DATA-INTEGRATED SUPPLY CHAIN CONTRACTS: LEARNING TO PRICE UNDER UNCERTAINTY

by

Xuejun Zhao

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Management

West Lafayette, Indiana

December 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL STATEMENT OF COMMITTEE APPROVAL

Dr. William B. Haskell, Chair

Krannert School of Management

Dr. J. George Shanthikumar

Krannert School of Management

Dr. Qi Annabelle Feng

Krannert School of Management

Dr. Mengshi Lu

Krannert School of Management

Approved by:

Dr. Yanjun Li

To my family

ACKNOWLEDGMENTS

First and foremost, I would like to express my greatest thanks and sincere gratitude to my advisor, Dr. William B. Haskell, who has continuously encouraged me, guided me, and supported me during my Ph.D. study. Dr. William B. Haskell has contributed a significant amount time and spent great efforts in advising me. I feel so fortunate to be advised by him through my Ph.D. journey.

I am also deeply grateful to my other dissertation committee members. Dr. J. George Shanthikumar's courses Stochastic Models and Stochastic Dynamic Programming built a solid foundation for me to do OM research. Dr. J. George Shanthikumar never hesitated to share his brilliant ideas and offer help through my Ph.D. journey. Dr. Qi Annabelle Feng and Dr. Mengshi Lu offered great help in both my research and my Ph.D. life.

I would also like to express my deep gratitude to my coauthors, Dr. Zheng Justin Jia, Dr. Hui Zhao, and Dr. Ruihao Zhu, for their significant contribution to our research, their constant efforts to improve the quality of our work, and their invaluable suggestions for me to develop my expertise. I have learned a lot from them and gained experiences, which are important to my Ph.D. research.

I would also like to thank the other Operations Management faculty members, especially Dr. Zhan Pang, Dr. Susan Lu, Dr. Pengyi Shi, Dr. Gökçe Esenduran. I benefited a lot from the brownbag discussions led by them.

I would also like to thank my peer students including Lei Li, Jianing Li, Jian Wu, Poyraz Bozkurt and many others who have accompanied and supported me.

Last but not least, my heartfelt appreciation goes to my whole family. I can always feel their love, inspiration, bless, and support behind me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PROTOCOLS

ABSTRACT

This study examines data-driven contract design in the small data regime and large data regime respectively, and the implications from contract pricing in the pharmaceutical supply chain. We provide below a brief description of the results obtained for the specific problems considered in this study.

In the first problem discussed in Chapter 2, we study supply chain contract design under uncertainty. In this problem, the retailer has full information about the demand distribution, while the supplier only has partial information drawn from past demand realizations and contract terms. The supplier wants to optimize the contract terms, but she only has limited data on the true demand distribution. We apply a distributionally robust optimization approach. We show that the classical approach for optimizing the contract terms is fragile in the small data-driven regime by identifying several cases where it incurs a large loss. We propose a robust model for contract design where the uncertainty set requires little prior knowledge from the supplier, and effectively combines the supplier's information from past demand realizations and past interactions with the retailer. We show how to optimize the supplier's worst-case profit based on this uncertainty set. In single product case, the worst-case order quantity can be found through bisection search. In the multi-product case, we give a cutting plane algorithm for finding the worst-case order quantity and the worst-case distribution. We also demonstrate the asymptotic optimality of our uncertainty set. Our model offers a versatile framework for combining different sources of information into a single distributionally robust optimization problem. We demonstrate the advantage of our robust model by comparing it against the classical data-driven approaches. This comparison sheds light on the value of information from interactions between agents in a game-theoretic setting, and suggest that such information should not be neglected in data-driven decision-making.

In the second problem discussed in Chapter 3, we investigate the supply chain contract design problem faced by a data-driven supplier who needs to respond to the inventory decisions of the downstream retailer. Both the supplier and the retailer are uncertain about the market demand and need to learn about it sequentially. The goal for the supplier is to develop data-driven pricing policies with sublinear regret bound under a wide range of

retailer's inventory learning policies for a fixed time horizon. To capture the dynamics induced by the retailer's learning strategy, we first make a connection to nonstationary online learning by following the notion of variation budget. The variation budget quantifies the impact of the retailer's learning strategy on the supplier's decision-making environments. We then propose dynamic pricing policies for the supplier for both discrete and continuous demand. We also note that our proposed pricing policy only requires access to the support of demand distribution, but critically, does not require the supplier to have any prior knowledge about the retailer's inventory policy or the demand realizations. We examine several well known data-driven policies for the retailer, including sample average approximation, distributionally robust optimization, and parametric approaches, and show that our pricing policies lead to sublinear regret bounds under these retailer policies. At the managerial level, we answer affirmatively that there is a pricing policy with a sublinear regret bound under a wide range of retailer's learning policies, even though she faces a learning retailer and an unknown demand distribution. Our work also provides a novel perspective in data-driven operations management where the principal has to learn to react to the learning policies employed by other agents in the system.

In the third problem discussed in Chapter 4, we investigate the implications from supply chain contract pricing from the pharmaceutical supply chain. This implication uncovers the intricate relationship between drug reimbursement shortage and drug shortage problem. Generic drug shortages have been a major challenge for the U.S. pharmaceutical industry and the government, causing severe consequences and drawing widespread attention. Recognizing that the reimbursement policies affect shortages through affecting supply chain parties' contract pricing decisions, we investigate the link between drug reimbursement policies and drug shortages for generic drugs by analyzing a drug supply model that captures the elements and tradeoffs in the drug wholesale price decisions under the reimbursement policies. We find that under the current policy, the interplay of two opposing effects, the free-ride effect and the coordination effect, determines the pricing decisions and shortage occurrences. We capture key factors influencing these effects and show that the current reimbursement policy actually possesses resilience to shortages of these generic drugs. In the end we also provide data-driven solutions to the contract pricing decisions for the GPOs under uncertainty to mitigate drug shortages.

1. INTRODUCTION

Contracting is a longstanding topic in OM. In supply chain, the seller (for example, manufaturer or upstream supplier) sells products to and get payment from the buyer (usually the downstream retailer) through setting up a contract. The contract terms specify how many units the buyer is going to purchase from the seller and how much the buyer is going to pay to the seller. Contracts guarantee the proper operations inside a supply chain, and also serve as tools for the supply chain parties to leverage for their own independent interest. When the supply chain parties design contract, they may face several sources of uncertainties. For example, they have uncertainty about the market demand, they are uncertain about their upstream suppliers or downstream retailers' decisions, and they are uncertain about their competitors' decisions. In this study, we examine contract designing problem under these sources of uncertainties.

In Chapter 2, we study data-driven approaches for the supplier to price a supply chain contract under market demand uncertainty. The development of this chapter is motivated by the fact that sometimes supply chain contract data is very limited because contract terms usually last for a long time [1]. Based on this fact, we identify senarios where the classical data-driven approaches do not work well, and we propose a distributionally robust optimization approach to effectively combine different sources of information into a versatile decision making framework for the supplier.

In Chapter 3, we study data-driven approaches for the supplier to price a supply chain contract under uncertainties in its downstream retailer's inventory learning policies as well as market demand. We consider a fixed time horizon, and design dynamic pricing policies for the supplier to maintain a balance between exploration of the uncertainties and the exploitation about a potentially good price. The proposed pricing policies lead to sublinear regret bound for the supplier under a wide range of retailers' inventory learning policies.

In Chapter 4, we study a practical supply chain contracting problem originated from the pharmaceutical supply chain, where the GPOs (representatives of healthcare provides) need to make pricing decisions in purchasing drugs from drug manufacturers. While the GPOs face intense competitions from each other and they are uncertain about each other's pricing strategies, we adopt Nash equilibrium as a solution concept. Under the Nash equilbrium, we uncover the impact of the reimbursement policy change taking place in 2005 for Medicare Part B drugs on drug shortages. In the end, we also provide data-driven solutions to the contract pricing decisions for the GPOs under supply uncertainty to mitigate drug shortages.

A brief introduction of this dissertation is provided below.

1.1 Data-driven Supply Chain Contract in the Small Data Regime

We study a two tier supply chain with uncertain market demand. The supply chain is decentralized and the retailer has better knowledge of market demand than the supplier. The supplier (she) and the retailer (he) set contracts for the transfer of money and goods, and the supplier is the price setter. This situation may arise, for example, when the supplier is a major corporation or monopoly that is far from the local market, and the retailer is small but can easily track local demand. For example, Ecuador and Columbia are widely recognized for their cut flower industry. Large vertically integrated firms in these countries like Floratrading grow flowers at the production bases, and then export them to rural florists in other countries [2]. In Columbia, 40 flower producers with a size greater than 50 hectares account for 50% of exports, which are distributed through various channels to small retailers including traditional florist shops, roadside vendors, and gas stations [3].

We study the supplier's pricing problem. One salient feature of supply chain contracts is their duration - the contract terms may extend for a long time. [1] collected data on $22,039$ unique contracts, for which the mean contract duration is 40.22 months with a standard deviation of 31.65 months, so the supplier may not have much past contract data. Thus, we focus on the situation where it is time consuming or costly for the supplier to obtain both market demand data and historical contract records.

1.2 Learning to Price Supply Chain Contract against a Learning Retailer

Rapid development of big data analytics has enabled data-driven supply chain management for companies in different industries. According to a survey conducted by [4] with 212 supply chain leaders from varying sections and company sizes, around 18% of the respondents have already pivoted to big data analytics and 61% of the respondents plan to adopt big data analytics in the next 12-36 months. Big data analytics has also automated the decision-making of companies, and strengthened the agility of the upstream supply chain (e.g., suppliers) to respond to downstream (e.g., retailers and market demand) changes. Motivated by this observation, we study the supply chain contract design problem faced by a data-driven supplier that needs to respond to a downstream retailer who is uncertain about market demand and employs big data analytics tools to make inventory decisions.

We study this problem through the lens of the supplier. The supplier (she) sells a product to a retailer (he) who faces uncertain market demand over a selling horizon of T periods, where the supplier sets a wholesale price (i.e., contract) for the retailer in each period. Then, the retailer makes a decision on the order quantity accordingly, which also determines the supplier's profit. The retailer does not know the market demand distribution in advance, and may employ a data-driven inventory learning policy that is *unknown* to the supplier. The supplier does not know the market demand distribution either, and she has to sequentially balance the trade-off between exploring the retailer's response to different prices and exploiting profitable prices found so far. This situation may arise in many scenarios. For example, when selling newly introduced products, both the supplier and the retailer are uncertain about the demand of the product and thus have to learn it on the fly.

The supplier's goal is to choose the price to maximize her total profit over the selling horizon. We measure her performance through the notion of regret with respect to a clairvoyant benchmark who has the same information as the retailer (and can predict his orders) and thus chooses the optimal wholesale prices in each period. This problem is challenging due to the following two sources of uncertainty:

1. **Unknown Market Demand:** In the full information case, when both the supplier and retailer have full knowledge about the market demand distribution, the supplier can directly infer the ordering decisions from the retailer using knowledge about the market demand (assuming the retailer is profit-driven). However, when neither the supplier nor retailer has information about the market demand, the retailer has to learn the demand distribution over time, and the supplier cannot directly infer the

retailer's ordering decisions in each period without knowing the retailer's observations and inventory learning policy.

2. **Uncertain Retailer Inventory Learning Policy:** In addition, uncertainty on the retailer's inventory learning policy makes it particularly challenging to optimize the supplier's profit function, since the retailer can employ a variety of learning policies, and each policy is a mapping from the information received by the retailer to an order quantity. That is, the retailer makes inventory decisions as a response to the supplier's wholesale prices, the observed demand realizations, and his particular inventory learning policy. In this case, even if the supplier *had* known the market demand (yet the retailer still does not know it), inferring the retailer's learning policy from his ordering decisions is not an easy task.

To this end, we ask the following main question: *Does there exist a pricing policy for the supplier with a sublinear regret bound that does not require knowledge of the specific data-driven inventory learning policy used by the retailer?* If there is such a pricing policy with a sublinear regret bound, then this policy will have no optimality gap with respect to a clairvoyant benchmark's profit asymptotically.

The setting in our paper is novel as well as relevant. The data-driven newsvendor has been studied extensively in the OM literature, but the impact of a data-driven newsvendor on its upstream supplier's decisions has not yet been thoroughly investigated. We approach the supplier's problem by formulating it as a non-stationary online optimization problem. However, the non-stationarity in our supplier's problem is different than in conventional single agent non-stationary online problems. In our problem, the non-stationarity lies in the retailer's inventory decisions which depend on his inventory learning policy and the information he receives. In our case, the non-stationarity of the problem is bounded sublinearly in T, but the non-stationarity of the retailer's decisions is not necessarily bounded sublinearly in T. In addition, the supplier's online problem has a continuous decision set instead of a finite one. The literature has studied non-stationary online problems with infinitely many decisions, but with the assumption that the objective function is strongly convex or at least continuous. However, we will see that the supplier's profit function in our problem is not necessarily convex/concave or even continuous. Due to these challenges, we need a novel data-driven policy to achieve sublinear regret for the supplier.

1.3 Contracting in the Pharmaceutical Supply Chain

Drug shortages have posed a major public health threat in the U.S. in recent years. The number of drug shortages tripled from 2005 to 2010 [5] and reached 456 in 2012 [6]. In 2011, 99.5% of U.S. hospitals experienced drug shortages and nearly 50% experienced more than 20 shortages in the prior 6 months [7]. Unfortunately, the situation seems to be quite persistent. After declining from a peak around 2012, the number of ongoing drug shortages has increased again [8] and has been generally above 200 at any point of time since 2018 according to the University of Utah Drug Information Service. A recent survey [9] of nearly 300 pharmacy directors, managers, and purchasing agents shows that 55% of the respondents experienced more than 20 drug shortages in the prior 6 months.

The persistent drug shortages have led to significant losses in healthcare outcomes due to medication safety issues or errors, delays or cancellations of patient care, patients receiving a less effective drug, and even deaths [9]–[11]. They have also led to significant annual labor cost for shortage management, estimated at $216 million in 2011 [12] and increased to $360 million by 2019 [13]. As a result, the drug shortages have spurred legislation efforts in Congress, a special FDA Drug Shortage Program around 2012, and an FDA task force in 2018 [14]–[16]. Indeed, the complexity of drug shortages has made it one of the biggest challenges facing the FDA and the pharmaceutical industry:

> "*As the Congressional letters noted, drug shortages, including those that arise during emergencies, have been a persistent problem despite public and private sector efforts to prevent and mitigate them.*" [8]

FDA has been working continuously to find long-term solutions.

The current drug shortages primarily concern generic drugs [5], [6], [8], [17]–[19], which are the focus of investigation in this study.[1] Given the severity of these drug shortages, a

[1]↑The limited number of brand-name drug shortages usually have quite different causes to which our analysis does not apply.

vast number of reports and news articles attempted to identify their causes, with one opinion linking the shortages to a change in Medicare's reimbursement policy that affected many of these drugs (see, e.g., [6], [18], [20]). In the supply chain of these drugs, healthcare providers (e.g., hospitals), typically represented by group purchasing organizations (GPOs), purchase drugs from manufacturers at wholesale prices negotiated by the GPOs, and then receive reimbursement from Medicare based on the reimbursement policy. Before 2005, Medicare employed an Average Wholesale Price (AWP) policy, where the reimbursement prices were set according to published list prices. These list prices were not related to actual wholesale prices and in fact often outdated and sometimes much higher than actual wholesale prices paid. Hence, the AWP policy was jokingly called "Ain't What's Paid" [21]. In 2005, with the enactment of the Medicare Prescription Drug Improvement and Modernization Act, Medicare changed to an Average Sales Price (ASP) policy that reimburses a drug in Medicare Part B (including most shortage drugs) at 106% of the drug's actual average wholesale price from the previous two quarters. Obviously, the ASP policy is more grounded on actual drug wholesale prices.

While the ASP policy better reflects healthcare providers' payment for drugs, some scholars and analysts link it to the increasing number of drug shortages. Those holding this opinion believe that the ASP policy leads to lower drug prices, and hence discourages manufacturers from producing these drugs, contributing to their shortages [e.g., 20], [22]. Some other scholars, however, disagree with this opinion, reasoning that the policy does not directly regulate drug prices, but rather requires the reimbursement price to be 6% above the average wholesale price, which preserves the drug supply chain's flexibility to adopt any desired wholesale price [e.g., 23]. While both opinions seem reasonable and contribute to a long-standing debate, few rigorous studies have been conducted to help clear the doubts on either side [see, e.g., the related discussion in 8, p. 84]. More importantly, there has been limited research on the impact of reimbursement policies on drug pricing decisions, which is the basis for analyzing the policy impact on drug shortages. Therefore, the objective of this study is to capture the essential elements and tradeoffs in drug wholesale pricing decisions under the reimbursement policies and investigate the policy impact on drug shortages.

Throughout our research, we have actively interacted with FDA staff, who have provided valuable comments and insights, some of which will be discussed in this chapter.

1.4 Organization of the Dissertation

In Chapter 2, we propose a distributionally robust optimization model for the supplier to seamlessly integrate the information obtained from demand realizations and from historical transactions with the retailer. We detail the solution approach for our robust optimization model, and also conduct numerical analysis on synthetic and semi-synthetic data set to compare the performance of our model with classical robust optimization models.

In Chapter 3, we develop pricing policies for the supplier under discrete and continuous demand distributions respectively. We derive an upper bound on the regret, and demonstrate that the pricing policies have sublinear regret bound under a wide range of retailer's inventory policies. We also compare our pricing policies with other algorithms designed for the non-stationary bandit problem, and show that our pricing policy leads to better performance.

In Chapter 4, we formulate a Nash competition model for the GPOs facing uncertainties in their competitors' pricing decisions when they purchase drugs from the drug manufacturer. We investigate the impact of reimbursement policy change in Medicare Part B drugs on drug shortages. In the end, we also provide data-driven solutions to the contract pricing decisions for the GPOs under supply uncertainty.

Finally, we conclude our study and provide future research directions in Chapter 5.

Chapter 2 is based on [24], Chapter 3 is based on [25], and Chapter 4 is based on [26]. I would like to express my sincere appreciation to my coauthors for their invaluable contributions.

2. DATA-DRIVEN DISTRIBUTIONALLY ROBUST SUPPLY CHAIN CONTRACTS WITH SMALL DATA

2.1 Synopsis

We study a two tier supply chain with uncertain market demand. The supply chain is decentralized and the retailer has better knowledge of market demand than the supplier. The supplier (she) and the retailer (he) set contracts for the transfer of money and goods, and the supplier is the price setter. We propose a robust approach to address the supplier's uncertainty in market demand, and we emphasize the *small data* regime where the supplier's data is limited. In the small data regime, we show that some widely studied policies can perform badly, and our proposed robust approach leads to better performance. We focus our study on wholesale price contracts [27]–[29]. We summarize our main findings in the following:

1. To motivate our robust approach, we show that SAA and parametric estimation can perform poorly in the small data regime. For instance, a supplier who uses SAA is likely to set the wholesale price equal to the retailer's selling price and earn zero profit. The parametric approach can fail to distinguish between two parametric distributions that are close to each other (in the Kullback-Leibler divergence) and then incur up to 100% relative loss.

2. Motivated by the shortcomings of the classical approach, we develop a distributionally robust model for the supplier's problem [30], [31]. Like classical DRO, our uncertainty set consists of distributions that are close to a *nominal* distribution. However, we differ from classical DRO because the nominal distribution in our uncertainty set must reflect both the demand data and the contract history. Although we focus on the small data regime, we show that our model is asymptotically optimal as the supplier collects more data.

3. We optimize the supplier's worst-case profit over the uncertainty set. In the single product case, the retailer's robust optimal order quantity can be found through bisec-

tion search. In the multi-product case, the retailer's robust order quantities can be found using cutting plane algorithms.

4. In our numerical experiments, we compare our model with benchmarks based on demand data only and contract history only. We find that our robust model gives a more accurate estimate of the supplier's actual profit, and that it is generally less conservative, than the two benchmarks. We also find that the relative benefit of our model to SAA is highest when the cost/price ratio c/s is of intermediate value.

2.1.1 Organization and Notation

This paper is organized as follows. In Section 2.3, we introduce the supplier's problem in the contract design when both the supplier and retailer have full information about the demand distribution. In Section 2.4, we consider the effectiveness of classical data-driven policies, and we demonstrate that these methods can perform poorly in the small data regime. In Section 2.5 we introduce the DRO approach for the supplier's problem in the data-driven regime. Then, in Sections 2.6 and 2.7, we solve the supplier's robust problem for the single product and multi-product cases, respectively. Section 2.8 reports numerical experiments, and we conclude the paper in Section 3.8. All proofs are organized in the Appendix.

Notation

We let $\mathbb{R}$ denote the real numbers and $\mathbb{R}_+$ denote the nonnegative real numbers. The $p-$norm on $\mathbb{R}^n$ for $1 \leq p \leq \infty$ is denoted $\|\cdot\|_p$. For any $\boldsymbol{x} \in \mathbb{R}^n$, we use $\delta_{\boldsymbol{x}}$ to denote the Dirac measure at $\boldsymbol{x}$. For an integer $n \geq 1$, let $[n] \triangleq \{1, 2, \ldots, n\}$ be the running index set. Throughout this paper, we adopt the conventions of extended arithmetic, whereby $\infty \times 0 = 0 \times \infty = 0/0 = 0$ and $\infty - \infty = -\infty + \infty = 1/0 = \infty$.

Let $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ be random variables with common support $\Xi \subset \mathbb{R}^n$, let $\Pi(\boldsymbol{\xi}, \boldsymbol{\xi}')$ be their joint distribution on $\Xi \times \Xi$, and let $\Pi(\Xi, \boldsymbol{\xi}')$ and $\Pi(\boldsymbol{\xi}, \Xi)$ be the corresponding marginal distributions of $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ respectively. For probability measures $\mathbb{P}^i$ defined on measurable spaces $(\Omega_i, \mathcal{F}_i)$ for $i \in [n]$, we use $\times_{i=1}^n \mathbb{P}^i$ to denote the product measure with marginal distributions $\{\mathbb{P}^i\}_{i \in [n]}$. For any $x \in \mathbb{R}$, let $\lceil x \rceil$ denote the smallest integer that is larger than or equal to x. Let $\mathcal{U}[a, b]$ denote the uniform distribution with support on $[a, b]$, and let $\mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ denote a truncated Gaussian distribution with mean μ and covariance matrix Σ truncated to Ξ.

Throughout this chapter, we write 'max' instead of 'sup' and 'min' instead of 'inf'. When the optimal solution to the optimization problem does not exist, an "optimal solution" means an $\epsilon-$optimal solution for $\epsilon > 0$ arbitrarily small.

2.2 Literature Review

Supply chain contracts and information asymmetry

Our work is related to contract design with information asymmetry [32]–[34]. The typical approach is to model the information asymmetry as an unknown parameter, and then to propose a mechanism design framework. Compared to this approach, we model the information asymmetry through an unknown *distribution* (which is essentially an infinite-dimensional parameter). Wholesale price contracts are particularly well studied. [29] study the price of anarchy for price-only contracts. [35] find that price-only contracts perform sufficiently well under information asymmetry about the demand distribution. [28] and [36] consider contracting with moral hazard, and prove the worst-case optimality of price-only contracts.

Robust game theory and robust contracts

Robust game theory has been studied in the literature [37]–[39]. The previous literature has modeled supply chain contracts with information asymmetry as a Stackelberg game. [40] studies wholesale price contracts when either the supplier or the retailer faces demand uncertainty with fixed first and second moments. [41] study design of a robust profit sharing contract where the retailer faces random demand and random selling price with moment uncertainty. They derived the relationship between the robust wholesale prices, order quantities, and profit shares. We also study supply chain contract design, but from the data-driven perspective, and we construct the uncertainty set based on the information that is revealed through the interactions between the supplier and the retailer.

Data-driven decision-making

[42] and [43] propose operational data analytics for the newsvendor problem with parametric demand distributions. [44] provide an exact finite sample analysis for the relative expected regret of SAA, and derive an optimal policy for the newsvendor problem in this setting. [45] study the advantage of data pooling for large-scale systems with unrelated optimization problems and limited data. [46] introduce bias corrected policies for optimization problems with a linear objective and uncertain parameters, based on Stein's Lemma in the small-data large-scale regime. [47] uses the estimated gradient of the optimal objective value to debias the out-of-sample performance of a class of affine plug-in policies for optimization problems with a linear objective when data is limited. [48] study tail behavior when data is limited using robust optimization and [49] study quantile optimization when simulations are expensive.

2.3 The Supplier's Classical Problem

The supplier offers a wholesale price contract to a retailer for the sale of n products with random demand. A wholesale price contract specifies the wholesale prices $\boldsymbol{w} = (w_i)_{i\in[n]}$ for all n products. The supplier's production costs are $\boldsymbol{c} = (c_i)_{i\in[n]}$, and the retailer's selling prices are $\boldsymbol{s} = (s_i)_{i\in[n]}$. We assume $\boldsymbol{c} \leq \boldsymbol{s}$ so that it is possible for both the retailer and supplier to make a profit, and we let $\mathcal{W} \subset [\boldsymbol{c}, \boldsymbol{s}]$ be the set of admissible wholesale prices.

Let $\boldsymbol{\xi} = (\xi_i)_{i\in[n]}$ be the vector of random demand for all n products defined on an underlying measurable space $(\Omega, \mathcal{F})$, where ξ_i is the demand for product $i \in [n]$. We assume the demand has compact support.

Assumption 2.3.1. *Demand $\boldsymbol{\xi}$ has support $\Xi \triangleq [0, \bar{\xi}]^n$ for $0 \leq \bar{\xi} < \infty$.*

We let $\mathcal{P}(\Xi)$ be the set of all probability measures on Ξ, with respect to the Borel $\sigma-$algebra on Ξ. A generic demand distribution is denoted $\mathbb{P} \in \mathcal{P}(\Xi)$, with ξ_i-marginal of $\mathbb{P}$ denoted $\mathbb{P}^i \in \mathcal{P}([0, \bar{\xi}])$. Throughout this paper, we let $\bar{\mathbb{P}} \in \mathcal{P}(\Xi)$ denote the true demand distribution of $\boldsymbol{\xi}$. The (true) i^{th} marginal of $\bar{\mathbb{P}}$ is denoted $\bar{\mathbb{P}}^i \in \mathcal{P}([0, \bar{\xi}])$.

Under a wholesale price contract, when the supplier offers $\boldsymbol{w}$ and the retailer orders $\boldsymbol{q}$, the supplier's profit is $(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}$ (which does not depend on the demand realization). Let $\pi_r(\boldsymbol{q}; \boldsymbol{\xi}, \boldsymbol{w}) \triangleq \sum_{i\in[n]} (s_i \min\{\xi_i, q_i\} - w_i q_i)$ be the retailer's profit when the demand realization is $\boldsymbol{\xi}$, and let $\mathbb{E}_{\bar{\mathbb{P}}} [\pi_r(\boldsymbol{q}; \boldsymbol{\xi}, \boldsymbol{w})]$ be the retailer's expected profit. For product $i \in [n]$, given wholesale price w_i, the retailer knows the true demand distribution is $\bar{\mathbb{P}}$ and will order $q_i(w_i; \bar{\mathbb{P}}^i) \triangleq \min_{q_i \geq 0}\{q : \bar{\mathbb{P}}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i\}$. We let $\boldsymbol{q}(\boldsymbol{w}; \bar{\mathbb{P}}) \triangleq (q_i(w_i; \bar{\mathbb{P}}^i))_{i\in[n]}$ denote the vector of retailer optimal order quantities corresponding to $\boldsymbol{w}$ and $\bar{\mathbb{P}}$.

In the classical approach, the supplier also knows that the true demand distribution is $\bar{\mathbb{P}}$ and so she knows how the retailer will respond to any wholesale price $\boldsymbol{w} \in \mathcal{W}$ via the mapping $\boldsymbol{q}(\boldsymbol{w}; \bar{\mathbb{P}})$. For wholesale prices $\boldsymbol{w}$, the supplier knows her corresponding profit will be $\pi(\boldsymbol{w}; \bar{\mathbb{P}}) \triangleq (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \bar{\mathbb{P}})$, and she wants to solve:

$$\max_{\boldsymbol{w}\in\mathcal{W}} \pi(\boldsymbol{w}; \bar{\mathbb{P}}). \tag{2.1}$$

We let $\boldsymbol{w}^* \in \mathcal{W}$ denote an optimal solution to Problem (2.1).

2.4 Classical Approach in the Data-Driven Regime

In this paper, we are concerned with the supplier's problem when she has historical data but does not precisely know the demand distribution. We first ask how the classical approach performs in the data-driven setting (for $n = 1$) based on SAA and the parametric approach. We consider the relative loss of the wholesale price w given by both methods with respect to the one given by the true demand distribution $\bar{\mathbb{P}}$, which is $(\pi(w^*; \bar{\mathbb{P}}) - \pi(w; \bar{\mathbb{P}}))/\pi(w^*; \bar{\mathbb{P}})$. We show that both SAA and the parametric approach can have high relative loss, especially when the number of samples is small.

2.4.1 Sample Average Approximation

The empirical distribution corresponding to historical demand realizations $(\xi^t)_{t\in[T]}$ is $\hat{\mathbb{P}}^{\mathrm{e}}_T \triangleq \frac{1}{T}\sum_{t\in[T]} \delta_{\xi^t}$. Based on $\hat{\mathbb{P}}^{\mathrm{e}}_T$, the supplier predicts the retailer will order $q(w; \hat{\mathbb{P}}^{\mathrm{e}}_T)$ when given wholesale price w. Let $\xi^{[k]}$ denote the k^{th} order statistic of $(\xi^t)_{t\in[T]}$, then the retailer's optimal order quantity given w and $\hat{\mathbb{P}}^{\mathrm{e}}_T$ is $q(w;\hat{\mathbb{P}}^{\mathrm{e}}_T) = \xi^{[\lceil(1-w/s)T\rceil]}$ [50], [51]. The supplier's corresponding profit is $\pi(w;\hat{\mathbb{P}}^{\mathrm{e}}_T) = (w-c)\xi^{[\lceil(1-w/s)T\rceil]}$, and the supplier solves:

$$w^{SAA}_T \in \arg\max_{w\in[c,s]} \pi(w;\hat{\mathbb{P}}^{\mathrm{e}}_T). \tag{2.2}$$

We show that if the supplier makes predictions using $\hat{\mathbb{P}}^{\mathrm{e}}_T$, then she is likely to set $w = s$ (when the set of admissible wholesale prices is $[c, s]$) and not make any profit.

In our first experiment, we suppose demand satisfies $\xi \sim \mathcal{TN}(\mu, \sigma^2, \Xi)$ where $\mu = 500$, $\sigma = 150$, and $\Xi = [0, 1000]$. We let selling price $s = 50$ and production cost $c = 10$. In Figure 2.1, we compare the profit $\pi(w;\bar{\mathbb{P}})$ under the true demand distribution and the profit $\pi(w;\hat{\mathbb{P}}^{\mathrm{e}}_T)$ based on the empirical distribution for different wholesale prices. We use 500 replications (where for each replication, $T = 5$ demand samples are randomly drawn) to compute the expected value of $\pi(w;\hat{\mathbb{P}}^{\mathrm{e}}_T)$. Figure 2.1 shows the average of the supplier's profit $\pi(w;\hat{\mathbb{P}}^{\mathrm{e}}_T)$ over the 500 replications. We see that the supplier's true profit function is smooth, but her SAA-based profit is piecewise linear. Indeed, this unusual feature of the SAA-based solution is not a coincidence for this particular demand distribution. The retailer's best response function $q(w;\hat{\mathbb{P}}^{\mathrm{e}}_T)$ is piecewise constant in w. However, the supplier's profit function is bilinear in w and q. This explains why the SAA-based profit function is piecewise linear in w. Figure 2.1 also shows that if the supplier maximizes the SAA-based profit, she will eventually set the wholesale price close to $w = s$, and make almost no profit.

In our second experiment, suppose demand satisfies $\xi \sim \mathcal{U}[0, 1]$. In Figure 2.2, we plot the relative loss of w^{SAA}_T as a function of T for different c/s ratios. Still we generate 500 replications (for each replication T demand realizations are drawn). For each replication, we calculate the relative loss of w^{SAA}_T and Figure 2.2 shows the averaged relative loss over the 500 sample paths. We see that when T is small (i.e., $T \leq 5$), the relative loss of w^{SAA}_T is 100%. As in our previous experiment, the supplier has a tendency to set $w = s$ based on $\hat{\mathbb{P}}^{\mathrm{e}}_T$, particularly for T small. Furthermore, this tendency to set $w = s$ is stronger for larger c/s ratios.

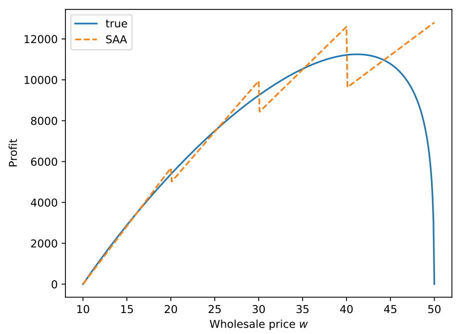


Figure 2.1. Comparison between SAA profit and true profit

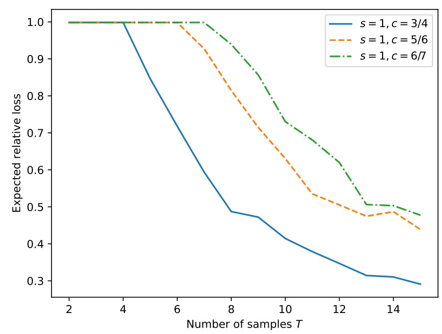


Figure 2.2. Expected relative loss of SAA

Notes: (Left) $\xi \sim \mathcal{TN}(\mu, \sigma^2, \Xi)$ where $\mu = 500$, $\sigma = 150$, and $\Xi = [0, 1000]$. (Right) $\xi \sim \mathcal{U}[0, 1]$.

2.4.2 Parametric Approach

Now suppose that demand follows either the exponential distribution $\mathbb{P}_\lambda$ (with mean $1/\lambda$ for $\lambda > 0$) or the Pareto distribution $\mathbb{P}_{(\theta,\alpha)}$ (with scale parameter $\theta > 0$ and shape parameter $\alpha > 0$). The supplier first needs to distinguish if the true demand distribution is exponential or Pareto, and then estimate the corresponding parameters. However, these two distributions can be close enough to each other that it is statistically hard to distinguish between them based on the demand samples, and the supplier can possibly choose the wrong distribution.

Under the exponential distribution, the retailer's optimal order quantity is $q(w; \mathbb{P}_\lambda) = \frac{1}{\lambda}\ln(s/w)$ [42]. In this case, the supplier's optimal wholesale price is the maximizer of $(w - c)\ln(s/w)$, which is independent of λ. Under the Pareto distribution, the retailer's optimal order quantity is $q(w; \mathbb{P}_{\theta,\alpha}) = \theta(s/w)^{1/\alpha}$ [52]. In this case, the supplier's optimal wholesale price is $w = s$ and the corresponding retailer order quantity is θ.

If the supplier sets prices based on the Pareto distribution when the true demand distribution is exponential, she will incur a 100% relative loss. Suppose the supplier wants to conduct a statistical test to see if the true distribution is exponential or Pareto based on demand samples $(\xi^t)_{t\in[T]}$. Table 2.1 shows the probability that the supplier fails to distinguish between the two as a function of T (we explain how to compute these probabilities in Appendix A.2.1). From Table 2.1, we see that when T is small (e.g., $T \le 10$), the probability of miss-specifying the distribution is close to 0.5, which is not much better than just flipping a coin.

Table 2.1. Probability of failing to distinguish between $\mathbb{P}_\lambda$ and $\mathbb{P}_{(\theta,\alpha)}$ ($\lambda = 0.1$, $\theta = 300$, $\alpha = 30$)

T	2	5	10	20	50	100
Probability	0.483	0.473	0.462	0.446	0.415	0.380

2.5 Robust Approach in the Data-Driven Regime

In practice, the retailer may get close access to the market (e.g., by launching marketing campaigns or learning from consumer sales data) and know the true demand distribution.

In contrast, the supplier may be far from the local market and so may experience uncertainty about the demand distribution. Then, the case where the retailer knows the demand distribution $\bar{\mathbb{P}}$, but the supplier has only partial demand information, is important.

The previous section demonstrated that the classical approach may not perform well in the data-driven setting because it does not account for the inherent estimation uncertainty in the small data regime. In this section, we propose a robust approach that explicitly accounts for this uncertainty, and additionally exploits the information from the retailer to help mitigate this uncertainty.

2.5.1 Uncertainty Set

The main ingredient to the robust approach is an *uncertainty set*, based on the supplier's data, that she believes contains all reasonable candidates for the demand distribution. Let w_i^t, q_i^t, and ξ_i^t be the wholesale price, corresponding retailer order quantity, and demand realization for product $i \in [n]$, respectively, in period t. The supplier's data consist of T historical data points $\{(\boldsymbol{w}^t, \boldsymbol{q}^t, \boldsymbol{\xi}^t)\}_{t\in[T]}$ where $\boldsymbol{w}^t = (w_i^t)_{i\in[n]}$ are the wholesale prices, $\boldsymbol{q}^t = (q_i^t)_{i\in[n]}$ are the retailer order quantities, and $\boldsymbol{\xi}^t = (\xi_i^t)_{i\in[n]}$ is the demand realization for period $t \in [T]$.

First we identify those distributions that are compatible with the past retailer data. The retailer has full knowledge of the demand distribution and will always order optimally. So, for each $t \in [T]$, we know that $\boldsymbol{q}^t$ has to be a maximizer of the retailer's expected profit corresponding to $\boldsymbol{w}^t$. We need only consider demand distributions $\mathbb{P}$ for which this is the case, i.e., $\boldsymbol{q}^t \in \arg\max_{\boldsymbol{q}\geq 0} \mathbb{E}_{\mathbb{P}}[\pi_r(\boldsymbol{q}, \boldsymbol{\xi}; \boldsymbol{w}^t)]$, which is a condition on $\mathbb{P}$. We then define $\mathcal{D}_T^o \triangleq \{\mathbb{P} \in \mathcal{P}(\Xi) : \boldsymbol{q}^t \in \arg\max_{\boldsymbol{q}\geq 0} \mathbb{E}_{\mathbb{P}}[\pi_r(\boldsymbol{q}, \boldsymbol{\xi}; \boldsymbol{w}^t)],\ \forall t \in [T]\}$ to be the set of demand distributions compatible with all past retailer orders.

We can characterize $\mathcal{D}_T^o$ explicitly using the retailer's first-order optimality conditions. We get simpler conditions under the following assumption (which is only a condition on the true demand distribution $\bar{\mathbb{P}}$, not on all admissible distributions).

Assumption 2.5.1. *The true demand distribution $\bar{\mathbb{P}}$ is continuous (i.e., there are no atoms).*

This assumption is not strictly necessary, since if $\bar{\mathbb{P}}$ is discrete, one can derive the first-order necessary and sufficient conditions using the perturbation method, see, e.g., [53, Theorem 3]. We impose this assumption just to simplify the exposition.

Under Assumption 2.5.1, the necessary and sufficient conditions for the retailer's optimal order quantities are:

$$\mathbb{P}^i(\xi_i \leq q_i) = 1 - w_i/s_i,\ \forall i \in [n]. \tag{2.3}$$

When $\boldsymbol{q}$ and $\boldsymbol{w}$ are fixed, Eq. (2.3) is a condition on $\mathbb{P}$, and we can write

$$\mathcal{D}_T^o = \left\{\mathbb{P} \in \mathcal{P}(\Xi) : \mathbb{P}^i(\xi_i \leq q_i^t) = 1 - w_i^t/s_i, \forall i \in [n], t \in [T]\right\}.$$

In particular, $\mathcal{D}_T^o$ is the intersection of the probability simplex $\mathcal{P}(\Xi)$ with linear constraints (and so it is convex).

We now choose a center distribution $\hat{\mathbb{P}}_T \in \mathcal{P}(\Xi)$, which represents our *nominal* belief about the true distribution. Our uncertainty set will consist of all distributions in $\mathcal{D}_T^o$ that are close enough to $\hat{\mathbb{P}}_T$, so we need a notion of distance on $\mathcal{P}(\Xi)$. We will use the Wasserstein distance.

Definition 2.5.1. *(Wasserstein distance, [54]) Let $C(\mathbb{P}, \mathbb{P}')$ denote the set of all joint distributions Π on $\Xi \times \Xi$ with marginals $\mathbb{P}$ and $\mathbb{P}'$, respectively. The Wasserstein distance $W_2 : \mathcal{P}(\Xi) \times \mathcal{P}(\Xi) \to \mathbb{R}_+$ with respect to the $2-$norm is:*

$$W_2(\mathbb{P}, \mathbb{P}') \triangleq \left(\inf_{\Pi \in C(\mathbb{P}, \mathbb{P}')} \int_{\boldsymbol{\xi}, \boldsymbol{\xi}' \in \Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2^2 \Pi(d\boldsymbol{\xi}, d\boldsymbol{\xi}')\right)^{1/2}. \tag{2.4}$$

We require $\hat{\mathbb{P}}_T$ to reflect both the retailer and the demand information. The retailer information is captured by $\mathcal{D}_T^o$, and the demand information is captured by the empirical distribution $\hat{\mathbb{P}}_T^{\mathrm{e}} \triangleq \frac{1}{T}\sum_{t\in[T]} \delta_{\boldsymbol{\xi}^t}$. To synthesize this information we take the projection of $\hat{\mathbb{P}}_T^{\mathrm{e}}$ onto $\mathcal{D}_T^o$:

$$\hat{\mathbb{P}}_T \in \arg\min_{\mathbb{P} \in \mathcal{D}_T^o} W_2^2(\mathbb{P}, \hat{\mathbb{P}}_T^{\mathrm{e}}). \tag{2.5}$$

Problem (2.5) is a convex optimization problem since W_2 is a metric and $\mathcal{D}_T^o$ is a convex set.

To complete the construction of our uncertainty set, we need a confidence parameter $\theta \geq 0$ which controls its size. When we are very confident, we would choose a small θ corresponding to a smaller uncertainty set, while when we lack confidence, we would choose a large θ corresponding to a larger one. The confidence parameter θ can be chosen using cross validation [e.g., 55], [56], see Appendix A.3.1.

Finally, we define our uncertainty set to be $\mathcal{D}_\theta(\hat{\mathbb{P}}_T) = \{\mathbb{P} \in \mathcal{D}_T^o : W_2(\mathbb{P}, \hat{\mathbb{P}}_T) \leq \theta\}$. The retailer data enters through $\mathcal{D}_T^o$, and the demand and retailer data both enter through $\hat{\mathbb{P}}_T$. We also note that $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$ is convex by convexity of W_2 (as a metric) and the fact that $\mathcal{D}_T^o$ is convex.

2.5.2 Supplier's Robust Problem

In our robust approach, the supplier's problem is to maximize the worst-case profit among admissible distributions in $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$. For fixed wholesale prices $\boldsymbol{w}$, the supplier's worst-case profit is

$$\pi_{T,\theta}(\boldsymbol{w}) \triangleq \min_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)} (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}). \tag{2.6}$$

Unlike classical DRO problems, the unknown distribution $\mathbb{P}$ does not directly enter the objective here. Rather, it indirectly influences the supplier's objective through the retailer's order quantity $\boldsymbol{q}(\boldsymbol{w}; \mathbb{P})$. When the minimum in Eq. (2.6) is attained by some $\mathbb{P}_{w,\theta} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)$, the supplier's worst-case profit is precisely $\pi_{T,\theta}(\boldsymbol{w}) = (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}_{w,\theta})$, i.e., the supplier's profit when the demand distribution is $\mathbb{P}_{w,\theta}$. In this case, we refer to $\mathbb{P}_{w,\theta}$ as the *worst-case distribution* corresponding to $\boldsymbol{w}$. In Theorem 2.6.1 (Section 2.6) and Theorem 2.7.2 (Section 2.7), we establish the existence of $\mathbb{P}_{w,\theta}$ (when θ is small enough) and show how to compute it.

The supplier's corresponding robust problem is:

$$\max_{\boldsymbol{w} \in \mathcal{W}} \min_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)} (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}). \tag{2.7}$$

In practice, the wholesale prices may be taken from a menu, and the monetary values themselves are discrete (e.g., $\$1, \$2, \ldots$), so we suppose $\mathcal{W}$ is finite. When $\mathcal{W}$ is finite (and moderately sized), we can solve Problem (2.7) by iterating over $\boldsymbol{w} \in \mathcal{W}$.

When $\mathcal{W}$ is continuous, it is known that even for $n = 1$ in the full information setting, the supplier's profit function is not necessarily concave and is difficult to optimize. If the demand distribution has the increasing generalized failure rate (IGFR) property, then the supplier's profit function is unimodal [27]. In our case, the worst-case distribution does not necessarily enjoy the IGFR property, so the worst-case profit is not guaranteed to be unimodal. [57] studies moment problems which require the worst-case distribution to have IGFR, but it is unclear how to incorporate their idea into our framework. Furthermore, we allow $n \geq 1$, so even unimodality in any single w_i does not guarantee tractability of the supplier's robust problem. When $\mathcal{W}$ is continuous, Problem (2.7) becomes a bilevel program (see Appendix A.3.2). In general, a deterministic bilevel program with linear upper and lower level objectives and constraints is NP-hard [58]. Thus, finding the robust optimal solution in our model when $\mathcal{W}$ is continuous can be rather challenging.

2.5.3 Reformulations of the Supplier's Robust Problem

Problem (2.7) emphasizes uncertainty from the perspective of the demand distribution. We can also emphasize uncertainty from the perspective of the retailer order quantities, which leads to two useful reformulations of Problem (2.7). These reformulations provide different interpretations of the distributionally robust Stackelberg game. In addition, they facilitate the computation of the worst-case profits. Through Reformulation (2.9), the worst-case profit for $n = 1$ can be found by bisection search. Through Reformulation (2.8), the worst-case profit can be found by a cutting plane algorithm (see Section 2.7).

Given $\boldsymbol{w}$, we define the set $\mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T) \triangleq \{\boldsymbol{q} : \exists \mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T) : \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i,\ \forall i \in [n]\}$ of retailer order quantities. The set $\mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ includes all retailer order quantities that are optimal for $\boldsymbol{w}$, for which $\mathbb{P}^i(\xi_i \leq q_i) = 1 - w_i/s_i$ holds for all $i \in [n]$ as per the retailer's

FOC (2.3). Since each $\mathbb{P}^i(\xi_i \le q_i)$ is non-decreasing in q_i, $\mathcal{Q}_\theta(\boldsymbol{w};\, \hat{\mathbb{P}}_T)$ includes all larger order quantities as well. We can then reformulate the worst-case profit equivalently as:

$$\min_{\boldsymbol{q} \in \mathcal{Q}_\theta\left(\boldsymbol{w};\hat{\mathbb{P}}_T\right)} (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}, \tag{2.8}$$

where we optimize over the retailer's order quantities. When the minimum is attained in Eq. (2.8) by some $\boldsymbol{q}(\boldsymbol{w}, \theta) \ge 0$, the supplier's worst-case profit is $\pi_{T,\theta}(\boldsymbol{w}) = (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}, \theta)$. We refer to $\boldsymbol{q}(\boldsymbol{w}, \theta)$ as the *worst-case retailer order quantity* corresponding to $\boldsymbol{w}$. Eq. (2.8) constructs an uncertainty set $\mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ for the possible retailer order quantities, and the worst-case profit is w.r.t. the uncertainty in retailer orders. The literature has also studied robust Stackelberg game when the uncertainty is w.r.t. the lower level decisions, for example, see [59], [60].

We can also model the uncertainty with respect to *both* the unknown demand distribution *and* the retailer's order quantities, which are connected by the FOC conditions. For given $\boldsymbol{w}$, we obtain:

$$\pi_{T,\theta}(\boldsymbol{w}) = \min_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T),\, \boldsymbol{q} \ge 0} \left\{ (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \mathbb{P}^i(\xi_i \le q_i) \ge 1 - w_i/s_i,\ \forall i \in [n] \right\}. \tag{2.9}$$

In Eq. (2.9), the supplier has uncertainty over the demand distribution as well as the retailer's order quantity.

The corresponding reformulations of the supplier's robust problem are:

$$\max_{\boldsymbol{w} \in \mathcal{W}} \min_{\boldsymbol{q} \ge 0} \left\{ (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \exists \mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T) \text{ s.t. } \mathbb{P}^i(\xi_i \le q_i) \ge 1 - w_i/s_i,\ \forall i \in [n] \right\}. \tag{2.10}$$

$$\max_{\boldsymbol{w} \in \mathcal{W}} \min_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T),\, \boldsymbol{q} \ge 0} \left\{ (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \mathbb{P}^i(\xi_i \le q_i) \ge 1 - w_i/s_i,\ \forall i \in [n] \right\}. \tag{2.11}$$

We connect all three formulations of the supplier's robust problem in the following proposition.

Proposition 2.5.1. *The optimal values of Problems (2.7), (2.10), and (2.11) are equal. Furthermore, if all problems attain their optimal values, then the sets of optimal wholesale prices of Problems (2.7), (2.10), and (2.11) are equal.*

Proposition 2.5.1 says that the three formulations of the supplier's robust problem are equivalent. Additionally, the worst-case profit can be obtained either from the worst-case distribution through Eq. (2.6), from the worst-case order quantity through Eq. (2.8), or from Eq. (2.9) which computes both.

2.5.4 Asymptotics

For $T \geq 1$ data points, we solve Problem (2.7) with optimal value $\pi_{T,\theta} = \max_{\boldsymbol{w} \in \mathcal{W}} \pi_{T,\theta}(\boldsymbol{w})$ and optimal solution $\boldsymbol{w}^*_{T,\theta} \in \arg\max_{\boldsymbol{w} \in \mathcal{W}} \pi_{T,\theta}(\boldsymbol{w})$. We will show that when θ is allowed to depend on $T \geq 1$ appropriately, $\pi_{T,\theta}$ converges almost surely to π^* and $\boldsymbol{w}^*_{T,\theta}$ has a subsequence that converges almost surely to an optimal solution of Problem (2.1) as $T \to \infty$.

We let $\alpha_T \in (0,1)$ be a pre-specified confidence level for all $T \geq 1$. Corresponding to each $T \geq 1$ and α_T, we define the confidence parameter:

$$\theta_T(\alpha_T) \triangleq \begin{cases} 2\left(\frac{\log(c_1 \alpha_T^{-1})}{c_2 T}\right)^{1/\max\{n,2\}}, & T \geq \frac{\log(c_1 \alpha_T^{-1})}{c_2}, \\ 2\left(\frac{\log(c_1 \alpha_T^{-1})}{c_2 T}\right), & T < \frac{\log(c_1 \alpha_T^{-1})}{c_2}. \end{cases} \tag{2.12}$$

The following theorem summarizes the asymptotic optimality of our model when the confidence parameters are chosen according to Eq. (2.12).

Theorem 2.5.1. *Suppose Assumptions 2.3.1 and 2.5.1 hold, and let $\{\alpha_T\}_{T\geq 1}$ be a sequence of confidence levels $\alpha_T \in (0,1)$ such that $\sum_{T=1}^{\infty} \alpha_T < \infty$ and $\lim_{T\to\infty} \theta_T(\alpha_T) = 0$.*

(i) $\pi_{T,\theta_T(\alpha_T)} \uparrow \pi^*$ as $T \to \infty$, $\bar{\mathbb{P}}^{\infty}-$almost surely.

(ii) Any accumulation point of $\{\boldsymbol{w}_{T,\theta_T(\alpha_T)}\}_{T\geq 1}$ is an optimal solution for Problem (2.1), $\bar{\mathbb{P}}^{\infty}-$almost surely.

2.6 Single Product Case

We develop the details of the single product case ($n = 1$) in this section. First we show how to construct the center and formulate our uncertainty set. Then, we show how to compute the worst-case profit $\pi_{T,\theta}(w)$ for given w.

When $n = 1$, WLOG we can arrange the past wholesale prices in descending order as $s \geq w^1 > \cdots > w^t > w^{t+1} > \cdots > w^T$ (where we assume that there are no repeats). Then, the retailer's corresponding past order quantities in ascending order are $0 \leq q^1 < \cdots < q^t < q^{t+1} < \cdots < q^T$.

We now explain how to solve Problem (2.5) when $n = 1$ to construct the center. While Problem (2.5) is infinite-dimensional (it is optimizing over $\mathcal{P}(\Xi)$), we can reformulate it as a tractable finite-dimensional convex optimization problem. The past orders $(q^t)_{t\in[T]}$ partition Ξ into $T+1$ subsets $[0, q^1]$, $(q^1, q^2], \ldots, (q^{T-1}, q^T]$, $(q^T, \bar{\xi}]$, that we index with $s = 1, \ldots, T+1$. We will show that the center has the form $\hat{\mathbb{P}}_T = \sum_{t=1}^{T} \sum_{s=1}^{T+1} \beta_{st} \delta_{p_{st}/\beta_{st}}$, and Problem (2.5) can be reformulated as:

$$
\begin{aligned}
\min_{\boldsymbol{\beta},\boldsymbol{p}} \quad & \sum_{t=1}^{T} \sum_{s=1}^{T+1} \beta_{st} \left(p_{st}/\beta_{st} - \xi^t\right)^2 && \text{(2.13a)}\\
\text{s.t.} \quad & \sum_{s=1}^{T+1} \beta_{st} = 1/T, \forall t \in [T], && \text{(2.13b)}\\
& \sum_{t=1}^{T} \sum_{s=1}^{t'} \beta_{st} = 1 - \frac{w^{t'}}{s}, \forall t' \in [T], && \text{(2.13c)}\\
& 0 \leq p_{1t} \leq q^1 \beta_{1t}, \forall t \in [T], && \text{(2.13d)}\\
& q^{s-1} \beta_{st} < p_{st} \leq q^s \beta_{st}, \forall 2 \leq s \leq T, t \in [T], && \text{(2.13e)}\\
& q^T \beta_{T+1,t} < p_{T+1,t} \leq \bar{\xi}\, \beta_{T+1,t}, \forall t \in [T], && \text{(2.13f)}\\
& \beta_{st} \geq 0, \forall s \in [T+1], t \in [T], && \text{(2.13g)}
\end{aligned}
$$

for $\boldsymbol{p} = (p_{st})_{s\in[T+1],t\in[T]}$ and $\boldsymbol{\beta} = (\beta_{st})_{s\in[T+1],t\in[T]}$. Problem (2.13) is convex, and the only nonlinearity appears in the objective function. We let $(\boldsymbol{\beta}^*, \boldsymbol{p}^*)$ denote an optimal solution of Problem (2.13). We then take the center to be

$$
\hat{\mathbb{P}}_T = \sum_{t=1}^{T} \sum_{s=1}^{T+1} \beta^*_{st} \delta_{p^*_{st}/\beta^*_{st}}. \tag{2.14}
$$

As shown in the following Proposition 2.6.1, the center has support on p^*_{st}/β^*_{st} for $s \in [T+1]$ with probability mass β^*_{st}, and each support point p^*_{st}/β^*_{st} falls into one of the subsets partitioned by $(q^t)_{t\in[T]}$ (see Eqs. (2.13d)-(2.13f)). The quantity β^*_{st} is the probability mass transported from ξ^t to p^*_{st}/β^*_{st} so that the total probability mass transported from ξ^t sums up to $1/T$ (see Eq. (2.13b)) and the first-order conditions are satisfied (see Eq. (2.13c)).

Proposition 2.6.1. *Suppose $n = 1$, and let $(\boldsymbol{\beta}^*, \boldsymbol{p}^*)$ be an optimal solution of Problem (2.13). Then, the optimal values of Problems (2.5) and (2.13) are equal, and $\hat{\mathbb{P}}_T$ as constructed in Eq. (2.14) is an optimal solution of Problem (2.5).*

Our uncertainty set is then: $\mathcal{D}_\theta(\hat{\mathbb{P}}_T) = \{\mathbb{P} \in \mathcal{P}(\Xi) : W_2(\mathbb{P}, \hat{\mathbb{P}}_T) \le \theta \text{ and } \mathbb{P}(\xi \le q^t) = 1 - w^t/s : t \in [T]\}$. Given w and θ, the supplier's worst-case profit for $n = 1$ is

$$\pi_{T,\theta}(w) = \min_{\mathbb{P}\in\mathcal{D}_\theta(\hat{\mathbb{P}}_T), q\ge 0} \{(w-c)q : \mathbb{P}(\xi \le q) \ge 1 - w/s\}. \tag{2.15}$$

In Eq. (2.15), we let $\mathbb{P}_{w,\theta}$ denote the worst-case demand distribution and $q(w,\theta)$ denote the worst-case retailer order quantity. The supplier's worst-case profit is then $\pi_{T,\theta}(w) = (w-c)q(w,\theta)$.

For each wholesale price w, we define a threshold $\bar{\theta}(w)$ on the parameter θ (see Eq. (A.9) in Appendix A.4). Based on the threshold $\bar{\theta}(w)$, when $\theta \ge \bar{\theta}(w)$, the worst-case order quantity and distribution do not exist, but we can still characterize the worst-case profit (which is not attained). Alternatively, when $\theta < \bar{\theta}(w)$, the worst-case order quantity and distribution exist and have closed form.

Theorem 2.6.1. *Fix $c \le w \le s$.*

(i) If $\theta \ge \bar{\theta}(w)$, then $\pi_{T,\theta}(w) = (w-c)q(w)$ where $q(w)$ is given in Eq. (2.16):

$$q(w) = \arg\max_{q\in(q^t)_{t\in[T]}} \left\{q : \bar{\mathbb{P}}(\xi \le q) < 1 - w/s\right\} \tag{2.16}$$

(ii) If $\theta < \bar{\theta}(w)$, then $\pi_{T,\theta}(w) = (w-c)q(w,\theta)$ where $q(w,\theta)$ is the unique solution to Eq. (A.10). In this case, $\mathbb{P}_{w,\theta}$ is given by Eq. (A.12) (see Appendix A.4).

Theorem 2.6.1 shows that the worst-case profit $\pi_{T,\theta}(w)$ is constant for $\theta \ge \bar{\theta}(w)$ ($q(w)$ is constant as θ increases). This indicates that even as θ becomes large, the worst-case order quantity is still lower bounded by $q(w)$. As mentioned before, θ reflects the supplier's confidence that the center $\hat{\mathbb{P}}_T$ is a good model, and larger θ corresponds to lower supplier confidence. The supplier's worst-case profit $\pi_{T,\theta}(w)$ is minimized over θ by choosing $\theta = \bar{\theta}(w)$. Thus, she can restrict her choice of confidence parameter to $\theta \in [0, \bar{\theta}(w)]$, since no additional robustness is obtained from larger $\theta > \bar{\theta}(w)$. For this reason, we call $\bar{\theta}(w)$ the *maximum level of uncertainty.*

When $\theta < \bar{\theta}(w)$, $q(w, \theta)$ can be efficiently computed using bisection search (see Eq. (A.10) in Appendix A.4). In the classical problem with known demand distribution $\bar{\mathbb{P}}$, the retailer's order quantity $q(w; \bar{\mathbb{P}})$ for any w is also computed using bisection search on the CDF of $\bar{\mathbb{P}}$. This indicates finding the worst-case order quantity $q(w, \theta)$ is no more difficult than finding $q(w; \bar{\mathbb{P}})$ under full information.

2.7 Multi-Product Case

Now we consider the multi-product case. In the supplier's classical problem with full information, the supplier can choose optimal wholesale prices separately for each product. However, the uncertainty set in our robust model is based on the joint demand distribution which couples the robust sub-problems for each product. We first allow dependent demand and propose a cutting plane algorithm for computing the worst-case profit. Then we consider the special case of independent demand, and we show that the center and worst-case profit have special structure and can be computed more easily.

2.7.1 Dependent Demand

In the multi-product setting, we will continue to work with centers of the form $\hat{\mathbb{P}}_T = \sum_{t=1}^{T}\sum_{s=1}^{S}\beta_{st}\delta_{\boldsymbol{p}_{st}/\beta_{st}}$. Problem (2.5) can then be reformulated as:

$$\min_{\boldsymbol{\beta}\geq 0,\boldsymbol{p}} \quad \sum_{t=1}^{T}\sum_{s=1}^{S}\beta_{st}\left\|\boldsymbol{p}_{st}/\beta_{st}-\boldsymbol{\xi}^t\right\|_2^2 \tag{2.17a}$$

$$\text{s.t.} \quad \sum_{s=1}^{S}\beta_{st}=1/T, \forall t\in[T], \tag{2.17b}$$

$$\sum_{t=1}^{T}\sum_{s=1}^{S}\beta_{st}z_{is}(\boldsymbol{q}^{t'})=\frac{s_i-w_i^{t'}}{s_i}, \forall i\in[n], t'\in[T], \tag{2.17c}$$

$$A_s\boldsymbol{p}_{st}\leq\beta_{st}\boldsymbol{d}_s, \forall s\in[S], t\in[T]. \tag{2.17d}$$

We give the detailed derivation of Problem (2.17) in Appendix A.5.1. We denote the optimal solutions of this problem by $\beta^*=(\beta^*_{st})_{s\in[S],t\in[T]}$ and $\boldsymbol{p}^*=(\boldsymbol{p}^*_{st})_{s\in[S],t\in[T]}$, and we construct the center:

$$\hat{\mathbb{P}}_T=\sum_{t=1}^{T}\sum_{s=1}^{S}\beta^*_{st}\delta_{\boldsymbol{p}^*_{st}/\beta^*_{st}}. \tag{2.18}$$

Proposition 2.7.1. *The optimal value of Problem (2.17) is equal to the optimal value of Problem (2.5). Furthermore, $\hat{\mathbb{P}}_T$ as constructed in Eq. (2.18) is an optimal solution of Problem (2.5).*

We propose a cutting plane algorithm (see Appendix A.5.2) based on Eq. (2.8) (for the worst-case order quantity) to compute the supplier's worst-case profit. Roughly speaking, we initialize with $\mathcal{Q}_\theta(\boldsymbol{w};\hat{\mathbb{P}}_T)=\Xi$. We then gradually prune order quantities in Ξ that are either sub-optimal or do not belong to $\mathcal{Q}_\theta(\boldsymbol{w};\hat{\mathbb{P}}_T)$. That is, each time we pick an order $\boldsymbol{q}$ that has not yet been pruned, and evaluate whether $\boldsymbol{q}\in\mathcal{Q}_\theta(\boldsymbol{w};\hat{\mathbb{P}}_T)$ by solving a convex optimization problem. If $\boldsymbol{q}\in\mathcal{Q}_\theta(\boldsymbol{w};\hat{\mathbb{P}}_T)$, then orders that yield higher profit than $\boldsymbol{q}$ cannot be optimal and are pruned. On the other hand, if $\boldsymbol{q}\notin\mathcal{Q}_\theta(\boldsymbol{w};\hat{\mathbb{P}}_T)$, then we can also discard a set of order quantities that do not belong to $\mathcal{Q}_\theta(\boldsymbol{w};\hat{\mathbb{P}}_T)$ according to the FOC (2.3).

Theorem 2.7.1. *For any $\epsilon>0$, Algorithm 4 will converge to an $\epsilon-$optimal solution to Eq. (2.8) after a finite number of iterations.*

2.7.2 Independent Demand

Now we consider the special case where the true demand distribution $\bar{\mathbb{P}}$ has independent marginal distributions. In this case the worst-case profit becomes a resource allocation problem, where the resource is the Wasserstein budget.

Assumption 2.7.1. *The true distributions of all products i and $j \neq i$ are independent, i.e., $\bar{\mathbb{P}} = \times_{i=1}^{n} \bar{\mathbb{P}}^i$, where $\bar{\mathbb{P}}^i \in \mathcal{P}([0, \bar{\xi}])$ is the marginal distribution for product $i \in [n]$.*

Assumption 2.7.1 is reasonable when all pairs of products are neither substitutes nor compliments, since then the demand for one product is generally not affected by the demand for the others. We introduce the set $\mathcal{D}_\times \triangleq \{\mathbb{P} \in \mathcal{P}(\Xi) : \mathbb{P} = \times_{i\in[n]}\mathbb{P}^i, \mathbb{P}^i \in \mathcal{P}([0, \bar{\xi}]), \forall i \in [n]\}$ of demand distributions on Ξ with independent marginals.

Under Assumption 2.7.1, we modify $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$ to restrict to distributions satisfying $\mathbb{P} \in \mathcal{D}_\times$. For each product $i \in [n]$, let $\hat{\mathbb{P}}_T^{\mathrm{e},i} \triangleq \frac{1}{T}\sum_{t\in[T]} \delta_{\xi_i^t}$ be the corresponding empirical marginal distribution.

We now modify the construction of the center so that $\hat{\mathbb{P}}_T \in \mathcal{D}_\times$ will hold. Notice that $\hat{\mathbb{P}}_T^{\mathrm{e}}$ does not necessarily belong to $\mathcal{D}_\times$, but $\hat{\mathbb{Q}}_T^{\mathrm{e}} \triangleq \times_{i\in[n]}\hat{\mathbb{P}}_T^{\mathrm{e},i}$ automatically belongs to $\mathcal{D}_\times$. We construct the center by projecting $\hat{\mathbb{Q}}_T^{\mathrm{e}}$ onto $\mathcal{D}_T^o \cap \mathcal{D}_\times$, instead of $\hat{\mathbb{P}}_T^{\mathrm{e}}$. We obtain the following projection problem:

$$\min_{\mathbb{P}\in\mathcal{P}(\Xi)\cap\mathcal{D}_\times} \left\{ W_2^2(\mathbb{P}, \hat{\mathbb{Q}}_T^{\mathrm{e}}) \,\text{s.t.}\, \mathbb{P}^i\left(\xi_i \le q_i^t\right) = 1 - w_i^t/s_i, \forall i \in [n], t \in [T] \right\}. \tag{2.19}$$

Problem (2.19) finds the distribution in $\mathcal{D}_T^o \cap \mathcal{D}_\times$ that has the smallest 2−Wasserstein distance to $\hat{\mathbb{Q}}_T^{\mathrm{e}}$. We let $\hat{\mathbb{Q}}_T$ denote the center found using Problem (2.19), to distinguish from $\hat{\mathbb{P}}_T$ (which is for the general case). The set $\mathcal{D}_\times$ is non-convex in $\mathbb{P}$ so Problem (2.19) is a non-convex optimization problem. However, we can decompose it into separate convex optimization problems for each $i \in [n]$ to solve it efficiently.

Proposition 2.7.2. *(i) Problem (2.19) decomposes into separate linear programming problems in $\mathbb{P}^i$ for each $i \in [n]$ (see Problem (2.20)):*

$$\min_{\mathbb{P}^i\in\mathcal{P}([0,\bar{\xi}])} \left\{ W_2^2\left(\mathbb{P}^i, \hat{\mathbb{P}}_T^{\mathrm{e},i}\right) \quad s.t. \quad \mathbb{P}^i\left(\xi_i \le q_i^t\right) = 1 - w_i^t/s_i, \forall t \in [T] \right\}. \tag{2.20}$$

(ii) Let $\hat{\mathbb{Q}}_{T,i}$ be an optimal solution of Problem (2.20) for each $i \in [n]$, then $\hat{\mathbb{Q}}_T = \times_{i\in[n]}\hat{\mathbb{Q}}_{T,i}$ is an optimal solution of Problem (2.19).

The modified uncertainty set $\mathcal{D}_{\theta,\times}(\hat{\mathbb{Q}}_T) \triangleq \{\mathbb{P} \in \mathcal{D}_T^o \cap \mathcal{D}_\times : W_2(\mathbb{P}, \hat{\mathbb{Q}}_T) \leq \theta\}$ requires admissible distributions to have independent marginals. We want to compute the corresponding overall worst-case profit $\pi_{T,\theta}(\boldsymbol{w})$. This computation depends on how much of the total uncertainty budget θ^2 is allocated to each product. For each $i \in [n]$, let $r_i \geq 0$ be the uncertainty budget allocated to product i, then we define the corresponding uncertainty sets $\mathcal{D}^i_{r_i}(\hat{\mathbb{Q}}_{T,i}) \triangleq \left\{\mathbb{P} \in \mathcal{P}([0,\bar{\xi}]) : W_2(\mathbb{P}, \hat{\mathbb{Q}}_{T,i}) \leq r_i : \mathbb{P}\left(\xi \leq q_i^t\right) = 1 - w_i^t/s, \forall t \in [T]\right\}$ for each product $i \in [n]$. The worst-case profit for product i, given budget r_i, is then:

$$\pi^i_{T,\sqrt{r_i}}(w_i) = \min_{\mathbb{P}^i \in \mathcal{D}^i_{\sqrt{r_i}}(\hat{\mathbb{Q}}_{T,i}),\, q_i \geq 0} \left\{(w_i - c_i)q_i \quad \text{s.t.} \quad \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i\right\}. \tag{2.21}$$

Notice that Eq. (2.21) matches the form of the single product case.

The following proposition characterizes the dependence of the worst-case profit on r_i.

Proposition 2.7.3. *For any w_i, $\pi^i_{T,\sqrt{r_i}}(w_i)$ is decreasing and convex in r_i for all $i \in [n]$.*

In total we have θ^2 uncertainty budget to allocate to all products. According to Proposition 2.7.3, the worst-case profit for each product i is decreasing in r_i. Therefore, finding the worst-case profit $\pi_{T,\theta}(\boldsymbol{w})$ is equivalent to finding an allocation plan $\boldsymbol{r} = (r_i)_{i=1}^n$ to minimize the total cost:

$$\pi_{T,\theta}(\boldsymbol{w}) = \min_{\boldsymbol{r} \geq 0} \left\{ \sum_{i\in[n]} \pi^i_{T,\sqrt{r_i}}(w_i) \quad \text{s.t.} \quad \sum_{i\in[n]} r_i \leq \theta^2 \right\}. \tag{2.22}$$

Problem (2.22) determines how much of the uncertainty budget θ^2 to allocate to each product. By the same reasoning as Theorem 2.6.1, $\pi^i_{T,\sqrt{r_i}}(w_i)$ is constant for any $r_i \geq \bar{\theta}_i^2(w_i)$ and so we can restrict to $r_i \leq \bar{\theta}_i^2(w_i)$ when solving Problem (2.22).

For any $\boldsymbol{w}$, Problem (2.22) is convex by Proposition 2.7.3, and can be solved by the analytic center cutting plane method (see Appendix A.5.6). We utilize the results from the single product case to derive the subgradient for the objective and generate specific cuts (see Proposition A.5.3).

We can use the special form of Problem (2.22) to identify the worst-case order quantities and worst-case distribution. That is, we first find the optimal allocation plan $\boldsymbol{r}^* = (r_i^*)_{i\in[n]}$ which solves Problem (2.22). Then, under $\boldsymbol{r}^*$, we derive the worst-case order quantity $q^i(w_i, \sqrt{r_i^*})$ and the worst-case distribution $\mathbb{P}^i_{w_i,\sqrt{r_i^*}}$ for product i as in the single product case.

Theorem 2.7.2. *Fix $\boldsymbol{w} \in \mathcal{W}$.*

(i) Problem (2.22) has an optimal solution $\boldsymbol{r}^* = (r_i^*)_{i\in[n]}$.

(ii) Let $(\mathbb{P}^i_{w_i,\sqrt{r_i^*}}, q^i(w_i, \sqrt{r_i^*}))$ be the worst-case distribution and order quantity corresponding to $\pi^i_{T,\sqrt{r_i^*}}(w_i)$ for all $i \in [n]$. Then, the worst-case distribution is $\mathbb{P}_{\boldsymbol{w},\theta} = \times_{i\in[n]}\mathbb{P}^i_{w_i,\sqrt{r_i^*}}$ and the worst-case profit is $\pi_{T,\theta}(\boldsymbol{w}) = \sum_{i\in[n]}(w_i - c_i)q^i(w_i, \sqrt{r_i^*})$.

2.8 Numerical Experiments

In this section, we explore the behavior of our robust model numerically on both simulated and real data sets. We pose three questions: (i) Does our robust model give a less conservative estimate of the true optimal profit compared with benchmark robust models based on retailer information only and demand information only? (ii) Will the supplier achieve higher actual profit if she uses our model instead of the benchmark models? (iii) How does the performance of our model compare with SAA as a function of the profit margin c/s?

Recall $\pi_{T,\theta}(\boldsymbol{w})$ is the worst-case profit based on $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$ (see Eq. (2.6)), and $\boldsymbol{w}^*_{T,\theta} \in \arg\max_{\boldsymbol{w}\in\mathcal{W}} \pi_{T,\theta}(\boldsymbol{w})$ are the optimal wholesale prices for our robust model. We now formally introduce the benchmarks. We first define the uncertainty set $\mathcal{D}^{\mathrm{e}}_\theta(\hat{\mathbb{P}}^{\mathrm{e}}_T) \triangleq \{\mathbb{P} \in \mathcal{P}(\Xi) : W_2(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_T) \le \theta\}$ corresponding to demand data only. Given $\boldsymbol{w} \in \mathcal{W}$, the worst-case profit corresponding to $\mathcal{D}^{\mathrm{e}}_\theta(\hat{\mathbb{P}}^{\mathrm{e}}_T)$ is

$$\pi^{\mathrm{e}}_{T,\theta}(\boldsymbol{w}) \triangleq \min_{\mathbb{P}\in\mathcal{D}^{\mathrm{e}}_\theta(\hat{\mathbb{P}}^{\mathrm{e}}_T),\boldsymbol{q}\ge 0} \left\{(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \mathbb{P}^i(\xi_i \le q_i) \ge 1 - w_i/s_i, \forall i \in [n]\right\},$$

and the supplier's corresponding DRO problem is $\pi^{\mathrm{e}}_{T,\theta} \triangleq \max_{\boldsymbol{w}\in\mathcal{W}} \pi^{\mathrm{e}}_{T,\theta}(\boldsymbol{w})$ with optimal wholesale prices $\boldsymbol{w}^{\mathrm{e}*}_{T,\theta} \in \arg\max_{\boldsymbol{w}\in\mathcal{W}} \pi^{\mathrm{e}}_{T,\theta}(\boldsymbol{w})$ (when they exist).

The set $\mathcal{D}_T^o$ captures all the supplier's information about the retailer. Given $\boldsymbol{w}$, the worst-case profit corresponding to $\mathcal{D}_T^o$ is

$$\pi_T^o(\boldsymbol{w}) \triangleq \min_{\mathbb{P}\in\mathcal{D}_T^o, \boldsymbol{q}\geq 0} \left\{ (\boldsymbol{w}-\boldsymbol{c})^\top \boldsymbol{q} : \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i, \, \forall i \in [n] \right\}.$$

The supplier's DRO problem based only on $\mathcal{D}_T^o$ is $\pi_T^o \triangleq \max_{\boldsymbol{w}\in\mathcal{W}} \pi_T^o(\boldsymbol{w})$ with optimal wholesale prices $\boldsymbol{w}_T^{o*} \in \arg\max_{\boldsymbol{w}\in\mathcal{W}} \pi_T^o(\boldsymbol{w})$ (when they exist). We first give an example showing that the uncertainty set based only on the retailer data does not necessarily enjoy asymptotic convergence.

(2.23) Suppose $n = 1$ and suppose the true demand distribution on $\Xi = [0, 1]$ is uniform so $\bar{\mathbb{P}}(\xi \leq x) = x$ for all $x \in [0, 1]$. Let $s = 1$ and $c = 1/2$, then the supplier's true optimal profit is $\pi^* = 1/16$. Suppose the past wholesale prices and retailer order quantities are $\{(w^t, q^t)\}_{t=1}^T = \{1 - t/(8T), t/(8T))\}_{t=1}^T$ for some $T \in \mathbb{N}$. Then, the worst-case profit is $\pi_T^o = 3/64$ for all $T \in \mathbb{N}$.

Example 2.23 shows that when the past wholesale prices are concentrated in certain regions (in this example, $7/8 \leq w^t \leq 1$ for all $t \in [T]$), then there may be a constant bias between $\lim_{T\to\infty} \pi_T^o$ and π^*. We infer that when one has no control over the data generating process, it is necessary to incorporate the demand data into the construction of the uncertainty set to correct this bias.

2.8.1 Simulated Data Set

Conservatism of the uncertainty sets

We first confirm that our model leads to less conservative estimates of the actual profit than the benchmarks. To show this, we fix $\boldsymbol{w}$ and compare the worst-case profit obtained from the different models to the true profit. We measure the degree of conservatism in terms of the absolute relative difference of the worst-case profit with respect to the true profit $\pi(\boldsymbol{w}; \bar{\mathbb{P}})$ via $(\pi(\boldsymbol{w}; \bar{\mathbb{P}}) - \pi_{T,\theta}(\boldsymbol{w}))/\pi(\boldsymbol{w}; \bar{\mathbb{P}})$, $(\pi(\boldsymbol{w}; \bar{\mathbb{P}}) - \pi_{T,\theta}^{\mathrm{e}}(\boldsymbol{w}))/\pi(\boldsymbol{w}; \bar{\mathbb{P}})$, and $(\pi(\boldsymbol{w}; \bar{\mathbb{P}}) - \pi_T^o(\boldsymbol{w}))/\pi(\boldsymbol{w}; \bar{\mathbb{P}})$. We use 100 replications to compute the averaged absolute relative difference. For each replication, we randomly draw T samples of past wholesale prices, order quantities, and demand realizations. The past order quantities are obtained for the corresponding wholesale prices under full information. Figures 2.3 - 2.7 compare the average absolute relative difference for different θ as T varies. Figures 2.3 - 2.5 report the results for $n = 1$, and Figures 2.6 - 2.7 report the results for $n = 2$ taking into account the asymmetry between the products.

As shown in Figures 2.3 - 2.7, our model leads to smaller absolute relative differences, especially for small T (e.g., $T \leq 20$). At the same time, all three models result in smaller absolute relative differences as T increases. The model based on demand information only is a possible exception. Here, when θ is large, the absolute relative difference first deceases when T increases to 10 but then stays constant as T increases further. In addition, the performance for the order only model depends heavily on how the past wholesale prices are distributed. For example, in Figures 2.3 - 2.5 parts (a)-(b), the order only model has a small relative difference. However, in Figures 2.3 parts (c)-(d), we see the order only model has a large relative difference. This difference in the performance of the order only model comes from the difference in the distribution of past wholesale prices.

Performance of the uncertainty sets

We now compare the actual profit earned (with respect to the true demand distribution) when the wholesale price is chosen according to each model. The confidence parameter θ is chosen by three-fold cross validation (see Section A.3.1). Figures 2.8 - 2.12 compare the actual profit for $\boldsymbol{w}^*_{T,\theta}$, $\boldsymbol{w}^{e*}_{T,\theta}$, and $\boldsymbol{w}^{o*}_T$ based on the average profit obtained from 100 replications. Here, again, each replication consists of T randomly drawn samples of past wholesale prices and demand realizations. We compare both the average robust optimal profits (left panel), as well as the standard deviation of the optimal profits (right panel) over these 100 replications.

In the first place, as T increases, the robust optimal profit increases and the standard deviation decreases for all three models. This is consistent with our intuition that as T increases, the supplier obtains more accurate and complete information and can thus make

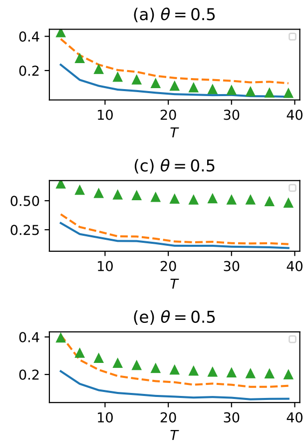


Figure 2.3. Absolute relative difference of both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 1$).

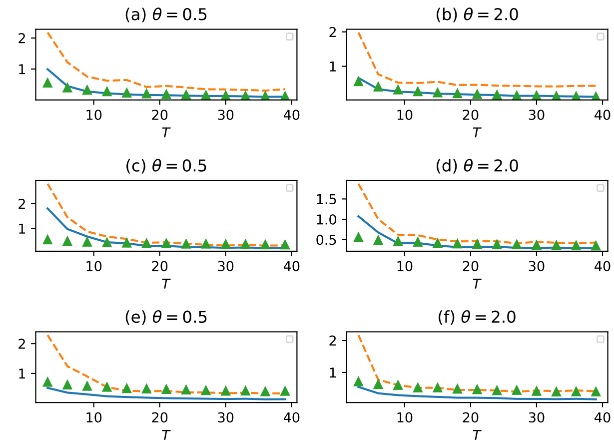


Figure 2.4. Absolute relative difference of both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 1$).

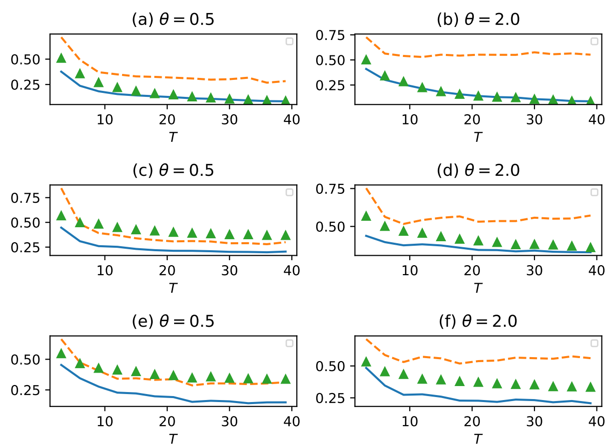


Figure 2.5. Absolute relative difference of both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 1$).

Notes: (Figure 2.3) $\xi \sim \mathcal{TN}(50, 20^2, \Xi)$ where $\Xi = [0, 100]$. (Figure 2.4) True profit $\pi(w; \bar{\mathbb{P}})$ is two modal. (Figure 2.5) $\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$. Past wholesale prices w have different modes in the subfigures.

better pricing decisions. As T increases, the robust optimal wholesale prices not only give higher actual profit, but also have less variation.

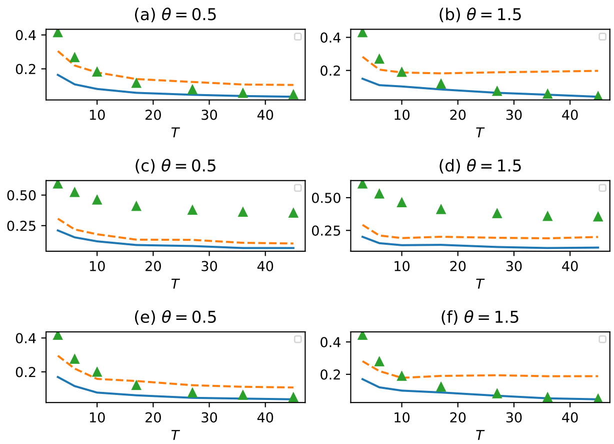


Figure 2.6. Absolute relative difference of both (blue solid line), demand (orange dashed line) and order (green triangle) $(n = 2)$.

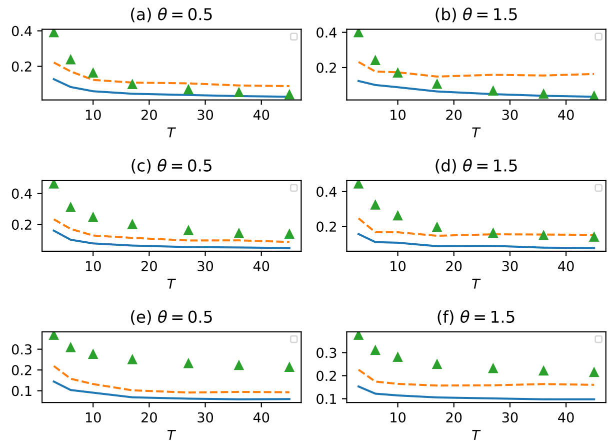


Figure 2.7. Absolute relative difference of both (blue solid line), demand (orange dashed line) and order (green triangle) $(n = 2)$.

Notes: (Figure 2.6) $\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100] \times [0, 100]$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$. (Figure 2.7) $\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100] \times [0, 100]$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$. The data for the subfigures of each figure are generated so that they have different asymmetries in mean demand level, profit margin, and past wholesale price distribution.

For these experiments, we have to simulate the past wholesale prices. Through comparing Figures 2.8 - 2.10 when the past wholesale prices are simulated following different distributions, we see that the performance of $\boldsymbol{w}_T^{o*}$ varies. For example, in Figures 2.8 (a)-(b), (e)-(f), Figures 2.9 (a)-(b), (c)-(d), and Figures 2.10 (a)-(b), (e)-(f), $\boldsymbol{w}_T^{o*}$ has superior or comparable performance (higher average actual profit and lower standard deviation) with respect to $\boldsymbol{w}_{T,\theta}^*$. On the other hand, in Figures 2.8 (c)-(d), Figures 2.9 (e)-(f), and Figures 2.10 (c)-(d), $\boldsymbol{w}_T^{o*}$ performs worst compared to the other two models. Recall that the model based only on order information can perform poorly asymptotically, depending on the distribution of past wholesale prices (see Example 2.23). When the past wholesale prices are generated uniformly, or when they are concentrated around the optimal wholesale price (this is the case for Figures 2.8 (a)-(b), (e)-(f), Figures 2.9 (a)-(b), (c)-(d), and Figures 2.10 (a)-(b), (e)-(f)), $\boldsymbol{w}_T^{o*}$ can have good performance.

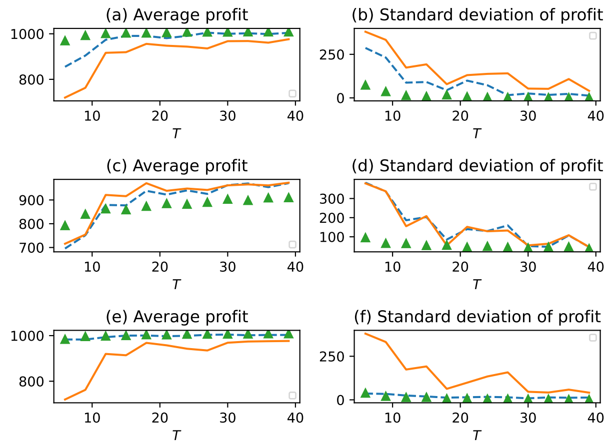


Figure 2.8. Average and standard deviation of profit for both (blue dashed line), demand (orange solid line) and order (green triangle) ($n = 1$).

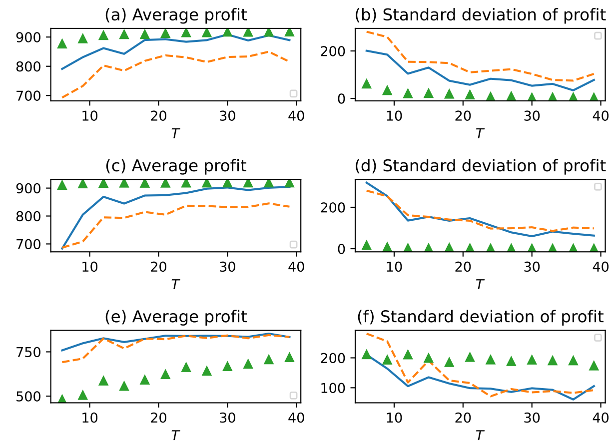


Figure 2.9. Average and standard deviation of profit for both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 1$).

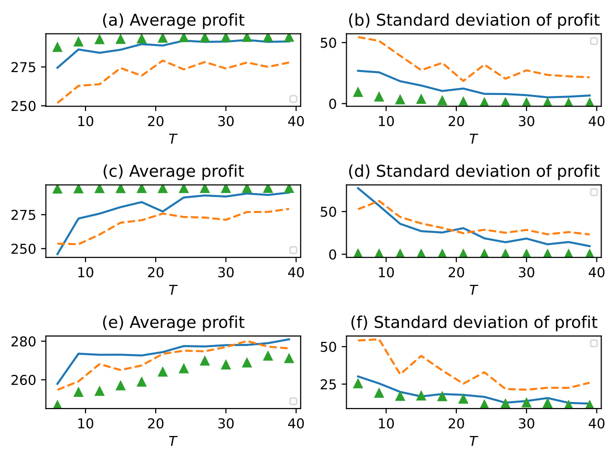


Figure 2.10. Average and standard deviation of profit for both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 1$).

Notes: (Figure 2.8) $\xi \sim \mathcal{TN}(50, 20^2, \Xi)$ where $\Xi = [0, 100]$. (Figure 2.9) True profit $\pi(w; \bar{\mathbb{P}})$ is two modal. (a) and (b) $w \sim \mathcal{U}[c, s]$. (Figure 2.10) $\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$. Past wholesale prices w have different modes in the subfigures.

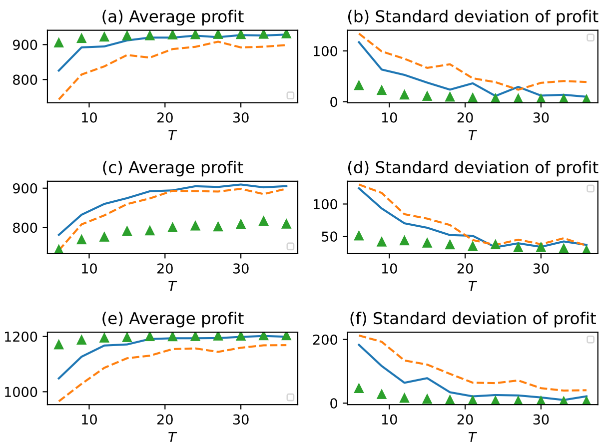


Figure 2.11. Average and standard deviation of profit for both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 2$).

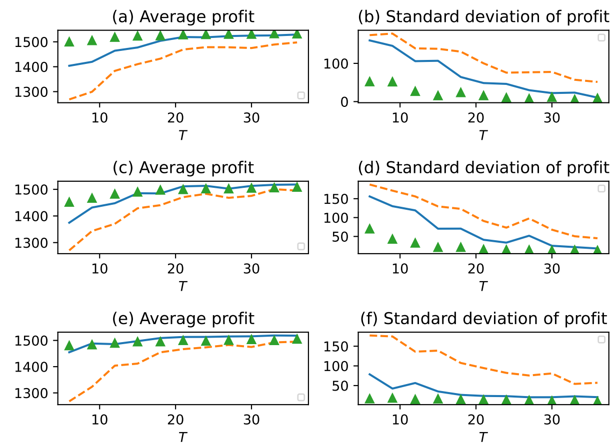


Figure 2.12. Average and standard deviation of profit for both (blue solid line), demand (orange dashed line) and order (green triangle) ($n = 2$).

Notes: (Figure 2.11) $\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100] \times [0, 100]$ and $\Sigma = \left(\begin{smallmatrix} 20^2 & 0 \\ 0 & 20^2 \end{smallmatrix}\right)$. (Figure 2.12) $\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100] \times [0, 100]$ and $\Sigma = \left(\begin{smallmatrix} 20^2 & 0 \\ 0 & 20^2 \end{smallmatrix}\right)$. The data for the subfigures of each figure are generated so that they have different asymmetries in mean demand level, profit margin, and past wholesale price distribution.

Given that the supplier does not know if the past wholesale prices are concentrated around the optimal wholesale price, we suggest they first conduct a goodness of fit test (for example, the Kolmogorov-Smirnov test) to see whether the past wholesale prices are generated uniformly. If there is insufficient evidence to reject the hypothesis that the past wholesale prices are uniformly distributed, the supplier can set wholesale prices $\boldsymbol{w}_T^{o*}$. However, if such evidence is found, the supplier is routed to use our main model.

2.8.2 Performance on Semi-synthetic Data Set

In this subsection, we evaluate the performance of our model on a data set of demand for dairy products obtained from the Economic Research Service of the U.S. Department of Agriculture (USDA). This data set contains the U.S. monthly domestic consumption (in millions of pounds) of dairy products from January 1995 to December 2010, for a total of

192 months. We focus on the dry skim dairy product. To simulate the retailer's ordering decisions, we suppose the retailer optimizes his profit based on the empirical demand distribution constructed from the demand data of the whole data set. Then we bootstrap T samples from the data set and take them to be the demand realizations seen by the retailer and the supplier. The past wholesale prices are generated randomly.

We compare the performance of $w^*_{T,\theta}$ with w^{o*}_T and w^{e*}_T as T increases (see the average profit and standard deviation calculated based on the bootstrap samples in Figures 2.13 - 2.14). In these experiments, we choose the confidence parameters by three-fold cross validation. By testing on this real demand data set, we get further confirmation that the performance of w^{o*}_T depends on the distribution of the past wholesale prices (see Figures 2.13-2.14). As can be seen in Figure 2.13, when past wholesale prices are generated uniformly, w^{o*}_T has the highest average profit, compared with the other two models. In contrast, in Figure 2.14, w^{o*}_T has an average profit that is biased below the other two models.

Now let $\boldsymbol{w}^{SAA*}_T \in \arg\max_{\boldsymbol{w}\in\mathcal{W}} \pi(\boldsymbol{w}; \hat{\mathbb{P}}^{e}_T)$ be the optimal wholesale price based on SAA. In Figure 2.15, we compare $w^*_{T,\theta}$ to w^{SAA*}_T and w^{o*}_T in terms of the relative differences with respect to the actual profit: $(\pi(w^*_{T,\theta}; \bar{\mathbb{P}}) - \pi(w^{SAA*}_T; \bar{\mathbb{P}}))/\pi(w^{SAA*}_T; \bar{\mathbb{P}})$ and $(\pi(w^*_{T,\theta}; \bar{\mathbb{P}}) - \pi(w^{o*}_T; \bar{\mathbb{P}}))/\pi(w^{o*}_T; \bar{\mathbb{P}})$, as the cost/price ratio c/s varies (see the relative difference in Figure 2.15).

In Figure 2.15, we see that the relative performance of $w^*_{T,\theta}$ with respect to w^{SAA*}_T first increases as c/s increases, and then it decreases as c/s increases further. We suspect that this is because as c/s first increases, the performance of SAA deteriorates as it has a tendency to choose an overly high wholesale price (see the discussion in Section 2.4). So, the relative benefit of setting $w^*_{T,\theta}$ at first increases. However, as c increases further, the relative benefit of $w^*_{T,\theta}$ decreases as the value of combining order information decreases (the relative performance of $w^*_{T,\theta}$ w.r.t. w^{o*}_T increases as c/s increase when c/s is large enough). In summary, we find that the relative benefit of our model is the highest when c/s is of intermediate value.

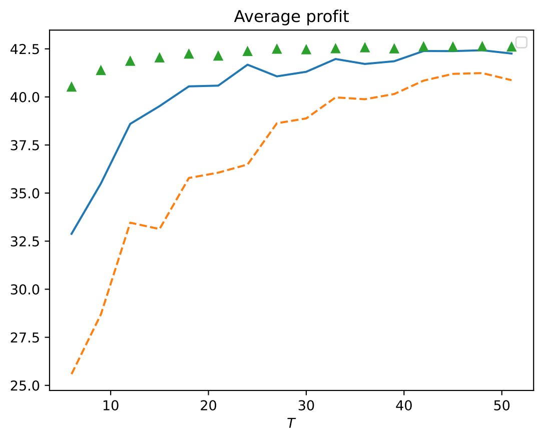


Figure 2.13. Average profit of both (blue solid line), demand (orange dashed line) and order (green triangle).

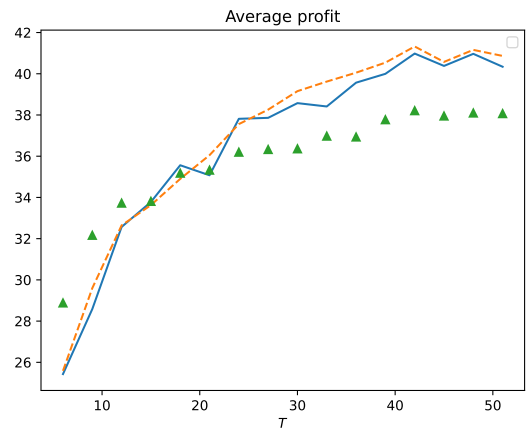


Figure 2.14. Average profit of both (blue solid line), demand (orange dashed line) and order (green triangle).

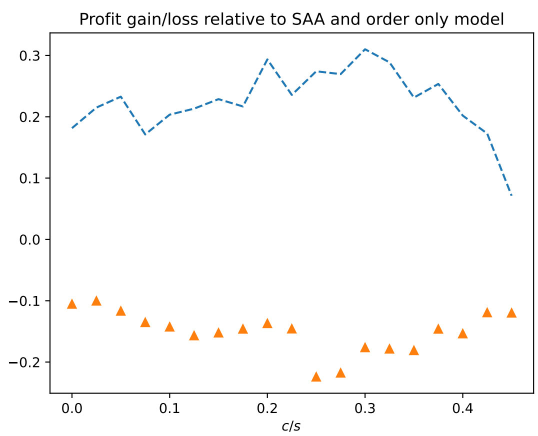


Figure 2.15. Relative profit to SAA (blue dashed line) and order only model (orange triangle).

Notes: (Figure 2.13) $w \sim \mathcal{U}[c,s]$ and $c/s = 1/2$. (Figure 2.14) $w \sim \mathcal{TN}(c + 0.2(s-c), 0.04(s-c)^2, [c,s])$ and $c/s = 1/2$. (Figure 2.15) $w \sim \mathcal{U}[c,s]$.

2.9 Conclusion

In this paper, we have proposed a new model for data-driven supply chain contract design, when the retailer has better knowledge about the demand distribution than the supplier. We found that classical data-driven approaches such as SAA and parametric estimation may perform poorly in this context. Instead, we proposed a distributionally robust optimization approach where the uncertainty set incorporates both past demand information and past retailer ordering information. Our uncertainty set does not require any prior knowledge about the demand distribution from the supplier, and is more informative than those based on demand information alone. We derived a closed form expression for the supplier's worst-case profit for the single product case, and designed a cutting plane algorithm to find the worst-case profit in the multi-product case. We compared its performance with benchmark uncertainty sets numerically, and found that it is less conservative than the benchmark based only on demand information.

This work suggests that in a game-theoretic setting, interactions between agents can reveal important information about the underlying uncertainty (in addition to observations of the uncertainty itself). Our framework combines different sources of information into a single distributionally robust optimization problem. In the contract design problem, we have shown that by incorporating the retailer order data into the construction of the uncertainty set (along with demand information), we improve performance over the classical DRO model based on demand information alone.

Our framework can be extended to other types of contracts, e.g., buy-back contracts. Supply chain coordination can be achieved if the buy-back price is appropriately chosen with full knowledge of the demand distribution. When the supplier has incomplete knowledge about the demand distribution, she can employ our method to learn about the demand distribution through interactions with the retailer. In this case, the supplier can observe the retailer's order quantity as well as the number of unsold units, since the supplier has to buy back unsold units from the retailer. The supplier can then directly infer the demand realization in each period and set the buy-back price to achieve better performance.

At the same time, we acknowledge some limitations of this work and directions for future research. First, it is worth investigating the supplier's problem when the retailer also only has partial information about the demand. Second, we leave development of algorithms for the supplier's robust problem when $\mathcal{W}$ is continuous to future work.

3. LEARNING TO PRICE SUPPLY CHAIN CONTRACTS AGAINST A LEARNING RETAILER

3.1 Synopsis

We study a supplier (she) selling a product to a retailer (he) who faces uncertain market demand over a selling horizon of T periods, where the supplier sets a wholesale price (i.e., contract) for the retailer in each period. Then, the retailer makes a decision on the order quantity accordingly, which also determines the supplier's profit. The retailer does not know the market demand distribution in advance, and may employ a data-driven inventory learning policy that is *unknown* to the supplier. The supplier does not know the market demand distribution either, and she has to sequentially balance the trade-off between exploring the retailer's response to different prices and exploiting profitable prices found so far.

The supplier's goal is to choose the price to maximize her total profit over the selling horizon. We measure her performance through the notion of regret with respect to a clairvoyant benchmark who has the same information as the retailer (and can predict his orders) and thus chooses the optimal wholesale prices in each period.

We provide an affirmative answer that there exist pricing policies for the supplier facing a learning retailer. We emphasize that our policy does not require the supplier to have any knowledge on the past demand realizations or the retailer's inventory policy. Instead, she only uses her past interactions with the retailer, and knowledge of the support of the demand distribution.

We propose the supplier's policies for both discrete and continuous demand distributions. When the demand distribution is discrete, the supplier's profit function admits a special structure that our policy exploits. When the demand distribution is continuous, this special structure vanishes, but we give a policy that approximates the supplier's profit function and still attains sublinear regret. We summarize the major results for this chapter as follows:

1. Note that the retailer's ordering decisions depend on the supplier's wholesale prices and also on the data-driven inventory learning policies used, this can create non-stationarity in the supplier's decision environment. To capture this effect, we follow the notion of

variation budget in the non-stationary bandit (see Section 3.2) to quantify the difficulty of the supplier's learning problem. With that, even if the retailer switches policies dynamically and/or use a mixture of them, we can encapsulate the impact through the variation budget. Different than prior literature on non-stationary bandits [61]–[63], we define the variation budget in terms of the Kolmogorov distance between the distributions that determine the retailer's inventory decisions. Here, the use of Kolmogorov distance turns out to be natural as it conveniently translates the variation on retailer's inventory decisions to the variation on the supplier's profit functions, and enables the development of pricing policies with provable regret bound for our setting. We also remark that Kolmogorov distance can be upper bounded by many other commonly used distance metrics or divergences, e.g., total variation distance, relative entropy, Helinger distance, Wasserstein distance, etc. [64].

2. We propose a pricing policy π_{LUNA} for the supplier that achieves sublinear regret when the market demand distribution is discrete. In this case, the supplier's profit function is discontinuous and non-stationary. In spite of this, we identify special structure in the supplier's profit function to resolve the challenge. We emphasize that our policy does not require any knowledge of the variation budget or the retailer's inventory learning policy. Instead, our policy automatically adjusts to a wide range of retailer policies and variation budgets.

3. When the market demand distribution is continuous, the unique structure in the supplier's profit function vanishes and one cannot directly apply π_{LUNA}. To overcome this challenge, we work on an approximation of the supplier's profit function. At a high level, our policy π_{LUNAC} for continuous demand is based on an approximate profit function for the supplier which inherits the desired structure. Then, our previous policy π_{LUNA} for discrete demand can be employed as a sub-routine for π_{LUNAC}.

4. We show that our proposed pricing policy leads to sublinear regret bounds for the supplier under a wide range of retailer inventory policies. We examine: (i) sample average approximation (SAA); (ii) distributionally robust optimization (DRO); and

(iii) some parametric approaches (maximum likelihood estimation (MLE), operational statistics, and Bayesian estimation). Under these policies, we compute the respective variation budgets and derive the corresponding regret bounds.

5. We also conduct numerical experiments to compare our pricing policy with several algorithms from the literature on non-stationary bandits, including the Exp3.S algorithm by [61], the deterministic non-stationary bandit algorithm proposed by [65], and the Master+UCB1 algorithm proposed by [66] where each price is treated as an arm to pull. We show that our pricing policy has the best performance among all these benchmarks. Our results demonstrate the importance of exploiting structural properties in data-driven operations.

6. At the managerial level, we establish that there is an asymptotically optimal policy for the supplier even though she faces a learning retailer and an unknown (possibly non-stationary) demand distribution More generally, our work shows the importance of data-driven operations management where the principal has to learn to react to the learning policies employed by other agents in the system. These results also further support the use of wholesale price contracts in practice.

3.1.1 Organization

This work is organized as follows. In Section 3.3, we introduce the problem formulation which consists of the supplier's dynamic pricing problem and the class of retailer inventory learning policies. In Section 3.4, we present a preliminary analysis of the regret of the supplier's pricing policy when the retailer has full knowledge about the demand distribution. Then, in Section 3.5 we develop the supplier's pricing policy and its regret upper bound under a learning retailer. We first develop the pricing policy for discrete distributions and then extend it to continuous distributions. In Section 3.6, we study several examples of retailer inventory policies under which our pricing policy achieves sublinear regret. In Section 3.7 we conduct numerical experiments and we conclude the paper in Section 3.8.

3.2 Literature Review

Contract Design under Uncertainty and MAB: Supply chain contract design is a longstanding topic, we refer to the survey by [67]. In particular, there is an increasing interest in studying contract design under uncertainty [28], [41]. We consider the design of wholesale price contracts. There have been efforts in the literature to justify the prevalence of wholesale price contracts in practice [28], [29], [35]. They suggest that wholesale price contracts are arguably the most natural form of contract for us to investigate when faced with a learning retailer.

Our work lies at the interface between contract design and multi-armed bandit (MAB) problems (see [68] and [69]). MAB problems have also been extensively studied. In particular, they have been used to model contract design problems. For example, [70] study repeated principle-agent interactions where the principle offers a contract to induce the efforts of i.i.d. arriving agents.

Dynamic Pricing and Inventory Control: Dynamic pricing and online revenue management has been studied widely in the OM literature [71]–[79]. Also see [80] for an overview of studies on dynamic pricing. More recently, a line of works also integrate inventory control into pricing decisions (see, e.g., [81], [82] and references therein). In this stream of literature, the decision maker is unknown about the demand function, and has to balance the trade-off of learning and earning while dynamically adjusting the pricing and/or inventory decisions.

Almost all the above works focus exclusively on the stationary demand environment, but in our case, due to the retailer's learning strategy, the decision environment could be dynamically changing. In this regime, [77], [83], [84] study dynamic pricing in a non-stationary environment. [62] study the online non-stationary newsvendor problem when the $L_2 - norm$ of the variation in mean demand is bounded. [77] study a dynamic joint inventory and pricing problem with perishable products where the price-demand relationship is piecewise stationary. They derive regret bound of $\tilde{O}(T^{2/3}(\log(T))^{1/2})$ for nonparametric noise distributions and $\tilde{O}(T^{1/2}(\log(T)))$ for parametric noise distributions, respectively.

Unlike the previous studies whose goals are to learn the unkown demand functions, the learning in our problem is with respect to the retailer's data-driven inventory learning policies. Furthermore, the non-stationarity in our problem is mostly driven by the learning policies of the self-interested retailer.

Non-stationary Online Learning: Many bandit problems are inherently non-stationary. One approach is to model the non-stationarity as a drifting environment, where some metric is used to measure the variation of the environments. The regret analysis is done by restricting to environments with bounded variation [63], [65], [85]–[90]. Different metrics have been considered, which result in different regret bounds. [85] study a K-armed bandit problem where the mean reward of the arms is changing. They derive a near-optimal policy with regret $\tilde{O}((KV)^{1/3}T^{2/3})$ when the *supremum norm* of the change in mean rewards is bounded by a known variation budget V. [86] study non-stationary stochastic optimization problems where the cost function is convex and the *supremum norm* of the deviations in the cost function in each period is bounded. [91] extends the previous work to use the $L_{p,q}-$*variational functional*, which better reflects local spatial and temporal changes in the objective cost functions. These works mostly require the DM to know the variation budget. In order to relax this requirement, [65] propose a restarting algorithm for the $K-$armed bandit problem that restarts whenever a large variation in the environment has been detected by a statistical test.

We build our supplier pricing policy based on the deterministic bandit setting in [65]. Their algorithm is epoch-based where each epoch consists of an exploration and an exploitation phase. In the exploration phase, the algorithm samples from each arm once and observes the noiseless bandit reward. In the exploitation phase, the algorithm randomly selects an arm to sample, where the sampling distribution is calibrated to balance the trade-off between exploration and exploitation. If the variation of the sampled arm is detected to be above some detection threshold $\mathcal{OV}_B$, then the algorithm starts the next epoch. Otherwise, the algorithm continues the exploitation phase. This algorithm relaxes the assumption that the DM knows the variation budget by sequentially decreasing the detection threshold $\mathcal{OV}_B$ in the exploitation phase.

In another approach, one can model non-stationarity in a piecewise fashion where the bandit remains stationary in each interval and varies across intervals. The total number of intervals is bounded by S, but the start and end time of each interval is unknown to the

DM. Some algorithms have been proposed for known S [89], [92]–[95] and unknown S [65], [77], [89], [96]–[99]. We note the difference between this approach for non-stationarity and the first one based on a variation budget. In the first approach, only a constraint on the total variation is imposed and the total number of intervals (where the bandit is stationary) can be linear in T as long as the total variation is bounded. On the other hand, the second approach requires the number of intervals to be bounded, but the variation within intervals can be substantial. Nevertheless, [66] generalize many reinforcement learning algorithms that work optimally in stationary environments to work optimally in non-stationary environments without any knowledge of the variation budget V or the total number of changes S. We also refer to [100], [101] for a discussion of the Markovian bandit and [102] for bandits with seasonality.

The non-stationary bandit is especially relevant to revenue management and dynamic pricing. [63], [90] propose a sliding window upper confidence bound algorithm for the linear bandit where the *Euclidean norm* of the variation in the cost coefficients is upper bounded (but the upper bound is unknown to the DM). Their results cover advertisement allocation, dynamic pricing, and traffic network routing.

Multi-Agent Learning: There is a rich literature on multi-agent learning, particularly focusing on online simultaneous games and online Stackelberg games. See [103] for an overview on multi-agent reinforment learning. In particular, [104] consider a platform on which multiple sellers offer products, where sellers' pricing decisions are incentivized by the platform's contract, and both the sellers and the platform do not have full knowledge about the demand price relationship. Unlike ours where the retailer has more information on market demand than the supplier and the latter needs to leverage her interactions with the former to learn the market demand and maximize profit, they focus on the information advantage of the platform over the sellers and study whether and when the platform should release its information advantage.

3.3 Problem Formulation

Throughout, we let $[N] \triangleq \{1, \ldots, N\}$ be the running index for any integer $N \geq 1$. We adopt the asymptotic notations $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$ [105]. When logarithmic factors are omitted, we use $\tilde{O}(\cdot)$, $\tilde{o}(\cdot)$, $\tilde{\Omega}(\cdot)$, and $\tilde{\Theta}(\cdot)$. We write 'max' instead of 'sup' and 'min' instead of 'inf'. When the optimal solution to the optimization problem does not exist, an "optimal solution" means an $\epsilon-$optimal solution for $\epsilon > 0$ arbitrarily small.

We consider a wholesale price contract between one supplier (she) and one retailer (he) for a single product, where the retailer faces random demand. Let c be the supplier's unit production cost and s be the retailer's unit selling price. We use $\mathcal{W} = [0, s]$ to denote the set of admissible wholesale prices, i.e., the supplier cannot sell for more than the retailer selling price (we extend to the case where the supplier's set of admissible decisions $\mathcal{W}$ has finite cardinality in the appendix). Notice that the supplier will gain a negative profit if she sells for less than her production cost c, however we allow this possibility since occasionally pricing for less than c may help the supplier explore.

The supplier and retailer interact over a series of time periods indexed by $t \in [T]$ with $T \geq 1$. Let ξ_t be the random demand in period $t \in [T]$ with support $\Xi \subset \mathbb{R}_+$. Let $\mathcal{P}(\Xi)$ be the set of probability distributions on Ξ. We denote the cumulative distribution function (CDF) of demand ξ_t as $F_t \in \mathcal{P}(\Xi)$ and whose density is f_t (if it exists). We introduce the shorthand $F_{1:t} \triangleq (F_i)_{i=1}^t$ for $t \in [T]$ for the sequences of true market demand distributions.

In each period t, the supplier first offers the retailer the wholesale price $w_t \in \mathcal{W}$. Then, the retailer observes w_t, determines his order quantity q_t, and the supplier earns profit $\varphi(w_t; q_t) \triangleq (w_t - c) q_t$. Finally, demand ξ_t is realized and the retailer earns profit $R(q_t; w_t, \xi_t) \triangleq s \min\{q_t, \xi_t\} - w_t q_t$.

The supplier only has access to past wholesale prices and corresponding retailer order quantities. We define $\mathcal{G}_t \triangleq \{(w_i, q_i)\}_{i=1}^t$ to be the history of prices and order quantities *by the end of* period t (we let $\mathcal{G}_0 \triangleq \emptyset$). The supplier's (possibly randomized) pricing policy is a sequence of mappings from $\mathcal{G}_t$ to the set of probability distributions on $\mathcal{W}$ (denoted $\mathcal{P}(\{W\})$). We denote the supplier's (possibly randomized) pricing policy by $\pi \triangleq (\pi_t)_{t=1}^T$

where $\pi_1 \in \mathcal{P}(\mathcal{W})$ and $\pi_t : \mathcal{G}_{t-1} \to \mathcal{P}(\mathcal{W})$ for all $t \geq 2$. The wholesale prices under π then follow:

$$w_1 \sim \pi_1, \tag{3.1a}$$

$$w_t \sim \pi_t\left(\mathcal{G}_{t-1}\right), \quad \forall t \geq 2. \tag{3.1b}$$

Now we characterize the retailer's policy. Let $\mathcal{H}_t \triangleq \{(w_i, q_i, \xi_i)\}_{i=1}^{t}$ be the retailer's information *by the end of* period t which consists of the history of wholesale prices, order quantities, and demand realizations up to period t (we simply let $\mathcal{H}_0 \triangleq \emptyset$). Then, the retailer has access to information $\mathcal{H}_{t-1} \cup \{w_t\}$ right before his ordering decision is made. We let $\mu = (\mu_t)_{t=1}^{T}$ denote the retailer's (possibly randomized) inventory learning policy, where $\mu_1 : w_1 \to \mathcal{P}(\Xi)$ and $\mu_t : \mathcal{H}_{t-1} \cup \{w_t\} \to \mathcal{P}(\Xi)$ for $t \in [2, T]$. The retailer's order quantities under μ are then determined by:

$$q_1^{\mu} \sim \mu_1(w_1), \tag{3.2a}$$

$$q_t^{\mu} \sim \mu_t\left(\mathcal{H}_{t-1} \cup \{w_t\}\right), \quad \forall t \geq 2. \tag{3.2b}$$

We write $q_t^{\mu}(w_t; \mathcal{H}_{t-1})$ to denote the retailer's period $t \in [T]$ response to wholesale price w_t under policy μ.

We measure the performance of the supplier's pricing policy π in terms of its *dynamic* regret. We use a clairvoyant benchmark who can predict the retailer's true order quantity given any wholesale price, and the clairvoyant does not necessarily know the true market demand distribution. For example, if the benchmark has the demand data received by the retailer and knows the policy in use by the retailer, then it can perfectly predict the retailer's order quantity given any wholesale price.

Since the retailer's policy is unknown, we identify a class $\mathcal{M}$ (which depends on T and other model parameters, we will specify $\mathcal{M}$ shortly) of reasonable retailer policies. If the retailer's policy is allowed to be completely arbitrary, then we cannot always expect to get a sublinear regret for the supplier. We then consider the supplier's worst-case regret over policies in $\mu \in \mathcal{M}$. Let

$$w_t^* \in \arg\max_{w \in \mathcal{W}} (w - c) q_t^\mu(w; \mathcal{H}_{t-1}) \tag{3.3}$$

be the benchmark's optimal wholesale price with full knowledge of how the retailer will respond under q_t^μ. Then, the regret in period t when the supplier prices at w_t is $(w_t^* - c)q_t^\mu(w_t^*; \mathcal{H}_{t-1}) - (w_t - c)q_t^\mu(w_t; \mathcal{H}_{t-1})$. The overall dynamic regret over the entire planning horizon is then:

$$\operatorname{Reg}(\pi, T) \triangleq \max_{\mu \in \mathcal{M}} \mathbb{E}\left[\sum_{t=1}^{T} ((w_t^* - c)q_t^\mu(w_t^*; \mathcal{H}_{t-1}) - (w_t - c)q_t^\mu(w_t; \mathcal{H}_{t-1})\right],$$

where the expectation is taken with respect to both the supplier and retailer's possibly randomized policies, and the underlying random demand. The clairvoyant benchmark in the dynamic regret is able to adjust its strategy dynamically in response to the non-stationarity of the retailer response functions $q_t^\mu(\cdot; \mathcal{H}_{t-1})$.

Dynamic regret is a stronger concept than stationary regret. In the definition of the stationary regret, the clairvoyant must set the same wholesale price $w^* \in \arg\max_{w \in \mathcal{W}} \sum_{t=1}^{T} (w - c)q_t^\mu(w; \mathcal{H}_{t-1})$ for the entire planning horizon and the supplier's stationary regret is

$$\operatorname{Reg}_{stat}(\pi, T) \triangleq \max_{\mu \in \mathcal{M}} \mathbb{E}\left[\sum_{t=1}^{T} ((w^* - c)q_t^\mu(w^*; \mathcal{H}_{t-1}) - (w_t - c)q_t^\mu(w_t; \mathcal{H}_{t-1}))\right].$$

It is immediate that the stationary regret is always upper bounded by the dynamic regret.

3.3.1 Retailer Model

Now we present a specific model for how the retailer makes his ordering decisions. For demand distribution $F \in \mathcal{P}(\Xi)$, given wholesale price w the retailer's expected profit from ordering q is $\mathbb{E}_F[R(q; w, \xi)]$. If the retailer believes the demand distribution is F, then his best response to wholesale price w is to order

$$q(w; F) \triangleq \arg\max_{q \geq 0} \mathbb{E}_F[R(q; w, \xi)],$$

or equivalently

$$q(w;F) = \min\{q : F(q) \geq 1 - w/s\}, \tag{3.4}$$

which maximizes his expected profit with respect to F.

In our setting the retailer does not know $F_{1:T}$, and has to implement some inventory learning policy as mentioned before. We now characterize μ by supposing that the retailer's ordering decisions are all best responses to a sequence of *perceived distributions*.

Assumption 3.3.1. *Let μ be the retailer's inventory learning policy. For all $t \in [T]$, there exists a perceived distribution $\hat{F}_t^{\mu}$ that is adapted to $\mathcal{H}_{t-1} \cup \{w_t\}$, such that $q_t^{\mu}(w_t; \mathcal{H}_{t-1}) = q(w_t; \hat{F}_t^{\mu})$.*

We let $\hat{F}_{1:t}^{\mu} \triangleq (\hat{F}_1^{\mu}, \ldots, \hat{F}_t^{\mu})$ for all $t \in [T]$ denote the sequences of perceived distributions. We say a *stationary retailer* is one who has full knowledge about the demand distribution, and the true distribution is stationary ($\hat{F}_t^{\mu} = F_t \triangleq F_0$ for all $t \in [T]$). Otherwise, we have a *learning retailer*. A learning retailer introduces non-stationarity into the supplier's decision-making environment, even if the true market demand distribution is stationary.

Assumption 3.3.1 says that, at any period $t \geq 1$, the retailer's order quantity $q(w_t; \hat{F}_t^{\mu})$ is a best response to some data-driven CDF $\hat{F}_t^{\mu}$ that only depends on the information that has been revealed to the retailer up to period t (i.e., $\mathcal{H}_{t-1} \cup w_t$). In other words, the retailer's ordering decisions and thus its entire policy are completely determined by $\hat{F}_{1:T}^{\mu}$. This assumption is without loss of generality, since q_t^{μ} must be adapted to $\mathcal{H}_{t-1} \cup w_t$ anyway. If we are given a rule for constructing q_t^{μ} directly from the data, we can always find $\hat{F}_t^{\mu}$ for which q_t^{μ} is the best response. There may exist more than one $\hat{F}_t^{\mu}$ in period t that satisfies Assumption 3.3.1.

We also assume that the retailer knows the support of the true sequence of demand distributions, and thus the retailer will construct $\hat{F}_t^{\mu}$ whose support is contained in the support of F_t.

Assumption 3.3.2. *Let μ be the retailer's inventory policy. For all $t \in [T]$, the support of $\hat{F}_t^{\mu}$ is contained in the support of F_t.*

In Section 3.6 we show that many inventory policies satisfy Assumptions 3.3.1 and 3.3.2. For example, under SAA, the retailer's perceived distribution is the empirical distribution of the observed demand samples.

3.3.2 Supplier's Regret

Since the retailer's order quantities are fully determined by $\hat{F}^{\mu}_{1:T}$, the supplier's task of minimizing regret is equivalent to learning $\hat{F}^{\mu}_{1:T}$. However, it is well known that if $\hat{F}^{\mu}_{1:T}$ can vary arbitrarily, then there is no pricing policy that achieves sublinear regret for the supplier. Our main question only makes sense if we restrict the retailer's inventory policy (or equivalently, the sequence of $\hat{F}^{\mu}_{1:T}$) to belong to a reasonable class. In this case, we expect the variation in $\hat{F}^{\mu}_{1:T}$ to be more limited since the retailer accumulates information about the demand distribution incrementally over time, and so their perceived distributions should not change too much from period to period.

We need a metric to quantify this variation in $\hat{F}^{\mu}_{1:T}$. Recall the Kolmogorov distance d_K between CDFs F and G with support on $\Xi \subset \mathbb{R}_+$ is defined by $d_K(F, G) = \max_{x\in\Xi} |F(x) - G(x)|$. The variation in the retailer's perceived distributions from period t to period $t+1$ is then $d_K(\hat{F}^{\mu}_t, \hat{F}^{\mu}_{t+1})$, and the total variation of the sequence $\hat{F}^{\mu}_{1:T}$ is $\sum_{t=1}^{T-1} d_K(\hat{F}^{\mu}_t, \hat{F}^{\mu}_{t+1})$. We note that d_K is computable for a wide range of possible perceived distributions. In addition, d_K is used in the well-known Kolmogorov-Smirnov test and thus has an intuitive appeal for measuring the similarity between two distributions. Furthermore, d_K can be upper bounded by many other distance metrics or divergences, e.g., total variation distance, relative entropy, Helinger distance, Wasserstein distance, etc. [64]. This feature of d_K greatly facilitates connecting the retailer's policy with the regret analysis for the supplier.

We restrict attention to the class of retailer policies for which the total variation of $\hat{F}^{\mu}_{1:T}$ (given by $\sum_{t=1}^{T-1} d_K(\hat{F}^{\mu}_t, \hat{F}^{\mu}_{t+1})$) is bounded. Specifically, let $V \geq 0$ (where V is a function of T) be a budget for the total variation and define the class of retailer policies:

$$\mathcal{M}(V,T) \triangleq \Bigg\{\mu : \text{for any } (w_t)_{t=1}^{T} \in \mathcal{W}, \text{exists } \hat{F}^{\mu}_{1:T} \text{ satisfying Assumptions 3.3.1 and 3.3.2} \\ \text{such that } \sum_{t=1}^{T-1} d_K(\hat{F}^{\mu}_t, \hat{F}^{\mu}_{t+1}) \leq V\Bigg\}.$$

The set $\mathcal{M}(V,T)$ includes all μ such that the total variation of $\hat{F}^{\mu}_{1:T}$ does not exceed V for any sequence of wholesale prices. Notice that $\mathcal{M}(V,T)$ also implicitly depends on $F_{1:T}$ (since the demand samples are generated from $F_{1:T}$), but we suppress this dependence for brevity.

With some abuse of the notation, we write the supplier's profit in period t as a function of w_t and $\hat{F}^{\mu}_t$ as $\varphi(w_t; \hat{F}^{\mu}_t) \triangleq (w_t - c)q(w_t; \hat{F}^{\mu}_t)$, when the retailer orders optimally based on the perceived distribution $\hat{F}^{\mu}_t$. The supplier's learning problem can then be framed in terms of the sequence of her profit functions $\{\varphi(w_t; \hat{F}^{\mu}_t)\}_{t=1}^{T}$.

The previous learning literature always considers bounded variation of the profit functions, and proposes learning algorithms specific to this type of variation [61], [85], [86]. In the following example, we show that the variation of $\hat{F}^{\mu}_{1:T}$ does not directly translate into the variation of $\{\varphi(w_t; \hat{F}^{\mu}_t)\}_{t=1}^{T}$. Thus, these previously proposed learning algorithms do not apply to our setting.

(3.5) Let $c = 0$ and $s = 1$. Let $\hat{F}^{\mu}_t$ be the CDF of a Bernoulli random variable which takes values 0 and 1 with probabilities p_t and $1 - p_t$, respectively, for all $t \in [T]$. Let

$$p_t = \begin{cases} \frac{1}{2} - \epsilon, & \text{for } t \text{ odd,} \\ \frac{1}{2} + \epsilon, & \text{for } t \text{ even,} \end{cases}$$

where $\epsilon = 1/T$. It is straightforward to see that $\sum_{t=1}^{T-1} d_K(\hat{F}_t^\mu, \hat{F}_{t+1}^\mu) \leq 2T\epsilon = 2$ but that

$$\sum_{t=1}^{T-1} \max_{w \in [c,s]} |\varphi(w; \hat{F}_t^\mu) - \varphi(w; \hat{F}_{t+1}^\mu)| = \frac{T}{2}.$$

We see that even if $\hat{F}_{1:T}^\mu$ has constant total variation that does not grow in T, the sequence of profit functions $(\varphi(w; \hat{F}_t^\mu))_{t=1}^T$ can have variation that grows linearly in T.

For the specific class $\mathcal{M}(V, T)$ of retailer policies, the overall regret is then:

$$\mathrm{Reg}(\pi, T) \triangleq \max_{\mu \in \mathcal{M}(V,T)} \mathbb{E}\left[\sum_{t=1}^{T} \Big((w_t^* - c)q(w_t^*; \hat{F}_t^\mu) - (w_t - c)q(w_t; \hat{F}_t^\mu)\Big)\right],$$

where the expectation is taken with respect to the supplier's possibly randomized policy and any randomization in $\hat{F}_t^\mu$ (i.e., $\hat{F}_t^\mu$ is random in general because it usually depends on the random demand realizations, and the retailer may also use a randomized policy). For comparison, we define the stationary regret to be:

$$\mathrm{Reg}_{stat}(\pi, T) \triangleq \max_{\mu \in \mathcal{M}(V,T)} \mathbb{E}\left[\sum_{t=1}^{T} \Big((w^* - c)q(w^*; \hat{F}_t^\mu) - (w_t - c)q(w_t; \hat{F}_t^\mu)\Big)\right].$$

If the demand distribution is stationary, then intuitively we expect $\hat{F}_{1:T}^\mu$ to exhibit some type of convergence since the retailer is accumulating more information about the same distribution. Yet, our results also apply when the true market demand distribution is changing over time. For example, the demand distribution may have a seasonal pattern. In this case, $\hat{F}_{1:T}^\mu$ may still have bounded variation that is sublinear in T.

3.4 Stationary Retailer

We first consider the special case of the supplier's problem for a stationary retailer to help us understand the structure of the supplier's profit function.

Assumption 3.4.1. *The true demand distribution is stationary, i.e., $\hat{F}_t^\mu = F_t = F_0$ for all $t \in [T]$.*

Since $\mathcal{W}$ is a continuum of allowable prices, this setting is a continuous bandit treating each price as an arm to pull. The literature typically imposes some structure on the DM's objective function, e.g., convexity or unimodality, Lipchitz continuity or Holder continuity, etc. (see [86], [106]). However, the supplier's profit function is not necessarily continuous in our setting. Nevertheless, we will show that we can find pricing policies that have sublinear regret bounds for general demand distributions.

We make the following boundedness assumption on the demand distribution to continue.

Assumption 3.4.2. *The sequence $F_{1:T}$ has bounded support on $\Xi = [0, \bar{\xi}]$ for $0 < \bar{\xi} < \infty$, and both the retailer and the supplier know $\bar{\xi}$.*

Under Assumption 3.5.2, we can always relate the values of the supplier's profit function at w_t and w'_t. Without loss of generality, suppose $w'_t < w_t$, then for any demand distribution F we have

$$\begin{aligned} \varphi(w_t; F) - \varphi(w'_t; F) &= (w_t - c)q(w_t; F) - (w'_t - c)q(w'_t; F) \\ &\leq (w_t - c)q(w_t; F) - (w'_t - c)q(w_t; F) \\ &\leq (w_t - w'_t)\bar{\xi}. \end{aligned} \tag{3.6}$$

This inequality holds regardless of whether the true distribution is discrete or continuous. In particular, we can discretize $\mathcal{W}$ with a finite set of prices, and then bound sub-optimality of this discretization using Eq. (3.6).

We consider the following simple pricing policy for the stationary retailer that we denote by π_{stat}. The supplier first discretizes $\mathcal{W}$ into $\lceil \sqrt{T} \rceil$ equally sized intervals, and then takes $\bar{\mathcal{W}}_{\lceil \sqrt{T} \rceil}$ to be the wholesale prices at the breakpoints of these intervals. In the first $\lceil \sqrt{T} \rceil$ periods, the supplier sets each price in $\bar{\mathcal{W}}_{\lceil \sqrt{T} \rceil}$ once and collects the corresponding profit. Let $w^*_{stat} \in \arg\max_{w \in \bar{\mathcal{W}}_{\lceil \sqrt{T} \rceil}} \varphi(w; F_0)$ be the wholesale price in $\bar{\mathcal{W}}_{\lceil \sqrt{T} \rceil}$ with the highest profit. In all of the remaining periods, the supplier sets the wholesale price $w_t = w^*_{stat}$ for all $t = T - \lceil \sqrt{T} \rceil + 1, \ldots, T$.

Since $\hat{F}_{1:T}$ is stationary in this case, the supplier's profit function does not change from period to period. Then, the dynamic and stationary regret coincide and are:

$$\operatorname{Reg}(\pi_{\text{stat}}, T) = \sum_{t=1}^{T} \mathbb{E}\left[(w^* - c)q(w^*; F_0) - (w_t - c)q(w_t; F_0)\right],$$

where $w^* \in \max_{w \in \mathcal{W}}(w - c)q(w; F_0)$. Our proposed policy π_{stat} gives the supplier $O(\sqrt{T})$ regret. We emphasize that this result holds for both discrete and continuous demand distributions.

Theorem 3.4.1. *Suppose Assumptions 3.4.1 and 3.5.2 hold. For all $T \geq 1$, we have $Reg(\pi_{stat}, T) = O(\sqrt{T})$.*

3.5 Supplier's Pricing Policy

The supplier's profit function is generally multi-modal, and it is also changing shape since the retailer is updating their perceived distribution. Towards solving the supplier's problem in this setting, we start with the case of a discrete demand distribution and then handle the continuous case via an approximation argument. In both cases, we establish a sublinear regret for the supplier's pricing policy.

3.5.1 Discrete Demand Distributions

We first assume that the demand in all time periods has common finite support.

Assumption 3.5.1. *The sequence $F^{\mu}_{1:T}$ has common support on M points: $y_M (\triangleq \bar{\xi}) > y_{M-1} > y_{M-2} > \cdots > y_1 > 0$. The support $\mathcal{Y}_M \triangleq \{y_m\}_{m \in [M]}$ is known to both the supplier and the retailer.*

Combined with Assumption 3.3.2, the sequence of perceived distributions $\hat{F}^{\mu}_{1:T}$ also has support on $\mathcal{Y}_M$. Let $p_{t,m} \triangleq \hat{F}^{\mu}_t(y_m)$ for $m \in [M]$ denote the values of the retailer's perceived distribution in period $t \in [T]$ (where $p_{t,0} \triangleq 0$ and $p_{t,M} \triangleq 1$). The retailer's order quantity given by Eq. (3.4) is then

$$q_t = \begin{cases} y_1, & \text{if } 0 \leq 1 - w_t/s \leq p_{t,1}, \\ y_m, & \text{if } p_{t,m-1} < 1 - w_t/s \leq p_{t,m}, \text{ for } m \in \{2, \ldots, M\}. \end{cases} \tag{3.7}$$

We see that the retailer's optimal order quantity is a piecewise constant function of the wholesale price w_t. According to Eq. (3.7), the retailer's order quantity will always be in the support of the discrete distribution and satisfy:

$$q_t \in \mathcal{Y}_M. \tag{3.8}$$

Consequently, the supplier's profit function in period t is:

$$\varphi(w_t; \hat{F}_t^{\mu}) = \begin{cases} (w_t - c)y_1, & \text{if } 0 \leq 1 - w_t/s \leq p_{t,1}, \\ (w_t - c)y_m, & \text{if } p_{t,m-1} < 1 - w_t/s \leq p_{t,m} \text{ for } m \in [M]. \end{cases} \tag{3.9}$$

We see it is a piecewise linear function of w_t, with discontinuities at the breakpoints $\{s(1 - p_{t,m})\}_{m \in [M]}$.

We propose a pricing policy for the supplier called π_{LUNA}, where LUNA stands for *Learning under a Nonstationary Agent.* It is based on the deterministic bandit algorithm proposed in [65]. π_{LUNA} works by taking advantage of the special structure of the supplier's profit function, as characterized in Eq. (3.9). In particular, the performance of the pricing policy depends on accurate estimation of the probabilities $\boldsymbol{p}_t \triangleq (p_{t,1}, \ldots, p_{t,M})$. π_{LUNA} indirectly estimates $\boldsymbol{p}_t$ by observing the retailer's order quantity and the supplier's profit through Eq. (3.9). A good policy is expected to maintain a somewhat accurate estimate of $\boldsymbol{p}_t$, but must also hedge against large variation in $\boldsymbol{p}_t$. If $\boldsymbol{p}_t$ has varied a lot, then the optimal wholesale price may be different, and a large regret will be incurred if the supplier fails to adapt.

The full details of π_{LUNA} are presented in Algorithm 1. It consists of multiple epochs, where each epoch consists of an exploration phase followed by an exploitation phase. We let $i \geq 1$ index epochs, but we usually omit dependence on i except when necessary since each epoch follows the same pattern. Let τ^0_{i+1} denote the last period of epoch $i \geq 1$ (where $\tau^0_1 = 0$). Then, epoch $i \geq 1$ covers periods $t \in [\tau^0_i + 1, \ldots, \tau^0_{i+1}]$. Each time π_{LUNA} starts the exploration phase of a new epoch, it discards all previous information and estimates $\boldsymbol{p}_t$ from scratch (since $\boldsymbol{p}_t$ is nonstationary, it is the estimate for some particular period t). It also constructs a set of exploratory wholesale prices, tries each price once, and records the optimal exploratory price which led to the highest observed supplier profit.

Once it has an initial estimate of $\boldsymbol{p}_t$ and an optimal exploratory wholesale price has been found, π_{LUNA} will enter the exploitation phase. It first constructs a new set of wholesale prices for the exploitation phase. Then, in each period, a wholesale price is drawn randomly from this set according to some distribution which balances the exploration-exploitation trade-off. Most of the time, π_{LUNA} prices at the nearly optimal wholesale price found in the exploration phase, while also occasionally detecting whether the previously identified optimal wholesale price is no longer optimal. If that is the case, then π_{LUNA} quantifies a lower bound on the variation of $\hat{F}^\mu_t$ in the current epoch and begins the next epoch.

Exploration Phase of π_{LUNA}

The goal of the exploration phase is to obtain an initial estimate of $\boldsymbol{p}_t$ and to find the optimal wholesale price corresponding to this initial estimate. As a first step, π_{LUNA} discretizes $\mathcal{W}$ into $K+1$ equal-length intervals with $K \geq 1$ equally spaced wholesale prices (where K is an input parameter to be specified later, which is the same for every epoch), and we let $\bar{\mathcal{W}}_K \triangleq \{\bar{w}_k\}_{k=1}^K$ where:

$$\bar{w}_k \leftarrow (k-1)\frac{(s-c)}{K} + c, \; \forall k \in [K]. \tag{3.10}$$

Then in each period $\tau^0_i + k$ for all $k \in [K]$, π_{LUNA} sets the wholesale price $\bar{w}_k$ and the corresponding retailer order quantity is $q(\bar{w}_k; \hat{F}^\mu_{\tau^0_i+k})$. Upon setting $\bar{w}_k$, π_{LUNA} earns profit φ_k given by:

$$\varphi_k \leftarrow (\bar{w}_k - c)q(\bar{w}_k; \hat{F}^\mu_{\tau^0_i+k}). \tag{3.11}$$

Let $k^* \in \arg\max_{k\in[K]} \varphi_k$, so $\bar{w}_{k^*}$ is the wholesale price that maximizes the observed profit among $\bar{\mathcal{W}}_K$. By Eq. (3.8), we must have $q(\bar{w}_{k^*}; \hat{F}^{\mu}_{\tau_i^0+k^*}) \in \mathcal{Y}_M$, so we also let $m^* \in [M]$ be such that $y_{m^*} \triangleq q(\bar{w}_{k^*}; \hat{F}^{\mu}_{\tau_i^0+k^*})$. With $\bar{w}_{k^*}$ and y_{m^*} in hand, we begin the exploitation phase of epoch i.

Exploitation Phase of $\pi_{\mathbf{LUNA}}$

If the retailer is stationary, then $\bar{w}_{k^*}$ will remain nearly optimal for the rest of the planning horizon (this is exactly the supplier's pricing policy for a stationary retailer, see Section 3.4). However, as the retailer is also learning the demand distribution, we expect $\hat{F}^{\mu}_t$ to vary. If $\hat{F}^{\mu}_t$ has varied a lot and the supplier still prices at $\bar{w}_{k^*}$, then she is likely to suffer a large regret. The supplier's pricing policy has to balance between exploitation (stick to the optimal $\bar{w}_{k^*}$ found in the exploration phase) and exploration (hedge against the risk that $\hat{F}^{\mu}_t$ has changed a lot since the exploration phase).

We show that this balance can be achieved during the exploitation phase by randomly choosing from a carefully constructed finite set of prices. This set of prices will in fact be different for each period t, and is constructed based on the structure of the supplier's profit function through Eq. (3.9). We first construct this set of prices, and then explain the intuition behind it.

We allow the pricing policy in period t to have a margin of sub-optimality $\Delta_t > 0$ compared with φ_{k^*}, the highest profit observed in the exploration phase. The sequence $\{\Delta_t\}_{t\geq 1}$ will be decreasing in t (in particular, we will take $\Delta_t = O(\sqrt{1/t})$). Then, in each period t, we construct a set of prices based on Δ_t to sample from.

There are two cases where $\bar{w}_{k^*}$ becomes sufficiently sub-optimal to end the current epoch, either: (i) the supplier's profit at some $w \neq \bar{w}_{k^*}$ has increased a lot since the exploration phase; or (ii) the supplier's profit at $\bar{w}_{k^*}$ has decreased a lot since the exploration phase. We discuss the details of these two cases separately.

Case one:

In the first case, some $w \in \mathcal{W}$ with $w \neq \bar{w}_{k^*}$ now earns greater profit for the supplier than $\bar{w}_{k^*}$. This w can be an arbitrary member of $\mathcal{W}$ as long as $w \neq \bar{w}_{k^*}$. However, we cannot check every price in $\mathcal{W}$, so we construct a specialized finite set of prices to check as follows.

We recall that $q_t \in \mathcal{Y}_M$ holds for all $t \in [T]$ under Assumption 3.5.1, so we ask the question: *Suppose the retailer's order quantity is $y_m \in \mathcal{Y}_M$ for some $m \in [M]$, then what wholesale price (denoted by w_m^t) would give the supplier a profit that is equal to $\varphi_{k^*} + \Delta_t$?* If the retailer's order quantity $q(w_m^t; \hat{F}_t^{\mu})$ under w_m^t turns out to be larger than (smaller than) y_m, then w_m^t will give a higher (lower) profit than $\varphi_{k^*} + \Delta_t$. We now construct a set of wholesale prices in this way corresponding to each $y_m \in \mathcal{Y}_M$. In period t, for each $m \in [M]$, we set the corresponding wholesale price according to:

$$(w_m^t - c)y_m = \varphi_{k^*} + \Delta_t + \frac{y_m s}{K}, \text{ which gives } w_m^t \triangleq \Big(\varphi_{k^*} + \Delta_t + \frac{y_m s}{K}\Big)/y_m + c, \tag{3.12}$$

where the term $\frac{y_m s}{K}$ is introduced to account for the error introduced by discretizing $\mathcal{W}$ to $\bar{\mathcal{W}}_K$.

If the supplier prices at w_m^t in period t, and if $q(w_m^t; \hat{F}_t^{\mu}) \geq y_m$, then $\bar{w}_{k^*}$ is no longer nearly optimal since the optimality gap now exceeds Δ_t (i.e., we have $(w_m^t - c)y_m \geq \varphi_{k^*} + \Delta_t$). We summarize this discussion in Lemma 3.5.1.

Lemma 3.5.1. *If $q(w_m^t; \hat{F}_t^{\mu}) \geq y_m$, then $p_{t,m-1} \leq 1 - w_m^t/s$ and $\varphi(w_m^t; \hat{F}_t^{\mu}) \geq (w_m^t - c)y_m = \varphi_{k^*} + \Delta_t + \frac{y_m s}{K}$. Otherwise, if $q(w_m^t; \hat{F}_t^{\mu}) < y_m$, then $p_{t,m-1} \geq 1 - w_m^t/s$ and $\varphi(w_m^t; \hat{F}_t^{\mu}) < (w_m^t - c)y_m = \varphi_{k^*} + \Delta_t + \frac{y_m s}{K}$.*

In addition, if $q(w_m^t; \hat{F}_t) \geq y_m$, then $\hat{F}_t^{\mu}$ has varied a lot since the exploration phase.

Lemma 3.5.2. *If $q(w_m^t; \hat{F}_t^{\mu}) \geq y_m$, then $\sum_{j \in [\tau_i^0 + 1, t-1]} d_K(\hat{F}_j^{\mu}, \hat{F}_{j+1}^{\mu}) \geq \Delta_t/(s\,\bar{\xi})$.*

Case two:

In the second case, the supplier's profit from pricing at $\bar{w}_{k^*}$ has decreased a lot since the exploration phase. This can happen if the retailer's order quantity $q(\bar{w}_{k^*}; \hat{F}_t^{\mu})$ in period t

during the exploitation phase is much smaller than the order quantity $q(\bar{w}_{k^*}; \hat{F}^{\mu}_{\tau_i^0+k^*})$ observed during the exploration phase. Since $\bar{w}_{k^*}$ is the optimal price found during the exploration phase, the policy should price frequently at $\bar{w}_{k^*}$ to exploit what is best. However, pricing at $\bar{w}_{k^*}$ does not give useful information about the variation in $\hat{F}_t$. Even if $\hat{F}_t$ has only varied by a small amount, the profit at $\bar{w}_{k^*}$ can still change drastically (recall Example 3.5). If we restart the epoch each time the profit at $\bar{w}_{k^*}$ has decreased a lot, we will end up with too many epochs and a high overall regret.

Therefore, instead of pricing at $\bar{w}_{k^*}$, we determine a surrogate price w_0^t which achieves two purposes: (i) the profit at w_0^t is not much lower than the profit at $\bar{w}_{k^*}$, so we can still exploit the optimality of $\bar{w}_{k^*}$ from the exploration phase (see Eq. (3.14)); and (ii) unlike $\bar{w}_{k^*}$, when the profit at w_0^t is sufficiently low, we can quantify a lower bound on the variation of $\hat{F}^{\mu}_t$ (see Lemma 3.5.4) and correctly restart the epoch. We define the surrogate price w_0^t in period t to satisfy:

$$(w_0^t - c)y_{m^*} = \varphi_{k^*} - \Delta_t \text{ and } w_0^t \geq 0, \text{ otherwise } w_0^t = 0,$$

which gives

$$w_0^t \triangleq \max\{\bar{w}_{k^*} - \Delta_t / y_{m^*}, 0\}. \tag{3.13}$$

Note we require $w_0^t \geq 0$ instead of $w_0^t \geq c$. By allowing $w_0^t < c$, the policy is able to detect variation of $\hat{F}^{\mu}_t$ that would otherwise not be detected.

By Eq. (3.6), the difference in profit between pricing at w_0^t and $\bar{w}_{k^*}$ is lower bounded by:

$$\varphi(w_0^t; \hat{F}^{\mu}_t) - \varphi(\bar{w}_{k^*}; \hat{F}^{\mu}_t) \geq -\Delta_t. \tag{3.14}$$

We make the following inferences based on the surrogate price.

Lemma 3.5.3. *If $q(w_0^t; \hat{F}^{\mu}_t) \geq y_{m^*}$, then $p_{t,m^*-1} \leq 1 - w_0/s$ and $\varphi(w_0^t; \hat{F}^{\mu}_t) \geq (w_0^t - c)y_{m^*} \geq \varphi_{k^*} - \Delta_t$. Otherwise, if $q(w_0^t; \hat{F}^{\mu}_t) < y_{m^*}$, then $p_{t,m^*-1} \geq 1 - w_0^t/s$ and $\varphi(w_0^t; \hat{F}^{\mu}_t) < (w_0^t - c)y_{m^*} = \varphi_{k^*} - \Delta_t$.*

If the retailer's order quantity satisfies $q(w_0^t; \hat{F}_t^{\mu}) < y_{m^*}$, then we know that $\hat{F}_t^{\mu}$ has varied a lot since the exploration phase.

Lemma 3.5.4. *If $q(w_0^t; \hat{F}_t^{\mu}) < y_{m^*}$, then $\sum_{j\in[\tau_i^0+1,t-1]} d_K(\hat{F}_j^{\mu}, \hat{F}_{j+1}^{\mu}) \geq \Delta_t/(s\,\bar{\xi})$.*

Algorithm and regret bound

In each period t, we construct the set of wholesale prices $\{w_0^t, w_1^t, \ldots, w_M^t\}$, and π_{LUNA} will randomly sample from $\{w_0^t, w_1^t, \ldots, w_M^t\}$ according to a distribution that is changing over time. Based on the discussion of the previous two cases, the exploitation phase continues until $q(w_m^t; \hat{F}_t^{\mu}) \geq y_m$ for some $m \in [M]$, or $q(w_0^t; \hat{F}_t^{\mu}) < y_{m^*}$. In both cases, $\hat{F}_t^{\mu}$ is guaranteed to have varied a lot since the exploration phase, and π_{LUNA} starts the next epoch. Let $\mathcal{U}([M])$ denote the uniform distribution on $\{1, 2, \ldots, M\}$.

Algorithm 1 Learning under Nonstationary Agent (LUNA)

Require: Time horizon T, supplier production cost c, retailer selling price s, support $\mathcal{Y}_M$, grid size K

Update current period $t \leftarrow 1$

Set epoch $i \leftarrow 1$ and $\tau_1^0 \leftarrow 0$

for epoch $i = 1, 2, \cdots$ **do**

Exploration:

Price at $\bar{w}_k$ (see Eq. (3.10)) and observe φ_k (see Eq. (3.11)) for the first K periods in epoch i

Let $k^* \in \arg\max_{k\in[K]} \varphi_k$ and m^* be such that $y_{m^*} = q\left(\bar{w}_{k^*}; \hat{F}^{\mu}_{\tau_i^0+k^*}\right)$

Exploitation:

In period t, set $\Delta_t \leftarrow \sqrt{M/(t-\tau_i^0)}$

Compute prices w_m^t for $m \in [M]$ and w_0^t according to Eq. (3.12) and Eq. (3.13), respectively

Select wholesale price $w_t \leftarrow w_{m_t}^t$ according to the distribution

$$m_t = \begin{cases} 0, & \text{w.p. } 1 - \sqrt{\frac{M}{t-\tau_i^0}}, \\ \mathcal{U}([M]), & \text{w.p. } \sqrt{\frac{M}{t-\tau_i^0}} \end{cases}$$

Observe retailer's order $q(w_t; \hat{F}_t^{\mu})$

if $q(w_{m_t}^t; \hat{F}_t^{\mu}) \geq y_{m_t}$ for $m_t \in [M]$ or $q(w_{m_t}^t; \hat{F}_t^{\mu}) < y_{m^*}$ for $m_t = 0$ **then**

$\tau_{i+1}^0 \leftarrow t$ and start the next epoch $i \leftarrow i + 1$

Theorem 3.5.1 upper bounds the regret of Algorithm 1 as a function of V and K (notice that π_{LUNA} does not need to know V, but K is an input). When in addition V is known, the decision maker can choose K optimally as a function of V to minimize the regret.

Theorem 3.5.1. *Suppose Assumption 3.5 holds.*

(i) For all $K \geq 1$, Reg$(\pi_{LUNA}, T) = \tilde{O}(\bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} M^{\frac{1}{3}} T^{\frac{2}{3}} + \frac{\bar{\xi}T}{K} + \bar{\xi}^{\frac{5}{3}} K V^{\frac{2}{3}} M^{-\frac{1}{3}} T^{\frac{1}{3}})$.

(ii) If the supplier knows V, then K can be chosen optimally as $K^* = \left\lceil T^{\frac{1}{3}} V^{-\frac{1}{3}} \bar{\xi}^{-\frac{1}{3}} \right\rceil$, and the minimized regret is Reg$(\pi_{LUNA}, T) = \tilde{O}(\bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} M^{\frac{1}{3}} T^{\frac{2}{3}})$.

(iii) If the supplier does not know V, then K can be chosen obliviously as $\hat{K} = \left\lceil \bar{\xi}^{-\frac{1}{3}} T^{\frac{1}{3}} \right\rceil$, and the regret is $Reg(\pi_{LUNA}, T) = \tilde{O}(\bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} M^{\frac{1}{3}} T^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} V^{\frac{2}{3}} M^{-\frac{1}{3}} T^{\frac{2}{3}})$.

According to Theorem 3.5.1, using our pricing policy π_{LUNA}, the supplier can achieve a sublinear regret bound if $V = \tilde{o}(T)$ and if $V = \tilde{o}(\sqrt{T})$ when the supplier knows V and does not know V respectively.

3.5.2 Proof Outline of Theorem 3.5.1 ($\pi_{\mathbf{LUNA}}$)

Here we overview the proof of Theorem 3.5.1, all detailed expressions and derivations referenced here appear in Appendix B.2. We first do the regret analysis for a single epoch i (which consists of periods $t \in [\tau_i^0 + 1, \tau_{i+1}^0]$), and then assemble these into an overall regret bound. To begin, we decompose the regret in epoch i into:

$$\sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \left\{ \varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu) \right\} = \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \left\{ \varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu) \right\} + \left\{ \varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu) \right\},$$

where the first part $\text{Reg}_i^c(\pi_{\text{LUNA}}) \triangleq \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \{\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\}$ (the superscript 'c' is for 'clairvoyant') is the regret incurred by always pricing at w_0^t compared to the clairvoyant benchmark, and the second part $\text{Reg}_i^0(\pi_{\text{LUNA}}) \triangleq \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \{\varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu)\}$ (the superscript '0' corresponds to the surrogate price) is the regret incurred compared with the benchmark of always pricing at the surrogate price w_0^t. We analyze these two parts separately.

Part I of the regret

To upper bound $\mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}})$, we define the following subset of periods of epoch i:

$$\mathcal{E}^i \triangleq \Big\{ t \in [\tau_i^0 + \max\{M+2, K\} + 1, \tau_{i+1}^0] : q(w_m^t; \hat{F}_t^\mu) < y_m,\ \forall m \in [M],\ \text{and} \\ q(w_0^t; \hat{F}_t^\mu) \geq y_{m^*} \Big\}.$$

If $t \in \mathcal{E}^i$, then $\hat{F}_t^\mu$ has not varied a lot within epoch i and pricing at w_0^t remains nearly optimal. On the other hand, if $t \notin \mathcal{E}^i$, then pricing at w_0^t is no longer nearly optimal either because the profit at w_0^t has gone down, or the profit at some $w_m^t \neq w_0^t$ has gone up. We can further decompose

$$\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu) = \Big(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\Big)\, \mathbb{1}(t \in \mathcal{E}^i) \\ + \Big(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\Big)\, \mathbb{1}(t \notin \mathcal{E}^i),$$

and then upper bound these expressions separately. First we upper bound the regret $\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)$ for periods $t \in \mathcal{E}^i$. The next result takes effect after period $\tau_i^0 + K$ and only applies to the exploitation phase.

Lemma 3.5.5. *For all $t \in [\tau_i^0 + \max\{M+2, K\} + 1, \tau_{i+1}^0] \cap \mathcal{E}^i$, we have $\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu) \leq 2\Delta_t + \frac{\bar{\xi}s}{K}$.*

Next we upper bound $E^i \triangleq \sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} \mathbb{1}(t \notin \mathcal{E}^i)$, the number of periods when $t \notin \mathcal{E}^i$ (during the exploitation phase of epoch i).

Lemma 3.5.6. *With probability at least $1 - 1/T^2$, $E^i \leq 2\log(T)\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)} + 1$.*

We combine Lemmas 3.5.5 and 3.5.6, and summarize the resulting bound on $\mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}})$ in Eq. (B.8) in Appendix B.2.

Part II of the regret

To upper bound $\mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}})$, we note

$$\varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu) = \left(\varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu)\right) \mathbb{1}(w_t \neq w_0^t)$$

for all t. That is, the supplier can only incur regret with respect to the benchmark of always pricing at w_0^t if $w_t \neq w_0^t$. Let

$$T_i(K) \triangleq \left|\{t \in [\tau_i^0 + K + 1, \tau_{i+1}^0] : m_t \neq 0\}\right|$$

be the number of periods in the exploitation phase of epoch i when π_{LUNA} does not select $m_t = 0$ (and price at w_0^t). The next lemma upper bounds $T_i(K)$.

Lemma 3.5.7. *([65, Lemma A.2]) For all $K \geq 1$, $T_i(K) \leq \sqrt{11 \log(T) M(\tau_{i+1}^0 - 1 - \tau_i^0)}$ with probability at least $1 - 1/T^2$.*

We summarize the resulting bound for $\mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}})$ in Eq. (B.9) in Appendix B.2.

Combining the two parts of the regret

We combine Eq. (B.8) and Eq. (B.9) to upper bound the regret for epoch i in Eq. (B.10). To derive the supplier's total regret over the entire planning horizon T, we also need an upper bound on the total number of epochs I. In Lemmas 3.5.2 and 3.5.4, we showed that when an epoch ends, $\hat{F}_t$ has varied a lot within the current epoch. Since the total variation of $\hat{F}_{1:T}$ over the entire planning horizon is bounded, the total number of epochs I must be bounded by the variation budget.

Lemma 3.5.8. *We have $I \leq (s\,\bar{\xi})^{\frac{2}{3}} V^{\frac{2}{3}} M^{-\frac{1}{3}} T^{\frac{1}{3}} + 1$ almost surely.*

Based on Lemma 3.5.8 and Eq. (B.10), we obtain our final regret bound in Eq. (B.11), concluding the proof.

3.5.3 Continuous Demand Distributions

We now turn to the continuous case. We may also use the upcoming approach when the support of demand is finite but very large, so we do not specifically require $F_{1:T}$ and $\hat{F}^{\mu}_{1:T}$ to have a density for this treatment.

We first make the following boundedness assumption on the demand distribution.

Assumption 3.5.2. *The sequence $F_{1:T}$ has bounded support on $[0, \bar{\xi}]$ for $0 < \bar{\xi} < \infty$, and both the retailer and the supplier know $\bar{\xi}$.*

When $\hat{F}^{\mu}_t$ is continuous, under Assumption 3.5.2, q_t can take any value in the interval $[0, \bar{\xi}]$. In contrast, when $\hat{F}^{\mu}_t$ has support on $\mathcal{Y}_M$, the retailer's order quantity always satisfies $q_t \in \mathcal{Y}_M$. π_{LUNA} used this fact to track the variation of $\hat{F}^{\mu}_{1:T}$ when demand has support on $\mathcal{Y}_M$, but it is much harder to infer the behavior of $\hat{F}^{\mu}_t$ in the continuous case.

Our strategy is based on approximating $[0, \bar{\xi}]$ with a finite subset of equally spaced points. Let $N \geq 1$ be the size of this subset, and let $\mathcal{Z}_N \triangleq \{z_n\}_{n \in [N]}$ be equally spaced points on $[0, \bar{\xi}]$ defined by: $z_1 = 0$ and $z_n = n\,\bar{\xi}/(N-1)$ for all $n \in [N-1]$.

We call our pricing policy for the continuous case *Learning under Nonstationary Agent with Continuous Distribution* (LUNAC), denoted π_{LUNAC}, which calls π_{LUNA} as a subroutine on $\mathcal{Z}_N$.

The details of π_{LUNAC} are outlined in Algorithm 2. In each period, π_{LUNAC} enacts the wholesale price w_t suggested by π_{LUNA} and receives the feedback $q(w_t; \hat{F}^{\mu}_t)$. π_{LUNAC} then maps the feedback $q(w_t; \hat{F}^{\mu}_t)$ to some $z_n \in \mathcal{Z}_N$, which is then given to π_{LUNA}, which then outputs a recommended price.

Algorithm 2 **L**earning **U**nder **N**onstationary **A**gent with **C**ontinuous Distribution (LUNAC)

Require: $\mathcal{Z}_N$ for $N \geq 1$
  Initialize π_{LUNA} with $\mathcal{Z}_N$
  while $t \leq T$ **do**
    Set the wholesale price w_t suggested by π_{LUNA} and observe retailer's order $q(w_t; \hat{F}^{\mu}_t)$
    Find n such that $z_{n-1} < q(w_t; \hat{F}^{\mu}_t) \leq z_n$ for some $2 \leq n \leq N$, and take z_n as the feedback to π_{LUNA}

The retailer's order quantities are determined by the perceived distributions $\hat{F}^{\mu}_{1:T}$ (which may have support on all of $[0, \bar{\xi}]$), while we are running π_{LUNA} as a subroutine on $\mathcal{Z}_N$. To analyze the behavior of the π_{LUNA} subroutine, we introduce a sequence of fictitious perceived distributions with support on $\mathcal{Z}_N$. Let $\tilde{F}_t$ be the fictional distribution on $\mathcal{Z}_N$ for period t, which satisfies

$$\tilde{F}_t(z_n) = \hat{F}^{\mu}_t(z_n),\ \forall n \in [N], \tag{3.15}$$

for all $t \in [T]$. We introduce the shorthand $\tilde{F}_{1:t} \triangleq (\tilde{F}_i)_{i=1}^t$ for $t \in [T]$ for the partial sequences of fictitious perceived distributions.

We will establish that, under Eq. (3.15), the wholesale prices output by π_{LUNAC} under $\hat{F}^{\mu}_{1:t}$ coincide with the wholesale prices output by π_{LUNA} on $\mathcal{Z}_N$ under $\tilde{F}_{1:t}$. Let ω be a sample path of the randomization of π_{LUNA}, and let Ω be the set of all such sample paths. All the randomization in π_{LUNAC} comes from the randomization in π_{LUNA} on $\mathcal{Z}_N$, so we can compare both algorithms on Ω. Let $w_t^{\text{LUNAC}}(\hat{F}^{\mu}_{1:t-1}; \omega)$ be the wholesale price output by π_{LUNAC} given the ex post distributions $\hat{F}^{\mu}_{1:t-1}$ under ω, and let $w_t^{\text{LUNA}}(\tilde{F}_{1:t-1}; \omega)$ be the wholesale price output by π_{LUNA} given the ex post distributions $\tilde{F}_{1:t-1}$ under ω (here, by 'ex post' we mean the entire sequences $\hat{F}^{\mu}_{1:t-1}$ and $\tilde{F}_{1:t-1}$ have been observed).

Lemma 3.5.9. *For all $t \in [T]$ and $\omega \in \Omega$, $w_t^{LUNAC}(\hat{F}^{\mu}_{1:t-1}; \omega) = w_t^{LUNA}(\tilde{F}_{1:t-1}; \omega)$.*

Loosely speaking, Proposition **??** says that π_{LUNAC} sets wholesale prices by approximating $\hat{F}^{\mu}_t$ with $\tilde{F}_t$ in each period. It then suggests the wholesale prices given by π_{LUNA}, which pretends the retailer's perceived distribution is actually $\tilde{F}_t$. This interpretation also suggests that if $\hat{F}^{\mu}_{1:T}$ has bounded variation, then $\tilde{F}_{1:T}$ should have bounded variation, as shown in Lemma 3.5.10.

Lemma 3.5.10. *For all $t \in [T-1]$, $d_K(\tilde{F}_t, \tilde{F}_{t+1}) \leq d_K(\hat{F}^{\mu}_t, \hat{F}^{\mu}_{t+1})$.*

Theorem 3.5.2 below bounds the regret of π_{LUNAC} (see Algorithm 2). π_{LUNAC} does not require the variation budget V as an input, but we get an improved regret bound with knowledge of V.

Theorem 3.5.2. *Suppose Assumptions 3.3.1 and 3.5.2 hold.*

(i) Suppose the supplier knows V, then N can be chosen optimally as $N^* = \left\lceil \bar{\xi}^{-\frac{1}{4}} V^{-\frac{1}{4}} T^{\frac{1}{4}} \right\rceil$, and Reg$(\pi_{LUNAC}, T) = \tilde{O}(\bar{\xi}^{\frac{5}{4}} V^{\frac{1}{4}} T^{\frac{3}{4}} + \bar{\xi}^{\frac{17}{12}} V^{\frac{3}{4}} T^{\frac{7}{12}})$.

(ii) Suppose the supplier does not know V. Then N can be chosen obliviously as $\hat{N} = \left\lceil \bar{\xi}^{-\frac{1}{4}} T^{\frac{1}{4}} \right\rceil$, and Reg$(\pi_{LUNAC}, T) = \tilde{O}(\bar{\xi}^{\frac{5}{4}} V^{\frac{1}{3}} T^{\frac{3}{4}} + \bar{\xi}^{\frac{17}{12}} V^{\frac{2}{3}} T^{\frac{7}{12}})$.

According to Theorem 3.5.2, the supplier can achieve a sublinear regret bound if $V = \tilde{o}(T)$ and if $V = \tilde{o}(T^{\frac{5}{8}})$ when the supplier knows V and does not know V respectively. Theorem 3.5.1 shows that without discrete approximation, the supplier has sublinear regret bound only if $V = \tilde{o}(T^{\frac{1}{2}})$. The improvement achieved from $V = \tilde{o}(T^{\frac{1}{2}})$ to $V = \tilde{o}(T^{\frac{5}{8}})$ lies in that through approximating the distribution, the supplier indirectly controls the number of epochs of the pricing policy, and thus the discrete approximation may give better regret bound when V is large. The results indicate that the supplier can use π_{LUNAC} not only when the distribution is continuous, but also when the supplier supposes the unknown V to be large under the discrete distribution.

Remark 3.5.1. *When the supplier does not know V, she can combine LUNAC with the BOB framework, which we refer to as LUNAC-N. The implementation details of $\pi_{LUNAC\text{-}N}$ are presented in Appendix B.4, and the upper bound on the regret of $\pi_{LUNAC\text{-}N}$ is presented in Theorem B.4.1.*

Remark 3.5.2. *We have shown in Theorem 3.5.1 and Theorem 3.5.2 that the pricing policies lead to sublinear regret bound only if the variation budget V grows sublinear in T. Here we comment that when the the variation budget is sublinear, we must have the Kolmogorov distance between F_t converges to 0.*

Lemma 3.5.11. *For each time horizon $T = 1, 2, \cdots$, let $\sum_{t=1}^{T} d_K(F_{t-1}, F_t) = V_T$ and $V_T = \tilde{O}(T^{\alpha})$ for some $0 \leq \alpha < 1$. Then we must have $\lim_{T \to \infty} d_K(F_{T-1}, F_T) = 0$.*

Before we end this section, we comment on the difference between our discretization approach and those approaches used for the continuous bandit. The decision set is continuous in a continuous bandit, and it is common to approximate the decision set with a finite set. Instead of finding the optimal decision in the original continuous decision set, an optimal decision is found from the finite set, and then the regret is established through a regularity assumption (e.g., Lipschitz or Holder continuity) on the reward/cost function. We do not directly approximate the continuous decision set. Instead, we approximate the supplier's profit function by finding an approximate distribution $\tilde{F}_t$ for $\hat{F}_t^{\mu}$ (albeit both the true profit function and $\hat{F}_t^{\mu}$ are unknown). Our approach relies on the bilinearity of the supplier's profit function in w and q, and it does not require the regularity assumptions on the objective from the continuous bandit literature.

3.6 Examples of Retailer's Strategies

We investigate several well known data-driven retailer inventory learning policies in this section, and show that our proposed pricing policies achieve sublinear regret for all of them. We emphasize that we do not need to know the retailer's exact inventory policy to achieve sublinear regret, we only mean to illustrate that these popular inventory policies satisfy our key assumption on the total variation of $\hat{F}_{1:T}^{\mu}$. In addition, the examples in this section suppose that the retailer does not have prior knowledge of V. Instead, these examples help provide guidance on refining V in practice.

3.6.1 Sample Average Approximation (SAA)

SAA is arguably the most widely studied approach for data-driven optimization [51], [107]. We let μ_{e} denote the retailer's inventory policy based on SAA. For all $t \in [T]$, let $\hat{F}_t^{\mathrm{e}}$ be the empirical CDF constructed from the (not necessarily i.i.d.) demand samples $(\xi_i)_{i=1}^{t-1}$ defined by $\hat{F}_t^{\mathrm{e}}(x) \triangleq \frac{1}{t-1}\sum_{i=1}^{t-1} \mathbb{1}(\xi_i \leq x)$ for all $x \geq 0$ (note that in period t, the retailer only got access to the demand realizations in the previous $t-1$ periods). Under μ_{e}, given wholesale price w_t, the retailer's order quantity satisfies $q_t^{\mathrm{e}} = q_t(w_t; \hat{F}_t^{\mathrm{e}})$.

We can upper bound the total variation V of the sequence $\hat{F}_{1:T}^{\mathrm{e}}$ for arbitrary $F_{1:T}$ (i.e., the true distribution can be changing arbitrarily).

Proposition 3.6.1. $\mu_e \in \mathcal{M}(\log(T) + 1, T)$.

Next we bound the supplier's regret under π_{LUNA} (if the distribution has finite support) or π_{LUNAC} (if the distribution has continuous support). The proof follows from Theorem 3.5.1 and Proposition 3.6.1.

Theorem 3.6.1. *Suppose the retailer follows μ_e.*

(i) Suppose $F_{1:T}$ have support on $\mathcal{Y}_M$, then Reg$(\pi_{LUNA}, T) = \tilde{O}(\bar{\xi}^{\frac{4}{3}} M^{\frac{1}{3}} T^{\frac{2}{3}})$.

(ii) Suppose $F_{1:T}$ have support on $[0, \bar{\xi}]$, then Reg$(\pi_{LUNAC}, T) = \tilde{O}(\bar{\xi}^{\frac{5}{4}} T^{\frac{3}{4}} + \bar{\xi}^{\frac{17}{12}} T^{\frac{7}{12}})$.

3.6.2 Distributionally Robust Optimization (DRO)

We suppose Assumption 3.4.1 holds for this subsection. The DRO approach is based on the worst-case expected profit over an uncertainty set of demand distributions. We let μ_{r} denote the retailer's inventory learning policy based on DRO. In each period t, the retailer has an uncertainty set $\mathcal{D}_t \subset \mathcal{P}(\Xi)$ that he believes contains the true market demand distribution. We consider uncertainty sets which consist of distributions that are "close" to the empirical distribution $\hat{F}_t^{\mathrm{e}}$, and we measure closeness on $\mathcal{P}(\Xi)$ with the $\phi-$divergence. Recall the $\phi-$divergence, denoted d_ϕ, for distributions F, $G \in \mathcal{P}(\Xi)$ with $F \ll G$ (where $F \ll G$ means F is absolutely continuous with respect to G) is defined by $d_\phi(F, G) = \int_\Xi \phi\,(dF/dG)\,dG$ for a convex function ϕ such that $\phi(1) = 0$. Let $\epsilon_t \geq 0$ be the retailer's confidence level in period t. The retailer's data-driven uncertainty sets under μ_{r} are $\mathcal{D}^{\phi}_{\epsilon_t}(\hat{F}_t^{\mathrm{e}}) \triangleq \{F \in \mathcal{P}(\Xi) : d_\phi(F, \hat{F}_t^{\mathrm{e}}) \leq \epsilon_t\}$. A retailer who is more confident that $\hat{F}_t^{\mathrm{e}}$ is close to the true distribution F_0 should choose a smaller ϵ_t, and vice versa.

The DRO literature has proposed multiple methods for choosing the confidence level ϵ_t, see the review by [108]. One way is to leverage the asymptotic or finite sample performance of the uncertainty set. In other words, we would choose ϵ_t so that the optimal value of the DRO problem gives a finite sample guarantee on the retailer's original stochastic optimization problem.

[109] uses the optimal value of a DRO problem based on smooth $\phi-$divergences to provide asymptotic confidence intervals for the optimal value of the original (full information) stochastic optimization problem. We will evaluate the performance of π_{LUNA} and π_{LUNAC} when ϵ_t is chosen by this method. Let χ_1^2 be a Chi-squared random variable with degree of freedom one.

Theorem 3.6.2. *([109, Theorem 4]) Suppose the following conditions hold:*

(i) The function $\phi : \mathbb{R}_+ \to \mathbb{R}$ is convex, three times differentiable in a neighborhood of 1, and satisfies $\phi(1) = \phi'(1) = 0$.

(ii) There exists a measurable function $M : \Xi \to \mathbb{R}_+$ such that for all $\xi \in \Xi$, $R(\cdot; w, \xi)$ is $M(\xi)-$Lipschitz with respect to some norm $\|\cdot\|$ on Ξ.

(iii) The function $R(\cdot; w, \xi)$ is proper and lower semi-continuous for F_0-almost all $\xi \in \Xi$. For any $\rho \geq 0$, let $\epsilon_t = \rho/t$ for all $t \geq 1$. Then,

$$\lim_{t\to\infty} \mathbb{P}\left(\max_{q\in\Xi} \mathrm{E}_{F_0}\left[R(q; w, \xi)\right] \geq l_t\right) = 1 - \frac{1}{2}\mathbb{P}(\chi_1^2 \geq \rho), \tag{3.16}$$

where $l_t \triangleq \max_{q\in\Xi} \inf_{F\in\mathcal{D}^{\phi}_{\epsilon_t}(\hat{F}^{e}_t)} \mathrm{E}_F\left[R(q; w, \xi)\right]$.

Theorem 3.6.2 says that the optimal value of the retailer's problem with knowledge of F_0 can be lower bounded by l_t (the optimal value of the DRO problem for uncertainty set $\mathcal{D}^{\phi}_{\epsilon_t}(\hat{F}^{\mathrm{e}}_t)$) with probability $1 - \frac{1}{2}\mathbb{P}(\chi_1^2 \geq \rho)$, as $t \to \infty$ for confidence levels $\epsilon_t = \rho/t$. Note that Assumption 3.4.1 must be satisfied for this result to hold.

Let $\alpha \in [0, 1]$ be a confidence level and $\chi^2_{1,\beta}$ denote the $\beta-$quantile of the χ_1^2 distribution. Theorem 3.6.2 suggests that in order to ensure the asymptotic coverage of the optimal value as in Eq. (3.16), the confidence levels should be chosen as

$$\epsilon_t = \chi^2_{1,1-2\alpha}/(t-1), \ \forall t \geq 2. \tag{3.17}$$

Confidence levels chosen in this way are usually overly conservative, and the retailer may choose a smaller ϵ_t in practice. In this case, we get a conservative estimate of V by this method.

Under μ_{r}, in period t the retailer orders

$$q^r_t \in \arg\max_{q\geq 0} \min_{F\in\mathcal{D}^{\phi}_{\epsilon_t}(\hat{F}^{\mathrm{e}}_t)} \mathrm{E}_F\left[R(q; w_t, \xi)\right]. \tag{3.18}$$

The objective in Eq. (3.18) is the retailer's worst-case expected profit. Then, corresponding to q_t^r, the perceived distribution

$$\hat{F}_t^{\mathrm{d}} \in \arg \min_{F \in \mathcal{D}_{\epsilon_t}^{\phi}(\hat{F}_t^{\mathrm{e}})} \mathrm{E}_F\left[R(q_t^r; w_t, \xi)\right],$$

is the distribution in $\mathcal{D}_{\epsilon_t}^{\phi}(\hat{F}_t^{\mathrm{e}})$ which attains the worst-case expected profit (which also depends on w_t).

We consider three widely used $\phi-$divergences: the KL-divergence d_{KL} (where $\phi(x) = x\log(x)$), the χ^2-distance d_{χ^2} (where $\phi(x) = (x-1)^2$), and the Hellinger distance d_H (where $\phi(x) = (\sqrt{x}-1)^2$). For these three, Proposition 3.6.2 upper bounds the total variation V of the sequence $\hat{F}_{1:T}^{\mathrm{d}}$ as a function of $\epsilon_{1:T} = (\epsilon_t)_{t=1}^T$.

Proposition 3.6.2. *Under Assumption 3.4.1, suppose the retailer follows μ_r with confidence levels $\epsilon_{1:T}$.*

(i) If $d_\phi = d_{KL}$, then $\mu_r \in \mathcal{M}\left(\log(T) + 1 + \sum_{t=1}^T \sqrt{2\epsilon_t}, T\right)$.

(ii) If $d_\phi = d_{\chi^2}$, then $\mu_r \in \mathcal{M}\left(\log(T) + 1 + \sum_{t=1}^T \sqrt{\epsilon_t}, T\right)$.

(iii) If $d_\phi = d_H$, then $\mu_r \in \mathcal{M}\left(\log(T) + 1 + 2\sum_{t=1}^T \epsilon_t, T\right)$.

Using the specific choice of $\epsilon_{2:T}$ in Eq. (3.17) and $\epsilon_1 = 1$, Theorem 3.6.3 characterizes the performance of π_{LUNA} and π_{LUNAC} for $d_\phi \in \{d_{KL}, d_{\chi^2}, d_H\}$. Theorem 3.6.3 follows directly from Theorem 3.5.1, Theorem 3.5.2, and Proposition 3.6.2.

Theorem 3.6.3. *Suppose Assumptions 3.4.1 and 3.5.2 hold, and suppose the retailer follows μ_r where $\epsilon_{1:T}$ are chosen as in Eq. (3.17).*

(i) If $d_\phi \in \{d_{KL}, d_{\chi^2}\}$, then $Reg(\pi_{LUNAC}, T) = \tilde{O}\big(\bar{\xi}^{\frac{17}{12}} T^{\frac{11}{12}}\big)$.

(ii) If $d_\phi = d_H$ and F_0 has support on $\mathcal{Y}_M$, then $Reg(\pi_{LUNA}, T) = \tilde{O}(\bar{\xi}^{\frac{4}{3}} M^{\frac{1}{3}} T^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} M^{-\frac{1}{3}} T^{\frac{2}{3}})$.

(iii) If $d_\phi = d_H$ and F_0 has support on $[0, \bar{\xi}]$, then $Reg(\pi_{LUNAC}, T) = \tilde{O}(\bar{\xi}^{\frac{5}{4}} T^{\frac{3}{4}} + \bar{\xi}^{\frac{17}{12}} T^{\frac{7}{12}})$.

3.6.3 Parametric Approach

We continue to suppose Assumption 3.4.1 is in force for this subsection. We additionally suppose that the retailer has a parametric model for F_0 determined by the parameter $\theta \in \mathbb{R}^d$.

Let $\Theta \subset \mathbb{R}^d$ be the set of admissible parameter values, and let $\{F_\theta\}_{\theta\in\Theta}$ be the corresponding parametric family. If F_0 belongs to a parametric family, then Assumption 3.5.2 is not likely to be satisfied (since many parametric distributions such as the normal and exponential distributions have unbounded support). In this case, we relax to the following assumption.

Assumption 3.6.1. *The retailer's order quantity is upper bounded by $\bar{q}$. That is, for any w_t and $\hat{F}_t^\mu$, the retailer's order satisfies $q(w_t; \hat{F}_t^\mu) = \min\{\min\{q : \hat{F}_t^\mu(q) \geq 1 - w_t/s\}, \bar{q}\}$.*

Assumption 3.6.1 states that even if the supplier's price w_t is low enough and the perceived distribution $\hat{F}_t^\mu$ has unbounded support, the retailer will not place arbitrarily large orders. Assumption 3.6.1 is also consistent with practical constraints, e.g., warehouse and transportation capacity.

We consider three specific methods for the parametric setting: (i) maximum likelihood estimation (MLE); (ii) operational statistics; and (iii) the parametric Bayesian approach.

Maximum likelihood Estimation (MLE)

We focus on MLE for the exponential family, where a distribution belongs to the exponential family if its probability density function $f(x;\theta)$ for x in its support can be written as:

$$f(x;\theta) = h(x) \exp\big(\eta(\theta)^T \cdot T(x) - A(\theta)\big). \tag{3.19}$$

In Eq. (3.19), $\eta(\theta)$ is the natural parameter, $T(x)$ is the sufficient statistic, $h(x)$ is the base measure, and $A(\theta)$ is the log-partition function which normalizes the density function. The exponential family includes the Poisson and Categorical distributions (for discrete demand), and the Normal and Exponential distributions (for continuous demand).

We let μ_{m} denote the retailer policy based on MLE. Under μ_{m}, the retailer produces an estimate θ_t of θ in each period t by maximizing the likelihood function of the past demand samples. This procedure has a special form for the exponential family. Let $\mu \triangleq \mathrm{E}_{F_\theta}[T(\xi)]$, then the MLE for μ based on demand samples $(\xi_i)_{i=1}^{t-1}$ is:

$$\mu_t = \frac{\sum_{i=1}^{t-1} T(\xi_i)}{t-1},\ t \geq 2.$$

One can then obtain the estimator θ_t for θ through the estimator μ_t for μ by the relationship between μ and θ, which depends on the particular distribution. Under μ_{m}, in each period $t \in [T]$, the retailer's perceived distribution is the fitted distribution

$$\hat{F}_t^{\mathrm{m}}(x) = \begin{cases} F_{\theta_t}(x), & 0 \leq x < \bar{q}; \\ 1, & x \geq \bar{q}. \end{cases} \tag{3.20}$$

This sequence satisfies Assumption 3.3.1 and the discontituity in $\hat{F}_t^m$ is introduced by Assumption 3.6.1. The retailer then orders $q_t^{\mathrm{m}} = q_t(w_t, \hat{F}_t^{\mathrm{m}})$ where $q_t(w_t, \hat{F}_t^{\mathrm{m}}) = \min\{\min\{q : F_{\theta_t}(q) \geq 1 - w/s\}, \bar{q}\}$.

We will investigate the variation of $\hat{F}_{1:T}^m$ for some canonical distributions in the exponential family. Let $\mathrm{P}(\lambda)$ denote the Poisson distribution with mean λ; let $\mathrm{C}(M)$ denote the categorical distribution with support size M; let $\mathrm{E}(\lambda)$ denote the exponential distribution with rate λ; and let $\mathrm{N}(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2.

In Proposition 3.6.3, we derive the total variation of $\hat{F}_{1:T}^m$. Note the categorical distribution has bounded support so it automatically satisfies Assumption 3.6.1.

Proposition 3.6.3. *Suppose the retailer follows μ_m.*

(i) If $F_0 = \mathrm{P}(\lambda)$, then $\mu_m \in \mathcal{M}\left((\ln(T) + 1)(4\ln(T) + 2\lambda), T\right)$ with probability at least $1 - 1/T$.

(ii) If $F_0 = \mathrm{C}(M)$, then $\mu_m \in \mathcal{M}(\ln(T+1), T)$.

(iii) If $F_0 = \mathrm{E}(\lambda)$, then $\mu_m \in \mathcal{M}(16\ln(2T^2) - 1 + (1 + 2\ln(2T^2))(\ln(T) + 1), T)$ with probability at least $1 - 1/T$.

(iv) If $F_0 = \mathrm{N}(\mu, \sigma^2)$, with σ^2 known and μ unknown to the retailer, then $\mu_m \in \mathcal{M}\left(1 + \frac{1}{\sigma}\sqrt{(\ln(T) + 1)(\mu^2 + 2\sigma^2\ln(2T^2))}, T\right)$ with probability at least $1 - 1/T$.

The next result on the supplier's regret bound follows directly from Theorem 3.5.1, Theorem 3.5.2, and Proposition 3.6.3.

Theorem 3.6.4. *Suppose the retailer follows μ_m.*

(i) Suppose $F_0 = \mathrm{P}(\lambda)$, then $Reg(\pi_{LUNAC}, T) = \tilde{O}(\bar{q}^{\frac{5}{4}}T^{\frac{3}{4}} + \bar{q}^{\frac{17}{12}}T^{\frac{7}{12}})$ with probability at least $1 - 1/T$.

(ii) Suppose $F_0 = \mathrm{C}(M)$*, then* $Reg(\pi_{LUNA}, T) = \tilde{O}(\bar{q}^{\frac{4}{3}} M^{\frac{1}{3}} T^{\frac{2}{3}})$*.*

(iii) Suppose $F_0 = \mathrm{E}(\lambda)$*, then* $Reg(\pi_{LUNAC}, T) = \tilde{O}(\bar{q}^{\frac{5}{4}} T^{\frac{3}{4}} + \bar{q}^{\frac{17}{12}} T^{\frac{7}{12}})$ *with probability at least* $1 - 1/T$*.*

(iv) Suppose $F_0 = \mathrm{N}(\mu, \sigma^2)$*, then* $Reg(\pi_{LUNAC}, T) = \tilde{O}(\bar{q}^{\frac{5}{4}} T^{\frac{3}{4}} + \bar{q}^{\frac{17}{12}} T^{\frac{7}{12}})$ *with probability at least* $1 - 1/T$*.*

Operational Statistics

Here we suppose $F_0 = \mathrm{E}(\lambda)$, the exponential distribution with an unknown rate $\lambda > 0$. [42], [43] propose the operational statistics approach for the retailer facing exponential demand with unknown rate. In this approach, the retailer first specifies a class of admissible policies, then finds a policy within this class to maximize his out-of-sample expected profit.

We let μ_{o} denote the retailer's inventory policy based on operational statistics, which is implemented as follows. According to [42], given wholesale price w_t and i.i.d. demand samples $(\xi_i)_{i=1}^{t-1}$, the retailer's order quantity that maximizes his out-of-sample expected profit is $(t-1)\left(\left(\frac{s}{w_t}\right)^{\frac{1}{t}} - 1\right) \frac{\sum_{i=1}^{t-1} \xi_i}{t-1}$ for $t \geq 2$. The order quantity is directly determined by the data, and its derivation does not involve estimating λ. However, we still can find a sequence of distributions $\hat{F}^{\mathrm{o}}_{2:T}$ satisfying Assumption 3.3.1 that map w_t to the order quantity under operational statistics. For all $t \geq 2$, define the rates λ_t so that $\frac{1}{\lambda_t} = \frac{(t-1)\left(\left(\frac{s}{w_t}\right)^{\frac{1}{t}} - 1\right)\left(\frac{\sum_{i=1}^{t-1} \xi_i}{t-1}\right)}{\ln\left(\frac{s}{w_t}\right)}$, and then set the perceived distribution to be

$$\hat{F}^{\mathrm{o}}_t(x) = \begin{cases} \mathrm{E}(\lambda_t)(x), & 0 \leq x < \bar{q}, \\ 1, & x \geq \bar{q}. \end{cases}$$

Then, under μ_{o} the retailer equivalently solves $q^{\mathrm{o}}_t = q(w_t, \hat{F}^{\mathrm{o}}_t)$ where $q(w_t, \hat{F}^{\mathrm{o}}_t) = \min\{\min\{q : \mathrm{E}(\lambda_t)(q) \geq 1 - w/s\}, \bar{q}\}$ for all $t \geq 2$.

We now derive the total variation of $\hat{F}^{\mathrm{o}}_{1:T}$.

Proposition 3.6.4. *Suppose* $F_0 = \mathrm{E}(\lambda)$*, and the retailer follows* μ_o*. Then,* $\mu_o \in \mathcal{M}\left(21 + 40 \ln(T) + 4(\ln(T))^2, T\right)$ *with probability at least* $1 - 1/T$*.*

Theorem 3.6.5. *Suppose $F_0 = \mathrm{E}(\lambda)$, and the retailer follows μ_o. Then, Reg$(\pi_{LUNAC}, T) = \tilde{O}(\bar{q}^{\frac{5}{4}}T^{\frac{3}{4}} + \bar{q}^{\frac{17}{12}}T^{\frac{7}{12}})$ with probability at least $1 - 1/T$.*

Parametric Bayesian approach

We now suppose demand is $\mathrm{E}(\Lambda)$ where the rate Λ is random and has a gamma prior distribution f_Λ with parameter $\alpha, \beta > 0$, i.e., $f_\Lambda(\lambda) = \frac{(\beta/\lambda)^{\alpha+1}}{\beta\Gamma(\alpha)} \exp\{-\beta/\lambda\}$. Exponential demand distributions with a gamma prior are widely studied in the OM literature [110].

We let μ_{b} denote the retailer's inventory learning policy under the Bayesian approach. Given w_t, the retailer orders

$$\arg\max_{q \geq 0} \int_0^\infty \mathrm{E}_{\mathrm{E}(\Lambda)}\left[R(q; w_t, \xi)\right] f_{\Lambda|(\xi_i)_{i=1}^{t-1}}(\lambda)\, d\lambda, \tag{3.21}$$

where $f_{\Lambda|(\xi_i)_{i=1}^{t-1}}$ is the posterior density function with respect to the demand samples from the previous $t-1$ periods. [111] show that the retailer's optimal order quantity as a solution to Eq. (3.21) is $\left(\beta + \sum_{i=1}^{t-1} \xi_i\right)\left(\left(\frac{s}{w_t}\right)^{1/(\alpha+t-1)} - 1\right)$. For all $t \in [T]$, define the rates λ_t so that

$$\frac{1}{\lambda_t} = \frac{\left(\beta + \sum_{i=1}^{t-1} \xi_i\right)\left((s/w_t)^{\frac{1}{\alpha+t-1}} - 1\right)}{\ln(s/w_t)},$$

and then set

$$\hat{F}_t^b = \begin{cases} \mathrm{E}(\lambda_t)(x), & 0 \leq x < \bar{q}, \\ 1, & x \geq \bar{q}. \end{cases}$$

Then, $\hat{F}_t^b$ is the retailer's perceived distribution that maps the supplier's wholesale price w_t to the retailer's order quantity q_t^b via $q_t^b = q(w_t, \hat{F}_t^b)$ where $q(w_t, \hat{F}_t^b) = \min\{\min\{q : \mathrm{E}(\lambda_t)(q) \geq 1 - w/s\}, \bar{q}\}$.

We derive the total variation of $\hat{F}_{1:T}^b$ similarly to Proposition 3.6.4, and so we omit the proof.

Proposition 3.6.5. *Suppose demand is exponentially distributed with mean $1/\lambda$ and the retailer follows μ_b. Then, $\mu_b \in \mathcal{M}(18(\alpha+1) + (40\alpha+39)\ln(T) + 4(\alpha+1)(\ln(T))^2, T)$ with probability at least $1 - 1/T$.*

Next we bound the regret of π_{LUNAC} when the retailer follows μ_{b}.

Theorem 3.6.6. *Suppose the retailer follows μ_b. Then, Reg$(\pi_{LUNAC}, T) = \tilde{O}(\bar{q}^{\frac{5}{4}} T^{\frac{3}{4}} + \bar{q}^{\frac{17}{12}} T^{\frac{7}{12}})$ with probability at least $1 - 1/T$.*

3.7 Numerical Experiments

3.7.1 Empirical Performance

We evaluate the empirical performance of π_{LUNA} when the true market demand distributions $F_{1:T}$ are discrete. In order to control the total variation of $\hat{F}^{\mu}_{1:T}$, we directly construct $\hat{F}^{\mu}_{1:T}$ as follows. For $t \in [T]$, we set $\hat{F}^{\mu}_{t}$ to be the CDF of a Bernoulli random variable which takes values 0 and 1 with probabilities $p_{t,0}$ and $p_{t,1} = 1 - p_{t,0}$, respectively, and let

$$p_{t,0} = \frac{1}{2} + \frac{3}{10} \sin \frac{5V\pi t}{3T}, \; t \in [T], \tag{3.22}$$

for fixed $V > 0$. Then, the total variation satisfies

$$\sum_{t=1}^{T-1} d_K(\hat{F}^{\mu}_{t}, \hat{F}^{\mu}_{t+1}) = C\, V,$$

for some constant $C > 0$.

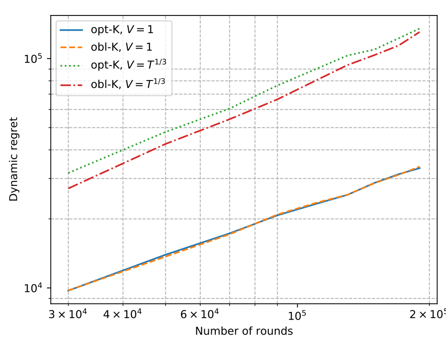


Figure 3.1. Performance of π_{LUNA} for discrete distributions

Figure 3.1 shows the dynamic regret of π_{LUNA} as a function of the number of rounds when K is optimally chosen (assuming the supplier knows V, with shorthand 'opt-K') and obliviously chosen (assuming the supplier does not know V, with shorthand 'obl-K'). We take different values of V to compare the growth rate of the regret for the same pricing policy. The regrets are plotted on a log-log scale, so the slope in this plot corresponds to the exponent of the regret, i.e., the slope is α if the regret grows in $\Theta(T^{\alpha})$.

We see that the slope roughly matches our theoretical results (see Theorem 3.5.1). When $V = 1$, the regret bounds corresponding to opt-K and obl-K overlap. When $V = T^{1/3}$, opt-K grows more slowly than obl-K but has a smaller constant term than opt-K (we did not optimize for the constant terms). Also notice that the gap in the regret bounds between opt-K and obl-K is small (i.e., the regret bounds for opt-K and obl-K are respectively $\tilde{O}(V^{\frac{1}{3}}T^{\frac{2}{3}})$ and $\tilde{O}(V^{\frac{2}{3}}T^{\frac{2}{3}})$, and the gap is $V^{\frac{1}{3}}$). Even when K is optimally chosen, obl-K can have better performance than opt-K within a large range of T (in our case, $T \leq 2 \times 10^5$) because of the constant terms dominating the regret bounds.

3.7.2 Comparison between Different Algorithms

In this subsection, we compare π_{LUNA} with some pricing policies that are designed for non-stationary bandits. Specifically, we compare π_{LUNA} with the following benchmarks:

1. The Exp3.S algorithm by [61], which is designed for non-stationary multi-armed bandits with known variation budget B_T. Notice we distinguish B_T from our variation budget V because B_T refers to the norm of variation in the mean bandit feedback and V refers to the total Kolmogorov variation in the sequence $\hat{F}^{\mu}_{1:T}$. The regret upper bound for Exp3.S is $\tilde{O}(d^{\frac{1}{3}}B_T^{\frac{1}{3}}T^{\frac{2}{3}})$ with d arms, see [61].

2. The deterministic non-stationary bandit algorithm proposed by [65] for multi-armed bandits with unknown variation budget B_T. The regret upper bound for this algorithm is $\tilde{O}(d^{\frac{1}{2}}T^{\frac{1}{2}} + d^{\frac{1}{3}}B_T^{\frac{1}{3}}T^{\frac{2}{3}})$.

3. The Master+UCB1 algorithm proposed by [66] for non-stationary stochastic bandits with unknown variation budget B_T. The regret upper bound is $\tilde{O}(B_T^{\frac{1}{3}}T^{\frac{2}{3}})$.

We note that all of these pricing policies are designed for problems with finitely many admissible decisions. For fair comparison, we developed a version of π_{LUNA} that works when $\mathcal{W}$ is finite which we call π_{LUNAF} (see Appendix B.5). Since Exp3.S requires the variation budget B_T as an input, we calculate V in our problem and simply let $B_T = V$.

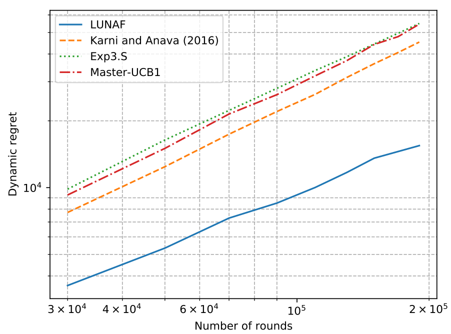


Figure 3.2. Performance comparison when $p_{m,t}$ is simulated according to (3.22)

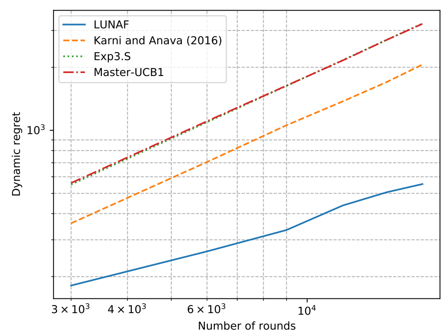


Figure 3.3. Performance comparison when F_t is simulated according to (3.23) and the retailer's policy is μ_{e}

We compare the regret of these benchmark algorithms with π_{LUNAF} in Figure 3.2 and Figure 3.3. In Figure 3.2, we directly simulated the retailer's ordering decisions by setting $\hat{F}_t^{\mu}$ as in Eq. (3.22). In Figure 3.3, we set the true distribution F_t to be Bernoulli which takes values 0 and 1 with probabilities p_t and $1 - p_t$ respectively where p_t is also determined by Eq. (3.22), i.e.,

$$p_t = \frac{1}{2} + \frac{3}{10} \sin \frac{5V\pi t}{3T}, \ t \in [T], \tag{3.23}$$

in which case the true market demand distribution is non-stationary. We also suppose the retailer follows μ_{e}. In both experiments, we let $\mathcal{W}$ contain $d = \left\lceil T^{\frac{1}{2}} \right\rceil$ equally spaced prices lying in $[0, s]$.

From Figures 3.2 and 3.3, we see that π_{LUNAF} outperforms the benchmarks, and the performance of the benchmarks is relatively close to each other. These results suggest that the supplier benefits from using the structure of the profit function in her pricing policy, instead of applying a black box algorithm. In addition, based on the results in Figure 3.3, we see π_{LUNAF} still performs well even when the true market demand distribution is non-stationary.

3.7.3 Experiment on Semi-synthetic Data Set

For our final experiment, we collected the weekly sales data of avocados in California from 2020 to 2022 [112]. Avocado sales can be non-stationary and vary from month to month, see, e.g., [62]. In order to approximate this non-stationarity, we first group the weekly sales data by month $m = \{1, \ldots, 12\}$. Then, to generate the daily sales in month m, we divide the weekly sales in month m by 7 and treat it as a sample of daily demand in month m. We repeat this procedure for all the weeks from 2020 to 2022 to get demand samples for each month of the year. Finally, we divide the daily sales by $1,000,000$ and round it to the nearest integer to build an approximate discrete daily demand distribution for avocados (in millions of units). Given the daily demand samples for each month, we then bootstrap the daily demand for a planning horizon of T days (assuming the first day in the horizon starts on Jan 1st). In this way, we generate random demand realizations for the retailer.

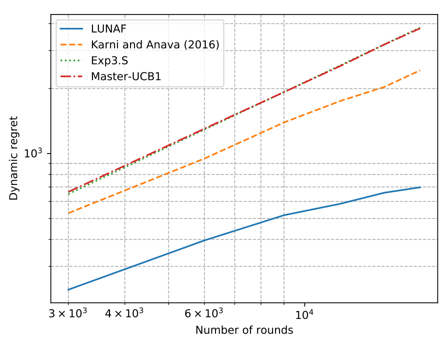


Figure 3.4. Performance comparison on semi-synthetic data

We suppose the retailer follows μ_{e}, and $\mathcal{W}$ has cardinality $d = \left\lceil T^{\frac{1}{2}} \right\rceil$. We compare the performance of the pricing policies for this setting in Figure 3.4. We see that π_{LUNAF} outperforms the other policies in this setting as well. Exp3.S and Master-UCB1 have almost identical performance here, and the deterministic non-stationary bandit algorithm by [65] outperforms both Exp3.S and Master-UCB1. This application further demonstrates that our pricing policy performs well even for non-stationary demand distributions.

3.8 Conclusion

In this paper, we studied the supplier's pricing problem facing a retailer who is learning the demand distribution and employs data-driven inventory learning policies. We model the non-stationarity of the retailer's inventory decisions through the non-stationarity of his "perceived" distributions. Then, we use the Kolmogorov distance to measure the variation of the retailer's perceived distributions and identify a tractable class of retailer policies. For both discrete and continuous demand distributions, we proposed pricing policies for the supplier and derived sublinear regret upper bounds. Our main conclusion is that the supplier can achieve asymptotically vanishing regret, even when the retailer is also learning the demand distribution, as long as the retailer's inventory policies belong to a reasonable class with bounded variation.

Much of the literature on optimization and learning in OM focuses on learning the random demand or unknown demand-price relationship. However, our work investigates the important problem of learning the learning policies implemented by a secondary agent in a multi-agent setting. This study brings new perspectives into learning in multi-agent problems in supply chain and inventory management, where the controller must learn to react to the learning policies by other agents in the system.

At the same time, we acknowledge some directions for future research. First, it is worth investigating information-theoretic lower bounds on the supplier's regret in our problem setting. Second, it may be possible to improve the supplier's regret bound when the non-stationarity has additional structure (e.g., seasonal demand patterns).

4. REIMBURSEMENT POLICY AND DRUG SHORTAGES: IMPLICATIONS FROM PHARMACEUTICAL SUPPLY CHAIN CONTRACTING

4.1 Synopsis

We investigate the link between shortage on generic injectable drugs and the reimbursement policy change in Medicare Part B taking place in 2005. The investigation of the policy impact entails a detailed examination of the drug supply chain, which involves multiple parties including drug manufacturers, GPOs, healthcare providers, and reimbursement programs (such as Medicare). Central to the drug supply chain are the GPOs, who represent their healthcare provider members to set wholesale prices with manufacturers for these generic drugs. Unlike for brand-name drugs, whose manufacturers generally dominate the wholesale pricing decisions due to their monopolistic position in the market, for generic drugs, GPOs possess high bargaining power, and hence generally dominate the wholesale pricing decisions (see Section 4.3 for a detailed discussion). The wholesale prices are key decisions affecting drug shortages. In modeling the drug supply chain, we capture important features such as GPOs' self-serving and member-serving characteristics, which derive from a major comment from the FDA staff. We also model supply allocations from drug manufacturers. The analytical model serves as a basis for examining the policy impact on drug shortages.

Based on the model, we first conduct theoretical analysis to uncover the underlying impact of the reimbursement policies on the wholesale pricing decisions and derive insights on the equilibrium drug wholesale prices and shortage statuses. We find that under the ASP policy, two effects, the *free-ride* effect and the *coordination* effect, work in opposite directions in determining the wholesale price decisions and the corresponding shortage statuses. The interplay between these effects determines the overall impact of the ASP policy on drug shortages. Further, we capture key factors that influence the interplay of these two effects. These factors, i.e., drug supply difficulty and demand-side market concentration, are shown to be highly influential on drug shortages. Overall, our results show that the ASP policy actually possesses resilience to drug shortages as a consequence of the interplay between the free-ride and coordination effects. In addition, we examine how the drug shortage changes with respect to drug-specific parameters and industry/policy parameters. While the former provides insights about characteristics that put a drug more at risk of shortages under different reimbursement policies, the latter provides important industry and policy implications.

In addition to the theoretical analysis, we further conduct numerical analysis that incorporates real data to deepen our insights. To overcome the challenge of non-transparency in pharmaceutical data, especially GPO-related data, we compile a novel dataset with the aid of machine learning techniques, based on which, we estimate the number of GPOs in the market of each shortage drug to understand the degree of the free-ride effect in the market. We further examine multiple levels of supply difficulty to understand the degree of the coordination effect. Given the limited work on quantifying the impact of the reimbursement policies, our numerical study provides quantitative insights on the policy impact, especially, the ASP policy's resilience to drug shortages.

Through the theoretical and numerical analyses, our study contributes to the understanding of the reimbursement policies, their impact on drug shortages, and the complex drug shortage problem in general. Furthermore, our study also examines how industry/policy parameters, such as the reimbursement markup percentage (currently set at 6%) and the GPOs' price consciousness, affect drug shortages, which provides important insights regarding the current industry practices and potential policy guidance, as discussed in the conclusion section.

4.2 Industry Context and Related Literature

While other industries often aim at revenue maximization and hence some supply shortages are commonly accepted if they lead to a higher revenue, persistent drug shortages in the pharmaceutical industry often lead to significant social welfare loss, such as delayed or canceled treatments, undesired treatment outcomes, and even patient deaths. This is especially true for many of the shortage drugs that are medically necessary and lack desirable substitutes (e.g., pertinent cancer drugs). Indeed, drug shortages have been one of the most challenging problems for the pharmaceutical industry and the government, drawing tremendous attention from the society. There have been extensive white papers and government reports discussing the situations and possible causes, e.g., [5]–[7], [14], [17], [113]–[119], and [120].

Academic literature on drug shortages has been limited although there has been increasing interest from the operations and supply chain field. [121] studied inventory policies to mitigate the impact of drug shortages. [122] provided a review of the drug shortages problem and used analytical and data analyses to show that shortages can be mitigated by properly designed Pareto-improving drug procurement contracts. [19] developed stochastic programs to compare several supply chain designs to reduce shortages. [123] empirically showed that the FDA's mandated reporting of foreseeable production disruptions alleviates shortages. The FDA encouraged more research on this subject [5], and launched a drug shortage program around 2012 and another task force in 2018 to continue efforts to resolve the problem [16]. Our interactions with FDA staff revealed that these efforts are only the starting point of the solution process since the problem is complex and many issues are not yet explored.

For the heavily regulated pharmaceutical industry, the Medicare drug reimbursement policy is one of the most important regulations affecting drug wholesale pricing decisions. Since the introduction of the ASP policy, there have been some analyses on its impact. A large body of these studies focused on evaluating the policy's impact on medical practices, e.g., [124], [125]. Some health economics studies examined the policy's impact on drug markets, e.g., [126], [127]. More related to our paper are the few studies that compared drug prices under the AWP and ASP policies, but they yielded mixed messages. Specifically, [128] analyzed price data and concluded that drug prices under ASP are substantially lower. [129], in contrast, argued that wholesale prices are higher under ASP than AWP because (1) drugs are reimbursed at fixed prices under AWP, hence incentivizing providers to lower wholesale prices to increase profit margins; and (2) this indicates a more intensive price competition under AWP compared to ASP, and hence the wholesale prices under ASP would be higher. However, this study lacks theoretical modeling, and the argument only includes part of the picture captured by our model. [130] conducted modeling and data analysis, also concluding that the wholesale prices under ASP are higher, but they attributed the reason to the lagged markup of the wholesale price as the reimbursement price.

While there is a lack of consensus on the impact of the reimbursement policy on drug prices, further inquiries about the link between the reimbursement policy and drug shortages are even scarcer. [18] concluded that the policy change leads to lower prices (for which, as just discussed, the literature still lacks consensus) and lower prices cause more shortages. However, the high-level economic model in [18] does not consider GPOs nor their interactions with other supply chain parties, and in fact does not differentiate wholesale prices and reimbursement prices, the two key parameters in the reimbursement policy. In contrast, recognizing that the reimbursement policy affects shortages through affecting supply chain parties' decisions, our model captures the details in the drug supply chain and the tradeoffs in the wholesale price decisions. This realistic granularity allows us to derive new insights, such as the free-ride and coordination effects and the key influential factors of supply difficulty and demand-side market concentration. Correspondingly, we believe that our model and analysis provide a more comprehensive picture of the reimbursement policy's impact on drug wholesale prices and shortages.

It is worth noting that our analysis focuses on shortage drugs reimbursed through the Medicare program. While shortage drugs may be reimbursed through private insurers, Medicare covers a significant portion of these drugs' reimbursement. For example, [17] examined a major type of shortage drugs, the sterile injectable oncology drugs, and stated "in most cases,..., reimbursement for these drugs is under Medicare Part B." The study then used Medicare Part B data to analyze the drug shortage problem. The close link between Medicare's drug reimbursement policy and drug shortages has also been discussed in reports such as [20] and [6], as well as a listening session organized by the FDA drug shortage task force in 2019. In addition, "*it is quite common for private insurers to mimic Medicare reimbursement*, albeit with a lag [18], [131]." Since private insurers' data are often unavailable and Medicare data are public, government and academic research often use Medicare's practice and data when examining drug reimbursement [e.g., 17]–[19], [122], [130].

Finally, our study is also related to the limited literature on pharmaceutical supply chains. For example, [132], [133], [134], and [135] provided extensive reviews of pharmaceutical supply chains. [136] and [137] investigated the impact of information transparency and information sharing through contracts in pharmaceutical distribution. [138] compared distribution agreements. [139] and [140] examined drug manufacturers' quality performance issues.

4.3 Model

In this section, we develop a parsimonious model to capture the most essential elements pertaining to the impact of the reimbursement policy on drug shortages. The model focuses on a generic drug on shortage reimbursed through Medicare. The model considers its supply chain, which consists of *drug manufacturers*, *GPOs*, which represent their respective *healthcare provider members* to determine the wholesale prices with the drug manufacturers, and the *government*, which provides the healthcare providers with reimbursement through the Medicare program according to the reimbursement policy. Central to the supply chain are the GPOs. While for brand-name drugs, manufacturers usually have pricing power, for the generic drugs that the shortages primarily concern, GPOs typically have high bargaining power as the top five GPOs in the U.S. possess 85-90% of the market [17]. Such high bargaining power "... allows GPOs to negotiate lower prices [6]" and, hence, the GPOs largely serve as the wholesale price setters for these generic drugs. Once the GPOs set the wholesale prices with manufacturers in drug procurement contracts, the wholesale prices usually remain the same during the contract durations that typically last 2-3 years [122], [141]. Since the wholesale prices are fixed over a relatively long period of time, we herein consider a single-period model, with the period corresponding to the contract duration.[1]

The reimbursement policy affects drug shortages through interactions of the supply chain parties: Under the government's reimbursement policy, the GPOs pool demands from their respective members and set the wholesale prices. Based on the wholesale prices, manufacturers determine their supply levels and allocations to the GPOs. The relationship between the supplies and demands determines the shortage occurrences. Figure 4.1 illustrates the drug supply chain model and parameters that will be detailed next. For convenience of exposition, we next introduce the model in a backward manner from downstream supply chain parties to upstream parties.

[1]↑Contracts in reality often have different start and end dates. For tractability, we focus on a single contract period, which is common in supply chain contract studies (see, e.g., the studies reviewed in [142]).

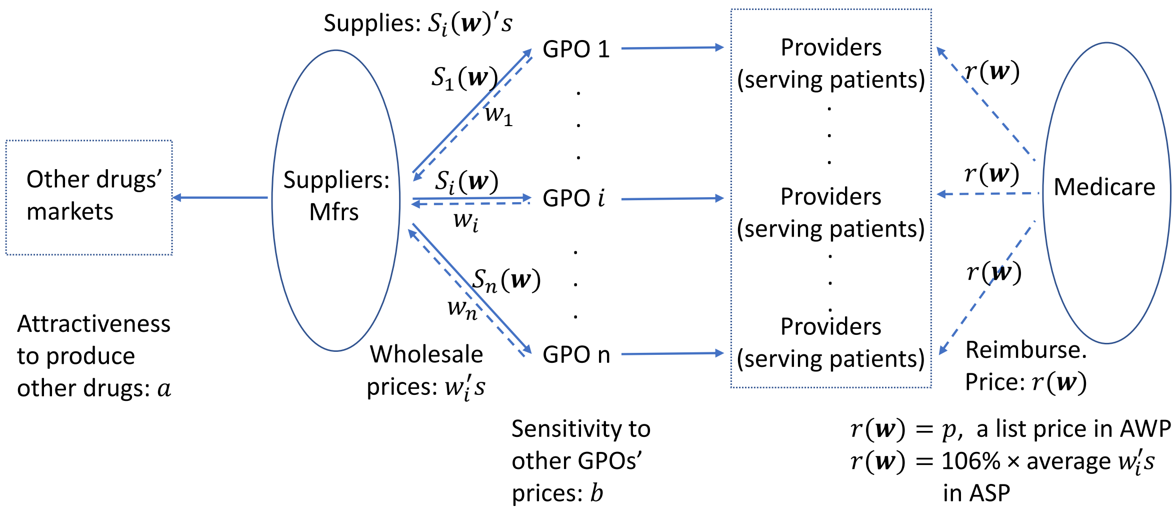


Figure 4.1. Model and Parameters

4.3.1 Government

The government reimburses drug sales at healthcare providers according to the reimbursement policy. Before 2005, the government used the AWP policy, where the reimbursement price of a drug was a fixed published price, denoted by p, which is independent of the drug's actual wholesale prices. After 2005, the government switched to the ASP policy, where the reimbursement price is 6% above the average of the drug's wholesale prices in the market. In other words, if $\bar{\boldsymbol{w}}$ denotes the average wholesale price, then the reimbursement price is $106\% \times \bar{\boldsymbol{w}}$. More generally, in our model, we write the reimbursement price under the ASP policy as $\theta\bar{\boldsymbol{w}}$, where θ stands for the reimbursement percentage. While $\theta = 106\%$ currently, this general expression allows flexibility to examine the impact of varying values of θ.

4.3.2 Healthcare Providers

Healthcare providers procure drugs through their respective GPOs' contracts to treat patients. Since most of the shortage drugs are medically necessary and lack substitutes (otherwise, shortages would not be a serious issue), their demand naturally lacks price-elasticity. The fact that drugs are typically not paid by patients but by patients' insurance or other reimbursement mechanisms further reduces patients' price sensitivity. Thus, drug demand is essentially more driven by need than price and it is reasonable to assume a price-independent demand in our analysis (see also [5], [17], [122], [143], [144] for similar discussions). In addition, demand for the medically necessary shortage drugs typically has low variability. [122] examined some representative shortage drugs and found that the average standard deviation to mean ratio (i.e., coefficient of variation) is only about 10%. Therefore, without loss of generality and for model parsimony, we assume a deterministic demand in our analysis.

Facing patients' demand, healthcare providers purchase the drug through their respective GPOs. In the long run (i.e., a typical contract duration of 2-3 years), the providers' purchase quantity matches the patients' demand. In reality, a drug's market is often dominated by a few large GPOs. We assume that n symmetric GPOs command the market, and each GPO faces a demand volume of d. This symmetry assumption maintains tractability and allows us to focus on the most important tradeoffs. Later, we will show that this symmetry consideration is consistent with real data and an asymmetry case would most likely enhance our results (Section 4.6).

4.3.3 GPOs

Facing the demand volume d, a GPO represents its members to set a wholesale price with manufacturers. It is important to recognize that in setting the wholesale price, a GPO must consider not only its own benefit, but also its members' benefit (an important point raised by FDA staff during our discussion of this research). Correspondingly, we model a GPO's objective function to comprise two parts, one reflecting its own profit and the other, its members' profit. A GPO's own profit mainly derives from a contract administration fee as a percentage commission on purchases of the GPO's members ([122], [132]; see also the "safe harbor" provisions of the 1987 amendment that governs the commission). Thus, a higher wholesale price and the resultant higher purchase payment would increase a GPO's own profit. In the meantime, considering its members' profit, a GPO must also be price conscious because a GPO's main function is to pool members' purchasing volumes to secure a favorable wholesale price for them, and the fulfillment of this function is tied to a GPO's reputation (see, e.g., the discussion in [6]). Notably, a higher price consciousness does not merely mean that a GPO should make the maximum effort to push down the wholesale price: An excessively low wholesale price hinders the GPO's members from obtaining sufficient supply (i.e., manufacturers may allocate the supply to members of other GPOs), hence hurting the members. Further, under the ASP policy, a lower wholesale price also implies a lower average wholesale price and, consequently, a lower reimbursement price. Therefore, while a GPO's motive to raise its own profit exerts an upward pressure on the wholesale price, the price consciousness is not a pure downward pressure on the wholesale price, but a motive for a GPO to secure a wholesale price considering its members' profit. Hence, to capture a comprehensive picture of a GPO's decision making in setting the wholesale price, we must incorporate the nuances in both motives. We do so through a *price consciousness* parameter, expressing a GPO's objective as a weighted sum of a GPO's own profit and its members' profit.

Specifically, let w_i denote the wholesale price set by GPO i, $\boldsymbol{w}$ denote the vector of the wholesale prices set by all GPOs, and $\bar{\boldsymbol{w}}$ denote the average wholesale price of all GPOs. Each GPO i faces a demand of d and obtains a supply of $S_i(\boldsymbol{w})$ for its members. $S_i(\boldsymbol{w})$ is random and dependent on the wholesale prices, whose detailed expression is delayed to the next subsection where we discuss how manufacturers allocate supplies. Thus, the total expected sales volume at GPO i is $\mathbb{E}[S_i(\boldsymbol{w}) \wedge d]$, where $\wedge$ represents the minimum of the components (i.e., sales are the minimum of supply and demand). Correspondingly, the total purchase payment made by GPO i's members is $w_i\mathbb{E}[S_i(\boldsymbol{w})\wedge d]$, the product of the unit wholesale price and the expected sales. The total profit of GPO i's members is $[r(\boldsymbol{w})-w_i]\mathbb{E}[S_i(\boldsymbol{w})\wedge d]$, where $r(\boldsymbol{w})$ represents the reimbursement price, so $r(\boldsymbol{w}) - w_i$ is the profit for the GPO's members per unit of sales. Clearly, under AWP, $r(\boldsymbol{w}) = p$, the exogenously published reimbursement price, and under ASP, $r(\boldsymbol{w}) = \theta\bar{\boldsymbol{w}}$, the marked-up average wholesale price.

Since a GPO's own profit is proportional to the purchase payment made by the GPO's members and a GPO's price consciousness is based on its members' total profit, we capture the two motives by modeling a GPO's payoff as a weighted sum:

$$\begin{aligned} &\alpha[r(\boldsymbol{w}) - w_i]\mathbb{E}[S_i(\boldsymbol{w}) \wedge d] + (1-\alpha)w_i\mathbb{E}[S_i(\boldsymbol{w}) \wedge d] \\ =&\alpha[r(\boldsymbol{w}) - (2\alpha - 1)w_i/\alpha]\mathbb{E}[S(\boldsymbol{w}) \wedge d], \end{aligned} \tag{4.1}$$

where $0 \leq \alpha \leq 1$ is the weight a GPO places on its members' profit and $1 - \alpha$ is the weight a GPO places on the total purchase payment.

Let $\pi_i(w_i, \boldsymbol{w}_{\text{-}i})$ denote GPO i's payoff when the GPO's wholesale price is w_i and the vector of other GPOs' wholesale prices is $\boldsymbol{w}_{\text{-}i}$. GPO i then sets the wholesale price w_i to maximize its expected payoff:

$$\max_{w_i \geq 0} \pi_i(w_i, \boldsymbol{w}_{-i}) = [r(\boldsymbol{w}) - \beta w_i]\mathbb{E}[S_i(\boldsymbol{w}) \wedge d], \tag{4.2}$$

which is strategically equivalent (i.e., yielding the same pricing decisions and shortage statuses) to equation (4.1) with $\beta = (2\alpha - 1)/\alpha$. We refer to β as a GPO's *price consciousness* parameter, with a higher β indicating a higher price consciousness. That $0 \leq \alpha \leq 1$ implies $\beta \leq 1$. In addition, $\beta > 0$ (otherwise, the payoff function (4.2) would be monotonically increasing in w_i, leading to an unrealistic, arbitrarily large wholesale price). To avoid triviality and maintain realisticity, we further require the unit profit and supply to be non-negative, i.e., $r(\boldsymbol{w}) - \beta w_i \geq 0$ and $S_i(\boldsymbol{w}) \geq 0$ everywhere, and that this constraint set contains at least one interior point. In addition, under the ASP policy, the coefficient of w_i in the unit profit $\theta\bar{\boldsymbol{w}} - \beta w_i$, i.e., $\theta/n - \beta$, must be negative; otherwise the payoff function would again be monotonically increasing in w_i, leading to trivial and unrealistic decisions. We relegate the technical details of these requirements to Appendix C.1.

4.3.4 Manufacturers

Given the wholesale prices set by the GPOs, $\boldsymbol{w}$, manufacturers decide the supply $S_i(\boldsymbol{w})$ allocated to each GPO i. This decision involves drug supply allocations, which can be quite complicated and not purely driven by profit maximization in reality. For example, [145]–[147], and [148] discussed a variety of principles guiding medical supply allocations in various practices. Given the complicated nature of the allocation, purely grounding the allocation on profit maximization is inappropriate, while an exact characterization of all allocation specifics through a tractable model also seems unachievable. Thus, we adopt an approximation and develop a simple model that captures the most essential dynamics. We note that manufacturers' supply to a GPO i, $S_i(\boldsymbol{w})$, should possess two basic properties: the supply should increase in GPO i's own wholesale price and decrease in other GPOs' wholesale prices, reflecting the fact that manufacturers are inclined to provide more supply to GPOs that offer higher wholesale prices. If we let $\bar{\boldsymbol{w}}_{\text{-}i} = \sum_{j \neq i} w_j/(n-1)$ represent the average wholesale price of the GPOs excluding GPO i, then a simple model that satisfies these basic properties is:

$$S_i(\boldsymbol{w}) = \underbrace{(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)}_{\text{allocation}} \times \underbrace{\epsilon_i}_{\text{random factor}}, \tag{4.3}$$

where $a > 0$, $0 < b < 1$, and ϵ_i is a nonnegative random factor with a mean of μ_i.

In this supply expression, the random factor ϵ_i captures the randomness beyond the dynamics explained by the supply allocation, and reflects the fact that, unlike drug demand, drug supply can be variable [17], [117]. The factor has a multiplicative relationship with the allocation component to ensure a proper form of supply: The allocation component could be small or large depending on the specific wholesale prices. If we used an additive random factor in the expression, then the scale of the additive factor might not match that of the allocation component and might even cause negative supply; The scale of the supply might not match that of the demand either. The multiplicative form resolves all these issues. Across different GPOs, the random factors ϵ_i's are assumed to be identically distributed (as previously discussed, we consider symmetric GPOs) and independent (i.e., the dependence between the random supplies $S_i(\boldsymbol{w})$'s is reflected in the allocation component, so the remaining randomnesses are independent). We assume that the random factors are continuous and have a survival function of $\bar{F}$, and let ϵ denote a generic random factor following the same distribution.

While real drug supply allocation practices are clearly more complicated, this tractable parsimonious approximation helps us capture the most essential elements and tradeoffs in the problem and uncover key effects and influential factors through the comparison between the AWP and ASP policies. Another benefit of the supply model (4.3) is that the parameters a and b have realistic meanings in drug shortages. The parameter a represents the attractiveness of producing other drugs. In drug production, manufacturers often make multiple drugs on the same lines via different batches [149]; in an example reported by [6], 30 to 50 different drugs were manufactured on one line. If other drugs have more attractive profits, then a manufacturer is more likely to allocate more capacity to other drugs and less capacity to the shortage drug. In fact, capacity issues are the direct cause for most drug shortages [17], [122]. Thus, in this model, the higher a is, the more attractive it is to produce other drugs, and the less capacity manufacturers would allocate to produce the focal drug. In other words, a GPO must set a sufficiently high wholesale price to ensure a desirable supply. The parameter b indicates manufacturers' price sensitivity, where a higher b indicates manufacturers' higher sensitivity to other GPOs' price offers. Further, the constraint $b < 1$ captures the fact that a GPO's own wholesale price has a higher impact than other GPOs' wholesale prices on the GPO's supply. Both a and b reflect the difficulty to secure the focal drug's supply, although from different perspectives, with a higher a or a higher b indicating a greater difficulty to secure the supply. We thus call these parameters *supply difficulty* parameters.

4.4 Theoretical Analysis

Given the formulated model, in this section, we derive the equilibrium to gain insights about the reimbursement policies' impact on drug shortages. We first show the existence and uniqueness of an equilibrium for GPOs' wholesale price decisions.

Theorem 4.4.1 (Existence and uniqueness of equilibrium)**.** *Under either the AWP or ASP policy, there exists a unique symmetric interior equilibrium wholesale price. Specifically, under the AWP policy, each GPO adopts an equilibrium wholesale price w that satisfies:*

$$[(b-2)\beta w + a\beta + p]\mathbb{E}\{[(1-b)w - a]\epsilon \wedge d\} - d(p - \beta w)\bar{F}(d/[(1-b)w - a]) = 0.$$

Under the ASP policy, each GPO adopts an equilibrium wholesale price w that satisfies:

$$\{(\theta/n - \beta)[(1-b)w - a] + (\theta - \beta)w\}\mathbb{E}\{[(1-b)w - a]\epsilon \wedge d\} - d(\theta - \beta)w\bar{F}(d/[(1-b)w - a]) = 0.$$

To further study the equilibrium shortages, we now define a shortage measure. Intuitively, a shortage occurs when supply is insufficient to meet demand. However, drug shortages in reality are of great concerns because they are systematic and chronic (see, e.g., [6], [8], [117]). The average drug shortage duration was over 9 months in 2011 and increased to nearly 14 months by 2015 [118], [119]. It is clear that merely supply randomness and minor insufficiency of the mean supply would not have caused the systematic and chronic (long-term) shortages. Those may cause backorders, for which manufacturers typically could inform buyers of estimated delivery times and buyers could often accommodate the waiting. However, the persistent shortages in reality are far more severe than backorders. Only systematic insufficiency in the mean supply (capacity) would cause the systematic and chronic shortages, see the discussions in [122], [150], [151]. Indeed, the causes of the mean supply/capacity insufficiency are what the FDA's shortage task force and much research try to identify and resolve. Correspondingly, our study investigates the long-term shortages across multiple years of a contract duration, and hence adopts a shortage measure based on the mean supply insufficiency: $\mathbb{I}\{(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)\mu_i \leq \gamma d\}$, where $0 < \gamma < 1$, and $\mathbb{I}$ is an indicator function that takes the value of 1 when the condition is satisfied (indicating a shortage occurrence) and 0 otherwise (indicating no shortage occurrence). This measure reflects that a (long-term) shortage occurs when the mean supply is considerably lower than the demand (i.e., lower than γ percent of the demand). In addition, this measure indicates shortage status (i.e., occurrence or no occurrence), not shortage magnitude, because our research question is on whether or not the policy is a major cause of the long-term drug shortages, instead of how great the shortages are for each drug. Finally, since GPOs are symmetric, the measure of one GPO's shortage suffices.

Based on the equilibrium in Theorem 4.4.1 and the defined shortage measure, we next analyze the policy impact on drug shortages. We note that different drugs have different parameter values, and hence their equilibrium shortage statuses are also different. In addition, the equilibrium shortage statuses are also clearly different under different policy parameter values. Therefore, in our analysis, we examine how the equilibrium shortage status changes with respect to different parameter values. To this end, we group the parameters into two categories: drug-specific parameters and industry parameters. The former category contains a, b and n, with a and b describing the difficulty in securing the drug's supply, and n representing the number of major GPOs in the drug's market. A small n implies a more concentrated demand-side market, and correspondingly, a larger weight of each GPO in the calculation of the *average* wholesale price under the ASP policy. Clearly, these parameters (a, b and n) take different values for different drugs, and by examining these parameters we can gain insights about characteristics that make a drug prone to shortages. The industry parameters contain θ and β, where θ is Medicare's reimbursement percentage (currently $\theta = 106\%$) and β represents GPOs' price consciousness. These parameters do not vary by drug, and the study of these parameters can yield important industry and policy implications.

We first examine the drug-specific parameters. Throughout the paper, we use the terms "increasing" and "decreasing" in a non-strict sense. We say shortage "increases" or "decreases" to indicate that shortage is more likely or less likely to occur under the specified situation.

Theorem 4.4.2 (Impact of drug-specific parameters)**.**

(i) At equilibrium, the wholesale price increases in a and b under either AWP or ASP. However, while shortage increases in a and b under AWP, it decreases in a and b under ASP.

(ii) At equilibrium, the wholesale price and shortage do not depend on n under AWP, while the wholesale price is lower and shortage is higher for a drug with a larger n under ASP.

As mentioned, the parameters a and b represent the difficulty in obtaining supplies. The first half of Theorem 4.4.2(i) states that the more difficult it is to obtain supplies, the higher the wholesale price buyers need to pay to secure their supplies. This result is intuitive and it holds for both AWP and ASP. The second half of Theorem 4.4.2(i), however, states that the shortage under AWP and that under ASP change in completely opposite directions with the change of a and b. While the supply difficulty aggravates shortage under AWP, it surprisingly alleviates shortage under ASP.

To understand these results, we first uncover two important effects under the ASP policy: the *free-ride* effect and the *coordination* effect. Given that the reimbursement price is marked up from the *average* wholesale price, each GPO has an incentive to set a low wholesale price, while hoping that other GPOs set high wholesale prices to sustain the reimbursement price. This is the free-ride effect. In other words, each GPO hopes to take a "free ride" on other GPOs' high wholesale prices. On the other hand, if all GPOs increase their wholesale prices, then all can benefit from the resultant higher average wholesale and reimbursement prices. This is the coordination effect. However, this coordination effect is usually not the natural outcome of an equilibrium, because at an equilibrium, any single GPO may have the incentive to deviate by lowering its wholesale price. Thus, the coordination effect requires some outside stimulus. The aforementioned supply difficulty (parameters a and b) serves as a stimulus for the coordination effect.

Specifically, supply difficulty affects the shortages in both direct and indirect ways. Directly, a higher supply difficulty makes it harder to secure supplies (e.g., manufacturers allocate more capacity to other more attractive drugs), thus *aggravating* shortages. Indirectly, however, a higher supply difficulty also drives up wholesale prices (as stated in the first half of Theorem 4.4.2i), which *reduces* shortages. From the result in Theorem 4.4.2(i), we know that under ASP, the indirect effect dominates the direct effect, i.e., the supply difficulty drives up the wholesale prices so much that shortages are reduced. This is due to the coordination effect: The supply difficulty incentivizes GPOs to increase wholesale prices to secure supplies, which raises the average wholesale price, and consequently, the reimbursement price. The higher reimbursement price then implies a higher profit margin, which motivates GPOs to further increase wholesale prices to secure more supplies. This reinforcing feedback loop that accelerates price increases mitigates drug shortages.

In addition to the supply difficulty, another influential factor is the number of GPOs in a drug's market, n. Different drugs typically have different n values (i.e., the number of GPOs dominating a drug's market varies from drug to drug). Theorem 4.4.2(ii) shows that while the AWP price and shortage are unaffected by the n value, the ASP wholesale price is lower and shortage is higher for drugs with larger n values. This is because for a drug with a larger n, each GPO has a smaller impact on the average wholesale price (and the reimbursement price). Hence, each GPO considers the reimbursement price more as exogenous and the free-ride effect would be stronger, leading to a lower wholesale price and a more likely shortage occurrence. On the other hand, for a drug with a small n, each GPO represents a large weight in the average wholesale price calculation and is thus less likely to take a free ride. In the extreme case when only one GPO exists, the average wholesale price is just this GPO's wholesale price, and there is no free ride at all.[2]

As previously discussed, the generic drugs that the shortages primarily concern have two prominent features: they usually have very few GPOs dominating their markets [i.e., a small n, as discussed in 5], [17, and our later numerical analysis] and they often exhibit greater supply difficulty (i.e., large a and b as these generic drugs are typically unattractive to manufacturers due to their low profit margins). According to Theorem 4.4.2, the small number of GPOs discourages the free-ride effect and the supply difficulty encourages the coordination effect, both of which help mitigate drug shortages under the ASP policy. In other words, the ASP policy possesses resilience to shortages of these generic drugs, which

[2]↑When examining the impact of n, a related question might be how a GPO's supply $S_i(\boldsymbol{w})$ and demand d scale with n: a larger n may lead to a smaller supply allocation to each GPO, as well as a smaller market share (demand) of each GPO. In this case, the supply and demand are both approximately scaled by a factor of $1/n$, so the coefficient $1/n$ can be factored out of a GPO's payoff (i.e., $(1/n\times\text{supply})\wedge(1/n\times\text{demand})=1/n\times(\text{supply}\wedge\text{demand})$). Correspondingly, a GPO's equilibrium payoff would become $1/n$ of its original payoff, but the equilibrium wholesale price decisions and shortage statuses would not be affected. It is also worth noting that our focus here is on how different drugs with different n values have different shortage situations (i.e., what drugs are more prone to shortages), instead of on how a change in the number of GPOs for a specific drug affects this drug's shortage situation. The latter involves market entry and exit of GPOs, a much longer-term and complicated process which is beyond the scope of this analysis. In addition, in the market of a mature generic drug that the current shortages primarily concern, n does not change much from year to year.

is a special property of the ASP policy that has not been discovered by any prior studies. Consequently, the ASP policy's impact on drug shortages may not be as high as some scholars originally thought.

Our analysis so far has yielded insights into the impacts of the policies on drug shortages. We next further analyze how some important industry/policy parameters affect drug shortages to obtain industry/policy implications. Such parameters include θ, Medicare's reimbursement percentage (currently $\theta = 106\%$), and β, GPOs' price consciousness.

Theorem 4.4.3 (Impact of industry/policy parameters)**.**

(i) Under either the AWP or ASP policy, the equilibrium wholesale price decreases and shortage increases as the GPOs' price consciousness parameter β increases.

(ii) Under the ASP policy, the equilibrium wholesale price increases and shortage decreases as the reimbursement percentage θ increases.

Theorem 4.4.3(i) shows that when GPOs are more price conscious, they set lower wholesale prices, which aggravate shortage. As discussed in Section 4.3.3, a GPO's self-serving motive exerts an upward force on the wholesale price, whereas its member-serving motive is not a pure downward pressure on the wholesale price. A larger β implies a smaller weight on the upward force exerted by the self-serving motive and hence reduces the wholesale price. In practice, it is commonly understood that the main function of GPOs is to secure lower wholesale prices and GPOs are often evaluated by the wholesale prices they bring to their members. This result shows that overemphasizing this function may increase shortage.

Theorem 4.4.3(ii) states that, as expected, a higher reimbursement markup percentage under ASP leads to higher wholesale prices and reduces shortage. Since the markup percentage is central in Medicare's reimbursement policy, it is critical to understand its influence on shortages. In the numerical analysis next, we will examine various levels of θ (as well as β) to gain additional industry/policy insights.

4.5 Numerical Analysis

In addition to our theoretical analysis which has revealed key effects and influential factors concerning the policy impact on drug shortages, in this section, we further conduct numerical analysis to quantify the impact and obtain additional insights. Such analysis is challenging because of the data non-transparency in the pharmaceutical industry; much of the data related to this analysis, especially GPO data, are unavailable. Although Medicare collects data on GPOs, such data are confidential as restricted by Section 1927(b)(3)(D) of the Social Security Act. In fact, the restriction has spurred advocacy for more transparency in the pharmaceutical industry [152]–[154]. To overcome these data challenges, we identify multiple public and private data sources, based on which we implement a multi-step process to compile a novel data set, with the aid of machine learning. This helps us to estimate the number of GPOs, n, in each shortage drug's market. As n is the underlying driver of the free-ride effect, this estimation sheds light on the degree of the effect in the market. In addition, the data also allow us to justify a major assumption in our model, i.e., the symmetry among GPOs (details in Section 4.6). We then evaluate the impact of the combinations of n and different levels of the supply difficulty parameters a and b (the underlying driver of the coordination effect) and characterize the sets of drugs prone to shortages under each policy. Finally, we test multiple levels of industry/policy parameters to derive policy implications. In the following, we first discuss our data sources and how we integrate the data before we detail the analysis and the results.

4.5.1 Data Sources and Data Integration

In this analysis, we focus on drugs that have experienced shortages since 2005, in which year drug shortages started to emerge as a serious issue (see, e.g., the discussion in [17]). We implement a multi-step process to merge four datasets. First, we identify the drugs on shortage since 2005 by using two authoritative drug shortage lists commonly used in prior studies, e.g., [18], [123]: the ASHP shortage list and the FDA shortage list, that is, we include a drug in our analysis if it shows up in either list. Here, drugs are defined by using the Healthcare Common Procedure Coding System (HCPCS) code provided by the Centers for

Medicare & Medicaid Services (CMS). We then identify and access the *Medicare Provider Utilization and Payment Data: Physician and Other Supplier Public Use File* (Medicare PUP), which contains reimbursement requests submitted by *physicians* for the usage of each shortage drug from 2012 to 2016. Further, we identify and acquire a dataset from the American Hospital Association (AHA), which surveyed *hospitals* for the GPOs they used from 2012 to 2016. This allows us to link the *physicians* in the Medicare PUP data to *hospitals* by using their addresses, and then link *hospitals* to *GPOs* through the hospital–GPO affiliation in the AHA data. This yields a dataset that contains the usage information of each shortage drug by GPOs.

In this data integration process, a challenge is that the AHA dataset is much smaller than the Medicare PUP dataset, and hence, only around 10% of the physicians can be assigned to a GPO. We overcome this challenge by using supervised machine learning to assign the remaining 90% of the physicians to GPOs, utilizing the GPO labels of the 10% already-assigned physicians. This learning is based on features such as provider type, place of service, entity code, average submitted charge amount, etc.. We test three common machine learning methods: random forest, support vector machine, and neural network, with their parameters tuned through cross validation. The support vector machine method delivers the best performance, and we thus adopt this method. The final outcome of this data integration process is the usage volume per drug per GPO for a total of 201 shortage drugs. Table 4.1 provides summary statistics for the data by years. Appendix C.5 contains additional details of this data integration process.

Table 4.1. Summary Statistics

	2016	2015	2014	2013	2012
Number of drugs	157	158	163	164	162
Average number of physicians per drug	968.2	938.4	898.3	881.5	909.5
Average number of services per drug	1758.6	1611.2	1402.9	1324.7	1329.7

4.5.2 Analysis and Results

Estimation of n for Free-Ride Effect

The GPO drug usage data that we assembled above allow us to estimate the number of GPOs in the market for each shortage drug. Our analysis discovers that on average, about 3 GPOs dominate 80% of the market for each drug, while the remaining 20% of the market is represented by various small GPOs that are used for ad-hoc or one-time purchases. This finding is in line with statistics in prior reports that the top 5 GPOs command 85-90% of the market of *all* shortage drugs [17], which further indicates that the number of GPOs dominating the market of *each* drug is likely to be even less than 5, since different GPOs may dominate the markets of different drugs. Our result is also consistent with the Pareto principle (i.e., the 80/20 rule). Since the 20% small GPOs play ad-hoc and minor roles in the policy effect, we focus on the major GPOs that account for 80% of the market. Figure 4.2 shows the number of major GPOs that dominate the markets of the shortage drugs.

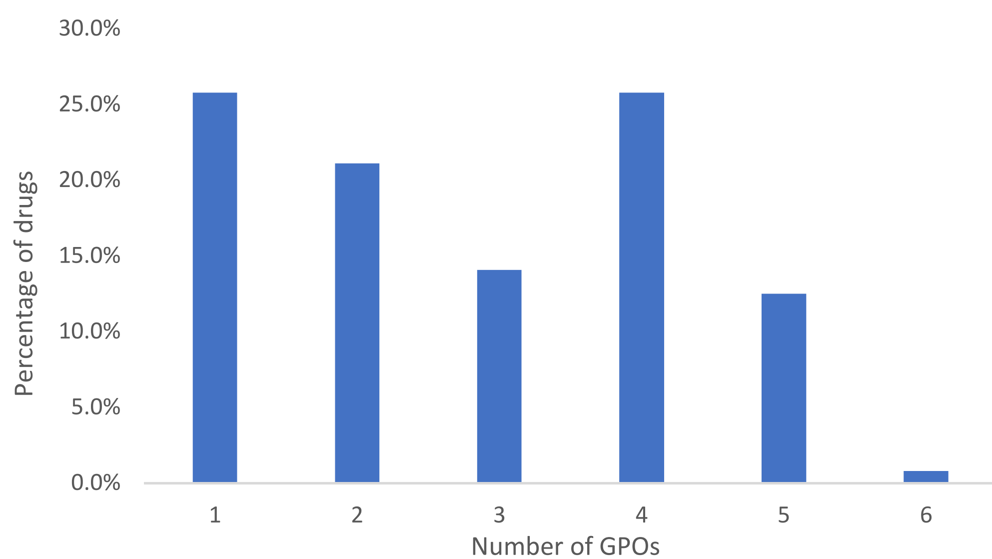


Figure 4.2. Number of GPOs

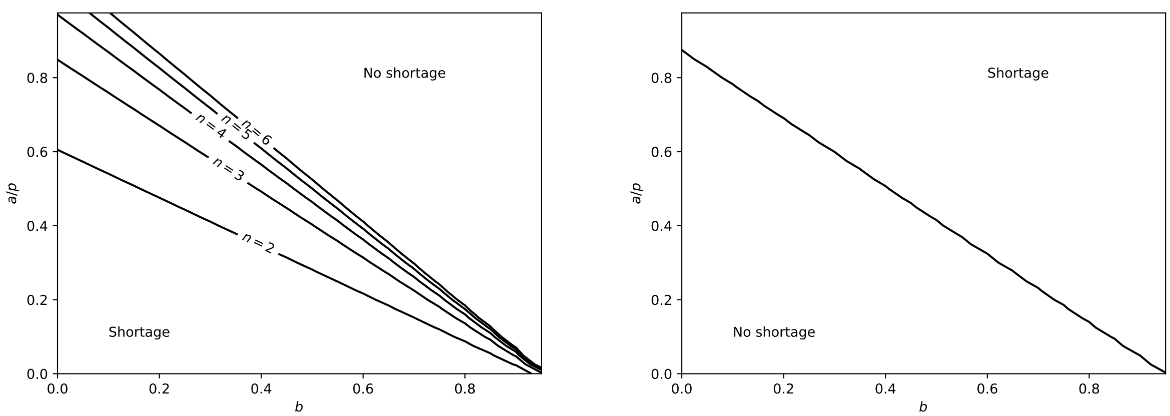
This figure well indicates a potentially low degree of free-ride effect in the markets of these shortage drugs. First, more than 25% of the shortage drugs only have one major GPO dominating their markets. As previously discussed, with only one major GPO, the free-ride effect would be very limited, ensuring a high wholesale price and, correspondingly, an unlikely shortage occurrence under ASP. Overall, on average, each shortage drug's market is dominated by only 3 major GPOs by volume, and the number of major GPOs rarely exceeds 5. Since the ASP policy possesses resilience to shortages of drugs with few GPOs, the result shows that the ASP policy is unlikely a major cause for the shortages of these drugs.

Examination of a and b for Coordination Effect

While we were able to find a way to estimate n from the data we assembled, the non-transparency of the pharmaceutical data makes the estimation of a and b, the supply difficulty parameters, prohibitive. Specifically, such an estimation would require data on each GPO's wholesale price, supply and demand, which, as previously discussed, are confidential as restricted by Section 1927(b)(3)(D) of the Social Security Act and challenging to estimate. Therefore, we test a large variety of the parameter values of a and b, and numerically evaluate whether there would be shortages under the combinations of these parameters based on our model of different reimbursement policies. This allows us to visualize the regions of the parameter values within which a drug would be prone to shortage under each policy. In doing this, we note that $0 < b < 1$, and we normalize a by the list price under AWP p (i.e., a/p), which also falls in about the same range according to our preliminary numerical analysis. As we know, higher values of a or b indicate a higher degree of supply difficulty.

Figure 4.3 shows the shortage regions corresponding to the ASP policy with different n values (Figure 4.3a) and the AWP policy (Figure 4.3b), which is unaffected by n. It is worth noting that these shortage regions *do not indicate* whether the corresponding drugs had shortages in reality. Instead, they show whether the corresponding drugs *would be on shortage* under a reimbursement policy based on our model (in this case, we say the drugs are *prone to shortages under the reimbursement policy*). If a drug already had shortages in reality, but our analysis found it not prone to shortages under a reimbursement policy, then the reimbursement policy would be unlikely the major cause of the drug's shortages in reality.

Figure 4.3a demonstrates that the shortage region under ASP (the area below the corresponding line) expands as n increases due to the higher free-ride effect under a larger n. When $n = 1$, the free-ride effect would be very limited (nominally zero), and hence the coordination effect dominates, leading to an empty shortage region (i.e., the line for $n = 1$ overlaps the horizontal axis). As n increases, the expansion of the shortage region is very fast initially and then slows down when n becomes larger, indicating a diminishing impact of n on shortages. The explanation is that as n increases, each GPO's weight in the average wholesale price calculation first drops sharply (e.g., from 1 to 1/2) and then decreases at a slower rate (e.g., from 1/4 to 1/5).



(a) ASP (b) AWP

Notes: plots generated with the setting: $\epsilon \sim N(\mu, \sigma)$ where $\mu = 20$ and $\sigma = 10$, $d = 40$, $p = 40$, $\theta = 106\%$, and $\beta = 1$.

Figure 4.3. Shortage Regions With Respect to Supply Difficulty Parameters

A comparison of Figure 4.3a and Figure 4.3b clearly shows the opposite impacts of supply difficulty (parameters a and b) on AWP and ASP, as stated in Theorem 4.4.2. While supply difficulty aggravates shortages under AWP as expected, it alleviates shortages under ASP due to the coordination effect and the resultant higher wholesale prices. Since drugs on shortage are primarily generic drugs with relatively low profit margins, manufacturers are naturally more attracted by producing other drugs (indicating a high value for a) and more sensitive to higher wholesale price offers from GPOs (indicating a high value for b). Drugs with such supply difficulty (i.e., high values for a and b), as shown in Figure 4.3a, tend not to have shortages under ASP. Hence, the current ASP policy is unlikely the major cause of the shortages of these generic drugs.

It is also worth noting that while in this figure, we set $\theta = 106\%$ and $\beta = 1$, we will next explore the impact of changes in the values of θ and β. By Theorem 4.4.3, we know that shortage decreases in θ and increases in β. Hence, the current θ and β values (i.e., $\theta = 106\%$ and $\beta = 1$) serve as the *worst case* for shortages.

Impact of Industry Parameters and Policy Implications

In addition to examining shortages under the current policy, we now explore whether changes in the industry/policy parameters (i.e., whether an increase in the reimbursement markup percentage, θ, and whether a decrease in GPOs' price consciousness, β) lead to *significant* shortage mitigation.

For the reimbursement percentage $\theta = 106\%$ under the current ASP policy, Figure 4.4 shows what if government is able to reimburse a higher percentage, i.e., $\theta = \{106\%, 107\%, 108\%, 109\%, 110\%, 111\%\}$. Note that θ is only a parameter in ASP and hence the figure only contains the ASP result. In this figure, we can see that an increase in θ reduces the shortage region, as stated in Theorem 4.4.3, and the shrinkage of the region slows down as θ becomes larger, indicating a likely diminishing return from increasing θ. Thus, while a policy adjustment that increases θ helps reduce shortages, government has to be cautious about whether the benefit from this increase could justify the presumably much higher Medicare spending for reimbursement due to this increase, especially at a much higher θ level.

We next test the impact of the GPOs' price consciousness β. We examine $\beta = \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, where $\beta = 0.5$ means that a GPO places $\alpha = 1/(2-\beta) = 2/3$ weight on its members' profit and $1/3$ on its own profit, and $\beta = 1$ indicates that a GPO places 100% weight on members' profit. Note that by definition, β cannot be too small: a GPO only caring about its own profit would attempt to adopt an arbitrarily high wholesale price to maximize its commission (technically, the payoff function in equation (4.2) would be monotonically increasing in the wholesale price), which is evidently unrealistic.

Figure 4.5 shows the result for β. In this figure, while a decrease in β from the full price consciousness case ($\beta = 1$) reduces the shortage regions under both AWP and ASP (as stated in Theorem 4.4.3), the two policies have very different sensitivities to β. The shortage region under ASP is very *sensitive* to the change in β, whereas the shortage region under AWP is

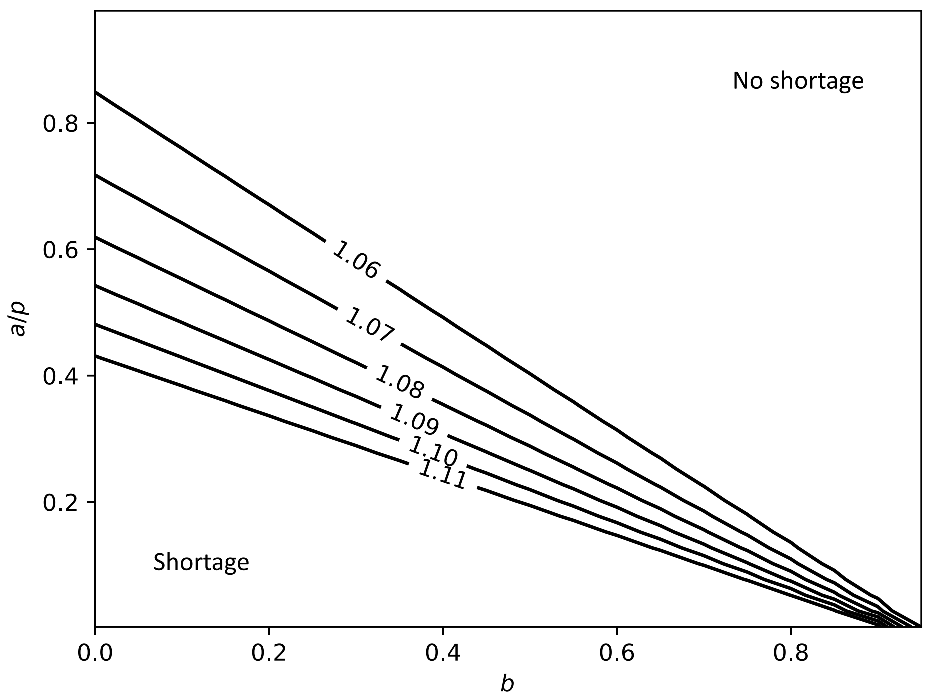
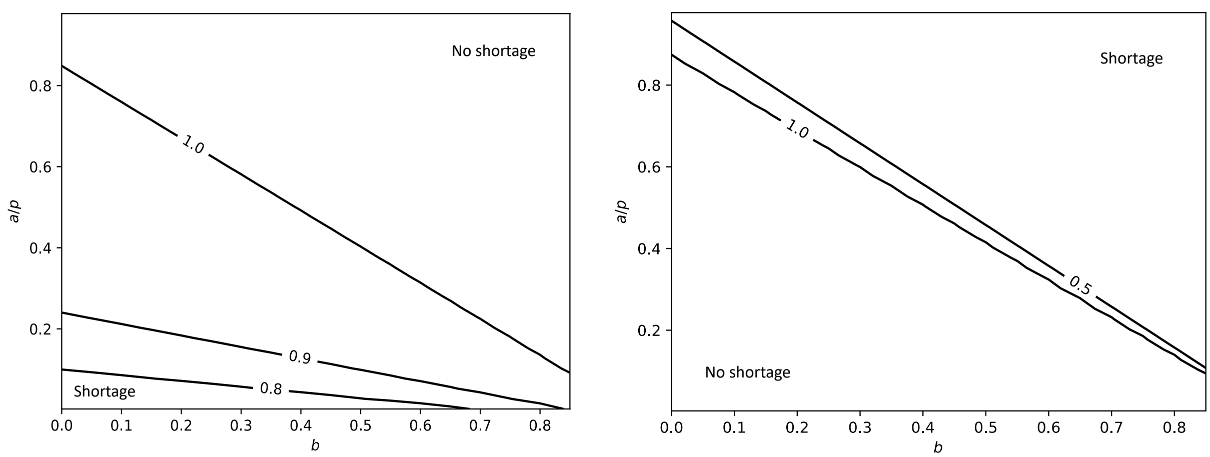


Figure 4.4. Impact of θ on Shortages under ASP ($n = 3$)

quite *insensitive* to the change. This indicates that β plays a critical role in determining the shortages under the ASP policy. Furthermore, the shortage region under ASP expands sharply as β increases. Thus, an overemphasis on GPOs' function to lower wholesale prices hinders shortage mitigation. On the other hand, ignoring the GPOs' function of securing low wholesale prices is also impractical, as it may incur significant welfare loss for GPO members. Thus, properly balancing these aspects is necessary. In practice, such balancing may be influenced by policy guidance. As discussed in [122], while GPOs and their healthcare provider members are currently mainly concerned with obtaining lower wholesale prices, incorporating more other metrics in their purchasing decisions, such as product quality as advocated by [8] and service quality (i.e., service levels suppliers can provide) as advocated in [122], would have a greater impact on shortage mitigation.



(a) ASP (b) AWP

Notes: In the (a) ASP plot, the line for $\beta = 0.7$ is very close to the horizontal axis and those for $\beta = 0.6, 0.5$ do not show up in the plot, indicating that no shortages would occur. In the (b) AWP plot, only lines for $\beta = 0.5$ and $\beta = 1$ are shown in order not to overcrowd the figure; lines for other β values are in between these two lines.

Figure 4.5. Impact of β on Shortages

4.6 Variations of the Model

While we believe our study captures the major elements and tradeoffs of the problem through a parsimonious model, we have made some simplifying assumptions to maintain tractability. In this section, we provide further justifications for some assumptions, while relaxing some other assumptions to extend our results.

Asymmetry Among GPOs

In our model, we have assumed that the major GPOs in a drug's market are symmetric. To check whether this is realistic, we estimate the degree of asymmetry among GPOs' volumes from our data by using two common asymmetry measures: the Gini index and the coefficient of variation (CV). The Gini index, widely used to measure inequality among the incomes of a nation's residents, ranges theoretically from 0 (complete equality) to 1 (complete inequality). For real countries, the range is between 0.24 and 0.63 [155]. The CV is the ratio of the standard deviation to the mean of GPOs' volume distribution and can assume any nonnegative value. Though lacking a consensus, a general rule of thumb for a low CV value is CV<1. In our data, the drug-average Gini index and CV of major GPOs (i.e., those taking 80% of the market) are 0.137 and 0.267, respectively, indicating a quite low degree of asymmetry, which validates our assumption.

It is worth noting that, with asymmetric GPOs, the ASP policy's resilience to shortages would likely be even stronger. The intuition is as follows. Imagine an asymmetric situation where one GPO has a dominant market share. This GPO then represents a dominant weight in the average wholesale price calculation and thus would be keen to raise the wholesale price. This higher wholesale price and the dominant GPO's large purchase volume would help shortage mitigation.

Price Consciousness Depending on Number of GPOs

In our model, GPOs' price consciousness is assumed to be a fixed value. In practice, GPOs' price consciousness may be related to the number of GPOs in the market: more GPOs in the market could intensify the competition between GPOs, which might result in GPOs' higher price consciousness. While we cannot verify this relationship due to the opacity of GPOs' data, we examine such a case with the price consciousness depending on n. Specifically, we replace the original price consciousness parameter β with a function $0 \leq g(n, \beta) \leq 1$, where $g(n, \beta)$ increases in the number of GPOs, n, and the price consciousness parameter, β, with

$0 \leq \beta \leq 1$. In this case, Proposition 4.6.1 next shows that our theoretical results still hold with some minor modifications.

Proposition 4.6.1. *When price consciousness increases in n, the existence and uniqueness of the equilibrium (wholesale price and shortage) captured in Theorem 4.4.1 and the monotonic impact of model parameters (a, b, n) captured in Theorem 4.4.2 still hold, except that the impact of n under AWP is updated as follows: the AWP equilibrium wholesale price is lower and shortage is higher for a drug with a larger n.*

The above proposition shows that our results on ASP still hold. This is because the dependence of the price consciousness on n would make the resilience of the ASP policy to shortages even more evident: if n is larger, then each GPO not only has an incentive to take a free ride by lowering the wholesale price, but also is more price conscious now, which would further lower the wholesale price. This would aggravate shortages for drugs with larger n values under ASP. On the other hand, since shortage drugs typically have small n values as previously discussed, the resilience of the ASP policy to these drugs' shortages would be more evident.

In addition, while the AWP equilibrium wholesale price and shortage do not depend on n in Theorem 4.4.2, n now affects the AWP equilibrium through affecting the price consciousness: when n becomes larger, GPOs are more price conscious, hence reducing their wholesale prices and aggravating shortages. The proof of Proposition 4.6.1 is in Appendix C.6.

4.7 Data-driven Analytics in the Drug Pricing Decisions

In the previous analysis, we have assumed that each GPO has perfect knowledge about a, b and the randomness ϵ in the supply function, as well as the demand facing each GPO. While the GPOs are able to directly observe demand from the past selling data, for example, through IMS or CMS data sources, it is unlikely they can observe a, b and ϵ for manufacturers' decisions. Therefore, in this section, we discuss how should the GPOs make pricing decisions given some historical observed transaction data with the drug manufacturers.

In the following discussion, we relax the assumption that GPOs are symmetric in their demand volume: Instead, we let d_i be the demand facing GPO i and we assume this is public information. We still assume that each GPO's demand d_i is constant across the years. Let $w_{i,t}$ be the wholesale price offered by GPO i in year $t = 1, \cdots, T$, and $w_{i,t}$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$ is public information. In addition to observing d_i and $w_{i,t}$, GPOs can also obtain $\mathbb{1}_{i,t}$ as an indicator of whether GPO i has a drug shortage in year t, where $\mathbb{1}_{i,t} = 1$ indicates a drug shortage and $\mathbb{1}_{i,t} = 0$ indicates no drug shortage, and GPOs can observe $\min\{S_{i,t}(\boldsymbol{w}_t), d_i\}$ which is the sales of GPO i in year t where $\boldsymbol{w}_t = (w_{i,t})_{i=1}^n$ are the wholesale prices and $S_{i,t}(\boldsymbol{w}_t)$ is the supply obtained by GPO i in year t. Besides that, GPO i can observe its own supply $S_{i,t}(\boldsymbol{w}_t)$. The unknown information to GPOs are therefore a, b and distribution of ϵ.

4.7.1 Application of Data-driven Robust Optimization Approach

In this subsection, we discuss how we can apply the robust optimization approach for the GPOs to make data-driven decisions. We will consider an estimate-then-optimize approach. That is, GPOs first use the past data to obtain an estimate for the unknown parameters, and then based on the unknown parameters, GPOs can optimize for data-driven prices by solving a robust optimization problem with uncertainty sets constructed for the unknown parameters.

In the first place, estimators $\hat{a}, \hat{b}$ and $\hat{F}$ for a, b, F respectively can be obtained based on past data. If the GPOs know the parametric form of F but are unknown about the parameters of it, then they can use maximization-expectation approach [156] for the estimation. If the GPOs are unknown of the parametric form of F, then they can resort to, for example, [157] for the nonparametric estimation approach with censored data.

Upon obtaining the estimates for $\hat{a}, \hat{b}$, and $\hat{F}$, GPOs can construct some uncertainty sets for their estimates, i.e., $a \in \mathcal{A}, b \in \mathcal{B}$, and the true distribution ϵ, $F \in \mathcal{D}(\hat{F})$ where $\mathcal{D}(\hat{F})$ contains distributions which the GPO believe to be candidates for the true distribution based on its estimate of $\hat{F}$. Then given $\boldsymbol{w}_{-i}$, the robust optimization model for GPO i's pricing decision is

$$\max_{w_i \geq 0} \min_{a \in \mathcal{A}, b \in \mathcal{B}, F \in \mathcal{D}(\hat{F})} [\boldsymbol{r}(\boldsymbol{w}) - \beta w_i] \mathbb{E}[S_i(\boldsymbol{w}) \wedge d_i].$$

We consider robust-optimization equilibrium as the solution to the game.

Definition 4.7.1 (Robust-optimization Equilibrium, [37]). *$\boldsymbol{w}^*$ is a robust-optimization equilibrium of the game if and only if for all $i \in \{1, 2, \cdots, n\}$,*

$$w_i^* \in \arg\max_{w_i \geq 0} \min_{a \in \mathcal{A}, b \in \mathcal{B}, F \in \mathcal{D}(\hat{F})} [\boldsymbol{r}(\boldsymbol{w}) - \beta w_i] \mathbb{E}[S_i(\boldsymbol{w}) \wedge d_i].$$

4.7.2 Application of Dynamic Learning Approach

The robust optimization approach sets the single period profit as the goal for the GPO to optimize. If the GPOs' goal is to optimize for a planning horizon of multiple periods, they can use the dynamic learning approach. The problem protocol for a multi-period game between GPOs is presented as follows.

Protocol 4.1. Repeated Game between GPOs

In each round $t = 1, 2, \cdots, T$:

1. Each GPO i simultaneously choose their optimal wholesale price $w_{i,t}$ to offer to the manufacturer.
2. Each GPO i obtains their supply $S_{i,t}(\boldsymbol{w}_t)$.
3. The wholesale prices $\boldsymbol{w}_t$, drug shortage information $(\mathbb{1}_{i,t})_{i=1}^n$ and all GPOs' sales information $\{\min\{S_{i,t}(\boldsymbol{w}_t), d_i\}\}_{i=1}^n$ becomes public information.

Let $\mathcal{H}_{\bar{t}} \triangleq \left\{ \boldsymbol{w}_t, (\mathbb{1}_{i,t})_{i=1}^n, \{\min\{S_{j,t}(\boldsymbol{w}_t), d_j\}\}_{j=1, j \neq i}^n, S_{i,t}(\boldsymbol{w}_t) \right\}_{t=1}^{\bar{t}}$ be the information that has been revealed to GPO i by the end of period $\bar{t}$. Then the GPO's pricing policy π^i sets a price $w_{i,t}^\pi$ according to the information that has been revealed to GPO i:

$$w_{i,t}^\pi \sim \pi^i(\mathcal{H}_t), \text{ for } i = 1, \cdots, n \text{ and } t \geq 2.$$

Each GPO's objective is to maximize the total profit it received through the planning horizon. Let $\boldsymbol{\pi}^{-i} = (\pi^j)_{j=1,j\neq i}^n$ be all GPOs' pricing policies except for GPO i's and $\boldsymbol{w}_t^{\boldsymbol{\pi}} = (w_{i,t}^{\pi})_{i=1}^n$. Then the total profit for GPO i across the planning horizon is

$$V_i(\pi^i, \boldsymbol{\pi}^{-i}) = \sum_{t=1}^{T} [\boldsymbol{r}(\boldsymbol{w}_t^{\boldsymbol{\pi}}) - \beta w_{i,t}^{\pi}]\mathbb{E}[S_{i,t}(\boldsymbol{w}_t^{\boldsymbol{\pi}}) \wedge d_i].$$

We consider the equilibrium concept as follows:

Definition 4.7.2 (Nash equilibrium of the dynamic game between GPOs)**.** *A joint policy $\boldsymbol{\pi}^* = (\pi^{1,*}, \pi^{2,*}, \cdots, \pi^{n,*})$ is a Nash equilibrium of the game if and only if for all $i \in \{1, 2, \cdots, n\}$,*

$$V_i(\pi^{i,*}, \boldsymbol{\pi}^{-i,*}) \geq V_i(\pi^i, \boldsymbol{\pi}^{-i,*}), \forall \textit{ admissable } \pi^i.$$

There is a rich literature on multi-agent dynamic learning. Interested readers can refer to the survey by [103].

4.8 Ending Remarks

Drug shortages remain a critical public health issue for the U.S. government and pharmaceutical industry. In this study, we investigate the impact of Medicare's drug reimbursement policy on drug shortages. To achieve this goal, we develop an analytical model of drug supply to capture the main tradeoffs and nuances in GPOs' pricing decisions, derive the equilibrium wholesale prices and shortage statuses, and analyze the impacts of influential parameters. Our analysis uncovers two opposing effects on the drug wholesale price decisions under the ASP policy: the free-ride effect (a GPO tends to lower its wholesale price while hoping other GPOs will maintain high wholesale prices to sustain a high reimbursement price) and the coordination effect (the reimbursement price is higher when all GPOs increase their wholesale prices). The interplay of these two effects determines the overall impact of ASP on wholesale prices and drug shortages. We capture key factors influencing these effects and show that the current ASP policy in fact possesses resilience to shortages of the generic drugs that the current shortages primarily concern.

In addition to the theoretical analysis, we further conduct numerical analysis for additional insights. We compile a novel dataset, integrating data from multiple sources, to estimate critical model parameters. We find that very few GPOs dominate markets of the drugs that experienced shortages, indicating a low degree of free-ride effect under which the ASP policy is unlikely to be the major cause for these shortages. In addition, we further visualize regions of the supply difficulty parameter values of drugs prone to shortages under each reimbursement policy. The visualization again confirms the ASP policy's resilience to shortages of the generic drugs which typically have a high degree of supply difficulty.

Furthermore, an examination of the impact of industry and policy parameters brings us additional insights. First, increasing the reimbursement percentage from the current 106% reduces shortages, but with likely diminishing returns. Given that such an increase would significantly increase government spending on reimbursement, a more detailed investigation is necessary before such a policy adjustment. Second, GPOs' price consciousness has a significant impact on shortages under ASP. Too much emphasis from GPOs and their healthcare provider members on lowering the wholesale prices leads to negative consequences. Interestingly, while [122] stated that GPOs and their members' focus on lowering wholesale prices may bring them a short-term benefit but hinder them from gaining a long-term benefit of a stable drug supply, our analysis indicates that such a focus may not even bring the short-term benefit: If all GPOs lower their wholesale prices, the reimbursement prices will be lower, which will in turn lower the income of the providers. In other words, too much focus on lowering wholesale prices hurts the providers not only in the long term due to drug shortages, but also in the short term due to lower reimbursement prices. GPOs and their members should shift some of the focus to non-price factors such as product quality, as advocated in [8], and suppliers' service quality (i.e., failure-to-supply clauses), as advocated in [122], to reduce shortages. The current FDA's initiatives in the development of a Quality Management Maturity (QMM) rating system, in collaboration with industry, academia, and other stakeholders, to facilitate such a focus shift in drug sourcing decisions [158] is a promising direction.

Finally, an interesting point arose during our discussion with FDA staff. As heard from many of the listening sessions organized by the FDA drug shortage task force in 2019, manufacturers blame GPOs for pushing the wholesale prices too low and GPOs blame providers for only focusing on low wholesale prices. While providers do care about patients as well as their budget, it seems that they have not considered these two aspects in a more integrated manner. Our research shows that these two aspects do not have to contradict each other — balancing between price and other non-price factors in drug procurement can lead to both short-term and long-term gains as discussed earlier. The main stakeholders of the drug supply chain (e.g., manufacturers, GPOs, and providers) need to adopt this new perspective to unite to tackle the drug shortage problem.

We conclude by noting that the non-transparency of pharmaceutical data (as restricted by Section 1927(b)(3)(D) of the Social Security Act) has made the numerical analysis challenging, as described. With the ongoing effort to improve transparency in the pharmaceutical industry [e.g., 152]–[154], an updated evaluation incorporating future available data would be helpful to refine our understanding of the reimbursement policies' impact on shortages.

5. CONCLUSION AND DIRECTION FOR FUTURE RESEARCH

This chapter concludes findings of this study and points out future research directions. In this study, we have examined data-integrated supply chain contract pricing problems under uncertainties. In comparison to data-driven or data-integrated decision making problems for a single agent, for example, the newsvendor problem, or a platform selling products to sequentially coming customers, the intricacy in the interactions between supply chain parties with the unknown environment underlying the data-integrated contract pricing problems have provided much more challenges as well as opportunities for us to explore. We investigate this problem under uncertainty in the market demand (Chapter 2), uncertainty in the downstream retailer's inventory policies (Chapter 3), and uncertainty in the competitors' pricing decisions as well as the uncertainty in the upstream manufacturer's supply decisions (Chapter 4). Specifically,

1. **Uncertainty in the Market Demand**: We study the data-driven contract design problem of a price-only contract for a supplier who has uncertainty in the market demand distribution. We propose a distributionally robust optimization model to seamlessly integrate the information obtained from the demand realizations and retailer's historical ordering information.

2. **Uncertainty in the Downstream Retailer's Inventory Learning Strategies**: We study the dynamic pricing strategies of a data-driven supplier in designing the price-only contract. Both the supplier and retailer are uncertain about the market demand. Besides that, the supplier also has uncertainty in the retailer's inventory learning policies. We propose dynamic pricing policies for the supplier and show that it leads to sublinear regret bound for the supplier under a wide range of retailer's inventory learning policies.

3. **Uncertainty in the Competitor's Pricing Decisions**: We study a practical problem of GPOs (representatives of healthcare providers) purchasing drugs from the manufacturers. GPOs have to decide contract prices appropriately considering their own and their healthcare providers' interest. In deciding their pricing decisions, GPOs

have to compete with each other and price under the uncertainty about each other's decisions. We adopt Nash equilibrium as the solution concept, and investigate the impact of reimbursement policy on drug shortages under Nash equilibrium of the pricing decisions.

We also provide several future research directions along this stream:

- **Data-driven Contract Design with Strategic Retailer**: In this study, we have mostly consider the situation that the retailer maximizes his single period profit, and the retailer does not manipulate the supplier's decisions. However, a strategic and intelligent retailer may send wrong signals to the supplier and induce the supplier to set wholesale prices favorable for the retailer. A future direction is to investigate the contract pricing problem along this line.

- **Data-driven Contract under Uncertainty in Risk Preferences**: In this study, we have considered that the decision makers are risk neural. However, there is an increasing evidence that decision makers are not risk neural, and furthermore, their risk attitudes may also be changing across time [159]. Therefore, how to elicit and hedge against the uncertainty under the unknown risk preferences of other supply chain parties is worth future investigation.

- **Data-driven Contract with Time Series Data**: In this problem, we have mostly assume that demand realizations are identically and independently distributed. However, in practice, demand observations are not identically nor independently distributed. With inter-correlated time series data, many of the classical results in the data-driven problems no longer hold. Therefore, it is challenging to study how time series data can influence the decision making framework for the contract problems.

- **Data-driven Contract with Dynamic Competition**: In source supplies or selling products, supply chain parties usually face competition from each other. In Chapter 4, we mainly consider this competition for a single period. However, there may be competitions over some periods of time. For a long time horizon, the supply chain parties are more aware of their long term profit and have to be more strategic in their decision making.

There are much more real life constraints and considerations we will face when we apply such methodologies in the real life problems. Such considerations make the direction more fruitful and challenging, which are worth our future exploration.

REFERENCES

[1] K. Ravindran, A. Susarla, D. Mani, and V. Gurbaxani, "Social capital and contract duration in buyer-supplier networks for information technology outsourcing," *Information Systems Research*, vol. 26, no. 2, pp. 379–397, 2015.

[2] X. Cai, J. Chen, Y. Xiao, X. Xu, and G. Yu, "Fresh-product supply chain management with logistics outsourcing," *Omega*, vol. 41, no. 4, pp. 752–765, 2013.

[3] H. Vega *et al.*, "The transportation costs of fresh flowers: A comparison between ecuador and major exporting countries," *Inter-American Development Bank, Washington DC*, 2008.

[4] EY Americas, *How the future of work will change the digital supply chain*, https://www.ey.com/en_us/consulting/how-the-future-of-work-will-change-the-digital-supply-chain, Accessed: 2022-10-04, 2020.

[5] FDA, "A review of FDA's approach to medical product shortages," U.S. Food and Drug Administration, Tech. Rep., 2011, p. 44.

[6] GAO, "Drug shortages: Public health threat continues, despite efforts to help ensure product availability," U.S. Government Accountability Office, Tech. Rep., 2014.

[7] AHA, "AHA survey on drug shortages," American Hospital Association, Tech. Rep., 2011.

[8] FDA, "Drug shortages: Root causes and potential solutions," U.S. Department of Health and Human Services, Tech. Rep., 2020.

[9] ISMP, "Drug shortages continue to compromise patient care," 2018.

[10] C. Cherici, J. Frazier, M. Feldman, B. Gordon, C. Petrykiw, and W. Russell, "Navigating drug shortages in American healthcare: A premier healthcare alliance analysis," *Premier Inc., March*, 2011.

[11] E. R. Fox, B. V. Sweet, and V. Jensen, "Drug shortages: A complex health care crisis," in *Mayo Clinic Proceedings*, Elsevier, vol. 89, 2014, pp. 361–373.

[12] R. Kaakeh, B. V. Sweet, C. Reilly, *et al.*, "Impact of drug shortages on us health systems," *American Journal of Health-System Pharmacy*, vol. 68, no. 19, pp. 1811–1819, 2011.

[13] AHA, *Survey: Drug shortages cost hospitals $360m annually*, 2019.

[14] FDA, *FDA drug shortage task force listening session*, Sep. 2018.

[15] S. Gottlieb, *FDA is advancing new efforts to address drug shortages*, Nov. 2018.

[16] J. Jia and H. Zhao, "Reflection on "mitigating the U.S. drug shortages through Pareto-improving contracts"," *Production and Operations Management*, 2019.

[17] ASPE, "Economic analysis of causes of drug shortages," U.S. Department of Health and Human Services, Tech. Rep., 2011.

[18] A. Yurukoglu, E. Liebman, and D. B. Ridley, "The role of government reimbursement in drug shortages," *American Economic Journal: Economic Policy*, vol. 9, no. 2, pp. 348–82, 2017.

[19] E. L. Tucker, M. S. Daskin, B. V. Sweet, and W. J. Hopp, "Incentivizing resilient supply chain design to prevent drug shortages: Policy analysis using two- and multi-stage stochastic programs," *IISE Transactions*, vol. 52, no. 4, pp. 394–412, Apr. 2020, ISSN: 2472-5854.

[20] M. Jacobson, A. Alpert, and F. Duarte, "Prescription drug shortages: Reconsidering the role of Medicare payment policies," Health Affairs, Tech. Rep., 2012.

[21] P. Mullen, "The arrival of average sale price," *Biotechnology healthcare*, vol. 4, no. 3, p. 48, 2007.

[22] E. J. Emanuel, "Shortchanging cancer patients," en-US, *The New York Times*, 2011.

[23] FDA, "Correspondence with the authors," Food and Drug Administration, Tech. Rep., 2019.

[24] X. Zhao, W. B. Haskell, and G. Yu, "Supply chain contracts in the small data regime," *Working paper*, 2022.

[25] X. Zhao, R. Zhu, and W. B. Haskell, "Learning to price supply chain contracts against a learning retailer," *arXiv preprint arXiv:2211.04586*, 2022.

[26] X. Zhao, J. Jia, and H. Zhao, "Reimbursement policy and drug shortages," *Working paper*, 2022.

[27] M. A. Lariviere and E. L. Porteus, "Selling to the newsvendor: An analysis of price-only contracts," *Manufacturing & service operations management*, vol. 3, no. 4, pp. 293–305, 2001.

[28] Y. Yu and X. Kong, "Robust contract designs: Linear contracts and moral hazard," *Operations Research*, vol. 68, no. 5, pp. 1457–1473, 2020.

[29] G. Perakis and G. Roels, "The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts," *Management Science*, vol. 53, no. 8, pp. 1249–1268, 2007.

[30] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1, pp. 115–166, 2018.

[31] R. Gao and A. J. Kleywegt, "Distributionally robust stochastic optimization with wasserstein distance," *arXiv preprint arXiv:1604.02199*, 2016.

[32] C. J. Corbett and C. S. Tang, "Designing supply contracts: Contract type and information asymmetry," in *Quantitative models for supply chain management*, Springer, 1999, pp. 269–297.

[33] S.-H. Kim and S. Netessine, "Collaborative cost reduction and component procurement under information asymmetry," *Management Science*, vol. 59, no. 1, pp. 189–206, 2013.

[34] H. Zhang, M. Nagarajan, and G. Sošić, "Dynamic supplier contracts under asymmetric inventory information," *Operations Research*, vol. 58, no. 5, pp. 1380–1397, 2010.

[35] B. Kalkanci, K.-Y. Chen, and F. Erhun, "Contract complexity and performance under asymmetric demand information: An experimental evaluation," *Management science*, vol. 57, no. 4, pp. 689–704, 2011.

[36] G. Carroll and D. Meng, "Robust contracting with additive noise," *Journal of Economic Theory*, vol. 166, pp. 586–604, 2016.

[37] M. Aghassi and D. Bertsimas, "Robust game theory," *Mathematical programming*, vol. 107, no. 1, pp. 231–273, 2006.

[38] X. V. Doan and T.-D. Nguyen, "Robust newsvendor games with ambiguity in demand distributions," *Operations Research*, vol. 68, no. 4, pp. 1047–1062, 2020.

[39] H. Jiang, S. Netessine, and S. Savin, "Robust newsvendor competition under asymmetric information," *Operations research*, vol. 59, no. 1, pp. 254–261, 2011.

[40] M. R. Wagner, "Robust purchasing and information asymmetry in supply chains with a price-only contract," *IIE Transactions*, vol. 47, no. 8, pp. 819–840, 2015.

[41] Q. Fu, C.-K. Sim, and C.-P. Teo, "Profit sharing agreements in decentralized supply chains: A distributionally robust approach," *Operations Research*, vol. 66, no. 2, pp. 500–513, 2018.

[42] L. H. Liyanage and J. G. Shanthikumar, "A practical inventory control policy using operational statistics," *Operations Research Letters*, vol. 33, no. 4, pp. 341–348, 2005.

[43] L. Y. Chu, J. G. Shanthikumar, and Z.-J. M. Shen, "Solving operational statistics via a bayesian analysis," *Operations Research Letters*, vol. 36, no. 1, pp. 110–116, 2008.

[44] O. Besbes and O. Mouchtaki, "How big should your data really be? data-driven newsvendor and the transient of learning," *Data-Driven Newsvendor and the Transient of Learning (March 15, 2021)*, 2021.

[45] V. Gupta and N. Kallus, "Data pooling in stochastic optimization," *Management Science*, vol. 68, no. 3, pp. 1595–1615, 2022.

[46] V. Gupta and P. Rusmevichientong, "Small-data, large-scale linear optimization with uncertain objectives," *Management Science*, vol. 67, no. 1, pp. 220–241, 2021.

[47] V. Gupta, M. Huang, and P. Rusmevichientong, "Debiasing in-sample policy performance for small-data, large-scale optimization," *arXiv preprint arXiv:2107.12438*, 2021.

[48] H. Lam and C. Mottet, "Tail analysis without parametric models: A worst-case perspective," *Operations Research*, vol. 65, no. 6, pp. 1696–1711, 2017.

[49] S. Wang, S. H. Ng, and W. B. Haskell, "A multilevel simulation optimization approach for quantile functions," *INFORMS Journal on Computing*, 2021.

[50] R. Levi, R. O. Roundy, and D. B. Shmoys, "Provably near-optimal sampling-based policies for stochastic inventory control models," *Mathematics of Operations Research*, vol. 32, no. 4, pp. 821–839, 2007.

[51] R. Levi, G. Perakis, and J. Uichanco, “The data-driven newsvendor problem: New bounds and insights,” *Operations Research*, vol. 63, no. 6, pp. 1294–1306, 2015.

[52] V. Ramamurthy, J. George Shanthikumar, and Z.-J. M. Shen, “Inventory policy with parametric demand: Operational statistics, linear correction, and regression,” *Production and Operations Management*, vol. 21, no. 2, pp. 291–308, 2012.

[53] J. Zhang, W. Xie, and S. C. Sarin, “Multiproduct newsvendor problem with customer-driven demand substitution: A stochastic integer program perspective,” *INFORMS Journal on Computing*, 2020.

[54] L. V. Kantorovich and S. Rubinshtein, “On a space of totally additive functions,” *Vestnik of the St. Petersburg University: Mathematics*, vol. 13, no. 7, pp. 52–59, 1958.

[55] M. J. Kim and A. E. Lim, “Robust multiarmed bandit problems,” *Management Science*, vol. 62, no. 1, pp. 264–285, 2016.

[56] J.-y. Gotoh, M. J. Kim, and A. E. Lim, “Calibration of distributionally robust empirical optimization models,” *Operations Research*, 2021.

[57] X. Chen, S. He, B. Jiang, C. T. Ryan, and T. Zhang, “The discrete moment problem with nonconvex shape constraints,” *Operations Research*, vol. 69, no. 1, pp. 279–296, 2021.

[58] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova, “Bilevel programming problems,” *Energy Systems. Springer, Berlin*, 2015.

[59] M. Besançon, M. F. Anjos, and L. Brotcorne, “Near-optimal robust bilevel optimization,” *arXiv preprint arXiv:1908.04040*, 2019.

[60] M. Jain, F. Ordónez, J. Pita, *et al.*, “Robust solutions in stackelberg games: Addressing boundedly rational human preference models,” in *Proc. of the AAAI 4th Multidiciplinary Workshop on Advances in Preference Handling*, 2008.

[61] O. Besbes, Y. Gur, and A. Zeevi, “Optimal exploration-exploitation in multi-armed-bandit problems with non-stationary rewards,” Columbia Business School Working paper, Tech. Rep., 2014.

[62] N. B. Keskin, X. Min, and J.-S. J. Song, “The nonstationary newsvendor: Data-driven nonparametric learning,” *Available at SSRN 3866171*, 2021.

[63] W. C. Cheung, D. Simchi-Levi, and R. Zhu, “Hedging the drift: Learning to optimize under nonstationarity,” *Management Science*, 2021.

[64] A. L. Gibbs and F. E. Su, “On choosing and bounding probability metrics,” *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.

[65] Z. S. Karnin and O. Anava, “Multi-armed bandits: Competing with optimal sequences,” *Advances in Neural Information Processing Systems*, vol. 29, pp. 199–207, 2016.

[66] C.-Y. Wei and H. Luo, “Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach,” in *Conference on Learning Theory*, PMLR, 2021, pp. 4300–4354.

[67] G. P. Cachon, "Supply chain coordination with contracts," *Handbooks in operations research and management science*, vol. 11, pp. 227–339, 2003.

[68] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári, "X-armed bandits.," *Journal of Machine Learning Research*, vol. 12, no. 5, 2011.

[69] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.

[70] C.-J. Ho, A. Slivkins, and J. W. Vaughan, "Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems," *Journal of Artificial Intelligence Research*, vol. 55, pp. 317–359, 2016.

[71] J. Broder and P. Rusmevichientong, "Dynamic pricing under a general parametric choice model," *Operations Research*, vol. 60, no. 4, pp. 965–980, 2012.

[72] K. J. Ferreira, D. Simchi-Levi, and H. Wang, "Online network revenue management using thompson sampling," *Operations research*, vol. 66, no. 6, pp. 1586–1602, 2018.

[73] N. B. Keskin and J. R. Birge, "Dynamic selling mechanisms for product differentiation and learning," *Operations research*, vol. 67, no. 4, pp. 1069–1089, 2019.

[74] A. V. den Boer and N. B. Keskin, "Discontinuous demand functions: Estimation and pricing," *Management Science*, vol. 66, no. 10, pp. 4516–4534, 2020.

[75] A. V. den Boer and N. B. Keskin, "Dynamic pricing with demand learning and reference effects," *Management Science*, 2022.

[76] G.-Y. Ban and N. B. Keskin, "Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity," *Management Science*, vol. 67, no. 9, pp. 5549–5568, 2021.

[77] N. B. Keskin, Y. Li, and J.-S. Song, "Data-driven dynamic pricing and ordering with perishable inventory in a changing environment," *Management Science*, vol. 68, no. 3, pp. 1938–1958, 2022.

[78] W. C. Cheung, D. Simchi-Levi, and H. Wang, "Dynamic pricing and demand learning with limited price experimentation," *Operations Research*, vol. 65, no. 6, pp. 1722–1731, 2017.

[79] H. Jia, C. Shi, and S. Shen, "Online learning and pricing for network revenue management with reusable resources," *Available at SSRN 4225832*, 2022.

[80] M. Chen and Z.-L. Chen, "Recent developments in dynamic pricing research: Multiple products, competition, and limited demand information," *Production and Operations Management*, vol. 24, no. 5, pp. 704–731, 2015.

[81] B. B. Chen, Y. Wang, and Y. Zhou, "Optimal policies for dynamic pricing and inventory control with nonparametric censored demands," *Management Science*, 2022.

[82] B. B. Chen, D. Simchi-Levi, Y. Wang, and Y. Zhou, "Dynamic pricing and inventory control with fixed ordering cost and incomplete demand information," *Management Science*, vol. 68, no. 8, pp. 5684–5703, 2022.

[83] O. Besbes and A. Zeevi, "On the minimax complexity of pricing in a changing environment," *Operations research*, vol. 59, no. 1, pp. 66–79, 2011.

[84] N. B. Keskin and A. Zeevi, “Chasing demand: Learning and earning in a changing environment,” *Mathematics of Operations Research*, vol. 42, no. 2, pp. 277–307, 2017.

[85] O. Besbes, Y. Gur, and A. Zeevi, “Stochastic multi-armed-bandit problem with non-stationary rewards,” *Advances in neural information processing systems*, vol. 27, 2014.

[86] O. Besbes, Y. Gur, and A. Zeevi, “Non-stationary stochastic optimization,” *Operations research*, vol. 63, no. 5, pp. 1227–1244, 2015.

[87] C.-Y. Wei, Y.-T. Hong, and C.-J. Lu, “Tracking the best expert in non-stationary stochastic environments,” *Advances in neural information processing systems*, vol. 29, 2016.

[88] L. Wei and V. Srivatsva, “On abruptly-changing and slowly-varying multiarmed bandit problems,” in *2018 Annual American Control Conference (ACC)*, IEEE, 2018, pp. 6291–6296.

[89] H. Luo, C.-Y. Wei, A. Agarwal, and J. Langford, “Efficient contextual bandits in non-stationary worlds,” in *Conference On Learning Theory*, PMLR, 2018, pp. 1739–1776.

[90] W. C. Cheung, D. Simchi-Levi, and R. Zhu, “Learning to optimize under non-stationarity,” in *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 1079–1087.

[91] X. Chen, Y. Wang, and Y.-X. Wang, “Nonstationary stochastic optimization under l p, q-variation measures,” *Operations Research*, vol. 67, no. 6, pp. 1752–1765, 2019.

[92] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, “The nonstochastic multi-armed bandit problem,” *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.

[93] A. Garivier and E. Moulines, “On upper-confidence bound policies for switching bandit problems,” in *International Conference on Algorithmic Learning Theory*, Springer, 2011, pp. 174–188.

[94] F. Liu, J. Lee, and N. Shroff, “A change-detection based framework for piecewise-stationary multi-armed bandit problem,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[95] Y. Cao, Z. Wen, B. Kveton, and Y. Xie, “Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit,” in *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 418–427.

[96] P. Auer, P. Gajane, and R. Ortner, “Adaptively tracking the best arm with an unknown number of distribution changes,” 2018.

[97] P. Auer, Y. Chen, P. Gajane, *et al.*, “Achieving optimal dynamic regret for non-stationary bandits without prior information,” in *Conference on Learning Theory*, PMLR, 2019, pp. 159–163.

[98] Y. Chen, C.-W. Lee, H. Luo, and C.-Y. Wei, “A new algorithm for non-stationary contextual bandits: Efficient, optimal and parameter-free,” in *Conference on Learning Theory*, PMLR, 2019, pp. 696–726.

[99] L. Besson and E. Kaufmann, "The generalized likelihood ratio test meets klucb: An improved algorithm for piece-wise non-stationary bandits," *Proceedings of Machine Learning Research vol XX*, vol. 1, p. 35, 2019.

[100] X. Zhou, Y. Xiong, N. Chen, and X. Gao, "Regime switching bandits," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[101] P. Auer, T. Jaksch, and R. Ortner, "Near-optimal regret bounds for reinforcement learning," *Advances in neural information processing systems*, vol. 21, 2008.

[102] N. Chen, C. Wang, and L. Wang, "Learning and optimization with seasonal patterns," *arXiv preprint arXiv:2005.08088*, 2020.

[103] K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," *Handbook of Reinforcement Learning and Control*, pp. 321–384, 2021.

[104] J. R. Birge, H. Chen, N. B. Keskin, and A. Ward, "To interfere or not to interfere: Information revelation and price-setting incentives in a multiagent learning environment," *Available at SSRN 3864227*, 2021.

[105] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2022.

[106] R. Kleinberg, A. Slivkins, and E. Upfal, "Bandits and experts in metric spaces," *arXiv preprint arXiv:1312.1277*, 2013.

[107] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello, "The sample average approximation method for stochastic discrete optimization," *SIAM Journal on Optimization*, vol. 12, no. 2, pp. 479–502, 2002.

[108] H. Rahimian and S. Mehrotra, "Distributionally robust optimization: A review," *arXiv preprint arXiv:1908.05659*, 2019.

[109] J. Duchi, P. Glynn, and H. Namkoong, "Statistics of robust optimization: A generalized empirical likelihood approach," *arXiv preprint arXiv:1610.03425*, 2016.

[110] K. S. Azoury, "Bayes solution to dynamic inventory models under unknown demand distribution," *Management science*, vol. 31, no. 9, pp. 1150–1160, 1985.

[111] A. E. Lim, J. G. Shanthikumar, and Z. M. Shen, "Model uncertainty, robust optimization, and learning," in *Models, Methods, and Applications for Innovative Decision Making*, INFORMS, 2006, pp. 66–94.

[112] Hass Avocado Board. "Volume data and category data." (Jul. 2022), [Online]. Available: https://hassavocadoboard.com/.

[113] E. R. Fox, A. Birt, K. B. James, H. Kokko, S. Salverson, and D. L. Soflin, "ASHP guidelines on managing drug product shortages in hospitals and health systems," *American Journal of Health-System Pharmacy*, vol. 66, no. 15, pp. 1399–1406, 2009.

[114] HSCA, "Group purchasing organizations (GPOs) work to maintain access to product supply for America's health care providers," Healthcare Supply Chain Association, Tech. Rep., 2011.

[115] ASA, "2012 ASA drug shortage survey results," American Society of Anesthesiologists (ASA), Tech. Rep., 2012.

[116] J. Woodcock and M. Wosinska, "Economic and technological drivers of generic sterile injectable drug shortages," *Clinical Pharmacology & Therapeutics*, vol. 93, no. 2, pp. 170–176, 2013.

[117] FDA, "Strategic plan for preventing and mitigating drug shortages," U.S. Department of Health and Human Services, Tech. Rep., 2013.

[118] GAO, "Drug shortages: FDA's ability to respond should be strengthened," U.S. Government Accountability Office, Tech. Rep., 2011.

[119] GAO, "Drug Shortages: Certain factors are strongly associated with this persistent public health challenge," U.S. Government Accountability Office, Tech. Rep., 2016.

[120] NORC, "Recent trends in hospital drug spending and manufacturer shortages," NORC at the University of Chicago, Tech. Rep., 2019.

[121] S. Saedi, O. E. Kundakcioglu, and A. C. Henry, "Mitigating the impact of drug shortages for a healthcare facility: An inventory management approach," *European Journal of Operational Research*, vol. 251, no. 1, pp. 107–123, May 2016, ISSN: 0377-2217.

[122] J. Jia and H. Zhao, "Mitigating the U.S. drug shortages through Pareto-improving contracts," en, *Production and Operations Management*, vol. 26, no. 8, pp. 1463–1480, 2017.

[123] J. Lee, H. S. Lee, H. Shin, and V. Krishnan, "Alleviating Drug Shortages: The Role of Mandated Reporting Induced Operational Transparency," *Management Science*, Jan. 2021.

[124] J. Siegel, "Impact of the Medicare prescription drug improvement and Modernization Act on the management of colorectal cancer," en, *American Journal of Health-System Pharmacy*, vol. 63, no. suppl 2, S18–S21, May 2006.

[125] M. Jacobson, C. C. Earle, M. Price, and J. P. Newhouse, "How Medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment," *Health Affairs*, vol. 29, no. 7, pp. 1391–1399, Jul. 2010.

[126] A. Alpert, H. Hsi, and M. Jacobson, "Evaluating the role of payment policy in driving vertical integration in the oncology market," *Health Affairs*, vol. 36, no. 4, pp. 680–688, 2017.

[127] E. R. Berndt, R. M. Conti, and S. J. Murphy, "The landscape of US generic prescription drug markets, 2004-2016," National Bureau of Economic Research, Tech. Rep., 2017.

[128] D. Levinson, "Medicaid drug price comparison: Average sales price to average wholesale price," *Washington, DC, Office of the Inspector General*, 2005.

[129] P. M. Danzon, G. R. Wilensky, and K. E. Means, "Alternative strategies for medicare payment of outpatient prescription drugs—part B and beyond," en, *The American Journal of Managed Care*, vol. 11, no. 3, p. 9, 2005.

[130] D. B. Ridley and C. Y. Lee, "Does Medicare reimbursement drive up drug launch prices?" *The Review of Economics and Statistics*, pp. 1–45, Jun. 2019, ISSN: 0034-6535.

[131] J. Clemens and J. D. Gottlieb, "In the shadow of a giant: Medicare's influence on private physician payments," *Journal of Political Economy*, vol. 125, no. 1, pp. 1–39, 2017, ISBN: 0022-3808 Publisher: University of Chicago Press Chicago, IL.

[132] L. R. Burns, *The health care value chain: Producers, purchasers, and providers*, eng, 1st ed.. San Francisco: Jossey-Bass, 2002, ISBN: 0787960217.

[133] L. Schwarz, "Healthcare-product supply chains: Medical surgical supplies, pharmaceuticals, and orthopedic devices: Flows of product, information, and dollars," in *Handbook of Healthcare Delivery Systems*, CRC Press, 2010, ISBN: 9781439803622.

[134] H. Zhao, "Chapter 35: Supply chain optimization in healthcare," in *Advances and Trends in Optimization with Engineering Applications*, T. Terlaky, M. F. Anjos, and S. Ahmed, Eds., SIAM, 2016, pp. 469–478.

[135] S.-H. Cho and H. Zhao, "Healthcare supply chain," in *Handbook of Healthcare Analytics: Theoretical Minimum for Conducting 21st Century Research on Healthcare Operations*, S. Tayur and T. Dai, Eds., John Wiley & Sons, 2018, pp. 159–185.

[136] L. B. Schwarz and H. Zhao, "The unexpected impact of information sharing on US pharmaceutical supply chains," *INFORMS Journal on Applied Analytics*, vol. 41, no. 4, pp. 354–364, Aug. 2011.

[137] H. Zhao, C. Xiong, S. Gavirneni, and A. Fein, "Fee-for-service contracts in pharmaceutical distribution supply chains: Design, analysis, and management," *Manufacturing & Service Operations Management*, vol. 14, no. 4, pp. 685–699, 2012.

[138] K. Iacocca, Y. Zhao, and A. Fein, "Resell versus direct models in brand drug distribution," en, *International Journal of Pharmaceutical and Healthcare Marketing*, vol. 7, no. 4, A. Mukherjee Yam Limbu, Ed., pp. 324–340, Nov. 2013.

[139] J. V. Gray, E. Siemsen, and G. Vasudeva, "Colocation still matters: Conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry," en, *Management Science*, vol. 61, no. 11, pp. 2760–2781, Nov. 2015.

[140] J. Gray and B. Tomlin, "Contract manufacturing and quality risk: Theory and empirical evidence," *SSRN Electronic Journal*, Jan. 2016.

[141] R. Navarro, *Managed care pharmacy practice*. Jones & Bartlett Learning, 2009.

[142] G. P. Cachon, "Supply chain coordination with contracts," *Handbooks in operations research and management science*, vol. 11, pp. 227–339, 2003.

[143] T. A. Hemphill, "U.S. Pharmaceutical Gray Markets: Why Do They Persist—and What to Do about Them?" *Business and Society Review*, vol. 121, no. 4, pp. 529–547, Dec. 2016, ISSN: 0045-3609.

[144] C. Stomberg, "Drug shortages, pricing, and regulatory activity," National Bureau of Economic Research, Working Paper 22912, Dec. 2016.

[145] T. L. Beauchamp and J. F. Childress, *Principles of biomedical ethics*. Oxford University Press, USA, 2001, ISBN: 0-19-514331-0.

[146] G. Persad, A. Wertheimer, and E. J. Emanuel, "Principles for allocation of scarce medical interventions," *The Lancet*, vol. 373, no. 9661, pp. 423–431, 2009, ISSN: 0140-6736.

[147] D. B. White, M. H. Katz, J. M. Luce, and B. Lo, "Who should receive life support during a public health emergency? Using ethical principles to improve allocation decisions," *Annals of Internal Medicine*, vol. 150, no. 2, pp. 132–138, 2009, ISSN: 0003-4819.

[148] L. P. Scheunemann and D. B. White, "The ethics and reality of rationing in medicine," eng, *Chest*, vol. 140, no. 6, pp. 1625–1632, Dec. 2011.

[149] C. L. Ventola, "The drug shortage crisis in the United States," *Pharmacy and Therapeutics*, vol. 36, no. 11, pp. 740–757, Nov. 2011.

[150] G. Wellman, "National supply-chain survey of drug manufacturer back orders," *American Journal of Health-System Pharmacy*, vol. 58, no. 13, pp. 1224–1228, Jul. 2001.

[151] L. Tyler, "Understanding drug and managing shortages," in *Proceedings of breakfast symposium, 37th ASHP Midyear Clinical Meeting*, Atlanta, Georgia: ASHP, 2002, pp. 1–13.

[152] Y. Abutaleb, "Trump administration proposes first rule on health-care cost transparency," en-US, *Washington Post*, Jul. 2019.

[153] N. Avenue, "Will President Trump's hospital price transparency plan lower costs?" en, *Forbes*, 2019.

[154] K. Keith, "Unpacking the executive order on health care price transparency and quality," *Health Affairs Blog. Retrieved June*, vol. 26, p. 2019, 2019.

[155] World Bank, "GINI index (World Bank estimate)," 2020. [Online]. Available: https://data.worldbank.org/indicator/SI.POV.GINI.

[156] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal processing magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[157] A. Lewbel and O. Linton, "Nonparametric censored and truncated regression," *Econometrica*, vol. 70, no. 2, pp. 765–779, 2002.

[158] FDA, "Quality Management Maturity: Essential for Stable U.S. Supply Chains of Quality Pharmaceuticals," Tech. Rep., 2022. [Online]. Available: https://www.fda.gov/media/157432/download.

[159] H. Schildberg-Hörisch, "Are risk preferences stable?" *Journal of Economic Perspectives*, vol. 32, no. 2, pp. 135–54, 2018.

[160] G. Weinberg, "Kullback–leibler divergence and the pareto–exponential approximation," *SpringerPlus*, vol. 5, no. 1, pp. 1–9, 2016.

[161] D. Maclaurin, D. Duvenaud, and R. Adams, "Gradient-based hyperparameter optimization through reversible learning," in *International conference on machine learning*, PMLR, 2015, pp. 2113–2122.

[162] L. Franceschi, M. Donini, P. Frasconi, and M. Pontil, "Forward and reverse gradient-based hyperparameter optimization," in *International Conference on Machine Learning*, PMLR, 2017, pp. 1165–1173.

[163] Z. Meng, C. Dang, R. Shen, and M. Jiang, "An objective penalty function of bilevel programming," *Journal of Optimization Theory and Applications*, vol. 153, no. 2, pp. 377–387, 2012.

[164] S. Dempe and S. Franke, "On the solution of convex bilevel optimization problems," *Computational Optimization and Applications*, vol. 63, no. 3, pp. 685–703, 2016.

[165] L. E. Ghaoui, M. Oks, and F. Oustry, "Worst-case value-at-risk and robust portfolio optimization: A conic programming approach," *Operations research*, vol. 51, no. 4, pp. 543–556, 2003.

[166] K. Natarajan, D. Pachamanova, and M. Sim, "Incorporating asymmetric distributional information in robust value-at-risk optimization," *Management Science*, vol. 54, no. 3, pp. 573–585, 2008.

[167] E. A. Ok, *Real analysis with economic applications*. Princeton University Press, 2011.

[168] S. Lang, *Real and functional analysis*. Springer Science & Business Media, 2012, vol. 142.

[169] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations research*, vol. 58, no. 3, pp. 595–612, 2010.

[170] K. Natarajan, M. Sim, and J. Uichanco, "Asymmetry and ambiguity in newsvendor models," *Management Science*, vol. 64, no. 7, pp. 3146–3167, 2018.

[171] H. Xu, Y. Liu, and H. Sun, "Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods," *Mathematical Programming*, vol. 169, no. 2, pp. 489–529, 2018.

[172] A. E. Lim and J. G. Shanthikumar, "Relative entropy, exponential utility, and robust dynamic pricing," *Operations Research*, vol. 55, no. 2, pp. 198–214, 2007.

[173] G. Bayraksan and D. K. Love, "Data-driven stochastic programming using phi-divergences," in *The Operations Research Revolution*, INFORMS, 2015, pp. 1–19.

[174] H. Lam, "Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization," *Operations Research*, vol. 67, no. 4, pp. 1090–1105, 2019.

[175] V. A. Nguyen, S. Shafieezadeh-Abadeh, D. Kuhn, and P. M. Esfahani, “Bridging bayesian and minimax mean square error estimation via wasserstein distributionally robust optimization,” *arXiv preprint arXiv:1911.03539*, 2019.

[176] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[177] R. Hettich and K. O. Kortanek, “Semi-infinite programming: Theory, methods, and applications,” *SIAM review*, vol. 35, no. 3, pp. 380–429, 1993.

[178] A. Shapiro, “On duality theory of conic linear problems,” in *Semi-infinite programming*, Springer, 2001, pp. 135–165.

[179] N. Fournier and A. Guillin, “On the rate of convergence in wasserstein distance of the empirical measure,” *Probability Theory and Related Fields*, vol. 162, no. 3, pp. 707–738, 2015.

[180] J. A. Adell and P. Jodrá, “Exact kolmogorov and total variation distances between some familiar discrete distributions,” *Journal of Inequalities and Applications*, vol. 2006, pp. 1–8, 2006.

[181] G. P. Cachon and S. Netessine, “Game theory in supply chain analysis,” in *Models, methods, and applications for innovative decision making*, INFORMS, 2006, pp. 200–233.

[182] L. R. Burns and J. A. Lee, “Hospital purchasing alliances: Utilization, services, and performance,” *Health Care Management Review*, vol. 33, no. 3, pp. 203–215, 2008.

[183] Q. Hu, L. Schwarz, and N. Uhan, “The impact of group purchasing organizations on healthcare-product supply chains,” *Manufacturing & Service Operations Management*, vol. 14, no. 1, pp. 7–23, 2012.

A. APPENDICES FOR CHAPTER 2

A.1 Additional Material for Section 2.3 (supplier's classical problem)

A.1.1 Zero Salvage Value

We may assume zero salvage value for the retailer, having a positive salvage value does not materially change our analysis. To see this, let b_i be the per unit salvage value for product i, then $\pi_r(\boldsymbol{q};\boldsymbol{\xi},\boldsymbol{w}) = \sum_{i\in[n]}(s_i \min\{\xi_i, q_i\} - w_i q_i - b_i \max\{0, q_i - \xi_i\}) = (s_i + b_i)\min\{\xi_i, q_i\} - (w_i + b_i)q_i = \sum_{i\in[n]}(s_i' \min\{\xi_i, q_i\} - w_i' q_i)$ where $s_i' = s_i + b_i$ and $w_i' = w_i + b_i$.

A.2 Additional Material for Section 2.4 (supplier's classical problem in the data-driven regime)

A.2.1 Hardness of Distinguishing the Exponential Distribution and the Pareto Distribution

Let $\mathbb{P}_\lambda$ denote the exponential distribution with mean $1/\lambda$ and let $\mathbb{P}_{\theta,\alpha}$ denote the Pareto distribution with scale parameter $\theta > 0$ and shape parameter $\alpha > 0$. The probability density function of $\mathbb{P}_\lambda$ is $f(x;\lambda) = \lambda \exp(-\lambda x)$ for $\lambda > 0$ and $x \geq 0$. The probability density function of $\mathbb{P}_{\theta,\alpha}$ is $f(x;\theta,\alpha) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}$ for $\alpha, \theta > 0$ and $x \geq \theta$.

We calculate the probability of failing to distinguish between the exponential distribution and the Pareto distribution, as a function of number of samples T. The Kullback-Leibler divergence between $\mathbb{P}_\lambda$ and $\mathbb{P}_{(\theta,\alpha)}$ is $D_{KL}\left(\mathbb{P}_{(\theta,\alpha)} \| \mathbb{P}_\lambda\right) = \log\left(\frac{\alpha}{\lambda\theta}\right) + \frac{\lambda\theta}{\alpha-1} - \frac{\alpha+1}{\alpha}$, see [160].

Let $\mathbb{P}_\lambda^T \triangleq \times_{t\in[T]}\mathbb{P}_\lambda$ be the distribution of $(\xi^t)_{t\in[T]}$ when the data are generated from the exponential distribution, and let $\mathbb{P}_{(\theta,\alpha)}^T \triangleq \times_{t\in[T]}\mathbb{P}_{(\theta,\alpha)}$ be the distribution of $(\xi^t)_{t\in[T]}$ when the data are generated from the Pareto distribution.

Let $H : (\xi^t)_{t\in[T]} \to \{0,1\}$ be a statistical test where $H\left((\xi^t)_{t\in[T]}\right) = 0$ indicates that the samples are drawn from an exponential distribution, and $H\left((\xi^t)_{t\in[T]}\right) = 1$ indicates that the samples are drawn from a Pareto distribution. Then, according to Pinsker's inequality, we have

$$\begin{aligned}\mathbb{P}_{\lambda}^{T}\left(H\left((\xi^{t})_{t\in[T]}\right)=1\right)+\mathbb{P}_{(\theta,\alpha)}^{T}\left(H\left((\xi^{t})_{t\in[T]}\right)=0\right) &\geq 1-\sqrt{\frac{1}{2}D_{KL}\left(\mathbb{P}_{(\theta,\alpha)}^{T}\|\mathbb{P}_{\lambda}^{T}\right)}\\ &= 1-\sqrt{\frac{T}{2}D_{KL}\left(\mathbb{P}_{(\theta,\alpha)}\|\mathbb{P}_{\lambda}\right)}.\end{aligned}$$

Therefore, the probability that the statistical test H fails to identify the true distribution is

$$\max\left\{\mathbb{P}_{\lambda}^{T}\left(H\left((\xi^{t})_{t\in[T]}\right)=1\right),\mathbb{P}_{(\theta,\alpha)}^{T}\left(H\left((\xi^{t})_{t\in[T]}\right)=0\right)\right\}\geq 1/2-1/2\sqrt{\frac{T}{2}D_{KL}\left(\mathbb{P}_{(\theta,\alpha)}\|\mathbb{P}_{\lambda}\right)}.$$

A.3 Additional Material for Section 2.5 (supplier's robust problem)

A.3.1 Cross-Validation

Let $\Theta\triangleq\{\theta_1,\ldots,\theta_m\}$ be a finite set of admissible confidence parameter values. We want to choose $\theta^{cv}\in\Theta$ to optimize the out-of-sample performance of the robust model.

We let $\mathcal{H}=\{(\boldsymbol{w}^t,\boldsymbol{q}^t,\boldsymbol{\xi}^t)\}_{t=1}^{T}$ denote the historical data consisting of T data points. To run $k-$fold cross validation, we first partition $\mathcal{H}$ into k subsets, where k is chosen as an input. For example, we choose $k=3$ in our numerical experiments. In each run, we hold out one subset to be the validation data set, and merge the remaining sets into the training data set. On the i^{th} run, using the training set, we first obtain the center distribution denoted $\hat{\mathbb{P}}_{-i}$. Then, for each value $\theta_j\in\Theta$, we compute the corresponding robust optimal decision $\boldsymbol{w}_i^*(\theta_j)\in\mathcal{W}$ and evaluate this decision using the empirical distribution constructed from the i^{th} validation data set.

For each validation data set, we obtain an out-of-sample profit for each $\theta_j\in\Theta$. We take the average of out-of-sample profit across the k runs, and then choose the θ_j which gives the best average out-of-sample profit.

Algorithm 3 illustrates the k-fold cross validation procedure to find the confidence parameter θ^{cv} that maximizes the out-of-sample performance for the robust model.

Algorithm 3 k-fold cross validation

Input: history data $\mathcal{H}$; finite set of confidence parameter values $\Theta = \{\theta_1, \ldots, \theta_m\}$

Shuffle $\mathcal{H}$, partition it into k subsets, and denote each subset by $\mathcal{H}_i$ for $i \in [k]$

for $i = 1, \cdots, k$ **do**

Hold-out $\mathcal{H}_i$ as the test set, and merge the remaining subsets as the training set, denoted as $\mathcal{H}_{-i}$

Let $\hat{\mathbb{P}}^{\mathrm{e}}_{-i}$ be the empirical distribution corresponding to demand samples in $\mathcal{H}_{-i}$

Find the center $\hat{\mathbb{P}}_{-i} \leftarrow \arg\min_{\mathbb{P} \in \mathcal{D}^o_T} W^2(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_{-i})$

for $j = 1, \cdots, m$ **do**

Find $\boldsymbol{w}_i^*(\theta_j) \leftarrow \arg\max_{\boldsymbol{w} \in \mathcal{W}} \min_{\mathbb{P} \in \mathcal{D}_{\theta_j}(\hat{\mathbb{P}}_{-i})} (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \mathbb{P})$

Let $\hat{\mathbb{P}}^{\mathrm{e}}_i$ be the empirical distribution corresponding to demand samples in $\mathcal{H}_i$

Evaluate the corresponding profit $\pi\left(\boldsymbol{w}_i^*(\theta_j); \hat{\mathbb{P}}^{\mathrm{e}}_i\right) \leftarrow (\boldsymbol{w}_i^*(\theta_j) - \boldsymbol{c})^\top \boldsymbol{q}\left(\boldsymbol{w}_i^*(\theta_j); \hat{\mathbb{P}}^{\mathrm{e}}_i\right)$

$\theta^{cv} \leftarrow \arg\max_{j \in [m]} 1/k \sum_{i=1}^k \pi\left(\boldsymbol{w}_i^*(\theta_j); \hat{\mathbb{P}}^{\mathrm{e}}_i\right)$.

A.3.2 Bilevel Programming Reformulation

We can reformulate Problem (2.7) (or Problem (2.10) or Problem (2.11)) as a bilevel program with upper level decision variable $\boldsymbol{w}$ and lower level decision variable $\boldsymbol{q}$.

Recall $\mathbb{P}_{\boldsymbol{w},\theta}$ is the worst-case distribution, i.e., the optimal solution to Eq. (2.6). Here we view $\mathbb{P}_{\boldsymbol{w},\theta}$ as a function that depends on $\boldsymbol{w}$. Let $\mathcal{S}(\boldsymbol{w}) \triangleq \arg\max_{\boldsymbol{q} \geq 0} \mathbb{E}_{\mathbb{P}_{\boldsymbol{w},\theta}} [\pi_r(\boldsymbol{q}; \boldsymbol{\xi}, \boldsymbol{w})]$ be the set of solutions of the lower level problem, we observe that $\mathcal{S}(\boldsymbol{w})$ is always a singleton. Notice that $\mathcal{S}(\boldsymbol{w})$ depends on $\boldsymbol{w}$ both through the retailer's objective function $\pi_r(\boldsymbol{q}; \boldsymbol{\xi}, \boldsymbol{w})$ and the mapping $\mathbb{P}_{\boldsymbol{w},\theta}$. We then obtain the bilevel programming problem:

$$\max_{\boldsymbol{c} \leq \boldsymbol{w} \leq \boldsymbol{s}} \left\{ (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \boldsymbol{q} \in \mathcal{S}(\boldsymbol{w}) \right\}, \tag{A.1}$$

where the dependence on $\boldsymbol{w}$ is absorbed into $\mathcal{S}(\boldsymbol{w})$.

Problem (A.1) is a bilevel programming problem with a convex lower level problem. To solve for the optimal wholesale price, we refer to the literature on bilevel programs where the lower level optimal solution is a singleton [161], [162], or where the lower level problem is convex [163], [164]. However, we point out that Problem (A.1) is still hard to solve because of the complex dependence introduced by the mapping $\mathbb{P}_{\boldsymbol{w},\theta}$.

A.3.3 Value-at-risk

For $n = 1$, we can connect the supplier's worst-case profit with the worst-case value-at-risk of demand. Recall the value-at-risk for loss ξ at level ϵ is $\mathrm{VaR}_\epsilon(\xi) \triangleq \min\{x : \mathbb{P}(x \leq \xi) \leq \epsilon\}$. For $n = 1$, Problem (2.6) is equivalent to:

$$\min_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)} q(w; \mathbb{P}) = \min_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)} \mathrm{VaR}_{w/s}(\xi), \tag{A.2}$$

which is the worst-case value-at-risk for ξ at level w/s with uncertainty set $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$ (see, e.g., [165], [166]). We can interpret $\mathcal{D}_T^o$ as the set of distributions for which we know the value-at-risk of the random loss at certain levels (since the retailer's optimal order quantity is a quantile function of the true demand distribution). Then, Problem (A.2) is the worst-case value-at-risk at level w/s among distributions in $\mathcal{D}_T^o$.

A.3.4 Proof of Proposition 2.5.1 (equivalence of supplier's robust problems)

It is sufficient to show that the inner minimization problems to compute the worst-case profit within Problems (2.7), (2.10), and (2.11) are all equivalent. For any wholesale price $\boldsymbol{w} \in \mathcal{W}$ and any $\delta > 0$, define

$$v_1(\boldsymbol{w}) \triangleq \inf_{\mathbb{P} \in \mathcal{D}} (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}), \tag{A.3}$$

with $\delta-$optimal solution $\mathbb{P}^*_{(1)}$; define

$$v_2(\boldsymbol{w}) \triangleq \inf_{\mathbb{P} \in \mathcal{D}, \boldsymbol{q} \geq 0} \{(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i, \forall i \in [n]\}, \tag{A.4}$$

with $\delta-$optimal solution $(\mathbb{P}^*_{(2)}, \boldsymbol{q}^*_{(2)})$; and define

$$v_3(\boldsymbol{w}) \triangleq \inf_{\boldsymbol{q} \geq 0} \{(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \exists \mathbb{P} \in \mathcal{D} : \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i, \forall i \in [n]\}, \tag{A.5}$$

with $\delta-$optimal solution $\boldsymbol{q}^*_{(3)} = (q^*_{(3),i})_{i\in[n]}$, and let $\mathbb{P}^*_{(3)} \in \mathcal{D}$ be the corresponding distribution such that $\mathbb{P}^{*,i}_{(3)}(\xi_i \leq q^*_{(3),i}) \geq 1 - w_i/s_i$ for all $i \in [n]$, where $\mathbb{P}^{*,i}_{(3)}$ is the $i-$th marginal distribution of $\mathbb{P}^*_{(3)}$. We will show the following parts:

(i) For any wholesale price $\boldsymbol{c} \leq \boldsymbol{w} \leq \boldsymbol{s}$, $v_1(\boldsymbol{w}) = v_2(\boldsymbol{w})$. Furthermore, $(\mathbb{P}^*_{(1)}, \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(1)}))$ is $\delta-$optimal to Problem (A.4) and $\mathbb{P}^*_{(2)}$ is $\delta-$optimal to Problem (A.3).

(ii) For any wholesale price $\boldsymbol{c} \leq \boldsymbol{w} \leq \boldsymbol{s}$, $v_2(\boldsymbol{w}) = v_3(\boldsymbol{w})$. Furthermore, $(\mathbb{P}^*_{(3)}, \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(3)}))$ is $\delta-$optimal to Problem (A.4) and $\boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(2)})$ is $\delta-$optimal to Problem (A.5).

To prove Part (i), notice that $(\mathbb{P}^*_{(1)}, \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(1)}))$ is feasible to Problem (A.4), since by definition $q_i(w_i; \mathbb{P}^{*,i}_{(1)}) = \min\{q \geq 0 : \mathbb{P}^{*,i}_{(1)}(\xi_i \leq q) \geq 1 - w_i/s_i, \forall i \in [n]\}$. Thus $v_1(\boldsymbol{w}) + \delta \geq v_2(\boldsymbol{w})$. At the same time, $\mathbb{P}^*_{(2)}$ is feasible to Problem (A.3) since $\mathbb{P}^*_{(2)} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)$. Furthermore, $\boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(2)}) \leq \boldsymbol{q}^*_{(2)}$ by the definition of $\boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(2)})$. Thus, we have $(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}; \mathbb{P}^*_{(2)}) \leq (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}^*_{(2)} \leq v_2(\boldsymbol{w}) + \delta$. Since $\delta > 0$ is arbitrary, the desired result follows.

To prove Part (ii), first notice that $\boldsymbol{q}^*_{(2)}$ is feasible to Problem (A.5) and thus $v_2(\boldsymbol{w}) + \delta \geq v_3(\boldsymbol{w})$. On the other hand, $(\mathbb{P}^*_{(3)}, \boldsymbol{q}^*_{(3)})$ is also feasible to Problem (A.4). Thus $v_3(\boldsymbol{w}) + \delta \geq v_2(\boldsymbol{w})$, and since δ was arbitrary this part follows.

A.3.5 Proof of Theorem 2.5.1 (asymptotics)

The sequence of i.i.d. demand samples is generated by $\bar{\mathbb{P}}^\infty \triangleq \times_{t=1}^\infty \bar{\mathbb{P}}$. Our uncertainty set $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$ is a random variable that depends on $\bar{\mathbb{P}}^T$ (through the supplier's first $T \geq 1$ data points). We want to estimate the probability that $\mathcal{D}_\theta(\hat{\mathbb{P}}_T) \ni \bar{\mathbb{P}}$. This is related to the probability that $W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq \theta$ (with respect to the sampling distribution $\bar{\mathbb{P}}^T$). By the triangle inequality (which is applicable because W_2 is a metric), we have $W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}^{\mathrm{e}}_T) + W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \hat{\mathbb{P}}_T)$. Since $\hat{\mathbb{P}}_T$ satisfies $W_2(\hat{\mathbb{P}}_T, \hat{\mathbb{P}}^{\mathrm{e}}_T) \leq W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}^{\mathrm{e}}_T)$ by construction, we have $W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq 2\, W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}^{\mathrm{e}}_T)$.

The next result (Theorem A.3.1) upper bounds the probability that $W_2(\hat{\mathbb{P}}_T, \bar{\mathbb{P}}) \geq \theta$.

Theorem A.3.1. *(Proof in Appendix II, A.7.5) Suppose Assumption 2.3.1 holds. Then for any* $\theta > 0$,

$$\bar{\mathbb{P}}^T \left\{ W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \geq \theta \right\} \leq \begin{cases} c_1 \exp\left(-c_2 T \left(\frac{\theta}{2}\right)^{max\{n,2\}} \right), & \theta/2 \leq 1, \\ c_1 \exp\left(-c_2 T \left(\frac{\theta}{2}\right)^{a} \right), & \theta/2 > 1, \end{cases}$$

where a, c_1, c_2 *are positive constants that only (possibly) depend on* n.

The Wasserstein ball centered at $\hat{\mathbb{P}}_T$ with radius $\theta_T(\alpha_T)$ includes the true distribution $\bar{\mathbb{P}}$ with probability no less than $1 - \alpha_T$. Specifically, this definition of $\theta_T(\alpha_T)$ ensures that $\bar{\mathbb{P}}^T\{W_2(\hat{\mathbb{P}}_T, \bar{\mathbb{P}}) \geq \theta_T(\alpha_T)\} \leq \alpha_T$ for all $T \geq 1$. Furthermore, $\lim_{T\to\infty} \theta_T(\alpha_T) = 0$. It then follows immediately from Theorem A.3.1 that the probability that $\mathcal{D}_{\theta_T(\alpha)}(\hat{\mathbb{P}}_T)$ contains the true distribution $\bar{\mathbb{P}}$ is lower bounded by

$$\begin{aligned} \bar{\mathbb{P}}^T \left\{ \mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T) \ni \bar{\mathbb{P}} \right\} &= \bar{\mathbb{P}}^T \left\{ \bar{\mathbb{P}} \in \mathcal{D}_T^o, W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq \theta_T(\alpha_T) \right\} \\ &= \bar{\mathbb{P}}^T \left\{ W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq \theta_T(\alpha_T) \right\} \geq 1 - \alpha_T, \end{aligned}$$

where the second equality follows since $\bar{\mathbb{P}} \in \mathcal{D}_T^o$ by definition.

Now we proceed to prove Theorem 2.5.1. Define the value functions

$$v(\boldsymbol{w}, \boldsymbol{q}, \mathbb{P}) = \begin{cases} (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}, & \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i, \quad \forall i \in [n], \\ +\infty, & \text{otherwise}, \end{cases}$$

and

$$v(\boldsymbol{w}, \mathbb{P}) = \min_{\boldsymbol{q} \geq 0} v(\boldsymbol{w}, \boldsymbol{q}, \mathbb{P}). \tag{A.6}$$

Also define $h(\boldsymbol{\lambda}, \boldsymbol{w}, \boldsymbol{q}, \boldsymbol{\xi}) \triangleq \sum_{i\in[n]} (w_i - c_i) q_i - \sum_{i\in[n]} \lambda_i \left(\mathbb{1}(\xi_i \leq q_i) - 1 + w_i/s_i \right)$.

According to [30, Lemma A.1], as $\mathbb{1}(\xi \leq q)$ is upper semicontinuous in ξ and $\mathbb{1}(\xi \leq q) \leq L(1 + |\xi|)$ for some $L > 0$, there exists a non-increasing sequence of Lipschitz continuous functions $(g_k(q,\cdot))_{k\in\mathbb{N}}$ that converges pointwise to $\mathbb{1}(\cdot \le q)$. We thus also define $h_k(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi}) \triangleq \sum_{i\in[n]}(w_i-c_i)q_i - \sum_{i\in[n]}\lambda_i\left(g_k(q_i,\xi_i)-1+w_i/s_i\right)$,

$$v_k(\boldsymbol{w},\boldsymbol{q},\mathbb{P}) = \begin{cases} (\boldsymbol{w}-\boldsymbol{c})^\top\boldsymbol{q}, & \mathbb{E}_{\mathbb{P}^i}[g_k(q,\xi)] \ge 1-w_i/s_i \quad \forall i\in[n], \\ +\infty, & \text{otherwise,} \end{cases}$$

and $v_k(\boldsymbol{w},\mathbb{P}) = \min_{\boldsymbol{q}\ge 0} v_k(\boldsymbol{w},\boldsymbol{q},\mathbb{P})$.

We need the following auxiliary lemma to prove Theorem 2.5.1.

Lemma A.3.1. *Suppose Assumptions 2.3.1 and 2.5.1 hold.*

(i) We have: $v(\boldsymbol{w},\mathbb{P}) = \inf_{\{q_i\}\ge 0}\sup_{\{\lambda_i\}\ge 0}\mathbb{E}_{\mathbb{P}}[h(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi})]$ *and* $v_k(\boldsymbol{w},\mathbb{P}) = \inf_{\{q_i\}\ge 0}\sup_{\{\lambda_i\}\ge 0}\mathbb{E}_{\mathbb{P}}[h_k(\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{q},\boldsymbol{\xi})]$.

(ii) $v(\boldsymbol{w},\bar{\mathbb{P}})$ *is continuous in* $\boldsymbol{w}$*.*

(iii) There exists $\boldsymbol{w}^* \ge \boldsymbol{c}$ *such that* $(\boldsymbol{w}^*-\boldsymbol{c})^\top\boldsymbol{q}(\boldsymbol{w}^*;\bar{\mathbb{P}}) = \pi^*$*.*

(iv) There exists an optimal solution $\boldsymbol{q}^*(\boldsymbol{w};\mathbb{P}) = (q_i^*(w_i;\mathbb{P}^i))_{i\in[n]} \ge 0$ *to Problem (A.6) that satisfies* $v(\boldsymbol{w},\boldsymbol{q}^*(\boldsymbol{w};\mathbb{P}),\mathbb{P}) = v(\boldsymbol{w},\mathbb{P})$*.*

(v) We have $\lim_{k\to\infty} v_k(\boldsymbol{w},\mathbb{P}) = v(\boldsymbol{w},\mathbb{P})$*.*

Proof of Lemma A.3.1. (i) This part follows since these are equivalent forms of the primal minimization problems, using an indicator function for constraint violation.

(ii) This part follows from the Berge Maximum Theorem [167], which applies because $(\boldsymbol{w}-\boldsymbol{c})^\top\boldsymbol{q}$ is continuous in $(\boldsymbol{w},\boldsymbol{q})$ and the correspondence

$$\Gamma(\boldsymbol{w}) = \left\{\boldsymbol{q} : \bar{\mathbb{P}}^i(\xi_i \le q_i) \ge 1-w_i/s_i,\ \forall i\in[n]\right\}$$

is continuous in $\boldsymbol{w}$, which holds because of the atomlessness of $\bar{\mathbb{P}}$.

(iii) Notice that $\pi^* = \sup_{\boldsymbol{c}\le\boldsymbol{w}\le\boldsymbol{s}} v(\boldsymbol{w},\bar{\mathbb{P}})$. By the continuity of $v(\cdot,\bar{\mathbb{P}})$ (from Part (ii)) and the compactness of the feasible region, there exists $\boldsymbol{w}^*\ge\boldsymbol{c}$ such that $v(\boldsymbol{w}^*,\bar{\mathbb{P}}) = \pi^*$.

(iv) Since $\mathbb{P}^i(\xi\le q)$ is upper semicontinuous in q, $\Gamma(\boldsymbol{w})$ is closed. Plus, we can restrict the feasible region of Problem (A.6) to Ξ. Then the existence of $\boldsymbol{q}^*(\boldsymbol{w};\mathbb{P})$ follows from Weierstrass theorem, which states that a continuous function attains its maximum/minimum over a compact set.

(v) Notice that $v(\boldsymbol{w}, \mathbb{P}) = \sum_{i\in[n]} \min_{q_i \geq 0}\{(w_i - c_i)q_i : \mathbb{P}^i(\xi_i \leq q_i) \geq 1 - w_i/s_i\}$ and $v_k(\boldsymbol{w}, \mathbb{P}) = \sum_{i\in[n]} \min_{q_i\geq 0}\{(w_i - c_i)q_i : \mathbb{E}_{\mathbb{P}^i}[g_k(q, \xi)] \geq 1 - w_i/s_i\}$. Thus, it suffices to show that $\lim_{k\to\infty} v_k(w, \mathbb{P}) = v(w, \mathbb{P})$ for the one-dimensional case.

Since $g_k(q, \xi) \geq \mathbb{1}(\xi \leq q)$, we always have $v_k(w, \mathbb{P}) \leq v(w, \mathbb{P})$. Now consider any order quantity $q' < q^*(w; \mathbb{P}) - \epsilon/(w - c)$. Thus, q' is not feasible to Problem (A.6) (for the one-dimensional case), since otherwise $q^*(w; \mathbb{P})$ would not be optimal. We then have $\mathbb{P}(\xi \leq q') < 1-w/s$. Let $\delta = 1-w/s-\mathbb{P}(\xi \leq q') > 0$. Since $\lim_{k\to\infty} \mathbb{E}_{\mathbb{P}}[g_k(q, \xi)] = \mathbb{P}(\xi \leq q)$ (which follows from the Monotone Convergence Theorem [168], which applies because $|\mathbb{E}_{\mathbb{P}}[g_k(q, \xi)]| < \infty$), then it follows that there exists K' such that for all $k \geq K'$, $0 \leq \mathbb{E}_{\mathbb{P}}[g_k(q', \xi)]-\mathbb{P}(\xi \leq q') < \delta$, implying that $\mathbb{E}_{\mathbb{P}}[g_k(q', \xi)] < \delta + \mathbb{P}(\xi \leq q') = 1 - w/s$. It follows that q' is infeasible to $v_k(w, \mathbb{P})$, and $v_k(w, \mathbb{P}) > (w - c)q' = v(w, \mathbb{P}) - \epsilon$. We have shown that for all $\epsilon > 0$, there exists $K' > 0$ such that $v_k(w, \mathbb{P}) > v(w, \mathbb{P}) - \epsilon$ for all $k \geq K'$.

□

The proof of Theorem 2.5.1 uses techniques from [30, Theorem 3.6] with some adjustments to our problem instance. For our results to be self-contained, we outline the complete proof here. We will also need the following results.

Theorem A.3.2. *([54]) For any distributions $\mathbb{P}, \mathbb{P}' \in \mathcal{P}(\Xi)$, we have*

$$W_2(\mathbb{P}, \mathbb{P}') = \sup_{f\in\mathcal{L}} \left\{ \int_\Xi f(\boldsymbol{\xi})\mathbb{P}(d\boldsymbol{\xi}) - \int_\Xi f(\boldsymbol{\xi})\mathbb{P}'(d\boldsymbol{\xi}) \right\},$$

where $\mathcal{L}$ denotes the space of all Lipschitz functions with $|f(\boldsymbol{\xi}) - f(\boldsymbol{\xi}')| \leq \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2$ for all $\boldsymbol{\xi}, \boldsymbol{\xi}' \in \Xi$.

Lemma A.3.2. *([30, Lemma 3.7]) Suppose Assumption 2.3.1 holds, that $\alpha_T \in (0, 1)$ for $T \in \mathbb{N}$ satisfies $\sum_{T\in\mathbb{N}} \alpha_T < \infty$, and that $\lim_{T\to\infty} \theta_T(\alpha_T) = 0$. Then, any sequence $\mathbb{P}_T \in \mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T)$ for $T \in \mathbb{N}$ converges under the Wasserstein metric to $\bar{\mathbb{P}}$, $\bar{\mathbb{P}}^\infty-$almost surely.*

Proof of Theorem 2.5.1. We first prove Part (i). As $\boldsymbol{c} \leq \boldsymbol{w}_{T,\theta_T(\alpha_T)} \leq \boldsymbol{s}$, we have $v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \bar{\mathbb{P}}) \leq \pi^*$. On the other hand, if $\bar{\mathbb{P}} \in \mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T)$, we have $v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \bar{\mathbb{P}}) \geq \inf_{\mathbb{P}\in\mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T)} v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \mathbb{P}) = \pi_T$. Thus by the choice of $\theta_T(\alpha_T)$, we have $\bar{\mathbb{P}}^T \left\{\pi_T \leq v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \bar{\mathbb{P}}) \leq \pi^*\right\} \geq$

$\bar{\mathbb{P}}^T \left\{ \bar{\mathbb{P}} \in \mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T) \right\} \geq 1 - \alpha_T$ for all $T \in \mathbb{N}$. As $\sum_{T\in\mathbb{N}} \alpha_T < \infty$, according to the Borel-Cantelli Lemma we have $\bar{\mathbb{P}}^\infty \left\{ \pi_T \leq v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \bar{\mathbb{P}}) \leq \pi^* \text{ for all sufficiently large } T \right\} = 1$.

To show that $\pi_T \to \pi^*$, it remains to show that $\liminf_{T\to\infty} \pi_T \geq \pi^*$. According to Lemma A.3.1 (iii), the optimal solution $\boldsymbol{w}^*$ to Problem (2.1) exists. For each $T \geq 1$, we can find the $\delta-$optimal distribution $\mathbb{P}_T^\delta \in \mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T)$ such that $v(\boldsymbol{w}^*, \mathbb{P}_T^\delta) \leq \inf_{\mathbb{P}\in\mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T)}$ $v(\boldsymbol{w}^*, \mathbb{P}) + \delta$. We then have

$$
\begin{aligned}
\liminf_{T\to\infty} \pi_T &\geq \liminf_{T\to\infty} \inf_{\mathbb{P}\in\mathcal{D}_{\theta_T(\alpha_T)}(\hat{\mathbb{P}}_T)} v(\boldsymbol{w}^*, \mathbb{P}) \\
&\geq \liminf_{T\to\infty} v(\boldsymbol{w}^*, \mathbb{P}_T^\delta) - \delta \\
&= \liminf_{T\to\infty} \inf_{\{q_i\}\geq 0} \sup_{\{\lambda_i\}\geq 0} \mathbb{E}_{\mathbb{P}_T^\delta}[h(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})] - \delta \\
&\geq \lim_{k\to\infty} \liminf_{T\to\infty} \inf_{\{q_i\}\geq 0} \sup_{\{\lambda_i\}\geq 0} \mathbb{E}_{\mathbb{P}_T^\delta}[h_k(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})] - \delta \\
&\geq \lim_{k\to\infty} \liminf_{T\to\infty} \inf_{\{q_i\}\geq 0} \sup_{\{\lambda_i\}\geq 0} \mathbb{E}_{\bar{\mathbb{P}}}[h_k(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})] - L_k W_2(\bar{\mathbb{P}}, \mathbb{P}_T^\delta) - \delta \\
&= \lim_{k\to\infty} \inf_{\{q_i\}\geq 0} \sup_{\{\lambda_i\}\geq 0} \mathbb{E}_{\bar{\mathbb{P}}}[h_k(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})] - \lim_{k\to\infty} \liminf_{T\to\infty} L_k W_2(\bar{\mathbb{P}}, \mathbb{P}_T^\delta) - \delta \\
&= \lim_{k\to\infty} \inf_{\{q_i\}\geq 0} \sup_{\{\lambda_i\}\geq 0} \mathbb{E}_{\bar{\mathbb{P}}}[h_k(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})] - \delta \qquad \bar{\mathbb{P}}^\infty - \text{almost surely} \\
&= \lim_{k\to\infty} v_k(\boldsymbol{w}^*, \bar{\mathbb{P}}) - \delta \\
&= v(\boldsymbol{w}^*, \bar{\mathbb{P}}) - \delta \\
&= \pi^* - \delta.
\end{aligned}
$$

In the above display, the first equality follows from Lemma A.3.1 (i). The third inequality holds because $h_k(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})$ converges from below to $h(\boldsymbol{\lambda}, \boldsymbol{w}^*, \boldsymbol{q}, \boldsymbol{\xi})$ by construction. The fourth inequality follows from Theorem A.3.2. The third equality follows from Lemma A.3.2. The fourth equality again follows from Lemma A.3.1 (i) and the second last equality follows from Lemma A.3.1 (v). Since δ is arbitrary, we have established that $\liminf_{T\to\infty} \pi_T \geq \pi^*$.

We now prove Part (ii). Fix an arbitrary realization of the stochastic process $(\hat{\xi}_T)_{T\in\mathbb{N}}$ such that $\lim_{T\to\infty} \pi_T = \pi^*$. From Part (i), for all sufficiently large T, we have $\pi_T \leq v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \bar{\mathbb{P}}) \leq \pi^*$, $\bar{\mathbb{P}}^\infty-$almost surely. Now by the closedness of $[\boldsymbol{c}, \boldsymbol{s}]$, $(\boldsymbol{w}_{T,\theta_T(\alpha_T)})_{T\in\mathbb{N}}$

has an accumulation point $\bar{\boldsymbol{w}} \geq \boldsymbol{c}$. Thus, we have $\pi^* \geq v(\bar{\boldsymbol{w}}, \bar{\mathbb{P}}) = \lim_{T\to\infty} v(\boldsymbol{w}_{T,\theta_T(\alpha_T)}, \bar{\mathbb{P}}) \geq \lim_{T\to\infty} \pi_T = \pi^*$, where the first equality follows from the continuity of $v(\cdot, \bar{\mathbb{P}})$ (see Lemma A.3.1). We conclude that $v(\bar{\boldsymbol{w}}, \bar{\mathbb{P}}) = \pi^*$.

$\square$

A.3.6 Alternative Uncertainty Sets

In classical DRO, the center of the uncertainty set is usually the empirical distribution $\hat{\mathbb{P}}_T = \hat{\mathbb{P}}^{\mathrm{e}}_T$. We can center our uncertainty set at $\hat{\mathbb{P}}^{\mathrm{e}}_T$ to obtain $\mathcal{D}_\theta(\hat{\mathbb{P}}^{\mathrm{e}}_T) = \{\mathbb{P} \in \mathcal{D}^o_T : W_2(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_T) \leq \theta\}$, where we also restrict to distributions in $\mathcal{D}^o_T$. However, $\hat{\mathbb{P}}^{\mathrm{e}}_T$ may not be compatible with the retailer's past orders so that $\hat{\mathbb{P}}^{\mathrm{e}}_T \notin \mathcal{D}^o_T$. In addition, it may be that $\mathcal{D}_\theta(\hat{\mathbb{P}}^{\mathrm{e}}_T)$ is empty when θ is too small (i.e., when $\theta < W(\hat{\mathbb{P}}^{\mathrm{e}}_T, \hat{\mathbb{P}}_T)$). Furthermore, $\hat{\mathbb{P}}_T$ rather than $\hat{\mathbb{P}}^{\mathrm{e}}_T$ better represents our 'nominal' belief about the true demand distribution. When the empirical distribution $\hat{\mathbb{P}}^{\mathrm{e}}_T$ is not consistent with $\mathcal{D}^o_T$, we know it will not be the true demand distribution. In contrast, we chose the center $\hat{\mathbb{P}}_T$ specifically to be the best match to $\hat{\mathbb{P}}^{\mathrm{e}}_T$ within $\mathcal{D}^o_T$.

The literature has considered other types of uncertainty sets in addition to the Wasserstein distance, e.g., moment-based [e.g., 169]–[171] and divergence-based [e.g., 55], [172]–[174]. Moment-based uncertainty sets constructed using mean and variance information can lead to conservative solutions [170], and divergence-based uncertainty sets restrict the support of the random variable in question to the observed samples. The Wasserstein distance [e.g., 30], [31] does not have these drawbacks and possesses other favorable properties. For example, the Wasserstein distance is a metric on $\mathcal{P}(\Xi)$ and it metrizes weak convergence [64]. Thus we consider the Wasserstein distance to be the most suitable for our model.

A.4 Additional Material for Section 2.6 (single product case)

A.4.1 Proof of Proposition 2.6.1 (center for $n = 1$)

Proposition 2.6.1 follows from the proof of the upcoming more general Proposition 2.7.1 which covers all $n \geq 1$.

A.4.2 Details of Theorem 2.6.1 (worst-case profit for $n = 1$)

We present the details about the worst-case order quantities $q(w, \theta)$ and $q(w)$, and the worst-case distribution $\mathbb{P}_{w,\theta}$ in this subsection. We can express the center more compactly by ignoring zero terms. Let $(\hat{\xi}^k)_{k \in [K]}$ be the support of $\hat{\mathbb{P}}_T$, and let $\eta^k = \hat{\mathbb{P}}_T(\xi = \hat{\xi}^k)$ for all $k \in [K]$ (in particular, $(\eta^k)_{k \in [K]}$ correspond to all the strictly positive elements of $(\beta^*_{st})_{s \in [T+1], t \in [T]}$). Then, we can write $\hat{\mathbb{P}}_T = \sum_{k=1}^K \eta^k \delta_{\hat{\xi}^k}$.

Fix $w \in [c, s]$, let

$$\xi(w) = \arg \min_{x \in (\hat{\xi}^k)_{k=1}^K} \{x : \hat{\mathbb{P}}_T(\xi \le x) \ge 1 - w/s\} \tag{A.7}$$

be the minimum element in the support of $\hat{\mathbb{P}}_T$ where the value of the CDF at this element is no smaller than $1 - w/s$, and let

$$q(w) = \arg \max_{q \in (q^t)_{t \in [T]}} \left\{q : \bar{\mathbb{P}}(\xi \le q) < 1 - w/s\right\} \tag{A.8}$$

be the maximum past order quantity such that the value of the CDF of $\bar{\mathbb{P}}$ at this order size is smaller than $1 - w/s$.

We introduce the threshold

$$\bar{\theta}(w) \triangleq \sqrt{\sum_{\{k : q(w) < \hat{\xi}^k < \xi(w)\}} \eta^k \left(q(w) - \hat{\xi}^k\right)^2 + \left(1 - w/s - \sum_{\{k : \hat{\xi}^k < \xi(w)\}} \eta^k\right) (q(w) - \xi(w))^2} \tag{A.9}$$

on the Wasserstein distance θ. When $\theta \ge \bar{\theta}(w)$, the worst-case order is $q(w)$. When $\theta < \bar{\theta}(w)$, the worst-case order quantity $q(w, \theta)$ is the unique solution to

$$1 - w/s = \sum_{\{k : \hat{\xi}^k < \xi(w)\}} \eta^k + \frac{\theta^2 - \sum_{\{k : q < \hat{\xi}^k < \xi(w)\}} \eta^k \left(q - \hat{\xi}^k\right)^2}{(q - \xi(w))^2}. \tag{A.10}$$

Note that the left hand side of Eq. (A.10) is constant and the right hand side of Eq. (A.10) is an increasing function in $q \in [q(w), \xi(w)]$. Thus, we can solve for $q(w,\theta)$ efficiently with bisection search.

For $q \geq 0$, let $l(q) \triangleq \arg\min_{x \in \{q^t\}_{t\in[T]}} \{x : x \geq q\}$ be the smallest past order quantity that is larger than q. Given θ, let $z(q,\theta) \in (q, l(q)] \cap (\hat{\xi}^k)_{k=1}^K$ denote the smallest element of the set $(q, l(q)] \cap (\hat{\xi}^k)_{k=1}^K$ (if this set is nonempty) that satisfies $\theta^2 \leq \sum_{\{k: q<\hat{\xi}^k \leq z(q,\theta)\}} \eta^k (q - \hat{\xi}^k)^2$. That is, for $z(q,\theta)$, the transportation budget θ^2 is just enough to transport the probability mass from samples lying between q and $z(q,\theta)$ to q.

According to the definition of $z(q,\theta)$, there must exist some k^* such that $z(q,\theta) = \hat{\xi}^{k^*}$. Let η^{k^*} be the corresponding value of the PDF of $\hat{\mathbb{P}}_T$ at $\hat{\xi}^{k^*}$, i.e., $\eta^{k^*} \triangleq \hat{\mathbb{P}}_T(\xi = \hat{\xi}^{k^*})$. Also let

$$\gamma(q) \triangleq \frac{\theta^2 - \sum_{\{k: q<\hat{\xi}^k < z(q,\theta)\}} \eta^k \left(q - \hat{\xi}^k\right)^2}{(q - z(q,\theta))^2}. \tag{A.11}$$

Then, the worst-case distribution is

$$\mathbb{P}_{w,\theta} = \sum_{\{k: 0 \leq \hat{\xi}^k < q\}} \eta^k \delta_{\hat{\xi}^k} + \left(\sum_{\{k: q \leq \hat{\xi}^k < z(q,\theta)\}} \eta^k + \gamma(q) \right) \delta_q + (\eta^{k^*} - \gamma(q)) \delta_{z(q,\theta)} + \sum_{\{k: \hat{\xi}^k > z(q,\theta)\}} \eta^k \delta_{\hat{\xi}^k}$$
$$\text{for } q = q(w,\theta). \tag{A.12}$$

A.4.3 Proof of Theorem 2.6.1 (worst-case profit for $n = 1$)

For order quantity $q \geq 0$, we define

$$\rho(q) \triangleq \max_{\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)} \mathbb{P}(\xi \leq q) \tag{A.13}$$

to be the worst-case overstock probability corresponding to the uncertainty set $\mathcal{D}_\theta(\hat{\mathbb{P}}_T)$. With some abuse of notation, let $\mathbb{P}_{q,\theta}$ be the distribution that attains the maximum in Problem (A.13) (we will show this maximizer exists in the proof of the upcoming Lemma A.4.1).

Using the definition of $\rho(q)$, the supplier's worst-case profit for wholesale price w can be expressed as $\pi_{T,\theta}(w) = \inf_{q\geq 0}\{(w-c)q : \rho(q) \geq 1 - w/s\}$.

For any w and θ, when we take $q = q(w,\theta)$ (the worst-case order quantity for the supplier), $\mathbb{P}_{q(w,\theta),\theta}$ is the distribution that induces the retailer to order $q(w,\theta)$. Therefore, $\mathbb{P}_{q(w,\theta),\theta}$ is the worst-case distribution for any w and θ.

We will give closed form expressions for $\rho(q)$ and $\mathbb{P}_{q,\theta}$, which are useful in the derivation of $q(w)$, $q(w,\theta)$, and $\mathbb{P}_{w,\theta}$. Let $w(q) \triangleq s - s\,\bar{\mathbb{P}}(\xi \leq q)$ be the wholesale price that induces the retailer to order q under the true demand distribution $\bar{\mathbb{P}}$ by the retailer first-order conditions Eq. (2.3). In particular, $w(l(q))$ is the wholesale price that induces the retailer to order $l(q)$. Lemma A.4.1 gives closed form expressions for $\rho(q)$ and $\mathbb{P}_{q,\theta}$.

Lemma A.4.1. *(Proof in Appendix II, A.7.1) Recall $\gamma(q)$ defined in Eq. (A.11). We have*

(i)

$$\rho(q) = \begin{cases} 1 - w(l(q))/s, & \theta^2 \geq \sum_{\{k: q < \hat{\xi}^k \leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2, \\ \sum_{\{k:\hat{\xi}^k < z(q,\theta)\}} \eta^k + \gamma(q), & \theta^2 < \sum_{\{k: q < \hat{\xi}^k \leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2. \end{cases}$$

(ii)

$$\mathbb{P}_{q,\theta} = \begin{cases} \sum_{\{k:0\leq \hat{\xi}^k < q\}} \eta^k \delta_{\hat{\xi}^k} + \sum_{\{k:q\leq \hat{\xi}^k \leq l(q)\}} \eta^k \delta_q \\ \qquad + \sum_{\{k:\hat{\xi}^k > l(q)\}} \eta^k \delta_{\hat{\xi}^k}, & \theta^2 \geq \sum_{\{k: q < \hat{\xi}^k \leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2, \\ \sum_{\{k:0\leq \hat{\xi}^k < q\}} \eta^k \delta_{\hat{\xi}^k} + \left(\sum_{\{k:q\leq \hat{\xi}^k < z(q,\theta)\}} \eta^k + \gamma(q)\right) \delta_q \\ \qquad +(\eta^{k^*} - \gamma(q))\delta_{z(q,\theta)} + \sum_{\{k:\hat{\xi}^k > z(q,\theta)\}} \eta^k \delta_{\hat{\xi}^k}, & \theta^2 < \sum_{\{k: q < \hat{\xi}^k \leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2. \end{cases}$$

Proof of Theorem 2.6.1. First we show that the worst-case order quantity satisfies $q(w,\theta) \in (q(w), \xi(w)]$. Notice $\rho(\xi(w)) \geq \hat{\mathbb{P}}_T(\xi \leq \xi(w)) \geq 1 - w/s$, and thus $\pi_{T,\theta}(w) \leq (w-c)\xi(w)$ and $q(w,\theta) \leq \xi(w)$. On the other hand, we have $\rho(q(w)) < 1 - w/s$ since $\mathbb{P}(\xi \leq q(w)) = \bar{\mathbb{P}}(\xi \leq q(w)) < 1 - w/s$ holds for any $\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)$. Thus, we have $q(w,\theta) > q(w)$. In summary, we have shown that $\pi_{T,\theta}(w) = \inf_{q \in (q(w),\xi(w)]}\{(w-c)q : \rho(q) \geq 1 - w/s\}$.

Now define the sets

$$\mathcal{Q}_1(w,\theta) \triangleq \left\{ q : 1 - w/s \le 1 - w(l(q))/s, \theta^2 \ge \sum_{\{k: q < \hat{\xi}^k \le l(q)\}} \eta^k (q - \hat{\xi}^k)^2 \right\}$$

and

$$\mathcal{Q}_2(w,\theta) \triangleq \left\{ q : 1 - w/s \le \sum_{\{k: \hat{\xi}^k < z(q,\theta)\}} \eta^k + \frac{\theta^2 - \sum_{\{k: q < \hat{\xi}^k < z(q,\theta)\}} \eta^k \left(q - \hat{\xi}^k\right)^2}{(q - z(q,\theta))^2}, \right.$$
$$\left. \theta^2 \le \sum_{\{k: q < \hat{\xi}^k \le l(q)\}} \eta^k (q - \hat{\xi}^k)^2 \right\}.$$

By plugging in the expression for $\rho(q)$ into $\pi_{T,\theta}(w)$, we have

$$\pi_{T,\theta}(w) = \inf_{q \in (q(w), \xi(w)]} \{(w-c)q : q \in \mathcal{Q}_1(w,\theta) \cup \mathcal{Q}_2(w,\theta)\}. \tag{A.14}$$

When $\theta < \bar{\theta}(w)$, we show that the optimal solution for Problem (A.14) is equal to the optimal solution $\tilde{q}(w,\theta)$ of the following problem:

$$\min_{q \in [q(w), \xi(w)]} \left\{ (w-c)q : 1 - w/s \le \sum_{\{k: \hat{\xi}^k < \xi(w)\}} \eta^k + \frac{\theta^2 - \sum_{\{k: q < \hat{\xi}^k < \xi(w)\}} \eta^k (q - \hat{\xi}^k)^2}{(q - \xi(w))^2} \right\}. \tag{A.15}$$

To streamline notation, we define $f(q,z) \triangleq \sum_{\{k: \hat{\xi}^k < z\}} \eta^k + \frac{\theta^2 - \sum_{\{k: q < \hat{\xi}^k < z\}} \eta^k (q - \hat{\xi}^k)^2}{(q-z)^2}$ and $c(q,z) \triangleq \sum_{\{k: q < \hat{\xi}^k \le z\}} \eta^k (q - \hat{\xi}^k)^2$. For any fixed $q \in (q(w), \xi(w)]$, $l(q)$ is also fixed, and we can just use the shorthand $l \triangleq l(q)$. We have $l \ge \xi(q)$, and so $f(q,z)$ is strictly increasing on $(q(w), \xi(w)]$ and $c(q,l)$ is strictly decreasing on $(q(w), \xi(w)]$.

Lemma A.4.2. *(Proof in Appendix A.7.2) Suppose $\theta < \bar{\theta}(w)$.*

(i) The optimal solution $\tilde{q}(w,\theta)$ to Problem (A.15) satisfies

$$1 - w/s = \sum_{\{k: \hat{\xi}^k < \xi(w)\}} \eta^k + \frac{\theta^2 - \sum_{\{k: \tilde{q}(w,\theta) < \hat{\xi}^k < \xi(w)\}} \eta^k (\tilde{q}(w,\theta) - \hat{\xi}^k)^2}{(\tilde{q}(w,\theta) - \xi(w))^2}.$$

(ii) The optimal solution $\tilde{q}(w,\theta)$ to Problem (A.15) is also an optimal solution of Problem (A.14).

By Lemma A.4.2, $q(w,\theta)$ is the unique solution to $1 - w/s = \sum_{\{k:\hat{\xi}^k<\xi(w)\}} \eta^k + \frac{\theta^2 - \sum_{\{k:q(w,\theta)<\hat{\xi}^k<\xi(w)\}} \eta^k (q(w,\theta)-\hat{\xi}^k)^2}{(q(w,\theta)-\xi(w))^2}$ for $\theta < \bar{\theta}(w)$.

Now we show that $\pi_{T,\theta}(w) = (w-c)q(w)$ when $\theta \geq \bar{\theta}(w)$. Notice that for arbitrary θ, $\pi_{T,\theta}(w) \geq (w-c)q(w)$ since any $q \leq q(w)$ is not feasible for Problem (A.14). At the same time, when $\theta \geq \bar{\theta}(w)$, $q(w)+\delta$ is feasible for Problem (A.14) for arbitrary $\delta > 0$. It follows that $\pi_{T,\theta}(w) \leq (w-c)(q(w)+\delta)$ for any $\delta > 0$, and we conclude that $\pi_{T,\theta}(w) = (w-c)q(w)$. □

A.5 Additional Material for Section 2.7 (multi-product case)

A.5.1 Proof of Proposition 2.7.1 (center for $n \geq 1$)

We show how to reformulate Problem (2.5) as a tractable finite-dimensional optimization problem, and then solve for $\hat{\mathbb{P}}_T$. This subsection covers both dependent and independent demand. WLOG, for each $i \in [n]$ we can arrange the past wholesale prices in descending order as $s_i \geq w_i^1 > \cdots > w_i^t > w_i^{t+1} > \cdots > w_i^T$ (where we assume there are no repeats), and we can arrange the retailer's corresponding past order quantities in ascending order as $0 \leq q_i^1 < \cdots < q_i^t < q_i^{t+1} < \cdots < q_i^T$. We set $q_i^0 \equiv 0$ and $q_i^{T+1} \equiv \bar{\xi}$ for all $i \in [n]$.

The Lagrangian dual to Problem (2.5) is:

$$\sup_{\boldsymbol{\mu},\boldsymbol{\gamma}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \gamma_t : \|\boldsymbol{\xi} - \boldsymbol{\xi}^t\|_2^2 + \sum_{i\in[n]}\sum_{t'=1}^{T} \mu_i^{t'} \left(\mathbb{1}(\xi_i \leq q_i^{t'}) - \frac{s_i - w_i^{t'}}{s_i} \right) \geq \gamma_t, \forall \boldsymbol{\xi} \in \Xi, t \in [T] \right\}, \tag{A.16}$$

where $\boldsymbol{\mu} \triangleq (\mu_i^t)_{i\in[n],t\in[T]}$ and $\boldsymbol{\gamma} \triangleq (\gamma_t)_{t=1}^T$ are the dual variables (we give the formal derivation of Problem (A.16) in the proof of Proposition A.5.1). Proposition A.5.1 establishes the existence of the optimal solution $\hat{\mathbb{P}}_T$ to Problem (2.5), and also demonstrates that strong duality holds between Problem (2.5) and its dual Problem (A.16).

Proposition A.5.1. *(Proof in Appendix A.7.3) (i) The minimizer $\hat{\mathbb{P}}_T$ in Problem (2.5) is attained. (ii) The optimal values of Problems (2.5) and (A.16) are equal.*

Problem (A.16) has a semi-infinite constraint indexed by $\boldsymbol{\xi} \in \Xi$. We can characterize this constraint more precisely as follows. The past orders $(\boldsymbol{q}^t)_{t=1}^T$ partition Ξ into $(T+1)^n$ squares $R_s \triangleq \{\boldsymbol{\xi} : \xi_i \in (q_i^{t_i-1}, q_i^{t_i}], \forall i \in [n]\}$ indexed by $s \in [(T+1)^n]$ where $(t_i)_{i=1}^n \in \{1, \dots, T+1\}^n$. Then, for some $\boldsymbol{\xi} \in R_s$, $\mathbb{1}(\xi_i \leq q_i^{t'}) = 0$ if $q_i^{t_i-1} \geq q_i^{t'}$ and $\mathbb{1}(\xi_i \leq q_i^{t'}) = 1$ if $q_i^{t_i} \leq q_i^{t'}$. In other words, $\mathbb{1}(\xi_i \leq q_i^{t'})$ is constant for $\boldsymbol{\xi} \in R_s$. We define $z_{is}(\boldsymbol{q}^{t'}) \triangleq \mathbb{1}(\xi_i \leq q_i^{t'} | \boldsymbol{\xi} \in R_s)$, and we represent R_s as $R_s = \{\boldsymbol{\xi} : A_s \boldsymbol{\xi} \leq \boldsymbol{d}_s\}$ where A_s is a matrix and $\boldsymbol{d}_s$ is a vector of appropriate dimension. Problem (A.16) can then be reformulated as:

$$
\begin{aligned}
\sup_{\boldsymbol{\mu}, \boldsymbol{\gamma}} \quad & \frac{1}{T} \sum_{t=1}^{T} \gamma_t, \\
\text{s.t.} \quad & \|\boldsymbol{\xi} - \boldsymbol{\xi}^t\|_2^2 + \sum_{i \in [n]} \sum_{t'=1}^{T} \mu_i^{t'} \left(z_{is}(\boldsymbol{q}^{t'}) - \frac{s_i - w_i^{t'}}{s_i} \right) \geq \gamma_t, \forall \boldsymbol{\xi} : A_s \boldsymbol{\xi} \leq \boldsymbol{d}_s, \forall t \in [T], \forall s \in [S],
\end{aligned}
\tag{A.17}
$$

where the constraints are grouped by subsets of the partition. The Lagrangian dual of Problem (A.17) is Problem (2.17), and the optimal solutions of Problem (2.17) are denoted $(\beta_{st}^*)_{s \in [S], t \in [T]}$ and $(\boldsymbol{p}_{st}^*)_{s \in [S], t \in [T]}$.

Proposition A.5.2. *(Proof in Appendix II, A.7.4) The optimal values of Problems (2.17) and (A.17) are equal.*

We now prove that $\hat{\mathbb{P}}_T$ defined by

$$
\hat{\mathbb{P}}_T = \sum_{t=1}^{T} \sum_{s=1}^{S} \beta_{st}^* \delta_{\boldsymbol{p}_{st}^* / \beta_{st}^*} \tag{A.18}
$$

is feasible and optimal for Problem (2.5). By Eq. (2.17d), we have $\boldsymbol{p}^*_{st}/\beta^*_{st} \in \Xi$. By Eq. (2.17b), $\hat{\mathbb{P}}_T = \sum_{t=1}^{T}\sum_{s=1}^{S} \beta^*_{st}\delta_{\boldsymbol{p}^*_{st}/\beta^*_{st}}$ is a valid probability distribution supported on Ξ. Furthermore, by Eq. (2.17c), $\hat{\mathbb{P}}_T$ satisfies the retailer's first-order conditions since

$$
\begin{aligned}
\hat{\mathbb{P}}_T\left(\xi_i \le q_i^{t'}\right) &= \sum_{s=1}^{S} \hat{\mathbb{P}}_T\left(\xi_i \le q_i^{t'} | \boldsymbol{\xi} \in R_s\right)\hat{\mathbb{P}}_T\left(\boldsymbol{\xi} \in R_s\right) \\
&= \sum_{s=1}^{S} \mathbb{E}_{\hat{\mathbb{P}}_T}\left[z_{is(\boldsymbol{q}^{t'})}\right]\hat{\mathbb{P}}_T(\boldsymbol{\xi} \in R_s) \\
&= \sum_{s=1}^{S}\sum_{t=1}^{T} \beta^*_{st} z_{is}(\boldsymbol{q}^{t'}) \\
&= \frac{s_i - w_i^{t'}}{s_i}.
\end{aligned}
$$

The first equality follows from the law of total probability since $(R_s)_{s=1}^S$ is a partition with empty pair-wise intersection, the second equality follows from the definition of $z_{is}(\boldsymbol{q}^{t'})$, the third equality follows because $z_{is}(\boldsymbol{q}^{t'})$ is constant and by the definition of $\hat{\mathbb{P}}_T$, and the last equality follows by Eq. (2.17c). Thus, we see that $\hat{\mathbb{P}}_T \in \mathcal{D}_T^o$ and so $\hat{\mathbb{P}}_T$ is feasible for Problem (2.5). Furthermore, by the definition of Wasserstein distance,

$$
W_2^2(\hat{\mathbb{P}}_T, \hat{\mathbb{P}}_T^{\mathrm{e}}) = \min_{(\lambda_{s,t,t'})\ge 0} \left\{ \sum_{t'=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{S} \lambda_{s,t,t'} \left\| \boldsymbol{p}^*_{st}/\beta^*_{st} - \boldsymbol{\xi}^{t'} \right\|_2^2 : \sum_{t'=1}^{T} \lambda_{s,t,t'} = \beta^*_{st}, \sum_{t=1}^{T}\sum_{s=1}^{S} \lambda_{s,t,t'} = 1/T \right\}.
$$

Let $\lambda^*_{s,t,t'} = 0$ if $t' \neq t$, and $\lambda^*_{s,t,t'} = \beta^*_{st}$ if $t' = t$. Then $(\lambda^*_{s,t,t'})_{s\in[S],t\in[T],t'\in[T]}$ is feasible to the above problem. Thus, we have $W_2^2(\hat{\mathbb{P}}_T, \hat{\mathbb{P}}_T^{\mathrm{e}}) \le \sum_{t'=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{S} \lambda^*_{s,t,t'} \left\| \boldsymbol{p}^*_{st}/\beta^*_{st} - \boldsymbol{\xi}^{t'} \right\|_2^2 = \sum_{t=1}^{T}\sum_{s=1}^{S} \beta^*_{st} \|\boldsymbol{p}^*_{st}/\beta^*_{st} - \boldsymbol{\xi}^t\|_2^2$.

At the same time, according to Propositions A.5.1 and A.5.2, the optimal values of Problem (2.5) and Problem (2.17) are equal. Thus, we have

$$
\inf_{\mathbb{P}\in\mathcal{D}_T^o} W_2^2(\mathbb{P}, \hat{\mathbb{P}}_T^{\mathrm{e}}) = \sum_{t=1}^{T}\sum_{s=1}^{S} \beta^*_{st} \left(\boldsymbol{p}^*_{st}/\beta^*_{st} - \boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{p}^*_{st}/\beta^*_{st} - \boldsymbol{\xi}^t\right).
$$

So, by the first part of this proposition, $\hat{\mathbb{P}}_T$ achieves the minimum in Problem (2.5) and so is an optimal solution to Problem (2.5).

A.5.2 Solving for Retailer's Worst-case Order Quantities (dependent demand)

We now show how to compute the worst-case distribution and worst-case order quantity for an arbitrary demand distribution $\mathbb{P}$. Specifically, we develop a cutting plane method for Problem (2.8). Given any $\boldsymbol{w}$, we can check if $\boldsymbol{q} \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ by checking feasibility of the following system of linear inequalities in $\mathbb{P}$:

$$\mathbb{P} \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T),\ \mathbb{P}^i(\xi_i \leq q_i) \geq \frac{s_i - w_i}{s_i},\ \forall i \in [n]. \tag{A.19}$$

We have $\boldsymbol{q} \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ if and only if Eq. (A.19) is satisfied. Eq. (A.19) consists of $nT+n+1$ linear inequalities/equalities in $\mathbb{P}$ (including the inequalities/equalities corresponding to $\mathbb{P} \in \mathcal{D}_T^o$ and $W_2(\mathbb{P}, \hat{\mathbb{P}}_T) \leq \theta$). In order to find a $\delta-$optimal solution for Problem (2.8), we can construct an ϵ-net $\hat{\Xi}$ of $[0, \bar{\xi}]^n$ for $\epsilon = \delta / \max_{i\in[n]}\{s_i - w_i\}$. Then, for any $\boldsymbol{\xi} \in \Xi$, there exists $\boldsymbol{\xi}' \in \hat{\Xi}$ with $\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_1 \leq \epsilon$. Since the worst-case order satisfies $q(\boldsymbol{w}, \theta) \in \Xi$, there must exist $\boldsymbol{q} \in \hat{\Xi}$ such that $\|\boldsymbol{q} - \boldsymbol{q}(\boldsymbol{w}, \theta)\|_1 \leq \epsilon$, and thus the worst-case profit satisfies $(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}(\boldsymbol{w}, \theta) \geq (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} - \delta$. It follows that there must exist a $\delta-$optimal solution in $\hat{\Xi}$ for Problem (2.8). We can check optimality of each $\boldsymbol{q} \in \hat{\Xi}$ which requires solving $|\hat{\Xi}|$ linear feasibility problems to identify a $\delta-$optimal solution. After constructing $\mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$, the supplier can compare the worst-case profit for all $\boldsymbol{q} \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$.

However, checking all elements in $\hat{\Xi}$ is computationally expensive when $|\hat{\Xi}|$ is large. To accelerate this process, we show that checking optimality of any order quantity $\boldsymbol{q} \in \Xi$ yields information about other potentially optimal order quantities. On one hand, if Eq. (A.19) is feasible, then we have identified a particular distribution $\tilde{\mathbb{P}}$ as a feasible solution to Eq. (A.19) and $\boldsymbol{q} \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$, and we must have $\pi_{T,\theta}(\boldsymbol{w}) \leq (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}$. Consequently, we can discard all order quantities in the set $\mathcal{X}(\boldsymbol{q}) \triangleq \left\{\boldsymbol{q}' \in \Xi : (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}' \geq (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}\right\}$ that yield higher profit for the supplier (since we are looking for the worst-case profit). On the other hand, if Eq. (A.19) is infeasible, we let $\mathcal{X}_\leq(\boldsymbol{q}) \triangleq \{\boldsymbol{q}' \in \Xi : \boldsymbol{q}' \leq \boldsymbol{q}\}$ be the set of orders that are component-wise smaller than $\boldsymbol{q}$.

Lemma A.5.1. *If Eq. (A.19) is infeasible, then* $\mathcal{X}_\leq(\boldsymbol{q}) \cap \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T) = \emptyset$.

Proof of Lemma A.5.1. Suppose there exists some $\boldsymbol{q}' \in \mathcal{X}_{\leq}(\boldsymbol{q})$ such that $\boldsymbol{q}' \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$. Then, there must exist some $\mathbb{P}' \in \mathcal{D}_\theta(\hat{\mathbb{P}}_T)$ such that $\boldsymbol{q}'$ satisfies Eq. (A.19) for $\mathbb{P} = \mathbb{P}'$. Since $\boldsymbol{q} \geq \boldsymbol{q}'$, we must have $\mathbb{P}^i(\xi_i \leq q_i) \geq \mathbb{P}^i(\xi_i \leq q_i') \geq \frac{s_i - w_i}{s_i}$ for all $i \in [n]$, a contradiction to the infeasibility of Eq. (A.19). □

To summarize this discussion, each time we check the inclusion $\boldsymbol{q} \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$, we either: (i) conclude that $\boldsymbol{q} \in \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ and discard a set of sub-optimal order quantities; or (ii) we conclude $\boldsymbol{q} \notin \mathcal{Q}_\theta(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ and we use this information to discard an alternative set of sub-optimal order quantities. We also notice that Eq. (A.19) is a linear feasibility problem involving the infinite-dimensional variable $\mathbb{P}$. However, we can reformulate it as a finite-dimensional nonlinear convex programming problem. The detailed procedure is summarized in Algorithm 4. For any two sets $\mathcal{X}, \mathcal{Y}$, we let $\mathcal{X} - \mathcal{Y} \triangleq \{x : x \in \mathcal{X} \text{ and } x \notin \mathcal{Y}\}$.

Algorithm 4 Demand correlation

Initialize: $\hat{\mathcal{O}} \leftarrow \hat{\Xi}$ contains the potential ϵ optimal solutions to Problem (2.8). Find the order $\boldsymbol{q}(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ corresponding to the center $\hat{\mathbb{P}}_T$; $\hat{\mathcal{O}} \leftarrow \hat{\mathcal{O}} - \mathcal{X}\left(\boldsymbol{q}(\boldsymbol{w}; \hat{\mathbb{P}}_T)\right)$. $\hat{\boldsymbol{q}} \leftarrow \boldsymbol{q}(\boldsymbol{w}; \hat{\mathbb{P}}_T)$ is the updated solution to Problem (2.8), and $\hat{\mathbb{P}} \leftarrow \hat{\mathbb{P}}_T$ is updated solution to Problem (2.6). Set upper bound: $\hat{v} \leftarrow (\boldsymbol{w} - \boldsymbol{c})^\top \hat{\boldsymbol{q}}$. Set lower bound $\underline{v} \leftarrow \min\{(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \boldsymbol{q} \in \hat{\mathcal{O}}\}$. Choose precision $\varepsilon > 0$.

while $\underline{v} - \hat{v} > \varepsilon$ **do**

  Take some $\boldsymbol{q} \in \hat{\mathcal{O}}$ and check feasibility of Eq. (A.19) under $\boldsymbol{q}$

  if (A.19) is infeasible **then**

    $\hat{\mathcal{O}} \leftarrow \hat{\mathcal{O}} - \mathcal{X}_{\leq}(\boldsymbol{q})$

  else

    Let $\tilde{\mathbb{P}}$ be a feasible distribution to Eq. (A.19)

    $\hat{\mathcal{O}} \leftarrow \hat{\mathcal{O}} - \mathcal{X}(\boldsymbol{q})$

    $\hat{v} \leftarrow (\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q}, \hat{\boldsymbol{q}} \leftarrow \boldsymbol{q}, \hat{\mathbb{P}} \leftarrow \tilde{\mathbb{P}}$

    Update $\underline{v} \leftarrow \min\{(\boldsymbol{w} - \boldsymbol{c})^\top \boldsymbol{q} : \boldsymbol{q} \in \hat{\mathcal{O}}\}$

Ouput: $\hat{v}, \hat{\boldsymbol{q}}$ as the optimal value and solution to Problem (2.8) respectively, and $\hat{\mathbb{P}}$ as the optimal solution to Problem (2.6)

Now we reformulate Eq. (A.19) in a more tractable form. We take the center $\hat{\mathbb{P}}_T$ as input, and let $(\hat{\boldsymbol{\xi}}^k)_{k\in[K]}$ be the support of $\hat{\mathbb{P}}_T$ and $\eta^k = \hat{\mathbb{P}}_T(\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^k)$ for all $k \in [K]$ (in particular, $(\eta^k)_{k\in[K]}$ correspond to all the strictly positive elements of $(\beta^*_{st})_{s\in[S],t\in[T]}$). Then, we have the equivalent expression $\hat{\mathbb{P}}_T = \sum_{k=1}^K \eta^k \delta_{\hat{\boldsymbol{\xi}}^k}$ for the center. Feasibility of Eq. (A.19) can then be checked by solving:

$$\min_{\mathbb{P}\in\mathcal{D}_T^o} \left\{ W_2^2(\mathbb{P}, \hat{\mathbb{P}}_T) : \mathbb{P}(\xi \leq q_i) \geq \frac{s_i - w_i}{s_i},\ \forall i \in [n] \right\}. \tag{A.20}$$

Eq. (A.19) is feasible if and only if the optimal value of Problem (A.20) is not larger than θ^2. Furthermore, Problem (A.20) can be reformulated in the same way as Problem (2.5), we omit the details here.

A.5.3 Proof of Proposition 2.7.2 (decomposition of projection problem)

Recall $\hat{\mathbb{Q}}_T$ is an optimal solution to Problem (2.19) and $\hat{\mathbb{Q}}_{T,i}$ is an optimal solution to Problem (2.20). In order to prove Proposition 2.7.2, it suffices to prove Lemma A.5.2 as below. Lemma A.5.2 shows that the center $\hat{\mathbb{Q}}_T$ obtained from Problem (2.19) and the center $\times_{i=1}^n \hat{\mathbb{Q}}_{T,i}$ obtained from Problem (2.20) are indeed equivalent.

Lemma A.5.2. *(i) We have* $W_2^2(\hat{\mathbb{Q}}_T, \hat{\mathbb{Q}}_T^{\mathrm{e}}) = \sum_{i\in[n]} W_2^2(\hat{\mathbb{Q}}_{T,i}, \hat{\mathbb{P}}_T^{\mathrm{e},i}) = W_2^2(\times_{i\in[n]}\hat{\mathbb{Q}}_{T,i}, \hat{\mathbb{Q}}_T^{\mathrm{e}})$*.*

(ii) For each $i \in [n]$*, the* i*-th marginal distribution of* $\hat{\mathbb{Q}}_T$*, denoted* $\hat{\mathbb{Q}}_T^i$*, solves Problem* (2.20)*.*

(iii) $\times_{i\in[n]}\hat{\mathbb{Q}}_{T,i}$ *is an optimal solution to Problem* (2.19)*.*

We need the following auxiliary result (Lemma A.5.3) to prove Lemma A.5.2. This result relates the Wasserstein distance between two probability distributions with independent marginals to the sum of the Wasserstein distances between their marginals.

Lemma A.5.3. *([175, Lemma 1.2]) Suppose* $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{D}_\times$ *have marginal distributions* $\{\mathbb{P}_1^i\}_{i\in[n]}$ *and* $\{\mathbb{P}_2^i\}_{i\in[n]}$ *respectively, then* $W_2^2(\mathbb{P}_1, \mathbb{P}_2) = \sum_{i\in[n]} W_2^2(\mathbb{P}_1^i, \mathbb{P}_2^i)$*.*

Proof of Lemma A.5.2. For Part (i), notice that $\times_{i=1}^n \hat{\mathbb{Q}}_{T,i}$ is a feasible solution to Problem (2.19), and thus we have $W_2^2(\hat{\mathbb{Q}}_T, \hat{\mathbb{Q}}_T^{\mathrm{e}}) \leq W_2^2(\times_{i=1}^n \hat{\mathbb{Q}}_{T,i}, \hat{\mathbb{Q}}_T^{\mathrm{e}}) = \sum_{i\in[n]} W_2^2(\hat{\mathbb{Q}}_{T,i}, \hat{\mathbb{P}}_T^{\mathrm{e},i})$, where $\hat{\mathbb{Q}}_T$ is optimal to Problem (2.19), and the second equality follows from Lemma A.5.3 and the fact that $\hat{\mathbb{Q}}_T^{\mathrm{e}} \in \mathcal{D}_\times$.

On the other hand, the marginal $\hat{\mathbb{Q}}_T^i$ is feasible to Problem (2.20) for all $i \in [n]$, and thus we have $W_2^2(\hat{\mathbb{Q}}_T, \hat{\mathbb{Q}}_T^{\mathrm{e}}) = \sum_{i\in[n]} W_2^2(\hat{\mathbb{Q}}_T^i, \hat{\mathbb{P}}_T^{\mathrm{e},i}) \geq \sum_{i\in[n]} W_2^2(\hat{\mathbb{Q}}_{T,i}, \hat{\mathbb{P}}_T^{\mathrm{e},i}) = W_2^2(\times_{i=1}^n \hat{\mathbb{Q}}_{T,i}, \hat{\mathbb{Q}}_T^{\mathrm{e}})$, where the first and last equalities follow from Lemma A.5.3 and the fact that $\hat{\mathbb{Q}}_T$, $\times_{i=1}^n \hat{\mathbb{Q}}_{T,i}$, $\hat{\mathbb{Q}}_T^{\mathrm{e}} \in \mathcal{D}_\times$. The inequality follows since, for each $i \in [n]$, $\hat{\mathbb{Q}}_T^i$ is feasible to Problem (2.20) while $\hat{\mathbb{Q}}_{T,i}$ is optimal to Problem (2.20). The conclusion of Part (i) follows.

Parts (ii) and (iii) then follow directly from Part (i).

$\square$

A.5.4 Proof of Proposition 2.7.3 (convexity of worst-case profit)

The argument is the same for all products so we can fix $i \in [n]$. It is immediate that $\pi^i_{T,\sqrt{r_i}}(w_i)$ (the worst-case profit function for the single product i alone), is decreasing in r_i. For $r_i \geq \bar{\theta}_i^2(w_i)$, $\pi^i_{T,\sqrt{r_i}}(w_i)$ is a constant, so we only need to show that $\pi^i_{T,\sqrt{r_i}}(w_i)$ is convex in r_i for $r_i \leq \bar{\theta}_i^2(w_i)$.

Define the function $f_i : [q_i(w_i), \xi_i(w_i)] \to [0, \bar{\theta}_i^2(w_i)]$ by

$$f_i(q_i) \triangleq \left(1 - w_i/s_i - \sum_{\{k:\hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k\right)(q_i - \xi_i(w_i))^2 + \sum_{\{k:q_i<\hat{\xi}_i^k<\xi_i(w_i)\}} \eta_i^k (q_i - \hat{\xi}_i^k)^2. \quad \text{(A.21)}$$

We have $f_i(q_i(w_i)) = \bar{\theta}_i^2(w)$ and $f_i(\xi_i(w_i)) = 0$. Since $f_i(q_i)$ is convex and strictly decreasing in $[q_i(w_i), \xi_i(w_i)]$, the inverse map $f_i^{-1} : [0, \bar{\theta}_i^2(w_i)] \to [q_i(w_i), \xi_i(w_i)]$ is well defined and is also convex on $r_i \in [0, \bar{\theta}_i^2(w_i)]$ [176]. Thus, the worst-case profit $\pi^i_{T,\sqrt{r_i}}(w_i) = (w_i - c_i) f_i^{-1}(r_i)$ is also convex in r_i.

A.5.5 Proof of Theorem 2.7.2 (supplier's worst-case distribution and order quantity)

Part (i) follows since Problem (2.22) is a convex optimization problem over compact support.

For Part (ii), we have $\pi_{T,\theta}(\boldsymbol{w}) = \sum_{i=1}^n \pi^i_{T,\sqrt{r_i^*}}(w_i) = \sum_{i=1}^n (w_i - c_i) q^i(w_i, \sqrt{r_i^*})$, where the first equality follows from Part (i) and Problem (2.22), and the second equality follows from Theorem 2.6.1.

To show $\mathbb{P}_{\boldsymbol{w},\theta} = \times_{i\in[n]} \mathbb{P}^i_{w_i,\sqrt{r_i^*}}$ is the worst-case distribution, we first show that $\mathbb{P}_{\boldsymbol{w},\theta} \in \mathcal{D}_{\theta,\times}(\hat{\mathbb{Q}}_T)$ and then show that it attains the worst-case profit in Problem (2.9). First notice that $\mathbb{P}_{\boldsymbol{w},\theta} \in \mathcal{D}_\times$. We also have $W_2^2(\mathbb{P}_{\boldsymbol{w},\theta}, \hat{\mathbb{Q}}_T) = W_2^2(\times_{i\in[n]} \mathbb{P}^i_{w_i,\sqrt{r_i^*}}, \hat{\mathbb{Q}}_T)$ $= \sum_{i=1}^n W_2^2(\mathbb{P}^i_{w_i,\sqrt{r_i^*}}, \hat{\mathbb{Q}}_T) \leq \sum_{i=1}^n r_i^* \leq \theta^2$, where the first equality follows from the definition of $\mathbb{P}_{\boldsymbol{w},\theta}$, the second equality follows from Lemma A.5.3, the first inequality follows from the definition of $\mathbb{P}^i_{w_i,\sqrt{r_i^*}}$ and the second inequality follows since $(r_i^*)_{i=1}^n$ is the optimal solution to Problem (2.22). In addition, $\mathbb{P}_{\boldsymbol{w},\theta} \in \mathcal{D}_T^o$ since $\mathbb{P}^i_{w_i,\sqrt{r_i^*}}(\xi_i \leq q_i^t) = 1 - w_i^t/s_i, \forall t \in [T]$. In summary we have $\mathbb{P}_{\boldsymbol{w},\theta} \in \mathcal{D}_{\theta,\times}(\hat{\mathbb{Q}}_T)$. The distribution $\mathbb{P}_{\boldsymbol{w},\theta}$ attains the worst-case profit because $(\mathbb{P}_{\boldsymbol{w},\theta}, (q^i(w_i, \sqrt{r_i^*}))_{i=1}^n)$ is feasible to Eq. (2.9) and attains the worst-case profit.

A.5.6 Cutting Plane Algorithm for Supplier's Worst-case Profit (independent demand)

We discuss how to solve Problem (2.22) and compute $\pi_{T,\theta}(\boldsymbol{w})$ for independent demand. First we characterize cuts for the objective of Problem (2.22), then we give the details for our cutting plane algorithm.

Proposition A.5.3. *Let $\tilde{\boldsymbol{r}} = (\tilde{r}_i)_{i\in[n]}$ be feasible for Problem (2.22) and satisfy $\tilde{r}_i \leq \bar{\theta}_i^2(w_i)$ for all $i \in [n]$. Let $q_i(w_i, \tilde{r}_i)$ be an optimal solution to Problem (2.21) for $\tilde{r}_i$ for each $i \in [n]$, and let $\boldsymbol{q}(\boldsymbol{w}, \tilde{\boldsymbol{r}}) = (q_i(w_i, \tilde{r}_i))_{i\in[n]}$ given $\tilde{\boldsymbol{r}}$. For each $i \in [n]$, let*

$$g_i = (w_i - c_i) \Big/ \Bigg\{ 2\Big(1 - w_i/s_i - \sum_{\{k:\hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k\Big)(q_i(w_i,\tilde{r}_i) - \xi_i(w_i)) + 2 \sum_{\{k: q_i(w_i,\tilde{r}_i) < \hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k (q_i(w_i,\tilde{r}_i) - \hat{\xi}_i^k) \Bigg\},$$

and let $\boldsymbol{g} = (g_i)_{i\in[n]}$. Then, for any $\boldsymbol{r}$ feasible to Problem (2.22) satisfying $\boldsymbol{g}^\top(\boldsymbol{r} - \tilde{\boldsymbol{r}}) \geq 0$, we have $\sum_{i\in[n]} \pi^i_{T,\sqrt{r_i}}(w_i) \geq \sum_{i\in[n]} \pi^i_{T,\sqrt{\tilde{r}_i}}(w_i)$.

Proof of Proposition A.5.3. To show that $\boldsymbol{g}$ is a valid objective cut at $\tilde{\boldsymbol{r}}$, it suffices to show that $g_i/(w_i - c_i)$ is a subgradient for $f_i^{-1}(r)$ (see Eq. (A.21)) at $r = \tilde{r}_i$ for all $i \in [n]$. To verify this claim, it is sufficient to show that

$$(w_i - c_i)/g_i = 2\left(1 - w_i/s_i - \sum_{\{k:\hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k\right)(q_i(w_i, \tilde{r}_i) - \xi_i(w_i)) \\ + 2 \sum_{\{k: q_i(w_i,\tilde{r}_i) < \hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k (q_i(w_i, \tilde{r}_i) - \hat{\xi}_i^k)$$

is a subgradient of $f_i(q)$ at $q = q_i(w_i, \tilde{r}_i)$. In particular, for any $r \leq \bar{\theta}_i^2(w_i)$, we can find $q \in [q_i(w_i), \xi_i(w_i)]$ such that $f_i(q) = r$. Also notice $f_i(q_i(w_i, \tilde{r}_i)) = \tilde{r}_i$. Then, if $(w_i - c_i)/g_i$ is a subgradient of $f_i(q)$ at $q = q_i(w_i, \tilde{r}_i)$, we have $f_i(q) \geq f_i\left(q_i(w_i, \tilde{r}_i)\right) + \left(q - q_i(w_i, \tilde{r}_i)\right)(w_i - c_i)/g_i$. Now plug $r = f_i(q)$ and $\tilde{r}_i = f_i\left(q_i(w_i, \tilde{r}_i)\right)$ into the above expression, and rearrange to get $r \geq \tilde{r}_i + g_i/(w_i - c_i)(r - \tilde{r}_i)$, which implies that $g_i/(w_i - c_i)$ is a subgradient of $f_i^{-1}(r)$ at $r = \tilde{r}_i$.

In the following, we show that $(w_i - c_i)/g_i$ is a subgradient of $f_i(q)$ at $q = q_i(w_i, \tilde{r}_i)$. Given arbitrary $q \in [q_i(w_i), \xi_i(w_i)]$, if $q \leq q_i(w_i, \tilde{r}_i)$, then

$$\begin{aligned} & f_i(q) - f_i(q_i(w_i, \tilde{r}_i)) - (q - q_i(w_i, \tilde{r}_i))(w_i - c_i)/g_i \\ = & \left(1 - w_i/s_i - \sum_{\{k:\hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k\right)(q - q_i(w_i, \tilde{r}_i))^2 + \sum_{\{k: q < \hat{\xi}_i^k \leq q_i(w_i,\tilde{r}_i)\}} \eta_i^k (q - \hat{\xi}_i^k)^2 \\ & + \sum_{\{k: q_i(w_i,\tilde{r}_i) < \hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k (q_i(w_i, \tilde{r}_i) - q_i(w_i, \tilde{r}_i))^2 \geq 0, \end{aligned}$$

and if $q \geq q_i(w_i, \tilde{r}_i)$, then

$$\begin{aligned}
&f_i(q) - f_i(q_i(w_i, \tilde{r}_i)) - (q - q_i(w_i, \tilde{r}_i))(w_i - c_i)/g_i \\
&\qquad = \left(1 - w_i/s_i - \sum_{\{k:\hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k\right)(q - q_i(w_i, \tilde{r}_i))^2 \\
&\qquad - \sum_{\{k:q_i(w_i,\tilde{r}_i) < \hat{\xi}_i^k \leq q\}} \eta^k (q_i(w_i, \tilde{r}_i) - \hat{\xi}_i^k)(2q - q_i(w_i, \tilde{r}_i) - \hat{\xi}_i^k) \\
&\qquad\qquad + \sum_{\{k:q < \hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k (q_i(w_i, \tilde{r}_i) - q_i(w_i, \tilde{r}_i))^2 \geq 0.
\end{aligned}$$

We have thus established that $q \geq f_i(q_i(w_i, \tilde{r}_i)) + (w_i - c_i)(q - q_i(w_i, \tilde{r}_i))/g_i$ for all $q \in [q_i(w_i), \xi_i(w_i)]$, for all $i \in [n]$. This establishes that $\boldsymbol{g}$ is a valid cut for Problem (2.22).

□

We now develop a cutting plane algorithm based on Proposition A.5.3. Similar to the single product case, we let $\xi_i(w_i) \in \arg\min_{x \in (\hat{\xi}_i^k)_{k=1}^K} \left\{x : \hat{\mathbb{P}}_T^i(\xi_i \leq x) \geq 1 - w_i/s_i\right\}$ be the smallest element in the support of $\hat{\mathbb{P}}_T^i$ where the value of the CDF is larger than $1 - w_i/s_i$, and we let $q_i(w_i) \in \arg\max_{q_i \in (q_i^t)_{t \in [T]}} \left\{q_i : \bar{\mathbb{P}}^i(\xi_i \leq q_i) < 1 - w_i/s_i\right\}$ be the largest past order quantity that corresponds to a wholesale price greater than w_i. Now let $\bar{\theta}_i(w_i) \triangleq \sqrt{\sum_{\{k:q_i(w_i) < \hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k \left(q_i(w_i) - \hat{\xi}_i^k\right)^2 + \left(1 - w_i/s_i - \sum_{\{k:\hat{\xi}_i^k < \xi_i(w_i)\}} \eta_i^k\right)(q_i(w_i) - \xi_i(w_i))^2}$ be a threshold for product i (which depends on $\hat{\mathbb{P}}_T^i$, w_i, and s_i for this product).

Proposition A.5.3 gives a valid cut for $\boldsymbol{r}$ with $r_i \leq \bar{\theta}_i^2(w_i)$ for all $i \in [n]$, so based on Proposition A.5.3 we can use the cutting plane method to solve Problem (2.22) (see Algorithm 5). In Problem (2.22), the feasible region is given explicitly by a set of simple linear inequalities, so the objective is the computational bottleneck. Our cutting plane algorithm is designed to cut off sub-optimal points.

Algorithm 5 An Analytic Center Cutting Plane Algorithm

Require: A tolerance $\varepsilon > 0$

Initialize: $r_i = 0$; $\mathcal{C}^{(0)} \triangleq \{\boldsymbol{r} : 0 \le r_i \le \bar{\theta}_i^2(w_i), \sum_{i\in[n]} r_i \le \theta^2\}$ is the set of valid cuts

Set upper bound and lower bound to Problem (2.22): UB $\leftarrow +\infty$ and LB $\leftarrow -\infty$

Set iterature counter: $j \leftarrow 0$

while UB $-$ LB $> \varepsilon$ **do**

$j \leftarrow j + 1$

Find the anaytical center $\tilde{\boldsymbol{r}}^{(j)}$ of $\mathcal{C}^{(j-1)}$

Find $\boldsymbol{q}(\boldsymbol{w}, \tilde{\boldsymbol{r}}^{(j)})$ and $\boldsymbol{g}^{(j)}$

Update the set of valid cuts: $\mathcal{C}^{(j)} \leftarrow \mathcal{C}^{(j-1)} \cap \{\boldsymbol{r} : \boldsymbol{g}^{(j)\top}(\boldsymbol{r} - \tilde{\boldsymbol{r}}^{(j)}) \le 0\}$

Update upper bound UB: UB $\leftarrow \min\{\text{UB}, \sum_{i\in[n]} \pi^i_{T,\sqrt{\tilde{r}_i^{(j)}}}(w_i)\}$

Update lower bound LB as the optimal value of

$$\min_{\gamma,(r_i)_{i\in[n]}} \max_{k\in[j]} \left\{ \sum_{i\in[n]} \pi^i_{T,\sqrt{\tilde{r}_i^{(k)}}}(w_i) + \boldsymbol{g}^{(k)\top}(\boldsymbol{r} - \tilde{\boldsymbol{r}}^{(k)}) : 0 \le r_i \le \bar{\theta}_i^2(w_i), \forall i \in [n], \sum_{i\in[n]} r_i \le \theta^2 \right\}$$

Output: $\boldsymbol{r} = (r_i)_{i\in[n]}$ as an optimal solution to Problem (2.22).

A.6 Additional Material for Section 2.8 (numerical experiments)

A.6.1 Details on the Numerical Experiments

We provide the details on the numerical experiments in Tables A.1 - A.4.

Table A.1. Simulation Details for Numerical Experiments of Figures 2.3- 2.5

Figure	Simulation details
Figure 2.3 (a) and (b)	$\xi \sim \mathcal{TN}(50, 20^2, [0, 100])$ where $\Xi = [0, 100]$ and $w \sim \mathcal{U}[c, s]$ and $w \sim \mathcal{U}[c, s]$
Figure 2.3 (c) and (d)	$\xi \sim \mathcal{TN}(50, 20^2, [0, 100])$ where $\Xi = [0, 100]$ and $w \sim \mathcal{TN}(c + 0.3(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.3 (e) and (f)	$\xi \sim \mathcal{TN}(50, 20^2, [0, 100])$ where $\Xi = [0, 100]$ and $w \sim \mathcal{TN}(c + 0.9(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.4 (a) and (b)	True profit $\pi(w; \bar{\mathbb{P}})$ is two modal and $w \sim \mathcal{U}[c, s]$ and $w \sim \mathcal{U}[c, s]$
Figure 2.4 (c) and (d)	True profit $\pi(w; \bar{\mathbb{P}})$ is two modal and $w \sim \mathcal{TN}(c + 0.5(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.4 (e) and (f)	True profit $\pi(w; \bar{\mathbb{P}})$ is two modal and $w \sim \mathcal{TN}(c + 0.9(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.5 (a) and (b)	$\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$ and $w \sim \mathcal{U}[c, s]$ and $w \sim \mathcal{U}[c, s]$
Figure 2.5 (c) and (d)	$\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$ and $w \sim \mathcal{TN}(c + 0.5(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.5 (e) and (f)	$\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$ and $w \sim \mathcal{TN}(c + 0.9(s - c), 0.01(s - c)^2, [c, s])$

Table A.2. Simulation Details for Numerical Experiments of Figures 2.6 - 2.7

Figure	Simulation details
Figure 2.6 (a) and (b)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [50, 50]$, $\boldsymbol{s} - \boldsymbol{c} = [20, 20]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{U}[c_2, s_2]$ independently
Figure 2.6 (c) and (d)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [50, 50]$, $\boldsymbol{s} - \boldsymbol{c} = [20, 20]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{TN}(c_1 + 0.2(s_2 - c_2), (0.1(s_2 - c_2))^2, [c_2, s_2])$. w_1 and w_2 are generated independently
Figure 2.6 (e) and (f)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [50, 50]$, $\boldsymbol{s} - \boldsymbol{c} = [15, 35]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{U}[c_2, s_2]$ independently
Figure 2.7 (a) and (b)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [30, 70]$, $\boldsymbol{s} - \boldsymbol{c} = [15, 35]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{U}[c_2, s_2]$ independently
Figure 2.7 (c) and (d)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [30, 70]$, $\boldsymbol{s} - \boldsymbol{c} = [15, 35]$, and $w_1 \sim \mathcal{TN}(c_1 + 0.2(s_1 - c_1), (0.1(s_1 - c_1))^2, [c_1, s_1])$, $w_2 \sim \mathcal{U}[c_2, s_2]$. w_1 and w_2 are generated independently
Figure 2.7 (e) and (f)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0, 100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ $\boldsymbol{\mu} = [30, 70]$, $\boldsymbol{s} - \boldsymbol{c} = [15, 35]$, $\boldsymbol{w} \sim \mathcal{TN}(\boldsymbol{\mu}_w, \Sigma_w, \Xi_w)$ where $\Xi_w = [0, 100]^2$, $\boldsymbol{w} = [c_1 + 0.2(s_1 - c_1), c_2 + 0.9(s_2 - c_2)]$ and $\Sigma_w = \begin{pmatrix} 0.1^2(s_1-c_1)^2 & 0.1^2(s_1-c_1)(s_2-c_2)/2 \\ 0.1^2(s_1-c_1)(s_2-c_2)/2 & 0.1^2(s_2-c_2)^2 \end{pmatrix}$

Table A.3. Simulation Details for Numerical Experiments of Figures 2.8 - 2.10

Figure	Simulation details
Figure 2.8 (a) and (b)	$\xi \sim \mathcal{TN}(50, 20^2, [0, 100])$ where $\Xi = [0, 100]$ and $w \sim \mathcal{U}[c, s]$ and $w \sim \mathcal{U}[c, s]$
Figure 2.8 (c) and (d)	$\xi \sim \mathcal{TN}(50, 20^2, [0, 100])$ where $\Xi = [0, 100]$ and $w \sim \mathcal{TN}(c + 0.3(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.8 (e) and (f)	$\xi \sim \mathcal{TN}(50, 20^2, [0, 100])$ where $\Xi = [0, 100]$ and $w \sim \mathcal{TN}(c + 0.9(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.9 (a) and (b)	True profit $\pi(w; \bar{\mathbb{P}})$ is two modal and $w \sim \mathcal{U}[c, s]$ and $w \sim \mathcal{U}[c, s]$
Figure 2.9 (c) and (d)	True profit $\pi(w; \bar{\mathbb{P}})$ is two modal and $w \sim \mathcal{TN}(c + 0.5(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.9 (e) and (f)	True profit $\pi(w; \bar{\mathbb{P}})$ is two modal and $w \sim \mathcal{TN}(c + 0.9(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.10 (a) and (b)	$\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$ and $w \sim \mathcal{U}[c, s]$ and $w \sim \mathcal{U}[c, s]$
Figure 2.10 (c) and (d)	$\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$ and $w \sim \mathcal{TN}(c + 0.5(s - c), 0.01(s - c)^2, [c, s])$
Figure 2.10 (e) and (f)	$\log(\xi) \sim \mathcal{TN}(3, 1, (-\infty, \log(100)])$ and $w \sim \mathcal{TN}(c + 0.9(s - c), 0.01(s - c)^2, [c, s])$

Table A.4. Simulation Details for Numerical Experiments of Figures 2.11 - 2.12

Figure	Simulation details
Figure 2.11 (a) and (b)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0,100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [50,50]$, $\boldsymbol{s} - \boldsymbol{c} = [20,20]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{U}[c_2, s_2]$ independently
Figure 2.11 (c) and (d)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0,100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [50,50]$, $\boldsymbol{s} - \boldsymbol{c} = [20,20]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{TN}(c_1 + 0.2(s_2 - c_2), (0.1(s_2 - c_2))^2, [c_2, s_2])$. w_1 and w_2 are generated independently
Figure 2.11 (e) and (f)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0,100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [50,50]$, $\boldsymbol{s} - \boldsymbol{c} = [15,35]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{U}[c_2, s_2]$ independently
Figure 2.12 (a) and (b)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0,100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [30,70]$, $\boldsymbol{s} - \boldsymbol{c} = [15,35]$, and $w_1 \sim \mathcal{U}[c_1, s_1]$, $w_2 \sim \mathcal{U}[c_2, s_2]$ independently
Figure 2.12 (c) and (d)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0,100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ and $\boldsymbol{\mu} = [30,70]$, $\boldsymbol{s} - \boldsymbol{c} = [15,35]$, and $w_1 \sim \mathcal{TN}(c_1 + 0.2(s_1 - c_1), (0.1(s_1 - c_1))^2, [c_1, s_1])$, $w_2 \sim \mathcal{U}[c_2, s_2]$. w_1 and w_2 are generated independently
Figure 2.12 (e) and (f)	$\boldsymbol{\xi} \sim \mathcal{TN}(\boldsymbol{\mu}, \Sigma, \Xi)$ where $\Xi = [0,100]^2$ and $\Sigma = \begin{pmatrix} 20^2 & 0 \\ 0 & 20^2 \end{pmatrix}$ $\boldsymbol{\mu} = [30,70]$, $\boldsymbol{s} - \boldsymbol{c} = [15,35]$, $\boldsymbol{w} \sim \mathcal{TN}(\boldsymbol{\mu}_w, \Sigma_w, \Xi_w)$ where $\Xi_w = [0,100]^2$, $\boldsymbol{w} = [c_1 + 0.2(s_1 - c_1), c_2 + 0.9(s_2 - c_2)]$ and $\Sigma_w = \begin{pmatrix} 0.1^2(s_1-c_1)^2 & 0.1^2(s_1-c_1)(s_2-c_2)/2 \\ 0.1^2(s_1-c_1)(s_2-c_2)/2 & 0.1^2(s_2-c_2)^2 \end{pmatrix}$

A.7 Proofs of Auxiliary Results

A.7.1 Proof of Lemma A.4.1 (worst-case profit and worst-case distribution in single-product case)

To derive $\rho(q)$, we consider the following modified problem:

$$\tilde{\rho}(q) \triangleq \max_{\mathbb{P}\in\mathcal{P}(\Xi)} \left\{\mathbb{P}(\xi \leq q) : W_2(\mathbb{P}, \hat{\mathbb{P}}_T) \leq \theta, \mathbb{P}(\xi \leq l(q)) = 1 - w(l(q))/s\right\}. \tag{A.22a}$$

Note that the constraint set of Problem (A.22) only includes the first-order conditions (2.3) for a *single* data point $(l(q), w(l(q)))$. Let $\tilde{\mathbb{P}}_{q,\theta}$ be the solution to Problem (A.22). In Lemma A.7.1, we will derive a closed form expression for $\tilde{\rho}(q)$. In Lemma A.7.2, we derive the optimal solution $\tilde{\mathbb{P}}_{q,\theta}$. Finally, to prove Lemma A.4.1, we will show $\tilde{\mathbb{P}}_{q,\theta}$ is also feasible to Problem (A.13), and thus $\rho(q) = \tilde{\rho}(q)$.

Lemma A.7.1. *For $q \geq 0$,*

$$\tilde{\rho}(q) = \begin{cases} 1 - w(l(q))/s, & \theta^2 \geq \sum_{\{k:q<\hat{\xi}^k\leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2, \\ \sum_{\{k:\hat{\xi}^k\leq z(q,\theta)\}} \eta^k + \frac{\theta^2 - \sum_{\{k:q<\hat{\xi}^k\leq z(q,\theta)\}} \eta^k (q-\hat{\xi}^k)^2}{(q-z(q,\theta))^2}, & \theta^2 < \sum_{\{k:q<\hat{\xi}^k\leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2. \end{cases}$$

Proof of Lemma A.7.1. We first show that $\tilde{\rho}(q) = 1-w(l(q))/s$ for $\theta^2 \geq \sum_{\{k:q<\hat{\xi}^k\leq l(q)\}} \eta_k(q-\hat{\xi}^k)^2$. Notice that $\tilde{\rho}(q)$ is always bounded from above by $1 - w(l(q))/s$ for $q \leq l(q)$, since any feasible distribution $\mathbb{P}$ to Problem (A.22) must satisfy $\mathbb{P}(\xi \leq l(q)) = 1 - w(l(q))/s$ for $q \leq l(q)$. Furthermore, the distribution

$$\hat{\mathbb{P}} = \sum_{\{k:0\leq\hat{\xi}^k<q\}} \eta^k \delta_{\hat{\xi}^k} + \sum_{\{k:q\leq\hat{\xi}^k\leq l(q)\}} \eta^k \delta_q + \sum_{\{k:\hat{\xi}^k>l(q)\}} \eta^k \delta_{\hat{\xi}^k} \tag{A.23}$$

is feasible to Problem (A.22) since according to the expression (A.23), $W_2^2(\hat{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq \sum_{\{k:q<\hat{\xi}^k\leq l(q)\}} \eta^k (q - \hat{\xi}^k)^2 \leq \theta^2$ and $\hat{\mathbb{P}}(\xi \leq q) = \sum_{\{k:\hat{\xi}^k\leq l(q)\}} \eta^k = 1 - w(l(q))/s$. Therefore, $\hat{\mathbb{P}}$ achieves the maximum of Problem (A.22), and thus we have $\tilde{\rho}(q) = 1 - w(l(q))/s$ for $\theta^2 \geq \sum_{\{k:q<\hat{\xi}^k\leq l(q)\}} \eta_k(q - \hat{\xi}^k)^2$.

We next consider the case where $\theta^2 < \sum_{\{k:q<\hat{\xi}^k \le l(q)\}} \eta_k (q-\hat{\xi}^k)^2$. We will formulate the Lagrangian dual to Problem (A.22) and then identify its solution. By strong duality between Problem (A.22) and its dual, we will then obtain the optimal value to Problem (A.22).

Introducing Lagrange multipliers $\lambda_1 \ge 0$ and $\lambda_2 \in \mathbb{R}$, the Lagrangian dual of Problem (A.22) is $\inf_{\lambda_1 \ge 0, \lambda_2} v(q,\lambda_1,\lambda_2) \triangleq \Big\{ \lambda_1\theta^2 - \lambda_2(1-w(l(q))/s)$ $+ \sum_{k=1}^K \eta^k \sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2)\Big\}$, where $f_k(\xi,\lambda_1,\lambda_2) \triangleq \mathbb{1}(\xi \in [0,1]) + \lambda_2 \mathbb{1}(\xi \in [0,l(q)]) - \lambda_1(\xi-\hat{\xi}_k)^2$. We can solve analytically for a dual optimal solution $(\lambda_1^*,\lambda_2^*)$ to obtain the dual optimal value $v(q,\lambda_1^*,\lambda_2^*)$.

For any fixed $\lambda_1 \ge 0$, $v(q,\lambda_1,\lambda_2)$ is increasing on $\lambda_2 \ge 0$ and therefore $\lambda_2^* \le 0$. For fixed λ_1 and $\lambda_2 \ge 0$, we have the following observations:

1.1 For $\hat{\xi}^k \le q$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(\hat{\xi}^k,\lambda_1,\lambda_2) = 1+\lambda_2$.

1.2 For $q < \hat{\xi}^k \le l(q)$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = \max\{\lambda_2, 1+\lambda_2-\lambda_1(q-\hat{\xi}^k)^2\} = \lambda_2 + \max\{0, 1-\lambda_1(q-\hat{\xi}^k)^2\}$.

1.3 For $\hat{\xi}^k > l(q)$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1^*,\lambda_2) = \max\{0, \lambda_2-\lambda_1(q_1-\hat{\xi}^k)^2, 1+\lambda_2-\lambda_1(q-\hat{\xi}^k)^2\}$.

It follows that for $\lambda_2 \ge 0$, we have

$$\begin{aligned} v(q,\lambda_1,\lambda_2) = \lambda_1\theta + \lambda_2 \left(1 - w(l(q))/s + \sum_{\{k:\hat{\xi}^k \le l(q)\}} \eta^k\right) + \sum_{\{k:\hat{\xi}^k \le l(q)\}} \eta^k \\ + \sum_{\{k:q<\hat{\xi}^k \le l(q)\}} \eta^k \max\{0, -\lambda_1(x-\hat{\xi}^k)^2\} \\ + \sum_{\{k:\hat{\xi}^k > l(q)\}} \eta^k \max\{0, \lambda_2-\lambda_1(q_1-\hat{\xi}^k)^2, 1+\lambda_2-\lambda_1(x-\hat{\xi}^k)^2\}. \end{aligned} \tag{A.24}$$

By construction of $\hat{\mathbb{P}}_T$, $\sum_{\{k:\hat{\xi}^k \le l(q)\}} \eta^k = 1-w(l(q))/s$, and $v(q,\lambda_1,\lambda_2)$ involves linear term of λ_2 with positive coefficients. Therefore $v(q,\lambda_1,\lambda_2)$ is non-decreasing in λ_2. Thus we must take $\lambda_2^* \le 0$ to minimize $v(q,\lambda_1,\lambda_2)$.

We now restrict attention to $\lambda_2^* \le 0$. For $\lambda_1 \ge 0$ and $\lambda_2 \le 0$, we can compute $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2)$ for the following cases of $\hat{\xi}^k$, λ_1, and λ_2:

2.1 If $\hat{\xi}^k \le q$ and $\lambda_2 \ge -1 - \lambda_1(l(q) - \hat{\xi}^k)^2$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(\hat{\xi}^k,\lambda_1,\lambda_2) = 1 + \lambda_2$.

2.2 If $\hat{\xi}^k \le q$ and $\lambda_2 < -1 - \lambda_1(l(q) - \hat{\xi}^k)^2$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(l(q),\lambda_1,\lambda_2) = -\lambda_1(l(q) - \hat{\xi}^k)^2$.

2.3 If $q < \hat{\xi}^k \le l(q), 1 - \lambda_1(q - \hat{\xi}^k)^2 \ge 0$ and $1 + \lambda_2 - \lambda_1(q - \hat{\xi}^k)^2 \ge -\lambda_1(l(q) - \hat{\xi}^k)^2$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(q) = 1 + \lambda_2 - \lambda_1(q - \hat{\xi}^k)^2$.

2.4 If $q < \hat{\xi}^k \le l(q), 1 - \lambda_1(q - \hat{\xi}^k)^2 < 0$ and $\lambda_2 \ge -\lambda_1(l(q) - \hat{\xi}^k)^2$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(\hat{\xi}^k,\lambda_1,\lambda_2) = \lambda_2$.

2.5 If $q < \hat{\xi}^k \le l(q), 1 + \lambda_2 - \lambda_1(q - \hat{\xi}^k)^2 < -\lambda_1(l(q) - \hat{\xi}^k)^2$ and $\lambda_2 < -\lambda_1(l(q) - \hat{\xi}^k)^2$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(l(q)) = -\lambda_1(l(q) - \hat{\xi}^k)^2$.

2.6 If $\hat{\xi}^k > l(q)$ and $1 + \lambda_2 - \lambda_1(q - \hat{\xi}^k)^2 \ge 0$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(q,\lambda_1,\lambda_2) = 1 + \lambda_2 - \lambda_1(q - \hat{\xi}^k)^2$.

2.7 If $\hat{\xi}^k > l(q)$ and $1 + \lambda_2 - \lambda_1(q - \hat{\xi}^k)^2 < 0$, $\sup_{\xi\in\Xi} f_k(\xi,\lambda_1,\lambda_2) = f_k(\hat{\xi}^k,\lambda_1,\lambda_2) = 0$.

Based on the above cases, we find

$$
\begin{aligned}
v(q,\lambda_1,\lambda_2) = \lambda_2 \Bigg\{ & -1 + w(l(q))/s + \sum_{\substack{k:\hat{\xi}^k \le q,\\ -1-\lambda_1(l(q)-\hat{\xi}^k)^2 \le \lambda_2}} \eta^k + \sum_{\substack{k:q<\hat{\xi}^k \le l(q),\\ -\lambda_1(l(q)-\hat{\xi}^k)^2 \le \lambda_2\\ +\max\{0,1-\lambda_1(q-\hat{\xi}^k)^2\}}} \eta^k \\
& + \sum_{\substack{k:\hat{\xi}^k > l(q),\\ 1+\lambda_2-\lambda_1(q-\hat{\xi}^k)^2 \ge 0}} \eta^k \Bigg\} + \lambda_1 \Bigg\{ \theta - \sum_{\substack{k:\hat{\xi}^k \le q,\\ -1-\lambda_1(l(q)-\hat{\xi}^k)^2 > \lambda_2}} \eta^k (l(q) - \hat{\xi}^k)^2 - \sum_{\substack{k:q<\hat{\xi}^k \le l(q),\\ 1-\lambda_1(q-\hat{\xi}^k)^2 \ge 0,\\ 1+\lambda_2-\lambda_1(q-\hat{\xi}^k)^2\\ \ge -\lambda_1(l(q)-\hat{\xi}^k)^2}} \eta^k (q - \hat{\xi}^k)^2 \\
& - \sum_{\substack{k:q<\hat{\xi}^k \le l(q),\\ -\lambda_1(l(q)-\hat{\xi}^k)^2 > \lambda_2\\ +\max\{0,1-\lambda_1(q-\hat{\xi}^k)^2\}}} \eta^k (l(q) - \hat{\xi}^k)^2 - \sum_{\substack{k:\hat{\xi}^k > l(q),\\ \lambda_2 \ge -1+\lambda_1(q-\hat{\xi}^k)^2}} \eta^k (q - \hat{\xi}^k)^2 \Bigg\} \\
& + \sum_{\substack{k:\hat{\xi}^k \le q,\\ -1-\lambda_1(l(q)-\hat{\xi}^k)^2 \le \lambda_2}} \eta^k + \sum_{\substack{k:q<\hat{\xi}^k \le l(q),\\ 1-\lambda_1(q-\hat{\xi}^k)^2 \ge 0\\ 1+\lambda_2-\lambda_1(q-\hat{\xi}^k)^2\\ \ge -\lambda_1(l(q)-\hat{\xi}^k)^2}} \eta^k + \sum_{\substack{k:\hat{\xi}^k > l(q),\\ \lambda_2 \ge -1+\lambda_1(q-\hat{\xi}^k)^2}} \eta^k.
\end{aligned}
$$

In the following, we establish that the dual optimal solution $(\lambda_1^*, \lambda_2^*)$ satisfies the inequalities $-1-\lambda_1^*(q-l(q))^2 \le \lambda_2^* \le -1+\lambda_1^*(q-l(q))^2$ in Step 1 and Step 2. Then, we show that $\max\{\lambda_2^*, \lambda_2^*+1-\lambda_1^*(q-l(q))^2\} = 0$ in Step 3. Finally, we will show that $\lambda_1^* = 1/(q-\hat{\xi}^k)^2$ for some $\hat{\xi}^k \in (q, l(q)]$ and $\lambda_2^* = 0$, and the desired result will follow.

Step 1: Show that $\lambda_2^* \le -1+\lambda_1^*(q-l(q))^2$.

To show that $\lambda_2^* \le -1+\lambda_1^*(q-l(q))^2$, for a contradiction suppose that there exists some $\lambda_2 > -1+\lambda_1^*(q-l(q))^2$ such that $v(q,\lambda_1^*, -1+\lambda_1^*(q-l(q))^2) > v(q,\lambda_1^*,\lambda_2)$. Then, we have

$$v(q,\lambda_1^*,-1+\lambda_1^*(q-l(q))^2) = \lambda_1^*\Bigg(\theta - \sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k(q-\hat{\xi}^k)^2\Bigg) + \sum_{\{k:\hat{\xi}^k\le q\}} \eta_k + \sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k,$$

and

$$\begin{aligned} v(q,\lambda_1^*,\lambda_2) = \lambda_2 &\sum_{\substack{k:\hat{\xi}^k>l(q),\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k + \lambda_1^*\Bigg(\theta - \sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k(q-\hat{\xi}^k)^2 \\ &- \sum_{\substack{k:\hat{\xi}^k>l(q),\\ \lambda_2\ge -1+\lambda_1^*(q-\hat{\xi}^k)^2}} \eta^k(q-\hat{\xi}^k)^2\Bigg) + \sum_{\{k:\hat{\xi}^k\le q\}} \eta_k + \sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k + \sum_{\substack{k:\hat{\xi}^k>l(q),\\ \lambda_2\ge -1+\lambda_1^*(q-\hat{\xi}^k)^2}} \eta^k. \end{aligned}$$

It then follows that $v(q,\lambda_1^*,-1+\lambda_1^*(q-l(q))^2) - v(q,\lambda_1^*,\lambda_2) = -\sum_{\substack{k:\hat{\xi}^k>l(q),\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k(1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2) \le 0$, which contradicts the assumption that $v(q,\lambda_1^*,-1+\lambda_1^*(q-l(q))^2) > v(q,\lambda_1^*,\lambda_2)$.

Step 2: Show that $\lambda_2^* \ge -1-\lambda_1^*(q-l(q))^2$.

To show that $\lambda_2^* \ge -1-\lambda_1^*(q-l(q))^2$, for a contradiction suppose that there exists some $\lambda_2 < -1-\lambda_1^*(q-l(q))^2$ such that $v(q,\lambda_1^*,-1-\lambda_1^*(q-l(q))^2) > v(q,\lambda_1^*,\lambda_2)$. Then, we have $v(q,\lambda_1^*,-1-\lambda_1^*(q-l(q))^2) = (-1-\lambda_1^*(q-l(q))^2)\big\{-1+w(l(q))/s+\sum_{\{k:\hat{\xi}^k\le q\}}\eta_k\big\} + \lambda_1^*\big\{\theta - \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta_k(l(q)-\hat{\xi}^k)^2\big\} + \sum_{\{k:\hat{\xi}^k\le q\}}\eta_k$ and $v(q,\lambda_1^*,\lambda_2) = \lambda_2\big\{-1+w(l(q))/s+$

$\sum_{\substack{k:\hat{\xi}^k\le q,\\ -1-\lambda_1^*(l(q)-\hat{\xi}^k)^2\le\lambda_2}} \eta^k\} + \lambda_1^*\{\theta - \sum_{\{k:\hat{\xi}^k\le q\}} \eta_k(l(q)-\hat{\xi}^k)^2$
$+\sum_{\substack{k:\hat{\xi}^k\le q,\\ -1-\lambda_1^*(l(q)-\hat{\xi}^k)^2\le\lambda_2}} \eta^k(l(q)-\hat{\xi}^k)^2 - \sum_{\{k:q<\hat{\xi}^k\le l(q)\}} \eta_k(l(q)-\hat{\xi}^k)^2\} + \sum_{\substack{k:\hat{\xi}^k\le q,\\ -1-\lambda_1^*(l(q)-\hat{\xi}^k)^2\le\lambda_2}} \eta^k$. It follows that

$$
\begin{aligned}
&v(q,\lambda_1^*,\lambda_2) - v(q,\lambda_1^*,-1-\lambda_1^*(q-l(q))^2)\\
=&(\lambda_2+1+\lambda_1^*(q-l(q))^2)\Bigg\{-1+w(l(q))/s+\sum_{\substack{k:\hat{\xi}^k\le q,\\ -1-\lambda_1^*(l(q)-\hat{\xi}^k)^2\le\lambda_2}} \eta^k\Bigg\} \ge 0,
\end{aligned}
$$

which contradicts the assumption that $v(q,\lambda_1^*,-1-\lambda_1^*(q-l(q))^2) > v(q,\lambda_1^*,\lambda_2)$.

Step 3: Show that $\max\{\lambda_2^*,\lambda_2^*+1-\lambda_1^*(q-l(q))^2\}=0$**.**

We have established the inequalities $-1-\lambda_1^*(q-l(q))^2 \le \lambda_2^* \le -1+\lambda_1^*(q-l(q))^2$, and hence $\max\{\lambda_2^*,\lambda_2^*+1-\lambda_1^*(q-l(q))^2\}\le 0$. In the following, we will show that $\max\{\lambda_2^*,\lambda_2^*+1-\lambda_1^*(q-l(q))^2\}=0$. Let λ_2' satisfy $\max\{\lambda_2',\lambda_2'+1-\lambda_1^*(q-l(q))^2\}=0$, and for a contradiction suppose that there exists some $\lambda_2\le 0$ such that $\max\{\lambda_2,\lambda_2+1-\lambda_1^*(q-l(q))^2\}<0$ and $v(q,\lambda_1^*,\lambda_2') > v(q,\lambda_1^*,\lambda_2)$. Then for $-1-\lambda_1^*(q-l(q))^2 \le \lambda_2,\lambda_2' \le -1+\lambda_1^*(q-l(q))^2$, we have

$$
v(q,\lambda_1^*,\lambda_2') = \lambda_1^*\Bigg\{\theta - \sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k(q-\hat{\xi}^k)^2\Bigg\} + \sum_{\{k:\hat{\xi}^k\le q\}} \eta_k + \sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0}} \eta^k,
$$

and

$$v(q,\lambda_1^*,\lambda_2)=\lambda_2\Bigg\{-1+w(l(q))/s+\sum_{\{k:\hat{\xi}^k\le q\}}\eta_k+\sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ -\lambda_1^*(l(q)-\hat{\xi}^k)^2\le\lambda_2\\ +\max\{0,1-\lambda_1^*(q-\hat{\xi}^k)^2\}}}\eta^k\Bigg\}$$

$$+\lambda_1^*\Bigg\{\theta-\sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0,\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\\ \ge-\lambda_1^*(l(q)-\hat{\xi}^k)^2}}\eta^k(q-\hat{\xi}^k)^2-\sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta_k(l(q)-\hat{\xi}^k)^2+\sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ -\lambda_1^*(l(q)-\hat{\xi}^k)^2\le\lambda_2\\ +\max\{0,1-\lambda_1^*(q-\hat{\xi}^k)^2\}}}\eta^k(l(q)-\hat{\xi}^k)^2\Bigg\}$$

$$+\sum_{\{k:\hat{\xi}^k\le q\}}\eta_k+\sum_{\substack{k:q<\hat{\xi}^k\le l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\ge 0\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\\ \ge-\lambda_1^*(l(q)-\hat{\xi}^k)^2}}\eta^k.$$

It follows that

$$v(q,\lambda_1^*,\lambda_2)-v(q,\lambda_1^*,\lambda_2') = \lambda_2(-1+w(l(q))/s+\sum_{\{k:\hat{\xi}^k\leq q\}}\eta_k)+\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\geq 0\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\geq\\ -\lambda_1^*(l(q)-\hat{\xi}^k)^2}}\eta^k(-\lambda_1^*(q-\hat{\xi}^k)^2+1)$$

$$+\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ -\lambda_1^*(l(q)-\hat{\xi}^k)^2\leq\lambda_2\\ +\max\{0,1-\lambda_1^*(q-\hat{\xi}^k)^2\}}}\eta^k(\lambda_2+\lambda_1^*(l(q)-\hat{\xi}^k)^2)-\lambda_1^*\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k(l(q)-\hat{\xi}^k)^2$$

$$-\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\geq 0}}\eta^k(1-\lambda_1^*(q-\hat{\xi}^k)^2) = \lambda_2(-1+w(l(q))/s+\sum_{\{k:\hat{\xi}^k\leq q\}}\eta_k)$$

$$-\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\geq 0\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\\ <-\lambda_1^*(l(q)-\hat{\xi}^k)^2}}\eta^k(-\lambda_1^*(q-\hat{\xi}^k)^2+1)+\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k(\lambda_2+\lambda_1^*(l(q)-\hat{\xi}^k)^2)$$

$$-\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ -\lambda^*(l(q)-\hat{\xi}^k)^2>\lambda_2,\\ -\lambda_1^*(l(q)-\hat{\xi}^k)^2>\lambda_2+1-\lambda^*(q-\hat{\xi}^k)^2}}\eta^k(\lambda_2+\lambda_1^*(l(q)-\hat{\xi}^k)^2)-\lambda_1^*\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k(l(q)-\hat{\xi}^k)^2$$

$$=\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ 1-\lambda_1^*(q-\hat{\xi}^k)^2\geq 0\\ 1+\lambda_2-\lambda_1^*(q-\hat{\xi}^k)^2\\ <-\lambda_1^*(l(q)-\hat{\xi}^k)^2}}\eta^k(\lambda_1^*(q-\hat{\xi}^k)^2-1)-\sum_{\substack{k:q<\hat{\xi}^k\leq l(q),\\ -\lambda^*(l(q)-\hat{\xi}^k)^2>\lambda_2,\\ -\lambda_1^*(l(q)-\hat{\xi}^k)^2>\lambda_2+1-\lambda^*(q-\hat{\xi}^k)^2}}\eta^k(\lambda_2+\lambda_1^*(l(q)-\hat{\xi}^k)^2)\geq 0,$$

which contradicts the assumption that $v(q,\lambda_1^*,\lambda_2')>v(q,\lambda_1^*,\lambda_2)$.

Step 4: Show that $\lambda_1^*\geq 1/(q-l(q))^2$ and $\lambda_2^*=0$.

We have shown that $\max\{\lambda_2^*,\lambda_2^*+1-\lambda_1^*(q-l(q))^2\}=0$. Now, for $\theta^2<\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k(q-\hat{\xi}^k)^2$, we show that $\lambda_1^*\geq 1/(q-l(q))^2$ and $\lambda_2^*=0$. Let $\lambda_1'=1/(q-l(q))^2, \lambda_2'=0$, $\lambda_1<1/(q-l(q))^2$, and $\lambda_2=-1+\lambda_1(q-l(q))^2$.

For a contradiction, suppose $v(q,\lambda_1,\lambda_2)<v(q,\lambda_1',\lambda_2')$, then $v(q,\lambda_1,\lambda_2)$ $=\lambda_1\{\theta-\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k(q-\hat{\xi}^k)^2\}+\sum_{\{k:\hat{\xi}^k\leq q\}}\eta_k+\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k$ and $v(q,\lambda_1',\lambda_2')=\lambda_1'\{\theta-\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k(q-\hat{\xi}^k)^2\}+\sum_{\{k:\hat{\xi}^k\leq q\}}\eta_k+\sum_{\{k:q<\hat{\xi}^k\leq l(q)\}}\eta_k$. It follows that $v(q,\lambda_1,\lambda_2)-$

$v(q,\lambda_1',\lambda_2') = (\lambda_1-\lambda_1')\big\{\theta - \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta_k\big\} \ge 0$, which contradicts the assumption that $v(q,\lambda_1,\lambda_2) < v(q,\lambda_1',\lambda_2')$.

In summary, we have shown that for $\theta^2 < \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta_k(q-\hat{\xi}^k)^2$, we have $\lambda_2^*=0$ and $\lambda_1^* \ge 1/(q-l(q))^2$. Define $\lambda_{1k} \triangleq 1/(q-\hat{\xi}^k)^2$ for all k with $q<\hat{\xi}^k\le l(q)$. For $\lambda_1 \in (\lambda_{1,k+1},\lambda_{1k}]$ for some $\hat{\xi}^k \in (q,l(q)]$, we have $v(q,\lambda_1,0) = \lambda_1\theta + \sum_{\{k:\hat{\xi}^k\le z(q,\theta)\}}\eta^k - \lambda_1\sum_{\{k:q<\hat{\xi}^k\le z(q,\theta)\}}\eta^k(q-\hat{\xi}^k)^2$. It then follows that $v(q,\lambda_1^*,\lambda_2^*) = \sum_{\{k:\hat{\xi}^k<z(q,\theta)\}}\eta^k$ $+ \frac{1}{(q-z(q,\theta))^2}\Big(\theta - \sum_{\{k:q<\hat{\xi}^k<z(q,\theta)\}}\eta^k(q-\hat{\xi}^k)^2\Big)$, which gives the optimal value $v(q,\lambda_1^*,\lambda_2^*)$. In particular, strong duality holds for Problem (A.22), which can be proved similarly as Proposition A.5.1, and thus we omit its proof. Thus, by strong duality we have $\tilde{\rho}(q) = v(q,\lambda_1^*,\lambda_2^*)$, which gives the statement in Lemma A.4.1 part (i). □

Lemma A.7.2 provides $\tilde{\mathbb{P}}_{q,\theta}$, the optimal solution to Problem (A.22).

Lemma A.7.2. *Recall $\gamma(q)$ defined in Eq. (A.11). Then*

$$\tilde{\mathbb{P}}_{q,\theta} = \begin{cases} \sum_{\{k:0\le\hat{\xi}^k<q\}}\eta^k\delta_{\hat{\xi}^k} + \sum_{\{k:q\le\hat{\xi}^k\le l(q)\}}\eta^k\delta_q \\ \qquad\qquad + \sum_{\{k:\hat{\xi}^k>l(q)\}}\eta^k\delta_{\hat{\xi}^k}, & \theta^2 \ge \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta^k(q-\hat{\xi}^k)^2, \\ \sum_{\{k:0\le\hat{\xi}^k<q\}}\eta^k\delta_{\hat{\xi}^k} + \Big(\sum_{\{k:q\le\hat{\xi}^k<z(q,\theta)\}}\eta^k + \gamma\Big)\delta_q \\ \qquad + (\eta^{k^*}-\gamma)\delta_{z(q,\theta)} + \sum_{\{k:\hat{\xi}^k>z(q,\theta)\}}\eta^k\delta_{\hat{\xi}^k}, & \theta^2 < \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta^k(q-\hat{\xi}^k)^2. \end{cases}$$

Proof of Lemma A.7.2. To prove that $\tilde{\mathbb{P}}_{q,\theta}$ is a worst-case distribution for Problem (A.22), we need to verify that $\tilde{\mathbb{P}}_{q,\theta}$ is feasible for Problem (A.22) and also that $\tilde{\mathbb{P}}_{q,\theta}$ achieves the optimal value $\tilde{\rho}(q)$.

For $\theta^2 \ge \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta^k(q-\hat{\xi}^k)^2$, $\tilde{\mathbb{P}}_{q,\theta}$ is feasible since $W_2^2(\tilde{\mathbb{P}}_{q,\theta},\hat{\mathbb{P}}_T) \le \sum_{\{k:q<\hat{\xi}^k\le l(q)\}}\eta^k(q-\hat{\xi}^k)^2 \le \theta^2$. Furthermore, $\tilde{\mathbb{P}}_{q,\theta}$ achieves $\tilde{\rho}(q)$ since we have $\tilde{\mathbb{P}}_{q,\theta}(\xi \le q) = \sum_{\{k:\hat{\xi}^k\le l(q)\}}\eta_k = 1 - w(l(q))/s$.

For $\theta^2 < \sum_{\{k:q<\hat{\xi}^k\le l(q)\}} \eta^k(q-\hat{\xi}^k)^2$, $\tilde{\mathbb{P}}_{q,\theta}$ is feasible since $W_2^2(\tilde{\mathbb{P}}_{q,\theta}, \hat{\mathbb{P}}_T) \le \sum_{\{k:q<\hat{\xi}^k<z(q,\theta)\}} \eta^k(q-\hat{\xi}^k)^2 + \gamma(q)(q-z(q,\theta))^2 = \theta^2$ and since

$$\begin{aligned}
\tilde{\mathbb{P}}_{q,\theta}(\xi \le q) &= \sum_{\{k:\hat{\xi}^k<z(q,\theta)\}} \eta^k + \gamma \\
&= \sum_{\{k:\hat{\xi}^k<z(q,\theta)\}} \eta^k + \frac{\theta - \sum_{\{k:q<\hat{\xi}^k<z(q,\theta)\}} \eta^k (q-\hat{\xi}^k)^2}{(q-z(q,\theta))^2} \\
&= \sum_{\{k:\hat{\xi}^k<z(q,\theta)\}} \eta^k + \frac{1}{(q-z(q,\theta))^2}\left(\theta - \sum_{\{k:q<\hat{\xi}^k<z(q,\theta)\}} \eta^k (q-\hat{\xi}^k)^2\right) \\
&= \sum_{\{k:\hat{\xi}^k\le z(q,\theta)\}} \eta^k + \frac{1}{(q-z(q,\theta))^2}\left(\theta - \sum_{\{k:q<\hat{\xi}^k\le z(q,\theta)\}} \eta^k (q-\hat{\xi}^k)^2\right),
\end{aligned}$$

implying $\tilde{\mathbb{P}}_{q,\theta}$ achieves the optimal value $\tilde{\rho}(q)$.

Thus we have show that $\tilde{\mathbb{P}}_{q,\theta}$ is feasible for Problem (A.22) for both cases $\theta^2 \ge \sum_{\{k:q<\hat{\xi}^k\le l(q)\}} \eta^k(q-\hat{\xi}^k)^2$ and $\theta^2 < \sum_{\{k:q<\hat{\xi}^k\le l(q)\}} \eta^k(q-\hat{\xi}^k)^2$, and that $\tilde{\mathbb{P}}_{q,\theta}$ achieves the optimal value $\tilde{\rho}(q)$. We conclude that $\tilde{\mathbb{P}}_{q,\theta}$ is an optimal solution for Problem (A.22). □

Proof of Lemma A.4.1. The conclusion of Lemma A.4.1 follows by noticing that $\tilde{\mathbb{P}}_{q,\theta}$ is also feasible to Problem (A.13). □

A.7.2 Proof of Lemma A.4.2

(i) The existence of $\tilde{q}(w,\theta)$ follows from the continuity of the mapping $q \to f(q,\xi(w))$, and the Weierstrass Theorem, which says a continuous function attains the maximum and minimum over a compact set. Given the range of θ (i.e., $\theta < \bar{\theta}(w)$), we have $[q(w),\xi(w)] \cap \{q : 1 - w/s = f(q,\xi(w))\} \ne \emptyset$. As $(w-c)q$ is strictly decreasing in q and $f(q,z)$ is strictly increasing in $q \in [q(w),\xi(w)]$, $\tilde{q}(w,\theta)$ must satisfy the constraint of Problem (A.15) with equality.

(ii) First we show

$$\pi_{T,\theta}(w) = \inf_{q\in\mathcal{Q}_2(w,\theta), q\in(q(w),\xi(w)]} (w-c)q \tag{A.25}$$

for $\theta < \bar{\theta}(w)$. Notice that $c(q,l)$ is continuous and strictly decreasing in $q \in (q(w), \xi(w)]$. If $\theta^2 < c(\xi(w), l)$, then $\theta^2 < c(q,l)$ for all $q \in (q(w), \xi(w)]$ and therefore $q(w,\theta) \notin \mathcal{Q}_1(w,\theta)$. Thus (A.25) holds. Otherwise, choose $q' \in (q(w), \xi(w)]$ to satisfy $\theta^2 = c(q', l)$ (q' exists given $\theta < \bar{\theta}(w)$). Then, $\theta^2 > c(q,l)$ for any $q \in (q(w), q')$, and we have $\inf_{q \in \mathcal{Q}_1(w,\theta)} (w-c)q = (w-c)q'$. At the same time, notice $q' \in \mathcal{Q}_2(w,\theta)$ since $f(q', l) \geq 1 - w/s$. Thus w.l.o.g. we have $q(w,\theta) \in \mathcal{Q}_2(w,\theta)$. Thus, it suffices to show that $\tilde{q}(w,\theta)$ is also optimal to Problem (A.25).

According to the definition of $\xi(w)$ and the characterization of $\tilde{q}(w,\theta)$, we have $\sum_{\{k:\tilde{q}(w,\theta)<\hat{\xi}^k<\xi(w)\}} \eta^k(\tilde{q}(w,\theta) - \hat{\xi}^k)^2 < \theta^2 \leq c(\tilde{q}(w,\theta), \xi(w))$, and thus $z(\tilde{q}(w,\theta),\theta) = \xi(w)$. Furthermore, $c(\tilde{q}(w,\theta), l) \geq c(\tilde{q}(w,\theta), \xi(w)) \geq \theta^2$. It follows that $\tilde{q}(w,\theta)$ is feasible to Problem (A.25). Then, $\tilde{q}(w,\theta)$ is optimal to Problem (A.25) because $f(q, z(q,\theta))$ is strictly increasing in q for any q that satisfies $\theta^2 \leq c(q,l)$ (notice $f(q, z(q,\theta)) = \rho(q)$ when $\theta^2 \leq c(q,l)$ and $\rho(q)$ is clearly increasing in q), and $\tilde{q}(w,\theta)$ satisfies the constraint $1 - w/s = f(\tilde{q}(w,\theta), \xi(w))$ with equality.

A.7.3 Proof of Proposition A.5.1 (strong duality for construction of the center distribution)

We derive the Lagrangian dual to Problem (2.5). Let $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ have distributions $\mathbb{P}_t$ and $\hat{\mathbb{P}}^{\mathrm{e}}_T$ respectively, and let $\mathbb{P}_t$ be the conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{\xi}' = \boldsymbol{\xi}^t$ for all $t \in [T]$. We first reformulate Problem (2.5) as follows:

$$
\begin{aligned}
\inf_{\mathbb{P} \in \mathcal{D}^o_T} W_2^2(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_T) &= \begin{cases} \inf_{\Pi, \mathbb{P}} \int_{\boldsymbol{\xi}, \boldsymbol{\xi}' \in \Xi} \mathrm{V}\, \boldsymbol{\xi} - \boldsymbol{\xi}' \|_2^2 \Pi(d\boldsymbol{\xi}, d\boldsymbol{\xi}') \\ \text{s.t.} \quad \Pi \in C(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_T), \\ \qquad\quad \mathbb{P} \in \mathcal{D}^o_T \end{cases} \\
&= \begin{cases} \inf_{\mathbb{P}_t \in \mathcal{P}(\Xi)} \frac{1}{T} \sum_{t=1}^T \int_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}^t\|_2^2 \mathbb{P}_t(d\boldsymbol{\xi}) \\ \text{s.t.} \qquad \frac{1}{T} \sum_{t=1}^T \mathbb{P}_t \in \mathcal{D}^o_T. \end{cases} \\
&= \begin{cases} \inf_{\mathbb{P}_t \in \mathcal{P}(\Xi)} \frac{1}{T} \sum_{t=1}^T \int_{\boldsymbol{\xi} \in \Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}^t\|_2^2 \mathbb{P}_t(d\boldsymbol{\xi}) \\ \text{s.t.} \qquad \frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\mathbb{P}_t} \left[\mathbb{1}(\xi_i \le q_i^{t'}) \right] = \frac{s_i - w_i^{t'}}{s_i}, \forall i \in [n], t' \in [T]. \end{cases}
\end{aligned}
$$

The second equality follows from the law of total probability (i.e., any joint probability distribution Π of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ can be constructed from the marginal distribution of $\boldsymbol{\xi}'$, which is the empirical distribution $\hat{\mathbb{P}}^{\mathrm{e}}_T$ in this case, and the conditional distribution $\mathbb{P}_t$ of $\boldsymbol{\xi}$ given $\boldsymbol{\xi}' = \boldsymbol{\xi}^t$ for all $t \in [T]$). The third equality follows from the definition of $\mathcal{D}^o_T$, where the constraints are the first-order conditions for the retailer's past ordering decisions.

Now we introduce Lagrange multipliers $(\mu_i^t)_{i\in[n],t\in[T]}$ for the first-order conditions, to obtain the dual problem

$$\sup_{\{\mu_i^t\}\in\mathbb{R}} \inf_{\{\mathbb{P}_t\}\in\mathcal{P}(\Xi)} \frac{1}{T}\sum_{t=1}^{T}\left\{\int_{\boldsymbol{\xi}\in\Xi}\|\boldsymbol{\xi}-\boldsymbol{\xi}^t\|_2^2\mathbb{P}_t(d\boldsymbol{\xi})+\sum_{i\in[n]}\sum_{t'=1}^{T}\mu_i^{t'}\left(\mathbb{E}_{\mathbb{P}_t}\left[\mathbb{1}(\xi_i\le q_i^{t'})\right]-\frac{s_i-w_i^{t'}}{s_i}\right)\right\}. \tag{A.26}$$

Because the Dirac measure satisfies $\delta\in\mathcal{P}(\Xi)$ and because the inner objective is linear in $\mathbb{P}_t$, the inner minimization of Problem (A.26) over all distributions in $\mathcal{P}(\Xi)$ is equivalent to minimization over elements in the support set Ξ. Thus, Problem (A.26) is equivalent to

$$\sup_{\{\mu_i^t\}\in\mathbb{R}} \frac{1}{T}\sum_{t=1}^{T}\inf_{\boldsymbol{\xi}\in\Xi}\left\{\|\boldsymbol{\xi}-\boldsymbol{\xi}^t\|_2^2+\sum_{i\in[n]}\sum_{t'=1}^{T}\mu_i^{t'}\left(\mathbb{1}(\xi_i\le q_i^{t'})-\frac{s_i-w_i^{t'}}{s_i}\right)\right\}. \tag{A.27}$$

We introduce auxiliary variables $(\gamma_t)_{t\in[T]}$ to reformulate Problem (A.27) as:

$$\begin{aligned}\sup_{\{\mu_i^t\}\in\mathbb{R},\{\gamma_t\}\in\mathbb{R}} \quad & -\sum_{i\in[n]}\sum_{t'=1}^{T}\mu_i^{t'}\frac{s_i-w_i^{t'}}{s_i}+\frac{1}{T}\sum_{t=1}^{T}\gamma_t,\\ \text{s.t.} \quad & \|\boldsymbol{\xi}-\boldsymbol{\xi}^t\|_2^2+\sum_{i\in[n]}\sum_{t'=1}^{T}\mu_i^{t'}\mathbb{1}(\xi_i\le q_i^{t'})\ge\gamma_t,\forall\boldsymbol{\xi}\in\Xi,t\in[T],\end{aligned} \tag{P}$$

which is a linear semi-infinite programming problem (LSIP).

We will leverage the duality theory for LSIP to establish strong duality for Problem (2.5). Take Problem (P) to be the primal. Problem (2.5) is the dual of Problem (P), and we refer to it as Problem (D). Let $\mathcal{Y}$ be the set of all bounded Borel measurable functions, and let $\mathcal{Y}^*$ be the set of all signed Borel measures on Ξ. The duality pairing between $\mathcal{Y}$ and $\mathcal{Y}^*$ is $\langle f,\nu\rangle\triangleq\int_\Xi f(\xi)\nu(d\xi)$. We also define the corresponding nonnegative cones $C^+(\mathcal{Y})=\{f\in\mathcal{Y}:f(\xi)\ge 0,\forall\xi\in\Xi\}$ and $C^+(\mathcal{Y}^*)=\{\nu\in\mathcal{Y}^*:\nu\ge 0\}$, which are dual to each other.

We use the shorthand $\tilde{\boldsymbol{c}}\triangleq\left(\left(-\frac{s_i-w_i^{t'}}{s_i}\right)_{i\in[n],t'\in[T]},1/T,\dots,1/T\right)$, $\boldsymbol{x}\triangleq\left(\left(\mu_i^{t'}\right)_{i\in[n],t'\in[T]},(\gamma_t)_{t\in[T]}\right)$, and $a_t(\boldsymbol{\xi})\triangleq\left(\left(\mathbb{1}(\xi_i\le q_i^{t'})\right)_{i\in[n],t'\in[T]},\mathbf{0}_{t-1},-1,\mathbf{0}_{T-t}\right)$, where $\mathbf{0}_t$ denotes a vector with t 0's. Also let $b_t(\boldsymbol{\xi})=-\|\boldsymbol{\xi}-\boldsymbol{\xi}^t\|_2^2$ for all $\boldsymbol{\xi}\in\Xi$ and $t\in[T]$. Then, we may rewrite Problem (P) as $\sup_{\boldsymbol{x}} \left\{ \tilde{\boldsymbol{c}}^\top \boldsymbol{x} : a_t(\boldsymbol{\xi})^\top \boldsymbol{x} \geq b_t(\boldsymbol{\xi}),\ \forall \boldsymbol{\xi} \in \Xi \text{ and } \forall t \in [T] \right\}$. Let us introduce the following cone where $(\mu(\boldsymbol{\xi}))_{\boldsymbol{\xi}\in\Xi}$ is a set of nonnegative multipliers:

$$M = \operatorname{cone}\{(a_t(\boldsymbol{\xi}), b_t(\boldsymbol{\xi})) : t \in [T],\ \boldsymbol{\xi} \in \Xi\} = \left\{ \boldsymbol{y} = \sum_{t\in[T],\boldsymbol{\xi}\in\Xi} \mu(\boldsymbol{\xi})\,(a_t(\boldsymbol{\xi}), b_t(\boldsymbol{\xi})) : \mu(\boldsymbol{\xi}) \in \mathbb{R}_+ \right\} \subset \mathbb{R}^{nT+T+1}.$$

We recall the following sufficient condition for strong duality in LSIP.

Theorem A.7.1. *([177, Theorem 6.5]) Suppose val(P) is finite and M is closed, then val(P) = val(D) and sol(D) is attained.*

Proof of Proposition A.5.1. By Theorem A.7.1, it is sufficient to show that val(P) is finite and M is closed. Finiteness of val(P) follows since, on one hand, val(P) $> -\infty$ because (P) is always feasible. On the other hand, by weak duality, val(P) $\leq$ val(D) and val(D) $< \infty$ since (D) is always feasible (in particular, $\mathcal{D}_T^o$ is nonempty since there is at least one distribution in this set, i.e., $\bar{\mathbb{P}} \in \mathcal{D}_T^o$).

Now we verify that M is closed. First notice that the set $\{a_t(\boldsymbol{\xi}) : t \in [T], \xi \in \Xi\}$ is finite. Since Ξ is compact and $b_t(\cdot)$ is continuous, the set $\{b_t(\boldsymbol{\xi}) : t \in [T], \boldsymbol{\xi} \in \Xi\}$ is also compact since the continuous image of a compact set is compact. Thus, M is closed.

□

A.7.4 Proof of Proposition A.5.2 (explicit form of the center distribution)

We first prove an auxiliary result (See Lemma A.7.3). In the following, we adopt the conventions of extended arithmetic where $\infty \times 0 = 0 \times \infty = 0/0 = 0$ and $\infty - \infty = -\infty + \infty = 1/0 = \infty$.

Lemma A.7.3. *Suppose $a \geq 0$, then the system $\{a\boldsymbol{\xi}^\top\boldsymbol{\xi} + \boldsymbol{b}^\top\boldsymbol{\xi} + c \geq 0, \forall \boldsymbol{\xi} : A\boldsymbol{\xi} \leq \boldsymbol{d}\}$ is feasible if and only if there exists $\boldsymbol{\tau} \leq 0$ such that $-(A^\top\boldsymbol{\tau} - \boldsymbol{b})^\top(A^\top\boldsymbol{\tau} - \boldsymbol{b})/(4a) + \boldsymbol{\tau}^\top\boldsymbol{d} + c \geq 0$.*

Proof of Lemma A.7.3. When $a > 0$, the proof directly follows from the duality theory for quadratic programming problems with linear constraints.

When $a=0$, on one hand, by the conventions of extended arithmetic, $-(A^\top\boldsymbol{\tau}-\boldsymbol{b})^\top(A^\top\boldsymbol{\tau}-\boldsymbol{b})/(4a)+\boldsymbol{\tau}^\top\boldsymbol{d}+c\geq 0$ implies that $\boldsymbol{\tau}^\top\boldsymbol{d}+c\geq 0$ and $A^\top\boldsymbol{\tau}-\boldsymbol{d}=0$, since otherwise $-(A^\top\boldsymbol{\tau}-\boldsymbol{b})^\top(A^\top\boldsymbol{\tau}-\boldsymbol{b})/(4a)+\boldsymbol{\tau}^\top\boldsymbol{d}+c=-\infty$. At the same time, the system $\{a\boldsymbol{\xi}^\top\boldsymbol{\xi}+\boldsymbol{b}^\top\boldsymbol{\xi}+c\geq 0, \forall\boldsymbol{\xi}: A\boldsymbol{\xi}\leq\boldsymbol{d}\}$ reduces to $\{\boldsymbol{b}^\top\boldsymbol{\xi}+c\geq 0,\forall\boldsymbol{\xi}: A\boldsymbol{\xi}\leq\boldsymbol{d}\}$, which is satisfied, by the duality theory for linear programming problems, if and only if there exists $\boldsymbol{\tau}\leq 0$ such that $\boldsymbol{\tau}^\top\boldsymbol{d}+c\geq 0$ and $A^\top\boldsymbol{\tau}-\boldsymbol{b}=0$, which gives the desired result. □

Proof of Proposition A.5.2. We prove the Proposition by establishing strong duality between Problems (A.17) and (2.17). Recall n_s is the number of linear inequalities characterizing the polyhedron $R_s=\{\boldsymbol{\xi}: A_s\boldsymbol{\xi}\leq\boldsymbol{d}_s\}$, i.e., $A_s\in\mathbb{R}^{n_s\times n}$ and $\boldsymbol{d}_s\in\mathbb{R}^{n_s}$. It follows from Lemma A.7.3 and Shur's complement that Problem (A.17) is equivalent to

$$
\begin{aligned}
&\sup_{\boldsymbol{\mu},\boldsymbol{\gamma},\boldsymbol{\tau}_{st}} \frac{1}{T}\sum_{t=1}^{T}\gamma_t,\\
&\text{s.t.} \begin{bmatrix} I_{n\times n} & -\left(A_s^\top\boldsymbol{\tau}+2\boldsymbol{\xi}^t\right)/2\\ -\left(A_s^\top\boldsymbol{\tau}+2\boldsymbol{\xi}^t\right)^\top/2 & \boldsymbol{\tau}_{st}^\top\boldsymbol{d}_s+\boldsymbol{\xi}^{t\top}\boldsymbol{\xi}^t+\sum_{i\in[n]}\sum_{t'=1}^{T}\mu_i^{t'}\left(z_{is}(\boldsymbol{q}^{t'})-\frac{s_i-w_i^{t'}}{s_i}\right)-\gamma_t\end{bmatrix}\succeq 0,\\
&\qquad\qquad\qquad \forall t\in[T], s\in[S],\\
&\quad\boldsymbol{\tau}_{st}\leq 0, \forall s\in[S], t\in[T],
\end{aligned}
\tag{A.28}
$$

where $I_{n\times n}$ stands for the identity matrix with size $n \times n$. The following Lagrangian dual to Problem (A.17) then follows from SDP duality:

$$
\begin{aligned}
\inf \quad & \sum_{t=1}^{T}\sum_{s=1}^{S} I_{n\times n} \bullet Q_{st} - 2\boldsymbol{p}_{st}^\top \boldsymbol{\xi}^t + \beta_{st}\boldsymbol{\xi}^{t\top}\boldsymbol{\xi}^t \\
\text{s.t.} \quad & \sum_{s=1}^{S} \beta_{st} = 1/T, \forall t \in [T], \\
& \sum_{t=1}^{T}\sum_{s=1}^{S} \beta_{st} z_{is}(\boldsymbol{q}^{t'}) = \frac{s_i - w_i^{t'}}{s_i}, \forall i \in [n], t' \in [T], \\
& A_s \boldsymbol{p}_{st} \leq \beta_{st}\boldsymbol{d}_s, \forall t \in [T], s \in [S], \\
& \begin{bmatrix} Q_{st} & \boldsymbol{p}_{st} \\ \boldsymbol{p}_{st}^\top & \beta_{st} \end{bmatrix} \succeq 0, \forall t \in [T], s \in [S], \\
& Q_{st} \in \mathbb{R}^{n\times n}, \boldsymbol{p}_{st} \in \mathbb{R}^n, \beta_{st} \in \mathbb{R},
\end{aligned}
\tag{A.29}
$$

where $\bullet$ is the Frobenius inner product between matrices, and $Q_{st} \in \mathbb{R}^{n\times n}$, $\boldsymbol{p}_{st} \in \mathbb{R}^n$, and $\beta_{st} \in \mathbb{R}$ are collectively the (matrix-valued) Lagrange multiplier to the first constraint of Problem (A.17).

We will simplify the dual problem by showing that $Q_{st}^* = \boldsymbol{p}_{st}^*\boldsymbol{p}_{st}^{*\top}/\beta_{st}^*$ for all $t \in [T]$ and $s \in [S]$. In particular, we have $Q_{st}^* = 0$ when $\beta_{st}^* = 0$, since $\boldsymbol{p}_{st}^* = 0$ and $0/0 = 0$. Indeed, $\beta_{st}^* = 0$ implies $\boldsymbol{p}_{st}^* = 0$ and $Q_{st}^* = 0$ as $Q_{st} \succeq 0$ must hold. When $\beta_{st}^* > 0$, Shur's complement shows that $Q_{st} \succeq \boldsymbol{p}_{st}\boldsymbol{p}_{st}^\top/\beta_{st}$ and so $Q_{st}^* = \boldsymbol{p}_{st}^*\boldsymbol{p}_{st}^{*\top}/\beta_{st}^*$ is an optimal solution. Substituting $Q_{st} = \boldsymbol{p}_{st}\boldsymbol{p}_{st}^\top/\beta_{st}$ in the objective of Problem (A.29), we obtain $\sum_{t=1}^{T}\sum_{s=1}^{S}\boldsymbol{p}_{st}^\top\boldsymbol{p}_{st}/\beta_{st} -$

$2\boldsymbol{p}_{st}^\top\boldsymbol{\xi}^t + \beta_{st}\boldsymbol{\xi}^{t\top}\boldsymbol{\xi}^t = \sum_{t=1}^T\sum_{s=1}^S \beta_{st}\left(\boldsymbol{p}_{st}/\beta_{st} - \boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{p}_{st}/\beta_{st} - \boldsymbol{\xi}^t\right)$, and so Problem (A.29) is equivalent to Problem (2.17) restated as follows:

$$
\begin{aligned}
\inf \quad & \sum_{t=1}^T\sum_{s=1}^S \beta_{st}\left(\boldsymbol{p}_{st}/\beta_{st} - \boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{p}_{st}/\beta_{st} - \boldsymbol{\xi}^t\right) \\
\text{s.t.} \quad & \sum_{s=1}^S \beta_{st} = 1/T, \forall t \in [T], \\
& \sum_{t=1}^T\sum_{s=1}^S \beta_{st} z_{is}(\boldsymbol{q}^{t'}) = \frac{s_i - w_i^{t'}}{s_i}, \forall i \in [n], t' \in [T], \\
& A_s \boldsymbol{p}_{st}/\beta_{st} \le \boldsymbol{d}_s, \forall t \in [T], s \in [S], \\
& \beta_{st} \ge 0, \boldsymbol{p}_{st} \in \mathbb{R}^n, \forall t \in [T], s \in [S].
\end{aligned}
$$

It is immediate that the relative interior of the feasible region of Problem (A.17) is non-empty as $\boldsymbol{\gamma}$ is a free variable. Then, based on the weaker version of [178, Proposition 3.4], strong duality holds between Problems (A.17) and (2.17). □

A.7.5 Proof of Theorem A.3.1 (convergence rate for the center distribution)

We first cite the following result on convergence with respect to the 2−Wasserstein distance.

Theorem A.7.2 ([179]). *Suppose there exists $a > 1$ such that $\mathbb{E}_{\bar{\mathbb{P}}}[\exp(\|\boldsymbol{\xi}\|_p^a)] < \infty$. Then,*

$$
\bar{\mathbb{P}}^T\left\{W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T^{\mathrm{e}}) \ge \theta\right\} \le \begin{cases} C_1\exp(-C_2T\theta^{\max\{n,2\}}), \theta \le 1, \\ C_1\exp\left(-C_2T\theta^a\right), \theta > 1, \end{cases}
$$

for all $T \ge 1$, and $\theta > 0$, where C_1, C_2 are positive constants that only depend on a and n.

Theorem A.7.2 requires the tail of the distribution $\bar{\mathbb{P}}$ to decay at an exponential rate. This condition is automatically satisfied when Ξ is compact. We also notice that Theorem A.7.2 holds when the Wasserstein ball is centered at the empirical distribution $\hat{\mathbb{P}}_T^{\mathrm{e}}$. In Section 2.7, when Assumption 2.7.1 holds, we constructed the center distribution $\hat{\mathbb{P}}_T$ as the projection of $\hat{\mathbb{Q}}^{\mathrm{e}}_T$ onto $\mathcal{D}_\times$. We obtain an analogous result to Theorem A.7.2 when the Wasserstein ball is centered at $\hat{\mathbb{Q}}^{\mathrm{e}}_T$ and when Assumption 2.7.1 holds.

Theorem A.7.3. *Suppose Assumptions 2.3.1 and 2.7.1 hold. Then,*

$$\bar{\mathbb{P}}^T\left\{W_2(\bar{\mathbb{P}}, \hat{\mathbb{Q}}^{\mathrm{e}}_T) \geq \theta\right\} \leq \begin{cases} c_1 \exp(-c_2 T\theta^2), \theta \leq 1, \\ c_1 \exp\left(-c_2 T\theta^a\right), \theta > 1, \end{cases}$$

for all $T \geq 1$, and $\theta > 0$, where c_1, c_2 and a are positive constants that only (possibly) depend on n.

Proof of Theorem A.7.3. Under Assumption 2.7.1, we have $\bar{\mathbb{P}} = \times_{i\in[n]}\bar{\mathbb{P}}^i$. Furthermore, we have $\hat{\mathbb{Q}}^{\mathrm{e}}_T = \times_{i\in[n]}\hat{\mathbb{P}}^{\mathrm{e},i}_T$ by construction. Then we have

$$\begin{aligned} \bar{\mathbb{P}}^T\{W_2(\bar{\mathbb{P}}, \hat{\mathbb{Q}}^{\mathrm{e}}_T) \geq \theta\} &= \bar{\mathbb{P}}^T\left\{\sum_{i\in[n]} W_2^2(\bar{\mathbb{P}}^i, \hat{\mathbb{P}}^{\mathrm{e},i}_T) \geq \theta^2\right\} \\ &\leq \bar{\mathbb{P}}^T\left\{W_2^2(\bar{\mathbb{P}}^i, \hat{\mathbb{P}}^{\mathrm{e},i}_T) \geq \theta^2/n \text{ for some } i \in [n]\right\} \\ &\leq \sum_{i\in[n]} \bar{\mathbb{P}}^T\left\{W_2(\bar{\mathbb{P}}^i, \hat{\mathbb{P}}^{\mathrm{e},i}_T) \geq \theta/\sqrt{n}\right\} \\ &\leq \begin{cases} c_1 \exp\left(-c_2 T\theta^2\right), \theta \leq 1, \\ c_1 \exp\left(-c_2 T\theta^a\right), \theta > 1, \end{cases} \end{aligned}$$

where c_1 and c_2 and a are positive constants that only (possibly) depend on n. In the above equations, the first equality follows from Lemma A.5.3, the second inequality follows from union law of probability, and the last inequality follows from Theorem A.7.2 for $n = 1$. $\square$

Proof of Theorem A.3.1. We first prove Theorem A.3.1 when the center distribution satisfies $\hat{\mathbb{P}}_T \in \arg\min_{\mathbb{P}\in\mathcal{D}^o_T} W_2^2(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_T)$. Since W_2 is a metric, we have $W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}^{\mathrm{e}}_T) + W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \hat{\mathbb{P}}_T) = W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \bar{\mathbb{P}}) + W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \hat{\mathbb{P}}_T)$, by the triangle inequality and symmetry. By construction, $\hat{\mathbb{P}}_T$ minimizes $W_2(\mathbb{P}, \hat{\mathbb{P}}^{\mathrm{e}}_T)$ over all $\mathbb{P} \in \mathcal{D}^o_T$, and so $W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \hat{\mathbb{P}}_T) \leq W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \bar{\mathbb{P}})$. Therefore, we must have $W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \leq 2W_2(\hat{\mathbb{P}}^{\mathrm{e}}_T, \bar{\mathbb{P}})$. It follows that $\bar{\mathbb{P}}^T\{W_2(\bar{\mathbb{P}}, \hat{\mathbb{P}}_T) \geq \theta\} \leq$

$\bar{\mathbb{P}}^T\{W_2(\hat{\mathbb{P}}^{\rm e}_T, \bar{\mathbb{P}}) \geq \theta/2\}$. Similarly, when $\hat{\mathbb{Q}}_T \in \arg\min_{\mathbb{P}\in\mathcal{D}^o_T} W_2^2(\mathbb{P}, \hat{\mathbb{Q}}^{\rm e}_T)$, we have $\bar{\mathbb{P}}^T\{W_2(\bar{\mathbb{P}}, \hat{\mathbb{Q}}_T) \geq \theta\} \leq \bar{\mathbb{P}}^T\{W_2(\hat{\mathbb{Q}}_T, \hat{\mathbb{Q}}^{\rm e}_T) \geq \theta/2\}$, and the desired result follows.

□

B. APPENDICES FOR CHAPTER 3

B.1 Additional material for Section 3.4

B.1.1 Proof of Theorem 3.4.1

We have

$$
\begin{aligned}
\operatorname{Reg}(\pi_{\text{stat}}, T) &= \sum_{t=1}^{T} \mathbb{E}\left[(w^* - c)q(w^*; F_0) - (w_t - c)q(w_t; F_0)\right] \\
&= \sum_{t=1}^{\lceil\sqrt{T}\rceil} \mathbb{E}\left[(w^* - c)q(w^*; F_0) - (w_t - c)q(w_t; F_0)\right] \\
&+ \sum_{t=\lceil\sqrt{T}\rceil+1}^{T} \mathbb{E}\left[(w^* - c)q(w^*; F_0) - (w^*_{stat} - c)q(w^*_{stat}; F_0)\right] \\
&\leq \left\lceil\sqrt{T}\right\rceil s\bar{\xi} + (T - \left\lceil\sqrt{T}\right\rceil)s\bar{\xi}/\left\lceil\sqrt{T}\right\rceil \\
&= O(\sqrt{T}),
\end{aligned}
$$

where the first inequality follows from Eq. (3.6) and the fact that $|w^* - w^*_{stat}| \leq s/\left\lceil\sqrt{T}\right\rceil$ by the discretization of $\mathcal{W}$.

B.2 Additional material for Section 3.5

B.2.1 Proof of Lemma 3.5.1

Suppose $q(w^t_m; \hat{F}^\mu_t) \geq y_m$, then it follows from Eq. (3.7) that $p_{t,m-1} \leq 1 - w^t_m/s$, and $\varphi(w^t_m; \hat{F}^\mu_t) \geq (w^t_m - c)y_m = \varphi_{k^*} + \Delta_t + \frac{y_m s}{K}$. The previous inequality follows from the fact that $q(w^t_m; \hat{F}^\mu_t) \geq y_m$ and the previous equality follows by construction of w^t_m (see Eq. (3.12)). The proof is similar for the case $q(w^t_m; \hat{F}^\mu_t) < y_m$.

B.2.2 Proof of Lemma 3.5.2

We will show that if $q(w^t_m; \hat{F}^\mu_t) \geq y_m$, then

$$
\max_{k\in[K]} \max_{x\in[0,\bar{\xi}]} |\hat{F}^\mu_t(x) - \hat{F}^\mu_{\tau^0_i+k}(x)| \geq \Delta_t/(s\,\bar{\xi}). \tag{B.1}
$$

It will then follow from the reverse triangle inequality that the total variation over epoch i is:

$$\sum_{t=\tau_i^0+1}^{\tau_{i+1}^0-1} \max_{x\in[0,\bar{\xi}]} |\hat{F}^\mu_{t+1}(x) - \hat{F}^\mu_t(x)| \geq \Delta_t/(s\bar{\xi}).$$

Recall $(p_{\tau_i^0+k,m})_{k\in[K]}$ are the values of $\hat{F}^\mu_t$ at y_m during the exploration phase in epoch i (i.e., for periods $t \in [\tau_i^0+1, \tau_i^0+K]$). To show Eq. (B.1), we start with the following Lemma B.2.1.

Lemma B.2.1. *During the exploration phase of epoch i, we must have*

$$\max_{k\in[K]} p_{\tau_i^0+k,m-1} \geq 1 - \frac{\varphi_{k^*}/y_m + c}{s} - \frac{1}{K}, \tag{B.2}$$

for all $m \in [M]$ with $m \neq m^*$, where the term $\frac{1}{K}$ is due to the discretization error.

Proof of Lemma B.2.1. For a contradiction, suppose there exists some $m' \in [M]$ with $m' \neq m^*$ such that

$$s(1 - p_{\tau_i^0+k,m'}) > \varphi_{k^*}/y_{m'} + c + \frac{s}{K}, \ \forall k \in [K]. \tag{B.3}$$

We let $k' \in [K]$ be such that $\bar{w}_{k'} \in [s(1-p_{\tau_i^0+k',m'-1}) - \frac{s}{K}, s(1-p_{\tau_i^0+k',m'-1})]$ (notice that such k' must exist by the discretization). We then have $\varphi(\bar{w}_{k'}; \hat{F}^\mu_{\tau_i^0+k'}) = (\bar{w}_{k'} - c)q(\bar{w}_{k'}; \hat{F}^\mu_{\tau_i^0+k'}) \geq (\bar{w}_{k'} - c)y_{m'} \geq (s - s\,p_{\tau_i^0+k,m'-1} - c - \frac{s}{K})y_{m'} > \varphi_{k^*}$ where the first inequality follows since $\bar{w}_{k'} \leq s(1 - p_{\tau_i^0+k',m'-1})$ and then according to Eq. (3.9), $q(\bar{w}_{k'}; \hat{F}^\mu_{\tau_i^0+k}) \geq y_{m'}$. The second inequality follows since $\bar{w}_{k'} \geq s(1 - p_{\tau_i^0+k',m'-1}) - \frac{s}{K}$ and the last inequality follows from Eq. (B.3). Thus we have shown that Eq. (B.3) is a contradiction of the fact that $k^* \in \arg\max_{k\in[K]} \varphi_k$. $\square$

On the other hand, since $q(w^t_m; F_t) \geq y_m$ for some $m \in [M]$, we have

$$p_{t,m-1} \leq 1 - w^t_m/s = 1 - \frac{(\varphi_{k^*} + \Delta_t)/y_m + 1/K + c}{s}, \tag{B.4}$$

where the inquality follows from Lemma 3.5.1 and the equality follows from Eq. (3.12). Combining Eq. (B.2) and Eq. (B.4) gives:

$$\max_{k\in[K]}(p_{\tau_i^0+k,m-1} - p_{t,m-1}) \geq \Delta_t/(s\,y_m) \geq \Delta_t/(s\bar{\xi}).$$

It follows that $\max_{k\in[K]} \max_{x\in[0,\bar{\xi}]} |\hat{F}^{\mu}_{\tau_0^i+k}(x) - \hat{F}^{\mu}_t(x)| \geq \Delta_t/(s\,\bar{\xi})$, and the Lemma holds.

B.2.3 Proof of Lemma 3.5.3

There are two cases: (i) $\bar{w}_{k^*} - \Delta_t/y_{m^*} \geq 0$; and (ii) $\bar{w}_{k^*} - \Delta_t/y_{m^*} < 0$. In the first case where $\bar{w}_{k^*} - \Delta_t/y_{m^*} \geq 0$, we have $w_0^t = \bar{w}_{k^*} - \Delta_t/y_{m^*}$ and the proof is similar to Lemma 3.5.1. In the second case where $\bar{w}_{k^*} - \Delta_t/y_{m^*} < 0$, $w_0^t = 0$ and we have $q(0; \hat{F}^{\mu}_t) = \bar{\xi} \geq y_{m^*}$ (the retailer will order as much as possible since the order cost is zero). In this case, we have $p_{t,m^*-1} \leq 1 - w_0^t/s = 1$ (by Eq. (3.7)) and $\varphi(w_0^t; \hat{F}^{\mu}_t) \geq (w_0^t - c)y_{m^*} \geq (\bar{w}_{k^*} - \Delta_t/y_{m^*} - c)y_{m^*} = \varphi_{k^*} - \Delta_t$.

B.2.4 Proof of Lemma 3.5.4

When $q(w_0^t; \hat{F}^{\mu}_t) < y_{m^*}$, we must have $w_0^t > 0$. Otherwise, if $w_0^t = 0$, the retailer will always order $q(0; \hat{F}^{\mu}_t) = \bar{\xi}$ since the order cost is zero. Thus we can restrict to $\bar{w}_{k^*} - \Delta_t/y_{m^*} > 0$.

We will show that if $q(w_0^t; \hat{F}^{\mu}_t) < y_{m^*}$, then

$$\max_{k\in[K]} \max_{x\in[0,\bar{\xi}]} |\hat{F}^{\mu}_{\tau_i^0+k}(x) - \hat{F}^{\mu}_t(x)| \geq \Delta_t/(s\,\bar{\xi}),$$

and consequently

$$\sum_{t=\tau_i^0+1}^{\tau_{i+1}^0-1} \max_{x\in[0,\bar{\xi}]} |\hat{F}^{\mu}_t(x) - \hat{F}^{\mu}_{t+1}(x)| \geq \Delta_t/(s\,\bar{\xi}).$$

First notice that according to the policy implementation, we have (by the relation $y_{m^*} = q(\bar{w}_{k^*}; \hat{F}^{\mu}_{\tau_i^0+k^*})$ and Eq. (3.9)) that:

$$p_{\tau_i^0+k^*,m^*-1} < 1 - \bar{w}_{k^*}/s. \tag{B.5}$$

Next, since $q(w_0^t; \hat{F}^{\mu}_t) < y_{m^*}$, we have

$$p_{t,m^*-1} \geq 1 - w_0^t/s = 1 - (\bar{w}_{k^*} - \Delta_t/y_{m^*})/s = 1 - \bar{w}_{k^*}/s + \Delta_t/(s\, y_{m^*}), \tag{B.6}$$

where the first inequality follows from Lemma 3.5.3 and the first equality follows from Eq. (3.13). Combining Eq. (B.5) and Eq. (B.6) gives

$$p_{t,m^*-1} - p_{\tau_i^0+k^*,m^*-1} \geq \Delta_t/(s\, y_{m^*}) \geq \Delta_t/(s\, \bar{\xi}),$$

and it follows that $\max_{k\in[K]} \max_{x\in[0,\bar{\xi}]} |\hat{F}^{\mu}_{\tau_0^i+k}(x) - \hat{F}^{\mu}_t(x)| \geq \Delta_t/(s\, \bar{\xi})$.

B.2.5 Proof of Theorem 3.5.1

Here we complete the details of the proof of Theorem 3.5.1. We use superscript i to denote quantities corresponding to epoch i since those quantities vary from epoch to epoch. That is, we use φ_k^i to denote the profit observed during period $k \in [K]$ of the exploration phase of epoch i, and $\varphi_{k^*}^i$ to denote the optimal profit observed during the exploration phase in epoch i.

Abusing notation, let $\mathrm{Reg}(\pi, \hat{F}^{\mu}_{t_1:t_2}) \triangleq \sum_{t=t_1}^{t_2} \varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_t; \hat{F}^{\mu}_t)$ be the regret incurred from periods $t \in [t_1, t_2]$ when the supplier follows π, given the sequence of perceived distributions $\hat{F}^{\mu}_{t_1:t_2}$. The regret incurred in epoch i (for periods $t \in [\tau_i^0 + 1, \tau_{i+1}^0]$) for π_{LUNA} is then:

$$\mathrm{Reg}\left(\pi_{\mathrm{LUNA}}, \hat{F}^{\mu}_{\tau_i^0+1:\tau_{i+1}^0}\right) = \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_0^t; \hat{F}^{\mu}_t) + \varphi(w_0^t; \hat{F}^{\mu}_t) - \varphi(w_t; \hat{F}^{\mu}_t).$$

We will bound the first part $\mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}}) \triangleq \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)$ and the second part $\mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}}) \triangleq \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu)$ separately.

Part I of the regret

To bound the first part $\mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}})$, recall the set (for epoch i):

$$\mathcal{E}^i = \Big\{t \in [\tau_i^0 + \max\{M+2, K\} + 1, \tau_{i+1}^0] : q(w_m^t; \hat{F}_t^\mu) < y_m, \forall m \in [M], \\ \text{and } q(w_0^t; \hat{F}_t^\mu) \geq y_{m^*}\Big\}.$$

Then, we have

$$\begin{aligned}
&\mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}}) \\
\leq &\sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} \Big(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\Big) + s\bar{\xi}\max\{M+2, K\} \\
\leq &\sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} \Bigg[\Big(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\Big)\mathbb{1}(t \notin \mathcal{E}^i) + \Big(\varphi(w_t^*; \hat{F}_t^\mu) \\
&\qquad - \varphi(w_0^t; \hat{F}_t^\mu)\Big)\mathbb{1}(t \in \mathcal{E}^i)\Bigg] + s\bar{\xi}\max\{M+2, K\} \\
\leq &\sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} \Big[s\bar{\xi}\mathbb{1}(t \notin \mathcal{E}^i) + \Big(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\Big)\mathbb{1}(t \in \mathcal{E}^i)\Big] \\
&\qquad + s\bar{\xi}\max\{M+2, K\}.
\end{aligned} \tag{B.7}$$

It then follows from Eq. (B.7), Lemma 3.5.5, and Lemma 3.5.6 that

$$
\begin{aligned}
&\mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}}) \\
\leq &\sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} \left[s\bar{\xi}\mathbb{1}(t \notin \mathcal{E}^i) + \left(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu)\right)\mathbb{1}(t \in \mathcal{E}^i)\right] \\
&\qquad + s\bar{\xi}\max\{M+2, K\} \\
\leq &\, s\bar{\xi}\left[\max\{M+2, K\} + 1 + 2\log(T)\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)}\right] \\
&\qquad + \sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} 2\Delta_t + (\tau_{i+1}^0 - 1 - \tau_i^0)s\bar{\xi}\left(\frac{1}{K}\right) \\
\leq &\, s\bar{\xi}\left[M + K + 3 + 2\log(T)\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)}\right] + 4\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)} \\
&\qquad + (\tau_{i+1}^0 - 1 - \tau_i^0)s\bar{\xi}\left(\frac{1}{K}\right)
\end{aligned}
\tag{B.8}
$$

where the last inequality follows since

$$
\sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} 2\Delta_t = \sum_{t=\tau_i^0+\max\{M+2,K\}+1}^{\tau_{i+1}^0} 2\sqrt{M/(t-\tau_i^0)} \leq 4\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)}.
$$

Part II of the regret

For the second part of the regret $\mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}})$, we have the equivalence

$$
\begin{aligned}
\mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}}) &= \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0} \varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu) \\
&= \left(\varphi(w_0^t; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu)\right)\mathbb{1}(w_t \neq w_0^t) \\
&\leq \bar{\xi}s\left[K + T_i(K)\right].
\end{aligned}
$$

Then, by Lemma B.2.3 we have

$$\mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}}) \leq \bar{\xi}s\left[K + \sqrt{11\log(T)M(\tau_{i+1}^0 - 1 - \tau_i^0)}\right], \tag{B.9}$$

with probability at least $1 - 1/T^2$.

Combining the two parts of the regret

Combining Eqs. (B.8) and (B.9), and using the union bound, we have with probability at least $1 - 1/T$ that (where we let I be the total number of epochs)

$$\begin{aligned}
&\mathrm{Reg}\left(\pi_{\mathrm{LUNA}}, \hat{F}^{\mu}_{1:T}\right) \\
&= \sum_{t=1}^{T} \varphi(w_t^*; \hat{F}_t^{\mu}) - \varphi(w_0^t; \hat{F}_t^{\mu}) + \varphi(w_0^t; \hat{F}_t^{\mu}) - \varphi(w_t; \hat{F}_t^{\mu}) \\
&= \sum_{i=1}^{I} \mathrm{Reg}_i^c(\pi_{\mathrm{LUNA}}) + \sum_{i=1}^{I} \mathrm{Reg}_i^0(\pi_{\mathrm{LUNA}}) \\
&\leq s\bar{\xi}(M + 2K + 3)I + \left(2s\bar{\xi}\log(T) + 4\right) \sum_{i=1}^{I} \sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)} + \sum_{i=1}^{I} (\tau_{i+1}^0 - 1 - \tau_i^0) s\bar{\xi}\frac{1}{K} \\
&\qquad + \sum_{i=1}^{I} \bar{\xi}s\sqrt{11\log(T)M(\tau_{i+1}^0 - 1 - \tau_i^0)} \\
&\leq s\bar{\xi}(M + 2K + 3)I + \left(2s\bar{\xi}\log(T) + 4 + \bar{\xi}\sqrt{11\log(T)}\right)\left(\sum_{i=1}^{I} \sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)}\right) + \frac{T}{K}.
\end{aligned} \tag{B.10}$$

It then follows that

$$
\begin{aligned}
&\mathrm{Reg}(\pi_{\mathrm{LUNA}}, T) \\
&= \sup_{\mu \in \mathcal{M}(V,T)} \mathbb{E}\left[\mathrm{Reg}\left(\pi_{\mathrm{LUNA}}, \hat{F}^{\mu}_{1:T}\right)\right] \\
&\leq s\bar{\xi}(M+2K+3)I + \left(2s\bar{\xi}\log(T) + 4 + \bar{\xi}\sqrt{11\log(T)}\right)\left(\sum_{i=1}^{I}\sqrt{M(\tau^0_{i+1} - 1 - \tau^0_i)}\right) + \frac{T}{K} \\
&\leq s\bar{\xi}(M+2K+3)I + \left(2s\bar{\xi}\log(T) + 4 + \bar{\xi}\sqrt{11\log(T)}\right)\sqrt{MTI} + \frac{T}{K}. \\
&= \tilde{O}\left(\bar{\xi}^{\frac{4}{3}}V^{\frac{1}{3}}M^{\frac{1}{3}}T^{\frac{2}{3}} + \frac{\bar{\xi}T}{K} + \bar{\xi}^{\frac{5}{3}}KV^{\frac{2}{3}}M^{-\frac{1}{3}}T^{\frac{1}{3}}\right),
\end{aligned}
\tag{B.11}
$$

where the second inequality follows by Jensen's inequality (using $\sum_{i=1}^{T}(\tau^0_{i+1} - \tau^0_i - 1) = T$), and the last equality follows by Lemma 3.5.8. The rest of the argument follows by setting $K^* = \left\lceil T^{\frac{1}{3}}V^{-\frac{1}{3}}\bar{\xi}^{-\frac{1}{3}}\right\rceil$ and $\hat{K} = \left\lceil \bar{\xi}^{-\frac{1}{3}}T^{\frac{1}{3}}\right\rceil$.

B.2.6 Proof of Lemma 3.5.5

Since $t \in \mathcal{E}^i \cap [\tau^0_i + \max\{M+2, K\} + 1, \tau^0_{i+1}]$, we have by Lemma 3.5.3 that

$$\varphi(w^t_0; \hat{F}^{\mu}_t) \geq (w^t_0 - c)y_{m^*} \geq \varphi^i_{k^*} - \Delta_t.$$

At the same time, by Eq. (3.9), the optimal supplier profit in period t satisfies:

$$\varphi(w^*_t; \hat{F}^{\mu}_t) = \max_{m \in [M-1]} (s - s\,p_{t,m} - c)y_{m+1}.$$

We then have

$$
\begin{aligned}
\varphi(w^*_t; \hat{F}^{\mu}_t) &= \max_{m \in [M-1]} (s - s\,p_{m,t} - c)y_{m+1} \\
&\leq \max_{m \in [M-1]} (w^t_{m+1} - c)y_{m+1} \\
&= \Delta_t + \frac{\bar{\xi}s}{K} + \varphi^i_{k^*},
\end{aligned}
$$

where the inequality follows from Lemma 3.5.1. Therefore,

$$\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_0^t; \hat{F}_t^\mu) = \varphi(w_t^*; \hat{F}_t^\mu) - \varphi_{k^*}^i + \varphi_{k^*}^i - \varphi(w_0^t; \hat{F}_t^\mu) \leq 2\Delta_t + \frac{\bar{\xi} s}{K}.$$

B.2.7 Proof of Lemma 3.5.6

We make use of the following supporting result.

Lemma B.2.2. *Let $(t_n)_{n=1}^N \in [\tau_i^0 + \max\{M+2, K\} + 1, \tau_{i+1}^0 - 1]$ for $N \geq 1$ be the sequence of periods where $t_n \in \mathcal{E}^i$. Let*

$$s = \arg\min_{s' \geq 1} \left\{ \sum_{n=1}^{s'} \frac{1}{\sqrt{M(t_n - \tau_i^0)}} \geq 2\log(T) \right\},$$

then $E^i \leq s$ with probability at least $1 - 1/T^2$.

The proof of Lemma B.2.2 follows from Lemma B.2.3 below and [65, Theorem A.4].

Lemma B.2.3. *If $t \notin \mathcal{E}^i$ and $t \geq \tau_i^0 + \max\{M+2, K\} + 1$, then epoch i ends in period t with probability at least $\sqrt{\frac{1}{M(t-\tau_i^0)}}$ given that epoch i has not ended before period t.*

Proof of Lemma B.2.3. If $t \notin \mathcal{E}^i$, then either $q(w_m^t; \hat{F}_t^\mu) > y_m$ for some $m \in [M]$ holds, or $q(w_0^t; \hat{F}_t^\mu) < y_{m^*}$ holds. In the first case, where $q(w_m^t; \hat{F}_t^\mu) > y_m$ for some $m \in [M]$, epoch i will end if $m_t = m$ which occurs with probability $\frac{1}{M}\sqrt{\frac{M}{t-\tau_i^0}} = \sqrt{\frac{1}{M(t-\tau_i^0)}}$. In other words, according to the algorithm implementation, with probability $\sqrt{\frac{1}{M(t-\tau_i^0)}}$ we choose $m_t = m$ and since $q(w_m^t; \hat{F}_t^\mu) > y_m$, we end the current epoch according to the pricing policy. In the second case, if $q(w_0^t; \hat{F}_t^\mu) < y_{m^*}$, then epoch i will end if $m_t = 0$ which occurs with probability $1 - \sqrt{\frac{M}{t-\tau_i^0}} \geq \sqrt{\frac{1}{M(t-\tau_i^0)}}$. In either case, epoch i will end with probability at least $\sqrt{\frac{1}{M(t-\tau_i^0)}}$ for all $t \geq \tau_i^0 + \max\{M+2, K\} + 1$ with $t \notin \mathcal{E}^i$.

□

Proof of Lemma 3.5.6. It follows from Lemma B.2.2 that with probability at least $1 - 1/T^2$, we have

$$\frac{E^i - 1}{\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)}} \leq \sum_{n=1}^{E^i - 1} \frac{1}{\sqrt{M(t_n - \tau_i^0)}} = \sum_{n=1}^{s-1} \frac{1}{\sqrt{M(t_n - \tau_i^0)}} \leq 2\log(T),$$

and so

$$E^i \leq 2\log(T)\sqrt{M(\tau_{i+1}^0 - 1 - \tau_i^0)} + 1.$$

□

B.2.8 Proof of Lemma 3.5.8

For π_{LUNA}, when epoch i ends in period $t = \tau_{i+1}^0$, we have $\Delta_t = \Delta_{\tau_{i+1}^0} = M^{\frac{1}{2}}(\tau_{i+1}^0 - \tau_i^0 - 1)^{-\frac{1}{2}}$ for every epoch $i \in [I-1]$. It then follows that

$$\begin{aligned} V &= \sum_{i=1}^{I} \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0 - 1} d_K(\hat{F}_t^\mu, \hat{F}_{t+1}^\mu) \geq \sum_{i=1}^{I-1} \sum_{t=\tau_i^0+1}^{\tau_{i+1}^0 - 1} d_K(\hat{F}_t^\mu, \hat{F}_{t+1}^\mu) \geq \sum_{i=1}^{I-1} \Delta_{\tau_{i+1}^0}/(s\,\bar{\xi}) \\ &= \sum_{i=1}^{I-1} \frac{1}{s\,\bar{\xi}} M^{\frac{1}{2}}(\tau_{i+1}^0 - \tau_i^0 - 1)^{-\frac{1}{2}} \geq \frac{M^{\frac{1}{2}}}{s\,\bar{\xi}}(I-1)^{\frac{3}{2}} T^{-\frac{1}{2}}. \end{aligned}$$

In the above display, we drop the last epoch I from the summation in the first inequality because we do not necessarily have $\sum_{t=\tau_I^0+1}^{\tau_{I+1}^0-1} d_K(\hat{F}_t^\mu, \hat{F}_{t+1}^\mu) \geq \Delta_I/(s\,\bar{\xi})$, i.e., epoch I does not necessarily end because $T \notin \mathcal{E}^I$. The second inequality follows from Lemma 3.5.2 and Lemma 3.5.4. The last inequality follows since

$$\sum_{i=1}^{I-1}(\tau_{i+1}^0 - \tau_i^0 - 1)^{-\frac{1}{2}} \geq (I-1)\left(\frac{1}{I-1}\sum_{i=1}^{I-1}(\tau_{i+1}^0 - \tau_i^0 - 1)\right)^{-\frac{1}{2}} \geq (I-1)\left(T/(I-1)\right)^{-\frac{1}{2}},$$

where the first inequality follows from Jensen's inequality and the second follows since $\sum_{i=1}^{I-1} \tau_{i+1}^0 - \tau_i^0 - 1 \leq T$. We conclude that $I \leq (s\,\bar{\xi})^{\frac{2}{3}} V^{\frac{2}{3}} M^{-\frac{1}{3}} T^{\frac{1}{3}} + 1$.

B.2.9 Proof of Lemma 3.5.9

Define the mapping $f : \mathbb{R} \to \mathbb{R}$ by $f(q) = z_n$ for $q \in (z_{n-1}, z_n]$. When $q = 0$, we have $f(0) \triangleq 0$.

We make use of the following result.

Lemma B.2.4. *For all $t \in [T]$, $q(w_t; \tilde{F}_t^\mu) = f(q(w_t; \hat{F}_t^\mu))$.*

Proof of Lemma B.2.4. When $q(w_t; \hat{F}_t^\mu) = 0$, we have $f(q(w_t; \hat{F}_t^\mu)) = 0$ and $\hat{F}_t^\mu(0) \geq 1 - w_t/s$ (by Eq. (3.4)). Since $\tilde{F}_t^\mu(0) = \hat{F}_t^\mu(0)$ by Eq. (3.15), we have $\tilde{F}_t^\mu(0) \geq 1 - w_t/s$ and thus $q(w_t; \tilde{F}_t^\mu) = 0$.

Now suppose $q(w_t; \hat{F}_t^\mu) \in (z_{n-1}, z_n]$ for some $n \geq 2$. Then, we have $\hat{F}_t^\mu(z_{n-1}) < 1 - w_t/s$, $\hat{F}_t^\mu(z_n) \geq 1 - w_t/s$, and $f(q(w_t; \hat{F}_t^\mu)) = z_n$. By Eq. (3.15), $\tilde{F}_t^\mu(z_{n-1}) < 1 - w_t/s$ and $\tilde{F}_t^\mu(z_n) \geq 1 - w_t/s$ both hold. Thus, $q(w_t; \tilde{F}_t^\mu) = z_n = f(q(w_t; \hat{F}_t^\mu)) = z_n$, and the claim holds.

$\square$

Proof of Lemma 3.5.9. The proof is by induction. The result clearly holds for period $t = 1$ since the first period wholesale price is fixed at $\bar{w}_1$ (recall π_{LUNAC} calls π_{LUNA} as a subroutine). Now suppose the claim holds up to some period $1 \leq t < T$, we will prove that it holds for period $t + 1$. For brevity, by the induction hypothesis we simply write $w_i \triangleq w_i^{\text{LUNAC}}(\hat{F}_{1:t-1}^\mu; \omega) = w_i^{\text{LUNA}}(\tilde{F}_{1:t-1}^\mu; \omega)$ for $i \in [t]$.

Fix the sample path ω, the history of wholesale prices $(w_i)_{i=1}^t$, and perceived distributions $\hat{F}_{1:t-1}^\mu$. In π_{LUNAC}, in each period the feedback $f(q(w_t; \hat{F}_t^\mu))$ is given to π_{LUNA} based on the actual order quantity $q(w_t; \hat{F}_t^\mu)$ (see Line 5 of Algorithm 2). Then, the wholesale price $w_{t+1}^{\text{LUNAC}}(\hat{F}_{1:t}^\mu; \omega)$ output by π_{LUNAC} is the wholesale price output by π_{LUNA} given the past wholesale prices $(w_i)_{i=1}^t$ and feedback $(f(q(w_t; \hat{F}_{1:i}^\mu)))_{i=1}^t$.

At the same time, according to the construction of $\tilde{F}_{1:t}^\mu$ (see Eq. (3.15)), given any wholesale price w_t we have $q(w_t; \tilde{F}_t^\mu) = f(q(w_t; \hat{F}_t^\mu))$ as shown in Lemma B.2.4. In other words, in each period $t \in [T]$, π_{LUNA} receives the feedback $f(q(w_t; \hat{F}_t^\mu)) = q(w_t; \tilde{F}_t^\mu)$. It then follows that $w_{t+1}^{\text{LUNAC}}(\hat{F}_{1:t}^\mu; \omega)$ is the price output by π_{LUNA} given past wholesale prices $(w_i)_{i=1}^t$ and orders $(q(w_t; \tilde{F}_t^\mu))_{i=1}^t$. Since π_{LUNA} will output $w_{t+1}^{\text{LUNA}}(\tilde{F}_{1:t}^\mu; \omega)$ given past wholesale prices $(w_i)_{i=1}^t$ and orders $(q(w_t; \tilde{F}_t^\mu))_{i=1}^t$, we have proved that $w_{t+1}^{\text{LUNAC}}(\hat{F}_{1:t}^\mu; \omega) = w_{t+1}^{\text{LUNA}}(\tilde{F}_{1:t}^\mu; \omega)$, and the induction step holds.

$\square$

B.2.10 Proof of Lemma 3.5.10

By definition, we have

$$d_K(\tilde{F}_t^\mu, \tilde{F}_{t+1}^\mu) = \max_{n\in[N]} |\tilde{F}_t^\mu(z_n) - \tilde{F}_{t+1}^\mu(z_n)| = \max_{n\in[N]} |\hat{F}_t^\mu(z_n) - \hat{F}_{t+1}^\mu(z_n)| \le d_K(\hat{F}_t^\mu, \hat{F}_{t+1}^\mu),$$

where the first equality follows since $\tilde{F}_t^\mu$ is supported on $\mathcal{Z}_N$, and the second equality follows by construction of $\tilde{F}_t^\mu$ in Eq. (3.15).

B.2.11 Proof of Theorem 3.5.2

Recall w_t^* defined in (3.3) is the optimal wholesale price in each period (regardless of whether the $\hat{F}_t^\mu$ is continuous or discrete). We can decompose the regret as

$$\begin{aligned}
&\mathrm{Reg}(\pi_{\mathrm{LUNAC}}, \hat{F}_{1:T}^\mu) \\
&= \sum_{t=1}^T \mathbb{E}\left[\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_t; \hat{F}_t^\mu)\right] \\
&= \sum_{t=1}^T \mathbb{E}\left[(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_t^*; \tilde{F}_t^\mu)) - (\varphi(w_t; \hat{F}_t^\mu) - \varphi(w_t; \tilde{F}_t^\mu)) + (\varphi(w_t^*; \tilde{F}_t^\mu) - \varphi(w_t; \tilde{F}_t^\mu))\right] \\
&\le \sum_{t=1}^T \mathbb{E}\Bigg[(\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_t^*; \tilde{F}_t^\mu)) - (\varphi(w_t; \hat{F}_t^\mu) - \varphi(w_t; \tilde{F}_t^\mu)) \\
&\qquad\qquad + \left(\sup_{w\in\mathcal{W}} \varphi(w; \tilde{F}_t^\mu) - \varphi(w_t; \tilde{F}_t^\mu)\right)\Bigg].
\end{aligned} \tag{B.12}$$

In the above display, both $\sum_{t=1}^T \mathbb{E}[\varphi(w_t^*; \hat{F}_t^\mu) - \varphi(w_t^*; \tilde{F}_t^\mu)]$ and $\sum_{t=1}^T \mathbb{E}\left[\varphi(w_t; \hat{F}_t^\mu) - \varphi(w_t; \tilde{F}_t^\mu)\right]$ represent the regret incurred by approximating $\hat{F}_t^\mu$ with $\tilde{F}_t^\mu$. According to Lemma 3.5.9, the wholesale prices $w_{1:T}$ output from π_{LUNAC} are just the pricing decisions of running the subroutine π_{LUNA} with distributions $\tilde{F}_{1:T}$. Thus the expression

$$\sum_{t=1}^T \mathbb{E}\left[\sup_{w\in\mathcal{W}} \varphi(w; \tilde{F}_t^\mu) - \varphi(w_t; \tilde{F}_t^\mu)\right]$$

is the regret from running π_{LUNA} with respect to $\tilde{F}_{1:T}$, i.e.,

$$\sum_{t=1}^{T} \mathbb{E}\left[\sup_{w\in\mathcal{W}} \varphi(w;\tilde{F}_t^{\mu}) - \varphi(w_t;\tilde{F}_t^{\mu})\right] = \text{Reg}(\pi_{\text{LUNA}}, T),$$

see Theorem 3.5.1.

Based on the approximation of $\hat{F}_t^{\mu}$ with $\tilde{F}_t^{\mu}$, for any $w \in \mathcal{W}$, we have

$$\begin{aligned} |\varphi(w,\hat{F}_t^{\mu}) - \varphi(w,\tilde{F}_t^{\mu})| &= (w-c)\left|\min\left\{q : \hat{F}_t^{\mu} \geq 1 - w/s\right\} - \min\left\{q : \tilde{F}_t^{\mu} \geq 1 - w/s\right\}\right| \\ &\leq (w-c)\bar{\xi}/N, \quad \text{(B.13)} \end{aligned}$$

where the inequality follows from Eq. (3.15). Then, we have

$$\begin{aligned} &\text{Reg}(\pi_{\text{LUNAC}}, \hat{F}_{1:T}^{\mu}) \\ &\leq \sum_{t=1}^{T} \mathbb{E}\Bigg[(\varphi(w_t^*;\hat{F}_t^{\mu}) - \varphi(w_t^*;\tilde{F}_t^{\mu})) - (\varphi(w_t;\hat{F}_t^{\mu}) - \varphi(w_t;\tilde{F}_t^{\mu})) \\ &\qquad + \left(\sup_{w\in\mathcal{W}} \varphi(w;\tilde{F}_t^{\mu}) - \varphi(w_t;\tilde{F}_t^{\mu})\right)\Bigg] \\ &\leq \sum_{t=1}^{T} \mathbb{E}\left[\left|\varphi(w_t^*;\hat{F}_t^{\mu}) + \varphi(w_t^*;\tilde{F}_t^{\mu})\right| + \left|\varphi(w_t;\hat{F}_t^{\mu}) - \varphi(w_t;\tilde{F}_t^{\mu})\right|\right] + \text{Reg}(\pi_{\text{LUNA}}, T). \end{aligned}$$

With Eq. (B.13) and Theorem 3.5.1, if the supplier knows V, then

$$\text{Reg}(\pi_{\text{LUNAC}}, \hat{F}_{1:T}^{\mu}) = \tilde{O}\left(\bar{\xi}\,T/N + \bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} N^{\frac{1}{3}} T^{\frac{2}{3}}\right)$$

and if the supplier does not know V, then

$$\text{Reg}(\pi_{\text{LUNAC}}, \hat{F}_{1:T}^{\mu}) = \tilde{O}\left(\bar{\xi}\,T/N + \bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} N^{\frac{1}{3}} T^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} V^{\frac{2}{3}} N^{-\frac{1}{3}} T^{\frac{2}{3}}\right).$$

It follows that if the supplier knows V, then with $N^* = \left\lceil \bar{\xi}^{-\frac{1}{4}} V^{-\frac{1}{4}} T^{\frac{1}{4}} \right\rceil$, the regret is $\text{Reg}(\pi_{\text{LUNAC}}, T) = \tilde{O}(\bar{\xi}^{\frac{5}{4}} V^{\frac{1}{4}} T^{\frac{3}{4}})$. If the supplier does not know V, then by choosing $\hat{N} = \left\lceil \bar{\xi}^{-\frac{1}{4}} T^{\frac{1}{4}} \right\rceil$ the regret is $\text{Reg}(\pi_{\text{LUNAC}}, T) = \tilde{O}(\bar{\xi}^{\frac{5}{4}} V^{\frac{1}{3}} T^{\frac{3}{4}} + \bar{\xi}^{\frac{17}{12}} V^{\frac{2}{3}} T^{\frac{7}{12}})$.

B.3 Additional material for Section 3.6

B.3.1 Proof of Proposition 3.6.1

Recall both $\hat{F}_t^{\mathrm{e}} = (\sum_{i=1}^{t-1} \mathbb{1}(\xi_i \leq x))/(t-1)$ and $\hat{F}_{t+1}^{\mathrm{e}} = (\sum_{i=1}^{t} \mathbb{1}(\xi_i \leq x))/t$ are empirical distributions, so for any $x \in [0, \bar{\xi}]$ we have

$$\left|\hat{F}_t^{\mathrm{e}}(x) - \hat{F}_{t+1}^{\mathrm{e}}(x)\right| \leq \left|\frac{\sum_{i=1}^{t-1} \mathbb{1}(\xi_i \leq x)}{t-1} - \frac{\sum_{i=1}^{t} \mathbb{1}(\xi_i \leq x)}{t}\right| = \begin{cases} \frac{\sum_{i=1}^{t-1} \mathbb{1}(\xi_i = x)}{t(t-1)}, \ \xi_t > x, \\ \frac{1}{t} - \frac{\sum_{i=1}^{t-1} \mathbb{1}(\xi_i = x)}{t(t-1)}, \ \xi_t \leq x. \end{cases} \tag{B.14}$$

It then follows that $d_K(\hat{F}_t^{\mathrm{e}}, \hat{F}_{t+1}^{\mathrm{e}}) \leq \frac{1}{t}$. Then $\sum_{t=1}^{T-1} d_K(\hat{F}_t^{\mathrm{e}}, \hat{F}_{t+1}^{\mathrm{e}}) \leq \log(T) + 1$ and thus $\mu_{\mathrm{e}} \in \mathcal{M}(\log(T) + 1, T)$.

B.3.2 Proof of Proposition 3.6.2

(i) For any $t \in [T-1]$, we have

$$\begin{aligned} d_K(\hat{F}_t^{\mathrm{d}}, \hat{F}_{t+1}^{\mathrm{d}}) &\leq d_K(\hat{F}_t^{\mathrm{d}}, \hat{F}_t^{\mathrm{e}}) + d_K(\hat{F}_{t+1}^{\mathrm{e}}, \hat{F}_t^{\mathrm{e}}) + d_K(\hat{F}_{t+1}^{\mathrm{e}}, \hat{F}_{t+1}^{\mathrm{d}}) \\ &\leq \frac{1}{t} + \sqrt{d_{KL}(\hat{F}_t^{\mathrm{d}}, \hat{F}_t^{\mathrm{e}})/2} + \sqrt{d_{KL}(\hat{F}_{t+1}^{\mathrm{e}}, \hat{F}_{t+1}^{\mathrm{d}})/2} \\ &\leq \frac{1}{t} + \sqrt{\epsilon_t/2} + \sqrt{\epsilon_{t+1}/2}, \end{aligned}$$

where the first inequality follows from triangle inequality, the second inequality follows from [64] (which states that $d_K(F, G) \leq \sqrt{d_{KL}(F, G)/2}$ for all $F, G \in \mathcal{P}$ with $F \ll G$) and Eq. (B.14). Then, we have

$$\sum_{t=1}^{T-1} d_K(\hat{F}_t^{d}, \hat{F}_{t+1}^{d}) \leq \sum_{t=1}^{T-1} \left(\frac{1}{t} + \sqrt{\epsilon_t/2} + \sqrt{\epsilon_{t+1}/2}\right) \leq \log(T) + 1 + \sum_{t=1}^{T} \sqrt{2\epsilon_t}.$$

(ii) For any $t \in [T-1]$, we have

$$
\begin{aligned}
d_K(\hat{F}_t^{\mathrm{d}}, \hat{F}_{t+1}^{\mathrm{d}}) &\leq d_K(\hat{F}_t^{\mathrm{d}}, \hat{F}_t^{\mathrm{e}}) + d_K(\hat{F}_{t+1}^{\mathrm{e}}, \hat{F}_t^{\mathrm{e}}) + d_K(\hat{F}_{t+1}^{\mathrm{d}}, \hat{F}_{t+1}^{\mathrm{e}}) \\
&\leq \frac{1}{t} + \sqrt{d_{\chi^2}(\hat{F}_t^{\mathrm{d}}, \hat{F}_t^{\mathrm{e}})/2} + \sqrt{d_{\chi^2}(\hat{F}_{t+1}^{\mathrm{e}}, \hat{F}_{t+1}^{\mathrm{d}})/2} \\
&\leq \frac{1}{t} + \sqrt{\epsilon_t}/2 + \sqrt{\epsilon_{t+1}}/2,
\end{aligned}
$$

where the second inequality follows from [64], (which states $d_K(F, G) \leq \sqrt{d_{\chi^2}(F, G)}/2$ for $F, G \in \mathcal{P}$ with $F \ll G$). Then, we have

$$
\sum_{t=1}^{T-1} d_K(\hat{F}_t^d, \hat{F}_{t+1}^d) \leq \sum_{t=1}^{T-1} \left(\frac{1}{t} + \sqrt{\epsilon_t}/2 + \sqrt{\epsilon_{t+1}}/2 \right) \leq \log(T) + 1 + \sum_{t=1}^{T} \sqrt{\epsilon_t}.
$$

(iii) We have

$$
\begin{aligned}
d_K(\hat{F}_t^{\mathrm{d}}, \hat{F}_{t+1}^{\mathrm{d}}) &\leq d_K(\hat{F}_t^{\mathrm{d}}, \hat{F}_t^{\mathrm{e}}) + d_K(\hat{F}_{t+1}^{\mathrm{e}}, \hat{F}_t^{\mathrm{e}}) + d_K(\hat{F}_{t+1}^{\mathrm{d}}, \hat{F}_{t+1}^{\mathrm{e}}) \\
&\leq \frac{1}{t} + d_H(\hat{F}_t^{\mathrm{d}}, \hat{F}_t^{\mathrm{e}}) + d_H(\hat{F}_{t+1}^{\mathrm{d}}, F_{t+1}^{\mathrm{e}}) \\
&\leq \frac{1}{t} + \epsilon_t + \epsilon_{t+1},
\end{aligned}
$$

where the second inequality follows from [64] (which states $d_K(F, G) \leq d_H(F, G)$ for $F, G \in \mathcal{P}$ with $F \ll G$). Then, we have

$$
\sum_{t=1}^{T-1} d_K(\hat{F}_t^d, \hat{F}_{t+1}^d) \leq \sum_{t=1}^{T-1} \left(\frac{1}{t} + \epsilon_t + \epsilon_{t+1} \right) \leq \log(T) + 1 + 2 \sum_{t=1}^{T} \epsilon_t.
$$

B.3.3 Proof of Proposition 3.6.3

To prove Proposition 3.6.3, we relate the Kolmogorov distance and the total variation distance d_{TV}. For two probaiblity distributions $F, G \in \mathcal{P}(\Xi)$ equipped with the $\sigma-$algebra $\mathcal{F}$, the total variation distance d_{TV} between F and G is defined by:

$$
d_{TV}(F, G) \triangleq \sup \{ |F(A) - G(A)| : A \in \mathcal{F} \}.
$$

According to [64],

$$d_{KL}(F,G) \le d_{TV}(F,G). \tag{B.15}$$

(i) According to [180], the total variation between Poisson distributions F_{λ_1} and F_{λ_2} with means λ_1 and λ_2 respectively (we assume that $\lambda_1 \le \lambda_2$) satisfies $d_{TV}\left(F_{\lambda_1}, F_{\lambda_2}\right) \le |\lambda_2 - \lambda_1|$. The MLE estimate of the mean of a Poisson distribution is $\lambda_t = \frac{\sum_{i=1}^{t-1} \xi_i}{t-1}$, so it follows that

$$\begin{aligned} d_{TV}(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) \le d_{TV}(F_{\lambda_t}, F_{\lambda_{t+1}}) &\le \left| \frac{\sum_{i=1}^{t} \xi_i}{t} - \frac{\sum_{i=1}^{t-1} \xi_i}{t-1} \right| \\ &= \left| \frac{t(\sum_{i=1}^{t-1} \xi_i) - (t-1)(\sum_{i=1}^{t-1} \xi_i + \xi_t)}{t(t-1)} \right| \le \frac{\max\left\{\sum_{i=1}^{t-1} \xi_i, (t-1)\xi_t\right\}}{t(t-1)}, \end{aligned} \tag{B.16}$$

where the first inequality follows by recalling from Eq. (3.20) that

$$\hat{F}_t^{\mathrm{m}}(x) = \begin{cases} F_{\lambda_t}(x), & 0 \le x < \bar{q}; \\ 1, & x \ge \bar{q}. \end{cases}$$

Poisson(λ) distribution has the following concentration inequality:

$$\mathbb{P}\left(\xi \ge \lambda + \epsilon\right) \le \exp\left(-\frac{\epsilon^2}{2(\lambda+\epsilon)}\right) \text{ for } \epsilon > 0, \tag{B.17}$$

and so $\mathbb{P}\left(\xi \le 4\ln(T) + 2\lambda\right) \ge 1 - 1/T^2$. By the union bound, we then have

$$\mathbb{P}\left(\xi_t \le 4\ln(T) + 2\lambda, \, \forall t \in [T]\right) \ge 1 - \frac{1}{T}.$$

It follows that

$$d_{TV}(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) \le \frac{4\ln(T) + 2\lambda}{t}$$

with probability at least $1-1/T$, and thus

$$\begin{aligned}\sum_{t=1}^{T-1} d_K(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) &\leq \sum_{t=1}^{T-1} d_{TV}(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) \\ &\leq \sum_{t=1}^{T-1} \frac{4\ln(T)+2\lambda}{t} \\ &\leq (4\ln(T)+2\lambda)\sum_{t=1}^{T-1}\frac{1}{t} \\ &\leq (\ln(T)+1)(4\ln(T)+2\lambda),\end{aligned}$$

where the first inequality follows from Eq. (B.15).

(ii) If the true demand distribution is the categorical distribution, and the retailer is using MLE, then $F_t = \hat{F}_t^{\mathrm{e}}$, the empirical distribution at time t. The argument then follows similarly to Proposition 3.6.1.

(iii) Let F_λ and $F_{\lambda'}$ be the CDFs of the $\mathrm{E}(\lambda)$ and $\mathrm{E}(\lambda')$ distributions, respectively, and suppose $\lambda < \lambda'$. Then we have

$$\begin{aligned}d_{TV}(F_\lambda, F_{\lambda'}) &= \frac{1}{2}\int_{x=0}^{\infty} |\lambda\exp(-\lambda x) - \lambda'\exp(-\lambda' x)|dx \\ &= \left(\frac{\lambda}{\lambda'}\right)^{\frac{\lambda}{\lambda'-\lambda}} - \left(\frac{\lambda}{\lambda'}\right)^{\frac{\lambda'}{\lambda-\lambda'}} \\ &\leq 1 - \frac{\lambda}{\lambda'} \\ &= \min\{\lambda, \lambda'\}\left|\frac{1}{\lambda} - \frac{1}{\lambda'}\right|.\end{aligned}$$

The MLE estimator for the rate is $\lambda_t = \frac{t-1}{\sum_{i=1}^{t-1}\xi_i}$, and so

$$\begin{aligned}d_{TV}(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) \leq d_{TV}(F_{\lambda_t}, F_{\lambda_{t+1}}) &\leq \lambda_t\left|\frac{\sum_{i=1}^{t-1}\xi_i}{t-1} - \frac{\sum_{i=1}^{t}\xi_i}{t}\right| \\ &\leq \frac{t-1}{\sum_{i=1}^{t-1}\xi_i}\frac{\left|\sum_{i=1}^{t-1}\xi_i - t\xi_t\right|}{t(t-1)} \\ &\leq \frac{1}{t} + \frac{t-1}{t}\frac{\xi_t}{\sum_{i=1}^{t-1}\xi_i}.\end{aligned}$$

Now according to the high probability bound for the $\mathrm{E}(\lambda)$ distribution, we have:

$$\mathbb{P}\left(\sum_{i=1}^{t} \xi_i \leq t/\lambda - \epsilon\right) \leq \exp\left(-\frac{\epsilon^2\lambda^2}{4t}\right),$$

which gives

$$\mathbb{P}\left(\sum_{i=1}^{t} \xi_i \geq \frac{t}{\lambda} - \frac{\sqrt{4t\ln\left(2T^2\right)}}{\lambda}\right) \geq 1 - \frac{1}{2T^2}.$$

On the other hand, according to the CDF of the exponential distribution, we have

$$\mathbb{P}\left(\xi \leq \frac{\ln\left(2T^2\right)}{\lambda}\right) \geq 1 - \frac{1}{2T^2}.$$

By the union bound, we have

$$\mathbb{P}\left(\sum_{i=1}^{t} \xi_i \geq \frac{t}{\lambda} - \frac{\sqrt{4t\ln\left(2T^2\right)}}{\lambda} \text{ and } \xi_t \leq \frac{\ln\left(2T^2\right)}{\lambda},\ \forall t \in [T-1]\right) \geq 1 - \frac{1}{T}. \tag{B.18}$$

It then follows that, with probability at least $1 - 1/T$, we have

$$\begin{aligned}
\sum_{t=1}^{T-1} d_K(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) &= \sum_{t=1}^{16\ln(2T^2)-1} d_K(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) + \sum_{t=16\ln(2T^2)}^{T-1} d_K(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) \\
&\leq \sum_{t=1}^{16\ln(2T^2)-1} d_K(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) + \sum_{t=16\ln(2T^2)}^{T-1} \left(\frac{1}{t} + \frac{t-1}{t}\frac{\xi_t}{\sum_{i=1}^{t-1}\xi_i}\right) \\
&\leq 16\ln(2T^2) - 1 + \sum_{t=16\ln(2T^2)}^{T-1} \frac{1 + 2\ln(2T^2)}{t} \\
&\leq 16\ln(2T^2) - 1 + \sum_{t=1}^{T-1} \frac{1 + 2\ln(2T^2)}{t} \\
&\leq 16\ln\left(2T^2\right) - 1 + \left(1 + 2\ln\left(2T^2\right)\right)\left(\ln\left(T\right) + 1\right),
\end{aligned}$$

where the second inequality follows from Eq. B.18 (which states that if $t \geq 16\ln(2T^2+1)$, then $\frac{\xi_t}{\sum_{i=1}^{t-1}\xi_i} \leq \frac{2\ln(2T^2)}{t}$ for all $t \geq 2$ with probability at least $1 - 1/T$).

(iv) Let F_μ and $F_{\mu'}$ be the CDFs of the $\mathrm{N}(\mu, \sigma^2)$ and $\mathrm{N}(\mu', \sigma^2)$ distributions, respectively (they have the same variance and possibly different means). The KL-divergence between F_μ and $F_{\mu'}$ is

$$d_{KL}(F_\mu, F_{\mu'}) = \frac{(\mu - \mu')^2}{2\sigma^2}.$$

The MLE estimator for the mean is

$$\mu_t = \frac{\sum_{i=1}^{t-1} \xi_i}{t-1},\ t \geq 2.$$

Thus, we have

$$\begin{aligned} d_{KL}(F_{\mu_t}, F_{\mu_{t+1}}) &= \frac{1}{2\sigma^2} \left(\frac{\sum_{i=1}^{t-1} \xi_i}{t-1} - \frac{\sum_{i=1}^{t} \xi_i}{t} \right)^2 \\ &\leq \frac{1}{2\sigma^2} \left(\frac{\sum_{i=1}^{t-1} \xi_i - (t-1)\xi_t}{t(t-1)} \right)^2 \\ &\leq \frac{1}{2\sigma^2} \left[\left(\frac{\sum_{i=1}^{t-1} \xi_i}{t(t-1)} \right)^2 + \left(\frac{\xi_t}{t} \right)^2 \right],\ t \geq 2. \end{aligned}$$

On one hand, when $\xi \sim \mathrm{Normal}(\mu, \sigma^2)$, we have

$$\mathbb{P}(\xi \geq \mu + \epsilon) \leq \exp\left(-\epsilon^2/2\sigma^2\right),$$

which gives

$$\mathbb{P}\left(\xi \leq \sigma\sqrt{2\log(2T^2)} + \mu\right) \geq 1 - 1/(2T^2).$$

On the other hand, according to Hoeffding's inequality,

$$\mathbb{P}\left(\sum_{i=1}^{t} \xi_i - t\mu \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2t\sigma^2}\right). \tag{B.19}$$

It then follows that

$$\mathbb{P}\left(\sum_{i=1}^{t}\xi_i \le t\mu + \sigma\sqrt{2t\ln(2T^2)}\right) \ge 1 - 1/(2T^2).$$

Using the union bound, we have

$$\mathbb{P}\left(\sum_{i=1}^{t-1}\xi_i \le (t-1)\mu + \sigma\sqrt{2(t-1)\ln(2T^2)} \text{ and } \xi_t \le \sigma\sqrt{2\ln(2T^2)} + \mu,\ \forall t \ge 2\right) \ge 1 - \frac{1}{T}. \tag{B.20}$$

We then have with probability at least $1 - 1/T$, for $t \ge 2$,

$$\begin{aligned}
d_{KL}(F_{\mu_t}, F_{\mu_{t+1}}) &\le \frac{1}{2\sigma^2}\left[\left(\frac{\sum_{i=1}^{t-1}\xi_i}{t(t-1)}\right)^2 + \left(\frac{\xi_t}{t}\right)^2\right] \\
&\le \frac{1}{2\sigma^2}\left[\left(\frac{(t-1)\mu + \sigma\sqrt{2(t-1)\ln(2T^2)}}{t(t-1)}\right)^2 + \left(\frac{\mu + \sigma\sqrt{2\ln(2T^2)}}{t}\right)^2\right] \\
&\le \frac{1}{\sigma^2}\left[\frac{2\mu^2}{t^2} + \frac{2\sigma^2\ln(2T^2)}{t(t-1)^2} + \frac{2\sigma^2\ln(2T^2)}{t^2}\right] \\
&\le \frac{2\mu^2 + 4\sigma^2\ln(2T^2)}{\sigma^2}\frac{1}{(t-1)^2}
\end{aligned}$$

where the second inequality follows by the high probability bound Eq. (B.20). The third inequality follows by the inequality $(x+y)^2 \le 2x^2 + 2y^2$ for arbitrary number x, y. Consequently,

$$\begin{aligned}
\sum_{t=1}^{T-1} d_K(\hat{F}_t^{\mathrm{m}}, \hat{F}_{t+1}^{\mathrm{m}}) &\le 1 + \sum_{t=2}^{T-1} d_K(F_{\mu_t}, F_{\mu_{t+1}}) \\
&\le 1 + \sum_{t=2}^{T-1}\sqrt{d_{KL}(F_{\mu_t}, F_{\mu_{t+1}})/2} \\
&\le 1 + \sum_{t=2}^{T-1}\frac{1}{t-1}\sqrt{\frac{\mu^2 + 2\sigma^2\ln(2T^2)}{\sigma^2}} \\
&\le 1 + \frac{1}{\sigma}\sqrt{(\ln(T)+1)\left(\mu^2 + 2\sigma^2\ln(2T^2)\right)}.
\end{aligned}$$

B.3.4 Proof of Proposition 3.6.4

Let F_λ and $F_{\lambda'}$ be the CDFs of the $\mathrm{E}(\lambda)$ and $\mathrm{E}(\lambda')$ distributions, respectively, and suppose $\lambda < \lambda'$. The total variation between F_λ and $F_{\lambda'}$ is

$$\begin{aligned}
d_{TV}(F_\lambda, F_{\lambda'}) &= \frac{1}{2}\int_{x=0}^{\infty} |\lambda \exp(-\lambda\, x) - \lambda' \exp(-\lambda' x)| dx \\
&= \left(\frac{\lambda}{\lambda'}\right)^{\frac{\lambda}{\lambda'-\lambda}} - \left(\frac{\lambda}{\lambda'}\right)^{\frac{\lambda'}{\lambda-\lambda'}} \\
&\leq 1 - \frac{\lambda}{\lambda'}.
\end{aligned}$$

For all $t \in [T]$, define the function f_t such that $f_t(w_t) \triangleq \frac{(t-1)\left((s/w_t)^{\frac{1}{t}}-1\right)}{\ln(s/w_t)}$. Then, we have $1/\lambda_t = f_t(w_t)\left(\sum_{i=1}^{t-1}\xi_i\right)/(t-1)$ and $f_t(w_t) \in [(t-1)/t, 1)$ for all $w_t \in \mathcal{W}$. It follows that

$$\begin{aligned}
d_K(\hat{F}_t^{\mathrm{o}}, \hat{F}_{t+1}^{\mathrm{o}}) \leq d_{TV}(\hat{F}_t^{\mathrm{o}}, \hat{F}_{t+1}^{\mathrm{o}}) \leq d_{TV}(F_{\lambda_t}, F_{\lambda_{t+1}}) \leq 1 - \frac{\min\{\lambda_t, \lambda_{t+1}\}}{\max\{\lambda_t, \lambda_{t+1}\}} \\
\leq \min\{\lambda_t, \lambda_{t+1}\}\left|\frac{1}{\lambda_t} - \frac{1}{\lambda_{t+1}}\right|.
\end{aligned}$$

Next, note that for $t \geq 2$,

$$\begin{aligned}
&\left|\frac{1}{\lambda_t} - \frac{1}{\lambda_{t+1}}\right| \\
&= \left| f_t(w_t)\frac{\sum_{i=1}^{t-1}\xi_i}{t-1} - f_{t+1}(w_{t+1})\frac{\sum_{i=1}^{t}\xi_i}{t}\right| \\
&= \left|\frac{t f_t(w_t)\sum_{i=1}^{t-1}\xi_i - (t-1)f_{t+1}(w_{t+1})\sum_{i=1}^{t}\xi_i}{t(t-1)}\right| \\
&= \left|\frac{(t-1)(f_t(w_t) - f_{t+1}(w_{t+1}))\left(\sum_{i=1}^{t-1}\xi_i\right) + f_t(w_t)\left(\sum_{i=1}^{t-1}\xi_i\right) - (t-1)f_{t+1}(w_{t+1})\xi_t}{t(t-1)}\right| \\
&\leq \frac{\left|(t-1)(f_t(w_t) - f_{t+1}(w_{t+1}))\left(\sum_{i=1}^{t-1}\xi_i\right)\right| + \left|f_t(w_t)\left(\sum_{i=1}^{t-1}\xi_i\right)\right| + |(t-1)f_{t+1}(w_{t+1})\xi_t|}{t(t-1)} \\
&\leq \frac{2t-1}{(t-1)t^2}\left(\sum_{i=1}^{t-1}\xi_i\right) + \frac{\xi_t}{t},
\end{aligned}$$

where the last inequality follows since $f_t(w_t) \in [(t-1)/t, 1)$ for all $w_t \in \mathcal{W}$. Clearly $\min\{\lambda_t, \lambda_{t+1}\} \leq \lambda_t = \frac{t-1}{f_t(w_t)\left(\sum_{i=1}^{t-1} \xi_i\right)}$ and thus for $t \geq 2$,

$$
\begin{aligned}
d_K(\hat{F}_t^{\mathrm{o}}, \hat{F}_{t+1}^{\mathrm{o}}) &\leq \lambda_t \left| \frac{1}{\lambda_t} - \frac{1}{\lambda_{t+1}} \right| \\
&\leq \frac{t-1}{f_t(w_t)\left(\sum_{i=1}^{t-1} \xi_i\right)} \left(\frac{2t-1}{(t-1)t^2} \left(\sum_{i=1}^{t-1} \xi_i \right) + \frac{\xi_t}{t} \right) \\
&\leq \frac{2t-1}{(t-1)t} + \frac{\xi_t}{\sum_{i=1}^{t-1} \xi_i} \\
&\leq \frac{2}{t-1} + \frac{\xi_t}{\sum_{i=1}^{t-1} \xi_i},
\end{aligned}
$$

where the third inequality again follows because $f_t(w_t) \in [(t-1)/t, 1)$ for all $w_t \in \mathcal{W}$. Since $(\xi_i)_{i=1}^T$ are i.i.d. Exponential(λ), by Eq. (B.18) we have

$$
\mathbb{P}\left(\sum_{i=1}^{t} \xi_i \geq \frac{t}{\lambda} - \frac{\sqrt{4t \ln(2T^2)}}{\lambda} \text{ and } \xi_{t+1} \leq \frac{\ln(2T^2)}{\lambda}, \forall t \in [T-1] \right) \geq 1 - \frac{1}{T}.
$$

Thus, with probability at least $1 - 1/T$, we have

$$
\begin{aligned}
\sum_{t=1}^{T-1} d_K(\hat{F}_t^{\mathrm{o}}, \hat{F}_{t+1}^{\mathrm{o}}) &\leq 1 + \sum_{t=2}^{T-1} \frac{2}{t-1} + 16 \ln(2T^2) + \sum_{t=2}^{T-1} \frac{2 \ln(2T^2)}{t-1} \\
&\leq 1 + 2 \ln T + 2 + 16 \ln(2T^2) + 2 \ln(2T^2)(\ln T + 1) \\
&\leq 21 + 40 \ln(T) + 4(\ln(T))^2.
\end{aligned}
$$

B.4 Additional material for Section 3.5.1

B.4.1 LUNAC-N

When the supplier does not know V, we show that we can further improve the regret bound by adopting the Bandit-over-Bandit (BOB) framework proposed by [63], [90] to sequentially adjust the approximation size N (we call this algorithm $\pi_{\text{LUNAC-N}}$). $\pi_{\text{LUNAC-N}}$ divides the time horizon into $\lceil T/H \rceil$ blocks (indexed by i) of equal length H. Inside block i, the discretization size N_i is chosen from a finite set $\mathcal{J} \subset [H]$. Based on the chosen N_i for each block, we run π_{LUNAC} for that block. After receiving the profits from each block, the algorithm sequentially adjusts the approximation size N_i from block to block.

The choice of N_i in each block is chosen according to the EXP3 algorithm [92] designed for the adversarial bandit. In this way, the overall procedure consists of a meta algorithm for choosing N_i in each block according to the profits collected from each block, and a sub-algorithm that is run inside each block, based on the chosen N_i for that block. Theorem B.4.1 presents the regret bound when the supplier does not have knowledge of V and N_i is chosen according to $\pi_{\text{LUNAC-N}}$.

Theorem B.4.1. *Suppose Assumption 3.5.2 holds, the supplier does not have knowledge of V, and $\{N_i\}$ are chosen according to $\pi_{LUNAC\text{-}N}$. Then, $Reg(\pi_{LUNAC\text{-}N}, T) = \tilde{O}(\bar{\xi}^{\frac{4}{3}} V^{\frac{3}{4}} T^{\frac{1}{3}} + \bar{\xi}^{\frac{5}{4}} V^{\frac{1}{3}} T^{\frac{3}{4}})$.*

Algorithm 6 presents the implementation details for $\pi_{\text{LUNAC-N}}$ for sequentially adjusting N when V is unknown. We initialize the EXP3 parameters as:

$$\gamma = \min\left\{1, \sqrt{\frac{(z+1)\ln(z+1)}{(\mathrm{e}-1)\lceil T/H\rceil}}\right\}, \; s_{j,1} = 1, \quad \forall j = 0, 1, \ldots, z. \tag{B.21}$$

Algorithm 6 LUNAC-N

Require: Time horizon T, production cost c, and selling price s

Initialize $H \leftarrow \left\lfloor \bar{\xi}^{-\frac{1}{4}} T^{\frac{1}{4}} \right\rfloor, z \leftarrow \lceil \ln H \rceil, \mathcal{J} \leftarrow \left\{ H^0, \left\lfloor H^{1/z} \right\rfloor, \ldots, H \right\}$; Set γ and $(s_{j,1})_{j=0}^{z}$ according to Eq. B.21

for $i = 1, 2, \ldots, \lceil T/H \rceil$ **do**

Define distributions $(\alpha_{j,i})_{j=0}^{z}$ as:

$$\alpha_{j,i} = (1-\gamma) \frac{s_{j,i}}{\sum_{u=0}^{z} s_{u,i}} + \frac{\gamma}{z+1}, \forall j = 0, \ldots, z$$

Choose $j_i \leftarrow j$ with probability $\alpha_{j,i}$ and set $N_i \leftarrow \left\lfloor H^{j_i/z} \right\rfloor$.

for $t = (i-1)H + 1, \ldots, (i \ldots H) \wedge T$ **do**

Run LUNAC-N with N_i.

$\sum_{t=(i-1)H+1}^{(iH)\wedge T} \varphi(w_t; F_t)$ is the profit collected during $t \in [(i-1)H + 1, (i \cdot H) \wedge T]$;

Update $s_{j,i+1}$ as:

$$s_{j_i,i+1} \leftarrow s_{j_i,i} \exp\left(\frac{\gamma}{(z+1)\alpha_{j_i,i}} \left(\frac{1}{2} + \frac{1}{2} \frac{\sum_{t=(i-1)H+1}^{(iH)\wedge T} \varphi(w_t; F_t)}{((iH) \wedge T - (i-1)H)(s-c)\bar{\xi}} \right) \right),$$

$$s_{j,i+1} \leftarrow s_{j,i}, \quad \text{if } j \neq j_i.$$

Proof of Theorem B.4.1. Let $N^\dagger$ be the optimally tuned approximation size and $w_t^\dagger$ be the corresponding wholesale price of π_{LUNA} when the approximation size satisfies $N = N^\dagger$. Notice that since each block has at most H rounds, we do not necessarily have $N^\dagger = N^*$ (where N^*

is the optimally chosen discretization size given the supplier knows V, see Theorem 3.5.2) since we need $N^\dagger \leq H$. The regret of running BOB on top of π_{LUNAC} can be decomposed as:

$$
\begin{aligned}
&\text{Reg}\left(\pi_{\text{LUNA-K}}; \hat{F}^{\mu}_{1:T}\right)\\
=&\sum_{t=1}^{T} \varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_t; \hat{F}^{\mu}_t)\\
=&\sum_{t=1}^{T} \varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_t^\dagger; \hat{F}^{\mu}_t) + \varphi(w_t^\dagger; \hat{F}^{\mu}_t) - \varphi(w_t; \hat{F}^{\mu}_t)\\
=&\sum_{i=1}^{\lceil T/H \rceil} \sum_{t=(i-1)H+1}^{i \cdot H \wedge T} \left\{\left(\varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_t^\dagger; \hat{F}^{\mu}_t)\right) + \left(\varphi(w_t^\dagger; \hat{F}^{\mu}_t) - \varphi(w_t; \hat{F}^{\mu}_t)\right)\right\}.
\end{aligned}
$$

where $\sum_{i=1}^{\lceil T/H \rceil} \sum_{t=(i-1)H+1}^{i \cdot H \wedge T} \left(\varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_t^\dagger; \hat{F}^{\mu}_t)\right)$ is the regret incurred by always discretizing at $N^\dagger$ and $\sum_{i=1}^{\lceil T/H \rceil} \sum_{t=(i-1)H+1}^{i \cdot H \wedge T} \left(\varphi(w_t^\dagger; \hat{F}^{\mu}_t) - \varphi(w_t; \hat{F}^{\mu}_t)\right)$ is the regret of learning $N^\dagger$.

Let $V(i)$ be the variation in block i:

$$
V(i) \triangleq \sum_{t=(i-1)H+1}^{(i \cdot H) \wedge T - 1} d_K(\hat{F}^{\mu}_t, \hat{F}^{\mu}_{t+1}).
$$

Then, the regret incurred by always discretizing at $N^\dagger$ can be upper bounded with:

$$
\begin{aligned}
&\sup_{\hat{F}^{\mu}_{1:T} \in \mathcal{M}(V,T)} \sum_{i=1}^{\lceil T/H \rceil} \sum_{t=(i-1)H+1}^{i \cdot H \wedge T} \mathbb{E}\left[\varphi(w_t^*; \hat{F}^{\mu}_t) - \varphi(w_t^\dagger; \hat{F}^{\mu}_t)\right]\\
=&\sum_{i=1}^{\lceil T/H \rceil} \tilde{O}\left(\bar{\xi}\, H/N^\dagger + \bar{\xi}^{\frac{4}{3}} V(i)^{\frac{1}{3}} N^{\dagger \frac{1}{3}} H^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} V(i)^{\frac{2}{3}} N^{\dagger - \frac{1}{3}} H^{\frac{2}{3}}\right)\\
=&\tilde{O}\left(\bar{\xi} T/N^\dagger + \bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} N^{\dagger \frac{1}{3}} T^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} V^{\frac{2}{3}} N^{\dagger - \frac{1}{3}} T^{\frac{1}{3}} H^{\frac{1}{3}}\right),
\end{aligned}
$$

where the first equality follows from Theorem 3.5.2 and the second equality follows from Holder's inequality.

The regret of learning $N^\dagger$ follows directly from the regret of running EXP3. Since the number of blocks for EXP3 is $\lceil T/H \rceil$, the number of possible values for K is $|\mathcal{J}|$ and the maximum regret in each block is $(s-c)\bar{\xi}H$, we have

$$\begin{aligned}
&\sup_{\mu\in\mathcal{M}(V,T)} \sum_{i=1}^{\lceil T/H \rceil} \sum_{t=(i-1)H+1}^{i\cdot H\wedge T} \mathbb{E}\left[\varphi(w_t^\dagger; F_t) - \varphi(w_t; F_t)\right] \\
=&\tilde{O}\left(\bar{\xi}H\sqrt{\frac{|\mathcal{J}|T}{H}}\right) \\
=&\tilde{O}\left(\bar{\xi}\sqrt{|\mathcal{J}|TH}\right).
\end{aligned}$$

Combining these bounds, we get

$$\mathrm{Reg}(\pi_{\mathrm{LUNAC}}, T) = \tilde{O}\left(\bar{\xi}T/N^\dagger + \bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} N^{\dagger\frac{1}{3}} T^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} V^{\frac{2}{3}} N^{\dagger-\frac{1}{3}} T^{\frac{1}{3}} H^{\frac{1}{3}} + \bar{\xi}\sqrt{|\mathcal{J}|TH}\right).$$

Following [63], [90], we consider the set $\mathcal{J}$ for possible choices of N:

$$\mathcal{J} = \{H^0, \left\lfloor H^{\frac{1}{z}} \right\rfloor, \left\lfloor H^{\frac{2}{z}} \right\rfloor, \ldots, H\}$$

where z is some positive integer. Since the choice of H cannot depend on V, we can set $H = \bar{\xi}^{\epsilon} T^{\alpha}$ for some $\alpha \in (0,1)$. We now discuss two cases depending on whether $N^* \geq H$ or not.

<u>Case 1</u>: $N^* \leq H$, then $V > T^{1-4\alpha}\bar{\xi}^{-1-4\epsilon}$. In this case, $N^\dagger$ can automatically adapt to the largest element in $\mathcal{J}$ that is smaller than N^* (i.e., $N^* H^{-\frac{1}{z}} \leq N^\dagger \leq N^* H^{\frac{1}{z}}$), and thus

$$\begin{aligned}
&\mathrm{Reg}(\pi_{\text{LUNA-N}}, T) \\
=&\tilde{O}\left(\bar{\xi}T/(N^* H^{-\frac{1}{z}}) + \bar{\xi}^{\frac{4}{3}} V^{\frac{1}{3}} (N^* H^{\frac{1}{z}})^{\frac{1}{3}} T^{\frac{2}{3}} + \bar{\xi}^{\frac{4}{3}} V^{\frac{2}{3}} (N^* H^{-\frac{1}{z}})^{-\frac{1}{3}} T^{\frac{1}{3}} H^{\frac{1}{3}} + \bar{\xi}\sqrt{|\mathcal{J}|TH}\right) \quad \text{(B.22)} \\
=&\tilde{O}\left(\bar{\xi}^{\frac{5}{4}} V^{\frac{1}{4}} T^{\frac{3}{4}} H^{\frac{1}{z}} + \bar{\xi}^{\frac{5}{4}} T^{\frac{3}{4}} V^{\frac{1}{4}} H^{\frac{1}{3z}} + \bar{\xi}^{\frac{17}{12}+\frac{\epsilon}{3}} H^{1/3z} V^{\frac{3}{4}} T^{\frac{1}{4}+\frac{\alpha}{3}} + \bar{\xi}^{1+\frac{\epsilon}{2}} z^{\frac{1}{2}} T^{\frac{\alpha+1}{2}}\right).
\end{aligned}$$

Case 2: $K^* \geq H$, then $V \leq T^{1-4\alpha}\bar{\xi}^{-1-4\epsilon}$. In this case, $M^\dagger = H$, and thus

$$\begin{aligned}\mathrm{Reg}(\pi_{\text{LUNA-N}}, T) &= \tilde{O}\left(\bar{\xi}T/H + \bar{\xi}^{\frac{-4}{3}}V^{\frac{1}{3}}H^{\frac{1}{3}}T^{\frac{2}{3}} + \bar{\xi}^{\frac{-4}{3}}V^{\frac{2}{3}}H^{-\frac{1}{3}}T^{\frac{1}{3}}H^{\frac{1}{3}} + \bar{\xi}\sqrt{|\mathcal{J}|TH}\right) \\ &= \tilde{O}\left(\bar{\xi}^{1-\alpha}T^{1-\alpha} + \bar{\xi}^{\frac{-4+\epsilon}{3}}V^{\frac{1}{3}}T^{\frac{2+\alpha}{3}} + \bar{\xi}^{\frac{-4}{3}}V^{\frac{2}{3}}T^{\frac{1}{3}} + z^{\frac{1}{2}}\bar{\xi}^{1+\frac{\epsilon}{2}}T^{\frac{1+\alpha}{2}}\right).\end{aligned} \tag{B.23}$$

According to Eqs. (B.22) and (B.23), we can set $z = \lfloor \ln H \rfloor$, $\epsilon = -\frac{1}{4}$ and $\alpha = \frac{1}{4}$ and the regret is

$$\mathrm{Reg}(\pi_{\text{LUNA-N}}, T) = \tilde{O}\left(\bar{\xi}^{\frac{-4}{3}}V^{\frac{3}{4}}T^{\frac{1}{3}} + \bar{\xi}^{\frac{-5}{4}}V^{\frac{1}{3}}T^{\frac{3}{4}}\right).$$

□

B.5 Additional materials for Section 3.7

B.5.1 Finite Decision Set

We can modify π_{LUNA} to handle finite $\mathcal{W}$, and we call this modified algorithm π_{LUNAF} (see the details in Algorithm 7). Let $d \triangleq |\mathcal{W}|$ be the number of admissible wholesale prices so $\mathcal{W} = \{w_j\}_{j=1}^d$ where WLOG we assume $w_1 < w_2 < \cdots < w_d$. Let $\mathrm{ceil}_{\mathcal{S}}(x)$ be the smallest element in a set $\mathcal{S}$ that is greater than or equal to x and $\mathrm{floor}_{\mathcal{S}}(x)$ be the largest element in $\mathcal{S}$ that is less than or equal to x. When $x \in \mathcal{S}$, then $\mathrm{ceil}_{\mathcal{S}}(x) = \mathrm{floor}_{\mathcal{S}}(x) = x$. In the exploration phase of π_{LUNAF}, the policy simply prices at each price in $\mathcal{W}$. Let j^* be the index of the optimal wholesale price $w_j^* \in \mathcal{W}$. Then, in each period in the exploitation phase, w_m^t for $m \in [M]$ and w_0^t are computed according to

$$(w_m^t - c)y_m = \varphi_{j^*} + \Delta_t + (w_{j^*+1} - w_{j^*})y_{m^*}, \tag{B.24}$$

which gives

$$w_m^t \triangleq \left(\varphi_{j^*} + (w_{j^*+1} - w_{j^*})y_{m^*} + \Delta_t\right)/y_m + c,$$

and

$$(w_0 - c)y_{j^*} = \varphi_{j^*} - \Delta_t \text{ and } w_0^t \geq 0, \text{ otherwise } w_t^0 = 0, \tag{B.25}$$

which gives

$$w_0^t \triangleq \max\{w_{j^*} - \Delta_t / y_{j^*}, 0\}.$$

However, since we do not necessarily have $w_{m_t}^t \in \mathcal{W}$ for $m_t \in [M]$, we need to project $w_{m_t}^t$ to $\mathcal{W}$. In this case, the dynamic regret follows as

$$\mathrm{Reg}(\pi, T) \triangleq \max_{\mu \in \mathcal{M}(V,T)} \mathbb{E}\left[\sum_{t=1}^{T} \left(\max_{w \in \mathcal{W}} \varphi(w; \hat{F}_t^{\mu}) - (w_t - c) q(w_t; \hat{F}_t^{\mu})\right)\right],$$

where the clairvoyant benchmark optimizes over the prices in the finite admissable set $\mathcal{W}$. We compare this regret for different algorithms numerically.

Algorithm 7 Learning under Nonstationary Agent with Finite decision set (LUNAF)

Require: Time horizon T, c, s, and admissible decisions $\mathcal{W}$

Update current time $t \leftarrow 1$

Set epoch $i \leftarrow 1$

for epoch $i = 1, 2, \ldots$ **do**

$\tau_i^0 \leftarrow t$

Exploration:

Price at $w_j \in \mathcal{W}, j \in [d-1]$ and observe φ_j for the first $d-1$ periods in epoch i

Let $j^* \in \arg\max_{j \in [J]} \varphi_j$ and m^* be such that $y_{m^*} = q\left(w_{j^*}; F_{\tau_i^0 + j^*}\right)$

Exploitation:

In period t, set $\Delta_t \leftarrow \sqrt{M/(t-\tau_i^0)}$

Compute prices w_m for $m \in [M]$ and w_0 according to Eq. (3.12) and Eq. (3.13), respectively

Select m_t according to the distribution

$$m_t = \begin{cases} 0, & \text{w.p. } 1 - \sqrt{\frac{M}{t-\tau_i^0}}, \\ \mathcal{U}\{1, \ldots, M\}, & \text{w.p. } \sqrt{\frac{M}{t-\tau_i^0}}; \end{cases}$$

if $m_t \geq 1$ **then** Set wholesale price at $w_t \leftarrow \text{ceil}_{\mathcal{W}}(w_m)$

else

Set wholesale price at $w_t \leftarrow \text{floor}_{\mathcal{W}}(w_0)$

Observe retailer's order $q(w_t; F_t)$

if $q(w_{m_t}; F_t) \geq y_{m_t}$ for $m_t \in [M]$ or $q(w_{m_t}; F_t) < y_{m^*}$ for $m_t = 0$ **then**

Start the next epoch $i \leftarrow i + 1$.

C. APPENDICES FOR CHAPTER 4

C.1 Additional Model Details

Some additional technical details of the model formulation are as follows. First, under the AWP policy, the wholesale price and supply satisfy the non-negativity constraints $p-\beta w_i \geq 0$ and $w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a \geq 0$, which implies $a + b\bar{\boldsymbol{w}}_{\text{-}i} \leq w_i \leq p/\beta$. A further requirement is that this constraint set is non-empty at the symmetric equilibrium derived in later analysis. If at the equilibrium, each GPO adopts a wholesale price w, then the constraint $w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a = (1-b)w - a \geq 0$, implying that $w \geq a/(1-b)$, and the constraint $p - \beta w_i = p - \beta w \geq 0$, implying that $w \leq p/\beta$. This further requires the model parameters to satisfy $a/(1-b) \leq p/\beta$, so that the model is meaningful and realistic. In our analysis, we assume by default that this relationship holds.

Under the ASP policy, the non-negativity constraints are $\theta\bar{\boldsymbol{w}} - \beta w_i \geq 0$ and $w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a \geq 0$, implying $a + b\bar{\boldsymbol{w}}_{\text{-}i} \leq w_i \leq \theta\bar{\boldsymbol{w}}/\beta$. Note that the constraint $w_i \leq \theta\bar{\boldsymbol{w}}/\beta$ is always satisfied at an symmetric equilibrium since $\theta > 1$ and $\beta \leq 1$. As stated in the main text, to maintain realisticity, an additional requirement for model parameters is $\beta > \theta/n$ (see the explanation after Eq. 4.2). We again assume by default that this relationship holds.

C.2 Proof of Theorem 4.4.1

We first show the existence and uniqueness of the equilibrium. Under the AWP policy, GPO i's payoff function (4.2) can be expressed as $\mathbb{E}[(p-\beta w_i)(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)\epsilon_i \wedge (p-\beta w_i)d]$, which is the expected minimum of a concave function in w_i and a linear function in w_i, and is clearly concave in w_i. The function is also continuous. The constraint set, $w_i \geq 0$, $p - \beta w_i \geq 0$ and $w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a \geq 0$, is clearly convex and compact. Further, the game is symmetric. Thus, there exists at least one symmetric pure strategy Nash Equilibrium.

In addition, it is straightforward to see that any boundary point of the constraint set renders zero payoff, while any interior point yields a positive payoff. Thus, the equilibria are interior.

Finally, the payoff function (4.2) is a product of a linear function and a concave function of $\boldsymbol{w}$, and hence is log-concave (and quasi-concave) in $\boldsymbol{w}$. Since the strategy space is convex and the equilibria are interior, by a univalent mapping argument (see, e.g., Theorem 6 in [181]), the equilibrium is unique.

Under the ASP policy, in GPO i's payoff function (4.2),

$$\theta\bar{\boldsymbol{w}} - \beta w_i = \theta \sum_{j \neq i} w_j/n - (\beta - \theta/n) w_i.$$

Since by assumption, $\beta - \theta/n > 0$, following the same procedure as that in the previous AWP proof, we can show the existence of a unique symmetric interior equilibrium.

We now derive the conditions that the equilibrium wholesale prices satisfy. Under the AWP policy, since ϵ_i has a continuous PDF, the payoff function (4.2) is differentiable. The payoff function can be written as:

$$(p - \beta w_i)(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)\mathbb{E}[\epsilon_i \wedge d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)].$$

The first order derivative with respect to w_i is:

$$[-\beta(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a) + (p - \beta w_i)]\mathbb{E}[\epsilon_i \wedge d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)] - (p - \beta w_i)\bar{F}(d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a))d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a).$$

As the equilibrium is symmetric, $w_i = w, \forall i$. Thus, the equilibrium w must satisfy:

$$[(b-2)\beta w + a\beta + p]\mathbb{E}\{[(1-b)w - a]\epsilon \wedge d\} - d(p - \beta w)\bar{F}(d/[(1-b)w - a]) = 0.$$

Similarly, under the ASP policy, the first order derivative with respect to w_i is:

$$\begin{aligned}[(\theta/n - \beta)(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a) + (\theta\bar{\boldsymbol{w}} - \beta w_i)]\mathbb{E}[\epsilon_i \wedge d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a)] \\ - (\theta\bar{\boldsymbol{w}} - \beta w_i)\bar{F}(d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a))d/(w_i - b\bar{\boldsymbol{w}}_{\text{-}i} - a).\end{aligned}$$

At the symmetric equilibrium,

$$\{(\theta/n-\beta)[(1-b)w-a]+(\theta-\beta)w\}\mathbb{E}\{[(1-b)w-a]\epsilon\wedge d\}-d(\theta-\beta)w\bar{F}(d/[(1-b)w-a])=0.$$

C.3 Proof of Theorem 4.4.2

We first show the monotonicity of the equilibrium wholesale price with respect to the parameters. Under the AWP policy, we define:

$$h=[(b-2)\beta w+a\beta+p][(1-b)w-a]\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}-d(p-\beta w)\bar{F}(d/[(1-b)w-a]),$$

that is, h represents the left-hand-side of the condition in Theorem 4.4.1 for the AWP policy. If we denote the equilibrium wholesale price under AWP by w^w, then clearly $h_{w=w^w}=0$. Thus,

$$\begin{aligned}\frac{\partial h}{\partial a}|_{w=w^w}=&\{\beta[(1-b)w-a]-[(b-2)\beta w+a\beta+p]\}\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\&+[(b-2)\beta w+a\beta+p]\bar{F}(d/[(1-b)w-a])d/[(1-b)w-a]\\&+d^2(p-\beta w)f(d/[(1-b)w-a])/[(1-b)w-a]^2|_{w=w^w}\\=&\{\beta[(1-b)w-a]-[(b-2)\beta w+a\beta+p]\}\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\&+\{[(b-2)\beta w+a\beta+p]^2/(p-\beta w)\}\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\&+d^2(p-\beta w)f(d/[(1-b)w-a])/[(1-b)w-a]^2|_{w=w^w}\\=&\{\beta^2[(1-b)w-a]^2/(p-\beta w)\}\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\&+d^2(p-\beta w)f(d/[(1-b)w-a])/[(1-b)w-a]^2|_{w=w^w}\geq 0,\end{aligned}$$

where the second equality follows from $h|_{w=w^w}=0$.

In addition, let $s=(1-b)w-a$, then

$$\frac{\partial h}{\partial a}=\frac{\partial h}{\partial s}\frac{\partial s}{\partial a},\ \frac{\partial h}{\partial b}=\frac{\partial h}{\partial s}\frac{\partial s}{\partial b},\ \text{and thus } \frac{\partial h}{\partial b}|_{w=w^w}=w\frac{\partial h}{\partial a}|_{w=w^w}\geq 0.$$

$$\frac{\partial h}{\partial w}=\frac{\partial h}{\partial s}\frac{\partial s}{\partial w}-\beta[(1-b)w-a]\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}+d\beta\bar{F}(d/[(1-b)w-a]).$$

By $h|_{w=w^w}=0$,

$$d\beta\bar{F}(d/[(1-b)w-a])|_{w=w^w}=\left[1-\frac{(1-b)w-a}{p-\beta w}\right]\beta[(1-b)w-a]\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}|_{w=w^w}.$$

Thus,

$$\frac{\partial h}{\partial w}|_{w=w^w}=-(1-b)\frac{\partial h}{\partial a}|_{w=w^w}-\frac{[(1-b)w-a]^2}{p-\beta w}\beta\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}|_{w=w^w}\leq 0.$$

By the implicit function theorem,

$$\frac{\partial w^w}{\partial a}=-(\frac{\partial h}{\partial w}|_{w=w^w})^{-1}\cdot\frac{\partial h}{\partial a}|_{w=w^w}\geq 0 \text{ and } \frac{\partial w^w}{\partial b}=-(\frac{\partial h}{\partial w}|_{w=w^w})^{-1}\cdot\frac{\partial h}{\partial b}|_{w=w^w}\geq 0.$$

Thus, w^w increases in a and b. Further, w^w is unaffected by n since h does not contain n.

Under the ASP policy, similarly, let

$$h=\{(\theta/n-\beta)[(1-b)w-a]+(\theta-\beta)w\}\mathbb{E}\{[(1-b)w-a]\epsilon\wedge d\}-d(\theta-\beta)w\bar{F}(d/[(1-b)w-a]),$$

that is, h represents the left-hand-side of the condition in Theorem 4.4.1 for the ASP policy. If we denote the equilibrium wholesale price under ASP by w^s, then clearly $h|_{w=w^s}=0$. Thus,

$$\begin{aligned}
&\frac{\partial h}{\partial a}|_{w=w^s}\\
=&-\{2(\theta/n-\beta)[(1-b)w-a]+(\theta-\beta)w\}\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\
&\qquad+\{(\theta/n-\beta)[(1-b)w-a]+(\theta-\beta)w\}\bar{F}(d/[(1-b)w-a])d/[(1-b)w-a]\\
&\qquad+d(\theta-\beta)wf(d/[(1-b)w-a])d/[(1-b)w-a]^2|_{w=w^s}\\
=&-\{2(\theta/n-\beta)[(1-b)w-a]+(\theta-\beta)w\}\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\
&\qquad+\{(\theta/n-\beta)[(1-b)w-a]+(\theta-\beta)w\}^2/[(\theta-\beta)w]\cdot\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\
&\qquad+d(\theta-\beta)wf(d/[(1-b)w-a])d/[(1-b)w-a]^2|_{w=w^s}\\
=&(\theta/n-\beta)^2[(1-b)w-a]^2/[(\theta-\beta)w]\cdot\mathbb{E}\{\epsilon\wedge d/[(1-b)w-a]\}\\
&\qquad+d(\theta-\beta)wf(d/[(1-b)w-a])d/[(1-b)w-a]^2|_{w=w^s}\geq 0,
\end{aligned}$$

where the second equality follows from $h|_{w=w^s} = 0$.

Again, let $s = (1-b)w - a$. By the same analysis as that in the AWP policy,

$$\frac{\partial h}{\partial b}|_{w=w^s} = w\frac{\partial h}{\partial a}|_{w=w^s} \geq 0 \text{ and } \frac{\partial h}{\partial n}|_{w=w^s} = -(\theta/n^2)[(1-b)w-a]\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\} \leq 0.$$

To show the monotonicity of h with respect to w, we note that in h, $\lim_{w\to\infty} \mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\} = d$ and $\lim_{w\to\infty} \bar{F}(d/[(1-b)w-a]) = 1$. Thus, $\lim_{w\to\infty} h = \lim_{w\to\infty}(\theta/n - \beta)[(1-b)w-a]d \leq 0$. Since h is continuous and there exists a unique w^s such that $h|_{w=w^s} = 0$, that $\lim_{w\to\infty} h \leq 0$ implies that $\partial h/\partial w|_{w=w^s} < 0$. A further application of the implicit function theorem similar to that in the AWP proof yields the desired result.

We next show the monotonicity of the equilibrium shortage status with respect to the parameters. We first note that the parameters, a, b, and n, affect the shortage only through the mean supply, $(1-b)w^w - a$ under AWP or $(1-b)w^s - a$ under ASP. Thus, investigating their impacts on the supply suffices. From the supply expressions, it may appear that the impacts of a and b are trivial. However, this is not the case since the equilibrium wholesale price w^w or w^s is also affected by a and b. Specifically, while large a and b values reduce the supply, they also boost the wholesale price which raises the supply. Therefore, identifying their impacts entails a detailed investigation.

To examine the impacts, we perform a transform: $s = (1-b)w - a$ and $c = a$, where s stands for the equilibrium mean supply. Similar to the proof of Theorem 4.4.1, it can be straightforwardly shown that in the transformed space, there also exists a unique symmetric equilibrium. We denote the equilibrium mean supply under AWP by s^w and that under ASP by s^s.

Under the AWP policy, the condition in Theorem 4.4.1 becomes:

$$[-\beta s - \beta(s+c)/(1-b) + p]\mathbb{E}\{s\epsilon \wedge d\} - d[p - \beta(s+c)/(1-b)]\bar{F}(d/s) = 0 \tag{C.1}$$

Let h be the left-hand-side of this equation, then

$$\begin{aligned}\frac{\partial h}{\partial c}|_{s=s^w} =& -\frac{\beta}{1-b}\mathbb{E}\{s\epsilon \wedge d\} + \frac{d\beta}{1-b}\bar{F}(d/s)|_{s=s^w}\\ =&\frac{-\beta}{(1-b)[p-\beta(s+c)/(1-b)]}\left[p-\frac{\beta(s+c)}{1-b}\right][\mathbb{E}\{s\epsilon\wedge d\} - d\bar{F}(d/s)]|_{s=s^w}\\ =&\frac{-\beta^2 s^2}{(1-b)[p-\beta(s+c)/(1-b)]}\mathbb{E}\{\epsilon\wedge d/s\}|_{s=s^w} \leq 0,\end{aligned}$$

where the third equality follows from (C.1).

In addition, if $w = (s+c)/(1-b)$ in (C.1), then clearly

$$\frac{\partial h}{\partial c} = \frac{\partial h}{\partial w}\frac{\partial w}{\partial c}, \frac{\partial h}{\partial b} = \frac{\partial h}{\partial w}\frac{\partial w}{\partial b}, \text{ and thus } \frac{\partial h}{\partial b}|_{s=s^w} = w\frac{\partial h}{\partial c}|_{s=s^w} \leq 0.$$

Further,

$$\begin{aligned}\frac{\partial h}{\partial s}|_{s=s^w} =& \left[-2\beta s - \frac{\beta(2s+c)}{1-b} + p\right]\mathbb{E}\{\epsilon\wedge d/s\} - \left[-\beta s - \frac{\beta(s+c)}{1-b} + p\right]\bar{F}(d/s)\frac{d}{s}\\ &+ \frac{d\beta}{1-b}\bar{F}(d/s) - d\left[p - \frac{\beta(s+c)}{1-b}\right]f(d/s)\frac{d}{s^2}|_{s=s^w}\\ =&\left[-\beta s - \frac{\beta(2s+c)}{1-b} + p\right]\left[\mathbb{E}\{\epsilon\wedge d/s\} - \bar{F}(d/s)\frac{d}{s}\right] - \beta s\mathbb{E}\{\epsilon\wedge d/s\}\\ &- \frac{d^2}{s^2}\left[p - \frac{\beta(s+c)}{1-b}\right]f(d/s)|_{s=s^w}\\ =&\frac{p-\beta s-\beta(2s+c)/(1-b)}{p-\beta(s+c)/(1-b)}\beta s\mathbb{E}\{\epsilon\wedge d/s\} - \beta s\mathbb{E}\{\epsilon\wedge d/s\}\\ &- \frac{d^2}{s^2}\left[p - \frac{\beta(s+c)}{1-b}\right]f(d/s)|_{s=s^w}\\ =& -\frac{\beta^2 s^2(2-b)}{[p-\beta(s+c)/(1-b)](1-b)}\mathbb{E}\{\epsilon\wedge d/s\} - \frac{d^2}{s^2}\left[p - \frac{\beta(s+c)}{1-b}\right]f(d/s)|_{s=s^w} \leq 0,\end{aligned}$$

where the third equality follows from (C.1).

Using the implicit function theorem and following the same procedure as that in the early part of this proof, we can show that under the AWP policy, the supply s decreases in a (i.e., c) and b. Further, the supply is unaffected by n since the condition (C.1) does not contain n.

Under the ASP policy, the condition in Theorem 4.4.1 becomes:

$$[(\theta/n-\beta)s+(\theta-\beta)(s+c)/(1-b)]\mathbb{E}\{s\epsilon\wedge d\}-d(\theta-\beta)[(s+c)/(1-b)]\bar{F}(d/s)=0 \quad \text{(C.2)}$$

Similarly, Let h be the left-hand-side of this equation, then

$$\begin{aligned}\frac{\partial h}{\partial c}|_{s=s^s} =&\frac{\theta-\beta}{1-b}\mathbb{E}\{s\epsilon\wedge d\}-\frac{d(\theta-\beta)}{1-b}\bar{F}(d/s)|_{s=s^s}\\ =&\frac{1}{s+c}\frac{(\theta-\beta)(s+c)}{1-b}[\mathbb{E}\{s\epsilon\wedge d\}-d\bar{F}(d/s)]|_{s=s^s}\\ =&\frac{-(\theta/n-\beta)s^2}{s+c}\mathbb{E}\{\epsilon\wedge d/s\}|_{s=s^s}\geq 0,\end{aligned}$$

where the third equality follows from (C.2) and the inequality follows from the requirement that $\beta>\theta/n$ (see the last paragraph of Section 4.3.3).

If we let $w=(s+c)/(1-b)$ in (C.2), then clearly

$$\frac{\partial h}{\partial c}=\frac{\partial h}{\partial w}\frac{\partial w}{\partial c},\frac{\partial h}{\partial b}=\frac{\partial h}{\partial w}\frac{\partial w}{\partial b}, \text{ and thus } \frac{\partial h}{\partial b}|_{s=s^s}=w\frac{\partial h}{\partial c}|_{s=s^s}\geq 0.$$

To show the monotonicity of h with respect to s, we note that in h, $\lim_{s\to\infty}\mathbb{E}\{s\epsilon\wedge d\}=d$ and $\lim_{s\to\infty}\bar{F}(d/s)=1$. Thus, $\lim_{s\to\infty}h=\lim_{s\to\infty}(\theta/n-\beta)sd\leq 0$ by the requirement that $\beta>\theta/n$. In addition, since h is continuous and there exists a unique s^s such that $h|_{s=s^s}=0$, that $\lim_{s\to\infty}h\leq 0$ implies $\partial h/\partial s|_{s=s^s}<0$. Again, by the implicit function theorem, the result holds.

C.4 Proof of Theorem 4.4.3

The proof is similar to that of Theorem 4.4.2. Let h be the same as that defined in the proof of Theorem 4.4.2. Under the AWP policy,

$$\begin{aligned}\frac{\partial h}{\partial \beta}|_{w=w^w} =&[(b-2)w+a]\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\} + dw\bar{F}(d/[(1-b)w-a])|_{w=w^w}\\ =&[(b-2)w+a]\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\}\\ &\qquad + \{[(b-2)\beta w + a\beta + p]w/(p-\beta w)\}\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\}|_{w=w^w}\\ =&-[(1-b)w-a]^2 p\mathbb{E}\{\epsilon \wedge d/[(1-b)w-a]\}/(p-\beta w)|_{w=w^w} \leq 0,\end{aligned}$$

where the second equality follows from $h|_{w=w^w}=0$. Similarly, under the ASP policy,

$$\begin{aligned}\frac{\partial h}{\partial \beta}|_{w=w^s} =&[(b-2)w+a]\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\} + dw\bar{F}(d/[(1-b)w-a])|_{w=w^s}\\ =&[(b-2)w+a]\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\}\\ &\qquad + \{(\theta/n-\beta)[(1-b)w-a]/(\theta-\beta)+w\}\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\}|_{w=w^s}\\ =&-(1-1/n)\theta[(1-b)w-a]^2\mathbb{E}\{\epsilon \wedge d/[(1-b)w-a]\}/(\theta-\beta)|_{w=w^s} \leq 0,\end{aligned}$$

where the second equality follows from $h|_{w=w^s}=0$. Applying the implicit function theorem in a similar manner to that in the proof of Theorem 4.4.2, we can show that both w^w and w^s decrease as β increases. Since the equilibrium supply $(1-b)w-a$ does not directly contain β, the decreasing wholesale price indicates that the supply decreases in β. Hence, shortage increases as the supply decreases.

In addition, under the ASP policy,

$$\begin{aligned}\frac{\partial h}{\partial \theta}|_{w=w^s} =&\{[(1-b)w-a]/n+w\}\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\} - dw\bar{F}(d/[(1-b)w-a])|_{w=w^s}\\ =&\{[(1-b)w-a]/n+w\}\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\}\\ &\qquad - \{(\theta/n-\beta)[(1-b)w-a]/(\theta-\beta)+w\}\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\}|_{w=w^s}\\ =&(1-1/n)\beta[(1-b)w-a]^2\mathbb{E}\{\epsilon \wedge d/[(1-b)w-a]\}/(\theta-\beta)|_{w=w^s} \geq 0.\end{aligned}$$

By a similar argument, we know that w^s increases and shortage decreases in θ.

C.5 Additional Details in Data Integration

When identifying shortage drugs, we use two authoritative lists: the ASHP shortage list and the FDA shortage list, where drugs are defined by using the Healthcare Common Procedure Coding System (HCPCS) code provided in the CMS (Centers for Medicare & Medicaid Services) data. If a drug has been listed as shortage for some period of time by either list since 2005, then the drug is identified as a shortage drug and included in our study. In the Medicare PUP data, each drug is identified by its HCPCS code, and each physician is identified by their National Provider Identifier. The AHA data surveyed on average 6266 hospitals a year from 2012 to 2016 and identified on average 107 GPOs every year. Most of these GPOs were small GPOs that were used by hospitals for ad hoc purchases. We integrate data from the four data sources (i.e., the ASHP and FDA shortage lists, Medicare PUP, and AHA data). In this data integration, we focus on U.S. healthcare providers who participate in Medicare, and drugs that are listed on Medicare Part B drug ASP file.

Since the data integration only matches about 10% of the physicians to GPOs. To match the remaining 90% of the physicians to GPOs, we test three common supervised learning methods: random forest, support vector machine, and neural network (multi-layer perceptron classifier), with their parameters tuned through cross validation. The features used in the classification include state, provider type, place of service, entity code, average Medicare allowed payment amount, average submitted charge amount, and average Medicare payment amount. We encode categorical variables by one-hot encoding. As noticed in previous studies as well as in our own data analysis, physicians often purchase a drug through a single GPO while using other small GPOs for ad hoc and one-time purchases (e.g., [182] and [183] stated that healthcare providers "route most of their purchases through a single national alliance" and "utilize (another) only for specific contracts in limited supply areas"). Thus, for each physician in the AHA hospital–GPO affiliation data, we identify the major GPO the physician uses (i.e., the one with the most frequent usage and largest purchase volume) and use it as the physician's GPO label.

As a common procedure, we normalize all features before the classification. Since the samples are unbalanced, we further adjust the sample weights of the training data to be inversely proportional to class frequencies when applicable. For each drug, we use five-fold cross validation to find the best tuning parameter values under a learning method, and then obtain the prediction accuracy with the best performing parameter values. The prediction accuracy is computed as the fraction of correct predictions (then averaged across all shortage drugs). For drugs with overly small training samples (i.e., training samples with less than 20 data points), we do not cross validate them but instead set appropriate tuning parameter values for them.

Our testing shows that the accuracy measures of random forest, support vector machine, and neural network are 46.6%, 49.2% and 48.0%, respectively. We thus adopt support vector machine since it has the highest accuracy. Through the data classification process, we are able to match all physicians to GPOs.

C.6 Proof of Proposition 4.6.1

We first note that GPOs' price consciousness, $g(n, \beta)$, is fixed under fixed n and β. Thus, all the original theoretical results hold except when we study the impacts of n and β. In other words, the existence and uniqueness of the equilibrium and the impacts of a, b and θ are examined with respect to fixed n and β (i.e., to examine the impact of a parameter, we have varied one parameter at a time), and hence the corresponding results stay the same. For the impacts of n and β, we first note that the impact of n and β on the equilibrium wholesale price and shortage under AWP and the impact of β under ASP can be straightforwardly shown by using the chain rule. We thus omit the details. We next analyze the impact of n under ASP. Let

$$\begin{aligned} h(n,\beta) = \{(\theta/n - g(n,\beta))[(1-b)w-a] + (\theta - g(n,\beta))w\}\mathbb{E}\{[(1-b)w-a]\epsilon \wedge d\} \\ - d(\theta - g(n,\beta))w\bar{F}(d/((1-b)w-a)) \end{aligned}$$

be the left hand side of the first order condition under ASP. Furthermore, with slight abuse of notation, let $g = g(n, \beta)$ and let

$$h(n,\beta,g) = \{(\theta/n-g)[(1-b)w-a]+(\theta-g)w\}\mathbb{E}\{[(1-b)w-a]\epsilon\wedge d\}-d(\theta-g)w\bar{F}(d/((1-b)w-a)).$$

Then by the rule of total derivative and chain rule, we have

$$\frac{\partial h(n,\beta)}{\partial n} = \frac{\partial h(n,\beta,g)}{\partial n} + \frac{\partial h(n,\beta,g)}{\partial g}\frac{\partial g(n,\beta)}{\partial n}.$$

One can verify that

$$\frac{\partial h(n,\beta,g)}{\partial n}|_{w=w^s} = -(\theta/n^2)[(1-b)w^s - a]\mathbb{E}\{[(1-b)w^s - a]\epsilon \wedge d\} \le 0,$$

$$\frac{\partial h(n,\beta,g)}{\partial g}|_{w=w^s} = -(1-1/n)\theta[(1-b)w^s - a]^2\mathbb{E}\{\epsilon \wedge d/[(1-b)w^s - a]\}/(\theta - g)|_{w=w^s} \le 0,$$

and by our assumption,

$$\frac{\partial g(n,\beta)}{\partial n} \ge 0, \text{ so } \frac{\partial h(n,\beta)}{\partial n}|_{w=w^s} \le 0.$$

Furthermore, it is straightforward to verify that $\partial h(n,\beta)/\partial w|_{w=w^s} < 0$. An application of the implicit function theorem shows that the equilibrium wholesale price under ASP decreases in n.

Since the equilibrium shortage status depends on the equilibrium supply $\mathbb{E}[(1-b)w^s - a]$, which further depends on w^s when a and b are fixed, the equilibrium shortage increases as n increases.

VITA

Xuejun Zhao was born and raised in Chongqing, China. She attended Chongqing Bashu High School. She received her B.E. degree from Xi'an Jiaotong University, Xi'an, China. Xuejun then obtained her M.S. in Mechanical Engineering from Purdue University. She started her doctoral study in 2017 and is currently a Ph.D. candidate in Operations Management at Purdue University under the supervision of Dr. William B. Haskell.