Learning Algorithms for Stochastic Dynamic Pricing and Inventory Control

by

Boxiao Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2016

Doctoral Committee:

Professor Xiuli Chao, Co-Chair
Professor Hyun-Soo Ahn, Co-Chair
Associate Professor Brian Denton
Assistant Professor Cong Shi
Associate Professor Xun Wu

ACKNOWLEDGEMENTS

First of all, I would like to thank my dissertation co-chairs Prof. Xiuli Chao and Prof. Hyun-Soo Ahn for their time and effort in guiding me through this Ph.D. journey. Without their support and help, this dissertation would not have been possible. One chapter in the dissertation also involves Prof. Cong Shi, from whom I have learned a lot. I would also like to thank my committee members Prof. Brian Denton and Prof. Xun Wu, for their helpful discussions and feedback.

My gratitude also goes to professors in the IOE department from whose lectures I have learned methodologies and techniques to conduct my research, including Prof. Marina Epelman, Prof. Jon Lee, Prof. Edwin Romeijn, Prof. Romesh Saigal and Prof. Robert Smith. And I also owe my thanks to staff members of the department for their assistance and help.

I appreciate the friendship from Prof. Xiuli Chao's research group, which includes former members Dr. Xiting Gong, Dr. Jingchen Wu, Dr. Gregory King and Dr. Majid Al-Gwaiz, and current members Huanan Zhang, Sentao Miao and Duo Xu. Their feedback and comments help me a lot in my processes of doing researches and job hunting. I would also like to thank all my other friends for their caring and sharing.

Finally, I would like to thank my dearest parents for their unconditional love that carries my through many difficulties during my Ph.D. study. Their encouragement and support have always been the best source of energy to keep me moving forward, not only in pursuing my Ph.D., but also in life.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Learning Algorithms for Stochastic Dynamic Pricing and Inventory Control

by

Boxiao Chen

Chair: Xiuli Chao, Hyun-Soo Ahn

This dissertation considers joint pricing and inventory control problems in which the customer's response to selling price and the demand distribution are not known a priori, and the only available information for decision-making is the past sales data. Data-driven algorithms are developed and proved to converge to the true clairvoyant optimal policy had decision maker known the demand processes a priori, and, for the first time in literature, this dissertation provides theoretical results on the convergence rate of these data-driven algorithms.

Under this general framework, several problems are studied in different settings. Chapter 2 studies the classical joint pricing and inventory control problem with backlogged demand, and proposes a nonparametric data-driven algorithm that learns about the demand on the fly while making pricing and ordering decisions. The performance of the algorithm is measured by regret, which is the average profit loss compared with that of the clairvoyant optimal policy. It is proved that the regret vanishes at the fastest possible rate as the planning horizon increases.

Chapter 3 studies the classical joint pricing and inventory control problem with lost-sales and censored demand. Major challenges in this study include the following:

First, due to demand censoring, the firm cannot observe either the realized demand or realized profit in case of a stockout, therefore only biased data is accessible; second, the data-driven objective function is always multimodal, which is hard to solve and establish convergence for. Chapter 3 presents a data-driven algorithm that actively explores in the inventory space to collect more demand data, and designs a sparse discretization scheme to jointly learn and optimize the multimodal data-driven objective. The algorithm is shown to be very computationally efficient.

Chapter 4 considers a constraint that only allows the firm to change prices no more than a certain number of times, and explores the impact of number of price changes on the quality of demand learning. In the data-driven algorithm, we extend the traditional maximum likelihood estimation method to work with censored demand data, and prove that the algorithm converges at the best possible rate for any data-driven algorithms.

CHAPTER I

Introduction

Firms often integrate inventory and pricing decisions to match demand with supply. For instance, a firm may offer a discounted price when there is excess inventory or raise the price when the inventory level is low. Since the seminal paper of *Whitin* (1955), the joint pricing and inventory control problems have attracted significant attention in the field (see, e.g., the survey papers by *Petruzzi and Dada* (1999), *Elmaghraby and Keskinocak* (2003), *Yano and Gilbert* (2003), *Chen and Simchi-Levi* (2012)). Almost all papers on this topic assume that the firm knows how the market responds to its selling prices and the exact distribution of uncertainty in customer demand, and the inventory and pricing decisions are made with full knowledge of the underlying demand process. However, in practice, the demand-price relationship is usually not known *a priori*. Indeed, even with past observed demand data (often censored in the lost-sales case), it remains difficult to select the most appropriate functional form and estimate the distribution of demand uncertainty (see *Huh and Rusmevichientong* (2009), *Huh et al.* (2011), *Besbes and Muharremoglu* (2013), *Shi et al.* (2015) for more discussions on censored demand in various other inventory systems).

In Chapter 2, we consider a firm (e.g., retailer) selling a single nonperishable product over a finite-period planning horizon. Demand in each period is stochastic and

price-dependent, and unsatisfied demands are backlogged. At the beginning of each period, the firm determines its selling price and inventory replenishment quantity, but it knows neither the form of demand dependency on selling price nor the distribution of demand uncertainty a priori, hence it has to make pricing and ordering decisions based on historical demand data. We propose a nonparametric data-driven policy that learns about the demand on the fly and, concurrently, applies learned information to determine replenishment and pricing decisions. The policy integrates learning and action in a sense that the firm actively experiments on pricing and inventory levels to collect demand information with the least possible profit loss. Besides convergence of optimal policies, we show that the regret, defined as the average profit loss compared with that of the clairvoyant optimal solution when the firm had complete information about the underlying demand, vanishes at the fastest possible rate as the planning horizon increases.

In Chapter 3, we consider the classical joint pricing and inventory control problem with lost-sales and censored demand in which the customer's response to selling price and the demand distribution are not known a priori, and the only available information for decision-making is the past sales data. Conventional approaches, such as stochastic approximation, online convex optimization, and continuum-armed bandit algorithms, cannot be employed since neither the realized values of the profit function nor its derivatives are known. A major difficulty of this problem lies in the fact that the estimated profit function from observed sales data is multimodal even when the expected profit function is concave. We develop a nonparametric data-driven algorithm that actively integrates exploration and exploitation through carefully designed cycles. The algorithm searches the decision space through a sparse discretization scheme to jointly learn and optimize a multimodal (sampled) profit function, and corrects the estimation biases caused by demand censoring. We show that the algorithm converges to the clairvoyant optimal policy as the planning hori-

zon increases, and obtain the convergence rate of regret. Numerical experiments show that the proposed algorithm performs very well.

In Chapter 4, we consider a firm selling a product over T periods. Demand in each period is random and price sensitive, and unsatisfied demands are lost and unobservable. The firm has limited prior knowledge about the demand process and needs to learn it through historical sales data. We consider the scenario where the firm is faced with the business constraint that prevents it from conducting extensive price experimentation. We develop data-driven algorithms for pricing and inventory decisions and evaluate their effectiveness using regret, which is the profit loss compared to a clairvoyant who has complete information about the demand process. We study three distinct scenarios and design algorithms that achieve the lowest possible regret rates: First, in a quite general case, when the number of price changes is bounded by a given number, the regret is $\mathcal{O}(T^{1/2})$. Second, in a special so-called well-separated case, when the number of price changes is limited to m, the regret is $\mathcal{O}(T^{1/m+1})$. Third, in the well-separated case allowing more frequent price changes that is limited to $\mathcal{O}(\log T)$, the regret is $\mathcal{O}(\log T)$. Numerical results show that these algorithms empirically perform very well.

CHAPTER II

Coordinating Pricing and Inventory Replenishment with Nonparametric Demand Learning

2.1 Introduction

Balancing supply and demand is a challenge for all firms, and failure to do so can directly affect the bottom-line of a company. From the supply side, firms can use operational levers such as production and inventory decisions to adjust inventory level in pace of uncertain demand. From the demand side, firms can deploy marketing levers such as pricing and promotional decisions to shape the demand to better allocate the limited (or excess) inventory in the most profitable way. With the increasing availability of demand data and new technologies, e.g., electronic data interchange, point of sale devices, click stream data etc., deploying both operational and marketing levers simultaneously is now possible. Indeed, both academics and practitioners have recognized that substantial benefits can be obtained from coordinating operational and pricing decisions. As a result, the research literature on joint pricing and inventory decisions has rapidly grown in recent years, see, e.g., the survey papers by *Petruzzi and Dada* (1999), *Elmaghraby and Keskinocak* (2003), *Yano and Gilbert* (2003), and *Chen and Simchi-Levi* (2012).

Despite the voluminous literature, the majority of the papers on joint optimization of pricing and inventory control have assumed that the firm knows how the market responds to its selling prices and the exact distribution of uncertainty in customer demand for any given price. This is not true in many applications, particularly with demand of new products. In such settings, the firm needs to learn about demand information during the dynamic decision making process and simultaneously tries to maximize its profit.

In this chapter, we consider a firm selling a nonperishable product over a finite-period planning horizon in a make-to-stock setting that allows backlogs. In each period, the firm sets its price and inventory level in anticipation of price-sensitive and uncertain demand. If the firm had complete information about the underlying demand distribution, this problem has been studied by, e.g., *Federgruen and Heching* (1999), among others. The point of departure this paper takes is that the firm possesses limited or even no prior knowledge about customer demand such as its dependency on selling price or the distribution of uncertainty in demand fluctuation. We develop a nonparametric data-driven algorithm that learns the demand-price relationship and the random error distribution on the fly. We also establish the convergence rate of the regret, defined as the average profit loss per period of time compared with that of the optimal solution had the firm known the random demand information, and that is fastest possible for any learning algorithm. This work is the first to present a nonparametric data-driven algorithm for the classic joint pricing and inventory control problem that not only shows the convergence of the proposed policies but also the convergence rate for regret.

2.1.1 Literature Review

Almost all early papers in joint pricing and inventory control, e.g., *Whitin* (1955), *Federgruen and Heching* (1999), and *Chen and Simchi-Levi* (2004a), among others,

assume that a firm has complete knowledge about the distribution of underlying stochastic demand for any given selling price. The complete information assumption provides analytic tractability necessary for characterizing the optimal policy. The extension to the parametric case (the firm knows the class of distribution but not the parameters) has been studied by, for example, *Subrahmanyan and Shoemaker* (1996), *Petruzzi and Dada* (2002), and *Zhang and Chen* (2006). *Chung et al.* (2011) also consider the problem of dynamic pricing and inventory planning with demand learning, and they develop learning algorithms using Bayesian method and Markov chain Monte Carlo (MCMC) algorithms, and numerically evaluate the importance of dynamic pricing. An alternative to the parametric approach is to model the firm's problem in a nonparametric setting. Under this framework, the firm does not make specific assumptions about underlying demand. Instead, the firm makes decisions solely based on the collected demand data, see *Burnetas and Smith* (2000). Our work falls into this category.

To our best knowledge, *Burnetas and Smith* (2000) is the only paper that considers the joint pricing and inventory control problem in a nonparametric setting. The authors consider a make-to-stock system for a *perishable* product with lost sales and linear costs, and propose an adaptive policy to maximize average profit. They assume that the price is chosen from a finite set and formulate the pricing problem as a multi-armed bandit problem, and show that the average profit under their approximation policy converges in probability. No convergence rate or performance bound is obtained for their algorithm.

Other approaches in the literature on developing nonparametric data-driven algorithms include online convex optimization (*Agarwal et al.* (2011), *Zinkevich* (2003), *Hazan et al.* (2007)), continuum-armed bandit problems (*Auer et al.* (2007), *Kleinberg* (2005), *Cope* (2009)), and stochastic approximation (*Kiefer and Wolfowitz* (1952b), *Lai and Robbins* (1981), and *Robbins and Monro* (1951)). In fact, *Burnetas and*

Smith (2000) is an example of implementing such algorithms to the joint pricing and inventory control problem. However, these methodologies require that the proposed solution be reachable in each and every period, which is not the case with our problem. This is because, in a demand learning algorithm of joint pricing/inventory control problem, in each period the algorithm utilizes the past demand data to prescribe a pricing decision and an order up-to level. However, if the starting inventory level of the period is already higher than the prescribed order up-to level, then the prescribed inventory level for the period cannot be reached. Actually, that is precisely the reason that *Burnetas and Smith* (2000) focused on the case of perishable product (hence the firm has no carry-over inventory and the inventory decision obtained by *Burnetas and Smith* (2000) based on multi-armed bandit process can be implemented in each period). *Agarwal et al.* (2011), *Auer et al.* (2007), and *Kleinberg* (2005) propose learning algorithms and obtain regrets that are not as good as ours in this chapter. *Zinkevich* (2003) and *Hazan et al.* (2007) present machine learning algorithms in which the the exact gradient of the unknown objective function at the current decision can be computed, and their results have been applied to dynamic inventory control in *Huh and Rusmevichientong* (2009). However, in the joint pricing and inventory control problem with unknown demand response, the gradient of the unknown objective function cannot be obtained thus the method cannot be applied.

2.1.2 Contributions and Comparison with Closely Related Literature

The closest related research works to ours are *Besbes and Zeevi* (2015), *Levi et al.* (2007) and *Levi et al.* (2011), offering nonparametric approaches to pure pricing problem (with no inventory) and pure inventory control problem (with no pricing), respectively.

Besbes and Zeevi (2015) consider a dynamic pricing problem in which a firm chooses its selling price to maximize expected revenue. The firm does not know

the deterministic demand curve (i.e., how the average demand changes in price) and learns it through noisy demand realizations, and the authors establish the sufficiency of linear approximations in maximizing revenue. They assume that the firm has infinite supply of inventory, or, alternatively, the seller has no inventory constraint. In this case, since the expected revenue in each period depends only on its mean demand, the distribution of random error is immaterial in their learning algorithm and analysis. On the other hand, in the dynamic newsvendor problem considered in *Levi et al.* (2007, 2010), the essence for effective inventory management is to strike a balance between overage cost and underage cost, for which the distribution of uncertain demand plays a key role. *Levi et al.* (2007) and *Levi et al.* (2011) apply Sample Average Approximation (SAA) to estimate the demand distribution and average cost function, and they analyze the relationship between sample sizes and accuracy of estimations and inventory decisions.

Our problem has both dynamic pricing and inventory control, and the firm knows neither the relationship between demand and selling price nor the distribution of demand uncertainty. In *Besbes and Zeevi* (2015), the authors only need to estimate the average demand curve in order to maximize revenue, and demand distribution information is irrelevant. In a remark, *Besbes and Zeevi* (2015) state that their method of learning the demand curve can be applied to maximizing more general forms of objective functions beyond the expected revenue which, however, does not apply to our setting. This is because, in the general form presented in *Besbes and Zeevi* (2015), the objective function still has to be a known function in terms of price and the demand curve for a given price and a given demand curve. Thus the firm must know the exact expression of the objective function when the estimate of a demand curve is given. In our problem, even with a given price and inventory level and a given demand curve, the objective function cannot be written as a known deterministic function. Indeed, this function contains the expected inventory holding and backorder costs

that depend on the distribution of demand fluctuation, which is also unknown to the firm. In fact, the latter is a major technical challenge encountered in this chapter because, as we will explain below, the estimation of the demand uncertainty, therefore also of the expected holding/shortage cost, cannot be decoupled with the estimation of the average demand curve, which is gathered through price experimentation.

Standard SAA method is implemented to the newsvendor problem by *Levi et al.* (2007) and *Levi et al.* (2011) which, however, cannot be applied to our setting for determining inventory decisions. In *Levi et al.* (2007) and *Levi et al.* (2011), dynamic inventory control is studied in which pricing is not a decision and it is assumed (implicitly) to be given. The only information the firm is uncertain about is the distribution of random fluctuation. Therefore, the firm can observe true realizations of demand fluctuation which are used to build an empirical distribution. In our model, however, the firm knows neither how average demand responds to the selling price (demand curve) nor the distribution of fluctuating demand, but both of them affect demand realizations. For any estimation of average demand curve, the error of this estimate will affect the estimation of distribution of random demand fluctuation. Hence, through the realization of random demand we are unable to obtain a true realization of random demand error without knowing the exact average demand function. As a result, the standard SAA analysis is not applicable in our setting because unbiased samples of the random error cannot be obtained.

Because the firm does not know the exact demand curve *a priori*, its estimate of error distribution using demand data is inevitably biased, and as a result, the data-driven optimization problem constructed to compute the pricing and ordering strategies is also biased. Because of this bias, it is no longer true that the solution of the data-driven problem using SAA must converge to the true optimal solution. Fortunately, we are able to show that as the learning algorithm proceeds, the biases will be gradually diminishing and that allows us to prove that our learning algorithm

still converges to the true optimal solution. This is done by establishing several important properties of the newsvendor problem that bound the errors of biased samples. One main contribution of this chapter is to explicitly prove that the solution obtained from a biased data-driven optimization problem still converges to the true optimal solution.

Finally, we highlight on the result of the convergence rate of regret. *Besbes and Zeevi* (2015) obtain a convergence rate of $T^{-1/2}(\log T)^2$ for their dynamic pricing problem, where T is the length of the planning horizon. For the pure dynamic inventory control problem, *Huh and Rusmevichientong* (2009) present a machine learning algorithm with convergence rate $T^{-1/2}$. For the joint pricing and inventory problem, we show that the regret of our learning algorithm converges to zero at rate $T^{-1/2}$, which is also the theoretical lower bound. Thus, this chapter strengthens and extends the existing work by achieving the tightest convergence rate for the problem with joint pricing and inventory control. One important implication of our finding is that the linear demand approximation scheme of *Besbes and Zeevi* (2015) actually achieves the best possible convergence rate of regret, which further improves the result of *Besbes and Zeevi* (2015). That is, nothing is lost in the learning algorithm in approximating the demand curve by a linear model.

2.1.3 Organization

The rest of this chapter is organized as follows. Section 2.2 formulates the problem and describes the data-driven learning algorithm for pricing and inventory control decisions. The following two sections (Sections 3 and 4) present our major theoretical results together with a numerical study, and the main steps of the technical proofs, respectively. The chapter concludes with a few remarks in Section 5. Finally, the details of the mathematical proofs are given in the Appendix.

2.2 Formulation and Learning Algorithm

We consider an inventory system in which a firm (e.g., a retailer) sells a nonperishable product over a planning horizon of T periods. At the beginning of each period t, the firm makes a replenishment decision, denoted by the order-up-to level, y_t, and a pricing decision, denoted by p_t, where $y_t \in \mathcal{Y} = [y^l, y^h]$ and $p_t \in \mathcal{P} = [p^l, p^h]$ for some known lower and upper bounds of inventory level and selling price, respectively. We assume $p^h > p^l$ since otherwise, the problem is the pure inventory control problem and learning algorithms have been developed in *Huh and Rusmevichientong* (2009), *Levi et al.* (2007), and *Levi et al.* (2011). During period t and when the selling price is set to p_t, a random demand, denoted by $\tilde{D}_t(p_t)$, is realized and fulfilled from on-hand inventory. Any leftover inventory is carried over to the next period, and in case the demand exceeds y_t, the unsatisfied demand is backlogged. The replenishment lead-time is zero, i.e., an order placed at the beginning of a period can be used to satisfy demand in the same period. Let h and b be the unit holding and backlog costs per period, and the unit purchasing cost is assumed, without loss of generality, to be zero.

The model as described above is the well-known joint inventory and pricing decision problem studied in *Federgruen and Heching* (1999), in which it is assumed that the firm has complete information about the distribution of $\tilde{D}_t(p_t)$. In this chapter we consider a setting where the firm does not have prior knowledge about the demand distribution.

In general, the demand in period t is a function of selling price p_t in that period and some random variable $\tilde{\epsilon}_t$, and it is stochastically decreasing in p_t. The most popular demand models in the literature are the additive demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t) + \tilde{\epsilon}_t$ and multiplicative demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\ \tilde{\epsilon}_t$, where $\tilde{\lambda}(\cdot)$ is a strictly decreasing deterministic function and $\tilde{\epsilon}_t, t = 1, 2, \ldots, T$, are independent and identically distributed random variables. In this chapter, we shall study both additive and the multiplicative demand models. However, the firm knows neither the function $\tilde{\lambda}(p_t)$

nor the distribution function of random variable $\tilde{\epsilon}_t$. The firm has to learn from historical demand data, that are the realizations of market responses to offered prices, and use that information as a basis for decision making. Suppose $\tilde{\epsilon}_t$ has finite support $[l, u]$, with $l \geq 0$ for the case of multiplicative demand.

To define the firm's problem, we let x_t denote the inventory level at the beginning of period t before replenishment decision. We assume that the system is initially empty, i.e., $x_1 = 0$. The system dynamics are $x_{t+1} = y_t - \tilde{D}_t(p_t)$ for all $t = 1, \ldots, T$. An admissible policy is represented by a sequence of prices and order-up-to levels, $\{(p_t, y_t), t \geq 1\}$, where (p_t, y_t) depends only on realized demand and decisions made prior to period t, and $y_t \geq x_t$, i.e., (p_t, y_t) is adapted to the filtration generated by $\{(p_s, y_s), \tilde{D}_s(p_s); s = 1, \ldots, t-1\}$. The firm's objective is to find an admissible policy to maximize its total profit.

If both the function of $\tilde{\lambda}(\cdot)$ and the distribution of $\tilde{\epsilon}_t$ are known a priori to the firm (complete information scenario), then the optimization problem the firm wishes to solve is

$$\max_{\substack{(p_t, y_t) \in \mathcal{P} \times \mathcal{Y} \\ y_t \geq x_t}} \sum_{t=1}^{T} \Big(p_t \mathbb{E}[\tilde{D}_t(p_t)] - h\mathbb{E}[y_t - \tilde{D}_t(p_t)]^+ - b\mathbb{E}[\tilde{D}_t(p_t) - y_t]^+ \Big), \tag{2.1}$$

where $\mathbb{E}$ stands for mathematical expectation with respect to random demand $\tilde{D}_t(p_t)$, and $x^+ = \max\{x, 0\}$ for any real number x. However, since in our setting the firm does not know the demand distribution, the firm is unable to evaluate the objective function of this optimization problem.

We develop a data-driven learning algorithm to compute the inventory control and pricing policy. It will be shown in Section 3 that the average profit of the algorithm converges to that of the case when complete demand distribution information is known a priori, and that the pricing and inventory control parameters also converge to that of the optimal control policy for the case with complete information as the

planning horizon becomes long. To save space we shall only present the algorithm and analytical results for the multiplicative demand model. The results and analyses for the additive demand case are analogous, and we only highlight the main differences at the end of this section.

Remark 1. For ease of exposition, in this chapter we assume the support of uncertainty $\tilde{\epsilon}_t$ is bounded. This can be relaxed, and all the results hold as long as we assume the moment generating functions of the relevant random variables are finite in a small neighborhood of 0, or light tailed.

Case of complete information about demand. In the case of complete information in which the firm knows $\tilde{\lambda}(\cdot)$ and the distribution of $\tilde{\epsilon}_t$, it follows from (2.1) that, if (p^*, y^*) is the optimal solution of each individual term

$$\max_{p\in\mathcal{P}, y\in\mathcal{Y}} \Big\{ p\mathbb{E}[\tilde{D}_t(p)] - h\mathbb{E}[y - \tilde{D}_t(p)]^+ - b\mathbb{E}[\tilde{D}_t(p) - y]^+ \Big\}. \tag{2.2}$$

and that this solution is reachable in every period, i.e., $x_t \le y^*$ for all t, then (p^*, y^*) is the optimal policy for each period. We refer to p^* and y^* as the optimal price and optimal order up-to level (or optimal base-stock level), respectively. It is clear that the reachability condition is satisfied if the system is initially empty, which we assume.

We find it convenient to analyze (2.2) using a slightly different but equivalent form. Taking logarithm on both sides of $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\tilde{\epsilon}_t$, we obtain

$$\log \tilde{D}_t(p_t) = \log \tilde{\lambda}(p_t) + \log \tilde{\epsilon}_t, \qquad t = 1, \ldots, T.$$

Denote $D_t(p_t) = \log \tilde{D}_t(p_t), \lambda(p_t) = \log \tilde{\lambda}(p_t)$ and $\epsilon_t = \log \tilde{\epsilon}_t$. Then, the logarithm of demand can be written as

$$D_t(p_t) = \lambda(p_t) + \epsilon_t, \qquad t = 1, \ldots, T. \tag{2.3}$$

We shall refer to $\lambda(\cdot)$ as the demand-price function (or demand-price curve) and ϵ_t as random error (or random shock). Clearly, $\lambda(\cdot)$ is also strictly decreasing in $p \in \mathcal{P}$. Hence, in the case of complete information, the firm knows the function $\lambda(\cdot)$ and the distribution of ϵ_t, and when the firm does not know function $\lambda(\cdot)$ and the distribution of ϵ_t, which is our case, the firm will need to learn about them. Without loss of generality, we assume $\mathbb{E}[\epsilon_t] = \mathbb{E}[\log \tilde{\epsilon}_t] = 0$. If this is not the case, i.e., $\mathbb{E}[\log \tilde{\epsilon}_t] = a \neq 0$, then $\mathbb{E}[\log(e^{-\alpha}\tilde{\epsilon}_t)] = 0$, thus if we let $\hat{\lambda}(\cdot) = e^a\tilde{\lambda}(\cdot)$ and $\hat{\epsilon}_t = e^{-a}\tilde{\epsilon}_t$, then $\tilde{D}_t(p_t) = \hat{\lambda}(p_t)\hat{\epsilon}_t$, and $\hat{\lambda}(\cdot)$ and $\hat{\epsilon}_t$ satisfy the desired properties.

For convenience, let ϵ be a random variable distributed as ϵ_1. In terms of $\lambda(\cdot)$ and ϵ, we define

$$G(p, y) = pe^{\lambda(p)}\mathbb{E}\left[e^{\epsilon}\right] - \left\{h\mathbb{E}\left[y - e^{\lambda(p)}e^{\epsilon}\right]^+ + b\mathbb{E}\left[e^{\lambda(p)}e^{\epsilon} - y\right]^+\right\}.$$

Then problem (2.2) can be re-written as

$$\textbf{Problem CI:} \quad \max_{p\in\mathcal{P}, y\in\mathcal{Y}} G(p, y) \tag{2.4}$$
$$= \max_{p\in\mathcal{P}} \left\{pe^{\lambda(p)}\mathbb{E}\left[e^{\epsilon}\right] - \min_{y\in\mathcal{Y}} \left\{h\mathbb{E}\left[y - e^{\lambda(p)}e^{\epsilon}\right]^+ + b\mathbb{E}\left[e^{\lambda(p)}e^{\epsilon} - y\right]^+\right\}\right\}.$$

The inner optimization problem (minimization) determines the optimal order-up-to level that minimizes the expected inventory and backlog cost for given price p, and we denote it by $\overline{y}\left(e^{\lambda(p)}\right)$. The outer optimization solves for the optimal price p. Let the optimal solution for (2.4) be denoted by p^* and y^*, then they satisfy $y^* = \overline{y}(e^{\lambda(p^*)})$.

The analysis above stipulates that the firm knows the demand-price curve $\lambda(p)$ and the distribution of ϵ, thus we refer to it as problem CI (complete information).

Learning algorithm. In the absence of the prior knowledge about the demand process, the firm needs to collect the demand information necessary to estimate $\lambda(p)$ and the empirical distribution of random error ϵ, thus price and inventory decisions not only affect the profit but also the demand information realized. The major dif-

ficulty lies in that, the estimations of demand-price curve $\lambda(p)$ and the distribution of random error cannot be decoupled. This is because, the firm only observes realized demands, hence with any estimation of demand-price curve, the estimation error transfers to the estimation of the random error distribution. Indeed, we are not even able to obtain unbiased samples of the random error ϵ_t.

In our algorithm below we approximate $\lambda(p)$ by an affine function, and construct an empirical (but biased) error distribution using the collected data. We divide the planning horizon into stages whose lengths are exponentially increasing (in the stage index). At the start of each stage, the firm sets two pairs of prices and order-up-to levels based on its current linear estimation of demand-price curve and (biased) empirical distribution of random error, and the collected demand data from this stage are used to update the linear estimation of demand-price curve and the biased empirical distribution of random error. These are then utilized to find the pricing and inventory decision for the next stage.

The algorithm requires some input parameters v, ρ and I_0, with $v > 1$, $I_0 > 0$, and $0 < \rho \leq 2^{-3/4}(p^h - p^l)I_0^{1/4}$. To initiate the algorithm, it sets $\{\hat{p}_1, \hat{y}_{11}, \hat{y}_{12}\}$, where $\hat{p}_1 \in \mathcal{P}, \hat{y}_{11} \in \mathcal{Y}, \hat{y}_{12} \in \mathcal{Y}$ are the starting pricing and order-up-to levels. For $i \geq 1$, let

$$I_i = \lfloor I_0 v^i \rfloor, \quad \delta_i = \rho(2I_{i-1})^{-\frac{1}{4}}, \text{ and } t_i = \sum_{k=1}^{i-1} 2I_k \text{ with } t_1 = 0, \tag{2.5}$$

where $\lfloor I_0 v^i \rfloor$ is the largest integer less than or equal to $I_0 v^i$.

The following is the detailed procedure of the algorithm. Recall that x_t is the starting inventory level at the beginning of period t, p_t is the selling price set for period t, and y_t ($\geq x_t$) is the order-up-to inventory level for period t, $t = 1, \ldots, T$. The number of learning stages is $n = \left\lceil \log_v \left(\frac{v-1}{2I_0 v} T + 1 \right) \right\rceil$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

Data-Driven Algorithm (DDA)

Step 0. Initialization. Choose $v > 1$, $\rho > 0$ and $I_0 > 0$, and $\hat{p}_1, \hat{y}_{11}, \hat{y}_{12}$. Compute $I_1 = \lfloor I_0 v \rfloor$, $\delta_1 = \rho(2I_0)^{-\frac{1}{4}}$, and $\hat{p}_1 + \delta_1$.

Step 1. Setting prices and order-up-to levels for stage i. For $i = 1, \ldots, n$, set prices p_t, $t = t_i + 1, \ldots, t_i + 2I_i$, to

$$\begin{aligned} p_t &= \hat{p}_i, \qquad t = t_i + 1, \ldots, t_i + I_i, \\ p_t &= \hat{p}_i + \delta_i, \qquad t = t_i + I_i + 1, \ldots, t_i + 2I_i; \end{aligned}$$

and for $t = t_i + 1, \ldots, t_i + 2I_i$, raise the inventory levels to

$$\begin{aligned} y_t &= \max\{\hat{y}_{i1}, x_t\}, \qquad t = t_i + 1, \ldots, t_i + I_i, \\ y_t &= \max\{\hat{y}_{i2}, x_t\}, \qquad t = t_i + I_i + 1, \ldots, t_i + 2I_i. \end{aligned}$$

Step 2. Estimating the demand-price function and random errors using data from stage i. Let $D_t = \log \tilde{D}_t(p_t)$ be the logarithm of demand realizations for $t = t_i + 1, \ldots, t_i + 2I_i$, and compute

$$(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = \underset{\alpha,\beta}{\operatorname{argmin}} \left\{ \sum_{t=t_i+1}^{t_i+2I_i} \Big(D_t - (\alpha - \beta p_t) \Big)^2 \right\}, \tag{2.6}$$

$$\eta_t = D_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t), \qquad \text{for } t = t_i + 1, \ldots, t_i + 2I_i. \tag{2.7}$$

Step 3. Defining and maximizing the proxy profit function, denoted by $G_{i+1}^{DD}(p, y)$. Define

$$G_{i+1}^{DD}(p,y) = p e^{\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p} \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\eta_t} - \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \left(h\left(y - e^{\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p} e^{\eta_t} \right)^+ + b\left(e^{\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p} e^{\eta_t} - y \right)^+ \right) \right\}.$$

Then the data-driven optimization is defined by

Problem DD:

$$\begin{aligned} & \max_{(p,y)\in\mathcal{P}\times\mathcal{Y}} G_{i+1}^{DD}(p,y) \qquad (2.8)\\ = \quad & \max_{p\in\mathcal{P}}\Bigg\{ p e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\eta_t} \\ & \qquad -\min_{y\in\mathcal{Y}}\Bigg\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\left(h\left(y-e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}e^{\eta_t}\right)^+ + b\left(e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}e^{\eta_t}-y\right)^+\right)\Bigg\}\Bigg\}. \end{aligned}$$

Solve problem DD and set the first pair of price and inventory level to

$$(\hat{p}_{i+1},\hat{y}_{i+1,1}) = \arg\max_{(p,y)\in\mathcal{P}\times\mathcal{Y}} G_{i+1}^{DD}(p,y),$$

and set the second price to $\hat{p}_{i+1}+\delta_{i+1}$ and the second order-up-to level to

$$\hat{y}_{i+1,2} = \arg\max_{y\in\mathcal{Y}} G_{i+1}^{DD}(\hat{p}_{i+1}+\delta_{i+1},y).$$

In case $\hat{p}_{i+1}+\delta_{i+1}\notin\mathcal{P}$, set the second price to $\hat{p}_{i+1}-\delta_{i+1}$.

Remark 2. When $\hat{\beta}_{i+1}>0$, the objective function in (2.8) after minimizing over $y\in\mathcal{Y}$ is unimodal in p. To see why this is true, let $d=e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}$ and thus $p=\frac{\hat{\alpha}_{i+1}-\log d}{\hat{\beta}_{i+1}}$ with $d\in\mathcal{D}=[d^l,d^h]$, where $d^l=e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p^h}$ and $d^h=e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p^l}$. Then the optimization problem (2.8) is equivalent to

$$\max_{d\in\mathcal{D}}\left\{d\frac{\hat{\alpha}_{i+1}-\log d}{\hat{\beta}_{i+1}}\left(\frac{1}{2I_i}\sum_{t=t_1+1}^{t_i+2I_i}e^{\eta_t}\right)-\min_{y\in\mathcal{Y}}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\left(h(y-de^{\eta_t})^+ + b(de^{\eta_t}-y)^+\right)\right\}\right\}.$$

The objective function of this optimization problem is jointly concave in (y,d) hence it is concave in d after minimizing over $y\in\mathcal{Y}$. Thus, it follows from $p=\frac{\hat{\alpha}_{i+1}-\log d}{\hat{\beta}_{i+1}}$ is strictly decreasing in d that the objective function in (2.8) (after minimization over y) is unimodal in $p\in\mathcal{P}$.

Remark 3. In Step 3 of DDA, the second price is set to $\hat{p}_{i+1}-\delta_{i+1}$ when $\hat{p}_{i+1}+\delta_{i+1} > p^h$. In this case our condition $\rho \leq 2^{-3/4}(p^h - p^l)I_0^{1/4}$ ensures that $\hat{p}_{i+1} - \delta_{i+1} \geq p^l$, thus $\hat{p}_{i+1} - \delta_{i+1} \in \mathcal{P}$. This is because, when $\hat{p}_{i+1} > p^h - \delta_{i+1}$, we have

$$\hat{p}_{i+1} - \delta_{i+1} > p^h - 2\delta_{i+1} \geq p^h - 2\delta_1 = p^h - 2\rho(2I_0)^{-1/4} \geq p^l,$$

where the last inequality follows from the condition on ρ.

Discussion of algorithm and its connections with the literature. In our algorithm above, iteration i focuses on stage i that consists of $2I_i$ periods. In Step 1, the algorithm sets the ordering quantity and selling price for each period in stage i, and they are derived from the previous iteration. In Step 2, the algorithm uses the realized demand data and least-squares method to update the linear approximation, $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p$, of $\lambda(p)$ and computes a biased sample η_t of random error ϵ_t, for $t = t_i+1, \ldots, t_i+2I_i$. Note that η_t is not a sample of the random error ϵ_t. This is because $\epsilon_t = D_t(p_t) - \lambda(p_t)$ and the (logarithm of) observed demand is $D_t(p_t)$. However as we do not know $\lambda(p)$, it is approximated by $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p$, therefore

$$\eta_t = D_t(p_t) - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p_t) \neq D_t(p_t) - \lambda(p_t) = \epsilon_t.$$

For the same reason, the constructed objective function for holding and shortage costs is not a sample average of the newsvendor problem.

In the traditional SAA, mathematical expectations are replaced by sample means, see e.g., *Kleywegt et al.* (2002). *Levi et al.* (2007) and *Levi et al.* (2011) apply SAA method in dynamic newsvendor problems. The argument above shows that the traditional analyses that show SAA leads to the optimal solution is not applicable to our setting. Indeed, in our inner layer optimization, we face a newsvendor problem for which the firm needs to balance holding and shortage cost, and the knowledge about demand distribution is critical. However, the lack of samples of random error

ϵ_t makes the inner loop optimization problem significantly different from *Levi et al.* (2007) and *Levi et al.* (2011), which consider pure inventory control problems and samples of random errors are available for applications of SAA result and analysis. Because of this, it is not guaranteed that the SAA method will lead to a true optimal solution.

The DDA algorithm integrates a process of earning (exploitation) and learning (exploration) in each stage. The earning phase consists of the first I_i periods starting at $t_i + 1$, during which the algorithm implements the optimal strategy for the proxy problem $G_i^{DD}(p, y)$. In the next I_i periods of learning phase that starts from $t_i + I_i + 1$, the algorithm uses a different price $\hat{p}_i + \delta_i$ and its corresponding order-up-to level. The purpose of this phase is to allow the firm to obtain demand data to estimate the rate of change of the demand with respect to the selling price. Note that, even though the firm deviates from the optimal strategy of the proxy problem in the second phase, the policies, $(\hat{p}_i + \delta_i, \hat{y}_{i,2})$ and $(\hat{p}_i, \hat{y}_{i,1})$, will be very close to each other as δ_i diminishes to zero. We will show that they both converge to the true optimal solution and the loss of profit from this deviation converges to zero.

The pricing part of our algorithm is similar to the pure pricing problem considered by *Besbes and Zeevi* (2015) as we also use linear approximation to estimate the demand-price function then maximize the resulting proxy profit function. Although our algorithm is heavily influenced by their work, there is a key difference. *Besbes and Zeevi* (2015) consider a revenue management problem and they only need to estimate the deterministic demand-price function, and the distribution of random errors is immaterial in their analysis. In our model, however, due to the holding and backlogging costs, the distribution of the random error is critical and that has to be learned during the decision process, but it cannot be separated from the estimation of demand-price curve, as discussed above.

Therefore, due to the lack of unbiased samples of random error and that the learn-

ing of demand-price curve and the random error distribution cannot be decoupld, we are not able to prove that the DDA algorithm converges to the true optimal solution by using the approaches developed in *Besbes and Zeevi* (2015) for the pricing problem and in *Levi et al.* (2007) for the newsvendor problem. To overcome this difficulty, we construct several intermediate bridging problems between the data-driven problem and the complete information problem, and perform a series of convergence analyses to establish the main results.

Performance Metrics. To measure the performance of a policy, we use two metrics proposed in *Besbes and Zeevi* (2015): *consistency* and *regret*. An admissible policy $\pi = ((p_t, y_t), t \geq 1)$ is said to be consistent if $(p_t, y_t) \to (p^*, y^*)$ in probability as $t \to \infty$. The average (per-period) regret of a policy π, denoted by $R(\pi, T)$, is defined as the average profit loss per period, given by

$$R(\pi, T) = G(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t, y_t)\right]. \tag{2.9}$$

Obviously, the faster the regret converges to 0 as $T \to \infty$, the better the policy.

In the next section, we will show that the DDA policy is consistent, and we will also characterize the rate at which the regret converges to zero.

2.3 Main Results

In this section, we analyze the performance of the DDA policy proposed in the previous section. We will show that under a fairly general assumption on the underlying demand process, which covers a number of well-known demand models including logit and exponential demand functions, the regret of DDA policy converges to 0 at rate $O(T^{-1/2})$. We also present a numerical study to illustrate its effectiveness.

Recall that the demand in period t is $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\tilde{\epsilon}_t$. As $\tilde{\lambda}(p)$ is strictly decreasing, it has strictly decreasing inverse function. Let $\tilde{\lambda}^{-1}(d)$ be the inverse function

of $\tilde{\lambda}(p)$, which is defined on $d \in [d^l, d^h] = \left[\tilde{\lambda}(p^h), \tilde{\lambda}(p^l)\right]$. We make the following assumption.

Assumption 1. The function $\tilde{\lambda}(p)$ satisfies the following conditions:

(i) The revenue function $d\tilde{\lambda}^{-1}(d)$ is concave in $d \in \left[d^l, d^h\right]$.

(ii) $0 < \dfrac{\tilde{\lambda}''(p)\tilde{\lambda}(p)}{(\tilde{\lambda}'(p))^2} < 2$ for $p \in \left[p^l, p^h\right]$.

The first condition is a standard assumption in the literature on joint optimization of pricing and inventory control (see e.g., *Federgruen and Heching* (1999), and *Chen and Simchi-Levi* (2004b)), and it guarantees that the objective function in problem CI after minimizing over y is unimodal in p. The second assumption imposes some shape restriction on the underlying demand function, and similar assumption has been made in *Besbes and Zeevi* (2015). Technically, this condition assures that the prices converge to a fixed point through a contraction mapping. Some examples that satisfy both conditions of Assumption 1 are given below.

Example 1. The following functions satisfy Assumption 1.

i) Exponential models: $\tilde{\lambda}(p) = e^{k-mp}, m > 0$.

ii) Logit models: $\tilde{\lambda}(p) = a\frac{e^{k-mp}}{1+e^{k-mp}}$ for $a > 0, m > 0$, and $k - mp < 0$ for $p \in \mathcal{P}$.

iii) Iso-elastic (constant elasticity) models: $\tilde{\lambda}(p) = kp^{-m}$ for $k > 0$ and $m > 1$.

We now present the main results of this chapter. Recall that p^* and y^* are the optimal pricing and inventory decisions for the case with complete information.

Theorem II.1. (Policy Convergence) *Under Assumption 1, the DDA policy is consistent, i.e.,* $(p_t, y_t) \to (p^*, y^*)$ *in probability as* $t \to \infty$.

Theorem II.1 states that both pricing and ordering decisions from the DDA algorithm converge to the *true* optimal solution (p^*, y^*) in probability. Note that the convergence of inventory decision $y_t \to y^*$ is stronger than the convergence of order

up-to levels $\hat{y}_{i,1} \to y^*$ and $\hat{y}_{i,2} \to y^*$. This is because, the order up-to levels may or may not be achievable for each period, thus the resulting inventory levels may "overshoot" the targeting order up-to levels. Theorem II.1 shows that, despite these overshoots, the realized inventory levels converge to the true optimal solution in probability.

Convergence of inventory and pricing decisions alone does not guarantee the performance of DDA policy is close to optimal. Our next result shows that DDA is asymptotically optimal in terms of maximizing the expected profit.

Theorem II.2. (Regret Convergence Rate) *Under Assumption 1, the DDA policy is asymptotically optimal. More specifically, there exists some constant $K > 0$ such that*

$$R(DDA, T) = G(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t, y_t)\right] \leq KT^{-\frac{1}{2}}. \tag{2.10}$$

Theorem II.2 shows that as the length of planning horizon, T, grows, the regret of DDA policy vanishes at the rate of $O\big(T^{-1/2}\big)$, hence DDA policy is asymptotically optimal as T goes to infinity. Thus, even though the firm does not have prior knowledge about the demand process, the performance of the data-driven algorithm approaches the theoretical maximum as the planning horizon becomes long. In *Keskin and Zeevi* (2014), the authors consider a *parametric* data-driven pricing problem (with no inventory decision) where the demand error term is additive and the average demand function is linear, and they prove that no learning algorithm can achieve a convergence rate better than $O(T^{-1/2})$. Our problem involves both pricing and inventory decisions, and the firm does not have prior knowledge about the parametric form of the underlying demand-price function or the distribution of the random error, and our algorithm achieves $O\big(T^{-1/2}\big)$, which is the theoretical lower bound. One interesting implication of this finding is that, linear model in demand learning achieves the best regret rate one can hope for, thus our result offers further evidence for the sufficiency of Besbes and Zeevi's linear model.

A numerical Study. We perform a numerical study on the performance of the DDA algorithm, and present our numerical results on the regret. We consider two demand curve environments for $\tilde{\lambda}(p)$:

1) exponential e^{k-mp}: $k \in [\underline{k}, \overline{k}], m \in [\underline{m}, \overline{m}]$, where $[\underline{k}, \overline{k}] = [0.1, 1.7], [\underline{m}, \overline{m}] = [0.3, 2]$,

2) logit $\frac{e^{k-mp}}{1+e^{k-mp}}$: $k \in [\underline{k}, \overline{k}], m \in [\underline{m}, \overline{m}]$, where $[\underline{k}, \overline{k}] = [-0.3, 1], [\underline{m}, \overline{m}] = [2, 2.5]$.

And we consider five error distributions for $\tilde{\epsilon}_t$:

i) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.1,

ii) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.25,

iii) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.35,

iv) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.5,

v) uniform on $[0.5, 1.5]$.

Here truncated normal on $[a, b]$ with mean μ and variance σ^2 is defined as random variable X conditioning on $X \in [a, b]$, where X is normally distributed with mean μ and variance σ^2.

Following *Besbes and Zeevi* (2015), for each combination of the above demand curve-error distribution specifications, we randomly draw 500 instances from the parameters k and m according to a uniform distribution on $[\underline{k}, \overline{k}]$ and $[\underline{m}, \overline{m}]$. For each draw, we compute the percentage of profit loss per period defined by

$$\frac{R(\pi, T)}{G(p^*, y^*)} \times 100\%.$$

Then we compute the average profit loss per period over the 500 draws and report them in Table 1. In all the experiments, we set $p^l = 0.51, p^h = 4, y^l = 0, y^h =$

$3, b = 1, h = 0.1, I_0 = 1$, and initial price $\hat{p}_1 = 1$, initial inventory order up-to level $\hat{y}_{11} = 1,\ \hat{y}_{12} = 0.3$. We test two values of ρ, $\rho = 0.5$ and $\rho = 0.75$, and two values of v, namely, $v = 1.3$ and $v = 2$.

Table 2.1: Exponential Demand

Time Periods		$T = 100$		$T = 500$		$T = 1000$		$T = 5000$		$T = 10000$	
	ρ	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$
Normal	0.5	6.83	6.21	3.39	2.46	2.54	1.71	1.25	0.86	0.87	0.62
$\sigma = 0.1$	0.75	6.84	6.31	3.65	2.59	2.89	1.84	1.39	1.06	0.95	0.76
Normal	0.5	15.36	12.75	8.73	6.55	6.74	4.76	3.48	2.31	2.67	1.69
$\sigma = 0.25$	0.75	11.70	9.74	6.48	4.58	5.12	3.39	2.60	1.78	1.82	1.27
Normal	0.5	18.20	15.12	11.04	8.09	8.65	5.77	4.55	3.03	3.39	2.24
$\sigma = 0.35$	0.75	13.62	10.83	7.64	5.18	5.91	3.76	3.08	2.03	2.26	1.51
Normal	0.5	20.03	16.55	12.07	9.47	9.40	6.87	5.11	3.54	3.88	2.64
$\sigma = 0.5$	0.75	14.84	12.15	8.41	6.12	6.59	4.44	3.51	2.41	2.54	1.76
Uniform	0.5	18.53	15.02	9.98	7.18	7.59	5.39	3.69	2.62	2.58	1.86
	0.75	14.08	11.11	8.12	5.57	6.49	4.22	3.41	2.54	2.40	1.85
Maximum		20.03	16.55	12.07	9.47	9.40	6.87	5.11	3.54	3.88	2.64
Average		14.00	11.58	7.95	5.78	6.19	4.22	3.21	2.22	2.34	1.62

Table 2.2: Logit Demand

Time Periods		$T = 100$		$T = 500$		$T = 1000$		$T = 5000$		$T = 10000$	
	ρ	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$	$v = 1.3$	$v = 2$
Normal	0.5	6.80	5.62	4.35	2.30	2.63	1.63	1.26	0.89	0.85	0.63
$\sigma = 0.1$	0.75	10.09	8.34	3.42	3.67	4.42	2.67	2.15	1.60	1.45	1.15
Normal	0.5	13.72	9.57	6.83	4.44	4.98	3.17	2.34	1.56	1.66	1.10
$\sigma = 0.25$	0.75	12.58	9.86	6.89	4.51	5.42	3.30	2.67	1.87	1.81	1.35
Normal	0.5	17.13	12.52	8.65	6.01	6.52	4.10	3.04	1.98	2.12	1.41
$\sigma = 0.35$	0.75	13.84	10.49	7.49	4.85	5.82	3.55	2.85	2.00	1.96	1.43
Normal	0.5	19.38	13.75	9.99	6.52	7.31	4.57	3.35	2.18	2.34	1.57
$\sigma = 0.5$	0.75	14.49	11.30	7.84	5.24	6.07	3.79	3.00	2.11	2.05	1.51
Uniform	0.5	21.20	15.29	9.51	6.20	7.16	4.46	3.36	2.39	2.29	1.72
	0.75	17.46	14.63	10.44	6.97	8.74	5.35	4.81	3.63	3.38	2.73
Maximum		21.20	15.29	10.44	6.97	8.74	5.35	4.81	3.63	3.38	2.73
Average		14.67	11.14	7.54	5.07	5.91	3.66	2.88	2.02	1.99	1.46

Table 2.1 summarizes the results when the underlying demand curve is exponential, and Table 2.2 displays the results when the underlying demand curve is logit. Combining both tables, one sees that when $T = 100$ periods, on average the profit loss from the DDA algorithm falls between 11% and 14% compared to the optimal profit under complete information, in which DDA starts with no prior knowledge about the underlying demand. When $T = 500$, the profit loss is further reduced to between 5% and 8%. The performance gets better and better when T becomes larger. Also, it

is seen from the table that the overall performance of algorithm is better when the variance of the demand is smaller, which is intuitive.

As mentioned earlier, Theorems II.1 and II.2 continue to hold for the additive demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t) + \tilde{\epsilon}_t$ with minor modifications. Specifically, we need to modify Assumption 1 to Assumption 1A below.

Assumption 1A. The demand-price function $\tilde{\lambda}(p)$ satisfy the following conditions:

(i′) $p\tilde{\lambda}(p)$ is unimodal in p on $p \in \mathcal{P}$.

(ii′) $-1 < \frac{\tilde{\lambda}''(p)\tilde{\lambda}(p)}{2(\tilde{\lambda}'(p))^2} < 1$, for all $p \in \mathcal{P}$.

Note that these are exactly the same assumptions made in *Besbes and Zeevi* (2015) for the revenue management problem, and examples that satisfy Assumption 1A include (a) linear with $\lambda(p) = k - mp$, $m > 0$, (b) exponential with $\lambda(p) = e^{k-mp}, m > 0$, and (c) logit with $\lambda(p) = \frac{e^{k-mp}}{1+e^{k-mp}}, m > 0, e^{k-mp} < 3$ for all $p \in \mathcal{P}$.

The learning algorithm for the additive demand model is similar to that of the multiplicative demand case, except that there is no need to transform it using the logarithm of the deterministic portion of demand and the logarithm of random demand error. Instead, the algorithm directly estimates $\tilde{\lambda}(p)$ using affine function and computes the biased samples of the random demand error in each iteration.

2.4 Sketches of the Proof

In this section, we present the main ideas and steps in proving the main results of this chapter. In the first subsection, we elaborate on the technical issues encountered in the proofs. The key ideas of the proofs are discussed in Subsection 4.2, and the major steps for the proofs of Theorems 1 and 2 are given in Subsections 4.3 and 4.4, respectively.

2.4.1 Technical Issues Encountered

To prove Theorem 1, we will need to show

$$\mathbb{E}\big[(\hat{p}_{i+1}-p^*)^2\big]\to 0,\quad \mathbb{E}\big[(\hat{p}_{i+1}+\delta_{i+1}-p^*)^2\big]\to 0,\quad \text{as } i\to\infty; \tag{2.11}$$

$$\mathbb{E}[(y^*-\hat{y}_{i+1,1})^2]\to 0,\quad \mathbb{E}[(y^*-\hat{y}_{i+1,2})^2]\to 0,\quad \text{as } i\to\infty, \tag{2.12}$$

where p^* is the optimal solution of

$$\max_{p\in\mathcal{P}} Q(p,\lambda(p)) = \max_{p\in\mathcal{P}}\Big\{pe^{\lambda(p)}\mathbb{E}\big[e^{\epsilon}\big]-J(\lambda(p))\Big\},$$

where $J(\lambda(p))$ is defined as

$$J(\lambda(p)) = \min_{y\in\mathcal{Y}}\left\{h\mathbb{E}\big[y-e^{\lambda(p)}e^{\epsilon}\big]^+ + b\mathbb{E}\big[e^{\lambda(p)}e^{\epsilon}-y\big]^+\right\}.$$

However, both $Q(\cdot,\cdot)$ and $J(\cdot)$ are *unknown* to the firm because all the expectations cannot be computed. To estimate $J(\cdot)$, in (2.8) of the learning algorithm we use the data-driven biased estimation of

$$J^{DD}_{i+1}(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p) = \min_{y\in\mathcal{Y}}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\left(h\left(y-e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}e^{\eta_t}\right)^+ + b\left(e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}e^{\eta_t}-y\right)^+\right)\right\},$$

and $\hat{p}_{i+1}$ is the optimal solution of

$$\max_{p\in\mathcal{P}} Q^{DD}_{i+1}(p,\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p) = \max_{p\in\mathcal{P}}\left\{pe^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\eta_t} - J^{DD}_{i+1}(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p)\right\},$$

in which $Q^{DD}_{i+1}(\cdot,\cdot)$ is *random* and is constructed based on *biased* random samples η_t.

To prove the convergence of the data-driven solutions to the true optimal solution, we face two major challenges. The first one is the comparison between $J^{DD}_{i+1}(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p)$ and $J(\lambda(p))$ as functions of p. In J^{DD}_{i+1}, the true demand-price function is

replaced by a linear estimation and, due to lack of knowledge about distribution of random error, the expectation is replaced by an arithmetic average from biased samples η_t not true samples of random error ϵ_t. To put it differently, the objective function for J_{i+1}^{DD} is not a sample average approximation, but a biased-sample average approximation. The second challenge lies in the comparison of $Q_{i+1}^{DD}(p, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p)$ and $Q(p, \lambda(p))$. Since Q_{i+1}^{DD} is a function of J_{i+1}^{DD} that is minimum of a biased-sample average approximation, the errors in replacing ϵ_t by η_t carry over to Q_{i+1}^{DD}, making it difficult to compare $(\hat{p}_{i+1}, \hat{y}_{i+1,1})$ and $(\hat{p}_{i+1}+\delta_{i+1}, \hat{y}_{i+1,2})$ with (p^*, y^*). To overcome the first difficulty, we establish several important properties of the newsvendor problem and bound the errors of biased samples (Lemmas A2, A3, A4, A8 in the Appendix). For the second, we identify high probability events in which uniform convergence of the data-driven objective functions can be obtained (Lemmas A1, A5, A6, and A7 in the Appendix).

We note that in the revenue management problem setting, *Besbes and Zeevi* (2015) also prove the convergence result (2.11). In *Besbes and Zeevi* (2015), p^* is the optimal solution of $\max_{p\in\mathcal{P}} Q(p, \lambda(p))$, and $\hat{p}_{i+1}$ is the optimal solution of $\max_{p\in\mathcal{P}} Q(p, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p)$, where $Q(\cdot,\cdot)$ is a *known* and *deterministic* function $Q(p, \lambda(p)) = p\lambda(p)$. As *Besbes and Zeevi* (2015) point out, their analysis extends to more general function $Q(p, \lambda(p))$ in which $Q(\cdot,\cdot)$ is a known deterministic function. This, however, is not true in our setting as $Q(\cdot,\cdot)$ is not known, and as a matter of fact, one cannot even find an unbiased sample average to estimate $Q(\cdot,\cdot)$. Therefore, the challenges discussed above were not present in *Besbes and Zeevi* (2015).

2.4.2 Main Ideas of the Proof

To compare the policy and the resultant profit of DDA algorithm with that of the optimal solution, we first note that these two problems differ along several dimensions. For example, in DDA we approximate $\lambda(p)$ by an affine function and estimate the

parameters of the affine function in each iteration, and we approximate the expected revenue and the expected holding and shortage costs using biased sample averages. These differences make the direct comparison of the two problems difficult. Therefore, we introduce several "intermediate" bridging problems, and in each step we compare two "adjacent" problems that differ just in one dimension.

For convenience, we follow *Besbes and Zeevi* (2015) to introduce notation

$$\breve{\alpha}(z) = \lambda(z) - \lambda'(z)z, \qquad \breve{\beta}(z) = -\lambda'(z), \qquad z \in \mathcal{P}. \tag{2.13}$$

We proceed to prove (2.11) as follows:

$$\begin{aligned} \mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \;\leq\; \mathbb{E}\Bigg[\Big(& \underbrace{\Big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|}_{\substack{\text{Comparison of problems CI and B1}\\ \text{Lemma A1}}} \\ & + \underbrace{\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\Big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\Big)\Big|}_{\substack{\text{Comparison of problems B1 and B2}\\ \text{Lemma A5}}} \\ & + + \underbrace{\Big|\tilde{p}_{i+1}\Big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\Big) - \hat{p}_{i+1}\Big|}_{\substack{\text{Comparison of problems B2 and DD}\\ \text{Lemma A6 and Lemma A7}}}\Big)^2\Bigg], \end{aligned} \tag{2.14}$$

where the two new prices $\overline{p}\big(\cdot,\cdot\big)$ and $\tilde{p}_{i+1}\left(\cdot,\cdot\right)$ are the optimal solutions of two bridging problems. Specifically, we let $\overline{p}\big(\alpha,\beta\big)$ denote the optimal solution for the first bridging problem B1 defined by

Bridging Problem B1:

$$\max_{p\in\mathcal{P}}\left\{pe^{\alpha-\beta p}\mathbb{E}\big[e^{\epsilon}\big] - \min_{y\in\mathcal{Y}}\left\{h\mathbb{E}\Big[y - e^{\alpha-\beta p}e^{\epsilon}\Big]^+ + b\mathbb{E}\Big[e^{\alpha-\beta p}e^{\epsilon} - y\Big]^+\right\}\right\} \tag{2.15}$$

while $\tilde{p}_{i+1}\left(\alpha,\beta\right)$ denotes the optimal solution for the second bridging problem B2

defined by

Bridging Problem B2:

$$\max_{p\in\mathcal{P}}\left\{pe^{\alpha-\beta p}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right)\right. \tag{2.16}$$
$$\left.-\min_{y\in\mathcal{Y}}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(y-e^{\alpha-\beta p}e^{\epsilon_t}\big)^+ + b\big(e^{\alpha-\beta p}e^{\epsilon_t}-y\big)^+\Big)\right\}\right\}.$$

Moreover, for given $p\in\mathcal{P}$, we let $\overline{y}(e^{\alpha-\beta p})$ denote the optimal order-up-to level for problem B1, and $\tilde{y}_{i+1}(e^{\alpha-\beta p})$ denote the optimal order-up-to level for problem B2. By Lemma A2 in the Appendix, the objective functions for problems B1 and B2 are unimodal in p after minimizing over $y\in\mathcal{Y}$ when $\beta>0$.

Comparing (2.15) with (2.4), it is seen that problem B1 simplifies problem CI by replacing the demand-price function $\lambda(p)$ by a linear function $\alpha-\beta p$, while problem B2 is obtained from problem B1 after replacing the mathematical expectations in problem B1 by their sample averages, i.e., problem B2 is the SAA of problem B1. Comparing (2.16) with (2.8), it is noted that problems B2 and DD differ in the coefficients of the linear function as well as the arithmetic averages. More specifically, in B2 the real random error samples ϵ_t, $t=t_i+1,\ldots,t_i+2I_i$, are used, while in problem DD, biased error samples η_t are used in place of ϵ_t, $t=t_i+1,\ldots,t_i+2I_i$. Furthermore, note that the optimal prices for problems CI and B1, p^* and $\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)$, are deterministic, but the optimal solutions of problems B2 and DD, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)$ and $\hat{p}_{i+1}$, are random. Specifically, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)$ is random because ϵ_t is random, while $\hat{p}_{i+1}$ is random due to demand uncertainty from periods 1 to t_{i+1}. Hence, to show the right hand side of (2.14) converges to 0, we will first develop an upper bound for $\big|p^*-\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\big|$ by comparing problems CI and B1, and the result is presented in Lemma A1. Since $\tilde{p}_{i+1}(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)$ is random, we compare the two problems B1 and B2 and show the probability that $\big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\big|$ exceeds some

small number diminishes to 0 in Lemma A5. Similarly, in Lemma A6 and Lemma A7 we compare problems B2 and DD and show the probability that $\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\breve{\beta}}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|$ exceeds some small number also diminishes to 0. Finally, we combine these several results to complete the proof of (2.11). The idea for proving (2.12) is similar, and that also relies heavily on the two bridging problems (Lemmas A6, A7, and A8). The detailed proofs for Theorem 1 and Theorem 2 are given in Subsections 4.3 and 4.4.

In the subsequent analysis, we assume that the space for feasible price, $\mathcal{P}$, and the space for order-up-to level, $\mathcal{Y}$, are large enough so that the optimal solutions p^* and optimal $\overline{y}(e^{\lambda(p)})$ over $\mathbb{R}_+$ for given $p \in \mathcal{P}$ for problem CI fall into $\mathcal{P}$ and $\mathcal{Y}$, respectively; and for given $q \in \mathcal{P}$, the optimal solutions $\overline{p}\big(\breve{\alpha}(q), \breve{\breve{\beta}}(q)\big)$ and $\overline{y}\big(e^{\breve{\alpha}(q)-\breve{\breve{\beta}}(q)p}\big)$ for given $p \in \mathcal{P}$ over $\mathbb{R}_+$ for problem B1 fall into $\mathcal{P}$ and $\mathcal{Y}$, respectively. Note that both problem CI and problem B1 depend only on primitive data and do not depend on random samples, hence these are mild assumptions. We remark that our results and analyses continue to hold even if these assumptions are not satisfied as long as we modify Assumption 1(ii) to $\left|\partial\overline{p}\big(\breve{\alpha}(z), \breve{\breve{\beta}}(z)\big)/\partial z\right| < 1$ for $z \in \mathcal{P}$. This condition reduces to Assumption 1(ii) if the optimal solutions for problem CI and problem B1 satisfy the feasibility conditions described above.

We end this subsection by listing some regularity conditions needed to prove the main theoretical results.

Regularity Conditions:

(i) $\overline{y}(e^{\lambda(q)})$ and $\overline{y}\big(e^{\breve{\alpha}(q)-\breve{\breve{\beta}}(q)p}\big)$ are Lipschitz continuous on q for given $p \in \mathcal{P}$, i.e., there exists some constant $K_1 > 0$ such that for any $q_1, q_2 \in \mathcal{P}$,

$$\left|\overline{y}(e^{\lambda(q_1)}) - \overline{y}(e^{\lambda(q_2)})\right| \leq K_1 \left|q_1 - q_2\right|, \tag{2.17}$$

$$\left|\overline{y}\big(e^{\breve{\alpha}(q_1)-\breve{\breve{\beta}}(q_1)p}\big) - \overline{y}\big(e^{\breve{\alpha}(q_2)-\breve{\breve{\beta}}(q_2)p}\big)\right| \leq K_1 \left|q_1 - q_2\right|. \tag{2.18}$$

(ii) $G(p, \bar{y}(e^{\lambda(p)}))$ has bounded second order derivative with respect to $p \in \mathcal{P}$.

(iii) $\mathbb{E}[D_t(p)] > 0$ for any price $p \in \mathcal{P}$.

(iv) $\lambda(p)$ is twice differentiable with bounded first and second order derivatives on $p \in \mathcal{P}$.

(v) The probability density function $f(\cdot)$ of $\tilde{\epsilon}_t$ satisfies $\min\{f(x), x \in [l, u]\} > 0$.

It can be seen that all the functions in Example 1 satisfy the regularity conditions above with appropriate choices of p^l and p^h.

2.4.3 Proof of Theorem 1

The proofs for the convergence results are technical and rely on several lemmas that are provided in the Appendix. In this subsection, we outline the main steps in establishing the first main result, Theorem 1.

Convergence of pricing decisions. To prove the convergence of pricing decisions, we continue the development in (2.14) as follows:

$$
\begin{aligned}
& \mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| \\
& \qquad + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big| \Big)^2\Big] \\
\leq\ & \mathbb{E}\Big[\Big(\gamma|p^* - \hat{p}_i| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big| \Big)^2\Big] \\
\leq\ & \left(\frac{1+\gamma^2}{2}\right) \mathbb{E}\left[(p^* - \hat{p}_i)^2\right] \\
& \qquad + K_2 \mathbb{E}\left[\left(\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\right)^2\right] \\
\leq\ & \left(\frac{1+\gamma^2}{2}\right) \mathbb{E}\left[(p^* - \hat{p}_i)^2\right] \\
& + K_3 \mathbb{E}\left[\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|^2\right] \\
& + K_3 \mathbb{E}\left[\Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|^2\right], \qquad (2.19)
\end{aligned}
$$

where the first inequality follows from the expansion in (2.14), the second inequality follows from Lemma A1, and the third inequality is justified by $\gamma < 1$ in Lemma A1 and some constant K_2, and the last inequality holds for some appropriately chosen K_3 because of the inequality $(a+b)^2 \leq 2(a^2+b^2)$ for any real numbers a and b.

To bound $\mathbb{E}\big[\big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\big|^2\big]$ in (2.19), by Lemma A5 one has, for some constant K_4,

$$\mathbb{E}\left[\left|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\right|^2\right] \leq K_4^2 \int_0^{+\infty} 5e^{-4I_i\xi^2} d\xi = \frac{5\pi^{\frac{1}{2}}K_4^2}{4I_i^{\frac{1}{2}}}. \quad (2.20)$$

And to bound $\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|^2\right]$ in (2.19), by Lemma A6 and Lemma A7, when i is large enough (greater than or equal to i^* defined in the proof of Lemma A7), for some positive constants K_5, K_6, and K_7 one has

$$\begin{aligned}
&\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|^2\right] \\
\leq\ & \mathbb{E}\Big[K_5^2\big(\big|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}\big| + \big|\breve{\alpha}(\hat{p}_i+\delta_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i+\delta_i) - \hat{\beta}_{i+1}\big|\big)^2\Big] \\
&\qquad + \frac{8}{I_i}\big(p^h - p^l\big)^2 \\
\leq\ & \mathbb{E}\left[K_6\left(|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + |\breve{\alpha}(\hat{p}_i+\delta_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i+\delta_i) - \hat{\beta}_{i+1}|^2\right)\right] \\
&\qquad + \frac{8}{I_i}\big(p^h - p^l\big)^2 \\
\leq\ & K_7 I_i^{-\frac{1}{2}}. \qquad (2.21)
\end{aligned}$$

Substituting (2.20) and (2.21) into (2.19), one has

$$\mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \leq \left(\frac{1+\gamma^2}{2}\right)\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + K_8 I_i^{-\frac{1}{2}}.$$

Letting $\frac{1+\gamma^2}{2} = \theta$, we further obtain

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \;\leq\; \theta^i(\hat{p}_1 - p^*)^2 + K_8 \sum_{j=0}^{i-1} \theta^j I_{i-j}^{-\frac{1}{2}} \leq K_9(v^{-\frac{1}{2}})^i \sum_{j=0}^{i-1} \theta^j (v^{\frac{1}{2}})^j. \quad (2.22)$$

We choose $v > 1$ that satisfies $\theta v^{\frac{1}{2}} < 1$, then there exists a positive constant K_{10} such that $\sum_{j=0}^{i-1} \theta^j (v^{\frac{1}{2}})^j \leq K_{10}$, therefore, for some constants K_{11} and K_{12},

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \leq K_{11}(v^{-\frac{1}{2}})^i \leq K_{12} I_i^{-\frac{1}{2}}. \quad (2.23)$$

Moreover, we have, for some positive constant K_{13},

$$\mathbb{E}\left[(\hat{p}_{i+1} + \delta_{i+1} - p^*)^2\right] \leq 2\mathbb{E}\left[(\hat{p}_{i+1} - p^*)^2\right] + 2\delta_{i+1}^2 \leq K_{13} I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty. \quad (2.24)$$

This completes the proof of (2.11). Because mean-square convergence implies convergence in probability, this shows that the pricing decisions from DDA converge to p^* in probability.

Convergence of inventory decisions. To prove y_t converges to y^* in probability, we first prove the convergence of order up-to levels (2.12). For some constant

K_{14}, we have

$$
\begin{aligned}
& \mathbb{E}\left[\left|y^* - \hat{y}_{i+1,1}\right|^2\right] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|\overline{y}\big(e^{\lambda(p^*)}\big) - \overline{y}(e^{\lambda(\hat{p}_{i+1})})\Big| + \Big|\overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big)\Big| \\
& \quad + \Big|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1}\big)-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\Big| \\
& \quad + \Big|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big) - \tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\Big| + \Big|\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big) - \hat{y}_{i+1,1}\Big|\Big)^2\Big] \\
\leq\ & K_{14}\mathbb{E}\Big[\Big(\underbrace{\Big|\overline{y}\big(e^{\lambda(p^*)}\big) - \overline{y}(e^{\lambda(\hat{p}_{i+1})})\Big|^2}_{\text{Difference between } p^* \text{ and } \hat{p}_{i+1}} + \underbrace{\Big|\overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big)\Big|^2}_{\text{Zero}} \qquad (2.25) \\
& \quad + \underbrace{\Big|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\Big|^2}_{\text{Difference between } \hat{p}_{i+1} \text{ and } \hat{p}_i} \\
& \quad + \underbrace{\Big|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big) - \tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\Big|^2}_{\substack{\text{Comparison of problems B1 and B2} \\ \text{Lemma A8}}} + \underbrace{\Big|\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big) - \hat{y}_{i+1,1}\Big|^2}_{\substack{\text{Comparison of problems B2 and DD} \\ \text{Lemma A6 and Lemma A7}}}\Big)\Big].
\end{aligned}
$$

We want to upper bound each term on the right hand side of (2.25). First, it follows from (2.17) that, for some constant K_{15} it holds

$$
\mathbb{E}\left[\left|\overline{y}\big(e^{\lambda(p^*)}\big) - \overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big)\right|^2\right] \leq K_{15}\mathbb{E}\left[\mid p^* - \hat{p}_{i+1} \mid^2\right].
$$

By definition of $\breve{\alpha}(p)$ and $\breve{\beta}(p)$ in (2.13) one has $\breve{\alpha}(\hat{p}_{i+1}) - \breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1} = \lambda(\hat{p}_{i+1})$, thus the second term on the right hand side of (2.25) vanishes. For the third term, we apply the Lipschitz condition on $\overline{y}\big(e^{\breve{\alpha}(q)-\breve{\beta}(q)p}\big)$ in (2.18) to obtain, for some constants K_{16} and K_{17},

$$
\begin{aligned}
\mathbb{E}\left[\left|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\right|^2\right] &\leq K_{16}\mathbb{E}\left[\mid \hat{p}_{i+1} - \hat{p}_i \mid^2\right] \\
&\leq K_{17}\mathbb{E}\left[\big(\mid p^* - \hat{p}_i \mid^2 + \mid p^* - \hat{p}_{i+1} \mid^2\big)\right].
\end{aligned}
$$

By Lemma A8, we have, for some constants K_{18} and K_{19},

$$\mathbb{E}\left[\left|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)-\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\right|^2\right] \;\leq\; K_{18}^2\int_0^{+\infty}2e^{-4I_i\xi}d\xi \leq \frac{K_{19}}{I_i} \tag{2.26}$$

and by Lemma A6 and Lemma A7 one has, for some constant K_{20},

$$\begin{aligned}
&\mathbb{E}\left[\left|\tilde{y}_{i+1}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}})-\hat{y}_{i+1,1}\right|^2\right]\\
\leq\;& K_{20}\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}|^2\right]\\
\leq\;& K_{20}I_i^{-\frac{1}{2}}.
\end{aligned}$$

Summarizing the analyses above we obtain, for some constants K_{21} and K_{22},

$$\begin{aligned}
&\mathbb{E}\left[\big(y^*-\hat{y}_{i+1,1}\big)^2\right]\\
\leq\;& K_{21}\mathbb{E}\big[\,|\,p^*-\hat{p}_{i+1}\,|^2+|\,p^*-\hat{p}_i\,|^2\,\big]+K_{21}I_i^{-\frac{1}{2}}\\
\leq\;& K_{22}I_i^{-\frac{1}{2}} \qquad (2.27)\\
\to\;& 0 \text{ as } i\to\infty,
\end{aligned}$$

where the second inequality follows from the convergence rate of the pricing decisions. Similarly, we obtain

$$\mathbb{E}\left[\big(y^*-\hat{y}_{i+1,2}\big)^2\right]\leq K_{22}I_i^{-\frac{1}{2}}\to 0, \qquad \text{as } i\to\infty.$$

We next show that $\mathbb{E}[(y^*-y_t)^2]\to 0$ as $t\to\infty$. It suffices to prove this for (a) $t\in\{t_{i+1}+1,\ldots,t_{i+1}+I_{i+1}\}$, $i=1,2,\ldots$, and for (b) $t\in\{t_{i+1}+I_{i+1}+1,\ldots,t_{i+1}+2I_{i+1}\}$, $i=1,2,\ldots$. We will only provide the proof for (a).

The inventory order up-to level prescribed from DDA for periods $t\in\{t_{i+1}+1,\ldots,t_{i+1}+I_{i+1}\}$ is $\hat{y}_{i+1,1}$. This, however, may not be achievable for some period

t. Consider the event that the second order up-to level of learning stage i, $\hat{y}_{i,2}$, is achieved during periods $\{t_i + I_i + 1, \ldots, t_i + 2I_i\}$. Since $\tilde{\lambda}(p^h)l \leq D_t \leq \tilde{\lambda}(p^l)u$, it follows from Hoeffding inequality[1] that for any $\zeta > 0$,

$$\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq \mathbb{E}\left[\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t\right] - \zeta\right\} \geq 1 - \exp\left(-\frac{2\zeta^2}{I_i(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l)^2}\right). \quad (2.28)$$

Let $\zeta = \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}}(\log I_i)^{\frac{1}{2}}$ in (2.28), then one has

$$\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq I_i\mathbb{E}\left[D_{t_i+I_i+1}\right] - \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}}(\log I_i)^{\frac{1}{2}}\right\} \geq 1 - \frac{1}{I_i^2}. \quad (2.29)$$

By regularity condition (iii), $\mathbb{E}\left[D_{t_i+I_i+1}\right] > 0$, thus when i is large enough, we will have

$$\frac{1}{2}I_i\mathbb{E}\left[D_{t_i+I_i+1}\right] \geq \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}}(\log I_i)^{\frac{1}{2}}.$$

Hence it follows from (2.29) that, when i is large enough, we will have

$$\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq \frac{1}{2}I_i\mathbb{E}\left[D_{t_i+I_i+1}\right]\right\} \geq 1 - \frac{1}{I_i^2}. \quad (2.30)$$

Define event

$$\mathcal{A}_1 = \left\{\omega : \sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq \frac{1}{2}I_i\mathbb{E}\left[D_{t_i+I_i+1}\right]\right\},$$

then (2.30) can be rewritten as

$$\mathbb{P}(\mathcal{A}_1) \geq 1 - \frac{1}{I_i^2}.$$

[1]If the random demand is not bounded, then the same result is obtained under the condition that the moment generating function of random demand is finite around 0.

Note that when i is large enough, $\frac{1}{2}I_i\mathbb{E}\left[D_{t_i+I_i+1}\right] > y^h - y^l$, which means that on the event $\mathcal{A}_1$, the accumulative demand during $\{t_i + I_i + 1, \ldots, t_i + 2I_i\}$ is high enough to consume the initial on-hand inventory of period $t_i + I_i + 1$ and $\hat{y}_{i,2}$ will be achieved. Therefore, for $t \in \{t_{i+1} + 1, \ldots, t_{i+1} + I_{i+1}\}$, y_t will satisfy $y_t \in [\hat{y}_{i,2}, \hat{y}_{i+1,1}]$ if $\hat{y}_{i+1,1} \geq \hat{y}_{i,2}$, and $y_t \in [\hat{y}_{i+1,1}, \hat{y}_{i,2}]$ otherwise. Thus,

$$\begin{aligned}
\mathbb{E}[(y^* - y_t)^2] &= \mathbb{P}(\mathcal{A}_1)\mathbb{E}[(y^* - y_t)^2 \big| \mathcal{A}_1] + \mathbb{P}(\mathcal{A}_1^c)\mathbb{E}[(y^* - y_t)^2 \big| \mathcal{A}_1^c] \\
&\leq \max\left\{\mathbb{E}\left[(y^* - \hat{y}_{i,2})^2\right], \mathbb{E}\left[(y^* - \hat{y}_{i+1,1})^2\right]\right\} + \frac{1}{I_i^2}\left(y^h - y^l\right)^2.
\end{aligned}$$

As shown above, $\mathbb{E}\left[(y^* - \hat{y}_{i,2})^2\right] \to 0$ and $\mathbb{E}\left[(y^* - \hat{y}_{i+1,1})^2\right] \to 0$ as $i \to \infty$. Hence it follows from $1/I_i^2 \to 0$ as $i \to \infty$ that $\mathbb{E}\left[(y^* - y_t)^2\right] \to 0$ for $t \in \{t_{i+1} + 1, \ldots, t_{i+1} + I_{i+1}\}$ as $i \to \infty$.

Similarly one can prove that $\mathbb{E}\left[(y^* - y_t)^2\right] \to 0$ for $t \in \{t_{i+1} + I_{i+1} + 1, \ldots, t_{i+1} + 2I_{i+1}\}$ as $i \to \infty$. This proves $\mathbb{E}[(y^* - y_t)^2] \to 0$ when $t \to \infty$. And again, since convergence in probability is implied by mean-square convergence, we conclude that inventory decisions y_t of DDA also converge to y^* in probability as $t \to \infty$. This completes the proof of Theorem 1.

2.4.4 Proof of Theorem 2

We next prove the second main result, the convergence rate of regret. By definition, the regret for the DDA policy is

$$R(DDA, T) = \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\Big(G(p^*, y^*) - G(p_t, y_t)\Big)\right].$$

We have

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{t=1}^{T}\big(G(p^*,y^*)-G(p_t,y_t)\big)\right] \\
\leq\ & \mathbb{E}\Bigg[\sum_{i=1}^{n}\Bigg(\sum_{t=t_i+1}^{t_i+I_i}\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big) \\
& \qquad +\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})+G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\Bigg)\Bigg] \\
=\ & \mathbb{E}\left[\sum_{i=1}^{n}I_i\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})\big)\right] \\
& +\mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+I_i}\big(G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big)+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\right)\right],
\end{aligned}
\tag{2.31}
$$

where n is the smallest number of stages that cover T, i.e., n is the smallest integer such that $2I_0\sum_{i=1}^{n}v^i \geq T$, and it satisfies $\log_v\left(\frac{v-1}{2I_0v}T+1\right) \leq n < \log_v\left(\frac{v-1}{2I_0v}T+1\right)+1$. The inequality in (2.31) follows from that the right hand side includes $2I_0\sum_{i=1}^{n}v^i$ periods which is greater than or equal to T.

The first expectation on the right hand side of (2.31) is with respect to the sum of the difference between profit values of DDA decisions and the optimal solution, hence its analysis relies on the convergence rate of DDA policies; these are demonstrated in (2.23), (2.24), and (2.27). The second expectation on the right hand side of (2.31) stems from the fact that in the process of implementing DDA, it may happen that the inventory decisions from DDA are not implementable. This issue arises in learning algorithms for nonperishable inventory systems and it presents additional challenges in evaluating the regret. We note that in *Huh and Rusmevichientong* (2009), a queueing approach is employed to resolve this issue for a pure inventory system with no pricing decisions.

To develop an upper bound for $G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})$ in (2.31), we first apply Taylor

expansion on $G(p, \overline{y}(e^{\lambda(p)})$ at point p^*. Using the fact that the first order derivative vanishes at $p = p^*$ and the assumption that the second order derivative is bounded (regularity condition (ii)), we obtain, for some constant $K_{23} > 0$, that

$$G\big(p^*, \overline{y}\big(e^{\lambda(p^*)}\big)\big) - G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) \leq K_{23}(p^* - \hat{p}_i)^2. \tag{2.32}$$

Noticing that $\overline{y}\big(e^{\lambda(\hat{p}_i)}\big)$ maximizes the concave function $G\big(\hat{p}_i, y\big)$ for given $\hat{p}_i$, we apply Taylor expansion with respect to y at point $y = \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)$ to yield that, for some constant K_{24},

$$G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) - G(\hat{p}_i, \hat{y}_{i,1}) \leq K_{24}\big(\overline{y}\big(e^{\lambda(\hat{p}_i)}\big) - \hat{y}_{i,1}\big)^2. \tag{2.33}$$

In addition, we have

$$\begin{aligned}
& \mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i)}) - \hat{y}_{i,1})^2\right] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|\overline{y}(e^{\lambda(\hat{p}_i)}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i})\Big| + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big| \\
& \quad + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big| + \Big|\tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \hat{y}_{i,1}\Big|\Big)^2\Big] \\
\leq\ & K_{25}\mathbb{E}\Big[\Big|\overline{y}(e^{\lambda(\hat{p}_i)}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i})\Big|^2 + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big|^2 \\
& \quad + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big|^2 + \Big|\tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \hat{y}_{i,1}\Big|^2\Big].
\end{aligned}$$

This is similar to the right hand side of (2.25) except that $i+1$ is replaced by i. Thus, using the same analysis as that for (2.25), we obtain

$$\mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i)}) - \hat{y}_{i,1})^2\right] \leq K_{26} I_{i-1}^{-\frac{1}{2}} \tag{2.34}$$

for some constant K_{26}.

Applying the results above, we obtain, for some constants K_{27}, K_{28}, and K_{29}, that

$$\begin{aligned}
& \mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i, \hat{y}_{i,1})\right] \\
= \; & \mathbb{E}\Big[\Big(G\big(p^*, \overline{y}\big(e^{\lambda(p^*)}\big)\big) - G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big)\Big) + \Big(G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) - G(\hat{p}_i, \hat{y}_{i,1})\Big)\Big] \\
\leq \; & K_{27}\Big(\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + \mathbb{E}\Big[\big(\overline{y}(e^{\lambda(\hat{p}_i)}) - \hat{y}_{i,1}\big)^2\Big]\Big) \\
\leq \; & K_{28}\left(K_{10} I_{i-1}^{-\frac{1}{2}} + K_{37} I_{i-1}^{-\frac{1}{2}}\right) \\
= \; & K_{29} I_{i-1}^{-\frac{1}{2}},
\end{aligned}$$

where the first inequality follows from (2.32) and (2.33), and the second inequality follows from the convergence rate of pricing decisions (2.23) and (2.34).

Similarly, we establish for some constants K_{30}, K_{31} and K_{32}, that

$$\begin{aligned}
\mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i + \delta_i, \hat{y}_{i,2})\right] \; & \leq \; K_{30}\left(\mathbb{E}\left[(p^* - \hat{p}_i - \delta_i)^2\right] + \mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i + \delta_i)}) - \hat{y}_{i,2})^2\right]\right) \\
& \leq \; K_{30}\left(\mathbb{E}\left[2(p^* - \hat{p}_i)^2 + 2\delta_i^2\right] + K_{31} I_{i-1}^{-\frac{1}{2}}\right) \\
& \leq \; K_{32} I_{i-1}^{-\frac{1}{2}}.
\end{aligned}$$

Note that, as seen from Lemma A7 in the Appendix, these results hold when i is greater than or equal to some number i^*.

Consequently, we have, for some constants K_{33}, K_{34} and K_{35},

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n}\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})\big)I_i\right]\\
=&\sum_{i=i^*+1}^{n}K_{33}I_{i-1}^{-\frac{1}{2}}I_i+\sum_{i=1}^{i^*}\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})\big)I_i\\
=&\sum_{i=i^*+1}^{n}K_{33}I_{i-1}^{\frac{1}{2}}+K_{34}\\
\le&\ K_{33}\sum_{i=2}^{n}I_{i-1}^{\frac{1}{2}}+K_{34}\\
=&\ K_{33}\frac{(2I_0)^{\frac{1}{2}}v^{\frac{1}{2}}}{v^{\frac{1}{2}}-1}\big(v^{\frac{n-1}{2}}-1\big)+K_{34}\\
\le&\ K_{33}\frac{(2I_0)^{\frac{1}{2}}v^{\frac{1}{2}}}{v^{\frac{1}{2}}-1}(v^{\log_v(\frac{v-1}{2I_0v}T+1)+1-1})^{\frac{1}{2}}+K_{34}\\
\le&\ K_{35}T^{\frac{1}{2}},
\end{aligned}\tag{2.35}$$

where $K_{34}=\sum_{i=1}^{i^*}\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})\big)I_i$.

We next evaluate the second term of (2.31), i.e.,

$$\mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+I_i}\big(G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big)+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\right)\right].\tag{2.36}$$

Recall from DDA that $p_t=\hat{p}_i$ for $t=t_i+1,\ldots,t_i+I_i$ and $p_t=\hat{p}_i+\delta_i$ for $t=t_i+I_i+1,\ldots,t_i+2I_i$, and DDA sets two order-up-to levels for stage i, $\hat{y}_{i1}$ and $\hat{y}_{i2}$, for the first and second I_i periods, respectively. The order-up-to levels may not be achievable, which happens when $x_t>\hat{y}_{i,1}$ for some $t=t_i+1,\ldots,t_i+I_i$, or $x_t>\hat{y}_{i,2}$ for some $t=t_i+I_i+1,\ldots,t_i+2I_i$. In such cases, $y_t=x_t$. If the inventory level before ordering at the beginning of the first I_i periods (in period t_i+1) or at the beginning of the second I_i periods (in period t_i+I_i+1) of stage i is higher than the DDA order-up-to level, then the inventory level will gradually decrease during the I_i periods until it drops to or below the order up-to level.

We start with the analysis of the first I_i periods of state i, i.e.,

$$\mathbb{E}\left[\sum_{t=t_i+1}^{t_i+I_i}\left(G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\right)\right].$$

A main issue with the analysis of this part is that, if $x_{t_i+1}>\hat{y}_i$, then $\hat{y}_i$ is not achievable. To resolve this issue, we apply a similar argument as that in the proof of the second part of Theorem 1 to show that, if this is the case, then with very high probability, after a (relatively) small number of periods, the prescribed inventory order up-to level will become achievable .

Consider the accumulative demands during periods t_i+1 to $t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor$. If these accumulative demands consume at least $x_{t_i+1}-\hat{y}_i$, then at period $t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor$, $\hat{y}_i$ will be surely achieved. Since $\tilde{\lambda}(p^h)l\le D_t\le\tilde{\lambda}(p^l)u$ for $t=1,\ldots,T$, by Hoeffding inequality, for any $\zeta>0$ one has

$$\mathbb{P}\left\{\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor}D_t\ge\mathbb{E}\left[\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor}D_t\right]-\zeta\right\}\ge 1-\exp\left(-\frac{2\zeta^2}{\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\left(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l\right)^2}\right). \quad (2.37)$$

Let $\zeta=\left(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l\right)\left(\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}$, then it follows from (2.37) that

$$\begin{aligned}&\mathbb{P}\left\{\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor}D_t\ge\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\mathbb{E}\left[D_{t_i+1}\right]-\left(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l\right)\left(\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\right\}\\ \ge\quad & 1-\frac{1}{\left\lfloor I_i^{\frac{1}{2}}\right\rfloor^2}. \quad (2.38)\end{aligned}$$

By regularity condition (iii), $\mathbb{E}\left[D_{t_i+1}\right]>0$. Thus, when i is large enough, say

greater than or equal to some number i^{**}, we will have

$$\begin{aligned} & \left\lfloor I_i^{\frac{1}{2}} \right\rfloor \mathbb{E}\left[D_{t_i+1}\right] - \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)\left(\left\lfloor I_i^{\frac{1}{2}} \right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}} \right\rfloor\right)^{\frac{1}{2}} \\ \geq \quad & \frac{1}{2}\left\lfloor I_i^{\frac{1}{2}} \right\rfloor \mathbb{E}\left[D_{t_i+1}\right] \geq y^h - y^l \geq x_{t_i+1} - \hat{y}_i. \end{aligned}$$

Based on (2.38), we define event $\mathcal{A}_2$ as

$$\mathcal{A}_2 = \left\{ \sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}} \right\rfloor} D_t \geq \left\lfloor I_i^{\frac{1}{2}} \right\rfloor \mathbb{E}\left[D_{t_i+1}\right] - \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)\left(\left\lfloor I_i^{\frac{1}{2}} \right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}} \right\rfloor\right)^{\frac{1}{2}} \right\}. \quad (2.39)$$

Then (2.38) can be restated as

$$\mathbb{P}(\mathcal{A}_2) \geq 1 - \frac{1}{\left\lfloor I_i^{\frac{1}{2}} \right\rfloor^2}. \quad (2.40)$$

On the event $\mathcal{A}_2$, the inventory order up-to level $\hat{y}_i$ will be achieved after periods $\left\{t_i+1, \ldots, t_i+\left\lfloor I_i^{\frac{1}{2}} \right\rfloor\right\}$. By (2.40), we have

$$\begin{aligned} & \mathbb{E}\left[\sum_{t=t_i+1}^{t_i+I_i}\left(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\right)\right] \\ = \quad & \mathbb{P}(\mathcal{A}_2)\mathbb{E}\left[\sum_{t=t_i+1}^{t_i+I_i}\left(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\right)\middle|\mathcal{A}_2\right] + \mathbb{P}(\mathcal{A}_2^c)\mathbb{E}\left[\sum_{t=t_i+1}^{t_i+I_i}\left(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\right)\middle|\mathcal{A}_2^c\right] \\ \leq \quad & \max\{h, b\}(y^h - y^l)\left\lfloor I_i^{\frac{1}{2}} \right\rfloor + \frac{1}{\left\lfloor I_i^{\frac{1}{2}} \right\rfloor^2}\max\{h, b\}(y^h - y^l)I_i \\ \leq \quad & 2\max\{h, b\}(y^h - y^l)I_i^{\frac{1}{2}}, \end{aligned}$$

where the first inequality follows from, for periods $t = t_i + 1, \ldots, t_i + I_i$, that

$$|G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)| = |G(\hat{p}_i, \hat{y}_{i,1}) - G(\hat{p}_i, y_t)| \leq \max\{h, b\}(y^h - y^l),$$

and $\mathbb{P}(\mathcal{A}_2^c) \leq 1 \Big/ \left\lfloor I_i^{1/2} \right\rfloor^2$. Similarly, for large enough i that is greater than or equal to i^{**}, we can establish

$$\mathbb{E}\left[\sum_{t=t_i+I_i+1}^{t_i+2I_i} \left(G(\hat{p}_i+\delta_i, \hat{y}_{i,2}) - G(p_t, y_t)\right)\right] \leq 2\max\{h,b\}(y^h - y^l) I_i^{\frac{1}{2}}.$$

Based on the analysis above, we upper bound (2.36). Let $K_{36} = \sum_{i=1}^{i^{**}} \max\{h,b\}(y^h - y^l) I_i$, it can be seen that there exist some constants K_{37} and K_{38} such that

$$\begin{aligned}
& \mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+I_i}\left(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\right) + \sum_{t=t_i+I_i+1}^{t_i+2I_i}\left(G(\hat{p}_i+\delta_i, \hat{y}_{i,2}) - G(p_t, y_t)\right)\right)\right] \\
\leq \quad & \sum_{i=1}^{i^{**}} \max\{h,b\}(y^h - y^l) I_i + \sum_{i=i^{**}+1}^{n} 4\max\{h,b\}(y^h - y^l) I_i^{\frac{1}{2}} \\
\leq \quad & K_{36} + 4\max\{h,b\}(y^h - y^l) I_0^{\frac{1}{2}} \frac{v^{\frac{1}{2}}(1 - (v^{\frac{1}{2}})^n)}{1 - v^{\frac{1}{2}}} \\
\leq \quad & K_{36} + K_{37}(v^{\frac{1}{2}})^{n+1} \\
\leq \quad & K_{36} + K_{37} v^{\log_v(\frac{v-1}{2I_0 v}T+1)^{\frac{1}{2}}} \\
\leq \quad & K_{38} T^{\frac{1}{2}}. \qquad (2.41)
\end{aligned}$$

By combining (2.35) and (2.41), we conclude

$$R(DDA, T) \leq \frac{1}{T}\left(K_{35} T^{\frac{1}{2}} + K_{38} T^{\frac{1}{2}}\right) \leq K_{39} T^{-\frac{1}{2}}$$

for some constant K_{39}. The proof of Theorem 2 is thus complete.

2.5 Conclusion

In this chapter, we consider a joint pricing and inventory control problem when the firm does not have prior knowledge about the demand distribution and customer response to selling prices. We impose virtually no explicit assumption about how

the average demand changes in price (other than the fact that it is decreasing) and on the distribution of uncertainty in demand. This chapter is the first to design a nonparametric algorithm data-driven learning algorithm for dynamic joint pricing and inventory control problem and present the convergence rate of policies and profits to those of the optimal ones. The regret of the learning algorithm converges to zero at a rate that is the theoretical lower bound $O(T^{-1/2})$.

There are a number of follow-up research topics. One is to develop an asymptotically optimal algorithm for the problem with lost-sales and censored data. In the lost-sales case, the DDA algorithm proposed here cannot be directly applied and the estimation and optimization problems are more challenging as the profit function of the data-driven problem is neither concave nor unimodal, and the demand data is censored. Another interesting direction for research is to develop a data-driven learning algorithm for dynamic pricing and stocking decisions for multiple products in an assortment.

2.6 Appendix

In this Appendix, we provide the technical lemmas and proofs omitted in the main context.

Lemma A1 compares the optimal solutions of problem CI and bridging problem B1, i.e., p^* and $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$.

Lemma A1. *Under Assumption 1, there exists some number $\gamma \in [0, 1)$ such that for any $\hat{p}_i \in \mathcal{P}$, we have*

$$\left| p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) \right| \leq \gamma \left| p^* - \hat{p}_i \right|.$$

Proof. First we make the observation that

$$p^* = \overline{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big). \tag{2.42}$$

This result links the optimal solutions of CI and B1 with parameters $\breve{\alpha}(p^*), \breve{\beta}(p^*)$, and it shows that p^* is a fixed point of $\overline{p}\big(\breve{\alpha}(z), \breve{\beta}(z)\big) = z$. To see why it is true, let

$$G(p, \lambda(p)) = pe^{\lambda(p)}\mathbb{E}[e^{\epsilon}] - \min_{y \in \mathcal{Y}} \left\{ h\mathbb{E}\big[y - e^{\lambda(p)}e^{\epsilon}\big]^+ + b\mathbb{E}\big[e^{\lambda(p)}e^{\epsilon} - y\big]^+ \right\}. \tag{2.43}$$

Then Assumption 1(i) implies that $G(p, \lambda(p))$ is unimodal in p. Assuming that G has a unique maximizer and that $\overline{p}(\breve{\alpha}(z), \breve{\beta}(z))$ is the unique optimal solution for problem B1 with parameters $\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$, then (2.42) follows from Lemma A1 of *Besbes and Zeevi* (2015) by letting their function G be (2.43).

When the optimal solution y over $\mathbb{R}_+$ for problem CI for a given p falls in $\mathcal{Y}$, $\overline{p}(\alpha, \beta)$ is the maximizer of $pe^{\alpha - \beta p}\mathbb{E}[e^{\epsilon}] - Ae^{\alpha - \beta p}$, where $A = \min_z \big\{ h\mathbb{E}[z - e^{\epsilon}]^+ + b\mathbb{E}[e^{\epsilon} - z]^+ \big\}$ is a constant. Thus $\overline{p}(\alpha, \beta)$ satisfies

$$\big(1 - \beta\overline{p}(\alpha, \beta)\big)\mathbb{E}[e^{\epsilon}] + A\beta = 0.$$

Letting $\alpha = \breve{\alpha}(z)$, $\beta = \breve{\beta}(z)$ and taking derivative of $\overline{p}\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$ with respect to z yield

$$\frac{d\overline{p}\big(\breve{\alpha}(z), \breve{\beta}(z)\big)}{dz} = \frac{\lambda''(z)}{(\lambda'(z))^2} = \frac{\tilde{\lambda}''(z)\tilde{\lambda}(z)}{(\tilde{\lambda}'(z))^2} - 1.$$

By Assumption 1(ii), we have $\left| \frac{d\overline{p}\left(\breve{\alpha}(z), \breve{\beta}(z)\right)}{dz} \right| < 1$ for any $z \in \mathcal{P}$. This shows that

$$\left| \overline{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big) - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) \right| \leq \gamma \left| p^* - \hat{p}_i \right|,$$

where $\gamma = \max_{z\in\mathcal{P}} \left| \frac{d\overline{p}\left(\breve{\alpha}(z),\breve{\beta}(z)\right)}{dz} \right| < 1$. This proves Lemma A1. □

To compare the optimal solutions of Problems B1 and B2, we need several technical Lemmas. To that end, we change the decision variables in B1 and B2. For given parameters α and $\beta > 0$, define $d = e^{\alpha-\beta p}, d \in \mathcal{D} = [d^l, d^h]$ where $d^l = e^{\alpha-\beta p^h}$ and $d^h = e^{\alpha-\beta p^l}$. Then problem B1 can be rewritten as

$$\max_{d\in\mathcal{D}} \left\{ d\frac{\alpha - \log d}{\beta}\mathbb{E}\left[e^{\epsilon}\right] - \min_{y\in\mathcal{Y}} \left\{ h\mathbb{E}\left[y - de^{\epsilon}\right]^+ + b\mathbb{E}\left[de^{\epsilon} - y\right]^+ \right\} \right\}.$$

Define

$$\overline{W}(d, y) = h\mathbb{E}\big(y - de^{\epsilon}\big)^+ + b\mathbb{E}\big(de^{\epsilon} - y\big)^+ \tag{2.44}$$

and

$$\overline{G}(\alpha, \beta, d) = d\frac{\alpha - \log d}{\beta}\mathbb{E}\left[e^{\epsilon}\right] - \min_{y\in\mathcal{Y}} \overline{W}(d, y) = d\frac{\alpha - \log d}{\beta}\mathbb{E}\left[e^{\epsilon}\right] - \overline{W}(d, \overline{y}(d)), \tag{2.45}$$

where $\overline{y}(d)$ is the optimal solution of (2.44) in $\mathcal{Y}$ for given d. Let $F(\cdot)$ be the cumulative distribution function (CDF) of e^{ϵ}, then it can be verified that

$$\overline{y}(d) = dF^{-1}\left(\frac{b}{b+h}\right), \tag{2.46}$$

where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Also, we let $\overline{d}\,(\alpha, \beta)$ denote the optimal solution of maximizing (2.45) in $\mathcal{D}$.

Similarly, we reformulate problem B2 with decision variables d and y as

$$\max_{d\in\mathcal{D}} \left\{ d\frac{\alpha - \log d}{\beta} \left(\frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t} \right) - \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \left(h\big(y - de^{\epsilon_t}\big)^+ + b\big(de^{\epsilon_t} - y\big)^+ \right) \right\} \right\}$$

Let

$$\tilde{W}_{i+1}(d,y) = \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(y-de^{\epsilon_t}\big)^+ + b\big(de^{\epsilon_t}-y\big)^+\Big), \tag{2.47}$$

and

$$\begin{aligned}\tilde{G}_{i+1}(\alpha,\beta,d) &= d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right) - \min_{y\in\mathcal{Y}}\tilde{W}_{i+1}(d,y)\\ &= d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right) - \tilde{W}_{i+1}(d,\tilde{y}(d)),\end{aligned} \tag{2.48}$$

where $\tilde{y}_{i+1}(d)$ denotes the optimal solution of $\tilde{W}_{i+1}(d,y)$ in (2.47) on $\mathcal{Y}$. Let $\tilde{d}_{i+1}(\alpha,\beta)$ be the optimal solution for $\tilde{G}_{i+1}(\cdot,\cdot,d)$ in (2.48) on $\mathcal{D}$. Also, let $\tilde{y}^u_{i+1}(d)$ denote the optimal order-up-to level for problem B2 on $\mathbb{R}_+$ for given $p\in\mathcal{P}$ (here the superscript "u" stands for "unconstrained"). Then

$$\tilde{y}^u_{i+1}(d) = \min\left\{de^{\epsilon_j}:\ \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\le e^{\epsilon_j}\right\}\ge\frac{b}{b+h}\right\}, \tag{2.49}$$

where $\mathbb{1}\{A\}$ is the indicator function taking value 1 if "A" is true and 0 otherwise. It can be checked that

$$\tilde{y}_{i+1}(d) = \min\left\{\max\left\{\tilde{y}^u_{i+1}(d), y^l\right\}, y^h\right\}. \tag{2.50}$$

Since $\tilde{y}_{i+1}(d)$ is random, it is possible for $\tilde{y}_{i+1}(d)$ to take value at the boundary, y^h or y^l.

We first compare the profit functions defined for the two problems (2.44), (2.45), and (2.47), (2.48). To this end, we need the following properties.

Lemma A2. If $\beta>0$, then both $\overline{G}(\alpha,\beta,d)$ and $\tilde{G}_{i+1}(\alpha,\beta,d)$ are concave in $d\in\mathcal{D}$, and both $\overline{G}(\alpha,\beta,e^{\alpha-\beta p})$ and $\tilde{G}_{i+1}(\alpha,\beta,e^{\alpha-\beta p})$ are unimodal in $p\in\mathcal{P}$.

Proof. It is easily seen that $\overline{W}(d,y)$ and $\tilde{W}_{i+1}(d,y)$ are both jointly convex in (d,y), hence $\min_{y\in\mathcal{Y}}\overline{W}(d,y)$ and $\min_{y\in\mathcal{Y}}\tilde{W}_{i+1}(d,y)$ are convex in d (Proposition B4 of *Heyman and Sobel* (1984)). Therefore, the results follow from that the first term of $\overline{G}$ (and $\tilde{G}_{i+1}$) is concave when $\beta > 0$.

The unimodality of $\overline{G}(\alpha,\beta,e^{\alpha-\beta p})$ and $\tilde{G}_{i+1}(\alpha,\beta,e^{\alpha-\beta p})$ follows from the concavity of $\overline{G}$ and $\tilde{G}_{i+1}$, and the fact that $e^{\alpha-\beta p}$ is strictly decreasing in p when $\beta > 0$. □

The following important result shows that, for any given d, $\overline{W}(d,\overline{y}(d))$ and $\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))$ are close to each other with high probability.

Lemma A3. There exists a positive constant K_{40} such that, for any $\xi > 0$,

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \le K_{40}\xi\right\} \ge 1 - 4e^{-2I_i\xi^2}.$$

Proof. By triangle inequality, we have

$$\begin{aligned} & \max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \\ \le\ & \max_{d\in\mathcal{D}}\left|\overline{W}\big(d,\overline{y}(d)\big) - \tilde{W}_{i+1}\big(d,\overline{y}(d)\big)\right| + \max_{d\in\mathcal{D}}\left|\tilde{W}_{i+1}\big(d,\overline{y}(d)\big) - \tilde{W}_{i+1}\big(d,\tilde{y}_{i+1}(d)\big)\right|. \end{aligned} \tag{2.51}$$

In what follows we develop upper bounds for $\max_{d\in\mathcal{D}}|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\overline{y}(d))|$ and $\max_{d\in\mathcal{D}}|\tilde{W}_{i+1}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))|$ separately.

For any $d\in\mathcal{D}$ and $y\in\mathcal{Y}$, we define $z = y/d$. Then, from (2.46), the optimal z to minimize $\overline{W}(d,dz)$ is

$$\overline{z} = \frac{\overline{y}(d)}{d} = F^{-1}\left(\frac{b}{b+h}\right).$$

Moreover, we have

$$\overline{W}(d,\overline{y}(d)) = \overline{W}(d,d\overline{z}) = d\Big(h\mathbb{E}\big(\overline{z} - e^{\epsilon}\big)^{+} + b\mathbb{E}\big(e^{\epsilon} - \overline{z}\big)^{+}\Big),$$

and

$$\tilde{W}_{i+1}(d,\overline{y}(d)) = \tilde{W}_{i+1}(d,d\overline{z}) = d\bigg(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(\overline{z}-e^{\epsilon_t}\big)^+ + b\big(e^{\epsilon_t}-\overline{z}\big)^+\Big)\bigg). \quad (2.52)$$

For $t \in \{t_i+1,\ldots,t_i+2I_i\}$, denote

$$\Delta_t = \big(h\mathbb{E}[\overline{z}-e^{\epsilon_t}]^+ + b\mathbb{E}[e^{\epsilon_t}-\overline{z}]^+\big) - \big(h(\overline{z}-e^{\epsilon_t})^+ + b(e^{\epsilon_t}-\overline{z})^+\big).$$

Then $\mathbb{E}[\Delta_t] = 0$. Since ϵ_t is bounded, so is Δ_t, thus we apply Hoeffding inequality (see Theorem 1 in *Hoeffding* (1963), and *Levi et al.* (2007) for its application in newsvendor problems) to obtain, for any $\xi > 0$,

$$\mathbb{P}\left\{d^h\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right| > d^h\xi\right\} = \mathbb{P}\left\{\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right| > \xi\right\} \le 2e^{-4I_i\xi^2}, \quad (2.53)$$

which deduces to

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\overline{y}(d))\right| > d^h\xi\right\} \le 2e^{-4I_i\xi^2}. \quad (2.54)$$

This bounds the first term on the right hand side of (2.51).

To bound the second term in (2.51), we use

$$\hat{F}(x) = \frac{1}{2I_i}\sum_{t=1}^{2I_i}\mathbb{1}\left\{e^{\epsilon_t}\le x\right\}, \qquad x\in[l,u]$$

to denote the empirical distribution of e^{ϵ_t}. For $\theta > 0$, we call $\hat{F}(\overline{z})$ a θ-estimate of $F(\overline{z})$ $(= b/(b+h))$, or simply a θ-estimate, if

$$\left|\hat{F}(\overline{z}) - \frac{b}{b+h}\right| \le \theta. \quad (2.55)$$

It can be verified that

$$\begin{aligned}\mathbb{P}\left\{\hat{F}(\overline{z}) < \frac{b}{b+h} - \theta\right\} &= \mathbb{P}\left\{\hat{F}(\overline{z}) < F(\overline{z}) - \theta\right\} \\ &= \mathbb{P}\left\{\hat{F}(\overline{z}) - F(\overline{z}) < -\theta\right\} \\ &\leq e^{-2I_i\theta^2},\end{aligned}$$

where the last inequality follows from Hoeffding inequality. Similarly, we have

$$\mathbb{P}\left\{\hat{F}(\overline{z}) > \frac{b}{b+h} + \theta\right\} \leq e^{-2I_i\theta^2}.$$

Combining the two results above we obtain

$$\mathbb{P}\left\{\left|\hat{F}(\overline{z}) - \frac{b}{b+h}\right| \leq \theta\right\} \geq 1 - 2e^{-2I_i\theta^2}.$$

Let $\mathcal{A}_3(\theta)$ represent the event that $\hat{F}(\overline{z})$ is a θ-estimate, then the result above states that

$$\mathbb{P}(\mathcal{A}_3(\theta)) \geq 1 - 2e^{-2I_i\theta^2}. \tag{2.56}$$

For $d \in \mathcal{D}$, let $\tilde{z}_{i+1}(d) = \frac{\tilde{y}_{i+1}(d)}{d}$ and $\tilde{z}^u_{i+1} = \frac{\tilde{y}^u_{i+1}(d)}{d}$, then it follows from (2.49) that

$$\tilde{z}^u_{i+1} = \min\left\{e^{\epsilon_j} : \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \leq e^{\epsilon_j}\right\} \geq \frac{b}{b+h}\right\}.$$

And it follows from (2.50) that

$$\tilde{z}_{i+1}(d) = \min\left\{\max\left\{\tilde{z}^u_{i+1}, \frac{y^l}{d}\right\}, \frac{y^h}{d}\right\}.$$

By $\tilde{y}^u_{i+1}(d) = d\ \tilde{z}^u_{i+1}$, we have $\tilde{W}_{i+1}(d, \tilde{y}^u_{i+1}(d)) = \tilde{W}_{i+1}(d, d\ \tilde{z}^u_{i+1})$. In the following, we develop an upper bound for $\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1})$ when $\hat{F}(\cdot)$ is a θ-estimate.

First, for any given $d \in \mathcal{D}$, if $\overline{z} \leq \tilde{z}^u_{i+1}$, then it follows from (2.52) that

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\overline{z}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(e^{\epsilon_t}-\overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\big\} \\
&\qquad\qquad +b\big(e^{\epsilon_t}-\overline{z}\big)\mathbb{1}\big\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z}-e^{\epsilon_t}\big)\mathbb{1}\big\{e^{\epsilon_t} \leq \overline{z}\big\}\Big] \\
&\leq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(e^{\epsilon_t}-\overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\big\} \\
&\qquad\qquad +b\big(\tilde{z}^u_{i+1}-\overline{z}\big)\mathbb{1}\big\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z}-e^{\epsilon_t}\big)\mathbb{1}\big\{e^{\epsilon_t} \leq \overline{z}\big\}\Big],
\end{aligned} \tag{2.57}$$

where the inequality follows from replacing e^{ϵ_t} in the second term by its upper bound $\tilde{z}^u_{i+1}$, and

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t}-\tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\} \\
&\qquad\qquad +h(\tilde{z}^u_{i+1}-e^{\epsilon_t})\mathbb{1}\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\} + h(\tilde{z}^u_{i+1}-e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \overline{z}\}\Big] \\
&\geq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t}-\tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\} + h(\tilde{z}^u_{i+1}-e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \overline{z}\}\Big],
\end{aligned} \tag{2.58}$$

with the inequality obtained by dropping the nonnegative middle term. Consequently when $\overline{z} \leq \tilde{z}^u_{i+1}$ we subtract (2.58) from (2.57) to obtain

$$\begin{aligned}
&\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \\
\leq\ & d\left(b(\tilde{z}^u_{i+1}-\overline{z})(1-\hat{F}(\tilde{z}^u_{i+1})) + b(\tilde{z}^u_{i+1}-\overline{z})(\hat{F}(\tilde{z}^u_{i+1})-\hat{F}(\overline{z})) + h(\overline{z}-\tilde{z}^u_{i+1})\hat{F}(\overline{z})\right) \\
=\ & d(\tilde{z}^u_{i+1}-\overline{z})\big(-(h+b)\hat{F}(\overline{z})+b\big) \\
\leq\ & d\big(\tilde{z}^u_{i+1}-\overline{z}\big)(b+h)\theta,
\end{aligned} \tag{2.59}$$

where the second inequality follows from $\hat{F}(\overline{z}) \geq \frac{b}{b+h} - \theta$ when $\hat{F}(\cdot)$ is a θ-estimate.

Similarly, if $\overline{z} > \tilde{z}^u_{i+1}$, then

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\overline{z}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t}-\overline{z})\mathbb{1}\{\overline{z}<e^{\epsilon_t}\} \\
&\qquad +h(\overline{z}-e^{\epsilon_t})\mathbb{1}\{\tilde{z}^u_{i+1}<e^{\epsilon_t}\le\overline{z}\}+h(\overline{z}-e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t}\le\tilde{z}^u_{i+1}\}\Big] \\
&\le \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t}-\overline{z})\mathbb{1}\{\overline{z}<e^{\epsilon_t}\} \\
&\qquad +h(\overline{z}-\tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1}<e^{\epsilon_t}\le\overline{z}\}+h(\overline{z}-e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t}\le\tilde{z}^u_{i+1}\}\Big],
\end{aligned} \tag{2.60}$$

where the inequality follows replacing e^{ϵ_t} in the second term by its lower bound $\tilde{z}^u_{i+1}$, and

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t}-\tilde{z}^u_{i+1})\mathbb{1}\{\overline{z}<e^{\epsilon_t}\} \\
&\qquad +b(e^{\epsilon_t}-\tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1}<e^{\epsilon_t}\le\overline{z}\}+h(\tilde{z}^u_{i+1}-e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t}\le\tilde{z}^u_{i+1}\}\Big] \\
&\ge \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t}-\tilde{z}^u_{i+1})\mathbb{1}\{\overline{z}<e^{\epsilon_t}\}+h(\tilde{z}^u_{i+1}-e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t}\le\tilde{z}^u_{i+1}\}\Big],
\end{aligned} \tag{2.61}$$

again the inequality follows from dropping the nonnegative second term. Subtracting (2.61) from (2.60), we obtain

$$\begin{aligned}
&\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \\
\le\ & d\left(b(\tilde{z}^u_{i+1}-\overline{z})(1-\hat{F}(\overline{z}))+h(\overline{z}-\tilde{z}^u_{i+1})(\hat{F}(\overline{z})-\hat{F}(\tilde{z}^u_{i+1}))+h(\overline{z}-\tilde{z}^u_{i+1})\hat{F}(\tilde{z}^u_{i+1})\right) \\
=\ & d(\overline{z}-\tilde{z}^u_{i+1})((h+b)\hat{F}(\overline{z})-b) \\
\le\ & d(\overline{z}-\tilde{z}^u_{i+1})(b+h)\theta,
\end{aligned} \tag{2.62}$$

where the last inequality follows from $\hat{F}(\overline{z}) \le \frac{b}{b+h}+\theta$ when $\hat{F}(\cdot)$ is a θ-estimate.

The results (2.59) and (2.62) imply that, when $\hat{F}(\cdot)$ is a θ-estimate, or (2.55) is satisfied, it holds that

$$\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \leq d\big|\overline{z} - \tilde{z}^u_{i+1}\big|(b+h)\theta.$$

As demand is bounded, $d\tilde{z}^u_{i+1}$ is bounded too, hence it follows from $d\overline{z} \in \mathcal{Y}$ that there exists some constant $K_{41} > 0$ such that $d\big|\overline{z} - \tilde{z}^u_{i+1}\big| \leq K_{41}$. Thus

$$\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \leq K_{41}(b+h)\theta.$$

Since $\tilde{z}^u_{i+1}$ is the unconstrained minimizer of $\tilde{W}_{i+1}(d, dz)$, it follows that

$$\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d)) \leq \tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \leq K_{41}(b+h)\theta.$$

As this inequality holds for any $d \in \mathcal{D}$, it implies that, when $\hat{F}(\cdot)$ is a θ-estimate, or on the event $\mathcal{A}_3(\theta)$,

$$\max_{d\in\mathcal{D}}\Big\{\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d))\Big\} \leq K_{41}(b+h)\theta. \tag{2.63}$$

Letting $\theta = \xi$ in (2.63) we obtain

$$\begin{aligned} & \mathbb{P}\left\{\max_{d\in\mathcal{D}}\Big(\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d))\Big) \leq K_{41}(b+h)\xi\right\} \\ \geq\ & \mathbb{P}(\mathcal{A}_3(\xi)) \\ \geq\ & 1 - 2e^{-2I_i\xi^2}, \end{aligned}$$

where the last inequality follows from (2.56). This proves, by noting $\tilde{W}_{i+1}(d, \overline{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \geq 0$ as $\tilde{y}_{i+1}(d)$ is the minimizer of $\tilde{W}_{i+1}$ on $\mathcal{Y}$, that

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\Big|\Big(\tilde{W}_{i+1}(d, \overline{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d))\Big)\Big| \leq K_{41}(b+h)\xi\right\} \geq 1 - 2e^{-2I_i\xi^2}. \tag{2.64}$$

Applying (2.54) and (2.64) in (2.51), we conclude that there exist a constant $K_{40} > 0$ such that for any $\xi > 0$, when I_i is sufficiently large,

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \le K_{40}\xi\right\} \ge 1 - 2e^{-2I_i\xi^2} - 2e^{-4I_i\xi^2} \ge 1 - 4e^{-2I_i\xi^2}.$$

This completes the proof of Lemma A3. □

Having compared functions $\overline{W}$ and $\tilde{W}_{i+1}$, we next compare $\overline{G}$ with $\tilde{G}_{i+1}$.

Lemma A4. Given parameters α and β, there exist a positive constant K_{42} such that, for any $\xi > 0$,

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{G}\left(\alpha,\beta,d\right) - \tilde{G}_{i+1}\left(\alpha,\beta,d\right)\right| \ge K_{42}\xi\right\} \le 5e^{-2I_i\xi^2}.$$

Proof. For any $d \in \mathcal{D}$, similar argument as that used in proving (2.53) of Lemma A2 shows that, for any $\xi > 0$,

$$\mathbb{P}\left\{\left|\mathbb{E}[e^{\epsilon_t}] - \left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le \xi\right\} \ge 1 - e^{-4I_i\xi^2},$$

where $\sigma = \sqrt{\mathrm{Var}(e^{\epsilon_t})}$. Let $r^* = \max_{d\in\mathcal{D}} \frac{|\alpha - \log d|}{\beta} d$, then we have

$$\begin{aligned}
& \mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}[e^{\epsilon_t}] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*\xi\right\} \\
= \; & \mathbb{P}\left\{r^*\left|\mathbb{E}[e^{\epsilon_t}] - \left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*\xi\right\} \\
\ge \; & 1 - e^{-4I_i\xi^2}. \qquad (2.65)
\end{aligned}$$

Hence, it follows from (2.45) and (2.48) that, for any $d \in \mathcal{D}$ and $\xi > 0$,

$$
\begin{aligned}
& \mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{G}(\alpha,\beta,d)-\tilde{G}_{i+1}(\alpha,\beta,d)\right| \le (K_{40}+r^*)\xi\right\} \\
= \;& \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\Bigg|\left(d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big]\right. \\
& \qquad \left.-\overline{W}(d,\overline{y}(d))\right)-\left(d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right)-\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right)\Bigg| \le (K_{40}+r^*)\xi\Bigg\} \\
\ge \;& \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big]-d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right)\right| \\
& \qquad +\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)-\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \le (K_{40}+r^*)\xi\Bigg\} \\
\ge \;& \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big]-d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right)\right| \le r^*\xi, \\
& \qquad \text{and } \max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)-\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \le K_{40}\xi\Bigg\} \\
= \;& 1-\mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big]-d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right)\right| > r^*\xi, \\
& \qquad \text{or } \max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d))-\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| > K_{40}\xi\Bigg\} \\
\ge \;& 1-\mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big]-d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}e^{\epsilon_t}\right)\right| > r^*\xi\Bigg\} \\
& \qquad -\mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)-\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| > K_{40}\xi\Bigg\} \\
\ge \;& 1-e^{-4I_i\xi^2}-4e^{-2I_i\xi^2} \\
\ge \;& 1-5e^{-2I_i\xi^2},
\end{aligned}
$$

where the last inequality follows from (2.65) and Lemma A2. Letting $K_{42} = K_{40}+2r^*\sigma$ completes the proof of Lemma A4. □

For any $\xi > 0$, we define event

$$
\mathcal{A}_4(\xi) = \left\{\omega : \max_{d\in\mathcal{D}}\left|\overline{G}(\alpha,\beta,d)-\tilde{G}_{i+1}(\alpha,\beta,d)\right| \le K_{42}\xi\right\}. \tag{2.66}
$$

Then Lemma A4 can be reiterated as $\mathbb{P}(\mathcal{A}_4(\xi)) \geq 1 - 5e^{-2I_i\xi^2}$.

With the preparations above, we are now ready to compare the optimal solutions of problems B1 and B2. Different from B1, in problem B2 the distribution of ϵ in the objective function is unknown, hence the expectations are replaced by their sample averages, giving rise to the SAA problem. Lemma A5 below presents a useful result that bounds the probability for the optimal solution of problem B2 to be away from that of problem B1. Since I_i tends to infinity as t goes to infinity, this shows that the probability that the two solutions, $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$, are significantly different converges to zero when the length of the planning horizon T increases.

Lemma A5. *For any $p \in \mathcal{P}$ and any $\xi > 0$,*

$$\mathbb{P}\left\{\left|\overline{p}\left(\breve{\alpha}(p), \breve{\beta}(p)\right) - \tilde{p}_{i+1}\left(\breve{\alpha}(p), \breve{\beta}(p)\right)\right| \geq K_{43}\xi^{\frac{1}{2}}\right\} \leq 5e^{-4I_i\xi^2}$$

for some positive constant K_{43}.

Proof. To slightly simplify the notation, for given parameters α and β, in this proof we let

$$\overline{G}(d) = \overline{G}(\alpha, \beta, d), \quad \tilde{G}(d) = \tilde{G}_{i+1}(\alpha, \beta, d), \quad \overline{d} = \overline{d}(\alpha, \beta), \quad \tilde{d} = \tilde{d}_{i+1}(\alpha, \beta).$$

By Taylor's expansion,

$$\overline{G}(\tilde{d}) = \overline{G}(\overline{d}) + \overline{G}'(\overline{d})(\tilde{d} - \overline{d}) + \frac{\overline{G}''(q)}{2}(\tilde{d} - \overline{d})^2, \tag{2.67}$$

where $q \in [\overline{d}, \tilde{d}]$ if $\overline{d} \leq \tilde{d}$ and $q \in [\tilde{d}, \overline{d}]$ if $\overline{d} > \tilde{d}$. Since we assume the minimizer of $\overline{W}(d, y)$ over $\mathbb{R}_+$ falls into $\mathcal{Y}$, it follows from (2.45) that $\overline{G}(d) = d^{\frac{\alpha - \log d}{\beta}}\mathbb{E}[e^{\epsilon}] - Ad$,

where $A = \min_z \{h\mathbb{E}(z - e^{\epsilon})^+ + b\mathbb{E}(e^{\epsilon} - z)^+\} > 0$ is a constant. Thus, we have

$$\overline{G}''(d) = -\frac{\mathbb{E}[e^{\epsilon}]}{\beta d}.$$

Since $\lambda(\cdot)$ is assumed to be strictly decreasing, it follows that $\breve{\beta}(\cdot)$ is bounded below by a positive number, say $\bar{a} > 0$. On $\beta \geq \bar{a}$, let $\min_{d\in\mathcal{D}} \frac{\mathbb{E}[e^{\epsilon}]}{\beta d} = m$ and it holds that $m > 0$, then it follows from (2.67) that

$$\overline{G}(\tilde{d}) \leq \overline{G}(\overline{d}) - \frac{m}{2}(\tilde{d} - \overline{d})^2. \tag{2.68}$$

Now we prove, on event $\mathcal{A}_4(\xi)$, that

$$\overline{G}(\tilde{d}) - \overline{G}(\overline{d}) \geq -2K_{42}\xi. \tag{2.69}$$

We prove this by contradiction. Suppose it is not true, i.e., $\overline{G}(\overline{d}) - \overline{G}(\tilde{d}) > 2K_{42}\xi$, then it follows from (2.66) that

$$\begin{aligned}
& \tilde{G}(\overline{d}) - \tilde{G}(\tilde{d}) \\
= \; & \big(\tilde{G}(\overline{d}) - \overline{G}(\overline{d})\big) + \big(\overline{G}(\overline{d}) - \overline{G}(\tilde{d})\big) + \big(\overline{G}(\tilde{d}) - \tilde{G}(\tilde{d})\big) \\
> \; & -K_{42}\xi + 2K_{42}\xi - K_{42}\xi \\
= \; & 0.
\end{aligned}$$

This leads to $\tilde{G}(\overline{d}) > \tilde{G}(\tilde{d})$, contradicting with $\tilde{d}$ being optimal for problem B2. Thus, (2.69) is satisfied on $\mathcal{A}_4(\xi)$.

Using (2.68) and (2.69), we obtain that, on event $\mathcal{A}_4(\xi)$,

$$\left|\tilde{d} - \overline{d}\right|^2 \leq \frac{4K_{42}}{m}\xi,$$

or equivalently, for some constant K_{44},

$$\left|\tilde{d} - \overline{d}\right| \leq K_{44}\xi^{\frac{1}{2}}.$$

Let $g(d) = \frac{\alpha - \log d}{\beta}$, then $\overline{p}(\alpha, \beta) = g(\overline{d})$ and $\tilde{p}_{i+1}(\alpha, \beta) = g(\tilde{d})$. Since the first order derivative of $g(d)$ with respect to $d \in \mathcal{D}$ is bounded, there exist constant $K_{45} > 0$, such that on $\mathcal{A}_4(\xi)$, it holds that

$$|\overline{p}(\alpha, \beta) - \tilde{p}_{i+1}(\alpha, \beta)| = |g(\overline{d}) - g(\tilde{d})| \leq K_{45}|\overline{d} - \tilde{d}| \leq K_{44} \times K_{45}\xi^{\frac{1}{2}}.$$

Letting $K_{43} = K_{44} \times K_{45}$, this shows that for any values of α and $\beta \geq \bar{a}$,

$$\mathbb{P}\left\{|\overline{p}(\alpha, \beta) - \tilde{p}_{i+1}(\alpha, \beta)| \leq K_{43}\xi^{\frac{1}{2}}\right\} \geq \mathbb{P}(\mathcal{A}_4(\xi)) \geq 1 - 5e^{-2I_i\xi^2}.$$

Substituting $\alpha = \breve{\alpha}(p)$ and $\beta = \breve{\beta}(p)$, we obtain the desired result in Lemma A5. $\square$

Lemma A6 shows that $(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$, $(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$ and $(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i))$ approach each other when i gets large.

Lemma A6. *There exists a positive constant K_{46} such that*

$$\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + |\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}|^2\right] \leq K_{46}I_i^{-\frac{1}{2}}.$$

Proof. The proof of this result bears similarity with that of *Besbes and Zeevi* (2015), hence here we only present the differences. For convenience we define

$$B_{i+1}^1 = \frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} \epsilon_t, \qquad B_{i+1}^2 = \frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} \epsilon_t.$$

Recall that $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$ are derived from the least-square method, and they are

given by

$$\hat{\alpha}_{i+1} = \frac{\lambda(\hat{p}_i) + \lambda(\hat{p}_i + \delta_i)}{2} + \frac{B^1_{i+1} + B^2_{i+1}}{2} + \hat{\beta}_{i+1}\frac{2\hat{p}_i + \delta_i}{2}, \tag{2.70}$$

$$\hat{\beta}_{i+1} = -\frac{\lambda(\hat{p}_i + \delta_i) - \lambda(\hat{p}_i)}{\delta_i} - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}). \tag{2.71}$$

Applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the second order for (2.71), we obtain

$$\begin{aligned}\hat{\beta}_{i+1} &= -\left(\lambda'(\hat{p}_i) + \frac{1}{2}\lambda''(q_i)\delta_i\right) - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}) \\ &= \breve{\beta}(\hat{p}_i) - \frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}),\end{aligned} \tag{2.72}$$

where $q_i \in [\hat{p}_i, \hat{p}_i + \delta_i]$. Substituting $\hat{\beta}_{i+1}$ in (2.70) by (2.72), and applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the first order, we have

$$\begin{aligned}\hat{\alpha}_{i+1} &= \lambda(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B^1_{i+1} + B^2_{i+1}}{2} - \lambda'(\hat{p}_i)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\ &\qquad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1})\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\ &= \breve{\alpha}(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B^1_{i+1} + B^2_{i+1}}{2} - \frac{1}{2}\lambda'(\hat{p}_i)\delta_i \\ &\qquad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1})\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right),\end{aligned} \tag{2.73}$$

where $q_i' \in [\hat{p}_i, \hat{p}_i + \delta_i]$.

Since the error terms ϵ_t are assumed to be bounded, we apply Hoeffding inequality to obtain

$$\mathbb{P}\left\{\left|-B^1_{i+1}\right| > \xi\right\} \leq 2e^{-2I_i\xi^2}, \qquad \mathbb{P}\left\{\left|B^2_{i+1}\right| > \xi\right\} \leq 2e^{-2I_i\xi^2}.$$

Hence,

$$\mathbb{P}\left\{\left|-B_{i+1}^1\right|+\left|B_{i+1}^2\right|>2\xi\right\}\leq\mathbb{P}\left\{\left|-B_{i+1}^1\right|>\xi\right\}+\mathbb{P}\left\{\left|B_{i+1}^2\right|>\xi\right\}\leq 4e^{-2I_i\xi^2}.$$

Therefore,

$$\mathbb{P}\left\{\left|-B_{i+1}^1+B_{i+1}^2\right|\leq 2\xi\right\}\geq\mathbb{P}\left\{\left|-B_{i+1}^1\right|+\left|B_{i+1}^2\right|\leq 2\xi\right\}\geq 1-4e^{-2I_i\xi^2}.$$

Similar argument shows

$$\mathbb{P}\left\{\left|B_{i+1}^1+B_{i+1}^2\right|\leq 2\xi\right\}\geq 1-4e^{-2I_i\xi^2}.$$

Since $\lambda'(\cdot)$ and $\lambda''(\cdot)$ are bounded and δ_i converges to 0, from (2.73) we conclude that there must exist a constant K_{47} such that, on the event $\left|B_{i+1}^1+B_{i+1}^2\right|\leq 2\xi$ and $\left|-B_{i+1}^1+B_{i+1}^2\right|\leq 2\xi$, it holds that

$$|\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i)|\leq K_{47}\left(\delta_i+\frac{\xi}{\delta_i}+\xi\right).$$

Therefore,

$$\begin{aligned}\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i)|\leq K_{47}\left(\delta_i+\frac{\xi}{\delta_i}+\xi\right)\right\} &\geq \mathbb{P}\left\{\left|B_{i+1}^1+B_{i+1}^2\right|\leq 2\xi,\left|-B_{i+1}^1\right|+\left|B_{i+1}^2\right|\leq 2\xi\right\}\\ &\geq 1-8e^{-2I_i\xi^2},\end{aligned}$$

which implies

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i)|^2\leq K_{48}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\right\}\geq 1-8e^{-2I_i\xi^2}. \tag{2.74}$$

From (2.72) we have

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|\leq K_{49}\left(\delta_i+\frac{\xi}{\delta_i}\right)\right\}\geq 1-4e^{-2I_i\xi^2},$$

which implies

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2\leq K_{50}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}\right)\right\}\geq 1-4e^{-2I_i\xi^2}. \tag{2.75}$$

Following the development of (2.74) and (2.75), we have

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i+\delta_i)|^2\leq K_{51}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\right\}\geq 1-8e^{-2I_i\xi^2}. \tag{2.76}$$

and

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i+\delta_i)\right|^2\leq K_{52}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}\right)\right\}\geq 1-4e^{-2I_i\xi^2}. \tag{2.77}$$

Combining(2.74), (2.75), (2.76), and (2.77), we obtain

$$\begin{aligned}\mathbb{P}\Big\{&|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i)|^2+\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2+|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i+\delta_i)|^2+\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i+\delta_i)\right|^2\\ &\leq K_{53}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\Big\} \qquad (2.78)\\ \geq\ & 1-24e^{-2I_i\xi^2},\end{aligned}$$

which is

$$\begin{aligned}\mathbb{P}\Big\{&\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right)^{-1}\Big(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2\\ &+|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}|^2-K_{53}\delta_i^2\Big)\geq\xi^2\Big\}<24e^{-2I_i\xi^2}.\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\Big[\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right)^{-1}\Big(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}|^2 \\
&\qquad\qquad +|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}|^2-K_{53}\delta_i^2\Big)\Big] \\
=\;&\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right)^{-1}\mathbb{E}\Big[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}|^2 \\
&\qquad\qquad +|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}|^2\Big]-\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right)^{-1}K_{53}\delta_i^2 \\
\leq\;&\int_0^{+\infty}24e^{-2I_i\xi}d\xi \\
=\;&\frac{12}{I_i}.
\end{aligned}$$

Hence one has

$$\begin{aligned}
&\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}|^2\right] \\
\leq\;&\left(\frac{12}{I_i}+\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right)^{-1}K_{53}\delta_i^2\right)\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right) \\
\leq\;&K_{46}I_i^{-\frac{1}{2}}. \qquad (2.79)
\end{aligned}$$

This completes the proof of Lemma A6. □

Lemma A7 bounds the difference between the solution for problem B2, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)$, and the solution for problem DD, $\hat{p}_{i+1}$. Comparing the two problems, we note that there are two main differences: First, problem DD has an affine function with coefficients $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$, while problem B2 has an affine function with coefficients $\breve{\alpha}(\hat{p}_i)$ and $\breve{\beta}(\hat{p}_i)$; second, in problem DD, the biased sample of demand uncertainty, η_t, is used, while in problem B2, an unbiased sample ϵ_t is used. Despite those differences, we have the following result.

Lemma A7. *There exists some positive constants K_{56} and i^* such that for any*

$i \geq i^*$ one has

$$\mathbb{P}\Bigg\{ \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big| \geq K_{56}\big(\big|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}\big| \\ + \big|\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}\big|\big)\Bigg\} \leq \frac{8}{I_i},$$

$$\mathbb{P}\Bigg\{ \Big|\tilde{y}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{y}_{i+1}\Big| \geq K_{56}\big(\big|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}\big| \\ + \big|\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}\big|\big)\Bigg\} \leq \frac{8}{I_i}.$$

Proof. To compare the solutions of these two problems, we introduce a general function based on the data-driven problem DD and problem B2: Given selling price $p_t = \hat{p}_i$ for $t = t_i+1, \ldots, t_i+I_i$ and $p_t = \hat{p}_i+\delta_i$ for $t = t_i+I_i+1, \ldots, t_i+2I_i$, logarithm demand data $D_t, t = t_i+1, \ldots, t_i+2I_i$, and two sets of parameters $(\alpha_1, \beta_1), (\alpha_2, \beta_2)$, define $\zeta_{t=t_i+1}^{t_1+I_i}(\alpha_1, \beta_1) = (\zeta_{t_i+1}, \ldots, \zeta_{t_i+I_i})$ and $\zeta_{t=t_i+I_i+1}^{t_i+2I_i}(\alpha_2, \beta_2) = (\zeta_{t_i+I_i+1}, \ldots, \zeta_{t_i+2I_i})$ by

$$\begin{aligned} \zeta_t &= D_t - (\alpha_1 - \beta_1 p_t) = \lambda(\hat{p}_i) + \epsilon_t - (\alpha_1 - \beta_1 \hat{p}_i), && t = t_i+1, \ldots, t_i+I_i, \\ \zeta_t &= D_t - (\alpha_2 - \beta_2 p_t) = \lambda(\hat{p}_i + \delta_i) + \epsilon_t - (\alpha_2 - \beta_2(\hat{p}_i + \delta_i)), && t = t_i+I_i+1, \ldots, t_i+2I_i. \end{aligned}$$

Then, we define a function H_{i+1} by

$$\begin{aligned} & H_{i+1}\Big(p, e^{\alpha_1-\beta_1 p}, \zeta_{t=t_i+1}^{t_1+I_i}(\alpha_1, \beta_1), \zeta_{t=t_i+I_i+1}^{t_i+2I_i}(\alpha_2, \beta_2)\Big) \qquad (2.80) \\ = \; & p e^{\alpha_1-\beta_1 p} \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\zeta_t} - \min_{y \in \mathcal{Y}} \Bigg\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big(h\big(y - e^{\alpha_1-\beta_1 p} e^{\zeta_t}\big)^+ + b\big(e^{\alpha_1-\beta_1 p} e^{\zeta_t} - y\big)^+ \Big) \Bigg\}. \end{aligned}$$

Consider the optimization of H_{i+1}, and let its optimal price be denoted by

$$p\big((\alpha_1, \beta_1), (\alpha_2, \beta_2)\big) = \arg\max_{p \in \mathcal{P}} H_{i+1}\left(p, e^{\alpha_1-\beta_1 p}, \zeta_{t=t_i+1}^{t_1+I_i}(\alpha_1, \beta_1), \zeta_{t=t_i+I_i+1}^{t_i+2I_i}(\alpha_2, \beta_2)\right) \quad (2.81)$$

and its optimal order-up-to level, for given price p, be denoted by

$$y\big(e^{\alpha_1-\beta_1 p},(\alpha_1,\beta_1),(\alpha_2,\beta_2)\big) = \arg\min_{y\in\mathcal{Y}}\bigg\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(y-e^{\alpha_1-\beta_1 p}e^{\zeta_t}\big)^+ + b\big(e^{\alpha_1-\beta_1 p}e^{\zeta_t}-y\big)^+\Big)\bigg\}. \quad (2.82)$$

Similar to Besbes and Zeevi (2015), we make the assumption that the optimal solutions $p\big((\alpha_1,\beta_1),(\alpha_2,\beta_2)\big)$ and $y\big(e^{\alpha_1-\beta_1 p},(\alpha_1,\beta_1),(\alpha_2,\beta_2)\big)$ are differentiable with respect to α_1,α_2 and β_1,β_2 with bounded first order derivatives. Then, $p\big((\alpha_1,\beta_1)(\alpha_2,\beta_2)\big)$ and $y\big(e^{\alpha_1-\beta_1 p},(\alpha_1,\beta_1),(\alpha_2,\beta_2)\big)$ are both Lipschitz and in particular, there exists a constant $K_{57}>0$ such that for any $\alpha_1,\alpha_2,\alpha_1',\alpha_2'$ and $\beta_1,\beta_2,\beta_1',\beta_2'$, it holds that

$$\Big|p\big((\alpha_1,\beta_1)(\alpha_2,\beta_2)\big)-p\big((\alpha_1',\beta_1')(\alpha_2',\beta_2')\big)\Big| \le K_{57}\Big(|\alpha_1-\alpha_1'|+|\beta_1-\beta_1'|+|\alpha_2-\alpha_2'|+|\beta_2-\beta_2'|\Big), \quad (2.83)$$

$$\Big|y\big(e^{\alpha_1-\beta_1 p},(\alpha_1,\beta_1),(\alpha_2,\beta_2)\big)-y\big(e^{\alpha_1'-\beta_1' p},(\alpha_1',\beta_1'),(\alpha_2',\beta_2')\big)\Big| \le K_{57}\Big(|\alpha_1-\alpha_1'|+|\beta_1-\beta_1'|+|\alpha_2-\alpha_2'|+|\beta_2-\beta_2'|\Big). \quad (2.84)$$

The optimization problem (2.80) will serve as yet another bridging problem between DD and B2. To see that, observe that when $\alpha_1=\alpha_2=\hat{\alpha}_{i+1}$ and $\beta_1=\beta_2=\hat{\beta}_{i+1}$, problem (2.81) is reduced to the data-driven problem DD. That is,

$$\hat{p}_{i+1}=p\big((\hat{\alpha}_{i+1},\hat{\beta}_{i+1}),(\hat{\alpha}_{i+1},\hat{\beta}_{i+1})\big). \quad (2.85)$$

On the other hand, when $\alpha_1=\breve{\alpha}(\hat{p}_i),\beta_1=\breve{\beta}(\hat{p}_i),\alpha_2=\breve{\alpha}(\hat{p}_i+\delta_i),\beta_2=\breve{\beta}(\hat{p}_i+\delta_i)$, we deduce from the definition of $\breve{\alpha}(\cdot)$ and $\breve{\beta}(\cdot)$ that for $t=t_i+1,\ldots,t_i+I_i$, we have

$$\zeta_t=D_t-(\alpha_1-\beta_1 p_t)=\lambda(\hat{p}_i)+\epsilon_t-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i)=\epsilon_t, \quad (2.86)$$

and for $t = t_i + I_i + 1, \ldots, t_i + 2I_i$, it holds that

$$\zeta_t = D_t - (\alpha_2 - \beta_2 p_t) = \lambda(\hat{p}_i + \delta_i) + \epsilon_t - (\breve{\alpha}(\hat{p}_i + \delta_i) - \breve{\beta}(\hat{p}_i + \delta_i)(\hat{p}_i + \delta_i)) = \epsilon_t. \quad (2.87)$$

This shows that when the parameters are $\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\big(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i)\big)$, problem (2.81) is reduced to bridging problem B2. This gives us

$$\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) = p\big(\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big), \big(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i)\big)\big). \quad (2.88)$$

The two results (2.85) and (2.88) will enable us to compare the optimal solutions of the data-driven optimization problem DD and bridging problem B2 through one optimization problem (2.81).

In Lemma A6, letting $\xi = (2I_i)^{-\frac{1}{2}}(\log 2I_i)^{\frac{1}{2}}$ in (2.78), we obtain

$$\begin{aligned} &\mathbb{P}\Big\{ |\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + |\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}|^2 \quad (2.89) \\ &\qquad\qquad \le K_{53}\left(I_i^{-\frac{1}{2}} + (2I_i)^{-\frac{1}{2}}(\log 2I_i) + (2I_i)^{-1}(\log 2I_i)\right)\Big\} \\ \ge\ & 1 - \frac{8}{I_i}. \end{aligned}$$

This implies

$$\begin{aligned} &\mathbb{P}\Big\{ |\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}, \\ &\quad |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}, \\ &\quad |\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}, \\ &\quad |\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\Big\} \\ \ge\ & 1 - \frac{8}{I_i}. \end{aligned} \quad (2.90)$$

For convenience, we define the event $\mathcal{A}_5$ by

$$\begin{aligned}\mathcal{A}_5 = \Big\{\omega : &|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\\ &|\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\\ &|\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\\ &|\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}| \le (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\Big\}. \end{aligned} \tag{2.91}$$

Then by (2.91) one has

$$\mathbb{P}(\mathcal{A}_5^c) \le \frac{8}{I_i}. \tag{2.92}$$

When $\beta_1 > 0$, similar to Remark 2 and Lemma A2, one can verify that $H_{i+1}(\cdot,\cdot,\cdot,\cdot)$ of (2.80) is unimodal in p thus its optimal solution is well-defined. Define

$$i^* = \max\Big\{\log_v \frac{e}{2I_0},\ \min\Big\{i \ \Big|\ (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}} < \min_{p\in\mathcal{P}} \breve{\beta}(p)\Big\}\Big\}, \tag{2.93}$$

where we need i^* to be no less than $\log_v \frac{e}{2I_0}$ to ensure that $(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}$ is decreasing on $i \ge i^*$. When $i \ge i^*$, it follows that $\hat{\beta}_{i+1} > 0$ on $\mathcal{A}_5$, hence on event $\mathcal{A}_5$, problem DD is unimodal in p after minimizing over y, and the optimal pricing is well-defined. These properties will enable us to prove that the convergence of parameters translates to convergence of the optimal solutions. Then the first part in Lemma A7 on p follows directly from (2.85), (2.88) and (2.83). From equations (2.82), (2.86), and (2.87), we conclude

$$\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big) = y\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}, \big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big), \big(\breve{\alpha}(\hat{p}_i+\delta_i), \breve{\beta}(\hat{p}_i+\delta_i)\big)\big),$$

and it follows from the DDA policy that

$$\hat{y}_{i+1,1} = y\big(e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}\hat{p}_{i+1}}, (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})\big).$$

Then, similar analysis as that in the proof of (2.83) can be used to prove (2.84). □

To prepare for the convergence proof of order-up-to levels in Theorem 1, we need another result. Recall that $\overline{y}\big(e^{\alpha-\beta p}\big)$ and $\tilde{y}_{i+1}\big(e^{\alpha-\beta p}\big)$ are the optimal y on $\mathcal{Y}$ for problem B1 and problem B2 respectively for given $p \in \mathcal{P}$. We have the following result.

Lemma A8. *T*here exists some constant K_{58} such that, for any $p \in \mathcal{P}$ and $\hat{p}_i \in \mathcal{P}$, for any $\xi > 0$, it holds that

$$\mathbb{P}\Big\{ \Big|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big) - \tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)\Big| \geq K_{58}\xi \Big\} \leq 2e^{-4I_i\xi^2}.$$

Proof. For $p \in \mathcal{P}$, the optimal solution for bridging problem B1 is the same as (2.46), $\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$. Thus

$$\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p} F^{-1}\left(\frac{b}{b+h}\right). \tag{2.94}$$

For given $p \in \mathcal{P}$, we follow (2.49) to define $\tilde{y}^u_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$ as the unconstrained optimal order-up-to level for problem B2 on $\mathbb{R}_+$, then it can be verified that

$$\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\ \min\left\{e^{\epsilon_j}:\ \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t} \leq e^{\epsilon_j}\right\} \geq \frac{b}{b+h}\right\}, \tag{2.95}$$

and, similar to (2.50), we have

$$\tilde{y}_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = \min\left\{\max\left\{\tilde{y}^u_{i+1}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}), y^l\right\} y^h\right\}.$$

It is seen that

$$\left|\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)-\tilde{y}_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)\right|\leq\left|\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)-\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)\right|. \quad (2.96)$$

Now, for any $z>0$, we have

$$\begin{aligned}
& \mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-\frac{b}{b+h}\leq -z\right\} \qquad (2.97)\\
= \;& \mathbb{P}\left\{\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}\\
\leq \;& \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}\geq\frac{b}{b+h}\right\}\\
= \;& \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}-\left(\frac{b}{b+h}-z\right)\geq z\right\},
\end{aligned}$$

where the first inequality follows from (2.95). Since $\mathbb{E}\left[\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}\right]=\frac{b}{b+h}-z$, we apply Hoeffding inequality to obtain

$$\mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}-\left(\frac{b}{b+h}-z\right)\geq z\right\}\leq e^{-4I_iz^2}.$$

Combining this with (2.94) and (2.97), we obtain

$$\mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)\leq -z\right\}\leq e^{-4I_iz^2}. \quad (2.98)$$

Similarly, we have

$$\mathbb{P}\left\{F\left(\tilde{y}_{i+1}^{u}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)\geq z\right\} \leq e^{-4I_i z^2}. \tag{2.99}$$

From regularity condition (v), the probability density function $f(\cdot)$ of e^{ϵ_t} satisfies $r=\min\{f(x), x\in[l,u]\}>0$. From calculus, it is known that, for any $x<y$, there exists a number $z\in[x,y]$ such that $F(y)-F(x)=f(z)(y-x)\geq r(y-x)$. Applying (2.98) and (2.99), for any $\xi>0$, we obtain

$$\begin{aligned}
& 2e^{-4I_i\xi^2} \\
\geq\ & \mathbb{P}\Bigg\{\Bigg|F\left(\tilde{y}_{i+1}^{u}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right) \\
& \qquad -F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)\Bigg|\geq\xi\Bigg\} \\
\geq\ & \mathbb{P}\left\{r\left|\tilde{y}_{i+1}^{u}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}-\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right|\geq\xi\right\} \\
=\ & \mathbb{P}\left\{\left|\tilde{y}_{i+1}^{u}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)-\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)\right|\geq\frac{1}{r}e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\xi\right\}.
\end{aligned}$$

Let $K_{58}=\max_{\hat{p}_i\in\mathcal{P},\hat{p}_{i+1}\in\mathcal{P}}\frac{1}{r}e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}$, then $K_{58}>0$. We have

$$\mathbb{P}\left\{\left|\tilde{y}_{i+1}^{u}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)-\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)\right|\geq K_{58}\xi\right\}\leq 2e^{-4I_i\xi^2},$$

and Lemma A8 follows from the inequality above and (2.96). □

CHAPTER III

Nonparametric Algorithms for Joint Pricing and Inventory Control with Lost-Sales and Censored Demand

3.1 Introduction

Different from Chapter 2 with backlogged demand, in this chapter, we consider lost-sales with censored demand.

3.1.1 Model Overview, Example and Research Issues

This paper studies a periodic-review joint pricing and inventory control problem with lost-sales over a finite horizon of T periods. At the beginning of each period, the firm makes pricing and inventory replenishment decisions. The demands across periods $t = 1, \dots T$ are denoted by $D_t(p_t) = \lambda(p_t) + \epsilon_t$, where $\lambda(p_t)$ is a decreasing deterministic function representing the average customer response rate to the selling price in period t, p_t, and ϵ_t, $t = 1, 2, \dots, T$, are independent and identically distributed (i.i.d.) random variables representing noises (see §3.2 for model details). For notational convenience, we use ϵ_t and ϵ interchangeably in this chapter, due to the i.i.d. assumption. Different from the related literature, the firm knows neither the form of $\lambda(\cdot)$ nor the distribution of ϵ_t *a priori*. Moreover, the firm only observes the

censored demand realizations over time, i.e., it observes the sales quantity in each period, which is the minimum of the realized demand and the on-hand inventory, and thus the lost-sales information is censored and not observed.

Even with complete information about the function $\lambda(\cdot)$ and the distribution of ϵ_t, this class of problems is known to be hard since the expected profit function, in general, fails to be jointly concave. Several papers in the literature present sufficient conditions under which the expected single-period profit function is unimodal (see, e.g., *Chen et al.* (2006), *Huh and Janakiraman* (2008), *Song et al.* (2009), *Wei* (2012), *Chen et al.* (2014b)). We also assume the expected single-period profit function is unimodal if the function $\lambda(\cdot)$ and the distribution of ϵ_t were known *a priori*. However, if one constructs an estimated profit function using past demand observations, the estimated (sampled) profit function, as we demonstrate in Example III.1 below, will be multimodal in price, which presents a major analytical barrier for learning and optimization (see Figure 3.2). In addition, the censored demand information adds further complexity to the problem because the estimators constructed from the observable demand data are also biased. As a result, the firm needs to actively explore the decision space in a cost-efficient manner so as to minimize the estimation errors while maximizing its profit on the fly.

We develop a nonparametric data-driven closed-loop control policy $\pi = (p_t, y_t \mid t \geq 1)$ where p_t and y_t are the pricing decision and the order-up-to level in period t, respectively; and we denote its total expected profit by $\mathscr{C}(\pi)$. Now, had the firm known the underlying demand-price function $\lambda(\cdot)$ and the distribution of ϵ_t *a priori*, there exists a *clairvoyant* optimal policy π^* with total expected profit denoted by $\mathscr{C}(\pi^*)$. We measure the performance of our proposed policy π through an average (per-period) regret $\mathcal{R}(\pi, T) \triangleq (\mathscr{C}(\pi^*) - \mathscr{C}(\pi))/T$. The main research question is to devise an effective nonparametric data-driven policy π that converges to π^* in probability and also drives the average regret $\mathcal{R}(\pi, T)$ to zero with a provable convergence

rate.

Example III.1. Let $\lambda(p) = 2.944 - 0.52p$, and the random error ϵ follows the truncated Normal distribution with mean 0 and standard deviation 0.5 on $[-1, 5]$. Set the price range $\mathcal{P} = [0, 3.6]$ and the inventory range $\mathcal{Y} = [0, 10]$. The clairvoyant's problem (Opt-CV) (i.e., with known $\lambda(p)$ and distribution of ϵ *a priori*) is given by

$$\max_{p\in[0,3.6]} \Big\{ p(2.944 - 0.52p + \mathbb{E}[\epsilon]) \\ - \min_{y\in[0,10]} \left\{ (b+p)\mathbb{E}\left[2.944 - 0.52p + \epsilon - y\right]^+ + h\mathbb{E}\left[y - (2.944 - 0.52p + \epsilon)\right]^+ \right\} \Big\}, \tag{3.1}$$

where $h = 1$ and $b = 1.1$ are the per-unit holding and lost-sales penalty costs, respectively.

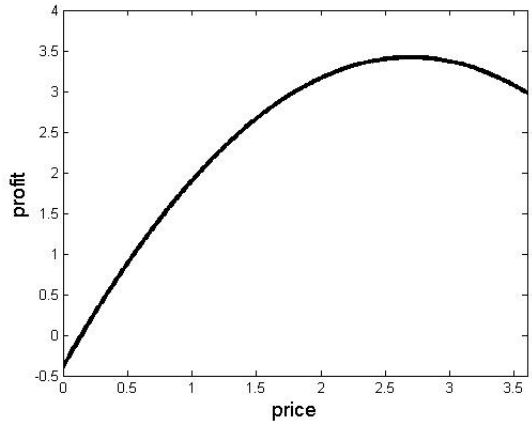

Figure 3.1: The clairvoyant's problem (Opt-CV)

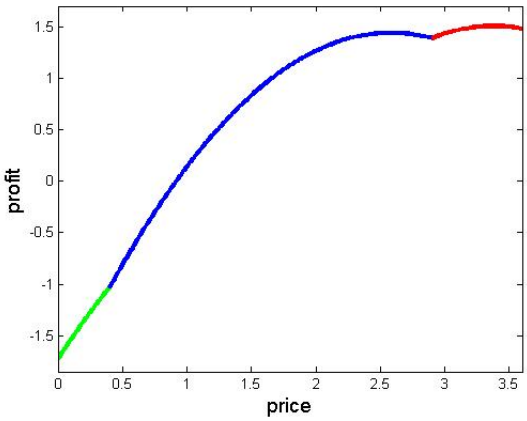

Figure 3.2: The sampled problem (Opt-SAA)

The objective profit function in (3.1) over $\mathcal{P}$ is plotted in Figure 3.1. It is clear that the clairvoyant's problem is unimodal (in fact concave) in p and the optimal price is around 2.7. However, if the demand information is not known *a priori*, then the estimated objective profit function constructed using demand samples can quickly become ill-structured (multimodal).

To illustrate, we for the moment hypothetically assume that the firm knows $\lambda(p)$, but does not know the distribution of ϵ and instead has observed 5 unbiased samples of ϵ as $\{-1, -1, -1, 0.85, 5\}$. (In the case that the firm does not know $\lambda(p)$, the samples of ϵ collected are usually biased in our model.) Using sample average, one can readily construct an estimated objective profit function as

$$\max_{p\in[0,3.6]}\left\{p\left(2.944-0.52p+\frac{1}{5}\sum_{j=1}^{5}\epsilon_j\right)-\min_{y\in[0,10]}\frac{1}{5}\left\{(b+p)\sum_{j=1}^{5}(\ 2.944-0.52p+\epsilon_j-y)^+ \right.\right.$$
$$\left.\left.+h\sum_{j=1}^{5}(y-(2.944-0.52p+\epsilon_j))^+\right\}\right\}. \quad (3.2)$$

Unfortunately, even for this simpler setting, as seen from Figure 3.2, the (sampled) objective profit function (3.2) is multimodal in p. More precisely, it is a piece-wise concave function with three pieces illustrated in different colors. There are two local maxima, i.e., 2.6 and 3.4, and the natural question arises as to how to choose from the multiple local maxima so that the convergence to the clairvoyant's maximum can be guaranteed. It can be seen that the number of local maxima increases in the number of sample points used to construct the sampled objective function. This poses significant challenges in both the algorithmic design and its performance analysis. □

We note again that the example above is in fact a simplified version of our problem with known $\lambda(\cdot)$ (for illustration purposes). The full-fledged problem needs to estimate $\lambda(\cdot)$, and therefore the unbiased samples of ϵ cannot be obtained (as they cannot be separated from the estimation of $\lambda(\cdot)$), making the problem considerably more difficult to analyze.

3.1.2 Main Results and Contributions

We propose the first nonparametric algorithm, called the *Data-Driven algorithm for Censored demand* (DDC for short), for the joint pricing and inventory control

problems with lost-sales and censored demand information. We show the convergence of pricing and inventory replenishment decisions to the clairvoyant optimal decisions in probability (Theorem III.3), and also characterize its rate of convergence (Theorem III.4). More specifically, we show that the average regret $\mathcal{R}(DDC, T)$ converges to zero at the rate of $\mathcal{O}(T^{-\frac{1}{5}}(\log T)^{\frac{1}{4}})$. We also conduct numerical experiments to demonstrate the effectiveness of the proposed algorithm.

Our proposed algorithm DDC builds upon the recent work of *Besbes and Zeevi* (2015) (which focused on a dynamic pricing problem without inventory replenishment decisions) and *Chen et al.* (2015) (which studied a joint pricing and inventory control problem with backlogging and full demand observation). In particular, DDC uses a linear approximation scheme to estimate the average demand function $\lambda(\cdot)$ using the least-square method. This elegant idea was originally put forth by *Besbes and Zeevi* (2015). One critical difference between our work and theirs is that with inventory replenishment as an operational decision, we also need to learn the underlying distribution of the random error ϵ_t using historical demand data in order to set the order-up-to levels in each period. Since $\lambda(\cdot)$ is not known *a priori* and is subject to estimation errors, true samples of the random error ϵ_t cannot be obtained. Furthermore, due to the lost-sales, the firm cannot observe the true realized cost whenever a stockout occurs in a period, as the lost-sales penalty cost depends on the lost-sales quantity which is not observed by the firm. Thus, in our problem the firm does not always observe the realized profit in a period, and due to lack of knowledge about $\lambda(p)$, nor can the firm assess the derivatives of the realized profit function. As a result, conventional approaches, such as stochastic approximation, online convex optimization, and continuum-armed bandit algorithms, cannot be applied or adapted to this setting, as these methods rely heavily on knowing either the realized objective value or its derivatives for a given decision.

Recently, *Chen et al.* (2015) studied a joint pricing and inventory control problem

with backlogging, and proposed a method for constructing a (sampled) proxy profit function. Our work is closely related to theirs, and also involves constructing a sampled proxy profit function using historical demand data, but is significantly different from that work in several aspects. It is well-known in the literature that the joint pricing and inventory control with lost-sales is much harder to analyze than its backorder counterpart, since the lost-sales problem is structurally much more complex even with known $\lambda(\cdot)$ and ϵ_t. This difficulty is further aggravated by censored demand information (i.e., the demand observations are truncated by the on-hand inventory levels). As noted in Example III.1, the sampled profit function for the lost-sales model is ill-behaved, which is a major difficulty not encountered in the backorder model. Therefore, the algorithmic design and analysis become highly nontrivial, requiring a multitude of new ideas and techniques.

In the following, we detail the three major challenges of our problem that did not exist in previous related works, and our high-level solution approaches.

(a) Active exploration of the inventory space. In the lost-sales model, the demand realizations are often truncated by the on-hand inventory levels (i.e., the lost-sales quantity is unobservable when customers find their desired items out of stock and walk away). This means that the firm does not know the lost-sales cost incurred during a period when a stockout occurs (as it does not know how many customers are lost). The censored data also create much difficulty in the algorithmic design since this unobservable lost-sales quantity is essential in estimating $\lambda(\cdot)$ and ϵ_t. To learn about the demand and maximize the profit, active exploration is needed to discover the lost-sales quantity that is otherwise unknown. Our algorithm DDC carries out active experimentation on the inventory space in carefully designed cycles. The algorithm raises the inventory level whenever there is a stockout (see Step 1 of DDC). The next immediate issue is how to use the observable sales data (censored demand realizations) to estimate $\lambda(\cdot)$ and the distribution of the error term. Using

the sales data clearly introduces downward biases in estimating the true demand, but we show that their long run impacts are negligible under our exploration strategy when the length of learning cycle increases (Lemmas III.7 and III.11).

(b) Correction of estimation bias. There are two sources of estimation biases we need to overcome in the performance analysis of DDC. First, since the demand-price function $\lambda(\cdot)$ is not known, we cannot obtain true samples of the random error ϵ_t based on demand observations. We instead use the *residual error* defined in (3.10), computed based on the linear estimate of $\lambda(\cdot)$, to approximate the true random error. Note that knowing the distribution of error term is crucial for making inventory decisions that strike a good balance between overage and underage costs. We show that this estimation bias vanishes as the algorithm proceeds (Lemmas III.8 and III.12). Second, we use sales data (censored demand) instead of full demand realizations to carry out our least-square estimation and sampled optimization (see Steps 2 and 3 of DDC), and this clearly introduces estimation biases that need to be overcome (Lemmas III.7 and III.11).

(c) Learning and optimizing a multimodal function. The most difficult (and also the unqiue) part of this lost-sales problem lies in the fact that the estimated (sampled) proxy profit function (Opt-SAA) using demand observations in the exploration phase is multimodal in p in Step 3 of DDC (see, e.g., Figure 3.2), even though the expected profit function is assumed to be unimodal in p (e.g., Figure 3.1). In contrast, the original objective and its (sampled) proxy function in the back-order model studied by *Chen et al.* (2015) are both unimodal in p, and therefore the optimal prices can be solved through a first-order condition, establishing that the convergence in parameters guarantees the convergence of decisions. Learning and optimizing a multimodal function is indeed a challenging task, which is a unique characteristic in the lost-sales setting. Moreover, the number of local maxima grows in the number of demand data points used. To resolve this issue, we develop a new

technique called *sparse discretization* to overcome the technical hurdle (Proposition III.6). More specifically, we optimize the multimodal (sampled) proxy profit function (Opt-SAA) on a sparse discretized set of prices. For any time horizon T, we only need to exhaustively check on the order of $T^{\frac{1}{5}}$ number of price points (which is very sparse). We show the (sampled) proxy profit function (Opt-SAA) is *uniformly close* to the linear approximated function (Approx-CV) over this sparse discretization (see Figure 3.5). We then establish the convergence result by exploiting some structural properties of the linear approximated function (Approx-CV). We believe the sparse discretization technique developed in this chapter can be useful in learning and optimizing multimodal functions of other settings where the original function has nice structures (e.g., concavity, unimodality) but the sampled proxy function is ill-behaved.

3.1.3 Literature Review

Our work is relevant to the following research streams.

Joint pricing and inventory control problems with lost-sales. The literature on joint pricing and inventory control problems has confined itself mainly to models that assume unmet demand is fully backlogged. The optimality of base-stock list-price or (s, S, p) policies for backorder models has been well established (see, e.g., *Federgruen and Heching* (1999), *Chen and Simchi-Levi* (2004a,b), *Huh and Janakiraman* (2008)). Compared with the classical backorder model, the difficulty in analyzing the lost-sales model is mainly due to the fact that the expected profit function fails to be jointly concave even when demand is linear in price p (see *Federgruen and Heching* (1999)), which is often a crucial property for characterizing optimal policies. Nevertheless, there is a stream of literature that extends the optimality of base-stock list-price or (s, S, p) policies to the lost-sales model with additive or multiplicative

demand (see, e.g., *Chen et al.* (2006), *Huh and Janakiraman* (2008), *Song et al.* (2009), *Chen et al.* (2014b)). These papers require that the expected single-period profit function to be unimodal (or quasiconcave) in p under certain technical conditions. Our work differs from the above literature by not taking the demand-price relationship as given. The firm needs to use observed demand data to learn the demand process on the fly while maximizing their expected profit. However, when the demand-price relationship is not known *a priori*, our (sampled) proxy profit functions, constructed using SAA methods, no longer preserve the unimodality property, which poses a significant challenge in the performance analysis.

Nonparametric algorithm for inventory models. *Huh and Rusmevichientong* (2009) proposed gradient descent based algorithm for lost-sales systems with censored demand. Subsequently, *Huh et al.* (2009) proposed algorithm for finding the optimal base-stock policy in lost-sales inventory systems with positive lead time. *Besbes and Muharremoglu* (2013) examined the discrete demand case and showed that active exploration is needed. *Huh et al.* (2011) applied the concept of Kaplan-Meier estimator to devise another data-driven algorithm for censored demand. *Shi et al.* (2015) proposed algorithm for multi-product inventory systems under a warehouse-capacity constraint with censored demand. Another nonparametric approach in the inventory literature is sample average approximation (SAA) (e.g., *Kleywegt et al.* (2002), *Levi et al.* (2007, 2011)) which uses the empirical distribution formed by *uncensored* samples drawn from the true distribution. Concave adaptive value estimation (e.g., *Godfrey and Powell* (2001), *Powell et al.* (2004)) successively approximates the objective cost function with a sequence of piecewise linear functions. The bootstrap method (e.g., *Bookbinder and Lordahl* (1989)) estimates the newsvendor quantile of the demand distribution. The infinitesimal perturbation approach (IPA) is a sampling-based stochastic gradient estimation technique that has been used to solve

stochastic supply chain models (see, e.g., *Glasserman* (1991)). *Eren and Maglaras* (2014) employed maximum entropy distributions to solve a stochastic capacity control problem. For parametric approaches in stochastic inventory systems, see, e.g., *Lariviere and Porteus* (1999) and *Chen and Plambeck* (2008) on Bayesian learning, and *Liyanage and Shanthikumar* (2005) and *Chu et al.* (2008) on operational statistics.

Nonparametric algorithm for dynamic pricing models. There is a growing literature on dynamic pricing problems with a demand learning approach (see, e.g, survey papers by *Aviv and Vulcano* (2012) and *den Boer* (2015)). The majority of the papers have adopted parametric models in which the firm knows the functional form of the underlying demand-price function (e.g., linear, logit, exponential). Popular approaches in this setting include Bayesian method (see, e.g., *Araman and Caldentey* (2009), *Farias and van Roy* (2010), *Harrison et al.* (2012)), Maximum Likelihood Estimation (see, e.g., *Broder and Rusmevichientong* (2012), *den Boer* (2014), *den Boer and Zwart* (2014, 2015)), Least Square method (see, e.g., *Bertsimas and Perakis* (2006), *Keskin and Zeevi* (2014)) and Thompson Sampling method (see, e.g., *Johnson et al.* (2015)). In contrast, there are only a few papers on nonparametric models. *Besbes and Zeevi* (2009, 2012) proposed simple "blind" policies to single-product and network revenue management models. *Wang et al.* (2014) and *Lei et al.* (2014) proposed generalized bisection search methods to produce a sequence of pricing intervals that converge to the optimal static price with a high probability and also obtained their convergence rates. On the methodological side, *Broadie et al.* (2011) derived general upper bounds on the mean-squared error for the Kiefer-Wolfowitz (KW) stochastic approximation algorithm. Closer to our work, *Besbes and Zeevi* (2015) used a linear approximation scheme to estimate the demand-price function, which gives (surprising) near-optimal performance.

Nonparametric algorithm for joint pricing and inventory control models. To the best of our knowledge, there have been only two papers that proposed nonparametric learning algorithms for the joint pricing and inventory control problem. *Burnetas and Smith* (2000) first developed a gradient descent type algorithm for ordering and pricing when inventory is perishable (i.e., without inventory carryover); they showed that the average profit converges to the optimal one but did not establish the rate of convergence. We also note that *Burnetas and Smith* (2000) did not even consider lost-sales penalty costs so they did not have the issue of not being able to observe the realized profit value. Recently, *Chen et al.* (2015) proposed a nonparametric data-driven algorithm for the joint pricing and inventory control problem with backorders. Our work contributes to the literature by considering a counterpart model with lost-sales and censored demand information, which is substantially harder to analyze. This is the first attempt in the literature to the best of our knowledge.

3.1.4 Organization and General Notation

The rest of this chapter is organized as follows. In §3.2, we formally describe our joint pricing and inventory control problem with lost-sales, and also characterize the clairvoyant optimal policy had the demand-price relationship known *a priori*. In §3.3, we propose a nonparametric data-driven algorithm called DDC under censored demand information. In §3.4, we state our main results and provide our proof strategies. The detailed proofs are deferred to the Appendix. In §3.5, we extend our model and results to the observable demand case and also the unbounded demand case.

Throughout this chapter, for any real numbers x and y, we denote $x^+ = \max\{x, 0\}$, $x \vee y = \max\{x, y\}$, and $x \wedge y = \min\{x, y\}$. We also use the notation $\lfloor x \rfloor$ and $\lceil x \rceil$ frequently, where $\lfloor x \rfloor$ is defined as the largest integer value which is smaller than or equal to x; and $\lceil x \rceil$ is the smallest integer value which is greater than or equal to x. The notation $\triangleq$ means "is defined as". We use LHS and RHS to denote "left hand

side" and "right hand side", respectively.

3.2 Joint Pricing and Inventory Control with Lost-Sales and Censored Demand

3.2.1 Problem Definition

We consider a periodic-review joint pricing and inventory control problem with lost-sales (see, e.g., *Chen et al.* (2006), *Huh and Janakiraman* (2008), *Song et al.* (2009)). Different from the conventional literature, the firm has no knowledge of the true underlying demand process *a priori*, and can make sequential pricing and inventory decisions only based on the past observed sales data (i.e., censored demand). We formally describe our problem below.

Demand process. For each period $t = 1, \ldots, T$, the demand in period t depends on the selling price p_t in period t and some random noise ϵ_t, and it is stochastically decreasing in p_t. The well-studied demand models in the literature are the additive demand model $D_t(p_t) = \lambda(p_t) + \epsilon_t$ and the multiplicative demand model $D_t(p_t) = \lambda(p_t)\epsilon_t$, where $\lambda(\cdot)$ is a non-increasing deterministic function and ϵ_t, $t = 1, 2, \ldots, T$, are independent and identically distributed (i.i.d.) random variables. We assume that ϵ_t is defined on a finite support $[l, u]$, but will later extend it to the case of unbounded support in §3.5. We denote the CDF of ϵ_t by $F(\cdot)$. For notational convenience, we use ϵ_t and ϵ interchangeably in this chapter, due to the i.i.d. assumption.

In this chapter we focus our attention on the additive demand model, and assume without loss of generality that $\mathbb{E}[\epsilon_t] = 0$. (We remark that the analysis and results for the multiplicative demand model are analogous.) The firm knows neither the function $\lambda(p_t)$ nor the distribution of the random term ϵ_t *a priori*, and thus it has to learn such demand-price information from the censored demand data collected over time while maximizing its profit on the fly. For convenience, we shall refer to $\lambda(\cdot)$ as

the demand-price function and ϵ_t as the random error.

System dynamics and objectives. We let x_t and y_t denote the inventory levels at the beginning of period t before and after an inventory replenishment decision, respectively. We assume that the system is initially empty, i.e., $x_1 = 0$. An admissible or feasible policy is represented by a sequence of prices and order-up-to levels, $\{(p_t, y_t), t \geq 1\}$ with $y_t \geq x_t$, where (p_t, y_t) depends only on the demand and decisions made prior to time t, i.e, (p_t, y_t) is adapted to the filtration generated by $\{(p_s, y_s), \min\{D_s(p_s), y_s\} : s = 1, \ldots, t-1\}$ under censored demand. We assume $y_t \in \mathcal{Y} = [y^l, y^h]$ and $p_t \in \mathcal{P} = [p^l, p^h]$ with known bounded support, and $\lambda(p^h) > 0$ and $y^h \geq \lambda(p^l) + u$.

Given any admissible policy π, we describe the sequence of events for each period t. (Note that x_t^π, y_t^π, p_t^π's are functions of π; for ease of presentation, we will make their dependence on π implicit.)

(a) At the beginning of period t, the firm observes the starting inventory level x_t.

(b) The firm decides to order a non-negative amount of inventory to bring the inventory level up to $y_t \in \mathcal{Y}$, and also sets the selling price $p_t \in \mathcal{P}$. We assume instantaneous replenishment.

(c) The demand $D_t(p_t)$ in period t realizes to be $d_t(p_t)$, and is satisfied to the maximum extent using on-hand inventory. Unsatisfied demand is *lost* and *unobservable*. In other words, the firm only observes the sales quantity $\min\{d_t(p_t), y_t\}$, instead of the full realized demand $d_t(p_t)$. The state transition is $x_{t+1} = (y_t - d_t(p_t))^+$.

(d) At the end of period t, the firm incurs a profit of

$$\begin{aligned} & p_t \min\{d_t(p_t), y_t\} - b(d_t(p_t) - y_t)^+ - h(y_t - d_t(p_t))^+ \qquad (3.3) \\ = \; & p_t d_t(p_t) - (b + p_t)\left(d_t(p_t) - y_t\right)^+ - h\left(y_t - d_t(p_t)\right)^+, \end{aligned}$$

where h and b are the per-unit holding and lost-sales penalty costs, respectively. We assume without loss of generality that the per-unit purchasing cost is zero (see *Zipkin* (2000)).

It is important to note that (3.3) is the *perceived profit* obtained by the firm; the firm cannot observe its true realized value if a stockout occurs. This is because, the lost-sales cost depends on the actual lost demand $(d_t(p_t) - y_t)^+$ which is not observed. However, this term does represent a true damage to the firm and it is part of the objective function that the firm wishes to optimize.

If the underlying demand-price function $\lambda(p)$ and the distribution of the error term ϵ_t were known and the firm could observe the lost demand, then the problem specified above could be formulated as an optimal control problem with state variables x_t, control variables (p_t, y_t), random disturbances ϵ_t, and the total profit given by

$$\max_{\substack{(p_t, y_t) \in \mathcal{P} \times \mathcal{Y} \\ y_t \geq x_t}} \sum_{t=1}^{T} \Big(p_t \mathbb{E}[D_t(p_t)] - (b + p_t)\mathbb{E}[D_t(p_t) - y_t]^+ - h\mathbb{E}[y_t - D_t(p_t)]^+ \Big). \quad (3.4)$$

However, in our setting, the firm does not know the demand information *a priori* and cannot observe the lost-sales quantity. Hence the firm is unable to evaluate the objective function of this optimization problem. The firm has to learn from historical sales data, revealing information about market responses to offered prices, and uses the learned information to estimate the objective profit function as a basis for optimization.

3.2.2 Clairvoyant Optimal Policy and Main Assumptions

We first characterize the *clairvoyant* optimal policy (as the benchmark of performance), had the firm known the demand-price function $\lambda(p)$ and the distribution of ϵ *a priori*. *Sobel* (1981) has shown that a myopic policy is optimal. Define the

single-period revenue function by

$$Q(p,y) \triangleq p\mathbb{E}[D_1(p)] - (b+p)\mathbb{E}[D_1(p)-y]^+ - h\mathbb{E}[y-D_1(p)]^+. \tag{3.5}$$

To find the optimal pricing and inventory decisions, it suffices to maximize the single-period revenue $Q(p,y)$, which is equivalent to solving

$$\begin{aligned}
&\max_{p\in\mathcal{P},y\in\mathcal{Y}} \left\{p\mathbb{E}[D_1(p)] - (b+p)\mathbb{E}[D_1(p)-y]^+ - h\mathbb{E}[y-D_1(p)]^+\right\} \\
=&\max_{p\in\mathcal{P},y\in\mathcal{Y}} \left\{p\lambda(p) - (b+p)\mathbb{E}\left[\lambda(p)+\epsilon-y\right]^+ - h\mathbb{E}\left[y-\lambda(p)-\epsilon\right]^+\right\} \\
=&\max_{p\in\mathcal{P}} \left\{p\lambda(p) - \min_{y\in\mathcal{Y}} \left\{(b+p)\mathbb{E}\left[\lambda(p)+\epsilon-y\right]^+ + h\mathbb{E}\left[y-\lambda(p)-\epsilon\right]^+\right\}\right\}.
\end{aligned}$$

Hence we write the clairvoyant optimization problem (Opt-CV) compactly as

$$\max_{p\in\mathcal{P},y\in\mathcal{Y}} Q(p,y) = \max_{p\in\mathcal{P}} \left\{\bar{G}(p,\lambda(p))\right\}, \tag{Opt-CV}$$

$$\text{where} \quad \bar{G}(p,\lambda(p)) \triangleq p\lambda(p) - \min_{y\in\mathcal{Y}} \left\{(b+p)\mathbb{E}\left[\lambda(p)+\epsilon-y\right]^+ + h\mathbb{E}\left[y-\lambda(p)-\epsilon\right]^+\right\}.$$

Let the optimal solution for (Opt-CV) be (p^*, y^*).

The inner optimization problem $\bar{G}(p,\lambda(p))$ determines the optimal order-up-to level for a given price p, and the outer optimization solves for the optimal price p. Clearly, in the clairvoyant problem, once we start below y^*, we are able to raise the inventory level to y^* at the beginning of all subsequent periods, and thus the expected profit in each period is $Q(p^*,y^*)$. This, however, is not the case when the underlying demand process is not known *a priori*.

Linear approximation. Next we introduce a linear approximation of (Opt-CV) that will be useful in developing our nonparametric algorithm. For given parameters

$\alpha, \beta > 0$, define the optimization problem (Approx-CV) by

$$\max_{p\in\mathcal{P}} \left\{ \bar{G}(p, \alpha - \beta p) \right\}, \tag{Approx-CV}$$

where $\bar{G}(p, \alpha - \beta p)$

$$\triangleq p(\alpha - \beta p) - \min_{y\in\mathcal{Y}} \left\{ (b+p)\mathbb{E}\left[(\alpha - \beta p) + \epsilon - y\right]^+ + h\mathbb{E}\left[y - (\alpha - \beta p) - \epsilon\right]^+ \right\}.$$

Note that (Approx-CV) replaces the demand-price function $\lambda(p)$ in (Opt-CV) by its linear approximation $\alpha - \beta p$, which serves as an intermediate benchmark. Let $\bar{p}(\alpha, \beta)$ be the optimal price for $\bar{G}(p, \alpha - \beta p)$. For any fixed $p \in \mathcal{P}$, let $\bar{y}(p, \alpha - \beta p)$ be the optimal order-up-to level for $\bar{G}(p, \alpha - \beta p)$.

We also conveniently write $y^* = \bar{y}(p^*, \lambda(p^*))$. And, for any $z \in \mathcal{P}$, we introduce notation

$$\breve{\alpha}(z) = \lambda(z) - \lambda'(z)z \quad \text{and} \quad \breve{\beta}(z) = -\lambda'(z). \tag{3.6}$$

Main assumptions. We state the main assumptions for our results in §3.4 to hold.

Assumption III.2. *(i)* $\bar{G}(p, \lambda(p))$ *is unimodal in* $p \in \mathcal{P}$.

(ii) $\bar{G}(p, \breve{\alpha}(z) - \breve{\beta}(z)p)$ *is strictly concave in* $p \in \mathcal{P}$ *for any* $z \in \mathcal{P}$.

(iii) $\max_{z\in\mathcal{P}} \left| \frac{d\bar{p}(\breve{\alpha}(z), \breve{\beta}(z))}{dz} \right| < 1$.

(iv) The random error ϵ *has a bounded support* $[l, u]$ *where* $l \le u < \infty$.

We remark that unimodality (or quasiconcavity) of the expected profit function is a predominant assumption in joint pricing and inventory control problems with lost-sales (see, e.g., *Chen et al.* (2006), *Huh and Janakiraman* (2008), *Song et al.* (2009), *Chen et al.* (2014b)). We provide sufficient conditions for a demand-price function $\lambda(\cdot)$ to satisfy Assumption III.2 in the Appendix. Assumption III.2 admits a large class of demand-price functions (e.g., linear, logarithmic, logit, and exponential). In

fact, Assumption III.2(iv) can be dropped, and we defer the detailed discussion to §3.5.

3.3 Nonparametric Data-Driven Algorithm

Without knowing both $\lambda(p)$ and the underlying distribution of ϵ *a priori*, the firm needs to experiment with prices and target inventory levels in order to collect demand data while maximizing the profit on the fly. We propose a nonparametric algorithm, called the Data-Driven algorithm for Censored demand (DDC for short), that strikes a good balance between learning (exploration) and earning (exploitation). At a high level, DDC estimates $\lambda(p)$ by an affine function, and constructs an empirical error distribution. The difficulty arises from the fact that the empirical distribution may be biased due to inaccurate estimation of $\lambda(p)$ as well as demand censoring. As a result, DDC needs to actively explore the decision space (especially the target inventory levels) whenever a stockout occurs, and also ensures that the sampling biases diminish to zero quickly. More strikingly, the sampled optimization problem (Opt-SAA) constructed using demand data loses unimodality, a key property utilized in *Chen et al.* (2015) to establish the desired convergence for the backorder counterpart model. To overcome this major difficulty, we develop a *sparse discretization* scheme to search for the optimal solution of (Opt-SAA) over a sparse discretized set of price points, and then develop a uniform convergence argument between the sampled profit function and the original profit function over this sparse discretization to establish our convergence results.

3.3.1 Data-Driven Algorithm for Censored Demands (DDC)

The DDC algorithm consists of learning stages with exponentially increasing length. Stage i has I_i periods in total, with the first $2L_i$ periods being the exploration phase, and the remaining $I_i - 2L_i$ periods being the exploitation phase. The algorithm

starts with initial parameters $\{\hat{p}_1, \hat{y}_{1,1}, \hat{y}_{1,2}\}$ where $\hat{p}_1 \in \mathcal{P}, \hat{y}_{1,1} \in \mathcal{Y}, \hat{y}_{1,2} \in \mathcal{Y}$, and four fixed parameters $I_0 > 0$, $v > 0$, $s > 0$ and $\rho > 0$. For each learning stage $i \geq 1$, we set

$$I_i = \lfloor I_0 v^i \rfloor, \quad L_i = \lfloor I_i^{\frac{4}{5}} \rfloor, \quad \delta_i = \rho L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}, \text{ and } t_i = \sum_{k=1}^{i-1} I_k \text{ with } t_1 = 0. \quad (3.7)$$

Then, stage $i > 1$ starts in period $t_i + 1$, and at the beginning of stage i, the algorithm proceeds with $\{\hat{p}_i, \hat{y}_{i,1}, \hat{y}_{i,2}\}$ that are derived in the preceding stage $i - 1$. Define a sparse discretized set of prices for stage i as

$$\mathcal{S}_i = \{p^l, p^l + \delta_i, p^l + 2\delta_i, \ldots, p^h\}, \quad (3.8)$$

which is the discrete search space for our pricing decisions in stage i.

Now we are ready to present the learning algorithm DDC under censored demand.

Step 0: Preparation. Input I_0, v, s, and ρ, and $\hat{p}_1$, $\hat{y}_{1,1}$, $\hat{y}_{1,2}$, δ_1, I_1, L_1.

Then, for each learning stage $i = 1, \ldots, n$ where $n = \left\lceil \log_v \left(\frac{v-1}{I_0 v} T + 1 \right) \right\rceil$, repeat the following steps.

Step 1: Setting prices and target levels for periods $t \in \{t_i + 1, \ldots, t_i + 2L_i\}$ in stage i.

Set prices p_t, $t = t_i + 1, \ldots, t_i + 2L_i$, to

$$p_t = \hat{p}_i, \text{ for all } t = t_i + 1, \ldots, t_i + L_i,$$
$$p_t = \hat{p}_i + \delta_i \text{ for all } t = t_i + L_i + 1, \ldots, t_i + 2L_i;$$

and for $t = t_i + 1, \ldots, t_i + 2L_i$, raise the inventory level of period t to y_t as follows:

(i) for $t = t_i + 1$, set $y_t = \hat{y}_{i,1} \vee x_t$;

(ii) for $t = t_i + 2, \ldots, t_i + L_i$,

$$y_t = \begin{cases} y_{t-1}, & \text{if } y_{t-1} > d_{t-1}; \\ ((1+s)y_{t-1}) \wedge y^h, & \text{otherwise;} \end{cases}$$

(iii) for $t = t_i + L_i + 1$, set $y_t = \hat{y}_{i,2} \vee x_t$;

(iv) for $t = t_i + L_i + 2, \ldots, t_i + 2L_i$,

$$y_t = \begin{cases} y_{t-1}, & \text{if } y_{t-1} > d_{t-1}; \\ ((1+s)y_{t-1}) \wedge y^h, & \text{otherwise.} \end{cases}$$

Step 2: Estimating the demand-price function and the error term.

Since the realized demand data d_t is not available in the event of stockouts, we instead use the sales data $d_t \wedge y_t$ (where $t \in \{t_i + 1, \ldots, t_i + 2L_i\}$), and solve the following least-square problem

$$(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = \text{argmin} \left\{ \sum_{t=t_i+1}^{t_i+2L_i} \Big[d_t \wedge y_t - (\alpha - \beta p_t) \Big]^2 \right\}, \text{ and} \tag{3.9}$$

$$\eta_t = d_t \wedge y_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t), \quad \text{for } t \in \{t_i + 1, \ldots, t_i + 2L_i\}. \tag{3.10}$$

Step 3: Maximize the proxy profit $Q^{SAA}(p, y)$.

We define the following sampled optimization problem (Opt-SAA).

$$\max_{(p,y)\in\mathcal{S}_{i+1}\times\mathcal{Y}} Q_{i+1}^{SAA}(p,y) \triangleq \max_{p\in\mathcal{S}_{i+1}}\left\{\hat{G}_{i+1}(p,\hat{\alpha}_{i+1},\hat{\beta}_{i+1})\right\}, \qquad \text{(Opt-SAA)}$$

$$\begin{aligned}\text{where}\quad \hat{G}_{i+1}(p,\hat{\alpha}_{i+1},\hat{\beta}_{i+1}) &\triangleq p\left(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p\right)\\ &\quad -\min_{y\in\mathcal{Y}}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i}\left\{(b+p)\left(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p+\eta_t-y\right)^+\right.\right.\\ &\qquad \left.\left.+\,h\left(y-(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p+\eta_t)\right)^+\right\}\right\}.\end{aligned}$$

Then, set the first pair of price and inventory level to

$$(\hat{p}_{i+1},\hat{y}_{i+1,1}) = \arg\max_{(p,y)\in\mathcal{S}_{i+1}\times\mathcal{Y}} Q_{i+1}^{SAA}(p,y),$$

and set the second pair to $(\hat{p}_{i+1}+\delta_{i+1},\hat{y}_{i+1,2})$, where

$$\hat{y}_{i+1,2} = \arg\max_{y\in\mathcal{Y}} Q_{i+1}^{SAA}(\hat{p}_{i+1}+\delta_{i+1},y).$$

In case that $\hat{p}_{i+1}+\delta_{i+1}\notin\mathcal{S}_{i+1}$, set the second price to be $\hat{p}_{i+1}-\delta_{i+1}$.

Step 4: Setting prices and target levels for periods $t\in\{t_i+2L_i+1,\ldots,t_i+I_i\}$ in stage i.

For $t=t_i+2L_i+1,\ldots,t_i+I_i$, set the price and target inventory level to

$$p_t=\hat{p}_{i+1}, \qquad y_t = x_t\vee\hat{y}_{i+1,1}.$$

3.3.2 Algorithmic Overview of DDC

The DDC algorithm integrates active learning (exploration) and earning (exploitation) in carefully designed cycles. We divide the planning horizon T into stages indexed by i, $i=1,2,\ldots,n$. The length of stage i is I_i, where I_i is an integer that is

exponentially increasing in i.

Step 1: The proposed algorithm DDC uses the first two L_i intervals of stage i to actively explore the target inventory levels in order to mitigate the negative effect caused by demand censoring. More specifically, during the first L_i periods, DDC sets the price as $\hat{p}_i$ and the order-up-to level as $\hat{y}_{i,1}$ (which are determined by its preceding stage $i-1$). Whenever a stockout occurs, DDC carries out an upward correction by increasing y_t by some fixed percentage $s > 0$ for the subsequent periods. A similar procedure is also carried out during the second L_i periods. The frequency of stockouts (and thus the upward corrections in setting target inventory levels) will decrease as L_i grows. Note that *Chen et al.* (2015) need not actively explore the inventory space in the backorder setting where all demand observations are uncensored.

During the active exploration phase, the target inventory levels may exceed the optimal inventory level. This is very different than *Huh et al.* (2011) and *Besbes and Muharremoglu* (2013) attempting to settle the order-up-to level around the true quantile solution of the newsvendor problem. Note that they consider a much simpler problem without pricing decisions and no inventory carryovers (the so-called "repeated" newsvendor problem), and hence need not learn the demand-price function. In our setting, the unobservable lost-sales data contain important information about the demand-price function, which is critical for making future pricing decisions.

Second, DDC also experiments with prices during the exploration phase. More specifically, DDC uses $\hat{p}_i$ during the first L_i periods where $\hat{p}_i$ is the optimal pricing decision based on the current belief about the demand-price function. Then DDC perturbs $\hat{p}_i$ by δ_i during the second L_i periods, so that the demand data collected at these two (nearby) price points can be used to carry out an affine estimation of the demand-price function using the least-square method in Step 2.

Step 2: The proposed algorithm DDC utilizes the sales information $d_t \wedge y_t$ collected from the $2L_i$ periods (since the true demand data d_t is not available in the

events of stockouts) to estimate the linear approximation of $\lambda(p)$ via the least-square method, and also to compute the *residual error* η_t. This step resembles *Besbes and Zeevi* (2015) in the dynamic pricing setting but in their problem the firm can always observe the complete demand data d_t and they need not estimate the random error ϵ_t, which is critical in our inventory setting. Our problem is more challenging for the following two reasons. First, the firm can only use the observable sales data (truncated demand data) to carry out the least-square estimation. Second, it is crucial to realize that η_t is not a true sample of ϵ_t, because the linear approximation $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p_t \neq \lambda(p_t)$ and thus $\eta_t = d_t \wedge y_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p_t) \neq d_t - \lambda(p_t) = \epsilon_t$. This poses significant challenges in estimating the distribution of the random error ϵ (when setting target inventory levels). In the traditional SAA approach (e.g., *Levi et al.* (2007)), true samples of a random variable are employed to construct its empirical distribution, and results are developed to show how many samples are needed for the empirical distribution to achieve a certain degree of accuracy. However, in our setting, true samples of ϵ_t are never available, because the demand-price function $\lambda(\cdot)$ is unknown, and the lost-sales quantity is censored. Therefore, the conventional SAA techniques cannot be applied to tackle our problem. Alternatively, our strategy at a high-level is to prove that, as i grows, $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p$ approaches the tangent line of $\lambda(p)$ at point $\hat{p}_i$, and the residual error η_t converges to ϵ_t with a high probability.

Step 3: The proposed algorithm DDC computes two new pairs of decisions, $(\hat{p}_{i+1}, \hat{y}_{i+1,1})$ and $(\hat{p}_{i+1}+\delta_{i+1}, \hat{y}_{i+1,2})$ using the sampled optimization problem (Opt-SAA). Note that (Opt-SAA) resembles (Opt-CV) except that (i) $\lambda(p)$ is replaced by a linear estimation $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p$, and (ii) the terms in the objective function involving ϵ_t are replaced by either empirical average or quantile of η_t. It is important to note that both $(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$ and η_t are computed using the sales data (i.e., censored demand realizations), thereby suffering from (downward) estimation bias. To correct such estimation bias, our strategy at a high-level is to show that the frequency of stockouts

drops as i grows, and the downward bias in estimating the terms involving $\lambda(p)$ and ϵ_t diminishes to zero. This challenge did not exist in *Chen et al.* (2015).

The most critical difference from its backorder counterpart model studied in *Chen et al.* (2015) is that the estimated (sampled) proxy profit function in (Opt-SAA) after minimizing over $y \in \mathcal{Y}$ is multimodal in p in the lost-sales setting, which stands as our major technical hurdle in the performance analysis of DDC (see Figure 3.2 for a simpler setting). To this end, we develop a new sparse discretization technique to jointly learn and optimize a multimodal function. More precisely, the optimization of (Opt-SAA) is conducted over a sparse discretized set of prices $\mathcal{S}_{i+1}$, instead of the original continuous set $\mathcal{P}$. The sparsity is on the order of $T^{\frac{1}{5}}$. We then develop a uniform convergence argument between the sampled objective function and the original objective function over $\mathcal{S}_{i+1}$, which is essential for obtaining the convergence in policy and the corresponding convergence rate. Optimizing a multimodal function on a sparse discrete set of price points significantly reduces the computational burden, and also makes the performance analysis tractable.

Step 4: In this exploitation phase, the proposed algorithm DDC implements the first pair of decisions $(\hat{p}_{i+1}, \hat{y}_{i+1,1})$ throughout the remaining $I_i - 2L_i$ periods in stage i (the earning phase). Note that $(\hat{p}_{i+1}, \hat{y}_{i+1,1})$ is optimal for the sampled optimization problem (Opt-SAA) in stage i.

3.3.3 Linear Approximation of (Opt-SAA), and Regularity Conditions

To analyze the performance of DDC, we need to compare the sampled problem (Opt-SAA) with the clairvoyant's problem (Opt-CV). However, the direct comparison or cost amortization between these two optimization problems are difficult. To alleviate such problem, we introduce two bridging optimization problems that serve as our intermediate benchmarks, with one already defined by (Approx-CV) replacing the demand-price function $\lambda(p)$ in (Opt-CV) by its linear approximation $\alpha - \beta p$. We

now define the other bridging problem (Approx-SAA) as follows.

$$\max_{p\in\mathcal{S}_{i+1}}\left\{\tilde{G}_{i+1}(p,\alpha-\beta p)\right\}, \tag{Approx-SAA}$$

$$\text{where}\quad \tilde{G}_{i+1}(p,\alpha-\beta p)\triangleq p\,(\alpha-\beta p) - \min_{y\in\mathcal{Y}}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i}\left\{(b+p)\,(\alpha-\beta p+\tilde{\epsilon}_t-y)^+ + h\,(y-(\alpha-\beta p+\tilde{\epsilon}_t))^+\right\}\right\},$$

where the truncated random error $\tilde{\epsilon}_t$ is defined by

$$\tilde{\epsilon}_t \triangleq d_t \wedge y_t - \lambda(p_t),\quad t\in\{t_i+1,\ldots,t_i+2L_i\}, \tag{3.11}$$

It is clear that the truncated random error $\tilde{\epsilon}_t=\epsilon_t$ only when $d_t\le y_t$. Let $\tilde{p}_{i+1}(\alpha,\beta)$ be the optimal price for $\tilde{G}_{i+1}(p,\alpha-\beta p)$, and for any fixed $p\in\mathcal{P}$, let $\tilde{y}_{i+1}(p,\alpha-\beta p)$ be the optimal order-up-to level for $\tilde{G}_{i+1}(p,\alpha-\beta p)$.

We have now established four key optimization problems needed to carry out our performance analysis, i.e., (Opt-CV), (Approx-CV), (Approx-SAA) and (Opt-SAA), in the order of requiring less and less demand information. The optimization problem (Opt-CV) assumes that the firm knows both the demand-price function $\lambda(p)$ and the distribution of ϵ, so the expectations can be readily computed. In (Approx-CV), the firm does not know $\lambda(p)$ and instead uses a linear function $\alpha-\beta p$ as a proxy to $\lambda(p)$. However, the distribution of ϵ is still available information. In (Approx-SAA), the firm knows neither $\lambda(p)$ nor the distribution of ϵ, but it could hypothetically access the truncated samples of ϵ_t (which does not incur the estimation error of $\lambda(p)$). Thus in addition to using a linear function $\alpha-\beta p$ as a proxy to $\lambda(p)$, the firm evaluates the expectations using truncated sample averages of ϵ in (Approx-SAA). Finally in (Opt-SAA), the firm estimates the coefficients of the linear demand-price function using historical censored demand data, and uses the (biased) residual errors η_t in place of the true random errors ϵ_t to construct the sample averages. The caveat here

is that the estimated $\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}$ from random demand realizations are random and subject to estimation errors, and so are the residual errors η_t.

We end this subsection by listing some mild *regularity conditions* for our results to hold.

(a) We assume Lipschitz condition for the single-period profit function $Q(p, y)$ in (3.5) on $p \in \mathcal{P}$ and $y \in \mathcal{Y}$, i.e., there exists some constant $K_1 > 0$ such that for any $p_1, p_2 \in \mathcal{P}$ and $y_1, y_2 \in \mathcal{Y}$,

$$|Q(p_1, y_1) - Q(p_2, y_2)| \leq K_1 \left(|p_1 - p_2| + |y_1 - y_2|\right). \tag{3.12}$$

We also assume Lipschitz conditions for $\bar{y}(q, \lambda(q))$ and $\bar{y}\big(p, \breve{\alpha}(q) - \breve{\beta}(q)p\big)$ on $q \in \mathcal{P}$ for any fixed $p \in \mathcal{P}$, i.e., there exists some constant $K_2 > 0$ such that for any $q_1, q_2 \in \mathcal{P}$,

$$|\bar{y}(q_1, \lambda(q_1)) - \bar{y}(q_2, \lambda(q_2))| \leq K_2 \left|q_1 - q_2\right|, \tag{3.13}$$

$$\left|\bar{y}\big(p, \breve{\alpha}(q_1) - \breve{\beta}(q_1)p\big) - \bar{y}\big(p, \breve{\alpha}(q_2) - \breve{\beta}(q_2)p\big)\right| \leq K_2 \left|q_1 - q_2\right|. \tag{3.14}$$

(b) The function $Q(p, \bar{y}(p, \lambda(p)))$ has a bounded second-order derivative with respect to $p \in \mathcal{P}$.

(c) The probability density function $f(\cdot)$ of ϵ satisfies $r = \min\{f(x), x \in [l, u]\} > 0$.

3.3.4 Numerical Experiment

An important question is how well the DDC algorithm performs computationally. We conduct a numerical study on its empirical performance, and present the numerical results below. The demand-price function is exponential, i.e., $\lambda(p) = e^{5.5-0.1p}$. The input parameters are initialized as follows: $p^l = 0, p^h = 20, y^l = 0, y^h = 120, I_0 = 2, v = 1.2, s = 0.1, \rho = 1$, the starting price is $\hat{p}_1 = 5$ and the starting target inventory

levels $\hat{y}_{1,1} = 80$ and $\hat{y}_{1,2} = 85$. We tested uniform and truncated normal random errors ϵ, given by

i) Uniform distribution on [-2.5, 2.5];

ii) Uniform distribution on [-5, 5];

iii) Truncated Normal distribution with mean 0 and standard deviation 1 on [-2.5, 2.5];

iv) Truncated Normal distribution with mean 0 and standard deviation 1 on [-5, 5].

And the cost parameters tested are $b \in \{2, 10, 20\}$ and $h \in \{1, 2\}$.

We measure the performance of DDC by the percentage of profit loss per period, compared with the clairvoyant optimal profit, defined by

$$\frac{Q(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} Q(p_t, y_t)\right]}{Q(p^*, y^*)} \times 100\%.$$

The results are averaged over 500 time periods and summarized in Table 3.1.

It can be seen from Table 3.1 that when $T = 10$, the regret is as high as 73.85%. When $T = 100$, the average regret is reduced to 8.69%, and then to 1.57% when $T = 1000$. Our numerical results show that DDC quickly converges to the clairvoyant optimal solution in computation.

3.4 Main Results and Performance Analysis

The average regret $\mathcal{R}(\pi, T)$ of a policy π is defined as the average profit loss per period compared with the clairvoyant optimal solution, given by

$$\mathcal{R}(\pi, T) = Q(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} Q(p_t, y_t)\right]. \tag{3.15}$$

			T=10	T=30	T=100	T=300	T=1000	T=3000	T=10000
U[-2.5, 2.5]	h=1	b=2	35.92	12.54	4.34	1.85	0.90	0.58	0.39
		b=10	64.15	22.01	7.19	2.77	1.14	0.62	0.39
		b=20	99.57	33.90	10.73	3.94	1.48	0.71	0.39
	h=2	b=2	36.03	12.78	4.59	2.09	1.23	0.92	0.67
		b=10	64.37	22.28	7.38	2.95	1.42	0.94	0.66
		b=20	99.81	34.12	10.88	4.12	1.70	0.99	0.66
U[-5, 5]	h=1	b=2	35.90	13.73	6.41	3.66	1.82	1.04	0.61
		b=10	64.13	23.31	9.39	4.51	2.07	1.09	0.61
		b=20	99.54	35.07	13.01	5.73	2.43	1.18	0.61
	h=2	b=2	35.91	13.83	6.59	3.81	2.05	1.32	0.87
		b=10	64.16	23.25	9.44	4.63	2.24	1.35	0.86
		b=20	99.80	35.34	13.02	5.77	2.52	1.43	0.87
$\mathbf{N}(0,1)$ on [-2.5, 2.5]	h=1	b=2	43.36	15.03	4.88	1.93	0.89	0.56	0.37
		b=10	78.07	26.65	8.38	3.08	1.20	0.64	0.39
		b=20	121.67	41.23	12.76	4.52	1.62	0.75	0.41
	h=2	b=2	43.40	15.26	5.15	2.20	1.24	0.91	0.64
		b=10	78.14	26.82	8.58	3.30	1.50	0.96	0.65
		b=20	121.87	41.48	12.94	4.71	1.87	1.05	0.67
$\mathbf{N}(0,1)$ on [-5, 5]	h=1	b=2	43.23	14.98	4.89	1.93	0.89	0.56	0.37
		b=10	78.12	26.65	8.39	3.10	1.21	0.64	0.39
		b=20	121.72	41.30	12.80	4.54	1.62	0.76	0.41
	h=2	b=2	43.43	15.29	5.18	2.19	1.23	0.90	0.64
		b=10	78.24	26.87	8.59	3.30	1.51	0.97	0.66
		b=20	121.93	41.49	12.96	4.73	1.88	1.05	0.67
average			73.85	25.63	8.69	3.56	1.57	0.91	0.58
maximum			121.93	41.49	13.02	5.77	2.52	1.43	0.87

Table 3.1: Percentage of Profit Loss (%)

We are now ready to present the main results of this chapter.

Theorem III.3. (Convergence of Decisions) *Under Assumption III.2, for the joint pricing and inventory control problem with lost-sales and censored demand, the DDC algorithm satisfies*

(a) $p_t \to p^*$ *in probability as* $t \to \infty$*;*

(b) $y_t \to y^*$ *in probability as* $t \to \infty$*.*

The above result asserts that the pricing and inventory decisions of DDC converge to the clairvoyant optimal solution (p^*, y^*) in probability under censored demand. The next result shows that the average regret of DDC converges to zero with a provable convergence rate.

Theorem III.4. (Convergence Rate of Regret) *Under Assumption III.2, for the joint pricing and inventory control problem with lost-sales and censored demand, there*

exists a constant $K_0 > 0$ *such that the average regret of the DDC policy satisfies*

$$\mathcal{R}(DDC, T) \leq K_0 \cdot T^{-\frac{1}{5}} (\log T)^{\frac{1}{4}}. \tag{3.16}$$

To the best of our knowledge, Theorems III.3 and III.4 provide the first asymptotic analysis on the joint pricing and inventory control problem with lost-sales and censored demand information, which significantly departs from its backorder counterpart model recently studied by *Chen et al.* (2015) in a number of ways (that are explicitly summarized in §3.1.2). We refer the readers to the detailed discussions in §3.3.2 (algorithmic design) and §3.4.1–§3.4.3 (technical analysis) for the key differences between our work and *Chen et al.* (2015). In particular, we explain the high-level ideas of the sparse discretization scheme, and the key factors that affect the above convergence rate in §3.4.1.

We also remark that traditional nonparametric approaches have been well studied in the literature of stochastic optimization, such as stochastic approximation (see *Kiefer and Wolfowitz* (1952a), *Robbins and Monro* (1951), *Nemirovski et al.* (2009) and references therein), online convex optimization (see *Zinkevich* (2003), *Hazan et al.* (2006), *Hazan* (2015) and references therein), and continuum-armed bandit algorithms (*Auer et al.* (2007), *Kleinberg* (2005), *Cope* (2009) and references therein). Many papers have adapted some of these ideas and techniques to the inventory and/or pricing settings (see, e.g., *Burnetas and Smith* (2000), *Levi et al.* (2007), *Huh and Rusmevichientong* (2009)). However, these standard approaches cannot be applied or adapted to establish convergence guarantees of our problem. The key reason is that in our setting the firm can observe neither the realized profit (because the lost-sales quantity, thus also the lost-sales cost, is unobservable), nor the realized derivatives of profit function (because the demand-price function is not known).

Remark on the convergence rate. *We first give an intuitive explanation on the*

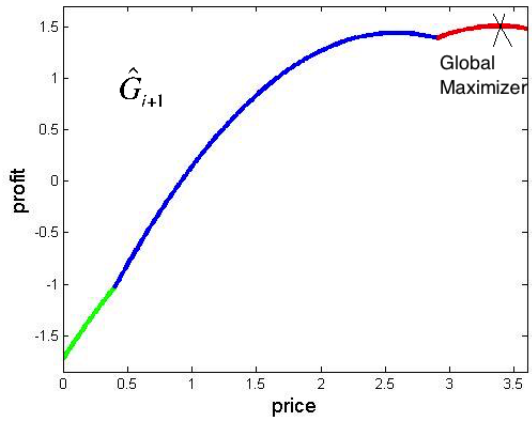

Figure 3.3: $\beta = 0.52$

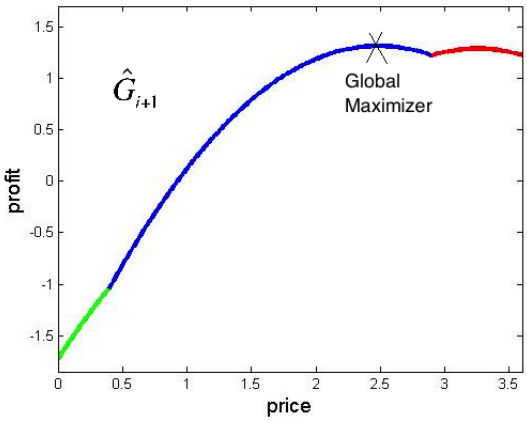

Figure 3.4: $\beta = 0.54$

key factor that affects the convergence rate, and will present a more technical discussion in §3.4.1. Figure 3.3 shows the same sampled objective function as in Example III.1, with parameters $\alpha = 2.944$ *and* $\beta = 0.52$*, and the global optimal solution is* $p = 3.4$*. Now we slightly perturb* β *to* 0.54 *(while holding all other parameters fixed), then the global optimal solution shifts drastically to* $p = 2.45$ *as depicted in Figure 3.4. This shows that even a slight perturbation in the parameters* α *and* β *can lead to a dramatic change in the global optimal solution (jumping from one region/piece of the multimodal function to another). In light of this simple numerical example of* $\hat{G}_{i+1}$ *in* (Opt-SAA)*, it is clear that the convergence in parameters (i.e.,* $\hat{\alpha}_{i+1} \to \breve{\alpha}(p^*)$ *and* $\hat{\beta}_{i+1} \to \breve{\beta}(p^*)$*) does not guarantee the convergence in the pricing decision (i.e.,* $\hat{p}_{i+1} \to p^*$*). We note that this convergence is satisfied (and needed) in both Besbes and Zeevi (2015) (see Condition 3 of their Appendix A) and Chen et al. (2015) as their optimal solution of* $\hat{G}_{i+1}$ *is Lipschitz in* (α, β)*. In these papers, the proxy objective function* $\hat{G}_{i+1}$ *is unimodal, and hence this "region switching phenomenon" of the optimal solution does not occur, and establishing the convergence in parameters guarantees the convergence in the pricing decision. Unfortunately, this is not true in the lost-sales model, and we will establish a uniform convergence argument through a sparse discretization technique in §3.4.1, which results in the final regret rate in*

Theorem III.4.

For the remainder of §3.4, we shall outline the main steps, ideas and techniques developed for the proofs of Theorems III.3 and III.4. We will also explain in details the key differences between this chapter and the closely related works, e.g., *Besbes and Zeevi* (2015) and *Chen et al.* (2015).

3.4.1 Key Ideas in Proving the Convergence of Pricing Decisions

We first discuss the key ideas and steps in proving Theorem III.3(a). It suffices to show $\mathbb{E}[(\hat{p}_{i+1} - p^*)^2] \to 0$, since convergence in L_2-norm implies convergence in probability.

Using $\bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ from (Approx-CV) as an intermediate benchmark, we have

$$(\hat{p}_{i+1} - p^*)^2 \leq \left(\underbrace{\left| p^* - \bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) \right|}_{\text{Proposition III.5}} + \underbrace{\left| \bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1} \right|}_{\text{Proposition III.6}} \right)^2 . \tag{3.17}$$

In the following, we develop upper bounds for the two terms on the RHS of (3.17).

Proposition III.5. *There exists some real number $\gamma \in [0, 1)$ such that, for any $\hat{p}_i \in \mathcal{P}$,*

$$\left| p^* - \bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) \right| \leq \gamma \left| p^* - \hat{p}_i \right| . \tag{3.18}$$

The proof of Proposition III.5 uses the same contraction mapping argument given in *Besbes and Zeevi* (2015). The key point is to establish $p^* = \bar{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big)$, which shows that p^* is a fixed point of $\bar{p}\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$ as a function of z. By further showing bounded derivative $\left| d\bar{p}\big(\breve{\alpha}(z), \breve{\beta}(z)\big)/dz \right| < 1$ under Assumption III.2(iii), we then obtain the desired result by contraction mapping. This result links the optimal solutions of (Opt-CV) and (Approx-CV) with parameters $\breve{\alpha}(\hat{p}_i)$ and $\breve{\beta}(\hat{p}_i)$. We remark that Proposition III.5 is the only technical result in which we followed the proof

techniques of *Besbes and Zeevi* (2015), and all subsequent analysis in this chapter requires new ideas.

We now develop an upper bound for the second term on the RHS of (3.17), which is one of the most critical results in our analysis. It bridges between (Approx-CV) and (Opt-SAA).

Proposition III.6. *There exist positive constants* $K_1^{\mathbf{P2}}$ *and* $K_2^{\mathbf{P2}}$ *such that, for any* $\hat{p}_i \in \mathcal{P}$,

$$\mathbb{P}\left\{\left|\bar{p}\left(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\right) - \hat{p}_{i+1}\right| \geq K_1^{\mathbf{P2}} L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}}\right\} \leq K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}. \quad (3.19)$$

The difficulty of establishing Proposition III.6 arises because the two sampled proxy objective functions $\tilde{G}_{i+1}$ in (Approx-SAA) and $\hat{G}_{i+1}$ in (Opt-SAA) lose unimodality, a key feature in the lost-sales problem. In contrast, $\tilde{G}_{i+1}$ and $\hat{G}_{i+1}$ are both unimodal in the backorder counterpart problem studied by *Chen et al.* (2015), and their proof strategy is to establish the "parameter" convergence, i.e., $\hat{\alpha}_{i+1} \to \breve{\alpha}(\hat{p}_i)$ and $\hat{\beta}_{i+1} \to \breve{\beta}(\hat{p}_i)$, which can be used to guarantee the convergence of $\hat{p}_{i+1} \to \tilde{p}_{i+1} \to \bar{p}$. However, this scheme of "parameter" convergence does not translate to "solution" convergence in the lost-sales case where unimodality is no longer preserved. This implies that we need to compare between proxy functions $\bar{G}$, $\tilde{G}_{i+1}$, and $\hat{G}_{i+1}$ from (Approx-CV), (Approx-SAA) and (Opt-SAA) directly. The two intermediate results below (Lemmas III.7 and III.8) show that, for any fixed price $p \in \mathcal{P}$, the two random proxy functions $\tilde{G}_{i+1}$ and $\hat{G}_{i+1}$ are very close to $\bar{G}$ with a high probability.

Lemma III.7. *There exists some positive constant* $K_1^{\mathbf{L1}}$ *such that, for any given* α, β, *and* $p \in \mathcal{P}$,

$$\mathbb{P}\left\{\left|\bar{G}\left(p, \alpha - \beta p\right) - \tilde{G}_{i+1}\left(p, \alpha - \beta p\right)\right| > K_1^{\mathbf{L1}} L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}\right\} \leq 4 L_i^{-4}. \quad (3.20)$$

In the function $\bar{G}(p, \alpha - \beta p)$ the distribution of ϵ is known and the expectation can be taken, whereas in the function $\tilde{G}_{i+1}(p, \alpha - \beta p)$ the expectation is replaced with the sample average of the sales data (i.e., truncated demand realizations), which suffers from (downward) estimation bias. In the proof of Lemma III.7, we show that the frequency of stockout (resulting in truncated demand realizations) decreases as L_i grows, and the estimation bias diminishes to zero as i grows.

Lemma III.8. *There exists a positive constant* $K_1^{\mathbf{L2}}$ *such that, for any given* $p \in \mathcal{P}$,

$$\mathbb{P}\left\{\left|\tilde{G}_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)\right| \geq K_1^{\mathbf{L2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}\right\} \leq 24 L_i^{-2}.$$

There are two main differences between $\tilde{G}_{i+1}$ and $\hat{G}_{i+1}$. First, $\tilde{G}_{i+1}$ is constructed using the truncated random error $\tilde{\epsilon}_t$ defined in (3.11), whereas $\hat{G}_{i+1}$ is constructed using the residual error η_t defined in (3.10). (Note that both $\tilde{\epsilon}_t$ and η_t are used to estimate the true random error ϵ_t, and both of them suffer from the estimation error due to demand censoring. The difference is that η_t also suffers from the estimation error of $\lambda(\cdot)$ while $\tilde{\epsilon}_t$ does not.)

Second, $\tilde{G}_{i+1}$ involves the parameters $\breve{\alpha}(\hat{p}_i)$ and $\breve{\beta}(\hat{p}_i)$, whereas $\hat{G}_{i+1}$ involves the parameters $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$. To compare $\tilde{G}_{i+1}$ and $\hat{G}_{i+1}$, we first make a simple yet key observation: since $\lambda(p) = \breve{\alpha}(p) - \breve{\beta}(p)p$ for any $p \in \mathcal{P}$, we can therefore write the truncated random error $\tilde{\epsilon}$ in $\tilde{G}_{i+1}$ and the residual error η_t in $\hat{G}_{i+1}$ as

$$\tilde{\epsilon}_t = d_t \wedge y_t - \lambda(\hat{p}_i) = d_t \wedge y_t - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i), \tag{3.21}$$

$$\eta_t = d_t \wedge y_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}\hat{p}_i), \tag{3.22}$$

which suggests that the difference between $\tilde{\epsilon}_t$ and η_t can be bounded by the difference between the associated parameters $(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$ and $(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$.We then show that both $\hat{\alpha}_{i+1} \to \breve{\alpha}(\hat{p}_i)$ and $\hat{\beta}_{i+1} \to \breve{\beta}(\hat{p}_i)$ in probability as i grows to obtain the desired result.

Sparse discretization scheme. To obtain the convergence of prices in Proposition III.6, we need to establish that $\bar{G}$ and $\hat{G}_{i+1}$ are uniformly close (with a high probability) across the continuous price set $\mathcal{P}$ (see Figure 3.5).

Definition III.9 (Uniform Closeness)**.** We say two functions $g_1(\cdot)$ and $g_2(\cdot)$ are uniformly close to each other over a domain $\mathcal{M}$, if there exist some (small) constants $\eta > 0$ and $\delta > 0$ such that the maximum deviation in their objective values over $\mathcal{M}$ is bounded by η with a (high) probability no less than $1 - \delta$, i.e., $\mathbb{P}\left(\max_{x \in \mathcal{M}} |g_1(x) - g_2(x)| \leq \eta\right) \geq 1 - \delta$.

Establishing uniform closeness between two functions over a continuous domain is a very challenging task, because one of them, the sampled objective function $\hat{G}_{i+1}$ in (Opt-SAA), is multimodal and ill-structured. Moreover, the number of local maxima increases when more demand data points are used to construct the sampled objective function $\hat{G}_{i+1}$.

To facilitate our performance analysis, we develop a *sparse discretization* scheme to jointly learn and optimize the multimodal profit function in (Opt-SAA). The basic idea is to carefully identify a sparse discrete set of pricing points (also referred to as the grid), and show that the sampled objective function $\hat{G}_{i+1}$ (multimodal) is uniformly close to the linear approximated objective function $\bar{G}$ (concave) over this grid (see (3.23)). We then exploit the strict concavity property of $\bar{G}$ to establish the desired convergence in price on the original continuous set of prices.

The choice of sparsity is delicate and non-trivial when constructing such a grid $\mathcal{S}_{i+1}$. We keep two factors in check: (1) Is the grid $\mathcal{S}_{i+1}$ sparse enough so that $\hat{G}_{i+1}$ is uniformly close to $\bar{G}$ over the discrete domain $\mathcal{S}_{i+1}$? The more sparse the grid is, the less outliers there will be on $\hat{G}_{i+1}$ that are far away from $\bar{G}$. (2) Is the grid $\mathcal{S}_{i+1}$ dense enough to guarantee the solution accuracy, i.e., is the optimal price of $\hat{G}_{i+1}$ on the grid $\mathcal{S}_{i+1}$ close enough to the true optimal price on the original continuous set $\mathcal{P}$? It turns out that the optimal choice of $\mathcal{S}_{i+1}$ is given by (3.8), since the chosen step

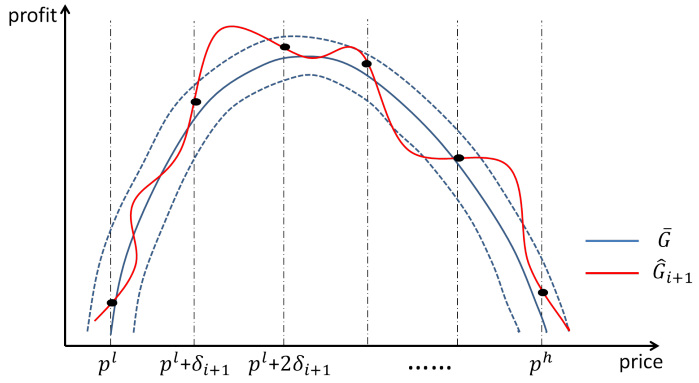

Figure 3.5: Sparse discretization and uniform closeness

size δ_{i+1} in (3.7) perfectly balances between the aforementioned two factors, yielding the best convergence results under this sparse discretization framework. Note that the number of price points in $\mathcal{S}_{i+1}$ is

$$\frac{p^h - p^l}{\delta_{i+1}} = \left(\frac{p^h - p^l}{\rho}\right) L_{i+1}^{\frac{1}{4}} (\log L_{i+1})^{-\frac{1}{4}} \leq \left(\frac{p^h - p^l}{\rho}\right) I_{i+1}^{\frac{1}{5}} \leq \left(\frac{p^h - p^l}{\rho}\right) T^{\frac{1}{5}}.$$

If, say, the planning horizon $T = 10^5$, the algorithm only needs to check no more than $(p^h - p^l)/\rho \times 10$ price points for each stage. Moreover, the number of stages $n \sim \log T = 5\log(10)$. As a result, our proposed algorithm DDC is computationally very efficient.

Using Lemmas III.7 and III.8, we first obtain the uniform closeness result (between $\bar{G}$ and $\hat{G}_{i+1}$) only on the sparse grid $\mathcal{S}_{i+1}$, i.e., for any $p \in \mathcal{S}_{i+1}$, we will show

$$\mathbb{P}\left\{\left|\bar{G}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)\right| \geq K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}\right\} \leq 28 L_i^{-2},$$

which leads to

$$\begin{aligned}\mathbb{P}\left\{\max_{p\in\mathcal{S}_{i+1}}\left|\bar{G}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)\right| \geq K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}\right\} \quad (3.23)\\ \leq 28 L_i^{-2}\left(\frac{p^h - p^l}{\delta_{i+1}}\right) \leq K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}},\end{aligned}$$

which says that $\bar{G}$ and $\hat{G}_{i+1}$ are uniformly close on the grid $\mathcal{S}_{i+1}$ (see Figure 3.5).

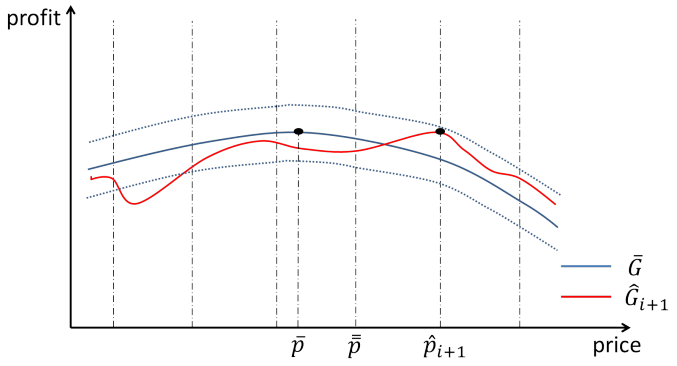

Figure 3.6: Choosing $\bar{\bar{p}}$ to be the closet point on the grid to $\bar{p}$, and also on the same side as $\hat{p}$ (relative to $\bar{p}$)

Since the uniform closeness result (3.23) is only established on the grid but the optimal price $\bar{p}$ for $\bar{G}$ in (Approx-CV) may not lie on the grid, we then choose an auxillary price point $\bar{\bar{p}} \in \mathcal{S}_{i+1}$ that is the closest point on the grid to $\bar{p}$ and also lies on the same side as $\hat{p}_{i+1}$ relative to $\bar{p}$ (see Figure 3.6). Because both $\bar{\bar{p}}$ and $\hat{p}_{i+1}$ lie on the grid, we can then apply (3.23) to obtain

$$\mathbb{P}\left\{\bar{G}(\bar{\bar{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\bar{p}}) - \bar{G}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) \leq 2K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}\right\} \quad (3.24)$$
$$\geq 1 - K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}.$$

We then show, by strict concavity of $\bar{G}$, that there exists some positive number $m > 0$ such that

$$\bar{G}(\bar{\bar{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\bar{p}}) - \bar{G}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) \geq m(\bar{\bar{p}} - \hat{p}_{i+1})^2. \quad (3.25)$$

Combining (3.24) and (3.25), there exists some constant $K_4^{\mathbf{P2}} > 0$ such that

$$\mathbb{P}\left\{|\bar{\bar{p}} - \hat{p}_{i+1}| \leq K_4^{\mathbf{P2}} L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}}\right\} \geq 1 - K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}. \quad (3.26)$$

By our choice of $\bar{\bar{p}}$ and the property of δ_{i+1}, we have

$$|\bar{p} - \bar{\bar{p}}| \leq \delta_{i+1}. \tag{3.27}$$

Combining (3.26) and (3.27), we obtain the desired result

$$\mathbb{P}\left\{|\bar{p} - \hat{p}_{i+1}| \leq K_1^{\mathbf{P2}} L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}}\right\} \geq 1 - K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}.$$

Remark. *To obtain the convergence rate of $\hat{p}_{i+1}$ to p^*, we need to bound the difference between $\hat{p}_{i+1}$ and $\bar{p}$. Due to the multimodality of the sampled proxy function $\hat{G}_{i+1}$ in* (Opt-SAA)*, a discretized search space is designed to show that $|\bar{p} - \hat{p}_{i+1}| \leq \mathcal{O}\left(L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}}\right)$ with a high probability, which largely determines the final regret rate of $T^{-\frac{1}{5}} (\log T)^{\frac{1}{4}}$ in Theorem III.4. We note that, because of the multimodality of $\hat{G}_{i+1}$, the above regret rate is the tightest possible under our current sparse discretization approach. In contrast, for the backorder counterpart model studied in Chen et al. (2015), the corresponding functions $\bar{G}$ and $\hat{G}_{i+1}$ are both* concave*, and thus their optimal prices $\bar{p}$ with $\hat{p}_{i+1}$ can be directly compared, which leads to a lower (in fact best possible) final regret rate of $T^{-\frac{1}{2}}$ in their Theorem 2.*

Now we are ready to prove Theorem III.3(a).

Proof of Theorem III.3(a). Using Proposition III.5 and some simple algebra, for some constant $K_1^{\mathbf{T1}}$,

$$\begin{aligned}
\mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] &\leq \mathbb{E}\Big[\Big(\Big|p^* - \bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\Big] \\
&\leq \mathbb{E}\Big[\Big(\gamma|p^* - \hat{p}_i| + \Big|\bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\Big] \\
&\leq \left(\frac{1+\gamma^2}{2}\right) \mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + K_1^{\mathbf{T1}} \mathbb{E}\left[\left|\bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|^2\right].
\end{aligned} \tag{3.28}$$

By Proposition III.6, we have

$$\mathbb{P}\left\{\left|\bar{p}\left(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\right)-\hat{p}_{i+1}\right|^2\geq(K_1^{\mathbf{P2}})^2L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\right\}\leq K_2^{\mathbf{P2}}L_i^{-\frac{7}{4}}(\log L_i)^{-\frac{1}{4}},\ (3.29)$$

It follows from (3.29) and the fact that $\bar{p}$ and $\hat{p}_{i+1}$ are bounded, that

$$\begin{aligned}\mathbb{E}\left[\left|\bar{p}\left(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\right)-\hat{p}_{i+1}\right|^2\right] &\leq K_2^{\mathbf{T1}}L_i^{-\frac{7}{4}}(\log L_i)^{-\frac{1}{4}}+K_3^{\mathbf{T1}}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\\ &\leq K_4^{\mathbf{T1}}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}.\end{aligned}\quad(3.30)$$

Substituting (3.30) into (3.28), we have that for $\frac{1+\gamma^2}{2}<1$,

$$\mathbb{E}\big[(p^*-\hat{p}_{i+1})^2\big]\leq\left(\frac{1+\gamma^2}{2}\right)\mathbb{E}\left[(p^*-\hat{p}_i)^2\right]+K_5^{\mathbf{T1}}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}.\quad(3.31)$$

Letting $\frac{1+\gamma^2}{2}=\theta$, we further obtain that

$$\mathbb{E}\big[(\hat{p}_{i+1}-p^*)^2\big]\ \leq\ \theta^i(\hat{p}_1-p^*)^2+K_6^{\mathbf{T1}}\sum_{j=0}^{i-1}\theta^jL_{i-j}^{-\frac{1}{4}}(\log L_{i-j})^{\frac{1}{4}}\leq K_7^{\mathbf{T1}}i^{\frac{1}{4}}(v^{-\frac{1}{5}})^i\sum_{j=0}^{i}\theta^j(v^{\frac{1}{5}})^j.\quad(3.32)$$

By choosing $v>1$ satisfying $\theta v^{\frac{1}{5}}<1$, there exists a positive constant $K_8^{\mathbf{T1}}$ such that $\sum_{j=0}^{i}\theta^j(v^{\frac{1}{5}})^j\leq K_8^{\mathbf{T1}}$. This implies that for some constants $K_9^{\mathbf{T1}}$ and $K_{10}^{\mathbf{T1}}$, we have

$$\mathbb{E}\big[(\hat{p}_{i+1}-p^*)^2\big]\leq K_9^{\mathbf{T1}}i^{\frac{1}{4}}(v^{-\frac{1}{5}})^i\leq K_{10}^{\mathbf{T1}}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\to 0,\ \text{as } i\to\infty.\quad(3.33)$$

Moreover, for some positive constant $K_{11}^{\mathbf{T1}}$, we have

$$\mathbb{E}\left[(\hat{p}_{i+1}+\delta_{i+1}-p^*)^2\right]\leq 2\mathbb{E}\left[(\hat{p}_{i+1}-p^*)^2\right]+2\delta_{i+1}^2\leq K_{11}^{\mathbf{T1}}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\to 0,\ \text{as } i\to\infty.\quad(3.34)$$

This completes the proof of Theorem III.3(a). □

3.4.2 Key Ideas in Proving the Convergence of Inventory Decisions

We next elaborate on the proof of Theorem III.3(b).

For any fixed $p \in \mathcal{P}$, recall that $\bar{y}(p, \lambda(p))$ and $\bar{y}(p, \alpha - \beta p)$ are optimal order-up-to levels for problems (Opt-CV) and (Approx-CV), respectively. By using the fact that $y^* = \bar{y}\big(p^*, \lambda(p^*)\big)$,

$$\mathbb{E}\left[\left|y^* - \hat{y}_{i+1,1}\right|^2\right] \leq \mathbb{E}\Big[\Big(\Big|\bar{y}\big(p^*, \lambda(p^*)\big) - \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))\Big| + \underbrace{\left|\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \hat{y}_{i+1,1}\right|}_{\text{Proposition III.10}}\Big)^2\Big]. \tag{3.35}$$

It follows from (3.13) that there exists some positive constant K_2 such that

$$\Big|\bar{y}\big(p^*, \lambda(p^*)\big) - \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\Big| \leq K_2 \left|p^* - \hat{p}_{i+1}\right|. \tag{3.36}$$

Thus, it suffices to bound the second term on the RHS of (3.35), which is more involved.

Proposition III.10. *There exists some positive constant $K_1^{\mathbf{P3}}$ such that,*

$$\mathbb{E}\left[(\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \hat{y}_{i+1,1})^2\right] \leq K_1^{\mathbf{P3}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}. \tag{3.37}$$

We provide some high-level ideas of proving Proposition III.10. By definitions of $\breve{\alpha}$ and $\breve{\beta}$, we have

$$\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) = \bar{y}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_{i+1}) - \breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}).$$

Then it follows that

$$\begin{aligned}
& \mathbb{E}\left[\left|\bar{y}(\hat{p}_{i+1},\lambda(\hat{p}_{i+1}))-\hat{y}_{i+1,1}\right|^2\right] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|\bar{y}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1})-\bar{y}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\Big| \\
& +\underbrace{\Big|\bar{y}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})-\tilde{y}_{i+1}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\Big|}_{\text{Lemma III.11}} \\
& +\underbrace{\Big|\tilde{y}_{i+1}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})-\hat{y}_{i+1,1}\Big|}_{\text{Lemma III.12}}\Big)^2\Big]. \qquad (3.38)
\end{aligned}$$

For the first term on the RHS of (3.38), by (3.14), there exists some positive constant K_2 such that

$$\begin{aligned}
& \Big|\bar{y}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}\big)-\bar{y}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)\Big| \qquad (3.39)\\
& \qquad\qquad \leq K_2\,|\hat{p}_{i+1}-\hat{p}_i| \leq K_2\big(\,|p^*-\hat{p}_i|+|p^*-\hat{p}_{i+1}|\,\big).
\end{aligned}$$

We then focus on the second and third terms on the RHS of (3.38) that both involve the optimal solution $\tilde{y}_{i+1}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$ for (Approx-SAA).

Lemma III.11. *There exists some positive constant $K_1^{\mathbf{L3}}$ such that, for any $\hat{p}_i,\hat{p}_{i+1}\in\mathcal{P}$,*

$$\mathbb{P}\Big\{\Big|\bar{y}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)-\tilde{y}_{i+1}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)\Big|\geq K_1^{\mathbf{L3}}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}\Big\}\leq 2L_i^{-1}.$$

For any given price $\hat{p}_{i+1}$ and parameters $\breve{\alpha}(\hat{p}_i)$ and $\breve{\beta}(\hat{p}_i)$, the inventory decision $\bar{y}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$ is the optimal solution for the inner newsvendor problem in (Approx-CV), which is a quantile solution of distribution $F(\cdot)$. On the other hand, $\tilde{y}_{i+1}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$ is computed using truncated random error $\tilde{\epsilon}_t$, which is the (downward biased) empirical newsvendor solution. To correct such estimation bias and prove Lemma III.11, we first compare $\tilde{y}_{i+1}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$ with the

unbiased empirical newsvendor solution (assuming uncensored demand data within $[t_i + 1, t_i + 2L_i]$ were available), and then compare this intermediate solution with $\bar{y}\big(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$.

Lemma III.12. *There exists some positive constant $K_1^{\mathbf{L4}}$ such that, for any $\hat{p}_i, \hat{p}_{i+1} \in \mathcal{P}$,*

$$\mathbb{P}\left\{ \left| \tilde{y}_{i+1}\big(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big) - \hat{y}_{i+1,1} \right| \geq K_1^{\mathbf{L4}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \right\} \leq 24 L_i^{-2}.$$

For any given price $\hat{p}_{i+1}$, the inventory target level $\tilde{y}_{i+1}\big(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$ in (Approx-SAA) is computed using the truncated random error $\tilde{\epsilon}_t$ and parameters $(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$, whereas the inventory target level $\hat{y}_{i+1,1}$ is computed using the residual error η_t and parameters $(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$. By (3.21) and (3.22), we can show that the difference between $\tilde{y}_{i+1}\big(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)$ and $\hat{y}_{i+1,1}$ can be bounded by the difference between their associated parameters $(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$ and $(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$.

While the sampled optimization problem (Opt-SAA) is multimodal in p, its *inner* sampled optimization problem (i.e., optimizing on y for a given price p) is concave and well-structured (see Figure 3.7). Hence, it can be shown that establishing the convergence in parameters guarantees the convergence in the inventory target levels. We emphasize again that such translation is not viable for establishing the convergence in pricing decisions, since (Opt-SAA) is multimodal in p.

Proof of Theorem III.3(b). By (3.35) and Proposition III.10, we have

$$\mathbb{E}\left[\big(y^* - \hat{y}_{i+1,1}\big)^2 \right] \leq K_{12}^{\mathbf{T1}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \to 0, \qquad \text{as } i \to \infty,$$

and similarly,

$$\mathbb{E}\left[\big(y^* - \hat{y}_{i+1,2}\big)^2 \right] \leq K_{13}^{\mathbf{T1}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \to 0, \qquad \text{as } i \to \infty.$$

This completes the proof of Theorem III.3(b). □

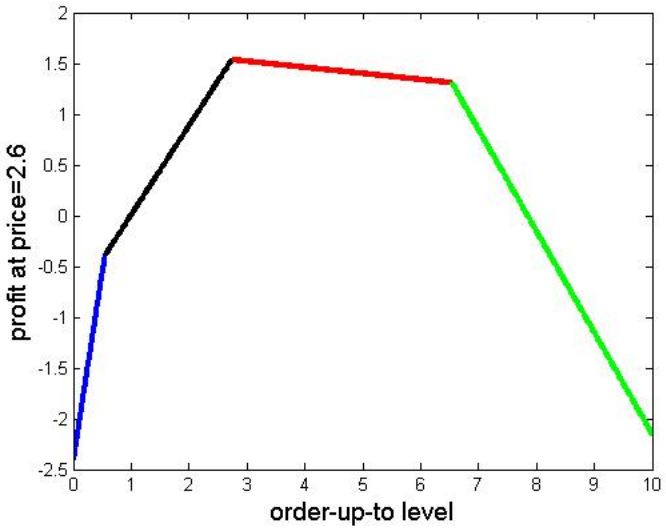

Figure 3.7: The sampled profit as a function of order-up-to level y (for a fixed price $p = 2.6$) in Example III.1

3.4.3 High Level Ideas in Proving the Regret Rate

To prove Theorem III.4, we break the time horizon T into n learning stages to obtain

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{t=1}^{T}\big(Q(p^*,y^*)-Q(p_t,y_t)\big)\right] \\
\leq\ & \mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+2L_i}\big(Q(p^*,y^*)-Q(p_t,y_t)\big)\right.\right. \\
& \left.\left.+\sum_{t=t_i+2L_i+1}^{t_i+I_i}\big(Q(p^*,y^*)-Q(\hat{p}_{i+1},\hat{y}_{i+1,1})+Q(\hat{p}_{i+1},\hat{y}_{i+1,1})-Q(p_t,y_t)\big)\right)\right] \\
=\ & \underbrace{\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+1}^{t_i+2L_i}\big(Q(p^*,y^*)-Q(p_t,y_t)\big)\right]}_{\text{regret from experimentation}}+\underbrace{\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+2L_i+1}^{t_i+I_i}\big(Q(p^*,y^*)-Q(\hat{p}_{i+1},\hat{y}_{i+1,1})\big)\right]}_{\text{regret from the estimation error}} \\
& +\underbrace{\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+2L_i+1}^{t_i+I_i}\big(Q(\hat{p}_{i+1},\hat{y}_{i+1,1})-Q(p_t,y_t)\big)\right]}_{\text{regret from missing inventory targets}}, \qquad (3.40)
\end{aligned}
$$

where the inequality follows from the construction of DDC. As shown in (3.40), the first part of the regret is due to pricing and ordering experimentation of each stage, which is not required in the observable demand case. To carry out upward corrections of order-up-to levels due to demand censoring, DDC keeps increasing the ordering level whenever a stockout occurs, resulting in some bounded profit loss. The second part of regret in (3.40) gives the difference between profit of "fictitious" DDC (that implements $(\hat{p}_{i+1}, \hat{y}_{i+1,1})$ exactly) and that of the clairvoyant optimal policy. We call it "fictitious" DDC because $\hat{y}_{i+1,1}$ may not be attained if the starting inventory level is already higher. The third part of regret in (3.40) reflects the profit loss of missing inventory targets from DDC, due to positive inventory carryover.

We note that for the classical inventory setting without dynamic pricing, *Huh and Rusmevichientong* (2009) developed a queueing approach to resolve the issue of missing inventory targets due to positive inventory carryovers; in contrast, we show that, with a very high probability, the prescribed target level $\hat{y}_{i+1,1}$ becomes achievable after a small number of periods. We also remark that other related works such as *Burnetas and Smith* (2000), *Huh et al.* (2011) and *Besbes and Muharremoglu* (2013) considered the so-called "repeated newsvendor problem" with no inventory carryovers, so they would not encounter this "overshooting" issue.

Next we develop upper bounds for each of the three terms on the RHS of (3.40).

The first term on the RHS of (3.40) can be bounded as follows: For some positive constants $K_2^{\mathbf{T2}}, K_3^{\mathbf{T2}}$, and $K_4^{\mathbf{T2}}$,

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+1}^{t_i+2L_i}\big(Q(p^*,y^*)-Q(p_t,y_t)\big)\right] \le \sum_{i=1}^{n} 2L_i K_2^{\mathbf{T2}} \le K_3^{\mathbf{T2}}\sum_{i=1}^{n} I_i^{\frac{4}{5}} \le K_4^{\mathbf{T2}} T^{\frac{4}{5}}, \quad (3.41)$$

where the first inequality holds because (p, y) is bounded on $\mathcal{P}\times\mathcal{Y}$ and $Q(p, y)$ is Lipschitz by (3.12), and the second inequality holds by the choice of the experimentation interval L_i in DDC.

We then focus on the second term on the RHS of (3.40), which has the following upper bound.

Proposition III.13. *There exists some positive constant* $K_1^{\mathbf{P4}}$ *such that*

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+2L_i+1}^{t_i+I_i}\left(Q\big(p^*,y^*\big)-Q(\hat{p}_{i+1},\hat{y}_{i+1,1})\right)\right]\leq K_1^{\mathbf{P4}}T^{\frac{4}{5}}(\log T)^{\frac{1}{4}}. \tag{3.42}$$

This proposition measures how good the target decisions of DDC are. We write

$$\begin{aligned} & \mathbb{E}\left[Q\big(p^*,\bar{y}\big(p^*,\lambda(p^*)\big)\big)-Q(\hat{p}_{i+1},\hat{y}_{i+1,1})\right] \\ \leq\ & \mathbb{E}\left[Q\big(p^*,\bar{y}\big(p^*,\lambda(p^*)\big)\big)-Q\big(\hat{p}_{i+1},\bar{y}\big(\hat{p}_{i+1},\lambda(\hat{p}_{i+1})\big)\big)\right] \\ & +\mathbb{E}\left[Q\big(\hat{p}_{i+1},\bar{y}\big(\hat{p}_{i+1},\lambda(\hat{p}_{i+1})\big)\big)-Q(\hat{p}_{i+1},\hat{y}_{i+1,1})\right], \end{aligned} \tag{3.43}$$

where the first term on the RHS of (3.43) can be bounded by the difference between p^* and $\hat{p}_{i+1}$ (from the analysis for Theorem III.3(a)), and the second term can be bounded by the difference between $\bar{y}\big(\hat{p}_{i+1},\lambda(\hat{p}_{i+1})\big)$ and $\hat{y}_{i+1,1}$ (from the analysis for Theorem III.3(b)).

Proposition III.14. *There exists some positive constant* $K_1^{\mathbf{P5}}$ *such that*

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+2L_i+1}^{t_i+I_i}\left(Q(\hat{p}_{i+1},\hat{y}_{i+1,1})-Q(p_t,y_t)\right)\right]\leq K_1^{\mathbf{P5}}T^{\frac{1}{2}}. \tag{3.44}$$

This part of the regret captures the profit loss of missing inventory targets, due to positive inventory carryover. More precisely, in the process of implementing DDC, the desired inventory order-up-to level $\hat{y}_{i+1,1}$ may not be reached if $x_t > \hat{y}_{i+1,1}$ for some periods $t = t_i + 2L_i + 1, \ldots, t_i + I_i$. This poses a challenge in bounding the regret.

Our strategy is to utilize the Hoeffding's inequality to show that, with a very high probability, after a small number of periods, the prescribed inventory order up-to

level $\hat{y}_{i+1,1}$ becomes achievable. By the construction of DDC, the same target level is prescribed for every period in $t = t_i + 2L_i + 1, \ldots, t_i + I_i$, which helps resolve the issue of missing inventory targets. During the small number of periods, although demand in each period can be zero, it is very likely that the cumulative demands during the small number of periods can consume the initial onhand inventory to a level lower than the inventory target. Then, the inventory target will always be reached from that point onwards.

Proof of Theorem III.4. The proof of Theorem III.4 follows directly from combining (3.41), and Propositions III.13 and III.14. □

3.5 Discussions

In this chapter we studied a joint pricing and inventory control problem with lost-sales and censored demand. The demand-price information is not known *a priori*, and the firm makes pricing and inventory decisions in each period based on past sales data. We developed the first nonparametric algorithm for such system, and showed that it converges to the optimal policy as the planning horizon increases. We also obtained the convergence rate at which the regret vanishes to zero.

Observable demand case. A natural question is whether the regret rate can be improved if the firm were able to observe the full demand realizations (including the lost-sales quantity). The answer turns out to be affirmative. Note that this *observable demand case* is in fact applicable in some applications, such as online retailing where the online system can keep track of the lost customers via clicks, queries and order submissions. We prescribe a different nonparametric algorithm called DDO in the Appendix, which has an improved regret rate of $\mathcal{O}(T^{-\frac{1}{4}}(\log T)^{\frac{1}{4}})$. The proofs are omitted since the arguments are very similar to that of the censored demand case. As we discussed in §4.1, the key reason behind the larger regret rate

(compared with the backorder model) is that the proxy objective profit functions constructed from the demand data are multimodal. Indeed, even when the lost-demand is observed, the sample-based single-period profit function is still multimodal. As a result, even though active exploration is no longer necessary to learn about the demand distribution, the sparse discretization approach is still needed to search for the approximate solution of the data-driven optimization problem, leading to the said regret rate.

Unbounded demand. In the preceding sections we have focused on the case with bounded demand. When the demand is unbounded, i.e., $\mathbb{P}\{D_t(p) > x\} > 0$ for all $x \geq 0$, we can extend our results after some minor modifications. We make the following mild technical assumptions:

(a) The random demand is light tailed, i.e., in a small neighborhood of 0, the moment generating function of the error term ϵ is finite, i.e., $\mathbb{E}[\exp(\kappa\epsilon)] < +\infty$ for κ near 0.

(b) The search region for inventory level is sufficiently large. More precisely, $y^h \geq K_1^{\mathbf{U}} \log T$ for some constant $K_1^{\mathbf{U}} > 0$.

(c) The optimal order-up-to level y^* is known to lie in some range $y^* \in [y_0^l, y_0^h]$ for some positive constants y_0^l and y_0^h. In addition, $r = \min\left\{f(x), x \in [z^l, z^u]\right\} > 0$, where

$$\begin{aligned} z^l &= \min_{y\in[y_0^l,y_0^h],p,q\in[p^l,p^h]} \Big(y - (\breve{\alpha}(p) - \breve{\beta}(p)q)\Big), \\ z^u &= \max_{y\in[y_0^l,y_0^h],p,q\in[p^l,p^h]} \Big(y - (\breve{\alpha}(p) - \breve{\beta}(p)q)\Big). \end{aligned}$$

When the demand is unbounded, at any inventory level, there will be stockouts from time to time. Assumption (b) ensures that the decision space during exploration phase for inventory order-up-to level is large enough so that the distribution of de-

mand can be adequately learned through experimentation. Indeed, since the demand can take very large values in the unbounded case, if the inventory decision space is tightly constrained, then one would not be able to learn about the necessary demand information in order to find the optimal inventory level. Assumption (c) is needed in the proof of Lemma 3 when bounding the difference between order quantities.

There are two changes in the DDC algorithm for the unbounded demand case. The first change is that, during the exploration phase of Step 1, every time a stockout occurs, we raise the inventory by certain percentage until the inventory level hits $K_1^{\mathbf{U}} \log T$; and the second change is that, in the data-driven optimization in Step 3, the feasible region for y is constrained to $[y_0^l, y_0^h]$. In the Appendix, we show that Theorems III.3 and III.4 continue to hold for the modified algorithm except that the regret rate in Theorem III.4 is changed to $\mathcal{O}(T^{-\frac{1}{5}} \log T)$.

3.6 Appendix

A: Sufficient Conditions for Assumption III.2

We present some sufficient conditions for the demand-price function to satisfy Assumption III.2. For notational convenience, let $\mathcal{D} = [d^l, d^h]$ where $d^l = \lambda(p^h)$ and $d^h = \lambda(p^l)$, and also let $\lambda^{-1}(\cdot)$ be the inverse function of $\lambda(\cdot)$.

For Assumption III.2(i) to hold, it suffices to require $R(d, y) \triangleq \lambda^{-1}(d)\mathbb{E}[\min\{d + \epsilon_t, y\}]$ to be jointly concave. Then the objective function after minimizing over y is concave in d, and as a result is unimodal in p. *Chen et al.* (2014b) proposed sufficient conditions for $R(d, y)$ to be jointly concave. Define $\varrho(d, y) = -\frac{\lambda^{-1}(d)}{(\lambda^{-1}(d))'} \frac{F'(y-d)}{\bar{F}(y-d)}$, and the two sufficient conditions are as follows.

(C1) $(\lambda^{-1}(d))''d + (\lambda^{-1}(d))' \leq 0$ for all $d \in \mathcal{D}$; and

(C2) $\varrho(d, y) \geq 1$ for all $d \in \mathcal{D}$ and $y \geq 0$.

The first condition (C1) is satisfied by a fairly large class of demand functions, which includes linear demand $\lambda(p) = k - mp$, log demand $\lambda(p) = \ln(k - mp)$, logit demand $\lambda(p) = \frac{e^{k-mp}}{1+e^{k-mp}}$, and exponential demand $\lambda(p) = e^{k-mp}$, for some parameter $m > 0$.

For the second condition (C2), note that $\varrho(d, y) = d \times \frac{-\lambda^{-1}(d)}{d(\lambda^{-1}(d))'} \times \frac{F'(y-d)}{\bar{F}(y-d)}$, where $\frac{-\lambda^{-1}(d)}{d(\lambda^{-1}(d))'}$ is the price elasticity of demand and $\frac{F'(y-d)}{\bar{F}(y-d)}$ is the hazard rate of the error distribution. Thus (C2) is satisfied if $d \geq 1$ and both the price elasticity and the hazard rate are no smaller than 1.

For Assumption III.2(ii) to hold, it suffices that for any $z \in \mathcal{P}$, any $p \in \mathcal{P}$, the second derivative

$$\frac{\partial \bar{G}^2(p, \breve{\alpha}(z) - \breve{\beta}(z)p)}{\partial p^2} = 2\lambda'(z) + h^2(b+p+h)^{-3} f\left(F^{-1}\left(\frac{b+p}{b+p+h}\right)\right)^{-1} < 0 \quad (3.45)$$

For Assumption III.2(iii) to hold, it suffices to require the absolute derivative

$$\left|\frac{d\bar{p}(\breve{\alpha}(z), \breve{\beta}(z))}{dz}\right| = \left|\frac{\lambda''(z)(2p-z)}{2\lambda'(z) + h^2(b+p+h)^{-3} f\left(F^{-1}\left(\frac{b+p}{b+p+h}\right)\right)^{-1}}\right| < 1 \quad (3.46)$$

to hold for any $z \in \mathcal{P}$ and for p satisfying $\lambda(z) - \lambda'(z)z + 2\lambda'(z)p = \mathbb{E}\left[\epsilon - F^{-1}\left(\frac{b+p}{b+p+h}\right)\right]^+$.

We provide several simple examples of demand-price functions that satisfy Assumption 1.

(1) Linear demand: $\lambda(p) = k - mp$ and ϵ is uniformly distributed over $[-n, n]$, where $m \geq 1$, $0 \leq p^l \leq p^h$ and $0 < n/8 < h \leq b$.

(2) Exponential demand: $\lambda(p) = e^{k-mp}$ and ϵ is uniform on $[-n, n]$, where $k > 5$, $0.01 < m < 0.2$, $0 \leq n < 3$, $0 \leq h < 0.076$, $b = 20h$, $p^l = 0$, and $p^h = 1.1/m$.

(3) Logit demand: $\lambda(p) = \frac{e^{k-mp}}{1+e^{k-mp}}$ and ϵ is uniform on $[-n, n]$, where $0 \leq k \leq 1.39$, $0.1 \leq m \leq 0.34$, $n = 2.56$, $h = 0.1$, $b = 2$, $p^l = 4$, $p^h = 6$, and $y^l = 25/9$, $y^h = 3$.

B: Technical Proofs for Theorem III.3(a)

Proof of Proposition III.5. We first show that

$$p^* = \bar{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big). \tag{3.47}$$

That is, p^* is a fixed point of $\bar{p}\big(\breve{\alpha}(z), \breve{\beta}(z)\big) = z$. Recall that

$$\bar{G}(p, \lambda(p)) = p\lambda(p) - \min_{y \in \mathcal{Y}} \Big\{ (b+p)\mathbb{E}\big[\lambda(p) + \epsilon - y\big]^+ + h\mathbb{E}\big[y - \lambda(p) - \epsilon\big]^+ \Big\}. \tag{3.48}$$

We know that $\bar{G}$ has a unique maximizer p^* by Assumption III.2(i), and also by definition of $\bar{p}$ that $\bar{p}(\breve{\alpha}(z), \breve{\beta}(z))$ is the unique optimal solution for (Approx-CV) with parameters $\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$. Then (3.47) follows immediately from Lemma A1 of *Besbes and Zeevi* (2015) by replacing their function G with our objective function (3.48).

In addition, by Assumption III.2(iii), we have $\left|\frac{d\bar{p}\left(\breve{\alpha}(z), \breve{\beta}(z)\right)}{dz}\right| < 1$ for any $z \in \mathcal{P}$, which implies that

$$\left|\bar{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big) - \bar{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\right| \leq \gamma \left|p^* - \hat{p}_i\right|,$$

where $\gamma = \max_{z \in \mathcal{P}} \left|\frac{d\bar{p}\left(\breve{\alpha}(z), \breve{\beta}(z)\right)}{dz}\right| < 1$. This completes the proof. □

Proof of Lemma III.7. For any given α, β, define the following newsvendor-type functions

$$\begin{aligned}
\bar{W}(p, y) &= h\mathbb{E}\left[y - (\alpha - \beta p) - \epsilon\right]^+ + (b+p)\mathbb{E}\left[\alpha - \beta p + \epsilon - y\right]^+, \\
\tilde{W}_{i+1}(p, y) &= \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \Big(h\big(y - (\alpha - \beta p) - \tilde{\epsilon}_t\big)^+ + (b+p)\big(\alpha - \beta p + \tilde{\epsilon}_t - y\big)^+ \Big),
\end{aligned}$$

where recall that the truncated random error $\tilde{\epsilon}_t$ is defined in (3.11).

For any given $p \in \mathcal{P}$, by the definition of G's and the triangle inequality, we have

$$\begin{aligned} & \left|\bar{G}(p, \alpha-\beta p)-\tilde{G}_{i+1}(p, \alpha-\beta p)\right| \\ = & \left|\bar{W}(p, \bar{y}(p, \alpha-\beta p))-\tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha-\beta p))\right| \\ \leq & \left|\bar{W}\big(p, \bar{y}(p, \alpha-\beta p)\big)-\tilde{W}_{i+1}\big(p, \bar{y}(p, \alpha-\beta p)\big)\right| \\ & +\left|\tilde{W}_{i+1}\big(p, \bar{y}(p, \alpha-\beta p)\big)-\tilde{W}_{i+1}\big(p, \tilde{y}_{i+1}(p, \alpha-\beta p)\big)\right| . \end{aligned} \tag{3.49}$$

It then suffices to bound the two terms on the RHS of (3.49) as follows, for any $\xi > 0$:

$$\mathbb{P}\left\{\left|\bar{W}(p, \bar{y}(p, \alpha-\beta p))-\tilde{W}_{i+1}(p, \bar{y}(p, \alpha-\beta p))\right| > K_3^{\mathbf{L1}} \xi+\frac{K_4^{\mathbf{L1}}}{2 L_i}\right\} \leq 2 e^{-4 L_i \xi^2}. \tag{3.50}$$

$$\mathbb{P}\left\{\left|\left(\tilde{W}_{i+1}(p, \bar{y}(p, \alpha-\beta p))-\tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha-\beta p))\right)\right| > K_5^{\mathbf{L1}}(b+p^h+h)\left(\xi+\frac{K_6^{\mathbf{L1}}}{2 L_i}\right)\right\} \leq 2 e^{-4 L_i \xi^2}. \tag{3.51}$$

Letting $\xi = L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}$, then the proof of Lemma III.7 follows from combining (3.49), (4.62) and (3.51).

We first focus on proving (4.62). For any $p \in \mathcal{P}$ and $y \in \mathcal{Y}$, let $z = y - (\alpha - \beta p)$. Then the optimal z that minimizes $\bar{W}(p, z+\alpha-\beta p)$ is $\bar{z} = \bar{y}(p, \alpha-\beta p) - (\alpha-\beta p) = F^{-1}\left(\frac{b+p}{b+p+h}\right)$. Moreover, we have

$$\bar{W}(p, \bar{y}(p, \alpha-\beta p)) = \bar{W}(p, \bar{z}+\alpha-\beta p) = h \mathbb{E}\big(\bar{z}-\epsilon\big)^+ + (b+p) \mathbb{E}\big(\epsilon-\bar{z}\big)^+, \tag{3.52}$$

$$\tilde{W}_{i+1}(p, \bar{y}(p, \alpha-\beta p)) = \tilde{W}_{i+1}(p, \bar{z}+\alpha-\beta p) = \frac{1}{2 L_i} \sum_{t=t_i+1}^{t_i+2 L_i}\left(h\big(\bar{z}-\tilde{\epsilon}_t\big)^+ + (b+p)\big(\tilde{\epsilon}_t-\bar{z}\big)^+\right), \tag{3.53}$$

$$\tilde{W}_{i+1}^A(p, \bar{y}(p, \alpha-\beta p)) = \tilde{W}_{i+1}^A(p, \bar{z}+\alpha-\beta p) = \frac{1}{2 L_i} \sum_{t=t_i+1}^{t_i+2 L_i}\left(h\big(\bar{z}-\epsilon_t\big)^+ + (b+p)\big(\epsilon_t-\bar{z}\big)^+\right) . \tag{3.54}$$

For $t \in \{t_i + 1, \ldots, t_i + 2L_i\}$, we denote

$$\Delta_t = \left(h\mathbb{E}[\bar{z} - \epsilon]^+ + (b+p)\mathbb{E}[\epsilon - \bar{z}]^+\right) - \left(h(\bar{z} - \epsilon_t)^+ + (b+p)(\epsilon_t - \bar{z})^+\right).$$

Then $\mathbb{E}[\Delta_t] = 0$ and Δ_t has a bounded support of some positive length $K_3^{\mathbf{L1}}$ (as ϵ_t is bounded). We then apply the Hoeffding's inequality (see Theorem 1 in *Hoeffding* (1963)) to obtain, for any $p \in \mathcal{P}$ and $\xi > 0$, $\mathbb{P}\left\{\left|\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i}\Delta_t\right| > K_3^{\mathbf{L1}}\xi\right\} \leq 2e^{-4L_i\xi^2}$, which implies

$$\mathbb{P}\left\{\left|\bar{W}(p, \bar{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}^A(p, \bar{y}(p, \alpha - \beta p))\right| > K_3^{\mathbf{L1}}\xi\right\} \leq 2e^{-4L_i\xi^2}. \tag{3.55}$$

Denote the set of periods whose target level is above the demand realization by

$$\mathcal{C}_i \triangleq \{t :\in [t_1 + 1, \ldots, t_i + 2L_i] : y_t > d_t\}.$$

Note that $\tilde{\epsilon}_t = \epsilon_t$ for $t \in \mathcal{C}_i$. Because $y_t \in [y^l, y^h]$, let $\tilde{n}$ be the number of order-up-to level raised during the two L_i intervals in Step 1 of DDC, one has $y^l(1+s)^{\tilde{n}} < y^h(1+s)$, which is $\tilde{n} < \log_{1+s}(y^h/y^l) + 1$. This implies that the number of stockout during these $2L_i$ intervals is thus bounded by a constant, i.e.,

$$2L_i - |\mathcal{C}_i| < 2\log_{1+s}(y^h/y^l) + 2, \tag{3.56}$$

where $|\mathcal{C}_i|$ denotes the cardinality of $\mathcal{C}_i$. In addition, because ϵ_t is bounded, and by (3.56), there exists a constant $K_4^{\mathbf{L1}} > 0$ such that

$$\left|\tilde{W}_{i+1}^A(p, \bar{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \bar{y}(p, \alpha - \beta p))\right| < \frac{K_4^{\mathbf{L1}}}{2L_i}. \tag{3.57}$$

Thus, (4.62) follows from (3.55) and (3.57).

Next we focus on proving (3.51). We denote the empirical distribution of ϵ_t by

$$\hat{F}^A(x) = \frac{1}{2L_i} \sum_{t=1}^{2L_i} \mathbb{1}\left\{\epsilon_t \leq x\right\}, \qquad x \in [l, u].$$

For $\xi > 0$, it can be verified that

$$\mathbb{P}\left\{\hat{F}^A(\bar{z}) < \frac{b+p}{b+p+h} - \xi\right\} = \mathbb{P}\left\{\hat{F}^A(\bar{z}) < F(\bar{z}) - \xi\right\} = \mathbb{P}\left\{\hat{F}^A(\bar{z}) - F(\bar{z}) < -\xi\right\} \quad \leq \quad e^{-4L_i\xi^2},$$

where the inequality is due to the Hoeffding's inequality. Similarly, we have

$$\mathbb{P}\left\{\hat{F}^A(\bar{z}) > \frac{b+p}{b+p+h} + \xi\right\} \quad \leq \quad e^{-4L_i\xi^2}.$$

Combining the above two inequalities, we have

$$\mathbb{P}\left\{\left|\hat{F}^A(\bar{z}) - \frac{b+p}{b+p+h}\right| \leq \xi\right\} \geq 1 - 2e^{-4L_i\xi^2}. \tag{3.58}$$

We then define a (biased) empirical distribution of ϵ_t using truncated demand data as follows,

$$\hat{F}(x) = \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\tilde{\epsilon}_t \leq x\right\}, \qquad x \in [l, u].$$

By (3.56), we have that for some positive constant $K_6^{\mathbf{L1}}$,

$$0 \leq \hat{F}(\bar{z}) - \hat{F}^A(\bar{z}) \leq \frac{K_6^{\mathbf{L1}}}{2L_i}. \tag{3.59}$$

For any given $p \in \mathcal{P}$ and any $\xi > 0$, define the event $\mathcal{A}_1(p, \xi)$ as follows,

$$\mathcal{A}_1(p, \xi) = \left\{\omega : \left|\hat{F}(\bar{z}) - \frac{b+p}{b+p+h}\right| \leq \xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right\}.$$

Combining (3.58) and (3.59), we have

$$\mathbb{P}(\mathcal{A}_1(p,\xi)) \geq 1 - 2e^{-4L_i\xi^2}. \tag{3.60}$$

For any given $p \in \mathcal{P}$, and any given α, β, we define $\tilde{y}^u_{i+1}(p, \alpha-\beta p)$ as the unconstrained optimal order-up-to level for (Approx-SAA) on $\mathbb{R}_+$, and let $\tilde{z}^u_{i+1} = \tilde{y}^u_{i+1}(p, \alpha - \beta p) - (\alpha - \beta p)$, then

$$\tilde{z}^u_{i+1} = \min\left\{\tilde{\epsilon}_j : \frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\tilde{\epsilon}_t \leq \tilde{\epsilon}_j\right\} \geq \frac{b+p}{b+p+h}\right\}.$$

Similarly, let $\tilde{z}_{i+1} = \tilde{y}_{i+1}(p, \alpha - \beta p) - (\alpha - \beta p)$. Because $\tilde{y}_{i+1}(p, \alpha - \beta p) \in [y^l, y^h]$, it holds that $\tilde{z}_{i+1} \in [y^l - (\alpha - \beta p), y^h - (\alpha - \beta p)]$, and by convexity of newsvendor functions, we have

$$\tilde{z}_{i+1} = \min\left\{\max\left\{\tilde{z}^u_{i+1}, y^l - (\alpha - \beta p)\right\}, y^h - (\alpha - \beta p)\right\}.$$

By $\tilde{y}^u_{i+1}(p, \alpha-\beta p) = \tilde{z}^u_{i+1} + \alpha - \beta p$, we have $\tilde{W}_{i+1}(p, \tilde{y}^u_{i+1}(p, \alpha-\beta p)) = \tilde{W}_{i+1}(p, \tilde{z}^u_{i+1} + \alpha - \beta p)$. It then suffices to develop an upper bound for $\tilde{W}_{i+1}(p, \bar{z}+\alpha-\beta p) - \tilde{W}_{i+1}(p, \tilde{z}^u_{i+1} + \alpha - \beta p)$ conditioning on the event $\mathcal{A}_1(p, \xi)$.

First, for any given $d \in \mathcal{D}$, if $\bar{z} \leq \tilde{z}^u_{i+1}$, it follows from (3.53) that

$$\begin{aligned}
& \tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) \\
= \; & \frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i}\Big[(b+p)\big(\tilde{\epsilon}_t - \bar{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < \tilde{\epsilon}_t\big\} \\
& + (b+p)\big(\tilde{\epsilon}_t - \bar{z}\big)\mathbb{1}\big\{\bar{z} < \tilde{\epsilon}_t \leq \tilde{z}^u_{i+1}\big\} + h\big(\bar{z} - \tilde{\epsilon}_t\big)\mathbb{1}\big\{\tilde{\epsilon}_t \leq \bar{z}\big\}\Big] \\
\leq \; & \frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i}\Big[(b+p)\big(\tilde{\epsilon}_t - \bar{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < \tilde{\epsilon}_t\big\} \\
& + (b+p)\big(\tilde{z}^u_{i+1} - \bar{z}\big)\mathbb{1}\big\{\bar{z} < \tilde{\epsilon}_t \leq \tilde{z}^u_{i+1}\big\} + h\big(\bar{z} - \tilde{\epsilon}_t\big)\mathbb{1}\big\{\tilde{\epsilon}_t \leq \bar{z}\big\}\Big],
\end{aligned} \tag{3.61}$$

where the inequality holds by replacing ϵ_t by its upper bound $\tilde{z}^u_{i+1}$, and

$$\begin{aligned}
& \tilde{W}_{i+1}(p, \tilde{z}^u_{i+1} + \alpha - \beta p) \\
= \; & \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \Big[(b+p)(\tilde{\epsilon}_t - \tilde{z}^u_{i+1})) \mathbb{1}\{\tilde{z}^u_{i+1} < \tilde{\epsilon}_t\} \\
& + h(\tilde{z}^u_{i+1} - \tilde{\epsilon}_t) \mathbb{1}\{\bar{z} < \tilde{\epsilon}_t \le \tilde{z}^u_{i+1}(p)\} + h(\tilde{z}^u_{i+1} - \tilde{\epsilon}_t) \mathbb{1}\{\tilde{\epsilon}_t \le \bar{z}\} \Big] \\
\ge \; & \frac{1}{2L_i} \sum_{t \in \mathcal{C}_i} \Big[(b+p)(\tilde{\epsilon}_t - \tilde{z}^u_{i+1})) \mathbb{1}\{\tilde{z}^u_{i+1} < \tilde{\epsilon}_t\} + h(\tilde{z}^u_{i+1} - \tilde{\epsilon}_t) \mathbb{1}\{\tilde{\epsilon}_t \le \bar{z}\} \Big], \quad (3.62)
\end{aligned}$$

where the inequality follows from dropping the nonnegative middle term. Consequently when $\bar{z} \le \tilde{z}^u_{i+1}$, we subtract (3.62) from (3.61) to obtain

$$\begin{aligned}
& \tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}^u_{i+1} + \alpha - \beta p) \\
\le \; & (b+p) \left(\tilde{z}^u_{i+1} - \bar{z}\right) \left(1 - \hat{F}(\tilde{z}^u_{i+1})\right) + (b+p)(\tilde{z}^u_{i+1} - \bar{z}) \left(\hat{F}(\tilde{z}^u_{i+1}) - \hat{F}(\bar{z})\right) \\
& + h(\bar{z} - \tilde{z}^u_{i+1}) \hat{F}(\bar{z}) \\
= \; & (\tilde{z}^u_{i+1} - \bar{z}) \left(-(h+b+p)\hat{F}(\bar{z}) + b + p\right) \\
\le \; & \big(\tilde{z}^u_{i+1} - \bar{z}\big)(b+p+h) \left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right), \quad (3.63)
\end{aligned}$$

where the second inequality follows from the definition of $\mathcal{A}_1(p, \xi)$.

Similarly, if $\bar{z} > \tilde{z}^u_{i+1}$, by the symmetric argument, we have

$$\tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}^u_{i+1} + \alpha - \beta p) \le (\bar{z} - \tilde{z}^u_{i+1})(b+p+h) \left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right). \quad (3.64)$$

Combining (3.63) and (3.64), we have that conditioning on $\mathcal{A}_1(p, \xi)$,

$$\tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}^u_{i+1} + \alpha - \beta p) \le \big|\bar{z} - \tilde{z}^u_{i+1}\big|(b+p+h) \left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right).$$

As the demand is bounded, so is $\tilde{z}^u_{i+1} + \alpha - \beta p$, and therefore it follows from $\bar{z}(p) +$

$\alpha - \beta p \in \mathcal{Y}$ that there exists some constant $K_5^{\mathbf{L1}} > 0$ such that $\left|\bar{z} - \tilde{z}_{i+1}^u\right| \leq K_5^{\mathbf{L1}}$. Thus

$$\tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq K_5^{\mathbf{L1}}(b + p + h)\left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right).$$

Since $\tilde{z}_{i+1}^u$ is the unconstrained minimizer of $\tilde{W}_{i+1}(p, z + \alpha - \beta p)$, it follows that

$$\begin{aligned} & \tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1} + \alpha - \beta p) \\ \leq \quad & \tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \\ \leq \quad & K_5^{\mathbf{L1}}(b + p + h)\left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right) \leq K_5^{\mathbf{L1}}(b + p^h + h)\left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right). \end{aligned}$$

For any given $p \in \mathcal{P}$, conditioning on the event $\mathcal{A}_1(p, \xi)$, we obtain

$$\tilde{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1} + \alpha - \beta p) \leq K_5^{\mathbf{L1}}(b + p^h + h)\left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right). \quad (3.65)$$

In addition, since $\tilde{y}_{i+1}(p, \alpha - \beta p)$ minimizes $\tilde{W}_{i+1}$, we have

$$\tilde{W}_{i+1}(p, \bar{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p)) \geq 0. \quad (3.66)$$

Thus, combining (3.65) and (3.66) with (3.60) yields

$$\begin{aligned} \mathbb{P}\left\{\left|\left(\tilde{W}_{i+1}(p, \bar{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p))\right)\right| \leq K_5^{\mathbf{L1}}(b + p^h + h)\left(\xi + \frac{K_6^{\mathbf{L1}}}{2L_i}\right)\right\} \\ \geq \mathbb{P}(\mathcal{A}_1(p, \xi)) \geq 1 - 2e^{-4L_i\xi^2}, \end{aligned}$$

which immediately implies (3.51). □

Proof of Lemma III.8. Define

$$B_{i+1}^1 = \frac{1}{L_i}\sum_{t=t_i+1}^{t_i+L_i} \tilde{\epsilon}_t, \qquad B_{i+1}^2 = \frac{1}{L_i}\sum_{t=t_i+L_i+1}^{t_i+2L_i} \tilde{\epsilon}_t.$$

Recall that $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$ are derived from the least-square method, and they are given by

$$\hat{\alpha}_{i+1} = \frac{\lambda(\hat{p}_i) + \lambda(\hat{p}_i + \delta_i)}{2} + \frac{B^1_{i+1} + B^2_{i+1}}{2} + \hat{\beta}_{i+1}\frac{2\hat{p}_i + \delta_i}{2}, \tag{3.67}$$

$$\hat{\beta}_{i+1} = -\frac{\lambda(\hat{p}_i + \delta_i) - \lambda(\hat{p}_i)}{\delta_i} - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}). \tag{3.68}$$

To measure the effectiveness of $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$ we define the (true) sample averages by

$$B^{1A}_{i+1} = \frac{1}{L_i}\sum_{t=t_i+1}^{t_i+L_i} \epsilon_t, \qquad B^{2A}_{i+1} = \frac{1}{L_i}\sum_{t=t_i+L_i+1}^{t_i+2L_i} \epsilon_t.$$

Let $\hat{\alpha}^A_{i+1}$ and $\hat{\beta}^A_{i+1}$ are derived from the least-square method, and they are given by

$$\hat{\alpha}^A_{i+1} = \frac{\lambda(\hat{p}_i) + \lambda(\hat{p}_i + \delta_i)}{2} + \frac{B^{1A}_{i+1} + B^{2A}_{i+1}}{2} + \hat{\beta}^A_{i+1}\frac{2\hat{p}_i + \delta_i}{2}, \tag{3.69}$$

$$\hat{\beta}^A_{i+1} = -\frac{\lambda(\hat{p}_i + \delta_i) - \lambda(\hat{p}_i)}{\delta_i} - \frac{1}{\delta_i}(-B^{1A}_{i+1} + B^{2A}_{i+1}). \tag{3.70}$$

Comparing (3.67) and (3.68) with (3.69) and (3.70), by (3.56), we have, for some constant $K_2^{\mathbf{L2}} > 0$,

$$\left|\hat{\alpha}_{i+1} - \hat{\alpha}^A_{i+1}\right| \le \frac{K_2^{\mathbf{L2}}}{L_i\delta_i}, \quad \left|\hat{\beta}_{i+1} - \hat{\beta}^A_{i+1}\right| \le \frac{K_2^{\mathbf{L2}}}{L_i\delta_i}. \tag{3.71}$$

Applying the Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the second order for (3.70), we obtain

$$\begin{aligned}\hat{\beta}^A_{i+1} &= -\left(\lambda'(\hat{p}_i) + \frac{1}{2}\lambda''(q_i)\delta_i\right) - \frac{1}{\delta_i}(-B^{1A}_{i+1} + B^{2A}_{i+1}) \\ &= \breve{\beta}(\hat{p}_i) - \frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^{1A}_{i+1} + B^{2A}_{i+1}),\end{aligned} \tag{3.72}$$

where $q_i \in [\hat{p}_i, \hat{p}_i + \delta_i]$. Substituting (3.72) into (3.69), and applying the Taylor's

expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the first order, we have, for $q_i' \in [\hat{p}_i, \hat{p}_i + \delta_i]$,

$$\begin{aligned}
\hat{\alpha}_{i+1}^A &= \lambda(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B_{i+1}^{1A} + B_{i+1}^{2A}}{2} - \lambda'(\hat{p}_i)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\
&\quad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B_{i+1}^{1A} + B_{i+1}^{2A})\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\
&= \breve{\alpha}(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B_{i+1}^{1A} + B_{i+1}^{2A}}{2} - \frac{1}{2}\lambda'(\hat{p}_i)\delta_i \\
&\quad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B_{i+1}^{1A} + B_{i+1}^{2A})\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right). \qquad (3.73)
\end{aligned}$$

Since the error terms ϵ_t are bounded, by the Hoeffding's inequality, we have that for any $\xi > 0$,

$$\mathbb{P}\left\{\left|-B_{i+1}^{1A}\right| > (u-l)\xi\right\} \le 2e^{-2L_i\xi^2}, \qquad \mathbb{P}\left\{\left|B_{i+1}^{2A}\right| > (u-l)\xi\right\} \le 2e^{-2L_i\xi^2}.$$

Hence, we have

$$\begin{aligned}
&\mathbb{P}\left\{\left|-B_{i+1}^{1A}\right| + \left|B_{i+1}^{2A}\right| > 2(u-l)\xi\right\} \\
&\qquad \le \mathbb{P}\left\{\left|-B_{i+1}^{1A}\right| > (u-l)\xi\right\} + \mathbb{P}\left\{\left|B_{i+1}^{2A}\right| > (u-l)\xi\right\} \le 4e^{-2L_i\xi^2},
\end{aligned}$$

which implies that

$$\mathbb{P}\left\{\left|-B_{i+1}^{1A} + B_{i+1}^{2A}\right| \le 2(u-l)\xi\right\} \ge \mathbb{P}\left\{\left|-B_{i+1}^{1A}\right| + \left|B_{i+1}^{2A}\right| \le 2(u-l)\xi\right\} \ge 1 - 4e^{-2L_i\xi^2}.$$

Similar argument shows

$$\mathbb{P}\left\{\left|B_{i+1}^{1A} + B_{i+1}^{2A}\right| \le 2(u-l)\xi\right\} \ge 1 - 4e^{-2L_i\xi^2}.$$

Since $\lambda'(\cdot)$ and $\lambda''(\cdot)$ are bounded and δ_i converges to 0, from (3.73) we conclude that there must exist a constant $K_3^{\mathbf{L2}}$ such that, on the event $\left|B_{i+1}^{1A} + B_{i+1}^{2A}\right| \le 2(u-l)\xi$

and $\left|-B_{i+1}^{1A}+B_{i+1}^{2A}\right| \leq 2(u-l)\xi$, it holds that

$$|\hat{\alpha}_{i+1}^{A}-\breve{\alpha}(\hat{p}_i)| \leq K_3^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\xi\right).$$

Therefore,

$$\begin{aligned} & \mathbb{P}\left\{\left|\hat{\alpha}_{i+1}^{A}-\breve{\alpha}(\hat{p}_i)\right| \leq K_3^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\xi\right)\right\} \qquad (3.74)\\ \geq \quad & \mathbb{P}\left\{\left|B_{i+1}^{1A}+B_{i+1}^{2A}\right| \leq 2(u-l)\xi, \left|-B_{i+1}^{1A}+B_{i+1}^{2A}\right| \leq 2(u-l)\xi\right\} \geq 1-8e^{-2L_i\xi^2}. \end{aligned}$$

By (3.72), we have

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}^{A}-\breve{\beta}(\hat{p}_i)\right| \leq K_4^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}\right)\right\} \geq 1-4e^{-2L_i\xi^2}. \qquad (3.75)$$

By (3.74) and (3.75), we have

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1}^{A}-\breve{\alpha}(\hat{p}_i+\delta_i)| \leq K_5^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\xi\right)\right\} \geq 1-8e^{-2L_i\xi^2}, \qquad (3.76)$$

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}^{A}-\breve{\beta}(\hat{p}_i+\delta_i)\right| \leq K_6^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}\right)\right\} \geq 1-4e^{-2L_i\xi^2}. \qquad (3.77)$$

Together with (3.71), we have

$$\begin{aligned} \mathbb{P}\Bigg\{ & |\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i)| \leq K_7^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\xi+\frac{1}{L_i\delta_i}\right), \left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right| \leq K_7^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\frac{1}{L_i\delta_i}\right),\\ & |\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i+\delta_i)| \leq K_7^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\xi+\frac{1}{L_i\delta_i}\right),\\ & \left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i+\delta_i)\right| \leq K_7^{\mathbf{L2}}\left(\delta_i+\frac{\xi}{\delta_i}+\frac{1}{L_i\delta_i}\right)\Bigg\}\\ & \leq 1-24e^{-2L_i\xi^2}. \qquad (3.78) \end{aligned}$$

To compare the two objective functions in Lemma III.8, we introduce a generalized problem called (Generalized-SAA) based on (Opt-SAA) and (Approx-SAA) as follows.

Given $p_t = \hat{p}_i$ for $t = t_i + 1, \ldots, t_i + L_i$ and $p_t = \hat{p}_i + \delta_i$ for $t = t_i + L_i + 1, \ldots, t_i + 2L_i$, the sales data for $t \in [t_i + 1, \ldots, t_i + 2L_i]$, and some given parameters $(\alpha_1, \beta_1), (\alpha_2, \beta_2)$, define the following two sets $\boldsymbol{\zeta}_i^1(\alpha_1, \beta_1) = (\zeta_t, t \in [t_i + 1, \ldots, t_i + L_i])$ and $\boldsymbol{\zeta}_i^2(\alpha_2, \beta_2) = (\zeta_t, t \in [t_i + L_i + 1, \ldots, t_i + 2L_i])$ with

$$\begin{aligned} \zeta_t &= d_t \wedge y_t - (\alpha_1 - \beta_1 p_t), \quad t \in [t_i + 1, \ldots, t_i + L_i], \\ \zeta_t &= d_t \wedge y_t - (\alpha_2 - \beta_2 p_t), \quad t \in [t_i + L_i + 1, \ldots, t_i + 2L_i]. \end{aligned}$$

We define a generalized function H_{i+1} by

$$\begin{aligned} &H_{i+1}\Big(p, \alpha_1 - \beta_1 p, \boldsymbol{\zeta}_i^1(\alpha_1, \beta_1), \boldsymbol{\zeta}_i^2(\alpha_2, \beta_2)\Big) \qquad \text{(Generalized-SAA)} \\ &= p(\alpha_1 - \beta_1 p) - \min_{y \in \mathcal{Y}} \left\{ \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \Big(h\big(y - (\alpha_1 - \beta_1 p + \zeta_t)\big)^+ + (b+p)\big(\alpha_1 - \beta_1 p + \zeta_t - y\big)^+ \Big) \right\}. \end{aligned}$$

Note that (Generalized-SAA) generalizes (Opt-SAA) and (Approx-SAA). To see this, by setting $\alpha_1 = \alpha_2 = \hat{\alpha}_{i+1}$ and $\beta_1 = \beta_2 = \hat{\beta}_{i+1}$, (Generalized-SAA) is reduced to (Opt-SAA), i.e.,

$$\hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right) = H_{i+1}\Big(p, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p, \boldsymbol{\zeta}_i^1(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \boldsymbol{\zeta}_i^2(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})\Big). \quad (3.79)$$

On the other hand, by setting $\alpha_1 = \breve{\alpha}(\hat{p}_i), \beta_1 = \breve{\beta}(\hat{p}_i), \alpha_2 = \breve{\alpha}(\hat{p}_i + \delta_i), \beta_2 = \breve{\beta}(\hat{p}_i + \delta_i)$, and using the fact that $\lambda(p) = \breve{\alpha}(p) - \breve{\beta}(p)p$, (Generalized-SAA) is reduced to (Approx-SAA), i.e.,

$$\begin{aligned} &\tilde{G}_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) \qquad (3.80) \\ &\qquad = H_{i+1}\Big(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p, \boldsymbol{\zeta}_i^1(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)), \boldsymbol{\zeta}_i^2(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i))\Big). \end{aligned}$$

Next, we see that $H_{i+1}\Big(p, \alpha_1 - \beta_1 p, \boldsymbol{\zeta}_i^1(\alpha_1, \beta_1), \boldsymbol{\zeta}_i^2(\alpha_2, \beta_2)\Big)$ is differentiable with respect to α_1, α_2 and β_1, β_2 with bounded first-order derivatives. In particular, there

exists a constant $K_8^{\mathbf{L2}} > 0$ such that for any $\alpha_1, \alpha_2, \alpha_1', \alpha_2'$ and $\beta_1, \beta_2, \beta_1', \beta_2'$, it holds that

$$\begin{aligned} & \left| H_{i+1}\Big(p, \alpha_1 - \beta_1 p, \boldsymbol{\zeta}_i^1(\alpha_1, \beta_1), \boldsymbol{\zeta}_i^2(\alpha_2, \beta_2)\Big) - H_{i+1}\Big(p, \alpha_1' - \beta_1' p, \boldsymbol{\zeta}_i^1(\alpha_1', \beta_1'), \boldsymbol{\zeta}_i^2(\alpha_2', \beta_2')\Big) \right| \\ \leq \quad & K_8^{\mathbf{L2}}\Big(|\alpha_1 - \alpha_1'| + |\beta_1 - \beta_1'| + |\alpha_2 - \alpha_2'| + |\beta_2 - \beta_2'|\Big). \end{aligned} \tag{3.81}$$

Now, by substituting (3.79) and (3.80) into (3.81), we see that the two objective functions $\tilde{G}_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right)$ and $\hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)$ differ only in their associated parameters. Consequently, Lemma III.8 follows from (3.78) by letting $\xi = L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}$. □

Proof of Proposition III.6. By Lemmas III.7 and III.8, we have that for any $p \in \mathcal{S}_{i+1}$,

$$\mathbb{P}\left\{\left|\bar{G}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)\right| \geq K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\right\} \leq 28 L_i^{-2},$$

which leads to

$$\begin{aligned} \mathbb{P}\left\{\max_{p \in \mathcal{S}_{i+1}}\left|\bar{G}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)\right| \geq K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\right\} \\ \leq 28 L_i^{-2}\left(\frac{p^h - p^l}{\delta_{i+1}}\right) \leq K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}}(\log L_i)^{-\frac{1}{4}}. \end{aligned}$$

Define event

$$\mathcal{A}_2 = \left\{\omega : \max_{p \in \mathcal{S}_{i+1}}\left|\bar{G}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)\right| < K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}\right\},$$

and we have that

$$\mathbb{P}(\mathcal{A}_2) > 1 - K_2^{\mathbf{P2}} L_i^{-\frac{7}{4}}(\log L_i)^{-\frac{1}{4}}.$$

Let $\bar{\hat{p}} \in \hat{\mathcal{S}}_{i+1}$ be the closest point on $\hat{\mathcal{S}}_{i+1}$ to $\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$ and

$$(\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)) - \bar{\hat{p}})(\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)) - \hat{p}_{i+1}) \geq 0,$$

where $\bar{\hat{p}}$ is chosen to be on the same side as $\hat{p}_{i+1}$ relative to $\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$ (see Figure 3.8).

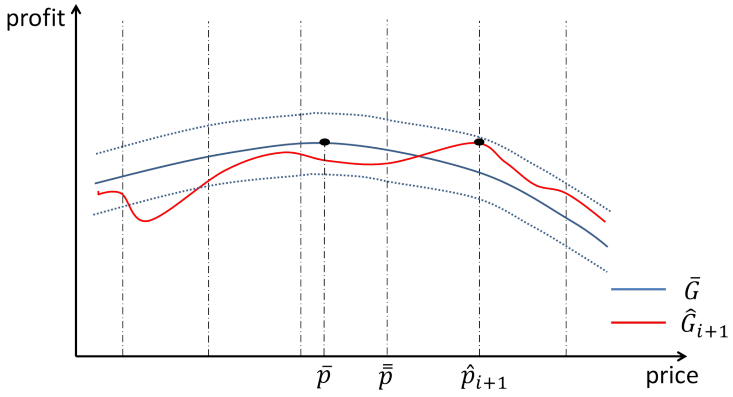

Figure 3.8: Choosing $\bar{\hat{p}}$ to be the closet point on the grid to $\bar{p}$, and also on the same side as $\hat{p}$ (relative to $\bar{p}$)

Applying the Taylor's expansion of $\bar{G}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1})$ at $\bar{\hat{p}}$, we obtain

$$\begin{aligned}
&\bar{G}(\bar{\hat{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\hat{p}}) - \bar{G}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) \\
=& -\bar{G}'(\bar{\hat{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\hat{p}})(\hat{p}_{i+1} - \bar{\hat{p}}) - \bar{G}''(q, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)q)(\hat{p}_{i+1} - \bar{\hat{p}})^2 \\
>& m(\hat{p}_{i+1} - \bar{\hat{p}})^2, \qquad (3.82)
\end{aligned}$$

where $m \triangleq \min_{q \in \mathcal{P}, \hat{p}_i \in \mathcal{P}} \left| \bar{G}''(q, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)q) \right| > 0$ due to strictly concavity of $\bar{G}(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p)$. The inequality follows also from the selection of $\bar{\hat{p}}$, which is chosen to be on the same side as $\hat{p}_{i+1}$ relative to $\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$. When $\hat{p}_{i+1} \geq \bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$, due to concavity of $\bar{G}(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p)$ in p, $\bar{G}'(\bar{\hat{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\hat{p}}) \leq 0$, and therefore

$$-\bar{G}'(\bar{\hat{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\hat{p}})(\hat{p}_{i+1} - \bar{\hat{p}}) \geq 0. \qquad (3.83)$$

When $\hat{p}_{i+1} < \bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i))$, (3.83) holds true by similar arguments.

On the other hand, conditioning on $\mathcal{A}_2$, we have

$$\begin{aligned}\bar{G}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) + K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} &> \hat{G}_{i+1}\left(\hat{p}_{i+1}, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right) \\ &\geq \hat{G}_{i+1}\left(\bar{\bar{p}}, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right) > \bar{G}(\bar{\bar{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) - K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}},\end{aligned}$$

where the first and the last inequalities follow from the definition of $\mathcal{A}_2$, and the second inequality holds because $\hat{p}_{i+1}$ is the maximizer for $\hat{G}_{i+1}\left(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}\right)$ on $\mathcal{S}_{i+1}$. Therefore,

$$\bar{G}(\bar{\bar{p}}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\bar{\bar{p}}) - \bar{G}(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) \leq 2K_3^{\mathbf{P2}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}. \quad (3.84)$$

Together with (3.82) and (3.84), one has

$$\left|\hat{p}_{i+1} - \bar{\bar{p}}\right| \leq K_4^{\mathbf{P2}} L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}},$$

which leads to, by conditioning on $\mathcal{A}_2$,

$$\begin{aligned}\left|\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)) - \hat{p}_{i+1}\right| &\leq \left|\bar{p}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)) - \bar{\bar{p}}\right| + \left|\bar{\bar{p}} - \hat{p}_{i+1}\right| \\ &\leq \delta_{i+1} + K_4^{\mathbf{P2}} L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}} \leq K_1^{\mathbf{P2}} L_i^{-\frac{1}{8}} (\log L_i)^{\frac{1}{8}}.\end{aligned}$$

This completes the proof of Proposition III.6. □

C: Technical Proofs for Theorem III.3(b)

Proof of Lemma III.11. For $p \in \mathcal{P}$, one has

$$\bar{y}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) = F^{-1}\left(\frac{b+p}{b+p+h}\right) + \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p. \quad (3.85)$$

For a given $p \in \mathcal{P}$, we define $\tilde{y}^u_{i+1}\big(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\big)$ as the unconstrained optimal target inventory level for (Approx-SAA) on $\mathbb{R}_+$, then it can be verified that

$$\begin{aligned}\tilde{y}^u_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) &= \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p \\ &+ \min\left\{\tilde{\epsilon}_j :\ j \in [t_i+1, \ldots, t_i+2L_i], \frac{\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\tilde{\epsilon}_t \le \tilde{\epsilon}_j\right\}}{2L_i} \ge \frac{b+p}{b+p+h}\right\}, \\ \tilde{y}_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) &= \min\left\{\max\left\{\tilde{y}^u_{i+1}(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p), y^l\right\} y^h\right\}.\end{aligned} \tag{3.86}$$

Then we have

$$\begin{aligned}&\left|\bar{y}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \tilde{y}_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right)\right| \\ \le\ &\left|\bar{y}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \tilde{y}^u_{i+1}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right)\right|.\end{aligned} \tag{3.87}$$

There exists a positive constant $K_2^{\mathbf{L3}}$ such that, by (3.56), for any $\xi > 0$

$$\begin{aligned}&\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\tilde{\epsilon}_t \le F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} \\ \le\ &\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\epsilon_t \le F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} + \frac{K_2^{\mathbf{L3}}}{2L_i}.\end{aligned} \tag{3.88}$$

Now, for any $\xi \geq K_2^{\mathbf{L3}}/L_i$, we have

$$\begin{aligned}
& \mathbb{P}\left\{F\left(\tilde{y}_{i+1}^u\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p)\right) - \frac{b+p}{b+p+h} \leq -\xi\right\} \qquad (3.89)\\
= \; & \mathbb{P}\left\{\tilde{y}_{i+1}^u\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p) \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\}\\
\leq \; & \mathbb{P}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\tilde{\epsilon}_t \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} \geq \frac{b+p}{b+p+h}\right\}\\
= \; & \mathbb{P}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\tilde{\epsilon}_t \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} - \left(\frac{b+p}{b+p+h} - \xi\right) \geq \xi\right\}\\
\leq \; & \mathbb{P}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\epsilon_t \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} + \frac{K_2^{\mathbf{L3}}}{2L_i} - \left(\frac{b+p}{b+p+h} - \xi\right) \geq \xi\right\}\\
\leq \; & \mathbb{P}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\epsilon_t \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} - \left(\frac{b+p}{b+p+h} - \xi\right) \geq \frac{\xi}{2}\right\},
\end{aligned}$$

where the first inequality follows from (3.86), the second inequality holds from using (3.88), and the last inequality holds as $\xi \geq K_2^{\mathbf{L3}}/L_i$, $\xi - K_2^{\mathbf{L3}}/(2L_i) \geq \frac{\xi}{2}$.

Since $\mathbb{E}\left[\mathbb{1}\left\{\epsilon_t \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\}\right] = \frac{b+p}{b+p+h} - \xi$, we apply the Hoeffding's inequality to obtain

$$\mathbb{P}\left\{\frac{1}{2L_i}\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\left\{\epsilon_t \leq F^{-1}\left(\frac{b+p}{b+p+h} - \xi\right)\right\} - \left(\frac{b+p}{b+p+h} - \xi\right) \geq \frac{\xi}{2}\right\} \leq e^{-L_i\xi^2}.$$

Together with (3.85) and (3.89), we have

$$\begin{aligned}
\mathbb{P}\Bigg\{ & F\left(\tilde{y}_{i+1}^u\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p)\right)\\
& -F\left(\bar{y}\left(p, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p)\right) \leq -\xi\Bigg\} \leq e^{-L_i\xi^2}. \qquad (3.90)
\end{aligned}$$

Similarly, we have

$$\mathbb{P}\Bigg\{ F\bigg(\tilde{y}^u_{i+1}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg) \\ -F\bigg(\bar{y}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg)\geq\xi\Bigg\}\leq e^{-L_i\xi^2}. \quad (3.91)$$

We have assumed the probability density function $f(\cdot)$ of ϵ_t satisfies $r=\min\{f(x), x\in[l,u]\}>0$. Then, for any $x<y$, there exists a number $z\in[x,y]$ such that $F(y)-F(x)=f(z)(y-x)\geq r(y-x)$. Applying (3.90) and (3.91), for any $\xi\geq K_2^{\mathbf{L3}}/L_i$, we obtain

$$\begin{aligned}
2e^{-L_i\xi^2} &\geq \mathbb{P}\Bigg\{\bigg|F\bigg(\tilde{y}^u_{i+1}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg) \\
&\qquad -F\bigg(\bar{y}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg)\bigg|\geq\xi\Bigg\} \\
&\geq \mathbb{P}\Bigg\{\bigg|F\bigg(\tilde{y}_{i+1}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg) \\
&\quad -F\bigg(\bar{y}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg)\bigg|\geq\xi\Bigg\} \\
&\geq \mathbb{P}\Bigg\{r\bigg|\tilde{y}_{i+1}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p) \\
&\qquad -\bar{y}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p)\bigg|\geq\xi\Bigg\} \\
&= \mathbb{P}\Bigg\{\bigg|\tilde{y}_{i+1}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-\bar{y}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)\bigg|\geq\frac{1}{r}\xi\Bigg\},
\end{aligned}$$

where the second inequality follows from (3.87). For constant $K_1^{\mathbf{L3}}=\frac{1}{r}$ we have, for $\xi\geq K_2^{\mathbf{L3}}/L_i$,

$$\mathbb{P}\Bigg\{\bigg|\tilde{y}_{i+1}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)-\bar{y}\Big(p,\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p\Big)\bigg|\geq K_1^{\mathbf{L3}}\xi\Bigg\}\leq 2e^{-L_i\xi^2}, \quad (3.92)$$

By letting $\xi=L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}$, Lemma III.11 follows from (3.92). □

Proof of Lemma III.12. Following the analysis of Lemma III.8, consider the

optimization of H_{i+1} in (Generalized-SAA), the inner optimization problem is convex in y, and therefore for given price p, we denote the optimal order-up-to level by (with ζ_t's defined in the proof of Lemma III.8)

$$\begin{aligned} & y_{i+1}\big(p, \alpha_1 - \beta_1 p, \boldsymbol{\zeta}_i^1(\alpha_1, \beta_1), \boldsymbol{\zeta}_i^2(\alpha_2, \beta_2)\big) \qquad (3.93) \\ = \ & \arg\min_{y \in \mathcal{Y}} \Bigg\{ \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \Big(h\big(y - (\alpha_1 - \beta_1 p + \zeta_t)\big)^+ + (b+p)\big(\alpha_1 - \beta_1 p + \zeta_t - y\big)^+ \Big) \Bigg\}. \end{aligned}$$

We see that $y_{i+1}\big(p, \alpha_1 - \beta_1 p, \boldsymbol{\zeta}_i^1(\alpha_1, \beta_1), \boldsymbol{\zeta}_i^2(\alpha_2, \beta_2)\big)$ are differentiable with respect to α_1, α_2 and β_1, β_2 with bounded first-order derivatives. In particular, there exists a constant $K_2^{\mathbf{L4}} > 0$ such that for any $\alpha_1, \alpha_2, \alpha_1', \alpha_2'$ and $\beta_1, \beta_2, \beta_1', \beta_2'$, it holds that

$$\begin{aligned} & \Big| y_{i+1}\big(p, \alpha_1 - \beta_1 p, \boldsymbol{\zeta}_i^1(\alpha_1, \beta_1), \boldsymbol{\zeta}_i^2(\alpha_2, \beta_2)\big) - y_{i+1}\big(p, \alpha_1' - \beta_1' p, \boldsymbol{\zeta}_i^1(\alpha_1', \beta_1'), \boldsymbol{\zeta}_i^2(\alpha_2', \beta_2')\big) \Big| \\ \leq \ & K_2^{\mathbf{L4}} \Big(|\alpha_1 - \alpha_1'| + |\beta_1 - \beta_1'| + |\alpha_2 - \alpha_2'| + |\beta_2 - \beta_2'| \Big). \qquad (3.94) \end{aligned}$$

In addition, we know that

$$\begin{aligned} & \hat{y}_{i+1,1} = y_{i+1}\big(\hat{p}_{i+1}, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}\hat{p}_{i+1}, \boldsymbol{\zeta}_i^1(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \boldsymbol{\zeta}_i^2(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})\big), \qquad (3.95) \\ & \tilde{y}_{i+1}\big(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big) \qquad (3.96) \\ & \qquad = y_{i+1}\big(\hat{p}_{i+1}, \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}, \boldsymbol{\zeta}_i^1\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big), \boldsymbol{\zeta}_i^2\big(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i)\big)\big). \end{aligned}$$

Thus, Lemma III.12 follows by combining (3.78), (3.94), (3.95), and (3.96) and letting $\xi = L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}$. □

Proof of Proposition III.10. From (3.38),

$$\begin{aligned}
&\mathbb{E}\left[\left|\bar{y}(\hat{p}_{i+1},\lambda(\hat{p}_{i+1}))-\hat{y}_{i+1,1}\right|^2\right]\\
\leq\ & K_2^{\mathbf{P3}}\mathbb{E}\Big[\left|\hat{p}_{i+1}-p^*\right|^2+\left|\hat{p}_i-p^*\right|^2\Big]\\
&+K_2^{\mathbf{P3}}\mathbb{E}\Big[\Big|\bar{y}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})-\tilde{y}_{i+1}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\Big|^2\Big]\\
&+K_2^{\mathbf{P3}}\mathbb{E}\Big[\Big|\tilde{y}_{i+1}(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})-\hat{y}_{i+1,1}\Big|^2\Big]. \qquad (3.97)
\end{aligned}$$

By Theorem III.3(a), we have

$$\mathbb{E}\left[\left|\hat{p}_{i+1}-p^*\right|^2+\left|\hat{p}_i-p^*\right|^2\right]\leq K_3^{\mathbf{P3}}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}}. \qquad (3.98)$$

And it follows from Lemma III.11 that

$$\begin{aligned}
&\mathbb{E}\left[\left|\bar{y}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)-\tilde{y}_{i+1}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)\right|^2\right] \qquad (3.99)\\
\leq\ & \frac{K_4^{\mathbf{P3}}}{L_i}+K_5^{\mathbf{P3}}L_i^{-1}\log L_i\leq K_6^{\mathbf{P3}}L_i^{-1}\log L_i. \qquad (3.100)
\end{aligned}$$

In Lemma III.12,

$$\begin{aligned}
&\mathbb{E}\left[\left|\bar{y}\big(\hat{p}_{i+1},\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)-\hat{y}_{i+1,1}\right|^2\right]\\
&\leq K_7^{\mathbf{P3}}L_i^{-2}+K_8^{\mathbf{P3}}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}\leq K_9^{\mathbf{P3}}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}. \qquad (3.101)
\end{aligned}$$

Proposition III.10 follows from (3.98), (3.99), and (3.101). □

D: Technical Proofs for Theorem III.4

Proof of Proposition III.13. To prove Proposition III.13, we proceed as follows. Using the fact that $y^* = \bar{y}\big(p^*, \lambda(p^*)\big)$, we have

$$\begin{aligned} & \mathbb{E}\left[Q\big(p^*, \bar{y}\big(p^*, \lambda(p^*)\big)\big) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\right] \\ \leq \quad & \mathbb{E}\Big[Q\big(p^*, \bar{y}\big(p^*, \lambda(p^*)\big)\big) - Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big) \\ & \qquad + Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\Big]. \end{aligned}$$

The first term $Q\big(p^*, \bar{y}\big(p^*, \lambda(p^*)\big)\big) - Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big)$ is bounded using Taylor expansion on $Q(p, \bar{y}(p, \lambda(p))$ at point p^*. Using the fact that the first order derivative vanishes at $p = p^*$ and bounded second-order derivative, we obtain, for some constant $K_2^{\mathbf{P4}} > 0$, that

$$\begin{aligned} & \mathbb{E}\left[Q\big(p^*, \bar{y}\big(p^*, \lambda(p^*)\big)\big) - Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big)\right] \leq \mathbb{E}[K_2^{\mathbf{P4}}(p^* - \hat{p}_{i+1})^2] \\ \leq \quad & K_3^{\mathbf{P4}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \leq K_4^{\mathbf{P4}} I_i^{-\frac{1}{5}} (\log I_i)^{\frac{1}{4}}. \end{aligned} \tag{3.102}$$

To bound the second term $Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})$, notice that $\bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)$ maximizes the concave function $Q\big(\hat{p}_{i+1}, y\big)$ for any given $\hat{p}_{i+1}$, we apply Taylor expansion with respect to y at point $y = \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)$ to yield that, for some constant $K_5^{\mathbf{P4}}$,

$$Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) \leq K_5^{\mathbf{P4}} \big(\bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big) - \hat{y}_{i+1,1}\big)^2, \tag{3.103}$$

which leads to

$$\begin{aligned} \mathbb{E}\left[Q\big(\hat{p}_{i+1}, \bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big)\big) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\right] \quad & \leq \quad K_5^{\mathbf{P4}} \mathbb{E}\left[\big(\bar{y}\big(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})\big) - \hat{y}_{i+1,1}\big)^2\right] \\ & \leq \quad K_6^{\mathbf{P4}} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}, \end{aligned} \tag{3.104}$$

where the second inequality follows from Proposition III.10.

By (3.102) and (3.104), we have

$$\begin{aligned} & \mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+2L_i+1}^{t_i+I_i}\left(Q\left(p^*,y^*\right)\right)-Q(\hat{p}_{i+1},\hat{y}_{i+1,1})\right)\right] \\ \leq \quad & \mathbb{E}\left[\sum_{i=1}^{n} I_i^{-\frac{1}{5}}(\log I_i)^{\frac{1}{4}} I_i\right] \leq K_7^{\mathbf{P4}} T^{\frac{4}{5}}(\log T)^{\frac{1}{4}}. \end{aligned}$$

This completes of the proof of Proposition III.13. □

Proof of Proposition III.14. Consider the accumulative demands during periods t_i+2L_i+1 to $t_i+2L_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor$. If these accumulative demands consume at least $x_{t_i+2L_i+1}-\hat{y}_{i+1}$, then at period $t_i+2L_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor+1$, $\hat{y}_{i+1}$ will be surely achieved. Since $\lambda(p^h)+l\leq D_t\leq\lambda(p^l)+u$ for all t, by Hoeffding inequality, for any $\zeta>0$,

$$\mathbb{P}\left\{\sum_{t=t_i+2L_i+1}^{t_i+2L_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t\geq\mathbb{E}\left[\sum_{t=t_i+2L_i+1}^{t_i+2L_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t\right]-\zeta\right\}\geq 1-\exp\left(-\frac{2\zeta^2}{\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\left(\lambda(p^l)+u-\lambda(p^h)-l\right)^2}\right). \tag{3.105}$$

Now choose $\zeta=\left(\lambda(p^l)+u-\lambda(p^h)-l\right)\left(\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}$, and define the event

$$\mathcal{A}_3=\left\{\sum_{t=t_i+2L_i+1}^{t_i+2L_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t\geq\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\mathbb{E}\left[D_{t_i+2L_i+1}\right]-\zeta\right\}. \tag{3.106}$$

Then it follows from (3.105) that $\mathbb{P}(\mathcal{A}_3)\geq 1-\left\lfloor I_i^{\frac{1}{2}}\right\rfloor^{-2}$.

Since $\lambda(p^h)>0$, $\mathbb{E}\left[D_{t_i+2L_i+1}\right]>0$. Then there exists some i^{**} such that whenever $i\geq i^{**}$,

$$\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\mathbb{E}\left[D_{t_i+2L_i+1}\right]-\zeta\geq\frac{1}{2}\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\mathbb{E}\left[D_{t_i+2L_i+1}\right]\geq y^h-y^l\geq x_{t_i+2L_i+1}-\hat{y}_i,$$

which suggests that on the event $\mathcal{A}_3$, the order-up-to level $\hat{y}_i$ will always be achieved after periods

$\left\{t_i + 2L_i + 1, \ldots, t_i + 2L_i + \left\lfloor I_i^{\frac{1}{2}} \right\rfloor\right\}$.

Now using $\mathbb{P}(\mathcal{A}_3^c) \leq \left\lfloor I_i^{1/2} \right\rfloor^{-2}$, we have

$$\begin{aligned}
& \mathbb{E}\left[\sum_{t=t_i+2L_i+1}^{t_i+I_i} \big(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\big)\right] \\
= \quad & \mathbb{P}(\mathcal{A}_3)\mathbb{E}\left[\sum_{t=t_i+2L_i+1}^{t_i+I_i} \big(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\big)\Big|\mathcal{A}_3\right] \\
& +\mathbb{P}(\mathcal{A}_3^c)\mathbb{E}\left[\sum_{t=t_i+2L_i+1}^{t_i+I_i} \big(G(\hat{p}_i, \hat{y}_{i,1}) - G(p_t, y_t)\big)\Big|\mathcal{A}_3^c\right] \\
\leq \quad & \max\{h, b+p^h\}(y^h - y^l)\left\lfloor I_i^{\frac{1}{2}} \right\rfloor + \left\lfloor I_i^{\frac{1}{2}} \right\rfloor^{-2} \max\{h, b+p^h\}(y^h - y^l) I_i \\
\leq \quad & 2\max\{h, b+p^h\}(y^h - y^l) I_i^{\frac{1}{2}}.
\end{aligned}$$

Then the result of Proposition III.14 follows. □

E: Nonparametric Algorithm for the Observable Demand Case

The algorithm starts with initial parameters $\{\hat{p}_1, \hat{y}_{1,1}, \hat{y}_{1,2}\}$, and three positive numbers, I_0, v and ρ, where $I_0 > 0$, $v > 1$, $\rho > 0$ and $\hat{p}_1 \in \mathcal{P}, \hat{y}_{1,1} \in \mathcal{Y}, \hat{y}_{1,2} \in \mathcal{Y}$. Let $I_1 = \lfloor I_0 v \rfloor$ and the first stage consists of $2I_1$ periods. For the first I_1 periods of stage 1 ($t = 1, \ldots, I_1$), the firm sets the price $\hat{p}_1$ and the order-up-to level $\hat{y}_{1,1}$; for the second I_1 periods of stage 1 ($t = I_1 + 1, \ldots, 2I_1$), the firm perturbs the price $\hat{p}_1$ by a small δ_1 amount (i.e., $\hat{p}_1 + \delta_1$), where $\delta_1 = \rho(2I_0)^{-\frac{1}{4}}(\log(2I_0))^{\frac{1}{4}}$. Then the firm re-sets the price $\hat{p}_1 + \delta_1$ and the order-up-to level $\hat{y}_{1,2}$. In general, for each learning stage $i \geq 1$,

$$I_i = \lfloor I_0 v^i \rfloor, \quad \delta_i = \rho(2I_{i-1})^{-\frac{1}{4}}(\log(2I_{i-1}))^{\frac{1}{4}}, \text{ and } t_i = \sum_{k=1}^{i-1} 2I_k. \tag{3.107}$$

Then, stage $i > 1$ starts in period $t_i + 1$, and at the beginning of stage i, the algorithm proceeds with $\{\hat{p}_i, \hat{y}_{i,1}, \hat{y}_{i,2}\}$ that are derived in the preceding stage $i - 1$. Define set $\mathcal{S}_i = [p^l, p^l + \delta_i, p^l + 2\delta_i, \ldots, p^h]$.

Now we present the learning algorithm DDO for the observable demand case.

Step 0: Preparation. Initialize I_0, v and ρ, and $\hat{p}_1$, $\hat{y}_{1,1}$, $\hat{y}_{1,2}$, δ_1, I_1 as shown above.

For each stage $i = 1, \ldots, n$ where $n = \left\lceil \log_v \left(\frac{v-1}{2I_0 v} T + 1 \right) \right\rceil$, repeat the three steps below:

Step 1: Setting prices and order-up-to level for stage i.

Set prices p_t, $t = t_i + 1, \ldots, t_i + 2I_i$, as follows,

$$p_t = \hat{p}_i, \text{ for all } t = t_i + 1, \ldots, t_i + I_i,$$
$$p_t = \hat{p}_i + \delta_i, \text{ for all } t = t_i + I_i + 1, \ldots, t_i + 2I_i;$$

and for $t = t_i + 1, \ldots, t_i + 2I_i$, raise the inventory level in period t to y_t as follows,

$$y_t = \hat{y}_{i,1} \vee x_t, \text{ for all } t = t_i + 1, \ldots, t_i + I_i,$$
$$y_t = \hat{y}_{i,2} \vee x_t, \text{ for all } t = t_i + I_i + 1, \ldots, t_i + 2I_i.$$

Step 2: Estimating the demand-price function and the error term.

Let d_t be demand realizations for $t = t_i + 1, \cdots, t_i + 2I_i$. Solve a least-square

problem

$$(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = \operatorname{argmin} \left\{ \sum_{t \in [t_i+1, t_i+2I_i]} \left[d_t - (\alpha - \beta p_t) \right]^2 \right\}, \text{ and} \tag{3.108}$$

$$\eta_t = d_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t) = \lambda(p_t) + \epsilon_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t) \text{ for } t = t_i + 1, \ldots, t_i + 2I_i. \tag{3.109}$$

Step 3: Maximize the proxy profit $Q^{SAA}(p, y)$.

We define the following sampled optimization problem (Opt-SAA-O).

$$\max_{(p,y) \in \mathcal{S}_{i+1} \times \mathcal{Y}} Q_{i+1}^{SAA}(p, y) \triangleq \max_{p \in \mathcal{S}_{i+1}} \left\{ \hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \right\}, \quad \text{(Opt-SAA-O)}$$

$$\begin{aligned} \text{where} \quad \hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) &\triangleq p \left(\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p \right) \\ &- \min_{y \in \mathcal{Y}} \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big\{ (b+p) \left(\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p + \eta_t - y \right)^+ \\ &\qquad + h \left(y - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p + \eta_t) \right)^+ \Big\}. \end{aligned}$$

Then, set the first pair of price and order-up-to level to

$$(\hat{p}_{i+1}, \hat{y}_{i+1,1}) = \arg \max_{(p,y) \in \mathcal{S}_{i+1} \times \mathcal{Y}} Q_{i+1}^{SAA}(p, y),$$

and compute $\hat{y}_{i+1,2}$ as

$$\hat{y}_{i+1,2} = \arg \max_{y \in \mathcal{Y}} Q_{i+1}^{SAA}(\hat{p}_{i+1} + \delta_{i+1}, y).$$

In case that $\hat{p}_{i+1} + \delta_{i+1} \notin \mathcal{S}_{i+1}$, set the second price to $\hat{p}_{i+1} - \delta_{i+1}$.

F: Analysis of Unbounded Demand

Under Assumption (a) of light tailed demand, there exist positive constants $K_2^{\mathbf{U}}, K_3^{\mathbf{U}}, K_4^{\mathbf{U}}$ such that $\mathbb{P}\{D_t(p^l) \geq x\} \leq K_2^{\mathbf{U}} \exp(-K_3^{\mathbf{U}} x)$ for any $x \geq K_4^{\mathbf{U}}$.

We assume the inventory search space ceiling y^h is at least $1/(2K_3^{\mathbf{U}}) \log T$ for the problem with planning horizon T. In our algorithm we explore the inventory space up to $1/(2K_3^{\mathbf{U}}) \log T$. For convenience we set $y^h = 1/(2K_3^{\mathbf{U}}) \log T$.

Most of the analyses for the unbounded demand case follow similar lines of arguments as those for the bounded demand case. The major difference lies in the estimation of average demand using sales data, which is the focus of our analysis below. To analyze the regret, we need to compute the error of estimation had the complete demand-price information been observed, and then study the impact of truncated demand data. The former follows from the Chebyshev's inequality and Assumption (a), that for some positive constant $K_5^{\mathbf{U}}$,

$$\mathbb{P}\left(-\sigma_1 L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \leq \frac{\sum_{t=t_i+1}^{t_i+L_i} D_t(p)}{L_i} - \mathbb{E}[D_t(p)] \leq \sigma_1 L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}\right) \geq 1 - \frac{K_5^{\mathbf{U}}}{L_i}, \tag{3.110}$$

where σ_1 is the standard deviation of $D_t(p)$, and similarly we have

$$\mathbb{P}\left(\frac{\sum_{t=t_i+1}^{t_i+L_i} (D_t(p) - y^h)^+}{L_i} - \mathbb{E}[(D_t(p) - y^h)^+] \leq \sigma_2 L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}\right) \geq 1 - \frac{K_6^{\mathbf{U}}}{L_i}, \tag{3.111}$$

where σ_2 is the standard deviation of $(D_t(p) - y^h)^+$. We first show that when T is large enough (because y^h grows linearly in $\log T$), we have $2\sigma_2 < \sigma_1$. To that end, we shall prove that $\sigma_2^2 = Var[(D_t(p) - y^h)^+]$ is continuous and decreasing in y^h from $\sigma_1^2 = Var[D_t(p)]$ to 0 as $y^h \to \infty$. Since y^h increases linearly in $\log T$, this implies that when T is large enough, we will have $\sigma_2 < \sigma_1/2$.

Let $F_{D(p)}$ and $f_{D(p)}$ denote the cdf and pdf of $D_t(p)$, respectively. Then

$$\sigma_2^2 = \int_{y^h}^{\infty} (x - y^h)^2 f_{D(p)}(x)dx - \Big(\int_{y^h}^{\infty} (x - y^h) f_{D(p)}(x)dx \Big)^2, \text{ and}$$

$$\frac{\partial \sigma_2^2}{\partial y^h} = -2F_{D(p)}(y^h) \int_{y^h}^{\infty} (x - y^h) f_{D(p)}(x)dx \leq 0,$$

and σ_2^2 is decreasing in y^h.

It is clear that σ_2^2 is continuous in y^h and $\sigma_2^2 = \sigma_1^2$ when $y^h = 0$. Furthermore, it follows from the monotone convergence theorem that $\sigma_2^2 \to 0$ as $y^h \to \infty$. Therefore, σ_2^2 is continuously decreasing in y^h from σ_1^2 to 0 as y^h goes from 0 to infinity.

To bound $\frac{1}{L_i} \sum_{t=t_i+1}^{t_i+L_i} \min\{D_t(p), y^h\} - \mathbb{E}[D_t(p)]$, we notice that $\min\{D_t(p), y^h\} = D_t(p) - (D_t(p) - y^h)^+$, and by Assumption (a), we have

$$\mathbb{E}[(D_t(p) - y^h)^+\}] = \int_0^{+\infty} \mathbb{P}(D_t(p) - y^h \geq x)dx \leq \int_0^{+\infty} K_2^{\mathbf{U}} e^{-K_3^{\mathbf{U}}(x+y^h)} dx = \frac{K_2^{\mathbf{U}}}{K_3^{\mathbf{U}}} e^{-K_3^{\mathbf{U}} y^h}.$$

Hence, by our choice of y^h, $\frac{K_2^{\mathbf{U}}}{K_3^{\mathbf{U}}} e^{-K_3^{\mathbf{U}} y^h} = \frac{K_2^{\mathbf{U}}}{K_3^{\mathbf{U}}} T^{-\frac{1}{2}} \leq \frac{K_2^{\mathbf{U}}}{K_3^{\mathbf{U}}} L_i^{-\frac{1}{2}}$ when T is large enough. Consequently for large enough i, we have

$$\mathbb{E}[(D_t(p) - y^h)^+] \leq \frac{K_2^{\mathbf{U}}}{K_3^{\mathbf{U}}} L_i^{-\frac{1}{2}} \leq \sigma_2 (L_i)^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}. \tag{3.112}$$

Combining (3.111) and (3.112) yields

$$\mathbb{P}\left(\frac{\sum_{t=t_i+1}^{t_i+L_i} (D_t(p) - y^h)^+}{L_i} \leq 2\sigma_2 (L_i)^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \right) \geq 1 - \frac{K_6^{\mathbf{U}}}{L_i}. \tag{3.113}$$

Let $\mathcal{A}_4$ be the event for (3.110), and $\mathcal{A}_5$ be the event for (3.113). Writing (3.110)

as

$$-\sigma_1(L_i)^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} - \frac{\sum_{t=t_i+1}^{t_i+L_i}(D_t(p)-y^h)^+}{L_i} \leq \frac{\sum_{t=t_i+1}^{t_i+L_i}\min\{D_t(p),y^h\}}{L_i} - \mathbb{E}[D_t(p)]$$
$$\leq \sigma_1(L_i)^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} - \frac{\sum_{t=t_i+1}^{t_i+L_i}(D_t(p)-y^h)^+}{L_i}. \tag{3.114}$$

Then it can be seen that on the event $\mathcal{A}_4 \cap \mathcal{A}_5$, it holds that

$$(-\sigma_1 - 2\sigma_2)(L_i)^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} \leq \frac{\sum_{t=t_i+1}^{t_i+L_i}\min\{D_t(p),y^h\}}{L_i} - \mathbb{E}[D_t(p)] \leq \sigma_1(L_i)^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}.$$

Thus by $-2\sigma_2 \geq -\sigma_1$, it follows from (3.110) and (3.113) that

$$\mathbb{P}\left(-2\sigma_1(L_i)^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} \leq \frac{\sum_{t=t_i+1}^{t_i+L_i}\min\{D_t(p),y^h\}}{L_i} - \mathbb{E}[D_t(p)] \leq \sigma_1(L_i)^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}\right)$$
$$\geq \mathbb{P}\left(\mathcal{A}_4 \cap \mathcal{A}_5\right) = 1 - \mathbb{P}\left(\mathcal{A}_4^c \cap \mathcal{A}_5^c\right) \geq 1 - \frac{K_5^{\mathbf{U}} + K_6^{\mathbf{U}}}{L_i}. \tag{3.115}$$

Comparing this result with (3.110) reveals that, estimating the average demand using sales data during the exploration phase leads to an error very similar to that using true demand data. This estimation error determines the regret from the exploitation phase, and it shows that the regret from the exploitation phase using sales data is similar in format to that in the bounded demand case. However, we note that in the proof of Theorem III.4, the regret is amplified by $y^h - y^l = \mathcal{O}(1)$ for the bounded demand case, whereas the regret is amplified by $y^h - y^l = \mathcal{O}(\log T)$ for the unbounded demand case, resulting in a total regret of $\mathcal{O}(T^{-\frac{1}{5}}\log T)$.

CHAPTER IV

Data-Driven Dynamic Pricing and Inventory Control with Censored Demand and Limited Price Changes

4.1 Introduction

Information about the demand distribution and its dependency on selling prices is critical for making pricing and inventory decisions. However, in many applications, such information is not known *a priori*, and the firm needs to learn through price experimentation. This is usually done through exploration and exploitation. In the exploration phase, the firm uses different price points to collect demand data, and then uses the obtained information to make decisions for implementation during the exploitation phase. When demand is censored, the firm can only observe the demand realization up to the inventory level, and any unsatisfied demand is lost and unobserved. This leads to partial and incomplete demand information. For a finite planning-horizon problem, there is a trade-off between these two phases: the longer the exploration phase, the more demand information the firm can extract, but the shorter the remaining time for exploitation to maximize profit. Thus, designing effective learning algorithms has been of great interest in the recent literature. See e.g., *Burnetas and Smith* (2000), *Besbes and Zeevi* (2009), *Broder and Rusmevichientong*

(2012), and *Besbes and Zeevi* (2015), to name just a few.

One common observation in practice is that a firm may be constrained from making frequent price changes. *Cheung et al.* (2015) discuss several practical reasons in price experimentations that forbid frequent price changes, such as customer's negative responses to frequent price changes (e.g., that may cause confusion and affect seller's brand reputation). One economic reason for not having frequent price change is the cost for making such changes (e.g., due to changing price labels, etc.), hence by limiting the number of price changes the firm would control the associated cost. Clearly, such constraint brings enormous additional complexity in demand learning.

In this chapter we consider a dynamic joint pricing and inventory replenishment problem over a finite planning horizon, where the firm has little prior knowledge about the demand distribution and needs to learn it through historical censored demand data. The firm needs to determine its inventory replenishment and pricing decisions in each period, subject to some constraint on the number of price changes, and the objective is to maximize total expected profit. We consider a setting where the demand distribution on an offered price is drawn from a family of distributions with unknown continuous parameters of dimension k. We develop data-driven algorithms to compute pricing and inventory replenishment decisions for various constraints on the number of price changes, and evaluate their performances by regret, which is the total profit loss compared to a clairvoyant who has complete information about the demand distribution and can change its selling prices as many times as it wishes. Three scenarios are considered: First, for a general case, when the number of price changes is limited to k, the regret is $\mathcal{O}(T^{1/2})$. Second, for a so-called well-separated case, when the number of price changes is limited to m, an arbitrarily given positive integer, the regret is $\mathcal{O}(T^{1/(m+1)})$. Third, also for the well-separated case, when the number of price changes is in the order of $\mathcal{O}(\log T)$, the regret is $\mathcal{O}(\log T)$. We further show that these bounds are the best possible in the sense that, they have the same

order of magnitude as the lower bound of regret for any learning algorithms of these problems. We also conduct numerical studies and show that these learning algorithms empirically perform very well.

Comparisons with the Literature. This chapter is related to the research literature dealing with limited demand information in stochastic inventory control, revenue management, and joint pricing and inventory control problems. For each category, the research literature is classified as either parametric (i.e., the firm knows the family of demand distribution but not the parameters of the distribution) or nonparametric approach. Our work falls into the parametric category of joint optimization of pricing and inventory control. Thus, in the following we briefly review related works on inventory and pricing using parametric demand estimation.

Earlier research papers on stochastic inventory control with parameter estimation include *Scarf* (1959, 1960), *Murray and Silver* (1966), *Azoury* (1985), *Lovejoy* (1990) for completely observable demand data; *Ding et al.* (2002), and *Lariviere and Porteus* (1999) for censored demand; and *Chen and Plambeck* (2008) for the case with multiple products. In these papers, price is static and exogenous, thus the firm is only concerned with inventory replenishment decisions. In the revenue management literature, early papers such as *Kalish* (1983), and *Gallego and van Ryzin* (1994) consider a firm's pricing problem when the firm has complete information about the underlying demand process. These papers have been extended to the parametric settings by, e.g., *Araman and Caldentey* (2009), *Aviv and Pazgal* (2005), *Broder and Rusmevichientong* (2012), *Carvalho and Puterman* (2005), *Farias and van Roy* (2010), *den Boer and Zwart* (2015), *Harrison et al.* (2012), and *Keskin and Zeevi* (2014), among others. *Cheung et al.* (2015) develop learning algorithms for a pricing problem with constraint on the number of price changes.

There have been numerous studies in the literature on joint pricing and inventory decisions. As in the above two categories, earlier papers in this area, including

Whitin (1955), *Karlin and Carr* (1962), *Thowsen* (1975), *Federgruen and Heching* (1999), and *Chen and Simchi-Levi* (2004b), assume that the firm has complete information about demand distribution. Refer to survey papers by *Chan et al.* (2004), *Elmaghraby and Keskinocak* (2003), *Yano and Gilbert* (2003), and *Chen and Simchi-Levi* (2004b) for more references. The extension to the parametric case has been studied by *Subrahmanyan and Shoemaker* (1996), and *Petruzzi and Dada* (2002).

The most closely related works to ours are *Broder and Rusmevichientong* (2012), *Cheung et al.* (2015), and *Besbes and Zeevi* (2009). *Broder and Rusmevichientong* (2012) consider a dynamic pricing problem with Bernoulli demand, and the firm learns unknown parameters by maximum likelihood estimation (MLE). They classify the customer response probability functions into a general case (which motivates our definition of identifiable demand distributions) and a well-separated case. In this chapter, our demand process follows a general distribution, and because of censored demand, the traditional MLE cannot be applied to our problem. We develop a modified MLE method to incorporate censored data and prove that it admits the same convergence rate as the traditional MLE. This development is new to the literature. Note that *Broder and Rusmevichientong* (2012) assume that the firm has infinite inventory available initially, hence there are no inventory replenishment decisions and all realized demands are fully observed. In contrast, our problem has joint pricing and inventory control, and unsatisfied demand is unobservable. It is clear that a lower starting inventory level gives higher chance of stockout hence less information about demand, implying that inventory replenishment decisions impact demand learning. Thus, besides experimenting with prices, we also need to explore in the replenishment space so that demand information can be collected. In addition, *Broder and Rusmevichientong* (2012) do not consider the business constraints on the number of price changes. For the well-separated case, we design two algorithms for the joint pricing and replenishment problem, with the first one achieving a regret of

$\mathcal{O}(T^{1/(m+1)})$ when the firm is constrained to changing prices at most m times, and our second algorithm admits a regret of $\mathcal{O}(\log T)$ when the firm can change price $\mathcal{O}(\log T)$ times (in contrast, the algorithm of *Broder and Rusmevichientong* (2012) for the pure pricing problem has a regret of $\mathcal{O}(\log T)$ by changing the price $\mathcal{O}(T)$ times); for the general case, we develop an algorithm that changes the price at most k times and achieves a regret of $\mathcal{O}(T^{1/2})$ with the knowledge of horizon length T (the learning algorithm of *Broder and Rusmevichientong* (2012) changes prices $\mathcal{O}(T^{1/2})$ times without knowing T). We further show that the regret rates of our algorithms are the lowest possible, i.e., they are the same magnitude as the lower bound of regret of any learning algorithm for the respective classes of problems.

Cheung et al. (2015) study a dynamic pure pricing problem with demand learning, in which the firm faces a constraint on the number of price changes. They consider a scenario where there is infinite initial inventory, and demand distribution belongs to a finite set of possibilities. *Cheung et al.* (2015) present an algorithm which changes prices no more than m times and show that it has a regret of $\mathcal{O}(\log^{(m)} T)$, which achieves the lower bound. In contrast, in our model the customer response is from a general parametric class of functions with unknown continuous parameters of dimension k, hence the set has an infinite and uncountable number of elements. We have a joint optimization of pricing and inventory control problem with non-perishable products, thus we face the issue of not being able to achieve inventory target in making replenishment decisions. We also have censored demand, adding additional complexity to the analysis. Methodology wise, our work is substantially different from Cheung et al. (2014). For parameter estimation, we develop the modified MLE method, while *Cheung et al.* (2015) implements the first moment estimation (using sample average to estimate mean of the demand). Note that the convergence analysis of the estimation method in *Cheung et al.* (2015) only works for a finite number of candidate functions, and they assume the difference between values of any two can-

didate functions at the testing price point is lower bounded by a positive constant. Even with an infinite and countable number of candidate functions, the method in *Cheung et al.* (2015) will no longer work. To develop the lower bound for regret, *Cheung et al.* (2015) apply change of measure as in *Lai and Robbins* (1985), while in our chapter, we apply the van Trees inequality (*Gill and Levit* (1995)) to establish the lower bound. It is worthwhile to note that, when no constraints are imposed on the number of price changes, the lower bound for regret of the dynamic pricing model of *Cheung et al.* (2015) is $\Omega(1)$ (constant); while the lower bounds for the regret of our dynamic pricing and inventory replenishment problem are $\Omega(\log T)$ and $\Omega(T^{1/2})$, respectively, for the well-separated and general cases.

Besbes and Zeevi (2009) study the revenue management problem with fixed initial inventory using both nonparametric and parametric approaches, thus they have no inventory decisions. For the parametric case, they prove that the lower bound for the regret of their algorithm is $\Omega(T^{1/2})$. In their k-unknown-parameter case (which is similar to our k-identifiable case), they propose an algorithm with regret $\mathcal{O}\big(T^{2/3}(\log T)^{1/2}\big)$; in their 1-unknown-parameter case (which is similar to our well-separated case), they obtain a regret of $\mathcal{O}\big(T^{1/2}(\log T)^{1/2}(\log\log T)\big)$.

Structure of This chapter. In the next section we formulate the joint pricing and inventory replenishment problem. In Section 3 we present the learning algorithms, for the well-separated case and the general case, respectively, as well as their regret rates. In Section 4 we conduct a numerical study and report the numerical results. We conclude the chapter in Section 5. Finally, some detailed proofs and background information are provided in the Appendix. Throughout the chapter, for a real number x, let $\lceil x \rceil$ denote the smallest integer that is greater than or equal to x, and we use $\|\cdot\|$ to denote the Euclidean norm, i.e., $\|\mathbf{x}\| = (\sum_{i=1}^{n} x_i^2)^{1/2}$ for $\mathbf{x} = (x_1, \ldots, x_n) \in R^n$.

4.2 Model Formulation and Preliminaries

A firm sells a product over a planning horizon of T periods. At the beginning of each period t, the firm sets a selling price $p_t \in \mathcal{P} = [p^l, p^h]$ and determines a replenishment decision, the order-up-to level, $y_t \in \mathcal{Y} = \{y^l, y^l + 1, \ldots, y^l\}$, $t = 1, \ldots, T$. During period t, a random number of customers $D_t(p_t, \mathbf{z})$ arrive, where $\mathbf{z} \in \mathcal{Z}$ is a parameter vector and $\mathcal{Z}$ is a compact and convex set. Suppose $D_t(\cdot, \cdot)$ takes integer value from $\mathcal{D}$ that ranges from $d^l \geq 0$ to d^h ($\geq d^l$), which may be infinity, and the average demand $\mathbb{E}[D_t(p, \mathbf{z})]$ at the true value $\mathbf{z}$ is positive at price $p \in \mathcal{P}$. Realized demands are satisfied as much as possible by on-hand inventory, and unsatisfied demands are lost. We consider the scenario with censored demand, i.e., the firm only observes sales data $\min\{D_t(p_t, \mathbf{z}), y_t\}$ in period t, but not the actual demand. The cost structure includes the unit holding cost h, unit shortage cost b, and the inventory ordering cost is normalized to zero. Suppose the inventory replenishment lead-time is zero. The objective of the firm is to dynamically determine its pricing and inventory replenishment decisions in each period to maximize the expected total profit.

The demand model described above is a parametric model, i.e., for a given $p \in \mathcal{P}$, the firm knows the probability mass function for $D_t(p, \mathbf{z})$, $f(\cdot; p, \mathbf{z})$, up to the unknown parameter vector $\mathbf{z}$. Assume $f(\cdot; p, \mathbf{z})$ is differentiable with respect to $\mathbf{z}$. Clearly, if the firm knew the values of $\mathbf{z}$, then this is a standard dynamic joint pricing and inventory control problem that has been extensively studied in the literature. However, in our setting, the firm does not know the parameter vector $\mathbf{z}$, thus it has to learn about the demand information from past sales data, which is obtained through price and ordering experimentations. Furthermore, this chapter is concerned with the case that the firm is faced with the business constraint that prevents it from conducting extensive price experimentations. Thus, the firm is subject to the constraint on the number of times it can change its selling price.

The objective of the firm is to develop a mechanism that learns the demand

information from sales data while satisfying the constraint on the number of price changes, and exploit the extracted information to maximize its expected total profit.

Remark 1. *In the subsequent analysis, we will focus on the case that the selling price is continuous and the demand and order quantities are discrete. However, we point out that the results, as well as all analyses, carry over to the case with continuous demands and ordering quantities, i.e., $D_t(p_t, \mathbf{z})$ is a continuous random variable and $\mathcal{Y} = [y^l, y^h] \subset \mathcal{R}^+$.*

The Complete Information Problem. Let x_t denote the inventory level at the beginning of period t before the replenishment decision, and suppose the initial inventory level is $x_1 = 0$. Given a pricing and inventory policy $\phi = ((p_1, y_1), (p_2, y_2), \ldots, (p_T, y_T))$, the total expected profit over the planning horizon is

$$V^{\phi}(T) \tag{4.1}$$
$$= \sum_{t=1}^{T} \left\{ p_t \mathbb{E}[D_t(p_t, \mathbf{z})] - \left\{ h\mathbb{E}\left[\max\{y_t, x_t\} - D_t(p_t, \mathbf{z})\right]^+ + (b + p_t)\mathbb{E}\left[D_t(p_t, \mathbf{z}) - \max\{y_t, x_t\}\right]^+ \right\} \right\}.$$

If the firm knows the parameters $\mathbf{z}$ and thus also the distribution of D_t a priori, then dynamic programming can be used to compute the optimal pricing and inventory replenishment decisions. In that case, and if in addition there is no constraint on the number of price changes, then it is known (see e.g. *Sobel* (1981)) that a myopic policy is optimal for problem (4.1). Let $G(p, y, \mathbf{z})$ denote the single-period profit function, i.e.,

$$G(p, y, \mathbf{z}) = p\mathbb{E}[D(p, \mathbf{z})] - h\mathbb{E}\left[y - D(p, \mathbf{z})\right]^+ - (b + p)\mathbb{E}\left[D(p, \mathbf{z}) - y\right]^+, \tag{4.2}$$

where $D(p, \mathbf{z})$ is the generic random demand when the true parameter is $\mathbf{z}$ and the selling price is p, and suppose it has a unique maximizer (p^*, y^*) on $\mathcal{P} \times \mathcal{Y}$. Then the optimal strategy ϕ^* for the firm is to order up to y^* and set the price at p^* in each period.

Definition of Regret. In our setting, the firm does not know the parameter vector $\mathbf{z}$ a priori, so it needs to develop an adaptive policy ϕ which determines the selling price p_t and replenishment level y_t for each period t based on historical information, i.e., past selling prices, order-up-to levels, and sales data, subject to the constraint on the number of price changes. To measure the performance of a policy ϕ, we define the regret as the total profit loss of policy ϕ compared with that of the optimal policy ϕ^* when complete information is available and there is no constraint on the number of price changes. That is,

$$R^{\phi}(T) = V^{\phi^*}(T) - V^{\phi}(T).$$

It is clear that $R^{\phi}(T) \geq 0$, and the smaller the regret, the better policy ϕ performs.

The Traditional Maximum Likelihood Estimation. To estimate the unknown parameters $\mathbf{z}$ of a distribution, a commonly used method is maximum likelihood estimation (MLE). For $1 \leq t_1 \leq t_2 < \infty$, let $\{p_{t_1}, p_{t_1+1}, \ldots, p_{t_2}\}$ be a sequence of given prices for periods $\{t_1, t_1+1, \ldots, t_2\}$, and if the corresponding realized demand $\{d_{t_1}, d_{t_1+1}, \ldots, d_{t_2}\}$ can be observed and there is no censored data, then an estimate of $\mathbf{z}$ can be computed using the standard MLE given by

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}\in\mathcal{Z}} \prod_{t=t_1}^{t_2} f(d_t; p_t, \mathbf{z}). \tag{4.3}$$

In our setting, however, the traditional MLE will not work due to censored demand data. Indeed, the true demand d_t is observed only when $d_t < y_t$. If $d_t \geq y_t$, then the firm observes the sales quantity y_t with the implication that the demand d_t is no less than y_t. Therefore, the likelihood function (4.3) cannot be applied under censored demand data.

In this chapter we modify the standard MLE to incorporate censored demand information. A key in this step in our analysis is to show that the modified estimator

possesses the desired convergence rate under the mean-squared error measure.

A Technical Result. We next develop an upper bound for regret from estimation error in a general setting, which will be used in our subsequent analysis. Suppose that a firm maximizes an objective function $H(p, y, \mathbf{z})$ over decision variables p and y without knowing the values of underlying parameters $\mathbf{z}$ a priori, where $\mathbf{z} \in \mathcal{Z} \subset R^{r_3}$ for some integer $r_3 \geq 1$. The objective function may be multimodal, and the decision variables $p \in \mathcal{P} \subset R^{r_1}$ for some integer $r_1 \geq 1$ and $y \in \mathcal{Y} \subset Z^{r_2}$ for some integer $r_2 \geq 1$. The firm learns the value of $\hat{\mathbf{z}}$ through some noisy observations during decision process. We impose the following regularity conditions.

Assumption A (Regularity Conditions).

i) There is a unique global maximizer on $\mathcal{P} \times \mathcal{Y}$, denoted by $(p^*(\mathbf{z}), y^*(\mathbf{z}))$ for $H(p, y, \mathbf{z})$, i.e.,

$$(p^*(\mathbf{z}), y^*(\mathbf{z})) = \arg \max_{p \in \mathcal{P}, y \in \mathcal{Y}} H(p, y, \mathbf{z}),$$

and it falls into the interior of $\mathcal{P} \times \mathcal{Y}$.

ii) For any $y \in \mathcal{Y}$, $H(p, y, \mathbf{z})$ is twice differentiable with respect to $p \in \mathcal{P}$ with bounded second order derivatives.

iii) $H(p, y, \mathbf{z})$ satisfies the Lipschitz condition on $\mathcal{P} \times \mathcal{Y}$, i.e., there exists some constant $K_1 > 0$ such that $\|H(p_1, y_1, \mathbf{z}) - H(p_2, y_2, \mathbf{z})\| \leq K_1(\|p_1 - p_2\| + \|y_1 - y_2\|)$ for any $p_1, p_2 \in \mathcal{P}$ and $y_1, y_2 \in \mathcal{Y}$.

iv) $p^*(\mathbf{z})$ is locally Lipschitz on $\mathcal{P}$ at the true underlying parameter $\mathbf{z}$. That is, there exist constants $\delta > 0$ and $K_2 > 0$ such that when $\|\mathbf{z}' - \mathbf{z}\| < \delta$, we have $\|p^*(\mathbf{z}') - p^*(\mathbf{z})\| \leq K_2\|\mathbf{z}' - \mathbf{z}\|$.

v) If $\mathbf{z}$ is the true underlying parameter, then there exists a constant $\delta > 0$ such that when $\|\mathbf{z}' - \mathbf{z}\| < \delta$, we have $y^*(\mathbf{z}') = y^*(\mathbf{z})$. (If y is continuous, then

there exist constants $\delta > 0$ and $K_3 > 0$ such that when $\|\mathbf{z}' - \mathbf{z}\| < \delta$, we have $\|y^*(\mathbf{z}') - y^*(\mathbf{z})\| \leq K_3\|\mathbf{z}' - \mathbf{z}\|$.)

Under these assumptions, we have the following basic result.

Theorem IV.1. (Regret from Estimation Error). *Suppose* $\hat{\mathbf{z}}$ *is an estimator of* $\mathbf{z}$ *using* c *data points, and for any* $\epsilon > 0$ *it satisfies*

$$\mathbb{P}\big\{\|\hat{\mathbf{z}} - \mathbf{z}\| \geq \epsilon\big\} \leq K_4 e^{-cK_5\epsilon^2} \tag{4.4}$$

for some constants $K_4 > 0$ *and* $K_5 > 0$. *Then, there exists a positive constant* K_6 *such that*

$$H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - \mathbb{E}[H(p^*(\hat{\mathbf{z}}), y^*(\hat{\mathbf{z}}), \mathbf{z})] \leq \frac{K_6}{c}.$$

This theorem will play an important role in proving the main results in this chapter. Its proof is provided in Appendix A.

4.3 Learning Algorithms

With censored demand, the firm only observes sales data $\min\{y_t, d_t\}$ in period t. If stockout occurs in period t, then the firm knows that the demand is at least y_t, but does not know the exact demand. This has two implications: One is that the incompleteness of demand data impedes parameter estimation, as the firm can no longer compute the MLE estimator in (4.3) in the usual way. The other is that, with censored demand, the collected demand information depends on the inventory level y_t, hence the quality of the observed demand data depends on the inventory replenishment decision. Indeed, it is intuitive that higher inventory level helps reveal more demand information as less likely stockout would occur. This implies that the firm needs to strategically integrate inventory (and pricing) decisions with demand

learning in maximizing its total profit.

Depending on the characteristics of the class of parametric demand models, in the following subsections we study two cases and design learning algorithms that achieve the lowest possible regret rate for each of them.

4.3.1 Well-Separated Case

We first consider the case that the parameter z is a scaler, i.e., $z \in \mathcal{Z} = [z^l, z^h] \subset R^1$ for some $z^l \leq z^h < \infty$, and the demand processes with different parameters z are relatively easy to differentiate. Recall that two probability mass functions are said to be identifiable if they are not identically the same.

Definition 1. The family of distributions $\{f(\cdot; p, z) : z \in \mathcal{Z}\}$ is called well-separated if for any $p \in \mathcal{P}$, the class of probability mass functions $\{f(\cdot; p, z) : z \in \mathcal{Z}\}$ is identifiable, i.e., $f(\cdot; p, z_1) \neq f(\cdot; p, z_2)$ for $z_1 \neq z_2$.

Identifiability is an important concept in mathematical statistics and it has been widely used in the literature, see Condition (A_0) in *Borovkov* (1998). If a family of distributions is well-separated, then no matter what selling price p the firm charges, the corresponding demand distribution differs for different parameter z, hence allowing the firm to learn about the parameter z at any selling price. This indicates that in the well-separated case, it is possible to combine exploration with exploitation to design an efficient learning algorithm. The well-separated demand distributions have been studied in the revenue management literature with infinite starting inventory (hence there is no censored demand and no inventory decision) in *Broder and Rusmevichientong* (2012) and *Chen et al.* (2014a), among others.

We make the following assumptions for the well-separated family of distributions $\{f(\cdot; p, z) : z \in \mathcal{Z}\}$.

Assumption 1.

(i) There exists some constant $\underline{c}_f > 0$ such that $\tilde{I}(p, z) \geq \underline{c}_f$ for all $p \in \mathcal{P}$ and

$z \in \mathcal{Z}$, where

$$\tilde{I}(p,z) = \frac{\left(\partial f(d^l;p,z)/\partial z\right)^2}{f(d^l;p,z)},$$

and there exists a constant $\bar{c}_f < +\infty$ such that the Fisher information $I_f(p,z)$, given by

$$I_f(p,z) = \sum_{d=d^l}^{d^h} \frac{\left(\partial f(d;p,z)/\partial z\right)^2}{f(d;p,z)},$$

satisfies $I(p,z) < \bar{c}_f$ for all $p \in \mathcal{P}$ and $z \in \mathcal{Z}$.

(ii) There exists a constant $\underline{f} > 0$ such that $f(d;p,z) \geq \underline{f}$ for all $p \in \mathcal{P}$, $z \in \mathcal{Z}$ and $d \in \{d^l, d^l+1, \ldots, y^h\}$.

(iii) For any $p \in \mathcal{P}$, $f(d^l;p,z)$ is strictly monotone in parameter $z \in \mathcal{Z}$.

Assumption 1 is satisfied by various demand distributions, and two are given below.

Example 1. The following examples satisfy Assumption 1: (a) Poisson random variable with rate $r(p,z)$, (b) Binomial random variable with total number of trials $d^h \geq 1$ and success probability $r(p,z)$. For both examples, $r(p,z)$ can be

1) linear function $r(p,z) = 2 - zp$ with $\mathcal{P} = [8/15, 2/3]$, $\mathcal{Z} = [2,3]$;

2) logit function $r(p,z) = \frac{e^{2-zp}}{1+e^{2-zp}}$ with $\mathcal{P} = [1/2, 2]$, $\mathcal{Z} = [1,5]$;

3) exponential function $r(p,z) = e^{2-zp}$ with $\mathcal{P} = [2,10]$, $\mathcal{Z} = [2,5]$.

As will be described in the algorithm shortly, we will estimate the unknown parameter z using a modified MLE method tailored for the well-separated case under censored demand data. Assumption 1 is imposed to guarantee that this modified MLE estimator will converge to the true value of z at the desired convergence rate.

We assume $y^* > d^l$, i.e., the true optimal order-up-to level is higher than the lower bound of random demand, which is a reasonable and weak assumption. During the learning process, the best solution under the updated estimate of z may be equal to d^l. That is, if $\hat{z}$ is the estimated parameter it may happen that $y^*(\hat{z}) = d^l$. If we implement inventory decision $y^*(\hat{z})$, then the inventory level will surely drop to zero at the end of the period, and the only demand information it yields is that demand is at least d^l, which is already known. This shows that, whenever $y^*(\hat{z}) = d^l$ occurs, we should modify the ordering decision to a quantity above d^l, say $d^l + \Delta$ for some small positive number Δ, in the algorithm so that some information about demand is to be revealed with positive probability.

Fixed Number of Price Changes. We first consider the setting where the number of price changes is limited to a given number, say $m \geq 1$. To develop a learning algorithm for this case, we divide the planning horizon T into $m+1$ stages, of which the ith stage consists of $I_i = \lceil T^{i/(m+1)} \rceil$ periods, $i = 1, \ldots, m$, and the last stage contains $T - \sum_{i=1}^{m} I_i$ periods. During stage $i \geq 2$, the algorithm sets a pricing and ordering decision that is constructed using data collected from the previous stage $i-1$. If the order-up-to level in the solution is above d^l, then both the pricing and ordering decision is implemented in stage i. Otherwise, and as we discussed above, the algorithm implements the pricing solution but raises the order-up-to level slightly. At the end of the stage, the algorithm applies the observed sales data to estimate parameter z using a modified MLE method, and then solve a data-driven version of optimization problem for (4.2) to obtain a new decision to be used for the subsequent stage $i+1$.

Let t_i denote the last period of stage $i-1$, $i = 2, \ldots, m+2$. To get the algorithm started, we need an initial pricing decision $\hat{p}_1 \in \mathcal{P}$ and initial ordering decision $\hat{y}_1 = d^l + \Delta$ for some constant $\Delta > 0$ such that $\hat{y}_1 \in \mathcal{Y}$. Let $F(d, p, z) = \sum_{x \leq d} f(x, p, z)$.

Algorithm I (m price changes for the well-separated case)

Step 0: Preparation

$I_i = \lceil T^{i/(m+1)} \rceil$, for $i = 1, \ldots, m$, and $I_{m+1} = T - \sum_{i=1}^{m} I_i$.

$t_1 = 0$, and $t_i = \sum_{j=1}^{i-1} I_j$ for $i = 2, \ldots, m+2$.

Step 1: Setting pricing and ordering decisions

For stage $i \leq m+1$, set the price and inventory level to

$$
\begin{aligned}
p_t &= \hat{p}_i, \quad t = t_i + 1, \ldots, t_{i+1}, \\
y_t &= \max\{x_t, \tilde{y}_i\}, \quad t = t_i + 1, \ldots, t_{i+1}, \\
x_{t+1} &= \max\{y_t - d_t, 0\}, \quad t = t_i + 1, \ldots, t_{i+1},
\end{aligned}
$$

where

$$
\tilde{y}_i = \begin{cases} \hat{y}_i, & \text{if } \hat{y}_i > d^l, \\ d^l + \Delta, & \text{if } \hat{y}_i = d^l. \end{cases}
$$

Step 2: Estimation

Compute the estimator for z by

$$
\hat{z}_i = \operatorname*{argmax}_{z \in \mathcal{Z}} \left\{ \prod_{\{t \in \{t_i+1, \ldots, t_{i+1}\} : d_t < y_t\}} f(d_t; \hat{p}_i, z) \cdot \prod_{\{t \in \{t_i+1, \ldots, t_{i+1}\} : d_t \geq y_t\}} \Big(1 - F(y_t - 1; \hat{p}_i, z)\Big) \right\}. \tag{4.5}
$$

Step 3: Data-driven optimization

Solve the data-driven optimization problem

$$
(\hat{p}_{i+1}, \hat{y}_{i+1}) = \arg \max_{(p,y) \in \mathcal{P} \times \mathcal{Y}} G(p, y, \hat{z}_i). \tag{4.6}
$$

Go to Step 1 with $i := i + 1$.

The intuition behind the learning algorithm above is the following. Since selling price cannot be changed more than m times, the planning horizon is divided into $m+1$ stages with each stage charging the same price. These stages are exponentially increasing in length since, as more data are collected, more accurate estimates of demand are obtained hence they can be used for longer time to extract profit.

For each stage $i \geq 2$, Step 1 reflects the tension between exploration and exploitation regarding the ordering decisions. At the new price decision $\hat{p}_i$, if the corresponding ordering decision $\hat{y}_i$ equals d^l, then implementing $\hat{y}_i$ will not yield any information about demand distribution. Hence the algorithm prescribes order-up-to level $d^l + \Delta$ instead, which will guarantee observing a non-censored demand realization with positive probability, thus providing information to update the estimate of parameter z. Note that in this case, the algorithm experiments an ordering decision at a loss of profit. Fortunately, as the learning process continues, the probability for having $\hat{y}_i = d^l$ will be diminishing because, if $\hat{z}_i$ approaches the true z, then $y^*(\hat{z}_i)$ will approach y^* which is greater than d^l. In Step 2, a modified MLE is employed to estimate z. For each period t, if $d_t < y_t$, then we can observe the true value of d_t, and the probability for that event is $f(d_t; \hat{p}_i, z)$; if $d_t \geq y_t$, then the firm only knows that the demand d_t is at least y_t, and the probability for this event is $1 - F(y_t - 1; \hat{p}_i, z)$, which is incorporated in the likelihood function. The optimization problem constructed in (4.5) resembles the traditional MLE method in (4.3). Finally, in Step 3, the data-driven optimization problem finds the optimal pricing and inventory decisions using the updated estimate $\hat{z}_i$ of parameter z, which will be implemented in the next iteration.

The following theorem gives the theoretical performance of Algorithm I.

Theorem IV.2. *For any problem instance of the well-separated case that satisfies Assumptions A and 1, for any initial values of $\hat{p}_1 \in \mathcal{P}$, $\Delta > 0$ and $\hat{y}_1 + \Delta \in \mathcal{Y}$, there exists a constant $K_7 > 0$ such that the regret of learning algorithm I with at most m*

price changes is upper bounded by

$$R(T) \leq K_7\, T^{\frac{1}{m+1}}.$$

Before presenting the proof of Theorem 1, we elaborate on the technical issues encountered in analyzing the algorithm. First, note that the objective function in (4.5) is different from the traditional MLE method, and there exists no result in the literature on convergence rate for estimators obtained by maximizing the modified likelihood function (4.5). To overcome this issue, we introduce a truncated random variable $\tilde{D}_{t,y_t}(\hat{p}_i, z)$ defined on $\{d^l, d^l+1, \ldots, y_t\}$ with probability mass function $\tilde{f}_{y_t}(\cdot; \hat{p}_i, z)$:

$$\tilde{f}_{y_t}(d; \hat{p}_i, z) = \begin{cases} f(d; \hat{p}_i, z), & \text{if } d^l \leq d < y_t, \\ 1 - F(y_t - 1; \hat{p}_i, z), & \text{if } d = y_t. \end{cases}$$

Then, $\tilde{D}_{t,y_t}(\hat{p}_i, z) = \min\{D_t(\hat{p}_i, z), y_t\}$. It is easily verified that (4.5) is exactly the maximum likelihood estimation for this truncated distribution. The main difficulty lies in that $\tilde{D}_{t,y_t}(\cdot; \hat{p}_i, z)$ are dependent across periods, because y_t depends on demand realizations and inventory levels of previous periods. Furthermore, $\tilde{D}_{t,y_t}(\hat{p}_i, z)$ are not identically distributed as y_t are not constant. As a result, the convergence rate result of MLE (see *Borovkov* (1998) Theorem 36.3), which requires samples to be independently and identically distributed, cannot be applied here.

Nonetheless, in Proposition 1 below, we show that $\hat{z}_i$ computed from (4.5), although involving dependent and non-identically distributed random variables, converges to the true z at the same rate as that of the traditional MLE method. This result is crucial for establishing the upper bound of regret in Theorem 2.

Proposition IV.3. *For any problem instance of the well-separated case that satisfies Assumptions A and 1, there exist constants $K_8 > 0$ and $K_9 > 0$ such that for any*

$\epsilon > 0$, $\hat{z}_i$ of (4.5) satisfies

$$\mathbb{P}\big\{|\hat{z}_i - z| \geq \epsilon\big\} \leq K_8 e^{-I_i K_9 \epsilon^2}.$$

Proposition IV.3 implies

$$\mathbb{E}[|\hat{z}_i - z|^2] = \int_0^{+\infty} P\{|\hat{z}_i - z|^2 \geq \epsilon\} d\epsilon \leq \int_0^{+\infty} K_8 e^{-I_i K_9 \epsilon} d\epsilon = \frac{K_{10}}{I_i}, \tag{4.7}$$

where $K_{10} = K_8/K_9$. This result will be utilized in proving Theorem IV.2.

Proof Sketch of Theorem IV.2. By the definition of regret, we have

$$\begin{aligned}
R(T) &= \sum_{t=1}^{T} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] \\
&= \sum_{t=t_1+1}^{t_2} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] + \sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] \\
&\leq \underbrace{\sum_{t=t_1+1}^{t_2} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right]}_{\text{regret from initial decision}} + \underbrace{\sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[G(p^*, y^*, z) - G(\hat{p}_i, \hat{y}_i, z)\right]}_{\text{regret from estimation error}} \\
&\quad + \underbrace{\sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, \tilde{y}_i, z)|\right]}_{\text{regret from exploration on the ordering decision}} + \underbrace{\sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \tilde{y}_i, z) - G(p_t, y_t, z)|\right]}_{\text{regret from missing inventory targets}}.
\end{aligned} \tag{4.8}$$

As marked in (4.8), the first term on the right hand side stems from the input initial solutions that may not be optimal, and its upper bound is proportional to the length of the first stage. The second term is due to the estimation error ($\hat{z}_i$ may not be equal z), and the closer $\hat{z}_i$ is to z, the smaller the second part of regret. The key drivers to obtain an upper bound for the second part of regret are, as pointed out by one reviewer, (i) the optimization error is linear in the estimation error, and (ii) the squared estimation error is inversely proportional to the sample size. The first one (i)

is established under some regularity condition concerning continuity of the function, while the second (ii) refers to (4.7). The upper bound for the second term is obtained by applying Theorem IV.1. The third term presents the regret from exploration in the inventory decision, and it is proportional to the probability for $\tilde{y}_i$ to be not equal to $\hat{y}_i$, which is also the probability for $\hat{y}_i$ to be equal to d^l. As discussed earlier, since $y^* > d^l$, as the data size increases, it is intuitive that the probability for $\hat{y}_i$ to be close to y^* will be high, hence the probability for this event will be small. The fourth and last term on the right hand side of (4.8) is contributed by the carry-over inventories between periods. We employ Hoeffding inequality to show that after a relative short number of periods, $\tilde{y}_i$ can be achieved with a high probability. A general result in bounding the fourth part of regret is shown in Proposition A1 in Appendix A.

An important question is whether there exists learning algorithm with m or fewer price changes but with lower regret rate than Algorithm I. The following result shows that is not possible at least for algorithms with predetermined price-change schedules.

Theorem IV.4. *There exist problem instances such that the regret for any learning algorithm for the joint inventory control and pricing problem with censored demand that changes price at most m times according to a predetermined schedule is lower bounded by $\Omega(T^{1/(m+1)})$. That is, there exists a constant $K_{11} > 0$ such that for any such learning algorithm ϕ,*

$$R^{\phi}(T) \geq K_{11}\, T^{\frac{1}{m+1}}.$$

Proof Sketch. To prove Theorem IV.4, we construct a problem instance in which the inventory order-up-to level for each period is fixed and high enough so that any realization of the demand can be satisfied under any price. Therefore, the effect of lost sales and censored data is eliminated and the original joint pricing and inventory control problem is reduced to a pure dynamic pricing problem with fixed inventory

control strategies. Because price for period t, p_t, is a function of the historical data from period 1 to period $t-1$, it can be considered as an "estimator" based on $t-1$ data points. By van Trees inequality (*Gill and Levit* (1995)), the performance of p_t is lower bounded as inversely proportional to $t-1$, which can be used to establish the lower bound of the regret. Because of the freedom of at most m price changes, this gives rise to a problem with $m+1$ variables, and we apply geometric inequality to prove the designed result.

The two theorems above show that our algorithm has achieved the lowest regret rate for the well-separated case with a fixed number and predetermined times of price changes under censored demand.

Remark 2. *The discussion following Algorithm I leads to a practically less interesting mathematical problem of what happens if the real optimal order-up-to level is very low, i.e., $y^* \leq d^l$? We have also studied this case, and as one can expect, it will become inevitable to have more tension between learning and earning because of the lack of information the learning phase can offer in exploring the true value of z. In that case learning algorithm can be developed with a higher regret rate of $O(T^{1/2})$.*

A More-frequent-price-change Case. In the analysis above it is assumed that the number of price changes is restricted up to a fixed number. In applications it may be the case that the firm cannot change the price too often, but it is allowed to make more price changes when the planning horizon is longer. In the following, we propose a learning algorithm for the joint pricing and inventory control problem which can change the price $\mathcal{O}(\log T)$ times, and we refer to it as the case with more-frequent-price-change. We show that the regret of our algorithm improves significantly, from polynomial $\mathcal{O}(T^{1/(m+1)})$ to $\mathcal{O}(\log T)$.

In our learning algorithm for the case with more-frequent-price-change, we again divide the time horizon into stages with exponentially increasing lengths. Let $I_0 > 0$

and $v > 1$ be given positive numbers, and let

$$I_i = \left\lceil I_0 v^i \right\rceil, \qquad i = 1, 2 \ldots, N, \tag{4.9}$$

denote the length of stage i, where

$$N = \left\lceil \log_v \left(v + \frac{v-1}{I_0} T \right) - 2 \right\rceil = \mathcal{O}(\log T)$$

is the number of price changes. The last stage, $N+1$, has $I_{N+1} = T - \sum_{i=1}^{N} I_i$ periods. Again we let t_i be the last period of stage $i-1$, i.e., $\sum_{j=1}^{i-1} I_j = t_i$, $i = 2, \ldots, N+2$, with $t_1 = 0$. Thus, stage i starts in period $t_i + 1$. The algorithm needs some initial input $\hat{p}_1 \in \mathcal{P}, \hat{y}_1 = d^l + \Delta \in \mathcal{Y}$ for the first stage. The algorithm runs in exactly the same manner as Steps 1 to 3 in Algorithm I, except that now the number of periods in stage i is given by (4.9) and there is a total of $N = \mathcal{O}(\log T)$ iterations.

Theorem IV.5. *For any problem instance of the well-separated case that satisfies Assumptions A and 1, for any initial values of $\hat{p}_1 \in \mathcal{P}$, $\Delta > 0$ and $\hat{y}_1 + \Delta \in \mathcal{Y}$, there exists a constant $K_{12} > 0$ such that the regret of the learning algorithm with $\mathcal{O}(\log T)$ price changes is upper bounded by*

$$R(T) \leq K_{12} \ \log T.$$

We remark that $\Omega(\log T)$ is also the lower bound for the regret of any algorithm for our problem in hand. As a matter of fact, even for the special case with no constraint on the number of price changes and no censored demand data, $\Omega(\log T)$ is the lower bound for the regret of any learning algorithm. Indeed, *Broder and Rusmevichientong* (2012) establish such a lower bound for the dynamic pricing problem with infinite initial inventory (thus there is no inventory replenishment decision and no censored data) and no constraint on the number of price changes, and they show that the regret

is lower bounded by $\Omega(\log T)$. *Broder and Rusmevichientong* (2012) obtain this lower bound by a pricing policy that changes price every period. As our problem is more general than theirs, the regret of our problem is also lower bounded by $\Omega(\log T)$. This shows that our algorithm has achieved the lowest possible regret rate for the problem with more-frequent-price-change.

To prove Theorem 4, we will not be able to apply Proposition A1 to bound the regret from missing inventory target $\tilde{y}_i$ as was done in the proof of Theorem 2. This is because, in the algorithm of Theorem 2, the stage is long, thus at the beginning of each stage we can allocate a relatively short phase to be the "depletion phase". During the deletion phase, the initial inventory of this stage can be consumed by the cumulative demands to below $\tilde{y}_i$ (with a high probability), and after which the target level $\tilde{y}_i$ is always achieved in this stage. However, in the algorithm in Theorem 4, the stage is short, even shorter than the required length of the "depletion phase", so the above method cannot be applied. The idea in proving Theorem 4 is to prove the initial inventory level of stage i is not very high compared with $\tilde{y}_i$. We obtain this by showing that $\tilde{y}_{i-1}$ and $\tilde{y}_i$ are very close (with a high probability), and once $\tilde{y}_{i-1}$ is finally achieved during stage $i-1$, the initial inventory level for stage i will be no higher than $\tilde{y}_{i-1}$.

4.3.2 The General Case

An important assumption in the previous subsection is that, the demand distribution is identifiable for any selling price p. This special demand structure allows the firm to learn about parameter z at any selling price, and therefore, the firm does not need to "sacrifice" revenue to "learn" demand information. More precisely, it allows the firm to combine "exploration" with "exploitation", leading to a small regret of the algorithm.

When that condition is not satisfied, there will be more tension between earning

and learning in making the pricing and inventory decisions: on one hand, the firm would like to set the prices as close as possible to estimated optimal price so that more profits can be earned, but on the other, that price may not be identifiable so that firm may not be able to learn more demand information at that price. Hence, firm has to intentionally create price dispersion so that the underlying demand-price relationship can be better learned. The latter, however, will result in profit loss. We consider this more general case in this subsection.

Suppose the parameter in probability mass function $f(\cdot; p, \mathbf{z})$ is a k-dimensional vector, i.e., $\mathbf{z} = (z_1, \ldots, z_k) \in \mathcal{Z} \subset R^k$ for some integer $k \geq 1$. To estimate $\mathbf{z}$, we need at least k prices for experimentation. In this subsection we assume $d_t \in \{d^l, d^l + 1, \ldots, d^h\}$ where $d^h \leq y^h < \infty$. The latter assumption is made to allow the firm to learn the demand distribution by raising inventory levels. For a set of given prices $\mathbf{p} = (p_1, \ldots, p_k) \in \mathcal{P}^k$, and correspondingly realized demands $\mathbf{d} = (d_1, \ldots, d_k) \in \{d^l, d^l + 1, \ldots, d^h\}^k$, we define

$$\mathcal{Q}^{\mathbf{p},\mathbf{z}}(\mathbf{d}) = \prod_{j=1}^{k} f(d_j; p_j, \mathbf{z}).$$

Definition 2. The family of distributions $\{\mathcal{Q}^{\mathbf{p},\mathbf{z}} : \mathbf{z} \in \mathcal{Z}\}$ is said to belong to the general case if there exist k price points $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_k) \in \mathcal{P}^k$ such that the family of distributions $\{\mathcal{Q}^{\bar{\mathbf{p}},\mathbf{z}} : \mathbf{z} \in \mathcal{Z}\}$ is identifiable, i.e., $\mathcal{Q}^{\bar{\mathbf{p}},\mathbf{z}_1}(\cdot) \neq \mathcal{Q}^{\bar{\mathbf{p}},\mathbf{z}_2}(\cdot)$ for any $\mathbf{z}_1 \neq \mathbf{z}_2$ in $\mathcal{Z}$.

The prominent difference between the general case and the well-separated case is that in the general case the likelihood function is known to be identifiable only at a set of prices $\mathbf{p}$; while in the well-separated case of the last subsection, the demand distribution is identifiable at any selling price. Thus, to learn about the true value of $\mathbf{z}$ in the general case, the firm has to consistently experiment at these prices, resulting in profit loss. This shows that it is inevitable that the learning algorithm will suffer

higher regret. For the above reason, we shall refer to $\mathbf{p}$ as the exploration prices.

To ensure that the unknown parameters $\mathbf{z}$ can be estimated using our modified maximum likelihood method, we make the following assumption for the general case.

Assumption 2. For any $\mathbf{z} \in \mathcal{Z}$,

i) there exists a constant $c_f > 0$ such that $\lambda_{\min}\{\mathbf{I}(\bar{\mathbf{p}}, \mathbf{z})\} \geq c_f$, where $\mathbf{I}(\bar{\mathbf{p}}, \mathbf{z})$ denotes the Fisher information matrix given by

$$[\mathbf{I}(\bar{\mathbf{p}}, \mathbf{z})]_{i,j} = \mathbb{E}_{\mathbf{z}}\left[-\frac{\partial^2}{\partial z_i \partial z_j} \log \mathcal{Q}^{\bar{\mathbf{p}},\mathbf{z}}(\mathbf{D})\right],$$

and $\lambda_{\min}\{\mathbf{I}(\bar{\mathbf{p}}, \mathbf{z})\}$ is the smallest eigenvalue of the Fisher information matrix $\mathbf{I}(\bar{\mathbf{p}}, \mathbf{z})$;

ii) there exists a constant $\underline{f} > 0$ such that $f(d; \bar{p}_j, \mathbf{z}) \geq \underline{f}$ for $1 \leq j \leq k$ and all d.

Similar conditions have been imposed and discussed in *Broder and Rusmevichientong* (2012), *Besbes and Zeevi* (2009), and *Chen et al.* (2014a). The following families of demand distributions have been verified to satisfy them.

Example 2. $D(p, z)$ is a binomial variable with a constant total number of trials $d^h \geq 1$ and success probability $r(p, z)$. Examples of $r(p, z)$ include

1) linear function $r(p, \mathbf{z}) = z_1 - z_2 p$ with $\mathcal{P} = [1/3, 1/2], \mathcal{Z} = [2/3, 3/4] \times [3/4, 1]$ and any $\bar{p}_1 \neq \bar{p}_2 \in \mathcal{P}$;

2) logit function $r(p, \mathbf{z}) = \frac{e^{-z_1 p - z_2}}{1 + e^{-z_1 p - z_2}}$ with $\mathcal{P} = [1/2, 2], \mathcal{Z} = [1, 2] \times [-1, 1]$ and any $\bar{p}_1 \neq \bar{p}_2 \in \mathcal{P}$;

3) exponential function $r(p, \mathbf{z}) = e^{-z_1 p - z_2}$ with $\mathcal{P} = [1/2, 1], \mathcal{Z} = [1, 2] \times [0, 1]$ and any $\bar{p}_1 \neq \bar{p}_2 \in \mathcal{P}$.

Because of censored data, the true demand realizations exceeding the on-hand inventory level cannot be observed. Thus, we design another variation of MLE to

estimate $\mathbf{z}$ in Algorithm II below. This algorithm divides the planning horizon T into two stages, i.e., an exploration stage which is of length $\lceil T^{1/2} \rceil$ followed by the exploitation stage. During the exploration stage, Algorithm II experiments in the inventory space. To guarantee that demand distribution is sufficiently explored, every time the firm observes a stockout, it increases the order-up-to level by some percentage. Let $I = \lceil T^{1/2}/k \rceil$, input $\bar{y} \in \mathcal{Y}$ for the initial inventory order-up-to level, and $s > 0$.

Algorithm II (k price change for the general case)

Step 0: Preparation

Let $I = \lceil T^{1/2}/k \rceil$.

Step 1: Exploration of prices and order-up-to levels for periods $t \in \{1, \ldots, kI\}$

Set price as follows: For $i = 1, \ldots, k$, set

$$p_t = \bar{p}_i, \text{ for } t = (i-1)I + 1, \ldots, iI.$$

Set inventory order-up-to level as follows:

(i) for $t = (i-1)I + 1$, set $y_t = \max\{x_t, \bar{y}\}$;

(ii) for $t = (i-1)I + 2, \ldots, iI$, set

$$y_t = \begin{cases} y_{t-1}, & \text{if } d_{t-1} < y_{t-1}; \\ \min\left\{(1+s)y_{t-1}, y^h\right\}, & \text{otherwise.} \end{cases}$$

And let

$$x_{t+1} = \max\{y_t - d_t, 0\}, \text{ for } t = (i-1)I + 1, \ldots, iI.$$

Step 2: Estimation

Estimate $\mathbf{z}$ by

$$\hat{\mathbf{z}} = \underset{\mathbf{z}\in\mathcal{Z}}{\operatorname{argmax}} \left\{ \prod_{\{t\in\{1,\ldots,kI\}:y_t>d_t\}} f(d_t; p_t, \mathbf{z}) \cdot \prod_{\{t\in\{1,\ldots,kI\}:y_t\leq d_t\}} \Big(1 - F(y_t - 1; p_t, \mathbf{z})\Big) \right\}. \tag{4.10}$$

Step 3: Data-driven optimization and exploitation

Solve the data-driven optimization problem

$$(\hat{p}, \hat{y}) = \max_{(p,\, y)\, \in\, \mathcal{P}\, \times\, \mathcal{Y}} \; G(p, y, \hat{\mathbf{z}}).$$

For periods $t = kI + 1, \ldots, T$, set the pricing and inventory level to

$$p_t = \hat{p}, \qquad y_t = \max\{x_t, \hat{y}\},$$

and let

$$x_{t+1} = \max\{0, y_t - d_t\}.$$

In Step 1, Algorithm II experiments at every exploration price for the same number of periods I during the exploration stage. At each price, the target inventory order-up-to level is first set to $\bar{y}$, but it is raised by some percentage s whenever a stockout is observed. The logic for this action is to explore more information about the demand distribution. In Step 2, the unknown parameter $\mathbf{z}$ is estimated as in (4.10), which is then used in Step 3 to compute the updated pricing and inventory decision $(\hat{p}, \hat{y})$, and that are implemented for the rest of the planing horizon.

The following theorem establishes the theoretical worst-case performance guarantee of Algorithm II.

Theorem IV.6. *Consider any problem instance of the general case satisfying Assumptions A and 2 with exploration prices $\bar{\mathbf{p}} \in \mathcal{P}^k$, for any $\bar{y} \in \mathcal{Y}$ in Algorithm II, there exists a constant $K_{12} > 0$ such that the regret is upper bounded by*

$$R(T) \leq K_{12} T^{\frac{1}{2}}.$$

A key in establishing the result above is the convergence rate of $\hat{\mathbf{z}}$ to the true parameter $\mathbf{z}$. Because of the censored demand data, the estimation of $\mathbf{z}$ is not the traditional MLE. However, since the demand in each period is upper bounded by d^h, the "raising inventory" action is performed for at most $\left\lceil \log_{1+s} \frac{d^h}{y^l} \right\rceil$ times, which is independent of the length of the exploration phase. In other words, there is a bounded number of stockout periods in the learning phase, and in the rest of at least $\lceil T^{1/2} \rceil - \left\lceil \log_{1+s} \frac{d^h}{y^l} \right\rceil$ periods, the firm observes complete demand realizations. During the stockout periods, demands are truncated up to the corresponding starting inventory levels, and are thus dependent and follow different distributions. Since the number of these truncated demands are upper bounded by a constant, when the time horizon grows, the impact of stockout periods diminishes. This allows us to show, in Proposition IV.7, that $\hat{\mathbf{z}}$ still converges to the true $\mathbf{z}$ at the same rate as the standard MLE method.

Proposition IV.7. *Consider any problem instance of the general case satisfying Assumptions A and 2 with exploration prices $\bar{\mathbf{p}} \in \mathcal{P}^k$, for any $\bar{y} \in \mathcal{Y}$ in Algorithm II, there exist some constants $K_{13} > 0$ and $K_{14} > 0$ such that for any $\epsilon > 0$, the estimator $\hat{\mathbf{z}}$ in Step 2 satisfies*

$$\mathbb{P}\big\{ \|\hat{\mathbf{z}} - \mathbf{z}\| \geq \epsilon \big\} \leq K_{13} e^{-IK_{14}\epsilon^2}.$$

The proof of Proposition IV.7 is given in Appendix A. The convergence rate of $\hat{\mathbf{z}}$ to $\mathbf{z}$ stated in Proposition IV.7 allows us to prove Theorem IV.6.

Proof Sketch of Theorem IV.6. The regret can be evaluated as

$$\begin{aligned} R(T) &= \sum_{t=1}^{T} \mathbb{E}[G(p^*, y^*, \mathbf{z}) - G(p_t, y_t, \mathbf{z})] \\ &\leq \underbrace{\sum_{t=1}^{kI} \mathbb{E}\left[G(p^*, y^*, \mathbf{z}) - G(p_t, y_t, \mathbf{z})\right]}_{\text{Exploration Regret}} \\ &\quad + \underbrace{\sum_{t=kI+1}^{T} \mathbb{E}\left[G(p^*, y^*, \mathbf{z}) - G(\hat{p}, \hat{y}, \mathbf{z})\right]}_{\text{Regret from Estimation Error}} + \underbrace{\sum_{t=kI+1}^{T} \left[|G(\hat{p}, \hat{y}, \mathbf{z}) - G(p_t, y_t, \mathbf{z})|\right]}_{\text{Regret from Missing Inventory Targets}}. \end{aligned} \tag{4.11}$$

In (4.11), the first part on the right hand side is the profit loss during the exploration phase. The second term comes from the estimation error of $\hat{\mathbf{z}}$, for which the convergence rate is provided in Proposition IV.7, thus the second term can be bounded using Theorem IV.1. The third term stems from the fact that $\hat{y}$ may not be achieved for some period t if $x_t > \hat{y}$, which happens when $x_{kI+1} > \hat{y}$, resulting in overshooting of inventory process. Its upper bound is presented in Proposition A1 of Appendix A.

We point out that, even when both the pricing and inventory decisions are allowed to change in each and every period (so there is no constraint on the number of price changes) and there is no censored demand data, the regret rate for any learning algorithm is lower bounded by $\Omega(T^{1/2})$. This lower bound is established in *Broder and Rusmevichientong* (2012) for a dynamic pricing problem with infinite initial inventory. Since that model is a special case of ours, the lower bound holds in our setting with joint pricing and inventory replenishment decisions as well. This shows that Algorithm II actually achieves the lowest possible regret rate.

Remark 3. *The problem instances for Theorems 2 to 5 can depend on the length of planning horizon T. Indeed, the lower bound developed in Broder and Rusmevichientong (2012) is also based on problem instances with parameters depending on T.*

4.4 Numerical Results

We consider time horizons of length $T = 100, 300, 1000, 3000, 10000$. The feasible region for order-up-to level is $\mathcal{Y} = \{0, 1, 2, 3, 4\}$, $b = 0.6$, and $h = 0.1$. For the well-separated case, demand follows the Poisson distribution with rate $r(p_t, z)$, and the function $r(p, z)$ and feasible region $\mathcal{P}$ for selling price p are described below. We consider two functions of $r(p_t, z)$:

i) Exponential function $r(p, z) = \exp(2 - zp)$ with true value $z = 4/5$, $\mathcal{P} = [1/10, 2]$, $\mathcal{Z} = [1/2, 1]$, the starting price $\hat{p}_1 = 1/10$, and the starting order-up-to level is $\hat{y}_1 = 2$.

ii) Logit function $r(p, z) = \exp(-zp)/(1 + \exp(-zp))$ with true value $z = 1$, $\mathcal{P} = [1/2, 3/2]$, $\mathcal{Z} = [1/5, 3/2]$, the starting price $\hat{p}_1 = 1$, and the starting order-up-to level is $\hat{y}_1 = 2$.

We conduct experiments when the number of price changes is constrained to 1, 2, 3, 4, 5, or $\lceil \log T \rceil$.

For the general case, demand follows the Binomial distribution $B(4, r(p_t, \mathbf{z}))$, and $s = 0.2$. We also consider two functions of $r(p_t, \mathbf{z})$:

1) Exponential function $r(p, \mathbf{z}) = \exp(-z_1 p - z_2)$ with $z_1 = 3/2$ and $z_2 = 1/2$, $\mathcal{P} = [1/2, 1]$ and $\mathcal{Z} = [1, 2] \times [0, 1]$. $\overline{p}_1 = 1/2$, $\overline{p}_2 = 1$, and $\overline{y} = 3$. The number of price changes is limited to 2.

2) Logit function $r(p, \mathbf{z}) = \exp(-z_1 p - z_2)/(1 + \exp(-z_1 p - z_2))$ with $z_1 = 1$ and $z_2 = -1$, $\mathcal{P} = [1/2, 2]$ and $\mathcal{Z} = [1/5, 2] \times [-1, 1]$. $\overline{p}_1 = 1/2$, $\overline{p}_2 = 3/2$, and $\overline{y} = 3$. The number of price changes is limited to 2.

To evaluate the performance of the algorithm, we consider the percentage profit loss compared with the complete information optimal profit when there is no con-

straint on the number of price changes, which is

$$\frac{R(T)}{T \times G(p^*, y^*, \mathbf{z})} \times 100\%.$$

We compute the percentage profit loss per period over 100 rounds, then calculate the average value. The results are summarized in Table 4.1.

Table 4.1: Numerical results

		Horizon Length	T=100	T=300	T=1000	T=3000	T=10000
Exponential Response Probability	Well-separated Case	m=1	13.15	9.57	7.06	5.57	5.25
		m=2	6.77	3.63	1.64	0.95	0.46
		m=3	5.77	2.74	1.07	0.49	0.2
		m=4	4.84	2.29	0.89	0.4	0.15
		m=5	5.01	1.99	0.82	0.35	0.14
		$\mathcal{O}(\log \mathcal{T})$	3.48	1.55	0.64	0.27	0.11
	General Case	m=2	12.66	9.87	8.09	7.39	4.36
Logit Response Probability	Well-separated Case	m=1	21.23	11.44	9.14	5.04	3.07
		m=2	16.22	8.57	5.25	2.96	1.97
		m=3	16.67	8.55	3.92	2.21	1.35
		m=4	15.62	8.58	3.90	2.18	1.15
		m=5	15.34	8.28	3.73	1.80	1.06
		$\mathcal{O}(\log \mathcal{T})$	17.11	8.67	3.82	1.77	1.06
	General Case	m=2	7.87	5.62	4.09	3.61	1.76

From Table 4.1, it is seen that, when $T = 300$ most percentages of profit loss are below 10% with only one exception, and when $T = 3000$, most percentages of profit loss are below 5% with four exceptions. For the well-separated case, within each column, it is seen that more significant improvement can be achieved by adding one more price change when there are initially very few price changes allowed; and allowing more price changes, though in general improves the performance of the algorithm, has diminishing effect.

4.5 Conclusion

In most real world applications it is unlikely that the firm has complete information of the distribution of customer demand, hence learning is an important task for the firm's decision making process. In this chapter we consider a dynamic joint pricing and inventory control problem in which the firm has little or no prior knowledge about the distribution of customer demand and that, due to business constraints or associated cost for making price changes, the firm is prevented from conducting extensive price experimentations. We consider several scenarios and develop learning algorithms that satisfy the constraints on the number of price experimentations. We derive the regrets for these learning algorithms and show that they are the best possible in the sense that, the rates of regrets have the same magnitude as the lower bounds. Numerical results show that the algorithms perform very well and quickly converge to that of the optimal solutions as the planning horizon becomes long.

After this work was completed, it was brought to our attention that *Broder* (2011) obtained similar results in his doctoral dissertation for a pure dynamic pricing problem[1]. *Broder* (2011) considers infinite initial inventory (thus no inventory decision and no censored demand) and a single customer arriving in each period (Bernoulli demand process), and applies the MLE to estimate the unknown parameters. In our model, we have a general demand process, the firm makes replenishment decision in addition to pricing decision in every period, and the demand is censored. Therefore, we have to explore the inventory space to learn the impact of inventory decision on demand parameter estimation. Because of demand censoring, the convergence result of the standard MLE cannot be applied to our model, hence we develop a modified MLE for censored data and show that it preserves the same convergence rate as the standard MLE method, that is new to the literature, to establish the regret rate of

[1]The authors are grateful to Profs. Omar Besbes and Paat Rusmevichientong for bringing this work to our attention.

our algorithm.

In this study we consider the scenario where the customer responses are drawn from a parametric class of distributions, which is possible if the firm has prior experience with similar and/or relevant products and has formed a knowledge base about the set of possible customer responses. If that is not the case, e.g., new product just released to the market, then the estimation of customer response will become a nonparametric problem. Nonparametric demand estimation for inventory control, revenue management, and joint pricing and inventory control problems have been studied in the literature, see e.g., *Levi et al.* (2007, 2010), *Huh and Rusmevichientong* (2009), *Huh et al.* (2011), *Besbes and Zeevi* (2009, 2015), and *Chen et al.* (2015). It is interesting future work to extend our study to the case with nonparametric customer responses to selling prices.

We end this section by elaborating on the technical issue with applying maximum likelihood method to dependent random samples. Recall that in our first algorithm we only use sales data from latest stage, instead of all the previous stages. There are a few papers in the operations literature that utilize all previous data points in MLE method and provide convergence rate result. In these studies the authors explore some special demand structure, and establish the results under specific conditions. For example, *Broder and Rusmevichientong* (2012) consider a revenue management problem with Bernoulli demand, and *den Boer and Zwart* (2014) impose conditions on the mean and variance of the demand distribution. In our problem, due to demand censoring and the general form of the demand process, we can establish the convergence rate of the modified MLE at stage i only when using data points from the most recent stage $i-1$, but not all previous stages $1,\ldots,i-1$. If we do include the data points from all previous stages, then it would require that $-\sum_{t=1}^{t_1}\log\tilde{f}_{y_t}(d_t;p_t,z)$, where $\tilde{f}_{y_t}(d_t;p_t,z)$ is $f(d_t;p_t,z)$ when $d_t<y_t$, and $1-F(y_t-1;p_t,z)$ when $d_t\geq y_t$, be convex in z for any $p_t\in\mathcal{P}$, $d_t\in\{d^l,d^l+1,\ldots\}$, and $1\leq t_1\leq T$ (a similar condition was imposed

in *Broder and Rusmevichientong* (2012) for the Bernoulli demand setting without demand censoring). This condition on $f(\cdot; p, z)$ for our general demand setting is clearly quite restrictive. It is an interesting future research to explore under what relaxed conditions the MLE of dependent and non-identically distributed samples enjoys similar convergence rate as that of MLE under i.i.d. assumptions.

4.6 Appendix

Appendix A

Proof of Theorem IV.1. From (4.4), let $\epsilon^2 = K_{15}^{-1} c^{-1} \log c$, then there exists a constant $K_{16} > 0$ such that

$$\mathbb{P}\left\{\|\hat{\mathbf{z}} - \mathbf{z}\|^2 < K_{15}^{-1} c^{-1} \log c\right\} > 1 - \frac{K_{16}}{c}.$$

Define event

$$\mathcal{A}_1 = \left\{\omega : \|\hat{\mathbf{z}} - \mathbf{z}\|^2 < K_{15}^{-1} c^{-1} \log c\right\}.$$

Then

$$\mathbb{P}(\mathcal{A}_1) > 1 - \frac{K_{16}}{c}. \tag{4.12}$$

One has

$$\begin{aligned} & \mathbb{E}[H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - H(p^*(\hat{\mathbf{z}}), y^*(\hat{\mathbf{z}}), \mathbf{z})] \\ = \quad & \mathbb{P}(\mathcal{A}_1)\mathbb{E}\left[H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - H(p^*(\hat{\mathbf{z}}), y^*(\hat{\mathbf{z}}), \mathbf{z})\middle|\mathcal{A}_1\right] \\ & + (1 - \mathbb{P}(\mathcal{A}_1))\,\mathbb{E}\left[H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - H(p^*(\hat{\mathbf{z}}), y^*(\hat{\mathbf{z}}), \mathbf{z})\middle|\mathcal{A}_1^c\right]. \end{aligned} \tag{4.13}$$

On $\mathcal{A}_1$, when c is large enough, one will have $\|\hat{\mathbf{z}} - \mathbf{z}\| < \delta$, thus by Assumption A v) one has

$$\begin{aligned} & \mathbb{P}(\mathcal{A}_1)\mathbb{E}\left[H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - H(p^*(\hat{\mathbf{z}}), y^*(\hat{\mathbf{z}}), \mathbf{z})\middle|\mathcal{A}_1\right] \\ = \quad & \mathbb{P}(\mathcal{A}_1)\mathbb{E}\left[H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - H(p^*(\hat{\mathbf{z}}), y^*(\mathbf{z}), \mathbf{z})\middle|\mathcal{A}_1\right]. \end{aligned} \tag{4.14}$$

To proceed, we apply Taylors expansion to $H(p, y^*(\mathbf{z}), \mathbf{z})$ at the maximizer $p = p^*(\mathbf{z})$. For a function $g(\mathbf{x}), \mathbf{x} \in R^n$, let $\mathcal{D}g(\mathbf{x})$ be the $1 \times n$ matrix of first order derivative of function $g(\mathbf{x})$, and $\mathcal{D}^2 g(\mathbf{x})$ be the Hessian Matrix of $g(\mathbf{x})$. Then for $p \in \mathcal{P}$, one has

$$\begin{aligned}
& H(p, y^*(\mathbf{z}), \mathbf{z}) \\
= \; & H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) + \mathcal{D}H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z})(p - p^*(\mathbf{z})) \\
& + \frac{3}{2}(p - p^*(\mathbf{z}))^T \int_0^1 (1-t)^2 \mathcal{D}^2 H(p^*(\mathbf{z}) + t(p - p^*(\mathbf{z})), y^*(\mathbf{z}), \mathbf{z}) dt \; (p - p^*(\mathbf{z})) \\
= \; & H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) \\
& + \frac{3}{2}(p - p^*(\mathbf{z}))^T \int_0^1 (1-t)^2 \mathcal{D}^2 H(p^*(\mathbf{z}) + t(p - p^*(\mathbf{z})), y^*(\mathbf{z}), \mathbf{z}) dt \; (p - p^*(\mathbf{z})),
\end{aligned} \tag{4.15}$$

where the equality holds because the first order derivative vanishes at the maximizer $p^*(\mathbf{z})$. Let $p = p^*(\hat{\mathbf{z}})$ in (4.15), then (4.14) satisfies, for some constants K_{17} to K_{20},

$$\begin{aligned}
& \mathbb{P}(\mathcal{A}_1)\mathbb{E}\left[H(p^*(\mathbf{z}), y^*(\mathbf{z}), \mathbf{z}) - H(p^*(\hat{\mathbf{z}}), y^*(\mathbf{z}), \mathbf{z}) \middle| \mathcal{A}_1\right] \\
= \; & \mathbb{P}(\mathcal{A}_1) \times \mathbb{E}\Bigg[\frac{3}{2}(p^*(\hat{\mathbf{z}}) - p^*(\mathbf{z}))^T \\
& \qquad \times \int_0^1 (1-t)^2 \mathcal{D}^2 H(p^*(\mathbf{z}) + t(p^*(\hat{\mathbf{z}}) - p^*(\mathbf{z})), y^*(\mathbf{z}), \mathbf{z}) dt \; (p^*(\hat{\mathbf{z}}) - p^*(\mathbf{z})) \middle| \mathcal{A}_1 \Bigg] \\
\leq \; & K_{17}\mathbb{P}(\mathcal{A}_1)\mathbb{E}\left[\|p^*(\hat{\mathbf{z}}) - p^*(\mathbf{z})\|^2 \middle| \mathcal{A}_1\right] \\
\leq \; & K_{18}\mathbb{P}(\mathcal{A}_1)\mathbb{E}\left[\|\hat{\mathbf{z}} - \mathbf{z}\|^2 \middle| \mathcal{A}_1\right] \\
\leq \; & K_{18}\mathbb{E}\left[\|\hat{\mathbf{z}} - \mathbf{z}\|^2\right] \\
\leq \; & K_{19} \int_{\epsilon=0}^{\infty} K_4 e^{-cK_5\epsilon} d\epsilon \\
= \; & \frac{K_{20}}{c},
\end{aligned} \tag{4.16}$$

where the first inequality follows the boundedness of second order derivative on $\mathcal{P}$ by Assumption A ii), and that for any real numbers a and b it holds that $ab \leq (a^2+b^2)/2$. The second inequality is justified by Assumption A iv), and the fourth inequality follows from (4.4).

Furthermore, we have, for some constant K_{21} and K_{22},

$$\begin{aligned} & (1-\mathbb{P}(\mathcal{A}_1))\,\mathbb{E}\left[H(p^*(\mathbf{z}),y^*(\mathbf{z}),\mathbf{z})-H(p^*(\hat{\mathbf{z}}),y^*(\hat{\mathbf{z}}),\mathbf{z})\big|\mathcal{A}_1^c\right] \\ \leq \quad & \frac{K_{21}}{c}(\|p^*(\mathbf{z})-p^*(\hat{\mathbf{z}})\|+\|y^*(\mathbf{z})-y^*(\hat{\mathbf{z}})\|) \\ \leq \quad & \frac{K_{22}}{c}, \end{aligned} \tag{4.17}$$

where the first inequality follows from (4.12) and Assumption A iii), and the second inequality is true because $\mathcal{P}$ and $\mathcal{Y}$ are bounded.

Combining (4.16) and (4.17) with (4.13), we complete the proof of Theorem IV.1. □

Proof of Proposition IV.3. Rewrite (4.5) as

$$\hat{z}_i=\operatorname*{argmax}_{z\in\mathcal{Z}}\prod_{t=t_i+1}^{t_i+I_i}\tilde{f}_{y_t}(\min\{d_t,y_t\};\hat{p}_i,z). \tag{4.18}$$

For $t\in\{t_i+1,\ldots,t_{i+1}\}$, one can see that y_t is nonincreasing and satisfies $y_t\geq\tilde{y}_i$. Let

$$\mathcal{C}_i=\{t:t_i+1\leq t\leq t_{i+1},y_t=\tilde{y}_i\} \tag{4.19}$$

and

$$\tilde{t}_i=\min\{t-1:t\in\mathcal{C}_i\}.$$

Next we analyze the following two cases separately: (1) $\tilde{t}_i\geq t_i+1$, and (2) $\tilde{t}_i=t_i$.

(1) If $\tilde{t}_i\geq t_i+1$, it means the initial inventory of stage i is greater than the target inventory order-up-to level, i.e., $x_{t_i+1}>\tilde{y}_i$. Then define

$$\tilde{\mathcal{C}}_i=\{t_i+1,\ldots,t_{i+1}\}-\mathcal{C}_i-\{\tilde{t}_i\},$$

and for any $t\in\tilde{\mathcal{C}}_i$, it can be seen that $d_t<y_t$, and this yields $\tilde{D}_{t,y_t}(\hat{p}_i,z)=D_t(\hat{p}_i,z)$, and

$$\tilde{f}_{y_t}(\min\{d_t,y_t\};\hat{p}_i,z)=f(d_t;\hat{p}_i,z),$$

which does not depend on y_t. Therefore, for given $\tilde{\mathcal{C}}_i$, $\tilde{D}_{t,y_t}(\hat{p}_i,z)=D(\hat{p}_i,z)$, $t\in\tilde{\mathcal{C}}_i$ are independent and each of them follows $f(\cdot;\hat{p}_i,z)$.

For any $t\in\mathcal{C}_i$, $y_t=\tilde{y}_i$, therefore $\tilde{D}_{t,y_t}(\hat{p}_i,z)=\min\{D_t(\hat{p}_i,z),\tilde{y}_i\}$ are independent,

following the probability mass function of $\tilde{f}_{\tilde{y}_i}(\cdot;\hat{p}_i,z)$.

For $t=\tilde{t}_i$, $\tilde{D}_{t,y_t}(\hat{p}_i,z)=\min\{D_t(\hat{p}_i,z),y_t\}$, and the probability mass function is

$$\tilde{f}_{y_t}(d_t;\hat{p}_i,z)=\begin{cases} f(d_t;\hat{p}_i,z) & \text{if } d_t<y_t,\\ 1-F(y_t-1;\hat{p}_i,z) & \text{if } d_t=y_t,\end{cases} \tag{4.20}$$

where $y_{\tilde{t}_i}=x_{t_i+1}-\sum_{j=t_i+1}^{\tilde{t}_i-1}d_j$ is random and it holds that $y_{\tilde{t}_i}>\tilde{y}_i$.

For any realization of $\tilde{\mathcal{C}}_i$ and $\mathcal{C}_i$, it can be seen that $\tilde{D}_{t,y_t}(\hat{p}_i,z), t\in\tilde{\mathcal{C}}_i\cup\mathcal{C}_i$ are independent random variables which follow two distinct distributions, and $\tilde{D}_{\tilde{t}_i,y_{\tilde{t}_i}}(\hat{p}_i,z)$ depends on $\tilde{D}_{t,y_t}(\hat{p}_i,z), t\in\tilde{\mathcal{C}}_i$.

Following the discussions above, (4.18) can be rewritten as

$$\hat{z}_i=\operatorname*{argmax}_{z\in\mathcal{Z}}\prod_{t\in\tilde{\mathcal{C}}_i}f(d_t;\hat{p}_i,z)\times\tilde{f}_{y_{\tilde{t}_i}}(\min(d_{\tilde{t}_i},y_{\tilde{t}_i}\};\hat{p}_i,z)\times\prod_{t\in\mathcal{C}_i}\tilde{f}_{\tilde{y}_i}(\min\{d_t,\tilde{y}_i\};\hat{p}_i,z). \tag{4.21}$$

Next we compare (4.21) with the following fictitious MLE formulation,

$$\tilde{z}_i=\operatorname*{argmax}_{z\in\mathcal{Z}}\prod_{t\in\tilde{\mathcal{C}}_i}f(d_t;\hat{p}_i,z)\times\prod_{t\in\{\tilde{t}_i\}\cup\mathcal{C}_i}\tilde{f}_{\tilde{y}_i}(\min\{d_t,\tilde{y}_i\};\hat{p}_i,z), \tag{4.22}$$

where $D_t(\hat{p}_i,z), t\in\tilde{\mathcal{C}}_i$ follows $f(\cdot;\hat{p}_i,z)$ and $\tilde{D}_{t,y_t}(\hat{p}_i,z), t\in\{\tilde{t}_i\}\cup\mathcal{C}_i$ follows $\tilde{f}_{\tilde{y}_i}(\cdot;\hat{p}_i,z)$, and they are all independent.

Comparing (4.22) and (4.21), the only difference is at period $t=\tilde{t}_i$. Divide the MLE formulation in (4.21) by that in (4.22) one has

$$\begin{aligned}&\frac{\prod_{t\in\tilde{\mathcal{C}}_i}f(d_t;\hat{p}_i,z)\times\tilde{f}_{y_{\tilde{t}_i}}(\min(d_{\tilde{t}_i},y_{\tilde{t}_i}\};\hat{p}_i,z)\times\prod_{t\in\mathcal{C}_i}\tilde{f}_{\tilde{y}_i}(\min\{d_t,\tilde{y}_i\};\hat{p}_i,z)}{\prod_{t\in\tilde{\mathcal{C}}_i}f(d_t;\hat{p}_i,z)\times\prod_{t\in\{\tilde{t}_i\}\cup\mathcal{C}_i}\tilde{f}_{\tilde{y}_i}(\min\{d_t,\tilde{y}_i\};\hat{p}_i,z)}\\ =\quad&\frac{\tilde{f}_{y_{\tilde{t}_i}}(\min(d_{\tilde{t}_i},y_{\tilde{t}_i}\};\hat{p}_i,z)}{\tilde{f}_{\tilde{y}_i}(\min(d_{\tilde{t}_i},\tilde{y}_i\};\hat{p}_i,z)}.\end{aligned}$$

If $d_{\tilde{t}_i}\in\{d^l,\ldots,\tilde{y}_i-1\}$, then one has

$$\frac{\tilde{f}_{y_{\tilde{t}_i}}(\min(d_{\tilde{t}_i},y_{\tilde{t}_i}\};\hat{p}_i,z)}{\tilde{f}_{\tilde{y}_i}(\min(d_{\tilde{t}_i},\tilde{y}_i\};\hat{p}_i,z)}=\frac{f(d_{\tilde{t}_i};\hat{p}_i,z)}{f(d_{\tilde{t}_i};\hat{p}_i,z)}=1.$$

If $d_{\tilde{t}_i} \geq \tilde{y}_i$, then by Assumption 1 ii), there exists constants $K_{23}, K_{24} > 0$ such that

$$K_{23} \leq \frac{\tilde{f}_{y_{\tilde{t}_i}}(\min(d_{\tilde{t}_i}, y_{\tilde{t}_i}\}; \hat{p}_i, z)}{\tilde{f}_{\tilde{y}_i}(\min(d_{\tilde{t}_i}, \tilde{y}_i\}; \hat{p}_i, z)} \leq K_{24}. \tag{4.23}$$

There exist constants $-\infty < K_{25} \leq K_{26} < +\infty$ such that, for any $p \in \mathcal{P}$, any $z \in \mathcal{Z}$, and any $d \in \{d^l, d^l + 1, \ldots, y^h\}$,

$$K_{25} \leq f'_z(d; p, z) \leq K_{26}. \tag{4.24}$$

Based on (4.21), we let

$$\begin{aligned}
\hat{Z}(z) &= \prod_{t \in \tilde{\mathcal{C}}_i} f(d_t; \hat{p}_i, z) \times \tilde{f}_{y_{\tilde{t}_i}}(\min(d_{\tilde{t}_i}, y_{\tilde{t}_i}\}; \hat{p}_i, z) \times \prod_{t \in \mathcal{C}_i} \tilde{f}_{\tilde{y}_i}(\min\{d_t, \tilde{y}_i\}; \hat{p}_i, z), \\
\hat{L}(z) &= \log \hat{Z}(z), \\
\hat{Z}(z, z+u) &= \frac{\hat{Z}(z+u)}{\hat{Z}(z)},
\end{aligned}$$

and based on (4.22), define

$$\begin{aligned}
\tilde{Z}(z) &= \prod_{t \in \tilde{\mathcal{C}}_i} f(d_t; \hat{p}_i, z) \times \prod_{t \in \{\tilde{t}_i\} \cup \mathcal{C}_i} \tilde{f}_{\tilde{y}_i}(\min\{d_t, \tilde{y}_i\}; \hat{p}_i, z), \\
\tilde{L}(z) &= \log \tilde{Z}(z), \\
\tilde{Z}(z, z+u) &= \frac{\tilde{Z}(z+u)}{\tilde{Z}(z)}.
\end{aligned}$$

Regardless of realizations of $d_t, t \in \{t_i + 1, \ldots, t_{i+1}\}$, $\tilde{t}_i$, $\tilde{\mathcal{C}}_i$, and $\mathcal{C}_i$, by (4.23) and (4.24), these exists constants $K_{27}, K_{28} > 0$ such that

$$E_z\left[\sqrt{\hat{Z}(z, z+u)}\right] \leq K_{27} E_z\left[\sqrt{\tilde{Z}(z, z+u)}\right], \tag{4.25}$$

and

$$|\hat{L}'(z+u)| \leq |\tilde{L}'(z+u)| + K_{28}. \tag{4.26}$$

(4.25) and (4.26) will serve as the key properties to analyze the performance of (4.21) through (4.22), as shown in what follows.

We will start with the definition of Hellinger distance $H_g(\theta_0, \theta)$, between two distributions $g(\cdot, \theta_0)$ and $g(\cdot, \theta)$, i.e.,

$$\begin{aligned} H_g(\theta_0, \theta) &= \sum_{\mathcal{R}} \left(\sqrt{g(x, \theta_0)} - \sqrt{g(x, \theta)}\right)^2 dx \\ &= 2\left(1 - \sum_{\mathcal{R}} \sqrt{g(x, \theta_0) g(x, \theta_1)} dx\right). \end{aligned}$$

Then clearly,

$$\sum_{\mathcal{R}} \sqrt{g(x, \theta_1) g(x, \theta_0)} dx = 1 - \frac{1}{2} H(\theta_0, \theta_1). \tag{4.27}$$

By *Borovkov* (1998) Theorem 31.3, if there exists constants $0 < K_{29} < K_{30} < +\infty$ such that the Fisher information for $g(\cdot, \theta)$, for any $\theta \in \Theta$, is bounded as follows:

$$K_{29} < I_g(\theta) = \sum_{x \in \mathcal{R}} \frac{(\partial g(x, \theta)/\partial \theta)^2}{g(x, \theta)} < K_{30},$$

and if the distribution is identifiable, then there exists some constant $a > 0$ such that for

$$H_g(\theta_0, \theta_1) = \sum_{x \in \mathcal{R}} \left(\sqrt{g(x, \theta_0)} - \sqrt{g(x, \theta)}\right)^2 \geq a(\theta_0 - \theta_1)^2.$$

By Assumption 1 (i) and the compactness of $\mathcal{P}$ and $\mathcal{Z}$, for any realizations of $\hat{p}_i \in \mathcal{P}$ and any $z \in \mathcal{Z}$, the Fisher information of $f(\cdot; \hat{p}_i, z)$ satisfies

$$\underline{c}_f < I_f(\hat{p}_i, z) = \sum_{d=d^l}^{+\infty} \frac{(\partial f(d; \hat{p}_i, z)/\partial z)^2}{f(d; \hat{p}_i, z)} < \bar{c}_f. \tag{4.28}$$

On the other hand,

$$\begin{aligned}
& \sum_{d=d^l}^{\tilde{y}_i} \frac{(\partial \tilde{f}_{\tilde{y}_i}(\min\{d,\tilde{y}_i\};\hat{p}_i,z)/\partial z)^2}{\tilde{f}(\min\{d,\tilde{y}_i\};\hat{p}_i,z)} \\
= & \sum_{d=d^l}^{\tilde{y}_i-1} \frac{(\partial f(d;\hat{p}_i,z)/\partial z)^2}{f(d;\hat{p}_i,z)} + \frac{\left(-\sum_{d=d^l}^{\tilde{y}_i-1}\partial f(d;\hat{p}_i,z)/\partial z\right)^2}{1-\sum_{d=d^l}^{\tilde{y}_i-1} f(d;\hat{p}_i,z)} \\
\leq & \sum_{d=d^l}^{\tilde{y}_i-1} \frac{(\partial f(d;\hat{p}_i,z)/\partial z)^2}{f(d;\hat{p}_i,z)} + \frac{2\sum_{d=d^l}^{\tilde{y}_i-1}(\partial f(d;\hat{p}_i,z)/\partial z)^2}{1-\sum_{d=d^l}^{\tilde{y}_i-1} f(d;\hat{p}_i,z)}.
\end{aligned}$$

By (4.28),

$$\sum_{d=d^l}^{\tilde{y}_i-1} \frac{(\partial f(d;\hat{p}_i,z)/\partial z)^2}{f(d;\hat{p}_i,z)} < \bar{c}_f,$$

then

$$\sum_{d=d^l}^{\tilde{y}_i-1} (\partial f(d;\hat{p}_i,z)/\partial z)^2 < \bar{c}_f,$$

and by Assumption 1 (ii),

$$\frac{\sum_{d=d^l}^{\tilde{y}_i-1}(\partial f(d;\hat{p}_i,z)/\partial z)^2}{1-\sum_{d=d^l}^{\tilde{y}_i-1} f(d;\hat{p}_i,z)} < \frac{\bar{c}_f}{\underline{f}}.$$

Therefore for any realizations of $\tilde{y}_i \in \{d^l+1,\ldots,y^h\}$, the Fisher information of $\tilde{f}_{\tilde{y}_i}(\cdot;\hat{p}_i,z)$ satisfies

$$\underline{c}_f < I_{\tilde{f}}(\hat{p}_i,z) = \sum_{d=d^l}^{\tilde{y}_i} \frac{(\partial \tilde{f}_{\tilde{y}_i}(\min\{d,\tilde{y}_i\};\hat{p}_i,z)/\partial z)^2}{\tilde{f}(\min\{d,\tilde{y}_i\};\hat{p}_i,z)} < \bar{c}_f + \frac{\bar{c}_f}{\underline{f}}. \tag{4.29}$$

By Assumption 1 (iii), both $f(\cdot;\hat{p}_i,z)$ and $\tilde{f}_{\tilde{y}_i}(\cdot;\hat{p}_i,z)$ are identifiable.
Hence there exists a constant $a>0$ such that one has

$$H_f(z,z+u) = \sum_{d=d^l}^{+\infty} \left(\sqrt{f(d;\hat{p}_i,z)} - \sqrt{f(d;\hat{p}_i,z+u)}\right)^2 \geq au^2, \tag{4.30}$$

and

$$H_{\tilde{f}}(z, z+u) = \sum_{d=d^l}^{\tilde{y}_i} \left(\sqrt{\tilde{f}_{\tilde{y}_i}(d; \hat{p}_i, z)} - \sqrt{\tilde{f}_{\tilde{y}_i}(d; \hat{p}_i, z+u)}\right)^2 \geq au^2. \tag{4.31}$$

Furthermore, it can be seen that

$$\begin{aligned} E_z\left[\sqrt{\frac{f(d; \hat{p}_i, z+u)}{f(d; \hat{p}_i, z)}}\right] &= \sum_{d=d^l}^{+\infty} \sqrt{\frac{f(d; \hat{p}_i, z+u)}{f(d; \hat{p}_i, z)}} f(d; \hat{p}_i, z) \\ &= \sum_{d=d^l}^{+\infty} \sqrt{f(d; \hat{p}_i, z+u) f(d; \hat{p}_i, z)} \\ &= 1 - \frac{1}{2} H_f(z, z+u), \end{aligned}$$

where the last equality follows from (4.27). Similarly one has

$$\begin{aligned} & E_z\left[\sqrt{\frac{\tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z+u)}{\tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z)}}\right] \\ = & \sum_{d=d^l}^{\tilde{y}_i} \sqrt{\frac{\tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z+u)}{\tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z)}} \tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z) \\ = & \sum_{d=d^l}^{\tilde{y}_i} \sqrt{\tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z+u) \tilde{f}_{\tilde{y}_i}(\min\{d, \tilde{y}_i\}; \hat{p}_i, z)} \\ = & 1 - \frac{1}{2} H_{\tilde{f}}(z, z+u). \end{aligned}$$

Because the demands in (4.22) are independent, and by the inequality $\log(1-x) \leq -x$ for $x < 1$, one has

$$\begin{aligned} E_z\left[\sqrt{\hat{Z}(z, z+u)}\right] &\leq K_2 E_z\left[\sqrt{\tilde{Z}(z, z+u)}\right] \\ &= K_2 \prod_{t=t_i+1}^{\tilde{t}_i} \left(1 - \frac{1}{2} H_f(\theta, z+u)\right) \prod_{t=\tilde{t}_i+1}^{t_{i+1}} \left(1 - \frac{1}{2} H_{\tilde{f}}(\theta, z+u)\right) \\ &= K_2 e^{\sum_{t=t_i+1}^{\tilde{t}_i} \log\left(1 - \frac{1}{2} H_f(z, z+u)\right) + \sum_{t=\tilde{t}_i+1}^{t_{i+1}} \log\left(1 - \frac{1}{2} H_{\tilde{f}}(z, z+u)\right)} \\ &\leq K_2 e^{-\frac{1}{2}\left(\sum_{t=t_i+1}^{\tilde{t}_i} H_f(z, z+u) + \sum_{t=\tilde{t}_i+1}^{t_{i+1}} H_{\tilde{f}}(z, z+u)\right)} \\ &\leq K_2 e^{-\frac{1}{2} a I_i u^2}, \end{aligned} \tag{4.32}$$

where the first inequality follows from (4.25), and the third inequality follows from (4.30) and (4.31).

For convenience, let

$$\hat{\kappa}(u) = \left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{3/4} = (\hat{Z}(z, z+u))^{3/4},$$

and

$$\tilde{\kappa}(u) = \left(\frac{\tilde{Z}(z+u)}{\tilde{Z}(z)} \right)^{3/4} = (\tilde{Z}(z, z+u))^{3/4}.$$

By Cauchy inequality $(E[|XY|])^2 \leq E[|X|^2 E[|Y|^2]$ for random variables X and Y, and $E_z[\tilde{Z}(z+u)/\tilde{Z}(z)] = 1$, it follows that there exists a constant $K_{31} > 0$ such that

$$\begin{aligned} E_z[\hat{\kappa}(u)] &\leq K_{31} E_z[\tilde{\kappa}(u)] \\ &= K_{31} E_z[(\tilde{Z}(z, z+u))^{1/2} (\tilde{Z}(z, z+u))^{1/4}] \\ &\leq K_{31} \Big(E_z[\tilde{Z}(z, z+u)]) \Big)^{1/2} \Big(E_z[(\tilde{Z}(z, z+u)])^{1/2}) \Big)^{1/2} \\ &= K_{31} \Big(E_z \left[(\tilde{Z}(z, z+u))^{1/2} \right] \Big)^{1/2} \\ &\leq K_{31} e^{-aI_i u^2/4}, \end{aligned} \tag{4.33}$$

where the first inequality follows from (4.25), and the last inequality follows from (4.32).

And note that

$$\begin{aligned} \hat{\kappa}'(u) &= \frac{3}{4} \left(\frac{\hat{Z}'(z+u)}{\hat{Z}(z+u)} \right) \left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{3/4} \\ &= \frac{3}{4} \hat{L}'(z+u) \left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{3/4}. \end{aligned}$$

Furthermore, one has

$$
\begin{aligned}
& E_z[|\hat{\kappa}'(u)|] \\
= \; & \frac{3}{4} E_z \left\{ \left[\left| \hat{L}'(z+u) \right| \left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{1/2} \right] \left[\left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{1/4} \right] \right\} \\
\leq \; & \frac{3}{4} \left(E_z \left[(\hat{L}'(z+u))^2 \frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right] E_z \left[\left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{1/2} \right] \right)^{1/2} \\
\leq \; & \frac{3}{4} \left(E_z \left[(|\tilde{L}'(z+u)| + K_{32})^2 K_{33} \frac{\tilde{Z}(z+u)}{\tilde{Z}(z)} \right] E_z \left[\left(\frac{\hat{Z}(z+u)}{\hat{Z}(z)} \right)^{1/2} \right] \right)^{1/2} \\
\leq \; & \frac{3}{4} \left(\left(E_{z+u} \left[(\tilde{L}'(z+u))^2 \right] + K_{34} \right) E_z \left[\sqrt{\hat{Z}(z, z+u)} \right] \right)^{1/2} \\
\leq \; & \frac{3}{4} K_{35} \left((\tilde{t}_i - 1) I_f(\hat{p}_i, z+u) + (I_i - \tilde{t}_i + 1) I_{\tilde{f}}(\hat{p}_i, z+u) \right)^{1/2} e^{-aI_i u^2/4} + K_{36} e^{-aI_i u^2/4} \\
\leq \; & K_{37} I_i^{1/2} e^{-aI_i u^2/4}, \qquad (4.34)
\end{aligned}
$$

where the second inequality follows from (4.25) and (4.26), the fourth inequality follows from (4.32) and the definition of $I_f(\hat{p}_i, z+u)$ and $I_{\tilde{f}}(\hat{p}_i, z+u)$, and last inequality is true by (4.28) and (4.29) when I_i is large enough.

In addition, for any $c > 0$, one has

$$
\hat{\kappa}(t) = \hat{\kappa}(c/\sqrt{I_i}) + \int_{c/\sqrt{I_i}}^{t} \hat{\kappa}'(u) du.
$$

Furthermore, for $x > 0$, it holds that

$$
1 - \Phi(x) \geq e^{-x^2/2},
$$

where Φ denotes the cumulative distribution function of the standard normal random variable.

Therefore, for any $t > 0$,

$$\begin{aligned}
E_z\left[\sup_{t\geq c/\sqrt{I_i}}\hat{\kappa}(t)\right] &= \mathbb{E}\left[\hat{\kappa}(c/\sqrt{I_i})\right] + \mathbb{E}\left[\sup_{t\geq c/\sqrt{I_i}}\int_{c/\sqrt{I_i}}^{t}\hat{\kappa}'(u)du\right] \\
&\leq \mathbb{E}\left[\hat{\kappa}(c/\sqrt{I_i})\right] + \mathbb{E}\left[\sup_{t\geq c/\sqrt{I_i}}\int_{c/\sqrt{I_i}}^{\infty}|\hat{\kappa}'(u)|du\right] \\
&\leq \mathbb{E}\left[\hat{\kappa}(c/\sqrt{I_i})\right] + \int_{c/\sqrt{I_i}}^{\infty}K_{37}I_i^{1/2}e^{-aI_iu^2/4}du \\
&= \mathbb{E}\left[\hat{\kappa}(c/\sqrt{I_i})\right] + 2K_{37}\sqrt{\frac{\pi K_{30}}{a}}\left(1-\Phi\big(c\sqrt{a/2}\big)\right) \\
&\leq e^{-ac^2/4} + 2K_{37}\sqrt{\frac{\pi K_{30}}{a}}e^{-ac^2/4} \\
&= \left(1+2K_{37}\sqrt{\frac{\pi K_{30}}{a}}\right)e^{-ac^2/4},
\end{aligned}$$

where the second inequality follows from (4.34), and the fourth inequality follows from (4.33).

Similar analysis shows that, for $t < 0$,

$$E_z\left[\sup_{t\leq -c/\sqrt{I_i}}\hat{\kappa}(t)\right] \leq \left(1+2K_{37}\sqrt{\frac{\pi K_{30}}{a}}\right)e^{-ac^2/4}.$$

This proves

$$E_z\left[\sup_{|t|\geq c/\sqrt{I_i}}\hat{\kappa}(t)\right] \leq \left(1+2K_{37}\sqrt{\frac{\pi K_{30}}{a}}\right)e^{-ac^2/4}.$$

Now one has

$$\begin{aligned} P_z\left\{\sup_{|t|\geq c/\sqrt{I_i}}\frac{\hat{L}(z+u)}{\hat{L}(z)}\geq 1\right\} &= P_z\left\{\sup_{|t|\geq c/\sqrt{I_i}}\left(\frac{\hat{L}(z+u)}{\hat{L}(z)}\right)^{3/4}\geq 1\right\} \\ &= P_z\left\{\sup_{|t|\geq c/\sqrt{I_i}}\hat{\kappa}(t)\geq 1\right\} \\ &\leq E_z\Big[\sup_{|t|\geq c/\sqrt{I_i}}\hat{\kappa}(t)\Big] \\ &\leq \left(1+2K_{37}\sqrt{\frac{\pi K_{30}}{a}}\right)e^{-ac^2/4}. \end{aligned}$$

Finally, we have the following relationship to complete the proof of the important result for maximum likelihood estimator:

$$\begin{aligned} P_z\{\sqrt{I_i}|z-\hat{z}_i|\geq c\} &= P\left\{\sup_{|t|\geq c/\sqrt{I_i}}\frac{\hat{L}(z+u)}{\hat{L}(z)}\geq \sup_{|t|\leq c/\sqrt{I_i}}\frac{\hat{L}(z+u)}{\hat{L}(z)}\right\} \\ &\leq P\left\{\sup_{|t|\geq c/\sqrt{I_i}}\frac{\hat{L}(z+u)}{\hat{L}(z)}\geq \frac{\hat{L}(z)}{\hat{L}(z)}=1\right\} \\ &\leq Ke^{-ac^2/4}. \end{aligned}$$

(2) If $\tilde{y}_i = t_i + 1$, then $x_{t_i+1} \leq \tilde{y}_i$, and $\tilde{y}_i$ is achieved for every period during stage i. Thus,

$$\hat{z}_i = \underset{z\in\mathcal{Z}}{\operatorname{argmax}}\prod_{t=I_i+1}^{t_i+I_i}\tilde{f}_{\tilde{y}_i}(\min\{d_t,\tilde{y}_i\};\hat{p}_i,z). \tag{4.35}$$

The probability mass function of $D_{t,\tilde{y}_i}(d_t;\hat{p}_i,z)$ is used for every period, and they are independent across periods. The standard MLE result in Borovkov (1998) Theorem 36.3 can be directly applied to prove the result in Proposition IV.3. Combing (1) and (2) completes the proof of Proposition IV.3. □

Proposition A1 below bounds the regret from missing inventory targets in Theorem IV.2 and Theorem IV.6.

Proposition A1 (Regret from Missing Inventory Targets). Consider a total of l periods, and the initial inventory level of period 1 is $x_1 \in \mathcal{Y}$. If for price $\hat{p} \in \mathcal{P}$ and any order-up-to level $\hat{y} \in \mathcal{Y}$, the algorithm sets

$$\begin{aligned} p_t &= \hat{p}, && t = 1,\ldots,l, \\ y_t &= \max\{x_t,\hat{y}\}, && t = 1,\ldots,l. \end{aligned}$$

Then, one has

$$\sum_{t=1}^{l} \mathbb{E}[|L(\hat{p}, \hat{y}) - L(p_t, y_t)|] \leq l^{\frac{1}{n}},$$

for any $n \geq 2$.

Proof. Let $l_1 = l^{\frac{1}{n}}$. Note that $D_t(\hat{p}, \mathbf{z}) \in [0, d^h]$, by Hoeffding's inequality (4.64) in Appendix B, let $\epsilon = d^h\left(\frac{n}{2} l_1 \log l_1\right)^{1/2}$ and we obtain

$$\mathbb{P}\left\{\sum_{t=1}^{l_1} D_t(\hat{p}, \mathbf{z}) \geq l_1\, \mathbb{E}[D_1(\hat{p}, \mathbf{z})] - d^h\left(\frac{n}{2} l_1 \log l_1\right)^{1/2}\right\} \geq 1 - \frac{1}{l_1^n}.$$

Because $\mathbb{E}[D_1(\hat{p}, \mathbf{z})] > 0$ for any $\hat{p} \in \mathcal{P}$, then when l is large enough, it will hold uniformly that

$$\frac{1}{2} l_1\, \mathbb{E}[D_1(\hat{p}, \mathbf{z})] \geq d^h\left(\frac{n}{2} l_1 \log l_1\right)^{1/2}.$$

Then, for large enough l,

$$\mathbb{P}\left\{\sum_{t=1}^{l_1} D_t(\hat{p}, \mathbf{z}) \geq \frac{1}{2} l_1\, \mathbb{E}[D_1(\hat{p}, \mathbf{z})]\right\} \geq 1 - \frac{1}{l_1^n}. \tag{4.36}$$

Define event

$$\mathcal{A}_2 = \left\{\omega : \sum_{t=1}^{l_1} D_t(\hat{p}, \mathbf{z}) \geq \frac{1}{2} l_1\, \mathbb{E}[D_1(\hat{p}, \mathbf{z})]\right\},$$

then (4.36) above can be restated as

$$\mathbb{P}\{\mathcal{A}_2\} \geq 1 - \frac{1}{l_1^n} = 1 - \frac{1}{l}. \tag{4.37}$$

Furthermore, when l is large enough, it will hold uniformly that

$$\frac{1}{2} l_1\, \mathbb{E}[D_1(\hat{p}, \mathbf{z})] \geq y^h - y^l.$$

Thus, when l is large enough, on the event $\mathcal{A}_2$ the inventory targets will be surely met after the first l_1 periods. To obtain the bound in Proposition A1, we proceed as follows:

$$\begin{aligned}
&\sum_{t=1}^{l}\mathbb{E}\big[|L(\hat{p},\hat{y})-L(p_t,y_t)|\big]\\
=\;&\mathbb{P}\left(\mathcal{A}_2\right)\sum_{t=1}^{l}\mathbb{E}\big[|L(\hat{p},\hat{y})-L(p_t,y_t)|\big|\mathcal{A}_2\big]\\
&+\left(1-\mathbb{P}\left(\mathcal{A}_2\right)\right)\sum_{t=1}^{l}\mathbb{E}\big[|L(\hat{p},\hat{y})-L(p_t,y_t)|\big|\mathcal{A}_2^C\big]. \qquad (4.38)
\end{aligned}$$

The first part in (4.38) is upper bounded by $K_{38}l_1$ for some constant K_{38} since on $\mathcal{A}_2$, inventory targets will be surely achieved after the first l_1 periods. By (4.37) the second part is upper bounded by $K_{39}l/l_1^n = K_{40}$. Hence one has, for some constant K_{41},

$$\sum_{t=1}^{l}\mathbb{E}\big[|L(\hat{p},\hat{y},z)-L(p_t,y_t,z)|\big]\le K_{38}l_1+K_{40}\le K_{41}l^{\frac{1}{n}}. \qquad (4.39)$$

Proposition A1 is thus proved. □

Proof of Theorem IV.2. By the definition of regret, we have

$$\begin{aligned}
R(T) &= \sum_{t=1}^{T}\mathbb{E}\left[G(p^*,y^*,z)-G(p_t,y_t,z)\right]\\
&= \sum_{t=t_1+1}^{t_2}\mathbb{E}\left[G(p^*,y^*,z)-G(p_t,y_t,z)\right]+\sum_{i=2}^{m+1}\sum_{t=t_i+1}^{t_{i+1}}\mathbb{E}\left[G(p^*,y^*,z)-G(p_t,y_t,z)\right]\\
&\le \underbrace{\sum_{t=t_1+1}^{t_2}\mathbb{E}\left[G(p^*,y^*,z)-G(p_t,y_t,z)\right]}_{\text{regret from initial decision}}+\underbrace{\sum_{i=2}^{m+1}\sum_{t=t_i+1}^{t_{i+1}}\mathbb{E}\left[G(p^*,y^*,z)-G(\hat{p}_i,\hat{y}_i,z)\right]}_{\text{regret from estimation error}}\\
&\quad+\underbrace{\sum_{i=2}^{m+1}\sum_{t=t_i+1}^{t_{i+1}}\mathbb{E}\left[|G(\hat{p}_i,\hat{y}_i,z)-G(\hat{p}_i,\tilde{y}_i,z)|\right]}_{\text{regret from exploration on the ordering decision}}+\underbrace{\sum_{i=2}^{m+1}\sum_{t=t_i+1}^{t_{i+1}}\mathbb{E}\left[|G(\hat{p}_i,\tilde{y}_i,z)-G(p_t,y_t,z)|\right]}_{\text{regret from missing inventory targets}}. \qquad (4.40)
\end{aligned}$$

By the existence of second order derivative, $\partial G(p,y,z)/\partial p$ is a continuous function

of p on $\mathcal{P}$. Then it follows from $\mathcal{P}$ is compact that

$$\alpha = \max_{p\in\mathcal{P}} \left| \frac{\partial G(p,y,z)}{\partial p} \right| < \infty.$$

Also it can be seen from (4.2) that

$$G(p,y,z) - G(p,y',z) \leq \max\{b,h\}\ |y-y'|. \tag{4.41}$$

The first term on the right hand side of (4.40) is bounded because, for some constant K_{42},

$$\begin{aligned} & \sum_{t=t_1+1}^{t_2} \mathbb{E}\left[G(p^*,y^*,z) - G(p_t,y_t,z)\right] \\ \leq & \sum_{t=t_1+1}^{t_2} \mathbb{E}\left(\alpha\left|p^*-p_t\right| + \max\{b,h\}\left|y^*-y_t\right|\right) \\ \leq & I_1\left(\alpha\left|p^h-p^l\right| + \max\{b,h\}\left|y^h-y^l\right|\right) \leq K_{42}I_1. \end{aligned} \tag{4.42}$$

To bound the second part in (4.40), we first analyze

$$\mathbb{E}\left[G(p^*,y^*,z) - G(\hat{p}_i,\hat{y}_i,z)\right].$$

According to Proposition IV.3 and Theorem IV.1,

$$\mathbb{E}\left[G(p^*,y^*,z) - G(\hat{p}_i,\hat{y}_i,z)\right] \leq \frac{K_{42}}{I_{i-1}},$$

which leads to

$$\sum_{i=2}^{m+1}\sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[G(p^*,y^*,z) - G(\hat{p}_i,\hat{y}_i,z)\right] \leq \sum_{i=2}^{m+1} \frac{K_{44}}{I_{i-1}} I_i \leq K_{43}T^{\frac{1}{m+1}}. \tag{4.43}$$

To bound the third part in (4.40), it can be seen that there exists a constant $\delta > 0$, such that

$$\mathbb{P}(\hat{y}_i \neq \tilde{y}_i) \leq \mathbb{P}(|\hat{z}_i - z| > \delta).$$

For any $i \geq 2$, when I_{i-1} grows, it will be true that $K_{45} I_{i-1}^{-1} \log I_{i-1} < \delta^2$. Therefore

$$\begin{aligned} \mathbb{P}(\hat{y}_i \neq \tilde{y}_i) &\leq \mathbb{P}(|\hat{z}_i - z| > \delta) \\ &\leq \mathbb{P}\{|\hat{z}_i - z|^2 \geq K_{45} I_{i-1}^{-1} \log I_{i-1}\} \\ &\leq \frac{K_{46}}{I_{i-1}}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} & \sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, \tilde{y}_i, z)|\right] \\ \leq & \sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \frac{K_{46}}{I_{i-1}} \mathbb{E}\left[\max\{b, h\} |\hat{y}_i - \tilde{y}_i|\right] \\ \leq & \sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \frac{K_{46} \max\{b, h\}}{I_{i-1}} \left(y^h - y^l\right) \\ \leq & K_{47} T^{\frac{1}{m+1}}. \end{aligned} \tag{4.44}$$

For each $i = 2, \ldots, m+1$, letting $n = i$ in Proposition A1 we obtain

$$\sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \tilde{y}_i, z) - G(p_t, y_t, z)|\right] \leq K_{48} I_i^{\frac{1}{i}} = K_{48} T^{\frac{1}{m+1}},$$

which renders

$$\sum_{i=2}^{m+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \tilde{y}_i, z) - G(p_t, y_t, z)|\right] \leq \sum_{i=2}^{m+1} K_{48} T^{\frac{1}{m+1}} = K_{48} m T^{\frac{1}{m+1}}. \tag{4.45}$$

Combing (4.42), (4.43),(4.44), and (4.45), one has

$$R(T) \leq K_{49} T^{\frac{1}{m+1}},$$

and Theorem IV.2 is proved. □

Proof of Theorem IV.4. We will consider the special case when $D(p, z)$ is binomial with success rate $r(p, z)$. Let $\mathcal{P} = [1/3, 1/2]$, $\mathcal{Y} = \{1\}$, $\mathcal{Z} = [2, 3]$, $r(p, z) = 1 - pz/2$, $h = 0$ and $b = 0$, and let $d^l = 0, d^h = 1$. In what follows, we prove that, for any joint pricing and inventory control policy ϕ that changes price no more than m times

$(m \geq 1)$ and any $T \geq 1$, there exists a $z \in \mathcal{Z}$ such that

$$R^{\phi}(T) \geq K_{50} T^{\frac{1}{m+1}}$$

for some positive constant $K_{50} > 0$.

Following *Broder and Rusmevichientong* (2012), we let z be a random variable with probability density function $f(z) = 2\{\cos(\pi(z - 5/2))\}^2$ on $\mathcal{Z}$. Recall that $p^*(z)$ is the optimal pricing policy for the complete information case. For any pricing and inventory control policy, the order-up-to level is 1, and because $h = 0$ and $b = 0$, both the holding cost and the shortage cost are 0. Therefore the firm only focuses on solving the revenue maximization problem of $\max_{p \in \mathcal{P}} p\, r(p, z)$.

With complete information of z, it can be seen that $p^*(z) = 1/z$, and $r(p^*(z), z) = 1/2 > 0$ for any $z \in \mathcal{Z}$, and

$$(p\, r(p, z))'_p = 1 - zp, \;\; (p\, r(p, z))''_p = -z. \tag{4.46}$$

In the data-driven optimization, inventory holding and shortage costs are both 0, therefore the firm only needs to learn z from historical data. Consider an arbitrary data-driven pricing and replenishment policy ϕ that allows at most m price changes. Let $\tau_i + 1$ denote the starting period for the i-th price with $\tau_1 = 0$, and $p_1 \in \mathcal{P}$ is the initial price at the beginning of period 1, then $1 \leq \tau_2 < \cdots < \tau_{m+1} \leq T - 1$ and the price at the beginning of period $\tau_i + 1$ is set at $p_{\tau_i + 1}$ for $i = 2, \ldots, m + 1$. For convenience let $\tau_{m+2} = T$. In case $p_{\tau_i + 1} = p_{\tau_{i-1} + 1}$ for some $i = 2, \ldots, m + 1$, then policy ϕ changes prices less than m times. For any $z \in \mathcal{Z}$, the regret of policy ϕ can be computed as

$$\begin{aligned} R^{\phi}(T) &= \sum_{t=1}^{T} \mathbb{E}[G(p^*(z), y^*(z), z) - G(p_t, y_t, z)] \\ &= \sum_{t=1}^{T} \mathbb{E}[p^*(z)\, r(p^*(z), z) - p_t\, r(p_t, z)] \\ &= \sum_{i=1}^{m+1} \sum_{t=\tau_i + 1}^{\tau_{i+1}} \mathbb{E}[p^*(z)\, r(p^*(z), z) - p_{\tau_i + 1}\, r(p_{\tau_i + 1}, z)], \end{aligned}$$

in which the expectation in the first equality is taken with respect to p_t, y_t and the binomial random variable, while the expectation in the second equality is taken with

respect to p_t. Hence

$$\begin{aligned}\sup_{z\in\mathcal{Z}} R^{\phi}(T) &\geq \sup_{z\in\mathcal{Z}} \sum_{i=1}^{m+1}\sum_{t=\tau_i+1}^{\tau_{i+1}} \mathbb{E}[p^*(z)\, r(p^*(z),z) - p_{\tau_i+1}\, r(p_{\tau_i+1},z)] \\ &\geq \sum_{i=1}^{m+1}\sum_{t=\tau_i+1}^{\tau_{i+1}} \mathbb{E}[p^*(z)\, r(p^*(z),z) - p_{\tau_i+1}\, r(p_{\tau_i+1},z)] \\ &= \mathbb{E}[p^*(z)\, r(p^*(z),z) - p_1\, r(p_1,z)]\tau_2 \\ &\quad + \sum_{i=2}^{m+1} \mathbb{E}[p^*(z)\, r(p^*(z),z) - p_{\tau_i+1}\, r(p_{\tau_i+1},z)](\tau_{i+1}-\tau_i), \quad (4.47)\end{aligned}$$

here the expectation in the first inequality is taken with respect to p_t, and the expectation in the second inequality is with respect to p_t and z, while in the equality, the first expectation is with respect to p_1 and z, and the second is with respect to p_t and z. Recall that z is distributed with pdf $f(z)$ on $\mathcal{Z}$.

Let

$$\gamma = \max_{z\in\mathcal{Z}}\big(p^*(z)\, r\,(p^*(z),z) - \mathbb{E}\big[p_1\, r(p_1,z)\big]\big) = \max_{z\in\mathcal{Z}}\left(\frac{1}{2z} - \mathbb{E}[p_1] + \frac{z\mathbb{E}[p_1^2]}{2}\right), \quad (4.48)$$

here the mathematical expectation is with respect to p_1 of policy ϕ. Since at the beginning of period 1, the firm has no information yet about customer response data, hence p_1 is not demand data-dependent, and it may be a random pricing policy. If z is known to the firm then p_1 is set to $1/z$ with probability one, then the right hand side of (4.48) (before maximizing over z) would be 0. Since z is not known a priori and that we take maximum over z in (4.48), we have $\gamma > 0$ ($\gamma = 0$ only if the firm knows the exact value of z a priori).

Denote

$$\overline{\mathcal{Z}} = \left\{z:\ \frac{1}{2z} - \mathbb{E}[p_1] + \frac{z\mathbb{E}[p_1^2]}{2} \geq \frac{\gamma}{2}\right\},$$

then $\mathbb{P}\{\overline{\mathcal{Z}}\} > 0$. Also, note that

$$\frac{1}{2z} - \mathbb{E}[p_1] + \frac{z\mathbb{E}[p_1^2]}{2} \geq \frac{1}{2z} - \mathbb{E}[p_1] + \frac{z(\mathbb{E}[p_1])^2}{2},$$

and the right hand side, as a function of $\mathbb{E}[p_1]$, is minimized when $\mathbb{E}[p_1] = 1/z$, at which point the right hand side is equal to 0. This shows that

$$\frac{1}{2z} - \mathbb{E}[p_1] + \frac{z\mathbb{E}[p_1^2]}{2} \geq 0.$$

Hence, by conditioning on $\mathcal{Z}$ and $\mathcal{Z}^C$, we obtain

$$\mathbb{E}[p^*(z)\, r(p^*(z), z) - p_1\, r(p_1, z)]\tau_2 \geq \frac{\gamma}{2}\mathbb{P}\{\overline{\mathcal{Z}}\}\tau_2. \tag{4.49}$$

Since the pricing problem in our specific example is the same as that of *Broder and Rusmevichientong* (2012), we follow the same argument as *Broder and Rusmevichientong* (2012) (see also *Gill and Levit* (1995) and *A.Goldenshluger and A.Zeevi* (2009)) to conclude that, there exists some constant $K_{51} > 0$ such that for any $t \geq 1$,

$$\mathbb{E}[(p^*(z) - p_{t+1})^2] \geq \frac{K_{51}}{t}.$$

Therefore,

$$\begin{aligned}
& \sum_{i=2}^{m+1} \mathbb{E}[p^*(z)\, r(p^*(z), z) - p_{\tau_i+1}\, r(p_{\tau_i+1}, z)](\tau_{i+1} - \tau_i) \\
\geq & \sum_{i=2}^{m+1} \mathbb{E}\left[K_{52}(p^*(z) - p_{\tau_i+1})^2\right](\tau_{i+1} - \tau_i) \\
\geq & \sum_{i=2}^{m+1} \frac{K_{53}}{\tau_i}(\tau_{i+1} - \tau_i) \\
= & \sum_{i=2}^{m+1} \left(K_{53}\frac{\tau_{i+1}}{\tau_i} - K_{53}\right),
\end{aligned} \tag{4.50}$$

where the first inequality follows from Taylor expansion at $p^*(z)$ to the second order, and by (4.46) the second order derivative of $pr(p, z)$ on p falls between $[-3, -2]$. Combining (4.49) and (4.50) with (4.47), we obtain, for some constant $K_2 > 0$, that

$$\begin{aligned}
& \sup_{z\in\mathcal{Z}} R^{\phi}(T) \\
\geq & \left(\frac{\gamma}{2}\mathbb{P}\{\overline{\mathcal{Z}}\}\tau_2 + \sum_{i=2}^{m+1} K_{53}\frac{\tau_{i+1}}{\tau_i}\right) - mK_{53} \\
\geq & (m+1)\left(\frac{\gamma}{2}\mathbb{P}\{\overline{\mathcal{Z}}\}\tau_2 \prod_{i=2}^{m+1}\left[K_{53}\frac{\tau_{i+1}}{\tau_i}\right]\right)^{\frac{1}{m+1}} - mK_{53} \\
\geq & K_{54}T^{\frac{1}{m+1}}
\end{aligned}$$

where the second inequality follows from arithmetic average is greater than or equal to geometric average for nonnegative real numbers. The proof for Theorem IV.4 is thus complete. □

Proof of Theorem IV.5. Similar as in Theorem IV.2, we divide the total regret as the following,

$$\begin{aligned} R(T) &= \sum_{t=1}^{T} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] \\ &= \sum_{t=t_1+1}^{t_2} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] + \sum_{i=2}^{N+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] \\ &\leq \underbrace{\sum_{t=t_1+1}^{t_2} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right]}_{\text{regret from initial decision}} + \underbrace{\sum_{i=2}^{N+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[G(p^*, y^*, z) - G(\hat{p}_i, \hat{y}_i, z)\right]}_{\text{regret from estimation error}} \\ &\quad + \underbrace{\sum_{i=2}^{N+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, \tilde{y}_i, z)|\right]}_{\text{regret from exploration on the ordering decision}} + \underbrace{\sum_{i=2}^{N+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \tilde{y}_i, z) - G(p_t, y_t, z)|\right]}_{\text{regret from missing inventory targets}}. \end{aligned} \tag{4.51}$$

The first part in (4.51) is bounded by, for some constant $K_{55} > 0$,

$$\sum_{t=t_1+1}^{t_2} \mathbb{E}\left[G(p^*, y^*, z) - G(p_t, y_t, z)\right] \leq K_{55} I_1 = K_{55} I_0 v. \tag{4.52}$$

We bound the second part in (4.51) by analyzing $\mathbb{E}\left[G(p^*, y^*, z) - G(\hat{p}_i, \hat{y}_i, z)\right]$. Based on similar analyses as Proposition IV.3, one has

$$\mathbb{P}\{|\hat{z}_i - z| \geq \xi\} \leq K_{56} e^{-K_{57} I_{i-1} \xi^2}.$$

Therefore by Theorem IV.1,

$$\mathbb{E}\left[G(p^*, y^*, z) - G(\hat{p}_i, \hat{y}_i, z)\right] \leq \frac{K_{58}}{I_{i-1}},$$

which leads to

$$\sum_{i=2}^{N+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[G(p^*, y^*, z) - G(\hat{p}_i, \hat{y}_i, z)\right] \leq \sum_{i=2}^{N+1} \frac{K_{58}}{I_{i-1}} I_i \leq \sum_{i=2}^{N+1} K_{58} v \leq K_{59} \log T. \tag{4.53}$$

The third part in (4.51) is upper bounded by $\mathbb{P}(\tilde{y}_i \neq \hat{y}_i) = \mathbb{P}\{|\hat{z}_{i-1} - z| > \delta\}$ for

some constant $\delta > 0$, and similar as in (4.44),

$$\mathbb{P}\{|\hat{z}_{i-1} - z| > \delta\} \leq \frac{2K_{60}}{I_{i-1}}.$$

Therefore,

$$\mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, \tilde{y}_i, z)|\right] \leq \frac{K_{61}}{I_{i-1}},$$

which renders

$$\sum_{i=2}^{N+1} \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, \tilde{y}_i, z)|\right] \leq \sum_{i=2}^{N+1} \frac{K_{61}}{I_{i-1}} I_i \leq K_{62} \log T. \tag{4.54}$$

The fourth part in (4.51) is upper bounded as the following. From the analyses of bounding the third part of regret, one has

$$\mathbb{P}\{\tilde{y}_{i-1} = \hat{y}_{i-1} = y^*\} \geq 1 - \frac{K_{63}}{I_{i-1}},$$

and

$$\mathbb{P}\{\tilde{y}_i = \hat{y}_i = y^*\} \geq 1 - \frac{K_{64}}{I_i}.$$

Therefore,

$$\mathbb{P}\{\hat{y}_{i-1} = y^*, \hat{y}_i = y^*\} \geq 1 - \frac{K_{65}}{I_{i-1}},$$

and accordingly we define

$$\mathcal{A}_3 = \{\omega : \hat{y}_{i-1} = y^*, \hat{y}_i = y^*\},$$

and one has

$$\mathbb{P}(\mathcal{A}_3) \geq 1 - \frac{K_{66}}{I_{i-1}}. \tag{4.55}$$

Next we consider the event that $\hat{y}_{i-1}$ is achieved during periods $t_{i-1}+1,\ldots,t_i$. By Hoeffding inequality (4.64), one has

$$\mathbb{P}\left\{\sum_{t=t_{i-1}+1}^{t_i} D_t(\hat{p}_{i-1},z) - I_{i-1}\mathbb{E}[D_t(\hat{p}_{i-1},z)] \geq d^h I_{i-1}^{1/2}(\log I_{i-1})^{1/2}\right\} \geq 1-\frac{1}{I_{i-1}^2}.$$

Because $\mathbb{E}[D_t(\hat{p}_{i-1},z)]>0$ for any $\hat{p}_{i-1}\in\mathcal{P}$, then when I_{i-1} is large enough,

$$\frac{1}{2}I_{i-1}\mathbb{E}[D_t(\hat{p}_{i-1},z)] > d^h I_{i-1}^{1/2}(\log I_{i-1})^{1/2},$$

therefore define

$$\mathcal{A}_4 = \left\{\omega : \sum_{t=t_{i-1}+1}^{t_i} D_t(\hat{p}_{i-1},z) \geq \frac{1}{2}I_{i-1}\,\mathbb{E}[D_t(\hat{p}_{i-1},z)]\right\},$$

and one has

$$\mathbb{P}(\mathcal{A}_4) \geq 1-\frac{1}{I_{i-1}^2}. \tag{4.56}$$

Moreover, when i is large enough, we will also have

$$\frac{1}{2}I_{i-1}\mathbb{E}[D_t(\hat{p}_{i-1},z)] \geq y^h - y^l,$$

hence the target inventory level in stage $i-1$ will eventually be met, and the starting inventory level at the beginning of stage i (in period t_i+1) is at most $\hat{y}_{i-1}$. This implies that the target inventory level in stage i, $\hat{y}_i$, will always be met if $\hat{y}_i \geq \hat{y}_{i-1}$, and the only possibility for ever missing target in stage i is when $\hat{y}_i < \hat{y}_{i-1}$. In this case, for $t_i+1 \leq t \leq t_{i+1}$, we argue that the following relationship holds:

$$|\hat{y}_i - y_t| \leq |\hat{y}_i - \hat{y}_{i-1}|. \tag{4.57}$$

This is because, if $\hat{y}_{i-1} \leq \hat{y}_i$ then the target $\hat{y}_i$ is always reached in stage i, hence $y_t = \hat{y}_i$ and the left hand side is equal to 0, thus (4.57) is obviously satisfied. On the other hand, if $\hat{y}_{i-1} > \hat{y}_i$, then the left hand side of (4.57) are not equal to 0 only when $\hat{y}_i$ is not reached hence $\hat{y}_{i-1} \geq y_t > \hat{y}_i$ (since the starting inventory level in stage i is no higher than $\hat{y}_{i-1}$), and in this case it is seen that (4.57) is also satisfied. Therefore,

one has

$$
\begin{aligned}
& \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(p_t, y_t, z)|\right] \\
= \quad & \mathbb{P}(\mathcal{A}_3 \cap \mathcal{A}_4) \left(\sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, y_t, z)| \big| \mathcal{A}_3 \cap \mathcal{A}_4\right] \right) \\
& +(1 - \mathbb{P}(\mathcal{A}_3 \cap \mathcal{A}_4)) \left(\sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, y_t, z)| \big| \mathcal{A}_3^c \cup \cap \mathcal{A}_4^c\right] \right) \\
\leq \quad & \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[|G(\hat{p}_i, \hat{y}_i, z) - G(\hat{p}_i, y_t, z)| \big| \mathcal{A}_3 \cap \mathcal{A}_4\right] + \frac{K_{67} I_i}{I_{i-1}} \\
\leq \quad & \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[\max\{h, b + \hat{p}_i\} |\hat{y}_i - y_t| \big| \mathcal{A}_3 \cap \mathcal{A}_4\right] + \frac{K_{67} I_i}{I_{i-1}} \\
\leq \quad & \sum_{t=t_i+1}^{t_{i+1}} \mathbb{E}\left[\max\{h, b + p^h\} |\hat{y}_i - \hat{y}_{i-1}| \big| \mathcal{A}_3 \cap \mathcal{A}_4\right] + \frac{K_{67} I_i}{I_{i-1}} \\
\leq \quad & K_{68},
\end{aligned}
$$

where the first inequality follows from (4.55) and (4.56), the third inequality follows from (4.57) when on event $\mathcal{A}_4$, and the last inequality is valid because on event $\mathcal{A}_3$ one has $\hat{y}_i = \hat{y}_{i-1}$.

Thus, the fourth part of regret in (4.51) is upper bounded by

$$\sum_{i=2}^{N+1} K_{68} \leq K_{69} \log T. \tag{4.58}$$

Combining (4.52), (4.53), (4.54) and (4.58) in (4.51) completes the proof of Theorem 4. □

Proof of Proposition IV.7.

Recall (4.10) is

$$\hat{\mathbf{z}} = \underset{\mathbf{z} \in \mathcal{Z}}{\operatorname{argmax}} \prod_{\{t \in \{1,\ldots,kI\}: y_t > d_t\}} f(d_t; p_t, \mathbf{z}) \cdot \prod_{\{t \in \{1,\ldots,kI\}: y_t \leq d_t\}} (1 - F(y_t - 1; p_t, \mathbf{z})), \tag{4.59}$$

and we will compare (4.59) with the following fictitious MLE formulation,

$$\tilde{\mathbf{z}} = \underset{\mathbf{z} \in \mathcal{Z}}{\operatorname{argmax}} \prod_{t=1}^{kI} f(d_t; p_t, \mathbf{z}), \tag{4.60}$$

where $D_t(p_t, \mathbf{z}), t \in \{1, \ldots, kI\}$ independently follows the distribution of $f(\cdot; p_t, \mathbf{z})$.

Let

$$\mathcal{C} = \{t :\in [1, \ldots, kI] : y_t \leq d_t\},$$

and let $|\mathcal{C}|$ denote the cardinality of $\mathcal{C}$. (4.59) and (4.60) are different only in periods $t \in \mathcal{C}$. Because demands are upper bounded by d^h, the "raising inventory" action is performed for at most $\left\lceil \log_{1+s} \frac{d^h}{y^l} \right\rceil$ times. In other words, there are at most $\left\lceil \log_{1+s} \frac{d^h}{y^l} \right\rceil$ stockout periods in the learning phase, hence one has

$$|\mathcal{C}| \leq \left\lceil \log_{1+s} \frac{d^h}{y^l} \right\rceil = K_{70}.$$

We will discuss the following two case separately, i.e., (1) $|\mathcal{C}| > 0$, and (2) $|\mathcal{C}| = 0$.

(1) If $|\mathcal{C}| > 0$, then divide (4.59) by (4.60) one has

$$\begin{aligned} & \frac{\prod_{\{t \in \{1,\ldots,kI\}: y_t > d_t\}} f(d_t; p_t, \mathbf{z}) \prod_{\{t \in \{1,\ldots,kI\}: y_t \leq d_t\}} (1 - F(y_t - 1; p_t, \mathbf{z}))}{\prod_{t=1}^{kI} f(d_t; p_t, \mathbf{z})} \\ = & \prod_{t \in \mathcal{C}} \frac{1 - F(y_t - 1; p_t, \mathbf{z})}{f(d_t; p_t, \mathbf{z})}. \end{aligned}$$

By Assumption 2 ii), it can also be seen that for any $t \in \mathcal{C}$, any $d_t \in \{d^l, \ldots, d^h\}$, any $p_t = \bar{p}_j, j = \{1, \ldots, k\}$, and $z \in \mathcal{Z}$, there exist constants $K_{71}, K_{72} > 0$, such that

$$K_{71} \leq \frac{1 - F(y_t - 1; p_t, \mathbf{z})}{f(d_t; p_t, \mathbf{z})} \leq K_{72},$$

therefore

$$K_{71} \leq \prod_{t \in \mathcal{C}} \frac{1 - F(y_t - 1; p_t, \mathbf{z})}{f(d_t; p_t, \mathbf{z})} \leq K_{70} K_{72},$$

which is parallel to (4.23).

There exist constants $-\infty < K_{73} \leq K_{74} < +\infty$ such that, for any $p \in \mathcal{P}$, any $z \in \mathcal{Z}$, and any $d \in \{d^l, \ldots, y^h\}$,

$$K_{73} \leq f'_z(d; p, \mathbf{z}) \leq K_{74},$$

which is in parallel to (4.24).

Then, based on (4.59) we construct

$$\begin{aligned}
\hat{Z}(z) &= \prod_{\{t\in\{1,\ldots,kI\}:y_t>d_t\}} f(d_t;p_t,\mathbf{z}) \cdot \prod_{\{t\in\{1,\ldots,kI\}:y_t\le d_t\}} (1-F(y_t-1;p_t,\mathbf{z})),\\
\hat{L}(z) &= \log\hat{Z}(z),\\
\hat{Z}(z,z+u) &= \frac{\hat{Z}(z+u)}{\hat{Z}(z)},
\end{aligned}$$

and based on (4.60), we define

$$\begin{aligned}
\tilde{Z}(z) &= \prod_{t=1}^{kI} f(d_t;p_t,\mathbf{z}),\\
\tilde{L}(z) &= \log\tilde{Z}(z),\\
\tilde{Z}(z,z+u) &= \frac{\tilde{Z}(z+u)}{\tilde{Z}(z)}.
\end{aligned}$$

The following analysis follows in parallel as that in proof of case (1) in Proposition IV.3.

(2) If $|\mathcal{C}| = 0$, then (4.59) is the same as (4.60), and the standard MLE result in Borovkov (1998) Theorem 36.3 can be directly applied. Combing (1) and (2), Proposition IV.7 is thus proved. □

Proof of Theorem IV.6. We evaluate the regret of the proposed policy as follows:

$$\begin{aligned}
R(T) &= \sum_{t=1}^{T}\mathbb{E}[G(p^*,y^*,\mathbf{z})-G(p_t,y_t,\mathbf{z})]\\
&= \sum_{t=1}^{I}\mathbb{E}\left[G(p^*,y^*,\mathbf{z})-G(p_t,y_t,\mathbf{z})\right]+\sum_{t=I+1}^{T}\mathbb{E}[G(p^*,y^*,\mathbf{z})-G(p_t,y_t,\mathbf{z})]\\
&\le \underbrace{\sum_{t=1}^{I}\mathbb{E}\left[G(p^*,y^*,\mathbf{z})-G(p_t,y_t,\mathbf{z})\right]}_{\text{Exploration Regret}}\\
&\quad+\underbrace{\sum_{t=I+1}^{T}\mathbb{E}\left[G(p^*,y^*,\mathbf{z})-G(\hat{p},\hat{y},\mathbf{z})\right]}_{\text{Regret from Estimation Error}}+\underbrace{\sum_{t=I+1}^{T}\left[|G(\hat{p},\hat{y},\mathbf{z})-G(p_t,y_t,\mathbf{z})|\right]}_{\text{Regret from Missing Inventory Targets}}.
\end{aligned} \tag{4.61}$$

Similar as developing (4.42), the first part in (4.61) is bounded by

$$\sum_{t=1}^{I} \mathbb{E}[G(p^*, y^*, \mathbf{z}) - G(p_t, y_t, \mathbf{z})] \leq I \left(\alpha \left| p^h - p^l \right| + \max\{b, h\} \left| y^h - y^l \right| \right) \leq K_{75} T^{\frac{1}{2}}. \tag{4.62}$$

By Proposition 2,

$$\mathbb{P}\big\{ \|\hat{\mathbf{z}} - \mathbf{z}\| \geq \epsilon \big\} \leq K_{76} e^{-I K_{77} \epsilon^2},$$

therefore the second part in (4.61) is bounded as

$$\sum_{t=I+1}^{T} \mathbb{E}\left[G(p^*, y^*, \mathbf{z}) - G(\hat{p}, \hat{y}, \mathbf{z}) \right] \leq \sum_{t=I+1}^{T} \frac{K_{78}}{I} \leq K_{78} T^{\frac{1}{2}}.$$

In Proposition A1, let $n = 2$, and the third part in (4.61) is upper bounded by $K_{78} T^{1/2}$. Combing the above analyses for the three parts of regret, we finish the proof of Theorem IV.6. □

Appendix B

Next we present Hoeffding inequality, which we include for convenience and it can be found in *Hoeffding* (1963) (and see *Levi et al.* (2007) for applications in inventory control).

Hoeffding Inequality. *Let $A_1, \ldots, A_l$ be independent random variables and $S_l = \sum_{i=1}^{l} A_i$. Assume that $A_i, i = 1, \ldots, l$, are almost surely bounded, i.e., $\mathbb{P}\{A_i \in [a_i^l, a_i^h]\} = 1$. Then, for any $\epsilon > 0$,*

$$\mathbb{P}\{S_l \leq \mathbb{E}[S_l] + \epsilon\} \geq 1 - \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{l} (a_i^h - a_i^l)^2} \right), \tag{4.63}$$

and

$$\mathbb{P}\{S_l \geq \mathbb{E}[S_l] - \epsilon\} \geq 1 - \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{l} (a_i^h - a_i^l)^2} \right). \tag{4.64}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Agarwal, A., D. Foster, D. Hsu, S. Kakade, and A. Rakhlin (2011), Stochastic convex optimization with bandit feedback, *Advances in Neural Information Processing Systems*, pp. 1035–1043.

A.Goldenshluger, and A.Zeevi (2009), Woodroofe's one-armed bandit problem revisited, *Ann. Appl. Probab.*, *19*(4), 1603–1633.

Araman, V. F., and R. Caldentey (2009), Dynamic pricing for nonperishable products with demand learning, *Operations Research*, *57*(5), 1169–1188.

Auer, P., R. Ortner, and C. Szepesvári (2007), Improved rates for the stochastic continuum-armed bandit problem, in *Proceedings of the 20th International Conference on Learning Theory (COLT)*, pp. 454–468.

Aviv, Y., and A. Pazgal (2005), A partially observed markov decision process for dynamic pricing, *Management Sci.*, *51*(9), 1400–1416.

Aviv, Y., and G. Vulcano (2012), *Dynamic List Pricing, R. Philips and O. Ozer, eds.*, Oxford University Press, Oxford.

Azoury, K. (1985), Bayes solution to dynamic inventory models under unknown demand distribution, *Management Sci.*, *31*(9), 1150–1160.

Bertsimas, D., and G. Perakis (2006), Dynamic pricing: A learning approach, in *Mathematical and Computational Models for Congestion Charging*, *Applied Optimization*, vol. 101, edited by S. Lawphongpanich, D. W. Hearn, and M. J. Smith, pp. 45–79, Springer US.

Besbes, O., and A. Muharremoglu (2013), On implications of demand censoring in the newsvendor problem, *Management Science*, *59*(6), 1407–1424.

Besbes, O., and A. Zeevi (2009), Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms, *Operations Research*, *57*(6), 1407–1420.

Besbes, O., and A. Zeevi (2012), Blind network revenue management, *Operations Research*, *60*(6), 1537–1550.

Besbes, O., and A. Zeevi (2015), On the surprising sufficiency of linear models for dynamic pricing with demand learning, *Management Science*, *61*(4), 723–739.

Bookbinder, J. H., and A. E. Lordahl (1989), Estimation of inventory re-order levels using the bootstrap statistical procedure, *IIE Transactions*, *21*(4), 302–312.

Borovkov, A. (1998), *Mathematics Statistics*, Gordon and Breach Science Publishers, Amsterdam.

Broadie, M., D. Cicek, and A. Zeevi (2011), General bounds and finite-time improvement for the kiefer-wolfowitz stochastic approximation algorithm, *Operations Research*, *59*(5), 1211–1224.

Broder, J. (2011), Online algorithms for revenue management, *Unpublished doctoral thesis*.

Broder, J., and P. Rusmevichientong (2012), Dynamic pricing under a general parametric choice model, *Operations Research*, *60*(4), 965–980.

Burnetas, A. N., and C. E. Smith (2000), Adaptive ordering and pricing for perishable products, *Operations Research*, *48*(3), 436–443.

Carvalho, A., and M. Puterman (2005), Learning and pricing in an internet environment with binomial demands, *Revenue and Pricing Management*, *3*(4), 320–336.

Chan, L., Z. Shen, D. Simchi-Levi, and J. Swann (2004), Coordination of pricing and inventory decisions: a survey and classification, *Handbook of Quantitative Supply Chain Analysis Modeling in the E-Business Era, Chapter 9*.

Chen, B., X. Chao, and H. Ahn (2015), Coordinating pricing and inventory replenishment with nonparametric demand learning, Working paper. Available at `http://ssrn.com/abstract=2694633`.

Chen, L., and E. L. Plambeck (2008), Dynamic inventory management with learning about the demand distribution and substitution probability, *Manufacturing & Service Operations Management*, *10*(2), 236–256.

Chen, Q., S. J. S, and I. Duenyas (2014a), Adaptive parametric and nonparametric multi-product pricing via self-adjusting controls, *Working Paper*.

Chen, X., and D. Simchi-Levi (2004a), Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The finite horizon case, *Operations Research*, *52*(6), 887–896.

Chen, X., and D. Simchi-Levi (2004b), Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The infinite horizon case, *Mathematics of Operations Research*, *29*(3), 698–723.

Chen, X., and D. Simchi-Levi (2012), *Pricing and inventory management, R. Philips and O. Ozer, eds.*, Oxford University Press, Oxford.

Chen, X., Z. Pang, and L. Pan (2014b), Coordinating inventory control and pricing strategies for perishable products, *Operations Research*, *62*(2), 284–300.

Chen, Y., S. Ray, and Y. Song (2006), Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales, *Naval Research Logistics*, *53*(2), 117–136.

Cheung, W., D. Simchi-Levi, and H. Wang (2015), Dynamic pricing and demand learning with limited price experimentation, *Working Paper*.

Chu, L. Y., J. G. Shanthikumar, and Z. M. Shen (2008), Solving operational statistics via a bayesian analysis, *Operations Research Letters*, *36*(1), 110 – 116.

Chung, B., J. Li, and T. Yao (2011), Dynamic pricing and inventory control with nonparametric demand learning, *Int. J. Services Operations and Informatics*, *6*(3), 259–271.

Cope, E. (2009), Regret and convergence bounds for a class of continuum-armed bandit problems, *Automatic Control, IEEE Transactions*, *54*(6), 1243–1253.

den Boer, A. V. (2014), Dynamic pricing with multiple products and partially specified demand distribution, *Mathematics of Operations Research*, *39*(3), 863–888.

den Boer, A. V. (2015), Dynamic pricing and learning: Historical origins, current research, and new directions, *Surveys in Operations Research and Management Science*, *20*(1), 1–18.

den Boer, A. V., and B. Zwart (2014), Simultaneously learning and optimizing using controlled variance pricing, *Management Science*, *60*(3), 770–783.

den Boer, A. V., and B. Zwart (2015), Dynamic pricing and learning with finite inventories, *Operations Research*, *63*(4), 965–978.

Ding, X., M. Puterman, and A. Bisi (2002), The censored newsvendor and the optimal acquisition of information, *Oper. Res.*, *50*(3), 517–527.

Elmaghraby, W., and P. Keskinocak (2003), Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions, *Management Science*, *49*(10), 1287–1309.

Eren, S., and C. Maglaras (2014), A maximum entropy joint demand estimation and capacity control policy, forthcoming, Production and Operations Management.

Farias, V. F., and B. van Roy (2010), Dynamic pricing with a prior on market response, *Operations Research*, *58*(1), 16–29.

Federgruen, A., and A. Heching (1999), Combined Pricing and Inventory Control Under Uncertainty, *Operations Research*, *47*(3), 454–475.

Gallego, G., and G. van Ryzin (1994), Optimal dynamic pricing of inventories with stochastic demand over finite horizon, *Management Sci.*, *40*(8), 999–1020.

Gill, R., and B. Levit (1995), Applications of the van trees inequality: a bayesian cramer-rao bound, *Bernoulli*, pp. 59–79.

Glasserman, P. (1991), *Gradient Estimation Via Perturbation Analysis*, Kluwer international series in engineering and computer science: Discrete event dynamic systems, Springer.

Godfrey, G. A., and W. B. Powell (2001), An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution, *Management Science*, *47*(8), 1101–1112.

Harrison, J. M., N. B. Keskin, and A. Zeevi (2012), Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution, *Management Science*, *58*(3), 570–586.

Hazan, E. (2015), Introduction to online convex optimization, Book Draft. Computer Science, Princeton University. Available at `http://ocobook.cs.princeton.edu/OCObook.pdf`.

Hazan, E., A. Kalai, S. Kale, and A. Agarwal (2006), Logarithmic regret algorithms for online convex optimization, in *In 19th COLT*, pp. 499–513.

Hazan, E., A. Agarwal, and S. Kale (2007), Logarithmic regret algorithms for online convex optimization, *Machine Learning*, *69*(2-3), 169–192.

Heyman, D., and M. Sobel (1984), *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization*, McGraw-Hill, New York.

Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, *58*(301), 13–30.

Huh, W. H., and P. Rusmevichientong (2009), A non-parametric asymptotic analysis of inventory planning with censored demand, *Mathematics of Operations Research*, *34*(1), 103–123.

Huh, W. H., P. Rusmevichientong, R. Levi, and J. Orlin (2011), Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator, *Operations Research*, *59*(4), 929–941.

Huh, W. T., and G. Janakiraman (2008), (s, S) optimality in joint inventory-pricing control: An alternate approach, *Operations Research*, *56*(3), 783–790.

Huh, W. T., G. Janakiraman, J. A. Muckstadt, and P. Rusmevichientong (2009), An adaptive algorithm for finding the optimal base-stock policy in lost sales inventory systems with censored demand, *Mathematics of Operations Research*, *34*(2), 397–416.

Johnson, K., D. Simchi-Levi, and H. Wang (2015), Online network revenue management using thompson sampling, working paper.

Kalish, S. (1983), Monopolist pricing with dynamic demand and production cost, *Marketing Sci.*, *2*(2), 135–159.

Karlin, S., and C. Carr (1962), Prices and optimal inventory policies, *Studies in Applied Probability and Management Science*, pp. 159–172.

Keskin, N. B., and A. Zeevi (2014), Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies, *Operations Research*, *62*(5), 1142–1167.

Kiefer, J., and J. Wolfowitz (1952a), Stochastic estimation of the maximum of a regression function, *The Annals of Mathematical Statistics*, *23*(3), pp. 462–466.

Kiefer, J., and J. Wolfowitz (1952b), Stochastic estimation of the maximum of a regression function, *Ann. Math. Statist.*, *23*, 462–466.

Kleinberg, R. (2005), Nearly tight bounds for the continuum-armed bandit problem, *Advances in Neural Information Processing Systems*, pp. 697–704.

Kleywegt, A. J., A. Shapiro, and T. H. de Mello (2002), The sample average approximation method for stochastic discrete optimization, *SIAM J. on Optimization*, *12*(2), 479–502.

Lai, T., and H. Robbins (1981), Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes, *Probability Theory and Related Fields*, *56*(3), 329–360.

Lai, T., and H. Robbins (1985), Asumptotically efficient adaptive allocation rules, *Advances in Applied Mathematics*, *6*(1), 4–22.

Lariviere, M. A., and E. L. Porteus (1999), Stalking information: Bayesian inventory management with unobserved lost sales, *Management Science*, *45*(3), 346–363.

Lei, Y., S. Jasin, and A. Sinha (2014), Near-optimal bisection search for nonparametric dynamic pricing with inventory constraint, working paper.

Levi, R., R. O. Roundy, and D. B. Shmoys (2007), Provably near-optimal sampling-based policies for stochastic inventory control models, *Mathematics of Operations Research*, *32*(4), 821–839.

Levi, R., G. Perakis, and J. Uichanco (2011), The data-driven newsvendor problem: new bounds and insights, mIT, Working Paper.

Liyanage, L. H., and J. G. Shanthikumar (2005), A practical inventory control policy using operational statistics, *Operations Research Letters*, *33*(4), 341 – 348.

Lovejoy, W. (1990), Myopic policies for some inventory models with uncertain demand distributions, *Management Sci.*, *36*(6), 724–738.

Murray, G., and E. Silver (1966), A bayesian analysis of the style goods inventory problem, *Management Sci.*, *12*(11), 785–797.

Nemirovski, A., A. Juditsky, G. Lan, and A. Shapiro (2009), Robust stochastic approximation approach to stochastic programming, *SIAM J. on Optimization*, *19*(4).

Petruzzi, N., and M. Dada (2002), Dynamic pricing and inventory control with learning, *Naval Res. Logist.*, *49*, 303–325.

Petruzzi, N. C., and M. Dada (1999), Pricing and the newsvendor problem: A review with extensions, *Operations Research*, *47*(2), 183–194.

Powell, W., A. Ruszczynski, and H. Topaloglu (2004), Learning algorithms for separable approximations of discrete stochastic optimization problems, *Mathematics of Operations Research*, *29*(4), 814–836.

Robbins, H., and S. Monro (1951), A stochastic approximation method, *Ann. Math. Statist.*, *22*(3), 400–407.

Shi, C., W. Chen, and I. Duenyas (2015), Nonparametric data-driven algorithms for multi-product inventory systems,, forthcoming, Operations Research.

Sobel, M. J. (1981), Myopic solutions of markov decision processes and stochastic games, *Operations Research*, *29*(5), 995–1009.

Song, Y., S. Ray, and T. Boyaci (2009), Optimal dynamic joint inventory-pricing control for multiplicative demand with fixed order costs and lost sales, *Operations Research*, *57*(1), 245–250.

Subrahmanyan, S., and R. Shoemaker (1996), Developing optimal pricing and inventory policies for retailers who face uncertain demand, *Journal of Retailing*, *72*(1), 7–30.

Thowsen, G. (1975), A dynamic, nonstationary inventory problem for a price/quantity setting firm, *Naval Res. Logist.*, *22*, 461–476.

Wang, Z., S. Deng, and Y. Ye (2014), Close the gaps: A learning-while-doing algorithm for single-product revenue management problems, *Operations Research*, *62*(2), 318–331.

Wei, Y. (2012), Optimization and optimality of a joint pricing and inventory control policy in periodic-review systems with lost sales, *OR Spectrum*, *34*(1), 243–271.

Whitin, T. M. (1955), Inventory control and price theory, *Management Science*, *2*(1), 61–68.

Yano, C., and S. M. Gilbert (2003), *Coordinated Pricing and Production/Procurement Decisions: A Review. J. Eliashberg, A. Chakravarty, eds.*, Kluwer, Norwell, MA.

Zhang, L., and J. Chen (2006), Bayesian solution to pricing and inventory control under unknown demand distribution, *Oper. Res. Lett.*, *34*(5), 517–524.

Zinkevich, M. (2003), Online convex programming and generalized infinitesimal gradient ascent, in *Proc. 20th Internat. Conf. Machine Learn.*, ICML.

Zipkin, P. (2000), *Foundations of Inventory Management*, McGraw-Hill, New York.