MICHIGAN ROSS

# Working Paper

# Coordinating Pricing and Inventory Replenishment with Nonparametric Demand Learning

Boxiao Chen
Department of Industrial and Operations Engineering
University of Michigan

Xiuli Chao
Department of Industrial and Operations Engineering
University of Michigan

Hyun-Soo Ahn
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper Series
Working Paper No. 1294
June 2015

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

**Keywords:** dynamic pricing, inventory control, demand learning, nonparametric estimation, nonperishable products, asymptotic optimality.

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

# 1 Introduction

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

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

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

## 1.1 Related literature

Almost all early papers in joint pricing and inventory control, e.g., Whitin (1955), Federgruen and Heching (1999), and Chen and Simchi-Levi (2004), among others, assume that a firm has complete knowledge about the distribution of underlying stochastic demand for any given selling price. The complete information assumption provides analytic tractability necessary for characterizing the optimal policy. The extension to the parametric case (the firm knows the class of distribution but not the parameters) has been studied by, for example, Subrahmanyan and Shoemaker (1996), Petruzzi and Dada (2002), and Zhang and Chen (2006). Chung et al. (2011) also consider the problem of dynamic pricing and inventory planning with demand learning, and they develop learning algorithms using Bayesian method and Markov chain Monte Carlo (MCMC) algorithms, and numerically evaluate the importance of dynamic pricing. An alternative to the parametric approach is to model the firm's problem in a nonparametric setting. Under this framework, the firm does not make specific assumptions about underlying demand. Instead, the firm makes decisions solely based on the collected demand data, see Burnetas and Smith (2000). Our work falls into this category.

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

Other approaches in the literature on developing nonparametric data-driven algorithms include online convex optimization (Agarwal et al. 2011, Zinkevich 2003, Hazan et al. 2006), continuum-armed bandit problems (Auer et al. 2007, Kleinberg 2005, Cope 2009), and stochastic approximation (Kiefer and Wolfowitz 1952, Lai and Robbins 1981, and Robbins and Monro 1951). In fact, Burnetas and Smith (2000) is an example of implementing such algorithms to the joint pricing and inventory control problem. However, these methodologies require that the proposed solution be reachable in each and every period, which is not the case with our problem. This is because, in a demand learning algorithm of joint pricing/inventory control problem, in each period the algorithm utilizes the past demand data to prescribe a pricing decision and an order up-to level. However, if the starting inventory level of the period is already higher than the prescribed order up-to level, then the prescribed inventory level for the period cannot be reached. Actually, that is precisely the reason that Burnetas and Smith (2000) focused on the case of perishable product (hence the firm has no carry-over inventory and the inventory decision obtained by Burnetas and Smith (2000) based on multi-armed bandit process can be implemented in each period). Agarwal et al. (2011), Auer et al. (2007), and Kleinberg (2005) propose learning algorithms and obtain regrets that are not as good as ours in this paper. Zinkevich (2003) and Hazan et al. (2007) present machine learning algorithms in which the the exact gradient of the unknown objective function at the current decision can be computed, and their results have been applied to dynamic inventory control in Huh and Rusmevichientong (2009). However, in the joint pricing and inventory control problem with unknown demand response, the gradient of the unknown objective function cannot be obtained thus the method cannot be applied.

### 1.2 Positioning of this paper

The closest related research works to ours are Besbes and Zeevi (2015), Levi et al. (2007) and Levi et al. (2010), offering nonparametric approaches to pure pricing problem (with no inventory) and pure inventory control problem (with no pricing), respectively.

Besbes and Zeevi (2015) consider a dynamic pricing problem in which a firm chooses its selling price to maximize expected revenue. The firm does not know the deterministic demand curve (i.e., how the average demand changes in price) and learns it through noisy demand realizations, and the authors establish the sufficiency of linear approximations in maximizing revenue. They assume that the firm has infinite supply of inventory, or, alternatively, the seller has no inventory constraint. In this case, since the expected revenue in each period depends only on its mean demand, the distribution of random error is immaterial in their learning algorithm and analysis. On the other hand, in the dynamic newsvendor problem considered in Levi et al. (2007, 2010), the essence for effective inventory management is to strike a balance between overage cost and underage cost, for which the distribution of uncertain demand plays a key role. Levi et al. (2007) and Levi et al. (2010) apply Sample Average Approximation (SAA) to estimate the demand distribution and average cost function, and they analyze the relationship between sample sizes and accuracy of estimations and inventory decisions.

Our problem has both dynamic pricing and inventory control, and the firm knows neither the relationship between demand and selling price nor the distribution of demand uncertainty. In Besbes and Zeevi (2015), the authors only need to estimate the average demand curve in order to maximize revenue, and demand distribution information is irrelevant. In a remark, Besbes and Zeevi (2015) state that their method of learning the demand curve can be applied to maximizing more general forms of objective functions beyond the expected revenue which, however, does not apply to our setting. This is because, in the general form presented in Besbes and Zeevi (2015), the objective function still has to be a known function in terms of price and the demand curve for a given price and a given demand curve. Thus the firm must know the exact expression of the objective function when the estimate of a demand curve is given. In our problem, even with a given price and inventory level and a given demand curve, the objective function cannot be written as a known deterministic function. Indeed, this function contains the expected inventory holding and backorder costs that depend on the distribution of demand fluctuation, which is also unknown to the firm. In fact, the latter is a major technical challenge encountered in this paper because, as we will explain below, the estimation of the demand uncertainty, therefore also of the expected holding/shortage cost, cannot be decoupled with the estimation of the average demand curve, which is gathered through price experimentation.

Standard SAA method is implemented to the newsvendor problem by Levi et al. (2007) and Levi et al. (2010) which, however, cannot be applied to our setting for determining inventory decisions. In Levi et al. (2007) and Levi et al. (2010), dynamic inventory control is studied in which pricing is not a decision and it is assumed (implicitly) to be given. The only information the firm is uncertain about is the distribution of random fluctuation. Therefore, the firm can observe true realizations of demand fluctuation which are used to build an empirical distribution. In our model, however, the firm knows neither how average demand responds to the selling price (demand curve) nor the distribution of fluctuating demand, but both of them affect demand realizations. For any estimation of average demand curve, the error of this estimate will affect the estimation of distribution of random demand fluctuation. Hence, through the realization of random demand we are unable to obtain a true realization of random demand error without knowing the exact average demand function. As a result, the standard SAA analysis is not applicable in our setting because unbiased samples of the random error cannot be obtained.

Because the firm does not know the exact demand curve *a priori*, its estimate of error distribution using demand data is inevitably biased, and as a result, the data-driven optimization problem constructed to compute the pricing and ordering strategies is also biased. Because of this bias, it is no longer true that the solution of the data-driven problem using SAA must converge to the true optimal solution. Fortunately, we are able to show that as the learning algorithm proceeds, the biases will be gradually diminishing and that allows us to prove that our learning algorithm still converges to the true optimal solution. This is done by establishing several important properties of the newsvendor problem that bound the errors of biased samples. One main contribution of this paper is to explicitly prove that the solution obtained from a biased data-driven optimization problem still converges to the true optimal solution.

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

### 1.3 Organization

The rest of this paper is organized as follows. Section 2 formulates the problem and describes the data-driven learning algorithm for pricing and inventory control decisions. The following two sections (Sections 3 and 4) present our major theoretical results together with a numerical study, and the main steps of the technical proofs, respectively. The paper concludes with a few remarks in Section 5. Finally, the details of the mathematical proofs are given in the Appendix.

## 2 Formulation and Learning Algorithm

We consider an inventory system in which a firm (e.g., a retailer) sells a nonperishable product over a planning horizon of $T$ periods. At the beginning of each period $t$, the firm makes a replenishment decision, denoted by the order-up-to level, $y_t$, and a pricing decision, denoted by $p_t$, where $y_t \in \mathcal{Y} = [y^l, y^h]$ and $p_t \in \mathcal{P} = [p^l, p^h]$ for some known lower and upper bounds of inventory level and selling price, respectively. We assume $p^h > p^l$ since otherwise, the problem is the pure inventory control problem and learning algorithms have been developed in Huh and Rusmevichientong (2009), Levi et al. (2007), and Levi et al. (2010). During period $t$ and when the selling price is set to $p_t$, a random demand, denoted by $\tilde{D}_t(p_t)$, is realized and fulfilled from on-hand inventory. Any leftover inventory is carried over to the next period, and in case the demand exceeds $y_t$, the unsatisfied demand is backlogged. The replenishment leadtime is zero, i.e., an order placed at the beginning of a period can be used to satisfy demand in the same period. Let $h$ and $b$ be the unit holding and backlog costs per period, and the unit purchasing cost is assumed, without loss of generality, to be zero.

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

In general, the demand in period $t$ is a function of selling price $p_t$ in that period and some random variable $\tilde{\epsilon}_t$, and it is stochastically decreasing in $p_t$. The most popular demand models in the literature are the additive demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t) + \tilde{\epsilon}_t$ and multiplicative demand model $\tilde{D}_t(p_t) = \tilde{\lambda}(p_t)\ \tilde{\epsilon}_t$, where $\tilde{\lambda}(\cdot)$ is a strictly decreasing deterministic function and $\tilde{\epsilon}_t, t = 1, 2, \ldots, T$, are independent and identically distributed random variables. In this paper, we shall study both additive and the multiplicative demand models. However, the firm knows neither the function $\tilde{\lambda}(p_t)$ nor the distribution function of random variable $\tilde{\epsilon}_t$. The firm has to learn from historical demand data, that are the realizations of market responses to offered prices, and use that information as a basis for decision making. Suppose $\tilde{\epsilon}_t$ has finite support $[l, u]$, with $l \geq 0$ for the case of multiplicative demand.

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

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

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

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

We develop a data-driven learning algorithm to compute the inventory control and pricing policy. It will be shown in Section 3 that the average profit of the algorithm converges to that of the case when complete demand distribution information is known a priori, and that the pricing and inventory control parameters also converge to that of the optimal control policy for the case with complete information as the planning horizon becomes long. To save space we shall only present the algorithm and analytical results for the multiplicative demand model. The results and analyses for the additive demand case are analogous, and we only highlight the main differences at the end of this section.

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

**Case of complete information about demand.** In the case of complete information in which the firm knows $\tilde{\lambda}(\cdot)$ and the distribution of $\tilde{\epsilon}_t$, it follows from (1) that, if $(p^*, y^*)$ is the optimal solution of each individual term

$$\max_{p\in\mathcal{P},y\in\mathcal{Y}}\Big\{p\mathbb{E}[\tilde{D}_t(p)] - h\mathbb{E}[y-\tilde{D}_t(p)]^+ - b\mathbb{E}[\tilde{D}_t(p)-y]^+\Big\}. \tag{2}$$

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

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

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

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

$$D_t(p_t) = \lambda(p_t) + \epsilon_t, \qquad t = 1,\ldots,T. \tag{3}$$

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

For convenience, let $\epsilon$ be a random variable distributed as $\epsilon_1$. In terms of $\lambda(\cdot)$ and $\epsilon$, we define

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

Then problem (2) can be re-written as

**Problem CI:**

$$\begin{aligned}&\max_{p\in\mathcal{P},y\in\mathcal{Y}} G(p,y) \\ =\ &\max_{p\in\mathcal{P}}\Big\{pe^{\lambda(p)}\mathbb{E}\big[e^{\epsilon}\big] - \min_{y\in\mathcal{Y}}\Big\{h\mathbb{E}\big[y-e^{\lambda(p)}e^{\epsilon}\big]^+ + b\mathbb{E}\big[e^{\lambda(p)}e^{\epsilon}-y\big]^+\Big\}\Big\}.\end{aligned} \tag{4}$$

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

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

**Learning algorithm.** In the absence of the prior knowledge about the demand process, the firm needs to collect the demand information necessary to estimate $\lambda(p)$ and the empirical distribution of random error $\epsilon$, thus price and inventory decisions not only affect the profit but also the demand information realized. The major difficulty lies in that, the estimations of demand-price curve $\lambda(p)$ and the distribution of random error cannot be decoupled. This is because, the firm only observes realized demands, hence with any estimation of demand-price curve, the estimation error transfers to the estimation of the random error distribution. Indeed, we are not even able to obtain unbiased samples of the random error $\epsilon_t$.

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

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

$$I_i = \lfloor I_0 v^i \rfloor, \quad \delta_i = \rho(2I_{i-1})^{-\frac{1}{4}}, \text{ and } t_i = \sum_{k=1}^{i-1} 2I_k \text{ with } t_1 = 0, \tag{5}$$

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

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

**Data-Driven Algorithm (DDA)**

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

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

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

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

$$y_t = \max\{\hat{y}_{i1}, x_t\}, \qquad t = t_i + 1, \dots, t_i + I_i,$$
$$y_t = \max\{\hat{y}_{i2}, x_t\}, \qquad t = t_i + I_i + 1, \dots, t_i + 2I_i.$$

**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 $t = t_i + 1, \dots, t_i + 2I_i$, and compute

$$(\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}) = \operatorname*{argmin}_{\alpha,\beta} \left\{ \sum_{t=t_i+1}^{t_i+2I_i} \Big( D_t - (\alpha - \beta p_t) \Big)^2 \right\}, \tag{6}$$

$$\eta_t = D_t - (\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p_t), \qquad \text{for } t = t_i + 1, \dots, t_i + 2I_i. \tag{7}$$

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

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

Then the data-driven optimization is defined by

**Problem DD:**

$$\max_{(p,y)\in\mathcal{P}\times\mathcal{Y}} G^{DD}_{i+1}(p, y) = \max_{p\in\mathcal{P}} \left\{ p e^{\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p} \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\eta_t} - \min_{y\in\mathcal{Y}} \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \left( h \left( y - e^{\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p} e^{\eta_t} \right)^+ + b \left( e^{\hat{\alpha}_{i+1} - \hat{\beta}_{i+1} p} e^{\eta_t} - y \right)^+ \right) \right\} \right\}. \tag{8}$$

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 $\hat{p}_{i+1}-\delta_{i+1}$.

**Remark 2.** When $\hat{\beta}_{i+1}>0$, the objective function in (8) after minimizing over $y\in\mathcal{Y}$ is unimodal in $p$. To see why this is true, let $d=e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p}$ and thus $p=\frac{\hat{\alpha}_{i+1}-\log d}{\hat{\beta}_{i+1}}$ with $d\in\mathcal{D}=[d^l,d^h]$, where $d^l=e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p^h}$ and $d^h=e^{\hat{\alpha}_{i+1}-\hat{\beta}_{i+1}p^l}$. Then the optimization problem (8) is equivalent to

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

The objective function of this optimization problem is jointly concave in $(y,d)$ hence it is concave in $d$ after minimizing over $y\in\mathcal{Y}$. Thus, it follows from $p=\frac{\hat{\alpha}_{i+1}-\log d}{\hat{\beta}_{i+1}}$ is strictly decreasing in $d$ that the objective function in (8) (after minimization over $y$) is unimodal in $p\in\mathcal{P}$.

**Remark 3.** In Step 3 of DDA, the second price is set to $\hat{p}_{i+1}-\delta_{i+1}$ when $\hat{p}_{i+1}+\delta_{i+1}>p^h$. In this case our condition $\rho\le 2^{-3/4}(p^h-p^l)I_0^{1/4}$ 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 $\rho$.

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

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

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

In the traditional SAA, mathematical expectations are replaced by sample means, see e.g., Kleywegt et al. 2001). Levi et al. (2007) and Levi et al. (2010)) apply SAA method in dynamic newsvendor problems. The argument above shows that the traditional analyses that show SAA leads to the optimal solution is not applicable to our setting. Indeed, in our inner layer optimization, we face a newsvendor problem for which the firm needs to balance holding and shortage cost, and the knowledge about demand distribution is critical. However, the lack of samples of random error $\epsilon_t$ makes the inner loop optimization problem significantly different from Levi et al. (2007) and Levi et al. (2010)), which consider pure inventory control problems and samples of random errors are available for applications of SAA result and analysis. Because of this, it is not guaranteed that the SAA method will lead to a true optimal solution.

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

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

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

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

$$R(\pi, T) = G(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t, y_t)\right]. \tag{9}$$

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

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

# 3 Main Results

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

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

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

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

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

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

**Example 1.** The following functions satisfy Assumption 1.

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

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

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

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

**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 both pricing and ordering decisions from the DDA algorithm converge to the *true* optimal solution $(p^*, y^*)$ in probability. Note that the convergence of inventory decision $y_t \to y^*$ is stronger than the convergence of order up-to levels $\hat{y}_{i,1} \to y^*$ and $\hat{y}_{i,2} \to y^*$. This is because, the order up-to levels may or may not be achievable for each period, thus the resulting inventory levels may "overshoot" the targeting order up-to levels. Theorem 1 shows that, despite these overshoots, the realized inventory levels converge to the true optimal solution in probability.

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

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

$$R(DDA, T) = G(p^*, y^*) - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} G(p_t, y_t)\right] \le KT^{-\frac{1}{2}}. \tag{10}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Then we compute the average profit loss per period over the 500 draws and report them in Table 1. In all the experiments, we set $p^l = 0.51, p^h = 4, y^l = 0, y^h = 3, b = 1, h = 0.1, I_0 = 1$, and initial price $\hat{p}_1 = 1$, initial inventory order up-to level $\hat{y}_{11} = 1$, $\hat{y}_{12} = 0.3$. We test two values of $\rho$, $\rho = 0.5$ and $\rho = 0.75$, and two values of $v$, namely, $v = 1.3$ and $v = 2$.

Table 1 summarizes the results when the underlying demand curve is exponential, and Table 2 displays the results when the underlying demand curve is logit. Combining both tables, one sees that when $T = 100$ periods, on average the profit loss from the DDA algorithm falls between 11%

Table 1: Exponential Demand

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

and 14% compared to the optimal profit under complete information, in which DDA starts with no prior knowledge about the underlying demand. When $T = 500$, the profit loss is further reduced to between 5% and 8%. The performance gets better and better when $T$ becomes larger. Also, it is seen from the table that the overall performance of algorithm is better when the variance of the demand is smaller, which is intuitive.

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

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

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

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

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

The learning algorithm for the additive demand model is similar to that of the multiplicative

Table 2: Logit Demand

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

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

# 4 Sketches of the Proof

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

## 4.1 Technical issues encountered

To prove Theorem 1, we will need to show

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \to 0, \quad \mathbb{E}\big[(\hat{p}_{i+1} + \delta_{i+1} - p^*)^2\big] \to 0, \quad \text{as } i \to \infty; \tag{11}$$

$$\mathbb{E}[(y^* - \hat{y}_{i+1,1})^2] \to 0, \quad \mathbb{E}[(y^* - \hat{y}_{i+1,2})^2] \to 0, \quad \text{as } i \to \infty, \tag{12}$$

where $p^*$ is the optimal solution of

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

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

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

However, both $Q(\cdot,\cdot)$ and $J(\cdot)$ are *unknown* to the firm because all the expectations cannot be computed. To estimate $J(\cdot)$, in (8) of the learning algorithm we use the data-driven biased estimation of

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

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

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

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

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

We note that in the revenue management problem setting, Besbes and Zeevi (2015) also prove the convergence result (11). In Besbes and Zeevi (2015), $p^*$ is the optimal solution of $\max_{p\in\mathcal{P}} Q(p,\lambda(p))$, and $\hat{p}_{i+1}$ is the optimal solution of $\max_{p\in\mathcal{P}} Q(p,\hat{\alpha}_{i+1} - \hat{\beta}_{i+1}p)$, where $Q(\cdot,\cdot)$ is a *known* and *deterministic* function $Q(p,\lambda(p)) = p\lambda(p)$. As Besbes and Zeevi (2015) point out, their analysis extends to more general function $Q(p, \lambda(p))$ in which $Q(\cdot,\cdot)$ is a known deterministic function. This, however, is not true in our setting as $Q(\cdot,\cdot)$ is not known, and as a matter of fact, one cannot even find an unbiased sample average to estimate $Q(\cdot,\cdot)$. Therefore, the challenges discussed above were not present in Besbes and Zeevi (2015).

### 4.2 Main ideas of the proof

To compare the policy and the resultant profit of DDA algorithm with that of the optimal solution, we first note that these two problems differ along several dimensions. For example, in DDA we approximate $\lambda(p)$ by an affine function and estimate the parameters of the affine function in each iteration, and we approximate the expected revenue and the expected holding and shortage costs using biased sample averages. These differences make the direct comparison of the two problems difficult. Therefore, we introduce several "intermediate" bridging problems, and in each step we compare two "adjacent" problems that differ just in one dimension.

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

$$\breve{\alpha}(z) = \lambda(z) - \lambda'(z)z, \qquad \breve{\beta}(z) = -\lambda'(z), \qquad z \in \mathcal{P}. \tag{13}$$

We proceed to prove (11) as follows:

$$\begin{aligned}\mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \;\leq\; & \mathbb{E}\Bigg[\Bigg( \underbrace{\Big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|}_{\substack{\text{Comparison of problems CI and B1}\\ \text{Lemma A1}}} \\ & + \underbrace{\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\Big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\Big)\Big|}_{\substack{\text{Comparison of problems B1 and B2}\\ \text{Lemma A5}}} + \underbrace{\Big|\tilde{p}_{i+1}\Big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\Big) - \hat{p}_{i+1}\Big|}_{\substack{\text{Comparison of problems B2 and DD}\\ \text{Lemma A6 and Lemma A7}}} \Bigg)^2\Bigg],\end{aligned} \tag{14}$$

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

**Bridging Problem B1:**

$$\max_{p \in \mathcal{P}} \left\{ p e^{\alpha - \beta p} \mathbb{E}\big[e^{\epsilon}\big] - \min_{y \in \mathcal{Y}} \left\{ h\mathbb{E}\Big[y - e^{\alpha-\beta p} e^{\epsilon}\Big]^+ + b\mathbb{E}\Big[e^{\alpha - \beta p} e^{\epsilon} - y\Big]^+ \right\} \right\}, \tag{15}$$

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

**Bridging Problem B2:**

$$\begin{aligned}\max_{p \in \mathcal{P}} \Bigg\{ & p e^{\alpha - \beta p} \left( \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t} \right) \\ & - \min_{y \in \mathcal{Y}} \left\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big( h\big(y - e^{\alpha-\beta p} e^{\epsilon_t}\big)^+ + b\big(e^{\alpha-\beta p}e^{\epsilon_t} - y\big)^+ \Big) \right\} \Bigg\}.\end{aligned} \tag{16}$$

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

Comparing (15) with (4), it is seen that problem B1 simplifies problem CI by replacing the demand-price function $\lambda(p)$ by a linear function $\alpha - \beta p$, while problem B2 is obtained from problem B1 after replacing the mathematical expectations in problem B1 by their sample averages, i.e., problem B2 is the SAA of problem B1. Comparing (16) with (8), it is noted that problems B2 and DD differ in the coefficients of the linear function as well as the arithmetic averages. More specifically, in B2 the real random error samples $\epsilon_t$, $t = t_i + 1, \ldots, t_i + 2I_i$, are used, while in problem DD, biased error samples $\eta_t$ are used in place of $\epsilon_t$, $t = t_i + 1, \ldots, t_i + 2I_i$. Furthermore, note that the optimal prices for problems CI and B1, $p^*$ and $\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$, are deterministic, but the optimal solutions of problems B2 and DD, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\hat{p}_{i+1}$, are random. Specifically, $\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ is random because $\epsilon_t$ is random, while $\hat{p}_{i+1}$ is random due to demand uncertainty from periods 1 to $t_{i+1}$. Hence, to show the right hand side of (14) converges to 0, we will first develop an upper bound for $\big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\big|$ by comparing problems CI and B1, and the result is presented in Lemma A1. Since $\tilde{p}_{i+1}(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)$ is random, we compare the two problems B1 and B2 and show the probability that $\big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\big|$ exceeds some small number diminishes to 0 in Lemma A5. Similarly, in Lemma A6 and Lemma A7 we compare problems B2 and DD and show the probability that $\big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\big|$ exceeds some small number also diminishes to 0. Finally, we combine these several results to complete the proof of (11). The idea for proving (12) is similar, and that also relies heavily on the two bridging problems (Lemmas A6, A7, and A8). The detailed proofs for Theorem 1 and Theorem 2 are given in Subsections 4.3 and 4.4.

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

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

**Regularity Conditions:**

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

$$\left|\overline{y}(e^{\lambda(q_1)}) - \overline{y}(e^{\lambda(q_2)})\right| \leq K_1 \left|q_1 - q_2\right|, \tag{17}$$

$$\left|\overline{y}\big(e^{\breve{\alpha}(q_1)-\breve{\beta}(q_1)p}\big) - \overline{y}\big(e^{\breve{\alpha}(q_2)-\breve{\beta}(q_2)p}\big)\right| \leq K_1 \left|q_1 - q_2\right|. \tag{18}$$

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

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

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

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

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

### 4.3 Proof of Theorem 1

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

**Convergence of pricing decisions.** To prove the convergence of pricing decisions, we continue the development in (14) as follows:

$$\begin{aligned}
&\mathbb{E}\big[(p^* - \hat{p}_{i+1})^2\big] \\
\leq\ & \mathbb{E}\Big[\Big(\Big|p^* - \overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\Big] \\
\leq\ & \mathbb{E}\Big[\Big(\gamma|p^* - \hat{p}_i| + \Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\Big] \\
\leq\ & \left(\frac{1+\gamma^2}{2}\right) \mathbb{E}\left[(p^* - \hat{p}_i)^2\right] \\
&\qquad + K_2 \mathbb{E}\left[\Big(\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big| + \Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|\Big)^2\right] \\
\leq\ & \left(\frac{1+\gamma^2}{2}\right) \mathbb{E}\left[(p^* - \hat{p}_i)^2\right] \\
& + K_3 \mathbb{E}\left[\Big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\Big|^2\right] + K_3 \mathbb{E}\left[\Big|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\Big|^2\right],
\end{aligned} \tag{19}$$

where the first inequality follows from the expansion in (14), the second inequality follows from Lemma A1, and the third inequality is justified by $\gamma < 1$ in Lemma A1 and some constant $K_2$, and the last inequality holds for some appropriately chosen $K_3$ because of the inequality $(a+b)^2 \leq 2(a^2+b^2)$ for any real numbers $a$ and $b$.

To bound $\mathbb{E}\big[\big|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\big|^2\big]$ in (19), by Lemma A5 one has, for some constant $K_4$,

$$\mathbb{E}\left[\left|\overline{p}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)\right|^2\right] \leq K_4^2 \int_0^{+\infty} 5e^{-4I_i\xi^2} d\xi = \frac{5\pi^{\frac{1}{2}} K_4^2}{4I_i^{\frac{1}{2}}}. \tag{20}$$

And to bound $\mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|^2\right]$ in (19), by Lemma A6 and Lemma A7, when $i$ is large enough (greater than or equal to $i^*$ defined in the proof of Lemma A7), for some positive constants $K_5$, $K_6$, and $K_7$ one has

$$\begin{aligned}
& \mathbb{E}\left[\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) - \hat{p}_{i+1}\right|^2\right] \\
\leq\ & \mathbb{E}\Big[K_5^2\big(\big|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}\big| + \big|\breve{\alpha}(\hat{p}_i+\delta_i) - \hat{\alpha}_{i+1}\big| + \big|\breve{\beta}(\hat{p}_i+\delta_i) - \hat{\beta}_{i+1}\big|\big)^2\Big] \\
& \qquad + \frac{8}{I_i}\big(p^h - p^l\big)^2 \\
\leq\ & \mathbb{E}\left[K_6\left(|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + |\breve{\alpha}(\hat{p}_i+\delta_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i+\delta_i) - \hat{\beta}_{i+1}|^2\right)\right] \\
& \qquad + \frac{8}{I_i}\big(p^h - p^l\big)^2 \\
\leq\ & K_7 I_i^{-\frac{1}{2}}.
\end{aligned} \tag{21}$$

Substituting (20) and (21) into (19), one has

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

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

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \quad \leq \quad \theta^i(\hat{p}_1 - p^*)^2 + K_8 \sum_{j=0}^{i-1} \theta^j I_{i-j}^{-\frac{1}{2}} \leq K_9 (v^{-\frac{1}{2}})^i \sum_{j=0}^{i-1} \theta^j (v^{\frac{1}{2}})^j. \tag{22}$$

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

$$\mathbb{E}\big[(\hat{p}_{i+1} - p^*)^2\big] \leq K_{11}(v^{-\frac{1}{2}})^i \leq K_{12} I_i^{-\frac{1}{2}}. \tag{23}$$

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

$$\mathbb{E}\left[(\hat{p}_{i+1} + \delta_{i+1} - p^*)^2\right] \leq 2\mathbb{E}\left[(\hat{p}_{i+1} - p^*)^2\right] + 2\delta_{i+1}^2 \leq K_{13} I_i^{-\frac{1}{2}} \to 0, \quad \text{as } i \to \infty. \tag{24}$$

This completes the proof of (11). Because mean-square convergence implies convergence in probability, this shows that the pricing decisions from DDA converge to $p^*$ in probability.

**Convergence of inventory decisions.** To prove $y_t$ converges to $y^*$ in probability, we first prove the convergence of order up-to levels (12). For some constant $K_{14}$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|y^* - \hat{y}_{i+1,1}\right|^2\right] \\
\leq\ & \mathbb{E}\Big[\Big(\left|\overline{y}\big(e^{\lambda(p^*)}\big) - \overline{y}\big(e^{\lambda(\hat{p}_{i+1})}\big)\right| + \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| \\
&\quad + \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| \\
&\quad + \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| + \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|\Big)^2\Big] \\
\leq\ & K_{14}\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}} \qquad (25) \\
&\quad + \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} \\
&\quad + \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{Lemma A8}}} + \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}_{\substack{\text{Comparison of problems B2 and DD} \\ \text{Lemma A6 and Lemma A7}}}\Big)\Big].
\end{aligned}$$

We want to upper bound each term on the right hand side of (25). First, it follows from (17) that, for some constant $K_{15}$ it holds

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

By definition of $\breve{\alpha}(p)$ and $\breve{\beta}(p)$ in (13) one has $\breve{\alpha}(\hat{p}_{i+1}) - \breve{\beta}(\hat{p}_{i+1})\hat{p}_{i+1} = \lambda(\hat{p}_{i+1})$, thus the second term on the right hand side of (25) vanishes. For the third term, we apply the Lipschitz condition on $\overline{y}\big(e^{\breve{\alpha}(q)-\breve{\beta}(q)p}\big)$ in (18) to obtain, for some constants $K_{16}$ and $K_{17}$,

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

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

$$\mathbb{E}\left[\left|\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big) - \tilde{y}_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\big)\right|^2\right] \leq K_{18}^2\int_0^{+\infty} 2e^{-4I_i\xi}d\xi \leq \frac{K_{19}}{I_i}, \qquad (26)$$

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

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

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

$$
\begin{aligned}
& \mathbb{E}\left[\left(y^* - \hat{y}_{i+1,1}\right)^2\right] \\
\leq \; & K_{21}\mathbb{E}\big[\,|\,p^* - \hat{p}_{i+1}\,|^2 + |\,p^* - \hat{p}_i\,|^2\,\big] + K_{21} I_i^{-\frac{1}{2}} \\
\leq \; & K_{22} I_i^{-\frac{1}{2}} \qquad (27) \\
\rightarrow \; & 0 \text{ as } i \rightarrow \infty,
\end{aligned}
$$

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

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

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

The inventory order up-to level prescribed from DDA for periods $t \in \{t_{i+1} + 1, \ldots, t_{i+1} + I_{i+1}\}$ is $\hat{y}_{i+1,1}$. This, however, may not be achievable for some period $t$. Consider the event that the second order up-to level of learning stage $i$, $\hat{y}_{i,2}$, is achieved during periods $\{t_i + I_i + 1, \ldots, t_i + 2I_i\}$. Since $\tilde{\lambda}(p^h)l \leq D_t \leq \tilde{\lambda}(p^l)u$, it follows from Hoeffding inequality[4] that for any $\zeta > 0$,

$$
\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq \mathbb{E}\left[\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t\right] - \zeta\right\} \geq 1 - \exp\left(-\frac{2\zeta^2}{I_i(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l)^2}\right). \qquad (28)
$$

Let $\zeta = \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}}(\log I_i)^{\frac{1}{2}}$ in (28), then one has

$$
\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq I_i\mathbb{E}\left[D_{t_i+I_i+1}\right] - \left(\tilde{\lambda}(p^l)u - \tilde{\lambda}(p^h)l\right)(I_i)^{\frac{1}{2}}(\log I_i)^{\frac{1}{2}}\right\} \geq 1 - \frac{1}{I_i^2}. \qquad (29)
$$

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

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

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

Hence it follows from (29) that, when $i$ is large enough, we will have

$$\mathbb{P}\left\{\sum_{t=t_i+I_i+1}^{t_i+2I_i} D_t \geq \frac{1}{2} I_i \mathbb{E}\left[D_{t_i+I_i+1}\right]\right\} \geq 1-\frac{1}{I_i^2}. \tag{30}$$

Define event

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

then (30) can be rewritten as

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

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

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

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

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

### 4.4 Proof of Theorem 2

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

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

We have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\big(G(p^*,y^*)-G(p_t,y_t)\big)\right]\\
\le\;&\mathbb{E}\Bigg[\sum_{i=1}^{n}\Bigg(\sum_{t=t_i+1}^{t_i+I_i}\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big)\\
&\qquad+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})+G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\Bigg)\Bigg]\\
=\;&\mathbb{E}\left[\sum_{i=1}^{n}I_i\big(G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})+G(p^*,y^*)-G(\hat{p}_i+\delta_i,\hat{y}_{i,2})\big)\right]\\
&+\mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+I_i}\big(G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big)+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\right)\right],\quad(31)
\end{aligned}
$$

where $n$ is the smallest number of stages that cover $T$, i.e., $n$ is the smallest integer such that $2I_0\sum_{i=1}^{n}v^i\ge T$, and it satisfies $\log_v\Big(\frac{v-1}{2I_0v}T+1\Big)\le n<\log_v\Big(\frac{v-1}{2I_0v}T+1\Big)+1$. The inequality in (31) follows from that the right hand side includes $2I_0\sum_{i=1}^{n}v^i$ periods which is greater than or equal to $T$.

The first expectation on the right hand side of (31) is with respect to the sum of the difference between profit values of DDA decisions and the optimal solution, hence its analysis relies on the convergence rate of DDA policies; these are demonstrated in (23), (24), and (27). The second expectation on the right hand side of (31) stems from the fact that in the process of implementing DDA, it may happen that the inventory decisions from DDA are not implementable. This issue arises in learning algorithms for nonperishable inventory systems and it presents additional challenges in evaluating the regret. We note that in Huh and Rusmevichientong (2009), a queueing approach is employed to resolve this issue for a pure inventory system with no pricing decisions.

To develop an upper bound for $G(p^*,y^*)-G(\hat{p}_i,\hat{y}_{i,1})$ in (31), we first apply Taylor expansion on $G(p,\overline{y}(e^{\lambda(p)})$ at point $p^*$. Using the fact that the first order derivative vanishes at $p=p^*$ and the assumption that the second order derivative is bounded (regularity condition (ii)), we obtain, for some constant $K_{23}>0$, that

$$G\big(p^*,\overline{y}\big(e^{\lambda(p^*)}\big)\big)-G\big(\hat{p}_i,\overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big)\le K_{23}(p^*-\hat{p}_i)^2. \qquad (32)$$

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

$$G\big(\hat{p}_i,\overline{y}\big(e^{\lambda(\hat{p}_i)}\big)\big)-G(\hat{p}_i,\hat{y}_{i,1})\le K_{24}\big(\overline{y}\big(e^{\lambda(\hat{p}_i)}\big)-\hat{y}_{i,1}\big)^2. \qquad (33)$$

In addition, we have

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

This is similar to the right hand side of (25) except that $i+1$ is replaced by $i$. Thus, using the same analysis as that for (25), we obtain

$$
\mathbb{E}\left[(\overline{y}(e^{\lambda(\hat{p}_i)}) - \hat{y}_{i,1})^2\right] \leq K_{26} I_{i-1}^{-\frac{1}{2}} \tag{34}
$$

for some constant $K_{26}$.

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

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

where the first inequality follows from (32) and (33), and the second inequality follows from the convergence rate of pricing decisions (23) and (34).

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

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

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

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

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

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

We next evaluate the second term of (31), i.e.,

$$
\mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+I_i}\big(G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big)+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\right)\right]. \tag{36}
$$

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

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

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

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

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

$$\mathbb{P}\left\{\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t \ge \mathbb{E}\left[\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t\right]-\zeta\right\} \ge 1-\exp\left(-\frac{2\zeta^2}{\left\lfloor I_i^{\frac{1}{2}}\right\rfloor(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l)^2}\right). \tag{37}$$

Let $\zeta=\left(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l\right)\left(\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}$, then it follows from (37) that

$$\begin{aligned}&\mathbb{P}\left\{\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t \ge \left\lfloor I_i^{\frac{1}{2}}\right\rfloor\mathbb{E}\left[D_{t_i+1}\right]-\left(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l\right)\left(\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\right\}\\ \ge\quad & 1-\frac{1}{\left\lfloor I_i^{\frac{1}{2}}\right\rfloor^2}.\end{aligned} \tag{38}$$

By regularity condition (iii), $\mathbb{E}\left[D_{t_i+1}\right]>0$. Thus, when $i$ is large enough, say greater than or equal to some number $i^{**}$, we will have

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

Based on (38), we define event $\mathcal{A}_2$ as

$$\mathcal{A}_2=\left\{\sum_{t=t_i+1}^{t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor} D_t \ge \left\lfloor I_i^{\frac{1}{2}}\right\rfloor\mathbb{E}\left[D_{t_i+1}\right]-(\tilde{\lambda}(p^l)u-\tilde{\lambda}(p^h)l)\left(\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\left(\log\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right)^{\frac{1}{2}}\right\}. \tag{39}$$

Then (38) can be restated as

$$\mathbb{P}(\mathcal{A}_2) \ge 1-\frac{1}{\left\lfloor I_i^{\frac{1}{2}}\right\rfloor^2}. \tag{40}$$

On the event $\mathcal{A}_2$, the inventory order up-to level $\hat{y}_i$ will be achieved after periods

$\left\{t_i+1,\ldots,t_i+\left\lfloor I_i^{\frac{1}{2}}\right\rfloor\right\}$. By (40), we have

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

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

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

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

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

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

$$\begin{aligned}
& \mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{t=t_i+1}^{t_i+I_i}\big(G(\hat{p}_i,\hat{y}_{i,1})-G(p_t,y_t)\big)+\sum_{t=t_i+I_i+1}^{t_i+2I_i}\big(G(\hat{p}_i+\delta_i,\hat{y}_{i,2})-G(p_t,y_t)\big)\right)\right] \\
\leq \quad & \sum_{i=1}^{i^{**}}\max\{h,b\}(y^h-y^l)I_i+\sum_{i=i^{**}+1}^{n}4\max\{h,b\}(y^h-y^l)I_i^{\frac{1}{2}} \\
\leq \quad & K_{36}+4\max\{h,b\}(y^h-y^l)I_0^{\frac{1}{2}}\frac{v^{\frac{1}{2}}(1-(v^{\frac{1}{2}})^n)}{1-v^{\frac{1}{2}}} \\
\leq \quad & K_{36}+K_{37}(v^{\frac{1}{2}})^{n+1} \\
\leq \quad & K_{36}+K_{37}v^{\log_v(\frac{v-1}{2I_0v}T+1)^{\frac{1}{2}}} \\
\leq \quad & K_{38}T^{\frac{1}{2}}. \qquad (41)
\end{aligned}$$

By combining (35) and (41), we conclude

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

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

# 5 Conclusion

In this paper, we consider a joint pricing and inventory control problem when the firm does not have prior knowledge about the demand distribution and customer response to selling prices. We impose virtually no explicit assumption about how the average demand changes in price (other than the fact that it is decreasing) and on the distribution of uncertainty in demand. This paper is the first to design a nonparametric algorithm data-driven learning algorithm for dynamic joint pricing and inventory control problem and present the convergence rate of policies and profits to those of the optimal ones. The regret of the learning algorithm converges to zero at a rate that is the theoretical lower bound $O(T^{-1/2})$.

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

*Acknowledgment:* The authors are grateful to the Department Editor, the Associate Editor, and two referees for their constructive comments on an earlier version of this paper, that have helped to significantly improve the clarity and exposition of this paper.

# 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] 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.

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

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

[5] 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.

[6] 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.

[7] Chung BD, Li J, Yao T (2011) Dynamic Pricing and Inventory Control with Nonparametric Demand Learning. *Int. J. Services Operations and Informatics* 6(3): 259-271.

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

[9] 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.

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

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

[12] Heyman D, Sobel M (1984) *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization.* McGraw-Hill, New York.

[13] Hoeffding W (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58:13-30.

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

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

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

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

[18] 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.

[19] 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.

[20] Levi R, Perakis G, Uichanco J (2010) The data-driven newsvendor problem: new bounds and insights. Working Paper. Massachusetts Institute of Technology, Cambridge, MA.

[21] 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.

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

[23] Petruzzi NC, Dada M (2002) Dynamic pricing and inventory control with learning. *Naval Res. Logist.* 49:303-325.

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

[25] Subrahmanyan S, Shoemaker R (1996) Developing optimal pricing and inventory policies for retailers who face uncertain demand. *Journal of Retailing* 72(1):7-30.

[26] Whitin TM (1955) Inventory control and price theory. *Management Sci.* 2(1):61-68.

[27] 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.

[28] Zhang L, Chen J (2006) Bayesian solution to pricing and inventory control under unknown demand distribution. *Oper. Res. Lett.* 34(5):517-524.

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

# Appendix

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

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

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

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

**Proof.** First we make the observation that

$$p^* = \overline{p}\big(\breve{\alpha}(p^*), \breve{\beta}(p^*)\big). \tag{42}$$

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

$$G(p, \lambda(p)) = pe^{\lambda(p)}\mathbb{E}[e^{\epsilon}] - \min_{y \in \mathcal{Y}} \Big\{ h\mathbb{E}\big[y - e^{\lambda(p)}e^{\epsilon}\big]^+ + b\mathbb{E}\big[e^{\lambda(p)}e^{\epsilon} - y\big]^+ \Big\}. \tag{43}$$

Then Assumption 1(i) implies that $G(p, \lambda(p))$ is unimodal in $p$. Assuming that $G$ has a unique maximizer and that $\overline{p}(\breve{\alpha}(z), \breve{\beta}(z))$ is the unique optimal solution for problem B1 with parameters $\big(\breve{\alpha}(z), \breve{\beta}(z)\big)$, then (42) follows from Lemma A1 of Besbes and Zeevi (2015) by letting their function $G$ be (43).

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

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

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

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

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

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

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 Lemma A1. $\square$

To compare the optimal solutions of Problems B1 and B2, we need several technical Lemmas. To that end, we change the decision variables in B1 and B2. For given parameters $\alpha$ 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

$$\overline{W}(d,y) = h\mathbb{E}\big(y - de^{\epsilon}\big)^{+} + b\mathbb{E}\big(de^{\epsilon} - y\big)^{+} \tag{44}$$

and

$$\overline{G}(\alpha,\beta,d) = d\frac{\alpha - \log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - \min_{y\in\mathcal{Y}}\overline{W}(d,y) = d\frac{\alpha - \log d}{\beta}\mathbb{E}\big[e^{\epsilon}\big] - \overline{W}(d,\overline{y}(d)), \tag{45}$$

where $\overline{y}(d)$ is the optimal solution of (44) in $\mathcal{Y}$ for given $d$. Let $F(\cdot)$ be the cumulative distribution function (CDF) of $e^{\epsilon}$, then it can be verified that

$$\overline{y}(d) = dF^{-1}\left(\frac{b}{b+h}\right), \tag{46}$$

where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Also, we let $\overline{d}\,(\alpha,\beta)$ denote the optimal solution of maximizing (45) in $\mathcal{D}$.

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

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

Let

$$\tilde{W}_{i+1}(d,y) = \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(y - de^{\epsilon_t}\big)^{+} + b\big(de^{\epsilon_t} - y\big)^{+}\Big), \tag{47}$$

and

$$\begin{aligned} \tilde{G}_{i+1}(\alpha,\beta,d) &= d\frac{\alpha - \log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right) - \min_{y\in\mathcal{Y}}\tilde{W}_{i+1}(d,y) \\ &= d\frac{\alpha - \log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right) - \tilde{W}_{i+1}(d,\tilde{y}(d)), \end{aligned} \tag{48}$$

where $\tilde{y}_{i+1}\,(d)$ denotes the optimal solution of $\tilde{W}_{i+1}(d,y)$ in (47) on $\mathcal{Y}$. Let $\tilde{d}_{i+1}\,(\alpha,\beta)$ be the optimal solution for $\tilde{G}_{i+1}(\cdot,\cdot,d)$ in (48) on $\mathcal{D}$. Also, let $\tilde{y}^{u}_{i+1}\,(d)$ denote the optimal order-up-to level for problem B2 on $\mathbb{R}_+$ for given $p \in \mathcal{P}$ (here the superscript "$u$" stands for "unconstrained"). Then

$$\tilde{y}^u_{i+1}(d) = \min\left\{ de^{\epsilon_j} : \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \le e^{\epsilon_j}\right\} \ge \frac{b}{b+h} \right\}, \tag{49}$$

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

$$\tilde{y}_{i+1}(d) = \min\left\{ \max\left\{ \tilde{y}^u_{i+1}(d), y^l \right\}, y^h \right\}. \tag{50}$$

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

We first compare the profit functions defined for the two problems (44), (45), and (47), (48). To this end, we need the following properties.

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

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

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

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

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

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

**Proof.** By triangle inequality, we have

$$\begin{aligned} &\max_{d\in\mathcal{D}} \left| \overline{W}(d, \overline{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \right| \\ \le\ &\max_{d\in\mathcal{D}} \left| \overline{W}\big(d, \overline{y}(d)\big) - \tilde{W}_{i+1}\big(d, \overline{y}(d)\big) \right| + \max_{d\in\mathcal{D}} \left| \tilde{W}_{i+1}\big(d, \overline{y}(d)\big) - \tilde{W}_{i+1}\big(d, \tilde{y}_{i+1}(d)\big) \right|. \end{aligned} \tag{51}$$

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

For any $d \in \mathcal{D}$ and $y \in \mathcal{Y}$, we define $z = y/d$. Then, from (46), the optimal $z$ to minimize $\overline{W}(d, dz)$ is

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

Moreover, we have

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

and

$$\tilde{W}_{i+1}(d,\overline{y}(d)) = \tilde{W}_{i+1}(d,d\overline{z}) = d\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Big(h\big(\overline{z}-e^{\epsilon_t}\big)^+ + b\big(e^{\epsilon_t}-\overline{z}\big)^+\Big)\right). \tag{52}$$

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

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

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

$$\mathbb{P}\left\{d^h\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right| > d^h\xi\right\} = \mathbb{P}\left\{\left|\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\Delta_t\right| > \xi\right\} \le 2e^{-4I_i\xi^2}, \tag{53}$$

which deduces to

$$\mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|\overline{W}(d,\overline{y}(d)) - \tilde{W}_{i+1}(d,\overline{y}(d))\right| > d^h\xi\right\} \le 2e^{-4I_i\xi^2}. \tag{54}$$

This bounds the first term on the right hand side of (51).

To bound the second term in (51), we use

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

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

$$\left|\hat{F}(\overline{z}) - \frac{b}{b+h}\right| \le \theta. \tag{55}$$

It can be verified that

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

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

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

Combining the two results above we obtain

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

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

$$\mathbb{P}(\mathcal{A}_3(\theta)) \geq 1 - 2e^{-2I_i\theta^2}. \tag{56}$$

For $d \in \mathcal{D}$, let $\tilde{z}_{i+1}(d) = \frac{\tilde{y}_{i+1}(d)}{d}$ and $\tilde{z}^u_{i+1} = \frac{\tilde{y}^u_{i+1}(d)}{d}$, then it follows from (49) that

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

And it follows from (50) that

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

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

First, for any given $d \in \mathcal{D}$, if $\overline{z} \leq \tilde{z}^u_{i+1}$, then it follows from (52) that

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\overline{z}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(e^{\epsilon_t} - \overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\big\} \\
&\qquad + b\big(e^{\epsilon_t} - \overline{z}\big)\mathbb{1}\big\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z} - e^{\epsilon_t}\big)\mathbb{1}\big\{e^{\epsilon_t} \leq \overline{z}\big\}\Big] \\
&\leq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b\big(e^{\epsilon_t} - \overline{z}\big)\mathbb{1}\big\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\big\} \\
&\qquad + b\big(\tilde{z}^u_{i+1} - \overline{z}\big)\mathbb{1}\big\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\big\} + h\big(\overline{z} - e^{\epsilon_t}\big)\mathbb{1}\big\{e^{\epsilon_t} \leq \overline{z}\big\}\Big],
\end{aligned} \tag{57}$$

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

$$\begin{aligned}
\tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\} \\
&\qquad + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{\overline{z} < e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \overline{z}\}\Big] \\
&\geq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1}))\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \overline{z}\}\Big],
\end{aligned} \tag{58}$$

with the inequality obtained by dropping the nonnegative middle term. Consequently when $\overline{z} \leq \tilde{z}^u_{i+1}$

we subtract (58) from (57) to obtain

$$
\begin{aligned}
& \tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \\
\leq\ & d\left(b(\tilde{z}^u_{i+1} - \overline{z})(1 - \hat{F}(\tilde{z}^u_{i+1})) + b(\tilde{z}^u_{i+1} - \overline{z})(\hat{F}(\tilde{z}^u_{i+1}) - \hat{F}(\overline{z})) + h(\overline{z} - \tilde{z}^u_{i+1})\hat{F}(\overline{z})\right) \\
=\ & d(\tilde{z}^u_{i+1} - \overline{z})\big(-(h+b)\hat{F}(\overline{z}) + b\big) \\
\leq\ & d\big(\tilde{z}^u_{i+1} - \overline{z}\big)(b+h)\theta, \qquad (59)
\end{aligned}
$$

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

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

$$
\begin{aligned}
\tilde{W}_{i+1}(d, d\overline{z}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \overline{z})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} \\
&\qquad + h(\overline{z} - e^{\epsilon_t})\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t} \leq \overline{z}\} + h(\overline{z} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\}\Big] \\
&\leq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \overline{z})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} \\
&\qquad + h(\overline{z} - \tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t} \leq \overline{z}\} + h(\overline{z} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\}\Big], \qquad (60)
\end{aligned}
$$

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

$$
\begin{aligned}
\tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) &= \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} \\
&\qquad + b(e^{\epsilon_t} - \tilde{z}^u_{i+1})\mathbb{1}\{\tilde{z}^u_{i+1} < e^{\epsilon_t} \leq \overline{z}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\}\Big] \\
&\geq \frac{d}{2I_i}\sum_{t=1}^{2I_i}\Big[b(e^{\epsilon_t} - \tilde{z}^u_{i+1})\mathbb{1}\{\overline{z} < e^{\epsilon_t}\} + h(\tilde{z}^u_{i+1} - e^{\epsilon_t})\mathbb{1}\{e^{\epsilon_t} \leq \tilde{z}^u_{i+1}\}\Big], \qquad (61)
\end{aligned}
$$

again the inequality follows from dropping the nonnegative second term. Subtracting (61) from (60), we obtain

$$
\begin{aligned}
& \tilde{W}_{i+1}(d, d\overline{z}) - \tilde{W}_{i+1}(d, d\tilde{z}^u_{i+1}) \\
\leq\ & d\left(b(\tilde{z}^u_{i+1} - \overline{z})(1 - \hat{F}(\overline{z})) + h(\overline{z} - \tilde{z}^u_{i+1})(\hat{F}(\overline{z}) - \hat{F}(\tilde{z}^u_{i+1})) + h(\overline{z} - \tilde{z}^u_{i+1})\hat{F}(\tilde{z}^u_{i+1})\right) \\
=\ & d(\overline{z} - \tilde{z}^u_{i+1})((h+b)\hat{F}(\overline{z}) - b) \\
\leq\ & d(\overline{z} - \tilde{z}^u_{i+1})(b+h)\theta, \qquad (62)
\end{aligned}
$$

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

The results (59) and (62) imply that, when $\hat{F}(\cdot)$ is a $\theta$-estimate, or (55) is satisfied, it holds that

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

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

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

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

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

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

$$\max_{d \in \mathcal{D}} \left\{ \tilde{W}_{i+1}(d, d\bar{z}) - \tilde{W}_{i+1}(d, d\tilde{z}_{i+1}(d)) \right\} \leq K_{41}(b+h)\theta. \tag{63}$$

Letting $\theta = \xi$ in (63) we obtain

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

where the last inequality follows from (56). This proves, by noting $\tilde{W}_{i+1}(d, \bar{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \geq 0$ as $\tilde{y}_{i+1}(d)$ is the minimizer of $\tilde{W}_{i+1}$ on $\mathcal{Y}$, that

$$\mathbb{P}\left\{ \max_{d \in \mathcal{D}} \left| \left( \tilde{W}_{i+1}(d, \bar{y}(d)) - \tilde{W}_{i+1}(d, \tilde{y}_{i+1}(d)) \right) \right| \leq K_{41}(b+h)\xi \right\} \geq 1 - 2e^{-2I_i\xi^2}. \tag{64}$$

Applying (54) and (64) in (51), we conclude that there exist a constant $K_{40} > 0$ such that for any $\xi > 0$, when $I_i$ is sufficiently large,

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

This completes the proof of Lemma A3. □

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

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

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

**Proof.** For any $d \in \mathcal{D}$, similar argument as that used in proving (53) of Lemma A2 shows that, for any $\xi > 0$,

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

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

$$\begin{aligned}
& \mathbb{P}\left\{\max_{d\in\mathcal{D}}\left|d\frac{\alpha-\log d}{\beta}\mathbb{E}[e^{\epsilon_t}] - d\frac{\alpha-\log d}{\beta}\left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*\xi\right\} \\
= \; & \mathbb{P}\left\{r^*\left|\mathbb{E}[e^{\epsilon_t}] - \left(\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} e^{\epsilon_t}\right)\right| \le r^*\xi\right\} \\
\ge \; & 1 - e^{-4I_i\xi^2}. \qquad (65)
\end{aligned}$$

Hence, it follows from (45) and (48) that, for any $d \in \mathcal{D}$ and $\xi > 0$,

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

where the last inequality follows from (65) and Lemma A2. Letting $K_{42} = K_{40} + 2r^*\sigma$ completes the proof of Lemma A4. □

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

$$\mathcal{A}_4(\xi) = \left\{ \omega : \max_{d \in \mathcal{D}} \left| \overline{G}(\alpha, \beta, d) - \tilde{G}_{i+1}(\alpha, \beta, d) \right| \leq K_{42}\xi \right\}. \tag{66}$$

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

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

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

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

*for some positive constant $K_{43}$.*

**Proof.** To slightly simplify the notation, for given parameters $\alpha$ and $\beta$, in this proof we let

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

By Taylor's expansion,

$$\overline{G}(\tilde{d}) = \overline{G}(\overline{d}) + \overline{G}'(\overline{d})(\tilde{d} - \overline{d}) + \frac{\overline{G}''(q)}{2}(\tilde{d} - \overline{d})^2, \tag{67}$$

where $q \in [\overline{d}, \tilde{d}]$ if $\overline{d} \leq \tilde{d}$ and $q \in [\tilde{d}, \overline{d}]$ if $\overline{d} > \tilde{d}$. Since we assume the minimizer of $\overline{W}(d, y)$ over $\mathbb{R}_+$ falls into $\mathcal{Y}$, it follows from (45) that $\overline{G}(d) = d^{\frac{\alpha - \log d}{\beta}}\mathbb{E}[e^\epsilon] - Ad$, where $A = \min_z\big\{h\mathbb{E}\big(z - e^\epsilon\big)^+ + b\mathbb{E}\big(e^\epsilon - z\big)^+\big\} > 0$ is a constant. Thus, we have

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

Since $\lambda(\cdot)$ is assumed to be strictly decreasing, it follows that $\breve{\beta}(\cdot)$ is bounded below by a positive number, say $\bar{a} > 0$. On $\beta \geq \bar{a}$, let $\min_{d \in \mathcal{D}} \frac{\mathbb{E}[e^\epsilon]}{\beta d} = m$ and it holds that $m > 0$, then it follows from (67) that

$$\overline{G}(\tilde{d}) \leq \overline{G}(\overline{d}) - \frac{m}{2}(\tilde{d} - \overline{d})^2. \tag{68}$$

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

$$\overline{G}(\tilde{d}) - \overline{G}(\overline{d}) \geq -2K_{42}\xi. \tag{69}$$

We prove this by contradiction. Suppose it is not true, i.e., $\overline{G}(\overline{d}) - \overline{G}(\tilde{d}) > 2K_{42}\xi$, then it follows from (66) that

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

This leads to $\tilde{G}(\overline{d}) > \tilde{G}(\tilde{d})$, contradicting with $\tilde{d}$ being optimal for problem B2. Thus, (69) is satisfied on $\mathcal{A}_4(\xi)$.

Using (68) and (69), we obtain that, on event $\mathcal{A}_4(\xi)$,

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

or equivalently, for some constant $K_{44}$,

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

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

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

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

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

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

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

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

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

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

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

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

$$\hat{\alpha}_{i+1} = \frac{\lambda(\hat{p}_i) + \lambda(\hat{p}_i + \delta_i)}{2} + \frac{B^1_{i+1} + B^2_{i+1}}{2} + \hat{\beta}_{i+1}\frac{2\hat{p}_i + \delta_i}{2}, \tag{70}$$

$$\hat{\beta}_{i+1} = -\frac{\lambda(\hat{p}_i + \delta_i) - \lambda(\hat{p}_i)}{\delta_i} - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}). \tag{71}$$

Applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the second order for (71), we obtain

$$\begin{aligned}\hat{\beta}_{i+1} &= -\left(\lambda'(\hat{p}_i) + \frac{1}{2}\lambda''(q_i)\delta_i\right) - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}) \\ &= \breve{\beta}(\hat{p}_i) - \frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1}),\end{aligned} \tag{72}$$

where $q_i \in [\hat{p}_i, \hat{p}_i + \delta_i]$. Substituting $\hat{\beta}_{i+1}$ in (70) by (72), and applying Taylor's expansion on $\lambda(\hat{p}_i + \delta_i)$ at point $\hat{p}_i$ to the first order, we have

$$\begin{aligned}\hat{\alpha}_{i+1} &= \lambda(\hat{p}_i) + \frac{1}{2}\lambda'(q'_i)\delta_i + \frac{B^1_{i+1} + B^2_{i+1}}{2} - \lambda'(\hat{p}_i)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\ &\qquad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1})\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right) \\ &= \breve{\alpha}(\hat{p}_i) + \frac{1}{2}\lambda'(q'_i)\delta_i + \frac{B^1_{i+1} + B^2_{i+1}}{2} - \frac{1}{2}\lambda'(\hat{p}_i)\delta_i \\ &\qquad + \left(-\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}(-B^1_{i+1} + B^2_{i+1})\right)\left(\hat{p}_i + \frac{\delta_i}{2}\right),\end{aligned} \tag{73}$$

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

Since the error terms $\epsilon_t$ are assumed to be bounded, we apply Hoeffding inequality to obtain

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

Hence,

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

Therefore,

$$\mathbb{P}\left\{\left|-B^1_{i+1} + B^2_{i+1}\right| \le 2\xi\right\} \ge \mathbb{P}\left\{\left|-B^1_{i+1}\right| + \left|B^2_{i+1}\right| \le 2\xi\right\} \ge 1 - 4e^{-2I_i\xi^2}.$$

Similar argument shows

$$\mathbb{P}\left\{\left|B^1_{i+1} + B^2_{i+1}\right| \le 2\xi\right\} \ge 1 - 4e^{-2I_i\xi^2}.$$

Since $\lambda'(\cdot)$ and $\lambda''(\cdot)$ are bounded and $\delta_i$ converges to 0, from (73) we conclude that there must exist a constant $K_{47}$ such that, on the event $\left|B^1_{i+1} + B^2_{i+1}\right| \le 2\xi$ and $\left|-B^1_{i+1} + B^2_{i+1}\right| \le 2\xi$, it holds that

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

Therefore,

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i)| \le K_{47}\left(\delta_i+\frac{\xi}{\delta_i}+\xi\right)\right\} \ge \mathbb{P}\left\{\left|B^1_{i+1}+B^2_{i+1}\right| \le 2\xi, \left|-B^1_{i+1}\right|+\left|B^2_{i+1}\right| \le 2\xi\right\}$$
$$\ge 1-8e^{-2I_i\xi^2},$$

which implies

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\breve{\alpha}(\hat{p}_i)|^2 \le K_{48}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\right\} \ge 1-8e^{-2I_i\xi^2}. \tag{74}$$

From (72) we have

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

which implies

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2 \le K_{50}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}\right)\right\} \ge 1-4e^{-2I_i\xi^2}. \tag{75}$$

Following the development of (74) and (75), we have

$$\mathbb{P}\left\{|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i+\delta_i)|^2 \le K_{51}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\right\} \ge 1-8e^{-2I_i\xi^2}. \tag{76}$$

and

$$\mathbb{P}\left\{\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i+\delta_i)\right|^2 \le K_{52}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}\right)\right\} \ge 1-4e^{-2I_i\xi^2}. \tag{77}$$

Combining(74), (75), (76), and (77), we obtain

$$\mathbb{P}\Big\{|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i)|^2+\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i)\right|^2+|\hat{\alpha}_{i+1}-\lambda(\hat{p}_i+\delta_i)|^2+\left|\hat{\beta}_{i+1}-\breve{\beta}(\hat{p}_i+\delta_i)\right|^2 \tag{78}$$
$$\le K_{53}\left(\delta_i^2+\frac{\xi^2}{\delta_i^2}+\xi^2\right)\Big\}$$
$$\ge 1-24e^{-2I_i\xi^2},$$

which is

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

Therefore,

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

Hence one has

$$\begin{aligned}
&\mathbb{E}\left[|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}|^2+|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}|^2+|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}|^2\right] \\
\le &\left(\frac{12}{I_i}+\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right)^{-1}K_{53}\delta_i^2\right)\left(\frac{K_{54}}{\delta_i^2}+K_{55}\right) \\
\le &K_{46}I_i^{-\frac{1}{2}}.
\end{aligned} \tag{79}$$

This completes the proof of Lemma A6. □

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

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

$$\begin{aligned}
&\mathbb{P}\Big\{\left|\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{p}_{i+1}\right| \ge K_{56}\big(\left|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}\right|+\left|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}\right| \\
&\qquad +\left|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}\right|+\left|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}\right|\big)\Big\} \le \frac{8}{I_i}, \\
&\mathbb{P}\Big\{\left|\tilde{y}_{i+1}\big(\breve{\alpha}(\hat{p}_i),\breve{\beta}(\hat{p}_i)\big)-\hat{y}_{i+1}\right| \ge K_{56}\big(\left|\breve{\alpha}(\hat{p}_i)-\hat{\alpha}_{i+1}\right|+\left|\breve{\beta}(\hat{p}_i)-\hat{\beta}_{i+1}\right| \\
&\qquad +\left|\breve{\alpha}(\hat{p}_i+\delta_i)-\hat{\alpha}_{i+1}\right|+\left|\breve{\beta}(\hat{p}_i+\delta_i)-\hat{\beta}_{i+1}\right|\big)\Big\} \le \frac{8}{I_i}.
\end{aligned}$$

**Proof.** To compare the solutions of these two problems, we introduce a general function based on the data-driven problem DD and problem B2: Given selling price $p_t = \hat{p}_i$ for $t = t_i + 1, \ldots, t_i + I_i$

and $p_t = \hat{p}_i + \delta_i$ for $t = t_i + I_i + 1, \ldots, t_i + 2I_i$, logarithm demand data $D_t, t = t_i + 1, \ldots, t_i + 2I_i$, and two sets of parameters $(\alpha_1, \beta_1), (\alpha_2, \beta_2)$, define $\zeta_{t=t_i+1}^{t_1+I_i}(\alpha_1, \beta_1) = (\zeta_{t_i+1}, \ldots, \zeta_{t_i+I_i})$ and $\zeta_{t=t_i+I_i+1}^{t_i+2I_i}(\alpha_2, \beta_2) = (\zeta_{t_i+I_i+1}, \ldots, \zeta_{t_i+2I_i})$ by

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

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

$$
\begin{aligned}
& H_{i+1}\Big(p, e^{\alpha_1 - \beta_1 p}, \zeta_{t=t_i+1}^{t_1+I_i}(\alpha_1, \beta_1), \zeta_{t=t_i+I_i+1}^{t_i+2I_i}(\alpha_2, \beta_2)\Big) \qquad (80) \\
= \; & p e^{\alpha_1 - \beta_1 p} \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} e^{\zeta_t} - \min_{y \in \mathcal{Y}} \bigg\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big( h\big(y - e^{\alpha_1 - \beta_1 p} e^{\zeta_t}\big)^+ + b\big(e^{\alpha_1 - \beta_1 p} e^{\zeta_t} - y\big)^+ \Big) \bigg\}.
\end{aligned}
$$

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

$$
p\big((\alpha_1, \beta_1), (\alpha_2, \beta_2)\big) = \arg\max_{p \in \mathcal{P}} H_{i+1} \Big(p, e^{\alpha_1 - \beta_1 p}, \zeta_{t=t_i+1}^{t_1+I_i}(\alpha_1, \beta_1), \zeta_{t=t_i+I_i+1}^{t_i+2I_i}(\alpha_2, \beta_2)\Big) \qquad (81)
$$

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

$$
y\big(e^{\alpha_1 - \beta_1 p}, (\alpha_1, \beta_1), (\alpha_2, \beta_2)\big) = \arg\min_{y \in \mathcal{Y}} \bigg\{ \frac{1}{2I_i} \sum_{t=t_i+1}^{t_i+2I_i} \Big( h\big(y - e^{\alpha_1 - \beta_1 p} e^{\zeta_t}\big)^+ + b\big(e^{\alpha_1 - \beta_1 p} e^{\zeta_t} - y\big)^+ \Big) \bigg\}. \qquad (82)
$$

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

$$
\begin{aligned}
& \Big| p\big((\alpha_1, \beta_1)(\alpha_2, \beta_2)\big) - p\big((\alpha_1', \beta_1')(\alpha_2', \beta_2')\big) \Big| \qquad (83) \\
& \qquad \leq K_{57} \Big( |\alpha_1 - \alpha_1'| + |\beta_1 - \beta_1'| + |\alpha_2 - \alpha_2'| + |\beta_2 - \beta_2'| \Big), \\
& \Big| y\big(e^{\alpha_1 - \beta_1 p}, (\alpha_1, \beta_1), (\alpha_2, \beta_2)\big) - y\big(e^{\alpha_1' - \beta_1' p}, (\alpha_1', \beta_1'), (\alpha_2', \beta_2')\big) \Big| \qquad (84) \\
& \qquad \leq K_{57} \Big( |\alpha_1 - \alpha_1'| + |\beta_1 - \beta_1'| + |\alpha_2 - \alpha_2'| + |\beta_2 - \beta_2'| \Big).
\end{aligned}
$$

The optimization problem (80) will serve as yet another bridging problem between DD and B2. To see that, observe that when $\alpha_1 = \alpha_2 = \hat{\alpha}_{i+1}$ and $\beta_1 = \beta_2 = \hat{\beta}_{i+1}$, problem (81) is reduced to the data-driven problem DD. That is,

$$
\hat{p}_{i+1} = p\big((\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}), (\hat{\alpha}_{i+1}, \hat{\beta}_{i+1})\big). \qquad (85)
$$

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

$$\zeta_t = D_t - (\alpha_1 - \beta_1 p_t) = \lambda(\hat{p}_i) + \epsilon_t - (\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_i) = \epsilon_t, \tag{86}$$

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

$$\zeta_t = D_t - (\alpha_2 - \beta_2 p_t) = \lambda(\hat{p}_i + \delta_i) + \epsilon_t - (\breve{\alpha}(\hat{p}_i + \delta_i) - \breve{\beta}(\hat{p}_i + \delta_i)(\hat{p}_i + \delta_i)) = \epsilon_t. \tag{87}$$

This shows that when the parameters are $\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big)$ and $\big(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i)\big)$, problem (81) is reduced to bridging problem B2. This gives us

$$\tilde{p}_{i+1}\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big) = p\Big(\big(\breve{\alpha}(\hat{p}_i), \breve{\beta}(\hat{p}_i)\big), \big(\breve{\alpha}(\hat{p}_i + \delta_i), \breve{\beta}(\hat{p}_i + \delta_i)\big)\Big). \tag{88}$$

The two results (85) and (88) will enable us to compare the optimal solutions of the data-driven optimization problem DD and bridging problem B2 through one optimization problem (81).

In Lemma A6, letting $\xi = (2I_i)^{-\frac{1}{2}}(\log 2I_i)^{\frac{1}{2}}$ in (78), we obtain

$$\begin{aligned}\mathbb{P}\Big\{ &|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}|^2 + |\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}|^2 + |\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}|^2 \\ &\leq K_{53}\left(I_i^{-\frac{1}{2}} + (2I_i)^{-\frac{1}{2}}(\log 2I_i) + (2I_i)^{-1}(\log 2I_i)\right)\Big\} \\ \geq \;& 1 - \frac{8}{I_i}.\end{aligned} \tag{89}$$

This implies

$$\begin{aligned}\mathbb{P}\Big\{ &|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}, \\ &|\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\Big\} \\ \geq \;& 1 - \frac{8}{I_i}.\end{aligned} \tag{90}$$

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

$$\begin{aligned}\mathcal{A}_5 = \Big\{\omega : &|\breve{\alpha}(\hat{p}_i) - \hat{\alpha}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i) - \hat{\beta}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}, \\ &|\breve{\alpha}(\hat{p}_i + \delta_i) - \hat{\alpha}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}},\ |\breve{\beta}(\hat{p}_i + \delta_i) - \hat{\beta}_{i+1}| \leq (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}\Big\}.\end{aligned} \tag{91}$$

Then by (91) one has

$$\mathbb{P}(\mathcal{A}_5^c) \leq \frac{8}{I_i}. \tag{92}$$

When $\beta_1 > 0$, similar to Remark 2 and Lemma A2, one can verify that $H_{i+1}(\cdot,\cdot,\cdot,\cdot)$ of (80) is unimodal in $p$ thus its optimal solution is well-defined. Define

$$i^* = \max\Big\{\log_v \frac{e}{2I_0},\ \min\Big\{i \;\Big|\; (3K_{53})^{\frac{1}{2}}(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}} < \min_{p\in\mathcal{P}} \breve{\beta}(p)\Big\}\Big\}, \tag{93}$$

where we need $i^*$ to be no less than $\log_v \frac{e}{2I_0}$ to ensure that $(2I_i)^{-\frac{1}{4}}(\log 2I_i)^{\frac{1}{2}}$ is decreasing on $i \geq i^*$. When $i \geq i^*$, it follows that $\hat{\beta}_{i+1} > 0$ on $\mathcal{A}_5$, hence on event $\mathcal{A}_5$, problem DD is unimodal in $p$ after minimizing over $y$, and the optimal pricing is well-defined. These properties will enable us to prove that the convergence of parameters translates to convergence of the optimal solutions. Then the first part in Lemma A7 on $p$ follows directly from (85), (88) and (83). From equations (82), (86), and (87), we conclude

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

and it follows from the DDA policy that

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

Then, similar analysis as that in the proof of (83) can be used to prove (84). $\square$

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

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

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

**Proof.** For $p \in \mathcal{P}$, the optimal solution for bridging problem B1 is the same as (46), $\overline{y}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$. Thus

$$\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}F^{-1}\left(\frac{b}{b+h}\right). \tag{94}$$

For given $p \in \mathcal{P}$, we follow (49) to define $\tilde{y}^u_{i+1}\big(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\big)$ as the unconstrained optimal order-up-to level for problem B2 on $\mathbb{R}_+$, then it can be verified that

$$\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) = e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p} \min\left\{e^{\epsilon_j} : \frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i} \mathbb{1}\left\{e^{\epsilon_t} \leq e^{\epsilon_j}\right\} \geq \frac{b}{b+h}\right\}, \tag{95}$$

and, similar to (50), we have

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

It is seen that

$$\left|\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) - \tilde{y}_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)\right| \leq \left|\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right) - \tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)p}\right)\right|. \tag{96}$$

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

$$\begin{aligned} & \mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-\frac{b}{b+h}\leq -z\right\} \qquad (97) \\ = & \mathbb{P}\left\{\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\} \\ \leq & \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}\geq\frac{b}{b+h}\right\} \\ = & \mathbb{P}\left\{\frac{1}{2I_i}\sum_{t=t_i+1}^{t_i+2I_i}\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}-\left(\frac{b}{b+h}-z\right)\geq z\right\}, \end{aligned}$$

where the first inequality follows from (95). Since $\mathbb{E}\left[\mathbb{1}\left\{e^{\epsilon_t}\leq F^{-1}\left(\frac{b}{b+h}-z\right)\right\}\right]=\frac{b}{b+h}-z$, we apply Hoeffding inequality to obtain

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

Combining this with (94) and (97), we obtain

$$\begin{aligned} & \mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)\leq -z\right\} \\ & \leq e^{-4I_iz^2}. \end{aligned} \qquad (98)$$

Similarly, we have

$$\begin{aligned} & \mathbb{P}\left\{F\left(\tilde{y}^u_{i+1}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)-F\left(\overline{y}\left(e^{\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}\right)e^{-(\breve{\alpha}(\hat{p}_i)-\breve{\beta}(\hat{p}_i)\hat{p}_{i+1})}\right)\geq z\right\} \\ & \leq e^{-4I_iz^2}. \end{aligned} \qquad (99)$$

From regularity condition (v), the probability density function $f(\cdot)$ of $e^{\epsilon_t}$ 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 (98) and (99), 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 $K_{58} = \max_{\hat{p}_i \in \mathcal{P}, \hat{p}_{i+1} \in \mathcal{P}} \frac{1}{r} e^{\breve{\alpha}(\hat{p}_i) - \breve{\beta}(\hat{p}_i)\hat{p}_{i+1}}$, then $K_{58} > 0$. We have

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

and Lemma A8 follows from the inequality above and (96). □