Chapter 2, we designed a bilateral energy sharing market which maximizes the resource utilization through high trade ratio with a unified price and evaluated the system by a PV market case study with parameters chosen from realistic statistics. We further provided an extension of such secondary market in price anticipating scenario, which could provide stochastic energy service in the Smart Grid. In Chapter 3, we developed a coupon incentive-based demand response market to encourage end-customers to shift their electricity usage from peak hours to off-peak hours. We simulated our system with Texas data and showed that a LSE can potentially attain substantial savings using our scheme. We modeled both market systems in a Mean Field Game framework and showed the existence of desirable equilibria. In last chapter, we presented the design and implementation of a real world experiment system as we discussed in Chapter 3. The experiment results indicated that to achieve same amount of consumption reduction, the effective cost of demand response providers can be reduced substantially through our system compared with other demand response schemes.

REFERENCES

[1] B. Xia, S. Shakkottai, and V. Subramanian, "Small-scale markets for bilateral resource trading in the sharing economy," *IEEE INFOCOM*, 2018.

[2] J. Li, B. Xia, X. Geng, H. Ming, S. Shakkottai, V. Subramanian, and L. Xie, "Mean field games in nudge systems for societal networks," *ACM Trans. Model. Perform. Eval. Comput. Syst.*, vol. 3, pp. 15:1–15:31, Aug. 2018.

[3] B. Xia, H. Ming, K.-Y. Lee, Y. Li, Y. Zhou, S. Bansal, S. Shakkottai, and L. Xie, "Energy-coupon: A case study on incentive-based demand response in smart grid," in *Proceedings of the Eighth International Conference on Future Energy Systems*, pp. 80–90, ACM, 2017.

[4] N. R. Darghouth, G. Barbose, and R. Wiser, "The impact of rate design and net metering on the bill savings from distributed pv for residential customers in California," *Energy Policy*, vol. 39, no. 9, pp. 5243–5253, 2011.

[5] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japan Journal of Mathematics*, 2007.

[6] K. Iyer, R. Johari, and M. Sundararajan, "Mean field equilibria of dynamic auctions with learning," *Management Science*, vol. 60, no. 12, pp. 2949–2970, 2014.

[7] R. Gummadi, P. Key, and A. Proutiere, "Optimal bidding strategies and equilibria in dynamic auctions with budget constraints," 2013.

[8] J. Li, R. Bhattacharyya, S. Paul, S. Shakkottai, and V. Subramanian, "Incentivizing sharing in realtime D2D streaming networks: A mean field game perspective," *IEEE/ACM Transactions on Networking*, 2016.

[9] M. H. Afrasiabi and R. Guérin, "Pricing strategies for user-provided connectivity services," in *INFOCOM*, pp. 2766–2770, IEEE, 2012.

[10] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 843–855, 2011.

[11] L. Gao, G. Iosifidis, J. Huang, L. Tassiulas, and D. Li, "Bargaining-based mobile data offloading," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1114–1125, 2014.

[12] D. Kalathil, C. Wu, K. Poolla, and P. Varaiya, "The sharing economy for the smart grid," *arXiv preprint arXiv:1608.06990*, 2016.

[13] "Electric Grid Test Case Repository." `https://electricgrids.engr.tamu.edu/`.

[14] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on power systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[15] C. Graham and S. Méléard, "Chaos hypothesis for a system interacting through shared resources," *Probability Theory and Related Fields*, vol. 100, no. 2, pp. 157–174, 1994.

[16] B. Xia, H. Ming, K.-Y. Lee, Y. Li, Y. Zhou, S. Bansal, S. Shakkottai, and L. Xie, "Energy-coupon: A case study on incentive-based demand response in smart grid," in *Proceedings of the Eighth International Conference on Future Energy Systems*, e-Energy '17, (New York, NY, USA), pp. 80–90, ACM, 2017.

[17] O. Hernandez-Lerma and J. Lasserre, *Further Topics on Discrete-Time Markov Control Processes*. Stochastic Modelling and Applied Probability.

[18] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, vol. 2. Cambridge University Press, 2009.

[19] M. Manjrekar, V. Ramaswamy, and S. Shakkottai, "A mean field game approach to scheduling in cellular systems," *eprint arXiv:1309.1220*.

[20] D. Llorens, *Do solar panels work in cloudy weather?*

[21] The Weather Company, LLC, 2017.

[22] M. Huang, P. E. Caines, and R. P. Malhame, "The NCE (mean field) principle with locality dependent cost interactions," *IEEE Transactions on Automatic Control*, vol. 55, pp. 2799–2805, Dec 2010.

[23] California Energy Commission, 2017.

[24] U.S. Department of Energy, 2016.

[25] U.S. Energy Information Administration, 2015.

[26] Quora, 2015. https://www.quora.com/How-much-would-I-make-selling-electricity-back-to-the-grid-with-a-1kw-system.

[27] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 627–636, 2014.

[28] D. Merugu, B. S. Prabhakar, and N. S. Rama, "An incentive mechanism for decongesting the roads: A pilot program in Bangalore," in *Proceedings of NetEcon, ACM Workshop on the Economics of Networked Systems*, July 2009.

[29] B. Prabhakar, "Designing large-scale nudge engines," in *Proceedings of the ACM SIGMETRICS*, pp. 1–2, 2013.

[30] ERCOT, "Electric Reliability Council of Texas (ERCOT)," 2014. Data set available at `http://www.ercot.com/`.

[31] Pecan-Street, 2014. Data set available at `https://dataport.pecanstreet.org/`.

[32] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the Econometric Society*, pp. 263–291, 1979.

[33] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[34] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice," *Science*, vol. 211, no. 4481, pp. 453–458, 1981.

[35] D. Kahneman and A. Tversky, "Choices, values, and frames.," *American psychologist*, vol. 39, no. 4, p. 341, 1984.

[36] S. Gao, E. Frejinger, and M. Ben-Akiva, "Adaptive route choices in risky traffic networks: A prospect theory approach," *Transportation research part C: emerging technologies*, vol. 18, no. 5, pp. 727–740, 2010.

[37] G. W. Harrison and E. E. Rutström, "Expected utility theory and prospect theory: One wedding and a decent funeral," *Experimental Economics*, vol. 12, no. 2, pp. 133–158, 2009.

[38] T. Li and N. B. Mandayam, "Prospects in a wireless random access game," in *Proceedings of Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2012.

[39] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden, "Tussle in cyberspace: defining tomorrow's Internet," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4, pp. 347–356, 2002.

[40] J. Yu, M. H. Cheung, and J. Huang, "Spectrum investment with uncertainty based on prospect theory," in *Proceedings of International conference on Communications (ICC)*, pp. 1620–1625, 2014.

[41] Y. Wang, W. Saad, N. B. Mandayam, and H. V. Poor, "Integrating energy storage into the smart grid: A prospect theoretic approach," in *Proceedings of Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7779–7783, 2014.

[42] L. Xiao, N. B. Mandayam, and H. V. Poor, "Prospect theoretic analysis of energy exchange among microgrids," *IEEE Transactions on Smart Grid*, vol. 6, pp. 63–72, January 2015.

[43] B. Jovanovic and R. W. Rosenthal, "Anonymous sequential games," *Journal of Mathematical Economics*, vol. 17, pp. 77–87, February 1988.

[44] M. Huang, R. P. Malhamé, and P. E. Caines, "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle," *Communications in Information & Systems*, vol. 6, no. 3, pp. 221–252, 2006.

[45] M. H. Albadi and E. F. El-Saadany, "A summary of demand response in electricity markets," *Electric Power Systems Research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[46] J. Li, B. Xia, X. Geng, M. Hao, S. Shakkottai, V. Subramanian, and X. Le, "Energy coupon: A mean field game perspective on demand response in smart grids," in *Proceedings of ACM SIGMETRICS*, pp. 455–456, 2015.

[47] M. Manjrekar, V. Ramaswamy, and S. Shakkottai, "A mean field game approach to scheduling in cellular systems," in *Proceedings of IEEE Infocom*, (Toronto, Canada), 2014.

[48] J. Li, R. Bhattacharyya, S. Paul, S. Shakkottai, and V. Subramanian, "Incentivizing sharing in realtime D2D streaming networks: A mean field game perspective," *IEEE/ACM Transactions on Networking*, 2016.

[49] J. Naritomi, *Consumers as Tax Auditors*. working paper, International Development Department and Institute of Public Affairs, London School of Economics, 2013.

[50] M. Poco, C. Lopes, and A. Silva, *Perception of Tax Evasion and Tax Fraud in Portugal: A Sociological Study*. working paper, 2015.

[51] S. Chen and J. Wang, "Tax evasion and fraud detection: A theoretical evaluation of Taiwan's business tax policy for internet auctions," *Asian Social Science*, vol. 6, no. 12, p. 23, 2010.

[52] P. Loiseau, G. A. Schwartz, J. Musacchio, S. Amin, and S. S. Sastry, "Incentive mechanisms for Internet congestion management: Fixed-budget rebate versus time-of-day pricing," *IEEE/ACM Transactions on Networking*, vol. 22, pp. 647–661, April 2014.

[53] H. Allcott and J. Kessler, "The welfare effects of nudges: A case study of energy use social comparisons," tech. rep., National Bureau of Economic Research, 2015.

[54] H. Zhong, L. Xie, and Q. Xia, "Coupon incentive-based demand response: Theory and case study," *IEEE Transactions on Power Systems*, vol. 28, pp. 1266–1276, May 2013.

[55] E. Bitar, "Coordinated aggregation of distributed demand-side resources," 2015. http://www.news.cornell.edu/stories/2015/03/adding-renewable-energy-power-grid-requires-flexibility.

[56] M. Hao and L. Xie, "Analysis of coupon incentive-based demand response with bounded consumer rationality," in *North American Power Symposium*, pp. 1–6, September 2014.

[57] G. A. Schwartz, H. Tembine, S. Amin, and S. S. Sastry, "Electricity demand shaping via randomized rewards: A mean field game approach," *Allerton Conference on Communication, Control, and Computing*, 2012.

[58] T. Qin, X. Geng, and T. Liu, "A new probabilistic model for rank aggregation," in *Advances in neural information processing systems*, pp. 1948–1956, 2010.

[59] J. A. Lozano and E. Irurozki, "Probabilistic modeling on rankings," 2012. available at http://www.sc.ehu.es/ccwbayes/members/ekhine/tutorial_ranking/info.html.

[60] D. R. Hunter, "MM algorithms for generalized bradley-terry models," *Annals of Statistics*, 2004.

[61] D. Gomes, J. Mohr, and R. Souza, "Discrete time, finite state space mean field games," *Journal de mathématiques pures et appliquées*, vol. 93, no. 3, pp. 308–328, 2010.

[62] S. Adlakha, R. Johari, and G. Weintraub, "Equilibria of dynamic games with many players: Existence, approximation, and market structure," *Journal of Economic Theory*, vol. 156, pp. 269–316, 2015.

[63] D. Prelec, "The probability weighting function," *Econometrica*, pp. 497–527, 1998.

[64] V. S. Borkar and R. Sundaresan, "Asymptotics of the Invariant Measure in Mean Field Models with Jumps," *Stochastic Systems*, vol. 2, no. 2, pp. 322–380, 2013.

[65] S. M. Ross, *Applied Probability Models with Optimization Applications*. Courier Corporation, 2013.

[66] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, 2009.

[67] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Aggregate flexibility of thermostatically controlled loads," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 189–198, 2015.

[68] OhmConnect, 2015. Online at `https://www.ohmconnect.com/`.

[69] J. Li, R. Bhattacharyya, S. Paul, S. Shakkottai, and V. Subramanian, "Incentivizing Sharing in Realtime D2D Streaming Networks: A Mean Field Game Perspective," *arXiv preprint arXiv:1604.02435*, 2016.

[70] U.S. Department of Energy, *Benefit of Demand Response in Electricity Market and Recommendations for Achieving Them*, 2006. [online] Available: `http://www.caiso.com/Documents/DemandResponseandProxyDemandResourcesFrequentlyAskedQuestions.pdf`.

[71] New York Independent System Operator (NYISO), *Day-ahead demand response program manual*, July 2009. [online] Available: `http://www.nyiso.com/public/webdocs/markets_operations/documents/Manuals_and_Guides/Manuals/Operations/dadrp_mnl.pdf`.

[72] California ISO, *Demand Response User Guide*, June 2015. [online] Available: `http://www.caiso.com/Documents/July20_2009InitialCommentsonPDAdoptingDRActivities_Budgets_2009-2011inDocketNos_A_08-06-001_etal_.pdf`.

[73] California ISO, *Demand response & proxy demand resource-frequently asked questions*, 2011. [online] Available: http://www.caiso.com/Documents/DemandResponseandProxyDemandResourcesFrequentlyAskedQuestions.pdf.

[74] D. Hurley, P. Peterson, and M. Whited, "Demand response as a power system resource," *Synapse Energy Economics Inc*, 2013.

[75] Airedale, *ACIS Building Energy Management System*, 2018. [online] Available: http://airedale.com.

[76] F. A. Wolak, "Residential customer response to real-time pricing: The anaheim critical peak pricing experiment," *Center for the Study of Energy Markets*, 2007.

[77] N. Ruiz, I. Cobelo, and J. Oyarzabal, "A direct load control model for virtual power plant management," *Power Systems, IEEE Transactions on*, vol. 24, no. 2, pp. 959–966, 2009.

[78] A. Gholian, H. M. Rad, and Y. Hua, "Optimal industrial load control in smart grid.," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2305–2316, 2016.

[79] M. Parvania, M. Fotuhi-Firuzabad, and M. Shahidehpour, "ISO's optimal strategies for scheduling the hourly demand response in day-ahead markets," *Power Systems, IEEE Transactions on*, vol. 29, no. 6, pp. 2636–2645, 2014.

[80] ClearlyEnergy.com, *Residential demand response programs*, 2018. [online] Available: https://www.clearlyenergy.com/residential-demand-response-programs.

[81] *Power2Choose.org*, 2018. [online] Available: http://www.powertochoose.org/.

[82] U.S. Department of Energy, *The Pecan Street Project: developing the electric utility system of the future*. PhD thesis, 2015.

[83] ENERNOC: An Enel Group Company, *https://www.enernoc.com/resources/datasheets-brochures/get-more-commercial-and-industrial-demand-response*.

[84] OhmConnect, *https://www.ohmconnect.com/*.

[85] H. Zhong, L. Xie, and Q. Xia, "Coupon incentive-based demand response (cidr) in smart grid," in *2012 IEEE Power and Energy Society General Meeting*, pp. 1–6, IEEE, 2012.

[86] H. Zhong, L. Xie, and Q. Xia, "Coupon incentive-based demand response: Theory and case study," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1266–1276, 2013.

[87] H. Ming and L. Xie, "Analysis of coupon incentive-based demand response with bounded consumer rationality," in *North American Power Symposium (NAPS), 2014*, pp. 1–6, IEEE, 2014.

[88] G. A. Schwartz, H. Tembine, S. Amin, and S. S. Sastry, "Electricity demand shaping via randomized rewards: A mean field game approach," *Allerton Conference on Communication, Control, and Computing*, 2012.

[89] J. Li, B. Xia, X. Geng, H. Ming, S. Shakkottai, V. Subramanian, and L. Xie, "Mean field games in nudge systems for societal networks," in *ACM Sigmetrics*, 2015.

[90] J. Li, B. Xia, X. Geng, H. Ming, S. Shakkottai, V. Subramanian, and L. Xie, "Energy coupon: A mean field game perspective on demand response in smart grids," *ACM SIGMETRICS Performance Evaluation Review*, vol. 43, no. 1, pp. 455–456, 2015.

[91] G. A. Schwartz, H. Tembine, S. Amin, and S. S. Sastry, "Demand response scheme based on lottery-like rebates," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 4584–4588, 2014.

[92] D. Merugu, B. S. Prabhakar, and N. Rama, "An incentive mechanism for decongesting the roads: A pilot program in bangalore," in *Proc. of ACM NetEcon Workshop*, Citeseer, 2009.

[93] *NuRide, Inc*. [online] Available: `https://www.nuride.com`.

[94] A. Ju, *Experiment makes energy savings a game*, January 2017.

[95] S. Pan, Y. Shen, Z. Sun, P. Mahajan, L. Zhang, and P. Zhang, "Demo abstract: saving energy in smart commercial buildings through social gaming," in *Proceedings of the 2013*

ACM conference on Pervasive and ubiquitous computing adjunct publication, pp. 43–46, ACM, 2013.

[96] Public Utilities Commission of Texas, *SmartMeterTexas.com*, April.

[97] H.-p. Chao, "Demand response in wholesale electricity markets: the choice of customer baseline," *Journal of Regulatory Economics*, vol. 39, no. 1, pp. 68–88, 2011.

[98] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espínola, "Forecasting next-day electricity prices by time series models," *IEEE Transactions on power systems*, vol. 17, no. 2, pp. 342–348, 2002.

[99] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "Arima models to predict next-day electricity prices," *IEEE transactions on power systems*, vol. 18, no. 3, pp. 1014–1020, 2003.

[100] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and arima models," *IEEE transactions on power systems*, vol. 20, no. 2, pp. 1035–1042, 2005.

[101] L. Wu and M. Shahidehpour, "A hybrid model for day-ahead price forecasting," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1519–1530, 2010.

[102] T. Jónsson, P. Pinson, H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Forecasting electricity spot prices accounting for wind power predictions," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 210–218, 2013.

[103] D. Muthirayan, D. Kalathil, K. Poolla, and P. Varaiya, "Mechanism design for self-reporting baselines in demand response," in *American Control Conference (ACC), 2016*, pp. 1446–1451, American Automatic Control Council (AACC), 2016.

[104] Wikipedia, *K-nearest Neighbors Algorithm*. [online] Available: `https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm`.

[105] Wikipedia, *Kernel Regression*. [online] Available: https://en.wikipedia.org/wiki/Kernel_regression.

[106] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the Fundamentals of Financial Decision Making: Part I*, pp. 99–127, World Scientific, 2013.

[107] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[108] R. Gonzalez and G. Wu, "On the shape of the probability weighting function," *Cognitive psychology*, vol. 38, no. 1, pp. 129–166, 1999.

[109] M. Abdellaoui, H. Bleichrodt, and O. lâĂŹHaridon, "A tractable method to measure utility and loss aversion under prospect theory," *Journal of Risk and uncertainty*, vol. 36, no. 3, p. 245, 2008.

[110] C. E. Commission, *Staff Forecast: Average Retail Electricity Prices 2005 to 2018*, 2018. [online] Available: http://www.energy.ca.gov/2007publications/CEC-200-2007-013/CEC-200-2007-013-SD.PDF.

[111] M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 1994.

[112] P. Billingsley, *Convergence of probability measures*. John Wiley & Sons, 2013.

APPENDIX A

PROOFS FROM CHAPTER 2

Proof of Lemma 1 Recall the definition of $\mathcal{A}(b) : \mathcal{D} \times \mathcal{D} \cap [0, b + s/(1+\alpha)]$. We claim the effective bid of a server, x_s, lies in a compact set with an upper bound $\bar{x}_s$, which follows from the argument that no server will bid above $\bar{x}_s < \infty$ such that the expected return is below c_{serve} (since this has to be paid in every time-frame). By (2.1), when a server bids x_s we have only the clients with budget $b \geq x_s - s/(1+\alpha)$ will respond so that the expected return is given by $\int_{x_s - s/(1+\alpha)}^{\infty} \hat{\pi}(\hat{b})(x_s + \alpha(x_s - \hat{b})^+)d\hat{b} \leq \int_{x_s - s/(1+\alpha)}^{\infty} \hat{\pi}(\hat{b})(x_s + s\alpha/(1+\alpha))d\hat{b}$ = $\mathbb{P}(\hat{b} \geq x_s - s/(1+\alpha))(x_s + s\alpha/(1+\alpha))$. However, all budget distributions in our system are stochastically dominated by the distribution obtained by transferring all wealth $s - c_{serve} + s\alpha/(1+\alpha)$ to the server in every time period. Note that the initial budget is given by the regeneration distribution that has support B_{init}. We will assume that B_{init} is bounded with upper-bound $\bar{b}_{init}$. Given the lifetime of an agent in the system is geometrically distributed with parameter $1 - \beta$, we have

$$\mathbb{P}\left(\hat{b} \geq x_s - \frac{s}{1+\alpha}\right) \leq \beta^{x_s - s/(1+\alpha) - \bar{b}_{init}}.$$

Thus, we have the expected return of the server

$$(x_s + s\alpha/1+\alpha)\mathbb{P}(\hat{b} \geq x_s - s/(1+\alpha)) \leq (x_s + s\alpha/1+\alpha)\beta^{x_s - s/(1+\alpha) - \bar{b}_{init}}, \text{ and}$$

$$\lim_{x_s \to \infty} (x_s + s\alpha/1+\alpha)\mathbb{P}(\hat{b} \geq x_s - s/(1+\alpha)) \leq \lim_{x_s \to \infty} (x_s + s\alpha/1+\alpha)\beta^{x_s - s/(1+\alpha) - \bar{b}_{init}} = 0.$$

Define the following

$$R \triangleq \max_{x_s \geq 0} (x_s + s\alpha/1+\alpha)\beta^{x_s - s/(1+\alpha) - \bar{b}_{init}}.$$

We will assume that the parameters (β, s, α) are chosen such that $R \geq c_{serve}$. Under this assumption we set $\bar{x}_s$ to be the largest root of the transcendental equation

$$c_{serve} = (x + s\alpha/1 + \alpha)\beta^{x - s/(1+\alpha) - \bar{b}_{init}}.$$

Meanwhile, as a client, given a finite budget b, the closed interval $[0, b + s/(1+\alpha)]$ is compact. Hence the effective action space lies in a compact set $\hat{\mathcal{A}}(b) \triangleq \mathcal{D} \cap [0, \bar{x}_s] \times \mathcal{D} \cap [0, \bar{x}_c]$, where $\bar{x}_c \triangleq b + s/(1+\alpha)$.

We define the reward per stage as

$$c(b, (x_s, x_c)) \triangleq p_s(1 - \hat{\rho}_c(x_s))(x_s - c_{serve}) + p_c\big(\hat{\rho}_s(x_c)(s - \mathbb{E}[\tilde{X}_s | \tilde{X}_s \leq x_c]) - (1 - \hat{\rho}_s(x_c))c_{lose}\big).$$

Since c_{serve}, s and c_{lose} are constants, we have $c(b, (x_s, x_c))$ is bounded and continuous in (x_s, x_c). Finally, the third result follows from the continuity of the transition kernel $\mathcal{Q}$ over discrete topology of $\hat{\mathcal{A}}(b)$.

Proof of Lemma 4 Suppose f_n is monotonically increasing and x_s^*, x_c^* maximize $T_{\hat{\rho},\hat{\pi}} f_n(b) = f_{n+1}(b)$. Let $b' > b$ we have

$$\begin{aligned}
&f_{n+1}(b') \\
&\geq p_s\bigg(\mathbb{E}_{\hat{\rho}}[\mathbf{1}_{\tilde{x}_c \geq x_s^*}(\mathbb{E}_{\hat{\pi}}[\beta f_n(b' + x_s - c_{serve} + \alpha(x_s - \hat{B})^+)] + x_s - c_{serve}) + \mathbf{1}_{\tilde{x}_c < x_s^*}\beta f(b')]\bigg) \\
&+ p_c\bigg(\mathbb{E}_{\hat{\rho}}[\mathbf{1}_{x_c^* \geq \tilde{x}_s}(\beta f_n(b' + s - \tilde{x}_s - \alpha(\tilde{x}_s - b')^+) + s - \tilde{x}_s) + \mathbf{1}_{x_c^* < \tilde{x}_s}(\beta f_n(b') - c_{lose})]\bigg) \\
&\geq p_s\bigg(\mathbb{E}_{\hat{\rho}}[\mathbf{1}_{\tilde{x}_c \geq x_s^*}(\mathbb{E}_{\hat{\pi}}[\beta f_n(b + x_s - c_{serve} + \alpha(x_s - \hat{B})^+)] + x_s - c_{serve}) + \mathbf{1}_{\tilde{x}_c < x_s^*}\beta f(b)]\bigg) \\
&+ p_c\bigg(\mathbb{E}_{\hat{\rho}}[\mathbf{1}_{x_c^* \geq \tilde{x}_s}(\beta f_n(b + s - \tilde{x}_s - \alpha(\tilde{x}_s - b)^+) + s - \tilde{x}_s) + \mathbf{1}_{x_c^* < \tilde{x}_s}(\beta f_n(b) - c_{lose})]\bigg) \\
&= f_{n+1}(b)
\end{aligned}$$

Proof of Lemma 5 Denote the expected trade ratio variable as κ. Consider the case in which servers bid multiple values, w.l.o.g., we assume $p_{\tilde{X}_s} = (p_{\tilde{X}_s}(k_1), p_{\tilde{X}_s}(k_2))$, where $c_{serve} < k_1 <$

$k_2 < s$ and all clients can afford k_2. By the four facts, we have the client will bid either k_1 or k_2. We denote the probabilities of clients placing such bids as $p_{\tilde{X}_c} = (p_{\tilde{X}_c}(k_1), p_{\tilde{X}_c}(k_2))$. This gives us

$$\begin{aligned}\kappa =& p_{\tilde{X}_c}(k_1)p_{\tilde{X}_s}(k_1) + p_{\tilde{X}_c}(k_2)(p_{\tilde{X}_s}(k_1) + p_{\tilde{X}_s}(k_2)) \\ =& p_{\tilde{X}_c}(k_1)p_{\tilde{X}_s}(k_1) + p_{\tilde{X}_c}(k_2)\end{aligned}$$

Now if the servers decide to change unilaterally to $p'_{\tilde{X}_s} = (p'_{\tilde{X}_s}(k'))$, where $k' \in [k_1, k_2]$, $p'_{\tilde{X}_s}(k') = 1$, the clients who were bidding k_2 will follow the new price k', since it leads to a higher payoff. Meanwhile, the clients who were bidding k_1 will choose a bid between 0 and k' by Fact 3. Bidding k' yields a lower but positive payoff compared with the earlier, however, bidding 0 yields a zero payoff. Thus, these clients will bid k' as well. The new trade ratio is then

$$\kappa' = (p_{\tilde{X}_c}(k_1) + p_{\tilde{X}_c}(k_2))p'_{\tilde{X}_s}(k') = p_{\tilde{X}_c}(k_1) + p_{\tilde{X}_c}(k_2) > \kappa.$$

The proof above implies within the clients' financial ability, merging two server's bids always increases the trade ratio, which induces the result of Lemma 5. Note that the proof also follows in the case that $p_{\tilde{X}_c}, p_{\tilde{X}_s}$ are p.d.fs by replacing the summations with integrals.

Proof of Lemma 6 By Lemma 4, we have $v^*_{\hat{\rho}}$ is monotonically increasing in b, which induces the monotonicity of v_{c_win} and v_{c_lose}. Therefore, v_{c_win} and v_{c_lose} are bounded increasing on the closed interval $[k - \frac{s}{1+\alpha}, k - \frac{s-k}{\alpha}]$. Define $g(b) \triangleq v_{c\_win}(b) - v_{c\_lose}(b)$ for $b \in [k - \frac{s}{1+\alpha}, k - \frac{s-k}{\alpha}]$. We have $g(b)$ is of bounded variation, i.e. $g(b)$ has finite total variation on $[k - \frac{s}{1+\alpha}, k - \frac{s-k}{\alpha}]$. Thus, the number of zero crossings of $g(b)$ over the closed interval is finite. Therefore, $\theta_{c,\hat{\rho}}(b)$ is piecewise constant on $[0, k - \frac{s-k}{\alpha}]$ with a finite number of constant intervals. Within each of the intervals, $\theta_{c,\hat{\rho}}(b)$ is constant and either 0 or k. At the boundaries of these intervals, where $v_{c\_win}(b) = v_{c\_lose}(b)$, we have $\theta_{c,\hat{\rho}}(b) = \{0, k\}$.

Proof of Lemma 7 From (2.5), for Borel set B, we have $\mathbb{P}(b[t+1] \in B | b[t] = b) \geq (1 - \beta)\Psi(B) > 0$, which satisfies the Doeblin condition. Then the budget chain is ergodic. The rest of the first part proof follows the results in Chapter 12, Meyn and Tweedie [18]. Since the

regeneration happens independently of the budget transition, between two regenerations, we could further derive the relationship between $\pi(B)$ and $\pi^{(\tau)}(B|b)$, where τ is the first regeneration time after $t = 0$ and $b(0) = b$ so that $\pi^{(\tau)}(\cdot|\cdot)$ is the τ-step transition function without any regenerations.

$$\begin{aligned}\pi(B) &= \sum_{\tau=0}^{\infty}(1-\beta)\beta^{\tau}\int \pi^{(\tau)}(B|b)d\Psi(b)\\ &= \sum_{\tau=0}^{\infty}(1-\beta)\beta^{\tau}\mathbb{E}_{\Psi}\big(\pi^{(\tau)}(B|B_{init})\big),\end{aligned}$$

where we use the short-hand $\mathbb{E}_{\Psi}\big(\pi^{(\tau)}(B|B_{init})\big)$ to mean $\int \pi^{(\tau)}(B|b)d\Psi(b)$.

Since $\pi(\cdot)$ is the invariant budget distribution through the transition kernel defined in (2.5), $\pi^{(\tau)}(B|b)$ is basically the τ -step transitions starting at $b_0 = b$ without regenerations. If B is a Lebesgue null-set, we have $\Psi(B) = 0$ and in each of the τ step $\pi^{(\tau)}(B|b) = 0$, therefore, $\pi(B) = 0$.

Proof of Lemma 8 We demonstrate the proof in two parts a) given $\hat{\pi}$, $v^*_{\hat{z},\hat{\pi}}$ is Lipschitz continuous in $\hat{z}$, and b) given z, $v^*_{\hat{z},\hat{\pi}}$ is Lipschitz continuous in $\hat{\pi}$. Then the proof completes by applying triangle inequality between the two parts.

a) For any given $\hat{z}$ and $\hat{\pi}$, by Theorem 1 there is a unique $v^*_{\hat{z},\hat{\pi}}(\cdot)$ which is the unique fixed point of the contraction mapping $T^j_{\hat{z},\hat{\pi}}$ with Lipschitz constant $\lambda \in (0,1)$. Rewriting (2.3) in terms of $\hat{z}$ and k, we have

$$\begin{aligned}&(T^j_{\hat{z},\hat{\pi}}f)(b)\\ &= \beta f(b) + p_s(1-\hat{z})(k - c_{serve} + \beta\Delta f_s(b,k,c_{serve},\hat{\pi}))+\\ &\max_{x_c\in\mathcal{A}_k(b)}\left(p_c x_c\frac{s-k+c_{lose}+\beta\Delta f_c(b,s,k,\alpha)}{k},0\right) - p_c c_{lose}\end{aligned}$$

where

$\mathcal{A}_k(b) = [0, b + s/(1+\alpha)] \cap \{0, k\}$,

$\Delta f_s(b,k,c_{serve},\hat{\pi}) = \int_0^{\infty}\hat{\pi}(\hat{b})f(b+k-c_{serve}+\alpha(k-\hat{b})^+)d\hat{b} - f(b)$, and $\Delta f_c(b,s,k,\alpha) = f(b+s-k-\alpha(k-b)^+) - f(b)$.

Taking the derivative with respect to $\hat{z}$ using the Envelope Theorem, we have $T^j_{\hat{z},\hat{\pi}}$ is Lipschitz

continuous in $\hat{z}$ with constant $k - c_{serve}$.

Pick $\hat{z}_1$ and $\hat{z}_2$, we have $\frac{||T^j_{\hat{z}_1,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}} - T^j_{\hat{z}_2,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}}||_\infty}{|\hat{z}_1 - \hat{z}_2|} \leq k - c_{serve}$. Since $v^*_{\hat{z}_2,\hat{\pi}}$ is the unique fixed point of the contraction mapping $T^j_{\hat{z}_2,\hat{\pi}}$, we have $T^j_{\hat{z}_2,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}} = v^*_{\hat{z}_2,\hat{\pi}}$. Then we have $\frac{||T^j_{\hat{z}_1,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}} - v^*_{\hat{z}_2,\hat{\pi}}||_\infty}{|\hat{z}_1 - \hat{z}_2|} \leq k - c_{serve}$. Applying $T^j_{\hat{z}_1,\hat{\pi}}$ n times, given the contraction parameter λ, we have $\frac{||T^{(n+1)j}_{\hat{z}_1,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}} - T^{nj}_{\hat{z}_1,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}}||_\infty}{|\hat{z}_1 - \hat{z}_2|} \leq \lambda^n (k - c_{serve})$. Also, we have the unique fixed point of $T^j_{\hat{z}_1,\hat{\pi}}$ being $v^*_{\hat{z}_1,\hat{\pi}}$. Letting $n \to \infty$, completes the proof as follows:

$$\begin{aligned}
\frac{||v^*_{\hat{z}_1,\hat{\pi}} - v^*_{\hat{z}_2,\hat{\pi}}||_\infty}{|\hat{z}_1 - \hat{z}_2|} &\leq \sum_{n=0}^{\infty} \frac{||T^{(n+1)j}_{\hat{z}_1,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}} - T^{nj}_{\hat{z}_1,\hat{\pi}} v^*_{\hat{z}_2,\hat{\pi}}||_\infty}{|\hat{z}_1 - \hat{z}_2|} \\
&\leq \frac{k - c_{serve}}{1 - \lambda}
\end{aligned}$$

b) We consider the Lipschitz continuity of $v^*_{\hat{z},\hat{\pi}}$ in $\hat{\pi}$. From the discussion in a), we only need to show $T^j_{\hat{z},\hat{\pi}}$ is Lipschitz continuous in $\hat{\pi}$, then the rest of the proof goes through in the similar way.

Recall the expression of $(T^j_{\hat{z},\hat{\pi}} f)(b)$, there is only one term of interest, $\Delta f_s(b, k, c_{serve}, \hat{\pi}) = \int_0^\infty \hat{\pi}(\hat{b}) f(b + k - c_{serve} + \alpha(k - \hat{b})^+) d\hat{b} - f(b)$. According to Lemma 4, f is monotonic increasing. Then, f is of bounded total variation on the closed interval $[b + k - c_{serve}, b + k - c_{serve} + \alpha k]$ for any b. Therefore, given f, $T^j_{\hat{z},\hat{\pi}}$ is Lipschitz continuous in $\hat{\pi}$ with constant $C = f(b + k - c_{serve} + \alpha k) - f(b + k - c_{serve})$.

Proof of Theorem 3 Given $\hat{z}$ and k, the optimal value function can be rewritten as

$$\begin{aligned}
&v^*_{\hat{z},\hat{\pi}}(b) \\
&= \beta v^*_{\hat{z},\hat{\pi}}(b) + p_s(1 - \hat{z})(k - c_{serve} + \beta \Delta v_s(b, k, c_{serve}, \hat{\pi}))+ \\
&\max_{x_c \in \mathcal{A}_k(b)} \left(p_c \big(\mathbf{1}_{x_c = k}(s - k + \beta \Delta v_c(b, s, k, \alpha)) - \mathbf{1}_{x_c = 0} c_{lose} \big) \right) \\
&= \beta v^*_{\hat{z},\hat{\pi}}(b) + p_s(1 - \hat{z})(k - c_{serve} + \beta \Delta v_s(b, k, c_{serve}, \hat{\pi}))+ \\
&\max_{x_c \in \mathcal{A}_k(b)} \left(p_c x_c \frac{s - k + c_{lose} + \beta \Delta v_c(b, s, k, \alpha)}{k}, 0 \right) - p_c c_{lose}
\end{aligned}$$

where

$\mathcal{A}_k(b) = [0, b + s/(1 + \alpha)] \cap \{0, k\}$,

$\Delta v_s(b,k,c_{serve},\hat{\pi}) = \int_0^\infty \hat{\pi}(\hat{b}) v^*_{\hat{z},\hat{\pi}}(b+k-c_{serve}+\alpha(k-\hat{b})^+)d\hat{b} - v^*_{\hat{z},\hat{\pi}}(b)$, and $\Delta v_c(b,s,k,\alpha) = v^*_{\hat{z},\hat{\pi}}(b+s-k-\alpha(k-b)^+) - v^*_{\hat{z},\hat{\pi}}(b)$.

Define the increasing piecewise linear convex function $h_{\hat{z}}(y)$:$\mathbb{R} \mapsto \mathbb{R}$ given by

$$h_{\hat{z}}(y) = \phi(\hat{z}) + \max_{x_c \in \mathcal{A}_k(b)} (x_c y)_+,$$

where $(\cdot)_+ := \max(\cdot, 0)$ and

$$\phi(\hat{z}) := \beta v^*_{\hat{z},\hat{\pi}}(b) + p_s(1-\hat{z})\Big(k - c_{serve} + \beta \Delta v_s(b,k,c_{serve},\hat{\pi})\Big) - p_c c_{lose}.$$

$$y = p_c \frac{s-k+c_{lose}+\beta\Delta v_c(b,s,k,\alpha)}{k}$$

By Lemma 8, we have $v^*_{\hat{z},\hat{\pi}}(\cdot)$ is continuous in $\hat{z}$ for all $x_c \in \mathcal{A}_k(b)$. Thus we have $\phi(\hat{z})$ is continuous in $\hat{z}$ for all $x_c \in \mathcal{A}_k(b)$. By Berge's Maximum Theorem we have the correspondence,

$$\mathcal{F}(y) := \arg\max_{x_c \in \mathcal{A}_k(b)} (x_c y)_+$$

is upper hemicontinuous in $\hat{z}$. Note that $\theta_{c,\hat{z}}(b)$ is given by

$$\theta_{c,\hat{z}}(b) = \mathcal{F}\left(p_c \frac{s-k+c_{lose}+\beta\Delta v_c(b,s,k,\alpha)}{k}\right). \tag{A.1}$$

Given the Lipschitz continuity of $v^*_{\hat{z},\hat{\pi}}(\cdot)$ in $\hat{z}$, we conclude for every state b, $\theta_{c,\hat{z}}(b)$ is upper hemicontinuous in $\hat{z}$.

Proof of Theorem 4 By Lemma 7, given $\hat{z}$ and $\hat{\pi}$, we have the Markov process of the budgets has a unique stationary distribution $\pi = \Pi(\hat{z},\hat{\pi})$. Here, we will prove the continuity of $\Pi(\hat{z},\hat{\pi})$ in $\hat{z}$ and $\hat{\pi}$. By the Portmanteau Theorem, we only need to show that for any uniform converging sequence $\hat{z}_n \to \hat{z}$, a sequence of density functions $\hat{\pi}_n \to \hat{\pi}$ and any open set B,

$\liminf_{n\to\infty} \pi_n(B) \geq \pi(B)$, where $\pi_n = \Pi(\hat{z}_n, \hat{\pi}_n)$. Thus, by Fatou's Lemma, we have

$$\begin{aligned}\liminf_{n\to\infty} \pi_n(B) &= \liminf_{n\to\infty} \sum_{\tau=0}^{\infty} (1-\beta)\beta^\tau \mathbb{E}_\Psi\big(\pi_n^{(\tau)}(B|B_{init})\big) \\ &\geq \sum_{\tau=0}^{\infty} (1-\beta)\beta^\tau \mathbb{E}_\Psi\big(\liminf_{n\to\infty} \pi_n^{(\tau)}(B|B_{init})\big)\end{aligned}$$

To complete the proof, we need to show that

$$\liminf_{n\to\infty} \pi_n^{(\tau)}(B|b) \geq \pi^{(\tau)}(B|b). \text{ for every } b \in B_{init}$$

The proof of the above holds by mathematical induction and the Skorokhod representation theorem. Details of a similar proof can be found in the appendix of [19].

Proof of Theorem 5 Define the single-point inverse (lower inverse) $\theta_{c,\hat{z}}^{-1}(0) = \{b \geq 0 : 0 \in \theta_{c,\hat{z}}(b)\}$ and also the upper inverse $\tilde{\theta}_{c,\hat{z}}^{-1}(0) = \{b \geq 0 : \theta_{c,\hat{z}}(b) = \{0\}\}$. By Lemma 6, we have $\theta_{c,\hat{z}}^{-1}(0)$ consists of a finite number of closed subintervals in $[0, k - \frac{s-k}{\alpha}]$, and $\tilde{\theta}_{c,\hat{z}}^{-1}(0)$ a finite number of open intervals (in $\mathbb{R}_+$) the closure of which is exactly $\theta_{c,\hat{z}}^{-1}(0)$, with the difference being only finitely many points. Since $\pi_{\hat{z},\hat{\pi}}$ is absolutely continuous with respect to Lebesgue measure we have

$$\pi_{\hat{z},\hat{\pi}}\left(\theta_{c,\hat{z}}^{-1}(0)\right) = \pi_{\hat{z},\hat{\pi}}\left(\tilde{\theta}_{c,\hat{z}}^{-1}(0)\right),$$

so that $\theta_{c,\hat{z}}^{-1}(0)$ is a continuity set of $\pi_{\hat{z},\hat{\pi}}$ for every $\hat{z}$.

From the definition of $\theta_{c,\hat{z}}(\cdot)$ we have

$$\begin{aligned}\theta_{c,\hat{z}}^{-1}(0) &= \left\{b : h(b) \geq 0, v_{\hat{z},\hat{\pi}}^*(h(b)) - v_{\hat{z}}^*(b) \leq \frac{k - s - c_{lose}}{\beta}\right\} \cup \{b : h(b) \leq 0\} \\ \tilde{\theta}_{c,\hat{z}}^{-1}(0) &= \left\{b : h(b) \geq 0, v_{\hat{z},\hat{\pi}}^*(h(b)) - v_{\hat{z}}^*(b) < \frac{k - s - c_{lose}}{\beta}\right\} \cup \{b : h(b) \leq 0\},\end{aligned}$$

where we have

$$h(b) = b + s - k - \alpha(k-b)^+ \text{and } \{b : h(b) \leq 0\} = \left[0, k - \frac{s}{1+\alpha}\right]$$

We know that $v^*_{\hat{z},\hat{\pi}}(\cdot)$ is Lipschitz continuous in $\hat{z}$ so that for all $\epsilon > 0$ we have

$$\|v_{\hat{z}',\hat{\pi}} - v_{\hat{z},\hat{\pi}}\|_\infty \leq L\epsilon \; \forall \hat{z}' \in [\hat{z} - \epsilon, \hat{z} + \epsilon] \cap [0,1]$$
$$\|v_{\hat{z}',\hat{\pi}} - v_{\hat{z},\hat{\pi}}\|_\infty < L\epsilon \; \forall \hat{z}' \in (\hat{z} - \epsilon, \hat{z} + \epsilon) \cap [0,1].$$

This then implies that for all $\{b \geq k - \frac{s}{1+\alpha} : h(b) \geq 0\}$ we have

$$\begin{aligned}
&v^*_{\hat{z},\hat{\pi}}(h(b)) - v^*_{\hat{z},\hat{\pi}}(b) - 2L\epsilon \leq v^*_{\hat{z}',\hat{\pi}}(h(b)) - v^*_{\hat{z}',\hat{\pi}}(b) \\
&\leq v^*_{\hat{z},\hat{\pi}}(h(b)) - v^*_{\hat{z},\hat{\pi}}(b) + 2L\epsilon \; \forall \hat{z}' \in [\hat{z} - \epsilon, \hat{z} + \epsilon] \cap [0,1] \\
&v^*_{\hat{z},\hat{\pi}}(h(b)) - v^*_{\hat{z},\hat{\pi}}(b) - 2L\epsilon < v^*_{\hat{z}',\hat{\pi}}(h(b)) - v^*_{\hat{z}',\hat{\pi}}(b) \\
&< v^*_{\hat{z},\hat{\pi}}(h(b)) - v^*_{\hat{z},\hat{\pi}}(b) + 2L\epsilon \; \forall \hat{z}' \in (\hat{z} - \epsilon, \hat{z} + \epsilon) \cap [0,1]
\end{aligned}$$

Therefore we have

$$\theta^{-1}_{c,\hat{z}'}(0) \subseteq F_{\hat{z}}(\epsilon), \; O_{\hat{z}}(\epsilon) \subseteq \tilde{\theta}^{-1}_{c,\hat{z}'}(0),$$

where the closed set $F_{\hat{z}}(\epsilon)$ and the open set $O_{\hat{z}}(\epsilon)$ are given by

$$F_{\hat{z}}(\epsilon) = \left\{b : h(b) \geq 0, v^*_{\hat{z},\hat{\pi}}(h(b)) - v^*_{\hat{z},\hat{\pi}}(b) \leq \frac{k - s - c_{lose}}{\beta} + 2L\epsilon\right\} \cup \{b : h(b) \leq 0\},$$
$$O_{\hat{z}}(\epsilon) = \left\{b : h(b) \geq 0, v^*_{\hat{z},\hat{\pi}}(h(b)) - v^*_{\hat{z}}(b) < \frac{k - s - c_{lose}}{\beta} - 2L\epsilon\right\} \cup \{b : h(b) \leq 0\}.$$

Given a sequence $\{\hat{z}_n\}_{n\geq 1}$ such that $\lim_{n\to\infty} \hat{z}_n = \hat{z}$, we know that $\pi_{\hat{z}_n,\hat{\pi}} \Rightarrow \pi_{\hat{z},\hat{\pi}}$. Next fix an

$\epsilon > 0$. Then by the Portmanteau theorem we have

$$\begin{aligned}
\pi_{\hat{z},\hat{\pi}}(O_{\hat{z}}(\epsilon)) &\leq \liminf_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(O_{\hat{z}}(\epsilon)) \\
&\leq \liminf_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(\tilde{\theta}^{-1}_{c,\hat{z}_n}(0)) = \liminf_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(\theta^{-1}_{c,\hat{z}_n}(0)) \\
&\leq \limsup_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(\tilde{\theta}^{-1}_{c,\hat{z}_n}(0)) = \limsup_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(\theta^{-1}_{c,\hat{z}_n}(0)) \\
&\leq \limsup_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(F_{\hat{z}}(\epsilon)) \\
&\leq \pi_{\hat{z},\hat{\pi}}(F_{\hat{z}}(\epsilon)).
\end{aligned}$$

Now the proof that $\lim_{n\to\infty} \pi_{\hat{z}_n,\hat{\pi}}(\theta^{-1}_{c,\hat{z}_n}(0)) = \pi_{\hat{z},\hat{\pi}}(\theta^{-1}_{c,\hat{z}}(0))$ follows by noticing that for $\epsilon_1 < \epsilon_2$ we have

$$\begin{gathered}
O_{\hat{z}}(\epsilon_2) \subseteq O_{\hat{z}}(\epsilon_1),\ F_{\hat{z}}(\epsilon_1) \subseteq F_{\hat{z}}(\epsilon_2), \\
\text{so } \lim_{\epsilon\downarrow 0} O_{\hat{z}}(\epsilon) = \tilde{\theta}^{-1}_{c,\hat{z}}(0),\ \lim_{\epsilon\downarrow 0} F_{\hat{z}}(\epsilon) = \theta^{-1}_{c,\hat{z}}(0),
\end{gathered}$$

and also the fact that $\pi_{\hat{z},\hat{\pi}}(\tilde{\theta}^{-1}_{c,\hat{z}}(0)) = \pi_{\hat{z},\hat{\pi}}(\theta^{-1}_{c,\hat{z}}(0))$.

APPENDIX B

PROOFS FROM CHAPTER 3

B.1 Properties of the optimal value function

Proof of Lemma 9

We first show that $T_\rho f \in \Phi$ for $\forall f \in \Phi$. The proof then follows through a verification of the conditions of Theorem $6.10.4$ in [111]. From the definition of T_ρ in (3.15), we have

$$|T_\rho f(x)| \leq |u(x)| + \max_{a(x)\in\mathcal{A}} \theta_{a(x)} + \beta \max(|f(x+w)|, |f(x-l)|).$$

From this it follows that

$$\sup_{x\in\mathbb{X}} \frac{|T_\rho f(x)|}{\Omega(x)} \leq \sup_{x\in\mathbb{X}} \frac{|u(x)|}{\Omega(x)} + \sup_{x\in\mathbb{X}} \frac{\max_{a(x)\in\mathcal{A}} \theta_{a(x)}}{\Omega(x)} + \beta \max\left(\sup_{x\in\mathbb{X}} \frac{|f(x+w)|}{\Omega(x)}, \sup_{x\in\mathbb{X}} \frac{|f(x-l)|}{\Omega(x)}\right).$$

Let x_+ be the unique positive surplus such that $u(x_+) = 1$ and x_- be the unique negative surplus such that $u(x_-) = -1$. Note that $\Omega(x)$ is non-decreasing for $x \geq x_-$ and non-increasing for $x \leq x_+$. To avoid cumbersome algebra we will assume $x_+ - w > 0$ and $x_- + l > 0$. Since $\Omega(x) \geq |u(x)| \geq 0$ and $\Omega(x) \geq 1$, the first two terms are bounded by 1 and $\max_{a(x)\in\mathcal{A}} \theta_{a(x)}$. For the last term we have

$$\sup_{x\in\mathbb{X}} \frac{|f(x+w)|}{\Omega(x)} \leq \|f\|_\Omega \sup_{x\in\mathbb{X}} \frac{\Omega(x+w)}{\Omega(x)}.$$

We have the following

$$\frac{\Omega(x+w)}{\Omega(x)} = \begin{cases} \frac{u(x+w)}{\max(u(x),1)} \leq \frac{u(x+w)}{u(x)}, & \text{if } x \geq x_+, \\ \frac{u(x+w)}{\max(u(x),1)} \leq \frac{u(x+w)}{u(x_+)}, & \text{if } x \in [x_+ - w, x_+], \\ 1, & \text{if } x \in [x_-, x_+ - w], \\ \frac{1}{|u(x)|} \leq 1, & \text{if } x \in [x_- - w, x_-], \\ \frac{u(x+w)}{u(x)} \leq 1, & \text{if } x \leq x_- - w. \end{cases}$$

For $x \geq x_+$, we know using monotonicity of $u(\cdot)$

$$\frac{u(x+w)}{u(x)} = 1 + \frac{u(x+w) - u(w)}{w} \frac{w}{u(x)} \leq 1 + \frac{u(x+w) - u(w)}{w} w.$$

Additionally, for $x \in [x_+ - w, x_+]$ we have

$$\frac{u(x+w)}{u(x_+)} = 1 + \frac{u(x+w) - u(x_+)}{x+w-x_+}(x+w-x_+) \leq 1 + \frac{u(x+w) - u(x_+)}{x+w-x_+} w.$$

For the analysis we assume that $u(\cdot)$ is Lipschitz such that $\sup_{x \in \mathbb{X}} u'(x) < +\infty$. Therefore, by the mean value theorem

$$\begin{aligned} &\frac{u(x+w) - u(x)}{w} = u'(\xi_1) \leq \sup_{x \geq x_+} u'(x), && \forall \xi_1, x \in [x_+, \infty), \\ &\frac{u(x+w) - u(x_+)}{x+w-x_+} = u'(\xi_2) \leq \sup_{x \in [x_+ - w, x_+]} u'(x), && \forall \xi_2, x \in [x_+ - w, x_+], \\ &\sup_{x \in \mathbb{X}} \frac{\Omega(x+w)}{\Omega(x)} \leq \|f\|_\Omega (1 + w \sup_{x \geq x_+ - w} u'(x)), && \forall x \in [x_+ - w, \infty). \end{aligned}$$

Similarly, we have

$$\sup_{x \in \mathbb{X}} \frac{|f(x-l)|}{\Omega(x)} \leq \|f\|_\Omega \sup_{x \in \mathbb{X}} \frac{\Omega(x-l)}{\Omega(x)}.$$

Now we have the following

$$\frac{\Omega(x-l)}{\Omega(x)} = \begin{cases} \frac{u(x-l)}{\min(u(x),-1)} \leq \frac{u(x-l)}{u(x)}, & \text{if } x \leq x_-, \\ \frac{u(x-l)}{\max(u(x),-1)} \leq \frac{u(x-l)}{u(x_-)}, & \text{if } x \in [x_-, x_- + l], \\ 1, & \text{if } x \in [x_- + l, x_+], \\ \frac{1}{u(x)} \leq 1, & \text{if } x \in [x_+, x_+ + l], \\ \frac{u(x-l)}{u(x)} \leq 1, & \text{if } x \leq x_+ l. \end{cases}$$

Using the same logic as before, we get

$$\sup_{x \in \mathbb{X}} \frac{\Omega(x-l)}{\Omega(x)} \leq \|f\|_\Omega (1 + l \sup_{x \in \mathbb{X}: x \leq x_-} u'(x)).$$

Since $u(\cdot)$ is Lipschitz, thus, there exists an $\alpha_0 \in (0, +\infty)$ such that $\|T_\rho f\|_\Omega \leq \alpha_0$.

Next, we need to verify the conditions of Theorem 6.10.4 in [111]. The lemma requires verification of the following three conditions. We set $x[k]$ to be the state variable denoting the surplus at time k. We need to show that $\forall x \in \mathbb{X}$, for some constants (independent of ρ) $\alpha_1 > 0$, $\alpha_2 > 0$ and $0 < \alpha_3 < 1$,

$$\sup_{a(x) \in \mathcal{A}} |u(x) - \theta_{a(x)}| \leq \alpha_1 \Omega(x), \tag{B.1}$$

$$\mathbb{E}_{x[1],a_0}[\Omega(x[1])|x[0] = x] \leq \alpha_2 \Omega(x), \quad \forall a_0 \in \mathcal{A}, \tag{B.2}$$

with the distribution of $x[1]$ chosen based on action a_0, and

$$\beta^J \mathbb{E}_{x[J],a_0,a_1,\ldots,a_{J-1}}[\Omega(x[J])|x[0] = x] \leq \alpha_3 \Omega(x), \tag{B.3}$$

for some $J > 0$ and all possible action sequences, i.e., $a_j \in \mathcal{A}$ for all $j = 0, 1, \ldots, J-1$ with the distribution of $x[J]$ chosen based on the action sequence $(a_0, a_1, \ldots, a_{J-1})$ chosen.

First consider (B.1). Since $\Omega(x) = \max(|u(x)|, 1)$, using the earlier analysis in Section 3.4,

(B.1) is true with $\alpha_1 = 1 + \max_{a \in \mathcal{A}} \theta_a$. Now consider (B.2). We have

$$\begin{aligned}\mathbb{E}_{x[1],a_0}[\Omega(x[1])|x[0] = x] &= \mathbb{E}_{\rho}[\phi(p_{\rho,a}(x))\Omega(x+w) + \phi(1-p_{\rho,a}(x)), \Omega(x-l)] \\ &\leq \max(\Omega(x+w), \Omega(x-l)),\end{aligned}$$

which is bounded by $\alpha_2\Omega(x)$ using our analysis from before.

Finally, (B.3) holds true using the properties of $\Omega(\cdot)$, the bounds on the probability of winning and losing (from Section 3.4) and our analysis from earlier in the proof as follows:

$$\begin{aligned}&\beta^J \mathbb{E}_{x[J],a_0,a_1,\ldots,a_{J-1}}[\Omega(x[J])|x[0] = x] \\ \leq &\beta^J \max(\phi(\overline{p}_W), \phi(1-\underline{p}_W))^J \max(\Omega(x+Jw), \Omega(x-Jl)) \\ \leq &(\beta \max(\phi(\overline{p}_W), \phi(1-\underline{p}_W))^J \alpha_4(J)\Omega(x),\end{aligned}$$

for some affine $\alpha_4(J) > 0$ using our analysis from before. It now follows that take J large enough we obtain an $\alpha_3 < 1$ that is also independent of ρ. Note that we can get a simpler bound of

$$\beta^J \mathbb{E}_{x[J],a_0,a_1,\ldots,a_{J-1}}[\Omega(x[J])|x[0] = x] \leq \beta^J \alpha_4(J)\Omega(x),$$

using just the properties of $\Omega(\cdot)$. Again we can take J large enough to obtain a $\alpha_3 < 1$ that is independent of ρ. This bound is useful when there is an action for which the probability of winning or losing is 1. Since all the conditions of Theorem 6.10.4 of [111] are met, then the first result in the lemma holds true. The second then follows immediately from (3.14).

Proof of Lemma 10

For any given ρ, from Lemma 9 we know that there is a unique $V_\rho(\cdot)$. Furthermore, it is the unique fixed point of operator T_ρ where T_ρ^J is a contraction mapping with constant α_3 that is independent of ρ. From (3.15), it follows that T_ρ^J is a continuous in ρ: computing derivatives using the envelope theorem and the expressions from Section 3.4, it is easily established that T_ρ^J is, in fact, Lipschitz with constant $(M-1)^J$ when the uniform norm is used for ρ.

Let ρ_1 and ρ_2 be two population/action profiles such that $\|\rho_1 - \rho_2\| \leq \epsilon$ (the choice of norm

is irrelevant as all are equivalent for finite dimensional Euclidean spaces). As T_ρ^J is continuous in ρ, there exists a $\delta > 0$ such that $\|T_{\rho_1}^J V_{\rho_2} - T_{\rho_2}^J V_{\rho_2}\|_\Omega \leq \delta$. However, since $T_{\rho_2}^J V_{\rho_2} = V_{\rho_2}$, we have shown that $\|T_{\rho_1}^J V_{\rho_2} - V_{\rho_2}\|_\Omega \leq \delta$. Applying $T_{\rho_1}^J$ n times and using the contraction property of $T_{\rho_1}^J$, we get

$$\|T_{\rho_1}^{(n+1)J} V_{\rho_2} - T_{\rho_1}^{nJ} V_{\rho_2}\|_\Omega \leq \alpha_3^n \delta.$$

The proof then follows since $\lim_{n\to\infty} \|T_{\rho_1}^{nJ} V_{\rho_2} - V_{\rho_1}\|_\Omega = 0$ so that

$$\|V_{\rho_1} - V_{\rho_2}\|_\Omega \leq \sum_{n=0}^{\infty} \|T_{\rho_1}^{(n+1)J} V_{\rho_2} - T_{\rho_1}^{nJ} V_{\rho_2}\|_\Omega \leq \frac{\delta}{1-\alpha_3}.$$

Furthermore, using the comment from above we can show that V_ρ is Lipschitz continuous in ρ.

B.2 The existence and uniqueness of stationary surplus distribution

Proof of Lemma 11

First, from the transition kernel (3.17), we satisfy the Doeblin condition as

$$\mathbb{P}(x[k] \in B | x[k-1] = x) \geq (1-\beta)\Psi(B),$$

where $0 < \beta < 1$, and Ψ is a probability measure for the regeneration process. Then from results in [18, Chapter 12], we have a unique stationary surplus distribution.

Next, let $-\tau$ be the last time before 0 that the surplus has a regeneration. Then we have

$$\zeta_{\rho\times\sigma}(B) = \sum_{k=0}^{\infty} \mathbb{P}(B, \tau = k) = \sum_{k=0}^{\infty} \mathbb{P}(B|\tau = k) \cdot \mathbb{P}(\tau = k). \tag{B.4}$$

Since the regeneration process happens independently of the surplus with inter-regeneration times geometrically distributed with parameter $(1-\beta)$, then $\mathbb{P}(\tau = k) = (1-\beta)\beta^k$. Also given $\tau = k$,

we have $X_{-k} \sim \Psi$. Therefore

$$\begin{aligned}\zeta_{\rho\times\sigma}(B) &= \sum_{k=0}^{\infty}(1-\beta)\beta^k\mathbb{P}(B|\tau=k) = \sum_{k=0}^{\infty}(1-\beta)\beta^k\mathbb{E}\left(\mathbb{E}\left(1_{x[0]\in B}|\tau=k, X_{-k}=X\right)|\tau=k\right)\\ &= \sum_{k=0}^{\infty}(1-\beta)\beta^k\mathbb{E}\left(\zeta^{(k)}_{\rho\times\sigma}(B|X)|\tau=k\right) = \sum_{k=0}^{\infty}(1-\beta)\beta^k\mathbb{E}_{\Psi}\left(\zeta^{(k)}_{\rho\times\sigma}(B|X)\right)\\ &= \sum_{k=0}^{\infty}(1-\beta)\beta^k\int\zeta^{(k)}_{\rho\times\sigma}(B|x)d\Psi(x). \end{aligned} \tag{B.5}$$

B.2.1 Existence of MFE

Proof of Lemma 12

Define the increasing and piecewise linear convex function

$$g_\rho(y) = \max_{a\in\mathcal{A}}\phi(p_{\rho,a})y - \theta_a = \max_{\sigma\in\Delta(|\mathcal{A}|)}\sum_{a\in\mathcal{A}}\sigma_a\big(\phi(p_{\rho,a})y-\theta_a\big), \tag{B.6}$$

where $\Delta(\mathcal{A})$ is the probability simplex on $A = |\mathcal{A}|$ elements. By the properties of the lottery and the weight function $\phi(\cdot)$, $\phi(p_{\rho,a})$ is continuous in ρ for all $a \in \mathcal{A}$. Using Berge's maximum theorem, we have

$$\arg\max_{\sigma\in\Delta(|\mathcal{A}|)}\sum_{a\in\mathcal{A}}\sigma_a\left(\phi(p_{\rho,a})y-\theta_a\right) \tag{B.7}$$

is upper semicontinuous in ρ.

Now let

$$\mathcal{A}(y) := \arg\max g(y) = \arg\max_{a\in\mathcal{A}}\phi(p_{\rho,a})y-\theta_a, \tag{B.8}$$

then set-valued function above is exactly $\Delta(|\mathcal{A}(y)|)$.

Hence, the optimal randomized policies at surplus x are a set-valued function $\Delta(|\mathcal{A}(y)|) = \Delta(|\mathcal{A}(V_\rho(x+w)-V_\rho(x-l))|)$, which is upper semicontinuous due to the Lipschitz continuity of $V_\rho(\cdot)$ in ρ and the u.s.c. of $\phi(p_{\rho,a})$ in ρ, i.e., for every state x, the action distribution $\sigma(x)$ is (pointwise) upper semicontinuous in ρ.

Proof of Lemma 13

The existence and uniqueness of $\zeta(x)$ for a given ρ and $\sigma(x)$, and the relationship between $\zeta(\cdot)$ and $\zeta^{(k)}(\cdot)$ are shown in Lemma 11. Now, we will prove the continuity of $\zeta_{\rho\times\sigma}$ in ρ and $\sigma(x)$ for every surplus $x \in \mathbb{X}$. For the assumed action distribution ρ on the finite set $\mathcal{A}$, we consider the topology of pointwise convergence which is equivalent to the uniform convergence by strong coupling results in [69]. For the randomized action distribution σ, corresponding to $\sigma(x)$ at each surplus $x \in \mathbb{X}$, we consider the topology with metric $\rho(\sigma^1, \sigma^2) = \sum_{j=1}^{\infty} 2^{-j} \min(\|\sigma^1(x_j) - \sigma^2(x_j)\|, 1)$, where $\|\cdot\|$ is any norm for $\mathbb{R}^{|\mathcal{A}|}$.

First, we will show that the surplus distribution $\zeta^{(k)}_{\rho\times\sigma}$ is continuous in ρ and σ. By Portmanteau theorem, we only need to show that for any sequence $\rho_n \to \rho$ uniformly, $\sigma_n \to \sigma$ pointwise, and any open set B, we have $\liminf_{n\to\infty} \zeta^{(k)}_{\rho_n\times\sigma_n}(B|x) \geq \zeta^{(k)}_{\rho\times\sigma}(B|x)$.

Lemma 17. $\liminf_{n\to\infty} \zeta^{(k)}_{\rho_n\times\sigma_n}(B|x) \geq \zeta^{(k)}_{\rho\times\sigma}(B|x)$.

Proof of Lemma 17

The proof proceeds by induction on k. For $k = 0$, $\zeta^{(0)}_{\rho_n\times\sigma_n}(B|x) = 1_{(x\in B)}$ is a point-mass at x irrespective of $\rho_n \times \sigma_n$, and in fact, for any $n \in \mathbb{N}_+$, we have $\zeta^{(0)}_{\rho_n\times\sigma_n}(B|x) = \zeta^{(0)}_{\rho\times\sigma}(B|x)$. Let $\rho_n \to \rho$ uniform, and $\sigma_n(x) \to \sigma(x)$ pointwise for every surplus x. We will show that $\zeta^{(k)}_{\rho_n\times\sigma_n}(B|x)$ converges pointwise to $\zeta^{(k)}_{\rho\times\sigma}(B|x)$.

We will refer to the measure and random variables corresponding to $\rho_n \times \sigma_n$ for the n^{th} system and those corresponding to $\rho \times \sigma$ as coming from the limiting system. We will prove that $\zeta^{(k)}_{\rho_n\times\sigma_n}(B|x)$ converges to $\zeta^{(k)}_{\rho\times\sigma}(B|x)$ pointwise using the metrics given above.

Suppose that the hypothesis holds true for $k-1$ where $k > 1$, i.e., $\zeta^{(k-1)}_{\rho_n\times\sigma_n}(B|x)$ converges pointwise to $\zeta^{(k-1)}_{\rho\times\sigma}(B|x)$. To prove this lemma, we only need to show that the hypothesis holds for k. Let $\mathbb{P}_{\rho\times\sigma,x}(\cdot)$ be the one-step transition probability measure of the surplus dynamics conditioned on the initial state of the surplus being x, and there is no regeneration. Then we have $\mathbb{P}_{\rho_n\times\sigma_n,x}(x+w) = \sum_{a\in\sigma_n(x)} p_{\rho_n\times\sigma_n,a}$, $\mathbb{P}_{\rho_n\times\sigma_n,x}(x-l) = 1 - \sum_{a\in\sigma_n(x)} p_{\rho_n\times\sigma_n,a}$ and $\mathbb{P}_{\rho\times\sigma,x}(x+w) = \sum_{a\in\sigma(x)} p_{\rho\times\sigma,a}$, $\mathbb{P}_{\rho\times\sigma,x}(x-l) = 1 - \sum_{a\in\sigma(x)} p_{\rho\times\sigma,a}$. By the properties of the lottery, $p_{\rho\times\sigma,a}$ is continuous in $\rho \times \sigma$ for all $a \in \mathcal{A}$, thus we have $p_{\rho_n\times\sigma_n,a}$ converges to $p_{\rho\times\sigma,a}$ pointwise, i.e., $\mathbb{P}_{\rho_n\times\sigma_n,x}(\cdot)$ converges to $\mathbb{P}_{\rho\times\sigma,x}(\cdot)$ pointwise. By the Skorokhod representation the-

orem [112], there exist random variables X_n and X on common probability space and a random integer N such that $X_n \sim \mathbb{P}_{\rho_n \times \sigma_n, x}(\cdot)$ for all $n \in \mathbb{N}$, and $X \sim \mathbb{P}_{\rho \times \sigma, x}(\cdot)$, and $X_n = X$ for $n \geq N$.

Then we have,

$$\begin{aligned}\liminf_{n\to\infty} \zeta^{(k)}_{\rho_n \times \sigma_n}(B|x) &= \liminf_{n\to\infty} \mathbb{E}\left(\zeta^{(k-1)}_{\rho_n \times \sigma_n}(B|X_n)\right) \geq \mathbb{E}\left(\liminf_{n\to\infty} \zeta^{(k-1)}_{\rho_n \times \sigma_n}(B|X_n)\right) \\ &\geq \mathbb{E}\left(\zeta^{(k-1)}_{\rho \times \sigma}(B|X)\right) = \zeta^{(k)}_{\rho \times \sigma}(B|x), \end{aligned} \tag{B.9}$$

where the second and third inequality hold due to Fatou's lemma and the induction hypothesis. Hence, for a given ρ and randomized policies $\sigma(x)$, the unique stationary surplus distribution $\zeta^{(k)}_{\rho_n \times \sigma_n}(B|x)$ converges pointwise to $\zeta^{(k)}_{\rho \times \sigma}(B|x)$.

Now by Lemma 11 and Equation (3.18), we need to show that $\liminf_{n\to\infty} \zeta_{\rho_n \times \sigma_n}(B) \geq \zeta_{\rho \times \sigma}(B)$. By Fatou's lemma, we have

$$\begin{aligned}&\liminf_{n\to\infty} \zeta_{\rho_n \times \sigma_n}(B) = \liminf_{n\to\infty} \sum_{k=0}^{\infty} (1-\beta)\beta^k \mathbb{E}_\Psi\left(\zeta^{(k)}_{\rho_n \times \sigma_n}(B|X_n)\right) \\ &\geq \sum_{k=0}^{\infty} (1-\beta)\beta^k \mathbb{E}_\Psi\left(\liminf_{n\to\infty} \zeta^{(k)}_{\rho_n \times \sigma_n}(B|X_n)\right) \geq \sum_{k=0}^{\infty} (1-\beta)\beta^k \mathbb{E}_\Psi\left(\zeta^{(k)}_{\rho \times \sigma}(B|X)\right) = \zeta_{\rho \times \sigma}(B). \end{aligned} \tag{B.10}$$

Thus, for a given ρ and the randomize policies $\sigma(x)$, the unique stationary surplus distribution $\zeta_{\rho_n \times \sigma_n}$ converges pointwise to $\zeta_{\rho \times \sigma}$. Then the stationary surplus distribution $\zeta_{\rho \times \sigma}$ is continuous in ρ and $\sigma(x)$ for every surplus $x \in \mathbb{X}$.

Proof of Lemma 14

Given the stationary surplus distribution $\zeta(x)$ and the action distribution $\sigma(x)$ at every surplus x, those will introduce a population profile based on the actions chosen at each point x, denoted that action distribution as ρ, and we have $\rho_a = \sum_{x \in \mathbb{X}} \zeta(x) \cdot \sigma_a(x)$, where $a \in \mathcal{A}$, $\mathbb{X}$ is a countable set and $\mathcal{A}$ is a finite set.

To show that ρ is continuous in $\zeta(x)$ and $\sigma(x)$, we only need to show that for any sequence $\{\zeta_n\}_{n=1}^{\infty}$ converging to ζ in uniform norm, $\{\sigma_n(x)\}_{n=1}^{\infty}$ converging to $\sigma(x)$ pointwise, we have

$\{\rho_n\}_{n=1}^{\infty}$ converges to ρ pointwise, which is equivalent to convergence in uniform norm as we have a finite set $\mathcal{A}$.

Since $\zeta_n \to \zeta$ uniformly, we have $\forall \epsilon_1 > 0$, $\exists N_1 \in \mathbb{N}$, so that $\forall n \geq N_1$, $\forall x \in \mathbb{X}$, $|\zeta_n(x) - \zeta(x)| \leq \epsilon_1$. Similarly, $\{\sigma_n(x)\}_{n=1}^{\infty}$ converges to $\sigma(x)$ pointwise, we have $\forall x \in \mathbb{X}$, and $\forall \epsilon_2 > 0$, $\exists N_2 \in \mathbb{N}$ so that $\forall n \geq N_2$, , $|\sigma_n(x) - \sigma(x)| \leq \epsilon_2$. Now consider $\forall \epsilon = \max(\epsilon_1, \epsilon_2)$, we can find an all but finite subset $\mathbb{X}_1$ of $\mathbb{X}$, such that $\sum_{x \in \mathbb{X}_1} \zeta(x) \leq \frac{\epsilon}{2}$. Let $N = \max(N_1, N_2)$, for $\forall x \in \mathbb{X} \setminus \mathbb{X}_1$, $\exists n > N$ large enough, such that $|\sigma_{n,a}(x) - \sigma_a(x)| \leq \frac{\epsilon}{2}$. Then $\forall x \in \mathbb{X}$, $\forall a \in \mathcal{A}$, we have

$$
\begin{aligned}
|\rho_{n,a} - \rho_a| &= \left| \sum_x \zeta_n(x)\sigma_{n,a}(x) - \sum_x \zeta(x)\sigma_a(x) \right| \\
&= \left| \sum_x \zeta_n(x)\sigma_{n,a}(x) - \sum_x \zeta_n(x)\sigma_a(x) + \sum_x \zeta_n(x)\sigma_a(x) - \sum_x \zeta(x)\sigma_a(x) \right| \\
&\leq \left| \sum_x \zeta_n(x)\sigma_{n,a}(x) - \sum_x \zeta_n(x)\sigma_a(x) \right| + \left| \sum_x \zeta_n(x)\sigma_a(x) - \sum_x \zeta(x)\sigma_a(x) \right| \\
&\leq \sum_x \zeta_n(x)|\sigma_{n,a}(x) - \sigma_a(x)| + \sum_x \sigma_a(x)|\zeta_n(x) - \zeta(x)| \\
&= \sum_{x \in \mathbb{X}_1} \zeta_n(x)|\sigma_{n,a}(x) - \sigma_a(x)| + \sum_{x \in \mathbb{X} \setminus \mathbb{X}_1} \zeta_n(x)|\sigma_{n,a}(x) - \sigma_a(x)| + \sum_x \sigma_a(x)|\zeta_n(x) - \zeta(x)| \\
&\overset{(a)}{\leq} \sum_{x \in \mathbb{X}_1} \zeta_n(x) \cdot 1 + \sum_{x \in \mathbb{X} \setminus \mathbb{X}_1} \zeta_n(x) \cdot \frac{\epsilon}{2} + \sum_x \sigma_a(x) \cdot \epsilon_1 \\
&\overset{(b)}{\leq} \frac{\epsilon}{2} \cdot 1 + 1 \cdot \frac{\epsilon}{2} + \epsilon_1 \cdot 1 \leq \epsilon \cdot 1 + \epsilon \cdot 1 = 2\epsilon,
\end{aligned} \tag{B.11}
$$

where (a) follows from the fact that $|\sigma_{n,a}(x) - \sigma_a(x)| \leq 1$ for $\forall x \in \mathbb{X}$, and $|\sigma_{n,a}(x) - \sigma_a(x)| < \frac{\epsilon}{2}$, for $x \in \mathbb{X} \setminus \mathbb{X}_1$ given $\epsilon > 0$ and n large enough, and the convergence of ζ_n. (b) follows from that $\sum_{x \in \mathbb{X}_1} \zeta(x) < \frac{\epsilon}{2}$ for $x \in \mathbb{X}_1$.

Therefore, $|\rho_{n,a} - \rho_a| < 2\epsilon$ for all $a \in \mathcal{A}$ and $\forall n \geq N$, hence $\rho_n \to \rho$ pointwise, which is equivalent to convergence in uniform norm as we have a finite set $\mathcal{A}$.

B.3 Characteristics of the best response policy

Proof of Lemma 15

First, we consider $x \in \mathbb{X}$ and $x \geq 0$. We have

$$\begin{aligned}
&u(x) - \theta_{a_2(x)} + \beta[p_{\rho,a_2}(x)V_\rho(x+w) + (1-p_{\rho,a_2}(x))V_\rho(x-l)] \\
&\gtrless u(x) - \theta_{a_1(x)} + \beta[p_{\rho,a_1}(x)V_\rho(x+w) + (1-p_{\rho,a_1}(x))V_\rho(x-l)] \\
&\Leftrightarrow \theta_{a_1(x)} - \theta_{a_2(x)} \gtrless \beta[(p_{\rho,a_1}(x) - p_{\rho,a_2}(x))V_\rho(x+w) \\
&\qquad\qquad + ((1-p_{\rho,a_1}(x)) - (1-p_{\rho,a_2}(x)))V_\rho(x-l)] \\
&\Leftrightarrow \theta_{a_1(x)} - \theta_{a_2(x)} \gtrless \beta(p_{\rho,a_1}(x) - p_{\rho,a_2}(x))[V_\rho(x+w) - V_\rho(x-l)]. \qquad \text{(B.12)}
\end{aligned}$$

As we assumed $\theta_{a_1(x)} > \theta_{a_2(x)}$, it follows that $p_{\rho,a_1}(x) > p_{\rho,a_2}(x)$. Also, since $w + l > 0$ and $V_\rho(x)$ is increasing in x, so both sides of the above inequality are non-negative. Since $V_\rho(x)$ is submodular when $x \geq -l$, the RHS is a decreasing function of x. Let $x^*_{a_1,a_2} \in \mathbb{X}$ be the smallest value such that LHS $\geq$ RHS, then for all $x > x^*_{a_1,a_2}$ action $a_2(x)$ is preferred to action $a_1(x)$, for all $x < x^*_{a_1,a_2}$ action $a_1(x)$ is preferred to action $a_2(x)$, and finally, if at $x^*_{a_1,a_2}$ LHS=RHS, then at $x^*_{a_1,a_2}$ the agent is indifferent between the two actions, and if instead LHS $>$ RHS, then action $a_2(x)$ is preferred to action $a_1(x)$. We call $x^*_{a_1,a_2}$ the threshold value of surplus for actions $a_1(x)$ and $a_2(x)$.

Similarly, for $x \in \mathbb{X}$ and $x \leq 0$, $V_\rho(x)$ is supermodular when $x \leq w$, which implies the existence of a threshold policy.

Proof of Lemma 16

First, let $f \in \Phi$, suppose that f is an increasing and submodular function. First we prove that $T_\rho f$ is increasing and submodular too. Let $a^*(x)$ be an optimal action in the definition of $T_\rho f(x)$ when the surplus is x, i.e., one of the maximizers from (3.15). Let $x_1 > x_2$, then

$$\begin{aligned}
&T_\rho f(x_1) - T_\rho f(x_2) = u(x_1) - u(x_2) - \theta_{a^*(x_1)} + \theta_{a^*(x_2)} + \beta\big[p_{\rho,a^*(x_1)}(x_1)f(x_1+w) + \\
&\qquad (1-p_{\rho,a^*(x_1)}(x_1))f(x_1-l) - p_{\rho,a^*(x_2)}(x_2)f(x_2+w) - (1-p_{\rho,a^*(x_2)}(x_2))f(x_2-l)\big] \\
&\geq u(x_1) - u(x_2) - \theta_{a^*(x_2)} + \theta_{a^*(x_2)} + \beta\big[p_{\rho,a^*(x_2)}(x_2)f(x_1+w)
\end{aligned}$$

$$+ (1 - p_{\rho,a^*(x_2)}(x_2))f(x_1 - l) - p_{\rho,a^*(x_2)}(x_2)f(x_2 + w) - (1 - p_{\rho,a^*(x_2)}(x_2))f(x_2 - l)\big]$$
$$= u(x_1) - u(x_2) + \beta\big[p_{\rho,a^*(x_2)}(x_2)(f(x_1 + w + a) - f(x_2 + w)$$
$$+ (1 - p_{\rho,a^*(x_2)}(x_2))(f(x_1 - l) - f(x_2 - l))\big] \geq 0.$$

The first inequality holds because $a^*(x_2)$ need not be an optimal action when the surplus is x_1.

Again, let $x_1 > x_2$ and let $x > 0$. Since $u(\cdot)$ is a concave function, it follows that it is submodular, i.e.,

$$u(x_1 + x) - u(x_1) \leq u(x_2 + x) - u(x_2) \Leftrightarrow u(x_1 + x) + u(x_2) \leq u(x_2 + x) + u(x_1).$$

Assuming that $f \in \Phi$ is submodular, we will now show that $T_\rho f$ is also submodular. Consider

$$T_\rho f(x_1 + x) + T_\rho f(x_2) = u(x_1 + x) + u(x_2) - \theta_{a^*(x_1+x)} - \theta_{a^*(x_2)}$$
$$+ \beta\big[p_{\rho,a^*(x_1+x)}(x_1 + x)f(x_1 + x + w) + p_{\rho,a^*(x_2)}(x_2)f(x_2 + w)$$
$$+ (1 - p_{\rho,a^*(x_1+x)}(x_1 + x))f(x_1 + x - l) + (1 - p_{\rho,a^*(x_2)}(x_2))f(x_2 - l)\big].$$

We assume without loss of generality that $p_{\rho,a^*(x_1+x)}(x_1 + x) \geq p_{\rho,a^*(x_2)}(x_2)$ and let δ be the difference; if $p_{\rho,a^*(x_1+x)}(x_1 + x) \leq p_{\rho,a^*(x_2)}(x_2)$, then a similar proof establishes the result. Using this we have the RHS (denoted by d) being

$$d = u(x_1 + x) + u(x_2) - \theta_{a^*(x_1+x)} - \theta_{a^*(x_2)} + \beta\big[p_{\rho,a^*(x_2)}(x_2)(f(x_1 + x + w) + f(x_2 + w))$$
$$+ (1 - p_{\rho,a^*(x_1+x)}(x_1 + x))(f(x_1 + x - l) + f(x_2 - l)) + \delta(f(x_1 + x + w) + f(x_2 - l))\big].$$

By submodularity of $f(\cdot)$ we have

$$f(x_1 + x + w) + f(x_2 + w) \leq f(x_2 + x + w) + f(x_1 + w),$$
$$f(x_1 + x - l) + f(x_2 - l) \leq f(x_2 + x - l) + f(x_1 - l),$$
$$f(x_1 + x + w) + f(x_2 - l) \leq f(x_2 + x + w) + f(x_1 - l).$$

With these and using the submodularity of $u(\cdot)$ we get

$$\begin{aligned}
d &\leq u(x_2 + x) + u(x_1) - \theta_{a^*(x_1+x)} - \theta_{a^*(x_2)} + \beta\big[p_{\rho,a^*(x_2)}(x_2)(f(x_2 + x + w) + f(x_1 + w)) \\
&\quad + (1 - p_{\rho,a^*(x_1+x)}(x_1 + x))(f(x_2 + x - l) + f(x_1 - l)) + \delta(f(x_2 + x + w) + f(x_1 - l))\big] \\
&= u(x_2 + x) - \theta_{a^*(x_1+x)} + \beta[p_{\rho,a^*(x_2)}(x_2)f(x_2 + x + w) + (1 - p_{\rho,a^*(x_2)}(x_2))f(x_2 + x - l)] \\
&\quad + u(x_1) - \theta_{a^*(x_2)} + \beta[p_{\rho,a^*(x_2)}(x_2)f(x_1 + w) + (1 - p_{\rho,a^*(x_1+x)}(x_1 + x))f(x_1 - l)] \\
&\leq T_\rho f(x_2 + x) + T_\rho f(x_1),
\end{aligned}$$

where the last inequality holds as using the optimal actions $(a^*(x_2 + x), a^*(x_1))$ yields a higher value as opposed to the sub-optimal actions $(a^*(x_1 + x), a^*(x_2))$ when the surplus is $x_2 + x$ and x_1.

Since both the monotonicity and submodularity properties are preserved when taking pointwise limits, choosing $f(\cdot) \equiv 0$ (or $u(\cdot)$) to start the value iteration proves that the value function $V_\rho(\cdot)$ is increasing and submodular.

Similarly, if $f \in \Phi$ is an increasing and supermodular function, following the same argument, we can prove that the value function $V_\rho(\cdot)$ is increasing and supermodular.