Submitted to *Operations Research*
manuscript OPRE-2011-11-587.R3

Dynamic Pricing and Learning with Finite Inventories

Arnoud V. den Boer
University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands,
a.v.denboer@utwente.nl

Bert Zwart
Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands,
VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands,
bert.zwart@cwi.nl

We study a dynamic pricing problem with finite inventory and parametric uncertainty on the demand distribution. Products are sold during selling seasons of finite length, and inventory that is unsold at the end of a selling season perishes. The goal of the seller is to determine a pricing strategy that maximizes the expected revenue. Inference on the unknown parameters is made by maximum likelihood estimation.

We show that this problem satisfies an endogenous-learning property, which means that the unknown parameters are learned on-the-fly if the chosen selling prices are sufficiently close to the optimal ones. We show that a small modification to the certainty equivalent pricing strategy - which always chooses the optimal price w.r.t. current parameter estimates - satisfies $\text{Regret}(T) = O(\log^2(T))$, where $\text{Regret}(T)$ measures the expected cumulative revenue loss w.r.t. a clairvoyant who knows the demand distribution. We complement this upper bound by showing an instance for which the regret of any pricing policy satisfies $\Omega(\log T)$.

Key words: dynamic programming/optimal control: Markov; marketing: estimation/statistical techniques, pricing;

1. Introduction

1.1. Introduction and Motivation

The emergence of the Internet as a sales channel has made it very easy for companies to experiment with selling prices. Where in the past costs and efforts were needed to change prices, for example by issuing a new catalogue or replacing price tags, and consequently prices were fixed for longer periods of time, nowadays a webshop can adapt its prices with a proverbial flick of the switch, without any additional costs or efforts. This flexibility in pricing is one of the main drivers for research on *dynamic pricing*: the study of determining optimal selling prices under changing circumstances.

A much-studied situation is a firm that sells limited amounts of products during finite selling seasons, after which all unsold products perish. Examples of products with this property are flight tickets, hotel rooms, car rental reservations, and concert tickets (Talluri and van Ryzin 2004). An important insight from the literature on dynamic pricing is that the optimal selling price of such products depends on the remaining inventory and the length of the remaining selling season, see e.g. Gallego and van Ryzin (1994). The optimal decision is thus not to use a single price but a collection of prices: one for each combination of remaining inventory and length of remaining selling season. To determine these optimal prices it is essential to know the relation between the demand and the selling price. In most literature from the nineties on dynamic pricing it is assumed that this relation is known to the seller, but in practice exact information on consumer behavior is generally not available. It is therefore not surprising that the review on dynamic pricing by Bitran and Caldentey (2003) mentions dynamic pricing with demand learning as an important future research direction. The presence of digital sales data enables a data-driven approach of dynamic pricing, where the selling firm not only determines optimal prices, but also learns how changing prices affects the demand. Ideally, this learning will eventually lead to optimal pricing decisions.

In this study we consider a pricing-and-learning problem motivated from the hotel industry (Talluri and van Ryzin 2004, section 10.2, Weatherford and Kimes 2003). In that context, a "product" corresponds to a combination of arrival-date and length-of-stay (possibly augmented by other features or requirements). These products are perishable (unsold opportunities cannot be held in stock), are sold during a finite time period, and the available capacity is finite. An important feature of this context is that a firm typically sells many different products with similar demand characteristics at the same time. This means that learning the demand characteristics of each product separately may not be very efficient; instead, the firm would want to learn about consumer behavior from all the sales data corresponding to products with similar demand characteristics.

This motivates the current study of dynamic pricing and learning for perishable products with finite initial inventory, during multiple finite selling seasons.

1.2. Contributions

We consider a parametric demand model which includes linear, exponential, and logit demand; these demand functions are frequently encountered in theory and practice (Talluri and van Ryzin 2004). The uncertainty in the demand is modeled by unknown parameters that can be estimated from historical sales data using maximum likelihood estimation. We propose a pricing strategy that is structurally very intuitive, and easy to understand by price managers: at each moment where prices can be changed the price manager calculates a statistical estimate of the unknown parameter; subsequently she determines the optimal price, assuming that the parameter estimate

is correct, and she uses this price until the next decision moment. In other words, at each decision moment the price manager acts as if she is certain about the parameter estimates. Only in the last period of a selling season for which inventory is still positive, a small deviation on this price may be prescribed by our pricing strategy.

This type of strategy for sequential decision problems under uncertainty is known under different names in the literature: certainty equivalent control, myopic control, passive learning, and the principle of estimation and control. There are problems for which certainty equivalent control is not a good strategy, e.g. the multi-period control problem (Anderson and Taylor 1976, Lai and Robbins 1982), and dynamic pricing with infinite inventory (Broder and Rusmevichientong 2012, Keskin and Zeevi 2014, den Boer and Zwart 2014b). In these two examples, passive learning is not sufficient to learn the parameters: the decision maker should actively account for the fact that she is not only optimizing prices, but also tries to "optimize" the learning process. This implies that sometimes decisions should be taken that seem suboptimal in the short term. In the dynamic pricing problem with infinite inventory, this can be accomplished by the controlled variance policy of den Boer and Zwart (2014b) or the MLE-cycle policy of Broder and Rusmevichientong (2012). The infinite-inventory setting is also closely related to several problems from the online convex-optimization, multi-armed bandit and stochastic approximation literature; see den Boer and Zwart (2014b) for references and a brief discussion on similarities and differences with dynamic pricing.

In the situation that we study in this article, dynamic pricing with finite inventory and finite selling seasons, certainty equivalent control does perform well: the parameter estimates converge with probability one to the correct values, and the prices converge to the optimal prices. The $\text{Regret}(T)$, which measures the expected amount of revenue loss in the first T selling seasons due to not using the optimal prices, is $O(\log^2(T))$. This growth rate is considerably smaller than $\sqrt{T}$, which is the best achievable growth rate of the regret for the problem with infinite inventory (in different settings, this is shown by Kleinberg and Leighton (2003), Besbes and Zeevi (2011), Broder and Rusmevichientong (2012) and Keskin and Zeevi (2013)), and moreover, this bound can hardly be improved. We show an instance for which *any* pricing strategy has $\text{Regret}(T) \geq K_0 \log(T)$, for some $K_0 > 0$ independent of T. This means that the upper bound $\log^2(T)$ on the regret is close to the best achievable growth rate. In Remark 4 we discuss the small gap between the lower and upper bound.

Thus, the regret, which can be interpreted as the "cost for learning", is structurally different in these two models: in our finite-inventory setting $\text{Regret}(T) = O(\log^2(T))$ is attainable, whereas in the infinite-inventory setting $\text{Regret}(T) = \Omega(\sqrt{T})$ for any policy.

This difference in qualitative behavior seems to be related to the presence of "uninformative prices" (Broder and Rusmevichientong 2012) or "indeterminate equilibria" (Harrison et al. 2012):

there are values of the parameter estimates such that the expected demand observed at the corresponding optimal price (optimal w.r.t. these estimates) is precisely equal to what these parameter estimates would predict; in other words, at these indeterminate equilibria, the observations seem to confirm the correctness of the (possibly incorrect) parameter estimates. The impact of indeterminate equilibra on achievable regret rates is, for general control problems, not yet fully understood. That they play an important rôle is apparent, for example, from Broder and Rusmevichientong (2012) who show that in the special case of "well-separable demand functions", which rules out indeterminate equilibria, the smallest achievable regret growth rate is $\log(T)$ instead of $\sqrt{T}$; moreover, this is achieved by a certainty-equivalent control rule, whereas the policies shown to achieve $O(\sqrt{T})$ regret in the general case all require active price experimentation to ensure consistency.

In our setting with finite inventories and finite selling seasons, the optimal price - optimal w.r.t. certain parameter estimates - is not a fixed number, but a collection of prices: one price for each pair of remaining inventory and remaining length of the selling season. Because both these quantities are constantly changing, a certainty equivalent policy induces dispersion in the selling prices. This price dispersion causes the parameter estimates to converge to the true value, and as a result, a (small modification of) certainty equivalent policy works well. The remarks following Theorem 1 further elaborate on the difference between the finite and infinite-inventory setting.

The main conceptual takeaway of our paper is that, in decision problems under uncertainty, a certainty equivalent strategy works well if it induces sufficient dispersion in the controls. We show this for a specific dynamic-pricing problem, but, as we argue in Section 5.3, the idea is also applicable in other decision problems.

1.3. Literature

Our work complements two streams of literature on dynamic-pricing-and-learning. First, in the infinite-capacity setting (Kleinberg and Leighton 2003, Broder and Rusmevichientong 2012, Keskin and Zeevi 2014, den Boer and Zwart 2014b, den Boer 2014) it is known that active price experimentation is necessary to achieve optimal regret; myopic policies have suboptimal performance (den Boer and Zwart 2014b, Section 3.1). In our finite-capacity setting, changes in the marginal-value-of-inventory causes endogenous price dispersion, which makes sure that learning the unknown parameters "takes care of itself", and which leads to a qualitatively much better performance than what is possible in the infinite-capacity setting.

Second, in the finite-capacity setting where demand and inventory level grow to infinity (Besbes and Zeevi 2009, Wang et al. 2014), active price experimentation is a key ingredient in all known asymptotically optimal policies; the amount of price dispersion induced by a certainty equivalent policy appears to be insufficient to ensure consistency and asymptotic optimailty. This asymptotic

regime may have practical value if demand, initial inventory, and the length of the selling season are relatively large. In the application that inspired the current study, pricing in the hotel industry, this is not the case: the average demand, initial capacity and length of a selling season are typically quite small, which makes this particular asymptotic regime not a suitable setting to study the performance of pricing strategies. We therefore consider a different asymptotic regime that allows for small initial inventories and short selling seasons, and we show that in this regime certainty equivalent control performs well. For a comprehensive overview of the literature on dynamic pricing and learning, we refer to den Boer (2013b).

From a methodological point of view, our work is related to the literature on adaptive control in Markov decision problems (Hernández-Lerma 1989, Kumar 1985, chapter 12 of Kumar and Varaiya 1986, Hernández-Lerma and Cavazos-Cadena 1990, Altman and Shwartz 1991, Burnetas and Katehakis 1997, Gordienko and Minjárez-Sosa 1998, Chang et al. 2005) and to the literature on partially observable Markov decision problems (Monahan 1982, Lovejoy 1991) that typically learns unknown parameters in a Bayesian fashion. The topic of combined statistical learning and optimal control is currently an important topic in operations research, and is studied e.g. in inventory control (Kunnumkal and Topaloglu 2008, Huh and Rusmevichientong 2014), assortment optimization (Sauré and Zeevi 2013), network revenue management (Besbes and Zeevi 2012), and many more application areas.

1.4. Organization

The rest of this paper is organized as follows. Section 2 introduces the model primitives and states convergence rates for the maximum likelihood estimator of β. The endogenous-learning property of the system is described in Section 3.1. Our pricing strategy is introduced in Section 3.2, the upper bound $\text{Regret}(T) = O(\log^2(T))$ is shown in Section 3.3, and the $\log(T)$ lower bound in Section 3.4. Numerical illustrations of the pricing strategy and its performance are provided in Section 4. To avoid heavy notation, we assume in these sections that different selling seasons have the same initial inventory and duration. Section 5.1 relaxes these assumptions and shows that $O(\log^2(T))$ regret still can be achieved. We also discuss extensions to non-stationary demand (Section 5.2) and applications of endogenous learning in other decision problems (Section 5.3). The e-companion to this paper contains the mathematical proofs of the theorems in this paper, as well as a number of auxiliary lemmas used in the proofs.

Notation If v is a vector then $||v||$ denotes the Euclidean norm, and v^T the transpose. If A is a square matrix then $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A. For $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer which is smaller than or equal to x. With $\mathbf{1}_E$ we denote the indicator of an event E.

2. Model Primitives

In this section we subsequently introduce the model, describe the characteristics of the demand distribution, discuss the optimal pricing policy under full information, introduce the regret as measure of pricing policies, and discuss convergence rates for the maximum likelihood estimator.

2.1. Model Formulation

We consider a monopolist seller of perishable products which are sold during consecutive selling seasons. Each selling season consists of $S \in \mathbb{N}$ discrete time periods: the i-th selling season starts at time period $1+(i-1)S$, and lasts until period iS, for all $i \in \mathbb{N}$. We write $SS_t = 1 + \lfloor (t-1)/S \rfloor$ to denote the selling season corresponding to period t, and $s_t = t - (SS_t - 1)S$ to denote the relative time in the selling period. At the start of each selling season the seller has $C \in \mathbb{N}$ discrete units of inventory at his disposal, which can only be sold during that particular selling season. At the end of a selling season, all unsold inventory perishes.

In each time period $t \in \mathbb{N}$ the seller has to determine a selling price $p_t \in [p_l, p_h]$. Here $0 < p_l < p_h$ denote the lowest and highest price admissible to the firm. After setting the price the seller observes a realization of demand, which takes values in $\{0,1\}$, and collects revenue $p_t d_t$. We let c_t, $(t \in \mathbb{N})$, denote the capacity or inventory level at the beginning of period $t \in \mathbb{N}$, and d_t the demand in period t. If $c_t = 0$ then $d_t = 0$: no demand is observed if the firm is out-of-stock. (The selling price p_t in these periods does not affect the revenue, and may be chosen arbitrarily). The distribution of d_t in case $c_t > 0$ is described in Section 2.2. The dynamics of $(c_t)_{t\in\mathbb{N}}$ are given by

$$\begin{array}{ll} c_t = C & \text{if } s_t = 1, \\ c_t = c_{t-1} - d_{t-1} & \text{if } s_t \neq 1. \end{array}$$

Notice that c_t can not become smaller than zero, since $d_{t-1} = 0$ if $c_{t-1} = 0$.

The pricing decisions of the seller are allowed to depend on previous prices and observed demand realizations, but not on future ones. More precisely, for each $t \in \mathbb{N}$ the set of possible histories $\mathcal{H}_t$ as

$$\mathcal{H}_t = \{(p_1, \ldots, p_t, d_1, \ldots, d_t) \in [p_l, p_h]^t \times \{0,1\}^t\},$$

with $\mathcal{H}_0 = \{\emptyset\}$. A pricing strategy $\psi = (\psi_t)_{t\in\mathbb{N}}$ is a collection of functions $\psi_t : \mathcal{H}_{t-1} \to [p_l, p_h]$, such that $p_1 = \psi_1(\emptyset)$, and for each $t \geq 2$ the seller chooses the price $p_t = \psi_t(p_1, \ldots, p_{t-1}, d_1, \ldots, d_{t-1})$.

The purpose of the seller is to find a pricing strategy ψ that maximizes the cumulative expected revenue earned after T selling seasons, $\sum_{i=1}^{TS} E_\psi[p_i d_i]$. Here we write E_ψ to emphasize that this expectation depends on the pricing strategy ψ.

2.2. Demand Distribution

The demand in a single time period with positive inventory, against selling price p, is a realization of the random variable $D(p)$. We assume that $D(p)$ is Bernoulli distributed with mean $E[D(p)] = h(\beta_0 + \beta_1 p)$, for all $p \in [p_l, p_h]$, some $(\beta_0, \beta_1) \in \mathbb{R}^2$, and some function h. The true value of β is denoted by $\beta^{(0)}$, and is unknown to the seller. Conditionally on selling prices, the demand in any two different time periods are independent. We assume that $\beta^{(0)}$ lies in the interior of a set $B := [\beta_{l,0}, \beta_{u,0}] \times [\beta_{l,1}, \beta_{u,1}] \subset \mathbb{R}^2$, for some known lower and upper bounds $\beta_{l,0}$, $\beta_{u,0}$, $\beta_{l,1}$, $\beta_{u,1}$ on β_0 and β_1, respectively, and with $\beta_{u,1} < 0$. Furthermore we assume that $h(z)$ is three times continuously differentiable in z, log-concave, $h(z) \in (0,1)$ and $\dot{h}(z) > 0$, for all $z \in \{\beta_0 + \beta_1 p \mid p \in [p_l, p_h], \beta \in B\}$; here $\dot{h}$ denotes the derivative of h.

Demand functions that fit into our framework (with appropriate conditions on B and $[p_l, p_h]$) include $h(z) = \exp(z)$, $h(z) = z$, and $h(z) = \text{logit}(z) = \exp(z)/(1+\exp(z))$.

2.3. Full-information Optimal Solution

If the value of β is known, the optimal prices can be determined by solving a Markov decision problem (MDP). Since each selling season corresponds to the same MDP, the optimal pricing strategy for multiple selling seasons is to repeatedly use the optimal policy for a single selling season. The state space of this MDP is $\mathcal{X} = \{(c,s) \mid c = 0, \ldots, C, s = 1, \ldots, S\}$, where (c,s) means that there are c units of remaining inventory at the beginning of the s-th period of the selling season, and the action space is the interval $[p_l, p_h]$. If action p is used in state (c,s), $s < S$, then with probability $h(\beta_0 + \beta_1 p)$ a state transition $(c,s) \to ((c-1)^+, s+1)$ occurs and reward $ph(\beta_0 + \beta_1 p)\mathbf{1}_{c>0}$ is obtained; with probability $1 - h(\beta_0 + \beta_1 p)$ a state transition $(c,s) \to (c, s+1)$ occurs and zero reward is obtained. The states (c,S) are terminal states; the reward using action p equals $ph(\beta_0 + \beta_1 p)\mathbf{1}_{c>0}$ with probability $h(\beta_0 + \beta_1 p)$, and zero with probability $1 - h(\beta_0 + \beta_1 p)$.

A (stationary deterministic) policy π is a matrix $(\pi(c,s))_{0 \le c \le C, 1 \le s \le S}$ in the policy space $\Pi = [p_l, p_h]^{(C+1)\times S}$. Given a policy $\pi \in \Pi$, let $V_\beta^\pi(c,s)$ be the expected revenue-to-go function starting in state $(c,s) \in \mathcal{X}$ and using the actions of π. Then $V_\beta^\pi(c,s)$ satisfies the following recursion:

$$\begin{aligned} V_\beta^\pi(c,s) &= (1 - h(\beta_0 + \beta_1 \pi(c,s))) \cdot V_\beta^\pi(c, s+1) \\ &\quad + h(\beta_0 + \beta_1 \pi(c,s)) \cdot (\pi(c,s) + V_\beta^\pi(c-1, s+1)), \quad (1 \le c \le C), \qquad (1) \\ V_\beta^\pi(0,s) &= 0, \qquad (2) \end{aligned}$$

for all $1 \le s \le S$, where we write $V_\beta^\pi(c, S+1) = 0$ for all $0 \le c \le C$.

By Proposition 4.4.3 of Puterman (1994), for each $\beta \in B$ there is a corresponding optimal policy $\pi_\beta^* \in \Pi$. This policy can be calculated using backward induction. Write $V_\beta(c,s) = V_\beta^{\pi_\beta^*}(c,s)$ for the

optimal revenue-to-go function. Then $V_\beta(c,s)$ and $\pi^*_\beta(c,s)$, for $1 \le c \le C$, $1 \le s \le S$, satisfy the following recursion:

$$\begin{aligned} V_\beta(c,s) &= \max_{p\in[p_l,p_h]} \big[p - \Delta V_\beta(c,s+1)\big] h(\beta_0+\beta_1 p) + V_\beta(c,s+1), \\ \pi^*_\beta(c,s) &= \arg\max_{p\in[p_l,p_h]} \big[p - \Delta V_\beta(c,s+1)\big] h(\beta_0+\beta_1 p), \end{aligned} \tag{3}$$

where we define $\Delta V_\beta(c,s) = V_\beta(c,s) - V_\beta(c-1,s)$, and $\Delta V_\beta(0,s) = 0$ for all $1 \le s \le S$. The price $\pi^*_\beta(0,s)$ can be chosen arbitrarily, since it has no effect on the reward. For $c \ge 1$, the prices $\pi^*_\beta(c,s)$ in (3) are uniquely defined; this follows from Lemma EC.1. The optimal reward of the MDP is equal to $V_\beta(C,1)$, and the true optimal reward is equal to $V_{\beta^{(0)}}(C,1)$.

We assume that p_l and p_h satisfy

$$p_l < \pi^*_{\beta_l}(C,S) \text{ and } \pi^*_{\beta_u}(1,1) < p_h, \tag{4}$$

where we write $\beta_l = (\beta_{l,0}, \beta_{l,1})$ and $\beta_u = (\beta_{u,0}, \beta_{u,1})$. By Lemma 1, this condition ensures that the optimal price decisions do not depend on the boundary prices p_l and p_h. Note that equation (4) is not difficult to check in practice; it only involves solving the MDP for $\beta = \beta_l$ and $\beta = \beta_u$.

LEMMA 1. *$\pi^*_\beta(c,s) \in (p_l, p_h)$, for all $\beta \in B$, $1 \le c \le C$, and $1 \le s \le S$.*

Proof of Lemma 1. Let $\beta \in B$, $1 \le c \le C$, $1 \le s \le S$, and let $p^*_{a,\beta}$ be as in Lemma EC.1. By application of Lemma EC.2(v), Lemma EC.2(iv), and Lemma EC.1(ii), it follows that

$$\pi^*_\beta(c,s) \le \pi^*_\beta(1,1) = p^*_{\Delta V_\beta(1,2),\beta} = p^*_{V_\beta(1,2),\beta} \le p^*_{V_{\beta_u}(1,2),\beta_u} = \pi^*_{\beta_u}(1,1) < p_h,$$

and

$$\pi^*_\beta(c,s) \ge \pi^*_\beta(C,S) = p^*_{0,\beta} \ge p^*_{0,\beta_l} = \pi^*_{\beta_l}(C,S) > p_l.$$

□

2.4. Regret Measure

The quality of the pricing decisions of the seller are measured by the regret: the expected amount of money lost due to not using optimal prices. The regret of pricing strategy ψ after the first T selling seasons is defined as

$$\text{Regret}(\psi,T) = T \cdot V_{\beta^{(0)}}(C,1) - \sum_{i=1}^{TS} E[p_i d_i], \tag{5}$$

where $(p_i)_{i\in\mathbb{N}}$ denote the prices generated by the pricing strategy ψ.

Maximizing the cumulative expected revenue is equivalent to minimizing the regret, but observe that the regret cannot directly be used by the seller to find the optimal strategy, since it depends on the unknown $\beta^{(0)}$.

2.5. Parameter Estimation

We can estimate the unknown parameter $\beta^{(0)}$ with maximum-likelihood estimation. Define the log-likelihood function

$$L_t(\beta) = \sum_{i=1}^{t} \log \left[h(\beta_0 + \beta_1 p_i)^{d_i} (1 - h(\beta_0 + \beta_1 p_i))^{1-d_i} \right] \mathbf{1}_{c_i > 0}, \tag{6}$$

and the score function (the derivative of $L_t(\beta)$ with respect to β)

$$l_t(\beta) = \sum_{i=1}^{t} \frac{\dot{h}(\beta_0 + \beta_1 p_i)}{h(\beta_0 + \beta_1 p_i)(1 - h(\beta_0 + \beta_1 p_i))} \begin{pmatrix} 1 \\ p_i \end{pmatrix} (d_i - h(\beta_0 + \beta_1 p_i)) \mathbf{1}_{c_i > 0}. \tag{7}$$

We define $\hat{\beta}_t$ to be a solution $\beta \in B$ of $l_t(\beta) = 0$; if multiple solutions exists, we choose the one that maximizes $L_t(\beta)$. If there is no solution of $l_t(\beta) = 0$ in B, we define $\hat{\beta}_t$ as the smallest maximizer (in the lexicographic ordering) of $L_t(\beta)$ on B; in this case $\hat{\beta}_t$ necessarily lies on the boundary of B. Note that $\hat{\beta}_t$ only depends on sales data of periods with positive inventory.

Remark 1. Because we allow for a general class of functions h (so-called nonnatural or general link functions, in the terminology of generalized linear models), the likelihood function $L_t(\beta)$ is not necessarily concave and the solution to $l_t(\beta) = 0$ is not necessarily unique; cf. the discussion in Section 4.1 of Fahrmeir and Kaufmann (1985). However, Proposition 1 in Section 2.6 guarantees that, for all sufficiently large t, $l_t(\beta) = 0$ has a solution in B, and provides a condition that ensures convergence to $\beta^{(0)}$. This is the reason that we define $\hat{\beta}_t$ as solution to $l_t(\beta) = 0$, instead of directly as maximizer of the log-likelihood function. If h is the logit function (the so-called canonical link function in this context), then $L_t(\beta)$ is concave and the solution to $l_t(\beta) = 0$ is unique.

2.6. Convergence Rates of Parameter Estimates

Understanding the asymptotic behavior of the maximum likelihood estimate, in particular the speed at which it converges to $\beta^{(0)}$, is important for studying the performance of pricing strategies. We include a result about these convergence rates based on den Boer and Zwart (2014a); in Section 3.3, this result is used to prove bounds on the regret of a pricing strategy.

The speed at which the estimates converge to $\beta^{(0)}$ turns out to be closely related to a certain measure of price dispersion: the more price dispersion, the faster the parameters converge. In particular, if we define the matrix

$$P_t = \sum_{i=1}^{t} \begin{pmatrix} 1 & p_i \\ p_i & p_i^2 \end{pmatrix} \mathbf{1}_{c_i > 0}, \quad (t \in \mathbb{N}), \tag{8}$$

then $\lambda_{\min}(P_t)$, the smallest eigenvalue of P_t, turns out to be a suitable measure for the amount of price dispersion in a sample.

The following proposition shows how $\lambda_{\min}(P_t)$ influences the convergence speed of the parameter estimates. To state the result, we define for all $\rho > 0$ the last-time random variable

$$T_\rho = \sup\left\{t \in \mathbb{N} \,|\, \text{there is no } \beta \in B \text{ with } \left|\left|\beta - \beta^{(0)}\right|\right| \leq \rho \text{ and } l_t(\beta) = 0\right\}, \tag{9}$$

PROPOSITION 1. *Suppose L is a non-random function on $\mathbb{N}$ such that $\lambda_{\min}(P_t) \geq L(t) > 0$ a.s., for all $t \geq t_0$ and some non-random $t_0 \in \mathbb{N}$, and such that $\inf_{t \geq t_0} L(t)t^{-\alpha} > 0$, for some $\alpha > 1/2$. Then there exists a $\rho_1 > 0$ such that for all $0 < \rho \leq \rho_1$ we have $T_\rho < \infty$ a.s., $E[T_\rho] < \infty$, and for each $t > T_\rho$ there is a $\hat{\beta}_t \in B$ with $l_t(\hat{\beta}_t) = 0$, and*

$$E\left[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t > T_\rho}\right] = O\left(\log(t)/L(t)\right).$$

By application of Theorem 1, Theorem 2, and Remark 2 in den Boer and Zwart (2014a), with $x_i = (1, p_i)^T \mathbf{1}_{c_i > 0}$ for all $i \in \mathbb{N}$, the statement follows.

3. Main Result: a Case of Endogenous Learning

The model described in the previous section satisfies an *endogenous-learning* property: if the decision maker does not deviate much from the optimal price policy, then the unknown parameters of the system are learned very fast. This is caused by a natural amount of price dispersion that appears when the optimal policy is used. This dispersion causes the estimates $\hat{\beta}_t$ to converge very quickly to the unknown parameters $\beta^{(0)}$, and as a result, the decision maker can use a simple myopic pricing policy to achieve a very good performance. This is the main takeaway of this paper.

The endogenous-learning property is formally stated in Section 3.1. In Section 3.2 we formulate a pricing strategy, which (apart from a small correction) is equal to a myopic strategy. The endogenous-learning property causes the regret of this pricing strategy to grow as $\text{Regret}(T) = O(\log^2(T))$; this is shown in Section 3.3. Remark 3 proposes an alternative myopic pricing strategy with estimates based on *completed* selling seasons, and argues that the same $O(\log^2(T))$ regret bound applies. These upper bounds are complemented by a lower bound in Section 3.4, where we show an instance for which no pricing strategy can achieve sub-logarithmic regret.

3.1. Endogenous Learning

The main result of this section is that $\lambda_{\min}(P_t)$ strictly increases if, during a selling season, prices are used that are close to those prescribed by π^*_β, for any $\beta \in B$. This means that a continuous use of prices close to $\pi^*_{\beta^{(0)}}$ leads to a linear growth rate of $\lambda_{\min}(P_t)$, which by Proposition 1 implies that the parameter estimates converges very fast to the true value, in particular with rate $E\left[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t > T_\rho}\right] = O\left(\log(t)/t\right)$.

THEOREM 1. *Let $1 < C < S$ and $k \in \mathbb{N}$. For each $\beta \in B$ there exist an open neighborhood $\mathcal{U}_\beta \subset \mathbb{R}^2$ containing β and a constant $v_0 > 0$ independent of β, such that, if*

$$p_{s+(k-1)S} = \pi^*_{\beta(s)}(c_{s+(k-1)S}, s)$$

for all $s = 1, \ldots, S$ and some sequence $\beta(1), \ldots, \beta(S) \in \mathcal{U}_\beta \cap B$, then there are $1 \leq s, s' \leq S$ with

$$|p_{s+(k-1)S} - p_{s'+(k-1)S}| \geq v_0/2, \quad c(s+(k-1)S)c(s'+(k-1)S) > 0 \tag{10}$$

and

$$\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S}) \geq \frac{1}{8} v_0^2 (1+p_h^2)^{-1}. \tag{11}$$

The condition $C < S$ in Theorem 1 is essential. The setting with $C \geq S$ can be interpreted as that $C - S$ items cannot be sold at all, and that each of the remaining S items can only be sold in a single, dedicated time period. Thus, there is no interaction between individual items, and the pricing problem is equivalent to S repetitions of the pricing problem with $C = 1$, $S = 1$, for which no price dispersion occurs. Phrased differently: if $C \geq S$ then the marginal-value-of-inventory remains constant throughout the selling season, and thus the optimal price is constant as well. Broder and Rusmevichientong (2012), den Boer and Zwart (2014b) and Keskin and Zeevi (2014) consider pricing strategies for this case, and show that the lack of endogenous learning means that active price experimentation is necessary to learn the unknown parameters. Section 4.4 numerically explores the effect of C and S on the amount of price dispersion.

If $C = 1$ then the firm may go out-of-stock in the first period of a selling season, resulting in a selling season with zero price dispersion. Consequently, it is not possible to find a strictly positive lower bound on the price dispersion per season that holds with probability one (but price dispersion may occur with probability between zero and one). Because our results rely on an a.s. strictly positive lower bound of the price dispersion, our results do not cover the case $C = 1$, and different proof techniques are required to analyze this case.

The proof of Theorem 1 is contained in the Appendix, Section EC.1.

3.2. Pricing Strategy

We propose a pricing strategy based on the following principle: in each period, estimate the unknown parameters, and subsequently use the action from the policy that is optimal with respect to this estimate.

Pricing strategy $\Phi(\epsilon)$

Initialization: Choose $0 < \epsilon < (p_h - p_l)/4$, and initial prices p_1, $p_2 \in [p_l, p_h]$, with $p_1 \neq p_2$.

For all $t \geq 2$: if $c_{t+1} = 0$, set $p_{t+1} \in [p_l, p_h]$ arbitrarily. If $c_{t+1} > 0$:

Estimation: Determine $\hat{\beta}_t$, and let $p_{\text{ceqp}} = \pi^*_{\hat{\beta}_t}(c_{t+1}, s_{t+1})$.

Pricing:

I) If

(a) $|p_i - p_j| < \epsilon$ for all $1 \leq i, j \leq t$ with $SS_i = SS_j = SS_{t+1}$, and

(b) $|p_i - p_{\text{ceqp}}| < \epsilon$ for all $1 \leq i \leq t$ with $SS_i = SS_{t+1}$, and

(c) $c_{t+1} = 1$ or $s_{t+1} = S$,

then choose $p_{t+1} \in \big(\{p_{\text{ceqp}} + 2\epsilon, p_{\text{ceqp}} - 2\epsilon\} \cap [p_l, p_h]\big)$.

II) Else, set $p_{t+1} = p_{\text{ceqp}}$.

Given a positive inventory level, the pricing strategy $\Phi(\epsilon)$ sets the price p_{t+1} equal to the price that is optimal according to $\hat{\beta}_t$, except possibly when the state (c_{t+1}, s_{t+1}) is in the set $\{(c, s) \mid c = 1 \text{ or } s = S\}$. This set contains all states that, with positive probability, are the last states in the selling season in which products are sold (either because the selling season almost finishes, or because the inventory consists of only a single product). In these states, the price p_{t+1} deviates from the certainty equivalent price p_{ceqp} if otherwise $\max\{|p_i - p_j| \mid SS_i = SS_{t+1}\} < \epsilon$. This deviation ensures that also for small t, when $\hat{\beta}_t$ may be far away from the true value $\beta^{(0)}$, a minimum amount of price dispersion is guaranteed.

3.3. Upper Bound on the Regret

The endogenous-learning property described in Section 3.1 implies that if $\hat{\beta}_t$ is sufficiently close to $\beta^{(0)}$ and ϵ is sufficiently small, then I) in the formulation of $\Phi(\epsilon)$ does not occur. As $\hat{\beta}_t$ converges to $\beta^{(0)}$, the pricing strategy $\Phi(\epsilon)$ eventually acts as a certainty equivalent pricing strategy. The pricing decisions in II) are driven by optimizing current season revenue, and do not reckon with the objective of optimizing the quality of the parameter estimates $\hat{\beta}_t$. The endogenous-learning property ensures that the unknown parameter values are learned on the fly, and that the pricing decisions converge quickly to the optimal pricing decisions. The following theorem shows that the regret of the strategy $\Phi(\epsilon)$ is $O(\log^2(T))$.

THEOREM 2. *Let $1 < C < S$, v_0 as in Theorem 1, and $\epsilon < v_0/2$. Then*

$$\textit{Regret}(\Phi(\epsilon), T) = O(\log^2(T)).$$

To prove Theorem 2, we construct a Markov decision problem with a state-space that consists of all sequences of possible demand realizations in a selling season. This ensures that, conditional on all prices and demand realizations before a selling season, $\Phi(\epsilon)$ corresponds to a stationary deterministic policy, where each state of the state-space is associated with a unique price prescribed by $\Phi(\epsilon)$. We subsequently prove several sensitivity results that enable us to quantify the effect of estimation errors $||\hat{\beta}_t - \beta^{(0)}||$ on the regret. Application of the convergence rates in Proposition 1 then imply the $O(\log^2(T))$ bound on the regret. The proof of the theorem is contained in the Appendix, Section EC.1.

An expression for v_0 is given in the proof of Theorem 1. This makes it possible to explicitly determine values of ϵ for which Theorem 2 is valid.

REMARK 2. The pricing strategy $\Phi(\epsilon)$ would be more elegant if $\epsilon = 0$ would be allowed; this would remove all the special cases in I) of the specification of $\Phi(\epsilon)$, and would result in a "purely" myopic strategy. Unfortunately, removing the requirement $\epsilon > 0$ creates technical difficulties in proving the upper bound on the regret. Concretely, an essential ingredient of the proof is a *deterministic* lower bound on $\lambda_{\min}(P_t)$; this enables us to apply Proposition 1 which ensures consistency and provides convergence rates for $\hat{\beta}_t$. Without the requirement $\epsilon > 0$ the existence of such deterministic lower bound is not ensured, and different proof techniques are necessary to prove the regret upper bound. A possible route could be to try to prove the conjecture $\lim_{t\to\infty} ||\hat{\beta}_t - \hat{\beta}_{t-1}|| = 0$ a.s., regardless how prices $p_t \in [p_l, p_h]$, $t \in \mathbb{N}$, are chosen. If this conjecture is true then, for all sufficiently large $k \in \mathbb{N}$, $\hat{\beta}_{1+(k-1)S}, \ldots, \hat{\beta}_{kS}$ all lie sufficiently close to each other to ensure by Theorem 1 that the myopic prices based on these estimates have a positive amount of price dispersion in selling season k. This would be a large step towards proving that Theorem 2 also holds for $\epsilon = 0$, i.e. for the "purely" myopic pricing strategy. Proving this conjecture seems far from trivial, however.

REMARK 3. An alternative approach to make a "purely" myopic strategy work, i.e. $\Phi(\epsilon)$ with $\epsilon = 0$, is the following: instead of updating the estimates $\hat{\beta}_t$ each time period, one could estimate the parameters solely at the beginning of each selling season. Concretely that means that $\hat{\beta}_t$ in the estimation step of $\Phi(\epsilon)$ is replaced by $\hat{\beta}_{(SS_t-1)S}$, with $\hat{\beta}_0$ chosen arbitrarily in B. By Theorem 1 the deterministic lower bound $\lambda_{\min}(P_{kS}) \geq (k-1) \cdot \frac{1}{8} v_0^2 (1+p_h^2)^{-1}$ is valid for all $k \in \mathbb{N}$, and it is not difficult to show along the same lines of Theorem 2 that this policy $\Phi(0)$ satisfies

$$\text{Regret}(\Phi(0), T) = O(\log^2(T)).$$

The potential downside of using $\Phi(0)$ instead of $\Phi(\epsilon)$, $\epsilon > 0$, is that some of the available sales data is neglected when forming estimates. Neglecting data generally leads to lower revenues (compare, for example, the numerical performance of the strategies MLE-CYCLE and MLE-CYCLE-S in

Broder and Rusmevichientong (2012)), although counterexamples to this intuition are also known (den Boer 2013a). In this paper we do not further elaborate on the drawbacks or benefits of using all available sales data.

3.4. Lower Bound on the Regret

In this section we complement the $O(\log^2(T))$ upper bound of Theorem 2 by a lower bound on the regret. In particular, we show an instance for which *any* pricing strategy has regret that grows at least logarithmically in T. This shows that the asymptotic growth rate of regret of $\Phi(\epsilon)$ is close to the best achievable asymptotic growth rate.

THEOREM 3. *Let $1 < C < S$, h the identity function, $[p_l, p_h] = [3/10, 8/10]$, and let $B = [5/8, 6/8] \times [-3/4, -9/16]$. There is a constant K_0 such that, for all pricing strategies ψ and all $T \in \mathbb{N}$,*

$$\sup_{\beta^{(0)} \in B} \mathit{Regret}(\psi, T) \geq K_0 \log(T).$$

The proof of Theorem 3 consists of two main steps. In the first step we show that the regret in a single selling season is bounded from below by a term proportional to the expected estimation error in a single time period. In the second step we further bound this term, using an adaptation of the van Trees inequality (Gill and Levit 1995) to our setting where β is estimated with a sample of random size (caused by the $\mathbf{1}_{c_i>0}$ terms in (7)). The proof of Theorem 3 is contained in the Appendix, Section EC.1.

REMARK 4 (GAP BETWEEN LOWER AND UPPER BOUND ON THE REGRET). Theorem 2 shows that the regret of our pricing strategy $\Phi(\epsilon)$ is $O(\log^2(T))$, and Theorem 3 shows that the regret of any pricing strategy grows at least as $\log(T)$. This "gap" between $\log^2(T)$ and $\log(T)$ points to the question whether Theorem 2 can be strengthened to $O(\log(T))$. This question turns out to be rather difficult to answer. The "additional" $\log(T)$ term is caused by the $\log(t)$ term in the convergence rates $E\left[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t>T_\rho}\right] = O(\log(t)/L(t))$ of Proposition 1. This $\log(t)$ term can be traced back to Proposition 2 of den Boer and Zwart (2014a), who extend the a.s. convergence rates of least-squares linear-regression estimators obtained by Lai and Wei (1982) to convergence rates in expectation. Nassiri-Toussi and Ren (1994) show that in some cases the $\log(t)$ term is really present in the behavior of least-squares estimates, and thus cannot simply be removed. On the other hand, if the design is non-random and the disturbance terms are normally distributed, it can be shown that this $\log(t)$-term in Proposition 2 of den Boer and Zwart (2014a) can be removed. It is not at all clear how to determine, for a particular adaptive design, whether the log-term plays a role in the asymptotic behavior of linear regression estimates. Consequently, it is very hard to determine whether the log-term in Theorem 2 is present in practice, or is merely a result of the used proof techniques. For practical applications this issue is fortunately not very important, as it is quite hard to determine from data if a functions grows like $\log(T)$ or like $\log^2(T)$.

4. Numerical Illustrations

To illustrate the analytical results that we have derived, we provide a number of numerical illustrations. We first offer a simple instance that illustrates strong consistency of the parameter estimates and convergence of the relative regret to zero. We also briefly consider the "gap" between the upper bound of Theorem 2 and the lower bound of Theorem 3. We subsequently look at an instance where we vary the level of initial inventory C and the duration of the selling season S, and look at the effect on the regret. In the last illustration we look at the effect of different values of (C, S) on the amount of price dispersion, and connect this with the asymptotic regime considered in Besbes and Zeevi (2009) and Wang et al. (2014). To speed up the simulations, parameter estimates were not updated during selling seasons (cf. Remark 3).

4.1. Basic Example

As a first example, we consider an instance with $C = 10$, $S = 20$, $p_l = 1$, $p_h = 20$, $\beta_0^{(0)} = 2$, $\beta_1^{(0)} = -0.4$, and $h(z) = \text{logit}(z)$. The optimal expected revenue per selling season, $V_{\beta^{(0)}}(C, 1)$, is equal to 47.8. We consider a time span of 100 selling periods, and run 100 simulations.

Figure 1 shows the simulation average of the regret after each selling season, and of the relative regret defined by

$$\text{Relative regret}(T) = \frac{\text{Regret}(T)}{T \cdot V_{\beta^{(0)}}(C, 1))} \times 100\%.$$

To show some light on the growth rate of the regret, we scale in Figure 2 the regret by a $\log(T)$ and a $\log^2(T)$ factor. Theorem 2 entails that $\text{Regret}(T)/\log^2(T)$ is bounded, which accords with the righthand plot in Figure 2. However, Theorem 3 suggests that the $O(\log^2(T))$ bound may be too conservative, and that in fact the regret may grow logarithmically (cf. the discussion in Remark 4). The lefthand plot of Figure 2 shows the regret scaled by a log-factor. This picture does not strongly support the assertion that $\text{Regret}(T)/\log(T)$ is bounded, but this may be caused by finite-horizon effects. Our numerical simulation thus does not give a conclusive answer on the question whether this "gap" really exists in practice, or merely is a consequence of used proof techniques. Different choices for $\beta^{(0)}$ show a similar picture.

4.2. Different Levels of Initial Inventory

In our second numerical example we illustrate the effect of initial inventory on the regret. We consider the same instance as in the previous example, but take $S = 10$ and $C \in \{1, 2, 3, \ldots, 9\}$, and run 100 simulations for each value of C. Table 1 shows for each C the optimal revenue per selling season, and the simulation average of the regret, the relative regret, and the estimation error at the end of the time horizon.

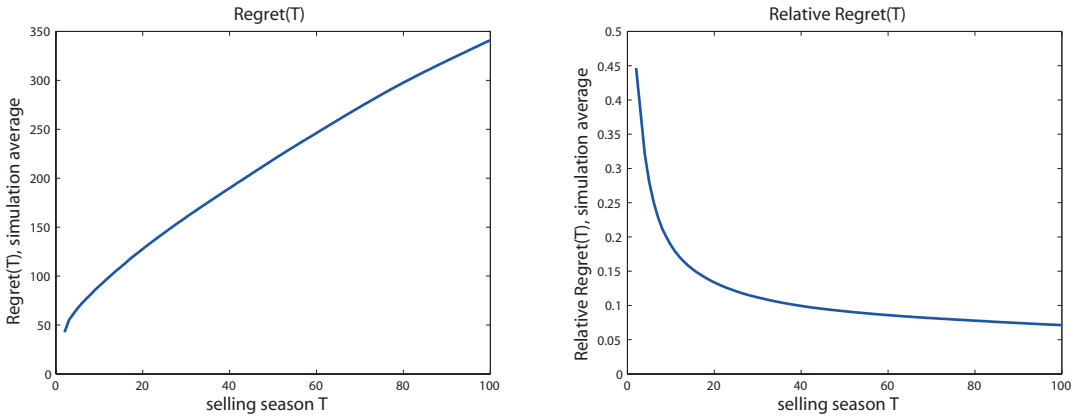

Figure 1 Simulation average of regret and relative regret

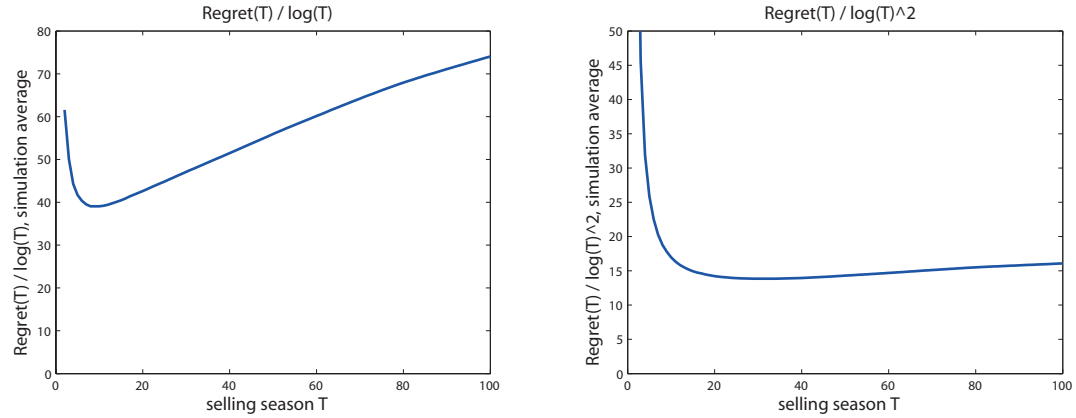

Figure 2 Simulation average of regret, scaled by $\log(t)$ and $\log^2(t)$.

Table 1 Simulation output for various choices of C

C	$V_{\beta^{(0)}}(C,1)$	Regret(100)	Relative regret(100)	$\|\|\hat{\beta}_{1000} - \beta^{(0)}\|\|$
1	8.00	37.01	4.63 %	0.517
2	13.79	49.38	3.58 %	0.478
3	18.06	73.59	4.07 %	0.522
4	21.10	109.0	5.16 %	0.566
5	23.10	199.5	8.64 %	0.753
6	24.24	308.7	12.7 %	1.08
7	24.78	352.5	14.2 %	1.20
8	24.96	395.5	15.9 %	1.33
9	25.00	392.2	15.7 %	1.32

The fourth column of Table 1 suggests that the relative regret is not monotone in C, but is minimal for some C strictly between 1 and S. This can intuitively be explained as follows. For larger values of C, the fraction of time that the firm is out-of-stock is small; this means that estimates are based on more data, which generally increases the quality of the parameter estimates. However, if C gets close to S then the amount of price dispersion induced by the myopic policy

decreases: for a substantial portion of a selling season there is hardly any variation in the marginal-value-of-inventory, and as result the optimal price for different states (c, s) in the state-space of the underlying MDP does not vary much. This behavior is reflected in the average estimation error at the end of the time horizon, shown in the fifth column of Table 1.

4.3. Different Length of Selling Season

In our third numerical illustration we consider the same instance as in the previous two illustrations, but fix the inventory level at $C = 5$, and vary the length of the selling season. We let $S \in \{6, 7, \ldots, 14\}$, and for each choice of S run 100 simulations. Table 2 shows for each S the optimal revenue per selling season, and the simulation average of the regret, the relative regret, and the estimation error at the end of the time horizon.

Table 2 Simulation output for various choices of S

S	$V_{\beta^{(0)}}(C, 1)$	Regret(100)	Relative regret(100)	$\|\|\hat{\beta}_{100S} - \beta^{(0)}\|\|$
6	14.94	243.7	16.3 %	1.246
7	17.25	256.8	14.9 %	1.216
8	19.38	247.6	12.8 %	1.091
9	21.33	231.9	10.9 %	0.946
10	23.10	207.5	8.98 %	0.780
11	24.70	156.0	6.31 %	0.635
12	26.17	120.6	4.61 %	0.529
13	27.51	119.0	4.33 %	0.500
14	28.74	106.2	3.70 %	0.442

The results from Table 2 show that the relative regret is decreasing in S. This is not surprising: larger values of S means that there are not only more opportunities to sell products, but also more opportunities to learn about customer behavior. This is reflected in the fifth column of the table, which shows that the simulation average of the estimation error at the end of the time horizon is decreasing in S.

4.4. Effect of C and S on Price Dispersion

The results from Section 4.2 indicate that the ratio between C and S influences the convergence speed of parameter estimates. Intuitively, the following happens: if C/S is close to zero, then the seller is relatively often out-of-stock; as a result less historical data is available to form estimates, which in general leads to larger estimation errors. If C/S is close to (but strictly smaller than) one, then the myopic policy induces less price dispersion; as long as the state (c, s) of the underlying MDP has $c/(S - s)$ "close to" one (we do not further quantify this statement here), the prices stay close to the price that is optimal for $C = S$, and do not generate much price dispersion.

To gain some insight on the influence of C and S on the growth rate of $\lambda_{\min}(P_t)$, we provide two numerical illustrations.

In the first, we take $p_l = 1$, $p_h = 100$, $\beta_0^{(0)} = 2$, $\beta_1^{(0)} = -0.4$, $h(z) = \text{logit}(z)$. We fix $C = 10$ and choose $S \in \{10, 20, 50, 100, 200, 500\}$. For a fair comparison, we let the number of selling seasons T be equal to $1000/S$; the total time horizon then consists of 1000 time periods, for each experiment. For each choice of S, we perform 100 simulations and record the price dispersion measured by $\lambda_{\min}(P_t)$, for $t = 1, \ldots, 1000$.

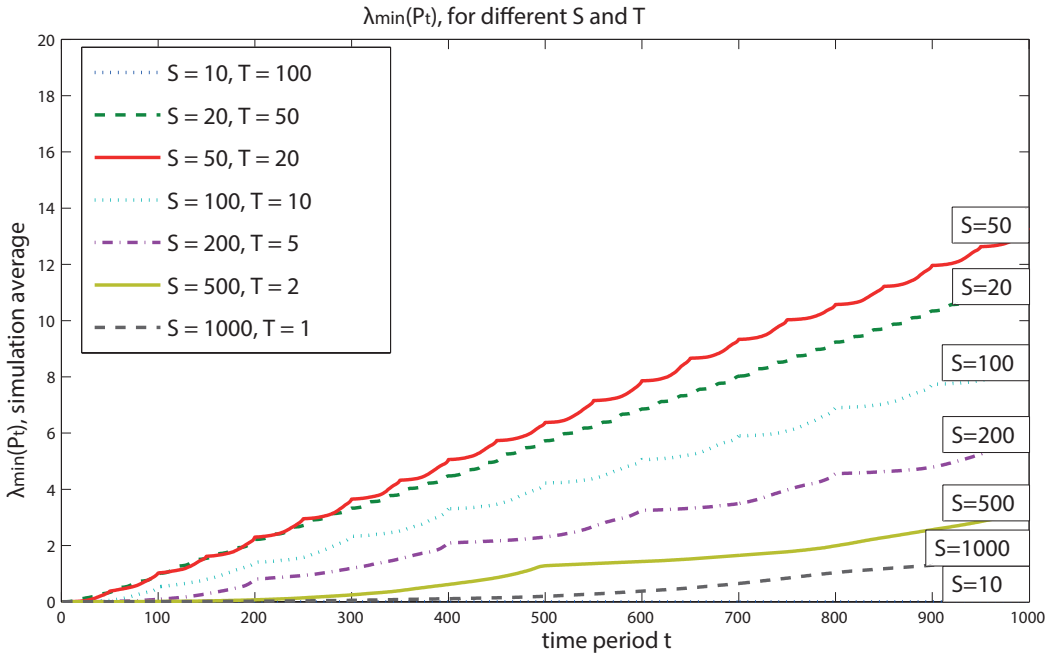

Figure 3 Price dispersion, for different values of S and T

Figure 3 shows the simulation average of $\lambda_{\min}(P_t)$ for $t = 1, \ldots, 1000$, for the different values of (S, T). For all experiments, $\lambda_{\min}(P_t)$ grows linearly in t. The magnitude of the growth rate (i.e. the slope of each graph in the figure) depends on the particular choice of S and T.

This magnitude affects the speed at which parameter estimates converge to the true value. Figure 4 shows for $S \in \{10, 20, 50, 1000\}$ the simulation average of the estimation error $||\hat{\beta}_t - \beta^{(0)}||$, where $\hat{\beta}_t$ is based on the available prices and demand realizations induced by the optimal policy. The figure shows that the estimation error $||\hat{\beta}_t - \beta^{(0)}||$ converges quicker to zero if the price dispersion $\lambda_{\min}(P_t)$ grows at a faster rate. For the case $S = 10$ the parameter estimates do not converge to the true value, and $\lambda_{\min}(P_t)$ does not grow to infinity. This is the case with $C = S$, where active price experimentation is necessary to ensure consistency (see our comments following Theorem 1).

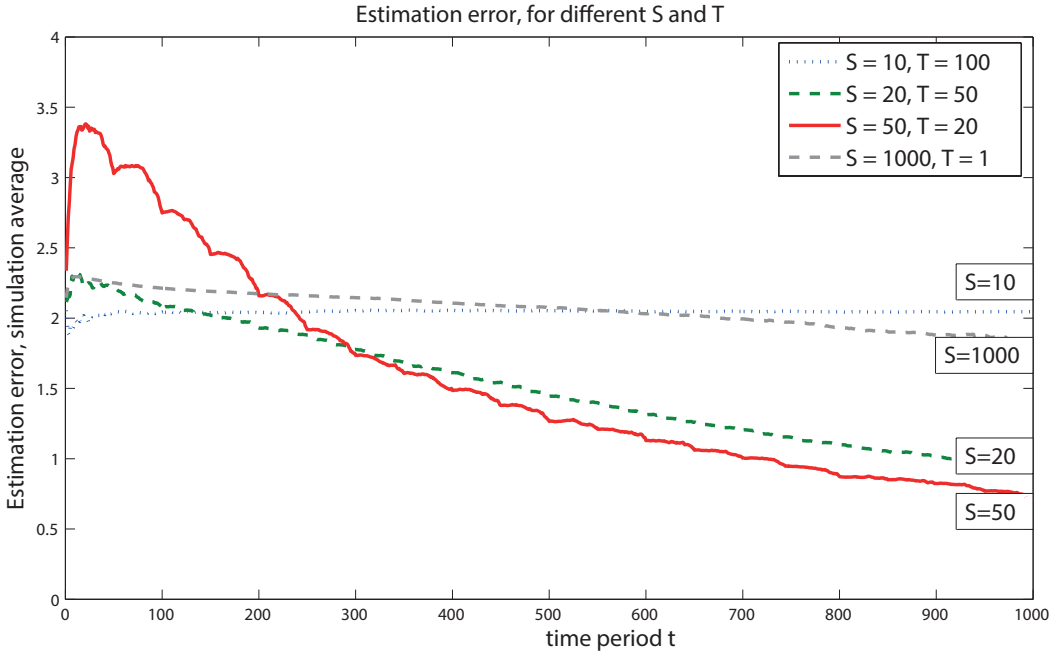

Figure 4 Estimation error $||\hat{\beta}_t - \beta^{(0)}||$, for different values of S and T

Figure 3 shows that the amount of price dispersion at the end of the time horizon, $\lambda_{\min}(P_t)$ at $t = 1000$, is not monotone in S: the largest growth rate is achieved at the experiment with $S = 50$, $T = 20$; for S larger than 50 it is decreasing in S, and for S smaller than 50 it is increasing in S. This is in accordance with the intuition outlined above, which says that the price dispersion grows slowly if C/S is close to zero or close to one.

In our second numerical illustration, we look at a scaling of C and S. We take the same instance as above (i.e. $p_l = 1$, $p_h = 100$, $\beta_0^{(0)} = 2$, $\beta_1^{(0)} = -0.4$, $h(z) = \text{logit}(z)$), and consider 100 experiments: the n-th experiment has $S = 10n$ and $C = 3n$, for $n = 1, 2, \ldots, 100$. For $n \to \infty$, this is the asymptotic regime considered in Besbes and Zeevi (2009) and Wang et al. (2014). Note that $C/S = 0.3$ for all n; we thus exclude the case where C/S gets close to zero or to one. For each experiment we run 1000 simulations, and record the price dispersion induced by the optimal policy after a single selling season, i.e. $\lambda_{\min}(P_S)$, when the prices of the optimal policy are used.

Figure 5 shows the simulation average of $\lambda_{\min}(P_S)$ as function of n (on the left), and as function of $\log(n)$ (on the right). It suggests that the amount of price dispersion, induced by the optimal pricing policy in a single selling season, grows as $\log(n)$. This slow growth rate explains why, in the asymptotic regime considered by Besbes and Zeevi (2009) and Wang et al. (2014), active price experimentation is necessary, whereas in our setting a myopic policy works well.

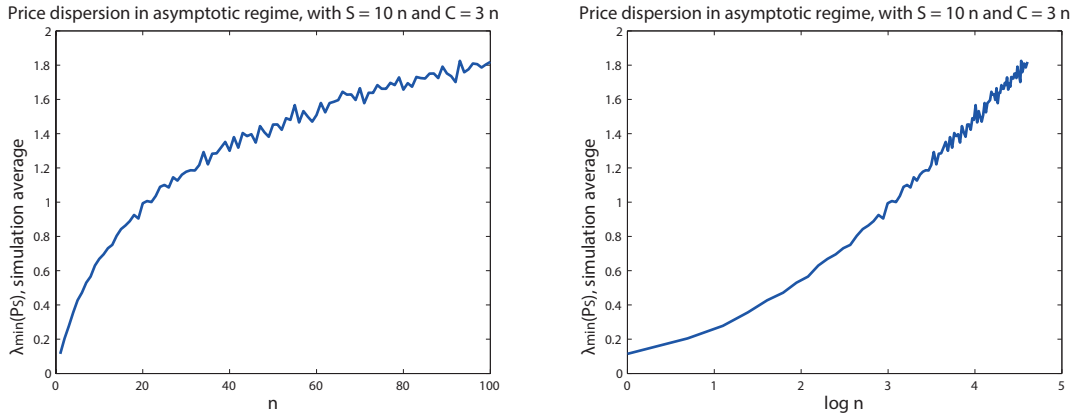

Figure 5 $\lambda_{\min}(P_S)$, for $S = 10n$, $C = 3n$

5. Extensions

5.1. Overlapping Selling Seasons with Varying C and S

To avoid heavy notation that obscures the message of this paper, we assume in the preceding sections that selling seasons are non-overlapping and have the same initial inventory and duration. In the application that motivates our study, dynamic pricing in the hotel industry, this might be too restrictive. In this section we consider the situation where different selling seasons may overlap, and may have different initial inventory and duration. We show that an adaptation of $\Phi(\epsilon)$, the pricing strategy defined in Section 3.2, has $\text{Regret}(T) = O(\log^2(T))$.

5.1.1. Setting. Let $C_j \in \mathbb{N}$ denote the initial inventory, $S_j \in \mathbb{N}$ the duration, and $t_j \in \mathbb{N}$ the first time period of the j-th selling season, for $j \in \mathbb{N}$. W.l.o.g. we assume that $t_1 = 1$ and $t_j \leq t_{j+1}$ for all $j \in \mathbb{N}$. Each selling season j corresponds to a product j that is sold, with corresponding prices and demand realizations. Let $c_{j,s}$ denote the inventory level, $p_{j,s}$ the selling price, and $d_{j,s}$ the demand of product j in stage s of its selling season, for $j \in \mathbb{N}$ and $s \in \{1, \ldots, S_j\}$. The dynamics of $c_{j,s}$ and $d_{j,s}$ are similar as in the "base case" discussed in Section 2: $d_{j,s}$ is Bernoulli distributed with mean $h(\beta_0 + \beta_1 p_{j,s})$ if $c_{j,s} > 0$, and $d_{j,s} = 0$ if $c_{j,s} = 0$. In addition, $c_{j,1} = C_j$ and $c_{j,s+1} = c_{j,s} - d_{j,s}$, for all $j \in \mathbb{N}$ and $s = 1, \ldots, S_j$. Prices $p_{j,s}$ lie in $[p_l, p_h]$ and are non-anticipating, i.e. they may depend on the history of prices and demand realizations $H_{j,s} = \{(d_{j',s'}, p_{j',s'}) \mid j' \in \mathbb{N}, 1 \leq s' \leq S_{j'}, t_{j'} - 1 + s' < t_j - 1 + s\}$, but not on future ones. A pricing strategy is a collection of functions $\psi = (\psi_{j,s})_{j \in \mathbb{N}, 1 \leq s \leq S_j}$, such that each $\psi_{j,s}$ generates for each possible history $H_{j,s}$ a price $p_{j,s} \in [p_l, p_h]$.

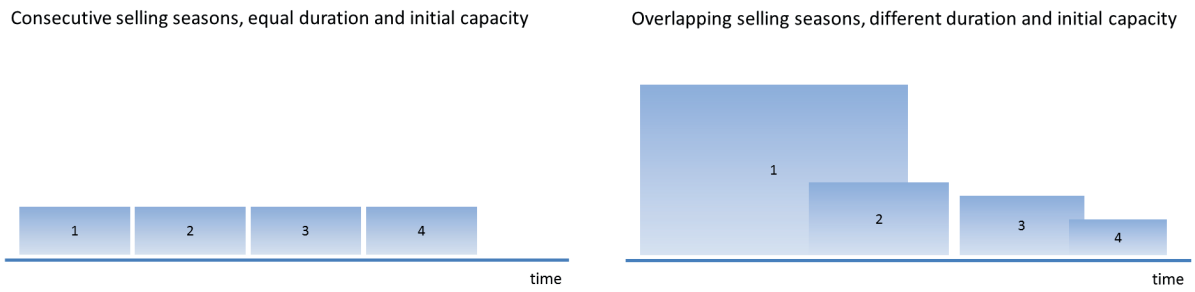

Figure 6 Fixed or varying initial capacity and duration

Let $V_{\beta^{(0)},S_j}(C_j,1)$ be the optimal reward in a selling season and let $\pi^*_{\beta^{(0)},C_j,S_j}$ be the optimal policy with $C=C_j$ and $S=S_j$, as defined in Section 2.3. Let $p_{j,s}$ be prices generated by pricing strategy ψ. The regret of ψ after T time periods is defined as

$$\text{Regret}(\psi,T)=\sum_{j:t_j-1+S_j<T}\Big\{V_{\beta^{(0)},S_j}(C_j,1)-\sum_{s=1}^{S_j}E[p_{j,s}d_{j,s}]\Big\},$$

i.e. as the cumulative regret over all selling seasons *completed* before period T.

5.1.2. Pricing strategy and regret bound. Theorem 2 shows in the case of non-overlapping selling seasons with $C_j=C$, $S_j=S$, and $t_j=1+(j-1)S$, for all j, that $\text{Regret}(\Phi(\epsilon),T)=O(\log^2(T))$. We now show that a modification $\Phi'(\epsilon)$ of this strategy has $\text{Regret}(T)=O(\log^2(T))$ in the more general setting described above.

Similarly as in the setting considered in Theorem 2, estimation of $\beta^{(0)}$ for selling season j is based on all sales data preceding t_j, and of all sales data generated by the product corresponding to selling season j. (Thus, for estimation we neglect sales data generated in selling seasons overlapping with season j; this is for technical reasons, and enables us to derive Theorem 4 analogously to the proof of Theorem 2). In particular, we define the set

$$\mathcal{J}_{j,s}=\{(j',s')\in\mathbb{N}^2\mid 1\le s'\le S_j, t_{j'}-1+s'\le t_j-1\}\cup\{(j,s')\in\mathbb{N}^2\mid 1\le s'<s\},$$

the likelihood function

$$L_{j,s}(\beta)=\sum_{(j',s')\in\mathcal{J}_{j,s}}\log\Big[h(\beta_0+\beta_1p_{j',s'})^{d_{j',s'}}(1-h(\beta_0+\beta_1p_{j',s'}))^{1-d_{j',s'}}\Big]\mathbf{1}_{c_{j',s'}>0},$$

and the score function

$$l_{j,s}(\beta)=\sum_{(j',s')\in\mathcal{J}_{j,s}}\frac{\dot h(\beta_0+\beta_1p_{j',s'})}{h(\beta_0+\beta_1p_{j',s'})(1-h(\beta_0+\beta_1p_{j',s'}))}\begin{pmatrix}1\\ p_{j',s'}\end{pmatrix}(d_{j',s'}-h(\beta_0+\beta_1p_{j',s'}))\mathbf{1}_{c_{j',s'}>0},$$

and define $\hat{\beta}_{j,s}$ as in Section 2.5.

We now formally define the pricing strategy $\Phi'(\epsilon)$. For notational convenience, write $\mathcal{I}(t) = \{(j,s) \in \mathbb{N}^2 \mid 1 \le s \le S_j, t_j - 1 + s = t\}$ for $t \in \mathbb{N}$.

Pricing strategy $\Phi'(\epsilon)$

Initialization: Choose $0 < \epsilon < (p_h - p_l)/4$, and initial prices p_1, $p_2 \in [p_l, p_h]$, with $p_1 \neq p_2$.

Set $p_{j,s} = p_1$ for all $(j,s) \in \mathcal{I}(1)$, and $p_{j,s} = p_2$ for all $(j,s) \in \mathcal{I}_2$.

For all $t \ge 2$, and for all $(j,s) \in \mathcal{I}(t+1)$:

Estimation: Determine $\hat{\beta}_{j,s}$.

Pricing:

If $c_{j,s} = 0$, set $p_{j,s} \in [p_l, p_h]$ arbitrarily.

If $c_{j,s} > 0$, let $p_{\text{ceqp}} = \pi^*_{\hat{\beta}_{j,s}, C_j, S_j}(c_{j,s}, s)$, and consider the following two cases:

I) If

(a) $|p_{j,s_1} - p_{j,s_2}| < \epsilon$ for all $1 \le s_1, s_2 < s$, and

(b) $|p_{j,s_1} - p_{\text{ceqp}}| < \epsilon$ for all $1 \le s_1 < s$, and

(c) $c_{j,s} = 1$ or $s = S_j$,

then choose $p_{j,s} \in \big(\{p_{\text{ceqp}} + 2\epsilon, p_{\text{ceqp}} - 2\epsilon\} \cap [p_l, p_h]\big)$.

II) Else, set $p_{j,s} = p_{\text{ceqp}}$.

The following theorem shows, under some conditions on t_j, C_j and S_j, that $\Phi'(\epsilon)$ has the same $O(\log^2(T))$ bound on the regret as $\Phi(\epsilon)$.

THEOREM 4. *Suppose* $\sup_{j\in\mathbb{N}} C_j < \infty$, $\sup_{j\in\mathbb{N}} S_j < \infty$, *and* $1 < C_j < S_j$ *for all* $j \in \mathbb{N}$. *In addition, suppose* $\sup_{t\in\mathbb{N}} |\{j \in \mathbb{N} : t_j = t\}| < \infty$. *Let* v_0 *as in Theorem 1, and* $0 < \epsilon < v_0/2$. *Then*

$$\textit{Regret}(\Phi'(\epsilon), T) = O(\log^2(T)).$$

The proof is sketched in the Appendix, Section EC.1. The conditions on C_j, S_j and t_j ensure that the initial inventories, the durations of the selling seasons, and the number of selling seasons starting in any time period, are all bounded.

5.2. Non-stationary Demand

Throughout the paper we assume that the market is stationary: the parameters $\beta^{(0)}$ do not change over time. In this section we explore what happens if this assumption is violated. We distinguish between two types of non-stationarity: (i) a "booking curve", meaning that demand depends on the stage s in the selling season, and (ii) a more general setting where $\beta^{(0)} = (\beta^{(0)}(t))_{t\in\mathbb{N}}$ is varying over time.

5.2.1. Booking curve. A booking curve can be handled by explicitly modeling the dependence of demand on the stage s. For example, one could assume that the demand in a period with positive inventory is Bernoulli distributed with mean $h(\beta_0^{(0)} + \beta_1^{(0)} p + \beta_2^{(0)} s)$, where p denotes the price, s the stage in the selling season, and $(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})$ are unknown parameters. Similarly as in Section 2.3 one can then define the optimal full-information solution $\pi^*_\beta(c,s)$, with $h(\beta_0 + \beta_1 p)$ in all relevant equations replaced by $h(\beta_0 + \beta_1 p + \beta_2 s)$. The design matrix (8) is then equal to

$$P_t = \sum_{i=1}^{t} \begin{pmatrix} 1 \\ p_i \\ s_i \end{pmatrix} (1, p_i, s_i) \mathbf{1}_{c_i > 0}.$$

To prove an endogenous-learning property similar to Theorem 1, one should show that for all β close to $\beta^{(0)}$, using the policy π^*_β in selling season k implies $\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S}) > \epsilon$, for all $k \in \mathbb{N}$ and some $\epsilon > 0$ independent of k and β. This means that the amount of price dispersion, measured by the smallest eigenvalue of the design matrix, strictly increases in each selling season, and that the maximum likelihood estimate of β converges a.s. to the true value. This guarantees that the prices generated by a (near-)myopic pricing strategy similar to $\Phi(\epsilon)$ converge to the true optimal prices.

In this particular model, with mean demand equal to $h(\beta_0^{(0)} + \beta_1^{(0)} p + \beta_2^{(0)} s)$, a sufficient condition for the endogenous-learning property to hold is that there are prices p_1, p_2, p_3 used in stage s_1, s_2, s_3, respectively, such that the vectors $\{(1, p_i, s_i)^T \mid i = 1, 2, 3\}$ are linearly independent, and such that $c_{s_1} c_{s_2} c_{s_3} > 0$. This implies

$$\begin{aligned} &\lambda_{\min}(P_{Sk}) - \lambda_{\min}(P_{S(k-1)}) \geq \lambda_{\min}\Bigg(\sum_{i=1}^{3} (1, p_i, s_i)^T (1, p_i, s_i) \mathbf{1}_{c_i > 0} \Bigg) \\ &\geq \frac{\det\big(\sum_{i=1}^{3} (1, p_i, s_i)^T (1, p_i, s_i) \mathbf{1}_{c_i > 0} \big)}{\mathrm{tr}\big(\sum_{i=1}^{3} (1, p_i, s_i)^T (1, p_i, s_i) \mathbf{1}_{c_i > 0} \big)^2} \geq \frac{(p_3(s_2 - s_1) + p_2(s_3 - s_1) + p_1(s_3 - s_2))^2}{(3 + 3S^2 + 3 \sup_{p \in [p_l, p_h]} p^2)^2} > 0, \end{aligned}$$

which implies the endogenous-learning property. In a similar way as Theorem 2 an upper bound on the regret of a (near-)myopic policy can then be obtained.

5.2.2. Time-varying parameters. A natural approach to estimate $\beta^{(0)}(t)$, for a particular $t \in \mathbb{N}$, is to use maximum-likelihood estimation based on sales data from the time periods $\{t - N, \ldots, t\}$, for some $N \in \mathbb{N}$. This approach is taken in a recent paper by Keskin and Zeevi (2013) in a dynamic pricing problem, and by Besbes et al. (2014) in a more general stochastic optimization setting. Both these papers show that the growth rate of the regret of pricing policies and the "optimal" choice of N depend on some measure of the volatility of the process $(\beta^{(0)}(t))_{t \in \mathbb{N}}$.

This approach can in principle also be taken in our setting of dynamic pricing with finite inventories, but there are several technical difficulties to overcome. For example, if estimation is based

on a sample of size N, the finiteness of T_ρ in Proposition 1 is not guaranteed, and it is not clear how one could obtain bounds on the mean square estimation error (obtaining such bounds are necessary to analyze the performance of pricing strategies). Overcoming these technical difficulties is an important and technically challenging direction for future research.

5.3. Endogenous Learning in other Decision Problems

The endogenous-learning property shown in Theorem 1 is the key result that leads to consistency of the myopic policy and to a regret that grows only $O(\log^2(T))$. This property seems not unique for the pricing problem under consideration, but may be satisfied by many other decision problems as well. We here briefly outline some types of problems for which this may be the case.

Consider a collection of discrete-time Markov decision problems (MDPs)

$$\{(X, \mathcal{A}, p(\cdot,\cdot,\cdot,\theta), r(\cdot,\cdot,\theta)) \mid \theta \in \Theta\},$$

parameterized by a finite-dimensional parameter θ contained in some set $\Theta \subset \mathbb{R}^d$. For each $\theta \in \Theta$, $(\mathcal{X}, \mathcal{A}, p(\cdot,\cdot,\cdot,\theta), r(\cdot,\cdot,\theta))$ corresponds to an MDP with statespace $\mathcal{X}$, action space $\mathcal{A}$, transition probabilities of going from state x to x' when action a is used denoted by $p(x,x',a,\theta)$, and the expected reward of using action a in state x denoted by $r(x,a,\theta)$, for $x, x' \in \mathcal{X}$ and $a \in \mathcal{A}$. (see Puterman (1994) for an introduction to MDPs). The goal of the decision maker may be to optimize the average reward or discounted reward, over a finite or infinite time horizon, without knowing the value of θ.

Suppose that each time that an action a is selected in state x, a realization y_i of a random variable Y is observed, the distribution of which depends on x, a, and θ. With an appropriate statistical model of Y, the value of the unknown θ may at each decision moment be inferred from the previously observed realizations, chosen actions, and visited stated, using an appropriate statistical technique (maximum likelihood estimation, (non)-linear regression, Bayesian methods). If $\hat{\theta}$ denotes the estimated value of θ, then a myopic policy is to always select the action that is optimal if $\hat{\theta}$ equals the true but unknown θ.

Strong consistency of an estimator (a.s. convergence of $\hat{\theta}$ to θ as the number of observations increases) typically presumes a minimum amount of variation/dispersion in the controls; see e.g. Skouras (2000) and Pronzato (2009) for nonlinear regression models, Chen et al. (1999) for generalized linear models, the classic Lai and Wei (1982) for linear regression models, and Hu (1996, 1998) for Bayesian regression models. The decision problems described above satisfy an endogenous-learning property if the myopic policy induces an amount of dispersion in the controls that guarantees strong consistency of the estimator. As a result, no active experimentation is then necessary to

eventually learn the unknown θ; learning "takes care of itself" by just simply using myopic actions. This contrasts with many other decision problems under uncertainty where deviating from the myopic policy is necessary to eventually learn the unknown parameters of the system (e.g. in various multi-armed bandit problems).

Acknowledgments

We thank Sandjai Bhulai for useful discussions and providing literature references. We also thank the referees, associate editor, and area editor; their constructive comments and suggestions have greatly improved the paper. Part of this research was done while the first author was affiliated Centrum Wiskunde & Informatica (CWI) Amsterdam, Eindhoven University of Technology, and University of Amsterdam. The research of Arnoud den Boer is supported by an NWO VENI grant. The research of Bert Zwart is partly supported by an NWO VIDI grant and an IBM faculty award.

References

Altman, E., A. Shwartz. 1991. Adaptive control of constrained Markov chains: criteria and policies. *Annals of Operations Research* **28**(1) 101–134.

Anderson, T. W., J. B. Taylor. 1976. Some experimental results on the statistical properties of least squares estimates in control problems. *Econometrica* **44**(6) 1289–1302.

Besbes, O., Y. Gur, A. Zeevi. 2014. Non-stationary stochastic optimization. Working paper, Columbia Business School, `http://ssrn.com/abstract=2296012`.

Besbes, O., A. Zeevi. 2009. Dynamic pricing without knowing the demand function: risk bounds and near-optimal algorithms. *Operations Research* **57**(6) 1407–1420.

Besbes, O., A. Zeevi. 2011. On the minimax complexity of pricing in a changing environment. *Operations Research* **59**(1) 66–79.

Besbes, O., A. Zeevi. 2012. Blind network revenue management. *Operations Research* **60**(6) 1537–1550.

Bhatia, R. 1997. *Matrix Analysis*. Springer Verlag, New York.

Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5**(3) 203–230.

Broder, J., P. Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Operations Research* **60**(4) 965–980.

Burnetas, A. N., M. N. Katehakis. 1997. Optimal adaptive policies for Markov decision processes. *Mathematics of Operations Research* **22**(1) 222–255.

Chang, H. S., M. C. Fu, J. Hu, S. I. Marcus. 2005. An adaptive sampling algorithm for solving Markov decision processes. *Operations Research* **53**(1) 126–139.

Chen, K., I. Hu, Z. Ying. 1999. Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *The Annals of Statistics* **27**(4) 1155–1163.

den Boer, A. V. 2013a. Does adding data always improve linear regression estimates? *Statistics & Probability Letters* **83**(3) 829–835.

den Boer, A. V. 2013b. Dynamic pricing and learning: historical origins, current research, and new directions. Working paper. Available at `http://ssrn.com/abstract=2334429`.

den Boer, A. V. 2014. Dynamic pricing with multiple products and partially specified demand distribution. *Mathematics of Operations Research* **39**(3) 863–888.

den Boer, A. V., B. Zwart. 2014a. Mean square convergence rates for maximum quasi-likelihood estimators. *Stochastic Systems* **4** 1–29.

den Boer, A. V., B. Zwart. 2014b. Simultaneously learning and optimizing using controlled variance pricing. *Management Science* **60**(3) 770–783.

Duistermaat, J. J., J. A. C. Kolk. 2004. *Multidimensional Real Analysis: Differentiation*. Series: Cambridge Studies in Advanced Mathematics (No. 86), Cambridge University Press, Cambridge.

Fahrmeir, L., H. Kaufmann. 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* **13**(1) 342–368.

Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science* **40**(8) 999–1020.

Gill, R. D., B. Y. Levit. 1995. Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli* **1**(1/2) 59–79.

Gordienko, E. I., J. A. Minjárez-Sosa. 1998. Adaptive control for discrete-time Markov processes with unbounded costs: average criterion. *Mathematical Methods of Operations Research* **48**(1) 37–55.

Harrison, J. M., N. B. Keskin, A. Zeevi. 2012. Bayesian dynamic pricing policies: learning and earning under a binary prior distribution. *Management Science* **58**(3) 570–586.

Hernández-Lerma, O. 1989. *Adaptive Markov control processes*. Springer-Verlag, New York.

Hernández-Lerma, O., R. Cavazos-Cadena. 1990. Density estimation and adaptive control of Markov processes: average and discounted criteria. *Acta Applicandae Mathematicae* **20**(3) 285–307.

Hu, I. 1996. Strong consistency of Bayes estimates in stochastic regression models. *Journal of Multivariate Analysis* **57**(2) 215–227.

Hu, I. 1998. Strong consistency of Bayes estimates in nonlinear stochastic regression models. *Journal of Statistical Planning and Inference* **67**(1) 155–163.

Huh, W. T., P. Rusmevichientong. 2014. Online sequential optimization with biased gradients: theory and applications to censored demand. *INFORMS Journal on Computing* **26**(1) 150–159.

Keskin, N. B., A. Zeevi. 2013. Chasing demand: Learning and earning in a changing environment. Working paper, University of Chicago, Booth School of Business. Available at `http://ssrn.com/abstract=2389750`.

Keskin, N. B., A. Zeevi. 2014. Dynamic pricing with an unknown linear demand model: asymptotically optimal semi-myopic policies. *Operations Research* **62**(5) 1142–1167.

Kleinberg, R., T. Leighton. 2003. The value of knowing a demand curve: bounds on regret for online posted-price auctions. *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*. 594–605.

Kumar, P. R. 1985. A survey of some results in stochastic adaptive control. *SIAM Journal on Control and Optimization* **23**(3) 329–380.

Kumar, P. R., P. Varaiya. 1986. *Stochastic systems: estimation, identification and adaptive control*. Prentice Hall, New Jersey.

Kunnumkal, S., H. Topaloglu. 2008. Using stochastic approximation methods to compute optimal base-stock levels in inventory control problems. *Operations Research* **56**(3) 646–664.

Lai, T. L., H. Robbins. 1982. Iterated least squares in multiperiod control. *Advances in Applied Mathematics* **3**(1) 50–73.

Lai, T. L., C. Z. Wei. 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics* **10**(1) 154–166.

Lovejoy, W. S. 1991. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research* **28**(1) 47–65.

Monahan, G. E. 1982. A survey of partially observable Markov decision processes: theory, models, and algorithms. *Management Science* **28**(1) 1–16.

Nassiri-Toussi, K., W. Ren. 1994. On the convergence of least squares estimates in white noise. *IEEE Transactions on Automatic Control* **39**(2) 364–368.

Pronzato, L. 2009. Asymptotic properties of nonlinear estimates in stochastic models with finite design space. *Statistics & Probability Letters* **79**(21) 2307–2313.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. 1st ed. Wiley, New York.

Sauré, D., A. Zeevi. 2013. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* **15**(3) 387–404.

Skouras, K. 2000. Strong consistency in nonlinear stochastic regression models. *The Annals of Statistics* **28**(3) 871–879.

Talluri, K. T., G. J. van Ryzin. 2004. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston.

Wang, Z., S. Deng, Y. Ye. 2014. Close the gaps: a learning-while-doing algorithm for a class of single-product revenue management problems. *Operations Research* **62**(2) 318–331.

Weatherford, L. R., S. E. Kimes. 2003. A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting* **19**(3) 401–415.

Proofs

This e-companion contains the mathematical proofs of the results in the paper. Section EC.1 contains the proof of Theorems 1, 2, 3 and 4. The proofs frequently refer to a number of auxiliary lemmas, which are formulated and proven in Section EC.2.

EC.1. Proofs of Main Theorems

Proof of Theorem 1

Let $\beta \in B$. Consider the k-th selling season, and write $c(1) = c_{1+(k-1)S}$, $c(2) = c_{2+(k-1)S}$, ..., $c(S) = c_{kS}$. The proof consists of two steps. In Step 1, we show that there is a $v_1(\beta) > 0$ such that if prices $\pi^*_{\beta^{(0)}}(c(s), s)$ are used in state $(c(s), s)$, for all $s = 1, \ldots, S$, then there are $1 \le s, s' \le S$ with $|\pi^*_\beta(c(s), s) - \pi^*_\beta(c(s'), s')| > v_1(\beta)$ and $c(s)c(s') > 0$. Since π^*_β is continuous in β (Lemma EC.2(vi)), there is an open neighborhood $\mathcal{U}_\beta \subset B$ around β such that, if price $\pi^*_{\beta^{(s)}}(c(s), s)$ is used in state $(c(s), s)$, for all $s = 1, \ldots, S$ and some sequence $(\beta(1), \ldots, \beta(S)) \in \mathcal{U}_\beta \cap B$, then there are $1 \le s, s' \le S$ such that $|\pi^*_{\beta^{(s)}}(c(s), s) - \pi^*_{\beta^{(s')}}(c(s'), s')| > v_1(\beta)/2$, $c(s) > 0$ and $c(s') > 0$. In Step 2 we show that $v_1(\beta)$ can be bounded from below by a constant $v_0 > 0$ independent of β. This proves (10). Equation (11) follows by application of Lemma EC.3.

Step 1: Occurence of price change. Define

$$\triangleleft = \{(c, s) \mid S + 1 - C \le s \le S, S + 1 - s \le c \le C\}. \tag{EC.1}$$

See Figure EC.1 for an illustration of $\triangleleft$ in the state space $\mathcal{X}$. Notice that since $(C, 1) \notin \triangleleft$ (by the assumption $C < S$), the path $(c(s), s)_{1 \le s \le S}$ may or may not hit $\triangleleft$. We show that, in both cases, at least two different selling prices occur on the path $(c(s), s)_{1 \le s \le S}$. Let $p^*_{a,\beta}$ be as in Lemma EC.1 and shorthand write $f^*_{a,\beta} = f_{a,\beta}(p^*_{a,\beta})$.

Case 1. The path $(c(s), s)_{1 \le s \le S}$ hits $\triangleleft$. Then there is an s such that $(c(s), s) \in \triangleleft$ and $(c(s), s-1) \in (L\triangleleft)$, where we define

$$(L\triangleleft) = \{(1, S-1), (2, S-2), \ldots, (C-1, S-C+1), (C, S-C)\}$$

as the set of points immediately left to $\triangleleft$ in Figure EC.1. The following two properties imply that a price change occurs when the path $(c(s), s)_{1 \le s \le S}$ hits $\triangleleft$.

(P.1) If $(c, s) \in \triangleleft$ then $\pi^*_\beta(c, s) = p^*_{0,\beta}$, $\Delta V_\beta(c, s+1) = 0$, and $V_\beta(c, s) = (S - s + 1) \cdot f^*_{0,\beta}$.

(P.2) If $(c, s) \in (L\triangleleft)$, then $\Delta V_\beta(c, s+1) \ge h(\beta_0 + \beta_1 p_h)^{c-1} f^*_{0,\beta}$ and $\pi^*_\beta(c, s) \ge p^*_{h(\beta_0+\beta_1 p_h)^{c-1} f^*_{0,\beta}, \beta} > p^*_{0,\beta}$.

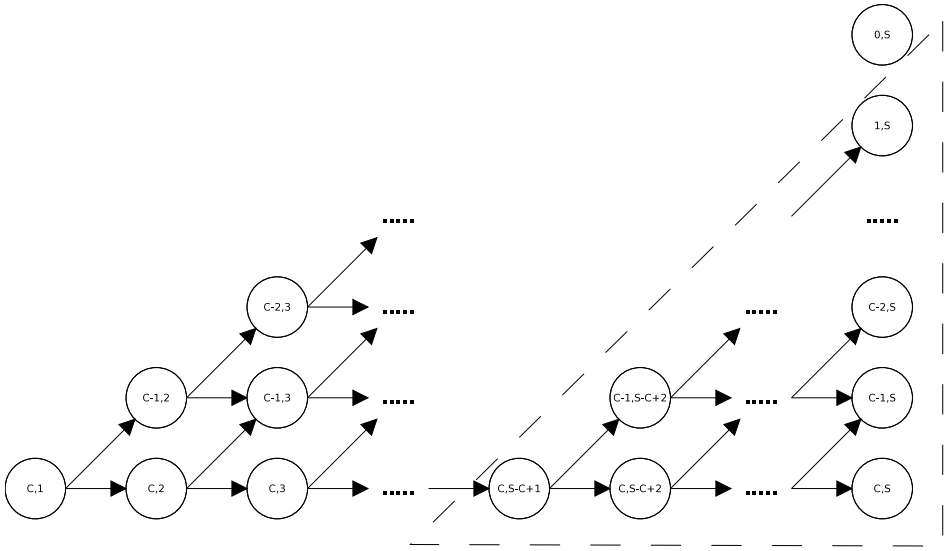

Figure EC.1 Schematic picture of $\triangleleft$

Proof of (P.1): Backward induction on s. If $s = S$ and $(c,s) \in \triangleleft$, then the assertions follow immediately. Let $s < S$. Then $\Delta V_\beta(c,s+1) = V_\beta(c,s+1) - V_\beta(c-1,s+1) = 0$, $\pi^*_\beta(c,s) = p^*_{0,\beta}$ and $V_\beta(c,s) = \max_{p\in[p_l,p_h]} ph(\beta_0+\beta_1 p) + V_\beta(c,s+1) = (S-s+1)\cdot V_\beta(1,S)$, by (3) and the induction hypothesis. This proves (P.1).

Proof of (P.2). Induction on c. If $c=1$ and $(c,s) \in (L\triangleleft)$, then $(c,s) = (1,S-1)$, $\Delta V_\beta(1,S) = f^*_{0,\beta}$ and $\pi^*_\beta(1,S-1) = p^*_{\Delta V_\beta(1,S)} = p^*_{f^*_{0,\beta}} > p^*_{0,\beta}$, since by Lemma EC.1(ii) $p^*_{a,\beta}$ is strictly increasing in a. If $c>1$ and $(c,s) \in (L\triangleleft)$ then $(c,s) = (c,S-c)$, and the induction hypothesis, together with the optimality of $\pi^*_\beta(c,S-c+1)$, implies

$$
\begin{aligned}
&\Delta V_\beta(c,S-c+1) = V_\beta(c,S-c+1) - V_\beta(c-1,S-c+1)\\
=&(\pi^*_\beta(c,S-c+1) - \Delta V_\beta(c,S-c+2))h(\beta_0+\beta_1\pi^*_\beta(c,S-c+1)) + V_\beta(c,S-c+2)\\
-&(\pi^*_\beta(c-1,S-c+1) - \Delta V_\beta(c-1,S-c+2))h(\beta_0+\beta_1\pi^*_\beta(c-1,S-c+1))\\
-&V_\beta(c-1,S-c+2)\\
\geq&\Delta V_\beta(c-1,S-c+2)h(\beta_0+\beta_1\pi^*_\beta(c-1,S-c+1))\\
\geq&h(\beta_0+\beta_1 p_h)^{c-1} f^*_{0,\beta},
\end{aligned}
$$

and $\pi^*_\beta(c,S-c) = p^*_{\Delta V_\beta(c,S-c+1),\beta} \geq p^*_{\Delta V_\beta(c,S-c+1),\beta} > p^*_{0,\beta}$, using again Lemma EC.1(ii). This proves (P.2), and concludes Case 1.

Case 2. The path $(c(s),s)_{1\leq s\leq S}$ does not hit $\triangleleft$. Then there is an $1 \leq s \leq S-2$ such that $c(s)=2$ and $c(s+1)=1$. We show by backward induction that

$$\Delta V_\beta(2,s) - \Delta V_\beta(1,s+1) \leq (V_\beta(1,1) - p_h)\cdot h(\beta_0+\beta_1 p_h), \tag{EC.2}$$

for all $2 \le s \le S-1$, and that $V_\beta(1,1) < p_h$. This implies $\Delta V_\beta(2,s+1) < \Delta V_\beta(1,s+2)$. By Lemma EC.1(ii) this implies

$$\pi^*_\beta(2,s) = p^*_{\Delta V_\beta(2,s+1),\beta} < p^*_{\Delta V_\beta(1,s+2),\beta} = \pi^*_\beta(1,s+1), \tag{EC.3}$$

and thus a price change occurs.

Let $2 \le s \le S-1$. The optimality of $\pi^*_\beta(1,s)$ and $\pi^*_\beta(1,s+1)$ implies

$$\begin{aligned}
&\Delta V_\beta(2,s) - \Delta V_\beta(1,s+1)\\
\le&(\pi^*_\beta(2,s) - \Delta V_\beta(2,s+1))h(\beta_0+\beta_1\pi^*_\beta(2,s)) + V_\beta(2,s+1)\\
&-(\pi^*_\beta(1,s) - \Delta V_\beta(1,s+1))h(\beta_0+\beta_1\pi^*_\beta(1,s)) - V_\beta(1,s+1)\\
&-(\pi^*_\beta(2,s) - \Delta V_\beta(1,s+2))h(\beta_0+\beta_1\pi^*_\beta(2,s)) - V_\beta(1,s+2)\\
=&\Big[\Delta V_\beta(2,s+1) - \Delta V_\beta(1,s+2)\Big]\Big[1 - h(\beta_0+\beta_1(\pi^*_\beta(2,s))\Big]\\
&-(\pi^*_\beta(1,s) - \Delta V_\beta(1,s+1))h(\beta_0+\beta_1\pi^*_\beta(1,s))\\
\le&-(p_h - V_\beta(1,s+1))h(\beta_0+\beta_1 p_h)\\
\le&-(p_h - V_\beta(1,1))h(\beta_0+\beta_1 p_h).
\end{aligned}$$

Here $\Big[\Delta V_\beta(2,s+1) - \Delta V_\beta(1,s+2)\Big] \le 0$ follows from the induction hypothesis if $s < S-1$, and follows from $V_\beta(2,S) - V_\beta(1,S) = 0$ if $s = S-1$. The last inequality follows from Lemma EC.1(ii). This proves (EC.2), and concludes Case 2.

We have shown that, on any path $(c(s),s)_{1\le s\le S}$ in $\mathcal{X}$ starting at $(C,1)$, the policy π^*_β induces a price-change. It follows that there exists a $v_1(\beta) > 0$ such that for all paths $(c(s),s)_{1\le s\le S}$,

$$|\pi^*_\beta(c(s),s) - \pi^*_\beta(c(s'),s')| \ge v_1(\beta).$$

Step 2: Lower bound on magnitude of price change. Property (P.2) in the proof of Theorem 1 shows that a price change of magnitude

$$|\pi^*_\beta(c(s),s) - \pi^*_\beta(c(s+1),s+1)| \ge |p^*_{h(\beta_0+\beta_1 p_h)^{C-1} f^*_{0,\beta},\beta} - p^*_{0,\beta}|$$

occurs in Case 1, and equation (EC.2) shows that a price change of magnitude

$$|\pi^*_\beta(c(s),s) - \pi^*_\beta(c(s+1),s+1)| \ge |p^*_{\Delta V_\beta(2,s+1),\beta} - p^*_{\Delta V_\beta(1,s+2),\beta}|$$

occurs, with $|\Delta V_\beta(2,s+1) - \Delta V_\beta(1,s+2)| < (V_\beta(1,1) - p_h)h(\beta+\beta_1 p_h) < 0$ (since $V_\beta(1,1) - p_h = \max_{p\in[p_l,p_h]} ph(\beta_0+\beta_1 p) - p_h < \max_{p\in[p_l,p_h]}(p-p_h) = 0$). The proof of Lemma EC.1(ii) implies that, for all $0 \le a_0, a_1 < p_h$ and all $\beta \in B$,

$$|p^*_{a_0,\beta} - p^*_{a_1,\beta}| \ge |a_0 - a_1| \cdot \inf_{0\le a\le p_h} \frac{-\beta_1 \dot{h}(\beta_0+\beta_1 p^*_{a,\beta})}{|\ddot{f}_{a,\beta}(p^*_{a,\beta})|} > 0,$$

by application of the Mean Value Theorem. If we define

$$\mathcal{C} = \inf_{\beta \in B} \inf_{0 \le a \le p_h} \frac{-\beta_1 \dot{h}(\beta_0 + \beta_1 p^*_{a,\beta})}{|\ddot{f}_{a,\beta}(p^*_{a,\beta})|} > 0$$

then the magnitude of the price change $v_1(\beta)$ on the path $(c(s), s)_{1 \le s \le S}$, is bounded from below by

$$v_1(\beta) \ge v_0 := \mathcal{C} \cdot \min\Big\{ \inf_{\beta \in B} h(\beta_0 + \beta_1 p_h)^{C-1} f^*_{0,\beta}, \quad \inf_{\beta \in B} |(V_\beta(1,1) - p_h) h(\beta + \beta_1 p_h)| \Big\}.$$

Proof of Theorem 2

Consider the k-th selling season, for some arbitrary fixed integer $k \ge 2$. The prices generated by $\Phi(\epsilon)$ are based on the estimates $\hat{\beta}_t$, which are determined by the historical prices and demand realizations. Now, different demand realizations can lead to the same state (c, s) of the MDP. For example, a sale in the first period of a selling season and no sale in the second period leads to state $(C-2, 3)$, but this state is also reached if there is no sale in the first period and a sale in the second period of the selling season. These two "routes" may lead to different estimates $\hat{\beta}_t$, and to different pricing decisions in state $(C-2, 3)$. Thus, with $\Phi(\epsilon)$, the prices in the k-th selling season are not determined by a stationary policy for the Markov decision problem described in Section 2.3.

To be able to compare the optimal revenue in a selling season with that obtained by $\Phi(\epsilon)$, we define a new Markov decision problem, in which the states are sequences of demand realizations in the selling season. Conditionally on all prices and demand realizations from before the start of the selling season, $\Phi(\epsilon)$ is then a stationary deterministic policy for this new MDP: each state is associated with a unique price prescribed by $\Phi(\epsilon)$. This enables us to calculate bounds on the regret obtained in a single selling season.

We define this new MDP for any $\beta \in B$. The state space $\tilde{\mathcal{X}}$ consists of all sequences of possible demand realizations in the selling season:

$$\tilde{\mathcal{X}} = \{(x_1, \ldots, x_s) \in \{0,1\}^s \mid 0 \le s \le S, \sum_{i=1}^{s} x_i \le C\},$$

where we denote the empty sequence by $(\emptyset)$. The action space is $[p_l, p_h]$. Using action p in state $(x_1, \ldots, x_s)$, with $0 \le s < S$ induces a state transition from $(x_1, \ldots x_s)$ to $(x_1, \ldots, x_s, 1)$ with probability $h(\beta_0 + \beta_1 p)\mathbf{1}_{\sum_{i=1}^{s} x_i < C}$ (corresponding to a sale, and inducing immediate reward p), and from $(x_1, \ldots x_s)$ to $(x_1, \ldots, x_s, 0)$ with probability $1 - h(\beta_0 + \beta_1 p)\mathbf{1}_{\sum_{i=1}^{s} x_i < C}$ (corresponding to no sale, and inducing zero reward). In the terminal state $(x_1, \ldots, x_S)$, no transitions occur, no reward is received, and no actions are taken. Note that the actions in states with zero inventory do not impact the reward or transitions.

It is easily seen that the MDP described in Section 2.3 corresponds to the one described here, except that there states are aggregated: all states $(x_1,\ldots,x_s)$ and $(x'_1,\ldots,x'_{s'})$ with $s=s'$ and $\sum_{i=1}^s x_i=\sum_{i=1}^{s'} x'_i$ are there taken together in the state $(C-\sum_{i=1}^s x_i, s+1)$.

Let $\tilde{\pi}=\{\tilde{\pi}(x)\,|\,x=(x_1,\ldots,x_s)\in\tilde{\mathcal{X}}, 0\le s<S\}$ be a stationary deterministic policy for this MDP with augmented state space (defining an action for all except the terminal states $(x_1,\ldots,x_S)$), and let $\tilde{V}_\beta^{\tilde{\pi}}(x)$ be the corresponding value function, for $\beta\in B$. For non-terminal states $x=(x_1,\ldots,x_s)\in\tilde{\mathcal{X}}$ we write $(x;1)=(x_1,\ldots,x_s,1)$ and $(x;0)=(x_1,\ldots,x_s,0)$. Then, for all non-terminal states $x=(x_1,\ldots,x_s)\in\tilde{\mathcal{X}}$, $s<S$, and all $\beta\in B$, $\tilde{V}_\beta^{\tilde{\pi}}(x)$ satisfies the backward recursion

$$\begin{aligned}\tilde{V}_\beta^{\tilde{\pi}}(x)=&(\tilde{\pi}(x)\mathbf{1}_{\sum_{i=1}^s x_i<C}+\tilde{V}_\beta^{\tilde{\pi}}(x;1))h(\beta_0+\beta_1\tilde{\pi}(x))\mathbf{1}_{\sum_{i=1}^s x_i<C}\\&+\tilde{V}_\beta^{\tilde{\pi}}(x;0)(1-h(\beta_0+\beta_1\tilde{\pi}(x))\mathbf{1}_{\sum_{i=1}^s x_i<C}),\end{aligned}$$

and $\tilde{V}_\beta^{\tilde{\pi}}(x)=0$ for all terminal states x.

Let $\tilde{\pi}^*_\beta$ be the optimal policy corresponding to $\beta\in B$, and write $\tilde{V}_\beta(x)=\tilde{V}_\beta^{\tilde{\pi}^*_\beta}(x)$. Then

$$\tilde{V}_\beta(x)=\max_{p\in[p_l,p_h]}\Big[p-\big(\tilde{V}_\beta(x;0)-\tilde{V}_\beta(x;1)\big)\Big]h(\beta_0+\beta_1 p)+\tilde{V}_\beta(x;0), \tag{EC.4}$$

$$\tilde{\pi}^*_\beta(x)=\operatorname*{arg\,max}_{p\in[p_l,p_h]}\Big[p-\big(\tilde{V}_\beta(x;0)-\tilde{V}_\beta(x;1)\big)\Big]h(\beta_0+\beta_1 p), \tag{EC.5}$$

for all states $(x_1,\ldots,x_s)$ with $\sum_{i=1}^s x_i<C$ and $s<S$, and $\tilde{V}_\beta(x)=0$ for all states $(x_1,\ldots,x_s)$ with $\sum_{i=1}^s x_i=C$ or $s=S$. Analogous to Lemma EC.2, it is not difficult to show that $\tilde{\pi}^*_\beta(x)$ is uniquely defined and lies in (p_l,p_h), for all $\beta\in B$ and all states $(x_1,\ldots,x_s)$ with $\sum_{i=1}^s x_i<C$ and $s<S$. In the notation of Lemma EC.1, this implies $(\tilde{V}_\beta(x;0)-\tilde{V}_\beta(x;1),\beta)\in\mathcal{U}_{AB}$..

Let $\mathcal{U}$ and v_0 be as in Theorem 1, ρ_1 as in Proposition 1, and choose $\rho\in(0,\rho_1)$ such that $\beta\in\mathcal{U}$ whenever $||\beta-\beta^{(0)}||\le\rho$. For all $l\in\mathbb{N}$, if $(l-1)S>T_\rho$ then $\hat{\beta}_t\in\mathcal{U}$ for all $t=1+(l-1)S,\ldots,S(l-1)S$, and Theorem 1 implies $\lambda_{\min}(P_{lS})-\lambda_{\min}(P_{(l-1)S})\ge\frac{1}{8}v_0^2(1+p_h^2)^{-1}\ge\frac{1}{2}\epsilon^2(1+p_h^2)^{-1}$, using $v_0/2\ge\epsilon$. If $(l-1)S\le T_\rho$, then I) of the pricing strategy $\Phi(\epsilon)$ guarantees that there are $1\le s,s'\le S$ such that $|p_{s+(l-1)S}-p_{s'+(l-1)S}|\ge\epsilon$, and Lemma EC.3 then implies $\lambda_{\min}(P_{lS})-\lambda_{\min}(P_{(l-1)S})\ge\frac{1}{2}\epsilon^2(1+p_h^2)^{-1}$. It follows that $\lambda_{\min}(P_{lS})\ge l\cdot\frac{1}{2}\epsilon^2(1+p_h^2)^{-1}$ for all $l\in\mathbb{N}$, and thus for all $t>S$,

$$\lambda_{\min}(P_t)\ge\lambda_{\min}(P_{(SS_t-1)S})\ge(SS_t-1)\cdot\frac{1}{2}\epsilon^2(1+p_h^2)^{-1}\ge t\cdot\frac{1}{4S}\epsilon^2(1+p_h^2)^{-1},$$

using $SS_t-1\ge t\frac{(SS_t-1)}{S\cdot SS_t}\ge\frac{t}{2S}$. (Recall the definition $SS_t=1+\lfloor(t-1)/S\rfloor$). By application of Proposition 1 with $t_0=S$ and $L(t)=t\cdot\frac{1}{4S}\epsilon^2(1+p_h^2)^{-1}$, we have

$$T_\rho<\infty\text{ a.s., }E[T_\rho]<\infty,\text{ and }E[||\hat{\beta}_t-\beta^{(0)}||^2\mathbf{1}_{t>T_\rho}]=O\left(\log(t)/t\right). \tag{EC.6}$$

In addition, $v_0/2 > \epsilon$ implies that I) of the pricing strategy $\Phi(\epsilon)$ does not occur for all t with $(SS_t - 1)S > T_\rho$. In particular, if $(k-1)S > T_\rho$, then

$$p_{1+s+(k-1)S} = \tilde{\pi}^*_{\hat{\beta}_{s+(k-1)S}}(d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S}), \tag{EC.7}$$

for all $1 \le s \le S-1$, and

$$p_{1+(k-1)S} = \tilde{\pi}^*_{\hat{\beta}_{(k-1)S}}(\emptyset). \tag{EC.8}$$

Let $H = (p_1, \ldots, p_{(k-1)S}, d_1, \ldots, d_{(k-1)S})$ denote the history of prices and demand up to and including time period $(k-1)S$. Conditionally on H, and given that $(k-1)S > T_\rho$, the parameter estimates $\hat{\beta}_{s+(k-1)S}$ in (EC.7) and (EC.8) are completely determined by the state $(d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S})$. Thus, for each state $(x_1, \ldots, x_s)$ with $\sum_{i=1}^s x_i < C$ and $s < S$, there is a uniquely associated price prescribed by $\Phi(\epsilon)$. Consequently, there is a stationary deterministic policy, denoted by $\tilde{\pi}^H$, such that $p_{1+(k-1)S} = \tilde{\pi}^H(\emptyset)$ and

$$p_{1+s+(k-1)S} = \tilde{\pi}^H(x)$$

when $x = (d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S})$, $1 \le s < S$, and $\sum_{i=1}^s d_{i+(k-1)S} < C$.

This enables us to bound the regret in the k-th selling season:

$$\begin{aligned}
&V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E[p_i d_i] \\
=&E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i d_i\right)\mathbf{1}_{(k-1)S \le T_\rho}\right] \\
+&E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i d_i\right)\mathbf{1}_{(k-1)S > T_\rho}\right] \\
\le&\tilde{V}_{\beta^{(0)}}(\emptyset) P((k-1)S \le T_\rho) \\
+&E\left[E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i d_i\right)\mathbf{1}_{(k-1)S > T_\rho} \mid H\right]\right] \\
\le&\tilde{V}_{\beta^{(0)}}(\emptyset)\frac{E[T_\rho]}{(k-1)S} && \text{(EC.9)} \\
+&E\left[E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \tilde{V}^{\tilde{\pi}^H}_{\beta^{(0)}}(\emptyset)\right)\mathbf{1}_{(k-1)S > T_\rho} \mid H\right]\right]. && \text{(EC.10)}
\end{aligned}$$

The term (EC.9) is finite because $E[T_\rho] < \infty$. To obtain an upper bound on the term (EC.10), we need two sensitivity results:

(S.1) Write $\vec{d}_s = (d_{1+(k-1)S}, \ldots, d_{s+(k-1)S})$ for $1 \le s \le S-1$, and set $\vec{d}_0 = (\emptyset)$. There is a $K_0 > 0$ such that, for all stationary deterministic policies $\tilde{\pi}$ and all $0 \le s \le S-1$,

$$\begin{aligned}&(\tilde{V}_{\beta^{(0)}}(\vec{d}_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s))\mathbf{1}_{\sum_{i=1}^{s} d_{i+(k-1)S} < C}\mathbf{1}_{(k-1)S > T_\rho}\\ \le K_0 &\sum_{\sigma=s}^{S-1}(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_\sigma) - \tilde{\pi}(\vec{d}_\sigma))^2\mathbf{1}_{\sum_{i=1}^{\sigma} d_{i+(k-1)S} < C} \cdot \mathbf{1}_{(k-1)S > T_\rho} \text{ a.s.}\end{aligned}$$

(S.2) There is a $K_3 > 0$ such that

$$|\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x)| \le K_3||\beta - \beta^{(0)}||, \tag{EC.11}$$

for all $\beta \in B$ with $||\beta - \beta^{(0)}|| \le \rho$, and all states $(x_1, \ldots, x_s)$ with $\sum_{i=1}^{s} x_i < C$ and $s < S$.

The proof of these two sensitivity properties is given below.

Application of (S.1), (S.2), and (EC.6) now gives

$$\begin{aligned}&E\left[E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \tilde{V}^{\tilde{\pi}^H}_{\beta^{(0)}}(\emptyset)\right)\mathbf{1}_{(k-1)S>T_\rho} \mid H\right]\right]\\ \le &E\left[E\left[K_0\sum_{\sigma=0}^{S-1}(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_\sigma) - \tilde{\pi}^H(\vec{d}_\sigma))^2\mathbf{1}_{\sum_{i=1}^{\sigma} d_{i+(k-1)S} < C} \cdot \mathbf{1}_{(k-1)S>T_\rho} \mid H\right]\right]\\ = &E\left[K_0\sum_{\sigma=0}^{S-1}(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_\sigma) - \tilde{\pi}^*_{\hat{\beta}_{\sigma+(k-1)S}}(\vec{d}_\sigma))^2\mathbf{1}_{\sum_{i=1}^{\sigma} d_{i+(k-1)S} < C} \cdot \mathbf{1}_{(k-1)S>T_\rho}\right]\\ \le &E\left[K_0K_3^2\sum_{\sigma=0}^{S-1}\left|\left|\beta^{(0)} - \hat{\beta}_{\sigma+(k-1)S}\right|\right|^2\mathbf{1}_{(k-1)S>T_\rho}\right]\\ \le &K_4\sum_{\sigma=0}^{S-1}\frac{\log(\sigma+(k-1)S)}{\sigma+(k-1)S},\end{aligned}$$

for some K_4 independent of k and S.

We thus have

$$\begin{aligned}&V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E[p_i \min\{d_i, c_i\}]\\ \le &\tilde{V}_{\beta^{(0)}}(\emptyset)E[T_\rho]\frac{1}{(k-1)S} + K_4\sum_{\sigma=0}^{S-1}\frac{\log(\sigma+(k-1)S)}{\sigma+(k-1)S}\\ \le &K_5\sum_{t=1+(k-1)S}^{kS}\frac{\log(t)}{t},\end{aligned} \tag{EC.12}$$

for some $K_5 > 0$, independent of k and S.

The proof of the theorem is complete by observing

$$\begin{aligned}\text{Regret}(\Phi(\epsilon), T) &= \sum_{k=1}^{T}\left[V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E[p_i d_i]\right]\\ &\le V_{\beta^{(0)}}(C,1) + \sum_{k=2}^{T} K_5 \sum_{t=1+(k-1)S}^{kS}\frac{\log(t)}{t} \le V_{\beta^{(0)}}(C,1) + K_5\sum_{t=1+S}^{TS}\frac{\log(t)}{t}\\ &= O(\log^2(T)).\end{aligned}$$

Proof of (S.1)

Backward induction on s. Let $s = S-1$. If $\sum_{i=1}^{S-1} d_{i+(k-1)S} < C$ then Lemma EC.1(iii) implies

$$\begin{aligned}\tilde{V}_{\beta^{(0)}}(\vec{d}_{S-1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_{S-1}) &= \max_{p\in[p_l,p_h]} ph(\beta_0^{(0)} + \beta_1^{(0)}p) - \tilde{\pi}(\vec{d}_{S-1})h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(\vec{d}_{S-1})) \\ \leq & K_0(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_{S-1}) - \tilde{\pi}(\vec{d}_{S-1}))^2 \text{ a.s.},\end{aligned}$$

and thus

$$\begin{aligned}&(\tilde{V}_{\beta^{(0)}}(\vec{d}_{S-1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_{S-1}))\cdot \mathbf{1}_{\sum_{i=1}^{S-1} d_{i+(k-1)S}<C}\cdot \mathbf{1}_{(k-1)S>T_\rho} \\ \leq\ & K_0(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_{S-1}) - \tilde{\pi}(\vec{d}_{S-1}))^2 \cdot \mathbf{1}_{\sum_{i=1}^{S-1} d_{i+(k-1)S}<C}\cdot \mathbf{1}_{(k-1)S>T_\rho} \text{ a.s.}\end{aligned}$$

Now let $0 \leq s < S-1$. If $\sum_{i=1}^{s} d_{i+(k-1)S} = C$ then $\tilde{V}_{\beta^{(0)}}(\vec{d}_s) = \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s) = 0$. If $\sum_{i=1}^{s} d_{i+(k-1)S} < C$, then, again using Lemma EC.1(iii),

$$\begin{aligned}&\tilde{V}_{\beta^{(0)}}(\vec{d}_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s) \\ =& \max_{p\in[p_l,p_h]} \big[p - (\tilde{V}_{\beta^{(0)}}(\vec{d}_s;0) - \tilde{V}_{\beta^{(0)}}(\vec{d}_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}p) + \tilde{V}(\vec{d}_s;0) \\ &-\big[\tilde{\pi}(\vec{d}_s) - (\tilde{V}_{\beta^{(0)}}(\vec{d}_s;0) - \tilde{V}_{\beta^{(0)}}(\vec{d}_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(\vec{d}_s)) \\ &+\big[\tilde{\pi}(\vec{d}_s) - (\tilde{V}_{\beta^{(0)}}(\vec{d}_s;0) - \tilde{V}_{\beta^{(0)}}(\vec{d}_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(\vec{d}_s)) \\ &-\big[\tilde{\pi}(\vec{d}_s) - (\tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s;0) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s;1))\big]h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(\vec{d}_s)) - \tilde{V}^{\tilde{\pi}}(\vec{d}_s;0) \\ \leq& K_0(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_s) - \tilde{\pi}(\vec{d}_s))^2 \\ &+\big(\tilde{V}_{\beta^{(0)}}(\vec{d}_s;0) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s;0)\big)\cdot\big(1 - h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(\vec{d}_s))\big) \\ &+\big(\tilde{V}_{\beta^{(0)}}(\vec{d}_s;1) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s;1)\big)\cdot(h(\beta_0^{(0)} + \beta_1^{(0)}\tilde{\pi}(\vec{d}_s)) \\ \leq& K_0(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_s) - \tilde{\pi}(\vec{d}_s))^2 + \big[\tilde{V}_{\beta^{(0)}}(\vec{d}_{s+1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_{s+1})\big] \text{ a.s.},\end{aligned}$$

and the induction hypothesis implies

$$\begin{aligned}&(\tilde{V}_{\beta^{(0)}}(\vec{d}_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(\vec{d}_s))\mathbf{1}_{\sum_{i=1}^{s} d_{i+(k-1)S}<C}\mathbf{1}_{(k-1)S>T_\rho} \\ \leq\ & K_0\sum_{\sigma=s}^{S-1}(\tilde{\pi}^*_{\beta^{(0)}}(\vec{d}_\sigma) - \tilde{\pi}(\vec{d}_\sigma))^2\mathbf{1}_{\sum_{i=1}^{\sigma} d_{i+(k-1)S}<C}\mathbf{1}_{(k-1)S>T_\rho} \text{ a.s.}\end{aligned}$$

Proof of (S.2)

Let $x = (x_1,\ldots,x_s) \in \tilde{\mathcal{X}}$ with $\sum_{i=1}^{s} x_i < C$ and $s < S$. We have

$$\begin{aligned}\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x) &= p^*_{\tilde{V}_\beta(x;0)-\tilde{V}_\beta(x;1),\beta} - p^*_{\tilde{V}_{\beta^{(0)}}(x;0)-\tilde{V}_{\beta^{(0)}}(x;1),\beta^{(0)}} \\ &= p^*_{a,\beta} - p^*_{a^{(0)},\beta^{(0)}},\end{aligned} \tag{EC.13}$$

in the notation of Lemma EC.1, with $a = \tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1)$ and $a^{(0)} = \tilde{V}_{\beta^{(0)}}(x;0) - \tilde{V}_{\beta^{(0)}}(x;1)$. Also we have that both $(a,\beta) \in \mathcal{U}_{AB}$ and $(a^{(0)},\beta^{(0)}) \in \mathcal{U}_{AB}$, since $p_l < \tilde{\pi}^*_\beta(x), \tilde{\pi}^*_{\beta^{(0)}}(x) < p_h$ (as noted above, in the remarks below equation (EC.5)). The set

$$\{\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \mid \beta \in B, \left|\left|\beta - \beta^{(0)}\right|\right| \le \rho, x \in \tilde{\mathcal{X}}\} \times \{\beta \in B \mid \left|\left|\beta - \beta^{(0)}\right|\right| \le \rho\}$$

is compact and contained in $\mathcal{U}_{AB}$. Since $p^*_{a,\beta}$ is continuously differentiable in a and β on $\mathcal{U}_{AB}$, it follows by a first order Taylor expansion that

$$|p^*_{a,\beta} - p^*_{a^{(0)},\beta^{(0)}}| \le K_6(|a - a^{(0)}| + ||\beta - \beta^{(0)}||), \tag{EC.14}$$

for a $K_6 > 0$ independent of a, $a^{(0)}$. It is not difficult to show by backward induction that for all $x \in \tilde{\mathcal{X}}$ there is a $K_x > 0$ such that, for all β with $||\beta - \beta^{(0)}|| \le \rho$,

$$\left|\tilde{V}_\beta(x) - \tilde{V}_{\beta^{(0)}}(x)\right| \le K_x \left|\left|\beta - \beta^{(0)}\right|\right|. \tag{EC.15}$$

Combining (EC.13), (EC.14), and (EC.15), we obtain

$$\begin{aligned}
&|\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x)| \\
\le& K_6(|a - a^{(0)}| + ||\beta - \beta^{(0)}||) \\
\le& K_6(|\tilde{V}_\beta(x;0) - \tilde{V}_{\beta^{(0)}}(x;0)| + |\tilde{V}_\beta(x;1) - \tilde{V}_{\beta^{(0)}}(x;1)| + ||\beta - \beta^{(0)}||) \\
\le& K_6(1 + 2\max_{x \in \tilde{\mathcal{X}}} K_x)||\beta - \beta^{(0)}||.
\end{aligned}$$

This proves (S.2).

Proof of Theorem 3

First note that h, B and $[p_l, p_h]$ satisfy the assumptions of Section 2.2.

The proof consists of three steps. In Step 1 we consider pricing strategies that select optimal prices whenever inventory is strictly below C. For these strategies we show that the regret in a single selling season is bounded from below by a term proportional to the mean squared estimation error at the end of the selling season, viz. equation (EC.19). In Step 2 we prove a lower bound on the expected value of this term, according to a probability density on $\beta^{(0)}$. The proof is analogous to the proof of the van Trees inequality in Gill and Levit (1995, Section 2). We cannot directly apply their result, however, because we are in a slightly different setting (discrete instead of continuous random variables, and non-deterministic number of observations upon which the estimate of β_0 is based); we therefore include a complete proof, resulting in equation (EC.24). In Step 3 we combine the results of Step 1 and Step 2 to show that the regret in the k-th single selling season is bounded

from below by a term proportional to $1/k$; this implies that the total regret after T selling seasons is bounded from below by a term proportional to $\log T$.

Throughout the proof, let ψ be an arbitrary price strategy.

Step 1. Let $\beta=(\beta_0,\beta_1)\in B$ be arbitrary and fixed, and let $k\in\mathbb{N}$, $k\geq 2$. For shorthand notation, write

$$h_\beta(p)=h(\beta_0+\beta_1 p).$$

Define

$$\varepsilon=\inf_{\beta\in B}\inf_{p\in[p_l,p_h]}\min\{h_\beta(p),1-h_\beta(p)\},$$

and note that $\varepsilon>0$. Let ψ' be the pricing strategy that for all $t\in\mathbb{N}$ coincides with ψ if $c_t=C$, and that equals the optimal price $\pi^*_\beta(c_t,s_t)$ if $c_t<C$. Let p_i be the prices generated by ψ'. For $1\leq j<S$ write $\vec{d}_j=(d_{1+(k-1)S},\ldots,d_{j+(k-1)S})\in\{0,1\}^j$ for the vector of demand realizations in the first j stages of season k, and set $\vec{d}_0=0$. For all $j\in\{1,\ldots,S-1\}$, we have the following inequality:

$$\begin{aligned}
&V(C,j)-E\Big[\sum_{i=j+(k-1)S}^{kS}p_i d_i\mid\vec{d}_{j-1}=0\Big]\\
=&h_\beta(\pi^*_\beta(C,j))\cdot(\pi^*_\beta(C,j)+V(C-1,j+1))+(1-h_\beta(\pi^*_\beta(C,j)))\cdot V(C,j+1)\\
&-E\Big[\sum_{i=j+(k-1)S+1}^{kS}p_i d_i\mid\vec{d}_{j-1}=0,d_{j+(k-1)S}=1\Big]P\left(d_{j+(k-1)S}=1\mid\vec{d}_{j-1}=0\right)\\
&-E\Big[\sum_{i=j+(k-1)S+1}^{kS}p_i d_i\mid\vec{d}_{j-1}=0,d_{j+(k-1)S}=0\Big]P\left(d_{1+(k-1)S}=0\mid\vec{d}_{j-1}=0\right)\\
&-E[p_{j+(k-1)S}\,h_\beta(p_{j+(k-1)S})\mid\vec{d}_{j-1}=0]\\
=&\quad h_\beta(\pi^*_\beta(C,j))\cdot(\pi^*_\beta(C,j)+V(C-1,j+1))+(1-h_\beta(\pi^*_\beta(C,j)))\cdot V(C,j+1)\\
&-E\Big[h_\beta(p_{j+(k-1)S})\cdot(p_{j+(k-1)S}+V(C-1,j+1))+(1-h_\beta(p_{j+(k-1)S}))\cdot V(C,j+1)\mid\vec{d}_{j-1}=0\Big]\\
&+E\Big[h_\beta(p_{j+(k-1)S})\mid\vec{d}_{j-1}=0\Big]\cdot\left(V(C-1,j+1))-E\Big[\sum_{i=j+(k-1)S+1}^{kS}p_i d_i\mid\vec{d}_{j-1}=0,d_{j+(k-1)S}=1\Big]\right)\\
&+E\Big[1-h_\beta(p_{j+(k-1)S})\mid\vec{d}_{j-1}=0\Big]\cdot\left(V(C,j+1))-E\Big[\sum_{i=j+(k-1)S+1}^{kS}p_i d_i\mid\vec{d}_{j-1}=0,d_{j+(k-1)S}=0\Big]\right)\\
\geq&\varepsilon\cdot\left(V(C,j+1)-E\Big[\sum_{i=j+(k-1)S+1}^{kS}p_i d_i\mid\vec{d}_j=0\Big]\right).
\end{aligned}\tag{EC.16}$$

In the first equality we use the recursive definition of $V(C,j)$, condition $E[\sum_{i=j+1+(k-1)S}^{kS}p_i d_i\mid\vec{d}_{j-1}=0]$ on the event $\{d_{j+(k-1)S}=1\mid\vec{d}_{j-1}=0\}$ and its complement, and use $E[p_{j+(k-1)S}\,d_{j+(k-1)S}\mid\vec{d}_{j-1}=0]=E[p_{j+(k-1)S}\,h_\beta(p_{j+(k-1)S})\mid\vec{d}_{j-1}=0]$. In the second inequality we add and substract $E[h_\beta(p_{j+(k-1)S})\cdot V(C-1,j+1)+(1-h_\beta(p_{j+(k-1)S}))\cdot V(C,j+1)\mid\vec{d}_{j-1}=0]$, and use $P(d_{j+(k-1)S}=$

$1 \mid \vec{d}_{j-1} = 0) = E[h_\beta(p_{j+(k-1)S}) \mid \vec{d}_{j-1} = 0]$. The inequality follows from the fact that $\pi^*_\beta(C,j)$, by definition, maximizes

$$h_\beta(p) \cdot (p + V(C-1, j+1)) + (1 - h_\beta(p)) \cdot V(C, j+1), \quad p \in [p_l, p_h],$$

and from

$$V(C-1,j+1)) - E\Bigg[\sum_{i=j+(k-1)S+1}^{kS} p_i d_i \mid \vec{d}_{j-1} = 0, d_{j+(k-1)S} = 1 \Bigg] = 0,$$

which follows from the fact that, by definition of ψ', optimal prices are chosen whenever inventory is strictly lower than C.

Repeated application of (EC.16) gives

$$V(C,1) - E\Bigg[\sum_{i=1+(k-1)S}^{kS} p_i d_i \Bigg] \geq \varepsilon^{S-1} \cdot \Big(V(C,S) - E\Big[p_{kS} d_{kS} \mid \vec{d}_{S-1} = 0 \Big] \Big). \tag{EC.17}$$

Write b_{kS} for p_{kS} multiplied by $-2\beta_1$, given that $\vec{d}_{S-1} = 0$:

$$b_{kS} = -2\beta_1 \cdot \psi'_{kS}(p_1, \ldots, p_{kS-1}, d_1, \ldots, d_{(k-1)S}, \underbrace{0, 0, \ldots, 0}_{S-1 \text{ zeros}}).$$

Note that b_{kS} (as well as all prices $p_{(k-1)S+1}, \ldots, p_{kS-1}$) only depends on prices $p_1, \ldots, p_{(k-1)S}$ and demand realizations $d_1, \ldots, d_{(k-1)S}$ up to selling season $k-1$; in particular, b_{kS} does not depend on the event $\{\vec{d}_{S-1} = 0\}$. Together with $\pi^*_\beta(C,S) = \beta_0/(-2\beta_1)$ this implies

$$\begin{aligned} V(C,S) - E\Big[p_{kS} d_{kS} \mid \vec{d}_{S-1} = 0 \Big] &= \pi^*_\beta(C,S) h_\beta(\pi^*_\beta(C,S)) - E\Big[p_{kS}\, h_\beta(p_{kS}) \mid \vec{d}_{S-1} = 0 \Big] \\ = -\beta_1 E\Big[(\pi^*_\beta(C,S) - p_{kS})^2 \mid \vec{d}_{S-1} = 0 \Big] &= \frac{-1}{4\beta_1} E\Big[(\beta_0 - b_{kS})^2 \Big]. \end{aligned} \tag{EC.18}$$

Combining (EC.17) and (EC.18) with $\inf_{\beta \in B} \frac{-1}{4\beta_1} = 1/3$ yields

$$V(C,1) - E\Bigg[\sum_{i=1+(k-1)S}^{kS} p_i d_i \Bigg] \geq \frac{1}{3} \varepsilon^{S-1} E\Big[(\beta_0 - b_{kS})^2 \Big]. \tag{EC.19}$$

Step 2. In this step we view β as a random variable $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$, drawn from B according to the probability density function λ, defined by

$$\lambda(\beta_0, \beta_1) = \frac{256}{3} \cos^2(8\pi\beta_0 - 11\pi/2), \quad (\beta_0, \beta_1) \in B.$$

We prove a lower bound on the term $E_\lambda\Big[(\beta_0 - b_{kS})^2 \Big]$ in (EC.19), where the subscript λ emphasizes that we take expectation not only with respect to the distribution of the demand, but also with respect to λ. To this end, we need to introduce some notation. Let $D = (D_1, \ldots, D_{(k-1)S})$, with D_i equal to the (random) demand in period i, $1 \leq i \leq (k-1)S$. Note that, conditional on $\boldsymbol{\beta} = \beta$, D_i is

Bernoulli with mean $h_\beta(p_i)$ if $c_i > 0$, and $D_i = 0$ with probability one if $c_i = 0$. For $d \in \{0,1\}^{(k-1)S}$, let $c_i(d) = C - \sum_{j<i, SS_j = SS_i} d_j$ be the inventory available at the beginning of period i given the vector of demand realizations d, and write $c(d) = (c_1(d), \ldots, c_{(k-1)S}(d))$. Note that D takes values in $\mathcal{D} = \{d \in \{0,1\}^{(k-1)S} \mid d \le c(d)\}$. The probability mass function $f(d \mid \beta)$ of D, conditional on $\boldsymbol{\beta} = \beta$, is given by

$$f(d \mid \beta) = \prod_{\substack{i=1,\\ c_i(d)>0}}^{(k-1)S} h_\beta(p_i(d))^{d_i} (1 - h_\beta(p_i(d))^{1-d_i},$$

where $p_i(d)$ is the price in period i, given demand realizations $d_1, \ldots, d_{i-1}$, under strategy ψ'. For $d \in \mathcal{D}$, write

$$b(d \mid \beta) = -2\beta_1 \psi'_{kS}(p_1(d), p_2(d), \ldots, p_{kS-1}, d_1, \ldots, d_{(k-1)S}, \underbrace{0, 0, \ldots, 0}_{S-1 \text{ zeros}})$$

for $-2\beta_1$ times the price in period kS given that demand is zero in periods $(k-1)S+1, \ldots, kS-1$ and demand in preceding selling seasons is given by d; all of this conditional on $\boldsymbol{\beta} = \beta$.

We now prove a lower bound on

$$E_\lambda[(b(D \mid \boldsymbol{\beta}) - \boldsymbol{\beta}_0)^2] = \int_{5/8}^{6/8} \int_{-3/4}^{-9/16} \sum_{d \in \mathcal{D}} f(d \mid \beta_0, \beta_1) \lambda(\beta_0, \beta_1) (b(d \mid \beta_0, \beta_1) - \beta_0)^2 \, d\beta_1 d\beta_0.$$

By the mean-value theorem, there is a $\tilde{\beta}_1 \in [-3/4, -9/16]$ such that

$$E_\lambda[(b(D \mid \boldsymbol{\beta}) - \boldsymbol{\beta}_0)^2] = \frac{3}{16} \int_{5/8}^{6/8} \sum_{d \in \mathcal{D}} f(d \mid \beta_0, \tilde{\beta}_1) \lambda(\beta_0, \tilde{\beta}_1) (b(d \mid \beta_0, \tilde{\beta}_1) - \beta_0)^2 \, d\beta_0.$$

For notational convenience we stop writing the dependence of f, λ, b on $\tilde{\beta}_1$. Applying Cauchy-Schwarz on the integral-sum and integrating by parts, we obtain

$$\begin{aligned}
&\frac{3}{16} \int_{5/8}^{6/8} \sum_{d \in \mathcal{D}} (b(d \mid \beta_0) - \beta_0)^2 f(d \mid \beta_0) \lambda(\beta_0) \, d\beta_0 \cdot \int_{5/8}^{6/8} \sum_{d \in \mathcal{D}} \Big(\frac{\partial}{\partial \beta_0} \log(f(d \mid \beta_0) \lambda(\beta_0)) \Big)^2 f(d \mid \beta_0) \lambda(\beta_0) \, d\beta_0 \\
\ge &\frac{3}{16} \int_{5/8}^{6/8} \sum_{d \in \mathcal{D}} (b(d \mid \beta_0) - \beta_0) \Big(\frac{\partial}{\partial \beta_0} \log(f(d \mid \beta_0) \lambda(\beta_0)) \Big) f(d \mid \beta_0) \lambda(\beta_0) \, d\beta_0 \\
= &\frac{3}{16} \sum_{d \in \mathcal{D}} b(d \mid \beta_0) \Big(f(d \mid 6/8) \lambda(6/8) - f(d \mid 5/8) \lambda(5/8) \Big) \\
&- \frac{3}{16} \sum_{d \in \mathcal{D}} \beta_0 \Big(f(d \mid 6/8) \lambda(6/8) - f(d \mid 5/8) \lambda(5/8) \Big) + \frac{3}{16} \int_{5/8}^{6/8} \sum_{d \in \mathcal{D}} f(d \mid \beta_0) \lambda(\beta_0) \, d\beta_0 = 1, \qquad \text{(EC.20)}
\end{aligned}$$

since $\lambda(5/8) = \lambda(6/8) = 0$ and $\int_{5/8}^{6/8} \lambda(\beta_0) \, d\beta_0 = \int_{5/8}^{6/8} \lambda(\beta_0, \tilde{\beta}_1) \, d\beta_0 = 16/3$.

We have

$$\int_{5/8}^{6/8} \Big(\frac{\partial}{\partial \beta_0} \log(\lambda(\beta_0)) \Big)^2 \lambda(\beta_0) \, d\beta_0 = \frac{4096\pi^2}{3}, \qquad \text{(EC.21)}$$

and

$$E\left[\frac{\partial}{\partial\beta_0}\log f(D\,|\,\beta_0)\right]=\sum_{d\in\mathcal{D}}\frac{\partial}{\partial\beta_0}f(d\,|\,\beta_0)=0\quad\text{for all }\beta_0, \tag{EC.22}$$

and

$$\begin{aligned}E\left[\left(\frac{\partial}{\partial\beta_0}\log f(D\,|\,\beta_0)\right)^2\right]&=E\left[\left(\sum_{\substack{i=1,\\ c_i(D)>0}}^{(k-1)S}\frac{D_i-h_{\beta_0}(p_i(D))}{h_{\beta_0}(p_i(D))(1-h_{\beta_0}(p_i(D)))}\right)^2\right]\\ &=\sum_{\substack{i=1,\\ c_i(D)>0}}^{(k-1)S}\frac{1}{h_{\beta_0}(p_i(D))(1-h_{\beta_0}(p_i(D)))}\le\frac{(k-1)S}{\varepsilon^2},\end{aligned} \tag{EC.23}$$

which follows from

$$E\left[\frac{D_i-h_{\beta_0}(p_i(D))}{h_{\beta_0}(p_i(D))(1-h_{\beta_0}(p_i(D)))}\cdot\frac{D_j-h_{\beta_0}(p_j(D))}{h_{\beta_0}(p_j(D))(1-h_{\beta_0}(p_j(D)))}\right]=0$$

for all $i\neq j$ with $c_i(D)>0$, $c_j(D)>0$, and

$$E\left[\left(\frac{D_i-h_{\beta_0}(p_i(D))}{h_{\beta_0}(p_i(D))(1-h_{\beta_0}(p_i(D)))}\right)^2\right]=\frac{1}{h_{\beta_0}(p_i(D))(1-h_{\beta_0}(p_i(D)))}$$

for all i with $c_i(D)>0$.

Plugging (EC.21), (EC.22) and (EC.23) into (EC.20), we obtain the following lower bound:

$$\begin{aligned}E_\lambda[(b(D\,|\,\boldsymbol{\beta})-\boldsymbol{\beta}_0)^2]&\ge\frac{16}{3}\left(\int_{5/8}^{6/8}\sum_{d\in\mathcal{D}}\Big(\frac{\partial}{\partial\beta_0}\log(f(d\,|\,\beta_0)\lambda(\beta_0))\Big)^2 f(d\,|\,\beta_0)\lambda(\beta_0)\,d\beta_0\right)^{-1}\\ &\ge\frac{1}{(k-1)S/\varepsilon^2+\frac{4096\pi^2}{3}}.\end{aligned} \tag{EC.24}$$

Step 3. Combining (EC.19) and (EC.24), we obtain

$$\begin{aligned}\sup_{\beta^{(0)}\in B}\text{Regret}(\psi,T)&\ge E_\lambda[\text{Regret}(\psi,T)]\\ &\ge\sum_{k=2}^{T}E_\lambda\left[V(C,1)-E\left[\sum_{i=1+(k-1)S}^{kS}p_id_i\right]\right]\\ &\ge\sum_{k=2}^{T}\frac{1}{3}\varepsilon^{S-1}E_\lambda[(b(D_1,\ldots,D_{(k-1)S}\,|\,\boldsymbol{\beta})-\boldsymbol{\beta}_0)^2]\\ &\ge\sum_{k=2}^{T}\frac{1}{3}\varepsilon^{S-1}\frac{1}{(k-1)S/\varepsilon^2+\frac{4096\pi^2}{3}}\\ &\ge K_0\log(T),\end{aligned}$$

where we define

$$K_0=\frac{1}{3}\varepsilon^{S-1}\frac{1}{\max\{S/\varepsilon^2,\frac{4096\pi^2}{3}\}}\cdot\frac{1}{2\log(2)}.$$

This completes the proof.

Proof of Theorem 4

The proof of Theorem 2 shows, for the base case of non-overlapping selling seasons with non-varying capacity and season length, that the regret of the k-th selling season ($k \geq 2$) satisfies

$$V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E[p_i \min\{d_i, c_i\}] \leq K_5 \sum_{t=1+(k-1)S}^{kS} \frac{\log(t)}{t}, \tag{EC.25}$$

for some $K_5 > 0$ independent of k; viz. Equation (EC.12).

For any season j, estimation of $\beta^{(0)}$ to determine the prices during season j is based only on (i) sales data preceding t_j and (ii) sales data generated during season j (thus, all sales data generated in other seasons overlapping with season j is neglected). Because of this assumption, we can use the same MDP as the one constructed in the proof of Theorem 2 to show, analogous to (EC.25), that the regret of the j-th selling season ($t_j > 1$) satisfies

$$\begin{aligned} V_{\beta^{(0)},S_j}(C_j,1) - \sum_{s=1}^{S_j} E[p_{j,s} d_{j,s}] &\leq K_5 \sum_{s=1}^{S_j} \frac{\log(t_j - 1 + s)}{t_j - 1 + s} \\ &\leq K_5 S_j \log(t_j)/t_j \end{aligned}$$

for some constant $K_5 = K_5(C_j, S_j)$. The constant $K_j(C_j, S_j)$ may depend on C_j and S_j. However, $\tilde{K}_5 := \sup_{j \in \mathbb{N}} S_j K_5(C_j, S_j)$ does not depend on (C_j, S_j), and $\tilde{K}_5 < \infty$ because it is simply a supremum over a finite set (note that C_j and S_j are bounded). It follows that

$$\begin{aligned} \text{Regret}(\Phi'(\epsilon), T) &= \sum_{j: t_j - 1 + S_j < T} \left\{ V_{\beta^{(0)},S_j}(C_j,1) - \sum_{s=1}^{S_j} E[p_{j,s} d_{j,s}] \right\} \\ &\leq \sum_{j: t_j = 1} V_{\beta^{(0)},S_j}(C_j,1) + \sum_{j: t_j > 1, t_j - 1 + S_j < T} \tilde{K}_5 \log(t_j)/t_j \\ &\leq V_{\beta^{(0)},S_j}(C_j,1) \cdot \sup_{t \in \mathbb{N}} |\{j \in \mathbb{N} : t_j = t\}| + \sum_{i=2}^{T} \sum_{j: t_j = i, t_j - 1 + S_j < T} \tilde{K}_5 \log(t_j)/t_j \\ &\leq V_{\beta^{(0)},S_j}(C_j,1) \cdot \sup_{t \in \mathbb{N}} |\{j \in \mathbb{N} : t_j = t\}| + \sum_{i=2}^{T} \tilde{K}_5 \sup_{t \in \mathbb{N}} |\{j \in \mathbb{N} : t_j = t\}| \cdot \log(i)/i \\ &= O(\log^2(T)). \end{aligned}$$

EC.2. Auxiliary Lemmas

LEMMA EC.1. *For all $a \in \mathbb{R}$ and $\beta \in B$, define the function $f_{a,\beta} : [p_l, p_h] \to \mathbb{R}$, $f(p) = (p-a)h(\beta_0 + \beta_1 p)$. Write $\dot{f}_{a,\beta}(p)$ and $\ddot{f}_{a,\beta}(p)$ for the first and second derivative of $f_{a,\beta}(p)$ with respect to p, and let $p^*_{a,\beta} = \arg\max_{p \in [p_l, p_h]} f_{a,\beta}(p)$. In addition, let*

$$\mathcal{U}_B = \left\{ (\beta_0, \beta_1) \in \times\mathbb{R} \times (-\infty, 0) \,\middle|\, 0 < h(z) < 1 \text{ and } \dot{h}(z) > 0, \text{ for all } z = \beta_0 + \beta_1 p,\ p \in [p_l, p_h] \right\}$$

and

$$\mathcal{U}_{AB} = \left\{(a,\beta_0,\beta_1) \in \mathbb{R}\times\mathcal{U}_B \,\middle|\, p^*_{a,\beta} \in (p_l,p_h)\right\}.$$

Then:

(i) For each $(a,\beta)\in\mathbb{R}\times\mathcal{U}_B$, $p^*_{a,\beta}$ *is uniquely defined, and for each* $(a,\beta)\in\mathcal{U}_{AB}$, $\dot{f}(p^*_{a,\beta})=0$ *and* $\ddot{f}(p^*_{a,\beta})<0$.

(ii) On $\mathcal{U}_{AB}$, $p^*_{a,\beta}$ *is continuously differentiable in* a *and* β, *strictly increasing in* a, *and nondecreasing in* β_0 *and* β_1. *On* $\mathbb{R}\times\mathcal{U}_B$, $p^*_{a,\beta}$ *is nondecreasing in* a, β_0 *and* β_1. *In addition, on* $\mathcal{U}_{AB}$, $f_{a,\beta}(p^*_{a,\beta})$ *is continuously differentiable in* a *and* β, *strictly decreasing in* a, *and strictly increasing in* β_0 *and* β_1.

(iii) There is a $K_0>0$ *such that* $f_{a,\beta}(p^*_{a,\beta})-f_{a,\beta}(p)\le K_0(p-p^*_{a,\beta})^2$ *for all* $p\in[p_l,p_h]$ *and* $(a,\beta)\in\mathcal{U}_{AB}$.

LEMMA EC.2. *Let* $\beta\in B$ *be arbitrary.*

(i) $\Delta V_\beta(c,s)\ge 0$ *for all* $1\le c\le C$ *and* $1\le s<S$.

(ii) $p_h\ge V_\beta(1,s)\ge V_\beta(1,s+1)\ge p_l$ *for all* $1\le s<S$.

(iii) $\Delta V_\beta(c,s)\ge\Delta V(c+1,s)$ *for all* $1\le c\le C$ *and* $1\le s\le S$.

(iv) $V_{\beta'}(1,s)\le V_\beta(1,s)$, *for all* $1\le s\le S$ *and all* $\beta'=(\beta'_0,\beta'_1)\in B$ *with* $\beta'_0\le\beta_0$ *and* $\beta'_1\le\beta_1$.

(v) $\pi^*_\beta(C,S)\le\pi^*_\beta(c,s)\le\pi^*_\beta(1,1)$, *for all* $1\le c\le C$ *and* $1\le s\le S$.

(vi) $\pi^*_\beta(c,s)$ *is continuous in* β, *for all* $1\le c\le C$ *and* $1\le s\le S$.

LEMMA EC.3. *Let* $k\in\mathbb{N}$ *and* $\delta>0$. *If there are* $s,s'\in\{1,\ldots,S\}$ *such that* $|p_{s+(k-1)S}-p_{s'+(k-1)S}|\ge\delta$ *and* $c(s+(k-1)S)c(s'+(k-1)S)>0$, *then* $\lambda_{\min}(P_{kS})\ge\lambda_{\min}(P_{(k-1)S})+\frac{1}{2}\delta^2(1+p_h^2)^{-1}$.

Proof of Lemma EC.1

(i) Let $(a,\beta)\in\mathbb{R}\times\mathcal{U}_B$. We have

$$\dot{f}(p)=h(\beta_0+\beta_1p)+(p-a)\beta_1\dot{h}(\beta_0+\beta_1p)$$

and

$$\ddot{f}(p)=2\beta_1\dot{h}(\beta_0+\beta_1p)+(p-a)\beta_1^2\ddot{h}(\beta_0+\beta_1p),$$

for all $p\in[p_l,p_h]$. Log-concavity of h implies

$$2-\frac{h(z)\ddot{h}(z)}{\dot{h}(z)^2}=1+\frac{\dot{h}(z)^2-h(z)\ddot{h}(z)}{h(z)^2}\cdot\frac{h(z)^2}{\dot{h}(z)^2}=1-\frac{h(z)^2}{\dot{h}(z)^2}\cdot\frac{\partial^2\log(h(z))}{\partial z^2}>0,$$

for all $z = \beta_0 + \beta_1 p$, $p \in [p_l, p_h]$. It follows that all $p \in [p_l, p_h]$ with $\dot{f}(p) = 0$ satisfy

$$\begin{aligned}\ddot{f}(p) &= 2\beta_1 \dot{h}(\beta_0 + \beta_1 p) + \frac{-h(\beta_0 + \beta_1 p)}{\beta_1 \dot{h}(\beta_0 + \beta_1 p)} \beta_1^2 \ddot{h}(\beta_0 + \beta_1 p) \\ &= \beta_1 \dot{h}(\beta_0 + \beta_1 p) \Big[2 - \frac{h(\beta_0 + \beta_1 p) \ddot{h}(\beta_0 + \beta_1 p)}{\dot{h}(\beta_0 + \beta_1 p)^2} \Big] < 0.\end{aligned}$$

This implies that either $f_{a,\beta}(p)$ is strictly monotone on $[p_l, p_h]$ (in which case the unique maximum of $f_{a,\beta}(p)$ is on the boundary of $[p_l, p_h]$, or $f_{a,\beta}(p)$ has a unique maximum $p^*_{a,\beta}$ on $[p_l, p_h]$ with $\ddot{f}(p^*_{a,\beta}) < 0$.

(ii) Note that $\mathcal{U}_{AB}$ is an open set, and that, for each $(a, \beta) \in \mathcal{U}_{AB}$, the equation $\dot{f}_{a,\beta}(p) = 0$ has a unique solution $p^*_{a,\beta} \in (p_l, p_h)$ with $\ddot{f}_{a,\beta}(p^*_{a,\beta}) < 0$. By the Implicit Function Theorem (see e.g. Duistermaat and Kolk 2004), $p^*_{a,\beta}$ is continuously differentiable in a and β on $\mathcal{U}_{AB}$, with derivatives given by

$$\left.\frac{\partial p^*_{a,\beta}}{\partial a}\right|_{a,\beta} = \frac{-1}{\ddot{f}_{a,\beta}(p^*_{a,\beta})} \cdot \left.\frac{\partial \dot{f}_{a,\beta}(p)}{\partial a}\right|_{p^*_{a,\beta}}$$

and

$$\left.\frac{\partial p^*_{a,\beta}}{\partial \beta_i}\right|_{a,\beta} = \frac{-1}{\ddot{f}_{a,\beta}(p^*_{a,\beta})} \cdot \left.\frac{\partial \dot{f}_{a,\beta}(p)}{\partial \beta_i}\right|_{p^*_{a,\beta}}, \quad (i = 1, 2).$$

For all $(a, \beta) \in \mathcal{U}_{AB}$ we have $\ddot{f}_{a,\beta}(p^*_{a,\beta}) < 0$,

$$\left.\frac{\partial \dot{f}_{a,\beta}(p)}{\partial a}\right|_{p^*_{a,\beta}} = -\beta_1 \dot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) > 0, \tag{EC.26}$$

$$\begin{aligned}\left.\frac{\partial}{\partial \beta_0} \dot{f}_{a,\beta}(p)\right|_{p^*_{a,\beta}} &= \dot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) + (p^*_{a,\beta} - a)\beta_1 \ddot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) \\ &= \dot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) \left(1 - \frac{h(\beta_0 + \beta_1 p^*_{a,\beta}) \ddot{h}(\beta_0 + \beta_1 p^*_{a,\beta})}{\dot{h}(\beta_0 + \beta_1 p^*_{a,\beta})^2} \right) \geq 0,\end{aligned} \tag{EC.27}$$

and

$$\begin{aligned}\left.\frac{\partial}{\partial \beta_1} \dot{f}_{a,\beta}(p)\right|_{p^*_{a,\beta}} &= p^*_{a,\beta} \dot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) + (p^*_{a,\beta} - a)\dot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) + p^*_{a,\beta}(p^*_{a,\beta} - a)\beta_1 \ddot{h}(\beta_0 + \beta_1 p^*_{a,\beta}) \\ &= p^*_{a,\beta} \left.\frac{\partial}{\partial \beta_0} \dot{f}(p)\right|_{p^*_{a,\beta}} + \frac{1}{-\beta_1} h(\beta_0 + \beta_1 p^*_{a,\beta}) \geq 0,\end{aligned} \tag{EC.28}$$

using $(p^*_{a,\beta} - a)\beta_1 = -h(\beta_0 + \beta_1 p^*_{a,\beta}) / \dot{h}(\beta_0 + \beta_1 p^*_{a,\beta})$ and log-concavity of h. It follows that $p^*_{a,\beta}$ is strictly increasing in a and nondecreasing in β_0 and β_1.

If $(a,\beta) \in \mathbb{R} \times \mathcal{U}_B$ and $p^*_{a,\beta} = p_l$, then $\dot{f}_{a,\beta}(p_l) \le 0$. By (EC.26), $\frac{\partial}{\partial a}\dot{f}_{a,\beta}(p_l) \ge 0$, which implies $\dot{f}_{a',\beta}(p_l) \le 0$ and $p^*_{a',\beta} = p_l$ for all $a' \le a$. By (EC.27) and (EC.28),

$$\begin{aligned}\frac{\partial}{\partial \beta_0}\dot{f}_{a,\beta}(p_l) &= \dot{h}(\beta_0+\beta_1 p_l) + (p_l - a)\beta_1 \ddot{h}(\beta_0+\beta_1 p_l) \\ &\ge \frac{\dot{h}(\beta_0+\beta_1 p_l)}{h(\beta_0+\beta_1 p_l)} \cdot (\dot{f}_{a,\beta}(p_l) - h(\beta_0+\beta_1 p_l)) \cdot \frac{\ddot{h}(\beta_0+\beta_1 p_l)h(\beta_0+\beta_1 p_l)}{\dot{h}(\beta_0+\beta_1 p_l)^2} \ge 0,\end{aligned}$$

and

$$\frac{\partial}{\partial \beta_1}\dot{f}_{a,\beta}(p_l) = p_l \frac{\partial}{\partial \beta_0}\dot{f}_{a,\beta}(p_l) + \frac{(\dot{f}_{a,\beta}(p_l) - h(\beta_0+\beta_1 p_l))}{\beta_1} \ge 0,$$

using $\dot{f}_{a,\beta}(p_l) - h(\beta_0+\beta_1 p_l) \le 0$ and log-concavity of h. This implies $\dot{f}_{a,\beta'}(p_l) \le 0$ and thus $p^*_{a,\beta'} = p_l$ for all $\beta' \in B$ with $\beta' \le \beta$ (in both coordinates). By similar arguments we can show that $p^*_{a',\beta'} = p_h$ if $p^*_{a,\beta} = p_h$, $a' \ge a$ and $\beta' \ge \beta$. These observations, combined with the fact that $p^*_{a,\beta}$ is nondecreasing in a, β_0 and β_1 if $p^*_{a,\beta} \in (p_l, p_h)$, show that $p^*_{a,\beta}$ is nondecreasing in a, β_0 and β_1 on $\mathbb{R} \times \mathcal{U}_B$.

The assertions on $f_{a,\beta}(p^*_{a,\beta})$ follow by observing that

$$\left.\frac{\partial f_{a,\beta}(p^*_{a,\beta})}{\partial a}\right|_{a,\beta} = \left.\frac{\partial f_{a,\beta}(p)}{\partial p}\right|_{p^*_{a,\beta}} \cdot \left.\frac{\partial p^*_{a,\beta}}{\partial a}\right|_{a,\beta} + \left.\frac{\partial f_{a,\beta}(p)}{\partial a}\right|_{p^*_{a,\beta}} = 0 - h(\beta_0+\beta_1 p^*_{a,\beta}) < 0,$$

$$\begin{aligned}\left.\frac{\partial f_{a,\beta}(p^*_{a,\beta})}{\partial \beta_0}\right|_{a,\beta} &= \left.\frac{\partial f_{a,\beta}(p)}{\partial p}\right|_{p^*_{a,\beta}} \cdot \left.\frac{\partial p^*_{a,\beta}}{\partial \beta_0}\right|_{a,\beta} + \left.\frac{\partial f_{a,\beta}(p)}{\partial \beta_0}\right|_{p^*_{a,\beta}} \\ &= 0 + (p^*_{a,\beta} - a)\dot{h}(\beta_0+\beta_1 p^*_{a,\beta}) = \frac{-h(\beta_0+\beta_1 p^*_{a,\beta})}{\beta_1} > 0,\end{aligned}$$

and

$$\begin{aligned}\left.\frac{\partial f_{a,\beta}(p^*_{a,\beta})}{\partial \beta_1}\right|_{a,\beta} &= \left.\frac{\partial f_{a,\beta}(p)}{\partial p}\right|_{p^*_{a,\beta}} \cdot \left.\frac{\partial p^*_{a,\beta}}{\partial \beta_1}\right|_{a,\beta} + \left.\frac{\partial f_{a,\beta}(p)}{\partial \beta_1}\right|_{p^*_{a,\beta}} \\ &= 0 + (p^*_{a,\beta} - a)p^*_{a,\beta}\dot{h}(\beta_0+\beta_1 p^*_{a,\beta}) = \frac{-h(\beta_0+\beta_1 p^*_{a,\beta})p^*_{a,\beta}}{\beta_1} > 0,\end{aligned}$$

again using $(p^*_{a,\beta} - a)\beta_1 = -h(\beta_0+\beta_1 p^*_{a,\beta})/\dot{h}(\beta_0+\beta_1 p^*_{a,\beta})$.

(iii) Let $K_0 = \sup_{(a,\beta,p)\in \mathcal{U}_{AB}\times[p_l,p_h]} -\ddot{f}_{a,\beta}(p)/2$. Since $\ddot{f}_{a,\beta}(p)$ is defined and continuous on the closure of $\mathcal{U}_{AB} \times [p_l, p_h]$ (which is compact) and since $\ddot{f}_{a,\beta}(p^*_{a,\beta}) < 0$ for all $(a,\beta) \in \mathcal{U}_{AB}$, it follows that $0 < K_0 < \infty$. A Taylor expansion then implies

$$f_{a,\beta}(p) \ge f_{a,\beta}(p^*_{a,\beta}) - K_0(p - p^*_{a,\beta})^2,$$

using $\dot{f}_{a,\beta}(p^*_{a,\beta}) = 0$ for all $(a,\beta) \in \mathcal{U}_{AB}$.

Proof of Lemma EC.2

Throughout the proof, fix $\beta \in B$.

(i) If $s = S$ then $\Delta V_\beta(c, S) = 0$ for $c > 1$ or $c = 0$, and $V_\beta(1, S) = \max_{p\in[p_l,p_h]} ph(\beta_0 + \beta_1 p) \geq 0$. If $s < S$ then by backward induction on s,

$$\begin{aligned}
\Delta V_\beta(c,s) &= (\pi^*_\beta(c,s) - \Delta V_\beta(c,s+1))h(\beta_0 + \beta_1\pi^*_\beta(c,s)) + V_\beta(c,s+1) \\
&\quad - (\pi^*_\beta(c-1,s) - \Delta V_\beta(c-1,s+1))h(\beta_0 + \beta_1\pi^*_\beta(c-1,s)) - V_\beta(c-1,s+1) \\
&\geq (\pi^*_\beta(c-1,s) - \Delta V_\beta(c,s+1))h(\beta_0 + \beta_1\pi^*_\beta(c-1,s)) + V_\beta(c,s+1) \\
&\quad - (\pi^*_\beta(c-1,s) - \Delta V_\beta(c-1,s+1))h(\beta_0 + \beta_1\pi^*_\beta(c-1,s)) - V_\beta(c-1,s+1) \\
&= \Delta V_\beta(c,s+1))(1 - h(\beta_0 + \beta_1\pi^*_\beta(c-1,s))) \\
&\quad + \Delta V_\beta(c-1,s+1))h(\beta_0 + \beta_1\pi^*_\beta(c-1,s)) \geq 0.
\end{aligned}$$

(ii) By backward induction on s it follows that $V(1,s) \in [p_l, p_h]$ for all $1 \leq s \leq S$. This implies

$$\begin{aligned}
V_\beta(1,s) &= h(\beta_0 + \beta_1\pi(1,s))\pi(1,s) + (1 - h(\beta_0 + \beta_1\pi(1,s)))V(1,s+1) \\
&\geq h(\beta_0 + \beta_1 V(1,s+1))V(1,s+1) + (1 - h(\beta_0 + \beta_1 V(1,s+1)))V(1,s+1) \\
&= V(1,s+1),
\end{aligned}$$

for all $1 \leq s < S$.

(iii) The proof is by backward induction on s, and mimics the proof of Proposition 5.2, page 238 of Talluri and van Ryzin (2004). For $s = S$ the assertion follows immediately from $\Delta V_\beta(1,s) \geq 0$ and $\Delta V_\beta(c,s) = 0$ for all $2 \leq c \leq C$. Now assume the assertion is true for $s+1$, and consider stage s, $1 \leq s < S$. For $2 \leq c \leq C$ we have

$$\begin{aligned}
&\Delta V_\beta(c+1,s) - \Delta V_\beta(c,s) \\
=&(\pi^*_\beta(c+1,s) - \Delta V_\beta(c+1,s+1)h(\beta_0 + \beta_1\pi^*_\beta(c+1,s)) + V_\beta(c+1,s+1) \\
&-(\pi^*_\beta(c,s) - \Delta V_\beta(c,s+1)h(\beta_0 + \beta_1\pi^*_\beta(c,s)) - V_\beta(c,s+1) \\
&-(\pi^*_\beta(c,s) - \Delta V_\beta(c,s+1)h(\beta_0 + \beta_1\pi^*_\beta(c,s)) - V_\beta(c,s+1) \\
&+(\pi^*_\beta(c-1,s) - \Delta V_\beta(c-1,s+1)h(\beta_0 + \beta_1\pi^*_\beta(c-1,s)) + V_\beta(c-1,s+1) \\
\leq&(1 - h(\beta_0 + \beta_1\pi^*_\beta(c+1,s)))(\Delta V_\beta(c+1,s+1) - \Delta V_\beta(c,s+1)) \\
&+h(\beta_0 + \beta_1\pi^*_\beta(c-1,s)))(\Delta V_\beta(c,s+1) - \Delta V_\beta(c-1,s+1)) \\
\leq&0,
\end{aligned}$$

using the optimality of $\pi^*_\beta(c,s)$ and the induction hypothesis, and for $c=1$ we have

$$\begin{aligned}
&\Delta V_\beta(2,s)-\Delta V_\beta(1,s)\\
\leq&(\pi^*_\beta(2,s)-\Delta V_\beta(2,s+1)h(\beta_0+\beta_1\pi^*_\beta(2,s))+V_\beta(2,s+1)\\
-&(\pi^*_\beta(1,s)-\Delta V_\beta(1,s+1)h(\beta_0+\beta_1\pi^*_\beta(1,s))-V_\beta(1,s+1)\\
\leq&(1-h(\beta_0+\beta_1\pi^*_\beta(2,s)))(\Delta V_\beta(2,s+1)-\Delta V_\beta(1,s+1))\leq 0,
\end{aligned}$$

using the induction hypothesis.

(iv) The proof is by backward induction on s.

First observe that for all fixed $p\in[p_l,p_h]$ and $0\leq a\leq p$,

$$\frac{\partial}{\partial\beta_0}\left\{ph(\beta_0+\beta_1p)+a(1-h(\beta_0+\beta_1p))\right\}=(p-a)\dot{h}(\beta_0+\beta_1p)\geq 0 \tag{EC.29}$$

and

$$\frac{\partial}{\partial\beta_1}\left\{ph(\beta_0+\beta_1p)+a(1-h(\beta_0+\beta_1p))\right\}=p(p-a)\dot{h}(\beta_0+\beta_1p)\geq 0. \tag{EC.30}$$

The optimality of $\pi^*_\beta(1,S)$, together with (EC.29) and (EC.30) applied to $p=\pi^*_\beta(1,S)$ and $a=0$, implies for any $\beta'=(\beta'_0,\beta'_1)\in B$ with $\beta'_0\leq\beta_0$ and $\beta'_1\leq\beta_1$,

$$\begin{aligned}
V_\beta(1,S)=&\pi^*_\beta(1,S)h(\beta_0+\beta_1\pi^*_\beta(1,S))\geq\pi^*_{\beta'}(1,S)h(\beta_0+\beta_1\pi^*_{\beta'}(1,S))\\
\geq&\pi^*_{\beta'}(1,S)h(\beta'_0+\beta'_1\pi^*_{\beta'}(1,S))=V_{\beta'}(1,S).
\end{aligned}$$

Let $1\leq s<S$, and assume the assertion is true for $s+1,\ldots,S$. Since $\Delta V_\beta(1,s+1)=V_\beta(1,s+1)\in(p_l,p_h)$ by (ii), it follows that $\pi^*_\beta(1,s)\geq V_\beta(1,s+1)$. The optimality of $\pi^*_\beta(1,s)$ and the induction hypothesis, together with equations (EC.29) and (EC.30) applied to $p=\pi^*_\beta(1,s)$ and $a=V_\beta(1,s+1)$, imply for any $\beta'=(\beta'_0,\beta'_1)\in B$ with $\beta'_0\leq\beta_0$ and $\beta'_1\leq\beta_1$,

$$\begin{aligned}
V_\beta(1,s)=&\pi^*_\beta(1,s)h(\beta_0+\beta_1\pi^*_\beta(1,s))+V_\beta(1,s+1)\cdot(1-h(\beta_u+\beta_u\pi^*_\beta(1,s)))\\
\geq&\pi^*_{\beta'}(1,s)h(\beta_0+\beta_1\pi^*_{\beta'}(1,s))+V_{\beta'}(1,s+1)\cdot(1-h(\beta_0+\beta_1\pi^*_{\beta'}(1,s)))\\
\geq&\pi^*_{\beta'}(1,s)h(\beta'_0+\beta'_1\pi^*_{\beta'}(1,s))+V_{\beta'}(1,s+1)\cdot(1-h(\beta'_0+\beta'_1\pi^*_{\beta'}(1,s)))\\
=&V_{\beta'}(1,s).
\end{aligned}$$

(v) For all $1\leq s<S$ and $1\leq c\leq C$, we have by (repeated) application of (i) and (ii),

$$0\leq\Delta V_\beta(c,s+1)\leq\Delta V_\beta(1,s+1)\leq\Delta V_\beta(1,2).$$

Let $p^*_{a,\beta}$ be as in Lemma EC.1. Since

$$\pi^*_\beta(C,S)=p^*_{0,\beta},\quad \pi^*_\beta(c,s)=p^*_{\Delta V_\beta(c,s+1),\beta},\quad \text{and } \pi^*_\beta(1,1)=p^*_{\Delta V_\beta(1,2),\beta},$$

it follows from Lemma EC.1(ii) that $\pi^*_\beta(C,S) \le \pi^*_\beta(c,s) \le \pi^*_\beta(1,1)$.

For $s = S$, note that $\pi^*_\beta(C,S) = \pi^*_\beta(c,s) = \pi^*_\beta(1,S) = p^*_{0,\beta} \le p^*_{\Delta V_\beta(1,2)} = \pi^*_\beta(1,1)$, for all $1 \le c \le C$.

(vi) Let $1 \le c \le C$ and $1 \le s \le S$. Since $\pi^*_\beta(c,s) = p^*_{\Delta V_\beta(c,s+1),\beta}$ and $\pi^*_\beta(c,s) \in [\pi^*_\beta(C,S), \pi^*_\beta(1,1)] \subset (p_l, p_h)$ (Lemma EC.2(v) and equation (4)), we have $(\Delta V_\beta(c,s+1),\beta) \in \mathcal{U}_{AB}$. The continuity assertion then follows from Lemma EC.1(ii).

Proof of Lemma EC.3

For any 2×2 positive definite matrix A with eigenvalues $0 < \lambda_1 \le \lambda_2$, we have $\lambda_2 \le \lambda_1 + \lambda_2 = \mathrm{tr}(A)$, $\det(A) = \lambda_1\lambda_2$, and consequentially $\lambda_1 = \det(A)/\lambda_2 \ge \det(A)/\mathrm{tr}(A)$. For $a, b \le p_h$ we thus have

$$\lambda_{\min}\begin{pmatrix} 2 & a+b \\ a+b & a^2+b^2 \end{pmatrix} \ge \frac{2a^2+2b^2-(a+b)^2}{2+a^2+b^2} \ge \frac{(a-b)^2}{2(1+p_h^2)}.$$

Since $\lambda_{\min}(P_t) \ge \lambda_{\min}(P_r) + \lambda_{\min}(P_{r'})$ for all $r, r', t \in \mathbb{N}$ with $r + r' = t$ (Bhatia 1997, Corollary III.2.2, page 63), we have

$$\begin{aligned}
\lambda_{\min}(P_{kS}) &\ge \lambda_{\min}(P_{(k-1)S}) + \lambda_{\min}\left(\sum_{1\le i\le S, i\notin\{s,s'\}} \begin{pmatrix} 1 \\ p_{i+(k-1)S} \end{pmatrix} (1, p_{i+(k-1)S}) \mathbf{1}_{c_{i+(k-1)S}>0}\right) \\
&\quad + \lambda_{\min}\left(\begin{pmatrix} 1 \\ p_{s+(k-1)S} \end{pmatrix} (1, p_{s+(k-1)S}) \mathbf{1}_{c_{s+(k-1)S}>0} + \begin{pmatrix} 1 \\ p_{s'+(k-1)S} \end{pmatrix} (1, p_{s'+(k-1)S}) \mathbf{1}_{c_{s'+(k-1)S}>0}\right) \\
&\ge \lambda_{\min}(P_{(k-1)S}) + \frac{(p_{s+(k-1)S} - p_{s'+(k-1)S})^2}{2(1+p_h^2)} \\
&\ge \lambda_{\min}(P_{(k-1)S}) + \frac{\delta^2}{2(1+p_h^2)}.
\end{aligned}$$