Title
Flexible Demand Management under Time-Varying Prices

Author
Liang, Yong

Publication Date
2012

Peer reviewed|Thesis/dissertation

Flexible Demand Management under Time-Varying Prices

by

Yong Liang

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Max Shen, Chair
Professor Shmuel S. Oren
Professor David I. Levine
Assistant Professor Ying-Ju Chen

Spring 2013

Flexible Demand Management under Time-Varying Prices

Abstract

Flexible Demand Management under Time-Varying Prices

by

Yong Liang

Doctor of Philosophy in Industrial Engineering and Operations Research

University of California, Berkeley

Professor Zuo-Jun Max Shen, Chair

In this dissertation, the problem of flexible demand management under time-varying prices is studied. This generic problem has many applications, which usually have multiple periods in which decisions on satisfying demand need to be made, and prices in these periods are time-varying. Examples of such applications include multi-period procurement problem, operating room scheduling, and user-end demand scheduling in the Smart Grid, where the last application is used as the main motivating story throughout the dissertation.

The current grid is experiencing an upgrade with lots of new designs. What is of particular interest is the idea of passing time-varying prices that reflect electricity market conditions to end users as incentives for load shifting. One key component, consequently, is the demand management system at the user-end. The objective of the system is to find the optimal trade-off between cost saving and discomfort increment resulted from load shifting. In this dissertation, we approach this problem from the following aspects: (1) construct a generic model, solve for Pareto optimal solutions, and analyze the robust solution that optimizes the worst-case payoffs, (2) extend to a distribution-free model for multiple types of demand (appliances), for which an *approximate dynamic programming* (ADP) approach is developed, and (3) design other efficient algorithms for practical purposes of the flexible demand management system.

We first construct a novel multi-objective flexible demand management model, in which there are a finite number of periods with time-varying prices, and demand arrives in each period. In each period, the decision maker chooses to either satisfy or defer outstanding demand to minimize costs and discomfort over a certain number of periods. We consider both the deterministic model, models with stochastic demand or prices, and when only partial information about the stochastic demand or prices is known. We first analyze the stochastic optimization problem when the objective is to minimize the expected total cost and discomfort, then since the decision maker is likely to be risk-averse, and she wants to protect herself from price spikes, we study the robust optimization problem to address the

risk-aversion of the decision maker. We conduct numerical studies to evaluate the price of robustness.

Next, we present a detailed model that manages multiple types of flexible demand in the absence of knowledge regarding the distributions of related stochastic processes. Specifically, we consider the case in which time-varying prices with general structures are offered to users, and an *energy management system* for each household makes optimal energy usage, storage, and trading decisions according to the preferences of users. Because of the uncertainties associated with electricity prices, local generation, and the arrival processes of demand, we formulate a stochastic dynamic programming model, and outline a novel and tractable ADP approach to overcome the *curses of dimensionality*. Then, we perform numerical studies, whose results demonstrate the effectiveness of the ADP approach.

At last, we propose another approximation approach based on Q-learning. In addition, we also develop another decentralization-based heuristic. Both the Q-learning approach and the heuristic make necessary assumptions on the knowledge of information, and each of them has unique advantages. We conduct numerical studies on a testing problem. The simulation results show that both the Q-learning and the decentralization based heuristic approaches work well. Lastly, we conclude the paper with some discussions on future extension directions.

To my daughter Naomi,
my wife Ye,
and my parents, Changhai Liang and Guimin Wan

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost I wish to acknowledge my adviser, Professor Zuo-Jun Max Shen, for his patient guidance, enthusiastic encouragement, and useful critiques throughout my study at Berkeley. Without his help and support, I would never have been able to finish this dissertation.

I would also like to express my deep gratitude to Professor David I. Levine, Professor Shmuel S. Oren and Professor Ying-Ju Chen. Their good advice, support and friendship has been invaluable on both an academic and a personal level, for which I am extremely grateful.

I have greatly benefited from the discussion, and feedback of Professor Shen's former and current students, Ye Xu, Ying Rong, Mengshi Lyu, Long He, and Tianhu Deng.

Last but by no means least, I would like to thank my wife, Ye Xu. Along the way of this long journey in pursuing my doctoral degree, she was always there cheering me up and standing by me through the good times and bad.

Chapter 1

Introduction

The main object being studied in this dissertation is *flexible demand*, the demand that is usually not time-sensitive and can be deferred for cost reduction. The *management* of flexible demand refers to problems that aim to find the best schedule of satisfying flexible demand in order to optimize certain objectives. Such problems generally consist of multiple periods in which prices (unit cost) for the resource to satisfy demand are time-varying and new flexible demand arrives in each period. Decisions on either satisfying or deferring the outstanding demand are made at the beginning of each period, and the objective is to minimize total cost and discomfort. Flexible demand management models have a variety of applications, such as emergency room planning, multi-periods procurement, optimal stopping problem, the demand management for the Smart Grid users with time-varying prices, etc. We use the demand management for the Smart Grid users as our motivating example to explain our models and insights throughout the chapter.

1.1 Current Situation and Motivation for Price-Based Demand Response

It is well-known that the current electricity grid is inefficient and leads to an increasing number of power outages because of the *supply follows demand* strategy being used today. It has been recognized that this strategy results in lack of coordination between demand and supply and costs significant waste because the fixed-rate price structure discourages users from reducing peak loads or using distributed electricity generation and storage devices. On the other hand, limitations on the supply side make it necessary to keep costly ancillary service in order to met demand at all times. Increasing uncertainties in supply due to the intermittency of renewable sources, such as wind, exacerbate the challenge ([36]).

As the reverse of *supply follows demand*, *demand follows supply* might fail as well, due to various political and social issues. Motivated by the desire to better coordinate supply and

demand and maintain grid reliability ([27]), numerous *demand response* (DR) mechanisms have been brought up following the idea of the famous work of [57]. DR mechanisms incentivize users to adjust their consuming habit and shift demands from peak to off-peak periods. As a result, the demand will be less fluctuating over time. Since the fuel consumption is a strictly increasing function of the power output [56], less fluctuating demand leads to lower fuel consumption, namely higher energy efficiency. Moreover, as argued in literature such as [44], it is a much more efficient way to improve supply security by having proper demand response on the demand-side than by extending generations capacities on the supply-side.

There are two types of DR, namely the price-based DR, see for example [19], and the incentive-based DR, see for example [20] and [63]. The price-based DR is believed to be able to incentivize users to adjust their consuming habit and shift demands from peak to off-peak periods. The optimal pricing strategy is one of the earliest research focuses regarding manipulating demand in the electricity market. Since 1950s, economists have proposed peak-load pricing model, which divided the cycle into several periods and distinct price values for the periods are announced ahead of time, aimed at maximizing social welfare (the sum of company profit and consumer surplus). [25] gives a survey on peak-load pricing problem. Other than peak-load pricing model, adaptive pricing strategy gives price value for each period in real-time based on the supply and demand. For example, [55] proposes a real-time pricing model for demand-side management in the Smart Grid to maximize the aggregate utility in the electricity market. There are mainly three kinds of rate structure for electricity pricing, namely time-of-use (TOU), critical peak pricing (CPP), and real-time pricing (RTP) ([33]). The first two structures give deterministic pricing rates for predetermined peak periods and off-peak periods, while RTP is a dynamic scheme with time-variant rate based on real-time electricity consumption and supply. According to [17], the long-run efficiency gained by adopting RTP structure in a competitive electricity market is significant even if the demand is of little elasticity, and it weighs much higher than that of adopting TOU structure. However, there are several encumbrances for applying the dynamic pricing structure in Smart Grid, and the design of proper demand response mechanism is one of them.

Recently, advances in technologies have enabled efficient communication between the users and the grid. However, the diffusion of DR is still extremely slow, and what prohibits effective DR in practice is the lack of an efficient control mechanism on the demand-side [37]. Indeed, manually turning on and off appliances according to time-varying prices can be extremely costly, and a bad control algorithm may hurt users instead of saving costs for them. Therefore, the main target of this dissertation is to model the flexible demand management problem and solve for optimal control strategies for Smart Grid users.

Early works on demand response to electricity price are mostly conducted by economists in view of price elasticity and consumer behavior under the TOU rate structure, see [22], [1], and [30]. Nevertheless, the optimal DR mechanism in the environment of real-time pricing can be terribly complicated due to the randomness and dynamics of price and demands, and

more advanced models and techniques in stochastic optimal control need to be developed. [45] designs an *Energy Box* to manage electricity usage in an environment of demand-sensitive real-time pricing. In this dissertation, we study a series of models, from a general one built to get insights on the impact of the deep penetration of flexible demand management, to a detailed model that is capable to take into consideration of multiple types of demand with only limited information about the stochastic processes is known. The following section briefly summarizes the main topic of each subsequent chapter.

1.2 Overview of the Dissertation

Chapter 2 starts with a novel multi-objective model for the flexible demand management problem. The objectives are minimizing the expected costs of electricity, and minimizing the expected discomfort resulting from shifting flexible demand. We analyze the policies that attain Pareto optimality. Then, motivated by the possible risk-aversion of decision makers when only partial information about the stochastic demand arrivals and prices is available, we formulate and solve the distributionally-robust robust optimization model for the flexible demand management, and shows that decision makers are potentially better off if they are confronted with stochastic prices compared to being charged with deterministic prices with values of the means of the stochastic ones.

Chapter 3 presents a detailed model, which does not require the knowledge of the distributions of demand arrivals. Flexible demand is first categorized into two types, namely additive demand, such as the demand for air-conditioning, and non-additive demand, such as the demand for washer and dryer. We develop separate treatment to the two types of flexible demand, and the model is solved by employing an approximate dynamic programming (ADP) approach to deal with the *curses of dimensionality* and the lack of demand distributions. Then, we demonstrate the effectiveness of the ADP approach using numerical experiments.

Chapter 4 proposes another two approaches to solve the problem formulated in Chapter 4. The first approach is a decentralized heuristic, which assumes the knowledge of demand arrivals. The other is a Q-learning based approach. The Q-learning approach works under more general settings compared to the heuristic, while the heuristic is able to deliver solutions in a much faster manner for regular sized problems.

Chapter 2

A General Model, Optimal Policies, and Robust Solutions

2.1 Introduction

As discussed in Chapter 1, there are three types of time-varying price structures that have been proposed in literature: Time-of-use (TOU), Critical-peak-pricing (CPP), and Real-time-pricing (RTP) ([33]). Much has been written about the advantages of price-based DR; see for example: [1], [23],and [30] on the TOU; [31] on the CPP; and [18], [17], and [38] on the RTP. A common feature of the price-based DR is that, it assumes that users response to different prices by adjusting their usage.

The hassle of manually adjusting usage according to prices usually outweighs the benefit from load shifting for users. As noted by [37], the diffusion of DR has been notably slow, and one of the major impedance is the lack of a demand management mechanism that achieves automatic control. Recently, the demand management problem for smart grid users has received increasing interests. [45] propose a smart energy management system, in which the problem is formulated as a stochastic dynamic program. However, the stochastic dynamic programming approach suffers from the "curses of dimensionally". To address this problem, [43] propose another model that integrates more features and aims at minimizing the total expected disutility of decision makers. They develop an approximate dynamic programming approach to solve the problem efficiently.

In addition to dynamic programming, two-stage stochastic programming has also been widely applied to model stochastic demand and prices, especially in the literature of unit commitment problems, see for example, [21], [50], [52], [58], and [61]. Two most common methods of solving stochastic programs are the stochastic approximation based approach and the scenario-based approach. The reliability of the approximation-based approach depends highly on the accuracy of forecasts. However in some cases it is challenging, if not impossible,

to obtain reasonable forecasts for demand distributions. Meanwhile for the scenario-based approaches, the scenarios are generated based on the forecasts of demand and supply. Even if there exists demand distribution forecasts, the size of the problems and the complexity of solving them increase dramatically as the number of scenarios selected increases.

While the objectives in the work mentioned above are to optimize the expected objectives, another stream of research focuses on the worst-case performance. Naturally, when there is limited knowledge about the randomness of data, or when decision makers are risk-averse, *robust* solutions, which optimize worst-case objectives, are desired. For example in the smart grid with real-time pricing, prices are affected by many stochastic factors such as weather conditions that influence the total demand, and the output of renewable sources that changes the total supply. Most of time there only exists partial information about these stochastic factors, and thus although users would like to lower their expected total disutility, they are generally more concerned with price spikes, such as the $\$3,000$ per magewatthour price in August 2011 in the Electric Reliability Council of Texas (ERCOT) wholesale market (compared with the $\$63.47$ per magewatthour yearly average in 2011).[1]

Robust optimization models are designed for the worst-case optimization problems. [59] was the first to study robust optimization problems. Recently, significant progress has been made for robust optimization. [6], [8], [7] formulate the linear problems with data uncertainty using ellipsoidal uncertainty sets to address the issue of over conservation. Later, [12] develop another framework, which allows decision makers to control the conservatism and provides probabilistic bounds on violating the constraints. Various recent work adopts the framework proposed by [12], for instance, [13] studies the robust inventory control, while [14] and [35]) applies the framework to unit commitment problems, in which it is assumed that system operators make decisions in order to prevent the worst-case outcome. Another recent thread of research on robust optimization focuses on the "distributionally-robust" optimization problems, for which it is assumed that only partial information, such as the moments, about the distributions of the stochastic parameters is known. In recent studies, [15], [26], [29], [49], and [48] formulate distributionally-robust optimization problems into tractable problems, some of which have received much attention in the last two decades.

In this chapter, we first construct a novel multi-objective model for the well-known flexible demand management problem, in which one objective is to minimize the cost, and the other is to minimize the discomfort from shifting demand. Then, we characterize the solutions that minimize the expected cost and discomfort. In addition, we formulate and solve distributionally-robust optimization models for flexible demand management problems, which has not been done in the literature. This chapter further contributes to the literature by showing the fairly counter-intuitive result that decision makers are potentially better off if they are confronted with stochastic prices, compared to being charged with deterministic

[1] Source: http://www.eia.gov/todayinenergy/detail.cfm?id=3010

prices with values equal to the first moments of the stochastic ones.

The remainder of the chapter is organized as follows. Section 2.2 presents the multi-objective programming formulation of the deterministic version of the demand-side control problem. Section 2.3 extends the deterministic model and discuss the case with stochastic demand. Section 2.4 turns to the case with price uncertainty. In this section, we consider robust models under different assumptions on the knowledge available regarding stochastic prices, and propose different approaches for these models. Section 2.5 provides simulation studies that benchmark the worst-case bounds derived from the robust optimization models with Monte-Carlo integration results obtained by using historical price data from wholesale markets. Section 2.6 concludes and discusses possible extensions for future research.

2.2 The Deterministic Model

We start with a deterministic model for the flexible demand management problem, and we use the example of demand response to introduce our model formulation. We assume that time-varying prices are announced and deterministic before making energy usage decisions. The demand arrivals are fixed and deterministic in terms of both arrival time and quantity. The decision maker can be either a single household, or an aggregator that aggregates the demand of multiple households. The control problem is to find the utility optimizing decisions.

Intuitively, when time-varying pricing is offered, decision makers can take advantage of low prices in some periods by shifting their demand. However, shifting demand causes discomfort from not being able to use energy immediately whenever there is demand. For instance, delay in satisfying the demand for air-conditioning leaves decision makers suffering uncomfortable room temperatures. Since decrease in cost can be achieved by lowering the comfort level of decision makers, it is natural that decision makers would like to find the optimal trade-off between comfort and cost savings.

The following example further illustrates the problem. Suppose that a decision maker needs to have a local storage device fully charged by the end of day. Then at the beginning of the first period, the outstanding demand x_1 is set as d_1, which represents the amount of energy required to fully charge the storage device. Then, the decision maker decides u_1 based on price p_1. Meanwhile, during this period, some energy may be extracted from the storage device — demand d_2 "arrives" at the beginning of the second period, and the outstanding demand x_2 equals to the new demand d_2, plus $(x_1 - u_1)$. Then the decision maker decides u_2, and the same process is repeated in every period. At last, because the storage has to be fully charged by the end of day (the n-th period), u_n equals to x_n.

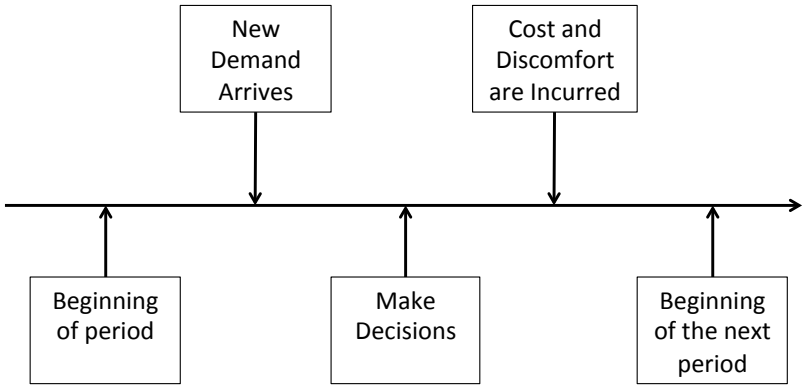


Figure 2.1: Sequence of Events

In the deterministic model and the following models with data uncertainties, we make the following common assumptions on the problem settings. Firstly, we assume that the planning horizon is one day discretized into n periods. Figure 2.1 describes the sequence of events. At the beginning of each period, demand arrives. Then, decisions are made and demand is satisfied. Next, the cost of electricity and the discomfort of delaying the unsatisfied demand are incurred, and the system evolves to the next period.

The second assumption we make is on quantifying cost and discomfort. Without loss of generality, we assume that cost incurred in each period equal to the unit price of electricity times the amount of energy consumed in that period, while the discomfort experienced by decision makers in each period is the product of a unit penalty and the amount of unsatisfied demand in that period. Outstanding demand at the beginning of each period equals to the unsatisfied demand from the last period, plus the new demand arrival. Our last assumption is that all demand needs to be satisfied by the end of day, that is, there should be no unsatisfied demand at the end of the n-th period.

To present the model, we first summarize the main notation in Table 2.1 for quick reference. Other symbols are defined as required throughout the text. In particular, boldface lowercase is used to denote vectors, while non-boldface is used to denote scalars. Boldface uppercase letters are used to denote polytopes.

The energy usage decision in period i is denoted as u_i. Auxiliary decision variables x_i represent the outstanding demand after the new arrivals in period i, and in the context of dynamic programming, x_i can be interpreted as the state status of the system. Users make decisions to minimize cost and discomfort over the entire planning horizon, and thus the multi-objective model for the deterministic problem ($\mathbf{P^D}$) can be formulated as follows:

Notation	Definition
Parameters:	
n:	Number of periods in the planning horizon. Let $N = \{1, 2, \ldots, n\}$
$\mathbf{p}$:	$\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is the vector of prices.
$\mathbf{d}$:	$\mathbf{d} = (d_1, d_2, \ldots, d_n)$ is the vector of demand.
$\mathbf{c}$:	$\mathbf{c} = (c_1, c_2, \ldots, c_n)$ is the vector of discomfort rate.
Decision Variables:	
$\mathbf{x}$	$\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the vector of outstanding demand after demand arrives.
$\mathbf{u}$	$\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is the vector of decisions on how much demand to be satisfied.

Table 2.1: Summary of Main Notation

$$
\begin{aligned}
(\mathbf{P^D}): \quad & \min_{\mathbf{x},\mathbf{u}} \quad \sum_{i \in N} p_i u_i \\
& \min_{\mathbf{x},\mathbf{u}} \quad \sum_{i \in N} c_i (x_i - u_i) \\
& s.t. \quad x_1 = d_1 && \text{(2.1a)} \\
& \qquad x_{i+1} - x_i + u_i = d_{i+1} \qquad \forall i = 1, 2, \ldots, n-1 && \text{(2.1b)} \\
& \qquad x_n - u_n = 0 && \text{(2.1c)} \\
& \qquad u_i \leq x_i \qquad \forall i = 1, 2, \ldots, n && \text{(2.1d)} \\
& \qquad u_i \geq 0 \qquad \forall i = 1, 2, \ldots, n && \text{(2.1e)}
\end{aligned}
$$

There are three sets of constraints: balance constraints, non-anticipating constraints, and non-negative constraints. Constraints (2.1a) - (2.1c) are the balance constraints. In particular, (2.1b) is the transition balance constraint, and (2.1c) makes sure there is no unsatisfied demand at the end of the n-th period. (2.1d) is the non-anticipating constraint, which enforces that no demand can be satisfied before its arrival. (2.1e) is the non-negativity constraint, which excludes the option of shorting. Although the above multi-objective problem can be solved by commercial solvers, the following lemma leads to a possible simplified model.

Lemma 1. *The efficient frontier of problem* $(\mathbf{P^D})$ *is (piecewise-linearly) convex. Then, by varying* w *in the objective of the following problem* $(\mathbf{P1})$ *from zero to positive infinity, all Pareto optimal solutions of problem* $(\mathbf{P^D})$ *can be obtained by solving* $(\mathbf{P1})$*, due to the*

convexity of the efficient frontier of ($\mathbf{P^D}$).

$$(\mathbf{P1}): \quad T(\mathbf{p},\mathbf{c},\mathbf{d}) = \min_{\mathbf{x},\mathbf{u}} \sum_{i\in N} [p_i u_i + wc_i(x_i - u_i)]$$
$$s.t. \quad \textit{Constraints (2.1a) - (2.1e)}$$

We reformulate the multi-objective program ($\mathbf{P^D}$) by combining the two objectives using a scalar w, as shown in ($\mathbf{P1}$). For every Pareto optimal solution of problem ($\mathbf{P^D}$), there exists a w such that the optimal solution or one of the optimal solutions of ($\mathbf{P1}$) generates the same cost and discomfort. Furthermore, w can be interpreted as the coefficient that converts discomfort into dollar-values, and a decision maker chooses w to reflect her preference over all Pareto optimal solutions of ($\mathbf{P^D}$). In particular, she chooses w so that the solution to ($\mathbf{P1}$) corresponds to the Pareto optimal solution of ($\mathbf{P^D}$) that she prefers. Similarly, the objective of ($\mathbf{P1}$) can be interpreted as the total dollar-valued disutility. Since the model with deterministic demand can be viewed as a special case of the model with stochastic demand, we defer the discussion of optimal solutions to problem ($\mathbf{P1}$) to the next subsection.

2.3 The Model with Stochastic Demand Arrivals

Most of the time, demand arrivals are stochastic and accurate demand forecasts are difficult, if not impossible, to obtain. It is non-trivial to decision makers how stochastic demand affects their expected cost and discomfort. It is also interesting to study the optimal control strategy when there is only limited information on demand arrivals. In this section, we formulate the demand-side control problem with demand uncertainty.

The Expectation Minimization Model

Similar to the deterministic model, the optimization problem with stochastic demand has two objectives: minimizing expected cost and minimizing expected discomfort. However, due to the balance and non-anticipating constraints, state status x_i and decision u_i depend on realized demand arrivals $\{d_j\}_{j=1}^{i}$. For notational convenience, we denote the dependence of both decisions and state status on demand realizations by defining the state status and the decision in period i as $x_i(\mathbf{d})$ and $u_i(\mathbf{d})$; however, note that both of them depend on only the realized demand arrivals. The multi-objective formulation with demand uncertainty follows

directly from the deterministic model:

$$
\begin{aligned}
(\mathbf{P}^{\mathbf{S_d}}): \quad & \min_{\mathbf{x}(\mathbf{d}),\mathbf{u}(\mathbf{d})} \quad \mathbb{E}_{\mathbf{d}}\left[\sum_{i\in N} p_i u_i(\mathbf{d})\right] \\
& \min_{\mathbf{x}(\mathbf{d}),\mathbf{u}(\mathbf{d})} \quad \mathbb{E}_{\mathbf{d}}\left[\sum_{i\in N} c_i(x_i(\mathbf{d}) - u_i(\mathbf{d}))\right] \\
& s.t. \quad x_1(\mathbf{d}) = d_1 && \text{(2.2a)} \\
& \qquad x_{i+1}(\mathbf{d}) - x_i(\mathbf{d}) + u_i(\mathbf{d}) = d_{i+1} \qquad \forall i = 1,2,\ldots,n-1 && \text{(2.2b)} \\
& \qquad x_n(\mathbf{d}) - u_n(\mathbf{d}) = 0 && \text{(2.2c)} \\
& \qquad u_i(\mathbf{d}) \le x_i(\mathbf{d}) \qquad \forall i = 1,2,\ldots,n && \text{(2.2d)} \\
& \qquad u_i(\mathbf{d}) \ge 0 \qquad \forall i = 1,2,\ldots,n && \text{(2.2e)}
\end{aligned}
$$

where constraints (2.2a) - (2.2e) are the three sets of constraints with stochastic demand. The main difficulty in solving problem $(\mathbf{P}^{\mathbf{S_d}})$ is that the optimal solutions $\mathbf{u}^*(\mathbf{d})$ and $\mathbf{x}^*(\mathbf{d})$ are functions of demand realizations. There are many possible families of control policies to which the optimal $\mathbf{u}^*(\mathbf{d})$ and $\mathbf{x}^*(\mathbf{d})$ may belong, and we define two of them as follows.

Definition 1. *For fixed* $\mathbf{p}$ *and* $\mathbf{c}$*, the Rationing policy and Threshold policy are defined as follows:*

- ***Rationing Policy****: A Rationing policy specifies a control sequence*

  $$\mathbf{\Pi}^R := \left[\boldsymbol{\pi}_1^R, \boldsymbol{\pi}_2^R, \ldots, \boldsymbol{\pi}_n^R\right]$$

  where $\boldsymbol{\pi}_i^R := \left[\pi_{ii}^R, \pi_{i(i+1)}^R, \ldots, \pi_{in}^R\right]$ *is a* $(n-i+1)$*-dimensional vector, indicating that* π_{ij}^R *percent of the demand that arrives in period* i *will be satisfied in period* j *(* $\forall j \ge i$*), that is,* $u_i(\mathbf{d}) = \sum_{k=1}^{i} \pi_{ki}^R d_k$*;*

- ***Threshold Policy****: A Threshold policy consists of a control sequence*

  $$\mathbf{\Pi}^T := \left[\pi_1^T, \pi_2^T, \ldots, \pi_n^T\right]$$

  indicating that the outstanding demand in period i *is satisfied up to* π_i^T*, and the excess demand is carried to the next period. Thus,* $u_i(\mathbf{d}) = \min(\pi_i^T, x_i(\mathbf{d}))$

In Appendix A.1, we provide an example to show that under some conditions, there exist both rationing and threshold policies that produce Pareto optimal solutions. We further show in Lemma 2 that in order to identify Pareto optimal solutions, we can limit our search in the family of rationing policies.

Lemma 2. *Without any assumption on the prior knowledge about the distributions of demand arrivals, for every Pareto optimal solution of problem* $(\mathbf{P}^{\mathbf{S_d}})$ *there exists at least one corresponding optimal rationing policy.*

Based on Lemma 2, we can show the following result, which echos Lemma 1 and allows us to combine the objectives of $(\mathbf{P}^{\mathbf{S_d}})$ to form a single objective stochastic optimization problem.

Lemma 3. *The efficient frontier of problem* $(\mathbf{P}^{\mathbf{S_d}})$ *is (piecewise-linearly) convex. Then, all Pareto optimal solutions of problem* $(\mathbf{P}^{\mathbf{S_d}})$ *by solving the following problem* $(\mathbf{P2})$ *with* w *in the objective being varied from zero to positive infinity.*

$$(\mathbf{P2}): \quad \min_{\mathbf{x}(\mathbf{d}),\mathbf{u}(\mathbf{d})} \quad \mathbb{E}_{\mathbf{d}}\left[\sum_{i\in N} p_i u_i(\mathbf{d}) + \sum_{i\in N} wc_i(x_i(\mathbf{d}) - u_i(\mathbf{d}))\right]$$
$$s.t. \quad \textit{Constraints (2.2a) - (2.2e)}$$

Recall that $(\mathbf{P1})$ is a special case of $(\mathbf{P2})$. Then, for a given scalar w that expresses the preference over Pareto optimal solutions, Proposition 1 characterizes the optimal solutions that solves both $(\mathbf{P1})$ and $(\mathbf{P2})$.

Proposition 1. *The optimal policy that minimizes the total expected disutility for both* $(\mathbf{P1})$ *and* $(\mathbf{P2})$ *is an All or Nothing (AON) policy, that is,* $u_i = x_i$ *or* $u_i = 0$ *for all* $1 = 1, 2, \ldots, n-1$. *Specifically,*

$$u_i = \begin{cases} x_i & \textit{if } \; p_i \leq wc_i + \Gamma_{i+1} \\ 0 & \textit{if } \; p_i > wc_i + \Gamma_{i+1} \end{cases}$$

where, $\Gamma_n = p_n$, *and* Γ_i *(for all* $1 = 1, 2, \ldots, n-1$*) satisfies:*

$$\Gamma_i = \min\{p_i, wc_i + \Gamma_{i+1}\}$$

Obviously, the *All or Nothing* (AON) policy belongs to the family of rationing policies. Moreover, the AON policy obtained in Proposition 1 depends only on prices $(\mathbf{p})$ and discomfort rates $(\mathbf{c})$. Intuitively, the decision on whether or not to satisfy demand in a certain period depends only on two values: the energy price in that period, and the dollar-valued discomfort rate plus the unit cost of satisfying demand in the subsequent periods. Whenever the price of the current period is low enough to incentivize decision makers to use energy, all outstanding demand should be satisfied in that period.

The Robust Optimization Model with Stochastic Demand Arrivals

The objective of (**P2**) is to minimize the expected total dollar-valued disutility. As discussed above, most of the time there is incomplete information about the stochastic demand arrivals and decision makers would like to know their worst-case total disutility over all possible demand distributions. Therefore, we introduce the following robust optimization model with stochastic demand arrivals.

Let $\mathcal{F}_{\mathbf{d}}$ be the set of all possible demand distributions. The robust optimization problem finds the optimal decision that minimizes the worst-case expected total disutility:

$$(\mathbf{R}-\mathbf{P2}) \quad \min_{(\mathbf{x}(\mathbf{d}),\mathbf{u}(\mathbf{d}))} \left\{ \max_{F_d \in \mathcal{F}_{\mathbf{d}}} \mathbb{E}_{F_d} \left[\sum_{i \in N} wc_i(x_i(\mathbf{d}) - u_i(\mathbf{d})) + \sum_{i \in N} p_i u_i(\mathbf{d}) \right] \right\}$$
$$s.t. \quad \text{Constraints (2.2a) - (2.2e)}$$

From the Stackelberg game's point of view, the above robust optimization problem can be interpreted as two players making sequential decisions. Player one first decides the energy usage decisions $(\mathbf{x}(\mathbf{d}), \mathbf{u}(\mathbf{d}))$ as functions of the realized demand to minimize the expected total disutility. Then, player two chooses the distribution of demand arrivals F_d to penalize player one. The following proposition characterizes the optimal solution to this robust optimization problem.

Proposition 2. *The optimal policy to the robust optimization model* $(\mathbf{R}-\mathbf{P2})$ *is again an AON policy. And it is the optimal policy of problem* $(\mathbf{P2})$ *with the same* $\mathbf{p}$ *and* $\mathbf{c}$.

Proposition 2 comes directly from the fact that demand-side decisions with demand uncertainty is not functions of demand. Note that the derivation of this result does not make any assumption on the set of possible demand distributions.

2.4 When Price Is Uncertain

There exists stronger motivation to study flexible demand management with stochastic prices. For instance, in the context of coupling flexible demand with renewable energy in the Smart Grid, the prices for electricity should be highly correlated with the output from renewable sources such as wind and solar, both of which are extremely unstable. In addition, it is not hard to see that demand uncertainty is endogenous information, about which decision makers have better knowledge, while prices are exogenous to decision makers. Therefore, decision makers tend to be more risk-averse about the prices.

Unlike the case with stochastic demand, there are two scenarios when the prices are stochastic. In one scenario, prices are stochastic, but their realization are announced to decision makers ahead of time. For instance, there are day-ahead markets in the wholesale electricity market, and similar mechanisms can be applied for flexible demand management problems. Consequently, when decision makers schedule the execution of demand, they face deterministic prices throughout their planning horizon. In the other scenario, stochastic prices are realized after making decisions on satisfying outstanding demand.

In the first scenario, the multi-objective problem can be formulated as the following, with $\mathbf{x}(\mathbf{p})$ and $\mathbf{u}(\mathbf{p})$ being the vector of decision variables for the vector of announced future prices $\mathbf{p}$:

$$
\begin{aligned}
(\mathbf{P}^{\mathbf{S_p}} - \mathbf{1}): \quad & \min_{\mathbf{x}(\mathbf{p}),\mathbf{u}(\mathbf{p})} \quad \mathbb{E}_{\mathbf{p}}\left[\sum_{i\in N} p_i u_i(\mathbf{p})\right] \\
& \min_{\mathbf{x}(\mathbf{p}),\mathbf{u}(\mathbf{p})} \quad \mathbb{E}_{\mathbf{p}}\left[\sum_{i\in N} c_i(x_i(\mathbf{p}) - u_i(\mathbf{p}))\right] \\
& s.t. \quad x_1(\mathbf{p}) = d_1 \\
& \qquad x_{i+1}(\mathbf{p}) - x_i(\mathbf{p}) + u_i(\mathbf{p}) = d_{i+1} \qquad \forall i = 1, 2, \ldots, n-1 \\
& \qquad x_n(\mathbf{p}) - u_n(\mathbf{p}) = 0 \\
& \qquad u_i(\mathbf{p}) \le x_i(\mathbf{p}) \qquad \forall i = 1, 2, \ldots, n \\
& \qquad u_i(\mathbf{p}) \ge 0 \qquad \forall i = 1, 2, \ldots, n
\end{aligned}
$$

It is not hard to see that for each of the possible price vector $\mathbf{p}$, $\mathbf{x}^*(\mathbf{p})$ and $\mathbf{u}^*(\mathbf{p})$ can be obtained by solving the corresponding deterministic problem $(\mathbf{P1})$, whose efficient frontier is convex. Therefore, the efficient frontier of problem $(\mathbf{P}^{\mathbf{S_p}} - \mathbf{1})$ is convex as taking expectation preserves convexity.

Define $(\mathbf{u}', \mathbf{x}') = \left(u_i'(\bar{\mathbf{p}}_{(i-1)}), x_i'(\bar{\mathbf{p}}_{(i-1)})\right)_{i=1}^{n}$, where $\bar{\mathbf{p}}_{(i-1)} = (p_1, p_2, \ldots, p_{i-1})$, and for $i = 1$, $\left(u_i'(\bar{\mathbf{p}}_{(i-1)}), x_i'(\bar{\mathbf{p}}_{(i-1)})\right) = (u_1', x_1')$. Then, the multi-objective problem for the second scenario,

on the other hand, can be formulated as the following:

$$
\begin{aligned}
(\mathbf{P}^{\mathbf{S_p}} - \mathbf{2}): \quad & \min_{(\mathbf{u}',\mathbf{x}')} \quad \mathbb{E}_{\mathbf{p}}\left[\sum_{i\in N} p_i u_i'(\bar{\mathbf{p}}_{(i-1)})\right] \\
& \min_{(\mathbf{u}',\mathbf{x}')} \quad \mathbb{E}_{\mathbf{p}}\left[\sum_{i\in N} c_i\big(x_i'(\bar{\mathbf{p}}_{(i-1)}) - u_i'(\bar{\mathbf{p}}_{(i-1)})\big)\right] \\
& s.t. \quad x_1' = d_1 && (2.3a)\\
& \quad x_{i+1}'(\bar{\mathbf{p}}_{(i)}) - x_i'(\bar{\mathbf{p}}_{(i-1)}) + u_i'(\bar{\mathbf{p}}_{(i-1)}) = d_{i+1} \qquad \forall i = 1,2,\ldots,n-1 && (2.3b)\\
& \quad x_n'(\bar{\mathbf{p}}_{(i-1)}) - u_n'(\bar{\mathbf{p}}_{(i-1)}) = 0 && (2.3c)\\
& \quad u_i'(\bar{\mathbf{p}}_{(i-1)}) \le x_i'(\bar{\mathbf{p}}_{(i-1)}) \qquad \forall i = 1,2,\ldots,n && (2.3d)\\
& \quad u_i'(\bar{\mathbf{p}}_{(i-1)}) \ge 0 \qquad \forall i = 1,2,\ldots,n && (2.3e)
\end{aligned}
$$

We can show that the efficient frontier of problem $(\mathbf{P}^{\mathbf{S_p}} - \mathbf{2})$ is convex, under weak conditions.

Lemma 4. *Suppose that prices take on a finite number of possible values and the joint distribution is known if prices are intertemporally correlated, the efficient frontier of problem* $(\mathbf{P}^{\mathbf{S_p}} - \mathbf{2})$ *is convex. Therefore, all Pareto optimal solutions of problem* $(\mathbf{P}^{\mathbf{S_p}} - \mathbf{2})$ *can be obtained by solving the following problem* $(\mathbf{P3})$ *with* w *being varied from zero to positive infinity.*

$$
\begin{aligned}
(\mathbf{P3}): \quad & \min_{\mathbf{x},\mathbf{u}} \quad \sum_{i\in N} \mathbb{E}_{\mathbf{p}}\Big[p_i u_i'(\bar{\mathbf{p}}_{(i-1)}) + w c_i\big(x_i'(\bar{\mathbf{p}}_{(i-1)}) - u_i'(\bar{\mathbf{p}}_{(i-1)})\big)\Big] \\
& s.t. \quad \textit{Constraints (2.3a) - (2.3e)}
\end{aligned}
$$

Suppose that prices are intertemporally independent, a similar *All or Nothing* policy is optimal for the above problem $(\mathbf{P3})$, as shown in the following proposition.

Proposition 3. *If prices are intertemporally independent, the optimal policy that minimizes the total expected disutility for problem* $(\mathbf{P3})$ *is an All or Nothing (AON) policy, that is,* $u_i = x_i$ *or* $u_i = 0$ *for all* $1 = 1,2,\ldots,n-1$. *Specifically,*

$$
u_i = \begin{cases} x_i & \textit{if } \mathbb{E}_{\mathbf{p}}[p_i] \le w c_i + \Gamma_{i+1} \\ 0 & \textit{if } \mathbb{E}_{\mathbf{p}}[p_i] > w c_i + \Gamma_{i+1} \end{cases}
$$

where, $\Gamma_n = \mathbb{E}_{\mathbf{p}}[p_n]$, *and* Γ_i *(for all* $1 = 1,2,\ldots,n-1$*) satisfies:*

$$
\Gamma_i = \min\{\mathbb{E}_{\mathbf{p}}[p_i], w c_i + \Gamma_{i+1}\}
$$

When the prices are intertemporally dependent, it is still possible to obtain the optimal policy when there exists complete information of the joint distribution of prices. As noted above, since there are many causes that make it difficult to infer price distributions, we are more interested in the robust policy that optimizes decision makers' payoffs when only partial information about the price distributions are available, compared to the optimal policy with intertemporally dependent prices. In the following sub-sections, we develop robust optimization models to address this issue. Based on assumptions on prices, we consider two different settings.

The Case When Prices are Symmetrically Distributed on Closed, Bounded Intervals

There have been arguments made on putting upper bounds on the prices to protect decision makers from price spikes. Our first robust optimization model tries to analyze the worst cast total disutility when the prices are bounded, and it is trivial that the worst case happens when prices take values of the upper bounds. However, this worst-case evaluation may be over conservative.

The over conservative can be addressed by allowing decision makers to control their preferred degree of robustness under one additional assumption on the prices. In particular, Let $\Gamma \in [0, n]$ be a scalar that represents the degree of robustness, and let $\mathcal{G}(\Gamma)$ be the set of feasible prices defined as the following. Price p_i in each period is symmetrically distributed on a known interval, that is, $p_i \in [\bar{p}_i - \hat{p}_i, \bar{p}_i + \hat{p}_i]$, where $\bar{p}_i$ is the median of the interval and $\hat{p}_i$ is the spread. The additional assumption is that prices in a total of $\lfloor \Gamma \rfloor$ periods are allowed to deviate freely from $\bar{p}_i$, and the price of another period is allowed to change by $(\Gamma - \lfloor \Gamma \rfloor)\hat{p}_i$. Then, consider the following problem:

$$
\begin{aligned}
(\mathbf{R}_{\mathbf{0}}^{\mathbf{S_P}}) \quad & \min_{\mathbf{x},\mathbf{u}} \quad \max_{\mathbf{p}\in\mathcal{G}(\Gamma)} \quad \sum_{i\in N} (p_i u_i + w c_i (x_i - u_i)) \\
& s.t. \quad \text{Constraints (2.1a) - (2.1e)}
\end{aligned}
$$

To solve the above problem, we use the similar treatment as in [12] to convert the above problem into a linear program. The next proposition finds the equivalent linear programming formulation.

Proposition 4. *Problem* $(\mathbf{R_0^{S_P}})$ *is equivalent to the following linear programming:*

$$
\begin{aligned}
(\mathbf{RLP-P^{S_P}}) \quad \min \quad & Y + \sum_{i \in N} [wc_i(x_i - u_i)] \\
s.t. \quad & x_1 = d_1 \\
& x_{i+1} - x_i + u_i = d_{i+1} \\
& \Gamma\lambda + \sum_{j \in N} \rho_j + \sum_{j \in N} \bar{p_j} u_j - Y \leq 0 \\
& \lambda + \rho_j \geq \hat{p_j} u_j \quad \forall j \\
& \lambda \geq 0, \ \boldsymbol{\rho} \geq 0 \\
& 0 \leq u_j \leq x_j \quad \forall j
\end{aligned}
\tag{2.4}
$$

Solving the above LP returns the robust solution with the robustness characterized by the parameter Γ. The probability bound on violating constraint (2.4) can be derived following the logic of the approach described in [12]. Moreover, Proposition 4 still holds if the distributions of prices are not symmetrical on the pre-announced intervals — only will the probability bound on violating constraint (2.4) fail. Problem $(\mathbf{R_0^{S_P}})$ and the solution approach described in Proposition 4 generate the "one-shot" robust solution and the corresponding worst-case total disutility. However, no conclusion on the long-term average total disutility can be drawn.

The Case When Only the First and Second Moments of Prices are Known

In a more generic setting, there should be no limitation on the support of prices. For instance, the electricity prices can even be negative when the real time supply overwhelms demand, which is justified by true stories that have happened in wholesale electricity markets[2]. As a result, the one shot worst-case total disutility goes unbounded and thus provides less useful information.

On the other hand, information about the long-term average worst-case total disutility is more valuable to risk-averse decision makers. We are able to calculate it when information such as the marginal moments of prices are known. The marginal moments can be obtained much easier when there is sufficient historical data. In the following analysis, we assume that only the first and second moments of prices are known, while the exact distributions are hidden from decision makers. Let $\mathcal{F}_\mathbf{p}$ denote the set of feasible distributions of prices,

[2]Source: U.S. Energy Information Administration.
URL: http://www.eia.gov/todayinenergy/detail.cfm?id=5110

defined as follows:

$$\mathcal{F}_{\mathbf{p}} = \left\{ F_{\mathbf{p}} \;\middle|\; \begin{array}{l} \int_{\mathbb{R}^n} dF_{\mathbf{p}}(\mathbf{p}) = 1 \\ \int_{\mathbb{R}^n} p_i dF_{\mathbf{p}}(\mathbf{p}) = \mu_i \quad \forall i = 1, 2, \ldots, n \\ \int_{\mathbb{R}^n} p_i^2 dF_{\mathbf{p}}(\mathbf{p}) = \mu_i^2 + \sigma_i^2 \quad \forall i = 1, 2, \ldots, n \\ F_{\mathbf{p}}(\mathbf{p}) \geq 0 \end{array} \right\}$$

Recall that there are two possible pricing schemes as discussed at the beginning of Section 2.4. We analyze robust solutions for each of them. The first scheme assumes that prices are realized after making decisions. Note that when joint distributions are not known, information about prices in the past has no value. Therefore, decisions $(\mathbf{u}, \mathbf{x})$ are not functions of past prices. Let set $\mathbf{X}$ be the set of feasible $(\mathbf{x}, \mathbf{u})$ defined by constraints (2.1a) - (2.1e), then the robust optimization problem under the first pricing scheme can be formulated as follows:

$$(\mathbf{R}^{\mathbf{S_P}} - \mathbf{1}) : \quad \min_{(\mathbf{x},\mathbf{u}) \in \mathbf{X}} \left\{ \max_{F_{\mathbf{p}} \in \mathcal{F}_{\mathbf{p}}} \quad \mathbb{E}_{F_{\mathbf{p}}} \left[\sum_{i \in N} c_i(x_i - u_i) + \sum_{i \in N} p_i u_i \right] \right\}$$

From the *Stackelberg game*'s point of view, $(\mathbf{R}^{\mathbf{S_P}} - \mathbf{1})$ indicates that decision makers make decisions first, then the invisible player chooses the price distributions $F_{\mathbf{p}}$ to penalize decision makers. In the optimization context, problem $(\mathbf{R}^{\mathbf{S_P}} - \mathbf{1})$ is a *min-max* problem, in which minimization is taken over set of feasible solutions, $\mathbf{X}$, and the maximization problem is to find the price distribution that maximizes the expectation of the total disutility over all distributions that have mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$. The optimal solution to this problem is trivial, as the expectation in the inner problem can be applied directly on the prices.

It is worthwhile to point out the major caveat of the first pricing scheme here. Under the first pricing schemes, decision makers do not know prices when making decisions, as prices are set to reflect real time demand, that is, prices should be functions of marginal generation cost (and some other factors). However under this pricing scheme, generators or service entities are able to exert market power by intentionally consuming massive energy in peak hours and drive up market clearing prices. Consequently, huge price spikes are created, and the reliability of the gird is undermined.

The second pricing scheme avoids most of the drawbacks of the first one. Besides, it is still possible to set RTP to reflect the balance between supply and demand, see for example, [47]. We briefly illustrate a possible pricing mechanism to justify our assumption: an aggregator receives an initial vector of prices and broadcasts it to decision makers. Decision makers take the prices as deterministic and solve problem $(\mathbf{P1})$. Next, the aggregator aggregates the usage information and sends it to the supplier as feedback. Then the supplier re-optimizes the prices based on the reported future usage and sends the new price vector to the aggregator.

By repeating this procedure, an equilibrium price vector can be attained and used as the final future prices. We maintain the assumption that equilibrium prices are drawn from some unknown distribution, where decision makers know only the first and second moments of prices. Then the robust problem of optimizing the long-term average worst-case total disutility can be formulated as follows:

$$
\begin{aligned}
(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2}): \quad & \max_{F_{\mathbf{p}}} \quad \mathbb{E}_{F_{\mathbf{p}}}\left[T(\mathbf{p}, \mathbf{c}, \mathbf{d})\right] \\
& s.t. \quad \int_{\mathbb{R}^n} dF_{\mathbf{p}}(\mathbf{p}) = 1 && (2.5)\\
& \int_{\mathbb{R}^n} p_i dF_{\mathbf{p}}(\mathbf{p}) = \mu_i \quad \forall i = 1, 2, \ldots, n && (2.6)\\
& \int_{\mathbb{R}^n} p_i^2 dF_{\mathbf{p}}(\mathbf{p}) = \mu_i^2 + \sigma_i^2 \quad \forall i = 1, 2, \ldots, n && (2.7)\\
& F_{\mathbf{p}}(\mathbf{p}) \geq 0
\end{aligned}
$$

where $T(\mathbf{p}, \mathbf{c}, \mathbf{d})$ is the optimal objective value of the deterministic problem $(\mathbf{P1})$. In problem $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2})$, decision makers make decisions after observing prices. Problem $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2})$ can be viewed as the max-min counterpart of problem $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{1})$. Therefore, it is expected from weak duality that, the optimal objective value of $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2})$ is no greater than that of problem $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{1})$. From game theory's point of view, this comes from the fact that the invisible player moves after observing the decisions of decision makers in problem $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{1})$; thus she has more information and is in better position than in problem $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2})$. With limited information on prices, $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2})$ is harder to solve. Next, we show how to solve this optimization problem.

Given the decisions of decision makers, the outer problem maximizes the expected total disutility over all distributions satisfying constraint (2.5) - (2.7). Therefore, the outer problem is an infinite dimensional linear program. Constraint (2.5) indicates that the decision variable $F_{\mathbf{p}}$ of the outer problem is the cumulative distribution function of the prices $\mathbf{p}$. Constraint (2.6) and (2.7) set the first and second moments for the prices. Let θ, $\boldsymbol{\rho}$, and $\boldsymbol{\eta}$ be the dual variables associated with constraints (2.5), (2.6), and (2.7), respectively. We first take the dual of the inner problem.

Proposition 5. *The optimal objective value of problem* $(\mathbf{R}^{\mathbf{S_p}} - \mathbf{2})$ *equals to that of the following optimization problem:*

$$
\begin{aligned}
\min_{\theta, \boldsymbol{\rho}, \boldsymbol{\eta}} \quad & \theta + \sum_{i \in N} \rho_i \mu_i + \sum_{i \in N} \eta_i (\mu_i^2 + \sigma_i^2) \\
s.t. \quad & \min_{(\mathbf{x}, \mathbf{u}) \in \mathbf{X}} \left\{ \max_{\mathbf{p} \in \mathbb{R}^n} \left[\sum_{i \in N} \left((u_i - \rho_i) p_i - \eta_i p_i^2 \right) \right] + \sum_{i \in N} c_i \left(x_i - u_i \right) \right\} \leq \theta && (2.8)
\end{aligned}
$$

where, set $\mathbf{X}$ *is defined as:*

$$\mathbf{X} = \left\{ (\mathbf{x}, \mathbf{u}) \;\middle|\; \begin{array}{l} x_1 = d_1;\; x_{i+1} - x_i + u_i = d_{i+1}, \quad \forall i = 1, \ldots, n-1; \\ u_n - x_n = 0;\; 0 \le u_i \le x_i \quad \forall i = 1, \ldots, n \end{array} \right\}$$

Compared to problem $(\mathbf{R^{S_P} - 2})$, we get rid of the integration over the unknown joint distribution in the dual problem. However, the dual problem is still difficult to solve. Nonetheless, we can first convert the dual into a min-max problem by eliminating θ. Notice that (2.8) must be binding at optimal, thus by replacing θ in the objective with the LHS of constraint (2.8) and combining the minimization operators, we arrive at the following optimization problem:

$$\min_{\boldsymbol{\rho}, \boldsymbol{\eta}, (\mathbf{x}, \mathbf{u}) \in \mathbf{X}} \left\{ \max_{\mathbf{p} \in \mathbb{R}^n} \left[\sum_{i \in N} \left((u_i - \rho_i) p_i - \eta_i p_i^2 \right) \right] + \sum_{i \in N} c_i (x_i - u_i) \right.$$
$$\left. + \sum_{i \in N} \rho_i \mu_i + \sum_{i \in N} \eta_i (\mu_i^2 + \sigma_i^2) \right\} \tag{2.9}$$

To solve the above min-max problem, we first analyze the sign of variable $\boldsymbol{\eta}$ at optimal, which leads us to the following useful results:

Proposition 6. *An optimal solution* $(\boldsymbol{\rho}^*, \boldsymbol{\eta}^*, \mathbf{x}^*, \mathbf{u}^*)$ *to problem (2.9) must satisfy the following conditions, for all* $i \in N$*:*

- $\eta_i^* \ge 0$*;*
- $u_i^* = \rho_i^*$ *if* $\eta_i^* = 0$*;*
- $\eta_i^* = 0$ *if* $u_i^* = \rho_i^*$*;*
- $p_i^* = \dfrac{(u_i - \rho_i)}{2\eta_i}$ *if* $\eta_i^* \ne 0$*, and* $p_i^* = 0$ *if* $\eta_i^* = 0$

.

With Proposition 6, we can derive an equivalent second order cone program for problem $(\mathbf{R^{S_P} - 2})$. However, The following proposition further characterizes the optimal solution, and leads to an important result.

Proposition 7. *The optimal solution to problem (2.9) must satisfy* $\boldsymbol{\eta}^* = \mathbf{0}$*, and* $\mathbf{u}^* = \boldsymbol{\rho}^*$

Let Z^M_{worst} be the optimal objective value of problem $(\mathbf{R^{S_P}} - \mathbf{2})$. Then solving for Z^M_{worst} boils down to solving the following deterministic problem:

$$Z^M_{\text{worst}} = T(\boldsymbol{\mu}, \mathbf{c}, \mathbf{d}) = \min_{\mathbf{x},\mathbf{u}} \quad \sum_{i \in N} [\mu_i u_i + w c_i (x_i - u_i)]$$
$$s.t. \quad \text{Constraints (2.1a) - (2.1e)}$$

This result suggests that when prices are uncertain and only the first two moments are known to decision makers, the worst-case long-term average total disutility of decision makers equals to $T(\boldsymbol{\mu}, \mathbf{c}, \mathbf{d})$, the optimal total disutility of a deterministic problem, the prices of which equal to the marginal first moments of the stochastic prices. Intuitively, although prices are drawn from some unknown distribution, decision makers always solve a deterministic problem, as they observe the prices first. On the other hand, the total disutility of the deterministic problem is piece-wise concave in prices, based on Proposition 1. Therefore, if the uncertain prices have means equal to the prices of the deterministic problem, by Jensen's inequality the long-term average total disutility should not be greater than the optimal objective value of the deterministic problem.

The most important implication of the above result states that decision makers are potentially better off if they are charged with stochastic prices under the second pricing scheme, instead of fixed prices with values equaling to the means of the stochastic ones. This result can help encourage the adoption of demand response programs.

The Case with Stochastic Demand and Prices

When both the prices and demand are stochastic, and the only partial information about them are know, we can again use robust optimization to analyze the worst-case long-term average total disutility. Based on previous discussion, the following setting is more likely to take place: day ahead prices are announced to decision makers, and prices are assumed to be drawn from unknown distribution $F_{\mathbf{p}} \in \mathcal{F}_{\mathbf{p}}$. After observing the prices, decision makers face stochastic demand arrivals, which are drawn from distribution $F_{\mathbf{d}} \in \mathcal{F}_{\mathbf{d}}$. Then the worst-case long-term average total disutility can be obtained by solving the following robust optimization problem:

$$\max_{F_{\mathbf{p}} \in \mathcal{F}_{\mathbf{p}}} \mathbb{E}_{F_{\mathbf{p}}} \left\{ \min_{(\mathbf{x}(\mathbf{d},\mathbf{p}),\mathbf{u}(\mathbf{d},\mathbf{p}))} \max_{F_{\mathbf{d}} \in \mathcal{F}_{\mathbf{d}}} \mathbb{E}_{F_{\mathbf{d}}} \left[\sum_{i \in N} c_i (x_i(\mathbf{d}, \mathbf{p}) - u_i(\mathbf{d}, \mathbf{p})) \right] \right\}$$

Following from previous results, the optimal solution and the optimal objective value of the above problem can be characterized by Corollary 1.

Corollary 1. *The solution that optimizes the worst-case long-term average total disutility and the corresponding worst-case objective value when both prices and demand arrivals are*

stochastic with incomplete information about the distributions are those of the deterministic problem $T(\mathbb{E}[\mathbf{p}], \mathbf{c}, \mathbb{E}[\mathbf{d}])$.

An Lower Bound on the Long-Term Average Total Disutility

Compared to stochastic dynamic programming approaches, the solution obtained via robust optimization approaches may perform badly on average, because the latter optimizes the objective in the worst case, and we are interested in the worst case because of the lack of knowledge about stochastic prices and demand. Obviously, when there is complete information, the robust optimization approach is unnecessary.

One drawback of the robust optimization approach is over conservative. To evaluate the worst-case solution, we provide an lower bound on the long-term average total disutility when only partial information about the uncertain prices is known. In particular, we assume the marginal distributions of prices are known. Obviously, this lower bound is tighter than the one with only information regarding the first and second moments of the prices in each period. Suppose that the marginal *pdf* and *cdf* of prices in period j are $f_j(p)$ and $F_j(p)$, and let $\widetilde{\mathcal{F}_p} := \{F_p | \int F_p(p_j, p_{-j}) dF_{-j}(p_{-j}) = F_j(p_j)\}$ be set of joint distributions of prices with marginal distributions F_j. Then, the problem that solves for the lower bound of the long-term average total disutility is:

$$Z^D_{\text{best}} = \min_{F_p \in \widetilde{\mathcal{F}_p}} \{T(\mathbf{p}, \mathbf{c}, \mathbf{d})\}$$

Let y_{ij} denote the probability that demand arrives in period i will be satisfied in period j, for all $j \geq i$. Denote $\sum_{k=i}^{j-1}$ as $\hat{c}_{ij}$. The optimal solution and optimal objective value of the above problem can be characterized by the following proposition.

Proposition 8. *When the marginal probability distribution functions of prices are* $f_j(p)$*, the lower bound on the long-term average total disutility* Z^D_{best} *is:*

$$Z^D_{best} = \sum_{i \in N} d_i \sum_{j \geq i} \int_0^{y^*_{ij}} (F_j^{-1}(t) + \hat{c}_{ij}) dt$$

where, $y^*_{ij} = F_j(-\hat{c}_{ij} - \lambda^*_i)$ *for all* $j \geq i$*, and* λ^*_i *is the solution to equation* $\sum_{j \geq i} F_j(-\hat{c}_{ij} - \lambda^*_i) = 1$*. Moreover, the lower bound* Z^D_{best} *is tight.*

The lower bound Z^D_{best} is attainable, and it is attained when joint distribution favors decision makers. Moreover, Z^D_{best} is calculated by minimizing over all possible joint distributions, which is extremely difficult to do. Proposition 8 suggests that we can instead calculate the lower bound by solving equations and taking integrations. The proof of Proposition 8 follows the ideas from the work of [46] and [49].

2.5 Numerical Study

In this section, we perform numerical experiments to test the robust solutions under wholesale electricity prices data. Qualitatively, the relationship between the worst case with marginal distribution and marginal moments, the best case with marginal distribution and marginal moments, and the expected total payoff with complete information satisfies the following inequality:

$$Z^M_{\text{best}} \leq Z^D_{\text{best}} \leq Z^*_{\text{complete}} \leq Z^D_{\text{worst}} \leq Z^M_{\text{worst}} \tag{2.10}$$

However, it is difficult to calculate the expected payoff with complete information (Z^*_{complete}) and the worst case with marginal distribution (Z^D_{worst}). Therefore, we turn to generate bounds for Z^*_{complete} and Z^D_{worst}, by simulating Z^D_{best} and Z^M_{worst}. The value of Z^M_{best} is less interesting when decision makers are not risk-seeking, hence is not analyzed in the chapter. For benchmark purpose, we use the joint distributions for the generation of prices to calculate the corresponding Z^*_{complete} value. Note that the joint distributions are not known in the problems that solve for Z^D_{best} and Z^M_{worst}. In particular, we seek to achieve the following four targets through numerical experiments:

- Show the effectiveness of demand scheduling in reducing electricity cost for household usage;
- Demonstrate the effectiveness of robust control strategy in hedging the risks of high losses when prices are undesirable;
- Evaluate the price of *robustness*;
- Quantify the impacts of discomfort penalty cost on the decision maker's energy usage behaviors.

The experiments are set up as follows: we simulate the case of energy usage in a typical US household for a whole day with 24 one-hour periods. The rated power of each appliance in a household is listed in Table 2.2 and the hourly aggregate load profile of these appliances based on typical household usage patterns is depicted in Figure 2.2. We assume that the load profile is fixed. In fact, not all the loads listed below are deferrable, such as lighting. In our study, we neglect these exceptions for the following two reasons: (1) the

Appliance	Power Consumptions (kW)
Air Conditioning	3.06
Small Appliance	0.76
Refrigerator	0.74
Dish Washer	0.86
Clothes Washer	0.81
Clothes Dryer	1
Miscellaneous Loads, Lighting	0.38
Charging of Electric Vehicle	1.7

Table 2.2: List of Appliances Specifications in a Typical Household in US

power consumptions of non-deferrable loads, such as lighting, computer charging power, are much smaller than those of deferrable loads such as air-conditioning, dish washer, etc., and (2) in the near future, people will be able to install small-scale distributed energy storage systems or utilize the energy in electric vehicle batteries to satisfy these non-deferrable loads to make their inflexible demands "flexible". We fix the dollar-valued discomfort rate at $0.01\$/KWh$. Meanwhile, we employ the locational marginal prices data of a grid connection node in ERCOT market (Texas, 2011) due to the lack of real-time electricity prices for end-users in retail markets. Figure 2.3 shows the average (wholesale) electricity prices in a day. Due to the nonlinearity in power system's operation, the locational marginal prices in wholesale markets vary with large variance in the same period of time in different days of a year. However, price variations are suspected to be small for end-users in retail markets in the future. Therefore, instead of directly using the variance of real locational marginal prices, we limit the coefficient of variation in a reasonable range and vary this ratio to see how price fluctuations affect demand scheduling decisions. For convenience, we denote this ratio as λ. Then, prices for simulation are calculated according to Equation (2.11).

$$p^j_{scale,t} = \frac{p^j_{real,t} - \bar{p}_{real,t}}{\sqrt{p_{real,t}}} \times \lambda \times \bar{p}_{real,t} + \bar{p}_{real,t},\ t = 1, 2, \cdots, 24,\ j = 1, 2, \cdots, 365 \tag{2.11}$$

First, we compute the electricity cost and disutility under two control strategies and three price profiles and seek to verify the benefits of demand scheduling in reducing electricity cost. The two control strategies are (1) no control, and (2) control with known price rates. The three price profiles are (1) Price 1: fixed constant price rates (average price in a day) (2) Price 2: deterministic varying price rates (price profile depicted in Figure 2.3) and (3) Price 3: stochastic time-varying price rates with $\lambda = 1$. The results are summarized in Table 2.3. The results indicate that with fixed constant price rates, no incentives (price differences) are provided to shift demand from high price periods to low price periods and therefore no cost savings are achieved. While under time-varying prices, the benefits of demand shifting appear (13.3% savings in electricity bill). When prices are stochastic, more cost savings (17.1% reduction) could be harvested through demand scheduling as in the case of Price 3.

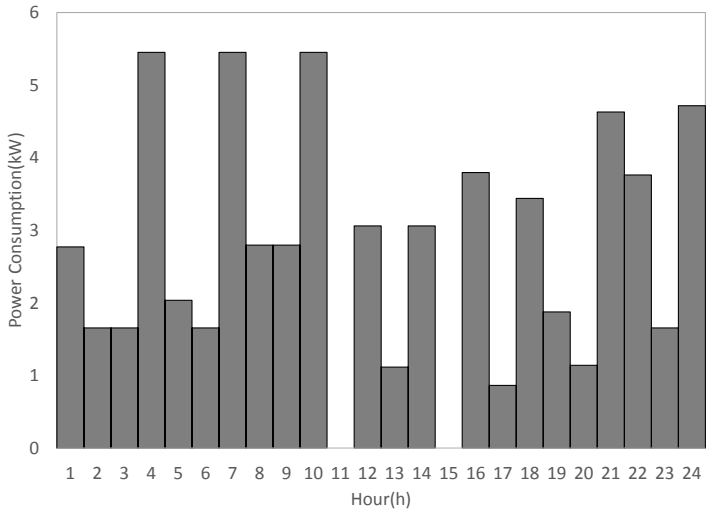


Figure 2.2: Hourly Power Consumption in a Day in a Typical US House

Rates	Without control		With Control	
	Elec. cost	Tot. Disutility	Elec. cost	Tot. Disutility
Price 1	2.9688	2.9688	2.9688	2.9688
Price 2	2.7178	2.7178	2.3559	2.4405
Price 3(Worst Case)	5.4357	5.4357	4.5078	4.7148

Table 2.3: The Electricity Cost($/day) and Disutility($/day) Comparisons Between Two Control Strategies and Three Price Rates

Second, we generate bounds for $Z^*_{complete}$ by simulating Z^D_{best} and Z^M_{worst}. Meanwhile, we approximate $Z^*_{complete}$ by calculating average payoff with 365 days of real data in year 2011 after scaling by Equation (2.11) to required variance and see whether it fits in the range of $[Z^D_{best}, Z^M_{worst}]$. We denote this value as $\hat{Z}^*_{complete}$.

In addition, considering the factors that will impact the prices are substantial, we assume that the marginal distribution of prices in each time period follows the normal distribution when calculating Z^D_{best}. The results are presented in Table 2.4 and Figure 2.4.

The results clearly show that the values of $\hat{Z}^*_{complete}$ under different standard deviations are successfully bounded by the values of Z^D_{best} and Z^M_{worst}. When the value of λ increases,

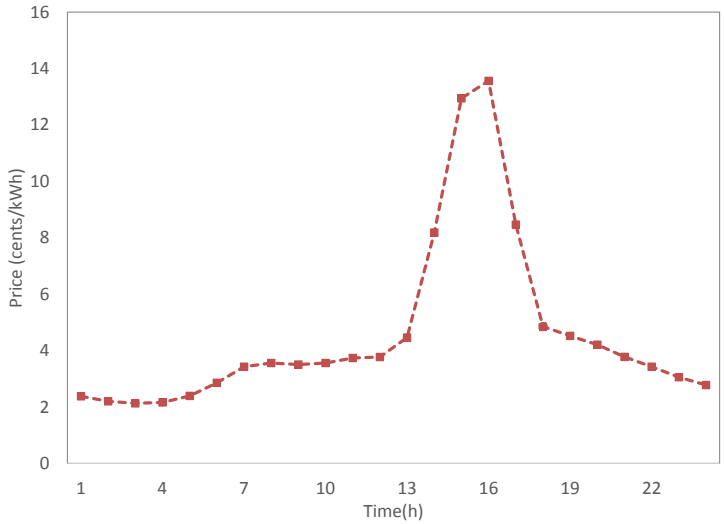


Figure 2.3: Hourly Average Electricity Prices

λ	Z^D_{best}	$\hat{Z}^*_{complete}$	Z^M_{worst}
0.0	2.4405	2.4405	2.4405
0.5	1.6134	2.3857	2.4405
1.0	-0.2637	2.2962	2.4405
1.5	-2.8656	2.1714	2.4405
2.0	-5.8407	2.0095	2.4405
2.5	-0.8995	1.8161	2.4405
3.0	-1.2239	1.5883	2.4405

Table 2.4: The Disutility Comparisons Between Z^D_{best}, $\hat{Z}^*_{complete}$ and Z^M_{worst} under Different λ

the gap between the two bounds expands, and it becomes harder to estimate the value of $Z^*_{complete}$. But the estimated $\hat{Z}^*_{complete}$ value based on one year real data is closer to Z^M_{worst} value. For the value of Z^D_{best}, specially, when λ increases above 1, it becomes negative. It is explained that in best case with large price variance, sometimes the price becomes negative. Users could possibly shift their demand to negative price periods to earn money and make their disutility in the long run negative. Meanwhile, the optimal results under $\mathbf{RLP} - \mathbf{P^{S_P}}$ are also computed and shown in Figure 2.4. It is necessary to note that this result is on daily base but not expected value in the long run. It increases with the increase in price variance. This indicates that with more uncertainties in prices, users' response to price changes are more conservative and higher disutility will be incurred.

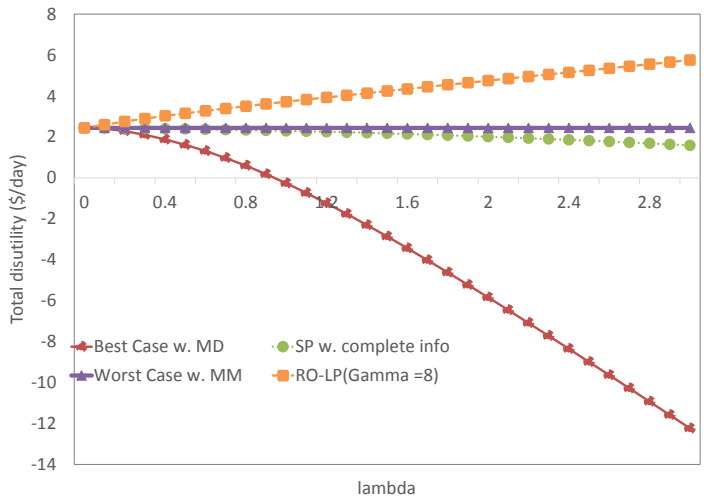


Figure 2.4: Cost Under Different Models With Varying λ

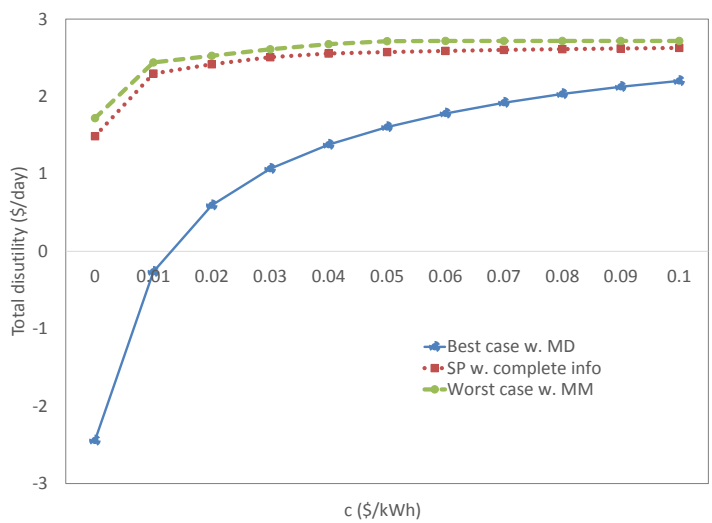


Figure 2.5: Sensitivity Analysis on **c**

Finally, in order to consider decision makers' different perceptions on *discomfort* when

energy demands are delayed, we carry out a sensitivity analysis on parameter $\mathbf{c}$. Still, it is reasonable to assume that the values of c_t in different time periods are the same in average case. For notational convenience, this value is denoted as c. Again, we assume that the value of λ equals to 1. The simulation results are shown in Figure 2.5. The gaps between the best case with marginal distribution (Z^D_{best}) and the worst case with marginal moments (Z^M_{worst}) decrease as the dollar-valued discomfort increases. This can be partly explained by that with increased discomfort dollar-valued cost, people will become less inclined to shift their load even with the low price incentives in other periods. The bounds generated by calculating Z^D_{best} and Z^M_{worst} converges. Additionally, under different discomfort rates, Z^M_{worst} again provides a better estimate for the value of $Z^*_{complete}$. It means that the price of robustness ($Z^M_{worst} - Z^*_{complete}$) with marginal distribution known is not high.

2.6 Summary

In this chapter, we study the *flexible demand management* problem. We start with a finite horizon deterministic multi-objective model with two objectives as minimizing total cost and minimizing total discomfort, followed by similar multi-objective problems with stochastic demand and prices. Then, we convert the multi-objective problems into single objective ones by exploiting the convexity of the efficient frontiers. For the deterministic problem, the problem with stochastic demand, and the problem with intertemporal independent stochastic prices, we develop *All or Nothing* policies and prove their optimality.

Next, we focus on the robust solutions that optimize the payoffs of decision makers when only partial information about the distributions of the stochastic demand or prices is available. We first show that the robust optimal solution when demands are stochastic is again an *All or Nothing* policy as a function of prices and discomfort rates. Then, we study the robust solutions when prices are stochastic and decision makers know only marginal first and second moments. The analytical results suggest that no matter whether the realized prices are announced to decision makers before making decisions or not, the worst-case long-term average total payoff is the same as the optimal payoff of deterministic problem, the prices of which take values of the marginal first moments of the stochastic prices. The result is extremely useful, in that it helps bound the worst-case payoff when only limited information is available to decision makers, and it shows that decision makers are potentially better off if stochastic prices are offered.

To get better understanding of the differences between the worst-case average payoff and the payoffs when more information is known or even with complete information, we analyze the best-case average payoff when marginal distributions of prices are available. Then, we conduct numerical study by calculating the best-case and worst-case payoffs and simulating the average payoffs with complete information. The numerical studies verify the relation-

ships between the bounds, and show the impact of different marginal price distributions on the payoffs of decision makers.

There are several extensions that worth further investigation in the future. First of all, it is interesting to understand the case in which there are capacity constraints on total outstanding demand at the end of each period. Secondly, we are not able to analytically calculate the robust solutions when marginal distributions of prices are available. Although we suspect that the benefit of having a better lower bound with more information is marginal, it is still interesting to answer the question of how extra information affects the robust solutions.

Chapter 3

A Complete Distribution-Free Model

3.1 Introduction.

In this chapter, we base our study on real-time pricing (RTP), because among the three pricing schemes, only the RTP is capable to reflect all changes in supply-demand balancing by prices. In addition, [32] show the improvement of short-run efficiency by adopting RTP; besides, [17] demonstrates that the long-run economic efficiency gains of using RTP is significant, and it weighs much higher than that of adopting TOU. Pilot studies also show the benefit of adopting RTP ([3]).

On the other hand, advances in technologies such as smart metering have made RTP possible. Obviously, the users can benefit from the RTP scheme if they trade off some comfort for cost saving by shifting their flexible demand to periods with lower electricity prices, or by storing electricity when prices are low. Nonetheless, the diffusion of RTP and DR is very slow. According to [37], one of the reason is the lack of effective and efficient energy management mechanisms that react to the time-varying prices. Indeed, it is implausible to keep users manually adjusting their consumption according to real-time prices. Therefore, in this chapter, we focus on automatic energy usage management based on the assumption that load serving entities (LSE) offer RTP to reflect the current or forecasted demand.

The problem of demand management mechanisms under the RTP has been studied in the literature, such as [24], [34], [45], [54], and [39]. In particular, [24] utilize a robust optimization approach to model the price uncertainty. Their model implicitly assumes that future demand is positively correlated with current decision on consumption. [34] propose a model that integrates a two-periods market and develop algorithms that solve for the optimal day-ahead and real-time energy procurement decision. [54] study the case when prices are uncertain but demand is known a priori. They also study the aggregate demand as a function of the prices. [39] study a control system to optimize electricity usage for residential users under supply capacity constraints. [45] is closely related to our work. They propose

an Energy Management System (EMS), which aims at shaving off peak demand in order to achieve cost saving for the users. They vision that the EMS supports smart appliances, local electricity generation, storage, and the sale of electricity back to the grid.

Most of the above work is based on stochastic dynamic programming (DP) ([5]). For instance, [45] formulate a stochastic dynamic programming model for the EMS and solve it by backward induction on discretized state space. Others such as [39] are also based on DP. Nevertheless, a large amount of similar problems suffers from the "curses of dimensionality" ([53]), which always make it computationally intractable to obtain optimal solutions, that is, in order to obtain the optimal solution by backward induction for the EMS, the states need to be highly aggregated, in that the complexity of backward induction grows exponentially in the size of input. However, the high level of state aggregation causes poor performance of the solution.

Consequently, *Approximate Dynamic Programming* (ADP) approaches have been developed to generate suboptimal policies for those problems that are originally hard to solve. There is a variety of approximation methods, and some are extremely powerful and efficient. For instance, [28] introduce a linear programming approach for infinite horizon problems with steady states. The approximation approach used in this chapter proceeds forward in time, simulates into the future and iteratively updates the estimation of the value-to-go term. For a more comprehensive survey of other ADP methods, we refer the readers to [10], [9], and [53]. [41], [51], and [39] among others implement ADP approaches. Particularly, [41] develop an ADP approach to compute the lower and upper bounds on the value of gas storage. [51] apply ADP to efficiently obtain control policies that couple deferrable demand with renewable sources of energy. [39] follow another ADP approach by limiting the decision space, and obtain estimation for the value-to-go function by sampling.

In this chapter, we study an important operations management problem: managing electricity usage for users confronted with RTP to minimize the expected total disutility. Our model is featured with the following unique properties: (1) it is able to adapt to various price structures, such as convex increasing price functions, which can be utilized to mitigate the "rebound effect" noted by [16] and [39], (2) it allows each appliance to have its own allowable delay, (3) it models additive demand, such as the demand for HVAC, in such a way that the decision variables are integers, and (4) it has only minimum requirement on the information regarding the random processes. The problem is hard to solve optimally due to the fact that the related stochastic processes are not fully known and the optimization problem is a non-linear integer problem. We propose an efficient approximate dynamic programming approach by parameterizing the *value-to-go function* based on the special property of the optimal value-to-go. The ADP approach learns the value-to-go without knowing the distribution of the stochastic processes. Since the approximate optimization problem is still hard to solve, we first convert it to an NP-hard Mixed-integer linear problem, which we call the *Variable Budget Precedence Constraints Knapsack Problem* (VBPCKP). Then, we outline

an algorithm that efficiently solves the approximate optimization problem. At last, we show by simulations that this efficient ADP approach provides comparable price elasticity to that of the backward induction approach, and its performance significantly outweighs a myopic approach and the traditional no-control approach.

The remainder of the chapter is organized as follows. Section 3.2 describes the characteristics of the problem and our model formulation. In Section 3.3, we provide an approximate approach that solves the problem efficiently using the ADP techniques. Then we run simulations to compare the performance of the ADP approach with those of other approaches in Section 3.4. Section 3.5 describes ways to generalize the proposed model for other supply chain management problems, provides future directions for research extensions, then concludes.

3.2 Model

We start this section by summarizing the main notation and terms in Table 3.1 and Table 3.2 for quick reference. Other symbols are defined as required throughout the text. In this chapter, calligraphy is used to denote sets, lowercase is used to denote scalars, while lowercase boldface denotes (column) vectors. In addition, unless otherwise noted, over line ($\bar{x}$) and hat ($\widehat{x}$) denote the forecast (or estimated) and the realization of random values (x), respectively.

Overview

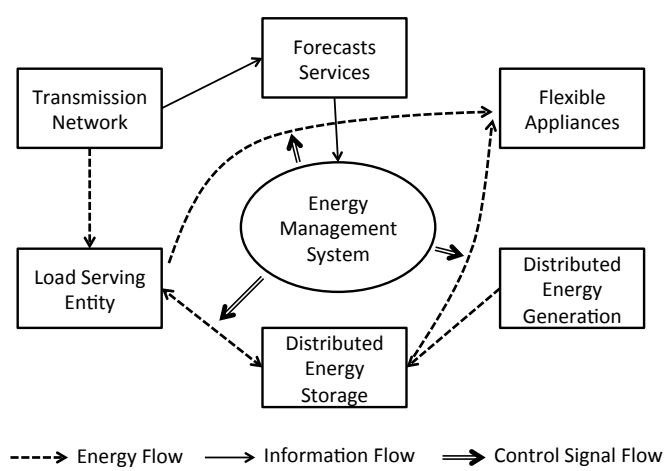


Figure 3.1: Diagram of the System, and the Energy, Information and Control Flow

Notation	Definition
Inputs:	
$\mathcal{A} := \{a : a = 1, 2, \ldots, \lvert\mathcal{A}\rvert\}$	the set of all demand tasks that are flexible. We refer to the demand of task a that will be due in n periods as a type $a^{[n]}$ demand.
$\mathcal{K} := \{k : k = 1, 2, \ldots, \lvert\mathcal{K}\rvert\}$	the set of all appliances. $\mathcal{S} \subset \mathcal{K}$ is the subset of all appliances on which demand is additive.
$\mathcal{A}_k \subset \mathcal{A}$	the subset of demand tasks that are associated with appliance $k \in \mathcal{K}$.
u_b, l_b, c_b	the maximum discharging, charging rate and the capacity of the battery
Random Information:	
$\mathbf{q}_j$	the vector of the demand arrivals in period j. $\mathbf{q}_j = (\mathbf{q}_{1j}^\intercal, \mathbf{q}_{2j}^\intercal, \ldots, \mathbf{q}_{\lvert\mathcal{A}\rvert j}^\intercal)^\intercal$, where $\mathbf{q}_{aj} = (q_{a^{[1]}j}, q_{a^{[2]}j}, \ldots, q_{a^{[N]}j})^\intercal$. $q_{a^{[n]}j}$ is the type $a^{[n]}$ demand (measured in kWh) that arrives in period j.
g_j	the local electricity generation (in kWh) in period j.
$p_{j+1} : \mathbb{R} \to \mathbb{R}$	period $j+1$'s price function.
$I_j := (\mathbf{q}_j, g_j, p_{j+1})$	period j's set of random information.
$\mathcal{I}_t = \{I_j\}_{j=t}^{t+T-1}$	the set of random information from period t to period $t+T-1$.
State Variables	
$\mathbf{d}_t$	the vector of outstanding demand at the beginning of period t. $\mathbf{d}_t = (\mathbf{d}_{1t}^\intercal, \mathbf{d}_{2t}^\intercal, \ldots, \mathbf{d}_{\lvert\mathcal{A}\rvert t}^\intercal)^\intercal$, where $\mathbf{d}_{at} = (d_{a^{[1]}t}, d_{a^{[2]}t}, \ldots, d_{a^{[N]}t})^\intercal$. $d_{a^{[n]}t}$ is the type $a^{[n]}$ outstanding demand (measured in kWh) at the beginning of period t.
b_t	the amount of electricity stored in the battery at the beginning of period t (assuming single battery for simplicity).
$R_t = (\mathbf{d}_t, b_t)$	the status of outstanding demand and local storage at the beginning of period t.
$\mathcal{H}_t := \left(\{\widehat{I}_j\}_{j=-\infty}^{t-1}, \{X_j\}_{j=-\infty}^{t-1}\right)$	the history information at the beginning of period t.
$S_t = (R_t, \mathcal{H}_t)$	the state status of the system at the beginning of period t.
Decision Variables	
$\mathbf{w}_t$	the vector of decisions on meeting outstanding demand in period t. $\mathbf{w}_t = (\mathbf{w}_{1t}^\intercal, \mathbf{w}_{2t}^\intercal, \ldots, \mathbf{w}_{\lvert\mathcal{A}\rvert t}^\intercal)^\intercal$, where $\mathbf{w}_{at} = (w_{a^{[1]}t}, w_{a^{[2]}t}, \ldots, w_{a^{[N]}t})^\intercal$. $w_{a^{[n]}t}$ denotes whether to meet the type $a^{[n]}$ demand in period t.
y_t	the amount of electricity to be extracted from the battery (negative if charging) in period t.
$X_t := (\mathbf{w}_t, y_t)$	the set of decision variables in period t.

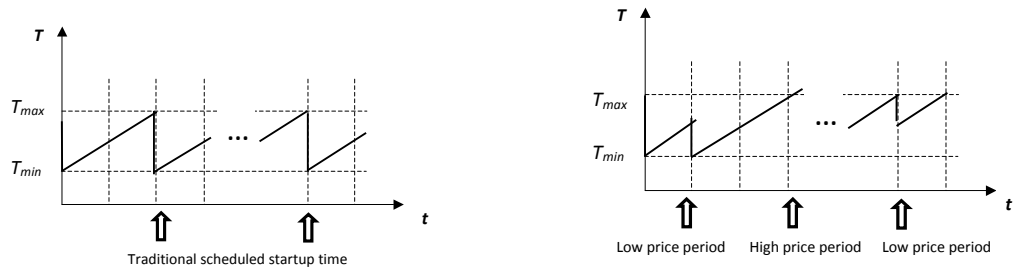
Table 3.1: Summary of Notation

We assume that an Energy Management System (EMS) controls energy usage on the household level, including smart appliances and local (distributed) energy generation as well as storage devices. Figure 3.1 describes how the EMS works. It is assumed that the EMS maintains a database of historical data, and it has access to web servers, which provide forecasts on weather conditions and electricity prices. The EMS monitors the status quo of the system, then makes decisions on which demand to satisfy immediately and which to defer. It also manages the charging and discharging of the local storage (such as a battery). It is worth mentioning that, besides delaying flexible demand, the EMS can meet some demand in advance to save cost. Take the control for refrigerators as an example. Figure

Notation	Definition
Coefficients:	
$\pi_{a^{[n]}}$:	the discomfort per kWh from deferring type $a^{[n]}$ demand.
$\pi'_{a^{[n]}}$:	the discomfort per kWh from failing to add type $a^{[n]}$ demand into the waiting queue.
$\pi''_{a^{[n]}}$:	the discomfort per kWh from unsatisfied type $a^{[n]}$ demand after its deadline.
β:	factor that converts discomfort into dollar value.
Balance Equations:	
$R_t^X = f(R_t, X_t)$	function $f : \mathbb{R}^{M+1} \times \mathbb{R}^{M+2} \to \mathbb{R}^{M+1}$ maps the state status of demand and storage R_t to the post-decision state R_t^X, where $M = \lvert\mathcal{A}\rvert \times N$.
$R_{t+1} = h(R_t^X, \widehat{\mathbf{q}}_t, \widehat{g}_t)$	function $h : \mathbb{R}^{M+1} \times \mathbb{R}^M \times \mathbb{R} \to \mathbb{R}^{M+1}$ maps the post-decision state R_t^X to the state status of demand and storage of the next period, R_{t+1}, where $M = \lvert\mathcal{A}\rvert \times N$.
One Period Costs:	
$C_t(R_t, X_t)$:	the amount of money paid to the grid for electricity.
$L_t(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t)$:	the discomfort from lost arrivals.
$\beta U_t(R_t, \mathcal{H}_t, X_t)$:	the dollar value of the total discomfort including $L_t(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t)$ and other terms.
Value Terms:	
$J_t^*(R_t, \mathcal{H}_t)$:	the optimal total cost, starting from state $(R_t, \mathcal{H}_t)$.
$V_{t+1}^*(R_t, \mathcal{H}_t, X_t)$:	the optimal expected value-to-go, starting from $(R_t, \mathcal{H}_t)$ and with decision X_t.
Approximations:	
$\overline{L}_t(\mathbf{d}_t^X; \mathbf{\Phi}_t^m, \zeta_t^m)$:	the approximation of $L_t(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t)$ based on coefficients $\mathbf{\Phi}_t^m = (\phi_{11t}^m, \phi_{12t}^m, \ldots, \phi_{1Nt}^m, \ldots, \phi_{\lvert\mathcal{A}\rvert Nt}^m)$, and ζ_t^m.
$\overline{V}_{t+1}(R_t^X; \mathbf{\Theta}_t^m, \psi_t^m, \eta_t^m)$:	the approximation of $V_{t+1}^*(R_t, \mathcal{H}_t, X_t)$ based on coefficients $\mathbf{\Theta}_t^m = (\theta_{11t}^m, \theta_{12t}^m, \ldots, \theta_{1Nt}^m, \ldots, \theta_{\lvert\mathcal{A}\rvert,N,t}^m)$, ψ_t^m, and η_t^m.

Table 3.2: Summary of Value Terms

3.2(a) describes that with traditional control, the refrigerator starts cooling every time the temperature reaches a pre-set threshold T_{max} regardless the price. While with an EMS, as shown in Figure 3.2(b), decisions on cooling are made according to time-varying prices. The detailed model of shifting demand forward requires the modeling techniques that will be presented in later sections, hence are delayed to Appendix A together with some insights.



(a) System scheduled startups (b) Controlled startups

Figure 3.2: Shifting Demand for Refrigerator Forward in Time

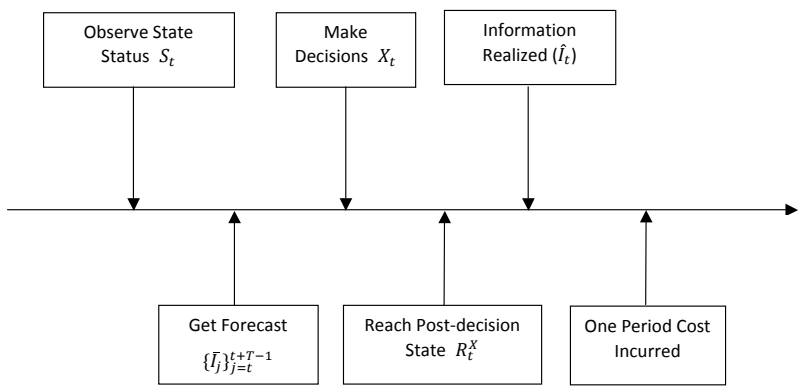


Figure 3.3: Sequence of Events in Period t

The idea of optimal control here is to save cost by shifting demand from high-price periods to low-price ones, at the cost of increased discomfort (or inconvenience). However, distributed energy generation, storage, and complex price structures complicate the trade-off between comfort and cost savings.

Figure 3.3 describes the sequence of events. We first discretize time into periods, and assume that the whole system works as follows. At the beginning of each period t, the EMS observes state status S_t, which is characterized by the outstanding demand and the battery level (denoted as R_t), and the information of history (denoted as $\mathcal{H}_t$). Then it obtains forecasts on the random information set $\bar{\mathcal{I}}_t$, including prices, local generation, and the arrival of demand[1]. The EMS makes decision X_t by solving a finite-horizon stochastic optimization problem, and the system evolves to the post-decision state R_t^X, defined as the following:

[1]Some or all of these random information processes depend on uncertain weather conditions

Definition 2. *The imaginary post-decision state R_t^X is the state of the system right after making the decision X_t, before the random information I_t is realized.*

After period t's random information is realized, one-period cost is incurred and the system moves on to the next period. The following sections present the details of the model.

Characterization of Demand, History and State Status

In general, demand can be categorized into (time) flexible demand and (time) inflexible demand, where the former can be shifted across periods, while the latter needs to be satisfied immediately. For instance, the demand for dishwasher is flexible, while the demand for lighting is inflexible. Since inflexible demand is uncontrollable in terms of load shifting, the EMS manages only flexible demand. We further categorize flexible demand into two groups: *additive* and *non-additive*. Additive demand is usually continuous and stackable. For example, demand on Air Conditioning (A.C.), refrigerator, and water heater is additive. On the contrary, non-additive demand is usually non-stackable and has to be satisfied separately, for instance, two loads of laundry require two separate runs of the wash machine.

When a demand is submitted, its allowable delay is specified at the same time by the user. Once the demand enters the waiting queue, it becomes an outstanding demand, characterized by its *task* and *allowable delay* - the time remaining before its deadline. The task associated with a demand is the assignment to be performed in order to satisfy that demand. Let $\mathcal{A} := \{a : a = 1, 2, \ldots, |\mathcal{A}|\}$ be the set of all tasks, and $\mathcal{A}_k \subset \mathcal{A}$ be the set of tasks associated with appliance $k \in \mathcal{K}$, where $\mathcal{K}$ is the set of all appliances. The allowable delay n ($n = 1, 2, \ldots, N$) of a demand is the number of periods between the current period t and its deadline[2]. We refer to the demand of task a that is due in n periods as a demand of *type* $a^{[n]}$. Demand is measured in kWh, for that cost only depends on the amount of electricity consumed. Let $d_{a^{[n]}t}$ represent the amount of electricity required by an outstanding demand of type $a^{[n]}$ at the beginning of period t. We further assume that the amount of electricity required by each demand is fixed and known[3]. For notational convenience, define $\mathbf{d}_{at}$ and $\mathbf{d}_t$ as follows:

$$\mathbf{d}_{at} = \begin{pmatrix} d_{a^{[1]}t} \\ d_{a^{[2]}t} \\ \vdots \\ d_{a^{[n]}t} \end{pmatrix}, \quad \mathbf{d}_t = \begin{pmatrix} \mathbf{d}_{1t} \\ \mathbf{d}_{2t} \\ \vdots \\ \mathbf{d}_{|\mathcal{A}|t} \end{pmatrix}$$

where $\mathbf{d}_{at}$ is the vector outstanding demand of task a, and $\mathbf{d}_t$ is the vector of all types of outstanding demand at the beginning of period t.

[2]It is assumed that all demand tasks have the same maximum allowable delay N

[3]This is achievable in application. For example, the wash machine may have options such as: "super cycle", "white and colors" and "by hand". The amount of energy required for each task can be estimated by multiplying the time and the power required to finish that task.

Random Information, Forecasts and Their Realizations

As introduced in section 3.2, the system obtains forecasts on future random information $\mathcal{I}_t := \left\{\bar{I}_j\right\}_{j=t}^{t+T-1}$ before making decisions in period t. In our model, $I_j := (\mathbf{q}_j, g_j, p_{j+1})$ denotes the random information associated with period j, where $\mathbf{q}_j$ is the vector of demand arrivals during period j; g_j represents the amount of local electricity generation in period j, and $p_{j+1} : \mathbb{R} \to \mathbb{R}$ is the price structure in period $j+1$. Recall that $\bar{I}_j = (\bar{\mathbf{q}}_j, \bar{g}_j, \bar{p}_{j+1})$ and $\widehat{I}_j = (\widehat{\mathbf{q}}_j, \widehat{g}_j, \widehat{p}_{j+1})$ are the forecast and realization of I_j.

There is hardly a common model for the random information processes. On the other hand, accurate ad hoc models are also hard to obtain. We assume in this chapter that the EMS is equipped with another module that generates enough sample paths of $\mathcal{I}_t$. Specifically, the module generates samples from both the forecasts obtained from web servers and the information of history recorded locally, where the information of history at the beginning of period t is represented by $\mathcal{H}_t$. $\mathcal{H}_t$ consists of the realization of all past random information at time t, $\left\{\widehat{I}_j\right\}_{j=-\infty}^{t-1}$, and past decisions $\left\{X_j\right\}_{j=-\infty}^{t-1}$. Since the discussion of sample generation mechanisms is beyond the scope of this chapter, we simply assume the existence of the module without going into details.

In terms of price structures, the most straightforward one is *flat-rate*, that is, when the price is independent of the amount of electricity consumed. However, the price structures under RTP can be complex. For instance, PG&E currently offers increasing block-rate with which the price is an increasing stepwise function of the amount of electricity consumed. Similarly, time varying block-rate may be used to mitigate the rebound effects studied by [16], by keeping users from using electricity aggressively in one period. Since the ultimate objective is to shift flexible demand, it suffices to price the flexible demand (and energy used to fill the storage). We also assume that users are price takers, that is, the decision of one end-user does not affect the electricity prices or price structures.

In a certain period t, $\widehat{\mathbf{q}}_t$, $\widehat{g}_t$, and the price structure $\widehat{p}_{t+1}$ are obtained after decision X_t is executed. Note that the price structure p_{t+1} is known to the user at the beginning of period $t+1$. We have this assumption because it is unreasonable to let users make energy purchase decisions before they know the price. At last, one-period cost is incurred, and the next period starts.

We assume additionally that R_t and $\mathcal{I}_t$ are conditionally independent given the information of history $\mathcal{H}_t$ and the decision X_t. Obviously for the underlying dynamic system, the optimal decision X_t^* depends on both R_t and $\mathcal{H}_t$. Therefore, by defining the state status S_t at the beginning of period t as $S_t := (R_t, \mathcal{H}_t)$, the system possesses Markovian property.

Decisions and State Transitions

Let $\mathbf{w}_t$ be the vector of decisions on meeting outstanding demand, and y_t the amount of electricity to be extracted from the battery. Then the decision to make in period t can be denoted as $X_t := (\mathbf{w}_t, y_t)$. The decisions have to satisfy the following constraints.

$$\mathbf{d}_t^\intercal \mathbf{w}_t - y_t - z_t = 0 \tag{3.1a}$$
$$y_t \leq b_t \tag{3.1b}$$
$$y_t \leq u_b \tag{3.1c}$$
$$-y_t \leq l_b \tag{3.1d}$$
$$b_t - y_t \leq c_b \tag{3.1e}$$
$$w_{a^{[n]}t} \in \{0, 1\} \qquad \forall a \in \mathcal{A},\ n \in \{1, 2, \ldots, N\} \tag{3.1f}$$

where z_t denotes the amount of energy to be purchased from the gird, and constraint (3.1a) is the energy balance constraint. (3.1b) requires that the amount of energy extracted from the battery must be less than or equal to the available energy in the battery. (3.1c) to (3.1e) are the charging rate, discharging rate, and capacity limits of the battery. At last, we assume that all types of demand must be satisfied in full, as in constraint (3.1f).

Recall that the system first reaches the imaginary state R_t^X after executing X_t. Let $f^d : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^M$ ($M = N \times |\mathcal{A}|$) be the function that maps $(\mathbf{d}_t, \mathbf{w}_t)$ to $\mathbf{d}_t^X$ in the following way:

$$\mathbf{d}_t^X = f^d(\mathbf{d}_t, \mathbf{w}_t) = \Pi^d(\mathbf{w}_t)\mathbf{d}_t \tag{3.2}$$

where $\Pi^d(X_t)$ is a transformation matrix with its (j, k)-th entry being:

$$\left(\Pi^d(\mathbf{w}_t)\right)_{jk} = \begin{cases} 1 - w_{amt} & \text{if } j = k = m + (a-1)N \\ 0 & \text{otherwise} \end{cases}$$

Note if there is no capacity constraint (c_a for all $a \in \mathcal{A}$, measured in kWh) for the waiting queues of tasks, then R_{t+1} is linear in the post-decision state R_t^X and new arrivals $\widehat{\mathbf{q}}_t$. In case we have capacity limits, R_{t+1} is non-decreasing concave in R_t^X and $\widehat{\mathbf{q}}_t$. Thus, let function $h^d : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^M$ be the state transition function from the post-decision state to the outstanding demand status at the beginning of the next period as follows:

$$\mathbf{d}_{t+1} = h^d\left(\mathbf{d}_t^X, \widehat{\mathbf{q}}_t\right) = \min\left\{\Sigma^d \mathbf{d}_t^X + \widehat{\mathbf{q}}_t, \mathbf{c}\right\} \tag{3.3}$$

where $\mathbf{c} = (c_1\mathbf{e}^\intercal, c_2\mathbf{e}^\intercal, \ldots, c_{|\mathcal{A}|}\mathbf{e}^\intercal)^\intercal$, $\mathbf{e}$ is a vector in $\mathbb{R}^N$ with all entries equal to 1, and Σ^d is a transformation matrix with its (j, k)-th entry being:

$$\left(\Sigma^d\right)_{jk} = \begin{cases} 1 & \text{if } j = k - 1 \text{ and } \mod(j, N) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Motivated by the fact in most applications that the excess energy from the solar PV panels is either stored in battery or fed into the grid, we further assume that locally produced electricity is first stored in the battery, because when making energy allocation decisions, the output from local generation is unknown, hence its usage is delayed. Then, the state transition of the battery is:

$$b_t^X = f^b(b_t, y_t) = b_t - y_t \tag{3.4}$$

$$b_{t+1} = h^b(b_t^X, \widehat{g}_t) = \min\left(b_t^X + \min\left(l_b, \widehat{g}_t - y_t\right), c_b\right) \tag{3.5}$$

where (3.4) is the transition to the post-decision state and (3.5) is the transition to next period's initial state. Similarly as (3.3), equation (3.5) comes from the fact that there is charging rate limit and capacity limit on the battery. For future convenience, we define $R_t^X = f(R_t, X_t)$ and $R_{t+1} = h(R_t^X, \widehat{\mathbf{q}}_t, \widehat{g}_t)$.

The Model of Additive Demand and Precedence Constraints

Note that although we assume that demand must be satisfied in full, it is straightforward to allow additive demand to have continuous state status and correspondingly, continuous decision variables. For instance, decisions on meeting demand on A.C. are supposed to be continuous as the room temperature is continuous itself. However, since our problem is reducible from knapsack problems as we will show later in this chapter, and knapsack problems with mixed integer decision variables are in general very hard to solve, we model the problem in such a way that all decision variables are binary.

In particular, we model additive demand by the following steps: (1) aggregate the continuous state status of additive demand into a finite set of discrete values; (2) assign a task for each of the aggregated state status; (3) make the decision variables for these new demand types binary; and (4) add precedence constraints.

Take the demand on A.C. for example, we first aggregate the continuous room temperature into integers, then define $\mathcal{A}_c := \{a : a = i(c), i(c)+1, \ldots, i(c)+|\mathcal{A}_c|-1\}$ as the set of tasks, representing how to cool the building, where $i : \mathcal{K} \to \mathcal{A}$ maps the index of each appliance to the smallest index among all demand tasks associated with it.[4] Specifically, if the user's preferred room temperature is $73°F$, then define task $(i(c)+j)$ as the demand of cooling the room temperature from $(73+j+1)°F$ to $(73+j)°F$. If the current room temperature is $75°F$, $2°F$ higher than the preferred one, (that is, the demand for the A.C. is immediate), then in addition to $i(c)^{[1]}$ (which represents lowering the room temperature from $74°F$ to $73°F$), another type of demand $(i(c)+1)^{[1]}$ representing "lowering the room temperature from $75°F$ to $74°F$" exists. Moreover, demand $(i(c)+1)^{[1]}$ must be met in advance. Thus, the following set of precedence constraints are needed for all types of additive

[4] We assume implicitly here that the tasks associated with one appliance are consecutively indexed.

demand:

$$w_{(i(s)+j)^{[n]}t} \leq w_{(i(s)+j+1)^{[n]}t} \qquad \forall j = 0, 1, \ldots, |\mathcal{A}_s| - 1, \quad \forall s \in \mathcal{S} \quad 1 \leq n \leq N \tag{3.6}$$

where $s \in \mathcal{S}$ is any appliance on which demand is additive, and $\mathcal{S}$ is the set of all such appliances. By this approach, the continuous decisions and continuous state status of additive demand are discretized, and the complexity of the problem is reduced at the cost of suboptimality (from discretization).

Objectives and Dynamic Programming Formulation

The EMS aims to optimally trade off comfort for cost saving. In particular, the EMS shifts demand to save cost. However, shifting demand degrades service level and causes discomfort. We set the objective of our problem as minimizing total disutility, and assume that user's total disutility is separable into (1) the cost of electricity, and (2) the discomfort resulting from load shifting. Without loss of generality, we convert discomfort into dollar values by coefficient β. Then, the unit for the total disutility is also dollar, and the one-period cost is the cost for electricity plus the dollar-value of discomfort. Due to the seasonality and the periodic nature of the whole system, the EMS minimizes the expected total disutility over T periods. Since the length of each period discussed in the current context are short, we assume no discount between different periods. In summary, the EMS solves the following stochastic program:

$$\begin{aligned} (\mathbf{P}) \quad \min_{X_1, X_2, \ldots, X_T} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \left(C_t\left(R_t, X_t\right) + \beta U_t(R_t, \mathcal{H}_t, X_t)\right)\right] \\ s.t. \quad & (3.1a) - (3.1f) \\ & (3.2) - (3.5) \\ & (3.6) \end{aligned} \tag{3.7}$$

where $C_t(R_t, X_t)$ is the cost of electricity and $\beta U_t(R_t, \mathcal{H}_t, X_t)$ is the dollar-value of discomfort. Note that $C_t(R_t, X_t)$ is allowed to be negative, indicating that, instead of buying, the EMS sells electricity back to the grid. Specifically,

$$\begin{aligned} C_t(R_t, X_t) =& z_t \cdot p_t\left(z_t\right) \\ =& \left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right) \cdot p_t(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t) \end{aligned} \tag{3.8}$$

In addition, we use three types of discomfort to capture the user's sensitivity on service level. One comes from deferring demand, and the other two from losing demand. Lost demand happens in two scenarios: (1) new arrivals fail to enter the waiting queue due to capacity limit, and (2) demands fail to be satisfied by the deadline. Assuming all types of discomfort are proportional to the level of demand, let $\pi_{a^{[n]}}$, $\pi'_{a^{[n]}}$, and $\pi''_{a^{[1]}}$ be the discomfort

per kWh associated with deferred demand, lost arrival, and unsatisfied demand, respectively. Therefore, the expected total discomfort during period t is:

$$\begin{aligned}
U_t(R_t, \mathcal{H}_t, X_t) =& \sum_{a\in\mathcal{A}} \sum_{1\le n\le N} \pi_{a^{[n]}} d^X_{a^{[n]}t} + \sum_{a\in\mathcal{A}} \pi''_{a^{[1]}} d^X_{a^{[1]}t} \\
&+ \mathbb{E}\left[\sum_{a\in\mathcal{A}} \sum_{1\le n\le N} \pi'_{a^{[n]}} \max\left\{0, d^X_{a^{[n]}t} + q_{a^{[n]}t} - c_a\right\} \middle| R_t, \mathcal{H}_t, X_t\right] \\
=& \sum_{a\in\mathcal{A}} \sum_{1\le n\le N} \pi_{a^{[n]}} d^X_{a^{[n]}t} + \sum_{a\in\mathcal{A}} \pi''_{a^{[1]}} d^X_{a^{[1]}t} + L_t\left(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t\right) \\
=& \sum_{a\in\mathcal{A}} \sum_{1\le n\le N} \widetilde{\pi}_{a^{[n]}} d^X_{a^{[n]}t} + L_t\left(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t\right) \qquad (3.9)
\end{aligned}$$

where $L_t\left(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t\right)$ is non-decreasing in $\mathbf{d}_t$ and non-increasing in $\mathbf{w}_t$. This term measures the discomfort from lost arrivals. To simplify the notation, define $\widetilde{\pi}_{a^{[n]}} := \pi_{a^{[n]}} + \pi''_{a^{[1]}}$. The distributions of the arrival processes of new demand are not known a priori, thus the expectation is hard to calculate. Therefore, we will approximate this term by a linear function of post-decision states.

$$\overline{L}_t\left(\mathbf{d}^X_t; \mathbf{\Phi}^m_t, \zeta^m_t\right) = (\mathbf{\Phi}^m_t)^\intercal \mathbf{d}^X_t + \zeta^m_t \qquad (3.10)$$

It is not optimal to simply satisfy demand at times exactly when prices are the lowest, because of the following: (1) the existence of energy storage device enables better energy saving opportunities, for instance, the battery could be full during high-price periods; (2) prices may deviate from the forecasts; (3) complex price structures, such as the increasing block-rate for each period makes satisfying all outstanding demand in one period inefficient; and (4) there is discomfort from both deferring demand and lost demand. It is straightforward to use Dynamic Programming to solve the original problem ($\mathbf{P}$), and the corresponding Bellman's equation for the EMS is formulated as follows:

$$\begin{aligned}
(\mathbf{P}^*) \qquad J^*_t(R_t, \mathcal{H}_t) =& \min_{X_t} \Big\{ C_t\left(R_t, X_t\right) + \beta U_t\left(R_t, \mathcal{H}_t, X_t\right) \\
&\qquad + \mathbb{E}_{\mathcal{I}_t}\left[J^*_{t+1}\left(R_{t+1}, \mathcal{H}_{t+1}\right) | R_t, \mathcal{H}_t, X_t\right] \Big\} \\
=& \min_{X_t} \Big\{ C_t\left(R_t, X_t\right) + \beta U_t\left(R_t, \mathcal{H}_t, X_t\right) + V^*_{t+1}\left(R_t, \mathcal{H}_t, X_t\right) \Big\} \qquad (3.11) \\
s.t. \quad & (3.1a) - (3.1f) \\
& (3.2) - (3.5) \\
& (3.6)
\end{aligned}$$

where $V^*_{t+1}\left(R_t, \mathcal{H}_t, X_t\right) = \mathbb{E}\left[J^*_{t+1}\left(R_{t+1}, \mathcal{H}_{t+1}\right) | R_t, \mathcal{H}_t, X_t\right]$ is the expected optimal value-to-go starting from $(R_t, \mathcal{H}_t)$ and following the decision X_t. The boundary condition is given

by: $J^*_{T+1}(R_{T+1}, \mathcal{H}_{T+1}) \equiv 0$. Recall that we choose a finite T, because: (1) there exists a discount factor that lowers the relevance of the periods far from now, (2) what happens at times more than T periods from now does not have enough influence on current decision, and (3) long-term forecast tends to be more inaccurate.

3.3 Approximate Dynamic Programming Approach

One of the main obstacles of solving problem ($\mathbf{P}^*$) is the calculation of value-to-go functions. First of all, it is hard to infer the probability distributions for future electricity prices, demand arrivals, and local generation. Due to the uniqueness of each household, it is necessary to have ad hoc models for the distributions of many of these random processes. Even if the distributions are known, the calculation of expectations may be computationally intractable. Moreover, the random information correlates both with each other and across time. For example, the demand for the A.C. correlates with local generation because both of them depend on weather conditions, and they also correlate across time, because future weather depends on past weather.

Even though the distributions of the random processes are known, calculating the value-to-go is still inarguably computationally intractable for high dimensional systems because of the "curses of dimensionality". In particular, the computational complexity for solving the DP grows exponentially with respect to the dimension of the state space, which consists of the whole information of history and the status quo of the whole system. Therefore, we look into the approximation to the original problem ($\mathbf{P}^*$).

Properties of ($\mathbf{P}^*$) and Value-to-go Function Approximation

We first notice that $J^*_t(R_t, \mathcal{H}_t)$ and $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ possess the following properties:

Proposition 9. *The optimal value function $J^*_t(R_t, \mathcal{H}_t)$ and expected optimal value-to-go $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ has the following properties (for all $a \in \mathcal{A}$ and $n \in \{1, 2, \ldots, N\}$):*

a. The optimal value function $J^*_t(R_t, \mathcal{H}_t)$ is non-decreasing in $d_{a^{[n]}t}$;

b. The expected optimal value-to-go $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ is non-decreasing in $d_{a^{[n]}t}$;

c. The optimal value function $J^*_t(R_t, \mathcal{H}_t)$ is non-increasing in b_t;

d. The expected optimal value-to-go $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ is non-increasing in b_t.

In the remainder of this section, we will present an *approximate dynamic programming* (ADP) approach for ($\mathbf{P}^*$) inspired by the properties described in Proposition 9. In short, we

solve $(\mathbf{P}^*)$ based on approximations of value-to-go functions. The approximations are obtained by learning via sample paths of the random processes. Specifically, we iterate through the following steps to update our approximation: (1) at time t, solve for the best decision X_t given the current approximate value-to-go functions; (2) following X_t, proceed forward in time to $t+1$, and repeat the same procedure as in step (1). The corresponding objective function value of $(\mathbf{P}^*)$ at time $t+1$ is then used as an approximation to $J^*_{t+1}(R_{t+1}, \mathcal{H}_{t+1})$; (3) update the prior belief of $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ based on $J^*_{t+1}(R_{t+1}, \mathcal{H}_{t+1})$ obtained in step (2).

This ADP approach differs from the backward induction approach in that it travels forward in time, and reduces computational time substantially by not visiting all states. However, ADP approaches usually cannot guarantee convergence to optimality ([53]). Nevertheless, the ADP approach used in this chapter has the following advantages. Firstly, it is model free in the sense that it does not require the distributions of random processes known a priori in order to solve $(\mathbf{P}^*)$ and to learn the value-to-go functions. Secondly, it makes no assumption on the parametric family of the distributions of the random processes. It does not even require the whole information of history, although as discussed later in this chapter, more information of history tends to improve the performance.[5] Lastly, the complexity of ADP approach mainly depends on the form of the stochastic optimization problem itself and the step-size rule being used, hence, grows much slower than that of the backward induction approach.

Recall that we assume that there will be another module that generates the sample paths used to update the approximation, and we focus on solving the approximate problem by using these sample paths but not on how to generate them. In addition, unless otherwise noted, we neglect $\mathcal{H}_t$ in the terms associated with the ADP approach for notational convenience, though we may choose sample paths according to the information of history.

In practice, it is likely that available data is limited compared to the complexity of the problem. For example, since the problem possesses seasonality, separate data sets are required for different times of year. As a result, we choose R^X_t to be the basis vector and use a linear function of R^X_t to approximate the value-to-go function, because there is only one coefficient to be estimated for each basis. Based on the above discussion, we define the approximate value-to-go function to be as follows:

$$\begin{aligned} V^*_{t+1}(R_t, \mathcal{H}_t, X_t) =& \mathbb{E}_{\mathcal{I}_t}\left[J^*_{t+1}(R_{t+1}, \mathcal{H}_{t+1}) \,|R_t, \mathcal{H}_t, X_t\right] \\ \approx& \overline{V}_{t+1}\left(R^X_t; \mathbf{\Theta}^m_t, \psi^m_t, \eta^m_t\right) \\ =& (\mathbf{\Theta}^m_t)^\intercal \mathbf{d}^X_t + \psi^m_t b^X_t + \eta^m_t \end{aligned} \tag{3.12}$$

where m is the iteration number. Let

$$\bar{\Gamma}_{t+1}(R^X_t; \widetilde{\mathbf{\Theta}}^m_t, \psi^m_t, \widetilde{\eta}^m_t) = \overline{L}_t(\mathbf{d}_t; \mathbf{\Phi}^m_t, \zeta^m_t) + \overline{V}_{t+1}\left(R^X_t; \mathbf{\Theta}^m_t, \psi^m_t, \eta^m_t\right)$$

[5] We conjecture that the more historical data which we use to learn, the more accurate the approximation of the value-to-go function will be.

Since both $\overline{V}_{t+1}\left(R_t^X;\mathbf{\Theta}_t^m,\psi_t^m,\eta_t^m\right)$ and $\overline{L}_t\left(\mathbf{d}_t;\mathbf{\Phi}_t^m,\zeta_t^m\right)$ in equation (3.10) are parameterized as linear functions of the post-decision state, $\bar{\Gamma}_{t+1}$ is also linear in R_t^X, with coefficients being $\tilde{\theta}_{i,n,t}^m=\theta_{i,n,t}^m+\beta\phi_{i,n,t}^m$ and $\tilde{\eta}_t^m=\eta_t^m+\beta\zeta_t^m$. Using the above approximations, the Bellman equation $(\mathbf{P}^*)$ can be approximated by the following problem $(\widetilde{\mathbf{P}})$:

$$
\begin{aligned}
(\widetilde{\mathbf{P}}) \qquad & \min_{X_t} \quad C_t(R_t,X_t)+\beta\sum_{a\in\mathcal{A}}\sum_{1\le n\le N}\tilde{\pi}_{a^{[n]}}d_{a^{[n]}t}^X+\left[\overline{\Gamma}_{t+1}(R_t^X;\widetilde{\mathbf{\Theta}}_t^m,\psi_t^m,\widetilde{\eta}_t^m)\right] && (3.13)\\
&=\min_{X_t} \quad C_t(R_t,X_t)+\beta\sum_{a\in\mathcal{A}}\sum_{1\le n\le N}\tilde{\pi}_{a^{[n]}}d_{a^{[n]}t}^X+\left[(\widetilde{\mathbf{\Theta}}_t^m)^\intercal\mathbf{d}_t^X+\psi_t^m b_t^X+\widetilde{\eta}_t^m\right] && (3.14)\\
& s.t. \quad (3.1a)-(3.1f)\\
& \qquad\ \ (3.2)-(3.5)\\
& \qquad\ \ (3.6)
\end{aligned}
$$

Following the approximation approach described above, we can get rid of the intractable calculation of the value-to-go functions. In addition, the non-linear state transitions are excluded from the model. This approximation strategy not only removes the dependence of the calculation on the knowledge of the distributions of the random processes, but also simplifies the resulting one-period optimization problem.

An Efficient Algorithm that Solves the One-period Problem

Even with the approximate value-to-go functions and the approximate discomfort terms (associated with lost arrivals), the one-period problem $(\widetilde{\mathbf{P}})$ is still hard to solve as it is a non-linear mixed integer program. Note that although the problem $(\widetilde{\mathbf{P}})$ is not separable in $\mathbf{w}_t$ and y_t, we can reformulate it in a lower-dimension decision space through the following steps: (1) for each feasible $\mathbf{w}_t$, minimize the objective over $y_t\in\mathcal{Y}_t$, where $\mathcal{Y}_t$ denotes the feasible region defined by constraints (3.1b) to (3.1e); (2) obtain a cost function $C_t^m:\mathbb{R}\to\mathbb{R}$, which only depends on the amount of electricity consumed by appliances, $\mathbf{d}_t^\intercal\mathbf{w}_t$; and (3) reformulate the problem to minimize the objective over $\mathbf{w}_t$. Recall that purchasing electricity from the grid costs $p_t\left(\mathbf{d}_t^\intercal\mathbf{w}_t-y_t\right)$, where $p_t:\mathbb{R}\to\mathbb{R}$ is the function that represents the price structure in period t. We need to solve the following problem for optimal $y_t^*(\mathbf{w}_t)$:

$$
\begin{aligned}
C_t^m\left(\mathbf{d}_t^\intercal\mathbf{w}_t\right) := \min_{y_t} \quad & \left(\mathbf{d}_t^\intercal\mathbf{w}_t-y_t\right)p_t\left(\mathbf{d}_t^\intercal\mathbf{w}_t-y_t\right)+\left(\psi_t^m\left(b_t-y_t\right)\right)\\
=\min_{y_t} \quad & \left(\mathbf{d}_t^\intercal\mathbf{w}_t-y_t\right)p_t\left(\mathbf{d}_t^\intercal\mathbf{w}_t-y_t\right)-\psi_t^m y_t+\tau && (3.15)\\
s.t. \quad & (3.1b)-(3.1e)
\end{aligned}
$$

With reasonable assumptions on the price function $p_t:\mathbb{R}\to\mathbb{R}$, we can show the following useful proposition:

Proposition 10. *$C_t^m\left(\mathbf{d}_t^\intercal\mathbf{w}_t\right)$ is convex increasing in $\mathbf{d}_t^\intercal\mathbf{w}_t$, when the price function $p_t:\mathbb{R}\to\mathbb{R}$ is either:*

(a) a non-decreasing stepwise function, that is given thresholds $\left\{b_t^1, b_t^2, \ldots, b_t^l\right\}$*, and the corresponding prices* $\left\{p_t^1, p_t^2, \ldots, p_t^l\right\}$ *with* $p_t^j \leq p_t^k$ *for all* $j \leq k$*,* $p_t\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right) = p_t^j$ *if and only if* $b_t^j \leq \mathbf{d}_t^\intercal \mathbf{w}_t - y_t < b_t^{j+1}$*; or*

(b) a twice-differentiable convex increasing function, and the second derivative of p_t *equals to zero for all* $x \leq 0$*.*

Moreover, in case (a), suppose $\exists j$ *such that* $p_t^{j-1} \leq -\psi_t^m < p_t^j$ *then* $y_t^*(\mathbf{w}_t)$*, which minimizes equation (3.15), can be obtained by solving:*

$$y_t^*(\mathbf{w}_t) = \underset{y}{\operatorname{argmin}} \left| \left(\mathbf{d}_t^\intercal \mathbf{w}_t - y\right) - b_t^j \right|$$

Similarly in case (b), $y_t^*(\mathbf{w}_t)$ *takes the following form:*

$$y_t^*(\mathbf{w}_t) = \begin{cases} \min\limits_{y \in \mathcal{Y}_t} y & \text{if } \min\limits_{y \in \mathcal{Y}_t} y > y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \\ y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) & \text{if } \min\limits_{y \in \mathcal{Y}_t} y \leq y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \leq \max\limits_{y \in \mathcal{Y}_t} y \\ \max\limits_{y \in \mathcal{Y}_t} y & \text{if } \max\limits_{y \in \mathcal{Y}_t} y < y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \end{cases} \tag{3.16}$$

where, $y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right)$ *is the solution to equation* $\left(y - \mathbf{d}_t^\intercal \mathbf{w}_t\right) p_t^{'}\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y\right) - p_t\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y\right) = \psi_t^m$*.*

Then, the one-period problem that solves for $\mathbf{w}_t$ is formulated as follows:

$$\begin{aligned} (\mathbf{P}') \qquad \min_{\mathbf{w}_t} \quad & \sum_{a \in \mathcal{A}} \sum_{1 \leq n \leq N} \beta \tilde{\pi}_{a^{[n]}} d^X_{a^{[n]}t} \\ & + \sum_{a \in \mathcal{A}} \sum_{1 \leq n \leq N} \tilde{\theta}^m_{a^{[n]}t} d^X_{a^{[n]}t} + C_t^m\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \\ = -\max_{\mathbf{w}_t} \quad & \sum_{a \in \mathcal{A}} \sum_{1 \leq n \leq N} \left(\beta \tilde{\pi}_{a^{[n]}} + \tilde{\theta}^m_{a^{[n]}t}\right) w_{a^{[n]}t} - C_t^m\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) + \varphi \\ s.t. \quad & (3.6) \\ & (3.1f) \end{aligned}$$

We can drop the constant term φ in the following discussion. Apparently, we can further reduce the dimension of decision variables by dropping items with zero outstanding demand.

Without loss of generality, re-define vectors v and u as follows:

$$v^m = \begin{pmatrix} \beta\tilde{\pi}_{11} + \tilde{\theta}^m_{11t} \\ \vdots \\ \beta\tilde{\pi}_{1N} + \tilde{\theta}^m_{1Nt} \\ \beta\tilde{\pi}_{21} + \tilde{\theta}^m_{21t} \\ \vdots \\ \beta\tilde{\pi}_{|\mathcal{A}|1} + \tilde{\theta}^m_{|\mathcal{A}|1t} \\ \vdots \\ \beta\tilde{\pi}_{|\mathcal{A}|N} + \tilde{\theta}^m_{|\mathcal{A}|Nt} \end{pmatrix}, \quad u = \begin{pmatrix} d_{11t} \\ \vdots \\ d_{1Nt} \\ d_{21t} \\ \vdots \\ d_{|\mathcal{A}|1t} \\ \vdots \\ d_{|\mathcal{A}|Nt} \end{pmatrix}$$

In addition, define a mapping $i' : \{1, 2, \ldots, |\mathcal{A}|\} \times \{1, 2, \ldots, N\} \to \{1, 2, \ldots, |\mathcal{A}| \times N\}$ that maps the demand types (characterized by $a^{[n]}$'s) to the indices of the newly defined vectors v^m and u. For future convenience, we also group the indices into K disjoint sets, such that all indices corresponding to additive demand appears in the same set as others related by precedence constraints, and the indices corresponding to non-additive demand are included in singleton sets.

Since $C^m_t(\mathbf{d}^\intercal_t \mathbf{w}_t)$ is non-decreasing in $\mathbf{w}_t$, when either of the two conditions in Proposition 10 holds, the optimal solution of problem $(\mathbf{P}')$ is also the optimal solution of problem $(\mathbf{P}'')$ defined as follows:

$$\begin{aligned} (\mathbf{P}'') \qquad \max_{\mathbf{w}_t, z} \quad & (v^m)^\intercal \mathbf{w}_t - C^m_t(z) \\ s.t. \quad & u^\intercal \mathbf{w}_t \le z \qquad (3.17) \\ & w_{i'(i(s)+j,n)} \le w_{i'(i(s)+j+1,n)} \qquad \forall j = 0, 1, \ldots, |\mathcal{A}_s| - 1 \quad \forall s \in \mathcal{S} \quad 1 \le n \le N \\ & w_{i'(a,n)}) \in \{0, 1\} \qquad \forall a \in \mathcal{A}, 1 \le n \le N \end{aligned}$$

Obviously, constraint (3.17) binds at the optimal solution. We name this problem $(\mathbf{P}'')$ *Variable Budget Precedence Constraints Knapsack Problem* (VBPCKP), in which the weight and the value of item a are v^m_a and u_a, respectively. Without the precedence constraints, problem $(\mathbf{P}'')$ becomes the *Variable Budget Knapsack Problem* (VBKP), which is reducible to the VBPCKP. Unlike the *knapsack problem* (KP), the capacity z of either the VBKP or the VBPCKP is not fixed. Instead, z is purchased at the cost of $C^m_t(z)$. To the knowledge of the authors, both the VBKP and the VBPCKP have not been studied before, and both the VBKP and the VBPCKP are harder than KP since budget z is also a decision variable.

A heuristic (greedy) algorithm for the VBKP exists when C^m_t is continuous and differentiable. Let $e_a = \dfrac{v^m_a}{u_a}$ be the "efficiency ratio" of item a. We rank the items in the order of decreasing efficiency ratios. Then the heuristic algorithm picks items from the one with the highest efficiency ratio until the cut-off item. The cut-off item is the one after picking which,

$\frac{C_t^m(z)}{dz}\Big|_{z=z_0}$ is greater than the efficiency ratio of the next item, where z_0 is the sum of the weights of picked items. For example, suppose $C_t^m(z)$ is linear in z, that is, $C_t^m(z) = p^0 z$, then we can simply select all items with efficiency ratios greater than the constant price rate p^0.

It is obvious that this heuristic algorithm generates the optimal solution if we relax the integer constraints. We name the relaxed problem *Continuous Variable Budget Knapsack Problem* (CVBKP). However, for the VBKP with integer decision variables, the above algorithm does not necessarily provide optimal solutions. In Appendix B, we present a counterexample in which the greedy algorithm fails. Moreover, there is no monotone policy for obtaining optimal $\mathbf{w}_t$.

Despite the existence of a polynomial time algorithm that solves the CVBKP as described above, it is challenging to solve the VBKP and the VBPCKP. The following theorem shows that both the latter two problems are NP-hard:

Theorem 1. *If the cost of budget can be evaluated in polynomial time, then the VBKP and the VBPCKP are NP-hard.*

In order to solve the VBPCKP efficiently, we first convert it to the *Variable Budget Multiple-Choice Knapsack Problem* (VBMCKP), by re-defining the value and weight of each item in each $\mathcal{A}_s$ as the following:

$$u'_{i'(i(s)+j,n)} = \sum_{k=j}^{|\mathcal{A}_s|-1} u_{i'(i(s)+k,n)} \quad \text{, and} \tag{3.18}$$

$$v'_{i'(i(s)+j,n)} = \sum_{k=j}^{|\mathcal{A}_s|-1} v_{i'(i(s)+k,n)} \quad \forall j = 0, 1, \ldots, |\mathcal{A}_s| - 1 \quad \forall s \in \mathcal{S}, 1 \leq n \leq N \tag{3.19}$$

Then, the precedence constraints are converted into the following multiple-choice constraints:

$$\sum_{k=0}^{|\mathcal{A}_s|-1} w'_{i'(i(s)+k,n)} = 1 \tag{3.20}$$

Equivalently speaking, constraint (3.20) forces that at most one item can be selected from each group, and $w'_{i(a_s+j,n)} = 1$ in the VBMCKP is equivalent with having the decision variables of the VBPCKP satisfy the following constraint:

$$w_{i'(i(s)+j,n)} = w_{i'(i(s)+j+1,n)} = \ldots = w_{i'(i(s)+|\mathcal{A}_s|-1,n)}$$

Meanwhile, the weights and values of the items that correspond to non-additive demand remain the same. The equivalence established above suggests that through above steps, the VBPCKP is converted to the VBMCKP. In addition, let $k = 1, 2, \ldots, K$ be the index of multiple-choice groups and let the groups of non-additive demand types be singleton. We outline an algorithm that can solve the VBMCKP efficiently (in pseudo polynomial time):

Algorithm 1 Pseudo Code for VBMCKP

for $B = 0 \rightarrow ||\mathbf{d}_t||_1$ **do**
  $f_0(B) \leftarrow -C_t(B)$
  for $k = 1 \rightarrow K$ **do**
    $$f_k(B) \leftarrow \max \left\{ f_{k-1}(B), \quad \begin{array}{ll} \max\limits_{a \in \mathcal{A}_k} & f_{k-1}(B - u'_a) + v'_a \\ s.t. & u'_a \leq B \end{array} \right\}$$
  end for
end for
Choose B that maximizes $f_K(B)$

The following proposition, which is obtained directly from the *Principle of Optimality*, suggests that the solution obtained by the above algorithm is optimal.

Proposition 11. *The algorithm shown in Algorithm 1 solves the VBMCKP optimally.*

Additionally, if all the weights are integral (or by normalization can be converted into integers), Algorithm 1 is a pseudo-polynomial time algorithm. In particular, the computational time is $O(K \sum_n \sum_a d_{a^{[n]}t})$. The proof is similar to that of the pseudo-polynomial time algorithm for the KP, thus, is omitted here.

Updating Rule and Exploration Rule

The approximation is updated by using the stochastic gradient method. Let $\Lambda_t = (\widetilde{\mathbf{\Theta}}_t, \psi_t, \tilde{\eta}_t)$. We minimize the following loss (potential) function over the coefficients Λ_t to minimize the expected difference between our estimation and the realization of the value-to-go plus the discomfort:

$$F(\widetilde{\mathbf{\Theta}}_t^m, \psi_t^m, \tilde{\eta}_t^m) = \frac{\mathbb{E}\left[\overline{\Gamma}_{t+1}(R_t^X; \widetilde{\mathbf{\Theta}}_t^m, \psi_t^m, \tilde{\eta}_t^m) - \widehat{\Gamma}_{t+1}\right]^2}{2}$$

Recall that $\overline{\Gamma}_{t+1}(R_t^X; \widetilde{\mathbf{\Theta}}_t^m, \mathbf{\Phi}_t^m, \tilde{\eta}_t^m)$ is the estimation, while $\widehat{\Gamma}_{t+1}$ is a realization. Note that in our case, we have relatively stronger temporal correlation between different periods. With the classic stochastic gradient updating rule, it may take multiple iterations in order to convey the effect of taking a specific action at later periods back to the post decision state at

the earlier periods. Therefore, *Temporal Difference* learning, also known as TD(λ) learning, is applied to get faster convergence for $\bar{\Lambda}_t^m$, for all $t \in T$:

$$\begin{cases} \widetilde{\mathbf{\Theta}}_t^{m+1} = \min\left(0, \widetilde{\mathbf{\Theta}}_t^m - \gamma_m \sum_{\tau=t}^{T} \lambda^{t-\tau}(\overline{\Gamma}_{\tau+1} - \widehat{\Gamma}_{\tau+1})\mathbf{d}_t^X\right) \\ \psi_t^{m+1} = \max\left(0, \psi_t^m - \gamma_m \sum_{\tau=t}^{T} \lambda^{t-\tau}(\overline{\Gamma}_{\tau+1} - \widehat{\Gamma}_{\tau+1})b_t^X\right) \\ \tilde{\eta}_t^{m+1} = \tilde{\eta}_t^m - \gamma_m \sum_{\tau=t}^{T} \lambda^{t-\tau}(\overline{\Gamma}_{\tau+1} - \widehat{\Gamma}_{\tau+1}) \end{cases} \tag{3.21}$$

We use *Harmonic stepsize* $\gamma_m = \dfrac{\gamma}{\gamma + m - 1}$ in the above updating rule, and γ is chosen carefully to guarantee fast convergence. The selected stepsize satisfies the three basic conditions for convergence of the stochastic gradient method; see [40] for the proof for sufficiency. Note that the rule above updates $\bar{\Lambda}_t^m$ based on only the realization at time $t+1$. However, $\bar{\Lambda}_t^m$ should depend on more future realizations. Especially in our case, we have relatively strong temporal correlation between different periods.

In addition, a modified mixed exploration strategy is applied. The rate of exploration is a piece-wise linear function $\rho(m)$, where m is the iteration number. We set $\rho(m)$ in such a way that the EMS explores more states and collect more information (by updating the initial approximation for more states) at early stages, and then exploit the collected information at later stages to get an approximation. The next subsection discuss the bounds on the performance of the ADP approach.

Upper bound for the ADP approach

Following the ADP approach, the EMS obtains solutions without having to infer the future price structures, the distributions of demand arrivals, and the distribution of future local generation, etc. Instead, the EMS directly learns the approximate value-to-go via sample paths of the future, which are generated by exploiting the local information of history. In addition, the linear parametrization in the ADP approach allows the EMS to find an efficient algorithm to solve the approximate one-period problem. However, the saving in computational effort and the reduction in dependence on knowledge of the random processes come at the cost of optimality. In this section, we provide an intuitive upper bound on the worst-case performance of the ADP approach.

We first introduce the following notation to ease the discussion. Despite the existence of random information, part of the one-period cost can be calculated accurately. More

specifically, given that the initial state status and the decisions are $(R_t, \mathcal{H}_t)$ and X_t, the cost plus the discomfort from deferring demand and unsatisfied demand, denoted as $E_t(R_t, X_t)$, equals to:

$$E_t(R_t, X_t) = C_t(R_t, X_t) + \sum_{a \in \mathcal{A}} \sum_{1 \le n \le N} \tilde{\pi}_{a^{[n]}} d^X_{a^{[n]}t} \tag{3.22}$$

Recall that $L_t(R_t, \mathcal{H}_t, \mathbf{w}_t)$ denotes the discomfort from lost arrivals, and $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ is the optimal expected value-to-go, starting from state $(R_t, \mathcal{H}_t)$ and making decision X_t. In addition, $\overline{L}_t\left(R^X_t; \mathbf{\Phi}^m_t, \zeta^m_t\right)$ and $\overline{V}_{t+1}(R^X_t; \widetilde{\mathbf{\Theta}}^m_t, \psi^m_t, \eta^m_t)$ are the estimation of $L_t(R_t, \mathcal{H}_t, \mathbf{w}_t)$ and $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$, and $\Gamma_{t+1}(R_t, \mathcal{H}_t X_t) = \beta L_t(R_t, \mathcal{H}_t, \mathbf{w}_t) + V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$, while $\overline{\Gamma}_{t+1}(R_t, X_t) = \beta \overline{L}_t\left(R^X_t; \mathbf{\Phi}^m_t, \zeta^m_t\right) + \overline{V}_{t+1}(R^X_t; \widetilde{\mathbf{\Theta}}^m_t, \psi^m_t, \eta^m_t)$. Let the true optimal control policy that minimizes $(\mathbf{P^*})$ and the policy that minimizes $\widetilde{\mathbf{P}}$ be μ^* and μ, respectively. Then denote $X^{\mu^*}_t$ and X^{μ}_t as the decisions following from the two policies, where the initial state is $(R_t, \mathcal{H}_t)$. We further define $J^{\mu}(R_t, \mathcal{H}_t)$ as the following:

Definition 3. *Let $J^{\mu}_t(R_t, \mathcal{H}_t)$ be the expected total disutility of starting from state $(R_t, \mathcal{H}_t)$ and following the policy generated by the ADP approach at the first step and then following the optimal policy thereafter. That is:*

$$J^{\mu}_t(R_t, \mathcal{H}_t) = E_t(R_t, X^{\mu}_t) + \Gamma_{t+1}(R_t, \mathcal{H}_t, X^{\mu}_t)$$

We then provide an upper bound on the performance of the ADP approach, based on the above notation. Specifically, the following proposition describes the upper bound of $J^{\mu}_t(R_t, \mathcal{H}_t) - J^*_t(R_t, \mathcal{H}_t)$.

Proposition 12. *Suppose that the initial state status is $(R_t, \mathcal{H}_t)$. Let*

$$B_t(R_t, \mathcal{H}_t) = \max_{X_t} \left|\Gamma_{t+1}(R_t, \mathcal{H}_t, X_t) - \overline{\Gamma}_{t+1}(R_t, X_t)\right|$$

That is, the ∞-norm of the difference between the vector of the optimal expected value-to-go and its approximation over all possible decisions. In addition, let $\widetilde{B}_t(R_t, \mathcal{H}_t)$ be as follows:

$$\begin{aligned} \widetilde{B}_t(R_t, \mathcal{H}_t) = \max_{X_1, X_2} \quad & \overline{\Gamma}_{t+1}(R_t, X_1) - \overline{\Gamma}_{t+1}(R_t, X_2) \\ s.t. \quad & E_t(R_t, X_1) + \overline{\Gamma}_{t+1}(R_t, X_1) \le E_t(R_t, X_2) + \overline{\Gamma}_{t+1}(R_t, X_2) \end{aligned}$$

Then, $0 \le J^{\mu}_t(R_t, \mathcal{H}_t) - J^*_t(R_t, \mathcal{H}_t) \le B_t(R_t, \mathcal{H}_t)$. Additionally, if $\overline{\Gamma}_{t+1}(R_t, X^{\mu^*}_t) - \overline{\Gamma}_{t+1}(R_t, X^{\mu}_t) \ge 0$ and $\Gamma_{t+1}(R_t, \mathcal{H}_t, X^{\mu^*}_t) - \Gamma_{t+1}(R_t, \mathcal{H}_t, X^{\mu}_t) \le 0$, then $0 \le J^{\mu}_t(R_t, \mathcal{H}_t) - J^*_t(R_t, \mathcal{H}_t) \le \min\left\{B_t(R_t, \mathcal{H}_t), \widetilde{B}_t(R_t, \mathcal{H}_t)\right\}$.

Intuitively, Proposition 12 validates the conjecture that, if the approximation of the sum of the value-to-go function and the discomfort of lost arrivals get uniformly closer to its true value, then the ADP approach is able to generate better decisions in terms of lower long term average total disutility. The following subsection summarizes the ADP approach.

Discussion and Summary of the ADP approach

It is worthwhile to note that even if we change the way in which the basis is defined, as long as the approximation of value-to-go functions is linear in the decisions, the above ADP approach and the corresponding analytical results still hold. For instance, instead of directly using $\mathbf{d}_t^X$ as the vector of basis for outstanding demand, we could have set $\left(\frac{\pi_{a^{[n]}}}{\min_{i,k} \pi_{i^{[k]}}}\right) \times d^X_{a^{[n]}t}$ as the basis for each of the outstanding demand to reflect their difference in discomfort from deferring. In practice, we suggest exploring through a variety of possible basis to find the one that works best. However, it is not necessary that the performance of the ADP approach will be better even if a more sophisticated basis is chosen.

We end this section by summarizing the algorithm that solves the stochastic optimal control problem using the ADP approach:

-*Step* 1 *At time t, observe state status, get forecasts for periods from t to $t+T$. Initialize the coefficients in the approximate value-to-go function;*

-*Step* 2 *Do for $m = 1 : M$, where M is the total iteration number;*

Do for $\tau = t : t+T$

Step 2.1 *Calculate $C^m_\tau(z)$ based on (3.15);*

Step 2.2 *Calculate u' and v' based on equations (3.18) and (3.19);*

Step 2.3 *Solve the one-period problem use the algorithm for the VBMCKP;*

Step 2.4 *Update the coefficients of $\bar{\Lambda}^m_{\tau-1}$ by (3.21);*

Step 2.5 *Proceed to next time period by state transition mechanisms;*

-*Step* 3 *Proceed to next time period by making decision based on $\bar{\Lambda}^M_t$ and the state transition mechanism.*

3.4 Numerical Study

In this section, we perform extensive numerical studies for the ADP approach. In particular, we compare the policy obtained by using the ADP approach, with the optimal policy obtained from backward induction, the myopic policy defined in Definition 4, and the traditional (no-control) policy. For notational convenience, denote these fore policies by ADP, EXDP, MYO, and TRD, respectively.

Definition 4. *The Myopic policy aims to minimize only the current period's total disutility. That is, it solves the following problem at the beginning of each period:*

$$\begin{aligned} \min_{X_t} \quad & C_t\left(R_t, X_t\right) + \beta U_t\left(R_t, \mathcal{H}_t, X_t\right) \\ s.t. \quad & (3.1a) - (3.1f),\ (3.2) - (3.5),\ (3.6) \end{aligned}$$

The backward induction approach solves for the optimal policy for the original problem ($\mathbf{P}^*$), thus its objective value serves as a lower bound on the total disutility. The MYO is essentially a greedy policy, because it minimizes the one-period total disutility. The TRD imposes no control, that is, it satisfies demand upon arrivals and consumes energy stored in the storage as much as possible. To make the comparison more competitive, we assume that all probability distributions are known to the EXDP and the MYO.

We perform numerical studies in order to answer the following questions: (1) whether the ADP approach is able to achieve load shifting, (2) how good the performance of the ADP approach is compared with the EXDP, the MYO and the TRD, and (3) what effects adjusting the parameters has, such as altering user's sensitivity on service level, the difference between peak and off-peak prices, and amount of local generation. We start answering these questions by first introducing the experiment setting at the beginning of the next subsection.

Experiment Setting

We use a generic setting in our numerical studies. In particular, the EMS controls three appliances and one storage device, which is linked with one local generator. The three flexible appliances have different demand profiles. In order to make it computationally tractable to obtain optimal solutions, we assume that all random information processes are markovian. We also assume that the distributions are not known to the ADP approach. Recall that the ADP approach does not require the random processes to possess the markovian property.

Figure 3.4 plots the averages of the random processes over 100 sample paths. We assume a linear price structure (flat-rate) in each period. The prices are generated by truncating the sum of a baseline price and normal white noises. The baseline price profile is convex increasing from off-peak to peak hours. This is consistent with the fact that the market clearing price, at which the convex increasing aggregate supply curve meets the relatively inelastic demand curve, is convex increasing from off-peak to peak hours. We add another valley period, which is symmetric to the peak period, in the middle the off-peak periods. The price is concave decreasing from ordinary off-peak periods to the valley period. This assumption enables us to investigate the scenario when extreme low price is offered to encourage users to dial up demand. This may happen when we want to couple demand with surging renewable sources, whose installed capacity keeps increasing in earnest recently. At last, it is assumed that there is only one peak and one valley period throughout the planning horizon.

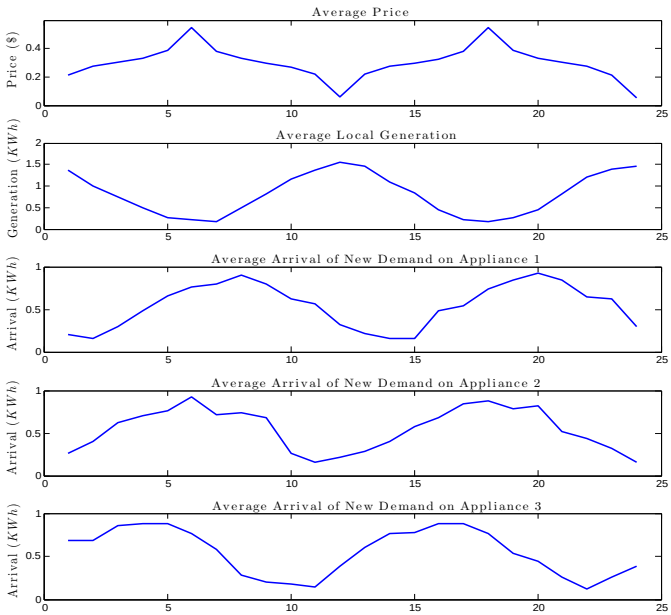


Figure 3.4: The Random Information Processes

At the beginning of each period, demand arrivals on each of the three flexible appliances are sampled from independent Bernoulli random variables. The arrival probabilities are the truncated sum of normal white noises and three sinusoidal shape baselines, which are assumed to have slightly different phase angle to capture the difference among appliances. Moreover, we assume that the baseline arrival probabilities are positively correlated with the prices. Demand also have random allowable delays, which are not plotted in Figure 3.4. The allowable delays are again sampled from different truncated normal random variables with different means and variances. Figure 3.4 plots the average arrival probabilities for the demand on these three flexible appliances.

We assume that local generation is negatively correlated with prices and demand arrivals. Samples are generated from the truncated sum of a sinusoidal shape baseline and normal white noises. This assumption of the negative correlation with prices on the local generation matches the fact that most local generation from renewable sources such as solar and wind always reach maximum output in the off-peak hours.

The convergence of the ADP approach is largely determined by the choice of stepsizes. [53] notes that in practice the stepsizes should be scaled so that the estimations of coefficients will not oscillate too much in early iterations. We adopt this guideline and the learning process in the ADP approach converges nicely within at most 2000 iterations. Our numerical study also shows that warm starting the learning process by initializing with the estimations of coefficients in the previous period leads to much more faster convergence. The value of λ

for the TD(λ) approach is chosen to be 0.3, because a greater value increases the instability during the learning, while a smaller value results in slower convergence.

At last, we set the length of planning horizon as 12 periods (that form one day). Monte-Carlo Integration is used to estimate the expected total disutility over the entire planning horizon. In particular, for each setting, the same randomly drawn 100 sample paths are used to test all the four policies. Note that to generate the policies, training sample set for the ADP approach is generated separately from the sample generator, which is hidden from the ADP approach. The number of look-ahead periods for both the backward induction approach and the ADP approach is set as 12 as well. The settings and simulation results are presented and discussed in the next subsection.

Simulation Results and Analysis

	Price ($)		Generation (kWh)		New Demand		Battery (kWh)		π ($/$kWh$)		
	p_{avg}	p_{spd}	g_{avg}	g_{spd}	q_{avg}	q_{spd}	c_b	$u_b(=l_b)$	π	π'	π''
Run 1	0.3	**0.1**	0	0	0.55	0.45	4	2	0.05	0.15	50
Run 2	0.3	**0.25**	0	0	0.55	0.45	4	2	0.05	0.15	50
Run 3	0.3	0.25	0	0	0.55	0.45	4	2	0.05	0.15	**0.75**
Run 4	0.3	0.25	0	0	0.55	0.45	4	2	0.05	**0.4**	0.75
Run 5	0.3	0.25	0	0	0.55	0.45	4	2	**0.25**	**0.5**	0.75
Run 6	0.3	0.25	**0.0.75**	**0.5**	0.55	0.45	4	2	0.05	0.4	0.75
Run 7	0.3	0.25	**1.5**	**0.5**	0.55	0.45	4	2	0.05	0.4	0.75
Run 8	0.3	0.25	0	0	0.55	**0.0.15**	4	2	0.05	0.4	0.75
Run 9	0.3	0.25	0	0	**0.3**	0.0.15	4	2	0.05	0.4	0.75
Run 10	0.3	0.25	0	0	0.55	0.45	**8**	**4**	0.05	0.4	0.75

Table 3.3: Summary of Experiment Settings

We consider a variety of settings and list ten most representative ones in Table 3.3. In general, we perform the following controlled experiments: (1) fixing the average (taken across the planning horizon) expected price (p_{avg}), and varying the spread between the peak and valley prices (p_{spd}), as in Run 1 and 2; (2) adjusting the sensitivities of the users on deferring demand, lost arrivals and unsatisfied demand, as in Run 3 to 5; (3) altering the output of local generation (similarly as p_{avg} and p_{spd}, we define g_{avg} and g_{spd} as the average local generation and the spread between peak and valley outputs), as in Run 6 and 7; (4) changing the new arrival probabilities (denote the average arrival probabilities and the their spread as q_{avg} and q_{spd}), as in Run 8 and 9, and (5) adjusting the capacity and the charging/discharging rate of the local storage device, as in Run 10. We also perform tests with various average allowable delay (before the deadlines of demand upon arrival), which are not listed in Table 3.3.

The total disutilities (across the planning horizon) returned by the four tested policies are summarized in Table 3.4. The results show that the total disutilities generated by the

	EXDP			ADP			MYO			TRD	
	Cost	Comfort	Total	Cost	Comfort	Total	Cost	Comfort	Total	Cost	Total
Run 1	0.1	5.4	5.5	0.6	5.1	5.7	-0.2	5.1	4.9	12.5	12.5
Run 2	-0.6	5.0	4.3	0.3	4.1	4.4	0.2	4.8	5.0	13.3	13.3
Run 3	1.1	3.4	4.6	1.3	3.3	4.6	0.8	4.5	5.3	13.2	13.2
Run 4	1.8	6.2	8.0	2.3	5.9	8.1	1.6	7.7	9.3	12.9	12.9
Run 5	9.4	2.3	11.7	10.0	1.7	11.6	5.0	10.0	15.0	13.0	13.0
Run 6	0.2	6.4	6.6	1.1	6.0	7.1	-0.2	7.9	7.7	11.4	11.4
Run 7	-2.2	6.6	4.4	-0.9	6.0	5.0	-2.7	8.0	5.4	9.0	9.0
Run 8	2.0	6.2	8.2	3.1	5.3	8.4	1.1	8.0	9.2	12.4	12.4
Run 9	1.6	3.2	4.8	3.6	1.8	5.5	0.9	5.2	6.1	7.7	7.7
Run 10	0.8	6.3	7.2	1.5	5.8	7.3	2.0	7.7	9.6	13.1	13.1

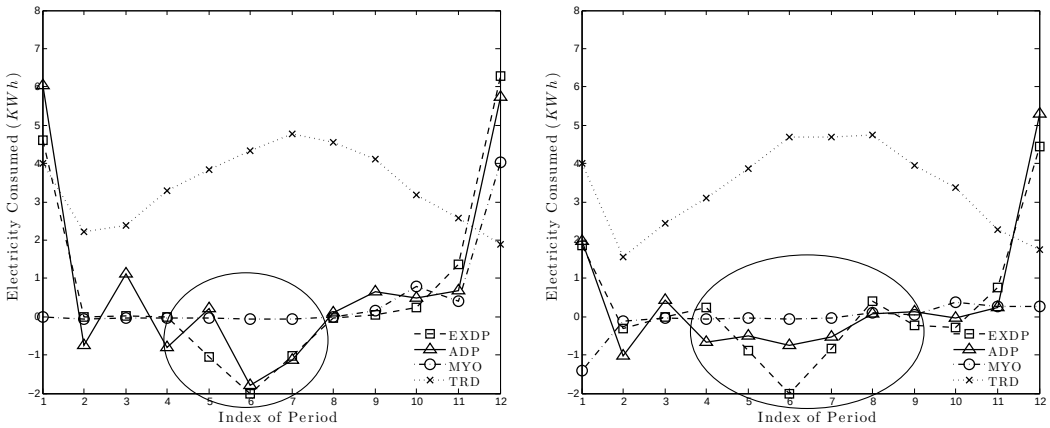
Table 3.4: Summary of Experiment Results (Units: $)

ADP approach are always close to those of the EXDP approach. In addition, the ADP approach dominates the other two. Namely, the results demonstrate the effectiveness of the ADP approach. The myopic policy also works well under many of the settings compared with the traditional policy, but as will be shown later, it reduces in the total disutility by inefficiently deferring demand, hence its performance is vulnerable to the user's sensitivity on service level and the average allowable delay of demand.

Figure 3.5 to Figure 3.7 plot the energy consumption profiles generated by the four policies under different settings. The consumption of the TRD perfectly reflects the prices of electricity, which are positively correlated with demand arrivals and TRD's demand executions. The figures show that the consumption profiles of the ADP match those of the EXDP, suggesting that the ADP is able to achieve almost the same load shifting as the EXDP. On the contrary, the relatively flat consumption profiles of the MYO suggest that the MYO fails to take advantage of off-peak low prices.

When prices are less volatile, there is less incentive for load shifting. As shown in Figure 3.5, the more volatile the prices are, the more loads the EXDP and the ADP shift. Specifically, when the prices are volatile, both the EXDP and the ADP consume more energy in periods 1 and 12 on average. When prices are less volatile, the ADP fails to sell as much energy back to the grid as the EXDP does. It should also be noted from Table 3.4 that both the EXDP and the ADP return lower total disutility under more volatile prices, while the MYO and the TRD generate more disutility for the users.

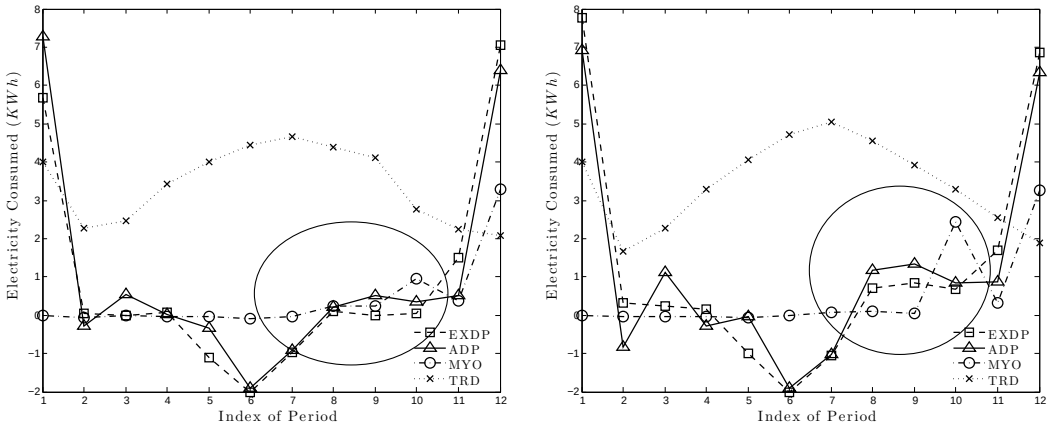
As we increase user's sensitivity (captured by parameters π, π', and π'') on service level, the total energy consumed increases. In particular, if the unit discomfort from unsatisfied demand (π'') is higher, as shown in Figure 3.6, energy consumption in peak hours (periods 5, 6, and 7) remains almost the same, while the consumption in off-peak periods increases. In addition, both the EXDP and the ADP create notable less total disutility and less discomfort



(a) $p_{avg} = \$0.3$, $p_{spd} = \$0.3$. When price is more volatile, the ADP and the EXDP reach minimum consumption at peak hours. (See circled part)

(b) $p_{avg} = \$0.3$, $p_{spd} = \$0.1$. When price is less volatile, the ADP's consumption is between those of the EXDP and the MYO at peak hours. (See circled part)

Figure 3.5: Energy Consumption under Different Price Volatilities.



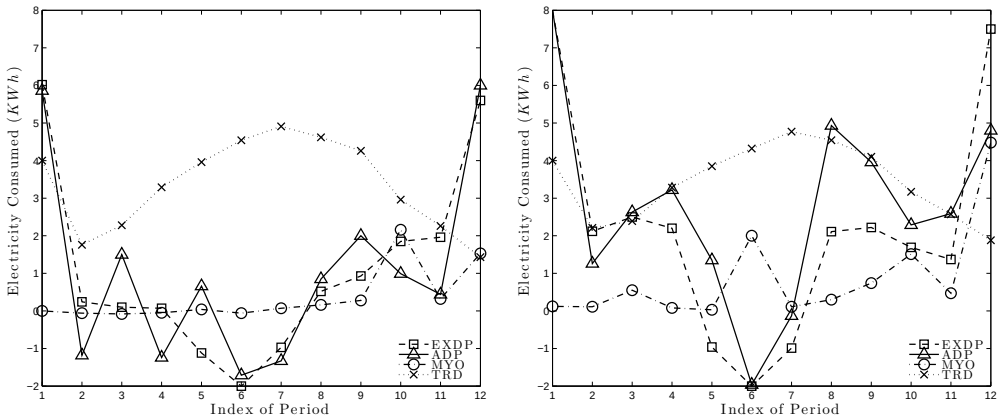
(a) $\pi'' = (\$0.5, \$0.5, \$0.5)$. When users are less sensitive on unsatisfied demand, consume less energy in off-peak hours. (See circled part)

(b) $\pi'' = (\$0.75, \$0.75, \$0.75)$. When users are more sensitive on unsatisfied demand, consume more energy in off-peak hours. (See circled part)

Figure 3.6: Energy Consumption under Different Sensitivity on Unsatisfied Demand

associated with unsatisfied demand than those of the MYO and TRD.

When the sensitivity associated with lost arrivals (π') increases, the EXDP and the ADP further increase off-peak consumption, while the MYO fails to shift some of the demand in peak hours, as shown in Figure 3.7. However, results in Table 3.4 suggest that the discomfort



(a) $\pi' = (\$0.15, \$0.15, \$0.15)$. When users are less sensitive on lost arrival, have less volatile consumption profiles.

(b) $\pi' = (\$0.6, \$0.6, \$0.6)$. When users are more sensitive on lost arrival, have more volatile consumption profiles.

Figure 3.7: Energy Consumption under Different Sensitivity on Lost Arrivals

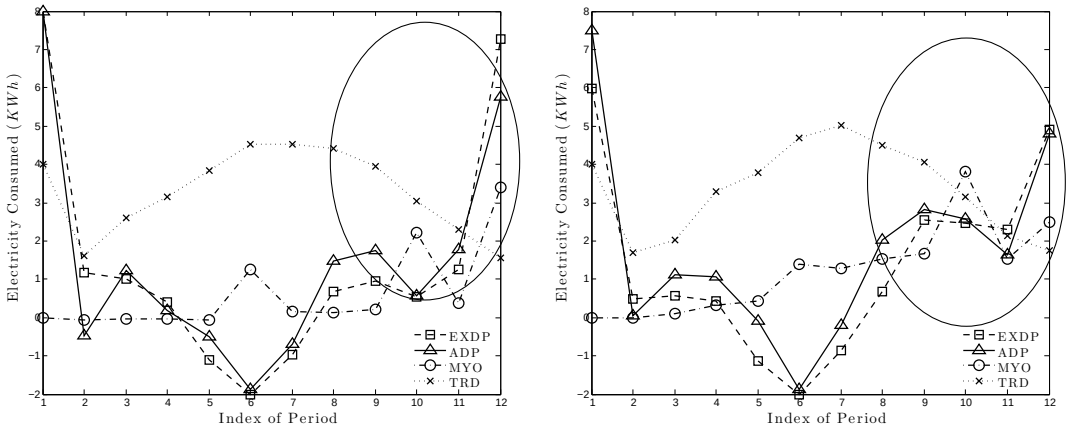
associated with lost arrivals further increases as well. In fact, as π' increases, more energy is consumed and less demand is lost. The discomfort increases simply because the increase in π' dominates the decrease in lost demand. When π' is high enough and passes a pivoting value, the EXDP and the ADP start to decrease the discomfort associated with lost arrivals. Above results demonstrate that, similar to the EXDP, the ADP is capable to capture user's sensitivity on services.

Having higher local generation capacity benefits all the four policies. When the output is limited by the charging rate of the local storage, the savings in cost determines the reduction in total disutility. Intuitively, the MYO and the TRD directly reduce cost by selling the local generation capacity back to the grid, while although the EXDP and the ADP arbitrage prices intertemporally via the storage, they further save cost in off-peak periods by charging the storage with locally generated electricity, thus the savings for the EXDP and the ADP are equivalent to those for the MYO and the TRD. Obviously, the EXDP and the ADP are able to save more if increasing block-rate or convex increasing prices are offered, or if the local generation capacity is greater than the charging rate limits.

Lowering the volatility of the new arrival probabilities does not cut the total utilities generated by the EXDP and the ADP, while it decreases those introduced by the MYO and the TRD. Intuitively, when demand arrives according to steady rates in every period, the limit on queue capacity leads to decline in demand flexibility, because shifting demand costs inefficient high discomfort from lost arrivals. The effect of lowering arrival volatility is very similar to that of increasing the discomfort associated with deferred demand. When the average arrival rates are lower, all the four policies incur less total disutility, simply because

less demand needs to be satisfied.

In addition, having extra storage and higher charging/discharging rate helps the EXDP and the ADP save more cost on electricity. Intuitively, more energy can be traded intertemporally to arbitrage. In addition, having local storage resembles the benefits of having local generation for both the EXDP and the ADP, but not for the MYO and the TRD. Sometimes, local storage comes at zero cost (such as the battery on a plug-in electric vehicle), hence is more cost attractive than expensive local generation devices.



(a) With Further Deadlines, have less volatile consumption profiles. (See circled part)

(b) With Nearer Deadlines, the profiles of the ADP and the EXDP get closer to that of the TRD. (See circled part)

Figure 3.8: Energy Consumption with Different Average Allowable Delay of Demand

At last, decreasing the average allowable delay has similar effects on the EXDP and the ADP as lowering the new arrival volatility. Obviously shorter allowable delays lead to fewer opportunities for load shifting. Figure 3.8 compares the cases with further and nearer deadlines. In the latter case, the consumption profiles generated by the EXDP and the ADP get closer to that of the TRD.

We conclude this section by providing some insights based on the numerical results with the ADP approach. Firstly, users should adjust the sensitivity settings to reflect how they evaluate the trade-off between comfort and cost savings. Secondly the users are protected against price hikes under a large range of settings, because the energy consumption in the peak periods is effectively controlled at the minimum level with the ADP approach. In addition, the ADP approach returns reasonable price elasticity. Therefore, to shave-off peak demand, creating price difference between periods is effective. Moreover, since the price elasticity of demand is negatively correlated with users' sensitivity on service, there is an opportunity for LSEs to adopt Ramsey-pricing. At last, our results also suggest that the

traditional way to model demand as a non-increasing function of the real-time price is not accurate when there is load shifting opportunities. Instead, demand depends on many factors such as past and future prices and demand arrivals. Our model can be used as a module to generate demand management signals in the analysis on pricing decisions.

3.5 Summary

In this chapter, we study an approximate dynamic programming approach to make energy usage decisions for users confronted with real-time pricing. The main features of our model include its ability to incorporate local energy generation, storage, and energy exchange between the user and the grid. The model adapts to stochastic arrival processes of flexible demand, which have random allowable delays. Moreover, our model takes into consideration both the cost to satisfy demand and the discomfort resulting from load shifting.

The straightforward way of modeling the problem requires both integer and continuous variables. In this chapter, we present a novel model that makes all decision variables integral, and then outline an ADP approach to overcome the "curses of dimensionality". The ADP approach is able to solve the problem efficiently, although at the cost of suboptimality. Then, we conduct numerical studies to test the performance of the ADP approach and study the effects of varying the parameters that reflect many of the important features of energy management problems in practice. The simulation results demonstrate that the performance of the ADP approach dominates those of a myopic policy and the traditional no-control policy. Our simulation results further suggest that the ADP approach is able to capture the sensitivity settings of users, and it protects the users from price hikes by providing sufficient price elasticity. Some other insights are also discussed.

Our model can be generalized to other dynamic resource allocation problems, in which there are arrival processes of flexible demand that can be shifted or ignored with a certain amount of discomfort. For instance, suppose a single assembly manufacture faces highly unstable demand, and the procurement prices for raw materials also vary quickly. If the manufacture applies an Available to Promise (ATP) policy, then with a similar model formulation and the corresponding ADP approach, the manufacture can quickly generate an estimate of the delivery dates (which are the times at which the demand can be satisfied) for the orders, without having to know the distributions of the random processes. In particular, the procurement schedule is determined jointly by the price forecasts and inventory cost, while the delivery dates are set at times that minimize the manufacturer's cost. Many other problems fit into the framework of this chapter, and the ADP approach can serve as either a practicable way for the originally difficult-to-solve problem, or a benchmark for better optimal solution approaches.

It is possible to extend our work in the following ways. Firstly, there are potentially

better ways to approximate the value-to-go terms. For instance, instead of fitting a linear function of the post-decision state, higher order polynomials may generate better results. Secondly, our model relies on the existence of a module that takes the information of history as well as the current state status, and generates sample paths for the future. However, the design of such a module is not discussed in this chapter. Last but not least, although our model and the corresponding ADP approach are capable to adapt price structures such as increasing block-rate and flat-rate, the problem of which pricing strategy maximizes social welfare has not studied.

Chapter 4

Other Efficient Algorithms

4.1 Introduction

Flexible demand management is a stochastic optimal control problem. Stochastic programming is one of the earliest approaches in solving stochastic optimal control problems, especially for unit commitment problems (the problem of deciding generation schedule of each generation unit for one period to minimize production costs or maximize profit), as in [21], [52], [58], and [51]. However, in stochastic programming models, scenarios need to be generated based on the forecasts of demand and supply, and the accuracy of optimal solutions highly depends on the quality of forecasts. On one hand, it is difficult to obtain accurate forecasts of the demand distribution; on the other hand, the complexity of solving the problem increases dramatically as the number of scenarios increases.

Stochastic dynamic programming is also a useful tool for stochastic optimal control. For example, [45] formulates a stochastic dynamic program for the EMS and solves it by backward induction. [39] and [51] also utilize dynamic programming models. Nevertheless, the dynamic programming method can be computationally intractable since the size of the state space, the outcome space, and the action space grow very quickly when the vector dimensions increase, which is known as the "three curses of dimensionality", see [53]. As a consequence, Approximate Dynamic Programming (ADP) approaches are developed and shown to be efficient in various applications.

ADP combines adaptive critic and reinforcement learning techniques with dynamic programming. The basic idea is to proceed forward in time, simulate into the future and iteratively update the value function estimations. ADP approaches can be classified into 4 groups based on their adaptive critic design: Heuristic Dynamic Programming (HDP), Dual Heuristic Dynamic Programming (DHDP), Action Dependent Heuristic Dynamic Programming (ADGDP, also known as Q-learning), and Action Dependent Dual Heuristic Dynamic Programming (ADDHDP). See [62], [10], [9], and [53] for comprehensive surveys of ADP

methods. ADP has been widely used in stochastic optimal control problems. For example, [51] formulates the problem of coupling renewable generation with deferrable demand to reduce supply fluctuation as a stochastic dynamic program, uses ADP algorithm to solve, and finds that the ADP approach gets near-optimal solutions. [**lai**] develops an ADP method to obtain the lower and upper bounds of the value of gas storage. [39] follows another ADP approach by limiting the decision space, and obtains estimation for the value-to-go function by sampling. Motivated by the high level of volatility and uncertainty in supply, demand and electricity price, ADP methodologies are especially useful in the control problems of the Smart Grid. For example, [4] develops an Adaptive Stochastic Control (ASC) system for load and source management of real-time Smart Grid operations. They use ADP algorithm to solve the ASC problem with thousands of variables, and demonstrate that the results are close to optimal.

Q-learning is a popular methodology in ADP, especially for those finite-horizon problems, and it has been proved to be an efficient ADP structure ([53]). [2] develops an on-line ADP technique based on Q-learning to solve the discrete-time zero-sum game problem with continuous state and action spaces. They show that the critic converges to the game value function and the action networks converge to the Nash equilibrium of the game. [62] transforms the multi-objective dynamic programming problem into quadratic normal problem by using incremental Q-learning method. [42] utilizes a modified Q-learning algorithm to solve the dynamic programming problem for an intelligent battery controller. They introduce a bias correction term in the learning process, which significantly reduces the bias in values estimation and gains better performance.

In this paper, we study the demand management problem. We first propose a dynamic programming model for this problem. In particular, the central controller faces demand uncertainties, schedules the fulfilling of the outstanding demands on all local appliances, and takes into account of both costs and comfort of users. Due to the "curses of dimensionality", we develop two different methods that trade in optimality for computational tractability and test their performances numerically. Our paper contributes to the literature in the following aspects: (1) we develop a novel model that permits users to specify the allowable delay for demands, (2) the decentralization based heuristic approach provides solutions in a significantly more efficient manner, while the Q-learning approach is able to deliver solutions under more general settings, and (3) both of the two methods for the central control problem perform close to optimal, and since they have their own advantages over each other under different settings, they can serve as complementary approaches in practice.

The remainder of the paper is organized as follows. In section 2, we describe our model formulation, and provide two different approaches that solve the problem efficiently: one is a decentralization based heuristic, and the other is a Q-learning approximate dynamic programming approach. In section 3, we run simulations to compare the performance of the two approaches with those of the exact optimal solutions and the no-control case. Lastly,

section 4 provides some future directions of research extensions and concludes.

4.2 Dynamic Programming Formulation of the Centralized Control Problem

Demands can be categorized into two groups: flexible and inflexible demands. Inflexible demands, such as demands for lighting and TV, cannot be shifted, while flexible demands, such as demands for air conditioning, space heating, and laundry appliances, are usually not time sensitive, thus can be shifted in time. In the presence of time-varying prices, flexible demands provide users with opportunities to hedge against high peak prices. Naturally, how to optimally shift flexible demands falls into the category of stochastic optimal control problems. In this section, we first present the assumptions of our model and our dynamic programming formulation, and then propose two approaches to solve this problem efficiently.

General Assumptions

The objective of our model is to design a general central controller that minimizes the total disutility of Smart Grid users. The disutility of users is assumed to be in dollar-value and contains two parts: the costs of electricity and the discomfort from deferring the demands and lost arrivals. The main trade-off is to save costs through load shifting, i.e., shifting the demand to off-peak hours when the price is low at the cost of discomfort from deferring the demands. From the view of the entire grid, achieving minimum total disutility will help shave-off the peak demand. Consequently, the required energy output from the electricity supplier is reduced, which eventually leads to savings on energy generation costs.

We also assume the existence of distribution estimations of demand arrivals and electricity prices. The first can be obtained by having statistical models to learn the behavior of users for a sufficient length of time. On the other hand, the time-varying price structures for the Smart Grid have been widely discussed. As for the distributions of prices, we expect that at equilibrium, the baseline electricity prices will be relatively easy to forecast based on historical data and market conditions. The randomness comes from the uncertainties associated with renewable sources, as well as the weather condition.

Model Formulation

We first introduce the general rules we follow in defining variables for quick reference. Boldface lowercase is used for vectors. Non-boldface lowercase is for scalars. Uppercase is used

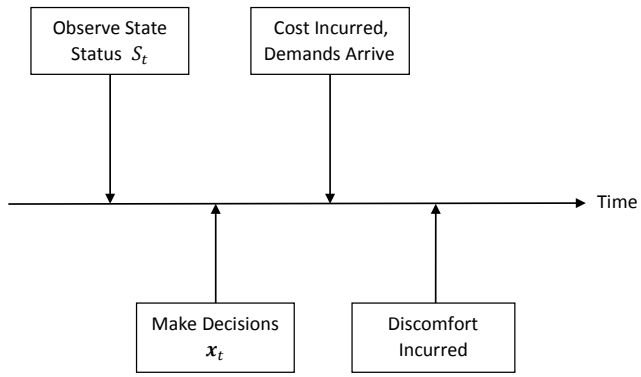


Figure 4.1: Sequence of Events in Period t

to represent functions.

We assume that the central controller works in the following way. According to sequence of events as plotted in Fig. 4.1, at the beginning of period t, the central controller observes outstanding demands and the forecasts of future prices as well as the forecasts of demand arrivals. Next, the controller makes energy usage decision by looking T periods ahead. Then, appliances satisfy demands according to the decisions, and costs are incurred. New demands arrive during this period, and some will become outstanding demands at the beginning of the next period, while some are lost immediately if there are already unsatisfied demands waiting. At last, discomfort is incurred and the system evolves to period $t+1$. Then, the whole decision making and execution process is repeated.

Users are equipped with a set I of smart appliances, such as smart laundry appliances, smart dish washer, smart refrigerator, and even smart printer, etc. We further assume that for the demand of each appliance $i \in I$, user can specify an allowable delay l_i. Then the demand must be satisfied within l_i periods from its arrival period. Thus, $l_i = 1$ implies that the demand must be satisfied immediately. There exists a maximum allowable delay, denoted as L_i, for each appliance $i \in I$.

The state of the system is characterized by outstanding demands, that is, it is implicitly assumed that the whole system is markovian. For non-markovian systems, it is possible to make them markovian by adding more information to the state variables, but it will increase the computational complexity and is beyond the context of this paper. Let the state of the system in period t be $\mathcal{S}_t$, and let $\mathbf{s}_{t,i}$ be the state status of appliance i in period t. Then $\mathcal{S}_t = (\mathbf{s}_{t,i})_{i \in I}$. In particular, the state status $\mathbf{s}_{t,i}$ is a vector in $\{0,1\}^{L_i}$, in which the j-th $(1 \leq j \leq L_i)$ element of $\mathbf{s}_{t,i}$ being one indicates that there exists demand on appliance i that must be satisfied in the next j periods. If all elements of $\mathbf{s}_{t,i}$ are zero, then there is no outstanding demand on appliance i. In addition, we assume that at most one element of $\mathbf{s}_{t,i}$

can be non-zero, that is, each appliance can allow at most one demand waiting. This assumption is intuitive, as for instance, the washing machine cannot have two loads of clothes waiting and automatically reload after finishing one load. Nonetheless, the model can be easily modified to increase the capacity on the number of demands waiting for each appliance.

We assume that new demand for each appliance in each period follows a Bernoulli distribution, and demands are independent across appliances. Relaxing the independence assumption on demand arrivals will not affect the main result of the paper. If demand arrivals have inter-temporal dependence, the markovian property of the system will be changed. Nonetheless as discussed before, adding historical demand arrivals to the state status is sufficient to address this issue. We assume that in period t, demand for appliance i arrives with probability $\lambda_{t,i}$. The allowable delay of the new demand is sampled from a discrete random variable $q_{t,i}$, which takes values of 1 to L_i with probability $(\mu_{t,i,1}, \mu_{t,i,2}, \ldots, \mu_{t,i,L_i})$, where $\mu_{t,i,j} > 0$ for all $j = 1, 2, \ldots, L_i$ and $\sum_{j=i}^{L_i} \mu_{t,i,j} = 1$.

After observing $\mathcal{S}_t$, the controller makes decisions on satisfying or deferring the demands. Let $x_{t,i}$ be the decision on satisfying demand of appliance i, and let e be a vector of ones in $\mathbb{R}^{L_i}$. Thus, if $e^T \mathbf{s}_{t,i} = 1$, then $x_{t,i} = 1$ implies satisfying demand and consumes ψ_i amount of energy. On the other hand, if $e^T \mathbf{s}_{t,i} = 1$ and $x_{t,i} = 0$, then some discomfort is incurred, and the demand is carried on to the next period. Nonetheless, if $\mathbf{s}_{t,i}[1] = 1$ then $x_{t,i}$ must be 1, as in this case the demand on i cannot be further delayed.

The decision of satisfying demand i leads to consuming ψ_i amount of energy, measured in kWh. Electricity price in period t is assumed to be a function of the total energy consumed in that period, denoted as $P_t(\cdot)$ ($\$/kWh$). These price structures on top of time-variability can effectively limit the total consumption in a single period, thus helping eliminates the rebound effect studied in [16]. In summary, the bellman equation for the controller's problem can be modeled as follows:

$$\begin{aligned} J_t(\mathcal{S}_t) \;=\; \min_{\mathbf{x}_t \in \mathcal{X}_t(\mathcal{S}_t)} & \underbrace{C_t\left(\mathbf{x}_t\right)}_{\text{one-period cost}} + \overbrace{\sum_{i \in I} U_{t,i}(\mathbf{s}_{t,i}, x_{t,i})}^{\text{one-period discomfort}} \\ & + \underbrace{\mathbb{E}_{\mathcal{D}_{t+1}}\left[J_{t+1}(\mathcal{S}_{t+1}) | \mathcal{S}_t, \mathbf{x}_t\right]}_{value-to-go} \end{aligned} \tag{4.1}$$

where, $C_t\left(\mathcal{S}_t, \mathbf{x}_t\right)$ is the one-period cost, $\sum_{i \in I} U_{t,i}(\mathbf{s}_{t,i}, x_{t,i})$ is the one-period expected discomfort, and $\mathbb{E}_{\mathcal{D}_{t+1}}\left[J_{t+1}(\mathcal{S}_{t+1}) | \mathcal{S}_t, \mathbf{x}_t\right]$ is the value-to-go term. Set $\mathcal{X}_t(\mathcal{S}_t)$ defines the set of feasible

decisions of $\mathbf{x}_t$. Then $\mathcal{X}_t(\mathcal{S}_t)$ can be expressed as follows:

$$\mathcal{X}_t(\mathcal{S}_t) = \Big\{ x_{t,i} \mid x_{t,i} \leq e^T \mathbf{s}_{t,i};\ x_{t,i} \geq \mathbf{s}_{t,i}[1]; \\ x_{t,i} \in \{0,1\};\ \forall i \in I \Big\} \tag{4.2}$$

Since the price is a function of the total energy usage, $C_t(\mathbf{x}_t) = P_t(\boldsymbol{\psi}^T \mathbf{x}_t) \cdot \boldsymbol{\psi}^T \mathbf{x}_t$. Discomfort comes from two distinctive sources: discomfort from deferring the satisfaction of demand and discomfort from lost arrivals. Deferring demands incurs discomfort $\phi_{t,i}$. Moreover, when there is an outstanding demand and the controller decides to defer it, new arrival of demand for the same appliance is lost because the appliance is occupied by the previous scheduled demand. In this case, each lost arrival incurs discomfort $\pi_{t,i}$. Therefore, the one period expected discomfort is the following:

$$\sum_{i \in I} U_{t,i}(\mathbf{s}_{t,i}, x_{t,i}) = \sum_{i \in I} \Big(\phi_{t,i} \left(e^T \mathbf{s}_{t,i} - x_{t,i} \right) + \\ \pi_{t+1,i} \lambda_{t+1,i} \left(e^T \mathbf{s}_{t,i} - x_{t,i} \right) \Big) \tag{4.3}$$

In the Bellman equation, the value-to-go term $\mathbb{E}_{\mathcal{D}_{t+1}}\left[J_{t+1}(\mathcal{S}_{t+1}) | \mathcal{S}_t, \mathbf{x}_t\right]$ is obtained by taking the expectation of the optimal value function of the next period over demand arrival, $\mathcal{D}_{t+1}$. The state transition is defined as follows:

$$\mathbf{s}_{t+1,i} = (1 - x_{t,i}) \mathcal{R}_i \mathbf{s}_{t,i} + \left(1 - e^T \mathbf{s}_{t,i} + x_{t,i}\right) \mathbf{d}_{t+1,i}$$

where, $\mathcal{R}_i$ is a $L_i \times L_i$ matrix, with only $r_{j,j+1} = 1$ for all $j = 1, 2, \ldots, L_i - 1$ and all other elements are zero. By multiplying matrix $\mathcal{R}_i$ from left, the allowable delay for the demands of appliance i is decreased by one. Demand arrival on appliance i is represented as $\mathbf{d}_{t+1,i}$, a vector contains either all zeros or $L_i - 1$ zeros and one 1. Recall that the probability for $\mathbf{d}_{t+1,i}$ to be non-zero is the probability of existing a demand arrival, that is, $\lambda_{t+1,i}$, and conditioning on the existence of a new demand, the probability of j-th element of $\mathbf{d}_{t+1,i}$ being 1 is $\mu_{t+1,i,j}$. For future convenience, denote $\mathcal{S}_{t+1} := H(\mathcal{S}_t, \mathbf{x}_t, \mathcal{D}_{t+1})$ as the function that calculates $\mathcal{S}_{t+1}$ from $\mathcal{S}_t$ and decision $\mathbf{x}_t$ following equation (4.2).

The last assumption we make is that the price structures are time-varying but deterministic. If the price functions are also stochastic but linear in the total energy usage, the stochastic prices can be replaced with their first moments. If the price functions can be decomposed into deterministic functions of the total energy usage plus random baselines, then again the random baselines can be replaced with their first moments, as well. But if the price functions are the multiplication of some random multipliers and deterministic functions, the same trick does not work anymore. In this case, the price structures need

to be included into the state status, and the computational complexity increases significantly.

Note that when describing the formulation of our model, we focus on only smart appliances. Nonetheless, our formulation can be generalized to take into account more devices. For example, we can model local generation devices as appliances on which demands are negative and always have allowable delays equal to 1, implying that the generated energy must be stored, used, or sold. To model local storage devices, we need to modify the way we define state status. In particular, the state of storage devices can be modeled as the level of storage. Charging and discharging decisions need to satisfy charging rate constraints and capacity constraints. Moreover, although the model is slightly more complicated after adding storage devices, the effectiveness of the two proposed approaches in this paper will not be affected.

The commonly used solution approach to the above Bellman equation is backward induction. In short, this approach visits all possible state vectors backwards in time to get one optimal solution for the current state. However, the main difficulty in solving this problem is the well-known "curses of dimensionality". For instance, if there are $|I|$ appliances, and each of which has a maximum allowable delay of L, then there will be $(L+1)^{|I|}$ possible state vectors. Therefore, solving this dynamic program for large-scale problems by the backward induction approach is computationally expensive. One approach to deal with the "curses of dimensionality" is to find a way to approximate the value functions. There is a vast of literature on the topic of approximate dynamic programming, as introduced at the beginning of the paper. Another approach is inspired by the idea of Lagrangian relaxation.

A Decentralization Based Heuristic

The main reason that we need to formulate and solve the centralized control problem is the existence of complex price structures. If the price structures are linear in every period, then the central control problem can be decomposed into decentralized ones, in which each appliance makes decisions for itself based on future prices and local information such as outstanding demands, dollar-valued discomforts, and the probabilities of demand arrivals. Obviously, the computational effort spent in solving $|I|$ decentralized control problems is much less than solving one centralized problem when $|I|$ is big. Motivated by this observation, we propose a similar decentralization based heuristic approach, and we will refer to it as *the heuristic* approach in the remainder of this paper for convenience. The heuristic includes the following steps: (1) the central controller decomposes the centralized problem into decentralized ones, (2) each appliance solves for its optimal decisions, (3) then the central controller aggregates the demands for each period and calculates the corresponding realized prices, (4) then it broadcasts the aggregation and new prices to all decentralized problems, and (5) each appliance updates its belief on the equilibrium prices and repeats from step (2), until the equilibrium prices are reached, where the equilibrium prices are defined as the

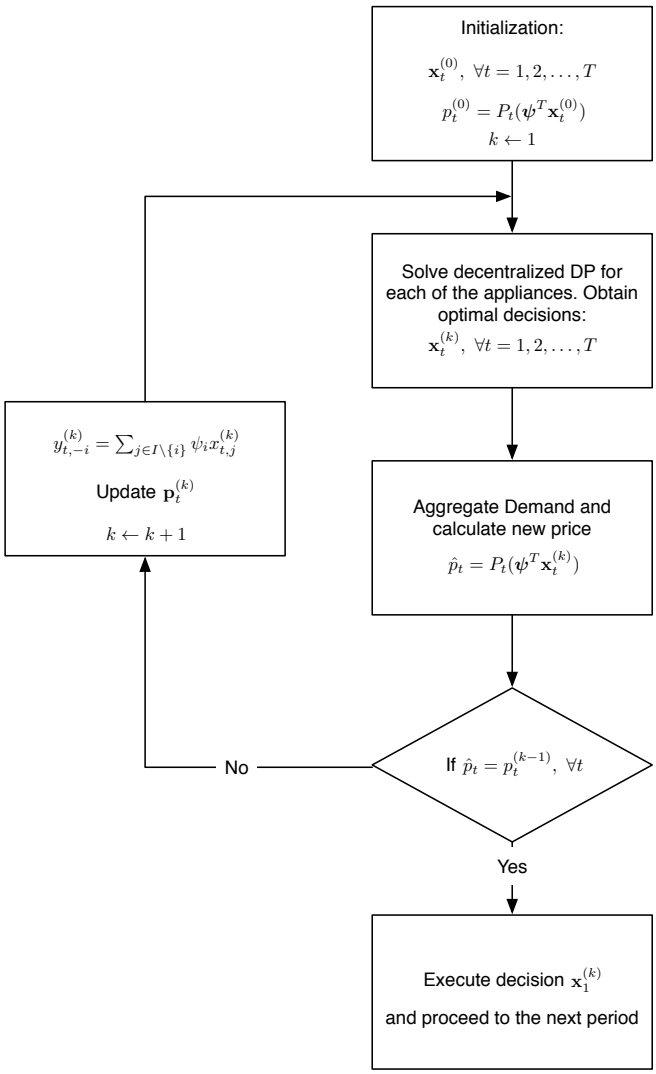


Figure 4.2: Flowchart of the Heuristic Algorithm

prices based on which the optimal decentralized decisions lead to the same prices. Fig. 4.2 summarizes the algorithm in a flowchart.

The detail of this heuristic is described as follows. The first step is to formulate and solve the decentralized dynamic programming. Denote the current total demand from all other appliances as $y_{t,-i}^{(k)}$[1], that is, $y_{t,-i}^{(k)} = \sum_{j \in I \setminus \{i\}} \psi_j x_{t,j}^{(k)}$. Then for each appliance i, the unit price of

[1]Here, (k) indicates that the $y_{t,-i}^{(k)}$ is the sum of energy usage by all appliances except for i in the m-th iteration.

electricity for it to satisfy the outstanding demand is $P_t(y_{t,-i}^{(k)} + \psi_i)$. Therefore, the Bellman equation for appliance i in period t is:

$$\begin{aligned} J_{t,i}(\mathbf{s}_{t,i}) \;=\; \min_{x_{t,i}\in\Omega_{t,i}} \Bigg\{ & P_t\left(y_{t,-i}^{(k)} + \psi_i\right)\psi_i x_{t,i} \\ & + \phi_{t,i}\left(e^T\mathbf{s}_{t,i} - x_{t,i}\right) \\ & + \pi_{t+1,i}\lambda_{t+1,i}\left(e^T\mathbf{s}_{t,i} - x_{t,i}\right) \\ & + \mathbb{E}_{\mathbf{d}_{t+1,i}}\left[J_{t+1,i}(\mathbf{s}_{t+1,i})|\mathbf{s}_{t,i}, x_{t,i}\right] \Bigg\} \end{aligned} \tag{4.4}$$

where

$$\Omega_{t,i} = \{x_{t,i}|x_{t,i} \in \{0,1\}, x_{t,i} \geq \mathbf{s}_{t,i}[1], x_{t,i} \leq e^T\mathbf{s}_{t,i}\}$$

Solving the above Bellman equation is much more time efficient than solving (4.1). Specifically, the number of possible state vectors for each of the DP problem is $L+1$, being much smaller compared to the DP for the centralized control problem, which grows exponentially in the number of appliances. In addition, The optimization for all appliances can be run in parallel to take the advantage of multi-core processors to save even more computational time.

Note that in (4.4), demands from other appliances in period t and all subsequent periods are taken as given, thus so are the prices. In practice, since the appliances make decentralized decisions in parallel, it is impossible to get real-time information on others energy usage decisions. Therefore, we first decompose the problem by breaking the dependence of price on the total demand, then update the prices iteratively towards an equilibrium price vector. Specifically in iteration k, each appliance i starts with an initial belief on the vector of equilibrium prices, $\mathbf{p}^{(k)}$ for all t, according to the most recent information on the energy usage of other appliances, and calculates its own optimal energy usage decisions, $x_{t,i}^{(k)}$, for all t. The new Bellman equation that each appliance i solves iteratively can be written as follows:

$$\begin{aligned} J_{t,i}(\mathbf{s}_{t,i}^{(k)}) \;=\; \min_{x_{t,i}^k\in\Omega_{t,i}^k} \Bigg\{ & p_t^{(k)}\psi_i x_{t,i}^{(k)} \\ & + \phi_{t,i}\left(e^T\mathbf{s}_{t,i}^{(k)} - x_{t,i}^{(k)}\right) \\ & + \pi_{t+1,i}\lambda_{t+1,i}\left(e^T\mathbf{s}_{t,i}^{(k)} - x_{t,i}^{(k)}\right) \\ & + \mathbb{E}_{\mathbf{d}_{t+1,i}}\left[J_{t+1,i}(\mathbf{s}_{t+1,i}^{(k)})|\mathbf{s}_{t,i}^{(k)}, x_{t,i}^{(k)}\right] \Bigg\} \end{aligned} \tag{4.5}$$

where

$$\Omega_{t,i}^k = \{x_{t,i}^k | x_{t,i}^{(k)} \in \{0,1\}, x_{t,i}^{(k)} \geq \mathbf{s}_{t,i}^{(k)}[1], x_{t,i}^{(k)} \leq e^T \mathbf{s}_{t,i}^{(k)}\} \tag{4.6}$$

Then, the controller aggregates the decisions and checks whether the realized prices ($\hat{p}_t = P_t\left(\boldsymbol{\psi}^T \mathbf{x}_t^{(k)}\right)$) equal to the prior belief on the equilibrium prices. If they are different, the new decisions of all appliances $\mathbf{x}_t^{(k)}$ are broadcasted to all appliances, and every appliance updates its belief on the equilibrium prices and gets $\mathbf{p}^{(k+1)}$. Then, they re-optimize by solving the Bellman equations again. There are two ways of updating the belief on prices. The first one is by taking a weighted average as the following:

$$p_t^{(k+1)} = (1 - \alpha^{(k)}) p_t^{(k)} + \alpha^{(k)} \hat{p}_t \tag{4.7}$$

where $\alpha^{(k)}$ is the stepsize used in iteration k. Other ways to update the prices include the following:

$$p_t^{(k+1)} = p_t^{(k)} + \beta^{(k)} \boldsymbol{\psi}^T (\mathbf{x}_t^{(k)} - \mathbf{x}_t^{(k-1)}) \tag{4.8}$$

and

$$p_t^{(k+1)} = P_t((1 - \gamma^{(k)}) \boldsymbol{\psi}^T \mathbf{x}_t^{(k-1)} + \gamma^{(k)} \boldsymbol{\psi}^T \mathbf{x}_t^{(k)}) \tag{4.9}$$

where $\beta^{(k)}$ and $\gamma^{(k)}$ are also stepsizes. Update rule (4.8) is mimic of that for updating the subgradient of the Lagrangian of mixed integer programs[2], and update rule (4.9) is similar to (4.7), with the exception of first taking a weighted average on the total energy consumption and then calculating the prices. All the above rules works well in numerical studies, and we focus on rule (4.7) as it involves the minimum number of operations per iteration.

In theory, if the stepsizes satisfy the following three conditions, namely (1) $\alpha^{(k)} \geq 0, \forall k$, (2) $\lim_{k\to\infty} \alpha^{(k)} \to 0$, and (3) $\lim_{k\to\infty} \sum_{i=1}^k \alpha^{(i)} \to \infty$, then the prices converge in limit. However, since the decisions on satisfying demand in our model (and also on most of the applications) have to be binary, it is not guaranteed that the optimal decentralized decisions will converge to the globally optimal centralized solutions. Plus, convergence in limit does not provide sufficient guideline to practice. In addition in practice, it is necessary to scale the stepsizes by some factor to avoid strong oscillation in convergence due to big stepsizes, and to avoid converging too quickly due to fast diminishing stepsizes. We conduct numerical studies to investigate the convergence of the heuristic algorithm. As will be shown later, the heuristic algorithm converges extremely fast and returns close to optimal objective values.

Q-Learning Approach

Although the heuristic looks very promising, it has several drawbacks. For instance, to formulate and solve the decentralized problems, it is assumed that the marginal distributions

[2]This can be readily seen if we add dummy variables as total consumption in every period, then relax those energy balance constraints.

of demand arrivals are known. However, in practice the distributions are hard to estimate. Even if the distributions are known a priori, it may be computationally intractable to calculate the expectations. Moreover, correlations between demands are not captured by the decentralized heuristic. Although it is possible to join correlated demands into groups and solve one decentralized problem for each group, the benefit of the decentralized heuristic is decreased by doing so, and it remains difficult to obtain joint distributions.

Q-learning, which belongs to the family of approximate dynamic programming approaches ([53]), is a good candidate to help address the above issues. In particular, Q-learning applies a sample-path based approximation approach to estimate the value-to-go of being a specific state and taking a specific decision, $Q_t(\mathcal{S}_t, \mathbf{x}_t)$, which is also known as *Q-factors*. Compared to other post-decision state based approximation approaches, Q-learning avoids subjective assumptions on the parametrization of the value-to-go's, thus it is capable to provide generic and robust solutions to different types of users. In contrary to the traditional backward induction approach that solves for the optimal expected value function for each of the possible state vectors backwards in time, Q-learning updates its estimation on Q-factors via iterative forward loops. In addition, unlike backward induction, Q-learning does not rely on the knowledge of probability distributions. It is also worthwhile to mention that, the complexity of the backward induction approach grows exponentially in the size of the problem, while the complexity of the Q-learning is not an explicit function of the problem's size. This means that when the size of problem is significant such that the backward induction approach is computationally intractable, efficient decision making is achievable via Q-learning, at the cost of sub-optimality. For more detailed description of the Q-learning approach, we refer the reader to [53] and [10].

In our model, the state space and feasible decision space are the same as those in the centralized control model. In each iteration, the Q-learning approach travels forward in time following one sample path of the demand arrival to update the estimations of Q-factors. However when making decision, the controller sees no realization of demand arrival. Specifically, Q-learning in our model works as follows: in period t of iteration k, if the state is $\mathcal{S}_t^{(k)}$, then decision $\mathbf{x}_t^{(k)}$ is obtained by:

$$\mathbf{x}_t^{(k)} = \underset{\mathbf{x}_t \in \mathcal{X}_t(\mathcal{S}_t^{(k)})}{\arg\min} \quad Q_t^{(k-1)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t) \tag{4.10}$$

where $\mathcal{X}_t(\mathcal{S}_t^{(k)})$ is the set of feasible decisions being at state $\mathcal{S}_t^{(k)}$. $Q_t^{(k-1)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t)$'s are the estimations of Q-factors from the $(k-1)$-th iteration. Then, following the k-th sample path of demand arrival, $\mathcal{D}_{t+1}^{(k)}$, the value to being at state $\mathcal{S}_t^{(k)}$ and taking action $\mathbf{x}_t^{(k)}$ is calculated as follows:

$$\begin{aligned}\hat{q} &= C_t(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}) + \sum_{i\in I} U_{t,i}(\mathbf{s}_{t,i}^{(k)}, x_{t,i}^{(k)}) \\ &\quad + V_{t+1}^{(k-1)}(\mathcal{S}_{t+1}^{(k)}|\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}, \mathcal{D}_{t+1}^{(k)}) \\ &= P_t(\boldsymbol{\psi}^T\mathbf{x}_t^{(k)}) \cdot \boldsymbol{\psi}^T\mathbf{x}_t^{(n)} + \sum_{i\in I}\Bigg(\phi_{t,i}\left(e^T\mathbf{s}_{t,i}^{(k)} - x_{t,i}^{(k)}\right) \\ &\quad + \pi_{t+1,i}\lambda_{t+1,i}\left(e^T\mathbf{s}_{t,i}^{(k)} - x_{t,i}^{(k)}\right)\Bigg) \\ &\quad + \ V_{t+1}^{(k-1)}(H(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}, \mathcal{D}_{t+1}^{(k)})) \end{aligned} \tag{4.11}$$

where,

$$\begin{aligned}&V_{t+1}^{(k-1)}(\mathcal{S}_{t+1}^{(k)}|\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}, \mathcal{D}_{t+1}^{(k)}) \\ &= V_{t+1}^{(k-1)}(H(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}, \mathcal{D}_{t+1}^{(k)})) \\ &= \min_{\mathbf{x}_{t+1}\in\mathcal{X}_{t+1}(\mathcal{S}_{t+1}^{(k)})} Q_t^{(k-1)}(H(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}, \mathcal{D}_{t+1}^{(k)}), \mathbf{x}_{t+1})\end{aligned}$$

which is also known as the optimal value-to-go of being at state $\mathcal{S}_{t+1}^{(k)}$. Then, the Q-factor $Q_t^{(k)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)})$ is updated by taking a weighted average of $Q_t^{(k-1)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)})$ and $\hat{q}$:

$$Q_t^{(k)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}) = (1-\alpha^{(k)})Q_t^{(k-1)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}) + \alpha^{(k)}\hat{q} \tag{4.12}$$

where $\alpha^{(k)}$ is the stepsize. Similarly with the decentralized heuristic, stepsizes for Q-learning need to be chosen carefully to avoid over oscillation or converging too quickly. The last step is to update the optimal value-to-go of being at state $\mathcal{S}_t^{(k)}$:

$$V_t^{(k)}(\mathcal{S}_t^{(k)}) = \min_{\mathbf{x}_t\in\mathcal{X}_t(\mathcal{S}_t^{(k)})} Q_t^{(k)}(\mathcal{S}_t^{(k)}, \mathbf{x}_t)$$

One of the problem of forward pass approximate dynamic programming approaches is that it may take a significant number of iterations to propagate the updates of value-to-go in periods close to the end to the beginning periods. On the other hand, propagating the updates is important as the value-to-go's in earlier periods include future costs. To have faster convergence and better decisions, we apply *temporal difference learning*, also known as *TD learning* (see [60], [53] for more details). In particular, *temporal difference* in our problem is defined as the following:

$$\begin{aligned}D_t \ = \ & C_t(\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}) + \textstyle\sum_{i\in I} U_{t,i}(\mathbf{s}_{t,i}^{(k)}, x_{t,i}^{(k)}) \\ & + V_{t+1}^{(k-1)}(\mathcal{S}_{t+1}^{(k)}|\mathcal{S}_t^{(k)}, \mathbf{x}_t^{(k)}, \mathcal{D}_{t+1}^{(k)})\mathcal{S}_t^{(k)} \\ & - V_t^{(k-1)}\mathcal{S}_t^{(k)}\end{aligned} \tag{4.13}$$

Here, the sum of first three terms is the observed value of being in state $\mathcal{S}_t^{(k)}$ while the last term is the corresponding old belief. The temporal difference defined in (4.13) measures the difference between our original estimation of being in state $\mathcal{S}_t^{(k)}$ and the observed value following one sample path. To propagate the difference back to the value-to-go estimations in all previous periods ($\tau < t$), the following step is taken once D_t is obtained:

$$V_\tau^{(k)}(\mathcal{S}_\tau^{(k)}) \;=\; V_\tau^{(k)}(\mathcal{S}_\tau^{(k)}) \;+\; \alpha^{(k)}\lambda^{t-\tau}D_t$$

where λ is the discount factor to reflect the fact that $\mathcal{S}_t^{(k)}$ is one of the possible future outcomes from some state $\mathcal{S}_\tau$ ($\tau < t$), and the probability of $\mathcal{S}_t^{(k)}$ to happen is smaller when τ is farther away from t.

4.3 Numerical Studies of the Control Approaches

We conduct the following controlled experiments to evaluate and compare the performances of the discussed approaches. Specifically, since we do not have real data, we test various combinations of parameters, such as the discomfort from deferring the satisfaction of demand, the discomfort from lost arrivals, demand arrival probabilities, and electricity pricing functions.

We focus on a typical experimental setting to analyze the performance of different approaches. In the experiments, we assume that a single controller manages a household with $I = 3$ appliances with the same maximum allowable delay of $L = 4$ periods. We also assume that the three appliances will consume $\boldsymbol{\psi} = [1, 1, 2]$ units of energy to satisfy one demand. At the beginning of each period, the controller makes energy usage decisions by looking $T = 8$ periods ahead, and we compare the total disutility returned by different approaches over $N = 8$ periods[3]. We use the Monte-Carlo Integration method to estimate the expected total disutlities by repeating the same experiment with 100 samples and the same initial state $S = [3, 1, 0]$, that is, at the beginning of the planning horizon, the first appliance has a demand that should be satisfied within 3 periods; the second appliance has a demand that should be satisfied immediately; and the third appliance does not have any outstanding demand yet.

We assume that demands arrive according to independent Bernoulli distributions, as shown in Fig. 4.3. Given a demand arrival, the demand is equally likely to have the allowable delays for 1 to L periods. If there is an unsatisfied demand, then demand arrival

[3]We also conducted tests on longer planning horizon with more appliances and longer allowable delays. Both the two proposed approaches worked well and delivered solutions close to optimal, however to obtain the exact optimal solutions, backward induction took so much computational resource that we were not able to run enough numerical studies for comparison purposes. Therefore, we limit the testing problem's dimension in our numerical study in this paper.

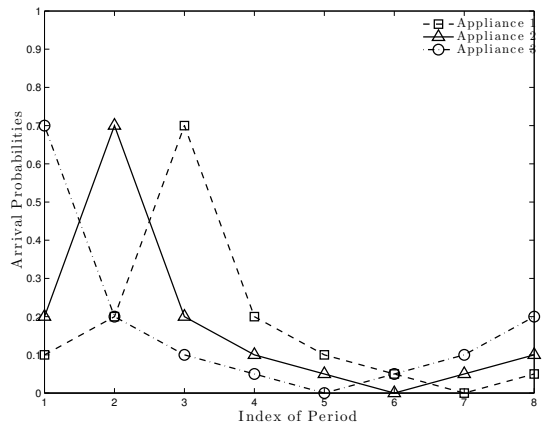


Figure 4.3: A Sample of the Arrival Probabilities

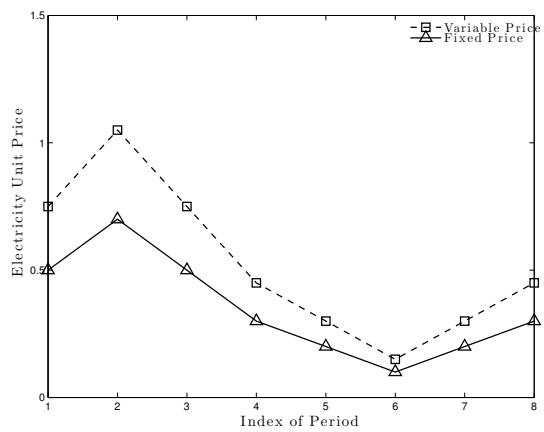


Figure 4.4: A Sample of the Parameters of the Price Structures

on the same appliance will be lost and a penalty in dollar-values for lost arrival is incurred. Similarly, for unsatisfied demands, another kind of discomfort measured in dollar-values for deferrals is charged. In our simulation, different appliances have different discomfort parameters but they are assumed to be time invariant. In particular, we choose the baseline settings of parameters as follows: (1) the arrival probabilities are periodic functions consisting of the following vector $[0.2, 0.7, 0.2, 0.1, 0.05, 0, 0.05, 0.1]$, and (2) the arrival probabilities of different appliances are the same periodic function shifting in time.

Moreover, the price structures are assumed to be time-varying but deterministic as mentioned above. Both linear and quadratic pricing structures are tested in our study. Since the prices are positively correlated with demand arrivals according to the fact that time-varying

prices should reflect real time demands, the unit price of electricity in period t is determined by a function of the total usage in that period. For instance, we choose $P_t(x) = a_t x + c_t$ and $P_t(x) = a_t x^2 + c_t$ as the linear and quadratic price functions, where a_t is the variable price coefficient and c_t is the fixed price coefficient. To facilitate describing our experiments, we apply similar treatment for varying the price structures. In particular, let a_t and c_t be periodic functions consisting of vectors $m_a \cdot [5, 7, 5, 3, 2, 1, 2, 3]$ and $m_c \cdot [5, 7, 5, 3, 2, 1, 2, 3]$, where m_a and m_c are the multipliers to be varied to change the volatility and the amplitude of prices. Fig. 4.4 shows a sample path of a_t's and c_t's.

Convergence Study

Both the heuristic and the Q-learning approaches iteratively obtain better solutions. Naturally, one of the main questions regarding these two approaches is how fast they converge. In this section, we discuss their convergence.

Convergence of the Heuristic

In the implementation of the heuristic, we stop the algorithm by the time either the solutions converge or the maximum allowed number of iterations (i.e. 1000 iterations) is reached. For each iteration k, the algorithm solves decentralized dynamic programs for all of the appliances, by taking the updated belief on prices $p_t^{(k)}$ $(\forall t = 1, 2, \ldots, T)$ from previous iteration as given. Based on the optimal decentralized solution $\mathbf{x}_t^{(k)}$ $(\forall t = 1, 2, \ldots, T)$ of iteration k, we have the corresponding realized price $\hat{p}_t = P_t(\boldsymbol{\psi}^T \mathbf{x}_t^{(k)})$ $(\forall t = 1, 2, \ldots, T)$ from the pricing function, as if we implement the decisions. The new belief on price $p_t^{(k+1)}$ is updated by following one of the updating rules described above. This vector $p_t^{(k+1)}$ $(\forall t = 1, 2, \ldots, T)$ is then passed to iteration $k + 1$.

We evaluate the convergence by measuring $conv(k+1) = \|p_t^{(k+1)} - p_t^{(k)}\|$. Fig. 4.5 shows that the heuristic converges very fast. In our experiment, the result is close to optimal after 400 iterations with proper parameter settings, especially the stepsizes chosen. If the stepsizes are too large, the results exhibits violent oscillations in price difference between iterations as shown in Fig. 4.5. Moreover, the solution might not converge within 1000 iterations. On the other hand, if we choose the stepsizes that diminishing too fast, it will result in a solution that is not optimal. Therefore, it is important to select proper stepsizes to achieve good convergent rate with plausible solution.

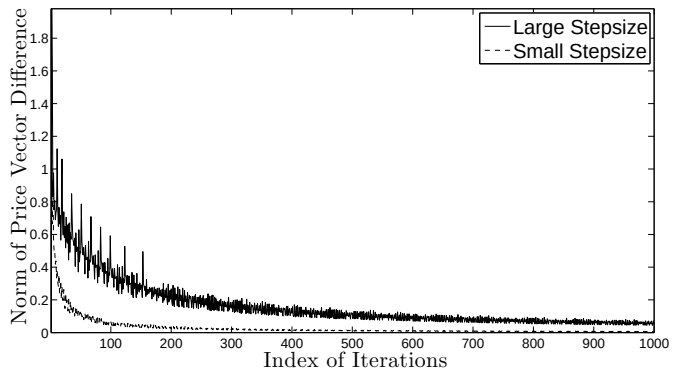


Figure 4.5: Convergence of the Heuristic Approach under Different Stepsizes

Convergence of Q-Learning

As suggested by [53], we also find that the choice of stepsize rules has impact on the convergence and the performance of the Q-learning approach. In our numerical study, we choose simple generalized harmonic stepsizes: $\alpha_k = \dfrac{a}{a+k-1}$. In addition, we scale the stepsizes such that the approximation of value-to-go's does not change dramatically (over 30%) in the first 20 iterations.

Fig. 4.6 presents the absolute difference in the approximation of value-to-go of the given initial state in subsequent iterations. We choose *Boltzmann exploration* rule in the learning process [53]. As shown in the figure, in the first half of the learning process, the algorithm explores the states and updates the approximation of the value-to-go frequently, while in the second half of the learning process, the algorithm exploits the value-to-go approximations and updates based on the corresponding optimal decisions. As a possible future extension, it is interesting to test other stepsize rules, such as the stochastic gradient adaptive rules to study how different rules affect the performance of the Q-learning approach.

Comparison of Different Approaches

As a benchmark, we solve the testing problems using backward induction for optimal solutions, and compare the performance of the heuristic and the Q-learning with the optimal solutions. To show how much better these control approaches are, we add in the performance of the traditional "no-control" case. We test these four approaches under various combinations of parameters, and Table 4.1 lists a selection of it. Specifically, we test both linear and quadratic price structures, and we vary the parameters of the price functions. In addition, we also vary the unit discomfort from deferral and lost arrivals. Table 4.2 summarizes the returned average total cost of electricity and average total disutility, where the

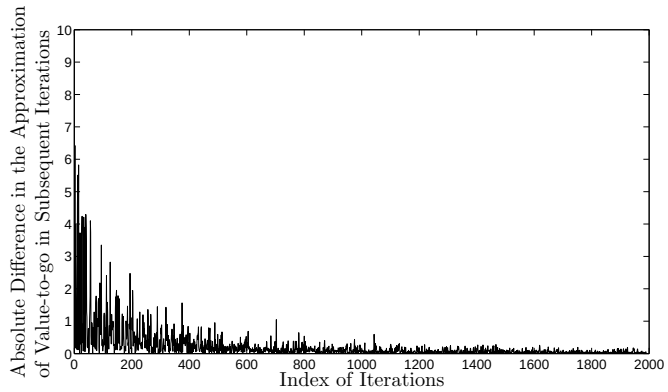


Figure 4.6: Convergence of the Q-Learning Approach

Table 4.1: Experiment Settings of Selected Runs

	Price Structure	Multiplier of Variable Price: m_a	Multiplier of Fixed Price: m_c	Unit Penalty of Lost Arrival: π	Unit Penalty of Deferral: ϕ
Run 1	Linear	0.5	0.2	$[1, 2, 1]$	$[0.1, 0.2, 0.1]$
Run 2	Linear	1	1	$[1, 2, 1]$	$[0.1, 0.2, 0.1]$
Run 3	Linear	1	1	$[5, 10, 5]$	$[1, 2, 1]$
Run 4	Linear	1	1	$[5, 10, 5]$	$[2, 4, 2]$
Run 5	Quadratic	1	1	$[5, 10, 5]$	$[1, 2, 1]$
Run 6	Quadratic	1	1	$[10, 20, 10]$	$[1, 2, 1]$

Table 4.2: Average Total Costs and Average Total Disutilities of Selected Runs

	Backward Induction		Heuristic		Q-Learning		No-Control
	Avg. cost	Avg. disU.	Avg. cost	Avg. disU.	Avg. cost	Avg. disU.	Avg. cost
Run 1	13.21	15.09	13.94	15.89	13.72	15.48	30.32
Run 2	39.52	41.25	40.56	42.52	40.60	42.37	79.16
Run 3	37.26	48.39	41.42	50.71	39.38	49.68	79.16
Run 4	43.40	56.24	48.90	58.67	42.94	56.43	80.18
Run 5	56.54	67.51	65.98	74.73	63.28	72.13	164.36
Run 6	61.50	74.51	66.40	79.56	62.98	77.08	147.14

total disutility is the sum of total cost and total discomfort. For the no-control case, there is no discomfort; therefore total cost is the total disutility.

It can be seen from Table 4.2 that the performance of either the heuristic or the Q-learning approach is close to optimal, with the average total disutilities roughly equaling to half of that of the no-control case. In particular, we can first analyze the effect of increasing prices from Run 1 and Run 2. In these two cases, the discomfort remains roughly the same when prices are increased, while costs spent by the backward induction, the heuristic and the Q-learning approaches increase. This is because in both cases, the cost of electricity

outweighs the discomfort from either deferring or lost arrivals, and further increase in prices does not have significant impact on the decisions.

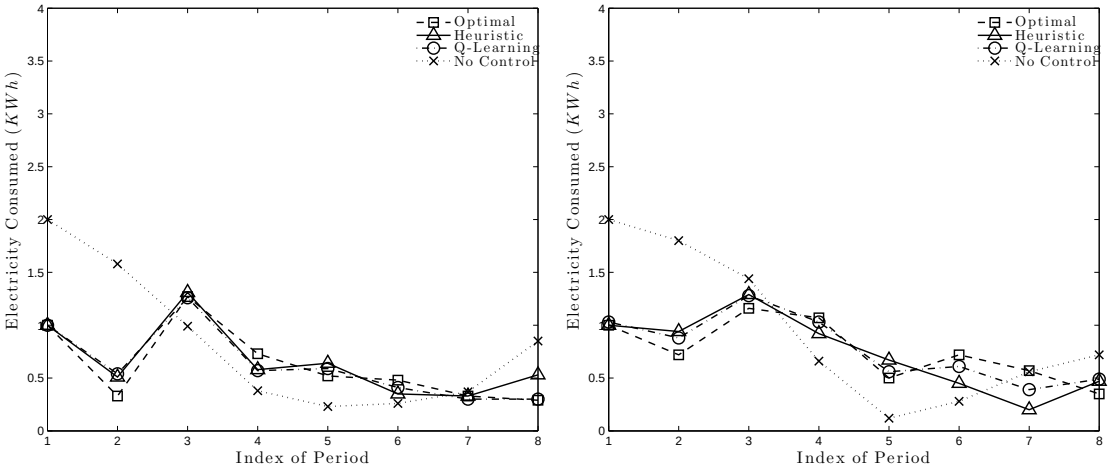
Compared to Run 2, Run 3 has higher discomfort per deferred demand and per lost arrival. In other word, Run 3 represents the case in which users are more sensitive to service level. It can be noticed that when discomfort per deferred demand is increased from the Run 2 to Run 3, although the total discomfort for each of the three control approaches increases, the total costs remain similar. This suggests that cost still dominate discomfort in Run 3. The main reason for this to happen is that the decisions are discrete, and in order for the control approaches to make different decisions, the discomfort per deferred demand of lost arrival has to be greater than some threshold. From Run 3 to Run 4 as we keep increasing the users' sensitivities on services, the total costs increase, while the total disutilities increase as well. This implies that as users becoming more sensitive to service, there is less load shifting. If we calculate the average total discomfort in Run 4, we can notice that the discomfort decreases from Run 3 to Run 4, although the unit discomfort increases, which verifies that more new demands are satisfied immediately.

As we change from linear price structures to quadratic structures, from Run 3 to Run 5, although **a** and **c** remain the same, the realized prices of the quadratic structure is higher for the same amount of usage. As a result, when quadratic prices are applied, the energy consumption profiles should be flatter. This can be seen from Fig. 4.7. Fig. 4.7 plots the energy consumption profiles for Run 3 and Run 5, the comparison verifies that quadratic functions leads to more load shifting and results in smoother consumption profiles. The insight here is that to overcome rebounds, steeper price structures can be applied. It is better than forcing all users to pay fixed higher rates (which are still time-varying), as higher rates lead to inefficient allocation of welfare.

Some insights from our simulation studies can be summarized as follows: (1) both the heuristic and the Q-learning approaches are able to shift demands and generate near optimal solutions, while consuming only limited computational resources, (2) the consumption profiles generated by the heuristic and the Q-learning approaches are different for different types of users, and the heterogeneity of users can be applied effectively by the control approaches to further smooth out peak demands, and (3) steeper price structures can be used to effectively overcome rebound effects.

Time Study of the Approaches

In this section, we compare the average CPU time consumed by Q-learning, the heuristic and backward induction approaches. The approaches are implemented in MATLAB R2009a



(a) Linear Price Structure (b) Quadratic Price Structure

Figure 4.7: Average Energy Consumption Profiles under Different Price Structures

Table 4.3: Time Study Summary (in seconds)

	$\|I\| = 3; L = 4$	$\|I\| = 3; L = 6$	$\|I\| = 4; L = 4$
Q-learning	28.83	28.98	29.43
Heuristic	4.51	7.15	6.07
Backward Induction	77.87	717.27	3639.50

(7.8.0.347) with Intel(R) Core(TM) i7 CPU 3.07 GHZ processor and 24.0 GB RAM[4]. We vary the number of appliances and the maximum allowed delay for each run with 100 replicates in the study. The results are summarized in Table 4.3:

Table 4.3 demonstrates that the heuristic outperforms other approaches in terms of average computation time in all three cases. The time grows almost linearly as the increase in the number of appliances and maximum allowed delay. It is because that, in each iteration, the heuristic solves $|I|$ decentralized dynamic programs for all appliances and each decentralized dynamic program has dimension of L. On the other hand, Q-learning approach shows stable performance among all three tested cases. Increase in number of appliances and maximum allowed delay has relatively small impact to Q-learning's performance in computation time. The main reason is that Q-learning approach solves the problem in $O(T)$ time per iteration and the number of iterations is fixed to be 2000 in our study. Lastly, as expected, the backward induction approach takes the longest time in computation. In addition, its computation time grows almost exponentially as the number of appliances and maximum allowed delay grows.

[4]To compare the consumption of computational resource, we did not parallelize the heuristic in this time study. However, it is worthwhile to note that the heuristic has the potential to achieve faster computation via parallelization.

4.4 Summary

In this paper, we study the energy usage control problem for Smart Grid users, who faces time-varying electricity prices. In particular, we formulate the stochastic control problem as a dynamic program, based on the assumptions that in the Smart Grid, users can specify the allowable delay for flexible demands and a central controller optimally schedules the time to satisfy those demands. Under some conditions on the uncertain information, the problem can be solved optimally by utilizing traditional backward induction approach.

However, the backward induction approach encounters the "curses of dimensionality" for large problems. Therefore, we aim to develop other efficient approaches for this problem. One is a decentralization-based heuristic that turns the centralized control problem into decentralized small sub-problems, and uses backward induction to solve each of them. Then, the decisions of each of the sub-problems are aggregated together, and prices, which are used as input for the sub-problems, are updated for primal feasibility. This heuristic works iteratively towards an equilibrium solution. The heuristic is numerically proved to be efficient and effective. Nonetheless, it also has several drawbacks. Therefore, we develop another alternate approach based on Q-learning. As an approximate dynamic programming approach, the Q-learning is also able to address the "curses of dimensionality". Our simulation study also demonstrates the effectiveness of the Q-learning approach. The potential problem of the Q-learning is that the dimension of Q-factors grows very fast in the size of the problem, and for big problems more iterations may be required. Therefore, the heuristic is potentially better for big problems.

Therefore, each of the two proposed approaches has some advantages over the other. The Q-learning approach works under more general settings, while the heuristic is able to deliver solutions in a much faster manner for regular sized problems (for example, at household level). These approaches are by no means the best for the control problem of Smart Grid users. As future extensions, it will be interesting to compare the performance of these approaches to others, such as the post-decision state based approximate dynamic programming. On the other hand, our approaches can be used as modules to analyze the pricing strategy in the Smart Grid. Last but not least, since users are in general risk-averse in costs, robust solutions and related robustness analysis under price and demand uncertainties can help better understand the pricing strategies in the Smart Grid and encourage the adoption of demand response mechanisms.

Appendix A

Appendix for Chapter 2

A.1 Comparison of Different Policies for $\mathbf{P^{S_d}}$

First of all, we present an example in which the two families of policies (rationing policies and threshold policies) are equivalent. Consider the case in which there are two periods. The unit prices of electricity are p_1 and p_2, with $p_1 > p_2 > 0$. The unit delay penalty is c, and $c > 0$. Demand are sampled from two *i.i.d.* uniform random variables, that is, $d_1, d_2 \sim Unif[0, 1]$. To find all pareto optimal solutions, we first re-formulate the problem as the following problem ($\mathbf{P_s^{U_D}}$)

$$
\begin{aligned}
(\mathbf{P_s^{U_d}}): \quad & \min_{\mathbf{x(d)},\mathbf{u(d)}} \quad \mathbb{E}\left[\sum_{i\in\{1,2\}} c_i(x_i(\mathbf{d}) - u_i(\mathbf{d}))\right] \\
& s.t. \quad \text{Constraints (2.2a) - (2.2e)}
\end{aligned}
$$

Then, vary C of the first constraint and for each C find the optimal solution, which will the pareto optimal solution for the original multi-objective problem.

Suppose we want to find the optimal threshold policy. Let the threshold for the first period be k, then $u_1 = d_1$ if $d_1 \leq k$, and $u_1 = k$ if $d_1 > k$. Then the decision variable of problem ($\mathbf{P_s^{U_d}}$) is converted to k. Obviously, as $C \geq \dfrac{p_1 + p_2}{2}$, the optimal objective function value is 0, and the optimal threshold is $k^* = 1$.

When $p_2 \leq C < \dfrac{p_1 + p_2}{2}$, the problem is feasible, but $k = 1$ is no longer a feasible solution. observe that when the threshold is k, the expected demand served in period 1 is $\mathbb{E}_{d_1}[u_1] = k_1 - \dfrac{k_1^2}{2}$, and the expected shifted demand that arrives in the first period is $\mathbb{E}_{d_1}[d_1 - u_1] = \dfrac{(1-k_1)^2}{2}$. In addition, let $C = \dfrac{p_1 + p_2}{2} - B$, where B denotes the shortage

in budget to have $k = 1$, then the optimization problem is the following:

$$\begin{aligned} \min_{k \geq 0} \quad & c\frac{(1-k_1)^2}{2} \\ s.t. \quad & p_1\left(k_1 - \frac{k_1^2}{2}\right) + p_2\left(\frac{(1-k_1)^2}{2} + \mathbb{E}_{d_2}[d_2]\right) \leq \frac{p_1+p_2}{2} - B \end{aligned}$$

The optimal solution of the above problem is $k^* = \dfrac{p_1 - p_2 - \sqrt{2B(p_1-p_2)}}{p_1 - p_2}$. And the penalty raised by the shortage B is:

$$\begin{aligned} c\frac{(1-k_1)^2}{2} &= \frac{c}{2}\left(1 - \frac{p_1 - p_2 - \sqrt{2B(p_1-p_2)}}{p_1-p_2}\right)^2 \\ &= \frac{B}{p_1 - p_2} \end{aligned}$$

That is, the increased penalty raised by B is proportional the the value of B itself. Obviously, the same objective function value can be achieved by the rationing policy. In particular, shifting the $\dfrac{2B}{p_1 - p_2}$ of the demand that arrives in the first period is a feasible rationing policy. In fact, it is trivial that this is also the optimal solution. Therefore, in this case, both the optimal threshold policy and the optimal rationing policy find the pareto optimal solutions for the two-period multi-objective problem.

A.2 Proofs

Proof. Proof of Lemma 1: All pareto optimal solutions that form the efficient frontier of problem $(\mathbf{P^D})$ can be obtained by solving the following problem by varying the right hand side parameter C in constraint (A.1).

$$\begin{aligned} (\widetilde{\mathbf{P^D}}): \qquad U(C) \;=\; \min_{\mathbf{x},\mathbf{u}} \quad & \sum_{i \in N} c_i(x_i - u_i) \\ s.t. \quad & \sum_{i \in N} p_i u_i \leq C \qquad \text{(A.1)} \\ & \text{Constraints (2.1a) - (2.1e)} \end{aligned}$$

where constraint (A.1) is the budget constraint on cost. Note that problem $(\widetilde{\mathbf{P^D}})$ is essentially a linear program. Since the objective function value is convex in the right hand side vector for linear programs, $U(C)$ is (picewise-liearly) convex in C.

□

Proof. Proof of Lemma 2 Note that first we can interchange the expectation and summation operator in the objective. Then to find all pareto optimal solution, we can set the first objective as a capacity constraint. Then the problem with stochastic demand is as follows:

$$\begin{aligned} \min_{\mathbf{x}(\mathbf{d}),\mathbf{u}(\mathbf{d})} \quad & \sum_{i\in N} c_i \mathbb{E}_{\mathbf{d}}\left[(x_i(\mathbf{d}) - u_i(\mathbf{d}))\right] \\ s.t. \quad & \sum_{i\in N} p_i \mathbb{E}_{\mathbf{d}}\left[u_i(\mathbf{d})\right] \leq C \\ & \text{Constraints (2.2a) - (2.2e)} \end{aligned} \tag{A.2}$$

With the rationing policies, $\mathbb{E}_{\mathbf{d}}[u_i]$ can be expressed as proportions of $\mathbb{E}_{\mathbf{d}}[x_i]$, where the proportions are the decision variables. In addition, $\mathbb{E}_{\mathbf{d}}[x_i]$ can be expressed as linear functions of $\mathbb{E}_{\mathbf{d}}[d_j]$, for $j = 1, 2, \ldots, i$. By replacing $\mathbb{E}_{\mathbf{d}}[x_i]$'s and $\mathbb{E}_{\mathbf{d}}[d_j]$'s with the decisions about proportions and $\mathbb{E}_{\mathbf{d}}[d_i]$'s, the above problem can be converted to one similar to $(\widetilde{\mathbf{P^D}})$, which is a LP. Then, it is trivial that constraint (A.2) is binding at optimal solution.

On the other hand, note that for each feasible policy U of the other families, there exists one corresponding rationing policy U^R, such that the expected amount of demand to be satisfied in each period decided by the two policy equal. This can be easily achieved by first calculating the expected amount of demand to be satisfied by U, then divided by the amount of expected outstanding demand at the beginning of each period. Obviously, U^R is a feasible rationing policy, and the objective function returned by U^R is bounded below by the optimal objective function value returned by the above optimal rationing policy.

□

Proof. Proof of Lemma 3 Notice that the main difficulty in the above formulation is the dependence of $u_i(\mathbf{d})$ on x_i, raised by the uncertainty of $\mathbf{d}$. Lemma 2 shows that there exists at least one rationing policy corresponds to each of the pareto optimal solution. Therefore, the proof for Lemma 3 consists of two steps: 1) rewrite $(\mathbf{P^{S_d}})$ using the rationing decisions, and 2) prove the efficient frontier is convex based on the new formulation.

To re-formulate the problem, we first define decision variable y_{ij} as the portion of demand arrives in period i to be satisfied in period j. Then one more set of constraints $\sum_{j=i}^{n} y_{i,j} = 1$, needs to be added. Then, the expected discomfort is $\mathbb{E}_{\mathbf{d}}\left[\sum_{i\in N} d_i \sum_{j=i}^{n} y_{ij} \left(\sum_{k=i}^{j-1} c_k\right)\right]$, and the expected cost is $\sum_{j=1}^{n} p_j \sum_{i=1}^{j} y_{ij} \mathbb{E}_{\mathbf{d}}[d_i]$.

Similar to the deterministic version of the problem, we put a cap C on the expected cost, and solve for the minimum expected discomfort for all feasible caps to find all pareto optimal solutions. After rewriting $u_i(\mathbf{d})$ and $x_i(\mathbf{d})$ using y_{ij}, it is readily seen that varying the cap C is again equivalent to changing the right hand side vector of a Linear Programming. Therefore, the efficient frontier is convex, following from similar argument made in the proof of Lemma 1.

□

Proof. Proof of Proposition 1 The Bellman equation for the problems is:

$$\begin{aligned} V_i(x_i) &= \min_{u_i} \quad p_i u_i + c_i\left(x_i - u_i\right) + \mathbb{E}_{d_{i+1}}\left[V_{i+1}(x_{i+1})\right] \\ & \quad s.t. \quad x_{i+1} = x_i - u_i + d_{i+1} \\ & \qquad\quad u_i \leq x_i \\ & \qquad\quad u_i \geq 0 \end{aligned}$$

where, $V_n(x_n) = p_n x_n = \Gamma_n x_n$, and:

$$\begin{aligned} V_{n-1}(x_{n-1}) &= \min_{u_{n-1} \in \mathcal{U}_{n-1}} \left\{p_{n-1}u_{n-1} + c_{n-1}\left(x_{n-1} - u_{n-1}\right) + \mathbb{E}_{d_n}\left[V_n(x_n)\right]\right\} \\ &= \min_{u_{n-1} \in \mathcal{U}_{n-1}} \left\{p_{n-1}u_{n-1} + c_{n-1}\left(x_{n-1} - u_{n-1}\right) + \mathbb{E}_{d_n}\left[p_n(x_n)\right]\right\} \\ &= \min_{u_{n-1} \in \mathcal{U}_{n-1}} \left\{p_{n-1}u_{n-1} + c_{n-1}\left(x_{n-1} - u_{n-1}\right) + \mathbb{E}_{d_n}\left[p_n(x_{n-1} - u_{n-1} + d_n)\right]\right\} \\ &= \min_{u_{n-1} \in \mathcal{U}_{n-1}} \left\{p_{n-1}u_{n-1} + c_{n-1}\left(x_{n-1} - u_{n-1}\right) + p_n(x_{n-1} - u_{n-1}) + \mathbb{E}_{d_n}\left[p_n(d_n)\right]\right\} \\ &= \min_{u_{n-1} \in \mathcal{U}_{n-1}} \left(p_{n-1} - c_{n-1} - p_n\right) u_{n-1} + \varphi \end{aligned}$$

where φ is a constant. It can be verified that the optimal $u_{n=1}$ for the above minimization problem satisfies:

$$u_{n-1} = \begin{cases} x_{n-1} & \text{if } \; p_{n-1} \leq c_{n-1} + \Gamma_n \\ 0 & \text{if } \; p_{n-1} > c_{n-1} + \Gamma_n \end{cases}$$

Therefore, $V_{n-1}(x_{n-1}) = \Gamma_{n-1}x_{n-1} + \mathbb{E}_{d_n}\left[\Gamma_n(d_n)\right]$. The rest of the proof follows mathematical induction steps and is trivial, and hence is omitted here.

Note that previous Lemmas prove that the efficient frontier is piece-wise linearly convex. Therefore, for any coefficient w that joins the two objectives into one, there are infinite number of optimal solutions. In particular, the vertices corresponds to the AON policy, that is, the solutions obtained according to this AON policy is optimal but not unique. Moreover, the linearity of the efficient frontier indicates that moving along the line segments on the frontier, reduction of the expected cost is achieved by gaining more discomfort at the same rate.

□

Proof. Proof of Proposition 2 For any instance of the demand distribution F_d, the objective minimizing policy is an AON policy. Moreover, the optimal AON policy $\mathbf{u}^*(\mathbf{p}, \mathbf{c})$ is only function of $(\mathbf{p}, \mathbf{c})$ based on Proposition 1. Therefore, without altering $\mathbf{p}$ and $\mathbf{c}$, modifying the distribution of demand will not affect the optimal policy.

Therefore, as the same policy remains optimal for any of the demand distribution $F_d \in \mathcal{F}_d$, the optimal solution to problem $(\mathbf{R} - \mathbf{P2})$ is the same AON policy.

□

Proof. Proof of Lemma 4 To find all pareto optimal solutions, we put a cap C on the expected cost, then find the minimum expected discomfort for each feasible C.

According to our assumption on the sequence of event, the decision is made before seeing the realization of prices. If prices are intertemporally dependent, then based on our assumption, the prices take on a finite number of possible values and the joint distributions are know. We can write the expected cost as the convex combination of the products of prices and the corresponding decision variables of each of the possible price vectors. Because the prices take on a finite number of possible outcomes, the number of such decision variables $(u'_(\bar{\mathbf{p}}_{(i-1)}), x'_i(\bar{\mathbf{p}}_{(i-1)}))$ for all i is finite. Then, after expanding the expected discomfort in a similar way, the problem is a linear program. Following from the same argument as in the proofs of Lemma 1 and Lemma 3, the efficient frontier is convex.

If prices in different periods are independent, each decision should be independent with the realized prices in previous periods. Then, interchanging the expectation and summation operators and based on the independence explained above, we have the following new formulation for the new budget constrained problem as follows.

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \sum_{i \in N} c_i(x_i - u_i) \\ s.t. \quad & \sum_{i \in N} \mathbb{E}_{\mathbf{p}}\left[p_i\right] u_i \leq C \\ & \text{Constraints (2.1a) - (2.1e)} \end{aligned}$$

Then it is readily seen that the objective function is convex in C. Hence the efficient frontier is convex.

□

Proof. Proof of Proposition 3 The proof can be obtained by directly replacing the deterministic prices $\mathbf{p}$ in the proof for Proposition 1 with the expected prices $\mathbb{E}_{\mathbf{p}}[\mathbf{p}]$, hence is omitted here.

□

Proof. Proof of Proposition 4 The proof follows directly from the derivation of [12]. First of all, problem ($\mathbf{R_0^{S_P}}$) is equivalent to the following problem, in which we define a dummy variable Y, and let Y be the upper bound on $\sum_{i \in N} p_i u_i$:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & Y + \sum_{i \in N} (c_i(x_i - u_i)) \\ s.t. \quad & \sum_{i \in N} p_i u_i \leq Y \\ & \text{Constraints (2.1a) - (2.1e)} \end{aligned} \tag{A.3}$$

Then, the left hand side (LHS) of constraint (A.3) can be re-written as:

$$\sum_{i \in N} \bar{p}_i u_i + \max_{\{\mathcal{S} \cup \{t\} | \mathcal{S} \subseteq N, t \in N \setminus \mathcal{S}, \mathcal{S}| = \lfloor \Gamma \rfloor\}} \left\{ \sum_{i \in \mathcal{S}} \hat{p}_i y_i + (\Gamma - \lfloor \Gamma \rfloor) \hat{p}_t y_t \right\} \tag{A.4}$$

where, $\mathcal{S} \cup \{t\}$ is the set of periods in which prices p_i's differs from $\bar{p}_i$'s. Parameter Γ, as defined, controls the robustness. y_i's satisfy $-y_i \leq u_i \leq y_i$ and $y_i \geq 0$. It is desired to get rid of the maximization operation in the constraints. Note that the maximization problem in (A.4) equals to the objective function of the following linear programming:

$$\begin{aligned} \max_{\mathbf{z}} \quad & \sum_{j=1}^{n} \hat{p}_j |u_j| z_j \\ s.t. \quad & \sum_{j=1}^{n} z_j \leq \Gamma & (\lambda) \\ & 0 \leq z_j \leq 1, \quad \forall j & (\rho_j) \end{aligned}$$

Taking the dual of the above LP yields a minimization linear programming, whose objective aligns with the objective of the original problem. Let the dual variables be λ and $\boldsymbol{\rho}$, the dual problem is:

$$\begin{aligned} \min_{\lambda, \boldsymbol{\rho}} \quad & \Gamma \lambda + \sum_{j=1}^{n} \rho_j \\ s.t. \quad & \lambda + \rho_j \geq \hat{p}_j |u_j|, \quad \forall j \\ & \lambda \geq 0 \\ & \boldsymbol{\rho} \geq 0 \end{aligned}$$

Based on weak duality, any feasible solution to the dual problem returns an objective function value greater than the optimal objective value of the primal problem. In addition,

since the objective of the original problem is minimization, constraint (A.3) can be replaced by the following set of constraints:

$$\begin{aligned}
\Gamma\lambda + \sum_{j=1}^{n} \rho_j + \sum_{j=1}^{n} \bar{p}_j u_j &\leq Y \\
\lambda + \rho_j &\geq \hat{p}_j y_j \quad \forall j \\
y_j &\geq |u_j| \quad \forall j \\
\lambda \geq 0,\ \boldsymbol{\rho} &\geq 0,\ \mathbf{y} \geq 0 \\
0 \leq u_j &\leq x_j \quad \forall j
\end{aligned}$$

Note that u_j's are forced to be positive. Thus, the robust LP, with stochastic prices, is:

$$\begin{aligned}
(\mathbf{RLP} - \mathbf{P}^{\mathbf{S_P}}) \quad \min \quad & Y + \sum_{i \in N} [c_i(x_i - u_i)] \\
s.t. \quad & x_1 = d_1 \\
& x_{i+1} = x_i - u_i + d_{i+1} \\
& \Gamma\lambda + \sum_{j=1}^{n} \rho_j + \sum_{j=1}^{n} \bar{p}_j u_j - Y \leq 0 \\
& \lambda + \rho_j \geq \hat{p}_j u_j \quad \forall j \\
& \lambda \geq 0,\ \boldsymbol{\rho} \geq 0 \\
& 0 \leq u_j \leq x_j \quad \forall j
\end{aligned}$$

□

Proof. Proof of Proposition 5 The first step is to take the dual of the original problem to get rid of the expectation over the unknown distribution $F_{\mathbf{p}}$. Let the Lagrangian multiplier corresponding to constraints (2.6) and (2.7) be $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$, respectively. The Lagrangian of the inner maximization problem is:

$$\begin{aligned}
&\max_{F_{\mathbf{p}}} \Bigg\{ \mathbb{E}_{F_{\mathbf{p}}} [T(\mathbf{p}, \mathbf{c}, \mathbf{d})] + \theta(1 - \int_{\mathbb{R}^n} dF_{\mathbf{p}}(\mathbf{p})) \\
&\qquad + \sum_{i \in N} \rho_i(\mu_i - \int_{\mathbb{R}^n} p_i dF_{\mathbf{p}}(\mathbf{p})) + \sum_{i \in N} \eta_i(\mu_i^2 + \sigma_i^2 - \int_{\mathbb{R}^n} p_i^2 dF_{\mathbf{p}}(\mathbf{p})) \Bigg\} \\
&= \max_{F_{\mathbf{p}}} \int_{\mathbb{R}^n} \left[T(\mathbf{p}, \mathbf{c}, \mathbf{d}) - \theta + \sum_{i \in N} \left(-\rho_i p_i - \eta_i p_i^2 \right) \right] dF_{\mathbf{p}}(\mathbf{p}) + \theta + \sum_{i \in N} \rho_i \mu_i + \sum_{i \in N} \eta_i(\mu_i^2 + \sigma_i^2)
\end{aligned}$$

For primal feasibility, we need the integrant of the integral to be non-positive. Thus, the dual of the original problem is:

$$\begin{aligned} \min_{\theta,\boldsymbol{\rho},\boldsymbol{\eta}} \quad & \theta + \sum_{i\in N} \rho_i \mu_i + \sum_{i\in N} \eta_i(\mu_i^2 + \sigma_i^2) \\ s.t. \quad & \max_{\mathbf{p}\in\mathbb{R}^n} \left\{ \min_{(\mathbf{x},\mathbf{u})\in\mathbf{X_p}} T(\mathbf{p},\mathbf{c},\mathbf{d}) + \sum_{i\in N} \left(-\rho_i p_i - \eta_i p_i^2\right) \right\} \leq \theta \end{aligned}$$

where, recall that $\min_{(\mathbf{x},\mathbf{u})\in\mathbf{X_p}} T(\mathbf{p},\mathbf{c},\mathbf{d})$ is the following linear programming:

$$\begin{aligned} (\mathbf{P1}): \quad T(\mathbf{p},\mathbf{c},\mathbf{d}) \;=\; \min_{\mathbf{x},\mathbf{u}} \quad & \sum_{i\in N} [p_i u_i + c_i(x_i - u_i)] \\ s.t. \quad & x_1 = d_1 && (\lambda_1) \\ & x_{i+1} - x_i + u_i = d_{i+1} \qquad \forall i = 1,2,\ldots,n-1 && (\lambda_{i+1}) \\ & u_n - x_n = 0 && (\lambda_{n+1}) \\ & u_i \leq x_i \qquad \forall i = 1,2,\ldots,n && (\psi_i) \\ & u_i \geq 0 \qquad \forall i = 1,2,\ldots,n \end{aligned}$$

In addition, strong duality holds for this problem, based on Theorem 2.2, [11]. Therefore, solving problemm $(\mathbf{R^{S_P}-1})$ is equivalent to solving its dual problem. Note that the constraint of the dual problem is a max-min problem itself. To align the two objectives, we first take the dual of the inner minimization linear programming. The dual of $(\mathbf{P1})$ is:

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\psi}} \quad & \sum_{i\in N} \lambda_i d_i \\ s.t. \quad & \lambda_i - \lambda_{i+1} - \psi_i = c_i \qquad \forall i = 1,\ldots,n \\ & \lambda_{i+1} + \psi_i \leq p_i - c_i \qquad \forall i = 1,\ldots,n \\ & \psi_i \leq 0 \qquad \forall i = 1,\ldots,n \end{aligned}$$

Let the projected unit cost for d_i, which arrives in period i, be the unit cost and dollar value of discomfort per unit of demand incurred to d_i. Then, by examining the objective of the dual, we can tell intuitively the optimal dual decision variables $\boldsymbol{\lambda}^*$ is the projected price for satisfying the demand at optimal. For instance, it is trivial that $\lambda_n^* = p_n$. Plug into the dual problem, it is not hard to see that λ_{n-1} needs to satisfy:

$$\lambda_{n-1} - p_n - \psi_n = c_{n-1}$$

and since $p_n + \psi_{n-1} \leq p_{n-1} - c_{n-1}$, we have the following two scenarios:

- if $p_n + c_{n-1} \leq p_{n-1}$, then intuitively, it suggests that decision makers are better-off if the outstanding demand of period $n-1$ is delayed to be satisfied in period n. On the other hand, the optimal dual variables are obtained by solving the following problem:

$$\max_{\psi_{n-1}\leq 0} p_n + \psi_{n-1} + c_{n-1} = \lambda_{n-1}^* = p_n + c_{n-1}$$

That is, the optimal projected unit cost for d_{n-1} is the unit cost of electricity in period n, plus the unit discomfort for deferring the demand from period $n-1$ to period n.

- if $p_n + c_{n-1} > p_{n-1}$, then let $\psi^*_{n-1} < 0$ such that $p_n + \psi^*_{n-1} = p_{n-1} - c_{n-1}$, then

$$\max_{\psi_{n-1} \leq 0} p_n + \psi_{n-1} + c_{n-1} = \lambda^*_{n-1} = p_{n-1}$$

which suggests that the projected unit cost for d_{n-1} is the unit cost of electricity in period $n-1$.

This backward induction can be repeated for all λ_i, and the remainder is omitted here. After taking the dual of the LP, the two objectives in the constraint of the dual of the original problem can be combined together. The combined LHS of the constraint is formed by the following optimization problem:

$$\begin{aligned}
\max_{\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\psi}} \quad & \sum_{i \in N} \left(\lambda_i d_i - \rho_i p_i - \eta_i p_i^2\right) \\
s.t. \quad & \lambda_i - \lambda_{i+1} - \psi_i = c_i \qquad \forall i = 1, \ldots, n && \text{(A.5)} \\
& \lambda_{i+1} + \psi_i \leq p_i - c_i \qquad \forall i = 1, \ldots, n && \text{(A.6)} \\
& \psi_i \leq 0 \qquad \forall i = 1, \ldots, n
\end{aligned}$$

Note that the objective is concave in its decision variables, and the constraints are linear, therefore, strong duality holds for this optimization problem. Assigning lagrangian multipliers $\mathbf{x}$ and $\mathbf{u} \geq \mathbf{0}$ to constraints (A.5) and (A.6), we can write the lagrangian of the above optimization problem as:

$$\begin{aligned}
\max_{\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\psi} \leq \mathbf{0}} \quad & \sum_{i \in N} \left(\lambda_i d_i - \rho_i p_i - \eta_i p_i^2\right) + \sum_{i \in N} x_i \left(c_i - \lambda_i + \lambda_{i+1} + \psi_i\right) \\
& + \sum_{i \in N} u_i \left(p_i - c_i - \lambda_{i+1} - \psi_i\right) \\
= \max_{\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\psi} \leq \mathbf{0}} \quad & \sum_{i \in N} \left(-\eta_i p_i^2 + (u_i - \rho_i) p_i\right) + (d_1 - x_1)\lambda_1 + \sum_{i=2}^{n} \left(d_i - x_i + x_{i-1} - u_{i-1}\right) \lambda_i \\
& + (x_n - u_n)\lambda_{n+1} + \sum_{i \in N} \left(x_i - u_i\right) \psi_i + \sum_{i \in N} \left(x_i - u_i\right) c_i
\end{aligned}$$

Since $\boldsymbol{\lambda}$ is free and $\boldsymbol{\psi} \leq \mathbf{0}$, for primal feasibility, we need $(\mathbf{x}, \mathbf{u}) \in \mathbf{X}$, where $\mathbf{X}$ is defined as:

$$\mathbf{X} = \left\{ (\mathbf{x}, \mathbf{u}) \;\middle|\; \begin{array}{c} x_1 = d_1;\; x_{i+1} - x_i + u_i = d_{i+1}, \quad \forall i = 1, \ldots, n-1; \\ u_n - x_n = 0;\; 0 \leq u_i \leq x_i \quad \forall i = 1, \ldots, n \end{array} \right\}$$

Plug the result back into the dual of the original problem, we have the following optimization problem:

$$\begin{aligned} \min_{\theta,\boldsymbol{\rho},\boldsymbol{\eta}} \quad & \theta + \sum_{i\in N} \rho_i \mu_i + \sum_{i\in N} \eta_i(\mu_i^2 + \sigma_i^2) \\ s.t. \quad & \min_{(\mathbf{x},\mathbf{u})\in\mathbf{X}} \left\{ \max_{\mathbf{p}\in\mathbb{R}^n} \left[\sum_{i\in N} \left((u_i - \rho_i)p_i - \eta_i p_i^2 \right) \right] + \sum_{i\in N} c_i \left(x_i - u_i \right) \right\} \leq \theta \end{aligned}$$

□

Proof. Proof of Proposition 6 This is a typical min-max problem. It can be found in most of the optimization textbooks that if we interpret the two optimization, one outer minimization and one inner maximization, as the decision of two players, as in a *Stackelberg game*, then the first player, who wish to minimize the objective, makes decisions first. That is, the second player has the advantage to take the decision of the first player as given and to make his/her decision to maximize the objective and penalize the first player. Nonetheless, since both players objectives are common knowledge, the first player anticipates the move of the second player, or equivalently speaking, the first player can write the second player's decision as a function of his/her decision, and calculate the optimal solution to a minimization problem. Our proof is based on the same logic.

We base our discussion on the sign of η_i. Note that there are the following three possibilities for η_i: $\eta_i < 0$, $\eta_i = 0$, and $\eta_i > 0$.

- ($\eta_i < 0$): if $\eta_i < 0$, then the objective function of the second player is a convex quadratic function. Then, by having p_i goes to either positive or negative infinity, the objective goes to infinity. Therefore obviously, the first player do not want choose $\eta_i < 0$.
- ($\eta_i = 0$): Similarly, if $\eta_i = 0$ and $u_i - \rho_i \neq 0$, the second player can again make the objective goes to infinity by letting $|p_i|$ goes to infinity and letting p_i have the same sign as $u_i - \rho_i$. When $u_i - \rho_i = 0$, on the other hand, the objective of the second player equals to 0
- ($\eta_i > 0$): At last, when $\eta_i > 0$, the objective function of the second player is a convex quadratic function. Thus to maximize the objective, the optimal decision of the second player is $p_i^* = \dfrac{(u_i - \rho_i)}{2\eta_i}$.

In addition, it can be shown that whenever $u_i - \rho_i = 0$, $\eta_i^* = 0$. This result is due to the fact that $(\mu_i^2 + \sigma_i^2) > 0$.

□

Proof. Proof of Proposition 7 Denote the optimal solution of any variable a as a^*. Firstly, note that $\mu_i^2 + \sigma_i^2 > 0$ for all i; therefore, $\mu_i^* = \rho_i^*$ implies η_i^*. Suppose the proposition is

not true, that is, there exists some i, such that the optimal ρ_i' and η_i' satisfy $\rho_i' \neq \mu_i^*$ and $\eta_i' > 0$. Then, without loss of generality, let $\mu_i^* - \rho_i' - \rho_i^* = \delta$. Since η_i' is positive, and the inner optimization over $\mathbf{p}$ is separable, it is obvious that $p_i^* = \dfrac{(u_i^* - \rho_i')}{2\eta_i'}$. Then, the change in objective is:

$$\frac{\delta^2}{4\eta_i'} + \delta\mu_i^* + \eta_i'(\mu_i^2 + \sigma_i^2) \tag{A.7}$$

If the optimal solutions ρ_i' and η_i' satisfy $\rho_i' \neq \mu_i^*$ and $\eta_i' > 0$, (A.7) needs to be negative for some δ. However, note that $\eta_i' > 0$, and:

$$\mu_i^2 - 4\frac{1}{4\eta_i'}\left[\eta_i'(\mu_i^2 + \sigma_i^2)\right] = \sigma_i^2 \geq 0$$

Contradicts with the assumption that ρ_i' and η_i' are optimal, as changing from ρ_i^* and η_i^* to ρ_i' and η_i' fails to improve the optimal objective value. Therefore, the optimal solution to problem 2.9 satisfies $\eta_i^* = 0$ and $\mu_i^* = \rho_i^*$

□

Proof. Proof of Proposition 1 From Proposition 2, the optimal solution to the inner minimization problem when prices are known is fixed for fixed $\mathbf{p}$ and $\mathbf{c}$. In particular, the solution is the same as that for the deterministic problem when prices, demand arrivals, and delay penalties are $\mathbf{p}$, $\mathbb{E}[\mathbf{d}]$, and $\mathbf{c}$.

Then remainder of the proof follows the proof of Proposition 5, 6, and 7, with $\mathbf{d}$ being replaced with $\mathbb{E}[\mathbf{d}]$.

□

Proof. Proof of Proposition 8 We first define new decision variables. Note that after prices are announced, decision makers follow an AON policy to minimize the total disutility. Therefore, redefine decision variables as $u_{ij} \in \{0, 1\}$ and let $u_{ij} = 1$ denote that the demand that arrives in period i is scheduled to be satisfied in period j. Then we further define $y_{ij}(p) = P(u_{ij}^* = 1|p_j = p)$, and $y_{ij} = P(u_{ij}^* = 1)$, where u_{ij}^* is the optimal solution for given prices, and u_{ij} satisfies $\sum_{j \geq i} u_{ij} = 1$. Let $\mathcal{U} := \{u_{ij}, (j \geq i)|u_{ij} \in \{0, 1\}, \sum_{j \geq i} u_{ij} = 1\}$. Thus, y_{ij} belongs the convex hull defined by the extreme points of $\mathcal{U}$. Then, the expected total disutlity for given

distribution $F_{\mathbf{p}}$ can be written as:

$$\begin{aligned}
\mathbb{E}_{F_{\mathbf{p}}}[T(\tilde{\mathbf{p}}, \mathbf{c}, \mathbf{d})] &= \mathbb{E}_{F_{\mathbf{p}}}\left[\sum_{i\in N} d_i \sum_{j\geq i} u^*_{ij}(\tilde{\mathbf{p}})\tilde{p}_j\right] + \mathbb{E}_{F_{\mathbf{p}}}\left[\sum_{i\in N} d_i \sum_{j\geq i} \hat{c}_{ij} u^*_{ij}(\tilde{\mathbf{p}})\right] \\
&= \sum_{i\in N} d_i \sum_{j\geq i} \int (p+\hat{c}_{ij})\mathbb{E}[u^*_{ij}(\tilde{\mathbf{p}})|\tilde{p}_j = p] f_j(p)dp \\
&= \sum_{i\in N} d_i \sum_{j\geq i} \int (p+\hat{c}_{ij})\mathbb{E}[u^*_{ij}(\tilde{\mathbf{p}})|\tilde{p}_j = p, u^*_{ij}(\tilde{\mathbf{p}}) = 1] P(u^*_{ij}(\tilde{\mathbf{p}}) = 1|\tilde{p}_j = p) f_j(p)dp \\
&= \sum_{i\in N} d_i \sum_{j\geq i} \int (p+\hat{c}_{ij}) y_{ij}(p) f_j(p)dp
\end{aligned}$$

Then, the lower bound on the best-case total disutility is:

$$\begin{aligned}
Z^* \geq \min \quad & \sum_{i\in N} d_i \sum_{j\geq i} \int (p+\hat{c}_{ij}) y_{ij}(p) f_j(p)dp \\
s.t. \quad & \int y_{ij}(p) f_j(p)dp = y_{ij} \qquad \forall i \in N, j \geq i \\
& 0 \leq y_{ij}(p) \leq 1 \qquad \forall i \in N, j \geq i,\ \forall p \\
& \mathbf{y} \in \mathcal{CH}(\mathcal{U})
\end{aligned}$$

Therefore, for given $\mathbf{y} \in \mathcal{CH}(\mathcal{U})$, above right hand side problem is separable in (i, j). After dropping index i from subscripts, the subproblem is:

$$\begin{aligned}
\min \quad & \int (p+\hat{c}_j) y_j(p) f_j(p)dp \\
s.t. \quad & \int y_j(p) f_j(p)dp = y_j \qquad \forall j \geq i \\
& 0 \leq y_j(p) \leq 1 \qquad \forall j \geq i,\ \forall p
\end{aligned}$$

After relaxing the first set of constraints in the above problem by introducing lagrangian multipliers λ, the dual problem is:

$$\begin{aligned}
\min \quad & \int (p+\hat{c}_j - \lambda) y_j(p) f_j(p)dp + \lambda y_j \\
s.t. \quad & 0 \leq y_j(p) \leq 1 \qquad \forall j \geq i,\ \forall p
\end{aligned}$$

For the dual problem, obviously $y_j(p)^* = 1$ only if $p + \hat{c}_j - \lambda \leq 0$. Let the optimal lagrangian multiplier be λ^*, and the corresponding dual optimal objective value be $\int_{-\infty}^{\lambda^*-\hat{c}_{ij}} (p+\hat{c}_{ij}) f_j(p)dp$. Then if we choose $y^*_{ij} = F_j(\lambda^* - \hat{c}_j)$, it can be shown that y^*_{ij} and $y^*_{ij}(p)$ are primal feasible and the corresponding primal objective value attains that of the optimal dual

objective, and thus they are primal optimal solutions. Hence, solving for the lower bound of the best-case long-term average total disutility reduces to solving the following:

$$\begin{aligned} \min \quad & \sum_{i\in N} d_i \sum_{j\geq i} \int_{-\infty}^{F^{-1}(y_{ij})} (p+\hat{c}_{ij}) f_j(p) dp \\ s.t. \quad & \mathbf{y} \in \mathcal{CH}(\mathcal{U}) \end{aligned}$$

Let $t = F_j(p)$, then the above optimization problem is equivalent to the following one:

$$\begin{aligned} \min \quad & \sum_{i\in N} d_i \sum_{j\geq i} \int_{0}^{y_{ij}} (F_j^{-1}(t)+\hat{c}_{ij}) dt \\ s.t. \quad & \sum_{j\geq i} y_{ij} = 1 \qquad \forall i \in N \\ & y_{ij} \geq 0 \qquad \forall i \in N,\ j \geq i \end{aligned}$$

Then, we can show the above problem is a convex optimization problem. Obviously, the constraints are linear. We verify the convexity of the objective by calculating its derivative. Firstly, since the problem is separable in i, it suffices to take derivative over the vector $\mathbf{y}_i := (y_{ii}, y_{i(i+1)}, \ldots, y_{in})$:

$$\frac{d}{d\mathbf{y}} \sum_{j\geq i} \int_{0}^{y_{ij}} (F_j^{-1}(t)+\hat{c}_{ij}) dt = \begin{pmatrix} F_i^{-1}(y_{ii}) \\ F_{i+1}^{-1}(y_{i(i+1)}) + \hat{c}_{i(i+1)} \\ \vdots \\ F_n^{-1}(y_{in}) + \hat{c}_{in} \end{pmatrix}$$

Since $F_j^{-1}(y_{ij})$ is non-decreasing in y_{ij}, the hessian is a diagonal matrix with non-negative entries, the objective is convex. Therefore, KKT conditions are necessary for the optimal solution, where the KKT conditions for each of the subproblem (for each $i \in N$) are:

$$\text{(KKT Conditions)}: \begin{cases} F_j^{-1}(y_{ij}^*) + \hat{c}_{ij} + \lambda_i^* = 0 \quad \forall j \geq i \\ \sum_{j\geq i} y_{ij}^* = 1 \end{cases}$$

where $\boldsymbol{\lambda}^*$ and $\mathbf{y}^*$, taking values as described in the proposition, are the optimal solution to the problem that solves for the lower bound on the best-case long-term average total disutility. Next, we show that this lower bound is attainable by constructing a joint distribution of prices.

Let $\mathcal{V}$ be the set of extreme points of $\mathcal{CH}(\mathcal{U})$, define probability distribution $\lambda_{\mathbf{y}}^*$ for all $\mathbf{y} \in \mathcal{V}$ as the following:

$$\begin{cases} \lambda_{\mathbf{y}}^* \geq 0 \\ \sum_{\mathbf{y}\in\mathcal{V}} \lambda_{\mathbf{y}}^* = 1 \\ y_{ij}^* = \sum_{\mathbf{y}\in\mathcal{V}: y_{ij}=1} \lambda_{\mathbf{y}}^* \end{cases}$$

Then, we generate the multivariate joint distribution by first choosing $\mathbf{y}$ from $\mathcal{V}$ with probability $\lambda^*_{\mathbf{y}}$, then generate the prices for each period j according to $\dfrac{f_j(p)\cdot \mathbf{1}_{\{p\le F_j^{-1}(y^*_{ij})\}}}{y_{ij^*}}$. Let the probability density function of the newly generated price in period j be $f'_j(p)$, then is not hard to show that:

$$\begin{aligned}
f'_j(p) &= \sum_{\mathbf{y}\in\mathcal{V}:y_{ij}=1} \lambda^*_{\mathbf{y}} \frac{f_j(p)\mathbf{1}_{\{p\le F_j^{-1}(y^*_{ij})\}}}{y_{ij^*}} + \sum_{\mathbf{y}\in\mathcal{V}:y_{ij}=0} \lambda^*_{\mathbf{y}} \frac{f_j(p)\mathbf{1}_{\{p\ge F_j^{-1}(y^*_{ij})\}}}{1-y_{ij^*}} \\
&= f_j(p)\mathbf{1}_{\{p\le F_j^{-1}(y^*_{ij})\}} + f_j(p)\mathbf{1}_{\{p\ge F_j^{-1}(y^*_{ij})\}} \\
&= f_j(p)
\end{aligned}$$

Then, we verify that equality holds at $f'_j(p)$. It can be shown that:

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[Z(\tilde{\mathbf{p}})\right] &\le \sum_{v\in\mathcal{V}} \lambda^*_v \left(\sum_{i\in N} d_i \sum_{j\ge i} \int (p+\hat{c}_{ij}) f'_j(p)dp \right) \\
&= \sum_{i\in N} d_i \sum_{j\ge i} \left(\sum_{\mathbf{y}\in\mathcal{V}:y_{ij}=1} \lambda^*_{\mathbf{y}} \right) \int (p+\hat{c}_{ij}) \frac{f_j(p)\mathbf{1}_{\{p\le F_j^{-1}(y^*_{ij})\}}}{y^*_{ij}} dp \\
&= \sum_{i\in N} d_i \sum_{j\ge i} \int_0^{y^*_{ij}} (F_j^{-1}(t)+\hat{c}_{ij})dt
\end{aligned}$$

Since the best-case long-term average total disutility is proved to be no-smaller than the right hand side, the lower bound Z^D_{best} that we derived above is attained.

□

Appendix B

Appendix for Chapter 3

B.1 Model Moving Demand Forward

To model the demand that can be executed in advance, we model a new set of additive demand on the refrigerator. Similarly with the demand on the A.C., each task on the refrigerator corresponds to lowering the temperature by $1°F$. Demand arrives at the beginning of each period, assuming the "preferred" temperature is T_{min}. The demand tasks on and above the margin (which corresponds to lower the temperature from T_{max})have high sensitivity on service, while all other demand tasks on the refrigerator have low sensitivity on service.

The insight of shifting demand forward in time is that, by following the above strategy, we are able to shift demand forward, in addition to shifting demand afterwards. The effect of this strategy is equivalent to shifting energy purchasing forward, which is achieved by owning a local storage device. When the capacity of local storage device is limited, or the marginal cost of local storage is high, having the ability to shift demand forward becomes more valuable.

B.2 A Counterexample that shows the Greedy Algorithm Fails to Solve the VBKP

As a counterexample, consider the case in which we have two items, with $v_1 = 11, u_1 = 1$ and $v_2 = 21.5, u_2 = 2$. The cost function $C_t(z)$ is defined as:

$$C_t(z) = \begin{cases} 10z & \text{if } z \leq 2 \\ 11.5z & \text{if } z \geq 2 \end{cases}$$

By ranking the efficiency ratios we pick item one first. Because picking item two does not change the objective value, we can either pick it or leave it, and the objective function values

are 1 in both cases. However, the optimal solution is to pick item two only and the corresponding objective function value is 1.5. This example illustrates that the greedy algorithm that solves the CVBKP fails to work for problems with piece-wise linear cost structures and mixed integer or integer decision variables.

B.3 Proofs

We first provide the following Lemma before proving Proposition 1:

Lemma 5. *Point-wise maximization (resp. minimization) or taking supremum (resp. infimum) preserves monotonicity.*

Proof of Lemma 5. Suppose $f : \mathbb{R}^u \times \mathbb{R}^v \to \mathbb{R}$ satisfies that $f(x, y) \geq f(z, y)$, $\forall x \leq z, \forall y$. We can show that function $g : \mathbb{R}^u \to \mathbb{R}$ defined as $g(x) = \sup\limits_y f(x, y)$ is non-increasing by contradiction.

Suppose g is not non-increasing in x, that is, $\exists\ g(x) < g(z)$ for some $x \leq z$, then:

$$g(x) = \sup_y f(x, y) < g(z) = \sup_y f(z, y) = f(z, \widehat{y}) \leq f(x, \widehat{y})$$

where the second inequality follows from the definition of function f. However, this contradicts with the definition of function g that $f(x, \widehat{y}) \leq \sup\limits_y f(x, y) = g(x)$. Therefore, point-wise maximization (or taking supremum) over non-increasing functions preserves monotonicity.

Similarly, suppose $f : \mathbb{R}^u \times \mathbb{R}^v \to \mathbb{R}$ satisfies $f(x, y) \leq f(z, y)$, $\forall x \leq z, \forall y$. We can show that function $g : \mathbb{R}^u \to \mathbb{R}$ defined as $g(x) = \sup_y f(x, y)$ is non-decreasing by contradiction.

Suppose g is not non-decreasing, that is, $\exists\ g(x) > g(z)$ for some $x \leq z$, then:

$$f(z, \widehat{y}) \geq f(x, \widehat{y}) = \sup_y f(x, y) = g(x) > g(z)$$

contradicts with $g(z) = \sup\limits_y f(z, y) \geq f(z, \widehat{y})$. Therefore, point-wise maximization (or taking supremum) over non-decreasing functions preserves monotonicity.

The proof for point-wise minimization or taking infimum preserves monotonicity is similar, and thus, is omitted here.

□

Based on Lemma 5, we can prove Proposition 1, which states the monotonicity of the value-to-go term, as follows:

Proof of Proposition 9. (Part (a) & (b)) Firstly, we show part (a) and (b) by mathematical induction.
For $t = T$, recall that we assume the following boundary condition: $J^*_{T+1}(R_{T+1}, \mathcal{H}_{T+1}) \equiv 0$ for all $(R_{T+1}, \mathcal{H}_{T+1})$, that is, $V^*_{T+1}(R_T, \mathcal{H}_T, X_T) \equiv 0$ for all $(R_T, \mathcal{H}_T, X_T)$. Thus, the optimal total cost at time T starting from state $(R_T, \mathcal{H}_T)$ is:

$$\begin{aligned} J^*_T(R_T, \mathcal{H}_T) = \min_{X_T} & \left[C_T(R_T, X_T) + \sum_{a \in \mathcal{A}} \sum_{1 \le n \le N} \widetilde{\pi}_{a^{[n]}} d^X_{a^{[n]}t} + L_T(\mathbf{d}_T, \mathcal{H}_T, \mathbf{w}_T) \right] \\ s.t. \quad & (3.1a) - (3.1f) \\ & (3.2) - (3.5) \\ & (3.6) \end{aligned}$$

Note that for any fixed b_T, $\mathcal{H}_T$, and feasible X_T, the costs paid to the grid $C_T(R_T, X_T)$ is non-decreasing in z_T, which is non-decreasing in each element of $\mathbf{d}_T$. Therefore, the first term in the objective function is non-decreasing in $\mathbf{d}_T$. Because the second term is linear in $d_{a^{[n]}t}$, and $\widetilde{\pi}_{a^{[n]}} > 0$, the second term is non-decreasing in $\mathbf{d}_T$ as well. At last, because $L_T(\mathbf{d}_T, \mathcal{H}_T, \mathbf{w}_T)$ is non-decreasing in $\mathbf{d}^X_T$, and $\mathbf{d}^X_T$ is linear non-decreasing in each element of $\mathbf{d}_T$, the last term is non-decreasing in $\mathbf{d}_T$. Then it follows from Lemma 5 that $J^*_T(R_T, \mathcal{H}_T)$ is non-decreasing in $\mathbf{d}_T$.

Then, suppose $J^*_{t+1}(R_{t+1}, \mathcal{H}_{t+1})$ is non-decreasing in $\mathbf{d}_t$. In order to calculate $J^*_t(R_t, \mathcal{H}_t)$, we need to calculate the following value-to-go function:

$$\begin{aligned} V^*_{t+1}(R_t, \mathcal{H}_t, X_t) &= \mathbb{E}_{\mathcal{I}_t}\left[J^*_{t+1}(R_{t+1}, \mathcal{H}_{t+1}) | R_t, \mathcal{H}_t, X_t \right] \\ &= \mathbb{E}_{\mathcal{I}_t}\left[J^*_{t+1}\left(h\left(R^X_t, \mathbf{q}_t, g_t\right), \mathcal{H}_{t+1}\right) | R_t, \mathcal{H}_t, X_t \right] \\ &= \mathbb{E}_{\mathcal{I}_t}\left[J^*_{t+1}\left(h\left(f\left(R_t, X_t\right), \mathbf{q}_t, g_t\right), \mathcal{H}_{t+1}\right) | \mathcal{H}_t, X_t \right] \end{aligned} \tag{B.1}$$

where the third equality holds because R_t is conditionally independent with $\mathcal{I}_t$ given $\mathcal{H}_t$ and X_t. We first note that $\mathbf{d}_{t+1} = h^d\left(\mathbf{d}^X_t, \mathbf{q}_t\right)$ is non-decreasing in $\mathbf{d}^X_t$. Secondly, because $\mathbf{d}^X_t = f^d(\mathbf{d}_t, X_t)$ is non-decreasing in $\mathbf{d}_t$, and taking the expectation of J^*_{t+1} over the random information processes $\mathcal{I}_t$, which is essentially takes the convex combination of non-decreasing functions, preserves monotonicity, $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ is non-decreasing in $\mathbf{d}_t$. In addition, following the proof in the previous part, the one-period cost, $C_t(R_t, X_t)$, $\sum_a \sum_n \widetilde{\pi}_{a^{[n]}} d^X_{a^{[n]}t}$, and $L_t(\mathbf{d}_t, \mathcal{H}_t, \mathbf{w}_t)$ are non-decreasing in $\mathbf{d}_t$. Therefore, the sum of the terms in the objective function for calculating j^*_t are non-decreasing in $\mathbf{d}_t$. Again by applying Lemma 5, we obtain that $J^*_t(R_t, \mathcal{H}_t)$ is non-decreasing in $\mathbf{d}_t$, that is, Part (a) holds.

Similarly, suppose $V^*_{t+1}(R_t, \mathcal{H}_t, X_t)$ is non-decreasing in $\mathbf{d}_t$, $J^*_t(R_t, \mathcal{H}_t)$ is non-decreasing in $\mathbf{d}_t$. Then, it follows that $V^*_t(R_{t-1}, \mathcal{H}_{t-1}, X_{t-1})$ is non-decreasing in $\mathbf{d}_{t-1}$.

(Part (c) & (d)) Let $R_t = (\mathbf{d}_t, b_t)$, $R'_t = (\mathbf{d}_t, b_t + \delta_b)$. We need to show for all $\delta_b > 0$, $J_t^*(R_t, \mathcal{H}_t) \geq J_t^*(R'_t, \mathcal{H}_t)$, for all $t \in T$. It follows directly that :

$$J_t^*(R'_t, \mathcal{H}_t) \leq J_t^*(R_t, \mathcal{H}_t) - \delta_b \min_z \{p_t(z)\} \leq J_t^*(R_t, \mathcal{H}_t)$$

where the first inequality comes from the fact that selling the difference in storage is feasible, but not necessarily optimal, and $\delta_b \min_z \{P_t(z)\}$ is the upper bound on the profit of selling the difference. The second inequality follows from the assumption that $\min_z \{P_t(z)\} >= 0$. Therefore, Part (c) is true. Part (d) follows directly from equality (B.1) and the state transitions of the battery.

□

Proof of Proposition 10.

(a) If $p_t(\cdot)$ is a non-decreasing stepwise function.
Suppose $\exists j$, such that $p_t^{j-1} \leq -\psi_t^m < p_t^j$, then:

1) if $\mathbf{d}_t^\intercal \mathbf{w}_t - y_t < b^j$, lowering y_t by $\delta_y > 0$ changes (3.15) by:

$$\begin{aligned}\psi_t^m \delta_y &+ (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t + \delta_y)\, p_t\, (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t + \delta_y) \\ &- (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t)\, p_t\, (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t) \leq \psi_t^m \delta_y + \delta_y p_t^{j-1} \leq 0\end{aligned}$$

Thus, reducing y_t when $\mathbf{d}_t^\intercal \mathbf{w}_t - y_t < b^j$ decreases (3.15);

2) if $\mathbf{d}_t^\intercal \mathbf{w}_t - y_t \geq b^j$, increasing y_t by $\delta_y > 0$ changes (3.15) by:

$$\begin{aligned}-\psi_t^m \delta_y &+ (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t - \delta_y)\, p_t\, (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t - \delta_y) \\ &- (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t)\, p_t\, (\mathbf{d}_t^\intercal \mathbf{w}_t - y_t) \leq -\psi_t^m \delta_y - \delta_y p_t^{j-1} < 0\end{aligned}$$

Thus, increasing y_t when $\mathbf{d}_t^\intercal \mathbf{w}_t - y_t \geq b^j$ decreases (3.15);

To sum up, setting y_t so as to let $(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t)$ be as close to b_t^j as possible minimizes $C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)$. On the other hand, increasing $\mathbf{d}_t^\intercal \mathbf{w}_t$ increases $C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)$. More specifically, for any $\epsilon > 0$:

$$\frac{C_t(\mathbf{d}_t^\intercal \mathbf{w}_t + \epsilon) - C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)}{\epsilon} =$$

$$\begin{cases} p_t^k,\ s.t.\ b_t^k \leq \mathbf{d}_t^\intercal \mathbf{w}_t - \min\limits_{y \in \mathcal{Y}_t} y \leq b_t^{k+1} & \text{, if } \mathbf{d}_t^\intercal \mathbf{w}_t - \min\limits_{y \in \mathcal{Y}_t} y < b_t^j \\ -\psi_t^m & \text{, if } \mathbf{d}_t^\intercal \mathbf{w}_t - \max\limits_{y \in \mathcal{Y}_t} y \leq b_t^j \leq \mathbf{d}_t^\intercal \mathbf{w}_t - \min\limits_{y \in \mathcal{Y}_t} y \\ p_t^k,\ s.t.\ b_t^k \leq \mathbf{d}_t^\intercal \mathbf{w}_t - \max\limits_{y \in \mathcal{Y}_t} y \leq b_t^{k+1} & \text{, if } \mathbf{d}_t^\intercal \mathbf{w}_t - \max\limits_{y \in \mathcal{Y}_t} y > b_t^j \end{cases}$$

Therefore, $C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)$ is convex and non-decreasing in $\mathbf{d}_t^\intercal \mathbf{w}_t$, when $p_t(\cdot)$ is a non-decreasing stepwise function.

(b) If $p_t(\cdot)$ is a twice-differentiable convex increasing function, and $p_t''(x) = 0$ for all $x \leq 0$. Note that:

$$\frac{d^2}{dy_t^2}\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right) p_t\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right) = 2p_t'\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right) + \left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right) p_t''\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y_t\right)$$

Then the problem that solves for $C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)$ is convex. By applying KKT condition, the optimal $y_t^*(\mathbf{w}_t)$ satisfies:

$$y_t^*(\mathbf{w}_t) = \begin{cases} \min\limits_{y \in \mathcal{Y}_t} y & \text{if } \min\limits_{y \in \mathcal{Y}_t} y > y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \\ y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) & \text{if } \min\limits_{y \in \mathcal{Y}_t} y \leq y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \leq \max\limits_{y \in \mathcal{Y}_t} y \\ \max\limits_{y \in \mathcal{Y}_t} y & \text{if } \max\limits_{y \in \mathcal{Y}_t} y < y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \end{cases}$$

where $y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right)$ is the solution of equation $\left(y - \mathbf{d}_t^\intercal \mathbf{w}_t\right) p_t'\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y\right) - p_t\left(\mathbf{d}_t^\intercal \mathbf{w}_t - y\right) = \psi_t^m$. Similarly for $C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)$ we have:

$$\frac{d}{d\mathbf{d}_t^\intercal \mathbf{w}_t} C_t(\mathbf{d}_t^\intercal \mathbf{w}_t) =$$
$$\begin{cases} \frac{d}{d\mathbf{d}_t^\intercal \mathbf{w}_t}\left(\mathbf{d}_t^\intercal \mathbf{w}_t - \min\limits_{y \in \mathcal{Y}_t} y\right) p_t'\left(\mathbf{d}_t^\intercal \mathbf{w}_t - \min\limits_{y \in \mathcal{Y}_t} y\right) & \text{if } \min\limits_{y \in \mathcal{Y}_t} y > y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \\ -\psi_t^m & \text{if } \min\limits_{y \in \mathcal{Y}_t} y \leq y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \leq \max\limits_{y \in \mathcal{Y}_t} y \\ \frac{d}{d\mathbf{d}_t^\intercal \mathbf{w}_t}\left(\mathbf{d}_t^\intercal \mathbf{w}_t - \max\limits_{y \in \mathcal{Y}_t} y\right) p_t'\left(\mathbf{d}_t^\intercal \mathbf{w}_t - \max\limits_{y \in \mathcal{Y}_t} y\right) & \text{if } \max\limits_{y \in \mathcal{Y}_t} y < y_t^0\left(\mathbf{d}_t^\intercal \mathbf{w}_t\right) \end{cases}$$

Therefore, $C_t(\mathbf{d}_t^\intercal \mathbf{w}_t)$ is convex and increasing when $p_t(\cdot)$ is a twice-differentiable convex increasing function, and $p_t''(x) = 0$ for all $x \leq 0$.

□

Proof of Theorem 1. We show that the decision problem of the variable budget knapsack problem is NP-complete by first showing that it is in the class of NP, followed by showing that the knapsack problem (KP) reduces to variable budget knapsack problems in polynomial time.

a) We first show the problem is in class of NP. Because the certificate consists of a realization of the decision sets and the budget, which is less than the maximum requirement on electricity, the certificate is polynomial in the size of input, which is the jobs and the requirement on electricity of each job. Since (1) the certificate checking algorithm verify the sum of electricity requirements of the jobs $a^{[n]}$'s with decision $w_{a^{[n]}t} = 1$ is less than or equal to the budget, which takes $O(n)$ where n is the total number of items, and (2) the algorithm checks the sum of the evaluation of these items (which

takes $O(n)$) minus the cost of the budget, which takes polynomial time since the cost of budget can also be evaluated in polynomial time, the problem is NP.

b) To show the decision form of VBKP is NP-complete, we show that the KP can reduce to the VBKP in polynomial time. Consider an arbitrary KP:

$$\begin{aligned} KP_1: \quad \max \quad & \sum_{a=1}^{k} v_a x_a \\ s.t. \quad & \sum_{a=1}^{k} w_a x_a \leq B \\ & x_a \in \{0,1\} \end{aligned}$$

We construct a corresponding instance of VBKP as the following. Let the convex cost function $C(y)$ take the form of:

$$C(y) = \begin{cases} 0 & \text{if } y \leq B \\ \infty & \text{if } y > B \end{cases}$$

Then by setting the weights and benefits of the items to be the same in this problem as in KP, we have the following VBKP:

$$\begin{aligned} VBKP_2: \quad \max \quad & \sum_{a=1}^{k} v_a x_a - C(y) \\ s.t. \quad & \sum_{a=1}^{k} w_a x_a \leq y \\ & x_a \in \{0,1\} \end{aligned}$$

It remains to show that the $VBKP_2$ is equivalent to KP_1. In $VBKP_2$, the value of y never exceeds B, for that whenever the budget constraint $\sum_{a=1}^{k} w_a x_a \leq B$ is violated, the objective function value of $VBKP_2$ is negative infinity. When the budget is less than or equal to B, $C(y)$ equals to zero, and thus $VBKP_2$ has the same objective function as KP_1 does. It follows that $VBKP_2$ solves exactly KP_1. Because the construction of $VBKP_2$ takes $O(k)$ time, KP reduces in polynomial time to VBKP and the decision form of VBKP is NP-complete.

To show the VBPCKP is NP-hard, it suffices to show that the VBKP is reducible to VBPCKP, which is obvious.

□

Proof of Proposition 11. The result is immediate from the Principal of Optimality. □

Proof of Proposition 12. Recall that the decisions $X_t^{\mu^*}$ and X_t^{μ} are the optimal solutions of (P^*) and the ADP approach, thus, we have:

$$E_t\big(R_t, X_t^{\mu}\big) + \overline{\Gamma}_{t+1}\big(R_t, {}_t^{\mu}\big) \leq E_t\big(R_t, {}_t^{\mu^*}\big) + \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) \tag{B.2}$$

$$E_t\big(R_t, X_t^{\mu}\big) + \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) \geq E_t\big(R_t, X_t^{\mu^*}\big) + \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big) \tag{B.3}$$

Re-organizing inequalities (B.2) and add to both sides term $\Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) - \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big)$, we obtain:

$$\begin{aligned}
&E_t\big(R_t, X_t^{\mu}\big) + \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) - \big(E_t\big(R_t, X_t^{\mu^*}\big) + \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big)\big)\\
&= J^{\mu}\big(R_t, \mathcal{H}_t\big) - J^*\big(R_t, \mathcal{H}_t\big)\\
&\leq \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) - \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big) - \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu}\big) + \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big)\\
&\leq 2\max_X \big|\Gamma_{t+1}\big(R_t, \mathcal{H}_t, X\big) - \overline{\Gamma}_{t+1}\big(R_t, X\big)\big|
\end{aligned} \tag{B.4}$$

Similarly, by taking inequality (B.2) off (B.3) and reorganizing the terms, we obtain the following necessary condition for all $X_t^{\mu} \neq X_t^{\mu^*}$:

$$\overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) - \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu}\big) \geq \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big) - \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) \tag{B.5}$$

Because $E_t\big(R_t, \mathcal{X}_t\big)$ is not necessarily monotone, we discuss the following three scenarios based on inequality (B.5):

- if $\overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) - \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu}\big) \geq 0$ and $\Gamma_{t+1}\big(R_t, \mathcal{H}_t, {}_t^{\mu^*}\big) - \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) \leq 0$, then by inequalities (B.2) and (B.4), the upper bound is tightened as:

$$J^{\mu}\big(R_t, \mathcal{H}_t\big) - J^*\big(R_t, \mathcal{H}_t\big) \leq \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) - \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu}\big) \tag{B.6}$$

  Meanwhile, the necessary condition holds, therefore, $J^{\mu}\big(R_t, \mathcal{H}_t\big) - J^*\big(R_t, \mathcal{H}_t\big) \leq \widetilde{B}_t\big(R_t, \mathcal{H}_t\big)$, where

$$\begin{aligned}
\widetilde{B}_t\big(R_t, \mathcal{H}_t\big) = \max_{X_1, X_2} \quad & \overline{\Gamma}_{t+1}\big(R_t, X_1\big) - \overline{\Gamma}_{t+1}\big(R_t, X_2\big)\\
s.t. \quad & E_t\big(R_t, X_1\big) + \overline{\Gamma}_{t+1}\big(R_t, X_1\big) \leq E_t\big(R_t, X_2\big) + \overline{\Gamma}_{t+1}\big(R_t, X_2\big)
\end{aligned}$$

- if $\overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) - \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu}\big) \geq 0$ and $\Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big) - \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) \geq 0$, or $\overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu^*}\big) - \overline{\Gamma}_{t+1}\big(R_t, X_t^{\mu}\big) \leq 0$ and $\Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu^*}\big) - \Gamma_{t+1}\big(R_t, \mathcal{H}_t, X_t^{\mu}\big) \leq 0$, no tighter bounds can be obtained.

□

Bibliography

[1] J.P. Acton and B.M. Mitchell. *The effect of time of use rates: Facts vs. opinions*. Public Utilities Fortnightly, 1981.

[2] Asma Al-Tamimi, Frank L. Lewis, and Murad Abu-Khalaf. "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control". In: *Automatica* 43.3 (2007), pp. 473 –481. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2006.09.019. URL: http://www.sciencedirect.com/science/article/pii/S0005109806004249.

[3] H. Allcott. "Real Time Pricing and Electricity Markets". working paper. Jan. 2012.

[4] R.N. Anderson et al. "Adaptive Stochastic Control for the Smart Grid". In: *Proceedings of the IEEE* 99.6 (June 2011), pp. 1098 –1115. ISSN: 0018-9219. DOI: 10.1109/JPROC.2011.2109671.

[5] R. E. Bellman and S. E. Dreyfus. *Applied Dynamic Programming*. Princeton University Press, 1962.

[6] A. Ben-Tal and A. Nemirovski. "Robust convex optimization". In: *Mathematics of Operations Research* 23 (1998), pp. 769–805.

[7] A. Ben-Tal and A. Nemirovski. "Robust solutions of linear programming problems contaminated with uncertain data". In: *Mathematica Programming* 88 (2000), pp. 411–424.

[8] A. Ben-Tal and A. Nemirovski. "Robust solutions to uncertain programs". In: *Operations Research Letters* 25 (1999).

[9] D. P. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

[10] D.P. Bertsekas. *Dynamic Programming and Optimal Contro*. New Hampshire: Athena Scientific, 1995.

[11] D. Bertsimas and I. Popescu. "Optimal inequalities in probability theory: A convex optimization approach". In: *SIAM Journal on Optimization* 15.3 (2005), pp. 780–804.

[12] D. Bertsimas and M. Sim. "The Price of Robustness". English. In: *Operations Research* 52.1 (2004), pp. 35–53. ISSN: 0030364X. URL: http://www.jstor.org/stable/30036559.

[13] D. Bertsimas and A. Thiele. "A robust optimization approach to inventory theory". In: *Operations Research* (2006).

[14] D. Bertsimas et al. "Adaptive robust optimization for the security constrained unit commitment problem". In: *Technical report, Ssubmitted to IEEE Transactions on Power Systems* (2011).

[15] D. Bertsimas et al. "Models for minimax stochastic linear optimization problems with risk aversion". In: *Mathematics of Operations Research* 35(3) (2010), pp. 580–602.

[16] J. W. Black and R. Tyagi. "Potential problems with large scale differential pricing programs". In: *Transmission and Distribution Conference and Exposition, 2010 IEEE PES*. Apr. 2010, pp. 1 –5. DOI: 10.1109/TDC.2010.5484309.

[17] S. Borenstein. "The Long-Run Efficiency of Real-Time Electricity Pricing". In: *The Energy Journal* 26.3 (2005), pp. 93–116.

[18] S. Borenstein and S.P. Holland. *On the Efficiency of Competitive Electricity Markets With Time-Invariant Retail Prices*. Working Paper 9922. National Bureau of Economic Research, Aug. 2003. URL: http://www.nber.org/papers/w9922.

[19] S. Borenstein, M. Jaske, and A. Rosenfeld. *Dynamic Pricing, Advanced Metering, and Demand Response in Electricity Markets*. Tech. rep. Berkeley, California: University of California Energy Institute, 2002.

[20] Peter Cappers, Charles Goldman, and David Kathan. "Demand response in U.S. electricity markets: Empirical evidence". In: *Energy* 35.4 (2010). ¡ce:title¿Demand Response Resources: the US and International Experience¡/ce:title¿ ¡xocs:full-name¿Demand Response Resources: the US and International Experience¡/xocs:full-name¿, pp. 1526 –1535. ISSN: 0360-5442. DOI: 10.1016/j.energy.2009.06.029. URL: http://www.sciencedirect.com/science/article/pii/S0360544209002461.

[21] P. Carpentier et al. "Stochastic optimization of unit commitment: a new decomposition framework". In: *IEEE Transactions on Power Systems* 11 (1996), pp. 1067–1073.

[22] Douglas W. Caves and Laurits R. Christensen. "Econometric analysis of residential time-of-use electricity pricing experiments". In: *Journal of Econometrics* 14.3 (1980), pp. 287 –306. ISSN: 0304-4076. DOI: 10.1016/0304-4076(80)90029-9. URL: http://www.sciencedirect.com/science/article/pii/0304407680900299.

[23] D.W. Caves, L.R. Christensen, and J.A. Herriges. "Consistency of residential customer response in time-of-use electricity pricing experiments". In: *Journal of Econometrics* 26.12 (1984), pp. 179 –203. ISSN: 0304-4076. DOI: 10.1016/0304-4076(84)90017-4. URL: http://www.sciencedirect.com/science/article/pii/0304407684900174.

[24] A.J. Conejo, J.M. Morales, and L. Baringo. "Real-Time Demand Response Model". In: *Smart Grid, IEEE Transactions on* 1.3 (Dec. 2010), pp. 236 –242. ISSN: 1949-3053. DOI: 10.1109/TSG.2010.2078843.

[25] Michael A. Crew, Chitru S. Fernando, and Paul R. Kleindorfer. "The theory of peak-load pricing: A survey". In: *Journal of Regulatory Economics* 8 (3 1995). 10.1007/BF01070807, pp. 215–248. ISSN: 0922-680X. URL: http://dx.doi.org/10.1007/BF01070807.

[26] E. Delage and Y. Ye. "Distributionally robust optimization under moment uncertainty with application to data-driven problems". In: *Operations Research* 58(3) (2010), pp. 595–612.

[27] DOE. *Benefit of Demand Response in Electricity Market and Recommendations for Achieving them*. Feb. 2006.

[28] D. P. de Farias and B. Van Roy. "The Linear Programming Approach to Approximate Dynamic Programming". In: *Operations Research* 51.6 (2003), pp. 850–865. DOI: 10.1287/opre.51.6.850.24925. eprint: http://or.journal.informs.org/content/51/6/850.full.pdf+html. URL: http://or.journal.informs.org/content/51/6/850.abstract.

[29] J. Goh and M. Sim. "Distributionally robust optimization and its tractable approximations". In: *Operations research* 58(4) (2010), pp. 902–917.

[30] Rob Hartway, Snuller Price, and C.K Woo. "Smart meter, customer choice and profitable time-of-use rate option". In: *Energy* 24.10 (1999), pp. 895 –903. ISSN: 0360-5442. DOI: 10.1016/S0360-5442(99)00040-7. URL: http://www.sciencedirect.com/science/article/pii/S0360544299000407.

[31] K. Herter. "Residential implementation of critical-peak pricing of electricity". In: *Energy Policy* 35.4 (2007), pp. 2121 –2130. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2006.06.019. URL: http://www.sciencedirect.com/science/article/pii/S0301421506002783.

[32] S. P. Holland and E. T. Mansur. "The Short-Run Effects of Time-Varying Prices in Competitive Electricity Markets." In: *Energy Journal* 27.4 (2006), pp. 127 –155. ISSN: 01956574.

[33] Veronica Irastorza. "New Metering Enables Simplified and More Efficient Rate Structures". In: *The Electricity Journal* 18.10 (2005), pp. 53 –61. ISSN: 1040-6190. DOI: 10.1016/j.tej.2005.10.008. URL: http://www.sciencedirect.com/science/article/pii/S104061900500134X.

[34] L. Jiang and S. Low. "Multi-period optimal procurement and demand responses in the presence of uncertain supply". In: *50th Conf. on Decision and Control*. Dec. 2011.

[35] R. Jiang et al. *Two-stage Robust Power Grid Optimization Problem*. submitted to Journal of Operations Research. 2010.

[36] P. L. Joskow. "Comparing the Costs of Intermittent and Dispatchable Electricity Generating Technologies". In: *The American Economic Review* 101.3 (2010), pp. 238–241. DOI: doi:10.1257/aer.100.3.238. URL: http://www.ingentaconnect.com/content/aea/aer/2011/00000101/00000003/art00042.

[37] Paul Joskow and Catherine Wolfram. "Dynamic Pricing of Electricity". In: *American Economic Review* 102.3 (2012).

[38] P.L. Joskow and J. Tirole. *Reliability and Competitive Electricity Markets*. Working Paper 10472. National Bureau of Economic Research, May 2004. URL: http://www.nber.org/papers/w10472.

[39] S. Kishore and L.V. Snyder. "Control Mechanisms for Residential Electricity Demand in SmartGrids". In: *Smart Grid Communications*. Oct. 2010, pp. 443–448. DOI: 10.1109/SMARTGRID.2010.5622084.

[40] H.J. Kushner and G.G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, 1997.

[41] G. Lai, F. Margot, and N. Secomandi. "An Approximate Dynamic Programming Approach to Benchmark Practice-Based Heuristics for Natural Gas Storage Valuation". In: *Operations Research* 58.3 (2010), pp. 564–582. DOI: 10.1287/opre.1090.0768. eprint: http://or.journal.informs.org/content/58/3/564.full.pdf+html. URL: http://or.journal.informs.org/content/58/3/564.abstract.

[42] Donghun Lee and W.B. Powell. "An Intelligent Battery Controller Using Bias-Corrected Q-Learning". In: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. Toronto, July 2012.

[43] Y. Liang and Z.-J. Shen. *Managing Electricity Usage for Smart Grid Users*. working paper, Univeristy of California at Berkekely. 2012.

[44] Mark G. Lijesen. "The real-time price elasticity of electricity". In: *Energy Economics* 29.2 (2007), pp. 249 –258. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2006.08.008. URL: http://www.sciencedirect.com/science/article/pii/S0140988306001010.

[45] D. Livengood and R. Larson. "The Energy Box: Locally Automated Optimal Control of Residential Electricity Usage". In: *Service Science* 1.1 (2009), pp. 1–16. DOI: 10.1287/serv.1.1.1. eprint: http://servsci.journal.informs.org/content/1/1/1.full.pdf+html. URL: http://servsci.journal.informs.org/content/1/1/1.abstract.

[46] Isaac Meilijson and Arthur Ndas. "Convex Majorization with an Application to the Length of Critical Paths". English. In: *Journal of Applied Probability* 16.3 (1979), pp. 671–677. ISSN: 00219002. URL: http://www.jstor.org/stable/3213097.

[47] A. Mohsenian-Rad et al. "Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid". In: *Smart Grid, IEEE Transactions on* 1.3 (Dec. 2010), pp. 320 –331. ISSN: 1949-3053. DOI: 10.1109/TSG.2010.2089069.

[48] K. Natarajan, C. Teo, and Z. Zheng. "Mixed 0-1 linear programs under objective uncertainty: A completely positive representation". In: *Operations Research* 59(3) (2011), pp. 713–728.

[49] Karthik Natarajan, Miao Song, and Chung-Piaw Teo. "Persistency Model and Its Applications in Choice Modeling". In: *Management Science* 55.3 (2009), pp. 453–469. DOI: 10.1287/mnsc.1080.0951. eprint: http://mansci.journal.informs.org/content/55/3/453.full.pdf+html. URL: http://mansci.journal.informs.org/content/55/3/453.abstract.

[50] A. Papavasiliou. "Coupling Renewable Energy Supply with Deferrable Demand". PhD thesis. University of California-Berkeley, 2011.

[51] A. Papavasiliou and S. S. Oren. "Supplying renewable energy to deferrable loads: Algorithms and economic analysis". In: *Power and Energy Society General Meeting, 2010 IEEE*. July 2010, pp. 1 –8. DOI: 10.1109/PES.2010.5589405.

[52] A. Philpott and R. Schultz. "Unit commitment in electricity pool markets". In: *Mathematical Programming* 108 (2006), pp. 313–337.

[53] W.B. Powell. *Approximate Dynamic Programming*. New Jersey: John Wiley and Sons, 2007.

[54] M. Roozbehani et al. "Load-Shifting under Perfect and Partial Information: Models, Robust Policies, and Economic Value". Feb. 2012.

[55] P. Samadi et al. "Optimal Real-Time Pricing Algorithm Based on Utility Maximization for Smart Grid". In: *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. Oct. 2010, pp. 415 –420. DOI: 10.1109/SMARTGRID.2010.5622077.

[56] A.M. Sasson and H.M. Merrill. "Some applications of optimization techniques to power systems problems". In: *Proceedings of the IEEE* 62.7 (July 1974), pp. 959 –972. ISSN: 0018-9219. DOI: 10.1109/PROC.1974.9548.

[57] F.C. Schweppe et al. "Homeostatic Utility Control". In: *Power Apparatus and Systems, IEEE Transactions on* PAS-99.3 (May 1980), pp. 1151–1163. ISSN: 0018-9510. DOI: 10.1109/TPAS.1980.319745.

[58] S. Sen, L. Yu, and T. Genc. "A stochastic programming approach to power portfolio optimization". In: *Operations Research* 54 (2006).

[59] A. L. Soyster. "Convex programming with set-inclusive constraints and applications to inexact linear programming". In: *Operations Research* 21 (1973), pp. 1154–1157.

[60] R.S. Sutton and Barto A.G. *Reinforcement Learning: An Introduction*. MIT press, 1998.

[61] S. Takriti, J. R. Birge, and E. Long. "A stochastic model for the unit commitment problem". In: *IEEE Transactions on Power Systems* 11 (1996), pp. 1497–1508.

[62] Qinglai Wei, Huaguang Zhang, and Jing Dai. "Model-free multiobjective approximate dynamic programming for discrete-time nonlinear systems with general performance index functions". In: *Neurocomputing* 72.7C9 (2009), pp. 1839 –1848. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2008.05.012. URL: http://www.sciencedirect.com/science/article/pii/S0925231208003287.

[63] H. Zhong, L. Xie, and Q. Xia. "Coupon Incentive-Based Demand Response (CIDR) in Smart Grid". In: *Proceedings of the IEEE Power and Energy Society General Meeting*. San Diego, 2012.