Reinforcement Learning for Real Options: Interpretable Planning Under Uncertainty and Limited Data

by

Seyed Danial Mohseni Taheri
M.S., University of Illinois at Chicago, Chicago, 2021

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Business Administration
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Defense Committee:
Selvaprabu Nadarajah, Chair and Advisor
Gilbert Bassett
Boxiao (Beryl) Chen
Daniel Jiang (University of Pittsburgh)
Aris M Ouksel
Theja Tulabandhula

ACKNOWLEDGMENT

First, I would like to offer my sincerest gratitude to my advisor Selva Nadarajah whose help and support accompanied me during my Ph.D. and shaped my knowledge about research. This thesis is the result of hours of discussion between us and was not possible without his passion and generosity in educating young minds.

I had wonderful collaborators in my Ph.D. that research could not be an exciting and rewarding experience without them. I want to thank Alessio Trivella, Theja Tulabandhula, Andreas Kleiven, Stein-Erik Fleten for their help and collaboration.

Words cannot begin to describe how grateful I am to my wife, Mehrnaz Amjadi. Her wisdom and patience were big parts of my success, and moments would have been tasteless without her. I feel so lucky to have her next to me on this journey.

I would also like to thank my thesis committee, Gilbert Bassett, Beryl Chen, Daniel Jiang, Aris Ouksel, and Theja Tulabandhula, for their constructive advice and comments to improve my thesis.

Finally, I would like to extend my gratitude and appreciation to my parents (Khadij and Majid) and my brother (Roozbeh), who taught me perseverance and optimism through their endless love and support.

SDMT

CONTRIBUTIONS OF AUTHORS

Chapter 1 presents an overview to position this dissertation in the literature and highlights the significance of its contribution. Chapter 2 represents a manuscript for which I had equal contribution with Dr. Alessio Trivella. I was the lead contributor in the model development and analysis which was the first attempt at formally studying corporate power purchase agreements in operations research and reinforcement learning as highlighted by [3]. My research coauthors Dr. Selva Nadarajah and Dr. Alessio Trivella helped me write the manuscript, conduct experiments, and develop methodologies for the paper. The work in this chapter also led to ned models for corporate procurement analytics in [183]. Chapter 3 represents a manuscript [127] for which I was the first author. I played a larger role in conducting experiments. My coauthors, Dr. Selva Nadarajah and Dr. Theja Tulabandhula helped me with the ideation and writing the manuscript. Chapter 4 represents a manuscript for which I was the first author. I focused on modeling, developing an algorithm, conducting numerical experiments, and writing parts of the manuscript. My co-author, Andreas Kleiven, contributed to the empirical section and helped me with writing the paper. Our research mentors, Dr. Selva Nadarajah and Stein-Erik Fleten, helped us with the idea and feedback regarding practice, respectively. Chapter 5 concludes the thesis with a discussion on future research directions in the field and an overall conclusion.

TABLE OF CONTENTS

CHAPTER PAGE

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

FIGURE **PAGE**

SUMMARY

Reinforcement learning (RL) addresses sequential decision making with the goal of computing a "near-optimal" policy that specifies a decision for each state of the world. Despite the progresses in RL, its deployment in real-world business problems is often hard. Some of the implementation challenges include the complexity of algorithms and/or policies which makes them less understandable to RL non-experts, computational challenges related to the curses of dimensionality, and limited data in modeling the operating environment accurately. In this thesis, we elaborate on these challenges while focusing on applications in the operations-finance area.

In Chapter 2, we study and model a new application in sustainable operations related to renewable power procurement. We focus on companies that have committed to procuring a specified percentage of their annual electricity demand from a renewable power source by a future date. The problem suffers from several **curses of dimensionality** due to the high dimensional state space, high dimensional expectations arising from multiple sources of uncertainty, and non-convexities resulting from business constraints requiring a minimum purchase quantity. We design a new rolling horizon policy based on information relaxation and duality theory with interpretable approximations, and in addition, account for uncertainty directly while computing decisions.

In Chapter 3, we focus on the problem faced by a firm providing services to store ethanol and analyze the behavior of users interacting with this storage provider. Users sign annual contracts

for storage space with a storage provider. However, users' underutilization of the capacity provides the chance for the storage provider to overbook its capacity. The risk exposure of the firm depends on an understanding of user behavior which is limited. Using **limited data** on past injection and withdrawals, we build a sample efficient model of user behavior using Gaussian processes in a non-standard manner that leverages interpretable characterizations of the optimal policy from the operations management literature.

In the final chapter, we investigate the long-term capacity investment problem faced by a hydropower plant. The long horizon of the problem and the presence of multiple underlying variables result in an intractable MDP. Furthermore, limited long-term market data about the evolution of uncertainties, including power price and inflow level, creates substantial planning risk. To tackle these challenges, we design an RL algorithm to hedge against long-term model misspecification risk while mitigating any financial losses due to being overly conservative. We also extend the interpretable reoptimization techniques.

CHAPTER 1

INTRODUCTION

In the last few years, reinforcement learning (RL) has achieved great milestones for planning under uncertainty, with an impressive performance on applications such as game playing and robotics [126, 146, 171]. RL addresses sequential decision making by formulating problems as Markov decision processes (MDP; [149]) with the goal of computing a "near-optimal" policy that specifies a decision for each state of the world that a decision maker can encounter in the future. Despite the strides made in RL, its deployment in real-world business problems is often hard. Some of the implementation challenges include the complexity of algorithms and/or policies which makes them less understandable to RL non-experts, computational challenges related to the curses of dimensionality, and limited data in modeling the operating environment accurately [20, 70, 148, 177]. We elaborate on these challenges below with a bias towards how they are typically encountered in the operations-finance area [10].

- Complex algorithms/policies and interpretability: The decision-making process in real systems is often owned and operated by humans. Interpretability of RL policies by users has received significant recent attention [128,182,190]. Another aspect of RL interpretability has to do with the algorithm itself. Specifically, algorithms with easily understandable approximations to obtain a policy facilitate adoption and confidence in practice. The latter reason is a major contributor to the popularity of model predictive control methods

in operations [51, 89, 112]. These methods are referred to as reoptimization approaches in the operations literature and rely on re-solving deterministic models based on forecasts of uncertain quantities.

- Curses of dimensionality: RL algorithms in research are usually tested and shown to be competitive on examples where the dimensionality of the MDP state and action spaces is small. Many of the traditional RL algorithms have scalability issues and fail to scale up as the complexity of the system increases, as discussed in [69]. Many real systems include complicated dynamics (i.e., multi-factor uncertain variables) with continuous and large-scale state and action spaces. Model-based RL approaches under the umbrella of approximate dynamic programming have been successful at solving structured MDPs [102, 135, 182] and RL algorithms [69, 112, 208].

- Limited data: The reduced cost of data acquisition has caused an explosion in the availability of data for machine learning tasks. This increase in data availability does not directly address the shortage of data for planning especially in long time horizons. Such timelines are common place in business when making investments and other strategic and irreversible decisions in the face of uncertainty. Calibrating models of uncertainty using historical data in these settings can lead to inaccurate representations of the future and thus potentially suboptimal decisions that have significant financial implications. An example is the investment in renewable energy assets where profits/costs depend on uncertainties related to the evolution of technology in storage and power production, climate change and its impact on energy supply/demand, and government policies and environ-

mental factors that effect the usage of energy assets. We highlight two streams of research in RL that focus on limited data. The first develops algorithms that require less data for optimization, that is, these algorithms are data efficient [58, 77, 116]. The second addresses the risk resulting from limited data by improving different notions of worst case performance [120, 178].

My thesis addresses the aforementioned challenges in the context of real options arising in contemporary applications at the interface of operations and finance, including renewable energy, management of storage, and the refurbishment of power plants. Real options studies the adaptation of flexibility to the unfolding of uncertainty, typically market variables such as price, supply, and demand [65]. The resulting decisions are usually taken at multiple connected dates, affected by multiple underlying variables and constraints, and are quantity-based [136, 167]. Such problems usually can be formulated by MDPs which are intractable. The three challenges mentioned earlier manifest themselves in important ways when deploying RL algorithms for real options applications studied in the three chapters of my thesis, as discussed below and highlighted in bold face.

- Chapter 2: We study and model a new application in sustainable operations related to renewable power procurement. We focus on companies that have committed to procuring a specified percentage of their annual electricity demand from a renewable power source by a future date. The problem suffers from several **curses of dimensionality** due to the high dimensional state space, high dimensional expectations arising from multiple sources of uncertainty, and non-convexities resulting from business constraints requiring a minimum

purchase quantity. Rolling planning horizon policies based on forecasts are commonplace as they are easy to understand and implement but do not directly capture uncertainty in planning and can thus be suboptimal. We design a new rolling horizon policy based on information relaxation and duality theory with **easy-to-interpret** approximations, and in addition, account for uncertainty directly while computing decisions.

- Chapter 3: We focus on the problem faced by a firm providing services to store ethanol and analyze the behavior of users interacting with this storage provider. The firm signs annual contracts for storage space and overbooks capacity. The risk exposure of the firm depends on an understanding of user behavior which is limited. Using **limited data** on past injection and withdrawals, we build a sample efficient model of user behavior using Gaussian processes in a non-standard manner that leverages **interpretable characterizations** of the optimal policy from the operations management literature.

- Chapter 4: We investigate the long-term capacity investment problem faced by a hydropower plant. The long horizon of the problem and the presence of multiple underlying variables result in an **intractable** MDP. Furthermore, **limited long-term market data** about the evolution of uncertainties, including power price and inflow level, creates substantial planning risk. To tackle these challenges, we design an RL algorithm to hedge against long-term model mis-specification risk while attempting to mitigate any financial losses due to being overly conservative. We also extend the **interpretable reoptimization techniques**.

This thesis focuses on developing near-optimal methods for solving complex real options problems arising in the aforementioned applications. We achieve this by leveraging ideas from model predictive control (specifically, rolling planning approaches), information relaxations and duality, Gaussian processes, operations management, and finance. The resulting methods are practical approaches that are scalable for commercial applications and novel in several aspects. In Chapter 2, we provide a new class of rolling horizon policies based on information relaxations. In Chapter 3, we use structured and interpretable decision rules from operations management within Gaussian processes to learn user behavior. Finally, in Chapter 4, we combine algorithms for solving MDPs, robust MDPs, and clustering techniques to address model misspecification in the long run.

By developing such approaches, this thesis also broadens the set of applications studied in real options and the operations-finance interface [10]. Specifically, in Chapter 2, we study a new corporate renewable power procurement application, which has not received prior attention in the operations management literature, as highlighted in [3]. In Chapter 3, we extend the study of storage management in the operations literature [136], which is from the perspective of a user who has contracted space, to consider the perspective of an owner who needs to decide how to allocate limited capacity among users with unknown behavior. In Chapter 4, we study a well known refurbishment problem but account for the risk arising from limited data, which has not been previously studied.

Next, we provide a more detailed overview of the applications that have been the cornerstones of this thesis and the related methodological developments. Nevertheless, it is worth

noting that the ideas and methodologies in the thesis have applicability beyond real options to broader applications in operations and finance, as well as, other business and engineering disciplines.

Overview of Applications and Summary of Results

Below, we provide a brief description of the application motivating each thesis chapter and then summarize the contributions.

Chapter 2: Meeting Corporate Renewable Power Targets.

Corporations are playing an increasing leadership role in promoting sustainability and social responsibility around the globe. Over half of the Fortune 500 companies have publicly announced commitments to meet sustainability and climate goals, which includes greenhouse gas emissions reduction, energy efficiency improvements, and renewable power procurement [46]. Prominent companies have committed to procuring a percentage of their power demand from renewable sources by a future date in the face of uncertain power demand and stochastic power and renewable energy certificate (REC) prices [28]. For example, Procter & Gamble and Intel have committed to RPPTs of 30% and 75%, respectively, by 2020 and 2025. Facebook along with a larger set of 138 global companies have pledged an RPPT of 100% as part of the RE100 initiative [45]. Despite these encouraging trends, a large proportion of companies have not yet committed to any renewable energy target due to the lack of strategic knowledge about renewable energy procurement [150]. We aim to take a meaningful step towards reducing this knowledge gap through studying procurement portfolios based on two dominant strategies to achieve this target: long-term procurement of power and RECs at a fixed price using corporate

power purchase agreements (CPPAs) and short-term purchases at volatile prices. We analyze a two-stage model to understand the behavior of procurement costs when using financial and physical CPPA variants employed in practice, which informs the structuring of these contracts. We subsequently formulate a Markov decision process (MDP) that optimizes the multi-stage procurement of power to reach and sustain a renewable procurement target. Our MDP is intractable because its action space is non-convex and its state space has high-dimensional endogenous and exogenous components. Although approximate methods to solve this MDP are limited, a procurement policy can be obtained using an easy-to-implement "primal" reoptimization strategy, which solves a deterministic model with stochastic quantities in the MDP replaced by forecasts. This approach does not, however, provide a lower bound on the optimal policy value. We propose a novel "dual" reoptimization heuristic which computes both procurement decisions and a lower bound while retaining the desirable implementation properties of primal reoptimization using the information relaxation and duality approach [8, 38, 91]. On realistic instances, the dual reoptimization policy is near-optimal and outperforms policies from primal reoptimization and other heuristics. Our numerical results also highlight the benefits of using CPPA contracts to meet a renewable target.

Chapter 3: Interpretable User Models via Decision-rule Gaussian Processes and Storage Overbooking Application.

In this chapter, we focus on the problem faced by a firm providing services to store ethanol – a real application that motivated this work. Suppose capacity (in gallons) is sold via annual contracts to N users. The contract of user $\mathfrak{n}$ specifies the maximum amount of ethanol that

can be stored. User behavior corresponds to the injection of ethanol and the withdrawal of previously injected ethanol, which can be modeled as a time series. The inventory in storage associated with user n at time t is the net of past injections and withdrawals. Models of user behavior are critical inputs in many prescriptive settings and can be viewed as decision rules that transform state information available to the user into actions. Gaussian processes (GPs), as well as nonlinear extensions thereof, provide a flexible framework to learn user models in conjunction with approximate Bayesian inference [14, 26]. However, the resulting models may not be interpretable in general. We propose decision-rule GPs (DRGPs) that apply GPs in a transformed space defined by decision rules that have immediate interpretability to practitioners. We illustrate this modeling tool on a real application and show that structural variational inference techniques can be used with DRGPs. We find that DRGPs outperform the direct use of GPs in terms of both out-of-sample performance and the quality of optimized decisions. These performance advantages continue to hold when DRGPs are combined with transfer learning.

Chapter 4: Repowering Power Plants Under Limited Long Term Information.

Hydropower plants are the dominant producers of renewable power worldwide, which constitutes over fifty percent of global renewable capacity [153]. Unlike intermittent renewable energy sources such as wind and solar, hydropower can sometimes be stored in reservoirs and is a flexible source of renewable power. In decentralized power markets, the decision to refurbish and upgrade is a firm-level decision and exhibits features such as long lifespan, partially unknown breakdown risk, irreversibility of the investment decision, and uncertainty with regard to climate and future markets. Moreover, cashflows from hydropower plants come primarily from

supply to organized markets. The firm aims to establish operational schedules that maximize the discounted expected profits over a given planning horizon, subject to a given capacity level and other relevant constraints. This means that investing in capacity needs to account for the value that results from changed operations. To address heterogeneous data availability over time, we propose to leverage useful information in the short- and medium-term to calibrate a stochastic model governing the evolution of prices as long as there are sufficient liquid markets to be used in real options valuation. To hedge against unreliable or non-existing data, we propose to consider policies aligned with worst-case scenarios. We model our problem as a Markov Decision Process (MDP) with a terminal value defined as a robust MDP. This formulation aims to maximize the sum of the expected revenues from electricity generation in the short term and worst-case revenues in the long term. The problem of combined operations and investment planning in the presence of uncertainty leads to a high-dimensional MDP with non-convex actions space which is very challenging to solve. Motivated by reoptimization of decisions on a regular basis after new information is revealed, and theoretical results on forecasts horizons which emphasizes that there is a diminishing effect on future data on initial decisions [47, 53], we propose to formulate an MDP, assuming full information about the data generating process as long as there is sufficient information available. In the long run, we switch to a robust framework and do not impose distributional information to avoid model misspecification.

We conduct numerical experiments based on the data of a real hydropower plant with a planning horizon of 20 years and apply the dual reoptimization heuristic (DRH) in [184] combined with robust value iteration to find a feasible MDP-policy and assess the value of the

policy against an upper bound [40]. Our results confirm the robustness of the expected cash flow of the policy achieved by this framework under various scenarios for the climate conditions and power markets.

CHAPTER 2

MEETING CORPORATE RENEWABLE POWER TARGETS

(Joint work with Alessio Trivella and Selvaprabu Nadarajah)

2.1 Introduction

Corporations are playing an increasing leadership role in promoting sustainability and social responsibility around the globe. Over half of the Fortune 500 companies have publicly announced commitments to meet sustainability and climate goals, which includes greenhouse gas emissions reduction, energy efficiency improvements, and renewable power procurement [46]. We focus on companies that have committed to renewable power purchase targets (RPPT), that is, they procure a specified percentage of annual electricity demand from renewable power sources by a future date and sustain this level of renewable procurement thereafter. For example, Procter & Gamble and Intel have committed to RPPTs of 30% and 75%, respectively, by 2020 and 2025. Facebook along with a larger set of 138 global companies have pledged an RPPT of 100% as part of the RE100 initiative [45]. To meet these targets, power purchases need to be coupled with renewable energy certificates (RECs), where each REC allows its owner to validate the use of one megawatt hour (MWh) of renewable power.

Corporate power purchase agreements (CPPAs) are used by companies to procure renewable power directly from the generator instead of going through a utility. A CPPA is a long-term bilateral contract between the company and a renewable generator to receive a fixed quantity

of power and RECs at a predetermined "strike" price for each year of the contract's tenor (usually between 5 and 25 years). The purchase of power using CPPAs has increased 20% from 2015 to 2017, reaching a record 7.2 gigawatts of power purchased with CPPAs in 2018 [13, 28]. Despite these encouraging trends, a large proportion of companies have not yet committed to any renewable energy target due to the lack of strategic knowledge about renewable energy procurement [150]. Moreover, the procurement problem faced by a firm that plans to setup and meet an RPPT has not been formally studied in the extant academic literature to the best of our knowledge. The goal of this paper is to take a meaningful step towards reducing this knowledge gap by studying two dominant power procurement options to satisfy an RPPT [29]: (i) enter into CPPAs, and (ii) buy short-term power from the grid, as needed, and supplement it with unbundled REC purchases[1].

CPPAs fall into two main categories: physical and synthetic [193]. Physical CPPAs, which we refer to as physical contracts (PCs), involve the delivery of power from the producer to the consumer. In contrast, synthetic CPPAs (henceforth synthetic contracts and abbreviated SCs) are financial agreements where the producer sells power to the grid, the firm buys power from the grid, and payments for differences relative to the strike price are made to ensure a price hedge. SCs account for the majority of CPPA contracts signed by corporations since 2015 [12]. The CPPA strike price is typically fixed, which helps a firm manage its procurement

[1] A short-term power purchase is akin to a spot purchase. The actual nature of a short-term purchase could vary by region. For instance, in the United States, such a purchase could represent power from an index-based pricing program [64].

costs. Some companies have also opted to use a variant of the fixed strike price, known as the interval strike price, that allows the strike to vary within a predefined interval similar to a collar option [1,192]. In this paper, we provide (i) analytical results to better understand the influence of the aforementioned CPPA structures and the RPPT on procurement costs and (ii) propose decision heuristics to help companies construct dynamic procurement policies to meet an RPPT.

Our analysis considers a two-stage setting where the long-term power purchase option involves a single CPPA. We characterize how the optimal expected procurement cost varies with the RPPT when using PCs and SCs, and determine that this cost under the latter contract is lower than with the former contract. Nevertheless, a drawback cited in the practitioner literature is that using SCs, as opposed to PCs, leads to increased variability in procurement costs [88]. We find that this drawback is indeed true when the same quantity of power is purchased via a PC and an SC. However, contrary to this sentiment, an optimized portfolio containing an SC reduces the quantity of short-term power and RECs purchased relative to a comparable portfolio with a PC, which in turn can decrease the variability of procurement costs under the former portfolio. In the same spirit, specifying an RPPT as a percentage of known past demand (i.e., 60% of 2016 demand) is easier to track and would seem preferable to the target being a fraction of uncertain future demand. However, we find that stochastic RPPTs can reduce procurement costs when demand exhibits a negative drift, which is likely for firms that are investing in energy efficiency initiatives. Finally, we show that interval strike prices employed in industry can reduce the procurement cost relative to a fixed strike price only in a

market where power prices are skewed. Hence, interval strike prices must be used with caution since the behavior of power prices in several markets change over time.

To compute dynamic procurement decisions, we formulate a Markov decision process (MDP) that minimizes the expected procurement cost to meet and sustain an RPPT. The planning horizon in this MDP is divided into a reach period where the target does not have to be fulfilled (but contracts can be signed) and a sustain period where the target must be satisfied. At each stage, the company decides whether to enter into new CPPAs. The strike prices of CPPAs are specified by a model that factors the effects of the expected power price and the return on investment required by the generator in a manner that is consistent with publicly available software from the National Renewable Energy Laboratory (NREL; [140]). The set of available CPPAs depends on the contracts offered by generators, which is unpredictable over time. Moreover, when entering a CPPA, the associated purchase quantity needs to satisfy minimum and maximum requirements. Modeling these features introduces non-convexities in the set of feasible procurement decisions. Approximate methods to solve our MDP are limited because of its non-convex action set and its state space including (i) the pipeline of power inventories from previously signed CPPAs, which defines a high-dimensional endogenous state component and action space, and (ii) multiple stochastic factors driving the evolution of power/RECs prices and demand, which is a high-dimensional exogenous state component.

Despite the aforementioned intractability, a procurement policy can be obtained using a "primal" reoptimization heuristic (PRH) that is extensively used in practice for stochastic control applications, often referred to as the rolling-planning approach [18, 47, 195]. PRH replaces

the stochastic quantities in the MDP by their forecasts and tackles a deterministic model at each stage, which involves solving a linear or mixed-integer program using a commercial optimization solver. It is thus easy to implement but does not compute a lower bound on the optimal policy cost, which is needed to assess the suboptimality of a policy. We develop a novel "dual" reoptimization heuristic (DRH) to simultaneously compute decisions and a lower bound while retaining the favorable implementation properties of PRH. DRH involves two key steps at a given stage and state: (i) it solves deterministic (hindsight) optimization models along sample paths in Monte Carlo simulation with costs corrected by a dual penalty term based on the information relaxation and duality approach [8,38,91]; and (ii) it extracts a non-anticipative decision from a distribution of anticipative actions across sample paths using a function that we refer to as a decision measure. Examples of a decision measure include the mean, median, and mode of a distribution. Repeating this procedure rolling forward in time provides the DRH policy. We specify when a decision measure leads to a feasible procurement policy and leverage the theory on information relaxations to show that this policy is optimal when using an ideal dual penalty.

We conduct numerical experiments on realistic instances with CPPA contract lengths ranging from 5 to 25 years and a planning horizon of 40 years. We calibrate stochastic processes for the evolution of uncertain quantities using market data and the practitioner literature. We compare the DRH procurement policy against the one from PRH and two additional problem-specific benchmarks. The first heuristic relies on short-term purchases of power and RECs alone, that is, it does not consider CPPAs. The second uses a single CPPA and renews this contract each time it expires, also allowing for short-term power purchases. The procurement

policies computed by DRH are near-optimal and result in lower procurement costs compared to the remaining benchmarks. In particular, the average DRH and PRH optimality gaps are 2.6% and 5.6%, respectively, for PCs, and 4.1% and 6.5% for SCs. The remaining benchmarks have average optimal gaps of 7.8–16.7% for PCs and 11.2–21.5% for SCs. The time required by both PRH and DRH to compute a procurement decision at a given stage and state is small – less than one second. Estimating policy values (i.e., upper bounds) is more time consuming, with PRH and DRH requiring 0.5 and 3.5 hours, respectively. Therefore, one incurs a higher computational burden when assessing the performance of the DRH policy but this overhead can be substantially reduced via parallelization.

Overall, our analytical results suggest that SCs and PCs yield significantly different procurement quantities and costs when included in a procurement portfolio – they are thus not merely physical and financial versions of CPPAs. Moreover, these findings provide insights relevant to structuring portfolios with CPPAs, in some cases, contrary to the intuition in practice. Our numerical experiments indicate that portfolios that include multiple CPPAs, in addition to short-term purchases, can significantly reduce procurement costs, especially for companies with aggressive RPPTs. Such portfolios also lead to procurement costs that are stable when contract availability and the market dynamics of REC prices change. Furthermore, DRH outperforms traditional rolling-planning methods (akin to PRH) used in practice and its near-optimal performance suggests that rolling-planning approaches can be extended to effectively compute dynamic procurement portfolios. The relevance of DRH extends beyond our specific

procurement setting to other applications that give rise to MDPs with continuous and high-dimensional state spaces and non-convex action sets.

2.1.1 Novelty and related work

We build on the extant literature that studies commodity procurement using spot purchases and forward contracts [34, 105, 114, 164] as well as procurement in supply chains via short- and long-term contracts, including dual- and multi-sourcing options [7, 122, 181, 189]. Our study of renewable power procurement adds to this line of work. Specifically, our focus on constructing procurement portfolios to meet an RPPT, related insights, and the DRH method are new to this literature. Moreover, the long-term contracts that we consider, that is CPPAs, have unique structure. For instance, CPPAs deliver power at each period over the tenure of the contract, which differs from the long-term contracts considered in the aforementioned papers.

Our work indeed contributes to the growing literature on renewable energy. Several studies in this area study important market level issues related to supply intermittency [2, 97, 204, 211], power supply equilibria [5, 176], support schemes and their impact on renewable energy investments [67, 99, 172], and market-based or equilibrium-based pricing of feed-in tariffs and CPPAs [6, 41, 154, 203]. We instead investigate a problem faced by a corporation, that is, a firm-level decision problem as opposed to a market level issue. We also do not focus on pricing CPPAs nor do we use an equilibrium model for this purpose. Rather, we obtain CPPA strike prices in our MDP by employing a modified net present value calculation consistent with the procedure in the publicly available software SAM from [140].

A more closely related research subarea focuses on individual players in the renewable power market. In particular, this stream of research studies the valuation and operations of renewable generators and operators of storage and transmission assets (see, e.g., [62,101,104,142,211], as well as the management of consumer incentive programs such as demand response [48,194], and references therein). To the best of our knowledge, a study of the power procurement problem faced by a corporation with an RPPT is new to the renewable energy literature.

The reoptimization methods we consider add to existing rolling-horizon planning approaches, also known as certainty equivalent control, which have been successfully used for decision making under uncertainty in engineering and business applications [18,47]. In the context of energy, reoptimization models are popular for determining the next day unit commitment and real-time economic dispatch of power generators [123,124,195]. They have also been used in real option settings, most notably for managing energy storage [108,131,205]. Our development of DRH introduces a new reoptimization scheme to this literature. Moreover, our extensive numerical study expands the set of applications for which reoptimization has been considered and shows that DRH can outperform PRH.

In addition to the reoptimization literature, DRH contributes to the active research on the information relaxation and duality approach [35–37,92,131,207], which does not directly provide control policies. Therefore, [63] design an auxiliary procedure to obtain decisions in this framework. Specifically, they estimate a value function approximation by regressing on value function estimates computed by solving dual optimization problems in Monte Carlo simulation. This approximation is then used along with the MDP Bellman operator to compute decisions.

It is not easy to extend the approach of [63] to our setting because estimating a value function approximation and computing decisions using the Bellman operator are both challenging due to the large endogenous state space in our MDP. Our development of DRH thus adds a direct way to obtain non-anticipative controls when using the information relaxation and duality approach.

More broadly, DRH adds to approximate dynamic programming (ADP; [18]), an area of stochastic optimization dealing with the solution of high-dimensional MDPs. Several ADP methods tackle MDPs where either the endogenous state or the exogenous state is high-dimensional. Well-known examples include least squares Monte Carlo [117, 185], approximate linear programming [57], and stochastic dual dynamic programming [143, 169]. However, methods to approximately solve MDPs with high-dimensional endogenous and exogenous state components are limited (see, e.g., [131, 158]) and approaches that handle non-convex action sets are even more scarce. DRH has potential value for solving MDPs with these complicating features, which arise beyond our specific procurement application.

2.1.2 Paper structure

In §2.2, we analyze CPPAs in a two-stage procurement setting. In §2.3, we formulate a strike price model and a multi-stage MDP to reach and sustain an RPPT. In §2.4, we present reoptimization methods to obtain policies and lower bounds for this MDP. We conduct an extensive numerical study and discuss our findings in §2.5. We conclude in §2.6. All proofs and additional material related to our models and numerical study can be found in an online supplement.

2.2 Corporate power purchase agreements

In this section, we analyze different CPPA structures using simplified models that capture key trade-offs. In §2.2.1, we characterize the behavior of procurement quantities and costs as functions of the RPPT. In §2.2.2, we study the impact of allowing the traditionally fixed strike price to vary within an interval.

2.2.1 Procurement costs and targets

We model a two-stage procurement problem with stages 0 (now) and 1 (future), where a company has committed to satisfy an RPPT at stage 1. Thus, the reach and sustain periods are each one stage. To fulfill the target, the company can (i) enter into a CPPA at stage 0 to receive power and RECs from a renewable generator at stage 1, and (ii) procure in stage 1 any unmet power demand and shortfall in the RPPT, after accounting for the stage 0 CPPA purchase, using grid power purchases and unbundled RECs, respectively. These procurement decisions depend on the power price, the REC price, and the power demand at stage $i \in \{0, 1\}$, which we denote by P_i (USD/MWh), R_i (USD/MWh), and D_i (MWh), respectively. To simplify notation we define $w_i := (P_i, R_i, D_i)$. Indeed, at stage 0, the vector w_1 is stochastic. We represent the RPPT as a fraction $\alpha \in [0, 1]$ of the stage-1 firm's power demand D_1; given that D_1 is stochastic at stage 0, the target αD_1 is also stochastic. In this setting, the firm determines a stage 0 CPPA quantity to minimize the expected cost of procuring long-term and short-term power to satisfy its power demand and the RPPT at stage 1. We formulate models considering physical and synthetic CPPA variants.

Signing a PC for z MWh results in the physical delivery of this power at stage 1 and a payment of K USD/MWh. PCs in wind and solar projects often have a "Take and Pay" structure, which obligates the off taker to pay for the contracted energy ([103, 141]). This feature may result in the CPPA delivering more power than the demand at stage 1, which is referred to as volume risk ([12]). In a symmetric fashion, we assume generators must deliver the contracted power, that is, PCs are firm contracts. Given an RPPT α and a strike price K, the stage-1 procurement cost as a function of z and w_1 is

$$\widetilde{C}_{PC}(z, w_1; \alpha, K) := K z + P_1(D_1 - z)_+ + R_1(\alpha D_1 - z)_+, \tag{2.1}$$

where the first term represents the stage 1 cost of procuring z MWh of power through a PC signed at stage 0, and the second and third terms represent the expected cost of fulfilling stage-1 shortfalls in meeting total demand and the RPPT, respectively, using the short-term market. The optimal expected procurement cost of a firm using a PC at stage 0 is thus

$$C_{PC}(\alpha, K) := \min_{z \geqslant 0} \mathbb{E}_0[\widetilde{C}_{PC}(z, w_1; \alpha, K)], \tag{2.2}$$

where we use $\mathbb{E}_0[\cdot] \equiv \mathbb{E}[\cdot|w_0]$ and $(\cdot)_+ \equiv \max\{\cdot, 0\}$ for notational convenience[1].

In contrast to a PC, an SC does not require the physical delivery of power. Instead, the generator sells z MWh to the grid and the company purchases the same amount of power from the grid. If the grid price P_1 is greater than the fixed strike price K, the generator pays

[1] The short-term procurement cost at stage 0 is excluded because it is a constant and does not affect the choice of z.

the company for each MWh the positive difference $P_1 - K$; otherwise, the company pays the generator $K - P_1$. Formally, the firm's stage-1 cost function is

$$\widetilde{C}_{SC}(z, w_1; \alpha, K) := P_1 D_1 + (K - P_1)z + R_1(\alpha D_1 - z)_+, \tag{2.3}$$

where the first term is the cost of purchasing the stage-1 power demand from the grid, the second is the cash flow resulting from difference payments between the generator and the company on the z MWh contracted via the SC when the grid price deviates from the strike price, and the third is the cost of procuring RECs to meet the RPPT shortfall. The optimal expected procurement cost when using an SC is

$$C_{SC}(\alpha, K) := \min_{z \geqslant 0} \mathbb{E}_0[\widetilde{C}_{SC}(z, w_1; \alpha, K)], \tag{2.4}$$

Next we compare models Equation 2.2 and Equation 2.4 under the following assumption.

Assumption 1. *It holds that (i) the strike price* K *belongs to the interval* $\big[\mathbb{E}_0[P_1], \mathbb{E}_0[P_1 + R_1]\big]$*; (ii) the power demand* D_1 *is uniformly distributed in the interval* $[a, b]$*, where* a *and* b *are positive scalars satisfying* $b > a$*; (iii) the power price* P_1 *follows a log-normal distribution; (iv) the expected REC price* $\mathbb{E}_0[R_1]$ *is positive; and (v) the power demand* D_1 *is independent of the prices* P_1 *and* R_1*.*

The domain of the strike price captures practically relevant values for the parameter K. The lower bound of $\mathbb{E}_0[P_1]$ avoids cases where the generator is better off selling its power directly to the grid as opposed to the company via a CPPA, while the upper bound of $\mathbb{E}_0[P_1 + R_1]$ removes situations where the company would save money from procuring power and RECs directly from the short-term market instead of using a CPPA. The log-normal assumption on the

stage-1 power price is consistent with the long-term components of common electricity price models such as one-factor and two-factor mean-reverting stochastic processes used in the literature, which consider the evolution of the logarithm of the power price[1] ([43,118]). We do not assume any specific disitrbutional form for the REC price but requires its mean to be positive, which is consistent with the behavior of REC prices across markets in the United States. The uniformly distributed power demand can be viewed as adding variability around a long-term demand forecast. Finally, our assumption of independence between the power price and the power demand stems from the power demand of an individual company not being large enough to affect the market price and companies in several sectors (e.g., high-tech and education) having limited flexibility to adjust their power consumption to fluctuations in the power price. The independence of demand and REC prices follows similar justification.

Lemma 2.2.1 characterizes the optimal CPPA procurement quantity. Let z^*_{PC} and z^*_{SC} denote the optimal solutions of models Equation 2.2 and Equation 2.4, respectively. Further, we define a target threshold $\bar{\alpha} := a\mathbb{E}_0[R_1]/(b\mathbb{E}_0[R_1] - (K - \mathbb{E}_0[P_1])(b-a))$.

Lemma 2.2.1. *Under Assumption 1, we have*

$$z^*_{PC} = \begin{cases} \alpha\left(b - \frac{K-\mathbb{E}_0[P_1]}{\mathbb{E}_0[R_1]}(b-a)\right), & \text{if } \alpha \leqslant \bar{\alpha}, \\ \frac{-K(b-a)+(\mathbb{E}_0[R_1]+\mathbb{E}_0[P_1])(b)}{\frac{1}{\alpha}\mathbb{E}_0[R_1]+\mathbb{E}_0[P_1]}, & \text{if } \alpha > \bar{\alpha}; \end{cases} \quad \textit{and} \quad z^*_{SC} = \alpha\left(b - \frac{K-\mathbb{E}[P_1]}{\mathbb{E}[R_1]}(b-a)\right).$$

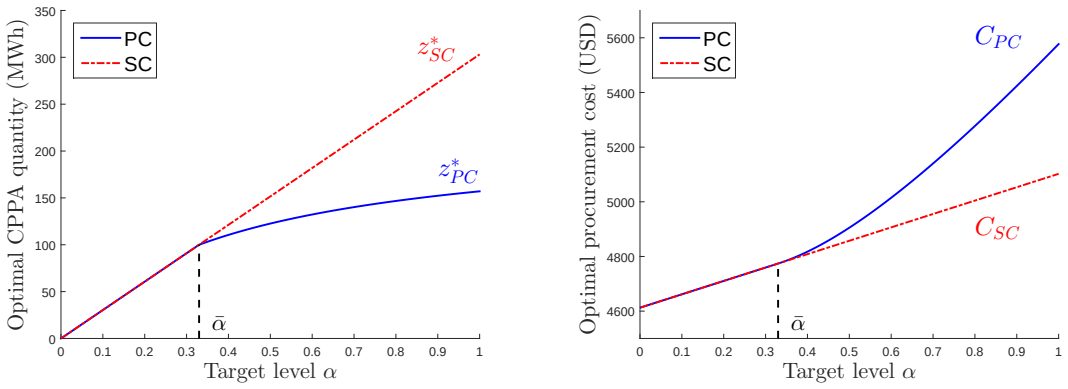
Both the optimal PC and SC procurement quantities are equal and vary linearly with α within the interval $[0, \bar{\alpha}]$. However, for α greater than $\bar{\alpha}$, z^*_{PC} and z^*_{SC} diverge. Specifically, z^*_{SC}

[1] Unlike short-term power prices, the long-term power prices that we model do not take negative values, which justifies considering the evolution of the logarithm of the power price.

continues to vary linearly with α while z^*_{PC} is an increasing concave function of the target. Example 1 and Figure 1(a) illustrate this behavior and show that the optimal PC procurement quantity can be substantially smaller than the SC optimal procurement quantity.

Example 1. *Suppose* D_1 *is uniformly distributed in the interval* $[100, 350]$, P_1 *is lognormal with* $\mathbb{E}_0[\mathsf{P}_1] = 20.5$ *USD/MWh, and* $\mathbb{E}_0[\mathsf{R}_1] = 8$ *USD/MWh. Moreover, the strike price* K *is* 22 *USD/MWh. In this setting, the target threshold* $\bar{\alpha}$ *equals* 0.32. *Figure 1(a) displays the optimal procurement quantities* z^*_{PC} *and* z^*_{SC} *as functions of* α. *For* α *equal to 1,* z^*_{SC} *(= 157 MWh) is roughly 93% smaller than* z^*_{PC} *(= 303 MWh).*

Figure 1: Optimal procurement quantities and costs as a function of the RPPT.
(a) Optimal CPPA procurement quantity (b) Optimal procurement cost



For $\alpha > \bar{\alpha}$, the conservative long-term procurement using a PC can be attributed to over-procurement risk, that is, the event when z is greater than the stage-1 demand. Such over-procurement risk is significant for large α in a PC because one needs to pay for the contracted

power even when all of it is not needed[1]. In contrast, this risk is mitigated when using an SC by the fact that the corporation purchases only the required power from the grid. Over-procurement risk has important practical implications on the procurement cost and its variance, which we study below.

Proposition 2.2.2 characterizes optimal expected procurement costs when using a PC and an SC.

Proposition 2.2.2. *Suppose Assumption 1 is true. The following hold:*

(a) $C_{PC}(\alpha, K)$ *is linear in* α *for* $\alpha \in [0, \bar{\alpha}]$*, and strictly convex in* α *for* $\alpha \in (\bar{\alpha}, 1]$*;*

(b) $C_{SC}(\alpha, K)$ *is linear in* α *for all* $\alpha \in [0, 1]$*;*

(c) $C_{SC}(\alpha, K) = C_{PC}(\alpha, K)$ *for* $\alpha \in [0, \bar{\alpha}]$*, and* $C_{SC}(\alpha, K) < C_{PC}(\alpha, K)$ *for* $\alpha \in (\bar{\alpha}, 1]$.

The behavior of the optimal procurement costs as a function of α are driven by the structure of the optimal procurement quantities in Lemma 2.2.1. Specifically, within the interval $[0, \bar{\alpha}]$, the procurement costs when using a PC and an SC are both equal and linear in α. For $\alpha > \bar{\alpha}$, the former cost is increasing convex while the latter remains linear. This difference can be attributed to an interplay between over-procurement (i.e., $z \geqslant D_1$) and under-procurement (i.e., $z < D_1$) risks. To elaborate, a PC procures less long-term power than an SC for large α due to over-procurement risk as already discussed above. The resulting

[1]Results analogous to the assumed Take-and-pay PC also hold when comparing the optimal procurement quantities of an SC and a Take-or-pay PC. However, here one would need to factor in the penalty a company pays for not taking delivery of the contracted power in the Take-or-pay PC when $z > D_1$. We omit this analysis as it does not provide sufficiently new insights but complicates the exposition of the key differences between a PC and an SC.

smaller z exposes PC to procuring more power and RECs from the short-term market, an expensive option, when an under-procurement event occurs, which amounts to higher expected costs. This finding suggests that an SC allows a company to manage expected procurement costs more efficiently for high RPPTs than a PC. Example 2 illustrates the preceding discussion by considering under- and over-procurement components of the expected stage-1 cost. For example, we have $\mathbb{E}_0[\widetilde{\mathsf{C}}_{\mathrm{PC}}(z^*_{\mathrm{PC}}, w_1; \alpha, \mathsf{K})] = \mathbb{E}_0[\widetilde{\mathsf{C}}_{\mathrm{PC}}(z^*_{\mathrm{PC}}, w_1; \alpha, \mathsf{K}) | z^*_{\mathrm{PC}} \geqslant \mathsf{D}_1] \Pr(z^*_{\mathrm{PC}} \geqslant \mathsf{D}_1) + \mathbb{E}_0[\widetilde{\mathsf{C}}_{\mathrm{PC}}(z^*_{\mathrm{PC}}, w_1; \alpha, \mathsf{K}) | z^*_{\mathrm{PC}} < \mathsf{D}_1] \Pr(z^*_{\mathrm{PC}} < \mathsf{D}_1)$, where we refer to the first and second terms in the right-hand-side of the equality as the over- and under-procurement components of the PC expected cost, respectively. Analogous definitions hold for the SC expected cost.

Example 2. *For the setting considered in Example 1, Figure 1(b) displays the expected procurement costs* $\mathsf{C}_{PC}(\alpha, \mathsf{K})$ *and* $\mathsf{C}_{SC}(\alpha, \mathsf{K})$ *as functions of* α*. The procurement cost when using a* PC *(= 5,570 USD) is roughly 9% greater than the analogous cost under an SC (= 5,102 USD) for* α *equals* 1*. The over- and under-procurement components of the PC procurement costs are 800 USD and 4,770 USD, respectively. The analogous cost components corresponding to SC equal 3720 and 1,382 USD, respectively. Therefore,* z^*_{SC} *being greater than* z^*_{PC} *leads to higher expected procurement costs when there is over-procurement but reduces the exposure of SC to the short-term market. This reduced exposure leads to a much smaller expected cost when there is under-procurement, which leads to* $\mathsf{C}_{SC}(1, \mathsf{K})$ *being strictly smaller than* $\mathsf{C}_{PC}(1, \mathsf{K})$*.*

The expected procurement cost measure discussed above does not consider an important motivation for companies entering into long-term contracts, which is to reduce or eliminate the variability of future costs – stable costs facilitate budgeting. A related sentiment in the

practitioner literature is that using an SC could increase the cash flow variability compared to a PC. Proposition 2.2.3 shows that this sentiment is in fact true when procuring the same quantity of power using both CPPA types. We use $\mathrm{Var}[\cdot]$ to represent the variance of a random variable.

Proposition 2.2.3. *Under Assumption 1, we have* $\mathrm{Var}[\widetilde{C}_{SC}(z, w_1; \alpha, K)] = \mathrm{Var}[\widetilde{C}_{PC}(z, w_1; \alpha, K)]$, *if* $\Pr(z > D_1) = 0$, *and* $\mathrm{Var}[\widetilde{C}_{SC}(z, w_1; \alpha, K)] > \mathrm{Var}[\widetilde{C}_{PC}(z, w_1; \alpha, K)]$, *if* $\Pr(z > D_1) > 0$.

To gain some intuition on this result, note that $\widetilde{C}_{PC}(z, w_1; \alpha, K)$ and $\widetilde{C}_{SC}(z, w_1; \alpha, K)$ are equal if $z \leqslant D_1$, that is, when there is under-procurement. Therefore, if the over-procurement risk is zero, we have the same variance of cash flows under a PC and an SC. Instead, if the over-procurement risk is positive, the comparison of cash flow variance becomes more involved but we can establish that the variance under an SC is greater than with a PC. Nevertheless, Proposition 2.2.3 may not hold for the optimized and unequal power procurement quantities computed by solving Equation 2.2 and Equation 2.4 for $\alpha > \bar{\alpha}$. Example 3 provides an instance where the variance of an optimized portfolio with an SC is smaller than an optimized portfolio containing a PC for large α. This example suggests that optimizing CPPA purchases is important because a suboptimal SC portfolio may lead to high variance in procurement costs but optimizing this portfolio can mitigate this effect.

Example 3. *Consider an instance with* $\alpha = 0.9$, *uniformly distributed demand* D_1 *in the interval* $[100, 200]$, *and strike price* K *equal to* 22 *USD/MWh. For simplicity, we assume a deterministic power price* P_1 *equal to* 20.5 *USD/MWh and a deterministic REC price* R_1 *of* 3.5 *USD/MWh. Invoking Lemma 2.2.1, we find* z^*_{PC} *and* z^*_{SC} *to be* 107 *MWh and* 141 *MWh, respectively. Moreover, we have* $\mathrm{Var}[\widetilde{C}_{PC}(z^*_{PC}, w_1; \alpha, K)]$ *(i.e., optimal PC cost vari-*

ance), $\mathrm{Var}[\widetilde{C}_{SC}(z^*_{PC}, w_1; \alpha, K)]$ *(i.e., variance of suboptimal SC cost evaluated at* z^*_{PC}*), and* $\mathrm{Var}[\widetilde{C}_{SC}(z^*_{SC}, w_1; \alpha, K)]$ *(i.e., optimal SC cost variance) equal to 445,840; 455,160; and 394,330, respectively. These values show that the suboptimal SC portfolio that procures long-term power equal to* z^*_{PC} *has higher variance than the optimal PC portfolio, which is consistent with Proposition 2.2.3. In contrast, in this example, the variance of an optimized SC portfolio is lower than that of an optimal PC portfolio.*

Finally, we discuss two possible choices for reference demand used in practice when specifying an RPPT. Thus far, we have assumed a "stochastic" RPPT, that is, a procurement target defined as a percentage of uncertain future demand. However, some companies define a "deterministic" RPPT with respect to known past demand as this is easier to track ([46]). We analyze these two target types in Online Supplement 2.7.2 and summarize our main finding here. While a deterministic RPPT seems intuitively beneficial, using a stochastic RPPT can lower expected procurement costs when future power is smaller than past demand, for instance, due to investments in energy efficiency improvements. In this case, specifying an $\alpha\%$ deterministic RPPT can lead to procuring more MWh of renewable power than specifying this percentage with respect to uncertain future demand that is unlikely to exceed its historic value. Our findings bode well for the use of stochastic RPPTs given recent efforts by companies to reduce their power consumption.

2.2.2 Interval strike price

An interval strike price is defined by a pair (K, δ), where K is the baseline strike price and δ the half-length of the interval $[K - \delta, K + \delta]$ in which the strike price is allowed to fluctuate.

In particular, if the power price P exceeds the upper bound $K + \delta$, then the generator pays the company the difference $P - K - \delta$ between the power price and this upper bound. Similarly, if the power price P is less than the lower bound $K - \delta$, then the company has to pay the generator the difference $K - \delta - P$. Instead, when P belongs to the interval $[K - \delta, K + \delta]$ no payment between parties occurs. Therefore, unlike the standard CPPA in which the cost of procuring a unit of power is fixed, the interval strike price has the following partially variable cost per unit:

$$K^{INT}(P; K, \delta) = \begin{cases} K + \delta & \text{if} \quad P > K + \delta; \\ P & \text{if} \quad P \in [K - \delta, K + \delta]; \\ K - \delta & \text{if} \quad P < K - \delta. \end{cases} \tag{2.5}$$

We are only aware of SC with interval strike prices and thus focus on this case here, although our results can be easily adapted to the PC setting. The analogue of the procurement problem (Equation 2.4) when using an interval strike price is

$$C_{SC}^{INT}(\alpha, K, \delta) := \min_{z \geqslant 0} \left\{ \mathbb{E}_0[\widetilde{C}_{SC}(z, w_1; \alpha, K^{INT}(P_1; K, \delta))] \right\}. \tag{2.6}$$

When $\delta = 0$, the interval strike price defined in Equation 2.5 becomes $K^{INT}(P; K, 0) = K$ and the optimization model (Equation 2.6) satisfies $C_{SC}^{INT}(\alpha, K, 0) = C_{SC}(\alpha, K)$. Proposition 2.2.4 compares the procurement cost function of interval versus fixed strike prices in SCs that share the same baseline strike price K, but where the former contract can be optimized by choosing the interval length δ. Consistent with Assumption 1, we assume that the power price is log-normally distributed. The mean and standard deviation of power price are $\mathbb{E}_0[P_1]$ and $\mathbb{E}_0[P_1]\sqrt{\exp(\sigma_P^2) - 1}$, respectively, where σ_P^2 denotes the variance of the natural logarithm of the stochastic power price P_1.

Proposition 2.2.4. *There exists a value* $\delta > 0$ *such that* $C_{SC}^{\mathrm{INT}}(\alpha, K, \delta) < C_{SC}(\alpha, K)$ *if and only if* $K > \mathbb{E}_0[P_1]\exp(-\sigma_P^2/2)$. *Moreover, if* $K > \mathbb{E}_0[P_1]\exp(-\sigma_P^2/2)$, *then* $C_{SC}^{\mathrm{INT}}(\alpha, K, \cdot)$ *attains its global minimum at* δ *equal to* $\sqrt{K^2 - \mathbb{E}_0[P_1]^2\exp(-\sigma_P^2)}$.

This proposition shows that an interval strike price can reduce the procurement cost relative to a fixed strike price when K exceeds $\mathbb{E}_0[P_1]\exp(-\sigma_P^2/2)$. This threshold is a decreasing function of the variance of the power price distribution. Thus, as σ_P increases sufficiently, it holds that $\exp(-\sigma_P^2/2) < 1$ implying that the interval strike price can reduce the cost even when the strike price is less than the expected power price. This behavior can be attributed to the positive skewness of the log-normal power price distribution, which is $\exp(\sigma_P^2)\sqrt{\exp(\sigma_P^2) - 1}$. As σ_P increases, the distribution becomes right-skewed, that is, lower power prices become more probable and an interval strike price contract with an appropriately defined half-length δ can benefit from it. Proposition 2.2.4 also characterizes the optimal interval length that minimizes the procurement cost.

Overall, our analysis unveils a potential advantage of interval strike price in SCs, but suggests some caution as this benefit is tied to the skewness of power prices, which can change over time due to changes in mean-reversion among other factors. Empirical evidence from the literature also suggests that the skewness of power prices could be both positive and negative ([44, 80, 106, 118]).

2.3 Dynamic procurement model

In this section, we discuss a dynamic procurement model to assist a firm in meeting a renewable energy target in a multi-period setting. In §2.3.1, we describe the CPPA strike price structure. In §2.3.2, we formulate an MDP that defines an optimal dynamic procurement policy.

2.3.1 CPPA strike price

A renewable power generator typically sets a CPPA strike price to recoup its project investment and maintenance costs as well as a return on investment [140]. In addition, historical data and models from NREL show that the CPPA strike price is affected by several factors including the average quantity of power produced as a fraction of installed capacity (i.e., capacity factor), tax credits, improvements in technology, the contract duration, and the expected power price over the tenor of the contract [66, 139, 202]. We describe below a model that accounts for these factors and determines a CPPA strike price for a given generator and contract. This model provides the strike prices of CPPAs used as input to the procurement model that we describe in §2.3.2.

Consider a renewable power generator that begins production at year i, where i belongs to a discrete set $\mathcal{I} := \{0, \ldots, I-1\}$ containing the years in our planning horizon. The generator has an expected lifetime of L^P years, a capacity factor equal to $\theta_i \in (0, 1]$, and incurs a cost of C_i^{INV} capturing the one-time installation and estimated maintenance costs associated with a MW of production capacity as well as any applicable investment tax credit[1]. We assume that

[1] An investment tax credit represents a one-time federal tax deduction equal to a pre-specified percentage of the installation cost of a renewable power project.

there is a production tax credit[1] of T_i USD per MWh for the next L_i^T years and that future cash flows are discounted at rate $r \in (0, 1]$, which can be chosen to also account for the generator's target return on investment. We begin by computing the fixed strike price $\hat{K}_i$ of a CPPA that spans the lifetime of the generator using the net present value (NPV) of the contract's cash flows. This approach is consistent with the System Advisor Model[2] (SAM; [140]). The NPV of 1 MW of installed capacity contracted via such a CPPA is

$$\text{NPV}_i = \sum_{l=1}^{L^P} r^l \theta_i \hat{K}_i + \sum_{l=1}^{L_i^T} r^l \theta_i T_i - C_i^{\text{INV}}.$$

Setting NPV_i to zero, we obtain the following strike price formula:

$$\hat{K}_i = \frac{1}{\sum_{l=1}^{L^P} r^l} \left[\frac{C_i^{\text{INV}}}{\theta_i} - \sum_{l=1}^{L_i^T} r^l T_i \right]. \tag{2.7}$$

Expression (Equation 2.7) captures the dependence of the strike price on generator vintage (i.e., the year that production begins) by treating the production tax credit T_i, investment cost C_i^{INV}, and capacity factor θ_i as time-dependent quantities. Currently, renewable power generators that start construction before 2020 are eligible for productions tax credits for 10 years from the date the facility starts production [66] but this status-quo is likely to change with government regulation. Investment costs and capacity factors typically decrease and increase, respectively, over time due to improvements in technology. The capacity factor, in addition, exhibits significant inter-region variation. For instance, in the case of wind power, capacity

[1] A production tax credit provides a per-kilowatt-hour tax credit for power generation for a fixed number of future years from the installation of a renewable power project.

[2] SAM is an open source performance and financial tool designed by NREL to access the feasibility of renewable energy projects (e.g., wind, solar, or biomass).

factors in the "internal" regions of the United States are significantly higher than those of coastal regions [202].

Next, we describe how the strike price $\hat{K}_i$ in Equation 2.7 can be modified to account for shorter contract lengths and the expected power price over the tenure of the contract. Consider a CPPA with a duration of m years that is less than the lifetime L^P of the generator. Shorter contracts result in additional cash flow risk over the period of the generator's life time for which they do not generate revenue [1]. We thus define a risk-adjusted strike price $\hat{K}_{i,m} := \hat{K}_i \cdot K^+_m$, where $K^+_m \geqslant 1$ is a risk factor that inflates the strike price if $m < L^P$ and equals 1 otherwise, that is, $\hat{K}_{i,m} = \hat{K}_i$ when the contract spans the life of the generator. The CPPA strike price is not solely determined by NPV but is also tied to the long-term expected power price because higher expected (future) power prices give the generator leverage to increase the CPPA price since the company's outside option is expensive [202]. To account for this effect, we lower bound the CPPA strike price by the average power price over the tenure of the contract, which is $\underline{K}_{i,m} := \left(\sum_{l=1}^{m} \gamma^l \, \mathbb{E}\left[P_{i+l} | P_i\right]\right) / \left(\sum_{l=1}^{m} \gamma^l\right)$, where P_i (USD/MWh) is the power price in year i and $\gamma \in (0, 1]$ a yearly discount factor. Our final strike price expression for a contract delivering power for m years starting in year i is

$$K_{i,m} := \max\left\{\hat{K}_{i,m}, \underline{K}_{i,m}\right\}. \tag{2.8}$$

2.3.2 Markov decision process

We formulate a Markov decision process (MDP) for the dynamic procurement of power to meet an RPPT and satisfy its annual power demand. A firm can enter into CPPAs at each year in the planning horizon represented by $\mathcal{I}$ and/or purchase power and RECs from the short-term

market. We assume that a stochastic RPPT is enforced from year I^R, that is, a percentage $\alpha \in (0, 1]$ of the annual demand in each year $i \in \mathcal{I}^S := \{I^R, \ldots, I-1\}$ must be satisfied. The renewable target α does not have to be fulfilled in the remaining part of the planning horizon (i.e., $\mathcal{I}^R := \{0, \ldots, I^{R-1}\}$) but CPPA contracts can be signed. We thus refer to the years in sets $\mathcal{I}^R$ and $\mathcal{I}^S$, respectively, as the reach and sustain periods.

At each stage i, the set of potentially available CPPAs are indexed by m with ground set $\mathcal{M}$. We assume these contracts are differentiated by their duration so that m can be interpreted as the length of a contract and $M = \max\{m \in \mathcal{M}\}$ is the length of the longest contract[1]. A CPPA of length m signed at stage i delivers power from stages $i+1$ to $i+m$. The strike price $K_{i,m}$ associated with contract m is given by Equation 2.8[2]. The company can choose to enter into a new CPPA of type m, if it is available, by determining a power procurement quantity that is within minimum and maximum allowable limits (often imposed by the generator) represented by $z_m^{\min}$ and $z_m^{\max}$, respectively. We model contract availability at stage i using a binary vector $a_i := (a_{i,m} \in \{0, 1\}, m \in \mathcal{M})$, where $a_{i,m}$ equals one, if contract m is available, and is zero, otherwise. The continuous-valued procurement decision vector is $z_i := \{z_{i,m}, m \in \mathcal{M}\}$, where

---

[1]The definition of the set of contracts in our formulation can be easily extended to differentiate contracts based on features other than length.

[2]Our MDP formulation and solution methodology are flexible to handle other strike price definitions.

$z_{i,m}$ is the size in MWh[1] of the CPPA of length m signed at stage i. Given a contract availability vector a_i, the vector z_i belongs to set

$$\mathcal{Z}_i(a_i) := \{z_i \in \mathbb{R}_+^{|\mathcal{M}|} | z_{i,m} = 0, \text{ if } a_{i,m} = 0, \text{ and } z_{i,m} \in \{0\} \cup [z_m^{\min}, z_m^{\max}], \text{ otherwise, } \forall m \in \mathcal{M}\},$$

which is non-convex when minimum purchase quantities are strictly positive (i.e., $z_m^{\min} > 0$).

The information required to make procurement decisions is described in the MDP state, which contains two components. The first component is a vector $x_i := (x_{i,l}, l \in \{0, \ldots, M-1\})$ representing the on-hand inventory of CPPA contracts, where $x_{i,l}$ is the total power in MWh delivered in year $i+l$ by the on-hand CPPAs. This component is affected by the firm's decisions and is thus referred to as the endogenous MDP state. The second component $w_i := (w_{i,k}, k \in \mathcal{K})$ contains the stochastic factors needed to determine the Markovian evolution of the power price P_i (USD/MWh), REC price R_i (USD/MWh), electricity demand D_i (MWh/year), and the stochastic availability of contracts a_i, where $\mathcal{K}$ denotes the index set of these factors. The complete stage i MDP state is represented by the pair $(x_i, w_i) \in \mathcal{X}_i \times \mathcal{W}_i$.

Executing procurement decisions $z_i \in \mathcal{Z}_i(a_i)$ at stage i and state $(x_i, w_i) \in \mathcal{X}_i \times \mathcal{W}_i$ results in an update of the endogenous state to

$$x_{i+1,l} = f_i(x_i, z_i)_l = \begin{cases} x_{i,l+1} + \sum_{m \in \mathcal{M}: m > l} z_{i,m}, & \text{if } l \in \{0, \ldots, M-2\}; \\ z_{i,M}, & \text{if } l = M-1, \end{cases} \tag{2.9}$$

[1]The MWh quantity is the product of the contracted capacity in MW, the duration of a period in hours, and the capacity factor of the generator. This is reasonable for long-term procurement planning.

where $f_i(x_i, z_i)$ is a vector transition function and $f_i(x_i, z_i)_l$ represents its l-th element. Entering into a CPPA of length m (i.e., $z_{i,m} > 0$) has an associated procurement cost $\sum_{l=1}^{L_{i,m}} \gamma^l K_{i,m} z_{i,m}$ over the tenure of the CPPA, where $L_{i,m} := \min\{m, I - i\}$ equals the number of periods of power delivery within the planning horizon and $\gamma \in [0, 1)$ denotes the discount factor. Demand not met by power from CPPAs at stage i, that is $u_i := \max\{D_i - x_{i,0}, 0\}$, is procured from the short-term market in both the reach and sustain periods at a price P_i USD per MWh. In addition, any shortfall $v_i := \max\{\alpha D_i - x_{i,0}, 0\}$ in meeting the renewable power target during the sustain period requires additional REC purchases at R_i USD per MWh. For each stage $i \in \mathcal{I}$, the cost accrued when entering into PCs is shown below.

$$\textbf{PCs:} \quad c_i(x_i, w_i, z_i) = \sum_{m \in \mathcal{M}} \sum_{l=1}^{L_{i,m}} \gamma^l K_{i,m} z_{i,m} + \begin{cases} P_i u_i, & \text{if } i \in \mathcal{I}^R; \\ P_i u_i + R_i v_i, & \text{if } i \in \mathcal{I}^S. \end{cases} \tag{2.10}$$

When using SCs, the cost incurred at stage i is instead:

$$\textbf{SCs:} \quad c_i(x_i, w_i, z_i) = \sum_{m \in \mathcal{M}} \sum_{l=1}^{L_{i,m}} \gamma^l K_{i,m} z_{i,m} + \begin{cases} P_i (D_i - x_{i,0}), & \text{if } i \in \mathcal{I}^R; \\ P_i (D_i - x_{i,0}) + R_i v_i, & \text{if } i \in \mathcal{I}^S. \end{cases} \tag{2.11}$$

Comparing Equation 2.10 and Equation 2.11 shows that the procurement cost at stage i is the same for PCs and SCs if the on-hand contracts delivering power at i do not exceed demand, i.e. $x_{i,0} \leqslant D_i$. On the other hand, if $x_{i,0} > D_i$, SCs allow the firm to purchase from the grid only the power that is needed to meet demand. In this case, the term $(D_i - x_{i,0})$ is negative. We assume that the terminal costs when employing PCs and SCs are $c_I(x_I, w_I) := P_I u_I + R_I v_I$ and $c_I(x_I, w_I) := P_I(D_I - x_{I,0}) + R_I v_I$, respectively. In other words, only short-term procurement of power and RECs is possible.

A stage i dynamic procurement policy π_i is a collection of stage-dependent decision rules $\{Z_j^{\pi_i}, j \in \mathcal{I}_i\}$, each mapping states to actions, where $\mathcal{I}_i := \{i, \ldots, I-1\}$. A decision rule $Z_i^{\pi_i}$ in stage i is feasible if it associates with each state $(x_i, w_i) \in \mathcal{X}_i \times \mathcal{W}_i$ an action $z_i(x_i, w_i)$ that belongs to $\mathcal{Z}_i(a_i)$. We denote by Π_i the set of all feasible stage i policies. Given an initial state (x_i, w_i) in stage i, an optimal policy in Π_i solves

$$V_i(x_i, w_i) := \min_{\pi_i \in \Pi_i} \mathbb{E}\left[\sum_{j \in \mathcal{I}_i} \gamma^{j-i} c_j(x_j^{\pi_i}, w_j, Z_j^{\pi_i}(x_j^{\pi_i}, w_j)) + \gamma^{I-i} c_I(x_I^{\pi_i}, w_I) \,\middle|\, x_i, w_i\right], \tag{2.12}$$

where $V_i(x_i, w_i)$ is the MDP value function at stage i and state (x_i, w_i), $\mathbb{E}$ is expectation with respect to the future exogenous states, and $x_j^{\pi_i}$ is the endogenous state reached in stage j when following the policy π_i starting from (x_i, w_i).

MDP (Equation 2.12) contains high-dimensional state and action spaces and is challenging to solve directly due to the well-known curse of dimensionality [19, 147]. Specifically, the endogenous state x_i and decision z_i are M- and $|\mathcal{M}|$-dimensional continuous vectors, respectively, and the exogenous state w_i may also be high dimensional when using a multi-factor stochastic model for the evolution of uncertainty. In addition to dimensionality issues, another source of intractability stems from the non-convex structure of $\mathcal{Z}_i(a_i)$. Proposition 2.3.1 provides conditions for which Equation 2.12 is a convex optimization problem with a convex value function in the endogenous state. The convexity of the value function plays an important role when approximating high-dimensional MDPs (see, e.g., [37, 131, 158].

Proposition 2.3.1. *Suppose (i)* $z_m^{min} = 0$ *and* $z_m^{max} < \infty$ *for all* $m \in \mathcal{M}$*, and (ii) expectations and iterated expectations of* P_i*,* R_i*, and* D_i *as well as their products are bounded at all stages*

$i \in \mathcal{I} \cup \{I\}$. *Then the value function* $V_i(\cdot, w_i)$ *is convex for each stage* $i \in \mathcal{I} \cup \{I\}$ *and exogenous state* $w_i \in \mathcal{W}_i$.

We show in Online Supplement 2.7.3 that a non-zero $z_m^{\min}$ leads to a non-convex action set $\mathcal{Z}_i(a_i)$, which can in turn result in a non-convex value function. Thus, computing approximations of MDP (Equation 2.12) using existing ADP methods is challenging, as they typically rely on convexity to handle high-dimensional endogenous states.

2.4 Reoptimization approaches

In this section, we present two reoptimization heuristics that approximate MDP (Equation 2.12) by solving math programs at each stage that are deterministic versions of this MDP. The computational burden of using both heuristics is thus tied to the difficulty of solving the deterministic form of our MDP, which is a favorable algorithmic property, especially in the presence of the non-convexities discussed in §2.3.2. In §2.4.1, we adapt a primal reoptimization heuristic popular in operations applications to our procurement setting. In §2.4.2, we propose a dual reoptimization heuristic based on the information relaxation and duality framework that extracts a non-anticipative policy, in addition to its common use in the literature for estimating a lower bound on the optimal policy cost.

2.4.1 Primal reoptimization heuristic

A primal reoptimization heuristic (PRH) computes procurement decisions as an optimal solution of a math program obtained by replacing random quantities in MDP (Equation 2.12) by their respective forecasts. Forecasts of random prices and demand are their respective expected values. In the case of contract availability, given the binary nature of this variable, we assign

a forecast of 1 if the contract is available with probability greater than 0.5, and 0 otherwise. Formally, the stage-j forecast for contract $m \in \mathcal{M}$ made at stage i, with $j \geqslant i$, is defined as $\bar{a}_{i,j,m} = 1$, if $\mathbb{E}[a_{j,m}|w_i] > 0.5$, and $\bar{a}_{i,j,m} = 0$ otherwise. At stage i and state $(x_i, w_i) \in \mathcal{X}_i \times \mathcal{W}_i$, PRH solves

$$\min_{(y_j, j\in\mathcal{I}_i\cup\{I\};\, z_j, j\in\mathcal{I}_i)} \sum_{j\in\mathcal{I}_i} \gamma^{j-i} c_j(y_j, \mathbb{E}[w_j|w_i], z_j) + \gamma^{I-i} c_I(y_I, \mathbb{E}[w_I|w_i]) \tag{2.13a}$$

$$\text{s.t.: } y_i = x_i, \tag{2.13b}$$

$$y_{j+1} = f_j(y_j, z_j), \quad \forall\, j \in \mathcal{I}_i, \tag{2.13c}$$

$$y_j \in \mathcal{X}_j, \quad \forall\, j \in \mathcal{I}_i \cup \{I\}, \tag{2.13d}$$

$$z_j \in \mathcal{Z}_j(\bar{a}_{i,j}), \quad \forall\, j \in \mathcal{I}_i. \tag{2.13e}$$

This math program computes procurement decisions z_j for stages j from i to I and also includes auxiliary variables y_j that track the endogenous MDP state. Its objective function (Equation 2.13a) is the sum of discounted procurement costs. Constraint Equation 2.13b initializes the stage i state to the current state x_i. Constraints Equation 2.13c enforce the feasibility of state transitions. Constraints Equation 2.13d–Equation 2.13e restrict decision variables to their respective feasible sets.

Using PRH over multiple periods involves reoptimization of the math program Equation 2.13. Specifically, solving this math program at stage i and state (x_i, w_i) provides the procurement decisions $\{z_j^*, j \in \mathcal{I}_i\}$. Among these decisions, we only implement z_i^* corresponding to the current stage, which results in a transition to a new inventory of power $x_{i+1} = f_i(x_i, z_i^*)$. Once new market information w_{i+1} is available at stage $i+1$, we recompute the expectations of uncertainty

$\mathbb{E}[w_j|w_{i+1}]$, $j = i+1, \ldots, I$, and solve an analogue of math program Equation 2.13 formulated using these updated expectations at state (x_{i+1}, w_{i+1}) to obtain z^*_{i+1}. We repeat this procedure until we reach stage I. Estimating the value of the PRH policy, that is an upper bound on the optimal policy cost, involves Monte Carlo simulation of this policy and averaging the resulting sum of discounted costs across sample paths. While PRH provides a policy and an upper bound, it does not provide a mechanism to obtain a lower bound on the optimal policy cost.

The computational burden of using the PRH policy is tied to the complexity of solving Equation 2.13. Since the objective (Equation 2.13a) is piecewise linear convex and the constraints (Equation 2.13b)-(Equation 2.13d) are linear, the optimization problem (Equation 2.13) has a linear programming representation when the constraints (Equation 2.13e) defined using set $\mathcal{Z}_j(\bar{a}_{i,j})$ take a polyhedral form. When $\mathcal{Z}_j(\bar{a}_{i,j})$ is instead a mixed integer set, for instance when minimum procurement quantities $z^{\min}_m > 0$ are enforced as discussed in §2.3.2, the math program (Equation 2.13) becomes a mixed integer program.

2.4.2 Dual reoptimization heuristic

The information relaxation and duality framework [8,38,91] is a popular technique to obtain dual (lower) bounds on the optimal policy cost of intractable MDPs and is applicable to MDP (Equation 2.12) as well. In its most commonly used form, a dual bound is estimated in Monte Carlo simulation by solving a deterministic variant of MDP (Equation 2.12) endowed with full information about future uncertainty and costs adjusted for this knowledge using a dual penalty. Let $q_i(x_i, z_i, W_i)$ denote the stage i dual penalty function, where $W_i := (w_i, w_{i+1}, \ldots, w_I)$ is a vector of realized stochastic factors for each stage from i to I. A feasible dual penalty function

satisfies $\mathbb{E}[q_i(x_i, z_i, W_i)|w_i] \geqslant 0$. Consider the following hindsight optimization problem given knowledge of W_i:

$$V_i^{IR}(x_i; W_i) = \min_{(y_j, j\in\mathcal{I}_i\cup\{I\};\, z_j, j\in\mathcal{I}_i)} \sum_{j\in\mathcal{I}_i} \gamma^{j-i}\Big[c_j(y_j, w_j, z_j) - q_j(y_j, z_j, W_j)\Big] + \gamma^{I-i} c_I(y_I, w_I) \tag{2.14a}$$

$$\text{s.t.: } y_i = x_i, \tag{2.14b}$$

$$y_{j+1} = f_j(y_j, z_j), \quad \forall\, j \in \mathcal{I}_i, \tag{2.14c}$$

$$y_j \in \mathcal{X}_j, \quad \forall\, j \in \mathcal{I}_i \cup \{I\}, \tag{2.14d}$$

$$z_j \in \mathcal{Z}_j(a_j), \quad \forall\, j \in \mathcal{I}_i. \tag{2.14e}$$

Constraints Equation 2.14b-Equation 2.14d are identical to constraints Equation 2.13b-Equation 2.13d in the math program solved by PRH. Constraints (Equation 2.14e) differ from Equation 2.13e in the availability vector used to define $\mathcal{Z}_j(\cdot)$. In the former case, we use the realization of the random contract availability vector on a given sample path while in the latter case we use the contract availability forecast vector described in §2.4.1. The objective (Equation 2.14a) can be obtained by modifying the PRH objective (Equation 2.13a) by subtracting dual penalty terms and replacing the forecasted uncertainty with the elements of W_i. The expectation $\mathbb{E}[V_i^{IR}(x_i; W_i)|w_i]$ taken with respect to the random variable $W_i|w_i$ defines a dual bound on the value function $V_i(x_i, w_i)$, that is the optimal policy value starting from stage i and state (x_i, w_i).

The quality of the dual bound depends on the choice of the dual penalty function in math program (Equation 2.14). Choosing this function to be zero, i.e., $q_i(\cdot,\cdot,\cdot) \equiv 0$, results in the dual bound being equivalent to the well known perfect information bound, which can be

weak. [38] show that the dual bound is instead equal to the optimal policy value when using the following ideal dual penalty based on the MDP value function:

$$q_i(x_i, z_i, W_i) = \gamma\{V_{i+1}(f_i(x_i, z_i), w_{i+1}) - \mathbb{E}[V_{i+1}(f_i(x_i, z_i), w_{i+1}) | w_i]\}. \tag{2.15}$$

Since this ideal dual penalty is not available, approximations are used instead by replacing the exact value function in (Equation 2.15) or using simple forms for penalties (see, e.g., [37, 131, 163]). Since computing a value function approximation in our application is challenging due its high-dimensional state and action spaces and non-convexities, we define the simple dual penalty function

$$q_i(x_i, z_i, W_i) := \sum_{m \in \mathcal{M}} z_{i,m} \sum_{l=1}^{m} \sum_{k \in \mathcal{K}} \gamma^l \theta \left(w_{i+l,k} - \mathbb{E}[w_{i+l,k} | w_i]\right), \tag{2.16}$$

in which the information gained when taking a decision is approximated by spreads between the value taken by the uncertainty in future stage $i+l$ and its expectation computed at stage i. This spread is multiplied by the CPPA purchase decisions. The dual penalty (Equation 2.16) is linear in z_i, does not depend on x_i, and is feasible because the expectation of $w_{i+l,k} - \mathbb{E}[w_{i+l,k}|w_i]$ equals zero. In addition to its simple form, linear penalties ensure that the math program (Equation 2.14) falls into the same complexity class as the deterministic version of MDP (Equation 2.12) and the math program (Equation 2.13) solved by PRH.

Traditionally, an operating policy is computed independent of the dual bound computation described above (see, e.g., [63] and the related discussion in §2.1.1). We now discuss an approach to define a non-anticipative decision directly during the dual bound estimation process, where being non-anticipative refers to a decision that only depends on the information available at

stage $\mathtt{i}$ (the decision resulting from solving Equation 2.14 is anticipative as it relies on future information on the sample path $W_\mathtt{i}$). Since the dual bound $\mathbb{E}[V_\mathtt{i}^{\mathrm{IR}}(x_\mathtt{i}; W_\mathtt{i})|w_\mathtt{i}]$ involves solving the math program (Equation 2.14) over multiple realizations of the random variable $W_\mathtt{i}|w_\mathtt{i}$, we also have a distribution of optimal solutions of this math program. We focus on stage $\mathtt{i}$ decisions obtained during this dual estimation process and represent them via a random decision $z_\mathtt{i}(W_\mathtt{i})$ that is a function of the random variable $W_\mathtt{i}|w_\mathtt{i}$. Our key idea is to define a functional that operates on the distribution of the random variable $z_\mathtt{i}(W_\mathtt{i})$ and returns a single non-anticipative decision. We call this functional a decision measure $\mathcal{H}_\mathtt{i}$ that maps a distribution $z_\mathtt{i}(W_\mathtt{i})$ to a vector of $\mathbb{R}^{|\mathcal{M}|}$. Proposition 2.4.1 establishes some useful properties of a decision measure with respect to optimality and feasibility.

Proposition 2.4.1. *The decision $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i}))$ is guaranteed to be feasible, that is $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i})) \in \mathcal{Z}_\mathtt{i}(a_\mathtt{i})$, if any of the following conditions hold:*

1. *$\mathcal{Z}_\mathtt{i}(a_\mathtt{i})$ is convex and $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i})) := \mathbb{E}[z_\mathtt{i}(W_\mathtt{i})|w_\mathtt{i}]$.*
2. *$\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i})) = \bar{z}_\mathtt{i}$ and there exists a realization $W_\mathtt{i}'$ of the variable $W_\mathtt{i}|w_\mathtt{i}$ such that $\bar{z}_\mathtt{i} = z_\mathtt{i}(W_\mathtt{i}')$.*

Moreover, if the decision $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i}))$ satisfies one of the conditions above and we use the ideal dual penalty (Equation 2.15) in math program (Equation 2.14), then $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i}))$ is an optimal solution to MDP (Equation 2.12).

Examples of $\mathcal{H}_\mathtt{i}$ that satisfy conditions 1 and 2 of Proposition 2.4.1 are $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i})) = \mathbb{E}[z_\mathtt{i}(W_\mathtt{i})|w_\mathtt{i}]$ and $\mathcal{H}_\mathtt{i}(z_\mathtt{i}(W_\mathtt{i})) = \mathbb{M}[z_\mathtt{i}(W_\mathtt{i})|w_\mathtt{i}]$, respectively, where $\mathbb{M}$ denotes the median oper-

ator of a distribution. It is important to note that condition (2) of Proposition 2.4.1 ensures the feasibility of $\mathcal{H}_i(z_i(W_i))$ even when $\mathcal{Z}_i(a_i)$ is non-convex. Thus, the median decision measure is more robust in terms of feasibility than the mean decision measure since the latter measure satisfies only condition (1) of Proposition 2.4.1. It also follows from this proposition that any feasible decision measure returns optimal decisions when using an ideal dual penalty. Indeed, decision measures other than the ones discussed above can be defined, for instance, the mode of the action distribution.

Computing a dual bound and decision at stage i and state (x_i, w_i) entails Monte Carlo simulation. Specifically, we generate H Monte Carlo sample paths of uncertainty $\{w_j^h, (j,h) \in \mathcal{I}_i \cup \{I\} \times \{1,\dots,H\}\}$ which provides a discrete approximation $\hat{W}_i|w_i$ of the random variable $W_i|w_i$. On this approximation, we estimate both a dual bound $\sum_{h=1}^{H} V_i^{IR}(x_i; W_i^h)/H$ and a decision $\mathcal{H}_i(z_i(\hat{W}_i))$, which requires the solution of H math programs of the type Equation 2.14. We apply the aforementioned decision to move to an endogenous state $x_{i+1} = f_i(x_i, \mathcal{H}_i(z_i(\hat{W}_i)))$. Then we observe the stage $i+1$ uncertainty w_{i+1} and repeat the same process at state (x_{i+1}, w_{i+1}) and keep moving forward in time until we reach the terminal stage. We call the resulting approach the dual reoptimization heuristic (DRH) and the policy computed in the process using decision measures as the DRH policy.

Dual bound estimation and the computation of the DRH policy may be expensive if the math program (Equation 2.14) is non-convex. To reduce this computational burden, one can solve a convex relaxation of math program (Equation 2.14), which is derived by replacing constraint (Equation 2.14e) by $z_j \in \text{conv}(\mathcal{Z}_j(a_j))$, for all $j \in \mathcal{I}_i$, where $\text{conv}(\mathcal{Z}_j(a_j))$ denotes the

convex hull of set $\mathcal{Z}_j(a_j)$. Let $\bar{V}_i^{IR}(x_i; W_i^h)$ and $\bar{z}_i(\hat{W}_i)$ denote the optimal objective function value and stage i decision obtained by solving this relaxation. Then it follows immediately that $\sum_{h=1}^{H} \bar{V}_i^{IR}(x_i; W_i^h)/H$ defines a valid lower bound estimate that is less than or equal to $\sum_{h=1}^{H} V_i^{IR}(x_i; W_i^h)/H$. However, $\mathcal{H}_i(\bar{z}_i(\hat{W}_i))$ may not belong to $\mathcal{Z}_i(a_i)$ since $\bar{z}_i(\hat{W}_i)$ is only guaranteed to be a part of $\text{conv}(\mathcal{Z}_i(a_i))$ and not necessarily in $\mathcal{Z}_i(a_i)$. Feasibility can be ensured by projecting the decision $\mathcal{H}_i(\bar{z}_i(\hat{W}_i))$ onto $\mathcal{Z}_i(a_i)$, potentially approximately.

2.5 Numerical study

In this section, we assess the performance of PRH, DRH, and simpler heuristics on realistic instances. We describe our instance sets in §2.5.1. In §2.5.2, we introduce benchmark policies and the computational setup. In §2.5.3, we discuss our findings regarding the relative performance of methods. In §2.5.4, we describe the procurement insights resulting from the numerical study.

2.5.1 Instances

We start by briefly introducing the stochastic processes we employed to model the evolution of the power price, the REC price, the power demand, and the CPPA contract availability. We model power price using a mean-reverting stochastic process with seasonality and jumps as this incorporates the main features of spot electricity markets [118, 199]. We employ a Jacobi diffusion process to forecast the REC price following [209]. Power demand evolves according to a geometric Brownian motion, which is a common choice in practice and in the procurement literature to describe the company's demand uncertainty [17, 107, 164]. Uncertainty in the availability of CPPA $m \in \mathcal{M}$ follows a Bernoulli random variable, where $p_m \in [0, 1]$ represents

the probability that this contract is available. We relegate to Online Supplement 2.7.4 a detailed discussion of these models and the calibration of power and REC model parameters using market data.

TABLE I: Parameters defining the baseline CPPA instance.

Name	Value	Unit	Name	Value	Unit	Name	Value	Unit
I	40	years	C_0^{INV}	1.7×10^6	USD/MW	α	90%	-
I^R	5	years	L^P	30	years	ξ	1%	-
$\mathcal{M}$	$\{5, 10, 15, 20, 25\}$	years	θ_i	3,066	hours/year	γ	0.97	-
z_m^{min}	$20\,\theta_i\ (\forall m \in \mathcal{M})$	MWh	L_i^T	10	years	r	0.94	-
p_m	$\{0.4, 0.5, 0.6, 0.7, 0.4\}$	-	T_i	23	USD/MWh	K_+^5	1.1	-

In Table I, we summarize the parameters defining our MDP of §2.3 in the baseline CPPA instance. We consider a 40 year planning horizon (I) and a 5 year reach period (I^R) to attain a stochastic RPPT that is 90% (α). Our set of contracts $\mathcal{M}$ and the minimum quantities z_m^{min} are consistent with the CPPA portfolio of Google [86], and we set a very loose upper bound on these quantities ($z_m^{max} = 1000\,\theta_i$ MWh). The corresponding availability factors p_m are chosen based on [202] and [12]: According to the first report, 15- to 25-year CPPAs are predominant, with 20-year contracts being the most common, while the second report indicates that CPPAs of length between 10 and 20 years are prevalent. Following [139], we use a functional form for C_i^{INV} that decreases over time by a fixed percentage ξ; specifically, it evolves according to a learning model $C_i^{INV} = C_0^{INV}(1-\xi)^i$. We chose the initial cost (C_0^{INV}) based on 2015–2016 wind projects in the U.S. [74] and the learning rate (ξ) based on the range of values in [139]. Wind turbines are usually designed to operate for 20–25 years but many remain operational for

a longer period of time [212], thus we select the lifetime (L^P) to be 30 years also to account for improving technology. The capacity factor (θ_i) of 35% is representative of the observed average for wind farms in the United States [73] and is assumed to be fixed throughout the planning horizon. The duration of the production tax credit (L_i^T) and its amount (T_i) are based on United States policy in 2016 [66]. Moreover, we assume the tax credit expires in 5 years, i.e., it is only granted to renewable energy facilities commencing construction at stages $i < 5$. The risk-free rate (γ) is chosen equal to the 10-year United States treasury rate in May 2018 [27], and we set the generator discount factor (r) such that its respective return on investment is roughly twice the risk-free interest rate. We use a maximum risk factor $K_+^m = 1.1$ for $m = 5$, i.e. a 10% premium for the 5-year CPPAs, which decreases linearly as m is increased.

In Table II, we perturb the parameters defining our baseline instance to obtain instance sets S1–S5, comprising of 17 instances in total. These instances allow us to analyze the robustness of methods and the behavior of procurement policies as market parameters change. We describe how these perturbed instances were obtained in §2.5.3.

TABLE II: Extended instance sets with the baseline-instance parameter superscripted by B.

Set	Modified parameter	Values
S1	Renewable energy target α	$\{0.6, 0.7, 0.8, 0.9^B, 1.0\}$
S2	CPPA availability p_m for all $m \in \mathcal{M}$	$\{-0.2, -0.1, 0^B, +0.1, +0.2\}$
S3	Long-term mean of power price	$\{20, 30, 39.7^B\}$ USD/MWh
S4	Long-term mean of RECs price	$\{5, 9.4^B, 20\}$ USD/MWh
S5	Generator discount factor r	$\{0.9, 0.91, 0.92, 0.93, 0.94^B\}$

2.5.2 Procurement heuristics and computational setup

Our computational study compares the reoptimization heuristics PRH and DRH described in §2.4.1 and §2.4.2, respectively, with two simpler procurement heuristics described below.

The first procurement heuristic involves only short-term procurement, that is, the entire power demand D_i is purchased on a short-term basis in each stage $i \in \mathcal{I} \cup \{I\}$. A portion αD_i of unbundled RECs is also procured in the sustain period $\mathcal{I}^S \cup \{I\}$ to meet the RPPT. This policy has no demand risk but is fully exposed to volatile power and REC prices. Since stages correspond to years in our numerical setting, here and in the rest of this section we refer to "short-term" power purchase to a yearly average purchase[1], as opposed to the long-term (i.e. multi-year) power delivery from CPPAs.

The second, referred to as the *block* heuristic (BH_m), is parameterized by $m \in \mathcal{M}$. BH_m uses a single CPPA of length m and renews it every m years, that is, each time a contract expires a new one of the same length is signed. Specifically, the first contract is entered at the last year of the reach period, I^{R-1}, and delivers renewable power during the first m years of the sustain period. The second contract is ordered one year before the first contract expires to ensure the continuous delivery of power from CPPAs. This process is repeated until the end of the planning horizon. The quantity $z_{i,m} \in \mathcal{Z}_i(a_i)$ associated with a new contract signed at stage i is obtained by solving a deterministic model that minimizes the procurement cost given forecasts of demand, CPPA availability, and power and REC prices over the delivery period

[1] Our MDP/methods can handle multiple settlements in a year, e.g. monthly, but will require higher simulation time to estimate costs.

of $\mathfrak{m}$ years. Any shortfall in meeting demand or the RPPT using the incumbent CPPA is addressed via purchases of short-term power and/or unbundled RECs.

TABLE III: Summary of methods.

	Method	Description
(Simple heuristics)	Spot	Spot purchase of power and RECs
	$\text{BH}_{\mathfrak{m}}$	Block heuristic with single CPPA $\mathfrak{m}$
(Reoptimization heuristics)	PRH	Primal reoptimization
	DRH-PI	Dual reoptimization with zero dual penalty
	DRH-LDP	Dual reoptimization with linear dual penalty

Table III summarizes the methods tested in our numerical study. We implement DRH using zero (labeled PI) and linear dual penalties (labeled LDP) in math program Equation 2.14. These two variants, dubbed DRH-PI and DRH-LDP, deliver both policies and dual bounds. We used linear dual penalties in DRH-LDP defined as in Equation 2.16 but that include power price spreads alone, which we found sufficient to obtain good-quality procurement policies and lower bounds.

All algorithms were programmed using C++ with Gurobi 8.0 as the math programming solver. We estimated the value of heuristic procurement policies (i.e., upper bounds on the optimal policy value) and the DRH lower bound in Monte Carlo simulation using 1000 evaluation sample paths (i.e., H equals 1000), as this choice resulted in standard errors below 1% of the mean. The DRH upper bound estimation process requires computing procurement decisions at each stage of an evaluation sample path. Therefore, for a fixed stage and evaluation sample path, we formulated and solved the math program (Equation 2.14) on 30 inner sample paths

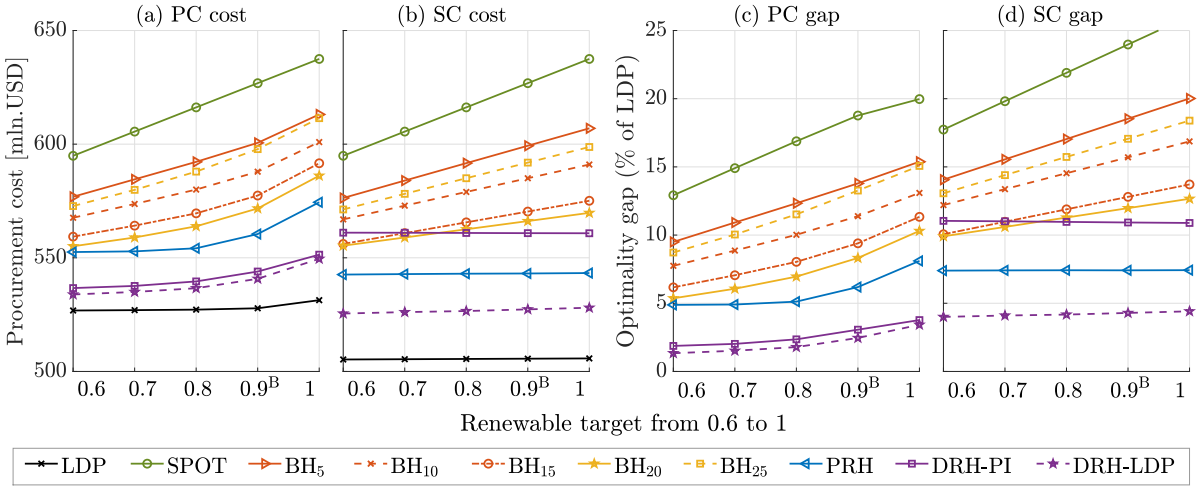
and then applied a decision measure to back out a non-anticipative control. We reported DRH results based on the mean decision measure but tested the median decision measure on the baseline instance and found its performance to be similar. When using the mean decision measure, we followed the strategy discussed at the end of §2.4.2 to ensure feasibility with respect to the non-convex action set.

2.5.3 Comparison of methods

Below we compare the performance of the methods in Table III on the instance sets S1–S5 summarized in Table II, also considering both PC and SC variants of MDP (Equation 2.12). For each instance, we report the expected procurement cost (i.e., the expected discounted total cost over the planning horizon) and the optimality gap with respect to the DRH-LDP lower bound, labeled LDP. We omit showing the DRH-PI lower bound because it is worse than the DRH-LDP lower bound on average by 5% and 50% for the PC and SC variants, respectively, and its standard error is also as high as 3.7%. The standard error of the remaining lower and upper bound estimates for PCs (SCs) is 0.6% (0.7%) on average and at most 0.7% (1%). In the remaining text, when discussing the performance of a method, we are referring to the quality of its procurement policy.

We begin by discussing the results for the instance set S1, which was obtained by varying the RPPT α from 60% to 100%. The corresponding results are displayed in Figure 2. DRH-LDP performs best on all the S1 instances and has average optimality gaps of 2.1% and 4.2% in the PC and SC contract settings, respectively. The DRH-PI optimality gap is similar to DRH-LDP when using PCs (2.6% on average) but substantially worse under SCs (11% on average).

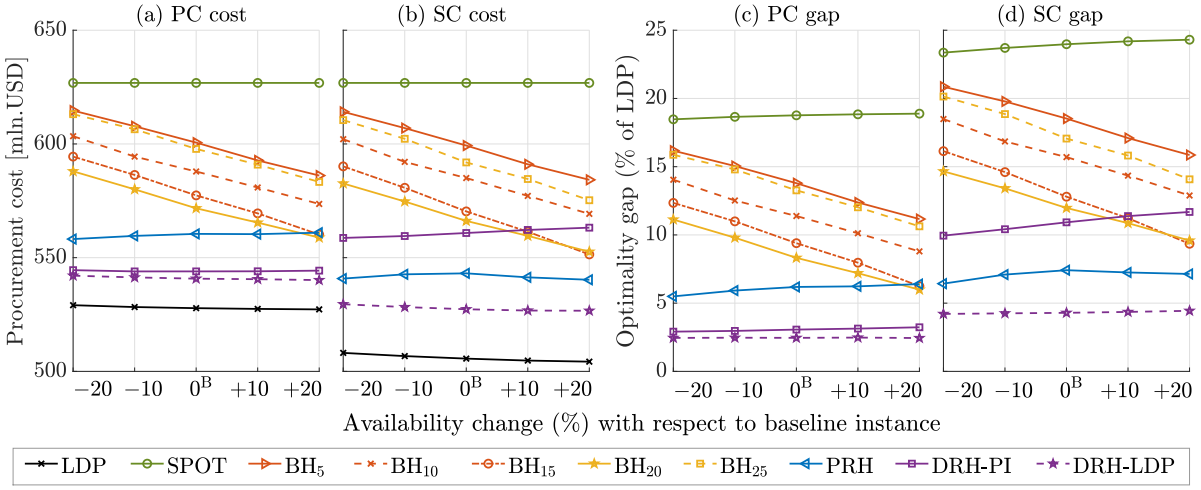
Figure 2: Procurement costs and optimality gaps for the S1 instance set.



While PRH optimality gaps are smaller than analogous DRH-PI gaps by 3.6% on average when using SCs, the former method is on average of 3.2% worse than the latter method under PCs. The performance of spot and block heuristics is largely inferior to reoptimization methods.

The S2 instances vary the probability $\mathfrak{p}_{\mathfrak{m}}$ of each contract $\mathfrak{m} \in \mathcal{M}$ from its base value between -20% and $+20\%$ to understand the effect of changing the contract availability on methods. The results corresponding to these instances are reported in Figure 3. The relative ranking of methods is similar to the S1 instances: DRH-LDP achieves the smallest optimality gaps across instances (2.4% and 4.3% on average for PCs and SCs, respectively). However, contract availability impacts single-contract methods (i.e., $\mathrm{BH}_{\mathfrak{m}}$) and multi-contract methods (PRH and DRH) in a markedly different manner. In particular, both the procurement cost and optimality gap of $\mathrm{BH}_{\mathfrak{m}}$ increase substantially in the presence of contract shortage. For instance, the BH_{20} procurement cost increases by more than 5% when decreasing contract availability from $+20\%$ to -20% relative to the baseline. In contrast, the costs associated with both PRH

Figure 3: Procurement costs and optimality gaps for the S2 instance set.



and DRH, which consider multiple CPPAs, are fairly stable under such availability changes (the maximum cost increase is less than 0.8%).

The instance sets S3–S4 are created by varying the long-term mean of power and REC prices. Specifically, our price model calibration results in a long-term mean of power and RECs of 39.7 USD/MWh and 9.4 USD/MWh, respectively. Market outlooks suggest that these long-term prices will decrease due to increasing penetration of renewable energy [125]. To understand this effect, we consider the instances S3 in which the long-term mean power price is reduced to 30 USD/MWh and further to 20 USD/MWh. In contrast to the power price, the average REC price can increase or decrease in the long-term due to regulatory changes [75]. To account for this effect, the long-term mean of the REC price is decreased to 5 USD/MWh and increased to 20 USD/MWh in the instance set S4. Results for the S3 and S4 instance sets are displayed in Figure 4 and Figure 5, respectively.

The relative performance of methods on these sets are consistent with our prior observations on the S1–S2 instances. If the long-term mean power price decreases, then the procurement

Figure 4: Procurement costs and optimality gaps for the S3 instance set.

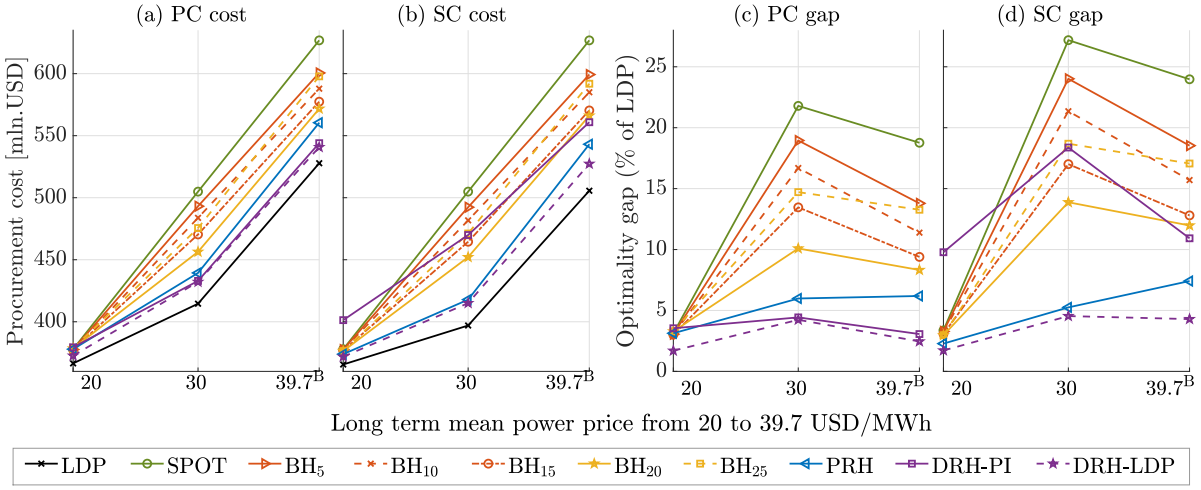
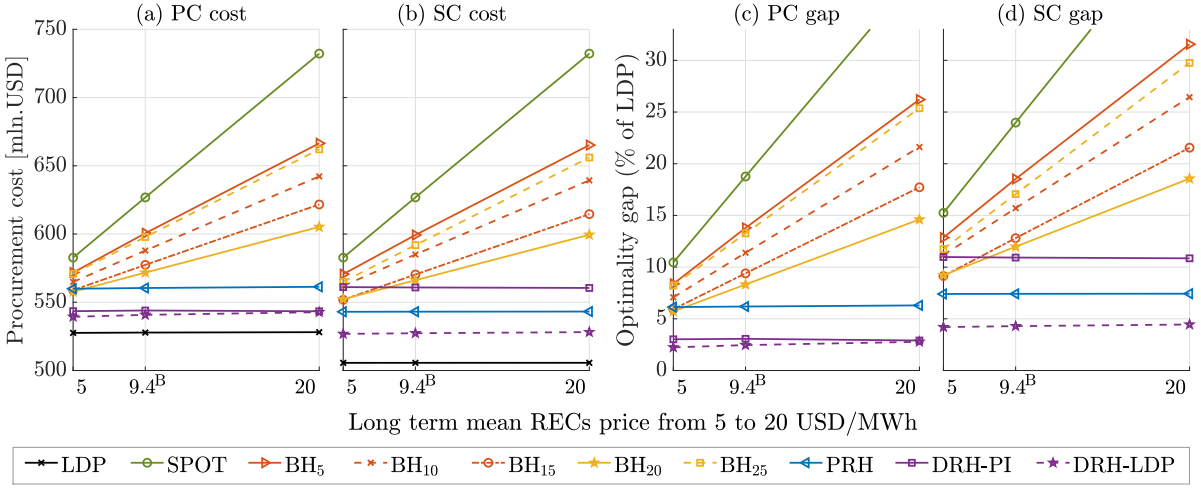


Figure 5: Procurement costs and optimality gaps for the S4 instance set.

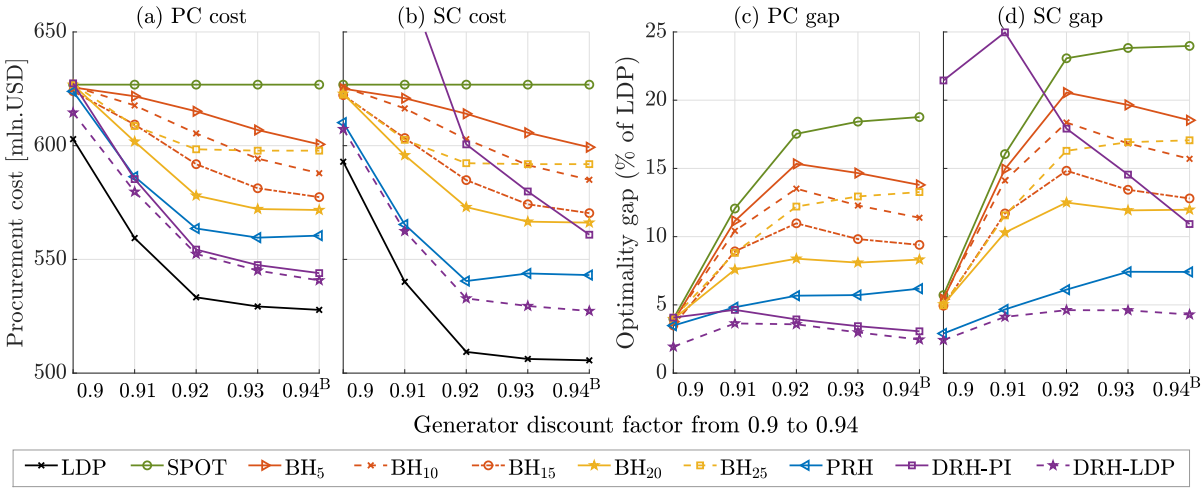


cost decreases substantially under all policies. In contrast, when the long-term mean of the REC price changes, the procurement costs and optimality gaps of the spot and BH policies are affected substantially, while the reoptimization methods, PRH and DRH, are stable across instances. This behavior is due to PRH/DRH purchasing a considerable amount of PCs/SCs

even when the REC prices are low, which insulates their procurement policies to REC price increases.

The instance set S5 considers changes in the CPPA strike price as a result of the generator varying r from 0.94 to 0.9, which models the return on investment changing between 6.4% and 11.1%. The corresponding results displayed in Figure 6 show that the procurement cost increases under all methods when r decreases, except for the spot policy. Equation 2.8 helps understand this behavior. From this equation it follows that reducing r raises the NPV component of the CPPA strike price and potentially the CPPA strike price itself. Therefore, BH_m, PRH, and DRH-LDP use less CPPAs and rely more on spot purchases as r decreases, which results in the performances of different methods becoming closer to each other.

Figure 6: Procurement costs and optimality gaps for the S5 instance set.



Overall, using simple procurement heuristics, such as spot and BH_m, on our instances results in higher procurement costs, whereas reoptimization methods work well, with DRH-

LDP outperforming PRH. This superior performance of DRH suggests that obtaining non-anticipative decisions by averaging away future information from anticipative actions is better than computing decisions based on a single (non-anticipative) forecast, as done in PRH. The near-optimal and stable performance of DRH-LDP across instance sets also indicates that the mean, a simple and inexpensive decision measure, is effective for use with DRH. Similarly, simple linear dual penalties appear sufficient to obtain high quality procurement decisions using DRH, in addition to good lower bounds.

Finally, we discuss the computational burden of using each method. The time taken to compute a decision at a given stage and state when using the spot and $\mathrm{BH}_{\mathtt{m}}$ policies is negligible. The analogous times for PRH and DRH are 0.04 and 0.8 seconds, respectively. Most of this time for DRH goes towards solving the dual optimization models on sample paths and the time required to extract the non-anticpative decision is negligible. Thus, for all practical purposes, implementing the DRH policy is inexpensive. In addition to computing a procurement decision, computing an upper bound of a policy requires applying it over the entire planning horizon in Monte Carlo simulation. When using 1000 sample paths this estimation time is on average 10 seconds for $\mathrm{BH}_{\mathtt{m}}$, and 28 and 216 minutes for PRH and DRH (each variant), respectively. Thus, estimating the value of DRH is more expensive than PRH because the former policy solves at each stage math programs for each of the 30 inner sample paths whereas the latter policy solves a single math program. Nevertheless, the solution of DRH math programs in the inner samples can be parallelized to substantially reduce this overhead. Finally, estimating the lower bound with DRH takes 10 minutes on average (each variant).

2.5.4 Procurement insights

Figure 2 shows that the procurement cost increases as expected with the RPPT α. Specifically, the procurement costs under PCs and SCs both vary linearly for $\alpha \leqslant 0.8$ and in a strictly convex and linear manner, respectively, for $\alpha > 0.8$. The procurement costs under SCs are in general lower than analogous costs under PCs (especially for high α values), which provides a cost incentive for using SCs in addition to their well-known advantage of being free from the physical delivery constraints associated with PCs. For instance, the procurement cost incurred under SCs is on average 2.5% lower than under PCs when using DRH-LDP. This comes at the expense of increased cash flow volatility under SCs as the coefficient of variation is on average 3.5% higher than PCs. These findings are largely in sync with our analytical results in §2.2.

Consistent with the extant procurement literature, we find that the inclusion of long-term contracts (i.e., CPPAs in our case) in procurement portfolios helps hedge against price uncertainty and reduces procurement costs. In addition, our results indicate that the use of CPPAs is more valuable to companies that have committed to an RPPT. This is seen, for instance in Figure 2, where the difference between the procurement costs of the spot-only and DRH-LDP policies increases with the RPPT. Therefore, as corporations become more aggressive with procuring renewable power, the use of CPPAs, in particular SCs, is likely to be higher, which is consistent with trends observed in practice [29]. Constructing portfolios with CPPAs is non-trivial as shown by the poor performance of the block heuristics that use a single contract type. The fairly flat procurement cost of the DRH-LDP policy to changes in contract availability suggests that dynamically constructed portfolios containing multiple CPPAs are robust

to such variability, which is a useful property. This observation suggests some level of substitutability between different subsets of CPPA contracts. However, individual contracts are not fully substitutable for another; if they were, the procurement costs of the block heuristics and DRH-LDP would be similar.

Portfolios computed by DRH-LDP on our baseline instance contain PCs with lengths 5, 10, 15, 20, and 25 years in the proportions 8.7%, 17.8%, 23.1%, 29.2%, and 21.2%, respectively, on average across the evaluation samples and stages. Analogous proportions when using DRH-LDP with SCs are 3.4%, 10.8%, 15.2%, 31.4%, and 39.2%, which shows that longer contracts are used more often due to SCs having lower over-procurement risk than PCs as also discussed in §2.2.1. We observed that the stage-averaged mix of CPPAs did not change significantly across instances. For example, on the S1 instances, the proportion of CPPAs of different lengths varied by at most 4%. However, this mix does change substantially over time, in a manner that is more pronounced in the reach period, with these changes remaining significant in the rest of the planning horizon. For example, the proportion of 5, 10, 15, 20, and 25 year CPPA contracts under DRH-LDP in the baseline instance fluctuates by as much as 11.8%, 18.4%, 17.6%, 22.8%, 13.9%, respectively, between years 10 and 40. Thus, the near-optimal portfolios computed by DRH-LDP indeed change dynamically over time.

Finally, using CPPAs for procurement is not always beneficial. For instance, if these contracts become expensive due to generators expecting a higher rate of return (see Figure 6) then spot procurement would displace signing CPPAs and the multi-stage procurement problem will reduce to procuring power and RECs as needed from the short-term market. This seems un-

likely given the increasing use of CPPAs [12] and decreasing production costs associated with renewable power [202]. Under low production costs, the strike price that a generator charges is likely to be highly correlated with the spot market, in particular, the expected spot price over the tenor of the contract, a feature that we try to capture in the strike price model of §2.3.1. In such an environment, generators would remain profitable even after current production tax credits expire resulting in CPPAs continuing to play an important role in a firm's renewable power procurement strategy. We find support for this statement in additional experiments that we conducted, where we removed the production tax credit in our base instance and found procurement costs to increase by only 1.3% when using PCs and by 1.7% with SCs.

2.6 Conclusion

Motivated by the recent global trend in corporate energy procurement, we study the problem of companies that have committed to satisfying a renewable power procurement target by a future date. We focus on CPPAs for long-term procurement and couple them with short-term purchases of power and RECs. We analyze a two-stage model to understand the behavior of procurement costs when using popular variants of CPPAs, namely financial and physical, and changing their structure following the practitioner literature. To facilitate decision making, we formulate a multi-period power procurement MDP model with short-term and long-term procurement options that is challenging to solve due to a non-convex action set and high-dimensional endogenous and exogenous state components. Heuristic procurement decisions can nevertheless be obtained using an easy-to-implement reoptimization method, PRH, but this approach does not provide a lower bound on the optimal cost. We thus develop DRH, a novel

scheme that combines reoptimization and the information relaxation and duality approach to extract non-anticipative decisions from anticipative-action distributions, in addition to estimating a lower bound. DRH retains the desirable implementation properties of PRH. Its procurement decisions are near-optimal on realistic instances and outperform PRH and problem-specific heuristics. DRH thus emerges as a promising way to tackle high-dimensional MDPs with non-convex action sets arising in our procurement setting as well as other applications. Moreover, our findings show that procurement portfolios with multiple CPPAs reduce power purchase costs significantly in the presence of an RPPT compared to using a single CPPA. In addition, such portfolios are effective at hedging against uncertainty in contract availability and REC prices.

2.7 Appendix

To ease notation, in sections 2.7.1 and 2.7.2 we represent $\mathbb{E}_0$ by $\mathbb{E}$ and the random variables D_1, P_1, and R_1 by D, P, and R, respectively. Furthermore, we denote by $\mathcal{N}(\mu, \sigma^2)$ a normal distribution with mean μ and variance σ^2, and by ϕ and Φ the probability distribution function (PDF) and cumulative distribution function (CDF), respectively, of a standard normal distribution $\mathcal{N}(0, 1)$.

2.7.1 Proofs

Lemma 2.7.1 is used in the proof of Lemma 2.2.1. Define $Q_{PC}(z; \alpha, K) := \mathbb{E}[\widetilde{C}_{PC}(z, w_1; \alpha, K)]$ and $Q_{SC}(z; \alpha, K) := \mathbb{E}[\widetilde{C}_{SC}(z, w_1; \alpha, K)]$ as the expected procurement costs under a PC and an SC, respectively.

Lemma 2.7.1. *Under Assumption 1, there exist optimal procurement decisions z^*_{PC} and z^*_{SC} minimizing, respectively, $Q_{PC}(z; \alpha, K)$ and $Q_{SC}(z; \alpha, K)$, that belong to the interval $[\alpha a, \alpha b]$.*

Proof. We start by considering PCs, and then move to SCs. In both cases, we will show that no unique optimal solution exists in $\Omega := [0, \alpha a) \cup (\alpha b, b]$, which establishes the desired result. Recall that the power demand D follows a uniform distribution in the interval $[a, b]$, and its CDF is equal to 0 for $z < a$, $(z - a)/(b - a)$ for $z \in [a, b]$, and 1 for $z > b$.

The derivative of $Q_{PC}(z; \alpha, K)$ with respect to z exists in the interval $[0, \alpha a]$ and is equal to

$$\frac{dQ_{PC}(z; \alpha, K)}{dz}\Big|_{z\in[0,\alpha a]} = K - \mathbb{E}[P] \Pr(D \geqslant z) - \mathbb{E}[R] \Pr(\alpha D \geqslant z) = K - \mathbb{E}[P] - \mathbb{E}[R] \leqslant 0,$$

where the second equality holds since $\Pr(D \geqslant z) = \Pr(\alpha D \geqslant z) = 1$ for $z \in [0, \alpha a]$, and the inequality follows from the strike price upper bound (i.e., $K \leqslant \mathbb{E}[P] + \mathbb{E}[R]$) in Assumption 1. Therefore, $Q_{PC}(z; \alpha, K)$ is non-increasing in $[0, \alpha a)$, has no unique optimal solution in this half-open interval, and we can focus on $z \geqslant \alpha a$ to search for an optimal solution. Similarly, the derivative exists in $[\alpha b, b]$ and is equal to

$$\frac{dQ_{PC}(z; \alpha, K)}{dz}\Big|_{z\in[\alpha b,b]} = K - \mathbb{E}[P] \Pr(D \geqslant z) \geqslant K - \mathbb{E}[P] \geqslant 0,$$

where the last inequality follows from the strike price lower bound (i.e., $K \geqslant \mathbb{E}[P]$) under Assumption 1. Therefore, $Q_{PC}(z; \alpha, K)$ is non-decreasing in $(\alpha b, b]$, has no unique optimal solution in this interval, and an optimal solution satisfies $z \leqslant \alpha b$. We thus can conclude that z^*_{PC} belongs to the interval $[\alpha a, \alpha b]$.

We proceed to show that an analogous result holds for the solution z^*_{SC} of $\min_{z\geqslant 0} Q_{SC}(z;\alpha,K)$. The derivative of $Q_{SC}(z;\alpha,K)$ with respect to z exists in the two intervals $[0,\alpha a]$ and $[\alpha b, b]$ with values, respectively, equal to

$$\frac{dQ_{SC}(z;\alpha,K)}{dz}\Big|_{z\in[0,\alpha a]} = \mathbb{E}[K-P] - \mathbb{E}[R]\Pr(\alpha D \geqslant z) = K - \mathbb{E}[P] - \mathbb{E}[R] \leqslant 0,$$
$$\frac{dQ_{SC}(z;\alpha,K)}{dz}\Big|_{z\in[\alpha b,b]} = \mathbb{E}[K-P] = K - \mathbb{E}[P] \geqslant 0,$$

where both inequalities follow from Assumption 1. Following the same argument used above in the case of PCs, we can conclude that z^*_{SC} belongs to the interval $[\alpha a, \alpha b]$. □

***Proof of Lemma** 2.2.1.* We start by determining the optimal procurement quantity under PCs, and then move to SCs.

From Lemma 2.7.1, we know that there exists an optimal solution to $\min_{z\geqslant 0} Q_{PC}(\cdot;\alpha,K)$ in the interval $[\alpha a, \alpha b]$, thus we can limit the search to this interval. The expected cost function $Q_{PC}(z;\alpha,K)$ is convex in the procurement size z in the interval $[\alpha a, \alpha b]$ since the second-order derivative is positive:

$$\frac{d^2Q_{PC}(z;\alpha,K)}{dz^2}\Big|_{z\in[\alpha a,\alpha b]} = \mathbb{E}[P]\frac{\mathbf{1}_{\{z\geqslant a\}}}{b-a} + \mathbb{E}[R]\frac{1}{\alpha b - \alpha a} > 0.$$

Therefore, the optimal quantity z^*_{PC} can be calculated using the first-order condition

$$\frac{dQ_{PC}(z;\alpha,K)}{dz} = K - \mathbb{E}[R]\Pr(\alpha D \geqslant z) - \mathbb{E}[P]\Pr(D \geqslant z) = 0. \tag{2.17}$$

We proceed by considering two cases.

Case 1: $\Pr(D \geqslant z) = 1$. Hence we have $z^*_{PC} \leqslant a$ in this case. The first-order condition Equation 2.17 simplifies to

$$K - \mathbb{E}[R]\,\frac{b - \frac{z}{\alpha}}{b-a} - \mathbb{E}[P] = 0 \iff z^*_{PC} = \alpha\Big(b - \frac{K - \mathbb{E}[P]}{\mathbb{E}[R]}(b-a)\Big). \tag{2.18}$$

Enforcing $z^*_{PC} \leqslant a$ using solution (Equation 2.18) results in an upper bound on α equal to

$$\bar{\alpha} = \frac{a\mathbb{E}[R]}{b\mathbb{E}[R] - (K - \mathbb{E}[P])(b-a)}. \tag{2.19}$$

Case 2: $\Pr(D \geqslant z) < 1$. Using $\Pr(D \geqslant z) = (b-z)/(b-a)$, we obtain $z^*_{PC} > a$ in this case. Moreover, solving the first order condition (Equation 2.17) gives

$$z^*_{PC} = \frac{-K(b-a) + (\mathbb{E}[R] + \mathbb{E}[P])\,b}{\frac{1}{\alpha}\mathbb{E}[R] + \mathbb{E}[P]}.$$

Note that $z^*_{PC} > a$ is satisfied when $\alpha > \bar{\alpha}$. In conclusion, the optimal PC procurement z^*_{PC} is given by

$$z^*_{PC} = \begin{cases} \alpha\Big(b - \frac{K-\mathbb{E}[P]}{\mathbb{E}[R]}(b-a)\Big), & \text{if } \alpha \leqslant \bar{\alpha}, \\ \frac{-K(b-a)+(\mathbb{E}[R]+\mathbb{E}[P])b}{\frac{1}{\alpha}\mathbb{E}[R]+\mathbb{E}[P]}, & \text{if } \alpha > \bar{\alpha}. \end{cases}$$

Next, consider an SC. As in the PC case, by Lemma 2.7.1, there exists an optimal solution to $\min_{z\geqslant 0} Q_{SC}(z;\alpha,K)$ in the interval $[\alpha a, \alpha b]$, thus we can limit our search space. This function is strictly convex in z because its second derivative is positive by Assumption 1:

$$\frac{d^2 Q_{SC}(z;\alpha,K)}{dz^2}\Big|_{z\in[\alpha a,\alpha b]} = \mathbb{E}[R]\,\frac{1}{\alpha b - \alpha a} > 0.$$

Thus, the optimal SC quantity z^*_{SC} can be calculated by applying the first-order condition:

$$\begin{aligned}
&\frac{dQ_{SC}(z;\alpha,K)}{dz}\Big|_{z\in[\alpha a,\,\alpha b]} = 0 \\
&\qquad\Longleftrightarrow\; K - \mathbb{E}[P] - \mathbb{E}\left[R\right]\Pr\left(\alpha D \geqslant z\right) = 0 \\
&\qquad\Longleftrightarrow\; z^*_{SC} = \alpha\Big(b - \frac{K-\mathbb{E}[P]}{\mathbb{E}[R]}(b-a)\Big).
\end{aligned}$$

□

***Proof of Proposition** 2.2.2.* We characterize the behavior of $C_{PC}(\alpha, K)$ and $C_{SC}(\alpha, K)$ as functions of α, respectively, in part (**a**) and part (**b**) of the proof, and compare them in part (**c**).

(**a**) When $\alpha \in [0, \bar{\alpha}]$, we have $z^*_{PC} \in [\alpha a, a]$, which implies $\Pr(D \geqslant z^*_{PC}) = 1$, and the optimal cost becomes

$$C_{PC}(\alpha, K) = K\, z^*_{PC} + \mathbb{E}[P(D - z^*_{PC})] + \mathbb{E}[R(\alpha\, D - z^*_{PC})_+].$$

Using the characterization of z^*_{PC} from Lemma 2.2.1, the derivative of this function with respect to α is

$$\begin{aligned}
\frac{dC_{PC}(\alpha,K)}{d\alpha}\Big|_{\alpha\in[0,\bar{\alpha}]} &= K\frac{z^*_{PC}}{\alpha} + \mathbb{E}[P]\Big(-\frac{z^*_{PC}}{\alpha}\Big) + \mathbb{E}\Big[R\Big(D - \frac{z^*_{PC}}{\alpha}\Big)_+\Big] \\
&= (K - \mathbb{E}[P])\frac{z^*_{PC}}{\alpha} + \mathbb{E}\Big[R\Big(D - \frac{z^*_{PC}}{\alpha}\Big)_+\Big].
\end{aligned}$$

Since z^*_{PC} is linear in α within the interval $[0, \bar{\alpha}]$, z^*_{PC}/α and thus $dC_{PC}(\alpha, K)/d\alpha$ are independent of α. Moreover, $dC_{PC}(\alpha, K)/d\alpha$ is non-negative because $K \geqslant \mathbb{E}[P]$ by assumption 1. Therefore, $C_{PC}(., K)$ is a linear function of α in the interval $[0, \bar{\alpha}]$.

We show next that $C_{PC}(\cdot, K)$ is convex increasing with the target when $\alpha \in (\bar{\alpha}, 1]$. In this case, $z^*_{PC} \in (a, \alpha b]$ and $\Pr(D \geqslant z^*_{PC}) < 1$. Expanding the definition of $C_{PC}(\alpha, K)$ gives

$$\begin{aligned}
C_{PC}(\alpha, K) &= K z^*_{PC} + \mathbb{E}[P(D - z^*_{PC})_+] + \mathbb{E}[R(\alpha D - z^*_{PC})_+] \\
&= K z^*_{PC} + \mathbb{E}[P] \int_{z^*_{PC}}^{b} \frac{D - z^*_{PC}}{b - a} \, dD + \mathbb{E}[R] \int_{z^*_{PC}/\alpha}^{b} \frac{\alpha D - z^*_{PC}}{b - a} \, dD \\
&= K z^*_{PC} + \frac{\mathbb{E}[P]}{b - a}\left(\frac{b^2}{2} - \frac{z^{*2}_{PC}}{2} - z^*_{PC}(b - z^*_{PC})\right) + \frac{\mathbb{E}[R]}{b - a}\left[\alpha\left(\frac{b^2}{2} - \frac{z^{*2}_{PC}}{2\alpha^2}\right) - z^*_2\left(b - \frac{z^*_{PC}}{\alpha}\right)\right] \\
&= K z^*_{PC} + \mathbb{E}[P]\frac{1}{2\,(b - a)}(z^*_{PC} - b)^2 + \mathbb{E}[R]\frac{1}{2\,\alpha(b - a)}(z^*_{PC} - \alpha b)^2.
\end{aligned}$$

The first derivative of $C_{PC}(\alpha, K)$ is

$$\frac{dC_{PC}(\alpha, K)}{d\alpha}\Big|_{\alpha \in (\bar{\alpha}, 1]} = \frac{\mathbb{E}[R]\,(-aK + b\,(K + \mathbb{E}[P](\alpha - 1)))\,(a\,K + b\,(-K + \mathbb{E}[P] + 2\mathbb{E}[R] + \alpha\,\mathbb{E}[P]))}{2\,(b - a)(\mathbb{E}[R] + \alpha\mathbb{E}[P])^2}.$$

The denominator as well as the third term of the product in the numerator are strictly positive due to the bounds on the strike price in Assumption 1. We show that the second term in the numerator is also strictly positive using the following chain of inequalities:

$$\begin{aligned}
-aK + b\,(K + \mathbb{E}[P](\alpha - 1)) &\geqslant -aK + b\,(K + \mathbb{E}[P](\bar{\alpha} - 1)) \\
&= \frac{(b - a)\,((bK - b\mathbb{E}[P])(\mathbb{E}[R] - K + \mathbb{E}[P]) + aK(K - \mathbb{E}[P]))}{b\mathbb{E}[R] - (K - \mathbb{E}[P])(b - a)} \\
&> 0.
\end{aligned}$$

The first inequality follows by lower bounding α by $\bar{\alpha}$, the first equality by replacing $\bar{\alpha}$ with its full expression given in Equation 2.19 and simplifying the resulting terms, and the second inequality results from both the numerator and denominator being positive under the bounds

on the strike price in Assumption 1. In addition to the first derivative being positive, the second derivative is also positive as it is equal to

$$\frac{d^2 C_{PC}(\alpha, K)}{d\alpha^2}\Big|_{\alpha\in(\bar{\alpha},1]} = \frac{\mathbb{E}[P]\mathbb{E}[R](aK + b\,(-K + \mathbb{E}[P] + \mathbb{E}[R]))^2}{(b-a)(\mathbb{E}[R] + \alpha\mathbb{E}[P])^3} > 0,$$

where the strict inequality holds due to Assumption 1. Thus, the procurement cost is a strictly convex increasing function in the target level for $\alpha \in [\bar{\alpha}, 1]$.

(**b**) From Lemma 2.2.1 we know that z^*_{SC} is linear in α. The slope of $C_{SC}(\alpha, K)$ with respect to α is

$$\frac{dC_{SC}(\alpha, K)}{d\alpha}\Big|_{\alpha\in[0,1]} = (K - \mathbb{E}[P])\frac{z^*_{SC}}{\alpha} + \mathbb{E}\Big[R\Big(D - \frac{z^*_{SC}}{\alpha}\Big)_+\Big],$$

which is independent of α. This implies that $C_{SC}(\alpha, K)$ is linear in α.

(**c**) From the proofs of parts (**a**) and (**b**) it follows that $dC_{SC}(\alpha, K)/d\alpha$ equals $dC_{PC}(\alpha, K)/d\alpha$ when $\alpha \leqslant \bar{\alpha}$. Therefore, since $C_{PC}(\alpha, K)$ and $C_{SC}(\alpha, K)$ coincide when $\alpha = 0$, these two costs are the same for $\alpha \in (0, \bar{\alpha}]$. Furthermore, $C_{PC}(\alpha, K)$ is strictly convex increasing in α for $\alpha \in (\bar{\alpha}, 1]$, while $C_{SC}(\alpha, K)$ remains linear with the same slope, which implies that the former cost is higher than the latter cost in this interval. □

***Proof of Proposition** 2.2.3.* Recall that $\widetilde{C}_{PC}(z, w_1; \alpha, K) = Kz + P(D - z)_+ + R(\alpha D - z)_+$. Defining $Y(z) = -P(z - D)_+$, we have $\widetilde{C}_{SC}(z, w_1; \alpha, K) = \widetilde{C}_{PC}(z, w_1; \alpha, K) + Y(z)$. We start by showing that $\mathrm{Var}[\widetilde{C}_{SC}(z, w_1; \alpha, K)] \geqslant \mathrm{Var}[\widetilde{C}_{PC}(z, w_1; \alpha, K)]$. Since $\mathrm{Var}[\widetilde{C}_{SC}(z, w_1; \alpha, K)] = \mathrm{Var}[\widetilde{C}_{PC}(z, w_1; \alpha, K)] + \mathrm{Var}[Y(z)] + 2\,\mathrm{Cov}[\widetilde{C}_{PC}(z, w_1; \alpha, K), Y(z)]$, and $\mathrm{Var}[Y(z)] \geqslant 0$, the proof

reduces to showing that $\mathrm{Cov}[\widetilde{C}_{PC}(z, w_1; \alpha, K), Y(z)] \geqslant 0$, which we illustrate using the following chain of inequalities:

$$\begin{aligned}
&\mathrm{Cov}[\widetilde{C}_{PC}(z, w_1; \alpha, K), Y(z)] \\
&\qquad = \mathrm{Cov}[Kz + P(D - z)_+ + R(\alpha D - z)_+, -P(z - D)_+] && \text{(2.20a)} \\
&\qquad = \mathrm{Cov}[P(D - z)_+ + R(\alpha D - z)_+, -P(z - D)_+] && \text{(2.20b)} \\
&\qquad = -\mathrm{Cov}[P(D - z)_+, P(z - D)_+] - \mathrm{Cov}[R(\alpha D - z)_+, P(z - D)_+] && \text{(2.20c)} \\
&\qquad = -\mathbb{E}[P^2(D - z)_+(z - D)_+] + \mathbb{E}[P(D - z)_+]\mathbb{E}[P(z - D)_+] \\
&\qquad\quad - \mathbb{E}[PR(z - D)_+(\alpha D - z)_+] + \mathbb{E}[R(\alpha D - z)_+]\mathbb{E}[P(z - D)_+] && \text{(2.20d)} \\
&\qquad = \mathbb{E}[P(D - z)_+]\mathbb{E}[P(z - D)_+] + \mathbb{E}[R(\alpha D - z)_+]\mathbb{E}[P(z - D)_+] && \text{(2.20e)} \\
&\qquad \geqslant 0, && \text{(2.20f)}
\end{aligned}$$

where Equation 2.20a follows from the definition of $\widetilde{C}_{PC}(z, w_1; \alpha, K)$, Equation 2.20b is a consequence of Kz being a constant, Equation 2.20c follows from the linearity of the covariance, and Equation 2.20d from the well-known property that $\mathrm{Cov}(A, B) = \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B]$ if A and B are two random variables. The equality in Equation 2.20e follows from $(D - z)_+ \cdot (z - D)_+ \equiv 0$ and $(z - D)_+ \cdot (\alpha D - z)_+ \equiv 0$, and Equation 2.20f is a consequence of all expectations involving only non-negative random variables.

If $\Pr(z > D) > 0$, then $\mathrm{Var}[Y(z)] > 0$, which combined with the non-negative covariance (Equation 2.20f) gives $\mathrm{Var}[\widetilde{C}_{SC}(z, w_1; \alpha, K)] > \mathrm{Var}[\widetilde{C}_{PC}(z, w_1; \alpha, K)]$. Instead, if $\Pr(z > D) = 0$, then $Y(z) \equiv 0$, which implies that $\mathrm{Var}[\widetilde{C}_{SC}(z, w_1; \alpha, K)] = \mathrm{Var}[\widetilde{C}_{PC}(z, w_1; \alpha, K)]$. □

Lemmas 2.7.2, 2.7.3, and 2.7.4 are used in the proof of Proposition 2.2.4.

Lemma 2.7.2. *Let* P *be a log-normal random variable with parameters* μ *and* σ^2*, and denote with* $f(P;\mu,\sigma^2)$ *its PDF. Moreover, consider* $Y = \log(P)$*, i.e.* $Y \sim \mathcal{N}(\mu,\sigma^2)$*, and a scalar* $B > 0$*. It holds that:*

$$\mathbb{E}\left[P\mathbf{1}_{\{P<B\}}\right] = e^{\mu+\frac{\sigma^2}{2}}\,\Phi\left(\frac{\ln(B)-(\mu+\sigma^2)}{\sigma}\right)$$

Proof. Using the relationship between P and Y, we establish the following equalities:

$$\begin{aligned}\mathbb{E}\left[P\mathbf{1}_{\{P<B\}}\right] &= \int_{-\infty}^{B} Pf(P;\mu,\sigma^2)dP = \int_{-\infty}^{B} e^{\ln P} f(P;\mu,\sigma^2)dP \\ &= \int_{-\infty}^{\ln(B)} e^{Y}\phi(Y;\mu,\sigma^2)dY = \int_{-\infty}^{\ln(B)} e^{Y}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(Y-\mu)^2}{2\sigma^2}}dY.\end{aligned}$$

By adding and subtracting $\sigma^4 + 2\sigma^2\mu$ to the numerator of the second exponential term, we obtain:

$$\int_{-\infty}^{\ln(B)} e^{Y}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(Y-\mu)^2}{2\sigma^2}}dY = e^{\mu+\frac{\sigma^2}{2}}\,\Phi\left(\frac{\ln(B)-\mu-\sigma^2}{\sigma}\right),$$

which proves the desired result. □

Lemma 2.7.3. *Given three scalars* $B > 0$*,* $\mu \geqslant 0$*, and* $\sigma > 0$*, it holds that:*

$$\frac{\exp(\mu+\sigma^2/2)}{\sigma}\left[\frac{1}{B}\phi\left(\frac{\ln(B)-(\mu+\sigma^2)}{\sigma}\right)\right] = \frac{1}{\sigma}\left[\phi\left(\frac{\ln(B)-\mu}{\sigma}\right)\right].$$

Proof.

$$\begin{aligned}
&\frac{\exp(\mu+\sigma^2/2)}{\sigma}\left[\frac{1}{B}\phi\left(\frac{\ln(B)-(\mu+\sigma^2)}{\sigma}\right)\right]\\
&\qquad=\frac{\exp(\mu+\sigma^2/2)}{\sigma}\left[\frac{1}{B\sqrt{2\pi}}\exp\left(\frac{-(\ln(B))^2-(\mu+\sigma^2)^2+2\ln(B)(\mu+\sigma^2)}{2\sigma^2}\right)\right]\\
&\qquad=\frac{\exp(\mu+\sigma^2/2)}{\sigma}\left[\frac{1}{B}\phi\left(\frac{\ln(B)-\mu}{\sigma}\right)\exp\big(-\sigma^2/2-\mu+\ln(B)\big)\right]\\
&\qquad=\frac{\exp(\mu+\sigma^2/2)}{\sigma}\left[\frac{1}{B}\phi\left(\frac{\ln(B)-\mu}{\sigma}\right)\exp\big(-\sigma^2/2-\mu\big)B\right]\\
&\qquad=\frac{1}{\sigma}\phi\left(\frac{\ln(B)-\mu}{\sigma}\right).
\end{aligned}$$

□

Lemma 2.7.4. *Given* $\alpha > 0$ *and* $\delta, \delta' > 0$*, if* $\mathbb{E}[K^{INT}(P;K,\delta)] < \mathbb{E}[K^{INT}(P;K,\delta')]$*, then it holds that* $C^{INT}_{SC}(\alpha,K,\delta) < C^{INT}_{SC}(\alpha,K,\delta')$.

Proof. Let $z^*_{SC}(\alpha,K,\delta)$ be an optimal solution to the expected procurement cost $Q^{INT}_{SC}(z;\alpha,K,\delta)$, which is defined by

$$\begin{aligned}
Q^{INT}_{SC}(z;\alpha,K,\delta) &:= \mathbb{E}[\widetilde{C}_{SC}(z,w_1;\alpha,K^{INT}(P;K,\delta))]\\
&= \mathbb{E}[K^{INT}(P;K,\delta)-P]z+\mathbb{E}[PD]+\mathbb{E}[R(\alpha D-z)_+].
\end{aligned}$$

From the hypothesis it follows that

$$\begin{aligned}
&Q^{INT}_{SC}(z^*_S(\alpha,K,\delta');\alpha,K,\delta)-Q^{INT}_{SC}(z^*_S(\alpha,K,\delta');\alpha,K,\delta')\\
&=\big(\mathbb{E}[K^{INT}(P;K,\delta)]-\mathbb{E}[K^{INT}(P;K,\delta')]\big)\cdot z^*_S(\alpha,K,\delta')<0.
\end{aligned}$$

As a consequence, the following relation between optimal procurement costs hold

$$\begin{aligned} C_{SC}^{INT}(\alpha, K, \delta) &= \min_{z \geqslant 0} Q_{SC}^{INT}(z; \alpha, K, \delta) \leqslant Q_{SC}^{INT}(z_S^*(\alpha, K, \delta'); \alpha, K, \delta) \\ &< Q_{SC}^{INT}(z_S^*(\alpha, K, \delta'); \alpha, K, \delta') = C_{SC}^{INT}(\alpha, K, \delta'), \end{aligned}$$

which proves the lemma. □

***Proof of Proposition** 2.2.4.* Recall that the power price P follows a log-normal distribution and that its natural logarithm follows $\ln(P) \sim \mathcal{N}(\mu_P, \sigma_P^2)$, i.e. it is a normal random variable with mean μ_P and standard deviation σ_P. Note that consequently $(\ln(P) - \mu_P)/\sigma_P \sim \mathcal{N}(0, 1)$. The expected unit cost with the interval strike price is defined for $\delta \in [0, K)$ and can be expressed using Lemma 2.7.2 as

$$\begin{aligned} \mathbb{E}[K^{INT}(P; K, \delta)] &= (K + \delta) \Pr(P > K + \delta) + (K - \delta) \Pr(P < K - \delta) + \mathbb{E}\left[P \mathbf{1}_{\{K-\delta \leqslant P \leqslant K+\delta\}}\right] \\ &= (K + \delta) \left[1 - \Phi\left(\frac{\ln(K + \delta) - \mu_P}{\sigma_P}\right)\right] + (K - \delta)\, \Phi\left(\frac{\ln(K - \delta) - \mu_P}{\sigma_P}\right) \\ &\quad + \exp\left(\mu_P + \frac{\sigma_P^2}{2}\right) \left[\Phi\left(\frac{\ln(K + \delta) - \mu_P - \sigma_P^2}{\sigma_P}\right) - \Phi\left(\frac{\ln(K - \delta) - \mu_P - \sigma_P^2}{\sigma_P}\right)\right]. \end{aligned}$$

The first derivative of $\mathbb{E}[K^{INT}(P; K, \delta)]$ with respect to δ is

$$\begin{aligned} \frac{\partial \mathbb{E}[K^{INT}(P; K, \delta)]}{\partial \delta} &= 1 - \Phi\left(\frac{\ln(K + \delta) - \mu_P}{\sigma_P}\right) - \frac{1}{\sigma_P} \phi\left(\frac{\ln(K + \delta) - \mu_P}{\sigma_P}\right) - \Phi\left(\frac{\ln(K - \delta) - \mu_P}{\sigma_P}\right) \\ &\quad - \frac{1}{\sigma_P} \phi\left(\frac{\ln(K - \delta) - \mu_P}{\sigma_P}\right) + \frac{\exp(\mu_P + \sigma_P^2/2)}{\sigma_P} \left[\frac{1}{K + \delta} \phi\left(\frac{\ln(K + \delta) - (\mu_P + \sigma_P^2)}{\sigma_P}\right)\right. \\ &\quad \left. + \frac{1}{K - \delta} \phi\left(\frac{\ln(K - \delta) - (\mu_P + \sigma_P^2)}{\sigma_P}\right)\right]. \end{aligned}$$

By reformulating the last term in this derivative using Lemma 2.7.3 and canceling opposite terms out, we obtain a simpler expression:

$$\frac{\partial \mathbb{E}[K^{\text{INT}}(P;K,\delta)]}{\partial \delta} = 1 - \Phi\left(\frac{\ln(K+\delta)-\mu_P}{\sigma_P}\right) - \Phi\left(\frac{\ln(K-\delta)-\mu_P}{\sigma_P}\right). \tag{2.21}$$

We structure the rest of proof as follows. We first show that $\exists \delta > 0$ such that $\mathbb{E}[K^{\text{INT}}(P;K,\delta)] < \mathbb{E}[K^{\text{INT}}(P;K,0)] = K$ if and only if $K > \mathbb{E}[P]\exp(-\sigma_P^2/2)$. By Lemma 2.7.4, this implies the validity of the first claim of the proposition, i.e., that $\exists \delta > 0$ such that $C_{SC}^{\text{INT}}(\alpha,K,\delta) < C_{SC}^{\text{INT}}(\alpha,K,0) = C_{SC}(\alpha,K)$ if and only if $K > \mathbb{E}[P]\exp(-\sigma_P^2/2)$. We subsequently prove that $\mathbb{E}[K^{\text{INT}}(P;K,\cdot)]$, and thus $C_{SC}^{\text{INT}}(\alpha,K,\cdot)$, attain global minimum at $\delta^* = \sqrt{K^2 - \exp(2\mu_P)} > 0$ if $K > \mathbb{E}[P]\exp(-\sigma_P^2/2)$.

Suppose $\mathbb{E}[K^{\text{INT}}(P;K,\delta)] < K$ for some δ. Then $\exists \bar{\delta} > 0$ such that the derivative (Equation 2.21) is negative at $\bar{\delta}$ because $\mathbb{E}[K^{\text{INT}}] = K$ at $\delta = 0$ and is a continuous function. The following implications hold.

$$\begin{aligned}
\frac{\partial \mathbb{E}[K^{\text{INT}}(P;K,\delta)]}{\partial \delta}\Big|_{\delta=\bar{\delta}} < 0 &\implies 1 - \Phi\left(\frac{\ln(K+\bar{\delta})-\mu_P}{\sigma_P}\right) < \Phi\left(\frac{\ln(K-\bar{\delta})-\mu_P}{\sigma_P}\right) \\
&\implies \ln(K+\bar{\delta}) - \mu_P > \mu_P - \ln(K-\bar{\delta}) \\
&\implies K^2 > \bar{\delta}^2 + \exp(2\mu_P) \\
&\implies K > \exp(\mu_P) = \mathbb{E}[P]\exp(-\sigma_P^2/2).
\end{aligned}$$

To prove the reverse direction of the iff result, suppose $K > \mathbb{E}[P]\exp(-\sigma_P^2/2) = \exp(\mu_P)$. The derivative (Equation 2.21) evaluated in $\delta = 0$ is negative because

$$\frac{\partial \mathbb{E}[K^{\text{INT}}(P;K,\delta)]}{\partial \delta}\Big|_{\delta=0} = 1 - 2\Phi\left(\frac{\ln(K)-\mu_P}{\sigma_P}\right) < 1 - 2\Phi\left(\frac{\ln(\exp(\mu_P))-\mu_P}{\sigma_P}\right) = 0,$$

where the last equality is a consequence of $\Phi(0) = 0.5$. Therefore, the expected unit cost is lower than K in the proximity of 0. In other words, there exists a sufficiently small $\delta > 0$ such that $\mathbb{E}[K^{\text{INT}}(P; K, \delta)] < K$. The first claim of the proposition is thus proven by invoking Lemma 2.7.4.

Next, we want to determine the global minimum δ^* of $\mathbb{E}[K^{\text{INT}}(P; K, \delta)]$ when $K > \exp(\mu_P)$. To this end, we provide a characterization of $\mathbb{E}[K^{\text{INT}}(P; K, \delta)]$ as a function of δ. First, we characterize a region for δ in which the function is convex as follows.

$$\frac{\partial^2 \mathbb{E}[K^{\text{INT}}(P; K, \delta)]}{\partial \delta^2} = \frac{-1}{(K+\delta)\sigma_P}\phi\left(\frac{\ln(K+\delta) - \mu_P}{\sigma_P}\right) + \frac{1}{(K-\delta)\sigma_P}\phi\left(\frac{\ln(K-\delta) - \mu_P}{\sigma_P}\right) \geqslant 0 \tag{2.22}$$

$$\iff \frac{1}{\sqrt{2\pi\,\sigma_P^2}}\left[\frac{-1}{(K+\delta)}\exp\left(\frac{-(\ln(K+\delta)-\mu_P)^2}{2\sigma_P^2}\right) + \frac{1}{(K-\delta)}\exp\left(\frac{-(\ln(K-\delta)-\mu_P)^2}{2\sigma_P^2}\right)\right] \geqslant 0$$

$$\iff -\frac{1}{(K+\delta)}\exp\left(\frac{-(\ln(K+\delta)-\mu_P)^2}{2\sigma_P^2}\right) + \frac{1}{(K-\delta)}\exp\left(\frac{-(\ln(K-\delta)-\mu_P)^2}{2\sigma_P^2}\right) \geqslant 0$$

$$\iff \exp\left(\frac{-\ln(K+\delta)^2 + \ln(K-\delta)^2 + 2\mu_P\ln(K+\delta) - 2\mu_P\ln(K-\delta)}{2\sigma_P^2}\right) \leqslant \frac{K+\delta}{K-\delta}$$

$$\iff \exp\left(\frac{-\ln(K^2-\delta^2) + 2\mu_P}{2\sigma_P^2}\ln\left(\frac{K+\delta}{K-\delta}\right)\right) \leqslant \frac{K+\delta}{K-\delta}$$

$$\iff -\ln\left(K^2-\delta^2\right) + 2\mu_P \leqslant 2\sigma_P^2$$

$$\iff \delta \leqslant \sqrt{K^2 - \exp(2\mu_P - 2\sigma_P^2)} =: \widehat{\delta}.$$

Therefore, the function $\mathbb{E}[K^{\text{INT}}(P; K, \cdot)]$ is convex for $\delta \in [0, \widehat{\delta}]$. Moreover, since $2\mu_P - 2\sigma_P^2 > -\infty$ and $\exp(\cdot)$ is a strictly increasing function, we have $\widehat{\delta} < K$. Hence, $\mathbb{E}[K^{\text{INT}}(P; K, \cdot)]$ is concave for $\delta \in [\widehat{\delta}, K)$. We further characterize $\mathbb{E}[K^{\text{INT}}(P; K, \delta)]$ by showing that it is increasing in δ when $\delta > \widehat{\delta}$. Since the second derivative (Equation 2.22) is negative in the interval $[\widehat{\delta}, K)$, then

the first derivative (Equation 2.21) is a decreasing function in $[\hat{\delta}, K)$. As a result, Equation 2.21 attains its infimum for $\delta \to K$ and the following inequalities hold:

$$\begin{aligned}
\frac{\partial \mathbb{E}[K^{INT}(P;K,\delta)]}{\partial \delta}\Big|_{[\hat{\delta},K)} &\geqslant \inf_{\delta \in [\hat{\delta},K)} \frac{\partial \mathbb{E}[K^{INT}(P;K,\delta)]}{\partial \delta} \\
&= \lim_{\delta \to K} 1 - \Phi\left(\frac{\ln(K+\delta)-\mu_P}{\sigma_P}\right) - \Phi\left(\frac{\ln(K-\delta)-\mu_P}{\sigma_P}\right) \\
&= 1 - \Phi\left(\frac{\ln(2K)-\mu_P}{\sigma_P}\right) \\
&> 0,
\end{aligned}$$

where the last inequality is strict because $\Phi\left((\ln(2K)-\mu_P)/\sigma_P\right) < 1$. Since $\mathbb{E}[K^{INT}(P;K,\delta)]$ is strictly concave increasing for $\delta \geqslant \hat{\delta}$ and convex when $\delta \leqslant \hat{\delta}$, its global minimum can be found using the first-order condition.

$$\begin{aligned}
\frac{\partial \mathbb{E}[K^{INT}(P;K,\delta)]}{\partial \delta} = 0 &\iff 1 - \Phi\left(\frac{\ln(K+\delta)-\mu_P}{\sigma_P}\right) = \Phi\left(\frac{\ln(K-\delta)-\mu_P}{\sigma_P}\right) \\
&\iff \ln(K+\delta) - \mu_P = \mu_P - \ln(K-\delta) \\
&\iff \delta = \sqrt{K^2 - \exp(2\mu_P)} = \sqrt{K^2 - \mathbb{E}_0[P_1]^2 \exp(-\sigma_P^2)} =: \delta^*,
\end{aligned}$$

where the second implication is a consequence of the symmetric property of the normal distribution, i.e. $\ln(K+\delta)$ and $\ln(K-\delta)$ have equal distance from the mean of $\ln(P)$. In conclusion, $\mathbb{E}[K^{INT}(P;K,\delta)]$ has global minimum at $\delta = \delta^*$ if $K > \mathbb{E}[P]\exp(-\sigma_P^2/2)$. From Lemma 2.7.4, δ^* is also global minimum of $C_{SC}^{INT}(\alpha, K, .)$ in this case. □

Next, we prove Proposition 2.3.1 by first formalizing its assumptions and then establishing Lemma 2.7.5.

Assumption 2. *It holds that* ***(i)*** $z_m^{min} = 0$ *and* $z_m^{max} < \infty$ *for all* $m \in \mathcal{M}$*;* ***(ii)*** $\mathbb{E}[P_j D_j | w_i] < \infty$, $\mathbb{E}[R_j D_j | w_i] < \infty$, *and* $\mathbb{E}[|P_j||w_i] < \infty$, $\forall j \in \mathcal{I}_{i+1} \cup \{I\}$*; and* ***(iii)*** $\mathbb{E}\left[\mathbb{E}[P_j D_j | w_{i+1}]|w_i\right] < \infty$, $\mathbb{E}\left[\mathbb{E}[R_j D_j | w_{i+1}]|w_i\right] < \infty$, *and* $\mathbb{E}\left[\mathbb{E}[|P_j||w_{i+1}]w_i\right] < \infty$, $\forall j \in \mathcal{I}_{i+2} \cup \{I\}$.

Lemma 2.7.5. *At stage* $i \in \mathcal{I} \cup I$*, given* $0 \leqslant w_i < \infty$*, if Assumption 2 holds, then* $|V_i(\cdot, w_i)| < \infty$ *and* $|\mathbb{E}\left[V_{i+1}(\cdot, w_{i+1})\middle| w_i\right]| < \infty$.

Proof. We start by showing that the value function $V_i(\cdot, w_i)$ is bounded from above for $i \in \mathcal{I} \cup \{I\}$. From the definition of MDP (Equation 2.12) at stage i, we know that for each policy $\pi_i \in \Pi_i$, we have

$$V_i(x_i, w_i) \leqslant \mathbb{E}\left[\sum_{j \in \mathcal{I}_i} \gamma^{j-i} c_j(x_j^{\pi_i}, w_j, Z_j^{\pi_i}(x_j^{\pi_i}, w_j)) + \gamma^{I-i} c_I(x_I^{\pi_i}, w_I) \,\middle|\, x_i, w_i\right].$$

By choosing π_i to be the policy that only procures from the short-term market (i.e. CPPA procurement decisions $z_{i,m}$ are always zero), we obtain the following inequalities:

$$V_i(x_i, w_i) \leqslant \mathbb{E}\left[\sum_{j \in \mathcal{I}_i \cup \{I\}} \gamma^{j-i}(P_j D_j + \alpha R_j D_j)\middle| w_i\right] \tag{2.23a}$$

$$\leqslant \sum_{j \in \mathcal{I}_i \cup \{I\}} \Big(\mathbb{E}\left[P_j D_j | w_i\right] + \alpha \mathbb{E}\left[R_j D_j | w_i\right]\Big), \tag{2.23b}$$

where Equation 2.23a is obtained by accounting for the costs of short-term procurement and Equation 2.23b by dropping the discount factor $\gamma \leqslant 1$ and the linearity of the expectation operator. The value function $V_i(\cdot, w_i)$ being bounded follows from Assumption 2. Next, we have

$$\mathbb{E}\left[V_{i+1}(x_{i+1}, w_{i+1})|w_i\right] \leqslant \sum_{j \in \mathcal{I}_{i+1} \cup \{I\}} \Big(\mathbb{E}\left[\mathbb{E}[P_j\, D_j | w_{i+1}]|w_i\right] + \alpha \mathbb{E}\left[\mathbb{E}[R_j\, D_j | w_{i+1}]|w_i\right]\Big),$$

where the individual terms in the right hand side are bounded by Assumption 2.

To show that $V_i(\cdot, w_i)$ is bounded from below, we distinguish between PCs and SCs. In case of PCs, $V_i(\cdot, w_i)$ is trivially bounded from below by zero because all terms in the cost function (Equation 2.10) are non-negative. Consequently, $\mathbb{E}\left[V_{i+1}(\cdot, w_{i+1})\middle|w_i\right]$ is also bounded from below by zero. In the case of SCs, the cost function $c_i(x_i, w_i, z_i)$ contains the term $P_i(D_i - x_{i,0})$ which can be negative, and it holds that $c_j(x_j, w_j, z_j) \geqslant -|P_j\, x_{j,0}| = -|P_j|\, x_{j,0}$ for each $j \in \mathcal{I}_i \cup \{I\}$. Since the actions at each stage are bounded, there exists a value N^u such that $x_{j,0} \leqslant N^u$ for all $j \in \mathcal{I}_i \cup \{I\}$. Hence, $c_j(x_j, w_j, z_j) \geqslant -|P_j|\, N^u$. Therefore, we obtain

$$V_i(x_i, w_i) \geqslant \mathbb{E}\left[\sum_{j\in\mathcal{I}_i\cup\{I\}} -\gamma^{j-i}|P_j\, x_{j,0}|\middle|\, x_i, w_i\right] \geqslant N^u \sum_{j\in\mathcal{I}_i\cup\{I\}} -\gamma^{j-i}\mathbb{E}\big[|P_j|\,\big|w_i\big],$$

which is bounded for a given $0 \leqslant w_i < \infty$ by Assumption 2. Therefore, $V_i(\cdot, w_i)$ is bounded from below. Finally, holds that

$$\begin{aligned}
\mathbb{E}\left[V_{i+1}(\cdot, w_{i+1})\middle|w_i\right] &\geqslant \sum_{j\in\mathcal{I}_{i+1}\cup\{I\}} -\gamma^{j-i}\mathbb{E}\left[\mathbb{E}\left[|P_j\, x_{j,0}|\middle|\, x_{i+1}, w_{i+1}\right]\middle|\, x_i, w_i\right] \\
&\geqslant N^u \sum_{j\in\mathcal{I}_{i+1}\cup\{I\}} -\gamma^{j-i}\mathbb{E}\left[\mathbb{E}\left[|P_j|\middle|, w_{i+1}\right]\middle|w_i\right] \\
&> -\infty,
\end{aligned}$$

where the last inequality holds by Assumption 2. Therefore, it holds that $|V_i(\cdot, w_i)| < \infty$ and $|\mathbb{E}\left[V_{i+1}(\cdot, w_{i+1})\middle|w_i\right]| < \infty$ for both PCs and SCs. □

Proof of Proposition *2.3.1.* Since $z_m^{\min} = 0$ for each $m \in \mathcal{M}$, we have that $\mathcal{Z}_i(a_i)$ is a convex set. Consider the Bellman recursion associated with Equation 2.12:

$$V_I(x_I, w_I) = c_I(x_I, w_I), \ \forall (x_I, w_I) \in \mathcal{X}_I \times \mathcal{W}_I,$$

$$V_i(x_i, w_i) = \min_{z_i \in \mathcal{Z}_i(a_i)} \Big\{ c_i(x_i, w_i, z_i) + \gamma \mathbb{E}\big[V_{i+1}\left(f_i(x_i, z_i), w_{i+1}\right) | w_i \big] \Big\}, \ \forall (i, x_i, w_i) \in \mathcal{I} \times \mathcal{X}_i \times \mathcal{W}_i.$$

At a given stage i, let $\mathcal{C}_i := \{(x_i, z_i) | x_i \in \mathcal{X}_i, z_i \in \mathcal{Z}\}$ denote the set of actions and states at i, and also define

$$G_i(x_i, w_i, z_i) := c_i(x_i, w_i, z_i) + \gamma \mathbb{E}[V_{i+1}(f(x_i, z_i), w_{i+1}) | w_i].$$

To prove the convexity of MDP (Equation 2.12), we start by showing that $G_{I-1}(x_{I-1}, w_{I-1}, z_{I-1})$ is jointly convex in the state x_{I-1} and action z_{I-1}, i.e. in set $\mathcal{C}_{I-1}$, at stage $I-1$. For a given w_{I-1}, we have:

$$V_{I-1}(x_{I-1}, w_{I-1}) = \min_{z_{I-1} \in \mathcal{Z}_{I-1}(a_{I-1})} \Big\{ c_{I-1}(x_{I-1}, w_{I-1}, z_{I-1}) + \gamma \mathbb{E}\big[V_I(f(x_{I-1}, z_{I-1}), w_I) | w_{I-1} \big] \Big\}.$$

We know from Lemma 2.7.5 that $V_{I-1}(x_{I-1}, w_{I-1})$ and $\mathbb{E}\big[V_I(f(x_{I-1}, z_{I-1}), w_I) | w_{I-1}\big]$ are finite quantities. Based on the definition in Equation 2.10–Equation 2.11, $c_{I-1}(x_{I-1}, w_{I-1}, z_{I-1})$ is piece-wise linear and convex in x_{I-1}. Furthermore, the terminal value function $V_I(x_I, w_I) = c_I(x_I, w_I)$ is convex in x_I for a given w_I based on the definition of terminal cost. Thus, the continuation function $\gamma \mathbb{E}[V_I(f(x_{I-1}, z_{I-1}), w_I) | w_{I-1}]$ is convex in x_{I-1}. In addition, the set $\mathcal{C}_{I-1}$ is convex since both $\mathcal{X}_{I-1}$ and $\mathcal{Z}_{I-1}(a_{I-1})$ are convex. As a consequence, for a realization of w_{I-1}, the function $G_{I-1}(z_{I-1}, x_{I-1}, w_{I-1})$ is convex in the set $\mathcal{C}_{I-1}$. The convexity of $V_{I-1}(x_{I-1}, w_{I-1})$ as a function of x_{I-1} now follows from Proposition B-4 in [94].

Suppose $V_{i+1}(\cdot, w_{i+1})$ is convex in x_{i+1}. Following the same steps above, we can show that $G_i(x_i, w_i, z_i)$ is convex in $\mathcal{C}_i$ and consequently $V_i(\cdot, w_i)$ is convex in x_i. Hence, the result follows from mathematical induction. □

Proof of Proposition *2.4.1.* We first prove the feasibility of $\mathcal{H}_i(z_i(W_i))$ under the two conditions of Proposition 2.4.1.

1. Since $\mathcal{Z}_i(a_i)$ is convex and bounded, for any $W_i|w_i$, the optimal decision $z_i(W_i)$ at stage i is finite. Moreover, the average of the actions for the random variable W_i given w_i, $\mathbb{E}[z_i(W_i)|w_i]$, belongs to the set $\mathcal{Z}_i(a_i)$; thus, $\mathbb{E}[z_i(W_i)|w_i]$ is feasible.

2. $\mathcal{H}_i(z_i(W_i))$ is feasible since $\bar{z}_i$ is feasible and finite based on the statement of the condition and boundedness of the set $\mathcal{Z}_i(a_i)$.

Next, assuming $\mathcal{H}_i$ satisfies one of these conditions, we establish its optimality when using ideal dual penalties in math program Equation 2.14a–Equation 2.14e. Specifically, we show that the decision at every sample path W_i is optimal. Our proof relies on the stochastic dynamic programming (SDP) reformulation of Equation 2.14a–Equation 2.14e at stage i, which is

$$V_I^{IR}(x_I; W_I) = c_I(x_I, w_I), \quad \forall x_I \in \mathcal{X}_I, \tag{2.24a}$$

$$V_j^{IR}(x_j; W_j) = \min_{z_j \in \mathcal{Z}_j(a_j)} \Big\{ c_j(x_j, w_j, z_j) - q_j(x_j, z_j, W_j) + \gamma V_{j+1}^{IR}\big(f_j(x_j, z_j)\big) \Big\}, \quad \forall j \in \mathcal{I}_i, \ \forall x_j \in \mathcal{X}_i. \tag{2.24b}$$

At stage I, it is true that $V_I^{IR}(x_I; W_I) = c_I(x_I, w_I) = V_I(x_I, w_I)$ for each $x_I \in \mathcal{X}_I$. By backward induction, we assume $V_{j+1}^{IR}(x_{j+1}; W_{j+1})$ in Equation 2.14 equals $V_{j+1}(x_{j+1}, w_{j+1})$ in Equation 2.12

and prove the equality for stage j. The following equation represents the stage-j step of SDP (Equation 2.24) when the ideal dual penalty (Equation 2.15) is used:

$$V_j^{IR}(x_j; W_j) = \min_{z_j \in \mathcal{Z}} \Big\{ c_j(x_j, w_j, z_j) - \gamma V_{j+1}\left(f_j(x_j, z_j), w_{j+1}\right) + \gamma \mathbb{E}\big[V_{j+1}\big(f_j(x_j, z_j), w_{j+1}\big)\big|w_j\big] + \gamma V_{j+1}^{IR}\left(f_j(x_j, z_j)\right) \Big\}, \quad \forall x_j \in \mathcal{X}_j. \tag{2.25a}$$

$$= \min_{z_j \in \mathcal{Z}} \Big\{ c_j(x_j, w_j, z_j) + \gamma \mathbb{E}\big[V_{j+1}\big(f_j(x_j, z_j), w_{j+1}\big)\big|w_j\big] \Big\}, \quad \forall x_j \in \mathcal{X}_j. \tag{2.25b}$$

$$= V_j(x_j, w_j), \quad \forall x_j \in \mathcal{X}_j, \tag{2.25c}$$

where Equation 2.25b follows from the induction hypothesis, and (Equation 2.25c) from the definition of the SDP associated with MDP Equation 2.12. The relation, $V_j^{IR}(x_j; W_j) = V_j(x_j, w_j)$, thus holds at the generic stage $j \in \mathcal{I}_i \cup \{I\}$ for the principle of mathematical induction. The optimality of the action for every W_i is immediate from this equality. Since $z_i(W_i)$ is equal to the optimal decision for any sample path, a feasible decision measure $\mathcal{H}_i$ that satisfies conditions 1 and 2 leads to an optimal solution to MDP (Equation 2.12). □

2.7.2 Deterministic renewable target

In this section, we analyze procurement costs for a PC and an SC under a deterministic RPPT and compare it with analogous costs under the stochastic RPPT discussed in §2.2.1. A deterministic RPPT involves satisfying a fraction α of known demand $\bar{D}$, that is, the RPPT is $\alpha\bar{D}$. Choosing a deterministic RPPT only affects the term corresponding to the shortfall in meeting the renewable target in the cost functions of PCs and SCs, i.e. Equation 2.1 and Equation 2.3, respectively. Specifically, the expression $\mathbb{E}[R(\alpha D - z)_+]$ in both functions is replaced by $\mathbb{E}[R](\alpha\bar{D} - z)_+$.

An analogous result to Proposition 2.2.2 holds for the expected procurement cost of PCs and SCs in the presence of a deterministic RPPT (we omit this result for brevity). However, the value of setting a deterministic target instead of a stochastic target is unclear and depends on both future and past power demands. Proposition 2.7.6 characterizes a region for $\bar{D}$ in which PCs and SCs with deterministic RPPTs can lead to higher procurement costs than with stochastic RPPTs. In (Equation 2.26a) and Equation 2.26b we define the optimal procurement cost, respectively, in PCs and SCs when the target is deterministic.

$$C_{PC,D}(\alpha, K) := \min_{z \geqslant 0} \left\{ Kz + \mathbb{E}\left[P(D - z)_+\right] + \mathbb{E}[R](\alpha\bar{D} - z)_+ \right\}; \tag{2.26a}$$

$$C_{SC,D}(\alpha, K) := \min_{z \geqslant 0} \left\{ \mathbb{E}[PD] + \mathbb{E}[(K - P)z] + \mathbb{E}[R](\alpha\bar{D} - z)_+ \right\}. \tag{2.26b}$$

Proposition 2.7.6. *Suppose Assumption 1 holds. Then, for each* $\alpha \in (0, 1]$, $C_{SC}(\alpha, K) < C_{SC,D}(\alpha, K)$ *if and only if* $\bar{D} > \frac{1}{2\alpha} z^*_{SC} + \frac{b}{2}$. *Moreover, assuming* $\mathbb{E}[R] \leqslant \mathbb{E}[P]$, *there exists an* $\alpha \in (0, 1]$ *such that* $C_{PC}(\alpha, K) < C_{PC,D}(\alpha, K)$ *if* $\bar{D} > \mathbb{E}[D] + \frac{(b-a)}{2} \frac{\mathbb{E}[P]}{\mathbb{E}[R]+\mathbb{E}[P]}$.

The proof of this proposition relies on Lemma 2.7.7.

Lemma 2.7.7. *Under Assumption 1, it holds that:*

(a) *The optimal PC procurement quantity* $z^*_{PC,D}$ *with deterministic RPPT,* $\alpha\bar{D}$, *is given by*

$$z^*_{PC,D} = \min \left\{ \alpha\bar{D}, b - \left(\frac{K - \mathbb{E}[R]}{\mathbb{E}[P]} \right) (b - a) \right\};$$

(b) $C_{PC,D}(\alpha, K)$ *is linear and increasing in* α *for* $\alpha = \left[0, a/\bar{D}\right) \cup \left[z'/\bar{D}, 1\right)$ *and strictly convex and increasing in* α *for* $\alpha \in \left[a/\bar{D}, \min\{z'/\bar{D}, 1\}\right)$, *where* $z' := b - \frac{K - \mathbb{E}[R]}{\mathbb{E}[P]}(b - a)$.

Proof. **(a)** The expected procurement cost for a PC with deterministic RPPT is

$$Q_{PC,D}(\alpha, K, z) = Kz + \mathbb{E}\left[P\left(D - z\right)_+\right] + \mathbb{E}\left[R\left(\alpha\bar{D} - z\right)_+\right]. \tag{2.27}$$

Similar to $Q_{PC}(\alpha, K, z)$, Equation 2.27 is also convex in the procurement quantity z. Therefore, the optimal quantity $z^*_{PC,D}$ minimizing Equation 2.27 can be determined by considering its first derivative:

$$\frac{dQ_{PC,D}(\alpha, K, z)}{dz} = K - \mathbb{E}\left[P\right] Pr\left(D \geqslant z\right) - \mathbb{E}\left[R\right] \mathbf{1}_{\{\alpha\bar{D} \geqslant z\}}.$$

If $z \geqslant \alpha\bar{D}$, then the indicator function in the last term is zero and the derivative is non-negative since $K \geqslant \mathbb{E}[P]$. Thus, the quantity $z^*_{PC,D}$ minimizing Equation 2.27 lies in the interval $[0, \alpha\bar{D}]$. In this interval, the indicator function is equal to 1 and the solution to $dQ_{P,D}(\alpha, K, z)/dz = 0$ is

$$z' = b - \left(\frac{K - \mathbb{E}[R]}{\mathbb{E}[P]}\right)(b - a).$$

This quantity is independent of α. Moreover, z' is greater than a because the strike price K is upper bounded by $\mathbb{E}[P] + \mathbb{E}[R]$ in Assumption 1. Since the optimal procurement quantity $z^*_{PC,D}$ must lie in the interval $[0, \alpha\bar{D}]$, we conclude that $z^*_{PC,D} = \min\{z', \alpha\bar{D}\}$.

(b) We distinguish three cases for the expected cost (Equation 2.27) depending on the value of α.

Case I): $\alpha\bar{D} < a$. Since $a < z'$, then $\alpha\bar{D} < z'$ and $z^*_{PC,D} = \min\{z', \alpha\bar{D}\} = \alpha\bar{D}$. The optimal cost is

$$\begin{aligned} C^I_{PC,D}(\alpha, K) :&= Q_{PC,D}(\alpha, K, z^*_{PC,D} = \alpha\bar{D}) \\ &= K\alpha\bar{D} + \mathbb{E}\left[P(D - \alpha\bar{D})_+\right] \\ &= K\alpha\bar{D} + \mathbb{E}\left[P(D - \alpha\bar{D})\right], \end{aligned}$$

where the last equality is due to $\alpha\bar{D} < a$. $C^I_{PC,D}(\alpha, K)$ is linear in α because its first derivative is independent of α:

$$\frac{dC^I_{PC,D}(\alpha, K)}{d\alpha} = (K - \mathbb{E}[P])\bar{D}.$$

Case II): $a \leqslant \alpha\bar{D} < z'$. In this case again it holds that $z^*_{PC,D} = \min\{z', \alpha\bar{D}\} = \alpha\bar{D}$, and the optimal procurement cost is

$$\begin{aligned} C^{II}_{PC,D}(\alpha, K) :&= Q_{PC,D}(\alpha, K, z^*_{PC,D} = \alpha\bar{D}) \\ &= K\alpha\bar{D} + \mathbb{E}\left[P\left(D - \alpha\bar{D}\right)_+\right]. \end{aligned}$$

The first and second derivative of this expression with respect to α are

$$\begin{aligned} \frac{dC^{II}_{PC,D}(\alpha, K)}{d\alpha} &= \left[K - \mathbb{E}[P]\frac{b - \alpha\bar{D}}{b - a}\right]\bar{D} > 0, \\ \frac{dC^{II^2}_{PC,D}(\alpha, K)}{d^2\alpha} &= \mathbb{E}[P]\frac{\bar{D}^2}{b - a} > 0. \end{aligned}$$

Therefore, the optimal procurement cost is convex increasing in α.

Case III): $z' < \alpha\bar{D}$. In this case, $z^*_{PC,D} = \min\{z', \alpha\bar{D}\} = z'$ and the optimal cost is

$$\begin{aligned} C^{III}_{PC,D}(\alpha, K) :&= Q_{PC,D}(\alpha, K, z^*_{PC,D} = z') \\ &= Kz' + \mathbb{E}[P]\frac{(z'-b)^2}{2(b-a)} + \mathbb{E}[R](\alpha\bar{D} - z'). \end{aligned}$$

Since the first derivative

$$\frac{dC^{III}_{PC,D}(\alpha, K)}{d\alpha} = \mathbb{E}[R]\bar{D}$$

does not depend on α, the procurement cost $C^{III}_{PC,D}(\alpha, K)$ increases linearly in the target level. Finally, the following relations hold between the slopes of the procurement costs in the aforementioned three cases.

$$\frac{dC^{I}_{PC,D}(\alpha, K)}{d\alpha} < \frac{dC^{II}_{PC,D}(\alpha, K)}{d\alpha} < \frac{dC^{III}_{PC,D}(\alpha, K)}{d\alpha}.$$

□

***Proof of Proposition** 2.7.6.* **(a)** The expected SC procurement cost with deterministic RPPT is

$$Q_{SC,D}(\alpha, K, z) = \mathbb{E}[PD] + \mathbb{E}[(K-P)z] + \mathbb{E}[R(\alpha\bar{D} - z)_+].$$

This function is continuous for $z \in \mathbb{R}_+$, and is differentiable for $z \in \mathbb{R}_+ \backslash \{\alpha\bar{D}\}$ with derivative

$$\frac{dQ_{SC,D}(\alpha, K, z)}{dz} = K - \mathbb{E}[P] - \mathbb{E}[R]\,\mathbf{1}_{\{\alpha\bar{D} \geqslant z\}}.$$

This expression as a function of z is a non-positive constant if $z \leqslant \alpha\bar{D}$, and a non-negative constant otherwise. In fact, in the former case the indicator function is 1 and it holds that $K - \mathbb{E}[P] - \mathbb{E}[R] \leqslant 0$ due to Assumption 1, whereas in the latter case the indicator function is

zero and it holds that $K - \mathbb{E}[P] \geqslant 0$ also due to Assumption 1. It follows that $z^*_{SC,D} = \alpha\bar{D}$ is an optimal procurement quantity, and the associated optimal cost is

$$C_{SC,D}(\alpha, K) = Q_{SC,D}(\alpha, K, z^*_{SC,D} = \alpha\bar{D}) = \mathbb{E}[PD] + \mathbb{E}[(K - P)]\alpha\bar{D}.$$

The slope of the optimal cost function with respect to the target is

$$\frac{dC_{SC,D}(\alpha, K)}{d\alpha} = (K - \mathbb{E}[P])\bar{D}.$$

Given that both functions $C_{SC}(\alpha, K)$ and $C_{SC,D}(\alpha, K)$ are linear increasing in α (see also Proposition 2.2.2) and are equal at $\alpha = 0$, it follows that $C_{SC}(\alpha, K) < C_{SC,D}(\alpha, K)$ in $\alpha \in (0, 1]$ if and only if the analogous condition on their slope holds, i.e., $dC_{SC}(\alpha, K)/d\alpha < dC_{SC,D}(\alpha, K)/d\alpha$. Below we establish a necessary and sufficient condition for this relation to be true.

$$\begin{aligned}\frac{dC_{SC}(\alpha, K)}{d\alpha} < \frac{dC_{SC,D}(\alpha, K)}{d\alpha} &\iff (K - \mathbb{E}[P])\frac{z^*_{SC}}{\alpha} + \mathbb{E}\left[R\left(D - \frac{z^*_{SC}}{\alpha}\right)_+\right] < (K - \mathbb{E}[P])\bar{D} \\ &\iff \bar{D} > \frac{(a - b)(K - \mathbb{E}[P])}{2\mathbb{E}[R]} + b = \frac{1}{2\alpha}z^*_{SC} + \frac{b}{2}.\end{aligned}$$

where the second implication is obtained by replacing z^*_{SC} with its expression given in Lemma 2.7.1 and simplifying the resulting term.

(b) The functions $C_{PC}(\alpha, K)$ and $C_{PC,D}(\alpha, K)$ have been characterized in Proposition 2.2.2 and Lemma 2.7.7, respectively. Both $C_{PC,D}(\alpha, K)$ and $C_{PC}(\alpha, K)$ are convex and increasing functions of α, and it holds that $C_{PC,D}(1, K) = C_{PC}(1, K)$ for

$$\bar{D} = D' := (b - a)\frac{\left(K - (\mathbb{E}[R] + \mathbb{E}[P])\right)^2 + \mathbb{E}[R](\mathbb{E}[P] + \mathbb{E}[R]) - (\mathbb{E}[P] + \mathbb{E}[R])^2}{2\mathbb{E}[P](\mathbb{E}[R] + \mathbb{E}[P])} + b.$$

This expression for D' belongs to the interval $[a, b]$ and can be derived as follows:

$$
\begin{aligned}
& C_{PC}(1, K) = C_{PC,D}(1, K) \\
& \iff K z^*_{PC} + \mathbb{E}[P]\frac{(z^*_{PC} - b)^2}{2(b-a)} + \mathbb{E}[R]\frac{(z^*_{PC} - b)^2}{2(b-a)} = K z' + \mathbb{E}[P]\frac{(z' - b)^2}{2(b-a)} + \mathbb{E}[R](\bar{D} - z') \\
& \iff \mathbb{E}[R]\bar{D} = K(z^*_{PC} - z') + \mathbb{E}[P]\frac{(z^*_{PC} - b)^2 - (z' - b)^2}{2(b-a)} + \mathbb{E}[R]\left(\frac{(z^*_{PC} - b)^2}{2(b-a)} + z'\right) \\
& \iff \bar{D} = \frac{z^*_{PC} - z'}{\mathbb{E}[R]}\left(K + \frac{\mathbb{E}[P](z^*_{PC} + z' - 2b)}{2(b-a)}\right) + \frac{(z^*_{PC} - b)^2}{2(b-a)} + z' \\
& \qquad = (b-a)\frac{\big(K - (\mathbb{E}[R] + \mathbb{E}[P])\big)^2 + \mathbb{E}[R](\mathbb{E}[P] + \mathbb{E}[R]) - (\mathbb{E}[P] + \mathbb{E}[R])^2}{2\mathbb{E}[P](\mathbb{E}[R] + \mathbb{E}[P])} + b \qquad (2.28) \\
& \qquad = D',
\end{aligned}
$$

where Equation 2.28 is obtained by replacing z^*_{PC} and z' by their expressions based on Lemmas 2.2.1 and 2.7.7, respectively, and simplifying. Since $C_{PC,D}(\alpha, K)$ is strictly increasing in the value of $\bar{D}$, we can claim that if $\bar{D} > D'$, $C_{PC,D}(1, K) > C_{PC}(1, K)$. Therefore, there exists $\alpha \in (0, 1]$ that satisfies $C_{PC,D}(\alpha, K) > C_{PC}(\alpha, K)$.

Finally, the value of D' can be bounded from above as follows:

$$
\begin{aligned}
D' &\leqslant (b-a)\frac{-\mathbb{E}[P]\big(\mathbb{E}[P] + \mathbb{E}[R]\big) + \mathbb{E}[R]^2}{2\mathbb{E}[P](\mathbb{E}[R] + \mathbb{E}[P])} + b \\
&\leqslant -(b-a)\frac{\mathbb{E}[R]}{2(\mathbb{E}[R] + \mathbb{E}[P])} + b \\
&= \mathbb{E}[D] + \frac{(b-a)}{2}\frac{\mathbb{E}[P]}{\mathbb{E}[R] + \mathbb{E}[P]},
\end{aligned}
$$

where the first inequality holds by replacing K with $\mathbb{E}[P]$ since $K \geqslant \mathbb{E}[P]$, and simplifying, and the second inequality results from the assumption $\mathbb{E}[P] \geqslant \mathbb{E}[R]$. Therefore, there exists $\alpha \in (0, 1]$ such that $C_{PC,D}(\alpha, K) > C_{PC}(\alpha, K)$ if $\bar{D} > \mathbb{E}[D] + \frac{(b-a)}{2}\frac{\mathbb{E}[P]}{\mathbb{E}[R]+\mathbb{E}[P]}$. □

Proposition 2.7.6 provides support to the fact that deterministic RPPTs are not always cost-beneficial and companies with such targets might incur higher procurement costs compared to a stochastic RPPT. In particular, stochastic RPPTs might be cost-beneficial for companies with a negative drift in their power consumption due to, for instance, investments in energy efficiency programs (see, e.g., the supplement of [46] for more details on companies with both an RPPT and energy efficiency initiatives). These companies might benefit from committing to a stochastic RPPT to lower the energy costs in spite of adding uncertainty to their RPPT fulfillment.

2.7.3 Non-convexity of the value function

We show that the value function $V_i(\cdot, w_i)$ of MDP (Equation 2.12) is non-convex in the endogenous state x_i by using a simple counter-example with two periods (i.e., 0 and 1), in which a PC can be entered at stage 0 with delivery at stage 1. We assume that $a_0 = 1$, i.e., the contract is available, and that minimum and maximum procurement quantities are $z^{\min} = 6$ MWh and $z^{\max} = 100$, respectively. Thus, a feasible action z_0 belongs to the action set $\mathcal{Z}_0(a_0) = \{0\} \cup [6, 100] \subset \mathbb{R}_+$. Moreover, we assume that demand is constant, i.e. $D_0 = D_1 = 10$ MWh, power and REC prices have values $P_0 = R_0 = 10$ USD/MWh and are martingales (i.e., $\mathbb{E}[P_1|P_0] = 10$ and $\mathbb{E}[R_1|R_0] = 10$), the strike price is $K = 11$ USD/MWh, and $\alpha = 0.8$. We consider the cost functions corresponding to a PC. Proceeding backward, the terminal value function (stage 1) is convex in the state x_1 as it is equal to

$$V_1(x_1, w_1) = P_1(D_1 - x_1)_+ + R_1(\alpha D_1 - x_1)_+,$$

which has the slope $P_1 + R_1$ for $x_1 \in [0, \alpha D_1]$, the smaller slope P_1 for $x_1 \in [\alpha D_1, D_1]$, and is equal to zero if $x_1 \geqslant 10$. Instead, the stage 0 value function is given by

$$\begin{aligned} V_0(x_0, w_0) &= \min_{z_0 \in \mathcal{Z}_0(a_0)} \left\{ K z_0 + \mathbb{E}_0 \left[P_1 (D_1 - x_0 - z_0) \mathbf{1}_{\{D_1 > x_0 + z_0\}} + R_1 (\alpha D_1 - x_0 - z_0) \mathbf{1}_{\{\alpha D_1 > x_0 + z_0\}} \right] \right\} \\ &= \min_{z_0 \in \mathcal{Z}_0(a_0)} \left\{ K z_0 + P_0 (D_0 - x_0 - z_0) \mathbf{1}_{\{D_0 > x_0 + z_0\}} + R_0 (\alpha D_0 - x_0 - z_0) \mathbf{1}_{\{\alpha D_0 > x_0 + z_0\}} \right\}, \end{aligned}$$

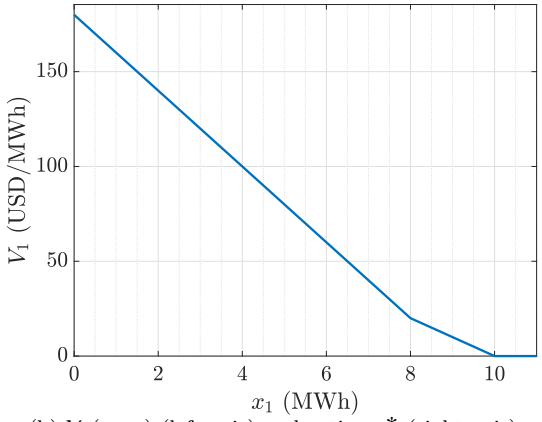
which is non-convex in x_0 as illustrated in Figure 7. In particular, Figure 7(a) displays the value function $V_1(\cdot, w_1)$ as a function of x_1 for $P_1 = R_1 = 10$, and Figure 7(b) plots $V_0(\cdot, w_0)$ and z_0^* as functions of x_0 for w_0, which is known. For $x_0 \in [4, 5.7]$, the value function $V_0(\cdot, w_0)$ takes a constant value because it is optimal to enter into a PC of size $z^{\min} = 6$, which causes over-procurement, and the expected short-term cost is zero. Instead, when $x_0 > 5.7$, the over-procurement cost exceeds the benefit of enjoying lower power price ($K = 11 < 20 = \mathbb{E}[P_1 + R_1 | P_0, R_0]$), and the optimal action jumps from $z_0^* = z^{\min} = 6$ to $z_0^* = 0$, i.e., it becomes optimal to use the short-term option.

2.7.4 Model of market dynamics and calibration

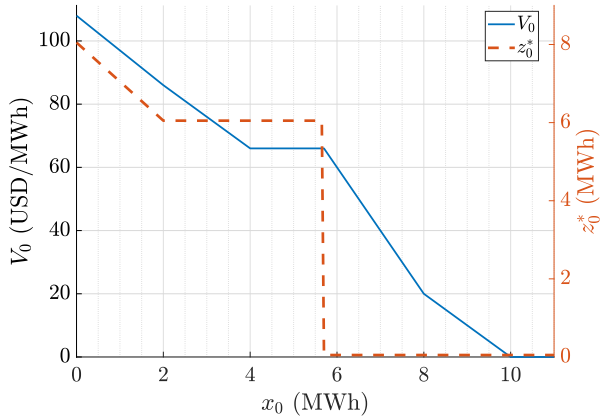
In this section we present in detail the stochastic processes used to describe the evolution of the power price, REC price, and power demand, and discuss their calibration using market data.

The evolution of electricity prices has been studied using various processes that capture features such as seasonality [118], mean-reversion and long-term trends ([161] and references therein), and jumps [43,76,168,198,199]. To obtain a power price model that captures the main features of spot electricity prices, we construct a mean-reverting stochastic process with jumps and seasonality. We use a continuous-time process for the power price $\{P_t,\, t \in \mathbb{R}_+\}$, and then

Figure 7: Stage 1 and stage 0 MDP value functions.
(a) Value function $V_1(\cdot, w_1)$ with $P_1 = R_1 = 10$.



(b) $V_0(\cdot, w_0)$ (left axis) and action z_0^* (right axis).



consider in our decision model discrete-time values $\{P_i, i \in \mathcal{I} \cup \{I\}\}$, which are the values taken by this process at the beginning of stages $i \in \mathcal{I} \cup \{I\}\}$ (we do the same for the REC price and demand process). Following [199], the power price model is written as:

$$\ln(P_t) = \chi_t + g(t), \tag{2.29a}$$

$$d\chi_t = (\nu_P - \mathcal{K}_P \chi_t)\, dt + \sigma_P\, dW_t + J(\mu_J, \sigma_J)\, d\Pi(\lambda), \tag{2.29b}$$

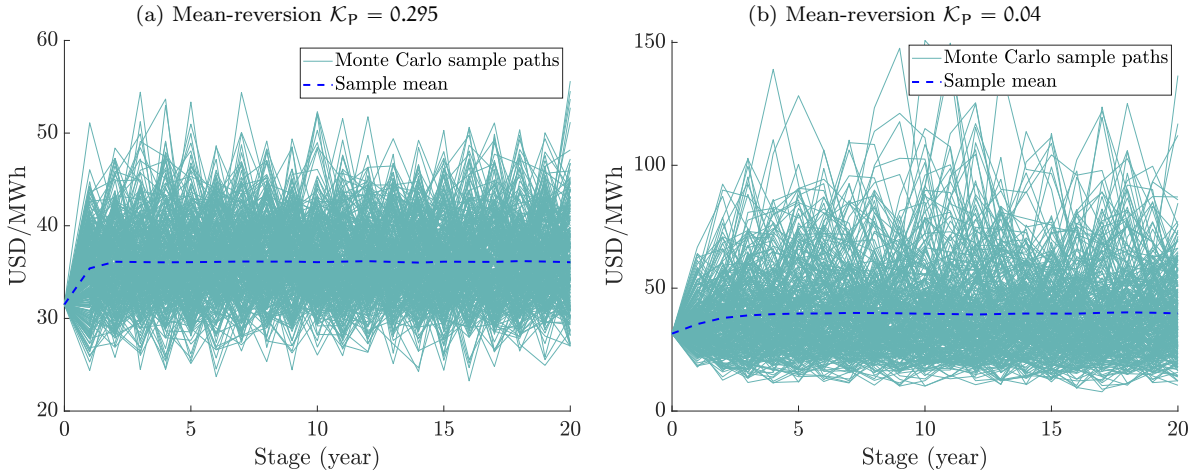
$$g(t) = \phi_0 + \sum_{k=1}^{12} \phi_k \hat{P}_t^k. \tag{2.29c}$$

Equation 2.29a describes the log power price process as the sum of a stochastic component χ_t and a deterministic component $g(t)$. The stochastic component evolves according to Equation 2.29b, where $\mathcal{K}_P$ is the speed of mean reversion, ν_P models the drift, σ_P is the volatility, and W_t is a standard Brownian motion. We model spikes in monthly prices by a jump diffusion process in which the jump size follows a normal distribution $J(\mu_J, \sigma_J)$ and the jump frequency a Poisson distribution $\Pi(\lambda)$ [43]. The deterministic function $g(t)$ in Equation 2.29c models the monthly price seasonality by using a constant ϕ_k for each month k, and binary values $\hat{P}_t^k$ equal to one if time t falls in month k and zero otherwise.

We calibrated the parameters of model (Equation 2.29) using historical monthly power price data from the Pennsylvania New Jersey Maryland Interconnection LLC (PJM) market during the period January 2010–August 2017. Analyzing power prices, we found that the jump frequency and intensity are small when considering monthly prices. We thus tested the number of jumps in the monthly power prices using the algorithm presented in [199, p. 1047], and found the jump diffusion parameters to be insignificant. Therefore, we removed jumps from the model and only focused on a mean-reverting process with seasonality. We first estimated the seasonality function $g(t)$ directly from the data using linear regression, resulting in the coefficients $\{\phi_k, k = 0 \ldots, 12\}$ = {3.519, 0.258,0.163,0.078,-0.026,0.003,0.039,0.174,0.013,0.009,-0.037,-0.038,0.000}. Then, we calibrated the mean reverting coefficients using maximum likelihood estimation. The resulting estimates were $\mathcal{K}_P = 0.295$, $\sigma_P = 0.178$ (both with a p-value below 0.001), and $\nu_P = 0$. We set $P_0 = 31.5$ USD/MWh which is the average power price observed in 2017. When plotting the sample paths using the calibrated parameters, we noticed

that the speed of mean reversion was too strong and so we decreased it from $\mathcal{K}_{\mathrm{P}} = 0.295$ to $\mathcal{K}_{\mathrm{P}} = 0.04$ for our numerical study. We illustrate this effect in Figure 8(a) and Figure 8(b), which display 300 power price sample paths generated in Monte Carlo simulation over 20 years using $\mathcal{K}_{\mathrm{P}} = 0.295$ and $\mathcal{K}_{\mathrm{P}} = 0.04$, respectively.

Figure 8: 300 Monte Carlo sample paths of power prices for different mean-reversion speeds.



The dynamics of REC prices has been less studied in the literature. Under renewable portfolio standard (RPS), which is one of the prominent support programs for renewable energy sources, the regulator requires producers, distributors, and consumers to purchase RECs. To forecast the evolution of REC prices, following [209], we use a Jacobi diffusion process to

generate values between zero and one, and obtain REC prices as the product between the output of this stochastic process and an upper bound threshold. The stochastic process is defined by

$$dr_t = (\nu_R - \mathcal{K}_R\, r_t)dt + \sigma_R \sqrt{r_t(1 - r_t)}\, dW_t, \tag{2.30a}$$

$$R_t = r_t\, \overline{R}. \tag{2.30b}$$

In Equation 2.30a, ν_R and $\mathcal{K}_R$ are the mean-reverting parameters, σ_R is the volatility, and W_t is a standard Brownian motion. The process (Equation 2.30a) generates values $r_t \in [0, 1]$, which are then scaled in the second Equation 2.30b by the threshold value $\overline{R}$ to ensure R_t belongs to the interval $[0, \overline{R}]$.

We chose an upper bound of $\overline{R} = 60$ USD/MWh that is representative of wind RECs [75]. We then estimated the parameters of the REC price model using monthly data for New Jersey REC prices between May 2015 and December 2017, and an adaptation of the maximum likelihood estimation method for Jacobi diffusion processes from [87]. We obtained the parameter estimates $\mathcal{K}_R = 0.448$, $\nu_R = 0.066$, and $\sigma_R = 0.109$ (p-values were below 0.001). We set $R_0 = 10$ USD/MWh, which is representative of the average REC price in our time series.

The electricity demand of a company is uncertain due to various factors including technology change, company expansion programs, energy efficiency programs, and environmental conditions. We model power demand uncertainty using a geometric Brownian motion, which is a common choice in the procurement literature to describe demand uncertainty [17, 107, 164]. The process is defined by

$$dD_t = \mu_D D_t\, dt + \sigma_D D_t\, dW_t, \tag{2.31}$$

where μ_D, σ_D, and W_t represent drift, volatility, and a standard Brownian motion, respectively.

To estimate the parameters of Equation 2.31, we use as reference the approximate power consumption of a Google data center in the United States. Considering the total Google annual power consumption [85] and the number of its data centers, we estimate the consumption of a facility with two data centers as $600,000$ MWh/year, and use this value as D_0. We assume $\mu_D = 0$, that is zero demand drift because of two opposing factors: (i) the increasing size and demand for such centers, which would suggest a positive drift; (ii) improving technology and energy efficiency initiatives implying a negative drift. We chose $\sigma_D = 0.05$ based on [164].

The calibrated models allow us to generate sample paths of the uncertainty in Monte Carlo simulation, which are needed to estimate the value of the procurement policies and dual bounds.

To construct the instance set S3, we changed the calibrated value for ν_P from 0 to -0.014 and -0.0342 to obtain long-term mean power prices equal to 30 and 20 USD/MWh, respectively. Similarly, to construct the instance set S4, we changed the calibrated value for ν_R from 0.066 to 0.035 and 0.14 to obtain long-term mean REC prices equal to 5 and 20 USD/MWh, respectively.

CHAPTER 3

INTERPRETABLE USER MODELS VIA DECISION-RULE GAUSSIAN PROCESSES AND STORAGE OVERBOOKING APPLICATION

(Joint work with Selvaprabu Nadarajah and Theja Tulabandhula)

3.1 Introduction

Models of user behavior are critical in many decision making problems and can be viewed as decision rules that transform state information (in set $\mathcal{S}$) available to the user to actions (in set $\mathcal{A}$). Formally, a user model is a function $f : \mathcal{S} \mapsto \mathcal{A}$. Gaussian processes (GPs) employed to learn functions on the action/target space (henceforth target GPs or TGPs for short) can thus be used to place a prior on user models and identify a posterior distribution over them supported by data in conjunction with approximate Bayesian inference techniques [14, 26].

TGPs for user modeling would assume that user actions at a given set of finite states follow a multivariate Gaussian. To capture non-Gaussian action distributions, one could apply GPs to learn functions in a transformed space that is not the target. Examples include warped and chained GPs proposed in [173] and [159], respectively. Extending this literature, we study the application of GPs in a transformed space defined by decision rules. Such rules are known in several applications and depend on functions themselves. Specifically, a user model based on a decision rule takes the form $g : \Pi_k \mathcal{P}_k \times \mathcal{S} \to \mathcal{A}$, where the arguments are obtained using functions $h_k : \mathcal{S} \mapsto \mathcal{P}_k$, $k = \{1, \ldots, K\}$ that map from $\mathcal{S}$ to transformed spaces $\mathcal{P}_k$, possibly different

from the target space $\mathcal{A}$. Each such function has immediate interpretability to a practitioner, and we model them using GPs. We refer to such a user model $\{g, h_1, ..., h_k\}$ as a decision-rule GP (DRGP).

To make the notion of DRGPs concrete in this short article, we focus on the problem faced by a firm providing services to store ethanol – a real application that motivated this work. Suppose capacity (in gallons) is sold via annual contracts to N users. The contract of user n specifies the maximum amount of ethanol that can be stored, denoted by C_n. User behavior corresponds to the injection of ethanol and the withdrawal of previously injected ethanol, which can be modeled as a time series. The inventory $I_{n,t}$ in storage associated with user n at time t is the net of past injections and withdrawals. A TGP approach would employ a GP to determine the next-period storage inventory level function $I_{n,t+1}$ directly. In contrast, we propose a DRGP that leverages a well-known decision rule based on injection and withdrawal threshold functions [49, 165]. These threshold functions are learned as GPs instead of the (relatively less interpretable) inventory function.

We focus on the following research questions in the context of the ethanol storage application: (Q1) Can existing exact and approximate Bayesian inference techniques be used for inference with DRGP? and (Q2) How does DRGP perform relative to TGP?

We answer these questions by executing numerical experiments based on real data of aggregated ethanol storage injection and withdrawals. For Q1, we show that sparse variational inference [93, 179], which can be applied to TGP on our data set, can also be used with DRGP, albeit heuristically, which is encouraging from an implementation standpoint. For Q2, we find

that DRGP implemented in this manner leads to lesser out-of-sample error than TGP on most of our datasets, in addition to being more interpretable to practitioners. This preliminary finding is promising and suggests that applying GPs in the interpretable space of the decision rule threshold functions has potential value, which adds to the growing literature on interpretable machine learning and optimization [21, 113]. In addition, the improvements we report are based on the heuristic use of sparse variational inference with DRGPs, which bodes well for additional potential improvements from the development of new inference techniques targeting DRGPs. Finally, several applications in energy, health care, and transportation, among other domains, have known interpratable decision rules, which can be leveraged in the DRGP framework proposed here.

3.2 Related work

[173] show that modeling data using a warped GP, which is a non-linear transformation (aka warping) of a GP, can enhance predictive performance. Inference using a warped GP can be performed in closed-form provided the warping function satisfies certain properties, such as being invertible. [110] consider the case where the warping function is not fixed a priori. DRGPs differ from warped GPs as they are based on a potentially non-invertible transformation of multiple GPs.

Chained GPs by [159] extend warped GPs by considering a likelihood function that factorizes across the data and is a general nonlinear transformation of multiple latent functions, each modeled as a GP. Exact inference of chained GPs is not tractable in general and thus approxi-

mate inference techniques are used instead. See [111,188]; and [130] for alternative approaches to handle multiple GPs.

Recent work has focused on finding a balance between the modeling generality (restrictiveness) of chained (warped) GPs and its associated challenging (straightforward) inference procedure. For example, [180] extend a warped GP using a composition of simple functions and retain closed form inference. DRGP is similar to a chained GP because its underlying decision rule is a nonlinear transformation $g(\cdot)$ of multiple GPs that model functions $h_1, \ldots, h_K$. However, unlike a chained GP, each function h_k is interpretable and not necessarily latent, which simplifies inference (see §3.3 for details). For instance, in our energy storage application (where $K = 2$), the functions h_1 and h_2 correspond to injection and withdrawal threshold functions, respectively, and are fully or partially observable.

3.3 DRGPs for energy storage

For each user $n \in N$, the most basic inventory update model capturing temporal dependencies can be written as: $I_{n,t+1} = f_n(I_{n,t}, X_{n,t}) + \epsilon_{n,t}$, where f_n is the user specific transition function, $X_{n,t}$ is an exogenous variable with information such as commodity price at time t and other observable user characteristics (e.g., contract size C_n), and $\epsilon_{n,t}$ is an i.i.d. zero mean Gaussian noise variable. We assume that the exogenous state evolves in a Markovian manner. Given sufficient historical inventory usage data for each user, we can infer a posterior on f_n for each user n separately (this is TGP).

While TGP can capture rich user behavior patterns, it is relatively less interpretable because the relationship between the previous inventory level (and other inputs) and the next

inventory level can turn out be highly nonlinear, and using the corresponding posterior belief in downstream overbooking decisions may become cumbersome. To alleviate this, we enhance the interpretibility by incorporating findings from prior literature [49]. In particular, it is known that a user (e.g., a merchant operator) makes injection-withdrawal decisions using a two threshold decision rule structure (also called a double base-stock policy) under reasonable assumptions on the stochasticity of the exogenous variable $X_{n,t}$:

$$I_{n,t+1} - I_{n,t} = \begin{cases} \min\{f^a_{n,t}(I_{n,t}, X_{n,t}) - I_{n,t}, G\} & \text{if } I_{n,t} \leqslant f^a_{n,t}(I_{n,t}, X_{n,t}) \\ 0 & \text{if } f^b_{n,t}(I_{n,t}, X_{n,t}) > I_{n,t} > f^a_{n,t}(I_{n,t}, X_{n,t}) \\ -\min\{I_{n,t} - f^b_{n,t}(I_{n,t}, X_{n,t}), G\} & \text{if } f^b_{n,t}(I_{n,t}, X_{n,t}) \leqslant I_{n,t}, \end{cases}$$

where f^a, f^b are two threshold functions and G is a known operational parameter. Because this two-threshold structure for user behavior is interpretable (user injects if below a given threshold, withdraws if above another threshold, and holds still in between), we use this to define DRGP as follows:

$$I_{n,t+1} = g(f^1_n, f^2_n, X_{n,t}) = \begin{cases} f^1_n(I_{n,t}, X_{n,t}) & \text{if } I_{n,t} \leqslant f^1_n(I_{n,t}, X_{n,t}) \\ I_{n,t} & \text{if } f^2_n(I_{n,t}, X_{n,t}) > I_{n,t} > f^1_n(I_{n,t}, X_{n,t}) \\ f^2_n(I_{n,t}, X_{n,t}) & \text{if } f^2_n(I_{n,t}, X_{n,t}) \leqslant I_{n,t}, \end{cases}$$

where GP beliefs are placed on the threshold functions f^1 and f^2 (with noise terms associated with each function suppressed to ease notation). Note that this composition of two functions f^1 and f^2 is non-invertible.

3.4 Computational experiments

3.4.1 Data

We use *aggregate* inventory level data (~ 100 observations over 2 years) provided by a US ethanol storage operator. The aggregate values are log-transformed and split into separate inventory levels for four users based on three different heuristics to simulate different types of injection-withdrawal behavior (see Appendix 3.6.1). As a result, we obtain three datasets with low, medium, and high variability of injection and withdrawal patterns. We also vary the number of data points across all users, T, between 200 and 400.

These data sets also include information about the exogenous state vector $X_{n,t}$ that includes: (i) the lease capacity of each user; (ii) the spot and prompt-week futures prices for ethanol; and (iii) the prompt-week futures prices for corn and natural gas. We obtain price data from Bloomberg.

3.4.2 Inference

At any time step, a user may inject, withdraw, or do nothing. When a user injects, the inventory level $X_{n,t+1}$ reached as a result of this injection is f_1 at $X_{n,t}$, and as a result, this threshold value is observed but the withdrawal threshold is not. Similarly, if there is a withdrawal action, f_2 at $X_{n,t}$ is observed while f_1 is not. In other words, f_1 and f_2 are partially observable over time. To avoid handling partial observability during inference, we partition the dataset based on when users inject and withdraw and learn the functions f_1 and f_2, respectively, on the resulting subsets. When computing posteriors in this manner, the ordering of f_1 and f_2 may not satisfy the condition $f_1 \leqslant f_2$ that is implicitly assumed in the DRGP model. To overcome this

issue, we train a classifier to first predict if a user's decision is either injection or withdrawal and then employ the corresponding threshold to determine the next stage inventory level.

We compute posteriors on f_1 and f_2 using sparse (GP) variational inference [179] with 10 inducing points and an Automatic Relevance Determination kernel (note that while one can also use exact GP regression here as an alternative, we chose the former for future scalability). We use a gradient boosting decision tree based classifier.

Both TGP and DRGP can be combined with transfer learning by assuming a common component across users and a user specific latent variable. We also consider such models and label them TGP-TL and DRGP-TL. Details of these models and their accompanying inference procedures can be found in Appendix 3.6.2.

3.4.3 Results

In the first experiment, we answer the question (Q2) laid out in the Introduction, which seeks to relate the empirical performance of DRGP when compared to TGP. In order to do so, we perform a training-validation partition of each dataset based on a $70\% - 30\%$ split. The training data is then used to obtain the posteriors, for instance on f_n^1 and f_n^2 in the case of DRGP, for each user $n = 1, ..., 4$. Subsequently these posteriors are used to predict the inventory levels in the validation data. The mean and standard deviation of the root mean squared errors (RMSEs) for TGP and DRGP are displayed in Figure 9 for two values of dataset size T. When T equals 200, the RMSE of DRGP is smaller than TGP across all datasets. As the number of data points T is increased to 400, this trend continues to be true for datasets 1 and 3 but

is reversed for dataset 2. Overall, we can conclude that DRGP obtains a lower RMSE than TGP in most cases, while also buying us interpretability.

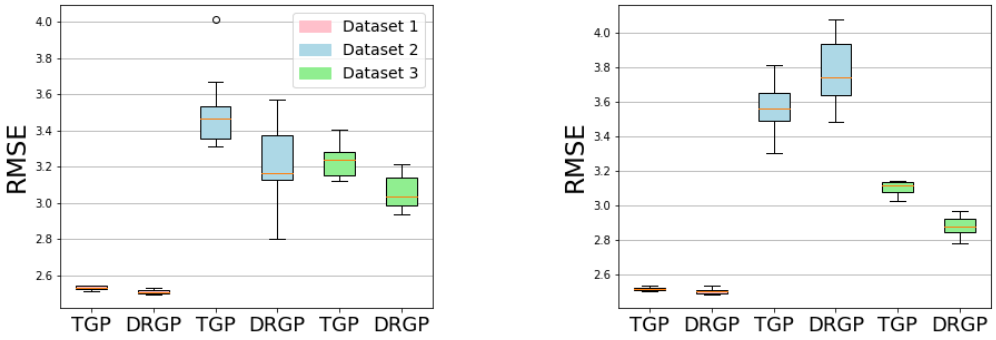


Figure 9: Out of sample RMSE for TGP and DRGP with T = 200 (left) and T = 400 (right). In each panel, there are 6 box plots corresponding to three pairs of datasets in order.

In the second experiment, we investigate the value of transfer learning (where users share common priors). Figure 10(a-c) compares the models with and without transfer learning (T = 400). We observe that incorporating transfer learning produces mixed results, suggesting that these datasets may lack a common user behavioral pattern that can be exploited. Further, in Figure 11 (a-d), we illustrate the quality of one-step predictions of the transfer learning models (for all users in Dataset 3, with T = 400) as a function of one of the exogenous variables (spot price) in the validation data. We observe that DRGP-TL can predict the out-of-sample log-inventory levels with higher accuracy and low uncertainty when compared to TGP-TL.

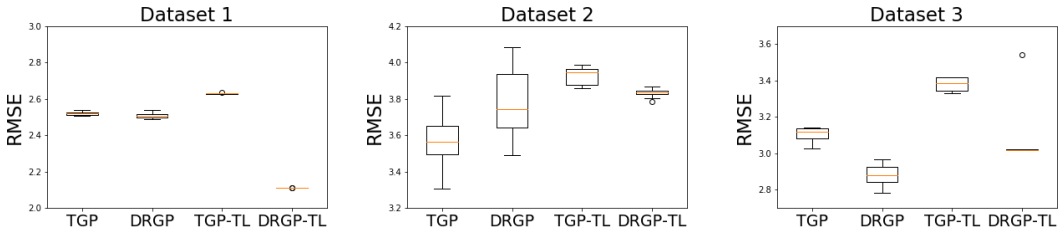


Figure 10: Predictive performance with & without transfer learning across the three datasets.

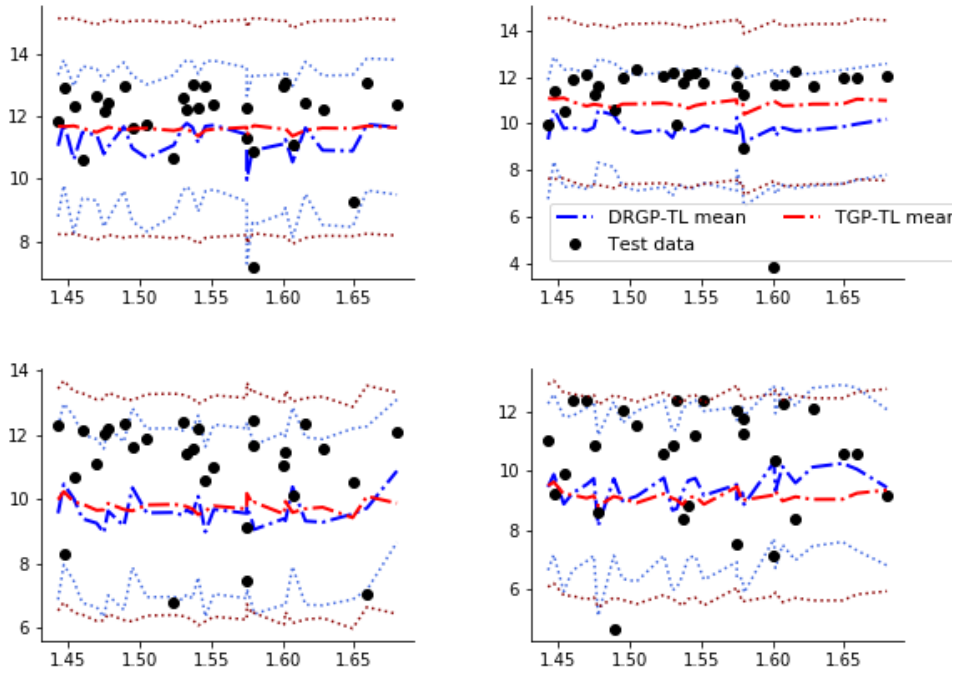


Figure 11: One step prediction of log(inventory) level for dataset 3. The means are shown in solid, and the standard deviations around them are shown using dotted curves.

3.5 Conclusion and future research

Our study of DRGPs shows that there is promise in leveraging decision rules to define non-linear transformations of GPs for user modeling in the ethanol storage application. Extending this investigation to other real-world applications and developing inference procedures tailored

to DRGP would be valuable. Other research directions being explored include: (i) the interplay between the structure of decision rules in a class of applications, their interpretability, and how this can be leveraged within inference procedures for DRGPs; and (ii) robust inference techniques for DRGP, where parameters are computed by optimizing a metric other than the (exact/approximate) likelihood function.

3.6 Appendix

3.6.1 Datasets

The ethanol application dataset contains the daily aggregate inventory level of a storage tank in the US, and the daily price of ethanol over a period of two years. We consider weekly inventory levels to model the behavior of users, as suggested by practitioners. There were 39 companies signed up in the system, with various contract sizes. Assuming that users cluster into groups that have similar injection and withdrawal patterns, we created four users (essentially user types/groups) by assigned these companies to each group based on their contract size. We break down the aggregate inventory levels to four user levels based on three heuristics. These three heuristics are designed to test the performance of the four approaches we have for user modeling; and they capture low, medium, and high variance of injection-withdrawal patterns of users.

The first dataset is created by assigning fractions of the aggregate inventory to each user proportional to their contracted capacity, and simulates a system where the users have low variability. In the second dataset, we simulate a setting where users have medium variability when interacting with the system. This is captured by ensuring that the users do not change

their inventory levels with probability 0.5, and change their inventory levels randomly between 0 and their rented capacity, again with probability 0.5. Finally, to simulate a system where users interact with the system with high volatility, we make the users change their inventory level randomly from 0 to their rented capacity in every period, such that the aggregate of these individual inventory levels is equal to the aggregate inventory level.

3.6.2 Approximations for Bayesian inference

Sparse GP for Scalability: For TGP and DRGP, we rely on variational sparse Gaussian process based inference procedure [93,179]. Inducing point methods involve introducing $M \ll T$ inducing points at locations $\mathbf{Z} = \{z_i\}_{i=1}^{M}$ with corresponding function values given by $\mathbf{u}_i = f(z_i)$ such that:

$$p(\mathbf{f}|\{I_{n,t}, X_{n,t}\}_{t=1}^{T}, \mathbf{Z}) = \int p(\mathbf{f}|\{I_{n,t}, X_{n,t}\}_{t=1}^{T}, \mathbf{u}) p(\mathbf{u}|\mathbf{Z}) d\mathbf{u},$$

where $\mathbf{f}$ is the vector of function evaluations at the T observation points. Using this approach, we are able to approximate the posterior GP with a variational distribution that only depends on the inducing points by obtaining a lower bound on the marginal likelihood.

Transfer Learning: TGP-TL modifies TGP by assuming that user specific latent variables and a common target function together drive the inventory updates of all users [54, 55, 191]. That is, $I_{n,t+1} = f(I_{n,t}, X_{n,t}, \gamma_n) + \epsilon_{n,t}$, where f is a common target function across users that maps the triple $(I_{n,t}, X_{n,t}, \gamma_n)$ to the next inventory level $I_{n,t+1}$, and γ_n is a user specific latent variable. We can jointly infer a posterior belief on f (we fix this to be a GP) and γ_n (which we take to be Gaussian distributed) using LVMOGP [54]. Common temporal patterns of all users can now be captured by f, while idiosyncratic aspects of each user can be captured using γ_n.

Similarly, DRGP can be extended to DRGP-TL, where we have common threshold functions for all users and user-specific latent variables to capture user heterogeneity. The graphical models for TGP-TL and DRGP-TL are illustrated in Figure 12.

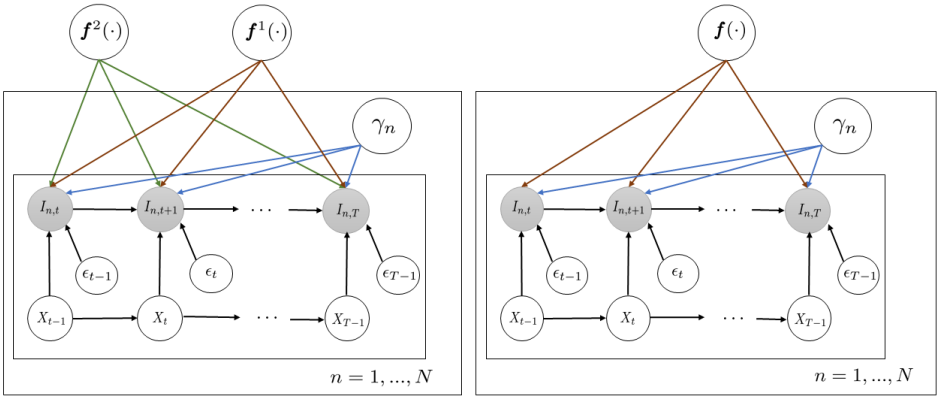


Figure 12: The DRGP-TL and TGP-TL models capturing common user behaviors.

The inference procedure for transfer learning extensions of TGP-TL and DRGP-TL involves handling the joint distribution with respect to the latent variables $\Gamma = \{\gamma_1, .., \gamma_N\}$ and the common function (two functions in DRGP-TL). The following independence assumption is made in the variational approximation for tractability:

$$p(\mathbf{f}, \Gamma) \approx q(\mathbf{f})q(\Gamma),$$

where $q(\mathbf{f})$ is a GP and $q(\Gamma) = \Pi_{n=1}^{N} \mathcal{N}(\gamma_n | \mu_n, \Sigma_n)$. Below, we show the evidence lower bound (ELBO) in TGP-TL:

$$\begin{aligned}
\log p(\{\{I_{n,t}\}_{t=1}^{T}\}_{n=1}^{N}) &\geqslant \mathbb{E}_{p(\Gamma,\mathbf{f})}\left[\log \frac{p(\{\{I_{n,t}\}_{t=1}^{T}\}_{n=1}^{N}, \Gamma, \mathbf{f})}{p(\Gamma, \mathbf{f})}\right] \\
&\geqslant \sum_{n=1}^{N}\sum_{t=1}^{T} \mathbb{E}_{q(\mathbf{f})}\left[\log p(I_{n,t}|\mathbf{f})\right] \\
&\quad - KL(q(\mathbf{u})||p(\mathbf{u})) - KL(q(\Gamma)||p(\Gamma)),
\end{aligned}$$

where we use Figure 12 in the second inequality. Following [96] and other prior works, we maximize the evidence lower bound (ELBO) which provides a lower bound for the log-marginal likelihood of observed data, and jointly optimize with respect to the model hyper-parameters and the variational parameters as suggested in [157].

CHAPTER 4

REPOWERING POWER PLANTS UNDER LIMITED LONG TERM INFORMATION

(Joint work with Andreas Kleiven, Selvaprabu Nadarajah, and Stein-Erik Fleten)

4.1 Introduction and related work

Hydropower plants are the dominant producers of renewable power around the globe which constitutes over fifty percent of global renewable capacity [153]. Unlike intermittent renewable energy sources such as wind and solar, hydropower can sometimes be stored in reservoirs and is a flexible source of renewable power. It can also supply power on short notice. Therefore, it plays a key role in the ongoing clean energy transition, offering flexible and cost-effective renewable power generation. However, most of the economically viable hydropower potential in developed regions, such as Europe, Canada and the United States, is already exploited [98] and were designed decades ago under different climate and market conditions, leading to inefficiency in power production [72]. In the last few years, there has been a extensive focus on the importance of repowering, i.e., modernization, refurbishment and reinvestment in capacity, of existing hydropower plants.

In decentralized power markets, the decision to refurbish and upgrade is a firm level decision, and exhibits features such as long lifespan, partially unknown breakdown risk, irreversibility of the investment decision, and uncertainty with regard to climate and future markets. Moreover,

cashflows from hydropower plants comes primarily from supply to organized markets, and the firm aims to establish operational schedules that maximizes the discounted expected profits over a given planning horizon, subject to a given capacity level and other relevant constraints. This means that investing in capacity needs to account for the value that results from changed operations. We aim to analyse the interaction between the flexibility to adopt the operating policy based on the evolution of operational uncertainty and uncertain near-term market prices along with the flexibility to invest in added capacity.

In classical valuation theory, it is assumed that markets are complete, arbitrage-free and perfectly liquid, and the value of a project is the expected cashflow under the risk-neutral measure, discounted at the risk-free rate [68, 156]. Empirically, futures contracts provide information about the price expectation under the risk-neutral measure, and historical futures contracts can be used to calibrate a stochastic model governing dynamics of risk-adjusted prices to be used in valuation of real options [65]. However, in practice, the planning horizon of long-term investments is often much longer than the longest lasting contract available the market, and liquidity varies. As an example, in Figure 13, the price and trading volume of the futures contract with maturity in 2022 is plotted against trading date in the Nordic electricity market. The contract has been available for 10 years, but in the period 2013 to 2017 it was barely traded. An NPV calculation of an investment project in 2013, using a price of slightly above 40 EUR/MWh as basis for long-term risk-neutral expectation of 2022-prices (and beyond) may therefore be a poor estimate of the NPV of the project. From 2017 onwards, liquidity increased, meaning that the assumptions behind the classical valuation theory is empirically backed for

around 5 years into the future in the case of the Nordic electricity market. We aim to address varying liquidity and varying information regarding exogenous factors affecting the decision to invest in our modeling framework.

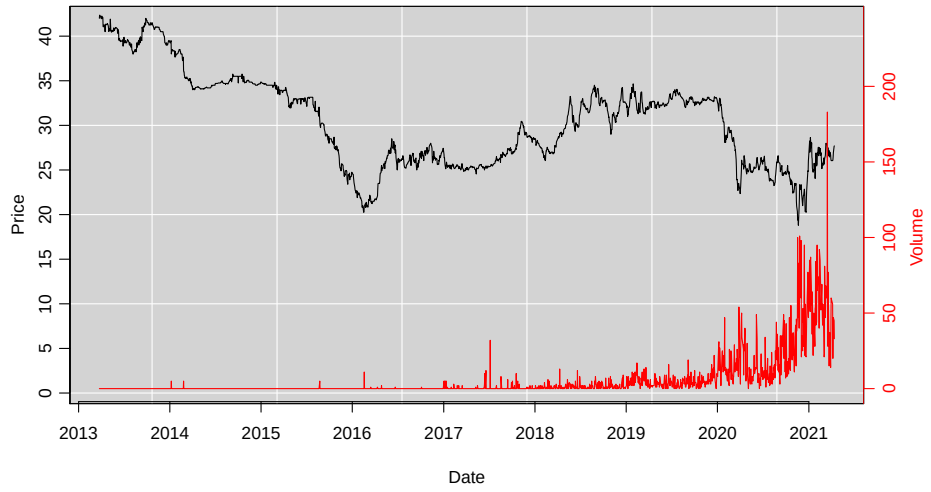


Figure 13: Electricity contract with delivery in 2022 from the Nordic electricity market.

To address heterogenous data availability over time, we propose to leverage useful information in the short- and medium-term to calibrate a stochastic model governing evolution of prices as long as there is sufficiently liquid markets, to be used in real options valuation. To hedge unreliable or non-existing data, we propose to take a worst-case approach. The division of timeline into two regions is illustrated in Figure 14. Our model can be interpreted as a Markov Decision Process (MDP) with a terminal value defined using a robust MDP, that maximizes the sum of the expected revenues from electricity generation in Region A, and worst-

case revenues in Region B. The problem of combined operations and investment planning in the presence of uncertainty leads to a high-deimensional MDP with non-convex actions space which is very challenging to solve.

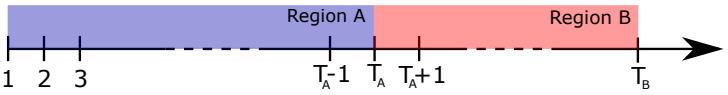


Figure 14: Timeline illustrating reliable data, region A, and data with unreliable information, or no data, region B.

Motivated by reoptimization of decisions on regular basis after new information is revealed, and theoretical results on forecasts horizons which emphasizes that there is a diminishing effect on future data on initial decisions [47, 53], we propose to formulate a MDP, assuming full information about the data generating process as long as there is sufficient information available, which for electricity prices means sufficiently liquid futures markets. In the long-run we switch to a robust framework and do not impose distributional information to avoid model misspecification. This combination leverages valuable short-term information, and it avoids overly conservative policies typically encountered when applying robust optimization. It also avoids long-run model misspecification due to the robust framework. We apply the dual reoptimization heuristic (DRH) in [184] combined with robust value iteration to find a feasible MDP-policy and assess the value of the policy against a dual upper bound [40]. We provide insights on how to construct uncertainty set using stationary distribution and discuss a new approach in building less conservative uncertainty set based on clustering to evaluate the long-term value of the asset.

We conduct numerical experiments based on the data of a real hydropower plant with a planning horizon of 40 years. We calibrate stochastic processes for the evolution of uncertain price quantities using market data from the Nordic electricity market and the practitioner literature. We compare hydropower plant operational policy using robust terminal values against the one that assumes that the long-run behavior of uncertainties can be explained by the short-term information (nominal). We show that the policies computed by DRH are near-optimal in all approaches and investigate the performance of different choices of decision measure choices in DRH for our application.

Our numerical results unveil the relation between the operational policies and patterns in the evolution of uncertainties. We show the presence of seasonal patterns in the operating policies, which is aligned with the policies incorporated in practice. Our results also confirm the theoretical results on forecast horizons and the diminishing effect of future data on initial decisions and highlight the differences between investment policies considering robust and nominal terminal values. Furthermore, we compare the performance of policies under various future scenarios and show that approaches with clustering-based robust terminal values can provide robust performance with competitive expected cashflows and lower volatility among other benchmarks.

4.1.1 Novelty and related work

We contribute to the growing literature on investment in renewable energy, where decisions are to be made at a firm level [30, 33, 78, 184]. Specifically we study the decision to repower hydropower plants, meaning to temporarily shut down the plant to refurbish, which includes increasing capacity if profitable [152]. We aim to analyze the investment decision in the presence

of operation schedules, which require both detailed short-term modelling of uncertain factors, and long planning horizon. This problem has been studied by [9]. However, jointly optimizing operations schedules and investment policies, which we illustrate the added value of in Section 4.2, in the presence of uncertain market prices, operational uncertainty, and breakdown risk, is new to this literature. [196] and [197] have studied investments in hydropower plants incorporating probability of failures, but did not consider operations schedules. Morover, these papers do not consider the change in data availability and its implications on model errors.

Our second contribution is to the literature on decision making in an uncertain dynamic environment where data is assumed to be partially known. The presence of multi-factor evolution of energy prices and operational uncertainty gives rise to an intractable Markov decision process (MDP). MDPs effectively describe sequential decision making in an uncertain dynamic environment, assuming full information of the parametrized model governing exogenous information [56]. The parametrized model governing the evolution of exogenous information is often estimated based on historical data or expert knowledge. However, in cases where available data is limited, the performance of optimal MDP-policy on real data may differ significantly from their performance using simulated environments because of parameter uncertainty and generalization error [119,121]. Papers that study the impact of inaccuracy in estimation on the optimal solutions include [31,151,155], among others. To cope with this, robust optimization (RO) is an alternative framework for decision making in the presence of uncertain data [15,174]. As opposed to stochastic optimization, which relies on the distributional specification of the underlying data, robust optimization requires only the support. The goal is to find solutions that per-

form well on worst-case parameter realizations contained in an a priori defined uncertainty set. A drawback with robust optimization is that it leads to very conservative solutions. Several papers have proposed alternatives to achieve less conservative decisions. For instance, [23] propose to adjust the degree of conservatism in terms of probabilistic bounds on constraint violation. Robust optimization leads to tractable formulations in a multistage setting [11, 59, 81, 137, 200], albeit often overly conservative policies are obtained, as addressed in [60]. Motivated by the same underlying issue as robust optimization addresses, other non-parametric approaches have been proposed to overcome the generalization error induced when committing to a parametrized distribution, often referred to as data-driven optimization [22, 25, 90, 119].

Distributionally robust optimization (DRO) ([160]) is another approach for decision making that assumes limited knowledge about the underlying uncertainties. DRO provides a potential tradeoff between the stochastic and robust frameworks. Parameters are considered as stochastic, but the distribution is not fully known and assumed to belong to an a priori defined ambiguity set which defines the notion of robustness. DRO hedges against ambiguity in the underlying data generating process by taking a worst-case approach, similar to RO, and also exploits partial information about uncertain model parameters. There is an extensive literature on DRO in the single-stage setting [42, 61, 71, 83, 201], and several recent works have studied tractable DRO formulations in the multi-stage setting [24, 50, 206]. We propose address the limitations of both the stochastic and robust frameworks, and proposing a potential trade-off between the frameworks. Moreover, our approach provides an alternative to the multistage distributionally

robust framework where the decision maker has partial information about the data generating process.

Moreover, we add to the literature on rolling horizon based, or reoptimization based methods, for solving problems for decision making under uncertainty, which have extensively been applied in energy applications [108, 184, 205]. We extend the work by [184] and propose a novel reoptimization algorithm combined with robust value iteration to solve the proposed model that can handle non-convex action set. More generally, we add to literature on approximation methods for solving high dimensional MDPs [148]. In energy applications, examples include approximate linear programming [115, 132], least squares Monte Carlo [32, 133, 134], and stochastic dual dynamic programming [82, 144, 170]. Literature on solving MDPs with non-convex action set is scarce. Decomposition based methods with theoretical convergence guarantees have been proposed [4, 145, 210, 214], and a few papers have successfully applied these methods in energy [95, 109, 213].

Finally, only a few other papers have studied the effect of ambiguity and robustness in the underlying data generating process in real world multi-stage setting. An example is [186] who take a distributionally robust approach in valuation of thermal power plants. We extend this literature by analysing the case of reservoir management and capacity investments in hydropower plants.

4.1.2 Paper structure

The paper is organized as follows: In Section 4.2 we present an empirical study for calibration of stochastic process under limited data, using electricity futures and discuss illustrative

examples showing how decisions optimized with respect to a misspecified model may materialize on real data. The mathematical formulation of the hydropower reservoir management and capacity investment problem is formulated in Section 4.3, and the method for addressing limited data and curse of dimensionality in the MDP is presented in Sections 4.4 and 4.5, respectively. Results from numerical experiments are provided in Section 4.6.

4.2 Empirical support and illustrative example

In this section we present the stochastic models governing exogenous information, and assess errors when calibrating the price model using limited data. We then present an illustrative example showing the effect on policies under long-term model misspecification.

4.2.1 Errors in training stochastic models using limited data

It is common in the literature to model the evolution of uncertainties with stochastic processes. We specify the price process as a two-factor model, as in [162]. The two factors corresponds to short-term and long-term price dynamics. Let $\ln S_t = \theta_t + \chi_t + \xi_t$, where

$$\theta_t = \alpha \cos(2\pi t) + \beta \sin(2\pi t) \tag{4.1}$$

$$d\chi_t = -\kappa_\chi \chi_t dt + \sigma_\chi dz_\chi \tag{4.2}$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi dz_\xi, \tag{4.3}$$

where dz_χ and dz_ξ are increments of standard Brownian motions processes with $dz_\chi dz_\xi = \rho_{\chi\xi} dt$. Following [162], a risk-neutral version of the spot-price process can be obtained by

introducing two constant risk-premia parameters, λ_χ and λ_ξ. The risk-neutral dynamics are given by

$$d\chi_t = -\kappa_\chi \left(\chi_t + \frac{\lambda_\chi}{\kappa_\chi} \right) dt + \sigma_\chi dz^*_\chi \tag{4.4}$$

$$d\xi_t = \mu^*_\xi dt + \sigma_\xi dz^*_\xi, \tag{4.5}$$

where, again, $dz^*_\chi dz^*_\xi = \rho_{\chi\xi} dt$, and $\mu^* = (\mu_\xi - \lambda_\xi)$. A discretization scheme for simulating the risk-neutral process is given in Appendix 4.8.7.

Other exogenous factors include one inflow factor and one deterioration factor. Inflow has strong seasonal variations, so we follow [170] and [82] and model inflow as deviations from the historical time t mean, $\mu^{\log}_t$. We let $\ln I_t = \mu^{\log}_t + \bar{\sigma}^{\log}_t \omega_t$, where

$$d\omega_t = -\kappa_\omega \omega_t + \sigma_\omega dz_\omega \tag{4.6}$$

For the evolution of the deterioration rate d_t we assume a gamma process which is commonly used in maintenance and reliability analysis [187]. We let d_t denote the deterioration at time t. The stationary gamma process has independent gamma distributed increments, d_t. In the stationary case we have

$$d_t - d_\tau \sim \mathrm{Ga}(a(t) - a(\tau), b). \tag{4.7}$$

To find the parameters of the price model, it is common to consider each factor as latent and apply Kalman filter and maximum likelihood estimation to obtain state and parameter estimates [162]. See Appendix 4.8.10 for calibration procedure and parameter estimates. As futures prices are expected spot prices under the risk-neutral measure, historical futures contracts can

be used for calibration. Let $\phi = (\kappa_\chi, \mu^*, \lambda_\chi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi})$ denote the vector of parameters in Equation 4.4-Equation 4.5. The expression for the model predicted log futures price is

$$\ln \hat{F}_{t,T}(\chi_t, \xi_t; \phi) = \mathbb{E}^*(\ln S_T)) + \frac{1}{2}\mathrm{Var}^*(\ln S_T) \tag{4.8}$$

$$= e^{-\kappa_\chi T}\chi_t + \xi_t + A(T), \tag{4.9}$$

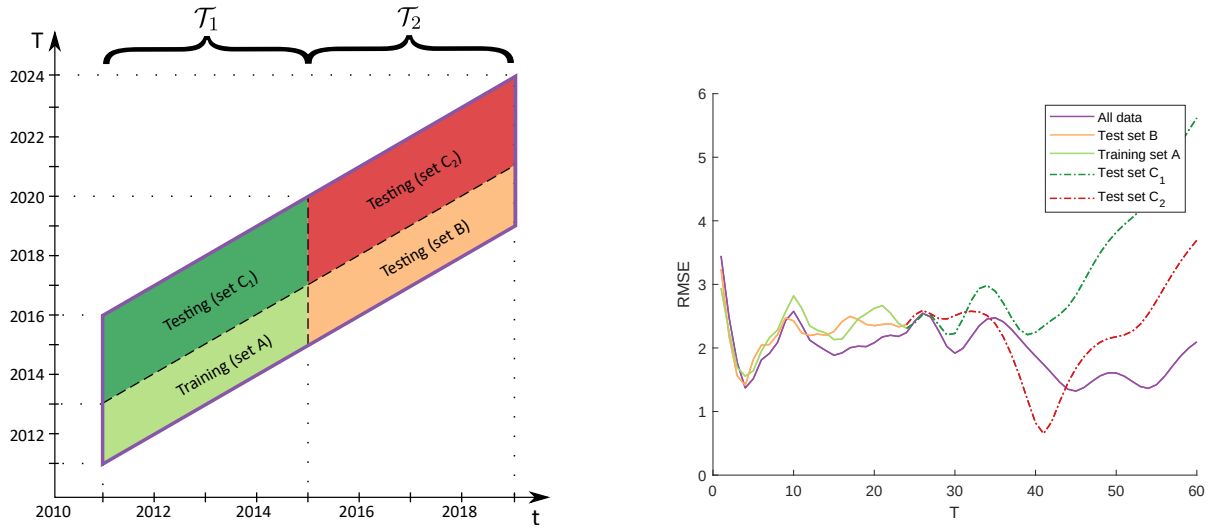
where

$$A(T) = \mu^* T - (1 - e^{\kappa} T)\frac{\lambda_\chi}{\kappa_\chi} + \frac{1}{2}\left((1 - e^{2\kappa_\chi T})\frac{\sigma_\chi^2}{2\kappa_\chi} + \sigma_\xi^2 T + 2(1 - e^{\kappa_\chi T})\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa_\chi}\right) \tag{4.10}$$

We conduct an empirical study for investigating the out-of-sample model performance by dividing the data set in three, as shown in Figure 15a. We use the root mean square error (RMSE) between the model predicted futures price, $\hat{F}_{t,T}(\hat{\chi}_t^{KF}, \hat{\xi}_t^{KF}; \hat{\phi}_A^{MLE})$, using latent states estimates $\hat{\chi}_t^{KF}$ and $\hat{\xi}_t^{KF}$ from applying the Kalman filter, and maximum likelihood estimates $\hat{\phi}_A^{MLE}$ from the training data set, and futures contracts $F_{t,T}$ traded at time t with maturity T. The RMSE expression is

$$\mathrm{RMSE}_T = \sqrt{\frac{1}{N}\sum_{t\in\mathcal{T}_i}\left(\hat{F}_{t,T}(\hat{\chi}_t^{KF}, \hat{\xi}_t^{KF}; \hat{\phi}_A^{MLE}) - F_{t,T}\right)^2}, \tag{4.11}$$

where $i \in \{1, 2\}$. The sets $\mathcal{T}_1$ and $\mathcal{T}_2$ include trading dates with weekly granularity, and are illustrated in Figure 15a. The first data set, *set* A, is the training set, consisting of futures contracts traded in the period 2011-2014, $\mathcal{T}_1$, with time to maturities from 1 to 24 months. We use subscript A in $\hat{\phi}_A^{MLE}$ to emphasize that parameter estimates are kept fixed when evaluating the RMSE for all data sets. The second set, *set B*, is a test set consisting of futures contracts



(a) Data sets.

(b) RMSE

Figure 15: Figure 15a shows the four data sets: The training set, and test sets A, B_1, and B_2. Figure 15b shows the RMSE between the model predicted futures price from (Equation 4.8), using MLE estimates based on data from the training set, and futures data from respective data sets. The purple line is a reference line, showing the RMSE when using all data to estimate model parameters.

traded in the period 2015-2018, $\mathcal{T}_2$, with maturities from 1 to 24 months. The Kalman filter is rerun for $\mathcal{T}_2$ with futures data from 1 to 24 months maturity, to obtain updated state estimates, while parameter estimates from $\mathcal{T}_1$ are kept fixed. Set B is used to evaluate how well the estimated model parameters can be used to predict futures prices with maturities in the same range as the training set. The third set is again split in two, *set* C_1 and *set* C_2. These sets are used to evaluate how well the model predicts long-maturity futures prices that are not used in estimation of parameters or latent states. Test set C_1 consists of futures contracts traded in the period 2011-2014, $\mathcal{T}_1$, with maturities from 25 to 60 months and test set C_2 consists of futures contracts traded in the period from 2015-2018, $\mathcal{T}_2$, with maturities from 25 to 60 months. Terminology in [52] is in-sample data set (set A), traditional out-of-sample set (set B), and extreme out-of-sample sets (set C_1 and C_2).

The RMSE for all data sets is plotted in Figure 15b. We observe that the RMSE of set B is comparable with the RMSE in the training set, set A. This indicates that model parameter estimates are not very sensitive to the trading period of futures contracts. We further observe that the RMSE quickly increases with T in sets C_1 and C_2. This means that the model estimates are very sensitive to which time-to-maturity contracts that are included during estimation and that extrapolating the model beyond the longest term maturity may lead to large errors. For comparison, the RMSE using all data is shown in purple, showing a smaller RMSE for high T.

Lack of sufficient information for estimating the uncertainties in the long-run thus can introduce errors in the model and result in suboptimal decisions. When calibrating a risk-neutral process to futures prices, the long-run behaviour cannot be explained by short-term futures contracts. For instance, [52] show that the long-run behaviour of oil price cannot be explained by short-term future contracts. Motivated by our observation about the errors of extrapolating the stochastic process of power price beyond the longest lasting futures contract as well as theoretical results on forecast horizons [47], we consider a robust scheme to estimate the long-term value of the asset. In practice, it is common to use rolling horizon for planning where the planning problem is to reoptimize as soon as uncertain parameters are revealed. Theoretical results on forecast horizons show that there is a diminishing effect of future data on initial decisions and state that small changes in the far future will not have a big impact on decisions implemented now. This naturally suggests to include statistical information in the first part of the planning horizon, $t \leqslant T_A$ to ensure the quality of short-term decisions, and consider a pure robust approach in the second part of the planning horizon, $t > T_A$ to hedge against

market shifts. Our model combines a Markov decision process with a robust terminal value, as shown in Figure 14. Intuitively, this combination leverages useful medium term information and handles model misspecification in the long run while avoiding overly conservative decisions typically encountered when using robust optimization.

4.2.2 Effect of long-term model misspecification on policies

For evaluation of capacity investments, a planning horizon of decades is needed. The longest maturity contracts available in the Nordic electricity market is 10 years, with low liquidity on longer term contracts, making it challenging to specify a risk-neutral price process over the entire horizon. In this section, we present an example, illustrating how decisions based on a misspecified model of exogenous factors may materialize on real data.

We present two approaches for solving the energy storage operations and investment problem and show that there is value in jointly optimizing operations and investments under the estimated data generating process (E-DGP). We then create a backtesting environment, a true data generating process (T-DGP), to evaluate policies from both solution methods, and we present a third solution method that hedges against a misspecified data generating process. In the end we compare policy performance from all three solution methods when two data generating processes are equally likely of being the true data generating process.

We consider the setting of a hydropower producer with one reservoir who needs to determine the water release policy, investment timing and production capacity choice over a planning period of $\mathcal{T} = \{1, 2, 3, 4, 5\}$ stages. We assume that the maximum reservoir limit is infinite, the initial reservoir volume is 5 MWh, and in every stage it deterministically flows 2.5 units of

water into the reservoir, in units of energy, MWh. The price at which the producer can produce evolves stochastically according to the binomial E-DGP in Figure 16a. We denote the set of price scenarios by $\mathcal{S}$. There is equal probability for up and down states. The price today is \$20. Numerical values of the instance are provided in Table IV. The total number of scenarios is $S = 2^4$, and the current production capacity is 3 MWh. To simplify the setting, decision maker can only choose to upgrade immediately or in stage $t = 3$ at cost 140 USD/MWh. We assume that the plant can always generate, i.e. no downtime associated with upgrading. We denote by $\mathcal{S}_s^t \in \mathcal{S}$ the set of scenarios that are indistinguishable from scenario s at time t, ensuring non-anticipativity of stage $t \in \mathcal{T}$. The problem is formulated in the appendix.

In the first solution approach, the fully integrated (FI) approach, we find the optimal solution by jointly optimizing operations and investments. In the second solution approach, the partially integrated (PI) approach, we heuristically find a feasible solution by optimizing production to all possible investment alternatives. That is, the investment decision is determined based on the combination of timing and capacity upgrade that gives the highest NPV, for an already optimized production schedule.

Policy performances are given in Table V. We observe that the FI approach outperforms the PI approach, as expected. In FI, it is optimal to delay production until stage 3 if prices go up twice, and then invest in a higher capacity. Then more water can be utilized, as apposed to the solution from the PI approach which starts producing in stage 1.

We create a backtesting environment. The true data generating process is given in Figure 16b. We denote the true data generating process by T-DGP. The first four stages is a binomial

TABLE IV: Parameter values for the illustrative example.

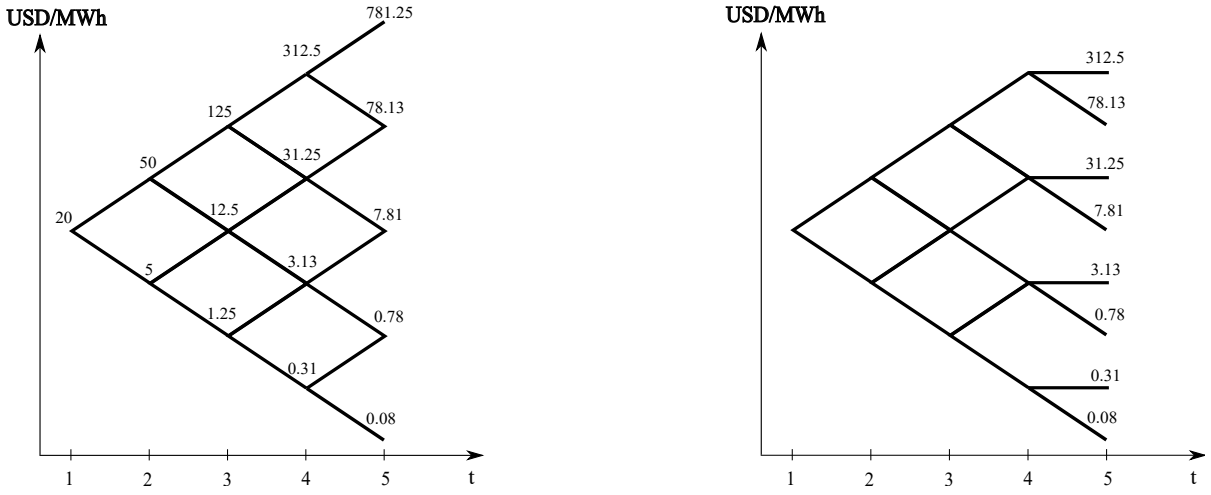
Name	Value	Unit	Description
p_1	20	USD/MWh	Initial price
l_1	5	MWh	Initial inventory
q_1	3	MWh	Initial production capacity
C_U	140	USD/MWh	Unit upgrade cost
i_t	2.5 $t \in \mathcal{T}$	MWh	Inflow at time t
δ	0.94	-	Discount factor
u	5/2	-	Price up factor, 50 % probability
d	1/4	-	Price down factor, 50 % probability

TABLE V: Overview of the optimal investment decision, optimal value and optimality gap for the FI and PI solution approach. The set $\mathcal{S}^3_{++} \in \mathcal{S}$ is the set of scenarios, that all have the property that they start with two up moves.

Solution approach	Investment decisions	Policy value	Optimality gap
FI	$\begin{cases} 2, & s \in \mathcal{S}^3_{++},\ t = 3 \\ 0, & \text{else} \end{cases}$	\$568	0%
PI	$0,\ s \in \mathcal{S},\ t \in \mathcal{T}$	\$540	5.3%

tree, equivalent to the E-DGP first four stages, but the transition to the last stage is different. We assume that in the last stage the up factor is 1, i.e. the price stays the same in the good scenario and is scaled by 0.25 in the bad scenario. The true data generating process is unknown to the decision maker.

We now evaluate the PI and FI policies in our backtesting environment, T-DGP, and performances are given in Table VI. From Table VI we observe that the policy from the PI approach, trained on E-DGP, gives value \$447 when evaluated on the T-DGP, and the policy from the FI approach, trained on E-DGP, gives \$424 when evaluated on the T-DGP. For this particular case, PI performs optimally on the T-DGP, while FI has optimality gap 5.3%. Because the



(a) Estimated data generating process (E-DGP).

(b) True data generating process (T-DGP).

Figure 16: Data generating processes. Up and down states happens each with probability 1/2. E-DGP and T-DGP are equivalent in the four first periods.

decision maker planned according to a misspecified DGP, i.e. E-DGP, the cost of upgrading 2 MWh can not be recovered from future revenues from production when evaluated on T-DGP. This means that heursistic planning can be superior to joint optimization if the parameters of the stochastic model governing the dynamics of exogenous information, or the stochastic model itself, is misspecified. This might happen if one does not have sufficient information about long-term data and extrapolate the stochastic model beyond the last data point. We illustrate this further in Section 4.4.

To address limited data, we present a hybrid stochastic and robust approach that leverages useful information in the short term and handles misspecification of the DGP in the long run. The idea is to plan according to the E-DGP in the first four stages and then assume that the price can take any value in the interval between 0.25 and 2.5 times the price in stage 4, as illustrated in Figure 17. Decisions are then optimized based on knowledge about the DGP in

the first four stages and the worst case realization in stage five. We call this solution the robust fully integrated (RFI) policy, and the RFI policy performance under the E-DGP and the T-DGP is given in the last column in Table VI. The optimality gap is 8.4% under E-DGP and 0.4 % under T-DGP.

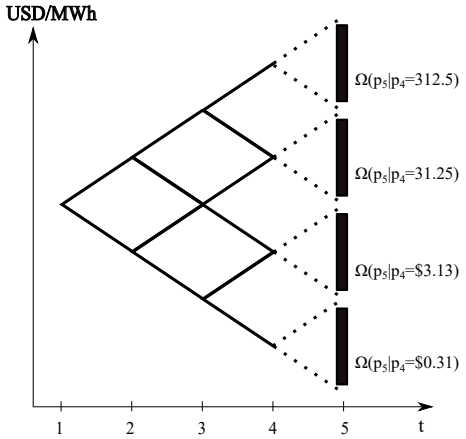


Figure 17: In the first four stages, we assume the E-DGP. In the fifth stage we define an uncertainty set $\Omega(p_5|p_4) = [\frac{5}{2}p_4, \frac{1}{4}p_4]$.

TABLE VI: Policy performance for each solution approach. The data generating process that the policy in FI and PI is trained on is given in parentheses. The data generating process used for evaluation is given in the leftmost column.

		Solution approach		
Data		FI (E-DGP)	PI (E-DGP)	RFI
E-DGP	Value	\$568	\$540	\$524
	Optimality gap	0%	5.3%	8.4 %
T-DGP	Value	\$424	\$447	\$445
	Optimality gap	5.1%	0%	0.4%

Finally, we perform an experiment where we assume that E-DGP and T-DGP are equally likely of being the process generating the data. In Table VII the expected profits and standard deviation of profits. We note that the robust approach leads to minimum profit variance among the three solution approaches, highlighting the effectiveness of robust optimization to hedge against the underlying dynamics of prices when data is scarce. However, a critique of robust optimization is the degree of conservatism of solution, and the boundaries of the uncertainty set where prices can belong, needs to be defined a priori, which is not straightforward. One alternative is to set the boundaries based on percentiles of the distribution of prices.

TABLE VII: Expectation and standard deviation of policies, assuming E-DGP and T-DGP are equally likely of being the process generating the data. The data generating process that the policy in FI and PI is subject to is given in parentheses

	Solution approach		
	FI (E-DGP)	PI (E-DGP)	RFI
Profit expectation	\$496	\$493.5	\$484.5
Profit standard deviation	\$101.8	\$65.8	\$55.9

There are two main takeaways from this illustrative example. Firstly, if the true data generating process is unknown and the estimated data generating process is misspecified, a simpler heuristic approach, PI, might outperform the optimal E-DGP solution, obtained by the FI approach, when evaluated on T-DGP. Secondly, if there is limited data, a robust approach can hedge against a misspecified DGP.

4.3 An MDP of hydropower capacity investment under limited long-term data

In this section, we formulate the hydropower reservoir management and capacity investment problem as an MDP. The set of all actions are indexed by k in the set $\mathcal{K}$. The set $\mathcal{K}$ includes {Generation (G), Upgrade (U), Refurbishment (R), Spillage (S)}. The condition of the plant is represented by a value in the interval $[0, 1]$ where 0 represents the condition of the plant after refurbishment and 1 is the threshold of failure. The decision vector at stage t is shown by $\mathbf{x}_t \in \mathcal{X}_t(\mathbf{x}_t := \{x_t^k, k \in \mathcal{K}\})$. While the action elements, x_t^k, corresponding to generation, spill and upgrade are continuous-valued the action element corresponding to the refurbishment is binary indicating refurbishment or no refurbishment.

Refurbishment and upgrading have some implication on the set in the following periods i.e., if the plant chooses the refurbishment or upgrades the system, the plant needs to be down for τ^R periods and it cannot generate during these periods. Therefore, choosing some actions will limit the possible actions of the future periods. Failure will lead to a downtime of τ^F periods. Upgrade requires refurbishment.

The information necessary for making a decision at stage t is described in the MDP state, which contains two components. The first component is a vector $\mathbf{v}_t := (l_t, c_t, q_t)$ containing the reservoir volume l_t, the condition of the plant c_t and the capacity of the plant q_t at time t. This component is affected by the decisions of the producer. For instance, the reservoir volume might decrease if the plant generates, and might increase otherwise. The increase is due to the non-negative exogenous inflow. However, the reservoir level cannot be more or less than L^{max} and L^{min}, respectively. Similarly, the condition of the plant can become worse with

time, due to exogenous deterioration. The plant condition can be improved by refurbishment. The second component is a vector $\mathbf{w}_t := (s_t, i_t, d_t)$ containing exogenous information, price, inflow, and deterioration rate, respectively. The MDP state at stage t is represented by the pair $(\mathbf{v}_t, \mathbf{w}_t) \in \mathcal{V}_t \times \mathcal{W}_t$. Below, we discuss components of the MDP.

Transition functions: Executing the decision $\mathbf{x}_t \in \mathcal{X}_t$ at stage t and state $(\mathbf{v}_t, \mathbf{w}_t) \in (\mathcal{V}_t, \mathcal{W}_t)$ results in an update of the reservoir volume l_t, plant condition c_t, and production capacity q_t in $\mathcal{T}_t := \{t, \cdots, T_B\}$

$$l_{t+1} = f(l_t, x_t) = l_t - x_t^G + i_t - x_t^S \qquad t \in \mathcal{T}_t \backslash \{T_B\} \tag{4.12a}$$

$$c_{t+1} = g(c_t, x_t) = c_t + d_t - c_t x_t^R \qquad t \in \mathcal{T}_t \backslash \{T_A + 1, \cdots, T_B\} \tag{4.12b}$$

$$q_{t+1} = h(q_t, x_t) = q_t + x_t^U \qquad t \in \mathcal{T}_t \backslash \{T_A + 1, \cdots, T_B\} \tag{4.12c}$$

The reservoir volume at time $t+1$, l_{t+1}, needs to be equal to the reservoir volume at time t plus the inflow during time t, i_t, minus the discharge and spill at time t, x_t^G and x_t^S. Furthermore, the condition of the plant at time $t+1$, c_{t+1}, deteriorates by d_t and is reset to state 0 if the plant is refurbished. The exogenous factor d_t is the rate of deterioration. The capacity of the plant at time $t+1$, q_{t+1}, can either be upgraded or stay the same.

Rewards: At each time step the reward is the based on the action at time t.

$$r(\mathbf{w}_t, \mathbf{v}_t, \mathbf{x}_t) = x_t^G p_t \eta - C_R x_t^R - C_U x_t^U - C_F \mathbf{1}_{\{c_t = 1\}},$$

where p_t is the price at time t, C_R is the cost of refurbishment, C_U is the cost of upgrading, C_F is the cost of failure, and η is a constant, representing water to electricity conversion rate.

Optimal policy: A stage j policy π_j is a collection of stage-dependent decision rules $\{X_t^\pi, t \in \mathcal{T}_j\}$ each mapping states to actions. A decision rule X_j^π in stage j is feasible if it associates with each state $(\mathbf{v}_j, \mathbf{w}_j) \in \mathcal{V}_j \times \mathcal{W}_j$ an action $\mathbf{x}_j(\mathbf{v}_j, \mathbf{w}_j)$ that belongs to the action set $\mathcal{X}_j$. We denote by Π_j the set of all feasible stage j policies. Given an initial stage $(\mathbf{v}_j, \mathbf{w}_j)$ in stage j, an optimal policy is found by solving

$$V_j(\mathbf{w}_j, \mathbf{v}_j) = \max_{\pi_j \in \Pi_j} \left\{ \mathbb{E}\left[\sum_{t=j}^{T_A-1} \delta^{t-j} r(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t, X_t^{\pi_j}(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t)) + \delta^{T_A} R^*(\mathbf{v}_{T_A}, \mathbf{w}_{T_A}) \Big| \mathbf{v}_j, \mathbf{w}_j \right] \right\},$$

$$R^*(\mathbf{v}_{T_A}, \mathbf{w}_{T_A}) := \mathbb{E}\left[\sum_{t=T_A}^{T_B} \delta^{t-T_A} r'(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t, X_t^{\pi_j}(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t)) \Big| \mathbf{v}_{T_A}, \mathbf{w}_{T_A} \right] \tag{4.13}$$

where $V_j(\mathbf{v}_j, \mathbf{w}_j)$ is the MDP value function at stage j and state $(\mathbf{v}_j, \mathbf{w}_j)$, $\mathbb{E}$ is the expectation with respect to future exogenous state, and $\mathbf{v}_t^{\pi_j}$ is the endogenous state reached at stage t following policy π_j with initial state $(\mathbf{v}_j, \mathbf{w}_j)$. where $R^*(\mathbf{v}_{T_A}, \mathbf{w}_{T_A})$ is the terminal value of the asset at time T_A. The parameter T_A sets the length of horizon A and $r'(.)$ is the reward function at horizon B:

$$r'(\mathbf{w}_t, \mathbf{v}_t, \mathbf{x}_t) = x_t^G p_t \eta \mathbf{1}_{\{c_t < 1\}} \tag{4.14}$$

The reward function at horizon B includes only the generation decisions and capture the potential of the asset in generating cashflow in this period, i.e., the terminal value of the asset.

Proposition 4.3.1. *For a given* $(j, w_j) \in (\{T_A, \cdots, T_B\} \times \mathcal{W}_j)$, $V'_j(\cdot, w_j)$ *is concave on* $\mathbb{R}^3_+$, *where* $V'_j(v_j, w_j) = \max_{x_j \in \mathcal{X}(v_j)} r'_j(x_j, w_j) + W'_j(v_j + \Delta v_j(x_j), w_j)$, *and* $W'_j(\cdot, w_j) = \delta \mathbb{E}(V'_{j+1}(\cdot, w_{j+1})|w_j)$, *and* $V'_j(\cdot, w_j)$, *and* $r'_j(x_j, v_j)$ *is defined in Equation 4.14.*

The MDP introduced at Equation 4.13 suffers from curse of dimensionality and as discussed in section 4.2, modeling the evolution of uncertainties in the long-term using historical data can lead to error and suboptimal decisions. In the following sections, we aim to provide solutions to overcome these issues.

4.4 Addressing limited data

In this section, we focus on estimating the value of an asset in horizon B given the limited information about the evolution of uncertainties. To alleviate problems due to the high dimensionality of the state space, we approximate the evolution of the uncertainties in horizon B with a stationary distribution.

$$\hat{R}^N(\mathbf{v}_{T_A}, \mathbf{w}_{T_A}) = \mathbb{E}\left[\sum_{t=T_A}^{T_B} \delta^{t-T_A} r'\left(\mathbf{v}_t, \mathbf{w'}_t, X_t(\mathbf{v}_t, \mathbf{w'}_t)\right) \middle| \mathbf{v}_{T_A}, \mathbf{w}_{T_A}\right], \tag{4.15}$$

where $\mathbf{w'}_t$ follows a stationary distribution and $\hat{R}^N$ is the estimation of the terminal value. We call this approach *nominal* and discuss details of this approximation in section 4.4.1.

Due to the limited long-term information about the stochastic environment, solution approaches in stochastic dynamic programming may underperform on real data due to the parameter uncertainty [121]. Robust MDP [100, 138] is a common approach to reduce the performance deviation of policies to model misspecification. In this approach, it is assumed that uncertain variables can be any member of an uncertainty set and solutions are chosen based on their performance on the worst case scenario.

To find the value of the terminal value, we model robust MDP. In this framework, it is assumed that the transition probability lie in an unknown uncertainty set and we seek for the

optimal value function $\hat{R}^{RO}(\mathbf{v}_{T_A}, \mathbf{w}_{T_A})$ that considers the worst case scenario. The robust MDP satisfies a Bellman recursion of the form [100, 138]

$$\hat{R}^{RO}(\mathbf{v}_{T_A}, \mathbf{w}_{T_A}) = \min_{w' \in \mathcal{U}} \mathbb{E}\left[\sum_{t=T_A}^{T_B} \delta^{t-T_A} r'\left(\mathbf{v}_t, \mathbf{w'}_t, X_t(\mathbf{v}_t, \mathbf{w'}_t)\right) \middle| \mathbf{v}_{T_A}, \mathbf{w}_{T_A} \right] \tag{4.16}$$

where $\mathcal{U}$ is the uncertainty set. Therefore, given an initial stage (v_j, w_j) in stage j, the optimal policy in this case is found by solving

$$V_j(\mathbf{v}_j, \mathbf{w}_j) = \max_{\pi_j \in \Pi_j} \left\{ \mathbb{E}_j \left[\sum_{t=j}^{T_A-1} \delta^{t-j} r(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t, X_t^{\pi_j}(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t)) + \delta^{T_A-j} \hat{R}^{RO}(\mathbf{v}_{T_A}^{\pi_j}, \mathbf{w}_{T_A}^{\pi_j}) \right] \right\} \tag{4.17}$$

We provide two approaches to model the uncertainty set. In the first approach, we build a rectangular uncertainty based on the stationary distribution of the uncertainties. To ensure the robustness of the approach, it is common to create an uncertainty set which results in conservative policies. A potential need for building conservative uncertainty set emerges because of the non-stationary behavior of the uncertainties. To overcome this challenge, we propose an approach to consider these non-stationary distributions using clustering techniques. In this approach, we generate a discrete uncertainty set based on the mean of points in each cluster.

4.4.1 Long-run distributions

Inflow and price variables are lognormally distributed in horizon A. Similarly, we model the long-run distributions of these variables as lognormal in horizon B, conditioned on information at time T_A. By letting t go to infinity in the stochastic models in Section 4.4, we obtain the following stationary distributions for price and inflow,

$$\ln I_t = \bar{\mu}_t^{\log} + \bar{\sigma}_t^{\log} \epsilon_t^I \qquad t \in \mathcal{T}_{T_A+1}^{AB} \tag{4.18}$$

$$\ln S_t = \xi_{T_A} + \alpha \cos(2\pi t) + \beta \sin(2\pi t) + \sigma^S \epsilon_t^S \qquad t \in \mathcal{T}_{T_A+1}^{AB}, \tag{4.19}$$

where ϵ_t^I and ϵ_t^S is standard normal distributed, and $\bar{\mu}_t^{log}$ and $\bar{\sigma}_t^{log}$ denote the mean and standard deviation of log inflow at time t. Since the short-term deviation is expected to revert to the long-term level in the long-run we use the stationary distribution for the short-term factor, and the distribution at the middle of the robust horizon for the long-term factor. The price level is set by ξ_{T_A} which is the long-term price at the end of the MDP. The expression for the variance parameter in Equation 4.19 is given by

$$\sigma^S = \sqrt{\frac{\sigma_\chi^2}{2\kappa_\chi} + \frac{T_B - T_A}{2}\sigma_\xi^2 + 2\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa_\chi}}. \tag{4.20}$$

Since the stochastic model for the long-term factor is non-stationary, we use the average variance in horizon B, where $\frac{T_B - T_A}{2}\sigma_\xi^2$. For the short-term factor we use the expression for the stationary variance. The stationary variance for inflow in a given month t is given by

$$\sigma_t^I = \bar{\sigma}_t^{log}\sqrt{\frac{\sigma_\chi^2}{2\kappa_\chi}}. \tag{4.21}$$

For the last exogenous factor in horizon A, deterioration d_t, we assume to be deterministic in horizon B, which d_{T_A} as initial condition i.e.

$$d_t = d_{T_A} + \mathbb{E}(d_{t+1} - d_t) \tag{4.22}$$

We consider the uncertainty set based on the distribution of the random variables w^i

$$\mathcal{U}_i^p = \left\{ w^i : \left| \ln w^i - \mu^i \right| \leqslant \rho\sigma^i \right\} \tag{4.23}$$

where we assume that $\ln w^i$ follows a normal distribution with mean μ_i and standard deviation σ_i. ρ considers a percentile of the distribution.

4.4.2 Discrete uncertainty set

The non-stationary dynamics of inflow and price evolution in the long run are important in estimating the asset's value under limited data. An approach used in the literature is to consider all historical data points to build a discrete uncertainty set. However, this usually results in conservative valuations and solutions with low out-of-sample performance. A natural solution is to classify similar scenarios in a group to decrease the dependence of the solutions to the outliers and consider multiple groups to capture different non-stationary dynamics. This approach assumes that the data is generated based on multiple non-stationary sources and aims to capture these sources based on the clustering approaches (K-means clustering and Gaussian mixture models (GMM)). Let $\boldsymbol{\omega} = (w_1, w_2, \cdots, w_t)$ and $\mathcal{U}_K^C$ denote of a set of d-dimensional vector and discrete uncertainty set based on K clusters, the uncertainty set $\mathcal{U}_K^C$ includes the set of $(\mu_1', , \cdots, \mu_K')$ where μ_k' is the mean of points in cluster c_i. In the case of K-means clustering, we can find μ_k' by considering the following objective function:

$$\mathcal{U}_K^C = \left\{ \boldsymbol{\mu}' : \underset{\mu_i', c_i}{\operatorname{argmin}} \sum_{i=1}^{K} \sum_{\boldsymbol{\omega} \in c_i} \|\boldsymbol{\omega} - \mu_i'\|^2 \right\}$$

Figure 18 illustrates how we find these non-stationary sources and include the mean of each cluster to build a discrete uncertainty set. We can control the degree of conservatism by choice of K. While $K = 1$ captures only the average-case scenario, $K = T$ considers the whole data set and provides more conservative solutions. While K-means clustering is a non-parametric clustering approach, GMM assumes that data points follow Gaussian distribution similar to the distribution assumption of the price and inflow in the presence of sufficient data (both follow a long-normal distribution).

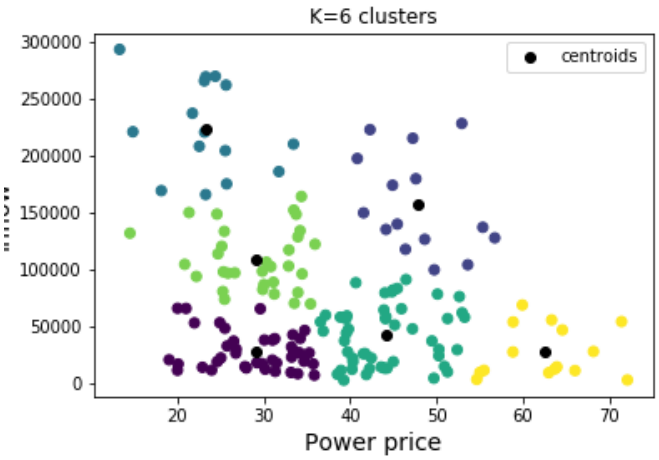


Figure 18: Clustering for monthly price and inflow data

4.5 Solution approach

The MDP in Equation 4.13 suffers from the curse of dimensionality because of large state and action spaces and non-convex action space. Furthermore, the long horizon of the problem introduces further complexities. Therefore, we first focus on assessing the terminal value. We use value and robust value iteration to find $\hat{R}^{N}(\mathbf{v}_{T_A}, \mathbf{w}_{T_A})$ and $\hat{R}^{RO}(\mathbf{v}_{T_A}, \mathbf{w}_{T_A})$ based on discretizing the state and action spaces. To solve the MDP, we need to approximate the value of the terminal state using a function that does not add to the complexity of the problem in horizon A. Piecewise linear functions are a good candidate that matches the property of the asset's terminal value as discussed in Proposition 4.3.1. We approximate the asset's terminal value based on two piece-wise regression models to find the optimal investment and operating policies in horizon A. Below, we create a simple approximation function based on $\hat{R}^{N}(\mathbf{s}), \forall \mathbf{s} \in \varphi$ ($\hat{R}^{RO}(\mathbf{s}), \forall \mathbf{s} \in \varphi$) as independent variables and $\mathbf{s}$ as dependent variables for nominal and robust approaches.

Piecewise regression model (PLR): In this approach, we estimate the terminal value using piecewise linear regression. The response variable is observations achieved by value iteration algorithm Y and a vector of m regressors $\mathbf{s}$. We use piecewise linear regression based on a decision tree to find an estimate $\hat{Y}$ of Y that minimizes squared error loss. The piecewise linear regression model estimates the $(\alpha, \boldsymbol{\beta})^b$ coefficients in set $J_b, \forall b \in \{1, \cdots, \mathcal{B}\}$.

$$\hat{Y} = \alpha^b + \boldsymbol{\beta}^b \mathbf{s}, \quad \mathbf{s} \in J_b, \quad \forall b \in \{1, \cdots, \mathcal{B}\}$$

The partitions of data, J_b will have following relations:

$$J_b \cap J_{b'} = \varnothing \quad b' \neq b, \quad \cup_{b \in \{1, \cdots, B\}} J_b = \text{Dataset}.$$

Piecewise regression decision tree requires minimum samples on each leaf not to face the lack of generalization to unseen samples. We can build the long-term investment framework using a piecewise regression model to estimate the value of the terminal state using $(\hat{R}(\mathbf{s}), \mathbf{s})$. The formulation introduces an integer variable for every class. This procedure requires the number of auxiliary integer variables equal to the number of classes, $\mathcal{B}$, leading to less time complexity than the previous model. However, the existing model disregards some properties of the problem, which might improve the optimization model's estimation performance and time complexity. Below, we use the concavity property of the value function in horizon B and develop an algorithm based on the decision tree that can construct a piecewise linear concave model.

Piecewise linear concave model (PLCR): Motivated by the concavity of the terminal state value, we propose a nonparametric method for multivariate regression subject to the concavity of the value function. We introduce a concave/convex adaptive regression tree (CARET)

which adaptively partitions the input space and fit a local linear regression on each partition. Therefore, unlike the piecewise regression model, the resulted model is a concave/convex continuous function. Given $\mathcal{B}$ subsets of linear model with coefficients $(\alpha, \beta)^k, \forall b \in \{1, \cdots, \mathcal{B}\}$, we estimate the target value, $\hat{Y}$ as:

$$\hat{Y} = \min_{b \in \{1, \cdots, \mathcal{B}\}} \alpha^b + \beta^b \mathbf{s}$$

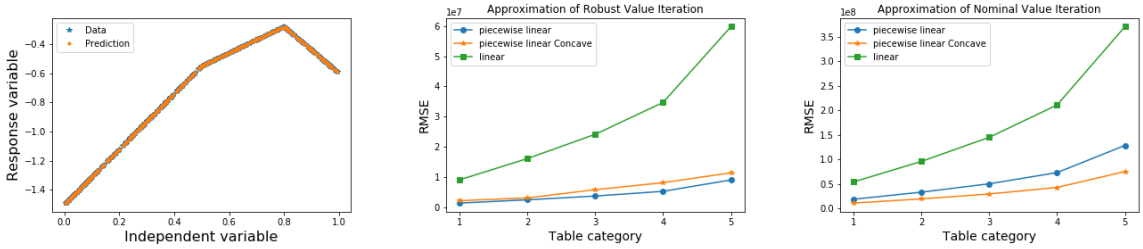
The partitioning happens in two stages. First, we create multiple classes such that it decreases the global least square error, and second, we reallocate each point, $s \in \mathbf{s}$, to one of constructed classes if $\alpha^{b'} + \beta^{b'} s \leqslant \alpha^b + \beta^b s, \forall b \in \{1, \cdots, \mathcal{B}\}$. In the first step, we split on a state variable in $\mathbf{s}$ to decrease the global least square error, increase the number of partitions, $\mathcal{B}$, and fit a linear regression on each subset to construct a hyperplane. In the second step, we reallocate each point to subsets, the hyperplane of which is dominant. This refit step places the hyperplanes in a closer alignment with observations in the dataset. The process continues until the reallocation of the observations does not create a change in the partitions, or all subsets have a minimal number of observations. Suppose $\mathbf{s} \in R^{N \times p}$. The algorithm is shown in 1.

Algorithm 1 adaptively partition the observations and fit hyperplanes on the resulted subsets. To ensure the concavity/convexity of the estimation, it refits the observations to the dominant hyperplane for each observation. We start the model with $\mathcal{B} = 1$, and split the data on the dimension l and threshold o into two subsets to minimize the square loss. In the next step we run a linear regression on each subset to find the corresponding coefficients (α, β). The algorithm accepts a split if the number of observations in a subset is more than the predefined minimum value n_{min}. Otherwise, we reject the split and move to the next subset. Once the

Algorithm 1: CARET

Result: $(\alpha, \beta)^b$
Set $\mathcal{B} = 1$, n_{min}, $\mathbf{s} \in J_1$, iter $= 1$, and Flag $=$ True;
while Flag $=$ True **do**
  $\mathcal{B}_{new} = \mathcal{B}$;
  counter $= 0$;
  for $b \in \{1, \cdots, \mathcal{B}\}$ **do**
    $\text{argmin}_{l,o} \left[\min_{c_1} \sum_{s_i^b \in J'_{b,iter}(l,o)} (y_i^b - c_1)^2 + \min_{c_2} \sum_{s_i^b \in J''_{b,iter}(l,o)} (y_i^b - c_2)^2\right]$, where $(i, l) \in (N, p)$, $s_i^b \in J_{b,iter}$, $J'_{b,iter}(l, o) = \{\mathbf{s} | \mathbf{s}_l \leqslant o\}$, $J''_{b,iter}(l, o) = \{\mathbf{s} | \mathbf{s}_l > o\}$, and $c_h = ave(y_i | \mathbf{s}_i \in J'_h(l, o))$;
    if $|J'_{b,iter}| > n_{min}$ and $|J''_{b,iter}| > n_{min}$ **then**
      $\mathcal{B}_{new} = \mathcal{B}_{new} + 1$;
      $(\alpha', \beta')^{counter+1} = (\mathbf{s}_{J'_b}{}^\intercal \mathbf{s}_{J'_b})^{-1} \mathbf{s}_{J'_b}{}^\intercal \mathbf{y}_{J'_b}$;
      $(\alpha', \beta')^{counter+1} = (\mathbf{s}_{J''_b}{}^\intercal \mathbf{s}_{J''_b})^{-1} \mathbf{s}_{J''_b}{}^\intercal \mathbf{y}_{J''_b}$;
    else
      $(\alpha', \beta')^{counter+1} = (\alpha, \beta)^b$;
  $(\alpha, \beta) = (\alpha', \beta')$;
  $\mathcal{B} = \mathcal{B}_{new}$;
  $J_{iter+1} = \{J'_{iter}, J''_{iter}\}$;
  $J_{b,iter+1} = \{\mathbf{s}_i : \alpha^b + \beta^b \mathbf{s}_i \leqslant \alpha^{b'} + \beta^{b'} \mathbf{s}_i, b' \neq b\}$;
  if $|J_{b,iter}| < n_{min}, \forall b \in \{1, \cdots, \mathcal{B}\}$ or $J_{b,iter+1} = J_{b,iter}, \forall b \in \{1, \cdots, \mathcal{B}\}$ **then**
    Flag $=$ False;
  else
    iter$+ = 1$
$\min_{b \in \{1, \cdots, \mathcal{B}\}} \alpha^b + \beta^b \mathbf{s}$

algorithm evaluates the cuts on all subsets, it refits using the partition induced by hyperplanes. The model stops either if there are not enough observations within each subset $|J_{b,iter}| < n_{min}$, or if there is no change in the subsets $J_{b,iter+1} = J_{b,iter}, \forall b \in \mathcal{B}$. Figure 19a shows the approximation results of the piecewise linear concave model on a toy example. In Figure 19, we compare the performance of approaches *PLR* and *PLCR* using (robust) value iteration data points under five various price distributions. We randomly select 70% of state valuations to constitute the training set, select 10% as the validation set to tune the hyperparameter of the decision



(a) Toy example

(b) RMSE of robust value iteration in test set

(c) RMSE of nominal value iteration in test set

Figure 19: Figure 19a shows the CARET performance in a toy example. Figures 19c and 19b illustrate the performance of approximating nominal (right) and robust (left) value iteration algorithms

tree and treat the remaining as the test. We use linear regression as a benchmark and measure the root mean square error (RMSE) of three methods on the test dataset. Figures 19b and 19c show that piecewise linear and piecewise linear regression perform better than linear regression.

Using the above algorithm, we can create an optimization framework for the long-term problem with a concave and continuous terminal value. The optimization model will be as following:

$$V_j(\mathbf{v}_j, \mathbf{w}_j) = \max_{\pi_j \in \Pi_j} \left\{ \mathbb{E}_j \left[\sum_{t=j}^{T_A-1} \delta^{t-j} r_t(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t, X_t^{\pi_j}(\mathbf{v}_t^{\pi_j}, \mathbf{w}_t)) + \delta^{T_A-j} \hat{R}(\boldsymbol{v}_{T_A}, \boldsymbol{w}_{T_A}) \right] \right\}$$

$$\text{s.t.} \quad \alpha^b + \boldsymbol{\beta}^b \mathbf{s}(\boldsymbol{v}_{T_A}, \boldsymbol{w}_{T_A}) \geqslant \hat{R}(\boldsymbol{v}_{T_A}, \boldsymbol{w}_{T_A}), \quad \forall b \in \{1, \cdots, \mathcal{B}\}$$

4.5.1 Dual reoptimization heuristic

In this section we propose an ADP-based algorithm for the hybrid MDP and terminal value formulation in Section 4.3. The heuristic is designed based on the Dual reoptimization heuristic (DRH) developed by [184]. The DRH is motivated by information relaxations and duality approach [40].

A dual bound is estimated in Monte Carlo simulation by solving a deterministic variant of MDP endowed with full information about future uncertainty and costs adjusted for this

knowledge using a dual penalty. Let $q(\mathbf{x}_t, \mathbf{W}_t, \mathbf{v}_t)$ denote the stage t dual penalty function, where $\mathbf{W}_t := (\mathbf{w}_t, t \in \{j, \cdots, T_A\})$ is a vector of realized stochastic factors for each stage from j to τ. If $\mathbf{E}_t[q_t(\mathbf{x}_t, \mathbf{W}_t, \mathbf{v}_t)] \leqslant 0$ holds then the dual penalty function is feasible. Given the knowledge of W_t, we define the following hindsight reward function:

$$r_t^{IR}(\mathbf{x}_t, \mathbf{W}_t, \mathbf{v}_t) = x_t^G p_t \eta - C_R x_t^R - C_U x_t^U - C_F \mathbf{1}_{\{c_t=1\}}$$

and consider the following deterministic optimization problem

$$V_t^{IR}(\mathbf{v}_t; \mathbf{W}_t) = \max_{\mathbf{x}_{t'}, \mathbf{y}_{t'}} \sum_{t'=t}^{T_A} \delta^{t'-t} \left[r_{t'}^{IR}(\mathbf{x}_{t'}, \mathbf{W}_{t'}, \mathbf{v}_{t'}) - q_{t'}(\mathbf{x}_{t'}, \mathbf{W}_{t'}, \mathbf{v}_{t'}) \right] + \delta^{T_A - t'} \hat{R}(\mathbf{v}_{T_A}, \mathbf{w}_{T_A}) \tag{4.24a}$$

$$\text{s.t.:} \quad \mathbf{y}_t = \mathbf{v}_t, \tag{4.24b}$$

$$\mathbf{y}_{t'+1} = f_{t'}(\mathbf{y}_{t'}, \mathbf{x}_{t'}), \quad \forall t' \in \{t, \cdots, T_A\}, \tag{4.24c}$$

$$\mathbf{y}_{t'} \in \mathcal{Y}_{t'}, \quad \forall t' \in \{t, \cdots, T_A\} \tag{4.24d}$$

$$\mathbf{x}_{t'} \in \mathcal{X}_{t'} \quad \forall t' \in \{t, \cdots, T_A\} \tag{4.24e}$$

The above math program compute decisions $\mathbf{x}_t, \forall t \in \mathcal{T}_A$ and includes auxiliary variables y_j to track endogenous MDP state. Its objective (Equation 4.24a) is the sum of discounted rewards over the horizon with full and limited information about evolution of uncertainties. Constraint Equation 4.24b initializes the stage t state to the current state $\mathbf{v}_t$. Constraints Equation 4.24c ensure the feasibility of state transitions. Constraints Equation 4.24e–Equation 4.24d restrict decision variables to their respective feasible sets.

The quality of the dual bound depends on the dual penalty function in Equation 4.24. As discussed by [40], the optimal policy value is attainable when using ideal dual penalty on the MDP value function:

$$q_t(\nu_t, W_t, x_t) = \gamma\{V_{t+1}(f(\nu_t, x_t), w_{t+1}) - \mathbb{E}[V_{t+1}(f(\nu_t, x_t), w_{t+1})]\}, \tag{4.25}$$

Since the exact value function is not available in Equation 4.25, one can approximate the value function. However, given the magnitude of our problem and the complexities that value function approximation based dual penalty can cause in solving the problem, we use simple dual penalties that do not rely on such approximations [39, 166]. We define the dual penalty function as:

$$\begin{aligned} q_t(\nu_t, W_t, x_t) := \delta\left[(\mathbb{E}[\xi_{t+1}] - \xi_{t+1}) + (\mathbb{E}[\chi_{t+1}] - \chi_{t+1}) + (\mathbb{E}[\omega_{t+1}] - \omega_{t+1})\right] x_t^{Gen} + \\ \sum_{t'>t+\tau_R} \delta^{t'-t}\left[(\mathbb{E}[\xi_{t'}] - \xi_{t'}) + (\mathbb{E}[\chi_{t'}] - \chi_{t'})\right] x_{t'}^{up} \end{aligned} \tag{4.26}$$

In equation Equation 4.26, the extra information for taking a decision at period t is approximated by the spreads between the value of uncertainties in the periods $(t+1, T_A]$ and their expectations. In particular, we consider the extra information at the next stage, $t+1$, and all stages in $(t+\tau_R, T_A]$ as the main component of taking generation and investment decisions. The dual penalty in Equation 4.26 is linear in x and is feasible because the expected value of spreads of uncertainties equal zero (e.g., $\mathbb{E}_t\left[\mathbb{E}_t[\xi_{t+1}] - \xi_{t+1}\right] = 0$).

Estimating the dual bound $\mathbb{E}[V_t^{IR}(\nu_t, \mathbf{W}_t)|w_t]$ involves solving the optimization problem on N sample paths $\{\mathbf{W}_t^n, n = 1, ..., N\}$. Although solving Equation 4.24 at stage t and state $(\mathbf{v}_t, \mathbf{W}_t|\mathbf{w}_t)$ provides a random decision $\mathbf{x}_t(\mathbf{W}_t)$, we can construct a distribution of the random

variable $\mathbf{x}_t(\mathbf{W}_t)$ and only implement a projected value of the distribution corresponding to the current state, $\mathbf{x}_t(\hat{\mathbf{W}}_t) = \mathcal{H}_t(\mathbf{x}_t(\mathbf{W}_t))$ [184]. Implementing this decision results in a transition to a new state $\mathbf{v}_{t+1} = f_t(\mathbf{y}_t, \mathbf{x}_t(\hat{\mathbf{W}}_t))$. Once new information at time $t+1$, $\mathbf{w}_{t+1}$ becomes available, we generate new sample paths $\mathbf{W}_{t+1}|\mathbf{w}_{t+1}$ and solve the analogue of math program Equation 4.24 to obtain $\mathbf{x}_t(\hat{\mathbf{W}}_t)$. We repeat this procedure until we reach stage T_A.

4.6 Numerical study

In sections 4.6.1 and 4.6.2, we describe our real instances and explain the computational setup.

4.6.1 Power plant instance

Table Table VIII summarizes the parameter values of our baseline. We use a 20 year planning horizon and set $T_A = 5$ years. Therefore, the problem includes 5 and 15 year MDP and robust period horizon. In the first 5 year we predict the evolution of uncertainties using the historical data. However, due to the limited information about uncertainties (e.g., price, inflow) in the next 15 year, horizon B, we assess the value of the asset considering worst case scenario. Instances in this paper illustrate properties of a real Hydro power plant. Features such as the down time period due to a failure and refurbishment is set based on experts' estimations.

We consider stochastic process for weekly power price and inflow in the MDP horizon. This is equivalent to the firm selling the commodity at the average weekly price at each stage and receiving average volume of water at each stage. We model inflow and the power price as discussed in section 4.2. Price parameters are estimated based on historical data obtained from Montel (see www.montel.com). Local inflow data are obtained from a real hydropower and

TABLE VIII: Parameters defining the baseline instance

Name	Value	Unit	Name	Value	Unit	Name	Value	Unit
T_A	260	weeks	q_0	$3.329 \cdot 10^4$	MWh/week	q^{max}	$2 * c_0$	MW/month
τ^F	24	weeks	C^F	$2 \cdot 10^6$	NOK	τ^R	8	weeks
C^U	$2 \cdot 10^3$	NOK	L^{max}	$3.813 \cdot 10^5$	MWh	γ	0.992	-
L^{min}	0	MWh						

the inflow process is calibrated by first normalizing log-inflow data and then fitting an AR(1)-process. This is similar procedure as in [82], but using log-inflow instead of inflow. Modeling log-inflow instead of inflow is common in the literature [170, 175]. We assume zero correlation between inflow and power price. The evolution of the deterioration rate follows a Gamma process, following [187]. Calibration of deterioration process is based on expert opinion and estimates in [196]. We model uncertainties in horizon B as discussed in sections 4.4.1 and 4.4.2.

4.6.2 Implementation and computational setup

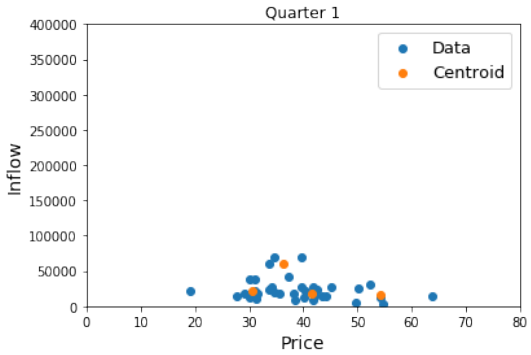
Our computational study compares policies based on robust terminal value estimations using uncertainty sets discussed in sections 4.4.1 and 4.4.2 with the policy when the terminal value is calculated using (Equation 4.15). This policy is based on a common assumption in the literature that information about uncertainties in horizon A can be extrapolated to horizon B. This model uses inflow and price distribution similar to what discussed in section 4.4.1, and we follow implementation details discussed in the next paragraph.

To find the asset's value at horizon B, we use features including the capacity, reservoir level, plant condition, and long-term mean of power price. We also consider the seasonal pattern of uncertainties by embedding time in the state space. We leverage the value iteration algorithm and discretize the state space. Inflow and price usually follow seasonal patterns. To capture

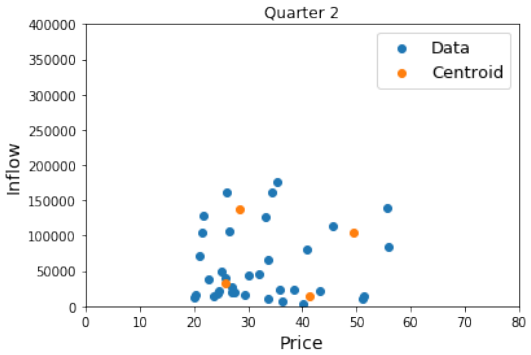
this effect, we incorporate time as one of the components in the state space in addition to the endogenous variables and model inflow and price based on monthly data. We also discretize the long-term mean of power price to five categories and adjust power price distribution in horizon B based on the power price long-term mean at time T_{A}.

We construct the discrete uncertainty sets based on monthly data as discussed in section 4.4.2. Clusters are calculated using the *Kmeans* and *GMM* functions. We remove outlier scenarios so that the set becomes less dependent on the data noise and outliers. Using these two clustering approaches, we create two discrete uncertainty sets on the monthly historical price and inflow data. In this way, we can capture their relations while not modeling the correlation between these sources of uncertainties in a parametric aspect. Therefore, the policies are only robust to the centroids of the clusters. We use the silhouette score to find the best number of clusters. In Figure Figure 20, we illustrate the monthly price and inflow data in quarters one and two and show a discrete uncertainty set based on K-means clustering and GMM with $\mathsf{K} = 4$ centers. We follow procedures discussed in section 4.4.1 to create a discrete uncertainty set. In this approach, we use features including the capacity, reservoir level, plant condition, and time to find the asset's value.

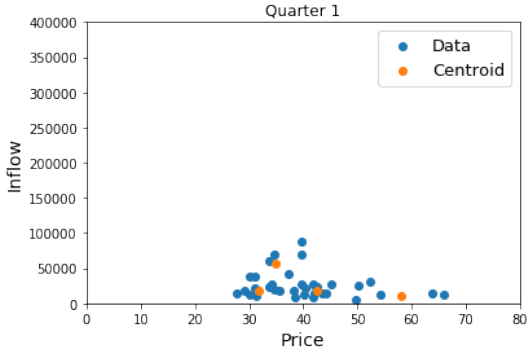
All algorithms are programmed using *Python 3.7* and *GUROBI 8.1* as the math programming solver. We use the median decision rule to find the policy in all approaches. We estimate the dual bound and the value of the investment and generation policies (i.e., expected discounted total revenue over the horizon) with distribution-based robust terminal value using 1000 and 100 Monte Carlo sample paths, respectively, with standard errors of upper/lower



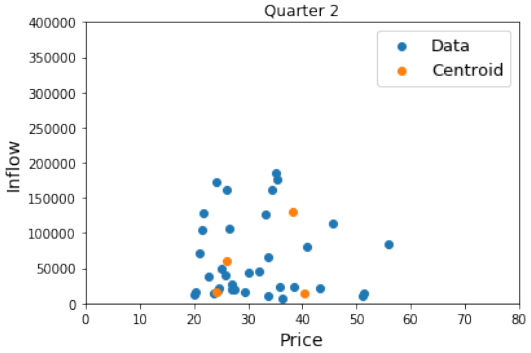
(a) A K-means clustering with K = 4 in quarter 1



(b) A K-means clustering with K = 4 in quarter 2



(c) A GMM clustering with K = 4 in quarter 1



(d) A GMM clustering with K = 4 in quarter 2

Figure 20: Illustration of methods to generate discrete uncertainty set on monthly (price, inflow) data. We used logarithm of data to find centroids for GMM.

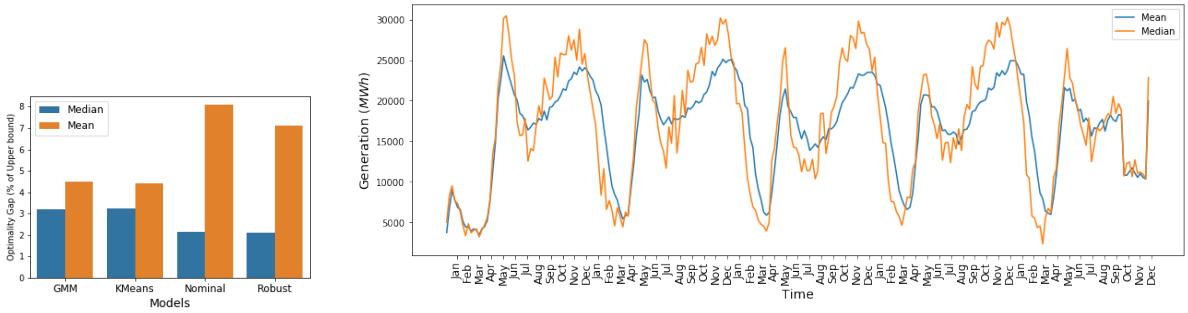
bound estimates of 0.8% and 3.05% of the mean. We calculate the optimality gap to be 2.12%. We also assess the dual bound and the value of investment and generation policies with nominal terminal value using 1000 and 100 sample paths with standard errors of upper/lower bound estimates of 3% and 4.5%. The optimality gap, in this case, is found to be 1.95%. We estimate the dual bound and the value of the investment and generation policies (i.e., expected discounted total revenue over the horizon) with Kmeans-based (GMM-based) robust terminal value using 1000 and 100 Monte Carlo sample paths, respectively, with standard errors of upper/lower bound estimates of 0.86%(0.88%) and 1.86%(1.9%) of the mean. We calculate the optimality gap to be 3.25%(3.2%). Let π_{robust}, π_{nominal}, π_{Kmeans}, and π_{GMM} denote the poli-

cies found by MDPR, MDPN, MDPKmeans, MDPGMM, respectively. We evaluate the performance of these policies using N sample trajectories discussed in section 4.8.10 and analyze its behavior on different scenarios of model misspecification in horizon B. Below, we evaluate the standard deviation and expected revenue of these policies over 50 random sample paths. We aim to compare the performance of π_{robust}, π_{nominal}, π_{Kmeans}, and π_{GMM}.

Below, we check the approaches' performance under moderate and extreme changes in the power market and climate. We optimize policies based on the calibrated parameters and obtain π_{robust} and π_{nominal} that maps elements in the set of states to action set elements. Finally, we evaluate their performance under different scenarios.

4.6.3 Impact of decision measures on the policies' performance

Below, we compare the performance of methods based on mean and median decision rules, $\mathcal{H}_{\text{t}}(\cdot)$, and report the optimality gap with respect to the upper bound of each model. The corresponding results are displayed in Figure 21a. The median decision rule results in policies with lower optimality gaps in all methods since the mean decision rule is usually affected by the outcome of rare scenarios (i.e., outliers). Figure 21b shows the average of generation quantities across samples in MDPGMM with different months using mean and median decision rules. Results show that the generation profile based on median changes more rapidly with the power price seasonal effect while the mean has a smoother generation profile. As a result, the policy based on the median is less prone to actions in rare scenarios and achieves higher expected revenue in all approaches.



(a) Optimality gaps (b) Generation profile of mean and median Decision rule in MDPGMM

Figure 21: Optimality gaps with different decision rules

4.6.4 Impact of the long-term asset valuation on policies

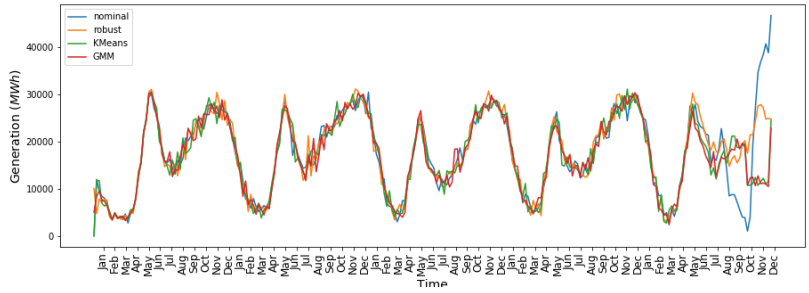
In Figure 22, we compare the expected operational and investment policies of the hydropower plant across sample paths using different approaches in horizon A. Figure 22a shows that differences in the long-term valuation of the asset has a small impact on the intermediate generation decisions, and expected generation profile of all approaches follow a similar pattern. Figures 22c and 22b illustrate seasonality patterns in the reservoir level and the spill volume of the power plant due to the high inflow input during the Spring season. As expected, the spill volume increases during periods when the amount of water in the reservoir reaches the peaks. DRH considers various future scenarios to find investment decisions. Investment on a larger number of sample paths states the optimality of investment under more scenarios. Figure 22d suggests that a robust approach results in more conservative investment decisions, whereas the nominal approach recommends refurbishment in a larger number of sample paths. Finally, Figure 22e shows that while the nominal approach finds a value in upgrading the plant's capacity, approaches with robust terminal valuations find it less plausible. This is due to the model of

the long-run evolution of uncertainties in the nominal approach, which finds the amount of inflow and the power prices high enough for an upgrade in the plant's capacity.

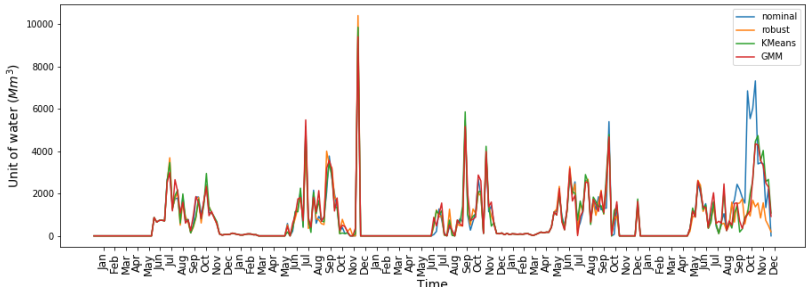
4.6.5 Performance of policies under climate change

Water as the primary source of power generation in a hydropower plant has a significant impact on its cash flow. Changes in the pattern of precipitation in the long-term can result in suboptimal short-term decisions. We compare the performance of the nominal and robust approaches under various inflow patterns to assess the performance of policies under model misspecification. In these sets of experiments, we assume that inflow can be forecasted based on historical data in the short term. However, observations deviate from the forecasting model in the long run.

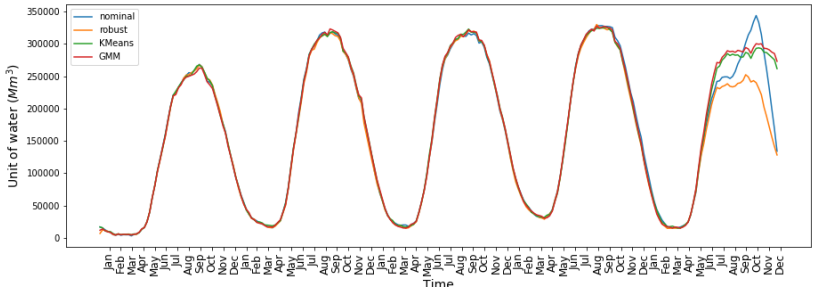
Inflow mean: We plot the expected discounted total revenue of the Hydropower plant obtained by π_{robust}, π_{nominal}, π_{GMM}, and π_{KMeans} under different means of inflow. In these figures, zero in the x-axis models the scenario when the long and short-term inflow dynamics remain the same. Negative and positive changes model environments with low (drought) and high (surplus) inflow patterns. Figures 23a and 23b show that while revenue achieved by π_{nominal} has a higher expected value in the presence of high inflow, it experiences lower expected revenue in scenarios with low inflow. Figure 23c shows the mean and standard deviation of revenue for each model across different environments. As expected, the nominal model results in higher expected revenue and standard deviation than other robust approaches. Robustness based on long-run distribution obtains the lowest expected and variance of revenue than others.



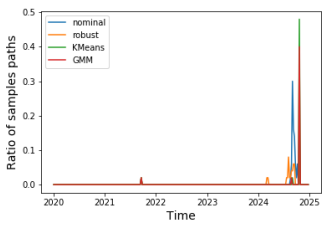
(a) Generation Policy in horizon A



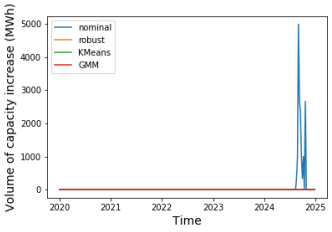
(b) Volume of Spill in horizon A



(c) Reservoir level as a result of operational and investment policies in horizon A

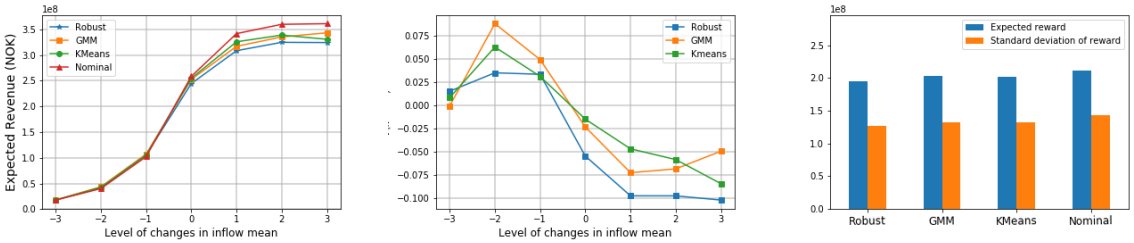


(d) Refurbishment Policy in horizon A



(e) Upgrade policy in horizon A

Figure 22: Illustration of hydropower plant policies in horizon A



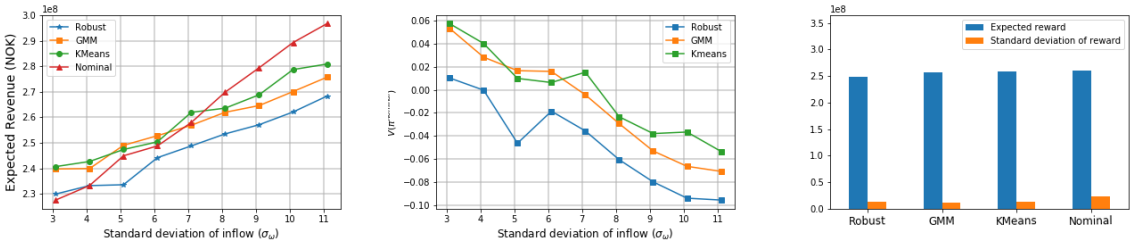
(a) Expected revenue under different inflow patterns

(b) Relative performance of robust approaches ($\frac{V(\pi)-V(\pi^N)}{V(\pi^N)}$)

(c) Expected and standard deviation of revenue across all scenarios

Figure 23: Performance of policies under different levels of inflow long-term mean

Inflow standard deviation: We plot the expected discounted total revenue of the Hydropower plant (from $N = 50$ out of sample trajectories) obtained by π_{robust}, π_{nominal}, π_{KMeans}, and π_{GMM} under different standard deviation of inflow σ_ω. Figure 24 shows that the performance of π_{robust}, π_{nominal}, π_{KMeans}, and π_{GMM} under different inflow volatility. Figure 24a illustrates that the expected revenue obtained by π_{nominal} is higher than robust policies for high volatility inflow scenarios and has a lower mean in cases of low inflow volatility. This is explainable by measuring the amount of spill suggested by the robust policy during high inflow volatility scenarios. Low future inflow anticipation in the robust models leads to having policies that tend to store water for generation during peak price periods. However, less predictability of inflow results in higher spill due to the limited reservoir capacity. Furthermore, Figure 24a shows that policy π_{robust} is the most conservative policy among approaches. Figure 24b shows that π_{GMM} and π_{KMeans} outperform π_{robust} in achieving better expected revenue and emphasize the importance of modeling the robustness. Figure 24c shows that robust approaches obtain lower volatility of expected revenue than the nominal approach.



(a) Expected revenue under different inflow patterns

(b) Relative performance of robust approaches ($\frac{V(\pi)-V(\pi^N)}{V(\pi^N)}$)

(c) Expected and standard deviation of revenue across all scenarios

Figure 24: Performance of policies under different levels of the inflow standard deviation

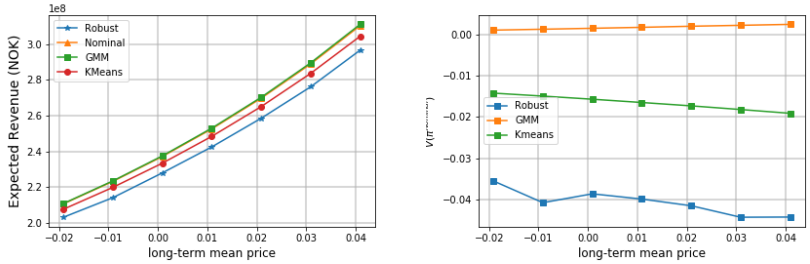
4.6.5.1 Performance of policies under changes in the power market

High penetration of the new sources of power generation in the electricity market can change the pricing structure of this commodity in the long term. Therefore, considering approaches capable of incorporating such changes can ensure the near-optimality of long-term decisions. Below, we consider various scenarios of the power market and compare the revenue distribution obtained by policies π_{nominal}, π_{robust}, π_{KMeans}, and π_{GMM}.

Long-term drift of power price: We plot the expected discounted revenue of the hydropower plant (from $N = 50$ out of sample trajectories) obtained by π_{robust}, π_{nominal}, π_{KMeans}, π_{GMM} under different drifts of the long-term price component. The base case has the long-term mean price equal to 0.01. Figure 25a shows that the revenue achieved by all policies increases with drifts of the long-term price component, and the relative performance of robust policies does not change significantly as illustrated in Figure 25b.

4.7 Conclusion

Motivated by the penetration of renewable power sources in the power generation portfolio of countries and the inefficiency of hydropower plants in power production, we study the re-



(a) Expected revenue under different price patterns

(b) Relative performance of robust approaches ($\frac{V(\pi)-V(\pi^{N})}{V(\pi^{N})}$)

Figure 25: Performance of policies under different levels of the price long-term mean

furbishment and capacity investment problem in these assets. We focus on finding the value of the asset in the long term and show that common assumption in practice for modeling the evolution of uncertainties in the long run (i.e., assuming that the short-term information can explain the long-run behavior of uncertainties) leads to high error and suboptimal decisions. To facilitate the decision making, we formulate a multi-period generation and investment MDP model that is challenging due to a non-convex action set and high dimensional state and action spaces. We introduce robust MDP for estimating the value of the hydropower plant under limited long-term information using a distribution- and clustering-based uncertainty set. Following practitioner literature, we use the dual reoptimization heuristic (DRH) to find operating and investment policies in addition to estimating an upper bound. Using real data of a hydropower plant, we show that the computed policies are near-optimal and provide comments on the choice of decision measures in DRH. Our findings suggest the presence of the annual pattern in the generation policy of hydropower plants. Furthermore, we show that estimating the value of an asset under limited long-term information through robust MDP using an un-

certainty set based on the clustering approach leads to a less conservative and robust stream of cash flows under various future environments.

4.8 Appendix

4.8.1 Proofs

***Proof of Proposition** 4.3.1.* No production and upgrade results in zero value, hence $V_j(v_j, w_j) \geqslant 0$, which implies $W_j(v_j, w_j) \geqslant 0$. An upper bound is provided by maximum upgrade and maximum production at every stage:

$$\begin{aligned} W_j(v_{j+1}, w_j) &= \mathbb{E}\left[\delta V_{j+1}(v_{j+1}, w_{j+1})|w_j\right] \\ &\leqslant \delta Q \left[\sum_{t=j+1}^{T_B} \mathbb{E}\left[p_{t+1}|p_j\right]\right] \end{aligned}$$

where $\mathbb{E}\left[p_{t+1}|p_j\right]$. Furthermore,

$$V_j = Q \min\{q_j, l_j\} + W_j(v_{j+1}, w_j).$$

Hence, the value function and continuation function are finite. Endogenous state at time j, v_j includes (c_j, l_j, q_j). We first show concavity with respect to c_j while other endogenous state variables are fixed. If $c_j \geqslant 1$ for all $t \in \{j, \cdots, T_B\}$, then $V_j(v_j, w_j) = 0$. If $c_j \leqslant 0$ for all $t \in \{j, \cdots, T_B\}$, then no costs is incurred, i.e. the immediate reward in $V_j(v_j, w_j)$ is $x_t^G p_t \eta$, which is constant in c_j. In the case of $c_j < 1$ for $j \in \{1, ..., T_F\}$ and $c_j > 1$ for $j \in \{T_F + 1, ..., J_B\}$, the value function can be reformulated to

$$V_j(v_j, w_j) = V_j''(v_j, w_j) - V_{T_F}''(v_{T_F}, w_{T_F}), \tag{4.27}$$

where $V_j''(v_j, w_j)$ has immediate reward $x_t^G p_t \eta$ instead of $x_t^G p_t \eta \mathbf{1}_{\{c_j > 1\}}$, as in the case of $c_j \leqslant 0$ for all $t \in \{j, \cdots, T_B\}$. We subtract $V_{T_F}''(v_{T_F}, w_{T_F})$, which is lost income from optimal operations

from time T_F. $V_j(v_j, w_j)$ is decreasing in T_F, and since $d_{j+1} \geqslant 0$ which implies $c_{j+1} \geqslant c_j$, which means that T_F is decreasing in c_j. Combined with the fact that $V_j(v_j, w_j) > 0$, which we showed earlier, $V_j(v_j, w_j) = V'_j(v_j, w_j) - V'_{T_F}(v_{T_F}, w_{T_F})$ is concave in c_j.

We proceed by showing that the terminal value is concave in $v'_j = (l_j, q_j)$ for a fixed w_j and for a fixed c_j. At stage $J-1$, for a given w_j, we have

$$V'_{J-1}(v'_{J-1}, w_{J-1}) = \max_{x_{J-1} \in \mathcal{X}(v'_{J-1})} r'_{J-1}(x_{J-1}, v'_{J-1}) \tag{4.28}$$

$$W'_J(v'_{J-1}, w_{J-1}) = 0 \tag{4.29}$$

This is a linear program where v'_{J-1} is the upper bound of the convex feasible set $\mathcal{X}(v'_{J-1})$. From standard linear programming results, $V_{J-1}(v'_{J-1}, w_{J-1})$ is piecewise linear concave in v'_{J-1}. The continuation function is zero and therefore also piecewise linear concave. By finiteness of the continuation function and the induction hypothesis, it is easy to verify that $W_j(\cdot, w_j)$ is piecewise linear concave in v'_j at stage j. Hence, for a feasible stage j action set $\mathcal{X}(v'_j)$ which is bounded by v_j, a linear reward function $r(x_j, w_j)$, and a concave continuation function, it follows that the value function $V_j(v'_j, w_j)$ is piecewise linear concave.

□

4.8.2 Notation

The set of periods from j to T^A is denoted $\mathcal{T}_j^A = \{j, ..., T^A\}$. The set of periods from $T^A + 1$ to T^B is denoted $\mathcal{T}^B = \{T^A + 1, ..., T^B\}$. By x_t^k, $k \in \mathcal{K}^A$ we denote decisions during periods $t \in \mathcal{T}_1^A$. By $(\mathbf{v}_t, \mathbf{w}_t) \in \mathcal{V}_t \times \mathcal{W}_t$ we denote vectors of endogenous and exogenous state variables during periods $t \in \mathcal{T}_j^A$. By $\mathbf{u}_t \in \mathcal{U}_t$ we denote the vector of exogenous states during periods

$t \in \mathcal{T}^B$.Note that the elements of the decision vector $\mathbf{x}_t = \{x_t^k,\ k \in \mathcal{K}^A\}$ for $t \in \mathcal{T}_1^A$ is different from the elements of the decision vector $\mathbf{y}_t = \{y_t^k,\ k \in \mathcal{K}^B\}$ for $t \in \mathcal{T}^B$. Also the elements of the state vector $(\mathbf{v}_t, \mathbf{w}_t)$ for $t \in \mathcal{T}_1^A$ is different from the elements of the state vector $\mathbf{u}_t$ in periods $t \in \mathcal{T}^B$. By π_j we denote a collection of of stage-dependent decision rules $\{X_t^\pi, t \in \mathcal{T}_j^A\}$, each mapping states to actions.

4.8.3 Sets

$\mathcal{T}_j^A = \{j, ..., T^A\}$: First set of time periods

$\mathcal{T}^B = \{T^A + 1, ..., T^B\}$: Second set of time periods

$\mathcal{K}^A = \{G, U, R, S\}$: Action indexing in time periods $t \in \mathcal{T}_1^A$.

$\mathcal{K}^B = \{G, L, S\}$: Action indexing in time periods $t \in \mathcal{T}^B$.

$\mathcal{X}_t \subset \mathbb{R}_+^3 \times \{0, 1\}$: Action set in time period $t \in \mathcal{T}_1^A$

$\mathcal{Y}_t \subset \mathbb{R}_+^3$: Action set in time periods $t \in \mathcal{T}^B$

$\mathcal{V}_t \subset \mathbb{R}_+^3$: Exogenous state set in time periods $t \in \mathcal{T}_1^A$

$\mathcal{W}_t \subset \mathbb{R}_+^3$: Endogenous state set in time periods $t \in \mathcal{T}_1^A$

$\mathcal{U}_t \subset \mathbb{R}_+^2$: Exogenous state set in time periods $t \in \mathcal{T}^B$

4.8.4 Parameters

L^{min}: Min storage volume (m^3)

L^{max}: Max storage volume (m^3)

T^A: End of first time horizon

T^B: End of second time horizon

τ^F: Downtime of the plant associated with failure ($time$)

τ^R: Downtime of the plant associated with refurbishment ($time$)

Q^{max}: Maximum upgrade capacity ($m^3/time$)

e: Energy coefficient (kWh/m^3)

δ: Discount rate

C_U: Cost of upgrading for unit of capacity (USD)

C_R: Cost of refurbishment (USD)

C_F: Cost of failure (USD)

4.8.5 Actions

$\mathbf{x}_t = (x_t^k, k \in \mathcal{K}^{\mathcal{A}}) \in \mathcal{X}_t$: Action vector at time $t \in \mathcal{T}_1^A$.

$\mathbf{y}_t = (y_t^k, k \in \mathcal{K}^{\mathcal{B}}) \in \mathcal{Y}_t$: Action vector at time $t \in \mathcal{T}^B$.

4.8.6 States

$\mathbf{v}_t = (s_t, i_t, d_t) \in \mathcal{V}_t$: Exogenous state vector at time $t \in \mathcal{T}_1^A$.

$\mathbf{w}_t = (l_t, c_t, q_t) \in \mathcal{W}_t$: Endogenous state vector at time $t \in \mathcal{T}_1^A$.

$\mathbf{u}_t = (s_t, i_t) \in \mathcal{U}_t$: Exogenous state vector at time $t \in \mathcal{T}^B$.

4.8.7 Discretization of price and inflow processes

$$I_t = \exp(\bar{\mu}_t^{\log} + \bar{\sigma}_t^{\log}\omega_t) \tag{4.30}$$

$$S_t = \exp\left(\alpha\cos\left(\frac{2\pi t}{52}\right) + \beta\sin\left(\frac{2\pi t}{52}\right) + \chi_t + \xi_t\right) \tag{4.31}$$

$$\xi_{t+\Delta t} = \xi_t + \mu_\xi^* dt + \sigma_\xi \epsilon_1 \tag{4.32}$$

$$\chi_{t+\Delta t} = \chi_t e^{-\kappa_\chi \Delta t} - \frac{\lambda_\chi}{\kappa_\chi}(1 - e^{-\kappa_\chi \Delta t}) + \sigma_\chi\sqrt{\frac{1 - e^{-2\kappa_\chi \Delta t}}{2\kappa_\chi}}\epsilon_2 \tag{4.33}$$

$$\omega_{t+\Delta t} = \omega_t e^{-\kappa_\omega \Delta t} + \sigma_\omega\sqrt{\frac{1 - e^{-2\kappa_\omega \Delta t}}{2\kappa_\omega}}\epsilon_3 \tag{4.34}$$

where ϵ_1 and ϵ_2 are correlated by $\rho_{\chi\xi}$ and can be generated by taking random draws $\epsilon_1 \sim \mathcal{N}(0,1)$ and $\epsilon_2 \sim \mathcal{N}(\rho_{\chi\xi}\epsilon_1, 1 - \rho_{\chi\xi}^2)$. ϵ_3 is a standard normally distributed variable.

4.8.8 Five-stage example formulation

The fully integrated problem is given by

$$z^F = \max_{\{x_{t,s}^G, l_{t,s}, q_{t,s}, x_{t,s}^U,\ t\in\mathcal{T},\ s\in\mathcal{S}\}} \sum_{s=1}^{S} \pi_s \left(p_1 x_{1,s}^G \eta + \sum_{t\in\mathcal{T}^1\cup\mathcal{T}^2} \delta^{t-1}(p_{t,s} x_{t,s}^G \eta - C_U x_{t,s}^U) \right) \tag{4.35}$$

s.t.

$$l_1 = 2 \tag{4.36a}$$

$$l_{t+1,s} = l_{t,s} - x^G_{t,s} + i_{t,s} \qquad \forall t \in \mathcal{T}\backslash\{T\}, s \in \mathcal{S} \tag{4.36b}$$

$$q_1 = 3 + x^u_{1,s} \qquad \forall s \in \mathcal{S} \tag{4.36c}$$

$$q_{t+1,s} = q_{t,s} + x^u_{t,s} \qquad t \in \mathcal{T} \backslash\{T\}, s \in \mathcal{S} \tag{4.36d}$$

$$x^G_{t,s} \leqslant \min\{l_{t,s}, q_{t,s}\} \qquad \forall t \in \mathcal{T}, \forall s \in \mathcal{S} \tag{4.36e}$$

$$x^G_{t,s} \geqslant 0 \qquad \forall t \in \mathcal{T}, \forall s \in \mathcal{S} \tag{4.36f}$$

$$x^u_{t,s} \geqslant 0 \qquad \forall t \in \mathcal{T}, \forall s \in \mathcal{S} \tag{4.36g}$$

$$x^u_{t,s} = 0 \qquad \forall t \in \mathcal{T}_I, \forall s \in \mathcal{S} \tag{4.36h}$$

$$x^G_{t,s} = x^G_{t,s'} \qquad \forall t \in \mathcal{T}; \forall s \in S, \forall s' \in \mathcal{S}^t_s \tag{4.36i}$$

$$q_{t,s} = q_{t,s'} \qquad \forall t \in \mathcal{T}; \forall s \in S, \forall s' \in \mathcal{S}^t_s \tag{4.36j}$$

$$l_{t,s} = l_{t,s'} \qquad \forall t \in \mathcal{T}; \forall s \in S, \forall s' \in \mathcal{S}^t_s \tag{4.36k}$$

$$x^u_{t,s} = x^u_{t,s'} \qquad \forall t \in \mathcal{T}; \forall s \in S, \forall s' \in \mathcal{S}^t_s \tag{4.36l}$$

Generation variables and upgrade variables at stage t and scenario s are $\{x^G_{t,s},\ t \in \mathcal{T},\ s \in \mathcal{S}\}$ and $\{x^u_{t,s},\ t \in \mathcal{T},\ s \in \mathcal{S}\}$, respectively. Moreover, capacities $\{q_{t,s},\ t \in \mathcal{T},\ s \in \mathcal{S}\}$ and reservoir levels are $\{l_{t,s},\ t \in \mathcal{T},\ s \in \mathcal{S}\}$. Constraints Equation 4.36b, and Equation 4.36e connect the reservoir level to the production, Equation 4.36c-Equation 4.36d adjust the capacity based on the investment decisions at stage one and two. We denote by $\mathcal{T}_I = \{1, 3\}$ periods in

which the investment can happen which is enforced by constraint Equation 4.36h. Constraints Equation 4.36i-Equation 4.36l ensure the non-anticipativity of stage $t \in \mathcal{T}$ decision across sample paths in the set $\mathcal{S}_s^t$, where $\mathcal{S}_s^t$ is defined as the set of all scenarios that are equivalent to scenario s at time t.

4.8.9 MDP constraints

Given state components $\mathbf{v}_t \in \mathcal{V}_t$, the vector of actions, $\mathbf{x}_t \in \mathcal{X}_t(\mathbf{v}_t)$, is feasible if it satisfies following constraints:

$$x_t^G - \min\{q_t, l_t + i_t\} \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.37}$$

$$x_t^G - M\xi_t^G \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.38}$$

$$x_t^U - Q^{max} x_t^R \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.39}$$

$$\xi_t^G + x_t^R \leqslant 1 \qquad \forall t \in \mathcal{T}_1^A \tag{4.40}$$

$$-x_t^R + \xi_t^F \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.41}$$

$$c_t - 0.99 - M\xi_t^F \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.42}$$

$$\sum_{t}^{\tau^R} x_t^G - M(1 - x_t^R) \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.43}$$

$$\sum_{t}^{\tau^F} x_t^G - M(1 - \xi_t^F) \leqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.44}$$

$$x_t^R, \xi_t^G, \xi_t^F \in \{0, 1\} \qquad \forall t \in \mathcal{T}_1^A \tag{4.45}$$

$$x_t^G \geqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.46}$$

$$x_t^U \geqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.47}$$

$$x_t^S \geqslant 0 \qquad \forall t \in \mathcal{T}_1^A \tag{4.48}$$

where $\xi_t^G = \mathbf{1}_{\{x_t^G > 0\}}$ and $\xi_t^F = \mathbf{1}_{\{c_t = 1\}}$. Constraints Equation 4.37-Equation 4.38 and Equation 4.39 limit the production and upgrade quantity, (Equation 4.40) prevents production in periods that refurbishment happens, Equation 4.41 force the refurbishment when failure happens, and Equation 4.43-Equation 4.44 determine the periods that the plant is down due to the refurbishment, upgrading, and failure, respectively.

4.8.10 Calibration of stochastic processes

For prices, since electricity futures have settlement over a given period, typically weeks, months, quarters and years, we first apply the smoothing method by [16]. After applying this method, we obtain smooth futures curves for each trading date in the period from 2011 to 2018 from which we can take monthly maturities and calibrate a stochastic model. An alternative method for constructing forward curves can be found in [79]. Nordic electricity futures data were obtained from the information provider Montel [129]. After obtained smooth forward curves, a Kalman filter and maximum likelihood approach were employed. Our implementation is based on [84]. The parameter estimates are provided in Table IX. We calibrate the price model to different historical periods to be able to assess policy performance when the underlying data generating process is unknown.

Numerical values For inflow parameters we present monthly estimates. The estimation is a two step procedure where we first deseasonalize and normalize inflow by estimating $\bar{\mu}_j^{\log}$ and $\bar{\sigma}_j^{\log}$, and then fit an AR(1) process to residuals obtaining estimates for κ_ω and σ_ω.

---

[1] ***: Significance 1%.

Parameter	Base case[1] 2011-2018	2011-2014	2012-2015	2013-2016	2014-2017	2011-2018
κ_χ	1.063***	0.966	0.771	0.615	0.980	1.077
σ_χ	0.403***	0.362	0.302	0.366	0.466	0.521
λ_χ	0.234***	0.153	0.060	0.115	0.170	0.281
μ_ξ	−0.041	-0.120	-0.118	0.189	-0.134	-0.015
μ_ξ^*	0.011***	0.013	0.008	0.022	0.009	0.022
σ_ξ	0.149***	0.102	0.090	0.111	0.190	0.194
$\rho_{\chi\xi}$	−0.03***	-0.129	-0.079	-0.361	-0.030	-0.047
α	0.163***					
β	−0.025***					

TABLE IX: Parameter estimates of the price model.

Month[2] j	$\bar{\mu}_j^{\log}$	$\bar{\sigma}_j^{\log}$	σ_ω	κ_ω
1	9.768	0.655	7.095	26.30
2	9.190	0.686	7.095	26.30
3	9.699	0.656	7.095	26.30
4	9.772	0.625	7.095	26.30
5	11.286	0.529	7.095	26.30
6	12.146	0.267	7.095	26.30
7	12.015	0.400	7.095	26.30
8	11.446	0.488	7.095	26.30
9	11.204	0.372	7.095	26.30
10	11.222	0.540	7.095	26.30
11	11.103	0.674	7.095	26.30
12	10.529	0.686	7.095	26.30
13	10.076	0.940	7.095	26.30

TABLE X: Base case inflow parameter estimates

For calibration of the deterioration process we base estimates on expert knowledge, since data for evolution of equipment states are scarce. The mean and variance of the deterioration process are given by

$$\mathbb{E}(d_t) = \frac{\kappa(t)}{\nu}, \qquad \mathrm{Var}(d_t) = \frac{\kappa(t)}{\nu^2}. \tag{4.49}$$

Furthermore, empirical studies suggest that the expected deterioration at time t is proportional to the power law [187]. We let $\kappa(t) = \theta t^{\gamma}$ which means parameters θ, γ and κ needs to be estimated. [1]

[2] We define let month to consist of 4 weeks, which then will give 13 months in a year.

[1] If $\gamma = 1$ the process is stationary.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This thesis studies the challenges in the deployment of reinforcement learning algorithms in real-world business problems while focusing on applications in the operations finance area. We focus on problems in sustainable energy procurement, management of storage, and refurbishment of power plants. While we tried to address some prominent research questions in each of these directions, many follow-up research works remained to explore. We briefly present an overview of potential future contributions.

Motivated by the recent global trend in corporate energy procurement, in Chapter 2, we study the problem faced by companies that have committed to satisfying a renewable power procurement target by a future date. We develop DRH, a novel scheme that combines reoptimization and the information relaxation and duality approach to provide non-anticipative procurement policies. Further investigations can unveil the impact of intermittency of renewable power supply on the optimal PPA portfolio. Also, more analysis on the pricing of these long-term agreements can help companies construct a portfolio of these options. This has already led to future work in [183].

In Chapter 3, we study a problem faced by a firm providing services to store ethanol and analyze the behavior of users interacting with this storage provider. Our study of DRGPs shows that there is promise in leveraging decision rules to define non-linear transformations of

GPs for user modeling in the ethanol storage application. Further investigations can introduce an inference mechanism for DRGPs in similar applications.

Finally, motivated by the penetration of renewable power sources in the power generation portfolio of countries, we study the refurbishment and capacity investment problem in these assets. We focus on finding the value of the asset in the long term and show that common assumption in practice for modeling the evolution of uncertainties in the long run leads to high error and suboptimal decisions. Our findings show the effectiveness of estimating the value of an asset under limited long-term information through a combination of standard and robust MDP to achieve a robust stream of cash flows under various future environments. Further analyses may unveil new approaches for quantifying uncertainties in long-term planning horizon problems.

LITERATURE

1. ACORE: Renewable energy ppa guidebook for corporate & industrial purchasers. Technical report, American Council On Renewable Energy (ACORE), 2016.

2. Aflaki, S. and Netessine, S.: Strategic investment in renewable energy sources: The effect of supply intermittency. Manufacturing & Service Operations Management, 19(3):489–507, 2017.

3. Agrawal, V. V. and Yücel, Ş.: Renewable energy sourcing. Responsible Business Operations: Challenges and Opportunities, pages 211–224, 2021.

4. Ahmed, S., Cabral, F. G., and da Costa, B. F. P.: Stochastic lipschitz dynamic programming. Mathematical Programming, pages 1–39, 2020.

5. Al-Gwaiz, M., Chao, X., and Wu, O. Q.: Understanding how generation flexibility and renewable energy affect power market competition. Manufacturing & Service Operations Management, 19(1):114–131, 2016.

6. Alizamir, S., de Véricourt, F., and Sun, P.: Efficient feed-in-tariff policies for renewable energy technologies. Operations Research, 64(1):52–66, 2016.

7. Allon, G. and Van Mieghem, J.: Global dual sourcing: Tailored base-surge allocation to near- and offshore production. Management Science, 56(1):110–124, 2010.

8. Andersen, L. and Broadie, M.: Primal-dual simulation algorithm for pricing multidimensional American options. Management Science, 50(9):1222–1234, 2004.

9. Andersson, A. M., Elverhøi, M., Fleten, S.-E., Fuss, S., Szolgayová, J., and Troland, O. C.: Upgrading hydropower plants with storage: timing and capacity choice. Energy Systems, 5(2):233–252, 2014.

10. Babich, V. and Birge, J. R.: Foundations and trends at the interface of finance, operations, and risk management. Operations, and Risk Management (September 10, 2020), 2020.

11. Bagnell, J. A., Ng, A. Y., and Schneider, J. G.: Solving uncertain markov decision processes. 2001.

12. Baker McKenzie: The rise of corporate PPAs - A new driver for renewables. Technical report, 2015.

13. Baker McKenzie: The rise of corporate PPAs 2.0. Technical report, 2018.

14. Beaumont, M. A.: Approximate bayesian computation. Annual Review of Statistics and its Application, 6:379–403, 2019.

15. Ben-Tal, A. and Nemirovski, A.: Robust solutions of uncertain linear programs. Operations research letters, 25(1):1–13, 1999.

16. Benth, F. E., Benth, J. S., and Koekebakker, S.: Stochastic modelling of electricity and related markets, volume 11. World Scientific, 2008.

17. Berling, P. and Rosling, K.: The effects of financial risks on inventory policy. Management Science, 51(12):1804–1815, 2005.

18. Bertsekas, D. P.: Dynamic programming and suboptimal control: A survey from ADP to MPC. European Journal of Control, 11(4-5):310–334, 2005.

19. Bertsekas, D. P.: Dynamic Programming and Optimal Control. Athena Scientific, Belmont, MA, USA, third edition, 2011.

20. Bertsekas, D. P. et al.: Dynamic programming and optimal control: Vol. 1. Athena scientific Belmont, 2000.

21. Bertsimas, D. and Dunn, J.: Optimal classification trees. Machine Learning, 106(7):1039–1082, 2017.

22. Bertsimas, D., Shtern, S., and Sturt, B.: A data-driven approach to multi-stage stochastic linear optimization. Preprint, 2018.

23. Bertsimas, D. and Sim, M.: The price of robustness. Operations research, 52(1):35–53, 2004.

24. Bertsimas, D., Sim, M., and Zhang, M.: Adaptive distributionally robust optimization. Management Science, 65(2):604–618, 2019.

25. Bertsimas, D. and Thiele, A.: Robust and data-driven optimization: modern decision making under uncertainty. In Models, methods, and applications for innovative decision making, pages 95–122. INFORMS, 2006.

26. Blei, D. M., Kucukelbir, A., and McAuliffe, J. D.: Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518):859–877, 2017.

27. Bloomberg: United States rates and bonds, 2018.

28. BNEF: Corporations already purchased record clean energy volumes in 2018, and it's not an anomaly, 2018. Bloomberg New Energy Finance (BNEF).

29. BNEF: Executing your corporate energy strategy. Technical report, Bloomberg New Energy Finance (BNEF), 2018.

30. Bøckman, T., Fleten, S.-E., Juliussen, E., Langhammer, H. J., and Revdal, I.: Investment timing and optimal capacity choice for small hydropower projects. European Journal of Operational Research, 190(1):255–267, 2008.

31. Bonnans, J. F. and Shapiro, A.: Perturbation analysis of optimization problems. Springer Science & Business Media, 2013.

32. Boogert, A. and De Jong, C.: Gas storage valuation using a monte carlo method. The journal of derivatives, 15(3):81–98, 2008.

33. Boomsma, T. K., Meade, N., and Fleten, S.-E.: Renewable energy investments under different support schemes: A real options approach. European Journal of Operational Research, 220(1):225–237, 2012.

34. Boyabatlı, O., Kleindorfer, P. R., and Koontz, S. R.: Integrating long-term and short-term contracting in beef supply chains. Management Science, 57(10):1771–1787, 2011.

35. Brown, D. B. and Haugh, M. B.: Information relaxation bounds for infinite horizon Markov decision processes. Operations Research, 65(5):1355–1379, 2017.

36. Brown, D. B. and Smith, J. E.: Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds. Management Science, 57(10):1752–1770, 2011.

37. Brown, D. B. and Smith, J. E.: Information relaxations, duality, and convex stochastic dynamic programs. Operations Research, 62(6):1394–1415, 2014.

38. Brown, D. B., Smith, J. E., and Sun, P.: Information relaxations and duality in stochastic dynamic programs. Operations Research, 58(4):785–801, 2010.

39. Brown, D. B. and Smith, J. E.: Information relaxations, duality, and convex stochastic dynamic programs. Operations Research, 62(6):1394–1415, 2014.

40. Brown, D. B., Smith, J. E., and Sun, P.: Information relaxations and duality in stochastic dynamic programs. Operations research, 58(4-part-1):785–801, 2010.

41. Bruck, M., Sandborn, P., and Goudarzi, N.: A levelized cost of energy (LCOE) model for wind farms that include power purchase agreements (PPAs). Renewable Energy, 122:131–139, 2018.

42. Calafiore, G. C. and El Ghaoui, L.: On distributionally robust chance-constrained linear programs. Journal of Optimization Theory and Applications, 130(1):1–22, 2006.

43. Cartea, A. and Figueroa, M. G.: Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. Applied Mathematical Finance, 12(4):313–335, 2005.

44. Cartea, A., Figueroa, M. G., and Geman, H.: Modelling electricity prices with forward looking capacity constraints. Applied Mathematical Finance, 16(2):103–122, 2009.

45. CDP: RE 100, 2018. Carbon Disclosure Project (CDP).

46. CDP, WWF, Investments, C., and Ceres: How the largest US companies are capturing business value while addressing climate change. Technical report, Carbon Disclosure Project (CDP), World Wide Fund for Nature (WWF), Coalition for Environmentally Responsible Economies (Ceres), 2017.

47. Chand, S., Hsu, V. N., and Sethi, S.: Forecast, solution, and rolling horizons in operations management problems: A classified bibliography. Manufacturing & Service Operations Management, 4(1):25–43, 2002.

48. Chao, X. and Chen, F. Y.: An optimal production and shutdown strategy when a supplier offers an incentive program. Manufacturing & Service Operations Management, 7(2):130–143, 2005.

49. Charnes, A., Dreze, J., and Miller, M.: Decision and horizon rules for stochastic planning problems: A linear example. Econometrica: Journal of the Econometric Society, pages 307–330, 1966.

50. Chen, Z., Yu, P., and Haskell, W. B.: Distributionally robust optimization for sequential decision-making. Optimization, 68(12):2397–2426, 2019.

51. Chua, K., Calandra, R., McAllister, R., and Levine, S.: Deep reinforcement learning in a handful of trials using probabilistic dynamics models. arXiv preprint arXiv:1805.12114, 2018.

52. Cortazar, G., Milla, C., and Severino, F.: A multicommodity model of futures prices: Using futures prices of one commodity to estimate the stochastic process of another. Journal of Futures Markets: Futures, Options, and Other Derivative Products, 28(6):537–560, 2008.

53. Cruise, J., Flatley, L., Gibbens, R., and Zachary, S.: Control of energy storage with market impact: Lagrangian approach and horizons. Operations Research, 67(1):1–9, 2019.

54. Dai, Z., Alvarez, M., and Lawrence, N.: Efficient modeling of latent information in supervised learning using gaussian processes. In Advances in Neural Information Processing Systems, pages 5131–5139, 2017.

55. Damianou, A. and Lawrence, N. D.: Semi-described and semi-supervised learning with gaussian processes. arXiv preprint arXiv:1509.01168, 2015.

56. Dantzig, G. B.: Linear programming under uncertainty. Management science, 1(3-4):197–206, 1955.

57. De Farias, D. P. and Van Roy, B.: The linear programming approach to approximate dynamic programming. Operations Research, 51(6):850–865, 2003.

58. Deisenroth, M. and Rasmussen, C. E.: Pilco: A model-based and data-efficient approach to policy search. In Proceedings of the 28th International Conference on machine learning (ICML-11), pages 465–472. Citeseer, 2011.

59. Delage, E. and Iancu, D. A.: Robust multistage decision making. In The operations research revolution, pages 20–46. INFORMS, 2015.

60. Delage, E. and Mannor, S.: Percentile optimization for markov decision processes with parameter uncertainty. Operations research, 58(1):203–213, 2010.

61. Delage, E. and Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations research, 58(3):595–612, 2010.

62. Denholm, P. and Sioshansi, R.: The value of compressed air energy storage with wind in transmission-constrained electric power systems. Energy Policy, 37(8):3149–3158, 2009.

63. Desai, V. V., Farias, V. F., and Moallemi, C. C.: Pathwise optimization for optimal stopping problems. Management Science, 58(12):2292–2308, 2012.

64. DirectEnergy: Index price electricity: Finding value in the marketplace, 2018.

65. Dixit, A. K., Dixit, R. K., and Pindyck, R. S.: Investment under uncertainty. Princeton university press, 1994.

66. DOE: Renewable electricity production tax credit (ptc), 2016. Department of Energy (DOE).

67. Drake, D. F., Kleindorfer, P. R., and Van Wassenhove, L. N.: Technology choice and capacity portfolios under emissions regulation. Production and Operations Management, 25(6):1006–1025, 2016.

68. Duffie, D.: Dynamic asset pricing theory. Princeton University Press, 2010.

69. Dulac-Arnold, G., Evans, R., van Hasselt, H., Sunehag, P., Lillicrap, T., Hunt, J., Mann, T., Weber, T., Degris, T., and Coppin, B.: Deep reinforcement learning in large discrete action spaces. arXiv preprint arXiv:1512.07679, 2015.

70. Dulac-Arnold, G., Mankowitz, D., and Hester, T.: Challenges of real-world reinforcement learning. arXiv preprint arXiv:1904.12901, 2019.

71. Dupavcova, J.: The minimax approach to stochastic programming and an illustrative application. Stochastics: An International Journal of Probability and Stochastic Processes, 20(1):73–88, 1987.

72. EIA: Hydroelectric generators are among the United States' oldest power plants - Today in Energy - U.S. Energy Information Administration (EIA), 2017.

73. EIA: Capacity factors for utility scale generators not primarily using fossil fuels, January 2013-March 2018, 2018. Energy Information Administration (EIA).

74. EIA: Wind generators' cost declines reflect technology improvements and siting decisions, 2018. Energy Information Administration (EIA).

75. EPA: Green power pricing, 2018. United States Environmental Protection Agency (EPA).

76. Escribano, A., Ignacio Pena, J., and Villaplana, P.: Modelling electricity prices: International evidence. Oxford Bulletin of Economics and Statistics, 73(5):622–650, 2011.

77. Finn, C., Abbeel, P., and Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning, pages 1126–1135. PMLR, 2017.

78. Fleten, S.-E., Maribu, K. M., and Wangensteen, I.: Optimal investment strategies in decentralized renewable power generation under uncertainty. Energy, 32(5):803–815, 2007.

79. Fleten, S.-E. and Lemming, J.: Constructing forward price curves in electricity markets. Energy Economics, 25(5):409–424, 2003.

80. Geman, H. and Roncoroni, A.: Understanding the fine structure of electricity prices. The Journal of Business, 79(3):1225–1261, 2006.

81. Georghiou, A., Tsoukalas, A., and Wiesemann, W.: Robust dual dynamic programming. Operations Research, 67(3):813–830, 2019.

82. Gjelsvik, A., Mo, B., and Haugstad, A.: Long-and medium-term operations planning and stochastic modelling in hydro-dominated power systems based on stochastic dual dynamic programming. In Handbook of power systems I, pages 33–55. Springer, 2010.

83. Goh, J. and Sim, M.: Distributionally robust optimization and its tractable approximations. Operations research, 58(4-part-1):902–917, 2010.

84. Goodwin, D.: Schwartz-smith 2-factor model - parameter estimation, 2020.

85. Google: Achieving our 100% renewable energy purchasing goal and going beyond. Technical report, 2016.

86. Google: Environmental report: 2017 progress update. Technical report, 2017.

87. Gouriéroux, C. and Valéry, P.: Estimation of a Jacobi process. 2004. Working paper, University of Toronto and Université de Montréal.

88. Green Power Partnership: Introduction to virtual power purchase agreements, 2016.

89. Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H., and Davidson, J.: Learning latent dynamics for planning from pixels. In International Conference on Machine Learning, pages 2555–2565. PMLR, 2019.

90. Hanasusanto, G. A. and Kuhn, D.: Robust data-driven dynamic programming. Advances in Neural Information Processing Systems, 26:827–835, 2013.

91. Haugh, M. B. and Kogan, L.: Pricing American options: A duality approach. Operations Research, 52(2):258–270, 2004.

92. Haugh, M. B. and Lacedelli, O. R.: Information relaxation bounds for partially observed Markov decision processes. Working paper, Imperial College London, 2018.

93. Hensman, J., Fusi, N., and Lawrence, N. D.: Gaussian processes for big data. arXiv preprint arXiv:1309.6835, 2013.

94. Heyman, D. P. and Sobel, M. J.: Stochastic Models in Operations Research: Stochastic Optimization, volume 2. Courier Corporation, 1982.

95. Hjelmeland, M. N., Zou, J., Helseth, A., and Ahmed, S.: Nonconvex medium-term hydropower scheduling by stochastic dual dynamic integer programming. IEEE Transactions on Sustainable Energy, 10(1):481–490, 2019.

96. Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J.: Stochastic variational inference. Journal of Machine Learning Research, 14:1303–1347, 2013.

97. Hu, S., Souza, G. C., Ferguson, M. E., and Wang, W.: Capacity investment in renewable energy technology with supply intermittency: Data granularity matters! Manufacturing & Service Operations Management, 17(4):480–494, 2015.

98. IRENA: Hydropower technology brief, 2015.

99. İşlegen, Ö. and Reichelstein, S.: Carbon capture by fossil fuel power plants: An economic analysis. Management Science, 57(1):21–39, 2011.

100. Iyengar, G. N.: Robust dynamic programming. Mathematics of Operations Research, 30(2):257–280, 2005.

101. Jiang, D. R. and Powell, W. B.: Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming. INFORMS Journal on Computing, 27(3):525–543, 2015.

102. Jiang, D. R. and Powell, W. B.: An approximate dynamic programming algorithm for monotone value functions. Operations Research, 63(6):1489–1511, 2015.

103. Johns, J. A., Lund, M. A., and Martin, J. H.: The Law of Solar Energy: A Guide to Business and Legal Issue, chapter Power Purchase Agreements: Distributed Generation Projects. Stoel Rives, fifth edition, 2017.

104. Kim, J. H. and Powell, W. B.: Optimal energy commitments with storage and intermittent supply. Operations research, 59(6):1347–1360, 2011.

105. Kleindorfer, P. R. and Wu, D. J.: Integrating long-and short-term contracting via business-to-business exchanges for capital-intensive industries. Management Science, 49(11):1597–1615, 2003.

106. Knittel, C. R. and Roberts, M.: An empirical examination of deregulated electricity prices. 2001. Working paper, Massachusetts Institute of Technology and University of Pennsylvania.

107. Kouvelis, P., Li, R., and Ding, Q.: Managing storable commodity risks: The role of inventory and financial hedge. Manufacturing & Service Operations Management, 15(3):507–521, 2013.

108. Lai, G., Margot, F., and Secomandi, N.: An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. Operations research, 58(3):564–582, 2010.

109. Lara, C. L., Siirola, J. D., and Grossmann, I. E.: Electric power infrastructure planning under uncertainty: stochastic dual dynamic integer programming (sddip) and parallelization scheme. Optimization and Engineering, pages 1–39, 2019.

110. Lázaro-Gredilla, M.: Bayesian warped gaussian processes. In Advances in Neural Information Processing Systems 25, eds, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, pages 1619–1627. Curran Associates, Inc., 2012.

111. Lázaro-Gredilla, M. and Titsias, M. K.: Variational heteroscedastic gaussian process regression. In International Conference on Machine Learning, pages 841–848, 2011.

112. Lazic, N., Lu, T., Boutilier, C., Ryu, M., Wong, E. J., Roy, B., and Imwalle, G.: Data center cooling using model-predictive control. 2018.

113. Letham, B., Rudin, C., McCormick, T. H., Madigan, D., et al.: Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. The Annals of Applied Statistics, 9(3):1350–1371, 2015.

114. Li, C. L. and Kouvelis, P.: Flexible and risk-sharing supply contracts under price uncertainty. Management Science, 45(10):1378–1398, 1999.

115. Lin, Q., Nadarajah, S., and Soheili, N.: Revisiting approximate linear programming: Constraint-violation learning with applications to inventory control and energy storage. Management Science, 66(4):1544–1562, 2020.

116. Lin, Q., Nadarajah, S., Soheili, N., and Yang, T.: A data efficient and feasible level set method for stochastic convex optimization with expectation constraints. Journal of Machine Learning Research, 21(143):1–45, 2020.

117. Longstaff, F. A. and Schwartz, E. S.: Valuing American options by simulation: A least-squares approach. Review of Financial Studies, 14(1):113–147, 2001.

118. Lucia, J. J. and Schwartz, E. S.: Electricity prices and power derivatives: Evidence from the Nordic power exchange. Review of Derivatives Research, 5(1):5–50, 2002.

119. Mandl, C., Nadarajah, S., Minner, S., and Gavirneni, S.: Structured data-driven operating policies for commodity storage. Available at SSRN 3398119, 2019.

120. Mankowitz, D., Mann, T., Bacon, P.-L., Precup, D., and Mannor, S.: Learning robust options. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.

121. Mannor, S., Simester, D., Sun, P., and Tsitsiklis, J. N.: Bias and variance approximation in value function estimates. Management Science, 53(2):308–322, 2007.

122. Martínez-de Albéniz, V. and Simchi-Levi, D.: A portfolio approach to procurement contracts. Production and Operations Management, 14(1):90–114, 2005.

123. Meibom, P., Larsen, H. V., Barth, R., Brand, H., Tuohy, A., and Ela, E.: Advanced unit commitment strategies in the united states eastern interconnection. Contract, 303:275–3000, 2011.

124. Milligan, M., Ela, E., Lew, D., Corbus, D., Wan, Y., Hodge, B., and Kirby, B.: Operational analysis and methods for wind integration studies. IEEE Transactions on Sustainable Energy, 3(4):612–619, 2012.

125. Mills, A., Seel, J., and Wiser, R.: As more solar and wind come onto the grid, prices go down but new questions come up, 2017.

126. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. nature, 518(7540):529–533, 2015.

127. Mohseni-Taheri, D., Nadarajah, S., and Tulabandhula, T.: Interpretable user models via decision-rule gaussian processes: Preliminary results on energy storage. 2019.

128. Montavon, G., Samek, W., and Müller, K.-R.: Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73:1–15, 2018.

129. Montel: Montel, 2021.

130. Moreno-Muñoz, P., Artés, A., and Álvarez, M.: Heterogeneous multi-output gaussian process prediction. In Advances in Neural Information Processing Systems, pages 6711–6720, 2018.

131. Nadarajah, S. and Secomandi, N.: Merchant energy trading in a network. Operations Research, 66(5):1304–1320, 2018.

132. Nadarajah, S., Margot, F., and Secomandi, N.: Relaxations of approximate linear programs for the real option management of commodity storage. Management Science, 61(12):3054–3076, 2015.

133. Nadarajah, S., Margot, F., and Secomandi, N.: Comparison of least squares monte carlo methods with applications to energy real options. European Journal of Operational Research, 256(1):196–204, 2017.

134. Nadarajah, S. and Secomandi, N.: Relationship between least squares monte carlo and approximate linear programming. Operations Research Letters, 45(5):409–414, 2017.

135. Nadarajah, S. and Secomandi, N.: Merchant energy trading in a network. Operations Research, 66(5):1304–1320, 2018.

136. Nadarajah, S. and Secomandi, N.: Real options in energy: A guided analysis of the operations literature. Working paper, 2021.

137. Nilim, A. and El Ghaoui, L.: Robust markov decision processes with uncertain transition matrices. Doctoral dissertation, University of California, Berkeley, 2004.

138. Nilim, A. and El Ghaoui, L.: Robust control of markov decision processes with uncertain transition matrices. Operations Research, 53(5):780–798, 2005.

139. NREL: Cost and performance assumptions for modeling electricity generation technologies. Technical Report NREL/SR-6A20-48595, National Renewable Energy Laboratory (NREL), November 2010.

140. NREL: System Advisor Model (SAM), 2017. National Renewable Energy Laboratory (NREL).

141. OPIC: Important features of bankable power purchase agreements for renewable energy power projects. Technical report, Overseas Private Investment Corporation (OPIC), 2014.

142. Pandvzic, H., Wang, Y., Qiu, T., Dvorkin, Y., and Kirschen, D. S.: Near-optimal method for siting and sizing of distributed storage in a transmission network. IEEE Transactions on Power Systems, 30(5):2288–2300, 2015.

143. Pereira, M. V. F. and Pinto, L. M. V. G.: Multi-stage stochastic optimization applied to energy planning. Mathematical Programming, 52(1-3):359–375, 1991.

144. Pereira, M. V. and Pinto, L. M.: Multi-stage stochastic optimization applied to energy planning. Mathematical programming, 52(1-3):359–375, 1991.

145. Philpott, A. B., Wahid, F., and Bonnans, J. F.: Midas: A mixed integer dynamic approximation scheme. Mathematical Programming, 181(1):19–50, 2020.

146. Pinto, L., Andrychowicz, M., Welinder, P., Zaremba, W., and Abbeel, P.: Asymmetric actor critic for image-based robot learning. arXiv preprint arXiv:1710.06542, 2017.

147. Powell, W. B.: Approximate Dynamic Programming: Solving the Curses of Dimensionality. John Wiley & Sons, Hoboken, NJ, USA, second edition, 2011.

148. Powell, W. B.: Approximate Dynamic Programming: Solving the curses of dimensionality, volume 703. John Wiley & Sons, 2007.

149. Puterman, M. L.: Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, 2014.

150. PWC: Corporate renewable energy procurement survey insights. Technical report, PricewaterhouseCoopers (PWC), 2016.

151. Rachev, S. T. and Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics. Mathematics of Operations Research, 27(4):792–818, 2002.

152. Rahi, O. and Chandel, A.: Refurbishment and uprating of hydro power plants—a literature review. Renewable and Sustainable Energy Reviews, 48:726–737, 2015.

153. REN21: RENEWABLES 2018 GLOBAL STATUS REPORT. Technical report, REN21 COMMUNITY, 2018.

154. Ritzenhofen, I., Birge, J. R., and Spinler, S.: The structural impact of renewable portfolio standards and feed-in tariffs on electricity markets. European Journal of Operational Research, 255(1):224–242, 2016.

155. Römisch, W.: Stability of stochastic programming problems. Handbooks in operations research and management science, 10:483–554, 2003.

156. Ross, S. A.: The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3):341–60, 1976.

157. Sæmundsson, S., Hofmann, K., and Deisenroth, M. P.: Meta reinforcement learning with latent variable gaussian processes. arXiv preprint arXiv:1803.07551, 2018.

158. Salas, D. F. and Powell, W. B.: Benchmarking a scalable approximate dynamic programming algorithm for stochastic control of grid-level energy storage. INFORMS Journal on Computing, 30(1):106–123, 2017.

159. Saul, A. D., Hensman, J., Vehtari, A., and Lawrence, N. D.: Chained gaussian processes. In Artificial Intelligence and Statistics, pages 1431–1440, 2016.

160. Scarf, H.: A min-max solution of an inventory problem. Studies in the mathematical theory of inventory and production, 1958.

161. Schwartz, E. S. and Smith, J. E.: Short-term variations and long-term dynamics in commodity prices. Management Science, 46(7):893–911, 2000.

162. Schwartz, E. and Smith, J. E.: Short-term variations and long-term dynamics in commodity prices. Management Science, 46(7):893–911, 2000.

163. Secomandi, N.: Merchant commodity storage practice revisited. Operations Research, 63(5):1131–1143, 2015.

164. Secomandi, N. and Kekre, S.: Optimal energy procurement in spot and forward markets. Manufacturing & Service Operations Management, 16(2):270–282, 2014.

165. Secomandi, N.: Optimal commodity trading with a capacitated storage asset. Management Science, 56(3):449–467, 2010.

166. Secomandi, N.: Merchant commodity storage practice revisited. Operations Research, 63(5):1131–1143, 2015.

167. Secomandi, N. and Seppi, D. J.: Real options and merchant operations of energy and other commodities. Nicola Secomandi and Duane J. Seppi," Real Options and Merchant Operations of Energy and Other Commodities," Foundations and Trends in Technology, Information and Operations Management, 6:3–4, 2014.

168. Seifert, J. and Uhrig-Homburg, M.: Modelling jumps in electricity prices: Theory and empirical evidence. Review of Derivatives Research, 10(1):59–85, 2007.

169. Shapiro, A.: Analysis of stochastic dual dynamic programming method. European Journal of Operational Research, 209(1):63–72, 2011.

170. Shapiro, A., Tekaya, W., da Costa, J. P., and Soares, M. P.: Risk neutral and risk averse stochastic dual dynamic programming method. European journal of operational research, 224(2):375–391, 2013.

171. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of go with deep neural networks and tree search. nature, 529(7587):484–489, 2016.

172. Singh, S. P. and Scheller-Wolf, A. A.: That's not fair: Tariff structures for electricity markets with rooftop solar. 2017. Working paper, Carnegie Mellon University.

173. Snelson, E., Ghahramani, Z., and Rasmussen, C. E.: Warped gaussian processes. In Advances in Neural Information Processing Systems, pages 337–344, 2004.

174. Soyster, A. L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations research, 21(5):1154–1157, 1973.

175. Stedinger, J.: Fitting log normal distributions to hydrologic data. Water Resources Research, 16(3):481–490, 1980.

176. Sunar, N. and Birge, J. R.: Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. Management Science., 2018. Forthcoming.

177. Sutton, R. S. and Barto, A. G.: Reinforcement learning: An introduction. MIT press, 2018.

178. Tamar, A., Mannor, S., and Xu, H.: Scaling up robust mdps using function approximation. In International Conference on Machine Learning, pages 181–189. PMLR, 2014.

179. Titsias, M.: Variational learning of inducing variables in sparse gaussian processes. volume 5 of Proceedings of Machine Learning Research, pages 567–574. PMLR, 16–18 Apr 2009.

180. Tobar, F. and Rios, G.: Compositionally-warped gaussian processes. arXiv preprint arXiv:1906.09665, 2019.

181. Tomlin, B. and Wang, Y.: On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. Manufacturing & Service Operations Management, 7(1):37–57, 2005.

182. Topaloglu, H. and Powell, W. B.: Dynamic-programming approximations for stochastic time-staged integer multicommodity-flow problems. INFORMS Journal on Computing, 18(1):31–42, 2006.

183. Trivella, A., Mohseni Taheri, D., and Nadarajah, S.: Meeting corporate renewable power targets. Available at SSRN 3294724, 2018.

184. Trivella, A., Mohseni Taheri, D., and Nadarajah, S.: Meeting corporate renewable power targets. Available at SSRN 3294724, 2020.

185. Tsitsiklis, J. N. and Van Roy, B.: Regression methods for pricing complex American-style options. IEEE Transactions on Neural Networks, 12(4):694–703, 2001.

186. van Ackooij, W., Escobar, D. D., Glanzer, M., and Pflug, G. C.: Distributionally robust optimization with multiple time scales: valuation of a thermal power plant. Computational Management Science, pages 1–29, 2019.

187. Van Noortwijk, J.: A survey of the application of gamma processes in maintenance. Reliability Engineering & System Safety, 94(1):2–21, 2009.

188. Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., and Vehtari, A.: Gpstuff: Bayesian modeling with gaussian processes. Journal of Machine Learning Research, 14(Apr):1175–1179, 2013.

189. Veeraraghavan, S. and Scheller-Wolf, A. A.: Now or later: A simple policy for effective dual sourcing in capacitated systems. Operations Research, 56(4):850–864, 2008.

190. Verma, A., Murali, V., Singh, R., Kohli, P., and Chaudhuri, S.: Programmatically interpretable reinforcement learning. In International Conference on Machine Learning, pages 5045–5054. PMLR, 2018.

191. Wang, C. and Neal, R. M.: Gaussian process regression with heteroscedastic or non-gaussian residuals. arXiv preprint arXiv:1212.6246, 2012.

192. WBCSD: IFRS accounting outline for power purchase agreements. Technical report, World Business Council For Sustainable Development (WBCSD), 2018.

193. WBCSD: Innovation in power purchase agreement structures. Technical report, World Business Council For Sustainable Development (WBCSD), 2018.

194. Webb, E., Wu, O. Q., and Cattani, K.: Mind the gap: Coordinating energy efficiency and demand response. Working paper, Indiana University, 2017.

195. Weber, C., Meibom, P., Barth, R., and Brand, H.: WILMAR: A stochastic programming tool to analyze the large-scale integration of wind energy. In Optimization in the energy industry, pages 437–458. Springer, 2009.

196. Welte, T.: Deterioration and maintenance models for components in hydropower plants. 2008.

197. Welte, T., Solvang, E., Øen, O., Jordet, R., Golberg, P., and Børresen, B.: Profitability analysis of hydropower maintenance and reinvestment projects. Hydropower'15, 2015.

198. Weron, R.: Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach, volume 403. John Wiley & Sons, West Sussex, Englend, 2007.

199. Weron, R.: Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4):1030–1081, 2014.

200. Wiesemann, W., Kuhn, D., and Rustem, B.: Robust markov decision processes. Mathematics of Operations Research, 38(1):153–183, 2013.

201. Wiesemann, W., Kuhn, D., and Sim, M.: Distributionally robust convex optimization. Operations Research, 62(6):1358–1376, 2014.

202. Wiser, R. and Bolinger, M.: 2016 wind technologies market report. Technical report, 2017.

203. Wu, O. Q. and Babich, V.: Unit-contingent power purchase agreement and asymmetric information about plant outage. Manufacturing & Service Operations Management, 14(2):245–261, 2012.

204. Wu, O. Q. and Kapuscinski, R.: Curtailing intermittent generation in electrical systems. Manufacturing & Service Operations Management, 15(4):578–595, 2013.

205. Wu, O. Q., Wang, D. D., and Qin, Z.: Seasonal energy storage operations with limited flexibility: The price-adjusted rolling intrinsic policy. Manufacturing & Service Operations Management, 14(3):455–471, 2012.

206. Xu, H. and Mannor, S.: Distributionally robust markov decision processes. Mathematics of Operations Research, 37(2):288–300, 2012.

207. Ye, F., Zhu, H., and Zhou, E.: Weakly coupled dynamic program: Information and Lagrangian relaxations. IEEE Transactions on Automatic Control, 63(3):698–713, 2018.

208. Zahavy, T., Haroush, M., Merlis, N., Mankowitz, D. J., and Mannor, S.: Learn what not to learn: Action elimination with deep reinforcement learning. arXiv preprint arXiv:1809.02121, 2018.

209. Zeng, Y., Klabjan, D., and Arinez, J.: Distributed solar renewable generation: Option contracts with renewable energy credit uncertainty. Energy Economics, 48:295–305, 2015.

210. Zhang, S. and Sun, X. A.: Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization. arXiv preprint arXiv:1912.13278, 2019.

211. Zhou, Y., Scheller-Wolf, A. A., Secomandi, N., and Smith, S.: Managing wind-based electricity generation in the presence of storage and transmission capacity. Production and Operations Management., 2018. Forthcoming.

212. Ziegler, L., Gonzalez, E., Rubert, T., Smolka, U., and Melero, J. J.: Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK. Renewable and Sustainable Energy Reviews, 82:1261–1271, 2018.

213. Zou, J., Ahmed, S., and Sun, X. A.: Multistage stochastic unit commitment using stochastic dual dynamic integer programming. IEEE Transactions on Power Systems, 34(3):1814–1823, 2018.

214. Zou, J., Ahmed, S., and Sun, X. A.: Stochastic dual dynamic integer programming. Mathematical Programming, 175(1):461–502, 2019.

VITA

NAME Seyed Danial Mohseni Taheri

EDUCATION M.S., Business Analytics, University of Illinois at Chicago, Chicago, IL, US, 2021

Exchange Scholar, Northwestern University, Evanston, IL, US, 2016

Exchange Scholar, University of Chicago, Chicago, IL, US, 2016

B.S., Industrial engineering, Amirkabir University of Technology, Tehran, IR, 2015

EXPERIENCE Graduate Teaching and Research Assistant, University of Illinois at Chicago (2015-2021)

HONORS ENRE Early Career Best Research Publication Award (INFORMS 2020)

Commodity and Energy Markets Association Best Paper Award (CEMA 2020-2021)

PUBLICATIONS Trivella, A., Mohseni Taheri, D., and Nadarajah, S.: "Meeting corporate renewable power targets." (SSRN.3294724, 2019).

Mohseni-Taheri, D., Nadarajah, S., and Tulabandhula, T.: "Interpretable User Models via Decision-rule Gaussian Processes: Preliminary Results on Energy Storage." (Openreview, 2019).

Amjadi, M., Mohseni Taheri, D., Tulabandhula, T.: "KATRec: Knowledge Aware aTtentive Sequential Recommendations." (arXiv preprint arXiv:2012.03323,2020).