Coordinating Pricing and Inventory Replenishment with Nonparametric Demand Learning

Boxiao Chen[1], Xiuli Chao[2] and Hyun-Soo Ahn[3]

Abstract

We consider a firm (e.g., retailer) selling a single non-perishable product over a finite-period planning horizon. Demand in each period is stochastic and price-sensitive, and unsatisfied demands are backlogged. At the beginning of each period, the firm determines its selling price and inventory replenishment quantity, with the objective of maximizing total profit, but it knows neither the average demand (as a function of price) nor the distribution of demand uncertainty *a priori*, hence it has to make pricing and ordering decisions based on observed demand data. We propose a nonparametric data-driven algorithm that learns about the demand on the fly and, concurrently, applies learned information to make replenishment and pricing decisions. The algorithm integrates learning and action in a sense that the firm actively experiments on pricing and inventory levels to collect demand information with minimum profit loss. Besides convergence of optimal policies, we show that the regret of the algorithm, defined as the average profit loss compared with that of the optimal solution had the firm known the underlying demand information, vanishes at the fastest possible rate as the planning horizon increases.

Keywords: dynamic pricing, inventory control, demand learning, nonparametric estimation, non-perishable products, asymptotic optimality.

1 Introduction

Balancing pricing and inventory replenishment decisions is a challenge for many firms, and failure to do so can directly affect the bottom-line of a company. With the fast developing information and operational technologies, a firm can adjust the pricing and ordering/production decisions very

[1]Department of Information and Decision Science, College of Business Administration, University of Illinois, Chicago, IL 60607. Email: bbchen@uic.edu

[2]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109. Email: xchao@umich.edu

[3]Department of Technology and Operations, Ross School of Business, University of Michigan, Ann Arbor, MI 48109. Email: hsahn@umich.edu

frequently, e.g., household products and daily necessities, digital products, and computer accessories. In the academic literature, this problem has been studied by many authors, see e.g., the survey papers by Petruzzi and Dada (1999), Elmaghraby and Keskinocak (2003), Yano and Gilbert (2003), and Chen and Simchi-Levi (2012). The majority of the papers on joint optimization of pricing and inventory control have assumed that the firm knows how the market responds to its selling prices and the exact distribution of uncertainty in customer demand *a priori*. This, however, may not be true in many applications.

In practice, a firm usually gains understanding of demand for a product from historical data. However, when a new generation of the product is released, such data is not available. Even for a product that has been in the market for a long time, data from the far past may not be very useful because demand may be affected by the economic environment, the state of the world, and many other factors, thus the demand will stay relatively stationary only for a certain period of time. Therefore, the scenario considered in this study is that the demand in the next given number of periods, referred to as planning horizon, is stationary, but the firm has limited prior information about the demand distribution. The firm needs to learn about the demand on the fly with the objective of maximizing total profit over the planning horizon. In recent years, data-driven optimization has become practical as many firms have the capability to collect demand data in short time periods (e.g., several times in an hour).

We consider a periodic-review non-perishable inventory system that sells a single product over a planning horizon of T periods. The demand in each period is random, with either multiplicative error (discussed in the main context) or additive error (discussed in Appendix D), but the firm does not have prior knowledge about average customer response to selling price or the distribution of the error term. Unsatisfied demand in each period is backlogged. For each period, the firm sets its price and makes inventory replenishment decisions in anticipation of price-sensitive and uncertain demand. If the firm has complete information about the underlying demand distribution *a priori*, this problem has been studied by a considerable number of authors, see e.g., Federgruen and Heching (1999). We develop a nonparametric data-driven algorithm that learns the demand distribution using observed demand data while at the same time maximizing total profit. We measure the performance of the algorithm by regret, which is defined as the average profit loss per period compared with that of the optimal solution had the firm known the demand distribution, and show that the regret of our algorithm converges to zero at the fastest possible rate of any learning algorithms for this class of problems.

Positioning of this paper. There are a number of papers studying revenue management or pure inventory control using learning, very few on both. Revenue management considers pricing problems with either an infinite or a fixed finite inventory. The former has no inventory constraint, among which Keskin and Zeevi (2014) study a parametric linear model and Besbes and Zeevi (2015) prove the (surprising) sufficiency of linear models in a more general nonparametric setting. The latter has knapsack constraint on total inventory and has been studied in Besbes and Zeevi (2009, 2012), Wang et al. (2014), and Ferreira et al. (2016), among others. A number of papers have studied demand learning in the repetitive newsvendor setting (i.e., no inventory carryovers), see e.g., Besbes and Muharremoglu

(2013) and Huh et al. (2011). Huh and Rusmevichientong (2009) consider a pure inventory control problem and obtain convergence rate for the regret of their online learning algorithm. Levi et al. (2007, 2015) apply sample average approximation (SAA) using true demand samples, and analyze the dependency of estimation accuracy on the number of samples. To the best of our knowledge, Burnetas and Smith (2000) is the only paper that considers the joint pricing and inventory control problem in a nonparametric setting, but the authors assume no inventory carryovers, i.e., any leftover inventory at the end of a period perishes, thus the authors are able to apply off-the-shelf existing algorithms from the unconstrained multi-armed bandit literature to solve it. Burnetas and Smith (2000) analyze the convergence of pricing and inventory decisions, but not the regret of their algorithm. This paper is the first to develop a learning algorithm for the joint pricing and inventory control problem of non-perishable products and obtain the rate of regret for the algorithm.

Challenges and contributions. The most closely related work is Besbes and Zeevi (2015), which develops a nonparametric learning algorithm for a dynamic pricing problem without inventory control decisions. We follow Besbes and Zeevi (2015) to use a linear approximation method to estimate the mean demand function, but there are several critical distinctions between our work and theirs. At a high level, the differences can be summarized from three perspectives:

(1) In the pure pricing optimization problem as studied in Besbes and Zeevi (2015), it is sufficient to estimate the average demand function. In contrast, when inventory is involved, we need to also estimate the entire distribution of the random demand. Indeed, in the special case where the optimal price is known *a priori*, our problem reduces to the classic inventory problem for which it is well-known that demand distribution is the key in determining the optimal inventory policy (e.g., base-stock level). Therefore in our paper, we have to estimate both the average demand function and the distribution that governs the randomness in demand. A major difficulty faced in this endeavor is that in our setting, i.i.d. samples of random error cannot be obtained and results from the literature on SAA cannot be applied. Hence new techniques have to be developed to resolve it.

(2) Different from Besbes and Zeevi (2015), the pricing decisions in our system not only depend on the pricing information but also on the past inventory levels. With inventory as a part of the system state, the pricing and inventory decisions are coupled over time, which introduces significant complexity. Specifically, in Besbes and Zeevi (2015), the pricing decision in a period only affects the profit of the current period. In contrast, when inventory replenishment is introduced, the decision in a period can impact the profits for many subsequent periods due to inventory carryovers.

(3) From the point of view of the learning literature, Besbes and Zeevi (2015) study an unconstrained optimization problem, i.e., the feasible region of a later decision is not affected by earlier decisions. In contrast, our problem is a constrained optimization problem due to the carry-over inventory constraints. In Huh and Rusmevichientong (2009), a queueing analysis is taken to evaluate the impact of the carry-over inventory constraint for a pure inventory management problem. They

establish a connection between the application of the stochastic gradient method and the waiting time process in a single-server GI/D/1 queue, whose service time parameter is given by the step size of the gradient descent method. This development is viewed as one of their primary contributions. In our algorithm, we keep an order-up-to target unchanged for longer and longer time intervals, and eventually the order-up-to target level is achieved with a high probability as demand gradually consumes the on-hand inventory.

We summarize the contributions of this paper, in reference to the relevant works in the existing literature, as follows:

(i) We establish the first convergence result of learning algorithms for joint pricing and inventory control problem with non-perishable products. Actually, even when pricing is not a decision, there exists no policy convergence result of any learning algorithm in the literature on non-perishable inventory control problem.

(ii) We also present the first result on convergence rate of regret for learning algorithms of the joint pricing and inventory control problem, which is known to be also the lower bound for the regret of any learning algorithm for this class of problems. One important implication of our finding is that the linear demand approximation scheme of Besbes and Zeevi (2015) actually achieves the best possible convergence rate of regret, hence their linear model is sufficient in developing learning algorithms not only for revenue management, but also for joint pricing and inventory control problems.

(iii) A major technical issue faced in this paper is that true samples of random demand error cannot be observed. Estimated and dependent error samples are constructed using realized demand data, and we refer to them as centered samples in this paper. A method similar to SAA is employed to construct proxy objective functions, but different from the traditional SAA in Levi et al. (2007, 2015), our construction is based on centered samples, so it suffers from data bias. Furthermore, Levi et al. (2007, 2015) consider a static optimization problem with T observations/samples, while we have an online optimization problem.

Approaches to designing nonparametric algorithms. We end this section with a brief overview of the main approaches in the literature on developing nonparametric data-driven algorithms, which include online convex optimization (Agarwal et al. 2011, Zinkevich 2003, Hazan et al. 2006), continuum-armed bandit problems (Auer et al. 2007, Kleinberg 2005, Cope 2009), stochastic approximation (Kiefer and Wolfowitz 1952, Lai and Robbins 1981, and Robbins and Monro 1951), and so on. These results are developed for general settings and have found many applications. However, most of these methodologies require that the proposed solution be reachable in each and every period, which is not the case with our problem due to the carry-over inventory constraint for non-perishable inventory systems.

There is an increasing interest in constrained data-driven optimization problems. In constrained stochastic approximation, there is an exogenous constraint that has to be satisfied by decisions in each period (see surveys in Kushner (2010) and Kushner and Yin (2003)), and normalized iterates are usually used to prove convergence of the interpolated process. Kushner and Yin (1997) assume that the true optimal solution falls in the interior of the constraint, and apply large deviation method to solve it. Buche and Kushner (2002) allow the true optimal solution to be on the boundary of the constraint and analyze the problem using a reflected diffusion process. The papers in this field first compute the cumulative effect of enforcing the constraints as a stochastic process and then show that this effect is small relative to the stochastic approximation perturbations. For our problem, the constraint in each step is random and depends on past realized demands and decisions, thus the constraint cannot be formulated as a deterministic fixed set as in this stream of literature. In our approach, we update the order-up-to targets less and less frequently, and as a result, the probability of not reaching the order-up-to levels due to inventory carryovers diminishes as time goes by.

In the literature on constrained multi-armed bandit problems, many papers have studied bandits with knapsacks, see e.g., Badanidiyuru et al. (2013) and Ding et al. (2013), among others. The learner incurs a cost when pulling an arm and it has a budget constraint requiring that the total costs of all rounds do not exceed the fixed budget. Both the objective function and constraints are linear. In Badanidiyuru et al. (2013), the algorithm explores as much as possible within a confidence set, which is made narrow enough to eliminate obviously suboptimal alternatives based on historical rewards and costs. In Ding et al. (2013), the algorithm picks an arm with the greatest upper confidence bound (UCB) and achieves a regret of $\mathcal{O}(T^{-1}\log T)$. Various extensions of the knapsack constraints have been made. Agarwal and Devanur (2014) consider a multi-armed bandit problem where the average of all realized outcomes is constrained to fall into a fixed given convex set. The authors extend the UCB method and apply the primal-dual paradigm using Fenchel duality. The result and analyses are then generalized in Agarwal and Devanur (2015) to a contextual bandit setting, in which the authors estimate a weight matrix connecting the observed contexts and outcomes and obtain a regret of $\mathcal{O}(T^{-1/2})$. Applications on online advertising, dynamic pricing with fixed inventory, dynamic procurement, crowdsourcing, etc., have been discussed, see e.g., Badanidiyuru et al. (2013) and Agarwal and Devanur (2014). Besides the knapsack constraint, Guha and Munagala (2009) consider metric switching costs whenever changing an arm and impose an upper budget constraint on that. Denardo et al. (2013) define $K+1$ types of rewards from pulling an arm, taking one type of reward as the objective function to maximize, and imposing a lower bound constraint for each of other types of rewards. They solve the problem by coupling the simplex method with column generation. Our problem can be considered as a constrained online learning problem, but the carry-over inventory constraints cannot be formulated in any of the above frameworks, and therefore need to be tackled using a new approach.

Organization of this paper. The next section formulates the research problem and describes our data-driven learning algorithm for pricing and inventory control decisions. Sections 3 and 4 present our theoretical and numerical results, respectively. Section 5 outlines the main steps and key ideas of the technical proofs. The paper concludes with a few remarks in Section 6. Finally, the detailed mathematical proofs are given in Appendix A, several results on pure inventory management of the non-perishable inventory system are provided in Appendix B, the technical differences from Besbes and Zeevi (2015) are discussed in Appendix C, detailed analyses for additive demand model are provided in Appendix D, and the case of unbounded demand is discussed in Appendix E.

2 Problem Formulation and Learning Algorithm

We consider a periodic review inventory and pricing system in which a firm (e.g., a retailer) sells a non-perishable product over a planning horizon of T periods. At the beginning of each period t, the firm makes a replenishment decision, denoted by the order-up-to level, y_t, and a pricing decision, denoted by p_t, where $y_t \in \mathcal{Y} = [y^l, y^h]$ and $p_t \in \mathcal{P} = [p^l, p^h]$ for some known lower and upper bounds of the inventory level and of the selling price, respectively. We assume $p^h > p^l$ since otherwise, the problem is the pure inventory control problem and learning algorithms have been developed in Huh and Rusmevichientong (2009), Levi et al. (2007), and Levi et al. (2015). During period t and when the selling price is set to p_t, a random demand, denoted by $\tilde{D}_t(p_t)$, is realized and fulfilled as much as possible from on-hand inventory. Any leftover inventory is carried over to the next period, and in case the demand exceeds y_t, the unsatisfied demand is backlogged. The replenishment leadtime is negligible, i.e., an order placed at the beginning of a period can be used to satisfy demand in the same period. Let h and b be the unit holding and backlog costs per period, and the unit purchasing cost is assumed, without loss of generality, to be zero. The objective is to maximize expected total profit.

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

In general, the demand in period t is a function of selling price p_t in that period and some random variable $\tilde{\epsilon}_t$, and it is stochastically decreasing in p_t. The most popular demand models in the literature are the additive demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t) + \tilde{\epsilon}_t$ and multiplicative demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\,\tilde{\epsilon}_t$, where $\tilde{\lambda}(\cdot)$ is a strictly decreasing deterministic function and $\tilde{\epsilon}_t, t = 1, 2, \ldots, T$, are independent and identically distributed random variables with probability density function $f(\cdot)$ and cumulative distribution function $F(\cdot)$. However, the firm knows neither the function $\tilde{\lambda}(p_t)$ nor the distribution of random variable $\tilde{\epsilon}_t$. The firm has to learn from observed demand data that are the realizations of market responses to offered prices, and use that information as a basis for decision making. Suppose $\tilde{\epsilon}_t$ has finite support $[l, u]$, with $l > 0$ for the case of multiplicative demand. Note here that we only allow price p_t to impact $\tilde{\lambda}(\cdot)$ but not the distribution of $\tilde{\epsilon}_t$.

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

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

$$\max_{\substack{(p_t, y_t) \in \mathcal{P} \times \mathcal{Y} \\ y_t \geq x_t}} \sum_{t=1}^{T} G(p_t, y_t) \tag{1}$$

where

$$G(p_t, y_t) := p_t \mathbb{E}[\tilde{D}_t(p_t)] - h\mathbb{E}[y_t - \tilde{D}_t(p_t)]^+ - b\mathbb{E}[\tilde{D}_t(p_t) - y_t]^+.$$

Here the notation " := " represents "defined as", $\mathbb{E}$ stands for the mathematical expectation with respect to random demand $\tilde{D}_t(p_t)$, and $a^+ = \max\{a, 0\}$ for any real number a. However, since in our setting the firm does not have prior knowledge about the demand distribution, the firm is unable to evaluate the objective function of this optimization problem.

We develop a data-driven learning algorithm to compute the inventory control and pricing policy. To save space we shall only present the algorithm and analytical results for the multiplicative demand model. The results and analyses for the additive demand case are analogous, and we only highlight the main differences at the end of this section and present the detailed analyses in Appendix D.

Remark 1. In our exposition we assume that the support of random error $\tilde{\epsilon}_t$ is bounded. This can be relaxed to the case that the moment generating function of $\tilde{\epsilon}_t$ is finite round zero, that is satisfied by many random variables including sub-exponential and sub-Gaussian (refer to Propositions 2.5.2 and 2.7.1 of Vershynin 2017). See Appendix E for details.

Clairvoyant's problem. The clairvoyant has complete information about $\tilde{\lambda}(\cdot)$ and the distribution of $\tilde{\epsilon}_t$. By (1), if (p^*, y^*) is the optimal solution of

$$\max_{p \in \mathcal{P}, y \in \mathcal{Y}} G(p, y) \tag{2}$$

and this solution is reachable in every period, i.e., $x_t \leq y^*$ for all t, then implementing (p^*, y^*) in every period $t = 1, 2, ..., T$ is optimal for (1). We refer to p^* and y^* as the optimal price and optimal order-up-to level (or optimal base-stock level), respectively. It is clear that the reachability condition is satisfied if the system is initially empty, which we assume.

We find it convenient to analyze (2) using a slightly different but equivalent form. Taking logarithm on both sides of $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\tilde{\epsilon}_t$, we obtain

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

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

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

We shall refer to $\lambda(\cdot)$ as the demand-price function and ϵ_t as the random error. Without loss of generality, we assume $\mathbb{E}[\epsilon_t] = \mathbb{E}[\log \tilde{\epsilon}_t] = 0$. If this is not the case, i.e., $\mathbb{E}[\log \tilde{\epsilon}_t] = a \neq 0$, we let $\hat{\lambda}(\cdot) = e^a \tilde{\lambda}(\cdot)$ and $\hat{\epsilon}_t = e^{-a}\tilde{\epsilon}_t$, then $\tilde{D}_t(p_t) = \hat{\lambda}(p_t)\hat{\epsilon}_t$, and $\hat{\lambda}(\cdot)$ and $\hat{\epsilon}_t$ satisfy the desired properties since $\mathbb{E}[\log(e^{-\alpha}\tilde{\epsilon}_t)] = 0$.

For convenience, let ϵ be a generic random variable distributed as ϵ_1. In terms of $\lambda(\cdot)$ and ϵ, the single-period profit function G can be written as

$$G(p, y) = pe^{\lambda(p)}\mathbb{E}\left[e^{\epsilon}\right] - \left\{h\mathbb{E}\left[y - e^{\lambda(p)}e^{\epsilon}\right]^+ + b\mathbb{E}\left[e^{\lambda(p)}e^{\epsilon} - y\right]^+\right\}. \tag{3}$$

Let $Q(p, e^{\lambda(p)}) := \max_{y\in\mathcal{Y}} G(p, y)$, then problem (2) can be re-written as

Problem CI:

$$\max_{p\in\mathcal{P}} Q(p, e^{\lambda(p)}) := \max_{p\in\mathcal{P}} \left\{pe^{\lambda(p)}\mathbb{E}\left[e^{\epsilon}\right] - \min_{y\in\mathcal{Y}} \left\{h\mathbb{E}\left[y - e^{\lambda(p)}e^{\epsilon}\right]^+ + b\mathbb{E}\left[e^{\lambda(p)}e^{\epsilon} - y\right]^+\right\}\right\}. \tag{4}$$

The inner optimization problem (minimization) determines the optimal order-up-to level that minimizes the expected holding and backlog cost for a given price p, and we denote it by $\overline{y}\left(e^{\lambda(p)}\right)$. The outer optimization solves for the optimal price p. Because (p^*, y^*) is the optimal solution for (4), they satisfy $y^* = \overline{y}(e^{\lambda(p^*)})$.

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

Learning algorithm. In the absence of the prior knowledge about the demand process, the firm needs to collect the demand information to estimate $\lambda(p)$ and the distribution of random error ϵ. This implies that the pricing and inventory decisions not only affect the profit in the current period but also the demand to be used for estimation. One important observation is that, the estimations of demand-price function $\lambda(p)$ and the distribution of random error cannot be decoupled. This is because, the firm only observes realized demand, which is a composition of $\lambda(p)$ and ϵ_t. Since $\lambda(p)$ is unknown at any price p, we are not able to obtain accurate samples of the random error ϵ_t just based on the values of $D_t(p_t)$. In Besbes and Zeevi (2015), the firm only needs to estimate $\lambda(p)$ but not the distribution of ϵ_t, hence they do not encounter the issue of inaccurate samples. Besbes and Zeevi (2015) remark that their method can be applied to a more general optimization problem where the objective needs to be a *known* function of $\lambda(p)$ and p. This however does not apply to our case as the dependency of our objective function, $Q(p, e^{\lambda(p)})$ in (4), on $\lambda(p)$ is not known due to the unknown random error distribution, and thus has to be estimated. This will lead to the main difference in analysis between Besbes and Zeevi (2015) and our work, see Appendix C for more discussions.

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

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

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

where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a.

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

Data-Driven Algorithm (DDA)

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

Main procedure. Repeat Steps 1, 2, and 3 for $i = 1, \ldots, n$.

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

$$p_t = \hat{p}_i, \quad t = t_i + 1, \ldots, t_i + I_i,$$
$$p_t = \hat{p}_i + \delta_i, \quad t = t_i + I_i + 1, \ldots, t_i + 2I_i;$$

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

$$y_t = \max \{\hat{y}_{i,1}, x_t\}, \quad t = t_i + 1, \ldots, t_i + I_i,$$
$$y_t = \max \{\hat{y}_{i,2}, x_t\}, \quad t = t_i + I_i + 1, \ldots, t_i + 2I_i.$$

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

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

$$\eta_t = D_t - \frac{1}{I_i} \sum_{t=t_i+1}^{t_i+I_i} D_t, \qquad \text{for } t = t_i+1, \ldots, t_i+I_i,$$

$$\eta_t = D_t - \frac{1}{I_i} \sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t, \qquad \text{for } t = t_i+I_i+1, \ldots, t_i+2I_i.$$

Step 3. Defining and maximizing the proxy profit function, denoted by $G^{DD}_{i+1}(p, y)$. The data-driven optimization problem is

Problem DD:

$$\max_{(p,y)\in\mathcal{P}\times\mathcal{Y}} G^{DD}_{i+1}(p, y) = \max_{p\in\mathcal{P}} Q^{DD}_{i+1}\big(p, e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}\big), \tag{5}$$

where

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

and

$$Q^{DD}_{i+1}\big(p, e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}\big) = \min_{y\in\mathcal{Y}} \ G^{DD}_{i+1}\big(p, y\big).$$

If $\hat{\beta}_{i+1} > 0$, then solve problem DD and set the first pair of price and inventory level to

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

otherwise, set

$$(\hat{p}_{i+1}, \hat{y}_{i+1,1}) = \left(\frac{p^l + p^h}{2}, \frac{y^l + y^h}{2} \right).$$

Set $\hat{p}_{i+1,2} = \hat{p}_{i+1} + \delta_{i+1}$ (in case $\hat{p}_{i+1} + \delta_{i+1} \notin \mathcal{P}$, set $\hat{p}_{i+1,2} = \hat{p}_{i+1} - \delta_{i+1}$), and

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

We have two remarks about DDA.

Remark 2. When $\hat{\beta}_{i+1} > 0$, the objective function in (5) after minimizing over $y \in \mathcal{Y}$ is unimodal in p. To see why this is true, let $d = e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}$ and thus $p = (\hat{\alpha}_{i+1} - \log d)/\hat{\beta}_{i+1}$ with $d \in \mathcal{D} = [d^l, d^h]$, where $d^l = e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p^h}$ and $d^h = e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p^l}$. Then the optimization problem (5) is equivalent to

$$\max_{d\in\mathcal{D}} \left\{ d\frac{\hat{\alpha}_{i+1} - \log d}{\hat{\beta}_{i+1}} \left(\frac{1}{2I_i} \sum_{t=t_1+1}^{t_i+2I_i} e^{\eta_t} \right) - \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \big(h(y - de^{\eta_t})^+ + b(de^{\eta_t} - y)^+ \big) \right\} \right\}. \tag{6}$$

The objective function of this optimization problem is jointly concave in (d, y), hence it is concave in d after minimizing over $y \in \mathcal{Y}$. Thus, it follows from $p = (\hat{\alpha}_{i+1} - \log d)/\hat{\beta}_{i+1}$ is strictly decreasing in d that the objective function in (5) (after minimizing over y) is unimodal in $p \in \mathcal{P}$. Based on the above analysis, to find the optimal p in (5), one can first maximize the concave function to find the optimal d and y in (6), then the optimal p is obtained by $p = (\hat{\alpha}_{i+1} - \log d)/\hat{\beta}_{i+1}$.

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

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

where the last inequality follows from the condition on ρ.

Algorithm overview and its connections with the literature. For $i = 1, 2, \ldots$ in the DDA algorithm, iteration i focuses on stage i that consists of $2I_i$ periods. In Step 1, the algorithm sets the ordering quantity and selling price for each period in stage i derived from the previous iteration. The first I_i periods (from $t_i + 1$ to $t_i + I_i$) try to implement order-up-to $\hat{y}_{i,1}$ policy while the second I_i periods try to implement order-up-to $\hat{y}_{i,2}$ policy. Because starting inventory level may be higher than the order-up-to level, $\hat{y}_{i,1}$ and $\hat{y}_{i,2}$ may not be achieved, and one challenge is to identify the impact of the carry-over inventory constraint on the performance of a learning algorithm. This is the reason that many papers in inventory control with learning only focus on repetitive newsvendor models, as discussed in Section 1, in which there is no such issue on carry-over inventory constraint.

In Step 2, the algorithm applies the realized demand data and least-square method to update the linear approximation, $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p$, of $\lambda(p)$ and computes a centered sample η_t of random error ϵ_t, for $t = t_i + 1, \ldots, t_i + 2I_i$. Note that η_t is not a sample of the random error ϵ_t. This is because $\epsilon_t = D_t(p_t) - \lambda(p_t)$ but $1/I_i \sum_{k=t_1+1}^{t_i+I_i} D_k \neq \lambda(p_t)$. For this reason, the constructed objective function for holding and shortage costs is not a sample average of the newsvendor problem. In the traditional SAA, mathematical expectations are replaced by true sample averages, see e.g., Kleywegt et al. (2001), Levi et al. (2007) and Levi et al. (2015). When only biased samples are available, techniques from statistics such as jackknife resampling can be applied to reduce bias for SAA (Wu 1986). In our work, samples of ϵ_t cannot be observed, however,

$$\eta_t = D_t(p_t) - \frac{1}{I_i}\sum_{k=t_1+1}^{t_i+I_i} D_k = \epsilon_t - \frac{1}{I_i}\sum_{k=t_1+1}^{t_i+I_i} \epsilon_k$$

can be obtained. Since $\mathbb{E}[\epsilon_k] = 0$, $1/I_i \sum_{k=t_1+1}^{t_i+I_i} \epsilon_k$ converges to 0 in probability as I_i grows, and one would expect $\eta_t \to \epsilon_t$ in probability as t grows. Thus, we use η_t in place of ϵ_t in computing proxy objectives. Since these samples are obtained from the original i.i.d. samples after subtracting the sample average, we call η_t *centered samples*, and $\{\eta_t, t = t_i + 1, \ldots, t_i + 2I_i\}$ are dependent.

In Step 3, a data-driven optimization problem is constructed. As discussed in Remark 2, when

$\hat{\beta}_{i+1} > 0$, the algorithm solves an optimization problem of a jointly concave function. Technical analyses in Appendix A show that the probability for $\hat{\beta}_{i+1} > 0$ converges to 1 as i grows.

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

Performance metrics. To measure the performance of a policy, we use two metrics proposed in Besbes and Zeevi (2015): *consistency* and *regret*. An admissible policy $\pi = ((p_t, y_t), t \geq 1)$ is said to be consistent if $(p_t, y_t) \to (p^*, y^*)$ in probability as $t \to \infty$ (note that y_t in our policy is not the order-up-to level in period t, but the realized inventory level after replenishment decision in period t). The average (per-period) regret of policy π, denoted by $R(\pi, T)$, is defined as the average profit loss per period. Since the system is initially empty, the clairvoyant optimal policy (p^*, y^*) is reached in each and every period, thus the clairvoyant average profit per period is $G(p^*, y^*)$. Under our policy, the inventory level (after replenishment) in period t is y_t, and the expected profit in period t with selling price p_t is $G(p_t, y_t)$. Therefore, the regret for policy π is given by

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

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

In the next section, we will show that the DDA policy is consistent, and we will also characterize the rate at which the regret converges to zero. For convenience, we follow Besbes and Zeevi (2015) to introduce the following notation

$$\breve{\alpha}(z) = \lambda(z) - \lambda'(z)z, \qquad \breve{\beta}(z) = -\lambda'(z), \qquad z \in \mathcal{P}. \tag{7}$$

Note here that $\breve{\alpha}(z) - \breve{\beta}(z)p$ is the tangent line of $\lambda(p)$ at point z, and $\breve{\beta}(z) > 0$.

3 Main Results

We will show that, under some technical conditions, the policy prescribed by DDA converges to the clairvoyant optimal policy, and we will also present the regret rate of the algorithm. To present the technical conditions, we need some preparations.

To compare the DDA policy with the clairvoyant optimal policy, i.e., the optimal solutions of problem DD (5) and problem CI (4), we note that these two objective functions have significant differences: In problem CI, both $\lambda(p)$ and the distribution of ϵ are known, but in problem DD, $\lambda(p)$ is approximated by a linear function and distribution of ϵ is estimated using centered samples instead of true samples. Therefore, to analyze DDA our approach is to introduce several "intermediate" bridging problems, and in each step we compare two "adjacent" problems that differ along only one dimension.

First, for parameters α and $\beta > 0$, we introduce bridging problem B1 defined by

Bridging Problem B1:

$$\max_{p\in\mathcal{P}} \overline{Q}(p, e^{\alpha-\beta p}) := \max_{p\in\mathcal{P}} \left\{ pe^{\alpha-\beta p}\mathbb{E}\big[e^{\epsilon}\big] - \min_{y\in\mathcal{Y}} \left\{ h\mathbb{E}\Big[y - e^{\alpha-\beta p+\epsilon}\Big]^+ + b\mathbb{E}\Big[e^{\alpha-\beta p+\epsilon} - y\Big]^+ \right\}\right\}. \tag{8}$$

It is easy to see that, the only difference between problem B1 and problem CI in (4) is that, in problem B1 we replace the demand-price function in CI by an affine function $\alpha-\beta p$. Let $\overline{p}\big(\alpha, \beta\big)$ denote the optimal price for problem B1, and for given $p \in \mathcal{P}$, we let $\overline{y}(e^{\alpha-\beta p})$ denote its optimal order-up-to level, which is the optimal solution for the inner minimization problem in (8).

The second bridging problem, B2, is defined for each iteration i of the DDA algorithm, and for any α and $\beta > 0$, it is given by

Bridging Problem B2:

$$\begin{aligned}\max_{p\in\mathcal{P}} \tilde{Q}_{i+1}(p, e^{\alpha-\beta p}) := \max_{p\in\mathcal{P}} \Bigg\{ & pe^{\alpha-\beta p}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right) \\ & - \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \Big(h\big(y - e^{\alpha-\beta p+\epsilon_t}\big)^+ + b\big(e^{\alpha-\beta p+\epsilon_t} - y\big)^+\Big)\right\}\Bigg\}.\end{aligned} \tag{9}$$

Compared with problem B1, it is seen that B2 is obtained from B1 after replacing the expectations in B1 by sample averages, hence B2 is the sample average approximation (SAA) of problem B1. Here ϵ_t, $t = t_i + 1, \dots, t_i + 2I_i$, represent the realizations of random errors during stage i. Let $\tilde{p}_{i+1}(\alpha, \beta)$ denote the optimal price and $\tilde{y}_{i+1}(e^{\alpha-\beta p})$ the optimal order-up-to level for problem B2, which is the optimal solution for the inner minimization problem in (9).

The third bridging problem B3 is a variation of problem B2, which replaces the true random error ϵ_t by a biased error sample ζ_t, $t = t_i + 1, \dots, t_i + 2I_i$. That is, for

$$\zeta_{t=t_i+1}^{t_1+I_i} = \big(\zeta_{t_i+1}, \dots, \zeta_{t_i+I_i}\big), \quad \zeta_{t=t_i+I_i+1}^{t_1+2I_i} = \big(\zeta_{t_i+I_i+1}, \dots, \zeta_{t_i+2I_i}\big),$$

and parameters α and $\beta > 0$, we define the third bridging problem B3 as

Bridging Problem B3:

$$\begin{aligned}&\max_{p\in\mathcal{P}} \breve{Q}_{i+1}\Big(p, e^{\alpha-\beta p}, \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\Big) \\ := \;& \max_{p\in\mathcal{P}} \left\{ pe^{\alpha-\beta p}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\zeta_t}\right) - \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \Big(h\big(y - e^{\alpha-\beta p+\zeta_t}\big)^+ + b\big(e^{\alpha-\beta p+\zeta_t} - y\big)^+\Big)\right\}\right\}.\end{aligned}$$

Note that when $(\alpha, \beta) = (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$, and $\zeta_t = \eta_t$ for $t = t_1 + 1, \ldots, t_i + 2I_i$, problem B3 reduces to problem DD (5) in the DDA algoroithm. Thus, problem B3 serves as a bridge between problem B2 and problem DD. We denote the optimal price of problem B3 by $\breve{p}_{i+1}\big((\alpha, \beta), \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big)$ and its optimal order-up-to level, for given price p, by $\breve{y}_{i+1}\big(e^{\alpha-\beta p}, \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big)$.

Based on their definitions, problem CI, bridging problems B1-B3, and problem DD, require less and less information about the demand process. Problem CI has complete information about both $\lambda(\cdot)$ and the distribution of ϵ; problem B1 does not know $\lambda(\cdot)$ but knows the distribution of ϵ; problem B2 does not know either $\lambda(\cdot)$ or the distribution of ϵ but has access to true samples of ϵ; problems B3 and DD do not have true samples and have to use biased samples. We prove convergence for each pair of adjacent problems, and eventually establish convergence of problem DD to problem CI.

Recall that the demand in period t is $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\tilde{\epsilon}_t$. As $\tilde{\lambda}(p)$ is strictly decreasing, it has a strictly decreasing inverse function, $\tilde{\lambda}^{-1}(d)$, defined on $d \in [d^l, d^h] = \big[\tilde{\lambda}(p^h), \tilde{\lambda}(p^l)\big]$. We make the following assumptions.

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

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

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

(iii) $G(p, \bar{y}(e^{\lambda(p)}))$ in (3) has bounded second order derivatives with respect to $p \in \mathcal{P}$.

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

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

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

(vii) The functions $\breve{p}_{i+1}\big((\cdot,\cdot),\cdot\big)$ and $\breve{y}_{i+1}\big((\cdot,\cdot),\cdot\big)$ for problem B3 satisfy the following Lipschitz condition: there exists some constant $K_1 > 0$ such that

$$\Big|\breve{p}_{i+1}\big((\alpha, \beta), \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big) - \breve{p}_{i+1}\big((\alpha', \beta'), \zeta_{t=t_i+1}^{t_1+I_i} + m_1\mathbf{1}_{I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i} + m_2\mathbf{1}_{I_i}\big)\Big| \le K_1\Big(|\alpha - \alpha'| + |\beta - \beta'| + |m_1| + |m_2|\Big), \tag{10}$$

and

$$\Big|\breve{y}_{i+1}\big(e^{\alpha-\beta p}, \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big) - \breve{y}_{i+1}\big(e^{\alpha'-\beta' p}, \zeta_{t=t_i+1}^{t_1+I_i} + m_1\mathbf{1}_{I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i} + m_2\mathbf{1}_{I_i}\big)\Big| \le K_1\Big(|\alpha - \alpha'| + |\beta - \beta'| + |m_1| + |m_2|\Big), \tag{11}$$

where $\mathbf{1}_{I_i}$ is the I_i-dimentional vector with all entries being 1.

(viii) The feasible regions for price, $\mathcal{P}$, and for order-up-to level, $\mathcal{Y}$, are large enough so that the optimal solutions p^* and $\overline{y}(e^{\lambda(p)})$ for problem CI over $\mathbb{R}_+$ for given $p \in \mathcal{P}$ fall into $\mathcal{P}$ and $\mathcal{Y}$, respectively; and for given $q \in \mathcal{P}$, the optimal solutions $\overline{p}\big(\breve{\alpha}(q), \breve{\beta}(q)\big)$ and $\overline{y}\big(e^{\breve{\alpha}(q)-\breve{\beta}(q)p}\big)$ for given $p \in \mathcal{P}$ for problem B1 fall into $\mathcal{P}$ and $\mathcal{Y}$, respectively.

Assumption 1(i) is a standard assumption in the literature on joint optimization of pricing and inventory control (see e.g., Federgruen and Heching 1999, and Chen and Simchi-Levi 2004), and it guarantees that the objective function in problem CI after minimizing over y is unimodal in p. Assumption 1(ii) imposes some shape restriction on the underlying demand function, and similar assumption has been made in Besbes and Zeevi (2015) in their Assumption 1(ii). Technically, this condition assures that the prices converge to a fixed point through a contraction mapping. Assumption 1(iii), (v) and (vii) are also similarly imposed by Besbes and Zeevi (2015) in their Appendix. Assumption 1(iv) assumes the decision maker only considers prices that yield a positive mean demand, and this is reasonable because the company can reorder in every period. Assumption 1(vi) is satisfied by almost all common distributions used in the literature. Assumption 1(viii) requires $\mathcal{P}$ and $\mathcal{Y}$ to be sufficiently large. Note that both problem CI and problem B1 depend only on primitive data and do not depend on random samples, hence these are mild assumptions.

Example 1. The following are some examples that satisfy Assumption 1 with appropriate choices of p^l and p^h, which includes those we used in our numerical studies.

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

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

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

We then present the theoretical results for the DDA algorithm. Recall that p^* and y^* are the optimal pricing and inventory decisions for the case with complete information, i.e., the maximizer of (4).

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

Theorem 1 states that, despite the lack of true samples of random demand error, both pricing and ordering decisions from the DDA algorithm converge to the clairvoyant optimal solution (p^*, y^*) in probability. Note that the convergence of actual inventory levels $y_t \to y^*$ is stronger than the convergence of order-up-to targets $\hat{y}_{i,1} \to y^*$ and $\hat{y}_{i,2} \to y^*$. This is because, the actual inventory level in a period can "overshoot" the order-up-to target when the latter is lower than the carry-over inventory. Theorem 1 shows that, despite these overshoots, the actual inventory levels y_t converge to the clairvoyant optimal solution in probability. We note that results on convergence of inventory decisions in the literature only consider convergence of order-up-to targets, and this result is the first on the convergence of actual inventory levels.

Our next result shows that DDA is asymptotically optimal in terms of maximizing the expected profit.

Theorem 2 (Regret Convergence Rate) *Under Assumption 1, the DDA policy is asymptotically optimal. More specifically, there exists some constant $K_2 > 0$ such that*

$$R(DDA, T) = G(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t, y_t)\right] \leq K_2 T^{-\frac{1}{2}}.$$

Theorem 2 shows that as the length of planning horizon, T, grows, the regret of DDA policy vanishes at rate $O\big(T^{-1/2}\big)$, hence DDA policy is asymptotically optimal as T goes to infinity. Thus, even though the firm does not have prior knowledge about the demand-price function or the error distribution, the performance of the data-driven algorithm approaches the theoretical maximum as the planning horizon becomes long.

In Keskin and Zeevi (2014), the authors consider a *parametric* data-driven pricing problem (with no inventory decision) where the demand error term is additive and the average demand function is linear, and they prove that no learning algorithm can achieve a convergence rate better than $O(T^{-1/2})$. Our problem involves both pricing and inventory decisions, and the firm does not have prior knowledge about the parametric form of the underlying demand-price function or the distribution of the random error, and our algorithm achieves $O\big(T^{-1/2}\big)$, which is the theoretical lower bound.

One important special case of our model is the dynamic pure pricing optimization problem. Letting $h = b = 0$, the inventory decision will no longer be relevant in our optimization and DDA solves the pure pricing problem. In Besbes and Zeevi (2015) for the pure pricing problem, the authors prove a regret $O\big(T^{-1/2}(\log T)^2\big)$. Theorem 2 implies that, through a refined analysis of the aggregated regret, we can obtain a slightly stronger result on the convergence rate. Since it is known that $\Omega\big(T^{-1/2}\big)$ is the lower bound for the regret of this class of problems, Theorem 2 closes the gap left in Besbes and Zeevi (2015) for learning in the pure pricing optimization problem. An important implication is that, the linear model of Besbes and Zeevi (2015) achieves the best possible regret rate. Hence, there is no loss in using Besbes and Zeevi's linear model in demand learning for revenue optimization as well as in joint pricing and inventory optimization problems.

As mentioned earlier, Theorems 1 and 2 continue to hold for the additive demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t) + \tilde{\epsilon}_t$ with some moderate modifications to the algorithm as well as Assumption 1. Examples that satisfy these conditions for the additive demand model include (a) linear with $\lambda(p) = k - mp$, $m > 0$, (b) exponential with $\lambda(p) = e^{k-mp}, m > 0$, and (c) logit with $\lambda(p) = \frac{e^{k-mp}}{1+e^{k-mp}}, m > 0, e^{k-mp} < 3$ for all $p \in \mathcal{P}$.

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

Remark 4. Condition for Assumption 1(ii) can be relaxed to $\left|\partial\overline{p}\big(\breve{\alpha}(z),\breve{\beta}(z)\big)/\partial z\right| < 1$ for $z \in \mathcal{P}$. This condition reduces to Assumption 1(ii) if the optimal solutions for problem CI and problem B1 satisfy the feasibility conditions assumed in (viii).

4 Numerical Study

In this section, we present the results for a numerical study on the performance of the DDA algorithm. For each problem instance, we run 500 rounds, and for each round, we compute the percentage of profit loss defined by

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

Then we compute the average profit loss over the 500 rounds. The results are reported in Tables 1 - 5. In all the experiments, we set $p^l = 0.5, p^h = 4, y^l = 0, y^h = 10, b = 1, h = 0.1$, and initial price $\hat{p}_1 = 1$, initial inventory order-up-to levels $\hat{y}_{11} = 1,\ \hat{y}_{12} = 0.3$ (the parameters are selected so that the largest demand is not too much higher than the upper bound of order-up-to level). In Tables 1 and 2, DDA parameters $I_0 = 1,\ v = 2$, and $\rho = 0.75$ for Table 1 and $\rho = 1.25$ for Table 2. DDA parameters for Tables 3 - 5 will be explained later. Some variations of the input parameters are also tested, and the numerical results are comparable.

In Table 1, we consider two demand-price functions for $\tilde{\lambda}(p)$:

1) exponential e^{w-mp}: in each round w is uniformly drawn from $[0.1, 1.7]$ and m is uniformly drawn from $[0.3, 2]$,
2) logit $\frac{e^{w-mp}}{1+e^{w-mp}}$: in each round w is uniformly drawn from $[-0.3, 1]$ and m is uniformly drawn from $[2, 2.5]$.

And for the error distribution of $\tilde{\epsilon}_t$, we consider truncated normal distributions with mean 1 and four different variances as well as a uniform distribution:

i) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.1,
ii) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.25,
iii) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.35,
iv) truncated normal on $[0.5, 1.5]$ with mean 1 and variance 0.5,
v) uniform on $[0.5, 1.5]$.

Here truncated normal on $[a, b]$ with mean μ and variance σ^2 is defined as random variable X conditioning on $X \in [a, b]$, where X is normally distributed with mean μ and variance σ^2. All the conditions in Assumption 1 are satisfied by these distributions.

Table 1: Results for Truncated Distributions

	T=100		T=500		T=1000		T=5000		T=10000	
	Exp	Logit	Exp	Logit	Exp	Logit	Exp	Logit	Exp	Logit
Normal $\sigma = 0.1$	6.31	8.34	2.59	3.67	1.84	2.67	1.06	1.60	0.76	1.15
Normal $\sigma = 0.25$	9.74	9.86	4.58	4.51	3.39	3.30	1.78	1.87	1.27	1.35
Normal $\sigma = 0.35$	10.83	10.49	5.18	4.85	3.76	3.55	2.03	2.00	1.51	1.43
Normal $\sigma = 0.5$	12.15	11.30	6.12	5.24	4.44	3.79	2.41	2.11	1.76	1.51
Uniform	11.14	14.68	5.60	7.03	4.08	5.25	2.52	3.62	1.89	2.75

As seen from Table 1, when the planning horizon is $T = 100$, the worst profit loss (compared with the clairvoyant optimal solution) is 14.68%, and for most problem instances it is below 10%. When $T = 500$, most of the results are around or below 5%. Given that the firm does not know the functional form of the demand-price function $\lambda(\cdot)$ or the error distribution of ϵ, the DDA algorithm performs satisfactorily well. As predicated by Theorem 2, the performance gets better when T becomes larger. Also, it is seen from Table 1 that the overall performance of DDA is better when the variance of the demand is smaller, which is intuitive as it is easier to learn the demand when it has smaller variance.

In Table 2, we consider the same exponential and logit functions as in Table 1, plus lognormal and the Weibull distributions, with the following parameters:

i) lognormal with $\mu = 1$ and $\sigma = 0.1$,
ii) lognormal with $\mu = 1$ and $\sigma = 0.25$,
iii) lognormal with $\mu = 1$ and $\sigma = 0.35$,
iv) lognormal with $\mu = 1$ and $\sigma = 0.5$,
v) Weibull with scale parameter 1 and shape parameter $k = 0.5$,
vi) Weibull with scale parameter 1 and shape parameter $k = 1$,
vii) Weibull with scale parameter 1 and shape parameter $k = 1.5$,
viii) Weibull with scale parameter 1 and shape parameter $k = 5$.

As seen from Table 2, for lognormal and Weibull distributions with unbounded support, their regrets converge to 0 but slower than those in Table 1. One observation from Table 2 is that, the heavier the tail of its distribution (larger σ for lognormal, smaller k for Weibull), the worse the performance of the learning algorithm.

Next, we test on the effect of DDA parameter ρ on the performance of the algorithm, and summarize the results in Table 3. In this experiment, we use the exponential demand-price function as in Tables

Table 2: Results for Unbounded Distributions

	T=100		T=500		T=1000		T=5000		T=10000	
	Exp	Logit	Exp	Logit	Exp	Logit	Exp	Logit	Exp	Logit
Lognormal $\sigma = 0.1$	9.95	13.21	4.31	7.08	3.18	5.52	2.04	3.72	1.51	2.76
Lognormal $\sigma = 0.25$	12.06	14.31	5.32	7.56	3.90	5.84	2.42	3.89	1.76	2.87
Lognormal $\sigma = 0.35$	13.55	15.57	6.39	8.05	4.71	6.17	2.79	3.99	2.03	3.24
Lognormal $\sigma = 0.5$	16.66	18.48	8.19	9.41	6.15	7.05	3.49	4.40	2.56	3.24
Weibull $k = 0.5$	77.52	296	50.87	138	42.37	88.27	23.50	28.48	18.43	17.61
Weibull $k = 1$	58.85	87.46	31.34	34.49	22.73	21.76	11.48	8.93	8.42	6.07
Weibull $k = 1.5$	29.30	32.57	16.24	14.37	12.63	10.34	6.73	5.64	4.96	4.13
Weibull $k = 5$	10.16	16.10	4.99	7.91	3.82	6.01	2.37	3.88	1.72	2.87

1 and 2. For the error distribution, we consider the four normal distributions in Table 1, and $I_0 = 1$, $v = 2$. From Table 3, we see that when $\sigma = 0.1, 0.25, 0.35, 0.5$, the best selections of ρ are $0.5, 0.75, 0.75, 1$ (especially for large T), which is increasing in the variance parameter σ. We offer the following intuition: If the variance of demand is higher, it becomes harder to learn the demand distribution, thus more active learning is necessary. A larger ρ results in a larger δ_i, which means that the exploration price $\hat{p}_i + \delta_i$ deviates by a larger amount from the profit maximizing price $\hat{p}_i$, translating into more aggressive learning and exploration.

In Tables 4 and 5, we present the results on the effects of the DDA parameters v and I_0, respectively, on the performance of the algorithm. Again, we consider the same exponential demand-price function as in Tables 1-3, as well as normal and uniform distributions, and $\rho = 0.75$. In Table 4, we set $I_0 = 1$, and in Table 5, $v = 2.75$.

The effect of v and I_0 on the regret of DDA is not clear. Note that the lengths of the learning stages I_i's are increasing while deviation quantities δ_i's are decreasing in v and I_0. Intuitively, if demands have large variance, then one would prefer large I_i (so that DDA will not need to update very frequently) and large δ_i (so that the exploration is sufficiently aggressive), and vice versa. But to increase I_i, one needs to raise v and I_0, and as a result δ_i will decrease. Therefore, the effects of v and I_0 are mixed and depend on other problem characteristics. This can be observed from the numerical results in Tables 4

Table 3: The Effect of ρ

		T=100	T=500	T=1000	T=5000	T=10000
Normal $\sigma = 0.1$	$\rho = 0.5$	6.37	2.51	1.82	0.89	0.62
	$\rho = 0.75$	6.18	2.53	1.83	1.02	0.74
	$\rho = 1$	6.98	3.12	2.31	1.45	1.06
	$\rho = 1.25$	8.19	3.87	2.93	1.94	1.44
Normal $\sigma = 0.25$	$\rho = 0.5$	11.91	6.02	4.56	2.35	1.70
	$\rho = 0.75$	9.60	4.73	3.50	1.84	1.33
	$\rho = 1$	9.26	4.33	3.24	1.89	1.38
	$\rho = 1.25$	9.97	4.95	3.76	2.38	1.76
Normal $\sigma = 0.35$	$\rho = 0.5$	14.70	7.50	5.74	3.03	2.21
	$\rho = 0.75$	10.55	5.28	3.82	2.05	1.47
	$\rho = 1$	10.18	4.99	3.68	2.01	1.48
	$\rho = 1.25$	10.31	5.12	3.85	2.40	1.75
Normal $\sigma = 0.5$	$\rho = 0.5$	15.92	8.21	6.18	3.30	2.42
	$\rho = 0.75$	11.18	5.82	4.40	2.27	1.66
	$\rho = 1$	10.91	5.17	3.79	2.19	1.63
	$\rho = 1.25$	10.96	5.40	4.03	2.46	1.81

and 5. Nevertheless, we note that the effect of v or I_0 on the performance of the DDA is rather robust, with most of the differences in regret between two values of either v or I_0 to be around or below 1% for the same problem instance and same planning horizon.

5 Sketches of Proofs

In this section we outline the proofs of our main results. In Sections 5.1 and 5.2, we discuss the major steps in proving the convergence of pricing and inventory decisions, respectively, of Theorem 1, and in Section 5.3, we present the main ideas in proving the convergence rate of regret of Theorem 2. Note that in our analysis of joint pricing and inventory decisions, we first find the optimal inventory level for each given price p, and then substitute that in the objective function to search for the optimal pricing decision. The structure of the multiplicative demand simplifies our analysis.

Table 4: The Effect of v

		T=100	T=500	T=1000	T=5000	T=10000
Normal $\sigma = 0.1$	v=2.5	5.39	2.57	2.02	0.71	0.56
	v=2.75	5.93	3.06	1.65	0.89	0.48
	v=3	7.42	2.48	2.03	0.74	0.61
Normal $\sigma = 0.25$	v=2.5	8.72	4.45	3.42	1.42	1.10
	v=2.75	8.93	5.04	3.14	1.65	1.02
	v=3	10.4	4.37	3.51	1.49	1.14
Normal $\sigma = 0.35$	v=2.5	10.97	5.39	4.08	1.69	1.23
	v=2.75	9.81	5.48	3.38	1.74	1.07
	v=3	10.77	4.59	3.73	1.64	1.28
Normal $\sigma = 0.5$	v=2.5	11.43	5.83	4.52	2.01	1.56
	v=2.75	10.93	5.84	3.71	2.02	1.35
	v=3	12.29	5.36	4.20	1.91	1.46
Uniform	v=2.5	10.08	5.58	4.65	1.82	1.47
	v=2.75	10.65	6.73	3.79	2.34	1.28
	v=3	11.49	5.11	4.36	1.75	1.55

5.1 Proof of Theorem 1: Convergence of Pricing Decision

To prove convergence of the pricing decision, we proceed as follows:

$$\mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \le \mathbb{E}\Bigg[\Bigg(\underbrace{\Big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|}_{\substack{\text{Comparison of problems CI and B1} \\ \text{Proposition 1}}} + \underbrace{\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\Big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\Big)\Big|}_{\substack{\text{Comparison of problems B1 and B2} \\ \text{Proposition 2}}} + \underbrace{\Big|\tilde{p}_{i+1}\Big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\Big) - \hat{p}_{i+1}\Big|}_{\substack{\text{Comparison of problems B2 and DD} \\ \text{Proposition 3}}} \Bigg)^2\Bigg], \tag{12}$$

where $\overline{p}(\cdot,\cdot)$ and $\tilde{p}_{i+1}(\cdot,\cdot)$ are the optimal solutions of the bridging problems B1 and B2, respectively.

As mentioned earlier, the difference in the first term in (12) is due to linear approximation of $\lambda(p)$, and that in the second term stems from the SAA. For the third term, comparing (9) with (5), it is noted that problems B2 and DD differ in the coefficients of the linear function as well as the arithmetic averages. More specifically, in B2 the real random error samples ϵ_t are used, while in problem DD, centered error samples η_t are used in place of ϵ_t. Note that the optimal prices for problems CI and B1, p^* and $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$, are deterministic, but the optimal solutions of problems B2 and DD, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\hat{p}_{i+1}$, are random as they depend on demand realizations in stage i. Thus, we shall develop probabilistic bounds for the second and third terms on the right hand side of

Table 5: The Effect of I_0

		T=100	T=500	T=1000	T=5000	T=10000
Normal $\sigma = 0.1$	I_0=0.5	6.10	2.30	1.91	0.70	0.58
	I_0=1	5.93	3.06	1.65	0.89	0.48
	I_0=1.5	8.82	2.87	2.23	0.78	0.61
	I_0=2	8.54	3.38	2.57	0.88	0.71
Normal $\sigma = 0.25$	I_0=0.5	9.46	3.95	3.13	1.40	1.05
	I_0=1	8.93	5.04	3.14	1.65	1.02
	I_0=1.5	11.80	4.56	3.44	1.35	1.03
	I_0=2	11.18	4.88	3.77	1.45	1.09
Normal $\sigma = 0.35$	I_0=0.5	10.98	4.73	3.73	1.62	1.24
	I_0=1	9.81	5.48	3.38	1.74	1.07
	I_0=1.5	12.92	5.42	4.19	1.79	1.35
	I_0=2	11.47	5.44	4.25	1.72	1.31
Normal $\sigma = 0.5$	I_0=0.5	10.96	5.01	3.92	1.70	1.30
	I_0=1	10.93	5.84	3.71	2.02	1.35
	I_0=1.5	13.32	5.38	4.06	1.67	1.28
	I_0=2	12.20	5.82	4.58	1.96	1.53
Uniform	I_0=0.5	17.38	6.00	4.73	1.81	1.46
	I_0=1	10.65	6.73	3.79	2.34	1.28
	I_0=1.5	12.34	5.16	4.36	1.74	1.37
	I_0=2	9.84	5.82	5.07	2.04	1.77

(12). The three terms are bounded in Propositions 1 - 3 respectively, and before that we need to first establish Lemma 1 regarding the shape of objective functions in problems B1, B2, and B3.

Lemma 1. *When $\beta > 0$, $\overline{Q}(p, e^{\alpha-\beta p})$, $\tilde{Q}_{i+1}(p, e^{\alpha-\beta p})$, and $\breve{Q}_{i+1}\big(p, e^{\alpha-\beta p}, \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_i+2I_i}\big)$ are all unimodal in $p \in \mathcal{P}$.*

This result shows that all the three objective functions are well-behaved, and thus their optimal solutions $\overline{p}(\cdot,\cdot)$, $\tilde{p}_{i+1}(\cdot,\cdot)$, and $\breve{p}_{i+1}\big((\cdot,\cdot),\cdot,\cdot\big)$ are well-defined. Unimodality is very important for our subsequent analyses, because we establish price convergence by first establishing convergence of the objective functions and parameters. Without unimodality, the latter cannot be conveniently translated into convergence of solutions (pricing decisions).

The first term on the right hand side of (12) measures the difference between the optimal solution of problem CI, p^*, and the optimal solution of problem B1, $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$. The key difference between them is that in B1, we use a linear approximation to replace the original demand-price function $\lambda(p)$. This term is bounded in Proposition 1 below.

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

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

The above result can be proved using first order condition, following similar arguments as that used in proving Lemma 1 of Besbes and Zeevi (2015). We point out that this is the only result in the paper that directly extends from Besbes and Zeevi (2015); all other results require new ideas and analyses.

SAA problem with i.i.d. error samples. The second term in (12) measures the difference in solutions between bridging problems B1 and B2. Different from B1, in problem B2 the distribution of ϵ in the objective function is unknown, hence the expectations are replaced by their sample averages, giving rise to the SAA problem. This term is upper bounded in Proposition 2 below.

Proposition 2. *For any $p \in \mathcal{P}$ and any $\xi > 0$,*

$$\mathbb{P}\left\{ \left| \overline{p}\left(\breve{\alpha}(p), \breve{\beta}(p)\right) - \tilde{p}_{i+1}\left(\breve{\alpha}(p), \breve{\beta}(p)\right) \right| \geq K_3 \xi^{\frac{1}{2}} \right\} \leq 6 e^{-4 I_i \xi^2}$$

for some positive constant K_3.

Proposition 2 presents a useful result that bounds the probability for the optimal solution of problem B2 to be away from that of problem B1. Since I_i tends to infinity as t goes to infinity, this shows that the probability for the two solutions, $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$, to be significantly different converges to zero exponentially fast when i increases.

In problem B1, the cumulative distribution function $F(\cdot)$ of $\tilde{\epsilon}$ is known. In problem B2, on the other hand, $F(\cdot)$ is not known and is estimated by an empirical distribution $\hat{F}(\cdot)$ using i.i.d. samples of ϵ_t as follows,

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

where $\mathbb{1}\{A\}$ is the indicator function taking value 1 if "A" is true and 0 otherwise. Therefore, the key to establish Proposition 2 is to bound the estimation error of this SAA method. For $\theta > 0$, we define $\hat{F}(\cdot)$ to be a θ-estimate of $F(\cdot)$, if

$$\left| \hat{F}\Big(F^{-1}\Big(\frac{b}{b+h}\Big)\Big) - \frac{b}{b+h} \right| \leq \theta. \tag{13}$$

We then prove that the probability for $\hat{F}(\cdot)$ to be a θ-estimate of $F(\cdot)$ is at least $1 - 2e^{-2I_i\theta^2}$. As it is seen, the longer the I_i, the larger probability for $\hat{F}(\cdot)$ to be a θ-estimate, which means the better the estimation.

The θ-estimate is defined based on the critical ratio $b/(b+h)$ because it helps to define the quantile solution of the newsvendor problem, which needs to be solved first in the inner minimization before solving the pricing problem (outer maximization). We prove that, if the empirical distribution $\hat{F}(\cdot)$ is a

θ-estimate, then the optimal newsvendor costs (holding and backlog costs) of problems B1 and B2 are different by at most $K_4\theta$, for some constant $K_4 > 0$. Based on these analyses, the objective functions of problems B1 and B2 can be bounded as

$$\mathbb{P}\Big\{ \max_{p\in\mathcal{P}} \left|\overline{Q}\left(p, e^{\breve{\alpha}(p)-\breve{\beta}(p)p}\right) - \tilde{Q}_{i+1}\left(p, e^{\breve{\alpha}(p)-\breve{\beta}(p)p}\right)\right| < K_5\theta \Big\} \geq 1 - 5e^{-2I_i\theta^2}, \tag{14}$$

for some constant $K_5 > 0$. (14) helps to bound the difference in pricing decisions in Proposition 2.

Bounding the impact of centered samples. Next we develop an upper bound for the third term on the right hand side of (12), which represents the discrepancy in pricing decisions between having true and i.i.d. samples ϵ_t in B2 and centered samples η_t in problem DD. There is another difference between these two problems: in B2 the tangent line $\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p$ is used, while in problem DD it is $\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p$. Proposition 3 bounds the difference between the optimal solution for problem B2, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$, and the optimal solution for problem DD, $\hat{p}_{i+1}$.

Proposition 3. *When i is large enough, there exists a positive constant K_6 such that*

$$\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|^2\right] \leq K_6 I_i^{-\frac{1}{2}}.$$

Problem B3, which serves as a bridging problem to link problems B2 and DD, plays a central role in proving Proposition 3. To see this, when $\hat{\beta}_{i+1} > 0$ (which can be proved to hold with a high probability), problem B3 will reduce to problem DD if $(\alpha, \beta) = (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})$ and $\zeta_t = \eta_t$ for $t = t_i + 1, \ldots, t_i + 2I_i$. Therefore, one has

$$\hat{p}_{i+1} = \breve{p}_{i+1}\big((\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \eta_{t=t_i+1}^{t_i+I_i}, \eta_{t=t_i+I_i+1}^{t_i+2I_i}\big).$$

On the other hand, when $(\alpha, \beta) = \big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\zeta_t = \epsilon_t$ for $t = t_i + 1, \ldots, t_i + 2I_i$, problem B3 reduces to problem B2, and we have

$$\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) = \breve{p}_{i+1}\big(\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big), \epsilon_{t=t_i+1}^{t_i+I_i}, \epsilon_{t=t_i+I_i+1}^{t_i+2I_i}\big).$$

It can be proved that parameters $\big((\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \eta_{t=t_i+1}^{t_i+I_i}, \eta_{t=t_i+I_i+1}^{t_i+2I_i}\big)$ will be very close to $\big(\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big), \epsilon_{t=t_i+1}^{t_i+I_i}, \epsilon_{t=t_i+I_i+1}^{t_i+2I_i}\big)$ with a high probability as i grows, and by the Lipschitz condition in Assumption 1 (vii), convergence in parameters leads to convergence in the pricing decisions. Proposition 3 can thus be proved based on the linkage of problem B3.

Combining the results from Propositions 1, 2, and 3, and summarizing them in (12), after some technical calculations we obtain

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \leq K_7 I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty, \tag{15}$$

for some positive constant K_7, and

$$\mathbb{E}\left[(\hat{p}_{i+1} + \delta_{i+1} - p^*)^2\right] \leq 2\mathbb{E}\left[(\hat{p}_{i+1} - p^*)^2\right] + 2\delta_{i+1}^2 \leq K_8 I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty,$$

for some positive constant K_8. The details can be found in Appendix A. Convergence of pricing decision can thus be proved.

5.2 Proof of Theorem 1: Convergence of Inventory Decision

We first prove convergence of the order-up-to targets $\hat{y}_{i+1,1}$ and $\hat{y}_{i+1,2}$, and after dealing with the overshooting issue, we prove that the actual inventory level y_t converges to y^* in mean-square when $t \to \infty$.

Convergence of $\hat{y}_{i+1,1}$ and $\hat{y}_{i+1,2}$. First note that, for some constant K_9, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|y^* - \hat{y}_{i+1,1}\right|^2\right] \\
\leq\ & K_9\mathbb{E}\Big[\Big(\underbrace{\left|\overline{y}\big(e^{\lambda(p^*)}\big) - \overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big)\right|^2}_{\text{Difference between } p^* \text{ and } \hat{p}_{i+1}} + \underbrace{\left|\overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big)\right|^2}_{\text{Zero}} \\
&+ \underbrace{\left|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i})-\breve{\beta}(\hat{p}_{i})\hat{p}_{i+1}}\big)\right|^2}_{\text{Difference between } \hat{p}_{i+1} \text{ and } \hat{p}_i} \\
&+ \underbrace{\left|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i})-\breve{\beta}(\hat{p}_{i})\hat{p}_{i+1}}\big) - \tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_{i})-\breve{\beta}(\hat{p}_{i})\hat{p}_{i+1}}\big)\right|^2}_{\substack{\text{Comparison of problems B1 and B2} \\ \text{Proposition 4}}} + \underbrace{\left|\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_{i})-\breve{\beta}(\hat{p}_{i})\hat{p}_{i+1}}\big) - \hat{y}_{i+1,1}\right|^2}_{\text{Comparison of problems B2 and DD}}\Big)\Big].
\end{aligned} \tag{16}$$

In the following analysis we shall upper bound each term on the right hand side of (16). Note that in the first term, $\overline{y}\big(e^{\lambda(p)}\big)$ is the newsvendor solution, and it is given by

$$\overline{y}\big(e^{\lambda(p)}\big) = e^{\lambda(p)}F^{-1}\left(\frac{b}{b+h}\right).$$

Because $\lambda(p)$ has bounded first order derivative in $\mathcal{P}$ (Assumption 1(v)), there must exist some constant K_{10} such that

$$\mathbb{E}\left[\left|\overline{y}\big(e^{\lambda(p^*)}\big) - \overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big)\right|^2\right] \leq K_{10}\mathbb{E}\left[|\ p^* - \hat{p}_{i+1}\ |^2\right].$$

By the definition of $\breve{\alpha}(p)$ and $\breve{\beta}(p)$ in (7), one has $\breve{\alpha}(\hat{p}_{i+1}) - \breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1} = \lambda(\hat{p}_{i+1})$, thus the second term on the right hand side of (16) vanishes. For the third term, because $\overline{y}\big(e^{\breve{\alpha}(q)-\breve{\beta}(q)p}\big) = e^{\breve{\alpha}(q)-\breve{\beta}(q)p}F^{-1}\left(\frac{b}{b+h}\right)$, and $\lambda(p)$ has bounded first and second order derivatives (Assumption 1(v)), there exists some constant K_{11} such that

$$\begin{aligned}
\mathbb{E}\left[\left|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}}\big) - \overline{y}\big(e^{\breve{\alpha}(\hat{p}_{i})-\breve{\beta}(\hat{p}_{i})\hat{p}_{i+1}}\big)\right|^2\right] &\leq K_{11}\mathbb{E}\left[|\ \hat{p}_{i+1} - \hat{p}_i\ |^2\right] \\
&\leq 2K_{11}\mathbb{E}\left[\left(\ |\ p^* - \hat{p}_i\ |^2 + |\ p^* - \hat{p}_{i+1}\ |^2\ \right)\right].
\end{aligned}$$

The fourth term measures the distance between the inventory targets in B1 and B2, which is upper bounded in Proposition 4 below.

Proposition 4. *There exists some constant K_{12} such that, for any $p \in \mathcal{P}$ and $\hat{p}_i \in \mathcal{P}$, and any $\xi > 0$, it holds that*

$$\mathbb{P}\Big\{ \Big|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big) - \tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)\Big| \geq K_{12}\xi \Big\} \leq 2e^{-4I_i\xi^2}.$$

Based on the properties of the SAA newsvendor problem in B2, one has

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

where $e^{\alpha-\beta p}\min\left\{e^{\epsilon_j} : \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \leq e^{\epsilon_j}\right\} \geq \frac{b}{b+h}\right\}$ is the unconstrained optimal SAA newsvendor solution. Instead of comparing $\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$ and $\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$ directly, we first compare $F\big(\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)\big)$ and $F\big(\tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)\big)$ using Hoeffding inequality, and prove that their difference can be upper bounded by a small value with a high probability.

To bound the fifth term of the right hand side of (16), we follow a similar approach as that in proving Proposition 3. We again use problem B3 as a bridging problem between problems B2 and DD, and by the Lipschitz condition in (11) we show that convergence of parameters $\big((\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), \eta_{t=t_i+1}^{t_i+I_i}, \eta_{t=t_i+I_i+1}^{t_i+2I_i}\big)$ to $\big((\breve{\alpha}(\hat{p}_i)\ , \breve{\beta}(\hat{p}_i)), \epsilon_{t=t_i+1}^{t_i+I_i}, \epsilon_{t=t_i+I_i+1}^{t_i+2I_i}\big)$ will lead to convergence in the order-up-to target decisions, and thus the following bound for the fifth term holds (for details please see Appendix A),

$$\mathbb{E}\left[\left|\tilde{y}_{i+1}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}) - \hat{y}_{i+1,1}\right|^2\right] \leq K_{13}I_i^{-\frac{1}{2}}.$$

Summarizing the analyses above we obtain, for some constants K_{14} and K_{15},

$$\mathbb{E}\left[\big(y^* - \hat{y}_{i+1,1}\big)^2\right] \leq K_{14}\mathbb{E}\big[\ |\ p^* - \hat{p}_{i+1}\ |^2 + |\ p^* - \hat{p}_i\ |^2\ \big] + K_{14}I_i^{-\frac{1}{2}} \leq K_{15}I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty, \tag{17}$$

where the second inequality follows from the convergence rate of the pricing decisions in (15). Similarly, we obtain

$$\mathbb{E}\left[\big(y^* - \hat{y}_{i+1,2}\big)^2\right] \leq K_{16}I_i^{-\frac{1}{2}} \to 0, \qquad \text{as } i \to \infty.$$

Convergence of $\hat{y}_{i+1,1}$ and $\hat{y}_{i+1,2}$ is thus obtained.

Dealing with overshooting. We next prove the convergence of actual inventory order-up-to levels y_t, i.e., $\mathbb{E}[(y^* - y_t)^2] \to 0$ as $t \to \infty$. It suffices to prove this for (a) $t \in \{t_{i+1}+1, \ldots, t_{i+1}+I_{i+1}\}$, $i = 1, 2, \ldots$, and for (b) $t \in \{t_{i+1}+I_{i+1}+1, \ldots, t_{i+1}+2I_{i+1}\}$, $i = 1, 2, \ldots$. We will only provide the proof for (a), as the proof for (b) is similar.

The inventory order-up-to level prescribed in DDA for periods $t \in \{t_{i+1}+1, \ldots, t_{i+1}+I_{i+1}\}$ is $\hat{y}_{i+1,1}$. This, however, may not be achieved for some period t. Whether $\hat{y}_{i+1,1}$ can be achieved or not largely depends on the value of $\hat{y}_{i,2}$ and whether it is achieved during periods $\{t_i+I_i+1, \ldots, t_i+2I_i\}$. Define event $\mathcal{A}_1$ to be the set of sample paths on which $\hat{y}_{i,2}$ is achieved during periods $\{t_i+I_i+1,\ldots,t_i+2I_i\}$, and we prove $\mathbb{P}(\mathcal{A}_1)\geq 1-1/I_i^2$ (for details please see Appendix A), which approaches 1 as i grows. This is because, as I_i becomes larger, accumulative demands will grow large enough to consume the initial inventory at the beginning of period t_i+I_i+1 and bring the inventory level to or below $\hat{y}_{i,2}$, from which point on $\hat{y}_{i,2}$ can be achieved.

If $\hat{y}_{i,2}$ is achieved during periods $\{t_i+I_i+1,\ldots,t_i+2I_i\}$, then the actual order-up-to levels during periods $\{t_{i+1}+1,\ldots,t_{i+1}+I_{i+1}\}$ is $\hat{y}_{i+1,1}$ if $\hat{y}_{i+1,1}$ is higher than or equal to $\hat{y}_{i,2}$, and between $\hat{y}_{i+1,1}$ and $\hat{y}_{i,2}$ if $\hat{y}_{i+1,1}$ is smaller than $\hat{y}_{i,2}$. Because both $\hat{y}_{i,2}$ and $\hat{y}_{i+1,1}$ converge to y^*, y_t, $t\in\{t_{i+1}+1,\ldots,t_{i+1}+I_{i+1}\}$, also converge to y^*. To be more specific, for $t=t_{i+1}+1,\ldots,t_{i+1}+I_{i+1}$, we have

$$\begin{aligned}\mathbb{E}[(y^*-y_t)^2] &= \mathbb{P}(\mathcal{A}_1)\mathbb{E}[(y^*-y_t)^2\big|\mathcal{A}_1]+\mathbb{P}(\mathcal{A}_1^c)\mathbb{E}[(y^*-y_t)^2\big|\mathcal{A}_1^c]\\ &\leq \max\left\{\mathbb{E}\left[(y^*-\hat{y}_{i,2})^2\right],\mathbb{E}\left[(y^*-\hat{y}_{i+1,1})^2\right]\right\}+\frac{1}{I_i^2}\left(y^h-y^l\right)^2\\ &\leq K_{17}I_{i-1}^{-\frac{1}{2}}\to 0,\ \text{as } i\to\infty, \end{aligned}\tag{18}$$

where the last inequality follows from (17). Similarly, one can prove that for $t\in\{t_{i+1}+I_{i+1}+1,\ldots,t_{i+1}+2I_{i+1}\}$,

$$\mathbb{E}\left[(y^*-y_t)^2\right]\leq K_{18}I_{i-1}^{-\frac{1}{2}}\to 0,\ \text{as } i\to\infty.$$

This shows $\mathbb{E}[(y^*-y_t)^2]\to 0$ when $t\to\infty$. And again, since convergence in probability is implied by mean-square convergence, we conclude that actual inventory decisions y_t of DDA also converge to y^* in probability as $t\to\infty$.

This completes the proof of convergence of the inventory decision.

5.3 Proof of Theorem 2

We next prove the second main result, the convergence rate of regret. We have

$$\begin{aligned}&\mathbb{E}\Bigg[\sum_{t=1}^{T}\big(G(p^*,y^*)-G(p_t,y_t)\big)\Bigg]\\ \leq\ &\mathbb{E}\Bigg[\sum_{i=1}^{n}\Bigg(\sum_{t=t_i+1}^{t_i+I_i}\big(G(p^*,y^*)-G(\hat{p}_i,y_t)\big)+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(p^*,y^*)-G(\hat{p}_i+\delta_i,y_t)\big)\Bigg)\Bigg],\end{aligned}\tag{19}$$

where n is the smallest number of stages that cover T, i.e., n is the smallest integer such that $2I_0\sum_{i=1}^{n}v^i\geq T$, and it satisfies $\log_v\Big(\frac{v-1}{2I_0v}T+1\Big)\leq n<\log_v\Big(\frac{v-1}{2I_0v}T+1\Big)+1$. Then, the proof of Theorem 2 follows from (19) and the following proposition.

Proposition 5. *There exist a positive integer i^* and a positive constant K_{19} such that for any $i\geq i^*$,*

$$\mathbb{E}\left[G(p^*,y^*)-G(\hat{p}_i,y_t)\right]\leq K_{19}I_{i-1}^{-\frac{1}{2}},\quad \text{for } t\in\{t_i+1,\ldots,t_i+I_i\},$$

and

$$\mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\right] \leq K_{19} I_{i-1}^{-\frac{1}{2}}, \quad \text{for } t \in \{t_i + I_i + 1, \ldots, t_i + 2I_i\}.$$

This result can be proved using Taylor expansion to the second order, then applying the convergence rates in (15), (17), and (18). To be more specific,

$$\begin{aligned}
& \mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i, y_t)\right] \\
= \;& \mathbb{E}\Big[\Big(G\big(p^*, \overline{y}\big(e^{\lambda(p^*)}\big)\big) - G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big)\Big) + \Big(G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) - G(\hat{p}_i, y_t)\Big)\Big] \\
\leq \;& K_{20}\Big(\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + \mathbb{E}\Big[\big(\overline{y}(e^{\lambda(\hat{p}_i)}) - y_t\big)^2\Big]\Big) \\
\leq \;& K_{21}\left(K_{22} I_{i-1}^{-\frac{1}{2}} + K_{23} I_{i-1}^{-\frac{1}{2}}\right),
\end{aligned}$$

where the first inequality is true because p^* is the maximizer of $G\big(p, \overline{y}\big(e^{\lambda(p)}\big)\big)$, $\overline{y}\big(e^{\lambda(\hat{p}_i)}\big)$ is the maximizer of $G(\hat{p}_i, y)$ under given $\hat{p}_i$, and the first order derivative equals zero when applying Taylor expansion. The second inequality is based on (15) and similar analyses in developing upper bound for (16).

The two important special cases of our model are i) pure inventory control problem of non-perishable products (which is the case where the selling price p is exogenously given), and ii) pure dynamic pricing problem (that assumes there are unlimited initial inventory at the beginning of the planning horizon with zero holding cost). For the pure inventory control problem with fixed price, it is not necessary (nor is it possible) to test two different prices in each stage, thus stage i will only contain I_i periods to implement the base-stock level $\hat{y}_i$. We refer to the resulting simplified algorithm as the modified data-driven algorithm (MDDA). The analyses in the proofs of Propositions 4 and 5 lead to the result that MDDA solves the pure inventory management problem with regret $O(\log T/T)$ (we provide the details in Appendix B). We note that for the pure inventory control problem, Huh and Rusmevichientong (2009), in their Subsection 3.5, present an adaptive optimization algorithm for *perishable* inventory system (repetitive newsvendor problem) and show that, under the conditions of this paper (for this special case the condition is Assumption 1(vi)), the algorithm has regret $O(\log T/T)$. Hence, our result complements that of Huh and Rusmevichientong (2009) by presenting an alternative algorithm for *non-perishable* inventory system with the same regret rate. Also note that, this regret rate is the best possible for learning algorithms of this inventory problems (see e.g., Hazan et al. 2006, and Besbes and Muharremoglu 2013).

The second special case of pure dynamic pricing problem has been studied in Besbes and Zeevi (2015). Since there is infinite initial inventory, inventory replenishment is no longer necessary. Using a linear demand model in each learning cycle, Besbes and Zeevi (2015) show that the regret for the pure pricing problem is $\mathcal{O}(T^{-1/2}(\log T)^2)$. Therefore, there are two main factors that contribute to the regret in this paper: inventory decision and pricing decision, but it is the pricing decision that ultimately determines the performance bound.

6 Conclusion

In this paper, we study the joint pricing and inventory control problem when the firm has limited prior knowledge about the demand distribution and customer response to selling prices. We design a nonparametric data-driven learning algorithm and show that its regret converges to zero at the non-improvable rate of $O(T^{-1/2})$.

There are a number of follow-up research topics. One is to develop an asymptotically optimal algorithm for the problem with lost-sales and censored data. In the lost-sales case, the DDA algorithm proposed here cannot be directly applied and the estimation and optimization problems are more challenging as the profit function of the data-driven problem is neither concave nor unimodal, and the demand data is censored. Another interesting direction for research is to develop a data-driven learning algorithm for dynamic pricing and stocking decisions for multiple products in an assortment. Finally, in this paper we have assumed that the firm has some prior information about the structure of the demand process, i.e., $\tilde{D}(p) = \tilde{\lambda}(p) + \tilde{\epsilon}$ for the additive model or $\tilde{D}(p) = \tilde{\lambda}(p)\tilde{\epsilon}$ for the multiplicative model. An interesting research is to develop learning algorithms when such structure is not known in advance, e.g., when the demand process is $\tilde{D}(p) = \tilde{\lambda}(p, \epsilon_t)$ for some unknown function $\tilde{\lambda}(\cdot, \cdot)$. It is not clear whether the linear model will still work in this more general setting.

Acknowledgment: The authors are very grateful to the Department Editor, the Associate Editor, and three referees for their constructive comments on earlier versions of this paper, that have helped to significantly improve both the content and exposition of this paper. The research of Xiuli Chao is supported in part by NSF grants CMMI-1131249, CMMI-1362619, and CMMI-1634676.

References

[1] Agarwal A, Foster DP, Hsu DJ, Kakade SM, Rakhlin A (2011) Stochastic convex optimization with bandit feedback. *Advances in Neural Information Processing Systems*1035-1043.

[2] Agarwal S, Devanur NR (2014) Bandits with concave rewards and convex knapsacks. *Proceedings of the fifteenth ACM conference on economics and computation* 989-1006.

[3] Agarwal S, Devanur NR (2015) Linear contextual bandits with global constraints and objective. *Working paper.*

[4] Auer P, Ortner R, Szepesvari C (2007) Improved rates for the stochastic continuum-armed bandit problem. *Proceedings of the 20th International Conference on Learning Theory (COLT)* 454-468.

[5] Badanidiyuru A, Kleinberg R, Slivkins A (2013) Bandits with knapsacks. *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on. IEEE* 207-216.

[6] Besbes O, Muharremoglu A (2013) On implications of demand censoring in the newsvendor problem. *Management Sci.* 59(6):1407-1424.

[7] Besbes O, Zeevi A (2009) Dynamic pricing without knowing the demand function: risk bounds and near-optimal algorithms. *Oper. Res.* 57(6):1407-1420.

[8] Besbes O, Zeevi A (2012) Blind network revenue management. *Oper. Res.* 60(6): 1537-1550.

[9] Besbes O, Zeevi A (2015) On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Sci.* 61(4): 723-739.

[10] Buche R, and Kushner HJ (2002) Rate of convergence for constrained stochastic approximation algorithms. *SIAM Journal on Control and Optimization* 40(4):1011-1041.

[11] Burnetas AN, Smith CE (2000) Adaptive ordering and pricing for perishable products. *Oper. Res.* 48(3):436-443.

[12] Chen X, Simchi-Levi D (2004) Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: the finite horizon case. *Oper. Res.* 52(6):887-896.

[13] Chen X, Simchi-Levi D (2012) Pricing and inventory management. Philips R and Ozalp O, eds. *The Handbook of Pricing Management*. Oxford University Press, 784-822.

[14] Cope EW (2009) Regret and convergence bounds for a class of continuum-armed bandit problems. *Automatic Control, IEEE Transactions* 54(6):1243-1253.

[15] Denardo EV, Feinberg EA, Rothblum UG (2013) The multi-armed bandit, with constraints. *Annals of Operations Research* 208(1): 37-62.

[16] Ding W, Qin T, Zhang XD, Liu TY (2013) Multi-armed bandit with budget constraint and variable costs. *AAAI* 2013 Jun 30.

[17] Elmaghraby W, Keskinocak P (2003) Dynamic pricing in the presence of inventory considerations: research overview, current practices, and future directions. *Management Sci.* 49(10):1287-1309.

[18] Federgruen A, Heching A (1999) Combined pricing and inventory control under uncertainty. *Oper. Res.* 47(3):454-475.

[19] Ferreira KJ, Simchi-Levi D, Wang H (2016) Online Network Revenue Management Using Thompson Sampling. *Working paper.*

[20] Guha S, Munagala K (2009) Multi-armed bandits with metric switching costs. *International Colloquium on Automata, Languages, and Programming* Jul 5: 496-507. Springer Berlin Heidelberg.

[21] Hazan E, Kalai A, Kale S, Agarwal A (2006) Logarithmic regret algorithms for online convex optimization. *Learning Theory.* Springer Berlin Heidelberg, 499-513.

[22] Huh WT, Levi R, Rusmevichientong P, Orlin JB (2011) Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* 59(4):929-941.

[23] Huh WT, Rusmevichientong P (2009) A nonparametric asymptotic analysis of inventory planning with censored demand. *Math. Oper. Res.* 34(1):103-123.

[24] Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: asymptotically optimal semi-myopic policies, *Oper. Res.* 62(5):1142-1167.

[25] Kiefer J, Wolfowitz J (1952) Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* (23):462-466.

[26] Kleinberg R (2005) Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*: 697-704.

[27] Kleywegt AJ, Shapiro A, Homem-de-mello T. (2001) The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* 12(2):479-502.

[28] Kushner (2010) Stochastic approximation: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(1): 87-96.

[29] Kushner HJ, Yin G (1997) *Stochastic approximation algorithms and applications.* Springer-Verlag, Berlin and New York, 1997.

[30] Kushner H, Yin G (2003). *Stochastic approximation and recursive algorithms and applications.* Springer Science & Business Media; 2003 Jul 17.

[31] Lai TL, Robbins H (1981) Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Probability Theory and Related Fields* 56(3):329-360.

[32] Levi R, Perakis G, Uichanco J (2015) The data-driven newsvendor problem: new bounds and insights. *Oper. Res.* 63(6):1294-1306.

[33] Levi R, Roundy RO, Shmoys DB (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. *Math. Oper. Res.* 32(4):821-839.

[34] Petruzzi NC, Dada M (1999) Pricing and the newsvendor problem: A review with extensions. *Oper. Res.* 47(2):183-194.

[35] Robbins H, Monro S (1951) A stochastic approximation method. *Ann. Math. Statist.* (22):400-407.

[36] Vershynin R (2017) *High Dimensional Probability: An Introduction with Applications to Data Science.* Lecture notes. Available at http:www-personal.umich.edu/ romanvpapersHDP-bookHDP-book.pdf.

[37] Wang Z, Deng S, Ye Y (2014) Close the gaps: a learning-while-doing algorithm for single-product revenue management problems. *Oper. Res.* 62(2):318-331.

[38] Wu CF (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics* 1261-1295.

[39] Yano CA and Gilbert SM (2003) Coordinated pricing and production/procurement decisions: A review. J. Eliashberg, A. Chakravarty, eds. *Managing Business Interfaces: Marketing, Engineering, and Manufacturing Perspectives.* Kluwer, Norwell, MA.

[40] Zinkevich M (2003) Online convex programming and generalized infinitesimal gradient ascent. *Proc. 20th Internat. Conf. Machine Learn. (ICML-2003)* Washington, D.C.

Online Appendices for
"Coordinating Pricing and Inventory Replenishment
with Nonparametric Demand Learning"

Boxiao Chen[4], Xiuli Chao[5] and Hyun-Soo Ahn[6]

Appendix A

In this Appendix, we provide all the proofs omitted in the main context. We first present the proof of Proposition 1, which compares the optimal solutions of problem CI and bridging problem B1, i.e., p^* and $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$.

Proof of Proposition 1. First we make the observation that

$$p^* = \overline{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big). \tag{20}$$

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

$$\overline{Q}(p, e^{\lambda(p)}) = pe^{\lambda(p)}\mathbb{E}[e^{\epsilon}] - \min_{y\in\mathcal{Y}} \left\{ h\mathbb{E}\big[y - e^{\lambda(p)}e^{\epsilon}\big]^+ + b\mathbb{E}\big[e^{\lambda(p)}e^{\epsilon} - y\big]^+ \right\}. \tag{21}$$

We show that Assumption 1(i) implies that $\overline{Q}(p, e^{\lambda(p)})$ is unimodal in p. Recall $\tilde{\lambda}(p) = e^{\lambda(p)}$. Letting $d = \tilde{\lambda}(p)$, we rewrite the above equation as

$$\overline{Q}(\tilde{\lambda}^{-1}(d), d) = d\tilde{\lambda}^{-1}(d)\mathbb{E}[e^{\epsilon}] - \min_{y\in\mathcal{Y}} \left\{ h\mathbb{E}\big[y - de^{\epsilon}\big]^+ + b\mathbb{E}\big[de^{\epsilon} - y\big]^+ \right\}.$$

It is clear that $h\mathbb{E}\big[y - de^{\epsilon}\big]^+ + b\mathbb{E}\big[de^{\epsilon} - y\big]^+$ is jointly convex in (d, y), hence

$$\min_{y\in\mathcal{Y}} \left\{ h\mathbb{E}\big[y - de^{\epsilon}\big]^+ + b\mathbb{E}\big[de^{\epsilon} - y\big]^+ \right\}$$

is convex in d (Proposition B4 of Heyman and Sobel 1984). Also, by Assumption 1(i), $d\tilde{\lambda}^{-1}(d)\mathbb{E}[e^{\epsilon}]$ is concave. Therefore, $\overline{Q}(\tilde{\lambda}^{-1}(d), d)$ is concave in d. The unimodality of $\overline{Q}(p, e^{\lambda(p)})$ follows from the concavity of $\overline{Q}(\tilde{\lambda}^{-1}(d), d)$, and the fact that $\tilde{\lambda}(p)$ is strictly decreasing in p.

We assume that $\overline{Q}$ has a unique maximizer and that $\overline{p}(\breve{\alpha}(z), \breve{\beta}(z))$ is the unique optimal solution for problem B1 with parameters $\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$, then (20) follows from Lemma A1 of Besbes and Zeevi (2015) by letting their function G be (21).

[4]Department of Information and Decision Science, College of Business Administration, University of Illinois, Chicago, IL 60607. Email: bbchen@uic.edu

[5]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109. Email: xchao@umich.edu

[6]Department of Technology and Operations, Ross School of Business, University of Michigan, Ann Arbor, MI 48109. Email: hsahn@umich.edu

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

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

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

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

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

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

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

To prove Lemma 1, we first change the decision variables in B1 and B2, and these transformations are also used in Lemmas A1 and A2 as well as in proving Proposition 2. For given parameters α and $\beta>0$, define $d=e^{\alpha-\beta p}$, $d\in\mathcal{D}=[d^l,d^h]$ where $d^l=e^{\alpha-\beta p^h}$ and $d^h=e^{\alpha-\beta p^l}$. Then problem B1 can be rewritten as

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

Define

$$\overline{W}(d,y)=h\mathbb{E}\big(y-de^\epsilon\big)^+ + b\mathbb{E}\big(de^\epsilon-y\big)^+,\tag{22}$$

and then

$$\overline{Q}\left(\frac{\alpha-\log d}{\beta},d\right)=d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^\epsilon\big]-\min_{y\in\mathcal{Y}}\overline{W}(d,y)=d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^\epsilon\big]-\overline{W}(d,\overline{y}(d)),\tag{23}$$

where $\overline{y}(d)$ is the optimal solution of (22) in $\mathcal{Y}$ for given d. $F(\cdot)$ is the cumulative distribution function (CDF) of e^ϵ, then it can be verified that

$$\overline{y}(d)=dF^{-1}\left(\frac{b}{b+h}\right),\tag{24}$$

where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Also, we let $\overline{d}\,(\alpha,\beta)$ denote the optimal solution of maximizing (23) over $\mathcal{D}$.

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

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

Let

$$\tilde{W}_{i+1}(d,y) = \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(y-de^{\epsilon_t}\big)^+ + b\big(de^{\epsilon_t}-y\big)^+\Big), \tag{25}$$

and then

$$\begin{aligned}\tilde{Q}_{i+1}\left(\frac{\alpha-\log d}{\beta}, d\right) &= d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right) - \min_{y\in\mathcal{Y}}\tilde{W}_{i+1}(d,y) \\ &= d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right) - \tilde{W}_{i+1}(d,\tilde{y}(d)), \end{aligned}\tag{26}$$

where $\tilde{y}_{i+1}(d)$ denotes the optimal solution of $\tilde{W}_{i+1}(d,y)$ in (25) on $\mathcal{Y}$. Let $\tilde{d}_{i+1}(\alpha,\beta)$ be the optimal solution for $\tilde{Q}_{i+1}(\frac{\alpha-\log d}{\beta}, d)$ in (26) on $\mathcal{D}$. Also, let $\tilde{y}^u_{i+1}(d)$ denote the optimal order-up-to level for problem B2 on $\mathbb{R}_+$ for given $p\in\mathcal{P}$ (here the superscript "u" stands for "unconstrained"). Then

$$\tilde{y}^u_{i+1}(d) = d\min\left\{e^{\epsilon_j} : \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\le e^{\epsilon_j}\right\} \ge \frac{b}{b+h}\right\}. \tag{27}$$

It can be checked that

$$\tilde{y}_{i+1}(d) = \min\left\{\max\left\{\tilde{y}^u_{i+1}(d), y^l\right\}, y^h\right\}. \tag{28}$$

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

Proof of Lemma 1. It is easily seen that $\overline{W}(d,y)$ and $\tilde{W}_{i+1}(d,y)$ are both jointly convex in (d,y), hence $\min_{y\in\mathcal{Y}}\overline{W}(d,y)$ and $\min_{y\in\mathcal{Y}}\tilde{W}_{i+1}(d,y)$ are convex in d (Proposition B4 of Heyman and Sobel (1984)). Also, the first term of $\overline{Q}$ (and $\tilde{Q}_{i+1}$) is concave when $\beta>0$. Therefore, if $\beta>0$, both $\overline{Q}(\frac{\alpha-\log d}{\beta}, d)$ and $\tilde{Q}_{i+1}(\frac{\alpha-\log d}{\beta}, d)$ are concave in $d\in\mathcal{D}$.

The unimodality of $\overline{Q}(p, e^{\alpha-\beta p})$ and $\tilde{Q}_{i+1}(p, e^{\alpha-\beta p})$ can be proved due to the concavity of $\overline{Q}(\frac{\alpha-\log d}{\beta}, d)$ and $\tilde{Q}_{i+1}(\frac{\alpha-\log d}{\beta}, d)$, and the fact that $e^{\alpha-\beta p}$ is strictly decreasing in p when $\beta>0$. Unimodality of $\breve{Q}_{i+1}(p, e^{\alpha-\beta p})$ can be proved following the same lines as in proving $\tilde{Q}_{i+1}(p, e^{\alpha-\beta p})$. $\square$

To prove Proposition 2, we need to establish Lemmas A1 and A2 first, and they are introduced after the changing-decision-variable procedure from p to d.

Lemma A1. *There exists a positive constant K_{24} such that, for any $\xi>0$,*

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \le K_{24}\xi\right\} \ge 1-4e^{-2I_i\xi^2}.$$

Proof. By triangle inequality, we have

$$\begin{aligned}&\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \\ \le\ &\max_{d\in\mathcal{D}}\left|\overline{W}\big(d,\overline{y}(d)\big) - \tilde{W}_{i+1}\big(d,\overline{y}(d)\big)\right| + \max_{d\in\mathcal{D}}\left|\tilde{W}_{i+1}\big(d,\overline{y}(d)\big) - \tilde{W}_{i+1}\big(d,\tilde{y}_{i+1}(d)\big)\right|. \end{aligned}\tag{29}$$

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

For any $d \in \mathcal{D}$ and $y \in \mathcal{Y}$, we define $z = y/d$. Then, from (24), the optimal z to minimize $\overline{W}(d, dz)$ is

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

Moreover, we have

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

and

$$\tilde{W}_{i+1}(d,\overline{y}(d)) = \tilde{W}_{i+1}(d, d\overline{z}) = d\Bigg(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(\overline{z} - e^{\epsilon_t}\big)^{+} + b\big(e^{\epsilon_t} - \overline{z}\big)^{+}\Big)\Bigg). \tag{30}$$

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

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

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

$$\mathbb{P}\left\{d^h\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right| > d^h\xi\right\} = \mathbb{P}\left\{\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right| > \xi\right\} \le 2e^{-2I_i\xi^2/(\Delta^h-\Delta^l)^2}, \tag{31}$$

which deduces to

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\overline{y}(d))\right| > d^h\xi\right\} \le 2e^{-2I_i\xi^2/(\Delta^h-\Delta^l)^2}. \tag{32}$$

This bounds the first term on the right hand side of (29). Note that when developing the bound, the separability of the terms involving d and the terms involving ϵ_t is crucial. Otherwise, because Hoeffding bounds can only be applied to a fixed d, to develop the result in (32) with "$\max_{d\in\mathcal{D}}$" is more challenging.

To bound the second term in (29), we use

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

to denote the empirical distribution of e^{ϵ_t}. For $\theta > 0$, it can be verified that

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

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

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

Combining the two results above we obtain

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

therefore the probability for $\hat{F}(\cdot)$ to be a θ-estimate is at least $1 - 2e^{-2I_i\theta^2}$. Let $\mathcal{A}_2(\theta)$ represent the event that $\hat{F}(\bar{z})$ is a θ-estimate, then the result above states that

$$\mathbb{P}(\mathcal{A}_2(\theta)) \geq 1 - 2e^{-2I_i\theta^2}. \tag{33}$$

For $d \in \mathcal{D}$, let $\tilde{z}_{i+1}(d) = \frac{\tilde{y}_{i+1}(d)}{d}$ and $\tilde{z}^u_{i+1} = \frac{\tilde{y}^u_{i+1}(d)}{d}$, then it follows from (27) that

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

And it follows from (28) that

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

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

First, for any given $d \in \mathcal{D}$, if $\overline{z} \leq \tilde{z}^u_{i+1}$, then it follows from (30) that

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\overline{z}) \quad &= \quad \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(e^{\epsilon_t} - \overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\big\} \\
&\qquad + b\big(e^{\epsilon_t} - \overline{z}\big)\mathbb{1}\big\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z} - e^{\epsilon_t}\big)\mathbb{1}\big\{e^{\epsilon_t} \leq \overline{z}\big\}\Big] \\
&\leq \quad \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(e^{\epsilon_t} - \overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\big\} \\
&\qquad + b\big(\tilde{z}^u_{i+1} - \overline{z}\big)\mathbb{1}\big\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z} - e^{\epsilon_t}\big)\mathbb{1}\big\{e^{\epsilon_t} \leq \overline{z}\big\}\Big],
\end{aligned} \tag{34}$$

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

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \quad &= \quad \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\} \\
&\qquad + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \overline{z}\}\Big] \\
&\geq \quad \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \overline{z}\}\Big],
\end{aligned} \tag{35}$$

with the inequality obtained by dropping the nonnegative middle term. Consequently when $\overline{z} \le \tilde{z}^u_{i+1}$ we subtract (35) from (34) to obtain

$$\begin{aligned}
&\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \\
\le\ & d\Big(b(\tilde{z}^u_{i+1} - \overline{z})(1 - \hat{F}(\tilde{z}^u_{i+1})) + b(\tilde{z}^u_{i+1} - \overline{z})(\hat{F}(\tilde{z}^u_{i+1}) - \hat{F}(\overline{z})) + h(\overline{z} - \tilde{z}^u_{i+1})\hat{F}(\overline{z})\Big) \\
=\ & d(\tilde{z}^u_{i+1} - \overline{z})\big(-(h+b)\hat{F}(\overline{z}) + b\big) \\
\le\ & d\big(\tilde{z}^u_{i+1} - \overline{z}\big)(b+h)\theta, \qquad (36)
\end{aligned}$$

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

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

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\overline{z}) \ =\ & \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \overline{z})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} \\
& \quad + h(\overline{z} - e^{\epsilon_t})\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t} \le \overline{z}\} + h(\overline{z} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \le \tilde{z}^u_{i+1}\}\Big] \\
\le\ & \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \overline{z})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} \\
& \quad + h(\overline{z} - \tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t} \le \overline{z}\} + h(\overline{z} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \le \tilde{z}^u_{i+1}\}\Big], \qquad (37)
\end{aligned}$$

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

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \ =\ & \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} \\
& \quad + b(e^{\epsilon_t} - \tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t} \le \overline{z}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \le \tilde{z}^u_{i+1}\}\Big] \\
\ge\ & \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \le \tilde{z}^u_{i+1}\}\Big], \qquad (38)
\end{aligned}$$

again the inequality follows from dropping the nonnegative second term. Subtracting (38) from (37), we obtain

$$\begin{aligned}
&\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \\
\le\ & d\Big(b(\tilde{z}^u_{i+1} - \overline{z})(1 - \hat{F}(\overline{z})) + h(\overline{z} - \tilde{z}^u_{i+1})(\hat{F}(\overline{z}) - \hat{F}(\tilde{z}^u_{i+1})) + h(\overline{z} - \tilde{z}^u_{i+1})\hat{F}(\tilde{z}^u_{i+1})\Big) \\
=\ & d(\overline{z} - \tilde{z}^u_{i+1})((h+b)\hat{F}(\overline{z}) - b) \\
\le\ & d(\overline{z} - \tilde{z}^u_{i+1})(b+h)\theta, \qquad (39)
\end{aligned}$$

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

The results (36) and (39) imply that, when $\hat{F}(\cdot)$ is a θ-estimate, or (13) is satisfied, it holds that

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

As demand is bounded, $d\tilde{z}_{i+1}^u$ is bounded too, hence it follows from $d\overline{z} \in \mathcal{Y}$ that there exists some constant $K_{25} > 0$ such that $d|\overline{z} - \tilde{z}_{i+1}^u| \leq K_{25}$. Thus

$$\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}^u) \leq K_{25}(b+h)\theta.$$

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

$$\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d)) \leq \tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}^u) \leq K_{25}(b+h)\theta.$$

As this inequality holds for any $d \in \mathcal{D}$, it implies that, when $\hat{F}(\cdot)$ is a θ-estimate, or on the event $\mathcal{A}_2(\theta)$,

$$\max_{d\in\mathcal{D}} \left\{ \tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d)) \right\} \leq K_{25}(b+h)\theta. \tag{40}$$

Letting $\theta = \xi$ in (40) we obtain

$$\begin{aligned} & \mathbb{P}\left\{ \max_{d\in\mathcal{D}} \left(\tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d)) \right) \leq K_{25}(b+h)\xi \right\} \\ \geq \ & \mathbb{P}(\mathcal{A}_2(\xi)) \\ \geq \ & 1 - 2e^{-2I_i\xi^2}, \end{aligned}$$

where the last inequality follows from (33). This proves, by noting $\tilde{W}_{i+1}(d, \overline{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \geq 0$ as $\tilde{y}_{i+1}(d)$ is the minimizer of $\tilde{W}_{i+1}$ on $\mathcal{Y}$, that

$$\mathbb{P}\left\{ \max_{d\in\mathcal{D}} \left| \left(\tilde{W}_{i+1}(d, \overline{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \right) \right| \leq K_{25}(b+h)\xi \right\} \geq 1 - 2e^{-2I_i\xi^2}. \tag{41}$$

Applying (32) and (41) in (29), we conclude that there exists a constant $K_{24} > 0$ such that for any $\xi > 0$, when I_i is sufficiently large,

$$\mathbb{P}\left\{ \max_{d\in\mathcal{D}} \left| \overline{W}(d, \overline{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \right| \leq K_{24}\xi \right\} \quad \geq \quad 1 - 4e^{-2I_i\xi^2}.$$

This completes the proof of Lemma A1. □

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

Lemma A2. *There exists a positive constant K_{26} such that, for any $\xi > 0$,*

$$\mathbb{P}\left\{ \max_{d\in\mathcal{D}} \left| \overline{Q}\left(\frac{\alpha - \log d}{\beta}, d \right) - \tilde{Q}_{i+1}\left(\frac{\alpha - \log d}{\beta}, d \right) \right| \geq K_{26}\xi \right\} \leq 6e^{-2I_i\xi^2}.$$

Proof. For any $d \in \mathcal{D}$, similar argument as that used in proving (31) of Lemma A1 shows that, for any $\xi > 0$,

$$\mathbb{P}\left\{ \left| \mathbb{E}[e^{\epsilon_t}] - \left(\frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t} \right) \right| \leq (u-l)\xi \right\} \geq 1 - 2e^{-4I_i\xi^2}.$$

Let $r^* = \max_{d\in\mathcal{D}} \frac{|\alpha-\log d|}{\beta} d$, then we have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}[e^{\epsilon_t}] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*(u-l)\xi\right\} \\
= \;& \mathbb{P}\left\{r^*\left|\mathbb{E}[e^{\epsilon_t}] - \left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*(u-l)\xi\right\} \\
\ge \;& 1 - 2e^{-4I_i\xi^2}.
\end{aligned} \tag{42}$$

Hence, it follows from (23) and (26) that, for any $d \in \mathcal{D}$ and $\xi > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{Q}\left(\frac{\alpha-\log d}{\beta}, d\right) - \tilde{Q}_{i+1}\left(\frac{\alpha-\log d}{\beta}, d\right)\right| \le (K_{24} + r^*(u-l))\xi\right\} \\
= \;& \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\Bigg|\left(d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - \overline{W}(d,\overline{y}(d))\right) - \left(d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right)\Bigg| \\
&\qquad\qquad \le (K_{24} + r^*(u-l))\xi\Bigg\} \\
\ge \;& \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| + \max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \\
&\qquad\qquad \le (K_{24} + r^*(u-l))\xi\Bigg\} \\
\ge \;& \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*(u-l)\xi, \\
&\qquad\qquad \text{and } \max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| \le K_{24}\xi\Bigg\} \\
= \;& 1 - \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| > r^*(u-l)\xi, \\
&\qquad\qquad \text{or } \max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| > K_{24}\xi\Bigg\} \\
\ge \;& 1 - \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| > r^*(u-l)\xi\Bigg\} \\
&\qquad\qquad - \mathbb{P}\Bigg\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d) - \tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right| > K_{24}\xi\Bigg\} \\
\ge \;& 1 - 2e^{-4I_i\xi^2} - 4e^{-2I_i\xi^2} \\
\ge \;& 1 - 6e^{-2I_i\xi^2},
\end{aligned}$$

where the last inequality follows from (42) and Lemma A1. Letting $K_{26} = K_{24} + r^*(u-l)$ completes the proof of Lemma A2. □

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

$$\mathcal{A}_3(\xi) = \left\{\omega : \max_{d\in\mathcal{D}} \left| \overline{Q}\left(\frac{\alpha - \log d}{\beta}, d\right) - \tilde{Q}_{i+1}\left(\frac{\alpha - \log d}{\beta}, d\right)\right| \le K_{26}\xi \right\}. \tag{43}$$

Then Lemma A2 can be reiterated as $\mathbb{P}(\mathcal{A}_3(\xi)) \ge 1 - 6e^{-2I_i\xi^2}$.

With preparations of Lemmas A1 and A2 above, we are now ready to compare the optimal solutions of problems B1 and B2 and prove Proposition 2.

Proof of Proposition 2. To slightly simplify the notation, for given parameters α and β, in this proof we let

$$\overline{Q}(d) = \overline{Q}\left(\frac{\alpha - \log d}{\beta}, d\right), \quad \tilde{Q}(d) = \tilde{Q}_{i+1}\left(\frac{\alpha - \log d}{\beta}, d\right), \quad \overline{d} = \overline{d}(\alpha, \beta), \quad \tilde{d} = \tilde{d}_{i+1}(\alpha, \beta).$$

By Taylor's expansion,

$$\overline{Q}(\tilde{d}) = \overline{Q}(\overline{d}) + \overline{Q}'(\overline{d})(\tilde{d} - \overline{d}) + \frac{\overline{Q}''(q)}{2}(\tilde{d} - \overline{d})^2, \tag{44}$$

where $q \in [\overline{d}, \tilde{d}]$ if $\overline{d} \le \tilde{d}$ and $q \in [\tilde{d}, \overline{d}]$ if $\overline{d} > \tilde{d}$. Since we assume the minimizer of $\overline{W}(d, y)$ over $\mathbb{R}_+$ falls into $\mathcal{Y}$, it follows from (23) that $\overline{Q}(d) = d\frac{\alpha - \log d}{\beta}\mathbb{E}[e^\epsilon] - Ad$, where $A = \min_z \{h\mathbb{E}(z - e^\epsilon)^+ + b\mathbb{E}(e^\epsilon - z)^+\} > 0$ is a constant. Thus, we have

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

Since $\lambda(\cdot)$ is assumed to be strictly decreasing, it follows that $\breve{\beta}(\cdot)$ is bounded below by a positive number, say $\bar{a} > 0$. On $\beta \ge \bar{a}$, let $\min_{d\in\mathcal{D}} \frac{\mathbb{E}[e^\epsilon]}{\beta d} = m$ and it holds that $m > 0$, then it follows from (44) that

$$\overline{Q}(\tilde{d}) \le \overline{Q}(\overline{d}) - \frac{m}{2}(\tilde{d} - \overline{d})^2. \tag{45}$$

Now we prove, on event $\mathcal{A}_3(\xi)$, that

$$\overline{Q}(\tilde{d}) - \overline{Q}(\overline{d}) \ge -2K_{26}\xi. \tag{46}$$

We prove this by contradiction. Suppose it is not true, i.e., $\overline{Q}(\overline{d}) - \overline{Q}(\tilde{d}) > 2K_{26}\xi$, then it follows from (43) that

$$\begin{aligned} & \tilde{Q}(\overline{d}) - \tilde{Q}(\tilde{d}) \\ = \ & \big(\tilde{Q}(\overline{d}) - \overline{Q}(\overline{d})\big) + \big(\overline{Q}(\overline{d}) - \overline{Q}(\tilde{d})\big) + \big(\overline{Q}(\tilde{d}) - \tilde{Q}(\tilde{d})\big) \\ > \ & -K_{26}\xi + 2K_{26}\xi - K_{26}\xi \\ = \ & 0. \end{aligned}$$

This leads to $\tilde{Q}(\overline{d}) > \tilde{Q}(\tilde{d})$, contradicting with $\tilde{d}$ being optimal for problem B2. Thus, (46) is satisfied on $\mathcal{A}_3(\xi)$.

Using (45) and (46), we obtain that, on event $\mathcal{A}_3(\xi)$,

$$\left|\tilde{d} - \overline{d}\right|^2 \leq \frac{4K_{26}}{m}\xi,$$

or equivalently, for some constant K_{27},

$$\left|\tilde{d} - \overline{d}\right| \leq K_{27}\xi^{\frac{1}{2}}.$$

Let $g(d) = \frac{\alpha - \log d}{\beta}$, then $\overline{p}(\alpha,\beta) = g(\overline{d})$ and $\tilde{p}_{i+1}(\alpha,\beta) = g(\tilde{d})$. Since the first order derivative of $g(d)$ with respect to $d \in \mathcal{D}$ is bounded, there exists a constant $K_{28} > 0$, such that on $\mathcal{A}_3(\xi)$, it holds that

$$|\overline{p}(\alpha,\beta) - \tilde{p}_{i+1}(\alpha,\beta)| = |g(\overline{d}) - g(\tilde{d})| \leq K_{28}|\overline{d} - \tilde{d}| \leq K_{28} \times K_{27}\xi^{\frac{1}{2}}.$$

Letting $K_3 = K_{28} \times K_{27}$, this shows that for any values of α and $\beta \geq \bar{a}$,

$$\mathbb{P}\left\{|\overline{p}(\alpha,\beta) - \tilde{p}_{i+1}(\alpha,\beta)| \leq K_3\xi^{\frac{1}{2}}\right\} \geq \mathbb{P}(\mathcal{A}_3(\xi)) \geq 1 - 6e^{-2I_i\xi^2}.$$

Substituting $\alpha = \breve{\alpha}(p)$ and $\beta = \breve{\beta}(p)$, we obtain the desired result in Proposition 2. □

To prove Proposition 3, we need two other results, Lemmas A3 and A4 below.

Lemma A3. *There exists a positive constant i^* such that when $i \geq i^*$,*

$$\mathbb{P}\left\{\hat{\beta}_{i+1} > 0\right\} \geq 1 - \frac{8}{I_i}.$$

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

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

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

$$\hat{\alpha}_{i+1} = \frac{\lambda(\hat{p}_i) + \lambda(\hat{p}_i + \delta_i)}{2} + \frac{B^1_{i+1} + B^2_{i+1}}{2} + \hat{\beta}_{i+1}\frac{2\hat{p}_i + \delta_i}{2}, \tag{47}$$

$$\hat{\beta}_{i+1} = -\frac{\lambda(\hat{p}_i + \delta_i) - \lambda(\hat{p}_i)}{\delta_i} - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}). \tag{48}$$

Applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the second order for (48), we obtain

$$\begin{aligned}\hat{\beta}_{i+1} &= -\left(\lambda'(\hat{p}_i) + \frac{1}{2}\lambda''(q_i)\delta_i\right) - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}) \\ &= \breve{\beta}(\hat{p}_i) - \frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}),\end{aligned} \tag{49}$$

where $q_i \in [\hat{p}_i, \hat{p}_i + \delta_i]$. Substituting $\hat{\beta}_{i+1}$ in (47) by (49), and applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the first order, we have

$$\begin{aligned}
\hat{\alpha}_{i+1} &= \lambda(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B_{i+1}^1 + B_{i+1}^2}{2} - \lambda'(\hat{p}_i)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\
&\qquad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B_{i+1}^1 + B_{i+1}^2)\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\
&= \breve{\alpha}(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B_{i+1}^1 + B_{i+1}^2}{2} - \frac{1}{2}\lambda'(\hat{p}_i)\delta_i \\
&\qquad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B_{i+1}^1 + B_{i+1}^2)\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right), \qquad (50)
\end{aligned}$$

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

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

$$\mathbb{P}\left\{\left|-B_{i+1}^1\right| > \xi\right\} \le 2e^{-2I_i\xi^2/(u-l)^2}, \qquad \mathbb{P}\left\{\left|B_{i+1}^2\right| > \xi\right\} \le 2e^{-2I_i\xi^2/(u-l)^2}. \qquad (51)$$

Hence,

$$\mathbb{P}\left\{\left|-B_{i+1}^1\right| + \left|B_{i+1}^2\right| > 2\xi\right\} \le \mathbb{P}\left\{\left|-B_{i+1}^1\right| > \xi\right\} + \mathbb{P}\left\{\left|B_{i+1}^2\right| > \xi\right\} \le 4e^{-2I_i\xi^2/(u-l)^2}.$$

Therefore,

$$\mathbb{P}\left\{\left|-B_{i+1}^1 + B_{i+1}^2\right| \le 2\xi\right\} \ge \mathbb{P}\left\{\left|-B_{i+1}^1\right| + \left|B_{i+1}^2\right| \le 2\xi\right\} \ge 1 - 4e^{-2I_i\xi^2/(u-l)^2}.$$

Similar argument shows

$$\mathbb{P}\left\{\left|B_{i+1}^1 + B_{i+1}^2\right| \le 2\xi\right\} \ge 1 - 4e^{-2I_i\xi^2/(u-l)^2}.$$

Since $\lambda'(\cdot)$ and $\lambda''(\cdot)$ are bounded and δ_i converges to 0, from (50) we conclude that there must exist a constant K_{29} such that, on the event $\left|B_{i+1}^1 + B_{i+1}^2\right| \le 2\xi$ and $\left|-B_{i+1}^1 + B_{i+1}^2\right| \le 2\xi$, it holds that

$$|\hat{\alpha}_{i+1} - \breve{\alpha}(\hat{p}_i)| \le K_{29}\left(\delta_i + \frac{\xi}{\delta_i} + \xi\right).$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left\{|\hat{\alpha}_{i+1} - \breve{\alpha}(\hat{p}_i)| \le K_{29}\left(\delta_i + \frac{\xi}{\delta_i} + \xi\right)\right\} &\ge \mathbb{P}\left\{\left|B_{i+1}^1 + B_{i+1}^2\right| \le 2\xi, \left|-B_{i+1}^1\right| + \left|B_{i+1}^2\right| \le 2\xi\right\} \\
&\ge 1 - 8e^{-2I_i\xi^2/(u-l)^2},
\end{aligned}$$

which implies

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1} - \breve{\alpha}(\hat{p}_i)|^2 \le K_{30}\left(\delta_i^2 + \frac{\xi^2}{\delta_i^2} + \xi^2\right)\right\} \ge 1 - 8e^{-2I_i\xi^2/(u-l)^2}. \qquad (52)$$

From (49) we have

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|\le K_{31}\left(\delta_i+\frac{\xi}{\delta_i}\right)\right\}\ge 1-4e^{-2I_i\xi^2/(u-l)^2},$$

which implies

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2\le K_{32}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}\right)\right\}\ge 1-4e^{-2I_i\xi^2/(u-l)^2}. \tag{53}$$

Combining(52) and (53), we obtain

$$\begin{aligned}&\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i)|^2+\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2\le K_{33}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\right\} \\ \ge\ & 1-12e^{-2I_i\xi^2/(u-l)^2}.\end{aligned} \tag{54}$$

Let $\xi=(2I_i)^{-\frac{1}{2}}(\log 2I_i)^{\frac{1}{2}}$ in (54), we obtain

$$\begin{aligned}&\mathbb{P}\left\{\ |\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2\le K_{33}\left(I_i^{-\frac{1}{2}}+(2I_i)^{-\frac{1}{2}}(\log 2I_i)+(2I_i)^{-1}(\log 2I_i)\right)\right\} \\ \ge\ & 1-\frac{8}{I_i}.\end{aligned}$$

This implies

$$\begin{aligned}&\mathbb{P}\left\{\ |\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|\le(3K_{33})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|\le(3K_{33})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\right\} \\ \ge\ & 1-\frac{8}{I_i}.\end{aligned}$$

For convenience, we define the event $\mathcal{A}_4$ by

$$\mathcal{A}_4=\left\{\omega:|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|\le(3K_{33})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|\le(3K_{33})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\right\}. \tag{55}$$

Then by (55) one has

$$\mathbb{P}(\mathcal{A}_4^c)\le\frac{8}{I_i}.$$

Define

$$i^*=\max\left\{\log_v\frac{e}{2I_0},\ \min\left\{i\ \middle|\ (3K_{33})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}<\min_{p\in\mathcal{P}}\breve{\beta}(p)\right\}\right\},$$

where we need i^* to be no less than $\log_v\frac{e}{2I_0}$ to ensure that $(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}$ is decreasing on $i\ge i^*$. When $i\ge i^*$, it follows that $\hat{\beta}_{i+1}>0$ on $\mathcal{A}_4$. Therefore

$$\mathbb{P}(\hat{\beta}_{i+1}>0)\ge 1-\frac{8}{I_i}.$$

Lemma A3 is thus proved. ☐

Lemma A4. *There exists a positive constant K_{34} such that*

$$\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}D_t-\lambda(\hat{p}_i)\right|^2+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}D_t-\lambda(\hat{p}_i+\delta_i)\right|^2\right]$$
$$\leq \big(K_{34}+4(u-l)^2\big)I_i^{-\frac{1}{2}}.$$

Proof. Starting from (54), one has

$$\mathbb{P}\left\{\left(\frac{K_{35}}{\delta_i^2}+K_{36}\right)^{-1}\left(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2-K_{33}\delta_i^2\right)\geq\xi^2\right\}<12e^{-2I_i\xi^2/(u-l)^2}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{K_{35}}{\delta_i^2}+K_{36}\right)^{-1}\left(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2-K_{33}\delta_i^2\right)\right]\\
=\;&\left(\frac{K_{35}}{\delta_i^2}+K_{36}\right)^{-1}\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2\right]-\left(\frac{K_{35}}{\delta_i^2}+K_{36}\right)^{-1}K_{33}\delta_i^2\\
\leq\;&\int_0^{+\infty}12e^{-2I_i\xi/(u-l)^2}d\xi\\
=\;&\frac{6(u-l)^2}{I_i}.
\end{aligned}$$

Hence one has

$$\begin{aligned}
&\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2\right]\\
\leq\;&\left(\frac{6(u-l)^2}{I_i}+\left(\frac{K_{35}}{\delta_i^2}+K_{36}\right)^{-1}K_{33}\delta_i^2\right)\left(\frac{K_{35}}{\delta_i^2}+K_{36}\right)\\
\leq\;&K_{37}I_i^{-\frac{1}{2}}. \qquad (56)
\end{aligned}$$

Because

$$\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}D_t-\lambda(\hat{p}_i)=\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}\epsilon_t$$

and

$$\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}D_t-\lambda(\hat{p}_i+\delta_i)=\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}\epsilon_t,$$

by (51), we have

$$\mathbb{P}\left\{\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}D_t-\lambda(\hat{p}_i)\right|>\xi\right\}\leq 2e^{-2I_i\xi^2/(u-l)^2},$$
$$\mathbb{P}\left\{\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}D_t-\lambda(\hat{p}_i+\delta_i)\right|>\xi\right\}\leq 2e^{-2I_i\xi^2/(u-l)^2}.$$

Therefore, we have

$$\mathbb{P}\left\{\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|^2 > 2\xi^2\right\} \le 4e^{-2I_i\xi^2/(u-l)^2},$$

which leads to

$$\begin{aligned}
& \mathbb{E}\left[\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|^2\right] \\
\le & \int_0^{+\infty} 4e^{-I_i\xi/(u-l)^2} \\
= & \frac{4(u-l)^2}{I_i}. \qquad (57)
\end{aligned}$$

Combining (56) and (57), we have

$$\begin{aligned}
& \mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|^2\right] \\
& \le (K_{37}+4(u-l)^2)I_i^{-\frac{1}{2}}.
\end{aligned}$$

This completes the proof of Lemma A4. □

Proof of Proposition 3. To bound $\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\right|^2\right]$, by Lemma A3 and Lemma A4, when i is large enough (greater than or equal to i^* defined in the proof of Lemma A3), for some positive constants K_{38}, and K_{39} one has

$$\begin{aligned}
& \mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\right|^2\right] \\
\le\ & \mathbb{E}\left[K_1^2\left(\left|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}\right| + \left|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}\right| + \left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right| + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|\right)^2\right] \\
& \qquad + \frac{8}{I_i}\big(p^h-p^l\big)^2 \\
\le\ & \mathbb{E}\left[K_{38}\left(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|^2\right)\right] \\
& \qquad + \frac{8}{I_i}\big(p^h-p^l\big)^2 \\
\le\ & K_{39}I_i^{-\frac{1}{2}}, \qquad (58)
\end{aligned}$$

where the first inequality follows from the Lipschitz condition (10) in Assumption 1 (vii), and the third inequality follows from Lemma A4. □

Proof of Theorem 1: Convergence of $\hat{p}_{i+1}$ and $\hat{p}_{i+1}+\delta_{i+1}$. We proceed as follows:

$$\begin{aligned}
& \mathbb{E}\big[(p^*-\hat{p}_{i+1})^2\big] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|p^*-\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\Big|\Big)^2\Big] \\
\leq\ & \mathbb{E}\Big[\Big(\gamma|p^*-\hat{p}_i| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\Big|\Big)^2\Big] \\
\leq\ & \left(\frac{1+\gamma^2}{2}\right)\mathbb{E}\left[(p^*-\hat{p}_i)^2\right] + K_{40}\mathbb{E}\left[\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\Big|^2\right] \\
& \qquad + K_{41}\mathbb{E}\left[\Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\Big|^2\right], \qquad (59)
\end{aligned}$$

where the first inequality follows from the expansion in (12), the second inequality follows from Proposition 1, and the third inequality is justified by $\gamma < 1$ in Proposition 1, and the last inequality holds for some appropriately chosen K_{40} and K_{41} because of the inequality $(a+b)^2 \leq 2(a^2+b^2)$ for any real numbers a and b.

To bound $\mathbb{E}\big[\big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\big|^2\big]$ in (59), by Proposition 2 one has, for some constant K_{42},

$$\mathbb{E}\left[\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)\Big|^2\right] \leq K_{42}^2\int_0^{+\infty} 5e^{-4I_i\xi^2}d\xi = \frac{5\pi^{\frac{1}{2}}K_{42}^2}{4I_i^{\frac{1}{2}}}. \qquad (60)$$

Substituting (60) and (58) into (59), one has

$$\mathbb{E}\big[(p^*-\hat{p}_{i+1})^2\big] \leq \left(\frac{1+\gamma^2}{2}\right)\mathbb{E}\left[(p^*-\hat{p}_i)^2\right] + K_{43}I_i^{-\frac{1}{2}}.$$

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

$$\mathbb{E}\big[(\hat{p}_{i+1}-p^*)^2\big] \leq \theta^i(\hat{p}_1-p^*)^2 + K_{43}\sum_{j=0}^{i-1}\theta^j I_{i-j}^{-\frac{1}{2}} \leq K_{44}(v^{-\frac{1}{2}})^i\sum_{j=0}^{i-1}\theta^j(v^{\frac{1}{2}})^j.$$

We choose $v>1$ that satisfies $\theta v^{\frac{1}{2}}<1$, then there exists a positive constant K_{45} such that $\sum_{j=0}^{i-1}\theta^j(v^{\frac{1}{2}})^j \leq K_{45}$, therefore, for some constants K_{46} and K_{47},

$$\mathbb{E}\big[(\hat{p}_{i+1}-p^*)^2\big] \leq K_{46}(v^{-\frac{1}{2}})^i \leq K_{47}I_i^{-\frac{1}{2}}.$$

Moreover, we have, for some positive constant K_{48},

$$\mathbb{E}\left[(\hat{p}_{i+1}+\delta_{i+1}-p^*)^2\right] \leq 2\mathbb{E}\left[(\hat{p}_{i+1}-p^*)^2\right] + 2\delta_{i+1}^2 \leq K_{48}I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i\to\infty.$$

Because mean-square convergence implies convergence in probability, this shows that the pricing decisions from DDA converge to p^* in probability. □

Proof of Proposition 4. For $p \in \mathcal{P}$, the optimal solution for bridging problem B1 is the same as (24), $\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$. Thus

$$\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p} F^{-1}\left(\frac{b}{b+h}\right). \tag{61}$$

For given $p \in \mathcal{P}$, we follow (27) to define $\tilde{y}^u_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$ as the unconstrained optimal order-up-to level for problem B2 on $\mathbb{R}_+$, then it can be verified that

$$\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p} \ \min\left\{e^{\epsilon_j} : \ \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \le e^{\epsilon_j}\right\} \ge \frac{b}{b+h}\right\}, \tag{62}$$

and, similar to (28), we have

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

It is seen that

$$\left|\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) - \tilde{y}_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)\right| \le \left|\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) - \tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)\right|. \tag{63}$$

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

$$\begin{aligned}
& \mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right) e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right) - \frac{b}{b+h} \le -z\right\} \qquad (64)\\
= \ & \mathbb{P}\left\{\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right) e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})} \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\}\\
\le \ & \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\} \ge \frac{b}{b+h}\right\}\\
= \ & \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\} - \left(\frac{b}{b+h} - z\right) \ge z\right\},
\end{aligned}$$

where the first inequality follows from (62). Since $\mathbb{E}\left[\mathbb{1}\left\{e^{\epsilon_t} \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\}\right] = \frac{b}{b+h} - z$, we apply Hoeffding inequality to obtain

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

Combining this with (61) and (64), we obtain

$$\begin{aligned}
\mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right) e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right) - F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right) e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right) \le -z\right\}\\
\le e^{-4I_i z^2}.
\end{aligned} \tag{65}$$

Similarly, we have

$$\mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)\geq z\right\} \leq e^{-4I_i z^2}. \tag{66}$$

From Assumption 1 (vi), the probability density function $f(\cdot)$ of e^{ϵ_t} satisfies $r = \min\{f(x), x \in [l,u]\} > 0$. From calculus, it is known that, for any $x < y$, there exists a number $z \in [x, y]$ such that $F(y) - F(x) = f(z)(y-x) \geq r(y-x)$. Applying (65) and (66), for any $\xi > 0$, we obtain

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

Let $K_{49} = \max_{\hat{p}_i\in\mathcal{P},\hat{p}_{i+1}\in\mathcal{P}} \frac{1}{r}e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}$, then $K_{49} > 0$. We have

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

and Proposition 4 follows from the inequality above and (63). □

Proof of Theorem 1: Convergence of y_t. To bound the fifth term of the right hand side of (16), we follow a similar approach as that in establishing Proposition 3. By Lipschitz condition (11) and Lemma A4, and following similar analyses as in the proof of Proposition 3, we obtain, for some constant K_{50},

$$\begin{aligned}
& \mathbb{E}\left[\left|\tilde{y}_{i+1}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}) - \hat{y}_{i+1,1}\right|^2\right] \\
\leq\ & 4K_1^2\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|^2\right] \\
& + \frac{8(y^h - y^l)^2}{I_i} \\
\leq\ & K_{50}I_i^{-\frac{1}{2}}.
\end{aligned}$$

Summarizing the analyses above we obtain, for some constants K_{51} and K_{52},

$$\mathbb{E}\left[\left(y^* - \hat{y}_{i+1,1}\right)^2\right] \leq K_{51}\mathbb{E}\left[\,|\,p^* - \hat{p}_{i+1}\,|^2 + |\,p^* - \hat{p}_i\,|^2\,\right] + K_{51}I_i^{-\frac{1}{2}} \leq K_{52}I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty,$$

where the second inequality follows from the convergence rate of the pricing decisions in (15).

Similarly, we obtain

$$\mathbb{E}\left[\left(y^* - \hat{y}_{i+1,2}\right)^2\right] \le K_{53} I_i^{-\frac{1}{2}} \to 0, \qquad \text{as } i \to \infty.$$

Convergence of $\hat{y}_{i+1,1}$ and $\hat{y}_{i+1,2}$ is thus obtained.

To deal with the overshooting problem, since $\tilde{\lambda}(p^h)l \le \tilde{D}_t \le \tilde{\lambda}(p^l)u$, it follows from Hoeffding inequality that for any $\zeta > 0$,

$$\mathbb{P}\Bigg\{ \sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \ge \mathbb{E}\Bigg[\sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \Bigg] - \zeta \Bigg\} \ge 1 - \exp\left(- \frac{2\zeta^2}{I_i\big(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\big)^2} \right). \tag{67}$$

Letting $\zeta = \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}} (\log I_i)^{\frac{1}{2}}$ in (67), then we obtain

$$\mathbb{P}\Bigg\{ \sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \ge I_i \mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] - \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}} (\log I_i)^{\frac{1}{2}} \Bigg\} \ge 1 - \frac{1}{I_i^2}. \tag{68}$$

By Assumption 1(iv), we have $\mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] \ge \min_{p\in\mathcal{P}} \mathbb{E}[D_t(p)] > 0$. This implies, when i is large enough, we will have

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

Define event

$$\mathcal{A}_5 = \left\{ \omega : \sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \ge \frac{1}{2} I_i \mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] \right\},$$

then it follows from (68) that, when i is large enough,

$$\mathbb{P}(\mathcal{A}_5) \ge 1 - \frac{1}{I_i^2}.$$

Note that when i is large enough, $\frac{1}{2} I_i \mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] > y^h - y^l$, which means that on the event $\mathcal{A}_5$, the cumulative demand during $\{t_i + I_i + 1, \ldots, t_i + 2I_i\}$ is high enough to consume all the initial on-hand inventory at the beginning of period $t_i + I_i + 1$ and $\hat{y}_{i,2}$ will be achieved.

Therefore, by definition of event $\mathcal{A}_1$ in the main paper, we have

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

The rest of proofs follow as explained in the main paper, and convergence of actual inventory levels y_t can thus be proved. □

Proof of Proposition 5. To develop an upper bound for $G(p^*, y^*) - G(\hat{p}_i, y_t)$, we first apply Taylor expansion on $G(p, \overline{y}(e^{\lambda(p)})$ at point p^*. Using the fact that the first order derivative vanishes at $p = p^*$

and the assumption that the second order derivative is bounded (Assumption 1 (iii)), we obtain, for some constant $K_{54} > 0$, that

$$G\big(p^*, \overline{y}\big(e^{\lambda(p^*)}\big)\big) - G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) \leq K_{54}(p^* - \hat{p}_i)^2. \tag{69}$$

Noticing that $\overline{y}\big(e^{\lambda(\hat{p}_i)}\big)$ maximizes the concave function $G\big(\hat{p}_i, y\big)$ for given $\hat{p}_i$, we apply Taylor expansion with respect to y at point $y = \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)$ to yield that, for some constant K_{55},

$$G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) - G(\hat{p}_i, y_t) \leq K_{55}\big(\overline{y}\big(e^{\lambda(\hat{p}_i)}\big) - y_t\big)^2. \tag{70}$$

In addition, we have

$$\begin{aligned}
& \mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i)}) - y_t)^2\right] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|\overline{y}(e^{\lambda(\hat{p}_i)}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i})\Big| + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big| \\
& \quad + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big| + \Big|\tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \hat{y}_{i,1}\Big| \\
& \quad + \Big|\hat{y}_{i,1} - y_t\Big|\Big)^2\Big] \\
\leq\ & 2\mathbb{E}\Big[\Big(\Big|\overline{y}(e^{\lambda(\hat{p}_i)}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i})\Big| + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big| \\
& \quad + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big| + \Big|\tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \hat{y}_{i,1}\Big|\Big)^2\Big] \\
& + 2\mathbb{E}\Big[\Big|\hat{y}_{i,1} - y_t\Big|^2\Big] \\
\leq\ & K_{56}\mathbb{E}\Big[\Big|\overline{y}(e^{\lambda(\hat{p}_i)}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i})\Big|^2 + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_i}) - \overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big|^2 \\
& \quad + \Big|\overline{y}(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i})\Big|^2 + \Big|\tilde{y}_i(e^{\breve{\alpha}(\hat{p}_{i-1})-\breve{\beta}(\hat{p}_{i-1})\hat{p}_i}) - \hat{y}_{i,1}\Big|^2\Big] \qquad (71) \\
& + 2\mathbb{E}\Big[\Big|\hat{y}_{i,1} - y_t\Big|^2\Big]. \qquad (72)
\end{aligned}$$

The expectation in (71) is similar to the right hand side of (16) except that $i+1$ is replaced by i. Thus, using the same analysis as that for (16), we obtain

$$\mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i)}) - \hat{y}_{i,1})^2\right] \leq K_{57} I_{i-1}^{-\frac{1}{2}} \tag{73}$$

for some constant K_{57}.

The expectation in (72) is bounded as the following:

$$\mathbb{E}\Big[\Big|\hat{y}_{i,1} - y_t\Big|^2\Big] \leq \mathbb{E}\Big[\Big|\hat{y}_{i,1} - y^*\Big|^2 + \Big|y^* - y_t\Big|^2\Big] \leq K_{58} I_{i-1}^{-\frac{1}{2}}, \tag{74}$$

where the second inequality follows from (17) and (18).

Therefore, combining (73) and (74), one has

$$\mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i)}) - y_t)^2\right] \leq K_{59} I_{i-1}^{-\frac{1}{2}}. \tag{75}$$

Applying the results above, we obtain, for some constants K_{60}, K_{61}, and K_{62}, that

$$\begin{aligned}
& \mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i, y_t)\right] \\
= \ & \mathbb{E}\bigg[\Big(G\big(p^*, \overline{y}\big(e^{\lambda(p^*)}\big)\big) - G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big)\Big) + \Big(G\big(\hat{p}_i, \overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big) - G(\hat{p}_i, y_t)\Big)\bigg] \\
\leq \ & K_{60}\Big(\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + \mathbb{E}\Big[\big(\overline{y}(e^{\lambda(\hat{p}_i)}) - y_t\big)^2\Big]\Big) \\
\leq \ & K_{61}\left(K_{10} I_{i-1}^{-\frac{1}{2}} + K_{86} I_{i-1}^{-\frac{1}{2}}\right) \\
= \ & K_{62} I_{i-1}^{-\frac{1}{2}},
\end{aligned}$$

where the first inequality follows from (69) and (70), and the second inequality follows from the convergence rate of pricing decisions (15) and (75).

Similarly, we establish for some constants K_{63}, K_{64} and K_{65}, that

$$\begin{aligned}
\mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\right] \ & \leq \ K_{63}\left(\mathbb{E}\left[(p^* - \hat{p}_i - \delta_i)^2\right] + \mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i + \delta_i)}) - y_t)^2\right]\right) \\
& \leq \ K_{64}\left(\mathbb{E}\left[2(p^* - \hat{p}_i)^2 + 2\delta_i^2\right] + K_{53} I_{i-1}^{-\frac{1}{2}}\right) \\
& \leq \ K_{65} I_{i-1}^{-\frac{1}{2}}.
\end{aligned}$$

Note that, as seen from Proposition 3, these results hold when i is greater than or equal to some number i^*. □

Proof of Theorem 2. We have, for some constants K_{66}, K_{67} and K_{68},

$$\begin{aligned}
& \mathbb{E}\bigg[\sum_{t=1}^{T}\big(G(p^*, y^*) - G(p_t, y_t)\big)\bigg] \\
\leq \ & \sum_{i=i^*+1}^{n} K_{66} I_{i-1}^{-\frac{1}{2}} I_i + \sum_{i=1}^{i^*}\left(\sum_{t=t_i+1}^{t_i+I_i}\big(G(p^*, y^*) - G(\hat{p}_i, y_t)\big) + \sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\big)\right) \\
= \ & \sum_{i=i^*+1}^{n} K_{66} I_{i-1}^{\frac{1}{2}} + K_{67} \\
\leq \ & K_{66}\sum_{i=2}^{n} I_{i-1}^{\frac{1}{2}} + K_{67} \\
= \ & K_{66}\frac{(2I_0)^{\frac{1}{2}} v^{\frac{1}{2}}}{v^{\frac{1}{2}} - 1}\big(v^{\frac{n-1}{2}} - 1\big) + K_{67} \\
\leq \ & K_{66}\frac{(2I_0)^{\frac{1}{2}} v^{\frac{1}{2}}}{v^{\frac{1}{2}} - 1}\big(v^{\log_v(\frac{v-1}{2I_0 v}T+1)+1-1}\big)^{\frac{1}{2}} + K_{67} \\
\leq \ & K_{68} T^{\frac{1}{2}},
\end{aligned}$$

where the first inequality follows from Proposition 5, and $K_{67} = \sum_{i=1}^{i^*} \Big(\sum_{t=t_i+1}^{t_i+I_i} \big(G(p^*, y^*) - G(\hat{p}_i, y_t)\big) + \sum_{t=t_i+I_i+1}^{t_i+2I_i} \big(G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\big)\Big)$. □

Appendix B: Analyses for the Pure Inventory Control Problem

In this Appendix, we modify the data-driven algorithm for the pure inventory control of non-perishable products and prove the convergence rate of its regret. The modified algorithm is straightforward: Since price p is exogenous, it is not a decision. Stage i has I_i periods, and realized demand data are collected from stage i to construct an SAA, which is solved to obtain an order-up-to level $\hat{y}_{i+1}$ for stage $i+1$. We refer to this modified algorithm as MDDA (Modified Data-Driven Algorithm). For $i \geq 1$, let

$$I_i = \lceil I_0 v^i \rceil, \ t_i = \sum_{k=1}^{i-1} I_k \text{ with } t_1 = 0,$$

and there are in total $n = \left\lceil \log_v \left(\frac{v-1}{I_0 v} T + 1 \right) \right\rceil$ stages.

Modified Data-Driven Algorithm (MDDA)

Initialization. Choose $v > 1$, $\rho > 0$, $I_0 > 0$, and $\hat{y}_1$. Compute $I_1 = \lfloor I_0 v \rfloor$.

Main procedure. Repeat Steps 1 and 2 for $i = 1, \ldots, n$.

Step 1. Setting order-up-to levels for stage i. In iteration i, raise the inventory levels to

$$y_t = \max\{\hat{y}_i, x_t\}, \qquad t = t_i + 1, \ldots, t_i + I_i.$$

Step 2. Defining and maximizing the proxy profit function, denoted by $G^{DD}_{i+1}(y)$. Let $\tilde{D}_t = \tilde{D}_t(p)$ be the demand realizations for $t = t_i + 1, \ldots, t_i + I_i$. Define

$$G^{DD}_{i+1}(y) = p\left(\frac{1}{I_i} \sum_{t=t_i+1}^{t_i+I_i} \tilde{D}_t \right) - \left\{ \frac{1}{I_i} \sum_{t=t_i+1}^{t_i+I_i} \left(h\left(y - \tilde{D}_t\right)^+ + b\left(\tilde{D}_t - y\right)^+ \right) \right\}.$$

Then the data-driven optimization is defined by

Problem DD:

$$\max_{y \in \mathcal{Y}} G^{DD}_{i+1}(y) = p\left(\frac{1}{I_i} \sum_{t=t_i+1}^{t_i+I_i} \tilde{D}_t \right) - \min_{y \in \mathcal{Y}} \left\{ \frac{1}{I_i} \sum_{t=t_i+1}^{t_i+I_i} \left(h\left(y - \tilde{D}_t\right)^+ + b\left(\tilde{D}_t - y\right)^+ \right) \right\}.$$

Solve problem DD and set the inventory level to

$$\hat{y}_{i+1} = \arg\max_{y \in \mathcal{Y}} G^{DD}_{i+1}(y).$$

Regret of MDDA. We want to show that, there exists a positive constant K_{69} such that

$$R(MDDA,T) \leq \frac{K_{69}\log T}{T}.$$

We have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\big(G(y^*)-G(y_t)\big)\right] \\
\leq\;& \mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=t_i+1}^{t_i+I_i}\big(G(y^*)-G(y_t)\big)\right] \\
\leq\;& K_{70}\sum_{i=1}^{n}\sum_{t=t_i+1}^{t_i+I_i}\mathbb{E}\left[\big(y^*-y_t\big)^2\right], \qquad (76)
\end{aligned}$$

where n, again, satisfies $\log_v\left(\frac{v-1}{I_0 v}T+1\right) \leq n < \log_v\left(\frac{v-1}{I_0 v}T+1\right)+1$.

Similar as the development of (68), we have

$$\mathcal{A}_6 = \left\{\omega : \sum_{t=t_{i-1}+1}^{t_i+I_{i-1}} \tilde{D}_t \geq I_{i-1}\mathbb{E}\left[\tilde{D}_{t_{i-1}+1}\right] - (\tilde{\lambda}(p)u - \tilde{\lambda}(p)l)I_{i-1}^{\frac{1}{2}}\left(\log I_{i-1}\right)^{\frac{1}{2}}\right\},$$

and

$$\mathbb{P}(\mathcal{A}_6) \geq 1 - \frac{1}{I_{i-1}^2}.$$

In event $\mathcal{A}_6$, $\hat{y}_{i-1}$ will be achieved during $\{t_{i-1}+1,...,t_{i-1}+I_{i-1}\}$.

Thus we have the following,

$$\begin{aligned}
&\mathbb{E}\left[\big(y^*-y_t\big)^2\right] \\
=\;& \mathbb{P}(\mathcal{A}_6)\mathbb{E}\left[\big(y^*-y_t\big)^2\Big|\mathcal{A}_6\right] + (1-\mathbb{P}(\mathcal{A}_6))\mathbb{E}\left[\big(y^*-y_t\big)^2\Big|(\mathcal{A}_6)^C\right] \\
\leq\;& \mathbb{E}\left[\max\left\{\big(y^*-\hat{y}_{i-1}\big)^2, \big(y^*-\hat{y}_i\big)^2\right\}\right] + \frac{(y^h-y^l)^2}{I_{i-1}^2}. \qquad (77)
\end{aligned}$$

Similar as developing Proposition 4, for any $\xi > 0$, one has

$$\mathbb{P}\Big\{\,|y^*-\hat{y}_{i-1}| \geq K_{71}\xi\Big\} \leq 2e^{-4I_{i-2}\xi^2},$$

and

$$\mathbb{P}\Big\{\,|y^*-\hat{y}_i| \geq K_{71}\xi\Big\} \leq 2e^{-4I_{i-1}\xi^2}.$$

Therefore,

$$\mathbb{P}\Big\{\max\left\{|y^*-\hat{y}_{i-1}|, |y^*-\hat{y}_i|\right\} \geq K_{71}\xi\Big\} \leq 4e^{-4I_{i-2}\xi^2},$$

which implies

$$\mathbb{P}\Big\{\max\Big\{|y^*-\hat{y}_{i-1}|^2, |y^*-\hat{y}_i|^2\Big\} \geq K_{71}^2\xi^2\Big\} \leq 4e^{-4I_{i-2}\xi^2}.$$

Therefore,

$$\mathbb{E}\left[\max\left\{\left(y^*-\hat{y}_{i-1}\right)^2, \left(y^*-\hat{y}_i\right)^2\right\}\right] \leq \frac{K_{72}}{I_{i-2}}. \tag{78}$$

Plugging (78) back into (77) we have that, for $t \in \{t_i+1, ..., t_i+2I_i\}$,

$$\mathbb{E}\left[\left(y^*-y_t\right)^2\right] \leq \frac{K_{73}}{I_{i-2}}. \tag{79}$$

Plugging (79) into (76), one has

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\left(G(y^*)-G(y_t)\right)\right] \\
\leq\ & K_{72}\sum_{i=1}^{n}\sum_{t=t_i+1}^{t_i+I_i}\mathbb{E}\left[\left(y^*-y_t\right)^2\right] \\
\leq\ & K_{72}\sum_{i=1}^{n}K_{66} \\
\leq\ & K_{69}\log T,
\end{aligned}$$

which leads to

$$R(MDDA, T) \leq \frac{K_{69}\log T}{T}.$$

Appendix C: A Technical Comparison with Besbes and Zeevi (2015)

In this subsection, we elaborate on the difference between our analysis and that of Besbes and Zeevi (2015). Since mean square convergence implies convergence in probability, to prove $p_t \to p^*$ in probability of Theorem 1, it suffices to show

$$\mathbb{E}\big[(p^*-\hat{p}_{i+1})^2\big] \to 0, \quad \mathbb{E}\big[(p^*-(\hat{p}_{i+1}+\delta_{i+1}))^2\big] \to 0, \quad \text{as } i \to \infty. \tag{80}$$

In our proof of inventory decisions $y_t \to y^*$ in probability, only showing

$$\mathbb{E}[(y^*-\hat{y}_{i+1,1})^2] \to 0, \quad \mathbb{E}[(y^*-\hat{y}_{i+1,2})^2] \to 0, \quad \text{as } i \to \infty \tag{81}$$

is not enough due to the overshooting (beyond target inventory levels) of non-perishable products. Our approach is to first prove (81), and then prove

$$\mathbb{E}[(y^*-y_t)^2] \to 0, \quad \text{as } t \to \infty$$

by taking care of the overshooting issue.

In their pricing optimization setting, Besbes and Zeevi (2015) also prove the convergence result (80). Their p^* and $\hat{p}_{i+1}$ are the maximizers of the following two optimization problems:

$$\max_{p\in\mathcal{P}} Q(p,\lambda(p)) = p\lambda(p), \quad \text{and} \quad \max_{p\in\mathcal{P}} Q(p,\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p) = p(\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p).$$

As Besbes and Zeevi (2015) indicate, their analysis extends to a more general function $Q(p,\lambda(p))$ in which $Q(p,\lambda(p))$ is a *known deterministic function* of p and $\lambda(p)$.

However, this argument does not apply to our setting. This is because, our objective function in (4) of the main paper is

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

Since the distribution of ϵ is not known *a priori*, $Q(p,e^{\lambda(p)})$ is *not* a known function of p and $\lambda(p)$. As a matter of fact, we cannot even find an i.i.d. sample of the random error to estimate $Q(\cdot,\cdot)$ for an arbitrary price p.

In our approach we construct an estimator $Q^{DD}_{i+1}(\cdot,\cdot)$ using centered samples discussed earlier, and bound the difference between the proxy objective function constructed using centered samples and that obtained from true error samples.

Appendix D: Analyses for Additive Demand Model

In the case with additive demand, for each period t we have $D_t(p_t) = \lambda(p_t) + \epsilon_t$, where ϵ_t is distributed on $[l,u]$ with CDF $F(\cdot)$ and PDF $f(\cdot)$, and without loss of generality, assume $\mathbb{E}[\epsilon_t] = 0$. In the multiplicative demand case, we need to take logarithm of the demand to transform it into an additive form, but in the additive demand case, there is no need to do any transformations. In the following, we will define problem CI$'$, problem DD$'$, and bridging problems B1$'$-B3$'$ in parallel with the multiplicative demand case. Then we will modify the algorithm, prove Lemma 1', Propositions $1'-5'$, and Theorems $1'$ and $2'$, also in parallel with the multiplicative demand case. It can be seen that, all the analyses and results of the multiplicative demand case can be applied to the additive demand case with minor modifications.

We define the single period objective function

$$G(p,y) = p\lambda(p) - \Big\{h\mathbb{E}\big[y-\lambda(p)-\epsilon\big]^{+} + b\mathbb{E}\big[\lambda(p)+\epsilon-y\big]^{+}\Big\}, \tag{82}$$

and

Problem CI$'$:

$$\max_{p\in\mathcal{P}} Q(p,\lambda(p)) := \max_{p\in\mathcal{P}}\Big\{p\lambda(p) - \min_{y\in\mathcal{Y}}\Big\{h\mathbb{E}\big[y-\lambda(p)-\epsilon\big]^{+} + b\mathbb{E}\big[\lambda(p)+\epsilon-y\big]^{+}\Big\}\Big\}. \tag{83}$$

In (83), the inner optimization problem (minimization) determines the optimal order-up-to level that minimizes the expected holding and backlog cost for a given price p, and we denote it by $\overline{y}\left(\lambda(p)\right)$. The outer optimization solves for the optimal price p. Because (p^*, y^*) is the optimal solution for (83), they satisfy $y^* = \overline{y}(\lambda(p^*))$.

The DDA algorithm can also be applied to the additive demand case with minor modifications. To be specific, Initialization and Step 1 remain the same. In Step 2, there is no need to take logarithm of the demand, and here D_t is just the realization of demand in period t. Step 3 is changed to the following:

Step 3. Defining and maximizing the proxy profit function, denoted by $G^{DD}_{i+1}(p, y)$. The data-driven optimization problem is

Problem DD′:

$$\max_{(p,y)\in\mathcal{P}\times\mathcal{Y}} G^{DD}_{i+1}(p, y) = \max_{p\in\mathcal{P}} Q^{DD}_{i+1}\big(p, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p\big), \tag{84}$$

where

$$\begin{aligned} G^{DD}_{i+1}\big(p, y\big) &= p\big(\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p\big) \\ &\quad - \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\left(h\left(y - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p + \eta_t)\right)^+ + b\left((\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p + \eta_t) - y\right)^+\right), \end{aligned}$$

and

$$Q^{DD}_{i+1}\big(p, \hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p\big) = \min_{y\in\mathcal{Y}} \; G^{DD}_{i+1}\big(p, y\big).$$

If $\hat{\beta}_{i+1} > 0$, then solve problem DD and set the first pair of price and inventory level to

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

otherwise, set

$$(\hat{p}_{i+1}, \hat{y}_{i+1,1}) = \left(\frac{p^l + p^h}{2}, \frac{y^l + y^h}{2}\right).$$

Set $\hat{p}_{i+1,2} = \hat{p}_{i+1} + \delta_{i+1}$ (in case $\hat{p}_{i+1} + \delta_{i+1} \notin \mathcal{P}$, set $\hat{p}_{i+1,2} = \hat{p}_{i+1} - \delta_{i+1}$), and

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

Note that if $\hat{\beta}_{i+1} > 0$, then $G^{DD}_{i+1}(p, y)$ is jointly concave in (p, y), and as we will prove later, $\hat{\beta}_{i+1} > 0$ with a high probability.

Next we define $\breve{\alpha}(z)$ and $\breve{\beta}(z)$ as follows (in the multiplicative case, $\lambda(\cdot)$ is the logarithm of the demand curve $\tilde{\lambda}(cdot)$, but here $\lambda(\cdot)$ is the demand curve),

$$\breve{\alpha}(z) = \lambda(z) - \lambda'(z)z, \qquad \breve{\beta}(z) = -\lambda'(z), \qquad z \in \mathcal{P}. \tag{85}$$

First, for parameters α and $\beta > 0$, we introduce bridging problem B1$'$ defined by

Bridging Problem B1$'$:

$$\max_{p\in\mathcal{P}} \overline{Q}(p, \alpha-\beta p) := \max_{p\in\mathcal{P}} \left\{ p(\alpha-\beta p) - \min_{y\in\mathcal{Y}} \left\{ h\mathbb{E}\Big[y - (\alpha - \beta p + \epsilon)\Big]^+ + b\mathbb{E}\Big[\alpha - \beta p + \epsilon - y\Big]^+ \right\}\right\}. \tag{86}$$

We let $\overline{p}(\alpha,\beta)$ denote the optimal price for problem B1$'$, and for given $p \in \mathcal{P}$, we let $\overline{y}(\alpha-\beta p)$ denote its optimal order-up-to level.

The second bridging problem, B2$'$, is defined for each iteration i of the DDA algorithm, and with α and $\beta > 0$, it is given by

Bridging Problem B2$'$:

$$\max_{p\in\mathcal{P}} \tilde{Q}_{i+1}(p, \alpha-\beta p) := \max_{p\in\mathcal{P}} \Bigg\{ p\big(\alpha-\beta p\big) \tag{87}$$
$$- \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big(h\big(y - (\alpha-\beta p + \epsilon_t)\big)^+ + b\big((\alpha - \beta p + \epsilon_t) - y\big)^+ \Big) \right\} \Bigg\}.$$

We let $\tilde{p}_{i+1}(\alpha,\beta)$ denote the optimal price and let $\tilde{y}_{i+1}(\alpha-\beta p)$ denote the optimal order-up-to level for problem B2$'$.

The third bridging problem B3$'$ is a variation of problem B2$'$, which replaces the true random error ϵ_t by a biased error ζ_t, $t = t_i+1,\ldots,t_i+2I_i$. That is, for

$$\zeta_{t=t_i+1}^{t_1+I_i} = \big(\zeta_{t_i+1},\ldots,\zeta_{t_i+I_i}\big), \quad \zeta_{t=t_i+I_i+1}^{t_1+2I_i} = \big(\zeta_{t_i+I_i+1},\ldots,\zeta_{t_i+2I_i}\big),$$

and parameters α and $\beta > 0$, we define the third bridging problem B3$'$ by

Bridging Problem B3$'$:

$$\max_{p\in\mathcal{P}} \breve{Q}_{i+1}\Big(p, \alpha-\beta p, \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\Big)$$
$$:= \max_{p\in\mathcal{P}} \left\{ p\big(\alpha-\beta p\big) - \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big(h\big(y - (\alpha-\beta p + \zeta_t)\big)^+ + b\big((\alpha - \beta p + \zeta_t) - y\big)^+ \Big) \right\} \right\}.$$

We denote optimal price of problem B3$'$ by $\breve{p}_{i+1}\big((\alpha,\beta), \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big)$ and its optimal order-up-to level, for given price p, by $\breve{y}_{i+1}\big(\alpha-\beta p, \zeta_{t=t_i+1}^{t_1+I_i}, \zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big)$.

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

(i) $p\lambda(p)$ is unimodal in p for $p \in \big[p^l, p^h\big]$, and admits a unique maximizer in the interior of $\mathcal{P}$.

(ii) $-2 < \frac{\lambda''(p)\lambda(p)}{(\lambda'(p))^2} < 2$ for $p \in \big[p^l, p^h\big]$.

(iii) $G(p, \lambda(p))$ in (82) has bounded second order derivatives with respect to $p \in \mathcal{P}$.

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

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

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

(vii) The functions $\breve{p}_{i+1}\big((\cdot,\cdot),\cdot\big)$ and $\breve{y}_{i+1}\big((\cdot,\cdot),\cdot\big)$ for problem B3′ satisfy the following Lipschitz condition: there exists some constant $K_{74} > 0$ such that

$$\Big|\breve{p}_{i+1}\big((\alpha,\beta),\zeta_{t=t_i+1}^{t_1+I_i},\zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big) - \breve{p}_{i+1}\big((\alpha',\beta'),\zeta_{t=t_i+1}^{t_1+I_i}+m_1\mathbf{1}_{I_i},\zeta_{t=t_i+I_i+1}^{t_1+2I_i}+m_2\mathbf{1}_{I_i}\big)\Big| \tag{88}$$
$$\leq K_{74}\Big(|\alpha-\alpha'|+|\beta-\beta'|+|m_1|+|m_2|\Big),$$

and

$$\Big|\breve{y}_{i+1}\big(\alpha-\beta p,\zeta_{t=t_i+1}^{t_1+I_i},\zeta_{t=t_i+I_i+1}^{t_1+2I_i}\big) - \breve{y}_{i+1}\big(\alpha'-\beta' p,\zeta_{t=t_i+1}^{t_1+I_i}+m_1\mathbf{1}_{I_i},\zeta_{t=t_i+I_i+1}^{t_1+2I_i}+m_2\mathbf{1}_{I_i}\big)\Big| \tag{89}$$
$$\leq K_{74}\Big(|\alpha-\alpha'|+|\beta-\beta'|+|m_1|+|m_2|\Big),$$

where $\mathbf{1}_{I_i}$ is the I_i-dimentional vector with all entries being 1.

(viii) The feasible regions for price, $\mathcal{P}$, and for order-up-to level, $\mathcal{Y}$, are large enough so that the optimal solutions p^* and $\overline{y}(\lambda(p)$ for problem CI′ over $\mathbb{R}_+$ for given $p \in \mathcal{P}$ fall into $\mathcal{P}$ and $\mathcal{Y}$, respectively; and for given $q \in \mathcal{P}$, the optimal solutions $\overline{p}\big(\breve{\alpha}(q),\breve{\beta}(q)\big)$ and $\overline{y}\big(\breve{\alpha}(q)-\breve{\beta}(q)p\big)$ for given $p \in \mathcal{P}$ for problem B1′ fall into $\mathcal{P}$ and $\mathcal{Y}$, respectively.

Examples that satisfy these conditions for the additive demand model include (a) linear with $\lambda(p) = k - mp$, $m > 0$, (b) exponential with $\lambda(p) = e^{k-mp}, m > 0$, and (c) logit with $\lambda(p) = \frac{e^{k-mp}}{1+e^{k-mp}}, m > 0, e^{k-mp} < 3$ for all $p \in \mathcal{P}$.

For the additive demand case, the DDA algorithm still has the following two results.

Theorem $1'$ **(Policy Convergence)** *Under Assumption* $1'$*, the DDA policy is consistent, i.e.,* $(p_t, y_t) \to (p^*, y^*)$ *in probability as* $t \to \infty$.

Theorem $2'$ **(Regret Convergence Rate)** *Under Assumption* $1'$*, the DDA policy is asymptotically optimal. More specifically, there exists some constant* $K_{75} > 0$ *such that*

$$R(DDA,T) = G(p^*,y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t,y_t)\right] \leq K_{75}T^{-\frac{1}{2}}.$$

To prove Theorems $1'$ and $2'$, statements in Lemma 1 and Propositions 1-5 all hold for the additive demand case, except that $\bar{y}(e^{\alpha-\beta p})$ is replaced by $\bar{y}(\alpha-\beta p)$, and $\tilde{y}_i(e^{\alpha-\beta p})$ is replaced by $\tilde{y}_i(\alpha-\beta p)$. We will call them Lemma $1'$, Propositions $1'-5'$, and will prove them in details shortly. Having proved those results, the rest of the procedures and proving logic in Section 5 all apply analogously to the additive demand case.

Proof of Lemma $1'$**.** For $\beta > 0$, because the objective functions of $B1', B2'$, and $B3'$ are jointly concave in (p, y), after the inner minimization over y, $\overline{Q}(p, \alpha - \beta p)$, $\tilde{Q}_{i+1}(p, \alpha - \beta p)$, and $\breve{Q}_{i+1}(p, \alpha - \beta p)$ are all concave thus unimodal in p. □

Proof of Proposition $1'$**.** First we make the observation that

$$p^* = \overline{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big). \tag{90}$$

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

$$\overline{Q}(p, \lambda(p)) = p\lambda(p) - \min_{y \in \mathcal{Y}} \Big\{ h\mathbb{E}\big[y - \lambda(p) - \epsilon\big]^+ + b\mathbb{E}\big[\lambda(p) + \epsilon - y\big]^+ \Big\}. \tag{91}$$

Then Assumption $1'$(i) implies that $\overline{Q}(p, \lambda(p))$ is unimodal in p. Assuming that $\overline{Q}$ has a unique maximizer and that $\overline{p}(\breve{\alpha}(z), \breve{\beta}(z))$ is the unique optimal solution for problem B1$'$ with parameters $\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$, then (90) follows from Lemma 1 of Besbes and Zeevi (2015) by letting their function G be (91).

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

$$\overline{p}(\alpha, \beta) = \frac{\alpha}{2\beta}.$$

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

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

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

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

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

To prove Proposition $2'$, we need to establish Lemmas A1$'$ and A2$'$ first. For convenience, we let

$$\overline{W}(\alpha - \beta p, y) := h\mathbb{E}\Big[y - (\alpha - \beta p + \epsilon)\Big]^+ + b\mathbb{E}\Big[(\alpha - \beta p + \epsilon) - y\Big]^+,$$

and

$$\tilde{W}_{i+1}(\alpha - \beta p, y) := \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big(h\big(y - (\alpha - \beta p + \epsilon_t)\big)^+ + b\big((\alpha - \beta p + \epsilon_t) - y\big)^+ \Big).$$

Lemma A1′. *There exists a positive constant K_{76} such that, for any $\xi > 0$,*

$$\mathbb{P}\left\{\max_{p\in\mathcal{P}}\left|\overline{W}(\alpha-\beta p,\overline{y}(\alpha-\beta p))-\tilde{W}_{i+1}(\alpha-\beta p,\tilde{y}_{i+1}(\alpha-\beta p))\right|\le K_{76}\xi\right\}\ge 1-4e^{-2I_i\xi^2}.$$

Proof. To simplify the presentation, let $\alpha - \beta p = d$. By triangle inequality, we have

$$\begin{aligned}&\max_{p\in\mathcal{P}}\left|\overline{W}(d,\overline{y}(d))-\tilde{W}_{i+1}(d,\tilde{y}_{i+1}(d))\right|\\ \le\ &\max_{p\in\mathcal{P}}\left|\overline{W}\big(d,\overline{y}(d)\big)-\tilde{W}_{i+1}\big(d,\overline{y}(d)\big)\right|+\max_{p\in\mathcal{P}}\left|\tilde{W}_{i+1}\big(d,\overline{y}(d)\big)-\tilde{W}_{i+1}\big(d,\tilde{y}_{i+1}(d)\big)\right|.\end{aligned}\tag{92}$$

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

For any $p\in\mathcal{P}$ and $y\in\mathcal{Y}$, we define $z=y-d$. Then the optimal z to minimize $\overline{W}(d,d+z)$ is

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

Moreover, we have

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

and

$$\tilde{W}_{i+1}(d,\overline{y}(d))=\tilde{W}_{i+1}(d,d+\overline{z})=\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(\overline{z}-\epsilon_t\big)^+ + b\big(\epsilon_t-\overline{z}\big)^+\Big)\right).\tag{93}$$

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

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

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

$$\mathbb{P}\left\{\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right|>\xi\right\}\le 2e^{-2I_i\xi^2/(\Delta^h-\Delta^l)^2},\tag{94}$$

which deduces to

$$\mathbb{P}\left\{\max_{p\in\mathcal{P}}\left|\overline{W}(\alpha-\beta p,\overline{y}(\alpha-\beta p))-\tilde{W}_{i+1}(\alpha-\beta p,\overline{y}(\alpha-\beta p))\right|>\xi\right\}\le 2e^{-2I_i\xi^2/(\Delta^h-\Delta^l)^2}.\tag{95}$$

This bounds the first term on the right hand side of (92).

To bound the second term in (92), we use

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

to denote the empirical distribution of ϵ_t. For $\theta > 0$, it can be verified that

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

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

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

Combining the two results above we obtain

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

therefore the probability for $\hat{F}(\cdot)$ to be a θ-estimate is at least $1 - 2e^{-2I_i\theta^2}$. Let $\mathcal{A}_7(\theta)$ represent the event that $\hat{F}(\overline{z})$ is a θ-estimate, then the result above states that

$$\mathbb{P}(\mathcal{A}_7(\theta)) \geq 1 - 2e^{-2I_i\theta^2}. \tag{96}$$

For $d \in \mathcal{D}$, let $\tilde{z}_{i+1}(d) = \tilde{y}_{i+1}(d) - d$ and $\tilde{z}^u_{i+1} = \tilde{y}^u_{i+1}(d) - d$, then one has

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

and then

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

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

First, for any given $d \in \mathcal{D}$, if $\overline{z} \leq \tilde{z}^u_{i+1}$, then it follows from (93) that

$$\begin{aligned}\tilde{W}_{i+1}(d, d + \overline{z}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(\epsilon_t - \overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < \epsilon_t\big\} \\ &\qquad + b\big(\epsilon_t - \overline{z}\big)\mathbb{1}\big\{\overline{z} < \epsilon_t \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z} - \epsilon_t\big)\mathbb{1}\big\{\epsilon_t \leq \overline{z}\big\}\Big] \\ &\leq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(\epsilon_t - \overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < \epsilon_t\big\} \\ &\qquad + b\big(\tilde{z}^u_{i+1} - \overline{z}\big)\mathbb{1}\big\{\overline{z} < \epsilon_t \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z} - \epsilon_t\big)\mathbb{1}\big\{\epsilon_t \leq \overline{z}\big\}\Big],\end{aligned} \tag{97}$$

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

$$\begin{aligned}\tilde{W}_{i+1}(d, d+\tilde{z}^u_{i+1}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(\epsilon_t - \tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < \epsilon_t\} \\ &\qquad + h(\tilde{z}^u_{i+1} - \epsilon_t)\mathbb{1}\{\overline{z} < \epsilon_t \le \tilde{z}^u_{i+1}\} + h(\tilde{z}^u_{i+1} - \epsilon_t)\mathbb{1}\{\epsilon_t \le \overline{z}\}\Big] \\ &\ge \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(\epsilon_t - \tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < \epsilon_t\} + h(\tilde{z}^u_{i+1} - \epsilon_t)\mathbb{1}\{\epsilon_t \le \overline{z}\}\Big], \end{aligned} \tag{98}$$

with the inequality obtained by dropping the nonnegative middle term. Consequently when $\overline{z} \le \tilde{z}^u_{i+1}$ we subtract (98) from (97) to obtain

$$\begin{aligned}&\tilde{W}_{i+1}(d, d+\overline{z}) - \tilde{W}_{i+1}(d, d+\tilde{z}^u_{i+1}) \\ \le\ & d\Big(b(\tilde{z}^u_{i+1} - \overline{z})(1 - \hat{F}(\tilde{z}^u_{i+1})) + b(\tilde{z}^u_{i+1} - \overline{z})(\hat{F}(\tilde{z}^u_{i+1}) - \hat{F}(\overline{z})) + h(\overline{z} - \tilde{z}^u_{i+1})\hat{F}(\overline{z})\Big) \\ =\ & d(\tilde{z}^u_{i+1} - \overline{z})\big(-(h+b)\hat{F}(\overline{z}) + b\big) \\ \le\ & d\big(\tilde{z}^u_{i+1} - \overline{z}\big)(b+h)\theta, \end{aligned} \tag{99}$$

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

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

$$\begin{aligned}\tilde{W}_{i+1}(d, d+\overline{z}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(\epsilon_t - \overline{z})\mathbb{1}\{\overline{z} < \epsilon_t\} \\ &\qquad + h(\overline{z} - \epsilon_t)\mathbb{1}\{\tilde{z}^u_{i+1} < \epsilon_t \le \overline{z}\} + h(\overline{z} - \epsilon_t)\mathbb{1}\{\epsilon_t \le \tilde{z}^u_{i+1}\}\Big] \\ &\le \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(\epsilon_t - \overline{z})\mathbb{1}\{\overline{z} < \epsilon_t\} \\ &\qquad + h(\overline{z} - \tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1} < \epsilon_t \le \overline{z}\} + h(\overline{z} - \epsilon_t)\mathbb{1}\{\epsilon_t \le \tilde{z}^u_{i+1}\}\Big], \end{aligned} \tag{100}$$

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

$$\begin{aligned}\tilde{W}_{i+1}(d, d+\tilde{z}^u_{i+1}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(\epsilon_t - \tilde{z}^u_{i+1})\mathbb{1}\{\overline{z} < \epsilon_t\} \\ &\qquad + b(\epsilon_t - \tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1} < \epsilon_t \le \overline{z}\} + h(\tilde{z}^u_{i+1} - \epsilon_t)\mathbb{1}\{\epsilon_t \le \tilde{z}^u_{i+1}\}\Big] \\ &\ge \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(\epsilon_t - \tilde{z}^u_{i+1})\mathbb{1}\{\overline{z} < \epsilon_t\} + h(\tilde{z}^u_{i+1} - \epsilon_t)\mathbb{1}\{\epsilon_t \le \tilde{z}^u_{i+1}\}\Big], \end{aligned} \tag{101}$$

again the inequality follows from dropping the nonnegative second term. Subtracting (101) from (100), we obtain

$$
\begin{aligned}
& \tilde{W}_{i+1}(d, d+\overline{z}) - \tilde{W}_{i+1}(d, d+\tilde{z}^u_{i+1}) \\
\leq \; & d\Big(b(\tilde{z}^u_{i+1} - \overline{z})(1-\hat{F}(\overline{z})) + h(\overline{z} - \tilde{z}^u_{i+1})(\hat{F}(\overline{z}) - \hat{F}(\tilde{z}^u_{i+1})) + h(\overline{z} - \tilde{z}^u_{i+1})\hat{F}(\tilde{z}^u_{i+1})\Big) \\
= \; & d(\overline{z} - \tilde{z}^u_{i+1})((h+b)\hat{F}(\overline{z}) - b) \\
\leq \; & d(\overline{z} - \tilde{z}^u_{i+1})(b+h)\theta, \qquad (102)
\end{aligned}
$$

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

The results (99) and (102) imply that, when $\hat{F}(\cdot)$ is a θ-estimate, it holds that

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

As demand is bounded, $d\tilde{z}^u_{i+1}$ is bounded too, hence it follows from $d\overline{z} \in \mathcal{Y}$ that there exists some constant $K_{77} > 0$ such that $d\big|\overline{z} - \tilde{z}^u_{i+1}\big| \leq K_{77}$. Thus

$$\tilde{W}_{i+1}(d, d+\overline{z}) - \tilde{W}_{i+1}(d, d+\tilde{z}^u_{i+1}) \leq K_{77}(b+h)\theta.$$

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

$$\tilde{W}_{i+1}(d, d+\overline{z}) - \tilde{W}_{i+1}(d, d+\tilde{z}_{i+1}(d)) \leq \tilde{W}_{i+1}(d, d+\overline{z}) - \tilde{W}_{i+1}(d, d+\tilde{z}^u_{i+1}) \leq K_{77}(b+h)\theta.$$

As this inequality holds for any $d \in \mathcal{D}$, it implies that, when $\hat{F}(\cdot)$ is a θ-estimate, or on the event $\mathcal{A}_7(\theta)$,

$$\max_{p\in\mathcal{P}}\Big\{\tilde{W}_{i+1}(\alpha-\beta p, \alpha-\beta p+\overline{z}) - \tilde{W}_{i+1}(\alpha-\beta p, \alpha-\beta p+\tilde{z}_{i+1}(\alpha-\beta p))\Big\} \leq K_{77}(b+h)\theta. \qquad (103)$$

Letting $\theta = \xi$ in (103) we obtain

$$
\begin{aligned}
& \mathbb{P}\left\{\max_{p\in\mathcal{P}}\Big(\tilde{W}_{i+1}(\alpha-\beta p, \alpha-\beta p+\overline{z}) - \tilde{W}_{i+1}(\alpha-\beta p, \alpha-\beta p+\tilde{z}_{i+1}(\alpha-\beta p))\Big) \leq K_{77}(b+h)\xi\right\} \\
\geq \; & \mathbb{P}(\mathcal{A}_7(\xi)) \\
\geq \; & 1 - 2e^{-2I_i\xi^2},
\end{aligned}
$$

where the last inequality follows from (96). This proves, by noting $\tilde{W}_{i+1}(\alpha-\beta p, \overline{y}(\alpha-\beta p)) - \tilde{W}_{i+1}(\alpha-\beta p, \tilde{y}_{i+1}(\alpha-\beta p)) \geq 0$ as $\tilde{y}_{i+1}(\alpha-\beta p)$ is the minimizer of $\tilde{W}_{i+1}$ on $\mathcal{Y}$, that

$$
\begin{aligned}
& \mathbb{P}\left\{\max_{p\in\mathcal{P}}\Big|\Big(\tilde{W}_{i+1}(\alpha-\beta p, \overline{y}(\alpha-\beta p)) - \tilde{W}_{i+1}(\alpha-\beta p, \tilde{y}_{i+1}(\alpha-\beta p))\Big)\Big| \leq K_{77}(b+h)\xi\right\} \\
\geq \; & 1 - 2e^{-2I_i\xi^2}. \qquad (104)
\end{aligned}
$$

Applying (95) and (104) in (92), we conclude that there exists a constant $K_{76} > 0$ such that for any $\xi > 0$, when I_i is sufficiently large,

$$\mathbb{P}\left\{\max_{p\in\mathcal{P}}\Big|\overline{W}(\alpha-\beta p, \overline{y}(\alpha-\beta p)) - \tilde{W}_{i+1}(\alpha-\beta p, \tilde{y}_{i+1}(\alpha-\beta p))\Big| \leq K_{76}\xi\right\} \;\; \geq \;\; 1 - 4e^{-2I_i\xi^2}.$$

This completes the proof of Lemma A1$'$. $\square$

The following result follows directly from of Lemma A1$'$, and its proof is omitted.

Lemma A2$'$. *There exists a positive constant K_{78} such that, for any $\xi > 0$,*

$$\mathbb{P}\left\{\max_{p\in\mathcal{P}}\left|\overline{Q}\left(p,\alpha-\beta p\right)-\tilde{Q}_{i+1}\left(p,\alpha-\beta p\right)\right|\leq K_{78}\xi\right\}\geq 1-4e^{-2I_i\xi^2}.$$

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

$$\mathcal{A}_8(\xi)=\left\{\omega:\max_{p\in\mathcal{P}}\left|\overline{Q}\left(p,\alpha-\beta p\right)-\tilde{Q}_{i+1}\left(p,\alpha-\beta p\right)\right|\leq K_{78}\xi\right\}. \tag{105}$$

Then Lemma A2$'$ can be reiterated as $\mathbb{P}(\mathcal{A}_8(\xi))\geq 1-4e^{-2I_i\xi^2}$.

With preparations of Lemmas A1$'$ and A2$'$ above, we are now ready to compare the optimal solutions of problems B1$'$ and B2$'$.

Proof of Proposition 2$'$. To slightly simplify the notation, for given parameters α and β, in this proof we let

$$\overline{Q}(p)=\overline{Q}\left(p,\alpha-\beta p\right),\quad \tilde{Q}(p)=\tilde{Q}_{i+1}\left(p,\alpha-\beta p\right),\quad \overline{p}=\overline{p}(\alpha,\beta),\quad \tilde{p}=\tilde{p}_{i+1}\left(\alpha,\beta\right).$$

By Taylor's expansion,

$$\overline{Q}(\tilde{p})=\overline{Q}(\overline{p})+\overline{Q}'(\overline{p})(\tilde{p}-\overline{p})+\frac{\overline{Q}''(q)}{2}(\tilde{p}-\overline{p})^2, \tag{106}$$

where $q\in[\overline{p},\tilde{p}]$ if $\overline{p}\leq\tilde{p}$ and $q\in[\tilde{p},\overline{p}]$ if $\overline{p}>\tilde{p}$. Since we assume the minimizer of $\overline{W}(\alpha-\beta p,y)$ over $\mathbb{R}_+$ falls into $\mathcal{Y}$, it follows from the definition in (86) that $\overline{Q}(p)=p(\alpha-\beta p)-A$, where $A=\min_z\{h\mathbb{E}(z-\epsilon)^+ + b\mathbb{E}(\epsilon-z)^+\}>0$ is a constant. Thus, we have

$$\overline{Q}''(p)=-2\beta.$$

Since $\lambda(\cdot)$ is assumed to be strictly decreasing, it follows that $\breve{\beta}(\cdot)$ is bounded below by a positive number, say $\bar{a}>0$. Therefore, $\max_{p\in\mathcal{P}}\overline{Q}''(p)\leq -2\bar{a}$, then it follows from (106) that

$$\overline{Q}(\tilde{p})\leq\overline{Q}(\overline{p})-\bar{a}(\tilde{p}-\overline{p})^2. \tag{107}$$

Now we prove, on event $\mathcal{A}_8(\xi)$, that

$$\overline{Q}(\tilde{p})-\overline{Q}(\overline{p})\geq -2K_{78}\xi. \tag{108}$$

We prove this by contradiction. Suppose it is not true, i.e., $\overline{Q}(\overline{p})-\overline{Q}(\tilde{p})>2K_{78}\xi$, then it follows from (105) that

$$\begin{aligned}
&\tilde{Q}(\overline{p})-\tilde{Q}(\tilde{p})\\
=\;&\big(\tilde{Q}(\overline{p})-\overline{Q}(\overline{p})\big)+\big(\overline{Q}(\overline{p})-\overline{Q}(\tilde{p})\big)+\big(\overline{Q}(\tilde{p})-\tilde{Q}(\tilde{p})\big)\\
>\;&-K_{78}\xi+2K_{78}\xi-K_{78}\xi\\
=\;&0.
\end{aligned}$$

This leads to $\tilde{Q}(\overline{p}) > \tilde{Q}(\tilde{p})$, contradicting with $\tilde{p}$ being optimal for problem B2. Thus, (108) is satisfied on $\mathcal{A}_8(\xi)$.

Using (107) and (108), we obtain that, on event $\mathcal{A}_8(\xi)$,

$$\left|\tilde{p} - \overline{p}\right|^2 \leq \frac{2K_{78}}{\bar{a}}\xi,$$

or equivalently, for some constant K_{79},

$$\left|\tilde{p} - \overline{p}\right| \leq K_{79}\xi^{\frac{1}{2}}.$$

That is, on $\mathcal{A}_8(\xi)$, it holds that

$$|\overline{p}(\alpha, \beta) - \tilde{p}_{i+1}(\alpha, \beta)| \leq K_{79}\xi^{\frac{1}{2}}.$$

This shows that for any values of α and $\beta \geq \bar{a}$,

$$\mathbb{P}\left\{|\overline{p}(\alpha, \beta) - \tilde{p}_{i+1}(\alpha, \beta)| \leq K_{79}\xi^{\frac{1}{2}}\right\} \geq \mathbb{P}(\mathcal{A}_8(\xi)) \geq 1 - 4e^{-2I_i\xi^2}.$$

Substituting $\alpha = \breve{\alpha}(p)$ and $\beta = \breve{\beta}(p)$, we obtain the desired result in Proposition 2′. □

To prove Proposition 3′, we need Lemmas $A3'$ and A4′ first.

Lemma A3′. *There exists a positive constant i^* such that when $i \geq i^*$,*

$$\mathbb{P}\Big\{\hat{\beta}_{i+1} > 0\Big\} \geq 1 - \frac{8}{I_i}.$$

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

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

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

$$\hat{\alpha}_{i+1} = \frac{\lambda(\hat{p}_i) + \lambda(\hat{p}_i + \delta_i)}{2} + \frac{B^1_{i+1} + B^2_{i+1}}{2} + \hat{\beta}_{i+1}\frac{2\hat{p}_i + \delta_i}{2}, \tag{109}$$

$$\hat{\beta}_{i+1} = -\frac{\lambda(\hat{p}_i + \delta_i) - \lambda(\hat{p}_i)}{\delta_i} - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}). \tag{110}$$

Applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the second order for (110), we obtain

$$\begin{aligned}\hat{\beta}_{i+1} &= -\left(\lambda'(\hat{p}_i) + \frac{1}{2}\lambda''(q_i)\delta_i\right) - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1})\\ &= \breve{\beta}(\hat{p}_i) - \frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}),\end{aligned} \tag{111}$$

where $q_i \in [\hat{p}_i, \hat{p}_i+\delta_i]$. Substituting $\hat{\beta}_{i+1}$ in (109) by (111), and applying Taylor's expansion on $\lambda(\hat{p}_i+\delta_i)$ at point $\hat{p}_i$ to the first order, we have

$$\begin{aligned}
\hat{\alpha}_{i+1} &= \lambda(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B_{i+1}^1 + B_{i+1}^2}{2} - \lambda'(\hat{p}_i)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\
&\quad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B_{i+1}^1 + B_{i+1}^2)\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\
&= \breve{\alpha}(\hat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + \frac{B_{i+1}^1 + B_{i+1}^2}{2} - \frac{1}{2}\lambda'(\hat{p}_i)\delta_i \\
&\quad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B_{i+1}^1 + B_{i+1}^2)\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right), \qquad (112)
\end{aligned}$$

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

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

$$\mathbb{P}\left\{\left|-B_{i+1}^1\right| > \xi\right\} \leq 2e^{-2I_i\xi^2/(u-l)^2}, \qquad \mathbb{P}\left\{\left|B_{i+1}^2\right| > \xi\right\} \leq 2e^{-2I_i\xi^2/(u-l)^2}. \qquad (113)$$

Hence,

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

Therefore,

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

Similar argument shows

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

Since $\lambda'(\cdot)$ and $\lambda''(\cdot)$ are bounded and δ_i converges to 0, from (112) we conclude that there must exist a constant K_{80} such that, on the event $\left|B_{i+1}^1 + B_{i+1}^2\right| \leq 2\xi$ and $\left|-B_{i+1}^1 + B_{i+1}^2\right| \leq 2\xi$, it holds that

$$|\hat{\alpha}_{i+1} - \breve{\alpha}(\hat{p}_i)| \leq K_{80}\left(\delta_i + \frac{\xi}{\delta_i} + \xi\right).$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left\{|\hat{\alpha}_{i+1} - \breve{\alpha}(\hat{p}_i)| \leq K_{80}\left(\delta_i + \frac{\xi}{\delta_i} + \xi\right)\right\} &\geq \mathbb{P}\left\{\left|B_{i+1}^1 + B_{i+1}^2\right| \leq 2\xi, \left|-B_{i+1}^1\right| + \left|B_{i+1}^2\right| \leq 2\xi\right\} \\
&\geq 1 - 8e^{-2I_i\xi^2/(u-l)^2},
\end{aligned}$$

which implies

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1} - \breve{\alpha}(\hat{p}_i)|^2 \leq K_{81}\left(\delta_i^2 + \frac{\xi^2}{\delta_i^2} + \xi^2\right)\right\} \geq 1 - 8e^{-2I_i\xi^2/(u-l)^2}. \qquad (114)$$

From (111) we have

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right| \leq K_{82}\left(\delta_i+\frac{\xi}{\delta_i}\right)\right\} \geq 1-4e^{-2I_i\xi^2/(u-l)^2},$$

which implies

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2 \leq K_{83}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}\right)\right\} \geq 1-4e^{-2I_i\xi^2/(u-l)^2}. \tag{115}$$

Combining (114) and (115), we obtain

$$\begin{aligned}&\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i)|^2+\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2 \leq K_{84}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\right\} \\ \geq \;& 1-12e^{-2I_i\xi^2/(u-l)^2}.\end{aligned} \tag{116}$$

Let $\xi=(2I_i)^{-\frac{1}{2}}(\log 2I_i)^{\frac{1}{2}}$ in (116), we obtain

$$\begin{aligned}&\mathbb{P}\left\{\,|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2 \leq K_{84}\left(I_i^{-\frac{1}{2}}+(2I_i)^{-\frac{1}{2}}(\log 2I_i)+(2I_i)^{-1}(\log 2I_i)\right)\right\} \\ \geq\;& 1-\frac{8}{I_i}.\end{aligned}$$

This implies

$$\begin{aligned}&\mathbb{P}\left\{\,|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}| \leq (3K_{84})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}| \leq (3K_{84})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\right\} \\ \geq\;& 1-\frac{8}{I_i}.\end{aligned}$$

For convenience, we define the event $\mathcal{A}_9$ by

$$\mathcal{A}_9=\left\{\omega:|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}| \leq (3K_{84})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}| \leq (3K_{84})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\right\}. \tag{117}$$

Then by (117) one has

$$\mathbb{P}(\mathcal{A}_9^c) \leq \frac{8}{I_i}.$$

Define

$$i^*=\max\left\{\log_v\frac{e}{2I_0},\ \min\left\{i\ \middle|\ (3K_{84})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}<\min_{p\in\mathcal{P}}\breve{\beta}(p)\right\}\right\},$$

where we need i^* to be no less than $\log_v\frac{e}{2I_0}$ to ensure that $(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}$ is decreasing on $i \geq i^*$. When $i \geq i^*$, it follows that $\hat{\beta}_{i+1}>0$ on $\mathcal{A}_9$. Therefore

$$\mathbb{P}(\hat{\beta}_{i+1}>0) \geq 1-\frac{8}{I_i}.$$

Lemma A3′ is thus proved. □

Lemma A4′. *There exists a positive constant* K_{85} *such that*

$$\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}D_t-\lambda(\hat{p}_i)\right|^2+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}D_t-\lambda(\hat{p}_i+\delta_i)\right|^2\right]$$
$$\leq \big(K_{85}+4(u-l)^2\big)I_i^{-\frac{1}{2}}.$$

Proof. Starting from (116), one has

$$\mathbb{P}\left\{\left(\frac{K_{86}}{\delta_i^2}+K_{87}\right)^{-1}\left(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2-K_{84}\delta_i^2\right)\geq\xi^2\right\}<12e^{-2I_i\xi^2/(u-l)^2}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{K_{86}}{\delta_i^2}+K_{87}\right)^{-1}\left(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2-K_{84}\delta_i^2\right)\right]\\
=&\left(\frac{K_{86}}{\delta_i^2}+K_{87}\right)^{-1}\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2\right]-\left(\frac{K_{86}}{\delta_i^2}+K_{87}\right)^{-1}K_{84}\delta_i^2\\
\leq&\int_0^{+\infty}12e^{-2I_i\xi/(u-l)^2}d\xi\\
=&\frac{6(u-l)^2}{I_i}.
\end{aligned}$$

Hence one has

$$\begin{aligned}
&\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2\right]\\
\leq&\left(\frac{6(u-l)^2}{I_i}+\left(\frac{K_{86}}{\delta_i^2}+K_{87}\right)^{-1}K_{84}\delta_i^2\right)\left(\frac{K_{86}}{\delta_i^2}+K_{87}\right)\\
\leq&K_{88}I_i^{-\frac{1}{2}}.
\end{aligned}\tag{118}$$

Because

$$\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}D_t-\lambda(\hat{p}_i)=\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}\epsilon_t$$

and

$$\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}D_t-\lambda(\hat{p}_i+\delta_i)=\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}\epsilon_t,$$

by (113), we have

$$\mathbb{P}\left\{\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i}D_t-\lambda(\hat{p}_i)\right|>\xi\right\}\leq 2e^{-2I_i\xi^2/(u-l)^2},$$
$$\mathbb{P}\left\{\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i}D_t-\lambda(\hat{p}_i+\delta_i)\right|>\xi\right\}\leq 2e^{-2I_i\xi^2/(u-l)^2}.$$

Therefore, we have

$$\mathbb{P}\left\{\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t-\lambda(\hat{p}_i)\right|^2+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t-\lambda(\hat{p}_i+\delta_i)\right|^2>2\xi^2\right\}\le 4e^{-2I_i\xi^2/(u-l)^2},$$

which leads to

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t-\lambda(\hat{p}_i)\right|^2+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t-\lambda(\hat{p}_i+\delta_i)\right|^2\right]\\
\le\ & \int_0^{+\infty} 4e^{-I_i\xi/(u-l)^2}\\
=\ & \frac{4(u-l)^2}{I_i}. \qquad (119)
\end{aligned}$$

Combining (118) and (119), we have

$$\begin{aligned}
&\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t-\lambda(\hat{p}_i)\right|^2+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t-\lambda(\hat{p}_i+\delta_i)\right|^2\right]\\
&\le (K_{88}+4(u-l)^2)I_i^{-\frac{1}{2}}.
\end{aligned}$$

This completes the proof of Lemma A4′. □

Proof of Proposition 3′. To bound $\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\right|^2\right]$, by Lemma A3′ and Lemma A4′, when i is large enough (greater than or equal to i^* defined in the proof of Lemma A3′), for some positive constants K_{89}, and K_{90} one has

$$\begin{aligned}
&\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\right|^2\right]\\
\le\ & \mathbb{E}\left[K_{74}^2\left(\left|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}\right|+\left|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}\right|+\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t-\lambda(\hat{p}_i)\right|+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t-\lambda(\hat{p}_i+\delta_i)\right|\right)^2\right]\\
&\qquad +\frac{8}{I_i}\left(p^h-p^l\right)^2\\
\le\ & \mathbb{E}\left[K_{89}\left(|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+\left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t-\lambda(\hat{p}_i)\right|^2+\left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t-\lambda(\hat{p}_i+\delta_i)\right|^2\right)\right]\\
&\qquad +\frac{8}{I_i}\left(p^h-p^l\right)^2\\
\le\ & K_{90}I_i^{-\frac{1}{2}}, \qquad (120)
\end{aligned}$$

where the first inequality follows from the Lipschitz condition (88) in Assumption 1′(vii), and the third inequality follows from Lemma A4′. □

Proof of Theorem 1′: Convergence of $\hat{p}_{i+1}$ and $\hat{p}_{i+1}+\delta_{i+1}$. We proceed as follows:

$$\begin{aligned}
& \mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\Big] \\
\leq\ & \mathbb{E}\Big[\Big(\gamma|p^* - \hat{p}_i| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\Big] \\
\leq\ & \left(\frac{1+\gamma^2}{2}\right)\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + K_{91}\mathbb{E}\left[\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|^2\right] \\
& \qquad + K_{92}\mathbb{E}\left[\Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|^2\right], \qquad (121)
\end{aligned}$$

where the second inequality follows from Proposition 1′, and the third inequality is justified by $\gamma < 1$ in Proposition 1′, and the last inequality holds for some appropriately chosen K_{91} and K_{92} because of the inequality $(a+b)^2 \leq 2(a^2+b^2)$ for any real numbers a and b.

To bound $\mathbb{E}\big[\big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\big|^2\big]$ in (121), by Proposition 2′ one has, for some constant K_{93},

$$\mathbb{E}\left[\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|^2\right] \leq K_{93}^2 \int_0^{+\infty} 5e^{-4I_i\xi^2} d\xi = \frac{5\pi^{\frac{1}{2}} K_{93}^2}{4I_i^{\frac{1}{2}}}. \qquad (122)$$

Substituting (122) and (120) into (121), one has

$$\mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \leq \left(\frac{1+\gamma^2}{2}\right)\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + K_{94} I_i^{-\frac{1}{2}}.$$

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

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \quad \leq \quad \theta^i(\hat{p}_1 - p^*)^2 + K_{94}\sum_{j=0}^{i-1}\theta^j I_{i-j}^{-\frac{1}{2}} \leq K_{95}(v^{-\frac{1}{2}})^i \sum_{j=0}^{i-1}\theta^j (v^{\frac{1}{2}})^j.$$

We choose $v > 1$ that satisfies $\theta v^{\frac{1}{2}} < 1$, then there exists a positive constant K_{96} such that $\sum_{j=0}^{i-1}\theta^j(v^{\frac{1}{2}})^j \leq K_{96}$, therefore, for some constants K_{97} and K_{98},

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \leq K_{97}(v^{-\frac{1}{2}})^i \leq K_{98} I_i^{-\frac{1}{2}}. \qquad (123)$$

Moreover, we have, for some positive constant K_{99},

$$\mathbb{E}\left[(\hat{p}_{i+1} + \delta_{i+1} - p^*)^2\right] \leq 2\mathbb{E}\left[(\hat{p}_{i+1} - p^*)^2\right] + 2\delta_{i+1}^2 \leq K_{99} I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty.$$

Because mean-square convergence implies convergence in probability, this shows that the pricing decisions from DDA converge to p^* in probability. □

Proof of Proposition $4'$. For $p \in \mathcal{P}$, the optimal solution for bridging problem B1$'$ is $\overline{y}\big(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\big)$. Thus

$$\overline{y}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) = \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p + F^{-1}\left(\frac{b}{b+h}\right). \tag{124}$$

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

$$\tilde{y}^u_{i+1}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) = \breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p + \min\left\{\epsilon_j : \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{\epsilon_t \le \epsilon_j\right\} \ge \frac{b}{b+h}\right\}, \tag{125}$$

and we have

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

It is seen that

$$\begin{aligned}&\left|\overline{y}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \tilde{y}_{i+1}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right)\right| \\ &\qquad \le \left|\overline{y}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right) - \tilde{y}^u_{i+1}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)p\right)\right|.\end{aligned} \tag{126}$$

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

$$\begin{aligned}
&\mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\right) - \frac{b}{b+h} \le -z\right\} \qquad (127)\\
= \;& \mathbb{P}\left\{\tilde{y}^u_{i+1}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\}\\
\le \;& \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{\epsilon_t \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\} \ge \frac{b}{b+h}\right\}\\
= \;& \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{\epsilon_t \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\} - \left(\frac{b}{b+h} - z\right) \ge z\right\},
\end{aligned}$$

where the first inequality follows from (125). Since $\mathbb{E}\left[\mathbb{1}\left\{\epsilon_t \le F^{-1}\left(\frac{b}{b+h} - z\right)\right\}\right] = \frac{b}{b+h} - z$, we apply Hoeffding inequality to obtain

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

Combining this with (124) and (127), we obtain

$$\begin{aligned}\mathbb{P}\Bigg\{&F\left(\tilde{y}^u_{i+1}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\right) - \\ &F\left(\overline{y}\left(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\right) - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\right) \le -z\Bigg\} \le e^{-4I_i z^2}.\end{aligned} \tag{128}$$

Similarly, we have

$$\mathbb{P}\Big\{F\Big(\tilde{y}^u_{i+1}\left(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\right)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\Big)- F\Big(\overline{y}\left(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\right)-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})\Big)\geq -z\Big\}\leq e^{-4I_iz^2}. \tag{129}$$

From Assumption $1'$(vi), the probability density function $f(\cdot)$ of ϵ_t satisfies $r=\min\{f(x),x\in[l,u]\}>0$. From calculus, it is known that, for any $x<y$, there exists a number $z\in[x,y]$ such that $F(y)-F(x)=f(z)(y-x)\geq r(y-x)$. Applying (128) and (129), for any $\xi>0$, we obtain

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

Let $K_{100}=\frac{1}{r}$, then $K_{100}>0$. We have

$$\mathbb{P}\Big\{\Big|\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)-\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)\Big|\geq K_{100}\xi\Big\}\leq 2e^{-4I_i\xi^2},$$

and Proposition $4'$ follows from the inequality above and (126). □

Convergence of $\hat{y}_{i+1,1}$ and $\hat{y}_{i+1,2}$. First note that, for some constant K_{101}, we have

$$\begin{aligned}
& \mathbb{E}\left[\left|y^*-\hat{y}_{i+1,1}\right|^2\right] \\
\leq\ & K_{101}\mathbb{E}\Big[\Big(\underbrace{\left|\overline{y}\big(\lambda(p^*)\big)-\overline{y}(\lambda(\hat{p}_{i+1}))\right|^2}_{\text{Difference between } p^* \text{ and } \hat{p}_{i+1}}+\underbrace{\left|\overline{y}\big(\lambda(\hat{p}_{i+1})\big)-\overline{y}\big(\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}\right|^2}_{\text{Zero}} \\
& +\underbrace{\left|\overline{y}\big(\breve{\alpha}(\hat{p}_{i+1})-\breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}\big)-\overline{y}\big(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)\right|^2}_{\text{Difference between } \hat{p}_{i+1} \text{ and } \hat{p}_i} \\
& +\underbrace{\left|\overline{y}\big(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)-\tilde{y}_{i+1}\big(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)\right|^2}_{\substack{\text{Comparison of problems B1}' \text{ and B2}' \\ \text{Proposition } 4'}}+\underbrace{\left|\tilde{y}_{i+1}\big(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)-\hat{y}_{i+1,1}\right|^2}_{\text{Comparison of problems B2}' \text{ and DD}'}\Big)\Big].
\end{aligned} \tag{130}$$

In the following analysis we shall upper bound each term on the right hand side of (130).

In the first term, one has

$$\mathbb{E}\left[\left|\overline{y}\big(\lambda(p^*)\big) - \overline{y}\big(\lambda(\hat{p}_{i+1})\big)\right|^2\right] \le K_{102}\mathbb{E}\left[|\ p^* - \hat{p}_{i+1}\ |^2\right].$$

By the definition of $\breve{\alpha}(p)$ and $\breve{\beta}(p)$ in (85), one has $\breve{\alpha}(\hat{p}_{i+1}) - \breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1} = \lambda(\hat{p}_{i+1})$, thus the second term on the right hand side of (130) vanishes. For the third term, because $\overline{y}\big(\breve{\alpha}(q) - \breve{\beta}(q)p\big) = \breve{\alpha}(q) - \breve{\beta}(q)p + F^{-1}\left(\frac{b}{b+h}\right)$, and $\lambda(p)$ has bounded first and second order derivatives (Assumption $1'$(v)), there exists some constant K_{103} such that

$$\begin{aligned}\mathbb{E}\left[\left|\overline{y}\big(\breve{\alpha}(\hat{p}_{i+1}) - \breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1}\big) - \overline{y}\big(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}\big)\right|^2\right] &\le K_{103}\mathbb{E}\left[|\ \hat{p}_{i+1} - \hat{p}_i\ |^2\right] \\ &\le 2K_{103}\mathbb{E}\left[\big(\ |\ p^* - \hat{p}_i\ |^2 + |\ p^* - \hat{p}_{i+1}\ |^2\ \big)\right].\end{aligned}$$

The fourth term is upper bounded by Proposition $4'$.

To bound the fifth term of the right hand side of (130), we follow a similar approach as that in establishing Proposition $3'$. By Lipschitz condition (89) and Lemma A$4'$, and following similar analyses as in the proof of Proposition $3'$, we obtain, for some constant K_{16},

$$\begin{aligned}&\mathbb{E}\left[\left|\tilde{y}_{i+1}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}) - \hat{y}_{i+1,1}\right|^2\right] \\ \le\ & 4K_{74}^2\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+1}^{t_i+I_i} D_t - \lambda(\hat{p}_i)\right|^2 + \left|\frac{1}{I_i}\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t - \lambda(\hat{p}_i+\delta_i)\right|^2\right] \\ &\qquad + \frac{8(y^h - y^l)^2}{I_i} \\ \le\ & K_{104}I_i^{-\frac{1}{2}}.\end{aligned}$$

Summarizing the analyses above we obtain, for some constants K_{105} and K_{106},

$$\mathbb{E}\left[\big(y^* - \hat{y}_{i+1,1}\big)^2\right] \le K_{105}\mathbb{E}\big[\ |\ p^* - \hat{p}_{i+1}\ |^2 + |\ p^* - \hat{p}_i\ |^2\ \big] + K_{105}I_i^{-\frac{1}{2}} \le K_{106}I_i^{-\frac{1}{2}} \to 0, \qquad \text{as } i \to \infty, \tag{131}$$

where the second inequality follows from the convergence rate of the pricing decisions in (123).

Similarly, we obtain

$$\mathbb{E}\left[\big(y^* - \hat{y}_{i+1,2}\big)^2\right] \le K_{107}I_i^{-\frac{1}{2}} \to 0, \qquad \text{as } i \to \infty.$$

Convergence of $\hat{y}_{i+1,1}$ and $\hat{y}_{i+1,2}$ is thus obtained.

Proof of Theorem $1'$: Convergence of y_t. To deal with the overshooting problem, since $\tilde{\lambda}(p^h) + l \le \tilde{D}_t \le \tilde{\lambda}(p^l) + u$, it follows from Hoeffding inequality that for any $\zeta > 0$,

$$\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \ge \mathbb{E}\left[\sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t\right] - \zeta\right\} \ge 1 - \exp\left(-\frac{2\zeta^2}{I_i\big(\tilde{\lambda}(p^l) + u - \tilde{\lambda}(p^h) - l\big)^2}\right). \tag{132}$$

Letting $\zeta = \left(\tilde{\lambda}(p^l) + u - \tilde{\lambda}(p^h) - l\right)(I_i)^{\frac{1}{2}} (\log I_i)^{\frac{1}{2}}$ in (132), then we obtain

$$\mathbb{P}\left\{ \sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \geq I_i \mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] - \left(\tilde{\lambda}(p^l) + u - \tilde{\lambda}(p^h) - l\right)(I_i)^{\frac{1}{2}} (\log I_i)^{\frac{1}{2}} \right\} \geq 1 - \frac{1}{I_i^2}. \quad (133)$$

By Assumption 1′(iv), we have $\mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] \geq \min_{p \in \mathcal{P}} \mathbb{E}[D_t(p)] > 0$. This implies, when i is large enough, we will have

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

Define event

$$\mathcal{A}_{10} = \left\{ \omega : \sum_{t=t_i+I_i+1}^{t_i+2I_i} \tilde{D}_t \geq \frac{1}{2} I_i \mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] \right\},$$

then it follows from (133) that, when i is large enough,

$$\mathbb{P}(\mathcal{A}_{10}) \geq 1 - \frac{1}{I_i^2}.$$

Note that when i is large enough, $\frac{1}{2} I_i \mathbb{E}\left[\tilde{D}_{t_i+I_i+1}\right] > y^h - y^l$, which means that on the event $\mathcal{A}_1^A$, the cumulative demand during $\{t_i + I_i + 1, \ldots, t_i + 2I_i\}$ is high enough to consume all the initial on-hand inventory at the beginning of period $t_i + I_i + 1$ and $\hat{y}_{i,2}$ will be achieved.

Therefore, by definition of event $\mathcal{A}_1$ in the main paper, we have

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

The rest of proofs is the same as the case of the multiplicative case as explained in the main paper. Convergence of actual inventory levels y_t can thus be proved, and one still has

$$\mathbb{E}\left[\left(y^* - y_t\right)^2\right] \leq K_{108} I_i^{-\frac{1}{2}} \to 0, \qquad \text{as } t \to \infty. \quad (134)$$

The proof of Theorem 1′ is now complete. □

Proof of Proposition 5′. To develop an upper bound for $G(p^*, y^*) - G(\hat{p}_i, y_t)$, we first apply Taylor expansion on $G(p, \overline{y}(\lambda(p))$ at point p^*. Using the fact that the first order derivative vanishes at $p = p^*$ and the assumption that the second order derivative is bounded (Assumption 1′(iii)), we obtain, for some constant $K_{109} > 0$, that

$$G\big(p^*, \overline{y}\big(\lambda(p^*)\big)\big) - G\big(\hat{p}_i, \overline{y}\big(\lambda(\hat{p}_i)\big)\big) \leq K_{109}(p^* - \hat{p}_i)^2. \quad (135)$$

Noticing that $\overline{y}\big(\lambda(\hat{p}_i)\big)$ maximizes the concave function $G\big(\hat{p}_i, y\big)$ for given $\hat{p}_i$, we apply Taylor expansion with respect to y at point $y = \overline{y}\big(\lambda(\hat{p}_i)\big)$ to yield that, for some constant K_{110},

$$G\big(\hat{p}_i, \overline{y}\big(\lambda(\hat{p}_i)\big)\big) - G(\hat{p}_i, y_t) \leq K_{110}\big(\overline{y}\big(\lambda(\hat{p}_i)\big) - y_t\big)^2. \quad (136)$$

In addition, we have

$$
\begin{aligned}
& \mathbb{E}\left[(\overline{y}(\lambda(\hat{p}_i)) - y_t)^2\right] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|\overline{y}(\lambda(\hat{p}_i)) - \overline{y}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i)\Big| + \Big|\overline{y}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i) - \overline{y}(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i)\Big| \\
& \quad + \Big|\overline{y}(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i) - \tilde{y}_i(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i)\Big| + \Big|\tilde{y}_i(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i) - \hat{y}_{i,1}\Big| \\
& \quad + \Big|\hat{y}_{i,1} - y_t\Big|\Big)^2\Big] \\
\leq\ & 2\mathbb{E}\Big[\Big(\Big|\overline{y}(\lambda(\hat{p}_i)) - \overline{y}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i)\Big| + \Big|\overline{y}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i) - \overline{y}(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i)\Big| \\
& \quad + \Big|\overline{y}(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i) - \tilde{y}_i(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i)\Big| + \Big|\tilde{y}_i(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i) - \hat{y}_{i,1}\Big|\Big)^2\Big] \\
& +2\mathbb{E}\Big[\Big|\hat{y}_{i,1} - y_t\Big|^2\Big] \\
\leq\ & K_{111}\mathbb{E}\Big[\Big|\overline{y}(\lambda(\hat{p}_i)) - \overline{y}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i)\Big|^2 + \Big|\overline{y}(\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i) - \overline{y}(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i)\Big|^2 \\
& \quad + \Big|\overline{y}(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i) - \tilde{y}_i(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i)\Big|^2 \\
& \quad + \Big|\tilde{y}_i(\breve{\alpha}(\hat{p}_{i-1}) - \breve{\beta}(\hat{p}_{i-1})\hat{p}_i) - \hat{y}_{i,1}\Big|^2\Big] \qquad (137) \\
& +2\mathbb{E}\Big[\Big|\hat{y}_{i,1} - y_t\Big|^2\Big]. \qquad (138)
\end{aligned}
$$

The expectation in (137) is similar to the right hand side of (130) except that $i+1$ is replaced by i. Thus, using the same analysis as that for (130), we obtain

$$
\mathbb{E}\left[(\overline{y}(\lambda(\hat{p}_i)) - \hat{y}_{i,1})^2\right] \leq K_{112} I_{i-1}^{-\frac{1}{2}} \qquad (139)
$$

for some constant K_{112}.

The expectation in (138) is bounded as the following:

$$
\mathbb{E}\Big[\Big|\hat{y}_{i,1} - y_t\Big|^2\Big] \leq \mathbb{E}\Big[\Big|\hat{y}_{i,1} - y^*\Big|^2 + \Big|y^* - y_t\Big|^2\Big] \leq K_{113} I_{i-1}^{-\frac{1}{2}}, \qquad (140)
$$

where the second inequality follows from (131) and (134).

Therefore, combining (139) and (140), one has

$$
\mathbb{E}\left[(\overline{y}(\lambda(\hat{p}_i)) - y_t)^2\right] \leq K_{114} I_{i-1}^{-\frac{1}{2}}. \qquad (141)
$$

Applying the results above, we obtain, for some constants K_{115}, K_{116}, and K_{117}, that

$$
\begin{aligned}
& \mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i, y_t)\right] \\
=\ & \mathbb{E}\Big[\Big(G\big(p^*, \overline{y}\big(\lambda(p^*)\big)\big) - G\big(\hat{p}_i, \overline{y}\big(\lambda(\hat{p}_i)\big)\big)\Big) + \Big(G\big(\hat{p}_i, \overline{y}\big(\lambda(\hat{p}_i)\big)\big) - G(\hat{p}_i, y_t)\Big)\Big] \\
\leq\ & K_{115}\Big(\mathbb{E}\left[(p^* - \hat{p}_i)^2\right] + \mathbb{E}\Big[\big(\overline{y}(e^{\lambda(\hat{p}_i)}) - y_t\big)^2\Big]\Big) \\
\leq\ & K_{116}\left(K_{98} I_{i-1}^{-\frac{1}{2}} + K_{114} I_{i-1}^{-\frac{1}{2}}\right) \\
=\ & K_{117} I_{i-1}^{-\frac{1}{2}},
\end{aligned}
$$

where the first inequality follows from (135) and (136), and the second inequality follows from the convergence rate of pricing decisions (123) and (141).

Similarly, we establish for some constants K_{118}, K_{119} and K_{120}, that

$$\begin{aligned}
\mathbb{E}\left[G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\right] &\leq K_{118}\left(\mathbb{E}\left[(p^* - \hat{p}_i - \delta_i)^2\right] + \mathbb{E}\left[(\overline{y}(\lambda(\hat{p}_i + \delta_i)) - y_t)^2\right]\right) \\
&\leq K_{118}\left(\mathbb{E}\left[2(p^* - \hat{p}_i)^2 + 2\delta_i^2\right] + K_{119} I_{i-1}^{-\frac{1}{2}}\right) \\
&\leq K_{120} I_{i-1}^{-\frac{1}{2}}.
\end{aligned}$$

Note that, as seen from Proposition $3'$, these results hold when i is greater than or equal to some number i^*. □

Proof of Theorem 2′. We have, for some constants K_{121}, K_{122} and K_{123},

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\left(G(p^*, y^*) - G(p_t, y_t)\right)\right] \\
\leq\ & \sum_{i=i^*+1}^{n} K_{121} I_{i-1}^{-\frac{1}{2}} I_i + \sum_{i=1}^{i^*}\left(\sum_{t=t_i+1}^{t_i+I_i}\left(G(p^*, y^*) - G(\hat{p}_i, y_t)\right) + \sum_{t=t_i+I_i+1}^{t_i+2I_i}\left(G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\right)\right) \\
=\ & \sum_{i=i^*+1}^{n} K_{121} I_{i-1}^{\frac{1}{2}} + K_{122} \\
\leq\ & K_{121} \sum_{i=2}^{n} I_{i-1}^{\frac{1}{2}} + K_{122} \\
=\ & K_{121} \frac{(2I_0)^{\frac{1}{2}} v^{\frac{1}{2}}}{v^{\frac{1}{2}} - 1}\left(v^{\frac{n-1}{2}} - 1\right) + K_{122} \\
\leq\ & K_{121} \frac{(2I_0)^{\frac{1}{2}} v^{\frac{1}{2}}}{v^{\frac{1}{2}} - 1}\left(v^{\log_v\left(\frac{v-1}{2I_0 v}T+1\right)+1-1}\right)^{\frac{1}{2}} + K_{122} \\
\leq\ & K_{123} T^{\frac{1}{2}},
\end{aligned}$$

where the first inequality follows from Proposition $5'$, and

$$K_{122} = \sum_{i=1}^{i^*}\left(\sum_{t=t_i+1}^{t_i+I_i}\left(G(p^*, y^*) - G(\hat{p}_i, y_t)\right) + \sum_{t=t_i+I_i+1}^{t_i+2I_i}\left(G(p^*, y^*) - G(\hat{p}_i + \delta_i, y_t)\right)\right).$$

□

Appendix E: The Case of Unbounded Error

In the main text of the paper, we have assumed that the random error $\tilde{\epsilon}$ has bounded support. In this section, we relax this assumption. For this case, impose the condition that the moment generating function of $\tilde{\epsilon}$ is finite around 0. That is, there exists a constant $\rho > 0$, such that for any $s \in (-\rho, \rho)$ it holds that

$$\mathbb{E}[e^{s\tilde{\epsilon}}] < +\infty. \tag{142}$$

Note that many distribution families satisfy this condition, including sub-exponential and sub-Gaussian distributions.

The learning algorithm and its regret rate for the case with unbounded error are changed slightly from that in the main text. The algorithm input parameters are v, ρ and I_0, with $v > 1$, $\rho > 0$ and $I_0 > 0$. To initiate the algorithm, it sets $\{\hat{p}_{11}, \hat{y}_{11}, \hat{y}_{12}\}$, where $\hat{p}_{11} \in \mathcal{P}, \hat{y}_{11} \in \mathcal{Y}, \hat{y}_{12} \in \mathcal{Y}$ are the starting pricing and order-up-to levels. For $i \geq 1$, let

$$I_i = \lfloor I_0 v^i \rfloor, \quad \delta_i = \rho(2I_{i-1})^{-\frac{1}{4}}(\log 2I_{i-1})^{\frac{1}{4}}, \text{ and } t_i = \sum_{k=1}^{i-1} 2I_k. \tag{143}$$

The following is the detailed procedure of the algorithm. Recall that x_t is the starting inventory level at the beginning of period t, p_t is the selling price set for period t, and y_t $(\geq x_t)$ is the inventory level of period t after replenishment decision, $t = 1, \ldots, T$. The number of learning stages is $n = \left\lceil \log_v \left(\frac{v-1}{2I_0 v} T \right) \right\rceil$.

Data-Driven Algorithm for Unbounded Error (DDAU)

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

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

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

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

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

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

$$(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = \operatorname*{argmin}_{\alpha,\beta} \left\{ \sum_{t=t_i+1}^{t_i+2I_i} \Big(D_t - (\alpha - \beta p_t) \Big)^2 \right\},$$
$$\eta_t = D_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t), \qquad \text{for } t = t_i + 1, \ldots, t_i + 2I_i.$$

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

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

Then the data-driven optimization is defined by

Problem DD:

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

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

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

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

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

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

The following result presents the regret rate for the case with unbounded error. Its proof is omitted for brevity.

Theorem 2″. *Under Assumption 1 and condition (142), the DDAU policy is asymptotically optimal. More specifically, there exists some constant $K_{124} > 0$ such that*

$$R(DDAU, T) = G(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t, y_t)\right] \le K_{124}T^{-\frac{1}{2}}(\log T)^{\frac{1}{2}}.$$

References

[1] Besbes O, Zeevi A (2015) On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Sci.* 61(4): 723-739.

[2] Heyman D, Sobel M (1984) *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization.* McGraw-Hill, New York.

[3] Hoeffding W (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58:13-30.

[4] Levi R, Roundy RO, Shmoys DB (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. *Math. Oper. Res.* 32(4):821-839.