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 \textit{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 \textit{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 optimal policy as the planning horizon increases, and obtain the convergence rate of regret. Numerical experiments show that the proposed algorithm performs very well.

\textit{Key words:} algorithm, joint pricing and inventory control, lost-sales, censored demand, asymptotic analysis

1. 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 \textit{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).

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 \( t = 1, \ldots, T \) are denoted by \( D_t(p_t) = \lambda_t + \epsilon_t \), where \( \lambda_t \) is a decreasing deterministic function representing the average customer response rate to the selling price in period \( t \), \( p_t \), and \( \epsilon_t, t = 1, 2, \ldots, T \), are independent and identically distributed (i.i.d.) random variables representing noises (see §2 for model details). For notational convenience, we use \( \epsilon_t \) and \( \epsilon \) interchangeably in this paper, 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 \( \epsilon_t \) 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 \( \epsilon_t \), 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. (2014)). We also assume the expected single-period profit function is unimodal if the function \( \lambda(\cdot) \) and the distribution of \( \epsilon_t \) 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 1 below, will be multimodal in price, which presents a major analytical barrier for learning and optimization (see Figure 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 \( \mathcal{C}(\pi) \). Now, had the firm known the underlying demand-price function \( \lambda(\cdot) \) and the distribution of \( \epsilon_t \) a priori, there exists a clairvoyant optimal policy \( \pi^* \) with total expected profit denoted by \( \mathcal{C}(\pi^*) \). We measure the performance of our proposed policy \( \pi \) through
an average (per-period) regret \( R(\pi, T) \triangleq (\mathcal{C}(\pi^*) - \mathcal{C}(\pi))/T \). The main research question is to devise an effective nonparametric data-driven policy \( \pi \) that converges to \( \pi^* \) in probability and also drives the average regret \( R(\pi, T) \) to zero with a provable convergence rate.

**Example 1.** Let \( \lambda(p) = 2.944 - 0.52p \), and the random error \( \epsilon \) 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 \( \text{(Opt-CV)} \) (i.e., with known \( \lambda(p) \) and distribution of \( \epsilon \) a priori) is given by

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

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

![Figure 1](image1.png) The clairvoyant’s problem (Opt-CV)  
![Figure 2](image2.png) The sampled problem (Opt-SAA)

The objective profit function in (1) over \( \mathcal{P} \) is plotted in Figure 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 \( \epsilon \) and instead has observed 5 unbiased samples of \( \epsilon \) as \([-1, -1, -1, 0.85, 5]\). (In the case that the firm does not know \( \lambda(p) \), the samples of \( \epsilon \) collected are usually biased in our model.) Using sample average, one can readily construct an estimated objective profit function as

\[
\max_{p \in [0, 3.6]} \left\{ p\left(2.944 - 0.52p + \frac{1}{5} \sum_{j=1}^{5} \epsilon_j \right) - \min_{y \in [0, 10]} \frac{1}{5} \left\{ (b + p)\sum_{j=1}^{5} (2.944 - 0.52p + \epsilon_j - y)^+ \right\} \right\}.
\]
Unfortunately, even for this simpler setting, as seen from Figure 2, the (sampled) objective profit function (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 $\epsilon$ cannot be obtained (as they cannot be separated from the estimation of $\lambda(\cdot)$), making the problem considerably more difficult to analyze.

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 1), and also characterize its rate of convergence (Theorem 2). More specifically, we show that the average regret $R(DDC,T)$ converges to zero at the rate of $O(T^{-\frac{1}{2}}(\log T)^{\frac{1}{4}})$. 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 $\epsilon_t$ 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 $\epsilon_t$ 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 $\epsilon_t$. 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 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 $\epsilon_t$. 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 1 and 3).

(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 $\epsilon_t$ based on demand observations. We instead use the residual error defined in (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 2 and 4). 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 1 and 3).

(c) Learning and optimizing a multimodal function. The most difficult (and also the unique) 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 2), even though the expected profit function is assumed to be unimodal in $p$ (e.g., Figure 1). In contrast, the original objective and its (sampled) proxy function in the backorder 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 2). 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 $T^{\frac{1}{2}}$ 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 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 paper 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.

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. (2014)). 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 \emph{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 \emph{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.

1.4. Organization and General Notation

The rest of this paper is organized as follows. In §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, we propose a nonparametric data-driven algorithm called DDC under censored demand information. In §4, we state our main results and
provide our proof strategies. The detailed proofs are deferred to the Appendix. In §5, we extend our model and results to the observable demand case and also the unbounded demand case.

Throughout this paper, 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.

2. Joint Pricing and Inventory Control with Lost-Sales and Censored Demand

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 $\epsilon_t$, 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 $\epsilon_t$, $t = 1, 2, \ldots, T$, are independent and identically distributed (i.i.d.) random variables. We assume that $\epsilon_t$ is defined on a finite support $[l, u]$, but will later extend it to the case of unbounded support in §5. We denote the CDF of $\epsilon_t$ by $F(\cdot)$. For notational convenience, we use $\epsilon_t$ and $\epsilon$ interchangeably in this paper, due to the i.i.d. assumption.

In this paper 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 $\epsilon_t$ 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 $\epsilon_t$ 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 $y_t \in \mathcal{Y} = [y', y^h]$ and $p_t \in \mathcal{P} = [p', p^h]$ with known bounded support, and $\lambda(p^h) > 0$ and $y^h \geq \lambda(p') + u$.

Given any admissible policy $\pi$, we describe the sequence of events for each period $t$. (Note that $x_t^r$, $y_t^r$, $p_t^r$’s are functions of $\pi$; for ease of presentation, we will make their dependence on $\pi$ 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

$$p_t \min \{d_t(p_t), y_t \} - b(d_t(p_t) - y_t)^+ - h(y_t - d_t(p_t))^+$$

(3)

$$= p_t d_t(p_t) - (b + p_t)(d_t(p_t) - y_t)^+ - h(y_t - d_t(p_t))^+,$$

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) 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 $\epsilon_t$ 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 $\epsilon_t$, and the total profit given by

$$\max_{(p_t, y_t) \in \mathcal{P} \times \mathcal{Y}} \sum_{t=1}^{T} \left( p_t \mathbb{E}[D_t(p_t)] - (b + p_t)\mathbb{E}[D_t(p_t) - y_t]^+ - h\mathbb{E}[y_t - D_t(p_t)]^+ \right).$$

(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.
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 $\epsilon$ a priori. Sobel (1981) has shown that a myopic policy is optimal. Define the single-period revenue function by

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

(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

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

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

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

(Opt-CV)

where

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

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

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

(Approx-CV)

where

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

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 $\bar{g}(\alpha, \beta)$ be the optimal price for $\bar{G}(p, \alpha - \beta p)$ . For any fixed $p \in P$, 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 $y^* = \tilde{y}(p^*, \lambda(p^*))$. And, for any $z \in \mathcal{P}$, we introduce notation

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

\textbf{Main assumptions.} We state the main assumptions for our results in §4 to hold.

\textsc{Assumption 1.}  
(i) $\tilde{G}(p, \lambda(p))$ is unimodal in $p \in \mathcal{P}$.
(ii) $\tilde{G}(p, \tilde{\alpha}(z) - \tilde{\beta}(z)p)$ is strictly concave in $p \in \mathcal{P}$ for any $z \in \mathcal{P}$.
(iii) $\max_{z \in \mathcal{P}} \left| \frac{d\tilde{\alpha}(z)}{dz} \right| < 1$.
(iv) The random error $\epsilon$ has a bounded support $[l, u]$ where $l \leq u < \infty$.

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. (2014)). We provide sufficient conditions for a demand-price function $\lambda(\cdot)$ to satisfy Assumption 1 in the Appendix. Assumption 1(iv) can be dropped, and we defer the detailed discussion to §5.

\section{Nonparametric Data-Driven Algorithm}

Without knowing both $\lambda(p)$ and the underlying distribution of $\epsilon$ 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 \textit{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.
3.1. 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 \( 2L_1 \) 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

\[
I_i = \lfloor I_0 v^i \rfloor, \quad L_i = \lfloor I_i^{\frac{3}{4}} \rfloor, \quad \delta_i = \rho L_i^{\frac{1}{2}} (\log L_i)^{\frac{1}{4}}, \quad \text{and} \quad t_i = \sum_{k=1}^{i-1} I_k \text{ with } t_1 = 0. \tag{7}
\]

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

\[
\mathcal{S}_i = \{p^i, p^i + \delta_i, p^i + 2\delta_i, \ldots, p^i + L_i\}, \tag{8}
\]

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 \( \rho \), and \( \hat{p}_1, \hat{y}_{1.1}, \hat{y}_{1.2}, \delta_1, I_1, L_1 \).

Then, for each learning stage \( i = 1, \ldots, n \) where \( n = \lfloor \log_v \left( \frac{v-I_0+1}{v-1} T + 1 \right) \rfloor \), 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 \( y_t = \hat{y}_{i.1} \lor x_i \);

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

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

(iii) for \( t = t_i + L_i + 1 \), set \( y_t = \hat{y}_{i.2} \lor x_i \);

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

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

\[
(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = \arg \min_{t_{i+1}+2L_i} \left\{ \sum_{t=t_{i+1}+1}^{t_{i+1}+2L_i} \left[ d_t \wedge y_t - (\alpha - \beta p_t) \right]^2 \right\}, \quad \text{and}
\]

\[
\eta_t = d_t \wedge y_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p_t), \quad \text{for } t \in \{t_i+1, \ldots, t_i+2L_i\}. \quad (9)
\]

Step 3: Maximize the proxy profit \( Q_{SAA}^i(p, y) \).

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

\[
\max_{(p, y) \in S_{i+1} \times Y} Q_{SAA}^i(p, y) \triangleq \max_{p \in S_{i+1}} \left\{ \hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \right\}, \quad \text{(Opt-SAA)}
\]

where

\[
\hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \triangleq p \left( \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p \right)
\]

\[
- \min_{y \in Y} \left\{ \frac{1}{2L_i} \sum_{t=t_{i+1}+1}^{t_{i+1}+2L_i} \left[ (b+p) \left( \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p + \eta_t - y \right)^+ + h \left( y - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p + \eta_t) \right)^+ \right] \right\}.
\]

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

\[
(\hat{p}_{i+1}, \hat{y}_{i+1,1}) = \arg \max_{(p, y) \in S_{i+1} \times Y} Q_{SAA}^i(p, y),
\]

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

\[
\hat{y}_{i+1,2} = \arg \max_{y \in Y} Q_{i+1}^S(\hat{p}_{i+1} + \delta_{i+1}, y).
\]

In case that \( \hat{p}_{i+1} + \delta_{i+1} \notin S_{i+1} \), 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

\[
p_t = \hat{p}_{i+1}, \quad y_t = x_t \lor \hat{y}_{i+1,1}.
\]

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 \( \delta_i \) 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 \( \eta_t \). 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 \( \epsilon_t \), 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 \( \eta_t \) is not a true sample of \( \epsilon_t \), 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 \( \epsilon \) (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 \( \epsilon_t \) 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 \( \eta_t \) converges to \( \epsilon_t \) 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 \( \epsilon_t \) are replaced by either empirical average or quantile of \( \eta_t \). It is important to note that both \((\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})\) and \( \eta_t \) 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 \( \epsilon_t \) 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 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. 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.

\[
\max_{p \in \mathcal{S}_{i+1}} \left\{ \tilde{G}_{i+1}(p, \alpha - \beta p) \right\}, \quad \text{(Approx-SAA)}
\]
where \( \tilde{G}_{i+1}(p, \alpha - \beta p) \triangleq p(\alpha - \beta p) \)
\[
- \min_{y \in \mathcal{Y}} \left\{ \frac{1}{2L_i} \sum_{i=t_{i+1}}^{t_i + 2L_i} \left\{ (b + p)(\alpha - \beta p + \hat{\epsilon}_t - y) + h(y - (\alpha - \beta p + \hat{\epsilon}_t))^2 \right\} \right\},
\]
where the truncated random error \( \hat{\epsilon}_t \) is defined by
\[
\hat{\epsilon}_t \triangleq d_t \land y_t - \lambda(p_t), \ t \in \{t_i + 1, \ldots, t_i + 2L_i\},
\]
It is clear that the truncated random error \( \hat{\epsilon}_t = \epsilon_t \) only when \( d_t \leq y_t \). 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 \( \epsilon \), 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 \( \epsilon \) is still available information. In (Approx-SAA), the firm knows neither \( \lambda(p) \) nor the distribution of \( \epsilon \), but it could hypothetically access the truncated samples of \( \epsilon_t \) (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 \( \epsilon \) 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 \( \eta_t \) in place of the true random errors \( \epsilon_t \) 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 \( \eta_t \).

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 (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 \( y_1, y_2 \in \mathcal{Y} \),
\[
|Q(p_1, y_1) - Q(p_2, y_2)| \leq K_1(|p_1 - p_2| + |y_1 - y_2|).
\]  
(12)

We also assume Lipschitz conditions for \( \tilde{y}(q, \lambda(q)) \) and \( \tilde{y}(p, \hat{\alpha}(q) - \hat{\beta}(q)p) \) 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 \( q_1, q_2 \in \mathcal{P} \),
\[
|\tilde{y}(q_1, \lambda(q_1)) - \tilde{y}(q_2, \lambda(q_2))| \leq K_2 |q_1 - q_2|, 
\]
\[
|\tilde{y}(p, \hat{\alpha}(q_1) - \hat{\beta}(q_1)p) - \tilde{y}(p, \hat{\alpha}(q_2) - \hat{\beta}(q_2)p)| \leq K_2 |q_1 - q_2|.
\]  
(13)
(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 \( \epsilon \) satisfies \( r = \min\{f(x), x \in [l, u]\} > 0 \).

### 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 \( \epsilon \), 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 1.

It can be seen from Table 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.

### 4. Main Results and Performance Analysis

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

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

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

**Theorem 1. (Convergence of Decisions)** Under Assumption 1, for the joint pricing and inventory control problem with lost-sales and censored demand, the DDC algorithm satisfies
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 2. (Convergence Rate of Regret)** Under Assumption 1, 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

\[
R(\text{DDC}, T) \leq K_0 \cdot T^{-\frac{1}{5}} (\log T)^{\frac{1}{4}}. 
\]  

\(^{(16)}\)

To the best of our knowledge, Theorems 1 and 2 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 §1.2). We refer the readers to the detailed discussions in §3.2 (algorithmic design) and §4.1–§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 §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 (1952), 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 cannot 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).

**Figure 3** \( \beta = 0.52 \)

**Figure 4** \( \beta = 0.54 \)

**Remark on the convergence rate.** We first give an intuitive explanation on the key factor that affects the convergence rate, and will present a more technical discussion in §4.1. Figure 3 shows the same sampled objective function as in Example 1, with parameters \( \alpha = 2.944 \) and \( \beta = 0.52 \), and the global optimal solution is \( p = 3.4 \). Now we slightly perturb \( \beta \) to 0.54 (while holding all other parameters fixed), then the global optimal solution shifts drastically to \( p = 2.45 \) as depicted in Figure 4. This shows that even a slight perturbation in the parameters \( \alpha \) and \( \beta \) can lead to a dramatic change in the global optimal solution (jumbling 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., \( \hat{\alpha}_{i+1} \to \hat{\alpha}(p^{*}) \) and \( \hat{\beta}_{i+1} \to \hat{\beta}(p^{*}) \)) 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.
as their optimal solution of $\hat{G}_{i+1}$ is Lipschitz in $(\alpha, \beta)$. 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 §4.1, which results in the final regret rate in Theorem 2.

For the remainder of §4, we shall outline the main steps, ideas and techniques developed for the proofs of Theorems 1 and 2. We will also explain in details the key differences between this paper and the closely related works, e.g., Besbes and Zeevi (2015) and Chen et al. (2015).

4.1. Key Ideas in Proving the Convergence of Pricing Decisions

We first discuss the key ideas and steps in proving Theorem 1(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 $\bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i))$ from (Approx-CV) as an intermediate benchmark, we have

$$
(\hat{p}_{i+1} - p^*)^2 \leq \left( \left| p^* - \bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) \right| + \left| \bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) - \hat{p}_{i+1} \right| \right)^2.
$$

In the following, we develop upper bounds for the two terms on the RHS of (17).

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

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

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

We now develop an upper bound for the second term on the RHS of (17), which is one of the most critical results in our analysis. It bridges between (Approx-CV) and (Opt-SAA).

**Proposition 2.** There exist positive constants $K_1^{P2}$ and $K_2^{P2}$ such that, for any $\hat{p}_i \in \mathcal{P}$,

$$
\mathbb{P} \left\{ \left| \bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) - \hat{p}_{i+1} \right| \geq K_1^{P2} L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \right\} \leq K_2^{P2} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}.
$$
The difficulty of establishing Proposition 2 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., \( \hat{\alpha}_{i+1} \to \tilde{\alpha}(\hat{p}_i) \) and \( \hat{\beta}_{i+1} \to \tilde{\beta}(\hat{p}_i) \), which can be used to guarantee the convergence of \( \hat{G}_{i+1} \) to \( \tilde{G}_{i+1} \) in the lost-sales case where unimodality is no longer preserved. This implies that we need to compare between proxy functions \( \bar{G} \) and \( \tilde{G} \). The two intermediate results below (Lemmas 1 and 2) show that, for any fixed price \( p \in P \), 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 1.** There exists some positive constant \( K_{i1}^{L1} \) such that, for any given \( \alpha, \beta, \) and \( p \in P \),

\[
P \left\{ \left| \bar{G}(p, \alpha - \beta p) - \hat{G}_{i+1}(p, \alpha - \beta p) \right| > K_{i1}^{L1}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{3}{4}} \right\} \leq 4L_i^{-4}. \tag{20} \]

In the function \( \bar{G}(p, \alpha - \beta p) \) the distribution of \( \epsilon \) is known and the expectation can be taken, whereas in the function \( \hat{G}_{i+1}(p, \alpha - \beta p) \) 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 1, 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 2.** There exists a positive constant \( K_{i2}^{L2} \) such that, for any given \( p \in P \),

\[
P \left\{ \left| \bar{G}_{i+1}(p, \hat{\alpha}(p) - \hat{\beta}(p)p) - \hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \right| \geq K_{i2}^{L2}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{3}{4}} \right\} \leq 24L_i^{-2}. \tag{21} \]

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 (11), whereas \( \hat{G}_{i+1} \) is constructed using the residual error \( \eta_t \) defined in (10). (Note that both \( \tilde{\epsilon}_t \) and \( \eta_t \) are used to estimate the true random error \( \epsilon_t \), and both of them suffer from the estimation error due to demand censoring. The difference is that \( \eta_t \) also suffers from the estimation error of \( \lambda(\cdot) \) while \( \tilde{\epsilon}_t \) does not.)

Second, \( \tilde{G}_{i+1} \) involves the parameters \( \hat{\alpha}(\hat{p}_i) \) and \( \hat{\beta}(\hat{p}_i) \), 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 \( \lambda(p) = \hat{\alpha}(p) - \hat{\beta}(p)p \) for any \( p \in P \), we can therefore write the truncated random error \( \tilde{\epsilon} \) in \( \tilde{G}_{i+1} \) and the residual error \( \eta_t \) in \( \hat{G}_{i+1} \) as

\[
\tilde{\epsilon}_t = d_t \land y_t - \lambda(\hat{p}_i) = d_t \land y_t - (\hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_i), \tag{21} \\
\eta_t = d_t \land y_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}\hat{p}_i), \tag{22}
\]
which suggests that the difference between $\tilde{\epsilon}_t$ and $\eta_t$ can be bounded by the difference between the associated parameters $(\tilde{\alpha}(\hat{p}_t), \tilde{\beta}(\hat{p}_t))$ and $(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$. We then show that both $\tilde{\alpha}_{i+1} \rightarrow \hat{\alpha}(\hat{p}_t)$ and $\tilde{\beta}_{i+1} \rightarrow \hat{\beta}(\hat{p}_t)$ in probability as $i$ grows to obtain the desired result.

**Sparse discretization scheme.** To obtain the convergence of prices in Proposition 2, 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 5).

**Definition 1 (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 $\eta$ with a (high) probability no less than $1 - \delta$, i.e., $\mathbb{P}(\max_{x \in \mathcal{M}} |g_1(x) - g_2(x)| \leq \eta) \geq 1 - \delta$.

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 (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.

![Figure 5](sparse_discretization_and_uniform_closeness.png)  
**Figure 5** Sparse discretization and uniform closeness

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 (8), since the chosen step size $\delta_{i+1}$ in (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

$$\frac{p^h - p^l}{\delta_{i+1}} = \left(\frac{p^h - p^l}{\rho}\right) L_{i+1}^{\frac{1}{4}} (\log L_{i+1})^{-\frac{1}{4}} \leq \left(\frac{p^h - p^l}{\rho}\right) L_{i+1}^{\frac{1}{2}} \leq \left(\frac{p^h - p^l}{\rho}\right) T^{\frac{1}{4}}.$$ 

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 1 and 2, 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

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

which leads to

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

which says that $\bar{G}$ and $\hat{G}_{i+1}$ are uniformly close on the grid $\mathcal{S}_{i+1}$ (see Figure 5).

---

**Figure 6** Choosing $\bar{p}$ to be the closest 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 (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 auxiliary 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 6). Because both \( \bar{p} \) and \( \hat{p}_{i+1} \) lie on the grid, we can then apply (23) to obtain

\[
\mathbb{P} \left\{ G(\bar{p}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i) \bar{p}) - G(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i) \hat{p}_{i+1}) \leq 2K_4^{P_2} L_i^{-\frac{1}{2}} (\log L_i)^{1/4} \right\} 
\geq 1 - K_4^{P_2} L_i^{-\frac{7}{2}} (\log L_i)^{-\frac{1}{2}}.
\]

We then show, by strict concavity of \( G \), that there exists some positive number \( m > 0 \) such that

\[
G(\bar{p}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i) \bar{p}) - G(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i) \hat{p}_{i+1}) \geq m(\bar{p} - \hat{p}_{i+1})^2.
\]

Combining (24) and (25), there exists some constant \( K_4^{P_2} > 0 \) such that

\[
\mathbb{P} \left\{ |\bar{p} - \hat{p}_{i+1}| \leq K_4^{P_2} L_i^{-\frac{1}{2}} (\log L_i)^{1/4} \right\} \geq 1 - K_4^{P_2} L_i^{-\frac{7}{2}} (\log L_i)^{-\frac{1}{2}}.
\]

By our choice of \( \bar{p} \) and the property of \( \delta_{i+1} \), we have

\[
|\bar{p} - \bar{p}| \leq \delta_{i+1}.
\]

Combining (26) and (27), we obtain the desired result

\[
\mathbb{P} \left\{ |\bar{p} - \hat{p}_{i+1}| \leq K_1^{P_1} L_i^{-\frac{1}{2}} (\log L_i)^{1/4} \right\} \geq 1 - K_2^{P_2} L_i^{-\frac{7}{2}} (\log L_i)^{-\frac{1}{2}}.
\]

**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 \( |\bar{p} - \hat{p}_{i+1}| \leq \mathcal{O}(L_i^{-\frac{1}{2}} (\log L_i)^{1/4}) \) with a high probability, which largely determines the final regret rate of \( T^{-\frac{1}{2}} (\log T)^{1/4} \) in Theorem 2. 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 \( G \) 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 1(a).

**Proof of Theorem 1(a).** Using Proposition 1 and some simple algebra, for some constant \( K_1^{P_1} \),

\[
\mathbb{E}[(p^* - \hat{p}_{i+1})^2] \leq \mathbb{E} \left[ \left( |p^* - \bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i))| + |\bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) - \hat{p}_{i+1}| \right)^2 \right] 
\leq \mathbb{E} \left[ (\gamma |p^* - \hat{p}_i| + |\bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) - \hat{p}_{i+1}|)^2 \right] 
\leq \left( 1 + \frac{\gamma^2}{2} \right) \mathbb{E}[(p^* - \hat{p}_i)^2] + K_1^{P_1} \mathbb{E} \left[ |\bar{p}(\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) - \hat{p}_{i+1}|^2 \right].
\]
By Proposition 2, we have
\[
\mathbb{P} \left\{ \left| \tilde{p} \left( \tilde{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i) \right) - \hat{p}_{i+1} \right|^2 \geq (K_1^{p_2})^2 L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \right\} \leq K_2^{p_2} L_i^{-\frac{1}{4}} (\log L_i)^{-\frac{1}{4}}, \tag{29}
\]

It follows from (29) and the fact that \( \bar{p} \) and \( \hat{p}_{i+1} \) are bounded, that
\[
\mathbb{E} \left[ \left| \tilde{p} \left( \tilde{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i) \right) - \hat{p}_{i+1} \right|^2 \right] \leq K_2^{T_1} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} + K_3^{T_1} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \leq K_4^{T_1} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}. \tag{30}
\]

Substituting (30) into (28), we have that for \( \frac{1+k_2^2}{2} < 1 \),
\[
\mathbb{E}[(p^* - \hat{p}_{i+1})^2] \leq \left( \frac{1 + \gamma^2}{2} \right) \mathbb{E}[(p^* - \hat{p}_i)^2] + K_5^{T_1} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}}. \tag{31}
\]

Letting \( \frac{1+k_2^2}{2} = \theta \), we further obtain that
\[
\mathbb{E}[(\hat{p}_{i+1} - p^*)^2] \leq \theta^i (\hat{p}_1 - p^*)^2 + K_6^{T_1} \sum_{j=0}^{i-1} \theta^j L_{i-j}^{-\frac{1}{4}} (\log L_{i-j})^{\frac{1}{4}} \leq K_7^{T_1} i \left( v^{-\frac{1}{2}} \right)^i \sum_{j=0}^{i} \theta^j (v^\frac{1}{2})^j. \tag{32}
\]

By choosing \( v > 1 \) satisfying \( \theta v^{\frac{1}{2}} < 1 \), there exists a positive constant \( K_8^{T_1} \) such that \( \sum_{j=0}^{i} \theta^j (v^{\frac{1}{2}})^j \leq K_8^{T_1} \). This implies that for some constants \( K_9^{T_1} \) and \( K_{10}^{T_1} \), we have
\[
\mathbb{E}[(\hat{p}_{i+1} - p^*)^2] \leq K_9^{T_1} i \left( v^{-\frac{1}{2}} \right)^i \leq K_{10}^{T_1} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \rightarrow 0, \text{ as } i \rightarrow \infty. \tag{33}
\]

Moreover, for some positive constant \( K_{11}^{T_1} \), we have
\[
\mathbb{E}[(\tilde{p}_{i+1} + \delta_{i+1} - p^*)^2] \leq 2\mathbb{E}[(\tilde{p}_{i+1} - p^*)^2] + 2\delta_{i+1}^2 \leq K_{11}^{T_1} L_i^{-\frac{1}{4}} (\log L_i)^{\frac{1}{4}} \rightarrow 0, \text{ as } i \rightarrow \infty. \tag{34}
\]

This completes the proof of Theorem 1(a). \( \square \)

4.2. Key Ideas in Proving the Convergence of Inventory Decisions

We next elaborate on the proof of Theorem 1(b).

For any fixed \( p \in P \), 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 \( y^* = \bar{y}(p^*, \lambda(p^*)) \),
\[
\mathbb{E} \left[ |y^* - \bar{y}_{i+1,1}|^2 \right] \leq \mathbb{E} \left[ \left( |\bar{y}(p^*, \lambda(p^*)) - \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))| + |\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \bar{y}_{i+1,1}| \right)^2 \right]. \tag{35}
\]
It follows from (13) that there exists some positive constant $K_2$ such that
\[
\left| \bar{y}(p^*, \lambda(p^*)) - \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) \right| \leq K_2 |p^* - \hat{p}_{i+1}|. \tag{36}
\]

Thus, it suffices to bound the second term on the RHS of (35), which is more involved.

**Proposition 3.** There exists some positive constant $K_1^{P3}$ such that,
\[
\mathbb{E}\left[ (\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \bar{y}_{i+1,1})^2 \right] \leq K_1^{P3} L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}. \tag{37}
\]

We provide some high-level ideas of proving Proposition 3. By definitions of $\hat{\alpha}$ and $\hat{\beta}$, we have
\[
\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) = \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}).
\]

Then it follows that
\[
\mathbb{E}\left[ (\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \bar{y}_{i+1,1})^2 \right] \leq \mathbb{E}\left[ \left( \left| \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) - \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) \right| \right. \right.
\]
\[
\left. + \left| \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) - \bar{y}_{i+1,1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) \right| \right]
\]
\[
\leq K_2 |\hat{p}_{i+1} - \hat{p}_i| \leq K_2( |p^* - \hat{p}_i| + |p^* - \hat{p}_{i+1}| ). \tag{38}
\]

For the first term on the RHS of (38), by (14), there exists some positive constant $K_2$ such that
\[
\left| \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) - \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) \right|
\]
\[
\leq K_2 |\hat{p}_{i+1} - \hat{p}_i| \leq K_2( |p^* - \hat{p}_i| + |p^* - \hat{p}_{i+1}| ). \tag{39}
\]

We then focus on the second and third terms on the RHS of (38) that both involve the optimal solution $\tilde{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1})$ for (Approx-SAA).

**Lemma 3.** There exists some positive constant $K_1^{L3}$ such that, for any $\hat{p}_{i+1}, \hat{p}_{i+1} \in \mathcal{P}$,
\[
\mathbb{P}\left\{ \left| \bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) - \tilde{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}) \right| \geq K_1^{L3} L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \right\} \leq 2L_i^{-1}.
\]

For any given price $\hat{p}_{i+1}$ and parameters $\hat{\alpha}(\hat{p}_{i+1})$ and $\hat{\beta}(\hat{p}_{i+1})$, the inventory decision $\bar{y}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1})$ 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, $\tilde{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1})$ is computed using truncated random error $\hat{\epsilon}_i$, which is the (downward biased) empirical newsvendor solution. To correct such estimation bias and prove Lemma 3, we first compare $\tilde{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_{i+1}) - \hat{\beta}(\hat{p}_{i+1})\hat{p}_{i+1})$ with
To prove Theorem 2, we break the time horizon into learning stages to obtain

\[ E \left[ \sum_{t=1}^{T} (Q(p^*, y^*) - Q(p_t, y_t)) \right] \]

\[ \leq E \left[ \sum_{i=1}^{n} \left( \sum_{t=t_i + 1}^{t_i + 2L_i} (Q(p^*, y^*) - Q(p_t, y_t)) \right) \right. 
+ \left. \sum_{t=t_i + 2L_i + 1}^{t_{i+1}} (Q(p^*, y^*) - Q(\hat{p}_{t+1}, \hat{y}_{t+1, 1}) + Q(\hat{p}_{t+1}, \hat{y}_{t+1, 1}) - Q(p_t, y_t)) \right] \]


**Lemma 4.** There exists some positive constant \( K_1^{L_4} \) such that, for any \( \hat{p}_i, \hat{p}_{i+1} \in \mathcal{P} \),

\[ P \left\{ \left| \hat{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_{i+1}) - \hat{y}_{i+1, 1} \right| \geq K_1^{L_4}L_i^{-\frac{1}{4}}(\log L_i)^{\frac{1}{4}} \right\} \leq 24L_i^{-2}. \]

For any given price \( \hat{p}_{i+1} \), the inventory target level \( \hat{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_{i+1}) \) in \( \text{Approx-SAA} \) is computed using the truncated random error \( \tilde{\varepsilon}_t \) and parameters \( (\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) \), whereas the inventory target level \( \hat{y}_{i+1, 1} \) is computed using the residual error \( \eta_t \) and parameters \( (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \). By (21) and (22), we can show that the difference between \( \hat{y}_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_{i+1}) \) and \( \hat{y}_{i+1, 1} \) can be bounded by the difference between their associated parameters \( (\hat{\alpha}(\hat{p}_i), \hat{\beta}(\hat{p}_i)) \) and \( (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \).

While the sampled optimization problem \( \text{Opt-SAA} \) 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 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 \( \text{Opt-SAA} \) is multimodal in \( p \).

**Proof of Theorem 1(b).** By (35) and Proposition 3, we have

\[ E \left[ (y^* - \hat{y}_{i+1, 1})^2 \right] \leq K_1^{L_4}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} \rightarrow 0, \quad \text{as } i \rightarrow \infty, \]

and similarly,

\[ E \left[ (y^* - \hat{y}_{i+1, 2})^2 \right] \leq K_1^{L_4}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} \rightarrow 0, \quad \text{as } i \rightarrow \infty. \]

This completes the proof of Theorem 1(b). \( \square \)

### 4.3. High Level Ideas in Proving the Regret Rate

To prove Theorem 2, we break the time horizon \( T \) into \( n \) learning stages to obtain
Figure 7  The sampled profit as a function of order-up-to level \( y \) (for a fixed price \( p = 2.6 \)) in Example 1

\[
\begin{align*}
\text{profit} &\approx \mathbb{E} \left[ \sum_{i=1}^{n} \sum_{t=t_{i}+1}^{t_{i}+2L_{i}} \left( Q(p^*, y^*) - Q(p_{t_{i}} y_{t_{i}}) \right) \right] + \mathbb{E} \left[ \sum_{i=1}^{n} \sum_{t=t_{i}+2L_{i}+1}^{t_{i}+I_{i}} \left( Q(p^*, y^*) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) \right) \right] \\
&\quad + \mathbb{E} \left[ \sum_{i=1}^{n} \sum_{t=t_{i}+2L_{i}+1}^{t_{i}+I_{i}} \left( Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) - Q(p_{t_{i}} y_{t_{i}}) \right) \right],
\end{align*}
\]

(40)

where the inequality follows from the construction of DDC. As shown in (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 (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 (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 (40).

The first term on the RHS of (40) can be bounded as follows: For some positive constants $K_2^{T_2}, K_3^{T_2}$, and $K_4^{T_2},$

$$
E \left[ \sum_{i=1}^{n} \sum_{t=t_i+1}^{t_i+2L_i} \left( Q(p^*, y^*) - Q(p_t, y_t) \right) \right] \leq \sum_{i=1}^{n} 2L_i K_2^{T_2} \leq K_3^{T_2} \sum_{i=1}^{n} I_i^{\frac{4}{5}} \leq K_4^{T_2} T^{\frac{4}{5}}, \quad (41)
$$

where the first inequality holds because $(p, y)$ is bounded on $P \times Y$ and $Q(p, y)$ is Lipschitz by (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 (40), which has the following upper bound.

**Proposition 4.** There exists some positive constant $K_1^{P_4}$ such that

$$
E \left[ \sum_{i=1}^{n} \sum_{t=t_i+1}^{t_i+2L_i+1} \left( Q(p^*, y^*) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) \right) \right] \leq K_1^{P_4} T^{\frac{4}{5}} (\log T)^{\frac{1}{4}}. \quad (42)
$$

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

$$
E \left[ Q(p^*, \bar{y}(p^*, \lambda(p^*))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) \right] \quad (43)
$$

$$
\leq E \left[ Q(p^*, \bar{y}(p^*, \lambda(p^*))) - Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) \right] + E \left[ Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) \right],
$$

where the first term on the RHS of (43) can be bounded by the difference between $p^*$ and $\hat{p}_{i+1}$ (from the analysis for Theorem 1(a)), and the second term can be bounded by the difference between $\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))$ and $\hat{y}_{i+1,1}$ (from the analysis for Theorem 1(b)).

**Proposition 5.** There exists some positive constant $K_1^{P_5}$ such that

$$
E \left[ \sum_{i=1}^{n} \sum_{t=t_i+1}^{t_i+2L_i+1} \left( Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) - Q(p_t, y_t) \right) \right] \leq K_1^{P_5} T^{\frac{1}{2}}. \quad (44)
$$

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 2.** The proof of Theorem 2 follows directly from combining (41), and Propositions 4 and 5.

### 5. Discussions

In this paper 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 $O(T^{-\frac{1}{4}}(\log T)^{\frac{1}{4}})$. 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 $\epsilon$ is finite, i.e., $\mathbb{E}[\exp(\kappa \epsilon)] < +\infty$ for $\kappa$ near 0.
- (b) The search region for inventory level is sufficiently large. More precisely, $y^h \geq K^U_1 \log T$ for some constant $K^U_1 > 0$. 
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, $r = \min \{f(x), x \in [z^l, z^u]\} > 0$, where

\[
z^l = \min_{y \in [y_0^l, y_0^h], p, q \in [p^l, p^h]} \left( y - (\hat{\alpha}(p) - \hat{\beta}(p)q) \right),
\]

\[
z^u = \max_{y \in [y_0^l, y_0^h], p, q \in [p^l, p^h]} \left( y - (\hat{\alpha}(p) - \hat{\beta}(p)q) \right).
\]

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 demand 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 $K_1^U \log T$; 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 1 and 2 continue to hold for the modified algorithm except that the regret rate in Theorem 2 is changed to $O(T^{-\frac{1}{5}} \log T)$.

References


APPENDIX

A: Sufficient Conditions for Assumption 1

We present some sufficient conditions for the demand-price function to satisfy Assumption 1. For notational convenience, let $D = [d_l, d_h]$ where $d_l = \lambda(p_l)$ and $d_h = \lambda(p_h)$, and also let $\lambda^{-1}(\cdot)$ be the inverse function of $\lambda(\cdot)$.

For Assumption 1(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. (2014) proposed sufficient conditions for $R(d, y)$ to be jointly concave. Define $g(d, y) = -\lambda^{-1}(d)\frac{F'(y-d)}{F(y-d)}$, and the two sufficient conditions are as follows.

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

(C2) $g(d, y) \geq 1$ for all $d \in D$ 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 $\lambda(p) = \frac{-k - mp}{1 + e^{k - mp}}$, and exponential demand $\lambda(p) = e^{k - mp}$, for some parameter $m > 0$.

For the second condition (C2), note that $g(d, y) = d \times \frac{-\lambda^{-1}(d)}{\lambda'(d)} \times \frac{F'(y-d)}{F(y-d)}$, where $\frac{-\lambda^{-1}(d)}{\lambda'(d)}$ is the price elasticity of demand and $\frac{F'(y-d)}{F(y-d)}$ 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 1(ii) to hold, it suffices that for any $z \in \mathcal{P}$, any $p \in \mathcal{P}$, the second derivative

$$\frac{\partial G^2(p, \alpha(z) - \beta(z)p)}{\partial p^2} = 2\lambda'(z) + h^2(b + p + h)^{-3}f\left(F^{-1}\left(\frac{b + p}{b + p + h}\right)\right)^{-1} < 0. \quad (45)$$

For Assumption 1(iii) to hold, it suffices to require the absolute derivative

$$\frac{d\bar{p}(\alpha(z), \beta(z))}{dz} = \left|\frac{\lambda''(z)(2p - z)}{2\lambda'(z) + h^2(b + p + h)^{-3}f\left(F^{-1}\left(\frac{b + p}{b + p + h}\right)\right)^{-1}}\right| < 1 \quad (46)$$

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 $\epsilon$ is uniformly distributed over $[-n, n]$, where $m \geq 1$, $0 \leq p' \leq p_h$ and $0 < n/8 < h \leq b$.

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

3. Logit demand: $\lambda(p) = \frac{e^{k - mp}}{1 + e^{k - mp}}$ and $\epsilon$ is uniform on $[-n, n]$, where $0 < k \leq 1.39$, $0.1 \leq m \leq 0.34$, $n = 2.56$, $h = 0.1$, $b = 2$, $p' = 4$, $p_h = 6$, and $y' = 25/9$, $y_h = 3$. 
B: Technical Proofs for Theorem 1(a)

Proof of Proposition 1. We first show that
\[ p^* = \bar{p}(\hat{\alpha}(p^*), \hat{\beta}(p^*)). \] (47)

That is, \( p^* \) is a fixed point of \( \bar{p}(\hat{\alpha}(z), \hat{\beta}(z)) = z \). Recall that
\[ \tilde{G}(p, \lambda(p)) = p\lambda(p) - \min_{y \in Y} \left\{ (b + p)\mathbb{E}[\lambda(p) + \epsilon - y]^+ + h\mathbb{E}[y - \lambda(p) - \epsilon]^+ \right\}. \] (48)

We know that \( \tilde{G} \) has a unique maximizer \( p^* \) by Assumption 1(i), and also by definition of \( \bar{p} \) that \( \bar{p}(\hat{\alpha}(z), \hat{\beta}(z)) \) is the unique optimal solution for (Approx-CV) with parameters \( (\hat{\alpha}(z), \hat{\beta}(z)) \). Then (47) follows immediately from Lemma A1 of Besbes and Zeevi (2015) by replacing their function \( P \) with our objective function (48).

In addition, by Assumption 1(iii), we have \( \left| \frac{dp(\hat{\alpha}(z), \hat{\beta}(z))}{dz} \right| < 1 \) for any \( z \in P \), which implies that
\[ \left| \bar{p}(\hat{\alpha}(p^*), \hat{\beta}(p^*)) - \bar{p}(\hat{\alpha}(\hat{p}_t), \hat{\beta}(\hat{p}_t)) \right| \leq \gamma |p^* - \hat{p}_t|, \]
where \( \gamma = \max_{z \in P} \left| \frac{dp(\hat{\alpha}(z), \hat{\beta}(z))}{dz} \right| < 1 \). This completes the proof.

Proof of Lemma 1. For any given \( \alpha, \beta \), define the following newsvendor-type functions
\[ \tilde{W}(p, y) = h\mathbb{E}[y - (\alpha - \beta p) - \epsilon]^+ + (b + p)\mathbb{E}[\alpha - \beta p + \epsilon - y]^+, \]
\[ \tilde{W}_{i+1}(p, y) = \frac{1}{2L_i} \sum_{t=i+1}^{t+2L_i} \left( h(y - (\alpha - \beta p) - \hat{\epsilon}_t)^+ + (b + p)(\alpha - \beta p + \hat{\epsilon}_t - y)^+ \right), \]
where recall that the truncated random error \( \hat{\epsilon}_t \) is defined in (11).

For any given \( p \in P \), by the definition of \( G \)'s and the triangle inequality, we have
\[ |\tilde{G}(p, \alpha - \beta p) - \tilde{G}_{i+1}(p, \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p))| \]
\[ \leq |\tilde{W}(p, \tilde{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p))| + |\tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p))|. \] (49)

It then suffices to bound the two terms on the RHS of (49) as follows, for any \( \xi > 0 \):
\[ \mathbb{P} \left\{ |\tilde{W}(p, \tilde{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \tilde{y}(p, \alpha - \beta p))| > K_5 L_1 \xi + \frac{K_4 L_1}{2L_i} \right\} \leq 2e^{-4L_i \xi^2}. \] (50)
\[ \mathbb{P} \left\{ |\tilde{W}_{i+1}(p, \tilde{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \tilde{y}_{i+1}(p, \alpha - \beta p))| > K_5 L_1 (b + p^h + h) \left( \xi + \frac{K_6 L_1}{2L_i} \right) \right\} \leq 2e^{-4L_i \xi^2}. \] (51)
Letting $\xi = L^{-\frac{1}{2}} (\log L)^{\frac{1}{2}}$, then the proof of Lemma 1 follows from combining (49), (50) and (51).

We first focus on proving (50). 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 $\bar{z} = \bar{y}(p, \alpha - \beta p) - (\alpha - \beta p) = F^{-1}\left(\frac{k + p}{b + p + h}\right)$. Moreover, we have

$$\bar{W}(p, \bar{y}(p, \alpha - \beta p)) = \bar{W}(p, \bar{z} + \alpha - \beta p) = h \mathbb{E}(\bar{z} - \epsilon)^+ + (b + p) \mathbb{E}(-\bar{z})^+, \tag{52}$$

$$\bar{W}^A_{i+1}(p, \bar{y}(p, \alpha - \beta p)) = \bar{W}^A_{i+1}(p, \bar{z} + \alpha - \beta p) = \frac{1}{2L_i} \sum_{t = t_i + 1}^{t_i + 2L_i} \left(h(\bar{z} - \bar{\epsilon}_t)^+ + (b + p)(\bar{\epsilon}_t - \bar{z})^\dagger\right), \tag{53}$$

$$\bar{W}^A_{i+1}(p, \bar{y}(p, \alpha - \beta p)) = \bar{W}^A_{i+1}(p, \bar{z} + \alpha - \beta p) = \frac{1}{2L_i} \sum_{t = t_i + 1}^{t_i + 2L_i} \left(h(\bar{z} - \epsilon_t)^+ + (b + p)(\epsilon_t - \bar{z})^\dagger\right). \tag{54}$$

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

$$\Delta_t = \left(h \mathbb{E}[\bar{z} - \epsilon]^+ + (b + p) \mathbb{E}[\epsilon - \bar{z}]^\dagger\right) - \left(h(\bar{z} - \epsilon_t)^+ + (b + p)(\epsilon_t - \bar{z})^\dagger\right).$$

Then $\mathbb{E}[\Delta_t] = 0$ and $\Delta_t$ has a bounded support of some positive length $K_3^{L_1}$ (as $\epsilon_t$ is bounded).

We then apply the Hoeffding’s inequality (see Theorem 1 in Hoeffding (1963)) to obtain, for any $p \in \mathcal{P}$ and $\xi > 0$, $\mathbb{P}\left\{\left|\sum_{t = t_i + 1}^{t_i + 2L_i} \Delta_t\right| > K_3^{L_1} \xi\right\} \leq 2e^{-4L_i \xi^2}$, which implies

$$\mathbb{P}\left\{\left|\bar{W}(p, \bar{y}(p, \alpha - \beta p)) - \bar{W}^A_{i+1}(p, \bar{y}(p, \alpha - \beta p))\right| > K_3^{L_1} \xi\right\} \leq 2e^{-4L_i \xi^2}. \tag{55}$$

Denote the set of periods whose target level is above the demand realization by

$$\mathcal{C}_t \triangleq \{t : [t_i + 1, \ldots, t_i + 2L_i] : y_i > d_i\}.$$

Note that $\bar{\epsilon}_t = \epsilon_t$ for $t \in \mathcal{C}_t$. Because $y_t \in [y', y^h]$, let $\bar{n}$ be the number of order-up-to level raised during the two $L_i$ intervals in Step 1 of DDC, one has $y'(1 + s)^{\bar{n}} < y^h(1 + s)$, which is $\bar{n} < \log_{1 + s}(y^h/y') + 1$. This implies that the number of stockout during these $2L_i$ intervals is thus bounded by a constant, i.e.,

$$2L_i - |\mathcal{C}_t| < 2 \log_{1 + s}(y^h/y') + 2, \tag{56}$$

where $|\mathcal{C}_t|$ denotes the cardinality of $\mathcal{C}_t$. In addition, because $\epsilon_t$ is bounded, and by (56), there exists a constant $K_4^{L_1} > 0$ such that

$$\left|\bar{W}^A_{i+1}(p, \bar{y}(p, \alpha - \beta p)) - \bar{W}_{i+1}(p, \bar{y}(p, \alpha - \beta p))\right| < \frac{K_4^{L_1}}{2L_i}. \tag{57}$$

Thus, (50) follows from (55) and (57).
Next we focus on proving (51). We denote the empirical distribution of $\epsilon_t$ by
\[
\hat{F}^A(x) = \frac{1}{2L_i} \sum_{t=1}^{2L_i} \mathbb{1}\{\epsilon_t \leq x\}, \quad x \in [l, u].\]
For $\xi > 0$, it can be verified that
\[
\mathbb{P}\left\{\hat{F}^A(\bar{z}) < \frac{b + p}{b + p + h} - \xi\right\} = \mathbb{P}\left\{\hat{F}^A(\bar{z}) < F(\bar{z}) - \xi\right\} = \mathbb{P}\left\{\hat{F}^A(\bar{z}) - F(\bar{z}) < -\xi\right\} \leq e^{-4L_i \xi^2},
\]
where the inequality is due to the Hoeffding’s inequality. Similarly, we have
\[
\mathbb{P}\left\{\hat{F}^A(\bar{z}) > \frac{b + p}{b + p + h} + \xi\right\} \leq e^{-4L_i \xi^2}.
\]
Combining the above two inequalities, we have
\[
\mathbb{P}\left\{\left|\hat{F}^A(\bar{z}) - \frac{b + p}{b + p + h}\right| \leq \xi\right\} \geq 1 - 2e^{-4L_i \xi^2}. \tag{58}
\]
We then define a (biased) empirical distribution of $\epsilon_t$ using truncated demand data as follows,
\[
\hat{F}(x) = \frac{1}{2L_i} \sum_{t=1}^{t_i+2L_i} \mathbb{1}\{\tilde{\epsilon}_t \leq x\}, \quad x \in [l, u].
\]
By (56), we have that for some positive constant $K^{L_1}_6$,
\[
0 \leq \hat{F}(\bar{z}) - \hat{F}^A(\bar{z}) \leq \frac{K^{L_1}_6}{2L_i}. \tag{59}
\]
For any given $p \in \mathcal{P}$ and any $\xi > 0$, define the event $A_1(p, \xi)$ as follows,
\[
A_1(p, \xi) = \left\{\omega : \left|\hat{F}(\bar{z}) - \frac{b + p}{b + p + h}\right| \leq \xi + \frac{K^{L_1}_6}{2L_i}\right\}.
\]
Combining (58) and (59), we have
\[
\mathbb{P}(A_1(p, \xi)) \geq 1 - 2e^{-4L_i \xi^2}. \tag{60}
\]
For any given $p \in \mathcal{P}$, and any given $\alpha, \beta$, 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
\[
\tilde{z}^u_{i+1} = \min \left\{\tilde{\epsilon}_j : \frac{1}{2L_i} \sum_{t=1}^{t_i+2L_i} \mathbb{1}\{\tilde{\epsilon}_t \leq \tilde{\epsilon}_j\} \geq \frac{b + p}{b + p + h}\right\}.
\]
Similarly, let \( \tilde{z}_{i+1} = \hat{y}_{i+1}(p, \alpha - \beta p) - (\alpha - \beta p) \). Because \( \hat{y}_{i+1}(p, \alpha - \beta p) \in [y^l, y^h] \), 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

\[
\tilde{z}_{i+1} = \min \left\{ \max \left\{ \tilde{z}_{i+1}^u, y^l - (\alpha - \beta p) \right\}, y^h - (\alpha - \beta p) \right\}.
\]

By \( \hat{y}_{i+1}^u(p, \alpha - \beta p) = \tilde{z}_{i+1}^u + \alpha - \beta p \), we have \( \check{W}_{i+1}(p, \hat{y}_{i+1}^u(p, \alpha - \beta p)) = \check{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \). It then suffices to develop an upper bound for \( \check{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \check{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \) conditioning on the event \( A_i(p, \xi) \).

First, for any given \( d \in D \), if \( \bar{z} \leq \tilde{z}_{i+1}^u \), it follows from (53) that

\[
\begin{align*}
\check{W}_{i+1}(p, \bar{z} + \alpha - \beta p) & = \frac{1}{2L_i} \sum_{t_i + 2L_i} (b + p)(\check{\epsilon}_t - \bar{z}) \mathbb{I} \{ \tilde{z}_{i+1}^u < \check{\epsilon}_t \} \\
& \quad + (b + p)(\check{\epsilon}_t - \bar{z}) \mathbb{I} \{ \bar{z} < \check{\epsilon}_t \leq \tilde{z}_{i+1}^u \} + h(\bar{z} - \check{\epsilon}_t) \mathbb{I} \{ \check{\epsilon}_t \leq \bar{z} \} \\
& \leq \frac{1}{2L_i} \sum_{t_i + 2L_i} (b + p)(\check{\epsilon}_t - \bar{z}) \mathbb{I} \{ \tilde{z}_{i+1}^u < \check{\epsilon}_t \} \\
& \quad + (b + p)(\tilde{z}_{i+1}^u - \bar{z}) \mathbb{I} \{ \bar{z} < \check{\epsilon}_t \leq \tilde{z}_{i+1}^u \} + h(\bar{z} - \check{\epsilon}_t) \mathbb{I} \{ \check{\epsilon}_t \leq \bar{z} \},
\end{align*}
\]

where the inequality holds by replacing \( \epsilon_t \) by its upper bound \( \tilde{z}_{i+1}^u \), and

\[
\begin{align*}
\check{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) & = \frac{1}{2L_i} \sum_{t_i + 2L_i} (b + p)(\check{\epsilon}_t - \tilde{z}_{i+1}^u)) \mathbb{I} \{ \tilde{z}_{i+1}^u < \check{\epsilon}_t \} \\
& \quad + h(\tilde{z}_{i+1}^u - \check{\epsilon}_t) \mathbb{I} \{ \tilde{z}_{i+1}^u \leq \check{\epsilon}_t \} + h(\tilde{z}_{i+1}^u - \check{\epsilon}_t) \mathbb{I} \{ \check{\epsilon}_t \leq \bar{z} \} \\
& \geq \frac{1}{2L_i} \sum_{t_i \in C_i} (b + p)(\check{\epsilon}_t - \tilde{z}_{i+1}^u)) \mathbb{I} \{ \tilde{z}_{i+1}^u < \check{\epsilon}_t \} + h(\tilde{z}_{i+1}^u - \check{\epsilon}_t) \mathbb{I} \{ \check{\epsilon}_t \leq \bar{z} \},
\end{align*}
\]

where the inequality follows from dropping the nonnegative middle term. Consequently when \( \bar{z} \leq \tilde{z}_{i+1}^u \), we subtract (62) from (61) to obtain

\[
\begin{align*}
\check{W}_{i+1}(p, \bar{z} + \alpha - \beta p) - \check{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) & \leq (b + p)(\tilde{z}_{i+1}^u - \bar{z}) \left( 1 - \hat{F}(\tilde{z}_{i+1}^u) \right) + (b + p)(\tilde{z}_{i+1}^u - \bar{z}) \left( \hat{F}(\tilde{z}_{i+1}^u) - \hat{F}(\bar{z}) \right) \\
& \quad + h(\tilde{z} - \tilde{z}_{i+1}^u)\hat{F}(\bar{z}) \\
& = (\tilde{z}_{i+1}^u - \bar{z}) \left( -(b + h) + \hat{F}(\tilde{z}) + b + p \right) \\
& \leq (\tilde{z}_{i+1}^u - \bar{z})(b + p + h) \left( \xi + \frac{K_6^{L_1}}{2L_i} \right),
\end{align*}
\]
where the second inequality follows from the definition of \(A_1(p, \xi)\).

Similarly, if \(\tilde{z} > \tilde{z}_{i+1}^u\), by the symmetric argument, we have

\[
\tilde{W}_{i+1}(p, \tilde{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq (\tilde{z} - \tilde{z}_{i+1}^u)(b + p + h) \left(\xi + \frac{K_{6}^{L_1}}{2L_i}\right).
\]  

Combining (63) and (64), we have that conditioning on \(A_1(p, \xi)\),

\[
\tilde{W}_{i+1}(p, \tilde{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq K_{5}^{L_1}(b + p + h) \left(\xi + \frac{K_{6}^{L_1}}{2L_i}\right).
\]

As the demand is bounded, so is \(\tilde{z}_{i+1}^u + \alpha - \beta p\), and therefore it follows from \(\tilde{z}(p) + \alpha - \beta p \in \mathcal{Y}\) that there exists some constant \(K_{5}^{L_1} > 0\) such that \(|\tilde{z} - \tilde{z}_{i+1}^u| \leq K_{5}^{L_1}\). Thus

\[
\tilde{W}_{i+1}(p, \tilde{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq K_{5}^{L_1}(b + p + h) \left(\xi + \frac{K_{6}^{L_1}}{2L_i}\right).
\]

Since \(\tilde{z}_{i+1}^u\) is the unconstrained minimizer of \(\tilde{W}_{i+1}(p, z + \alpha - \beta p)\), it follows that

\[
\tilde{W}_{i+1}(p, \tilde{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq \tilde{W}_{i+1}(p, \tilde{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq K_{5}^{L_1}(b + p + h) \left(\xi + \frac{K_{6}^{L_1}}{2L_i}\right).
\]

For any given \(p \in \mathcal{P}\), conditioning on the event \(A_1(p, \xi)\), we obtain

\[
\tilde{W}_{i+1}(p, \tilde{z} + \alpha - \beta p) - \tilde{W}_{i+1}(p, \tilde{z}_{i+1}^u + \alpha - \beta p) \leq K_{5}^{L_1}(b + p^h + h) \left(\xi + \frac{K_{6}^{L_1}}{2L_i}\right).
\]  

\[ (65) \]

In addition, since \(\bar{y}_{i+1}(p, \alpha - \beta p)\) minimizes \(\tilde{W}_{i+1}\), we have

\[
\tilde{W}_{i+1}(p, \bar{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \bar{y}_{i+1}(p, \alpha - \beta p)) \geq 0.
\]  

\[ (66) \]

Thus, combining (65) and (66) with (60) yields

\[
\mathbb{P} \left\{ \left| (\tilde{W}_{i+1}(p, \bar{y}(p, \alpha - \beta p)) - \tilde{W}_{i+1}(p, \bar{y}_{i+1}(p, \alpha - \beta p))) \right| \leq K_{5}^{L_1}(b + p^h + h) \left(\xi + \frac{K_{6}^{L_1}}{2L_i}\right) \right\} \geq \mathbb{P}(A_1(p, \xi)) \geq 1 - 2e^{-4L_i\xi^2},
\]

which immediately implies (51).

**Proof of Lemma 2.** Define

\[
B_{i+1}^1 = \frac{1}{L_i} \sum_{t=t_i+1}^{t_{i}+L_i} \tilde{\epsilon}_t, \quad B_{i+1}^2 = \frac{1}{L_i} \sum_{t=t_i+L_i+1}^{t_{i}+2L_i} \tilde{\epsilon}_t.
\]
Recall that $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$ are derived from the least-square method, and they are given by

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

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

To measure the effectiveness of $\hat{\alpha}_{i+1}$ and $\hat{\beta}_{i+1}$ we define the (true) sample averages by

$$
B_{i+1}^{1} = \frac{1}{L_i} \sum_{t = t_i}^{t_i + L_i} \epsilon_t,
$$

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

Let $\hat{\alpha}_{i+1}^A$ and $\hat{\beta}_{i+1}^A$ are derived from the least-square method, and they are given by

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

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

Comparing (67) and (68) with (69) and (70), by (56), we have, for some constant $K_2^{1L^2} > 0$,

$$
|\hat{\alpha}_{i+1} - \hat{\alpha}_{i+1}^A| \leq \frac{K_2^{1L^2}}{L_i \delta_i},
$$

$$
|\hat{\beta}_{i+1} - \hat{\beta}_{i+1}^A| \leq \frac{K_2^{1L^2}}{L_i \delta_i}.
$$

Applying the Taylor’s expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the second order for (70), we obtain

$$
\hat{\beta}_{i+1}^A = -\left(\lambda'(\hat{p}_i) + \frac{1}{2} \lambda''(q_i) \delta_i\right) - \frac{1}{\delta_i}(-B_{i+1}^{1A} + B_{i+1}^{2A})
$$

$$
= \hat{\beta}(\hat{p}_i) - \frac{1}{2} \lambda''(q_i) \delta_i - \frac{1}{\delta_i}(-B_{i+1}^{1A} + B_{i+1}^{2A}),
$$

where $q_i \in [\hat{p}_i, \hat{p}_i + \delta_i]$. Substituting (72) into (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 $q'_i \in [\hat{p}_i, \hat{p}_i + \delta_i]$,

$$
\hat{\alpha}_{i+1}^A = \lambda(\hat{p}_i) + \frac{1}{2} \lambda'(q'_i) \delta_i + \frac{B_{i+1}^{1A} + B_{i+1}^{2A}}{2} - \lambda(\hat{p}_i) \left(\hat{p}_i + \frac{\delta_i}{2}\right)
$$

$$
+ \left(-\frac{1}{2} \lambda''(q_i) \delta_i - \frac{1}{\delta_i}(-B_{i+1}^{1A} + B_{i+1}^{2A})\right) \left(\hat{p}_i + \frac{\delta_i}{2}\right)
$$

$$
= \hat{\alpha}(\hat{p}_i) + \frac{1}{2} \lambda'(q'_i) \delta_i + \frac{B_{i+1}^{1A} + B_{i+1}^{2A}}{2} - \frac{1}{2} \lambda'(\hat{p}_i) \delta_i
$$

$$
+ \left(-\frac{1}{2} \lambda''(q_i) \delta_i - \frac{1}{\delta_i}(-B_{i+1}^{1A} + B_{i+1}^{2A})\right) \left(\hat{p}_i + \frac{\delta_i}{2}\right).
$$

Since the error terms $\epsilon_t$ are bounded, by the Hoeffding’s inequality, we have that for any $\xi > 0$, 

$$
P\{B_{i+1}^{1A} > (u - l)\xi\} \leq 2e^{-2L_i \xi^2},
$$

$$
P\{B_{i+1}^{2A} > (u - l)\xi\} \leq 2e^{-2L_i \xi^2}.
Hence, we have
\[
\mathbb{P} \left\{ -B_{i+1}^{1A} + B_{i+1}^{2A} > 2(u-l)\xi \right\} \leq \mathbb{P} \left\{ -B_{i+1}^{1A} > (u-l)\xi \right\} + \mathbb{P} \left\{ B_{i+1}^{2A} > (u-l)\xi \right\} \leq 4e^{-2L_i\xi^2},
\]
which implies that
\[
\mathbb{P} \left\{ -B_{i+1}^{1A} + B_{i+1}^{2A} \leq 2(u-l)\xi \right\} \geq \mathbb{P} \left\{ -B_{i+1}^{1A} \right\} + \mathbb{P} \left\{ B_{i+1}^{2A} \leq 2(u-l)\xi \right\} \geq 1 - 4e^{-2L_i\xi^2}.
\]

Similar argument shows
\[
\mathbb{P} \left\{ |B_{i+1}^{1A} + B_{i+1}^{2A}| \leq 2(u-l)\xi \right\} \geq 1 - 4e^{-2L_i\xi^2}.
\]

Since \( \lambda'(\cdot) \) and \( \lambda''(\cdot) \) are bounded and \( \delta_i \) converges to 0, from (73) we conclude that there must exist a constant \( K_3^{L_2} \) such that, on the event \( |B_{i+1}^{1A} + B_{i+1}^{2A}| \leq 2(u-l)\xi \) and \( -B_{i+1}^{1A} + B_{i+1}^{2A} \leq 2(u-l)\xi \), it holds that
\[
|\hat{\alpha}_{i+1}^A - \bar{\alpha}(\hat{\rho}_i)| \leq K_3^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \xi \right).
\]

Therefore,
\[
\mathbb{P} \left\{ |\hat{\alpha}_{i+1}^A - \bar{\alpha}(\hat{\rho}_i)| \leq K_3^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \xi \right) \right\} \geq \mathbb{P} \left\{ |B_{i+1}^{1A} + B_{i+1}^{2A}| \leq 2(u-l)\xi \right\} \geq 1 - 8e^{-2L_i\xi^2}. \quad (74)
\]

By (72), we have
\[
\mathbb{P} \left\{ |\hat{\beta}_{i+1}^A - \bar{\beta}(\hat{\rho}_i)| \leq K_4^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} \right) \right\} \geq 1 - 4e^{-2L_i\xi^2}. \quad (75)
\]

By (74) and (75), we have
\[
\mathbb{P} \left\{ |\hat{\alpha}_{i+1}^A - \bar{\alpha}(\hat{\rho}_i + \delta_i)| \leq K_5^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \xi \right) \right\} \geq 1 - 8e^{-2L_i\xi^2}, \quad (76)
\]
\[
\mathbb{P} \left\{ |\hat{\beta}_{i+1}^A - \bar{\beta}(\hat{\rho}_i + \delta_i)| \leq K_6^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} \right) \right\} \geq 1 - 4e^{-2L_i\xi^2}. \quad (77)
\]

Together with (71), we have
\[
\mathbb{P} \left\{ |\hat{\alpha}_{i+1} - \bar{\alpha}(\hat{\rho}_i)| \leq K_7^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \xi + \frac{1}{L_i\delta_i} \right), |\hat{\beta}_{i+1} - \bar{\beta}(\hat{\rho}_i)| \leq K_8^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \frac{1}{L_i\delta_i} \right), 
\]
\[
|\hat{\alpha}_{i+1} - \bar{\alpha}(\hat{\rho}_i + \delta_i)| \leq K_9^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \xi + \frac{1}{L_i\delta_i} \right), |\hat{\beta}_{i+1} - \bar{\beta}(\hat{\rho}_i + \delta_i)| \leq K_1^{L_2} \left( \delta_i + \frac{\xi}{\delta_i} + \frac{1}{L_i\delta_i} \right) \right\} \nonumber \quad \leq 1 - 24e^{-2L_i\xi^2}. \quad (78)
\]
To compare the two objective functions in Lemma 2, we introduce a generalized problem called (Generalized-SAA) based on (Opt-SAA) and (Approx-SAA) as follows. Given \( p_t = \hat{p}_t \) for \( t = t_i + 1, \ldots, t_i + L_i \) and \( p_t = \bar{p}_t + \delta_t \) 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 \( \zeta^1_t(\alpha_1, \beta_1) = (\zeta_t, t \in [t_i + 1, \ldots, t_i + L_i]) \) and \( \zeta^2_t(\alpha_2, \beta_2) = (\zeta_t, t \in [t_i + L_i + 1, \ldots, t_i + 2L_i]) \) with
\[
\zeta_t = d_t \wedge y_t - (\alpha_1 - \beta_1 p_t), \quad t \in [t_i + 1, \ldots, t_i + L_i],
\]
\[
\zeta_t = d_t \wedge y_t - (\alpha_2 - \beta_2 p_t), \quad t \in [t_i + L_i + 1, \ldots, t_i + 2L_i].
\]

We define a generalized function \( H_{i+1} \) by
\[
H_{i+1}(p, \alpha_1 - \beta_1 p, \zeta^1_t(\alpha_1, \beta_1), \zeta^2_t(\alpha_2, \beta_2)) = p(\alpha_1 - \beta_1 p) - \min_{y \in \mathbb{R}} \left\{ \frac{1}{2L_i} \sum_{t = t_i + 1}^{t_i + 2L_i} \left( h(y - (\alpha_1 - \beta_1 p + \zeta_t)) + (b + p)(\alpha_1 - \beta_1 p + \zeta_t - y) \right) \right\}.
\]

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.,
\[
\hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = H_{i+1}(p, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p, \zeta^1_t(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \zeta^2_t(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})).
\]

On the other hand, by setting \( \alpha_1 = \tilde{\alpha}(\hat{p}_i), \beta_1 = \tilde{\beta}(\hat{p}_i), \alpha_2 = \bar{\alpha}(\hat{p}_i + \delta_i), \beta_2 = \bar{\beta}(\hat{p}_i + \delta_i) \), and using the fact that \( \lambda(p) = \bar{\alpha}(p) - \bar{\beta}(p)p \), (Generalized-SAA) is reduced to (Approx-SAA), i.e.,
\[
\tilde{G}_{i+1}(p, \tilde{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) = H_{i+1}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p, \zeta^1_t(\tilde{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)), \zeta^2_t(\tilde{\alpha}(\hat{p}_i + \delta_i), \tilde{\beta}(\hat{p}_i + \delta_i))).
\]

Next, we see that \( H_{i+1}(p, \alpha_1 - \beta_1 p, \zeta^1_t(\alpha_1, \beta_1), \zeta^2_t(\alpha_2, \beta_2)) \) is differentiable with respect to \( \alpha_1, \alpha_2 \) and \( \beta_1, \beta_2 \) with bounded first-order derivatives. In particular, there exists a constant \( K_{i+1}^{L_2} > 0 \) such that for any \( \alpha_1, \alpha_2, \alpha'_1, \alpha'_2 \) and \( \beta_1, \beta_2, \beta'_1, \beta'_2 \), it holds that
\[
\left| H_{i+1}(p, \alpha_1 - \beta_1 p, \zeta^1_t(\alpha_1, \beta_1), \zeta^2_t(\alpha_2, \beta_2)) - H_{i+1}(p, \alpha'_1 - \beta'_1 p, \zeta^1_t(\alpha'_1, \beta'_1), \zeta^2_t(\alpha'_2, \beta'_2)) \right| \leq K_{i+1}^{L_2} \left( |\alpha_1 - \alpha'_1| + |\beta_1 - \beta'_1| + |\alpha_2 - \alpha'_2| + |\beta_2 - \beta'_2| \right).
\]

Now, by substituting (79) and (80) into (81), we see that the two objective functions \( \tilde{G}_{i+1}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) \) and \( \hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) \) differ only in their associated parameters. Consequently, Lemma 2 follows from (78) by letting \( \xi = L_i^{1/2} (\log L_i)^{1/2} \).
Proof of Proposition 2. By Lemmas 1 and 2, we have that for any \( p \in S_{i+1} \),

\[
P \left\{ \left| \bar{G} \left( p, \hat{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p \right) - \hat{G}_{i+1} \left( p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1} \right) \right| \geq K_3^{P_2} L_i^{-\frac{3}{4}} (\log L_i)^{\frac{1}{4}} \right\} \leq 28 L_i^{-2},
\]

which leads to

\[
P \left\{ \max_{p \in S_{i+1}} \left| \bar{G} \left( p, \hat{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p \right) - \hat{G}_{i+1} \left( p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1} \right) \right| \geq K_3^{P_2} L_i^{-\frac{3}{4}} (\log L_i)^{\frac{1}{4}} \right\} \leq 28 L_i^{-2} \left( \frac{p^h - p^l}{\delta_i + 1} \right) \leq K_2^{P_2} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}.
\]

Define event

\[ A_2 = \{ \omega : \max_{p \in S_{i+1}} \left| \bar{G} \left( p, \hat{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p \right) - \hat{G}_{i+1} \left( p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1} \right) \right| < K_3^{P_2} L_i^{-\frac{3}{4}} (\log L_i)^{\frac{1}{4}} \}, \]

and we have that

\[ P(A_2) > 1 - K_2^{P_2} L_i^{-\frac{7}{4}} (\log L_i)^{-\frac{1}{4}}. \]

Let \( \bar{p} \in \hat{S}_{i+1} \) be the closest point on \( \hat{S}_{i+1} \) to \( \bar{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) \) and

\[ (\bar{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) - \bar{p})(\bar{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) - \hat{p}_{i+1}) \geq 0, \]

where \( \bar{p} \) is chosen to be on the same side as \( \hat{p}_{i+1} \) relative to \( \bar{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) \) (see Figure 8).

![Figure 8 Choosing \( \bar{p} \) to be the closest point on the grid to \( \bar{p} \), and also on the same side as \( \hat{p} \) (relative to \( \bar{p} \))](image)

Applying the Taylor’s expansion of \( \bar{G}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)\hat{p}_{i+1}) \) at \( \bar{p} \), we obtain

\[
\bar{G}(\bar{p}, \hat{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)\bar{p}) - \bar{G}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)\hat{p}_{i+1})
\]
\[ G(p, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)p) \leq \hat{G}(\hat{p}_i, \tilde{\alpha}(\hat{p}_i)) + \hat{G}'(\hat{p}_i, \tilde{\beta}(\hat{p}_i)p) \hat{p} \leq \hat{G}(\hat{p}_i, \tilde{\alpha}(\hat{p}_i)) + \hat{G}'(\hat{p}_i, \tilde{\beta}(\hat{p}_i)p) \hat{p} = 0. \]

when \( \hat{p}_i \leq \hat{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) \), (83) holds true by similar arguments.

On the other hand, conditioning on \( A_2 \), we have

\[ \hat{G}(\hat{p}_i, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_i) + K_3^p L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \hat{G}(\hat{p}_i, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_i) \geq \hat{G}(\hat{p}_i, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_i) + K_3^p L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \]

where the first and the last inequalities follow from the definition of \( A_2 \), and the second inequality holds because \( \hat{p}_i \) is the maximizer for \( \hat{G}(\hat{p}_i, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_i) \) on \( S_{i+1} \). Therefore,

\[ \hat{G}(\hat{p}_i, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}) - \hat{G}(\hat{p}_i, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_i) \leq 2K_3^p L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}. \]

Together with (82) and (84), one has

\[ \hat{p}_{i+1} - \hat{p} \leq K_4^p L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}, \]

which leads to, by conditioning on \( A_2 \),

\[ |\hat{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) - \hat{p}_{i+1}| \leq |\hat{p}(\hat{\alpha}(\hat{p}_i), \tilde{\beta}(\hat{p}_i)) - \hat{p}| + |\hat{p} - \hat{p}_{i+1}| \leq \delta_{i+1} + K_4^p L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} \leq K_1^p L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}. \]

This completes the proof of Proposition 2. \( \square \)

C: Technical Proofs for Theorem 1(b)

Proof of Lemma 3. For \( p \in \mathcal{P} \), one has

\[ \tilde{y} (p, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)p) = F^{-1} \left( \frac{b + p}{b + p + h} \right) + \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)p. \]
For a given $p \in P$, we define $	ilde{y}^u_{i+1}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p)$ as the unconstrained optimal target inventory level for (Approx-SAA) on $\mathbb{R}_+$, then it can be verified that

$$
\tilde{y}^u_{i+1}\left(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p\right) = \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p
$$

$$
+ \min\left\{ \tilde{\epsilon}_j : j \in [t_i + 1, \ldots, t_i + 2L_i], \frac{\sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\{\tilde{\epsilon}_t \leq \tilde{\epsilon}_j\}}{2L_i} \geq \frac{b + p}{b + p + \tilde{h}} \right\}, \tag{86}
$$

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

Then we have

$$
\left| \tilde{y}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) - \tilde{y}_{i+1}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) \right| \\
\leq \left| \tilde{y}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) - \tilde{y}_{i+1}^u(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) \right|.
$$

There exists a positive constant $K_2^{\mathbb{L}_3}$ such that, by (56), for any $\xi > 0$

$$
\frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\{\epsilon_t \leq F^{-1}\left(\frac{b + p}{b + p + h} - \xi\right)\} \leq \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\{\epsilon_t \leq F^{-1}\left(\frac{b + p}{b + p + h} - \xi\right)\} + \frac{K_2^{\mathbb{L}_3}}{2L_i}. \tag{88}
$$

Now, for any $\xi \geq K_2^{\mathbb{L}_3}/L_i$, we have

$$
\mathbb{P}\left\{ F\left(\tilde{y}_{i+1}^u(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) - (\tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) \right) - \frac{b + p}{b + p + h} \leq -\xi \right\} \tag{89}
$$

$$
= \mathbb{P}\left\{ \tilde{y}_{i+1}(p, \tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) - (\tilde{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)p) \leq F^{-1}\left(\frac{b + p}{b + p + h} - \xi\right) \right\} \leq \mathbb{P}\left\{ \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\{\epsilon_t \leq F^{-1}\left(\frac{b + p}{b + p + h} - \xi\right)\} - \frac{b + p}{b + p + h} \leq -\xi \right\}
$$

$$
\leq \mathbb{P}\left\{ \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\{\epsilon_t \leq F^{-1}\left(\frac{b + p}{b + p + h} - \xi\right)\} - \frac{b + p}{b + p + h} \geq \xi \right\}
$$

$$
\leq \mathbb{P}\left\{ \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} \mathbb{1}\{\epsilon_t \leq F^{-1}\left(\frac{b + p}{b + p + h} - \xi\right)\} + \frac{K_2^{\mathbb{L}_3}}{2L_i} - \frac{b + p}{b + p + h} \geq \xi \right\}
$$

where the first inequality follows from (86), the second inequality holds from using (88), and the last inequality holds as $\xi \geq K_2^{\mathbb{L}_3}/L_i$, $\xi - \frac{K_2^{\mathbb{L}_3}}{2L_i} \geq \frac{\xi}{2}$. 
Since \( \mathbb{E} \left[ 1 \{ \epsilon_t \leq F^{-1} \left( \frac{b+p}{b+p+h} - \xi \right) \} \right] = \frac{b+p}{b+p+h} - \xi \), we apply the Hoeffding’s inequality to obtain

\[
P \left\{ \frac{1}{2L_i} \sum_{t=t_i+1}^{t_i+2L_i} 1 \{ \epsilon_t \leq F^{-1} \left( \frac{b+p}{b+p+h} - \xi \right) \} - \left( \frac{b+p}{b+p+h} - \xi \right) \geq \frac{\xi}{2} \right\} \leq e^{-L_i \xi^2}.
\]

Together with (85) and (89), we have

\[
P \left\{ F \left( \tilde{y}_{i+1} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right\} - F \left( \tilde{y} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right) \leq -\xi \right\} \leq e^{-L_i \xi^2}. \tag{90}
\]

Similarly, we have

\[
P \left\{ F \left( \tilde{y}_{i+1} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right\} - F \left( \tilde{y} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right) \geq \xi \right\} \leq e^{-L_i \xi^2}. \tag{91}
\]

We have assumed the probability density function \( f(\cdot) \) of \( \epsilon_t \) 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 (90) and (91), for any \( \xi \geq K_2^{L_3} / L_i \), we obtain

\[
2e^{-L_i \xi^2} \geq P \left\{ \left| F \left( \tilde{y}_{i+1} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right| \geq \xi \right\} \\
\geq P \left\{ \left| F \left( \tilde{y}_{i+1} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right| \geq \xi \right\} \\
\geq P \left\{ \left| \tilde{y} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) - \left( \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right| \geq \xi \right\} \\
= P \left\{ \left| \tilde{y}_{i+1} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) - \tilde{y} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right| \geq \frac{1}{r} \xi \right\},
\]

where the second inequality follows from (87). For constant \( K_1^{L_3} = \frac{1}{r} \) we have, for \( \xi \geq K_2^{L_3} / L_i \),

\[
P \left\{ \left| \tilde{y}_{i+1} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) - \tilde{y} \left( p, \hat{\alpha}(\tilde{p}_i) - \tilde{\beta}(\tilde{p}_i)p \right) \right| \geq K_1^{L_3} \xi \right\} \leq 2e^{-L_i \xi^2}, \tag{92}
\]
By letting \( \xi = L_i^{-\frac{1}{2}}(\log L_i)\frac{1}{2} \), Lemma 3 follows from (92).

**Proof of Lemma 4.** Following the analysis of Lemma 2, 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 \( \zeta_i \)'s defined in the proof of Lemma 2)

\[
y_{i+1}(p, \alpha_1 - \beta_1 p, \zeta_i^1(\alpha_1, \beta_1), \zeta_i^2(\alpha_2, \beta_2)) = \arg\min_{y \in \mathcal{Y}} \left\{ \frac{1}{2L_i} \sum_{t=i+1}^{t=2L_i} \left( h(y - (\alpha_1 - \beta_1 p + \zeta_i))^+ + (b + p)(\alpha_1 - \beta_1 p + \zeta_i - y)^+ \right) \right\}.
\]  

(93)

We see that \( y_{i+1}(p, \alpha_1 - \beta_1 p, \zeta_i^1(\alpha_1, \beta_1), \zeta_i^2(\alpha_2, \beta_2)) \) are differentiable with respect to \( \alpha_1, \alpha_2 \) and \( \beta_1, \beta_2 \) with bounded first-order derivatives. In particular, there exists a constant \( K_2^{LA} > 0 \) such that for any \( \alpha_1, \alpha_2, \alpha'_1, \alpha'_2 \) and \( \beta_1, \beta_2, \beta'_1, \beta'_2 \), it holds that

\[
\left| y_{i+1}(p, \alpha_1 - \beta_1 p, \zeta_i^1(\alpha_1, \beta_1), \zeta_i^2(\alpha_2, \beta_2)) - y_{i+1}(p, \alpha'_1 - \beta'_1 p, \zeta_i^1(\alpha'_1, \beta'_1), \zeta_i^2(\alpha'_2, \beta'_2)) \right|
\leq K_2^{LA} \left| |\alpha_1 - \alpha'_1| + |\beta_1 - \beta'_1| + |\alpha_2 - \alpha'_2| + |\beta_2 - \beta'_2| \right|.
\]  

(94)

In addition, we know that

\[
y_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_{i+1}) = y_{i+1}(\hat{p}_{i+1}, \hat{\alpha}(\hat{p}_i) - \hat{\beta}(\hat{p}_i)\hat{p}_{i+1}, \zeta_i^1(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \zeta_i^2(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})),
\]  

(95)

Thus, Lemma 4 follows by combining (78), (94), (95), and (96) and letting \( \xi = L_i^{-\frac{1}{2}}(\log L_i)\frac{1}{2} \).

**Proof of Proposition 3.** From (38),

\[
\mathbb{E} \left[ \left| \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \bar{y}_{i+1,1} \right|^2 \right] \leq K_2^{P3} \mathbb{E} \left[ \left| \hat{p}_{i+1} - p^* \right|^2 + \left| \hat{p}_i - p^* \right|^2 \right] + K_2^{P3} \mathbb{E} \left[ \left| \bar{y}(\hat{p}_{i+1}, \bar{\alpha}(\hat{p}_i) - \bar{\beta}(\hat{p}_i)\hat{p}_{i+1}) - \bar{y}_{i+1}(\hat{p}_{i+1}, \bar{\alpha}(\hat{p}_i) - \bar{\beta}(\hat{p}_i)\hat{p}_{i+1}) \right|^2 \right]
\]  

(97)

By Theorem 1(a), we have

\[
\mathbb{E} \left[ \left| \hat{p}_{i+1} - p^* \right|^2 + \left| \hat{p}_i - p^* \right|^2 \right] \leq K_3^{P3} L_i^{-\frac{1}{2}}(\log L_i)\frac{1}{2}.
\]  

(98)

And it follows from Lemma 3 that

\[
\mathbb{E} \left[ \left| \bar{y}(\hat{p}_{i+1}, \bar{\alpha}(\hat{p}_i) - \bar{\beta}(\hat{p}_i)\hat{p}_{i+1}) - \bar{y}_{i+1}(\hat{p}_{i+1}, \bar{\alpha}(\hat{p}_i) - \bar{\beta}(\hat{p}_i)\hat{p}_{i+1}) \right|^2 \right]
\]  

(99)
\[
\frac{K_{\text{P}3}L_i}{P_i} + K_{\text{P}3}L_i^{-1} \log L_i \leq K_{\text{P}3}L_i^{-1} \log L_i.
\] (100)

In Lemma 4,
\[
\mathbb{E}\left[\left|\bar{y}(\hat{p}_{i+1}, \bar{\alpha}(\hat{p}_i) - \tilde{\beta}(\hat{p}_i)\hat{p}_{i+1}) - \hat{y}_{i+1,1}\right|^2\right] \leq K_{\text{P}3}L_i^{-2} + K_{\text{P}3}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}} \leq K_{\text{P}3}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}. \tag{101}
\]

Proposition 3 follows from (98), (99), and (101).

D: Technical Proofs for Theorem 2

Proof of Proposition 4. To prove Proposition 4, we proceed as follows. Using the fact that
\[y^* = \bar{y}(p^*, \lambda(p^*)),\]
we have
\[
\mathbb{E}\left[Q(p^*, \bar{y}(p^*, \lambda(p^*))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\right] \\
\leq \mathbb{E}\left[Q(p^*, \bar{y}(p^*, \lambda(p^*))) - Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) + Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\right].
\]

The first term \(Q(p^*, \bar{y}(p^*, \lambda(p^*))) - Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})))\) 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 \(K_{\text{P}4} > 0\), that
\[
\mathbb{E}\left[Q(p^*, \bar{y}(p^*, \lambda(p^*))) - Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})))\right] \leq \mathbb{E}\left[K_{\text{P}4}(p^* - \hat{p}_{i+1})^2\right]
\leq K_{\text{P}4}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}. \tag{102}
\]

To bound the second term \(Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\), notice that \(\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))\) maximizes the concave function \(Q(\hat{p}_{i+1}, y)\) for any given \(\hat{p}_{i+1}\), we apply Taylor expansion with respect to \(y\) at point \(y = \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))\) to yield that, for some constant \(K_{\text{P}4}\),
\[
Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1}) \leq K_{\text{P}4}\left(\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \hat{y}_{i+1,1}\right)^2, \tag{103}
\]
which leads to
\[
\mathbb{E}\left[Q(\hat{p}_{i+1}, \bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1}))) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})\right] \leq K_{\text{P}4}\mathbb{E}\left[(\bar{y}(\hat{p}_{i+1}, \lambda(\hat{p}_{i+1})) - \hat{y}_{i+1,1})^2\right]
\leq K_{\text{P}4}L_i^{-\frac{1}{2}}(\log L_i)^{\frac{1}{2}}, \tag{104}
\]
where the second inequality follows from Proposition 3.
By (102) and (104), we have

\[
\mathbb{E} \left[ \sum_{i=1}^{n} \sum_{t=t_i+2L_i+1}^{t_i+L_i} (Q(p^*_t, y^*_t) - Q(\hat{p}_{i+1}, \hat{y}_{i+1,1})) \right] \\
\leq \mathbb{E} \left[ \sum_{i=1}^{n} I_i^{1/2} (\log I_i)^{1/2} I_i \right] \\
\leq K T^{1/2} \mathbb{E} (\log T)^{1/2}.
\]

This completes the proof of Proposition 4.

Proof of Proposition 5. Consider the accumulative demands during periods \(t_i + 2L_i + 1\) to \(t_i + 2L_i + \left\lfloor I_i^{1/2} \right\rfloor\). If these accumulative demands consume at least \(x_{t_i+2L_i+1} - \hat{y}_{i+1}\), then at period \(t_i + 2L_i + \left\lfloor I_i^{1/2} \right\rfloor + 1\), \(\hat{y}_{i+1}\) will be surely achieved. Since \(\lambda(p') + l \leq D_t \leq \lambda(p') + u\) for all \(t\), by Hoeffding inequality, for any \(\zeta > 0\),

\[
\mathbb{P} \left\{ \sum_{t=t_i+2L_i+1}^{t_i+2L_i+\left\lfloor I_i^{1/2} \right\rfloor} D_t \geq \mathbb{E} \left[ \sum_{t=t_i+2L_i+1}^{t_i+2L_i+\left\lfloor I_i^{1/2} \right\rfloor} D_t \right] - \zeta \right\} \geq 1 - \exp \left( - \frac{2\zeta^2}{\mathbb{E} \left[ D_{t_i+2L_i+1} \right] - \zeta} \right),
\]

Now choose \(\zeta = (\lambda(p') + u - \lambda(p') - l) \left( \left\lfloor I_i^{1/2} \right\rfloor \right)^{1/2} \left( \log \left\lfloor I_i^{1/2} \right\rfloor \right)^{1/2}\), and define the event \(A_3\)

\[
A_3 = \left\{ \sum_{t=t_i+2L_i+1}^{t_i+2L_i+\left\lfloor I_i^{1/2} \right\rfloor} D_t \geq \left\lfloor I_i^{1/2} \right\rfloor \mathbb{E} \left[ D_{t_i+2L_i+1} \right] - \zeta \right\}.
\]

Then it follows from (105) that \(\mathbb{P}(A_3) \geq 1 - \left\lfloor I_i^{1/2} \right\rfloor^{-2}\).

Since \(\lambda(p') > 0\), \(\mathbb{E} \left[ D_{t_i+2L_i+1} \right] > 0\). Then there exists some \(i^{**}\) such that whenever \(i \geq i^{**}\),

\[
\left\lfloor I_i^{1/2} \right\rfloor \mathbb{E} \left[ D_{t_i+2L_i+1} \right] - \zeta \geq \frac{1}{2} \left\lfloor I_i^{1/2} \right\rfloor \mathbb{E} \left[ D_{t_i+2L_i+1} \right] \geq y^h - y^l \geq x_{t_i+2L_i+1} - \hat{y}_i,
\]

which suggests that on the event \(A_3\), the order-up-to level \(\hat{y}_i\) will always be achieved after periods \(\{t_i + 2L_i + 1, \ldots, t_i + 2L_i + \left\lfloor I_i^{1/2} \right\rfloor\}\).

Now using \(\mathbb{P}(A_3^c) \leq \left\lfloor I_i^{1/2} \right\rfloor^{-2}\), we have

\[
\mathbb{E} \left[ \sum_{t=t_i+2L_i+1}^{t_i+I_i} (G(\hat{p}_t, \hat{y}_{i,1}) - G(p_t, y_t)) \right] \\
= \mathbb{P}(A_3) \mathbb{E} \left[ \sum_{t=t_i+2L_i+1}^{t_i+I_i} (G(\hat{p}_t, \hat{y}_{i,1}) - G(p_t, y_t)) \left| A_3 \right. \right] \\
+ \mathbb{P}(A_3^c) \mathbb{E} \left[ \sum_{t=t_i+2L_i+1}^{t_i+I_i} (G(\hat{p}_t, \hat{y}_{i,1}) - G(p_t, y_t)) \left| A_3^c \right. \right]
\]
Initialize

Step 0: Preparation.

\[
\begin{align*}
\max \{h, b + p^h\} (y^h - y') \left[ I_{t_0}^i \right] + \left[ I_{t_0}^i \right]^{-2} & \max \{h, b + p^h\} (y^h - y') I_i \\
\leq 2 \max \{h, b + p^h\} (y^h - y') I_i^2.
\end{align*}
\]

Then the result of Proposition 5 follows.

\[\square\]

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 \(\rho\), 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 \(I_1 = \lfloor I_0v \rfloor\) and the first stage consists of \(2I_1\) periods. For the first \(I_1\) periods of stage 1 \((t = 1, \ldots, I_1)\), the algorithm proceeds with initial parameters \(\hat{p}_1\) 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 \(\delta_1\) amount (i.e., \(\hat{p}_1 + \delta_1\)), where \(\delta_1 = \rho(2I_0)^{-\frac{1}{4}}(\log(2I_0))^\frac{1}{2}\). 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 \(i \geq 1\),

\[I_i = \lfloor I_0v^i \rfloor, \quad \delta_i = \rho(2I_{i-1})^{-\frac{1}{4}}(\log(2I_{i-1}))^\frac{1}{2}, \quad \text{and} \quad t_i = \sum_{k=1}^{i-1} 2I_k. \tag{107}\]

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 \(S_i = [p^i, p^i + \delta_i, p^i + 2\delta_i, \ldots, p^h]\).

Now we present the learning algorithm DDO for the observable demand case.

Step 0: Preparation. Initialize \(I_0, v\) and \(\rho\), and \(\hat{p}_1, \hat{y}_{1,1}, \hat{y}_{1,2}, \delta_1, I_1\) as shown above.

For each stage \(i = 1, \ldots, n\) where \(n = \left\lfloor \log_v \left( \frac{v^i - 1}{2I_0} T + 1 \right) \right\rfloor\), repeat the three steps below:

Step 1: Setting prices and order-up-to level for stage 1.

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,

\[y_t = \hat{y}_{i,1} \lor x_t, \text{ for all } t = t_i + 1, \ldots, t_i + I_i,\]

\[y_t = \hat{y}_{i,2} \lor x_t, \text{ for all } t = t_i + I_i + 1, \ldots, t_i + 2I_i.\]

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

Let \(d_t\) be demand realizations for \(t = t_i + 1, \ldots, t_i + 2I_i\). Solve a least-square problem

\[\widehat{\alpha}_{t+1}, \widehat{\beta}_{t+1} = \arg\min \sum_{t = [t_i + 1, t_i + 2I_i]} [d_t - (\alpha - \beta p_t)]^2, \tag{108}\]

\(\square\)
\[ \eta_t = d_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t) = \lambda(p_t) + \epsilon_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t) \text{ for } t = t_i + 1, \ldots, t_i + 2I_i. \] (109)

**Step 3: Maximize the proxy profit** \( Q_{SAA}(p, y) \).

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

\[
\max_{(p, y) \in S_i \times Y} Q_{SAA}^{i+1}(p, y) \triangleq \max_{p \in S_i+1} \left\{ \hat{G}_{i+1}(p, \hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \right. \\
- \min_{y \in Y} \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \left( (b + p) \left( \hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p + \eta_t - y \right) + h \left( y - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p + \eta_t) \right) \right) \right\}.
\]

Then, set the first pair of price and order-up-to level to

\[ (\hat{p}_{i+1}, \hat{y}_{i+1}, 1) = \arg \max_{(p, y) \in S_i+1 \times Y} Q_{SAA}^{i+1}(p, y), \]

and compute \( \hat{y}_{i+1,2} \) as

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

In case that \( \hat{p}_{i+1} + \delta_{i+1} \not\in S_i+1 \), 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 \( K_2^U, K_3^U, K_4^U \) such that

\[ \mathbb{P}\{D_t(p) \geq x\} \leq K_2^U \exp(-K_3^U x) \text{ for any } x \geq K_4^U. \]

We assume the inventory search space ceiling \( y^h \) is at least \( 1/(2K_3^U) \log T \) for the problem with planning horizon \( T \). In our algorithm we explore the inventory space up to \( 1/(2K_3^U) \log T \). For convenience we set \( y^h = 1/(2K_3^U) \log T \).

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 \( K_5^U \),

\[ \mathbb{P}\left(-\sigma_i L_i^{-\frac{1}{2}} \left( \log L_i \right)^{\frac{1}{2}} \leq \frac{\sum_{t=t_i+1}^{t_i+L_i} D_t(p)}{L_i} - \mathbb{E}[D_t(p)] \leq \sigma_i L_i^{-\frac{1}{2}} \left( \log L_i \right)^{\frac{1}{2}} \right) \geq 1 - \frac{K_5^U}{L_i}, \] (110)
where $\sigma_1$ is the standard deviation of $D_i(p)$, and similarly we have

$$
P\left(\frac{\sum_{t=i}^{i+L_i} (D_i(p) - y^h)^+}{L_i} - \mathbb{E}[(D_i(p) - y^h)^+] \leq \sigma_2 L_i^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}\right) \geq 1 - \frac{K^U_6}{L_i}, \quad (111)$$

where $\sigma_2$ is the standard deviation of $(D_i(p) - y^h)^+$. 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 $\sigma_2^2 = \text{Var}[(D_i(p) - y^h)^+]$ is continuous and decreasing in $y^h$ from $\sigma_2^2 = \text{Var}[D_i(p)]$ 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 $F_{D_i(p)}$ and $f_{D_i(p)}$ denote the cdf and pdf of $D_i(p)$, respectively. Then

$$
\sigma_2^2 = \int_{y^h}^{\infty} (x - y^h)^2 f_{D_i(p)}(x) dx - \left( \int_{y^h}^{\infty} (x - y^h) f_{D_i(p)}(x) dx \right)^2, \quad \text{and}
$$

$$
\frac{\partial \sigma_2^2}{\partial y^h} = -2 F_{D_i(p)}(y^h) \int_{y^h}^{\infty} (x - y^h) f_{D_i(p)}(x) dx \leq 0,
$$

and $\sigma_2^2$ is decreasing in $y^h$.

It is clear that $\sigma_2^2$ 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, $\sigma_2^2$ is continuously decreasing in $y^h$ from $\sigma_1^2$ to 0 as $y^h$ goes from 0 to infinity.

To bound $\frac{1}{L_i} \sum_{t=i}^{i+L_i} \min\{D_i(p), y^h\} - \mathbb{E}[D_i(p)]$, we notice that $\min\{D_i(p), y^h\} = D_i(p) - (D_i(p) - y^h)^+$, and by Assumption (a), we have

$$
\mathbb{E}[(D_i(p) - y^h)^+] = \int_0^{\infty} \mathbb{P}(D_i(p) - y^h \geq x) dx \leq \int_0^{\infty} K^U_2 e^{-K^U_4 (x+y^h)} dx = \frac{K^U_2}{K^U_3} e^{-K^U_4 y^h}.
$$

Hence, by our choice of $y^h$, $\frac{K^U_2}{K^U_3} e^{-K^U_4 y^h} = \frac{K^U_2}{K^U_3} T^{-\frac{1}{2}} \leq \frac{K^U_2}{K^U_3} L_i^{-\frac{1}{2}}$ when $T$ is large enough. Consequently for large enough $i$, we have

$$
\mathbb{E}[(D_i(p) - y^h)^+] \leq \frac{K^U_2}{K^U_3} L_i^{-\frac{1}{2}} \leq \sigma_2(L_i)^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}. \quad (112)
$$

Combining (111) and (112) yields

$$
P\left(\frac{\sum_{t=i}^{i+L_i} (D_i(p) - y^h)^+}{L_i} \leq 2\sigma_2(L_i)^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}}\right) \geq 1 - \frac{K^U_6}{L_i}. \quad (113)$$

Let $\mathcal{A}_4$ be the event for (110), and $\mathcal{A}_5$ be the event for (113). Writing (110) as

$$
-\sigma_1(L_i)^{-\frac{1}{2}} (\log L_i)^{\frac{1}{2}} - \frac{\sum_{t=i}^{i+L_i} (D_i(p) - y^h)^+}{L_i} \leq \sum_{t=i}^{i+L_i} \min\{D_i(p), y^h\} - \mathbb{E}[D_i(p)] \quad (114)
$$
\begin{align*}
\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}.
\end{align*}

Then it can be seen that on the event \( A_4 \cap A_5 \), it holds that
\begin{align*}
(-\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}}. \quad (115)
\end{align*}

Thus by \(-2\sigma_2 \geq -\sigma_1\), it follows from (110) and (113) that
\begin{align*}
\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) \quad (116)
\end{align*}
\begin{align*}
\geq \mathbb{P}(A_4 \cap A_5) = 1 - \mathbb{P}(A_4^c \cap A_5^c) \geq 1 - \frac{K_5^U + K_6^U}{L_i}.
\end{align*}

Comparing this result with (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 2, the regret is amplified by \( y^h - y' = \mathcal{O}(1) \) for the bounded demand case, whereas the regret is amplified by \( y^h - y' = \mathcal{O}(\log T) \) for the unbounded demand case, resulting in a total regret of \( \mathcal{O}(T^{-\frac{1}{2}} \log T) \).