Memorandum 2029 (December 2013). ISSN 1874-4850. Available from: http://www.math.utwente.nl/publications
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Dynamic Pricing and Learning with Finite Inventories

Arnoud den Boer[1], Bert Zwart[2,3]

[1]University of Twente, P.O. Box 217, 7500 AE Enschede

[2]Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG Amsterdam

[3]VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam

July 7, 2013

Abstract

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 propose a pricing strategy for this problem, and show that the Regret - which is the expected revenue loss due to not using the optimal prices - after T selling seasons is $O(\log^2(T))$. Apart from a small modification, our pricing strategy is a certainty equivalent pricing strategy, which means that at each moment, the price is chosen that is optimal w.r.t. the current parameter estimates. The good performance of our strategy is caused by an endogenous-learning property: using a pricing policy that is optimal w.r.t. a certain parameter sufficiently close to the optimal one, leads to a.s. convergence of the parameter estimates to the true, unknown parameter. We also show an instance in which the regret for all pricing policies grows as $\log(T)$. This shows that our upper bound on the growth rate of the regret is close to the best achievable growth rate.

1 Introduction

1.1 Introduction, Motivation, Literature

The emergence of Internet as a sales channel has made it very easy for companies to experiment with selling prices. Where in the past costs and effort were needed to change prices, for example by issuing a new catalogue or replacing price tags, and consequently prices where fixed for longer periods of time, nowadays a webshop can adapt their 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 who sells limited amounts of products during finite selling periods, after which all unsold products perish. Examples of products with this property are flight tickets, hotel rooms, car rental reservations, and concert tickets. Various dynamic pricing models are already applied in these branches (see Talluri and van Ryzin, 2004). Other products that fall in this framework but for which dynamic pricing is not (yet) commonplace, are newspapers, magazines, and food at a grocery store. The emergence of digital price tags however may change this in the near future, see Kalyanam et al. (2006).

An important insight from the literature on dynamic pricing is that the optimal selling price of these type of products depends on the remaining inventory and the length of the remaining selling period, 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 remaining length of the selling period. 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 exactly 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.

Since then, a considerable number of studies on this subject have appeared, most of which are reviewed in Araman and Caldentey (2011). We also mention the related studies by Kleinberg and Leighton (2003), Broder and Rusmevichientong (2012), den Boer and Zwart (2010), Harrison et al. (2012), who consider dynamic pricing in a slightly different setting, namely with infinite inventory. This significantly changes the structure of the learning behavior, as further discussed in Section 4.

A common feature of the studies on dynamic pricing with finite inventory is the restriction to a single selling season during which learning and optimization takes place. To assess the performance of proposed pricing strategies, one often considers an asymptotic regime where the demand rate and the initial amount of inventory grow to infinity (e.g. Besbes and Zeevi, 2009, Wang et al., 2011). Such an asymptotic regime may have practical value if demand, initial inventory, and the length of the selling season are relatively large. In many situations, however, this is not the case. For example, in the hotel rooms industry (Talluri and van Ryzin, 2004, section 10.2, Weatherford and Kimes, 2003), a product may be modeled as a combination of arrival date and length-of-stay. Different products may have different, overlapping selling periods, and similar demand characteristics. It would therefore be unwise to learn the consumer behavior for each product and selling period separately. In addition, the average demand, initial capacity and length of a selling period may be quite low, which makes this particular asymptotic regime not a suitable setting to study the performance of pricing strategies.

These considerations motivate the present study dynamic pricing of perishable products with finite initial inventory, during multiple *consecutive* selling seasons of finite and fixed duration.

1.2 Contributions

We consider a parametric demand model which includes practically all demand function that are used in practice. The uncertainty in the demand is modeled by unknown parameters, which can be estimated from historical sales data using maximum quasi-likelihood estimation.

We propose a pricing strategy that is structurally very intuitive, and easy to understand by price managers. At every moment where prices can be changed, the firm calculates a statistical estimate of the unknown parameter. Subsequently, the price is determined that would be optimal if this parameter estimate were correct, and this price is used until the next decision moment. In other words, at each decision moment the firm acts as if being 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, Harrison et al., 2011, den Boer and Zwart, 2010). In these two examples, passive learning is not sufficient to learn the parameters: the decision maker should actively account for the fact that he is not only optimizing prices, but also tries to 'optimize' the learning process. This implies that sometimes decisions should be taken that seem suboptimal on a short term. In the dynamic pricing problem with infinite inventory, this can be accomplished by the controlled variance policy of den Boer and Zwart (2010) 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 (2010) 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 periods, 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 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 bound is considerably better 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 moreover, this bound can hardly be improved. We show an instance for which *any* pricing strategy has Regret(T) $\geq K \log(T)$, for some K independent of T and of the pricing strategy. This means that the upper bound $\log^2(T)$ on the regret is close to the best achievable growth rate $\log(T)$. In Section 7.3 we discuss the small gap between the lower and upper bound.

Thus, the regret, which can be interpreted as the 'cost for learning', behaves structurally different in these two models. This difference in qualitative behavior can be explained as follows. In the infinite inventory model, prices and parameter estimates can get stuck in what Harrison et al. (2012) call

an 'indeterminate equilibrium'. This means that for some values of the parameter estimates, the expected observed demand at the certainty equivalent price is equal to what the parameter estimates predict; in other words, the observations confirm the correctness of the (incorrect) parameter estimates. As a result, certainty equivalent control induces insufficient dispersion in the chosen selling prices to eventually learn the true value of the parameters.

Such cannot occur in the setting with finite inventories and finite selling seasons. An optimal price - optimal w.r.t. certain parameter estimates - is namely not a fixed number, but changes depending on the remaining inventory and the remaining length of the selling season. Thus, an optimal policy naturally induces endogenous price dispersion, and prices cannot get stuck in an 'indeterminate equilibrium'. On the contrary, the large amount of price dispersion implies that the unknown parameters are learned quite fast, and consequently that the Regret(T) is only $O(\log^2(T))$.

The main conceptual takeaway of our paper is that, in decision problems under uncertainty, a passive-learning 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 7.2, the idea is also applicable in other decision problems. 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, Harrison et al., 2011, den Boer and Zwart, 2010) it is known that active price experimentation is necessary to achieve optimal regret; myopic policies have suboptimal performance. 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., 2011), active price experimentation is also known to be necessary to achieve optimal performance. The reason is that, in this asymptotic regime, the amount of price dispersion induced by the myopic policy decreases to zero. We consider a different asymptotic regime in which changes in the marginal-value-of-inventory keeps inducing price dispersion in the asymptotic regime; as a result, no active price experimentation is necessary, and the myopic strategy performs very well.

Our work is also connected to the field of adaptive control in Markov decision problems (Hernández-Lerma, 1989, Kumar, 1985, chapter 12 of Kumar and Varaiya, 1986). An important feature that distinguishes our work from many previous literature in this area, is the following. Hernández-Lerma and Cavazos-Cadena (1990), Gordienko and Minjárez-Sosa (1998) assume that the "next" state x_{t+1} at period $t+1$ is determined by the "current" state x_t, action a_t, and a random component ξ_t. These random components are assumed to be *independent and identically distributed*. In our setting, the randomness in state transitions is completely determined by the demand realizations. These are neither identically distributed (their distribution depends on the chosen prices), nor independent (chosen prices may depend on all previously chosen prices and observed demand realizations, and, consequentially, demand in different time periods is not independent). In other literature, such as Altman and Shwartz (1991), unknown transition probabilities are estimated by the empirically observed relative frequencies. In our setting, all uncertainty is captured by an unknown parameter, and transition probabilities are estimated simultaneously. Furthermore, we

consider a compact continuous action space, in contrast to e.g. Burnetas and Katehakis (1997), Chang et al. (2005) who assume a finite action space, which links the adaptive control problem to the multi-armed bandit problem.

Summarizing, the contributions of this paper are as follows:

(i) We formulate the problem of dynamic pricing with finite inventories during multiple, consecutive selling seasons of finite duration, with parametric uncertainty in the demand function.

(ii) We propose a simple and intuitive pricing strategy, based on the idea of subsequently estimating the unknown parameters and choosing the selling price that would be optimal if this parameter estimate were correct.

(iii) We show that the problem satisfies an endogenous-learning property, which means that the use of policies that are optimal w.r.t. parameter estimates automatically induces a certain amount of price dispersion.

(iv) We prove that this leads to convergence of the parameter estimates to the true value, and we show $\text{Regret}(T) = O(\log^2(T))$.

(v) We provide an instance for which *any* pricing strategy has $\text{Regret}(T)$ that grows at least logarithmically in T, implying that the $O(\log^2(T))$ upper bound on the regret is close to the best achievable growth rate.

(vi) We provide numerical examples to illustrate our results, and discuss various extensions of our model.

1.3 Organization

The rest of this paper is organized as follows. Section 2 discusses the mathematical model, the structure of the demand distribution, the full-information optimal solution, and the regret measure. Section 3 shows how the unknown parameters of the model can be estimated, and contains a result concerning the speed at which parameter estimates converge to the true value. The endogenous-learning property of the system is described in Section 4. Our pricing strategy is introduced in Section 5.1, the upper bound $\text{Regret}(T) = O(\log^2(T))$ is shown in Section 5.2, and the $\log(T)$ lower bound in Section 5.3. Numerical illustrations of the pricing strategy and its performance are provided in Section 6. A discussion of the results and possible extensions of this paper is provided in Section 7. The mathematical proofs of the main results in this paper are contained in Section 8. A number of auxiliary results are formulated and proven in Section 9.

Notation The interior of a set $U \subset \mathbb{R}^n$ is denoted by $\text{int}(U)$. If v is a vector then $||v||$ denotes the Euclidean norm, and v^T the transpose. If A is an $m \times n$ matrix, $||A|| = \max_{x \in \mathbb{R}^n, ||x||=1} ||Ax||$ denotes the induced matrix norm of A, and $\lambda_{\min}(A)$ 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.

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, and introduce the regret as quality measure of pricing policies.

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. 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. The dynamics of $(c_t)_{t\in\mathbb{N}}$ are given by

$$\begin{aligned} c_t &= C, && \text{if } s_t = 1, \\ c_t &= \max\{c_{t-1} - d_{t-1}, 0\}, && \text{if } s_t \neq 1. \end{aligned}$$

The pricing decisions of the seller are allowed to depend on previous prices and demand realizations, but not on future ones. More precisely, for each $t \in \mathbb{N}$ we define 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 revenue collected in period t equals $p_t \min\{c_t, d_t\}$. The purpose of the seller is to find a pricing strategy ψ that maximizes the cumulative expected revenue earned after T selling seasons, $\sum_{i=1}^{TS} E_\psi[p_i \min\{d_i, c_i\}]$. 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 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.

To ensure existence and uniqueness of revenue-maximizing selling prices, we make a number of assumptions on h and β. First, we assume that $\beta^{(0)}$ lies in the interior of a compact set $B \subset \mathbb{R}^2$ known to the seller, and assume that $\beta_1 < 0$ for all $\beta \in B$. Second, we assume that h is three times continuously differentiable, log-concave, $h(\beta_0 + \beta_1 p) \in (0,1)$ for all $\beta \in B$ and $p \in [p_l, p_h]$, and the derivative $\dot{h}(z)$ of $h(z)$ is strictly positive. This last assumption, together with $\beta_1 < 0$ for all $\beta \in B$, implies that expected demand is decreasing in p, for all $\beta \in B$.

Write $r^* = \max_{p \in [p_l, p_h]} p \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p)$, and for $(a, \beta, p) \in \mathbb{R} \times B \times [p_l, p_h]$, define

$$g_{a,\beta}(p) = -(p-a)\beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)}.$$

We assume that $g_{a,\beta^{(0)}}(p_l) < 1$, $g_{a,\beta^{(0)}}(p_h) > 1$, and $g_{a,\beta^{(0)}}(p)$ is strictly increasing in p, for all $0 \le a \le r^*$. These conditions, which for $a = 0$ coincide with the assumptions in Lariviere (2006, page 602), ensure that the function which maps p to $(p-a)h(\beta_0^{(0)} + \beta_1^{(0)} p)$ has a unique maximizer in (p_l, p_h).

Practically all demand functions that are used in practice fit into our framework. Some examples (with appropriate conditions on B and $[p_l, p_h]$) are $h(z) = \exp(z)$, $h(z) = z$, 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. If action p is used in state (c,S), then with probability one a state transition $(c,s) \mapsto (C,1)$ occurs; the obtained reward 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 $1 \leq s \leq S$, where we write $V_\beta^\pi(c, S+1) = 0$ for all $0 \leq c \leq C$.

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 $1 \leq c \leq C$, $1 \leq s \leq S$, satisfy the following recursion:

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

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 $1 \leq s \leq S$. The price $\pi_\beta^*(0,s)$ can be chosen arbitrarily, since it has no effect on the reward.

The optimal average reward of the MDP is equal to $V_\beta(C,1)$, and the true optimal average reward is equal to $V_{\beta^{(0)}}(C,1)$.

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

$$\text{Regret}(\psi, T) = T \cdot V_{\beta^{(0)}}(C,1) - \sum_{i=1}^{TS} E[p_i \min\{d_i, c_i\}], \tag{4}$$

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 $\beta^{(0)}$. Also note that we calculate the regret over a number of selling seasons, and not over a number of time periods. The reason is that the optimal policy $\pi_{\beta^{(0)}}^*$ is optimized over an entire selling season, and not over each individual state of the underlying MDP: a price p_t may induce a higher instant reward in a certain state (c_t, s_t) than the optimal price $\pi_{\beta^{(0)}}^*(c_t, s_t)$. This effect is averaged out by looking at the optimal expected reward in an entire selling season.

For small T the optimal policy under incomplete information can in theory be calculated exactly, by solving a MDP with state-space that includes all possible demand realizations. This MDP however is computationally intractable for even moderate values of T. It is therefore common in the literature on dynamic pricing to study the asymptotic growth rate of Regret(T) as T grows large, and search for pricing strategies that have the lowest possible growth rate on the regret.

3 Parameter Estimation

3.1 Maximum-likelihood Estimation

The value of $\beta^{(0)}$ can be estimated with maximum-likelihood estimation. In particular, given a sample of prices $p_1, \ldots, p_t$ and demand realizations $d_1, \ldots d_t$, the log-likelihood function $L_t(\beta)$ equals

$$L_t(\beta) = \sum_{i=1}^{t} \log \left[h(\beta_0 + \beta_1 p_i)^{d_i} (1 - h(\beta_0 + \beta_1 p_i))^{1-d_i} \right].$$

The score function, the derivative of $L_t(\beta)$ with respect to β, equals

$$l_t(\beta) = \sum_{i=1}^{t} \frac{\dot{h}(\beta_0 + \beta_1 p_i)}{h(\beta_0 + \beta_1 p_i)(1 - h(\beta_0 + \beta_1 p_i))} \begin{pmatrix} 1 \\ p_i \end{pmatrix} (d_i - h(\beta_0 + \beta_1 p_i)). \tag{5}$$

We let $\hat{\beta}_t$ be a solution to $l_t(\beta) = 0$. If no solution exists, we define $\hat{\beta}_t = \beta^{(1)}$, for some predefined $\beta^{(1)} \in B$. If a solution to $l_t(\beta) = 0$ exists but lies outside B, we define $\hat{\beta}_t$ as the projection of this solution on B. For most choices of h there is no explicit formula for the solution of $l_t(\beta) = 0$, and numerical methods have to be deployed to calculate it.

3.2 Convergence Rates of Parameter Estimates

Understanding the asymptotic behavior of the maximum quasi-likelihood estimate $\hat{\beta}_t$, in particular the speed at which it converges to $\beta^{(0)}$, is important to study the performance of pricing strategies. We here quote a result from den Boer and Zwart (2011) about these convergence rates; in Section 5.2, 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

$$P_t = \begin{pmatrix} t & \sum_{i=1}^{t} p_i \\ \sum_{i=1}^{t} p_i & \sum_{i=1}^{t} p_i^2 \end{pmatrix}, \quad (t \in \mathbb{N}), \tag{6}$$

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 the last-time random variable

$$T_\rho = \sup \left\{ t \in \mathbb{N} \mid \text{there is no } \beta \in B \text{ with } \left|\left| \beta - \beta^{(0)} \right|\right| \leq \rho \text{ and } l_t(\beta) = 0 \right\}, \tag{7}$$

for $\rho > 0$.

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*

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

This proposition follows directly from Theorem 1, Theorem 2, and Remark 2 in den Boer and Zwart (2011).

4 Main Result: a Case of Endogenous Learning

Proposition 1 shows how the growth rate of $\lambda_{\min}(P_t)$ influences the speed at which the parameter estimates converge to the true value. 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 that prescribed by $\pi^*_{\beta^{(0)}}$. 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 $E\left[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t>T_\rho}\right] = O\left(\log(t)/t\right)$.

This result can be interpreted as the system having an *endogenous-learning* property: the unknown parameters are learned very fast when a policy close to the optimal policy is used. This is the main takeaway of this paper. In Section 5.2 this property will be exploited to prove upper bounds on our proposed pricing strategy.

Theorem 1. *Let $1 < C < S$ and $k \in \mathbb{N}$. There exist a constant $v_0 > 0$, and an open neighborhood $\mathcal{U} \subset B$ containing $\beta^{(0)}$, 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 $\beta(1), \ldots, \beta(S) \in \mathcal{U}$, then

$$\min_{1 \leq s,s' \leq S} |p_{s+(k-1)S} - p_{s'+(k-1)S}| \geq v_0/2. \tag{8}$$

and

$$\lambda_{min}(P_{kS}) - \lambda_{min}(P_{(k-1)S}) \geq \frac{1}{8} v_0^2 (1 + p_h^2)^{-1}, \tag{9}$$

The condition $1 < C < S$ in Theorem 1 makes sure that price dispersion occurs during a selling season. If $C = 1$ then the firm may go out-of-stock in the first period of the selling season, which implies that only a single price is charged during that selling season and thus no price dispersion occurs. The $C \geq S$ case 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. As a result, there is no interaction between individual items, and the pricing problem is equivalent to S repetitions of the pricing problem with $C = 1$, $S = 1$, which means that 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 (2010), Harrison et al. (2012) 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. For $1 < C < S$, Section 7.4 discusses in more detail the effect of C and S on the amount of price dispersion.

In Remark 1, stated directly after the proof of Theorem 1, we compute an explicit, positive lower bound on v_0.

The proof of Theorem 1 makes use of a number of auxiliary lemmas, which are formulated and proven in Section 9.

5 Pricing Strategy and Performance Bounds

5.1 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 $p_{t+1} \in [p_l, p_h]$ arbitrary. 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 $SS_i = SS_{t+1}$, 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 the available parameter estimates $\hat{\beta}_t$, 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.

5.2 Upper Bound on the Regret

The endogenous-learning property described in Section 4 implies that if $\hat{\beta}_t$ is sufficiently close to $\beta^{(0)}$ and ϵ is sufficiently small, then I) 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 instant revenue, and do not reckon with the objective of optimizing the quality of the parameter estimates $\hat{\beta}_t$. The endogenous-learning property makes sure that learning the parameter values happens on the fly, without active effort.

As a result, the parameter estimates converge quickly to their true values, and the pricing decisions quickly to the optimal pricing decisions. The following theorem shows that the regret of the strategy $\Phi(\epsilon)$ is $O(\log^2(T))$ in the number of selling seasons 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 $\left|\left|\hat{\beta}_t - \beta^{(0)}\right|\right|$ on the regret. The endogenous-learning property of Theorem 1, combined with the "small t correction" in I) of the description of $\Phi(\epsilon)$, implies that $\lambda_{\min}(P_t)$ grows linearly in t. Using Proposition 1 this enables us to prove the $O(\log^2(T))$ bound on the regret.

In Remark 1, stated directly after the proof of Theorem 1, we compute an explicit, positive lower bound on v_0.

5.3 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 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 $C = 1$, $S = 2$, h the identity function, $[p_l, p_h] = [3/8, 17/16]$, and let $B = [5/8, 6/8] \times [-3/4, -9/16]$. Then, for all pricing strategies ψ and all $T \in \mathbb{N}$, we have*

$$\sup_{\beta \in B} \textit{Regret}(\psi, T) \geq \frac{9245}{25600(80 + 64\pi^2)} \log(T).$$

The theorem is proven by applying a generalization of the van Trees inequality, (Gill and Levit, 1995), along the same lines of (Lemma 4.6 Broder and Rusmevichientong, 2012). Note that the goal of the theorem is not to provide the best constant before the $\log(T)$ term, but to show the qualitative result that there is no pricing strategy with $\text{Regret}(T) = o(\log(T))$.

6 Numerical Illustration

To illustrate the analytical results that we have derived, we provide a number of numerical illustrations. We first offer a simple instance that illustrate strong consistency of the parameter estimates and convergence of the relative regret to zero. We also briefly consider the 'gap' between the upper bound $O(\log^2(T))$ 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 look at the effect on the regret. In the last illustration we fix C but vary S, to look at the effect of the length of the selling season on the regret.

A. 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

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

To show some light on the growth rate of the regret, we scale in Figure 2 the regret by a $\log(n)$ and a $\log^2(n)$ factor. Theorem 2 entails that $\text{Regret}(n)/\log^2(n)$ is bounded, which accords with the righthand plot in Figure 2. However, Theorem 3 suggests that the $O(\log^2(n))$ bound may be too conservative, and that in fact the regret may grow logarithmically (cf. the discussion in Section 7.3). The lefthand plot of Figure 2 shows the regret scaled by a log-factor. This picture does not strongly support the assertion that $\text{Regret}(n)/\log(n)$ 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.

B. 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 example A., 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.

The fourth column of Table 1 suggests that the relative regret is not monotone in C, but is minimal

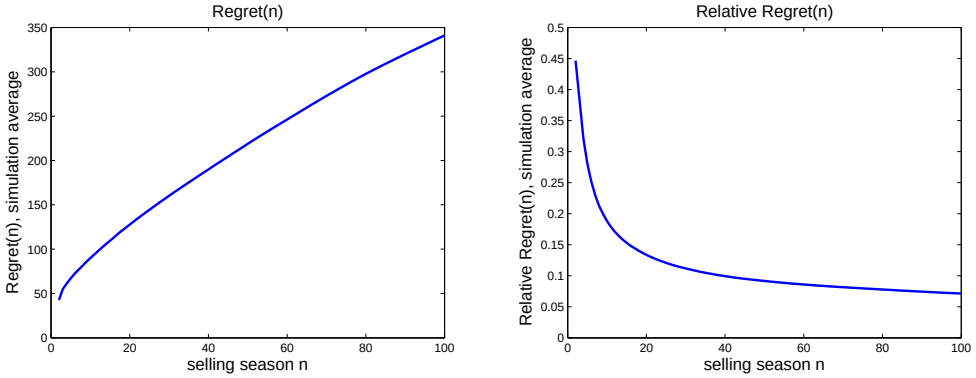

Figure 1: Simulation average of regret and relative regret

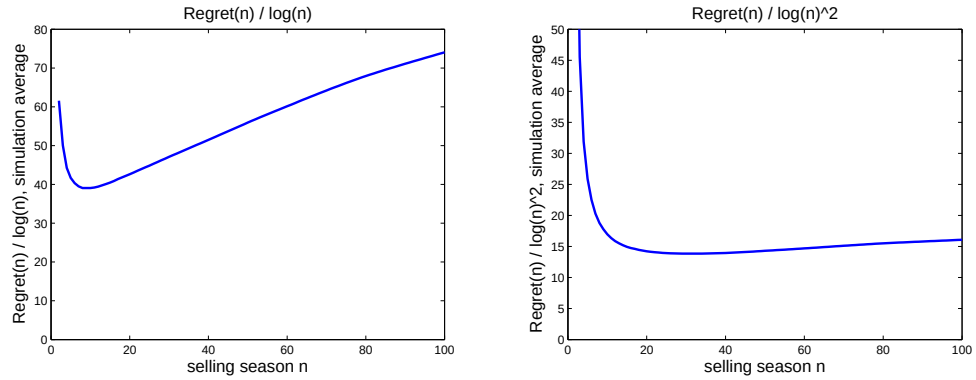

Figure 2: Simulation average of regret, scaled by $\log(n)$ and $\log^2(n)$.

Table 1: Simulation output for various choices of C

C	$V_{\beta^{(0)}}(C,1)$	Regret(100)	Relative regret(100)	$\left\|\hat{\beta}_{1000} - \beta^{(0)}\right\|$
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

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.

C. Different length of selling season

In our third numerical illustration we consider the same instance as in A. and B., 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)	$\left\|\hat{\beta}_{100S} - \beta^{(0)}\right\|$
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.

7 Discussion

7.1 Extensions to Other Demand Models

To facilitate analysis we impose some assumptions on the demand function: it depends on only two unknown parameters (β_0, β_1), and is of the form $E[D(p)] = h(\beta_0 + \beta_1 p)$. Conceptually our results do not hinge on these assumptions, and may still hold if one considers a demand model that involves more than two unknown parameters, or where demand depends on the stage in the selling season.

As an example, suppose that $E[D(p)] = h(\beta_0 + \beta_1 p + \beta_2 p^2 + \ldots + \beta_m p^m)$, for some $m \in \mathbb{N}$ and unknown parameters $(\beta_0, \ldots, \beta_m)$. Similarly as in Section 2.3 one can define the optimal full-information solution $\pi^*_\beta(c, s)$, with $h(\beta_0 + \beta_1 p)$ in all relevant equations replaced by $h(\beta_0 + \beta_1 p + \beta_2 p^2 + \ldots + \beta_m p^m)$. The design matrix (6) is then equal to the $(m+1) \times (m+1)$ matrix

$$P_t = \sum_{i=1}^{t} (1, p_i, p_i^2, \ldots, p_i^m)^T (1, p_i, p_i^2, \ldots, p_i^m).$$

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 as a result, the maximum likelihood estimate of β converge a.s. to the true value.

For this particular demand model, the endogenous-learning property can be guaranteed if a.s. $m+1$ distinct prices $p_1, \ldots, p_{m+1}$ are used during a selling season, under policy π^*_β. (Compare this to the proof of Theorem 1, where we show that at least *two* different prices occur a.s. during a selling season). If this is the case, then

$$\begin{aligned}
&\lambda_{\min}(P_{Sk}) - \lambda_{\min}(P_{S(k-1)}) \geq \lambda_{\min}\Big(\sum_{i=1}^{m+1} (1, p_i, \ldots, p_i^m)^T (1, p_i, \ldots, p_i^m) \Big) \\
\geq &\frac{\det\big(\sum_{i=1}^{m+1} (1, p_i, \ldots, p_i^m)^T (1, p_i, \ldots, p_i^m)\big)}{\mathrm{tr}\big(\sum_{i=1}^{m+1} (1, p_i, \ldots, p_i^m)^T (1, p_i, \ldots, p_i^m)\big)^m} \geq \frac{\prod_{1 \leq i < j \leq m+1} (p_i - p_j)^2}{\big(\sup_{p \in \mathcal{P}} \sum_{i=0}^{m} p^{2i}\big)^m} > 0,
\end{aligned}$$

which implies the endogenous-learning property.

Another example is $E[D(p, s)] = h(\beta_0 + \beta_1 p + \beta_2 s)$. Here the demand explicitly depends on the stage s of the selling season, which models changing demand during a selling season. Again, similarly as in Section 2.3 one can define the optimal full-information solution $\pi^*_\beta(c, s)$ of the pricing problem, 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 (6) is equal to

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

To prove an endogenous-learning property similar to Theorem 1, one should show that, for β close to $\beta^{(0)}$, using the policy π^*_β in selling season k implies that $\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S} > \epsilon$, for all $k \in \mathbb{N}$ and some $\epsilon > 0$ independent of β. This again implies strong consistency of the maximum likelihood estimate of β.

In this demand model, a sufficient condition for the endogenous-learning property to hold is if there are prices p_1, p_2, p_3 used in stage s_1, s_2, s_3, respectively, such that $(p_3(s_2 - s_1) + p_2(s_3 - s_1) + p_1(s_3 - s_2))^2 > 0$. (This condition ensures that the vectors $\{(1, p_i, s_i)^T \mid i = 1, 2, 3\}$ are

linearly independent). If this holds, then

$$\lambda_{\min}(P_{Sk}) - \lambda_{\min}(P_{S(k-1)}) \geq \lambda_{\min}\Big(\sum_{i=1}^{3}(1,p_i,s_i)^T(1,p_i,s_i)\Big)$$
$$\geq \frac{\det\big(\sum_{i=1}^{3}(1,p_i,s_i)^T(1,p_i,s_i)\big)}{\mathrm{tr}\big(\sum_{i=1}^{3}(1,p_i,s_i)^T(1,p_i,s_i)\big)^2} \geq \frac{(p_3(s_2-s_1)+p_2(s_3-s_1)+p_1(s_3-s_2))^2}{(3+3S^2+3\sup_{p\in\mathcal{P}}p^2)^2} > 0,$$

which implies the endogenous-learning property.

We believe that for these alternative demand models an endogenous-learning property can be shown. Formally proving the needed price-dispersion conditions can however be somewhat tedious; the proof of Theorem 1 for the simpler demand model $E[D(p)] = h(\beta_0 + \beta_1 p)$ is already quite delicate. Numerical simulations show that many different prices are occur during a selling season, and not only two different prices as guaranteed by Theorem 1. This suggest that the endogenous-learning property may also holds in the two demand models discussed above. Formalizing this property for these (and other) demand models is an interesting direction for future research.

7.2 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, nonparametric 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

increase) typically presumes a minimum amount of variation/dispersion in the controls; see e.g. Skouras (2000), 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 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 multi-armed bandit problems).

7.3 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 $E\left[\left|\left|\hat{\beta}_t - \beta^{(0)}\right|\right|^2 \mathbf{1}_{t>T_\rho}\right] = O(\log(t)/L(t))$ of Proposition 5. This $\log(t)$ term can be traced back to Proposition 2 of den Boer and Zwart (2011), 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 (2011) 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 $\log^2(T)$. For a discussion on this topic in a related pricing-and-learning problem, we refer to Section 5.2 of den Boer (2011).

7.4 Effect of C and S on Price Dispersion

The results from section 6, in particular example B, 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 n 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 $t = 1, \ldots, 1000$.

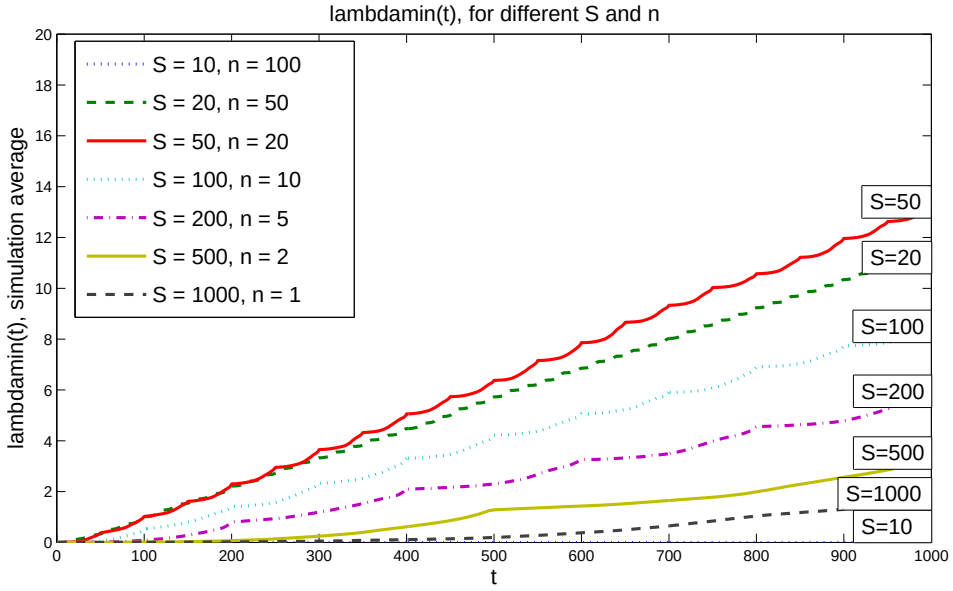

Figure 3: Price dispersion, for different values of S and n

Figure 3 shows the simulation average of $\lambda_{\min}(P_t)$ for $t = 1, \ldots, 1000$, for the different values of (S, n). 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 n.

This magnitude effects 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 $\left|\left|\hat{\beta}_t - \beta^{(0)}\right|\right|$, 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 $\left|\left|\hat{\beta}_t - \beta^{(0)}\right|\right|$ 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 $C = S$, which means that active price experimentation is necessary (see our comments following Theorem 1).

Table 3 lists the values of $\lambda_{\min}(P_t)$ at $t = 1000$, the end of the time horizon. It shows that the amount of price dispersion is not monotone in S: the largest growth rate is achieved at the experiment with $S = 50$, $n = 20$; for S larger than 50 it is decreasing in S, and for S smaller than

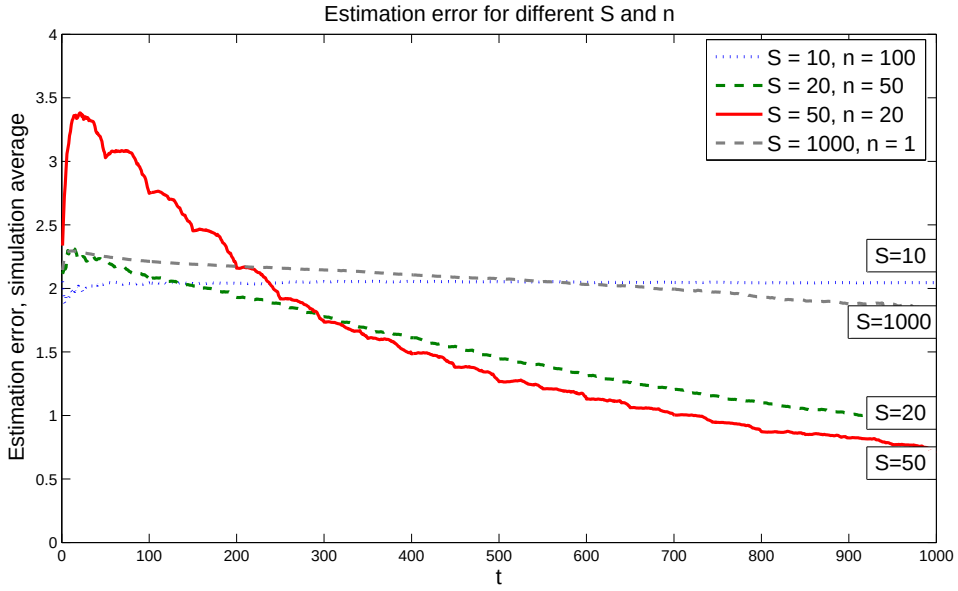

Figure 4: Estimation error $\left|\left|\hat{\beta}_t - \beta^{(0)}\right|\right|$, for different values of S and n

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.

Table 3: Price dispersion, for different values of S and n

S	n	$\lambda_{\min}(P_{1000})$
10	100	0.000
20	50	11.43
50	20	13.31
100	10	8.629
200	5	5.891
500	2	3.370
1000	1	2.003

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. (2011). 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. (2011), active price experimentation is necessary, whereas in our setting a myopic policy works well.

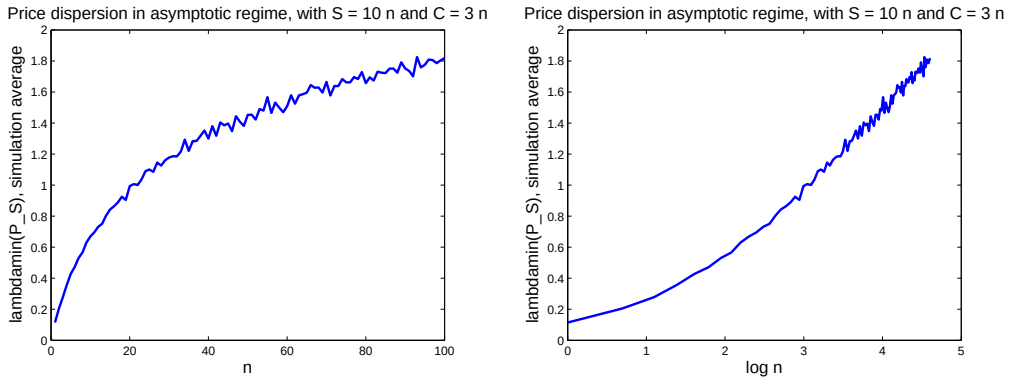

Figure 5: $\lambda_{\min}(P_S)$, for $S = 10n$, $C = 3n$

8 Proofs

In this section we prove the main theorems of the paper. The proofs frequently refer to a number of auxiliary lemmas, which are formulated and proven in Section 9.

Proof of Theorem 1

Consider the k-th selling season, and write $c(1) = c_{1+(k-1)S}$, $c(2) = c_{2+(k-1)S}$, ..., $c(S) = c_{kS}$. We show that there is a $v_0 > 0$ 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 $1 \leq s, s' \leq S$ with $|\pi^*_{\beta^{(0)}}(c(s), s) - \pi^*_{\beta^{(0)}}(c(s'), s')| > v_0$. Since π^*_β is continuous in β around $\beta^{(0)}$ (Lemma 3), this implies that there is an open neighborhood $\mathcal{U} \subset U_B$ around $\beta^{(0)}$ 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 $(\beta(1), \ldots, \beta(S)) \in \mathcal{U}$, then there are $1 \leq s, s' \leq S$ such that $|\pi^*_{\beta(s)}(c(s), s) - \pi^*_{\beta(s')}(c(s'), s')| > v_0/2$. This proves (8). Equation (9) follows by application of Lemma 4.

Define

$$\triangleleft = \{(c, s) \mid S + 1 - C \leq s \leq S, S + 1 - s \leq c \leq C\}. \tag{10}$$

See Figure 6 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 $(c(s), s)_{1 \leq s \leq S}$ may or may not hit $\triangleleft$. We show that in both cases, at least two different selling prices occur on the path $(c(s), s)_{1 \leq s \leq S}$.

Case 1. The path $(c(s), s)_{1 \leq s \leq S}$ hits $\triangleleft$. Then there is an s such that $(c(s), s) \in \triangleleft$ and $(c(s), s-1) \notin \triangleleft$. In particular, $(c(s), s-1) \in (L\triangleleft) = \{(1, S-1), (2, S-2), \ldots, (C-1, S-C+1), (C, S-C)\}$, where $(L\triangleleft)$ denotes the points (c, s) immediately left to $\triangleleft$ in Figure 6. We show that the sets $\triangleleft$ and $(L\triangleleft)$ satisfy the following properties:

(P.1) If $(c, s) \in \triangleleft$ then $\Delta V_{\beta^{(0)}}(c, s+1) = 0$, $\pi^*_{\beta^{(0)}}(c, s) = \arg\max_{p \in [p_l, p_h]} ph(\beta_0^{(0)} + \beta_1^{(0)} p)$, and $V_{\beta^{(0)}}(c, s) = (S - s + 1) \cdot V_{\beta^{(0)}}(1, S)$.

(P.2) If $(c, s) \in (L\triangleleft)$, then $\pi^*_{\beta^{(0)}}(c, s) \neq \pi^*_{\beta^{(0)}}(c, s+1)$ and $\Delta V_{\beta^{(0)}}(c+1, S-c) \neq 0$ (provided

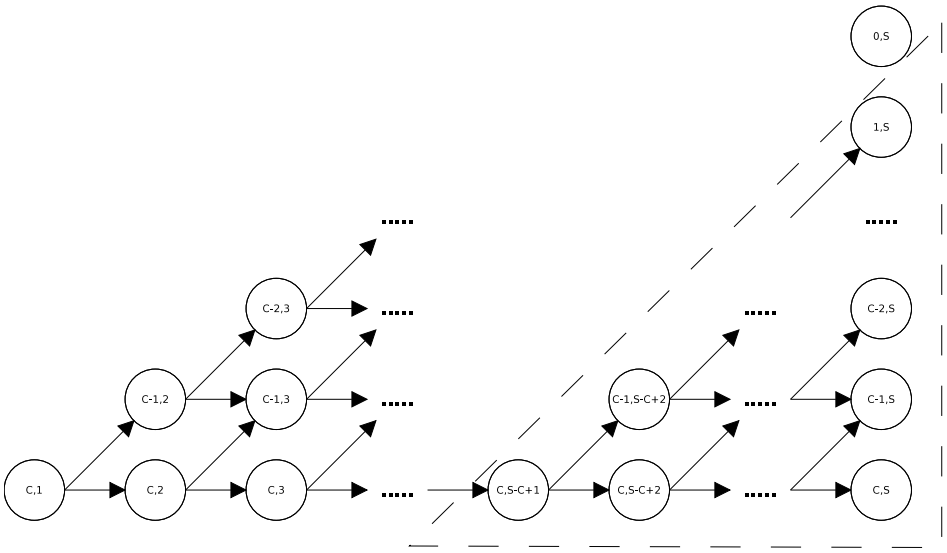

Figure 6: Schematic picture of $\triangleleft$

$c < C$). More precisely, $\Delta V_{\beta^{(0)}}(1,S) = f^*_{0,\beta^{(0)}}$ and

$$\Delta V_{\beta^{(0)}}(c+1,S-c) = f^*_{0,\beta^{(0)}} - f^*_{\Delta V_{\beta^{(0)}}(c,S-c+1),\beta^{(0)}} > 0, \tag{11}$$

for $1 < c < C$, where f and p^* are as in Lemma 2, and we shorthand write $f^*_{a,\beta^{(0)}} = f_{a,\beta^{(0)}}(p^*_{a,\beta^{(0)}})$.

Property (P.2) implies that a price change occurs when the path $(c(s),s)_{1\le s\le S}$ hits $\triangleleft$. Property (P.1) is used in the proof of property (P.2).

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 $\Delta V_{\beta^{(0)}}(c,s+1) = V_{\beta^{(0)}}(c,s+1) - V_{\beta^{(0)}}(c-1,s+1) = 0$, $\pi^*_{\beta^{(0)}}(c,s) = \arg\max_{p\in[p_l,p_h]} ph(\beta_0^{(0)}+\beta_1^{(0)}p)$ and $V_{\beta^{(0)}}(c,s) = \max_{p\in[p_l,p_h]} ph(\beta_0^{(0)}+\beta_1^{(0)}p)+V_{\beta^{(0)}}(c,s+1) = (S-s+1)\cdot V_{\beta^{(0)}}(1,S)$, 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 $(c,s) = (1,S-1)$. Since $\Delta V_{\beta^{(0)}}(1,S) = V_{\beta^{(0)}}(1,S) = f^*_{0,\beta^{(0)}} > 0$, Lemma 2 and equation (3) imply $\pi^*_{\beta^{(0)}}(1,S-1) \neq \pi^*_{\beta^{(0)}}(1,S)$. In addition,

$$V_{\beta^{(0)}}(2,S-1) = \max_{p\in[p_l,p_h]} \Big((p-\Delta V_{\beta^{(0)}}(2,S))h(\beta_0^{(0)}+\beta_1^{(0)}p) + V_{\beta^{(0)}}(2,S)\Big), \tag{12}$$

$$V_{\beta^{(0)}}(1,S-1) = \max_{p\in[p_l,p_h]} \Big((p-\Delta V_{\beta^{(0)}}(1,S))h(\beta_0^{(0)}+\beta_1^{(0)}p) + V_{\beta^{(0)}}(1,S)\Big). \tag{13}$$

Property $(P.1)$ implies $V_{\beta^{(0)}}(2,S) = V_{\beta^{(0)}}(1,S)$ and $\Delta V_{\beta^{(0)}}(2,S) = 0$. Furthermore, $\Delta V_{\beta^{(0)}}(1,S) = V_{\beta^{(0)}}(1,S) > 0$, and thus by Lemma 2,

$$\Delta V_{\beta^{(0)}}(2,S-1) = V_{\beta^{(0)}}(2,S-1) - V_{\beta^{(0)}}(1,S-1) = f^*_{0,\beta^{(0)}} - f^*_{\Delta V_{\beta^{(0)}}(1,S),\beta^{(0)}} > 0,$$

since $f^*_{a,\beta^{(0)}}$ is strictly decreasing in a.

Let $c > 1$ and $(c, s) \in (L\triangleleft)$. Then $(c, s) = (c, S - c)$. By the induction hypothesis we have $\Delta V_{\beta^{(0)}}(c, S - c + 1) > 0$, and thus

$$\pi^*_{\beta^{(0)}}(c, S - c) = \arg\max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(c, S - c + 1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) \tag{14}$$

$$\neq \arg\max_{p\in[p_l,p_h]} ph(\beta_0^{(0)} + \beta_1^{(0)} p) = \pi^*_{\beta^{(0)}}(c, S - c + 1), \tag{15}$$

where we used Lemma 2 for the first inequality, and (P.1) for the second equality. It remains to show

$$\Delta V_{\beta^{(0)}}(c + 1, S - c) = f^*_{0,\beta^{(0)}} - f^*_{\Delta V_{\beta^{(0)}}(c,S-c+1),\beta^{(0)}} > 0,$$

when $c < C$. Note that

$$V_{\beta^{(0)}}(c + 1, S - c) = \max_{p\in[p_l,p_h]} (p - \Delta V_{\beta^{(0)}}(c + 1, S - c + 1)) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(c + 1, S - c + 1),$$

and

$$V_{\beta^{(0)}}(c, S - c) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(c, S - c + 1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(c, S - c + 1).$$

Since $(c+1, S-c+1) \in \triangleleft$ and $(c, S-c+1) \in \triangleleft$, (P.1) implies $V_{\beta^{(0)}}(c+1, S-c+1) = V_{\beta^{(0)}}(c, S-c+1)$. In addition, $c < C$ implies $(c + 1, S - c) \in \triangleleft$, and thus $\Delta V_{\beta^{(0)}}(c + 1, S - c + 1) = 0$ by (P.1). It follows that

$$\Delta V_{\beta^{(0)}}(c + 1, S - c) = V_{\beta^{(0)}}(c + 1, S - c) - V_{\beta^{(0)}}(c, S - c) = f^*_{0,\beta^{(0)}} - f^*_{\Delta V_{\beta^{(0)}}(c,S-c+1),\beta^{(0)}} > 0,$$

where the strict positiveness follows by the induction hypothesis from the fact that $\Delta V_{\beta^{(0)}}(c, S - c + 1) > 0$, together with the fact that $f^*_{a,\beta^{(0)}}$ is strictly decreasing in a (Lemma 2(ii)).

This proves (P.2), and shows that a price-change occurs when $\triangleleft$ is entered.

This concludes case 1.

Case 2. The path $(c(s), s)_{1\le s\le S}$ does not hit $\triangleleft$. Then there is an s such that $c(s) = 2$ and $c(s + 1) = 1$. We show $\pi^*_{\beta^{(0)}}(2, s) \neq \pi^*_{\beta^{(0)}}(1, s + 1)$, for all $1 \le s \le S - 2$.

$$\pi^*_{\beta^{(0)}}(2, s) = \arg\max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(2, s + 1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p), \tag{16}$$

$$\pi^*_{\beta^{(0)}}(1, s + 1) = \arg\max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(1, s + 2)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p). \tag{17}$$

By Lemma 2, and the fact that $\pi^*_{\beta^{(0)}}(2, s)$ and $\pi^*_{\beta^{(0)}}(1, s + 1)$ are both contained in (p_l, p_h), it

suffices to show $\Delta V_{\beta^{(0)}}(2, s+1) \neq \Delta V_{\beta^{(0)}}(1, s+2)$. We show by backward induction that

$$V_{\beta^{(0)}}(2,s) - V_{\beta^{(0)}}(1,s) \leq V_{\beta^{(0)}}(1,s+1) - p_h(1 - \max_{p\in[p_l,p_h]} h(\beta_0^{(0)} + \beta_1^{(0)} p)) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p_h), \quad (18)$$

for all $2 \leq s \leq S-1$. Since $\max_{p\in[p_l,p_h]} h(\beta_0^{(0)} + \beta_1^{(0)} p) < 1$, this proves $\Delta V_{\beta^{(0)}}(2, s+1) \neq \Delta V_{\beta^{(0)}}(1, s+2)$, and that in case 2 a price change occurs.

Let $2 \leq s \leq S-1$. Then

$$V_{\beta^{(0)}}(2,s) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(2,s+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(2,s+1), \quad (19)$$

$$V_{\beta^{(0)}}(1,s) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(1,s+1)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(1,s+1), \quad (20)$$

$$V_{\beta^{(0)}}(1,s+1) = \max_{p\in[p_l,p_h]} \big(p - \Delta V_{\beta^{(0)}}(1,s+2)\big) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p) + V_{\beta^{(0)}}(1,s+2). \quad (21)$$

Using

$$V_{\beta^{(0)}}(1,s+1) \geq \Big[(\pi^*_{\beta^{(0)}}(2,s) - \Delta V_{\beta^{(0)}}(1,s+2)) h(\beta_0^{(0)} + \beta_1^{(0)} \pi^*_{\beta^{(0)}}(2,s)) + V_{\beta^{(0)}}(1,s+2)\Big],$$

we have

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

The last inequality is implied by $V_{\beta^{(0)}}(2,s+1) - V_{\beta^{(0)}}(1,s+1) - V_{\beta^{(0)}}(1,s+2) \leq 0$, which for $s = S-1$ follows from (P.1), and for $s < S-1$ follows from the induction hypothesis.

The proof of Lemma 3 shows $V_{\beta^{(0)}}(1,s+1) = \Delta V_{\beta^{(0)}}(1,s+1) \leq \max_{p\in[p_l,p_h]} p h(\beta_0^{(0)} + \beta_1^{(0)} p) \leq p_h \max_{p\in[p_l,p_h]} h(\beta_0^{(0)} + \beta_1^{(0)} p)$. This implies

$$\begin{aligned}
V_{\beta^{(0)}}(1,s) &\geq (p_h - V_{\beta^{(0)}}(1,s+1)) \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p_h) + V_{\beta^{(0)}}(1,s+1)\\
&\geq p_h[1 - \max_{p\in[p_l,p_h]} h(\beta_0^{(0)} + \beta_1^{(0)} p)] \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p_h) + V_{\beta^{(0)}}(1,s+1),
\end{aligned}$$

and thus

$$V_{\beta^{(0)}}(2,s) - V_{\beta^{(0)}}(1,s) - V_{\beta^{(0)}}(1,s+1) \leq V_{\beta^{(0)}}(1,s+1) - V_{\beta^{(0)}}(1,s)$$
$$\leq -p_h[1 - \max_{p\in[p_l,p_h]} h(\beta_0^{(0)} + \beta_1^{(0)} p)] \cdot h(\beta_0^{(0)} + \beta_1^{(0)} p_h),$$

i.e. equation (18). This concludes case 2.

We have shown that, on any path $(c(s), s)_{1\leq s\leq S}$ in $\mathcal{X}$ starting at $(C, 1)$, the policy $\pi^*_{\beta^{(0)}}$ induces a price-change. It follows that there exists a $v_0 > 0$ such that for all paths $(c(s), s)_{1\leq s\leq S}$,

$$|\pi^*_{\beta^{(0)}}(c(s), s) - \pi^*_{\beta^{(0)}}(c(s'), s')| \geq v_0.$$

Remark 1. Equations (11) and (18) enable us to provide a lower bound on the price change v_0. Let $\beta \in U_B$, $a, a' \in U_a$, and $a > a'$, where U_a, U_B are as in Lemma 1. A Taylor expansion of $g_{a,\beta}(p)$ yields

$$g_{a,\beta}(p^*_{a',\beta}) = g_{a,\beta}(p^*_{a,\beta}) + \frac{\partial g_{a,\beta}}{\partial p}(\tilde{p})(p^*_{a',\beta} - p^*_{a,\beta}), \tag{22}$$

for some $\tilde{p}$ between $p^*_{a',\beta}$ and $p^*_{a,\beta}$, and, for any $p \in [p_l, p_h]$,

$$g_{a,\beta}(p) = g_{a',\beta}(p) + \beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)}(a - a').$$

In particular, choosing $p = p^*_{a',\beta}$,

$$g_{a,\beta}(p^*_{a',\beta}) = g_{a',\beta}(p^*_{a',\beta}) + \beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p^*_{a',\beta})}{h(\beta_0 + \beta_1 p^*_{a',\beta})}(a - a'), \tag{23}$$

and thus, by combining (22) and (23) and using $g_{a,\beta}(p^*_{a,\beta}) = g_{a',\beta}(p^*_{a',\beta}) = 1$, we obtain

$$1 - \beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p^*_{a',\beta})}{h(\beta_0 + \beta_1 p^*_{a',\beta})}(a' - a) = 1 + \frac{\partial g_{a,\beta}}{\partial p}(\tilde{p})(p^*_{a',\beta} - p^*_{a,\beta}),$$

which implies

$$\left|\frac{p^*_{a',\beta} - p^*_{a,\beta}}{a' - a}\right| = \left|\frac{-\beta_1 \frac{\dot{h}(\beta_0+\beta_1 p^*_{a',\beta})}{h(\beta_0+\beta_1 p^*_{a',\beta})}}{\frac{\partial g_{a,\beta}}{\partial p}(\tilde{p})}\right|. \tag{24}$$

Thus, writing

$$\mathcal{C} = \min_{\beta\in B} \frac{-\beta_1 \min_{p\in[p_l,p_h]} \frac{\dot{h}(\beta_0+\beta_1 p)}{h(\beta_0+\beta_1 p)}}{\max_{p\in[p_l,p_h], a\in U_a} \left|\frac{\partial g_{a,\beta}}{\partial p}(p)\right|},$$

we have

$$|p^*_{a',\beta} - p^*_{a,\beta}| \geq \mathcal{C} \cdot |a' - a|. \tag{25}$$

Write $x_{1,\beta} = f^*_{0,\beta}$ and define recursively

$$x_{c+1,\beta} = x_{1,\beta} - f^*_{x_{c,\beta},\beta}, \quad 1 \leq c \leq C-1.$$

Equation (11) implies

$$|\pi^*_{\beta^{(0)}}(c,s) - \pi^*_{\beta^{(0)}}(c,s+1)| \geq \mathcal{C} \cdot x_{c,\beta^{(0)}}$$

for all $(c,s) \in (L\triangleleft)$, and equation (18) implies

$$|\pi^*_{\beta^{(0)}}(2,s) - \pi^*_{\beta^{(0)}}(1,s+1)| \geq \mathcal{C} \cdot \min_{\beta \in B} \left| p_h(1 - \max_{p \in [p_l,p_h]} h(\beta_0 + \beta_1 p)) \cdot h(\beta_0 + \beta_1 p_h) \right|$$

for $1 \leq s \leq S-2$. As a result, it follows that v_0 satisfies

$$v_0 \geq \mathcal{C} \cdot \min_{\beta \in B} \min \left\{ x_{1,\beta}, x_{2,\beta}, \ldots, x_{C,\beta}, \left| p_h(1 - \max_{p \in [p_l,p_h]} h(\beta_0 + \beta_1 p)) \cdot h(\beta_0 + \beta_1 p_h) \right| \right\}. \tag{26}$$

Proof of Theorem 2

Consider the k-th selling season, for some arbitrary fixed $k \in \mathbb{N}$. 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:

$$\tilde{\mathcal{X}} = \{(x_1, \ldots, x_s) \in \{0,1\}^s \mid 0 \leq s \leq S\},$$

where we denote the empty sequence by $(\emptyset)$. The action space is $[p_l, p_h]$. Using action p in state $(x_1, \ldots, x_s)$, for $0 \leq s < S$, induces a state transition from $(x_1, \ldots x_s)$ to $(x_1, \ldots, x_s, 1)$ with probability $h(\beta_0+\beta_1 p)$ (corresponding to a sale, and inducing immediate reward $ph(\beta_0+\beta_1 p)\mathbf{1}_{\sum_{i=1}^s x_i < C}$), and from $(x_1, \ldots x_s)$ to $(x_1, \ldots, x_s, 0)$ with probability $1 - h(\beta_0 + \beta_1 p)$ (corresponding to no sale, and inducing zero reward). There are no state transitions in the terminal states $(x_1, \ldots, x_S) \in \tilde{\mathcal{X}}$.

It is easily seen that the MDP described in section 2.3 is the same as 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.

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

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

where we write $\tilde{V}_\beta^{\tilde{\pi}}(x; 1) = \tilde{V}_\beta^{\tilde{\pi}}(x; 0) = 0$ for all terminal states $(x_1, \ldots, x_S) \in \tilde{\mathcal{X}}$.

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

$$\tilde{V}_\beta(x) = \max_{p \in [p_l, p_h]} \left[p\mathbf{1}_{\sum_{i=1}^{s} x_i < C} - \left(\tilde{V}_\beta(x; 0) - \tilde{V}_\beta(x; 1) \right) \right] h(\beta_0 + \beta_1 p) + \tilde{V}_\beta(x; 0), \tag{27}$$

$$\tilde{\pi}^*_\beta(x) = \arg\max_{p \in [p_l, p_h]} \left[p\mathbf{1}_{\sum_{i=1}^{s} x_i < C} - \left(\tilde{V}_\beta(x; 0) - \tilde{V}_\beta(x; 1) \right) \right] h(\beta_0 + \beta_1 p). \tag{28}$$

Using the same line of reasoning as Lemma 2 and 3, it can easily be shown that $\tilde{\pi}^*_\beta((x_1, \ldots, x_s))$ is unique if and only if $\sum_{i=1}^{s} x_i < C$. For all x with $\sum_{i=1}^{s} x_i \geq C$, choose $\tilde{\pi}^*_\beta(x) = p_h$. In this way $\tilde{\pi}^*_\beta(x)$ is uniquely defined for all $x \in \tilde{\mathcal{X}}$.

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 $||\beta - \beta^{(0)}|| \leq \rho$.

If $(k-1)S > T_\rho$, then $\hat{\beta}_t \in \mathcal{U}$ for all $t = 1 + (k-1)S, \ldots, S(k-1)S$, and Theorem 1 implies $\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S}) \geq \frac{1}{8} v_0^2 (1 + p_h^2)^{-1}$. If $(k-1)S \leq T_\rho$, then I) of the pricing strategy $\Phi(\epsilon)$ guarantees that there are $1 \leq s, s' \leq S$ such that $|p_{s+(k-1)S} - p_{s'+(k-1)S}| \geq \epsilon$. By Lemma 4 this implies $\lambda_{\min}(P_{kS}) - \lambda_{\min}(P_{(k-1)S}) \geq \frac{1}{2}\epsilon^2 (1 + p_h^2)^{-1}$. Since $\epsilon^2 \leq v_0^2/4$, this means that $\lambda_{\min}(P_{kS}) \geq k \cdot \frac{1}{2}\epsilon^2 (1 + p_h^2)^{-1}$ for all $k \in \mathbb{N}$, and thus for all $t > S$,

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

using $SS_t - 1 \geq t\frac{(SS_t - 1)}{S \cdot SS_t} \geq \frac{t}{2S}$. (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 $T_\rho < \infty$ a.s., $E[T_\rho] < \infty$, and $E[||\hat{\beta}_t - \beta^{(0)}||^2 \mathbf{1}_{t > T_\rho}] = O(\log(t)/t)$.

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

$$p_{1+s+(k-1)S} = \tilde{\pi}^*_{\hat{\beta}_{s+(k-1)S}}(d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S}), \tag{29}$$

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

$$p_{1+(k-1)S} = \tilde{\pi}^*_{\hat{\beta}_{(k-1)S}}(\emptyset). \tag{30}$$

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 (29) and (30) 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 $x \in \tilde{\mathcal{X}}$ there is a uniquely associated price prescribed by $\Phi(\epsilon)$. Consequently, there is a stationary deterministic policy, denoted by $\tilde{\pi}^H$, such that

$$\begin{aligned} p_{1+s+(k-1)S} &= \tilde{\pi}^H(x), \quad \text{when } x = (d_{1+(k-1)S}, d_{2+(k-1)S}, \ldots, d_{s+(k-1)S}),\ 1 \le s \le S-1, \\ p_{1+(k-1)S} &= \tilde{\pi}^H(\emptyset). \end{aligned}$$

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

$$\begin{aligned} & V_{\beta^{(0)}}(C,1) - \sum_{i=1+(k-1)S}^{kS} E[p_i \min\{d_i, c_i\}] \\ =& E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i \min\{d_i, c_i\}\right) \mathbf{1}_{(k-1)S \le T_\rho}\right] \\ +& E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i \min\{d_i, c_i\}\right) \mathbf{1}_{(k-1)S > T_\rho}\right] \\ \le& \tilde{V}_{\beta^{(0)}}(\emptyset) P((k-1)S \le T_\rho) \\ +& E\left[E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \sum_{i=1+(k-1)S}^{kS} p_i \min\{d_i, c_i\}\right) \mathbf{1}_{(k-1)S > T_\rho} \mid H\right]\right] \\ \le& \tilde{V}_{\beta^{(0)}}(\emptyset) \frac{E[T_\rho]}{(k-1)S} \qquad (31) \\ +& E\left[E\left[\left(\tilde{V}_{\beta^{(0)}}(\emptyset) - \tilde{V}^{\tilde{\pi}^H}_{\beta^{(0)}}(\emptyset)\right) \mathbf{1}_{(k-1)S > T_\rho} \mid H\right]\right]. \qquad (32) \end{aligned}$$

The term (31) is finite because $E[T_\rho] < \infty$. To obtain an upper bound on the term (32), we need a number of sensitivity results:

(S.0) For all $\beta \in U_B$ and x such that $(x;0), (x;1) \in \tilde{\mathcal{X}}$, we have

$$0 \le \tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \le \max_{p \in [p_l, p_h]} p \cdot h(\beta_0 + \beta_1 p). \qquad (33)$$

(S.1) Write $Y_s = (d_{1+(k-1)S}, \ldots, d_{s+(k-1)S})$ for $1 \le s \le S-1$, and $Y_0 = (\emptyset)$. There is a $K_0 > 0$ such that, for all stationary deterministic policies $\tilde{\pi}$ and all $0 \le s \le S-1$,

$$(\tilde{V}_{\beta^{(0)}}(Y_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s)) \mathbf{1}_{(k-1)S > T_\rho} \le K_0 \sum_{\sigma=s}^{S-1} (\tilde{\pi}^*_{\beta^{(0)}}(Y_\sigma) - \tilde{\pi}(Y_\sigma))^2 \mathbf{1}_{(k-1)S > T_\rho} \text{ a.s.} \qquad (34)$$

(S.2) There is a $K_3 > 0$ such that for all $\beta \in B$ with $||\beta - \beta^{(0)}|| \leq \rho$, and all $x \in \tilde{\mathcal{X}}$,

$$|\tilde{\pi}^*_{\beta}(x) - \tilde{\pi}^*_{\beta^{(0)}}(x)| \leq K_3||\beta - \beta^{(0)}||. \tag{35}$$

The proof of these three sensitivity properties is given below.

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

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

for some K_4 independent of k and S.

We then have

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

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

The proof of the theorem is complete by observing

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

Proof of (S.0)

We prove the assertion for all $(x_1,\ldots,x_{s-1}) \in \tilde{\mathcal{X}}$, $s = 1,\ldots,S$, by backward induction on s. If

$x = (x_1, \ldots, x_{S-1}) \in \tilde{\mathcal{X}}$ then $\tilde{V}_\beta(x;0) = \tilde{V}_\beta(x;1) = 0$.

Let $x \in \mathcal{X}$. If $\sum_{i=1}^s x_i \geq C$ then $\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) = 0$. If $\sum_{i=1}^s x_i < C$ then the induction hypothesis implies

$$
\begin{aligned}
&\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \\
&= \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;0)) + \tilde{V}_\beta(x;0;0) \\
&- \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;1)) - \tilde{V}_\beta(x;1;0) \\
&\geq \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;1)) + \tilde{V}_\beta(x;0;0) \\
&- \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;1)) - \tilde{V}_\beta(x;1;0) \\
&= \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)(1 - h(\beta_0 + \beta_1 \pi^*_\beta(x;1))) \\
&+ \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big) h(\beta_0 + \beta_1 \pi^*_\beta(x;1)) \\
&\geq 0,
\end{aligned}
$$

and

$$
\begin{aligned}
&\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \\
&= \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;0)) + \tilde{V}_\beta(x;0;0) \\
&- \Big[\pi^*_\beta(x;1) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;1)) - \tilde{V}_\beta(x;1;0) \\
&\leq \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;0)) + \tilde{V}_\beta(x;0;0) \\
&- \Big[\pi^*_\beta(x;0) - \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big)\Big]\, h(\beta_0 + \beta_1 \pi^*_\beta(x;0)) - \tilde{V}_\beta(x;1;0) \\
&= \big(\tilde{V}_\beta(x;0;0) - \tilde{V}_\beta(x;0;1)\big)(1 - h(\beta_0 + \beta_1 \pi^*_\beta(x;0))) \\
&+ \big(\tilde{V}_\beta(x;1;0) - \tilde{V}_\beta(x;1;1)\big) h(\beta_0 + \beta_1 \pi^*_\beta(x;0)) \\
&\leq \max_{p \in [p_l, p_h]} p \cdot h(\beta_0 + \beta_1 p).
\end{aligned}
$$

Proof of (S.1)

Backward induction on s. If $s = S-1$ then Lemma 2(iii) implies

$$
\begin{aligned}
&\tilde{V}_{\beta^{(0)}}(Y_{S-1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_{S-1}) \\
=& \max_{p \in [p_l, p_h]} p \mathbf{1}_{\sum_{i=1}^{S-1} Y_i < C} h(\beta_0^{(0)} + \beta_1^{(0)} p) - \tilde{\pi}(Y_{S-1}) \mathbf{1}_{\sum_{i=1}^{S-1} Y_i < C} h(\beta_0^{(0)} + \beta_1^{(0)} \tilde{\pi}(Y_{S-1})) \\
\leq& K_0 (\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_{S-1}))^2 \text{ a.s.},
\end{aligned}
$$

and thus

$$
(\tilde{V}_{\beta^{(0)}}(Y_{S-1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_{S-1})) \cdot \mathbf{1}_{(k-1)S > T_\rho} \leq K_0 (\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_{S-1}))^2 \cdot \mathbf{1}_{(k-1)S > T_\rho} \text{ a.s.}
$$

If $0 \leq s < S-1$, then

$$
\begin{aligned}
&\tilde{V}_{\beta^{(0)}}(Y_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s) \\
=&\max_{p\in[p_l,p_h]} \big[p\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1))\big] h(\beta_0^{(0)} + \beta_1^{(0)} p) \\
&-\big[\tilde{\pi}(Y_s)\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1))\big] h(\beta_0^{(0)} + \beta_1^{(0)} \tilde{\pi}(Y_s)) \\
&+\big[\tilde{\pi}(Y_s)\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1))\big] h(\beta_0^{(0)} + \beta_1^{(0)} \tilde{\pi}(Y_s)) \\
&-\big[\tilde{\pi}(Y_s)\mathbf{1}_{\sum_{i=1}^s Y_i<C} - (\tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;0) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;1))\big] h(\beta_0^{(0)} + \beta_1^{(0)} \tilde{\pi}(Y_s)) \\
&+\tilde{V}(Y_s;0) - \tilde{V}^{\tilde{\pi}}(Y_s;0) \\
\leq& K_0(\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_s))^2 \\
&+\big(\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;0)\big) \cdot \big(1 - h(\beta_0^{(0)} + \beta_1^{(0)} \tilde{\pi}(Y_s))\big) \\
&+\big(\tilde{V}_{\beta^{(0)}}(Y_s;1) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s;1)\big) \cdot (h(\beta_0^{(0)} + \beta_1^{(0)} \tilde{\pi}(Y_s)) \\
=& K_0(\tilde{\pi}^*_{\beta^{(0)}}(Y_s) - \tilde{\pi}(Y_s))^2 + \big[\tilde{V}_{\beta^{(0)}}(Y_{s+1}) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_{s+1})\big] \text{ a.s.}
\end{aligned}
$$

Here the first inequality follows from Lemma 2(iii), observing that (S.0) implies $\tilde{V}_{\beta^{(0)}}(Y_s;0) - \tilde{V}_{\beta^{(0)}}(Y_s;1) \in U_a$. The induction hypothesis now implies

$$
(\tilde{V}_{\beta^{(0)}}(Y_s) - \tilde{V}^{\tilde{\pi}}_{\beta^{(0)}}(Y_s))\mathbf{1}_{(k-1)S>T_\rho} \leq K_0 \sum_{\sigma=s}^{S-1} (\tilde{\pi}^*_{\beta^{(0)}}(Y_\sigma) - \tilde{\pi}(Y_\sigma))^2 \mathbf{1}_{(k-1)S>T_\rho} \text{ a.s.}
$$

Proof of (S.2)

If $\sum_{i=1}^s x_i \geq C$ then $\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x) = p_h - p_h = 0$. If $\sum_{i=1}^s x_i < C$, then

$$
\begin{aligned}
\tilde{\pi}^*_\beta(x) - \tilde{\pi}^*_{\beta^{(0)}}(x) &= p^*_{\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1), \beta} - p^*_{\tilde{V}_{\beta^{(0)}}(x;0) - \tilde{V}_{\beta^{(0)}}(x;1), \beta^{(0)}} \\
&= p^*_{a,\beta} - p^*_{a^{(0)},\beta^{(0)}},
\end{aligned} \tag{36}
$$

in the notation of Lemma 2, with $a = \tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1)$ and $a^{(0)} = \tilde{V}_{\beta^{(0)}}(x;0) - \tilde{V}_{\beta^{(0)}}(x;1)$.

By (S.0) we have $a, a^{(0)} \in U_a$ and $\beta \in U_B$, and thus by Lemma 2, $\tilde{\pi}^*_\beta(x)$ is continuously differentiable. The set $\{\beta \in B \mid ||\beta - \beta^{(0)}|| \leq \rho\}$ is compact, and so is the set $\{\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \mid ||\beta - \beta^{(0)}|| \leq \rho, \beta \in B\}$. As a result, the derivative of $p^*_{a,\beta}$ w.r.t. (a,β) is bounded on the set $(a,\beta) \in \{\tilde{V}_\beta(x;0) - \tilde{V}_\beta(x;1) \mid ||\beta - \beta^{(0)}|| \leq \rho, \beta \in B\} \times \{\beta \in B \mid ||\beta - \beta^{(0)}|| \leq \rho\}$. It follows by a first-order Taylor expansion that there is a $K_6 > 0$ such that for all such (a,β),

$$
|p^*_{a,\beta} - p^*_{a^{(0)},\beta^{(0)}}| \leq K_6(|a - a^{(0)}| + ||\beta - \beta^{(0)}||). \tag{37}
$$

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 $||\beta - \beta^{(0)}|| \leq \rho$,

$$
\left|\tilde{V}_\beta(x) - \tilde{V}_{\beta^{(0)}}(x)\right| \leq K_x \left|\left|\beta - \beta^{(0)}\right|\right|. \tag{38}
$$

Combining (36), (37), and (38), we obtain

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

This proves (S.2).

Proof of Theorem 3

Let ψ be a pricing strategy, and define $B' = [5/8, 6/8] \times \{-10/16\}$. For $\beta \in B'$, let ψ^β be the pricing strategy that coincides with ψ if $t \mod S = 1$, and that equals the optimal price $\pi^*_\beta(c_t, s_t)$ if $t \mod S \neq 1$; we thus replace the pricing decisions in the second time periods of each selling season by the optimal price w.r.t. β. For $T = 1$ the statement of the theorem is trivial; let $T \geq 2$.

By the principle of optimality we have

$$\begin{aligned}
&\sup_{\beta\in B} \text{Regret}(\psi, T) \geq \sup_{\beta\in B'} \text{Regret}(\psi, T) \geq \sup_{\beta\in B'} \text{Regret}(\psi^\beta, T) \\
&= \sup_{\beta\in B'} \sum_{i=1}^{T} E\left[\begin{array}{l} [\pi^*_\beta(1,1) - \Delta V_\beta(1,2)](\beta_0 + \beta_1 \pi^*_\beta(1,1)) \\ -[p_{1+2(i-1)} - \Delta V_\beta(1,2)](\beta_0 + \beta_1 p_{1+2(i-1)}) \end{array} \right],
\end{aligned} \tag{39}$$

since the regret of ψ^β is determined by the pricing decisions in the first periods of selling seasons, $p_{1+2(i-1)}$, $i = 1, \ldots, T$.

For all $\beta \in B'$, we have

$$\begin{aligned}
\Delta V_\beta(1,2) = V_\beta(1,2) &= \max_{p\in[p_l,p_h]} p(\beta_0 + \beta_1 p) = \frac{\beta_0^2}{-4\beta_1}, \\
\pi^*_\beta(1,1) &= \arg\max_{p\in[p_l,p_h]} [p - \Delta V_\beta(1,2)](\beta_0 + \beta_1 p) = \frac{\beta_0}{-2\beta_1} + \frac{1}{2}\frac{\beta_0^2}{-4\beta_1},
\end{aligned}$$

and

$$\begin{aligned}
&[\pi^*_\beta(1,1) - \Delta V_\beta(1,2)](\beta_0 + \beta_1 \pi^*_\beta(1,1)) - [p_{1+2(i-1)} - \Delta V_\beta(1,2)](\beta_0 + \beta_1 p_{1+2(i-1)}) \\
=& -\beta_1 (p_{1+2(i-1)} - \pi^*_\beta(1,1))^2,
\end{aligned}$$

and thus

$$\sup_{\beta\in B} \text{Regret}(\psi, T) \geq \sup_{\beta\in B'} \frac{10}{16} \sum_{i=1}^{T} E[(p_{1+2(i-1)} - \pi^*_\beta(1,1))^2].$$

We proceed by showing that $E[(p_{1+2(i-1)} - \pi^*_\beta(1,1))^2] \geq \frac{1}{ai+b}$, for some $a, b > 0$ and all $2 \leq i \leq T$. This inequality follows from a generalization of the van Trees inequality, derived in Gill and Levit

(1995, equation (4)). Similar approaches to derive regret lower bounds are found in Goldenshluger and Zeevi (2009), Harrison et al. (2011), and Broder and Rusmevichientong (2012). Our proof closely follows the proof of Lemma 4.6 in the e-companion to Broder and Rusmevichientong (2012).

Fix $2 \le i \le T$. define the sample space $\mathcal{D}_{(i-1)S} = \{0,1\}^{(i-1)S}$, let $\mathbf{D}_{(i-1)S} = (d_1, \ldots, d_{(i-1)S}) \in \mathcal{D}_{(i-1)S}$ be the random variable denoting the demand in periods 1 to $(i-1)S$, and write $\mathbf{d}_{(i-1)S}$ for a realization of $\mathbf{D}_{(i-1)S}$. Define the family of distributions $\{Q^{\beta_0} \mid \beta \in [\beta_0' - \delta, \beta_0' + \delta]\}$, where

$$Q^{\beta_0}(\mathbf{d}_{(i-1)S}) = \prod_{t=1}^{(i-1)S} (\beta_0 + \beta_1 p_t)^{d_t} (1 - (\beta_0 + \beta_1 p_t))^{1-d_t}, \quad \mathbf{d}_{(i-1)S} \in \mathcal{D}_{(i-1)S},$$

is the distribution of demand realizations in time periods 1 to $(i-1)S$.

Let $\lambda_0(\theta) = \cos^2(\pi\theta/2)\mathbf{1}_{|\theta|\le 1}$, define the density λ on $[5/8, 6/8]$ as

$$\lambda(\beta_0) = \frac{1}{1/8}\lambda_0\left(\frac{\beta_0 - 11/16}{(1/8)}\right), \quad \beta_0 \in [5/8, 6/8],$$

cf. Goldenshluger and Zeevi (2009, page 1632), and let Z be a random variable supported on $[5/8, 6/8]$ with density λ.

Then by the van Trees inequality, we have

$$E[(p_{1+2(i-1)} - \pi_\beta^*(1,1))^2] \ge \frac{(E[\frac{\partial}{\partial\beta_0}\pi_\beta^*(1,1)])^2}{E[(\frac{\partial}{\partial\beta_0}\log Q^{\beta_0}(\mathbf{D}_{(i-1)S}))^2] + E[(\frac{\partial}{\partial\beta_0}\log\lambda(\beta_0))^2]},$$

where the expectations are with respect to the joint distribution of Q^{β_0} and λ.

Now

$$E\Big[(\frac{\partial}{\partial\beta_0}\log Q^{\beta_0}(\mathbf{D}_{(i-1)S}))^2 \mid \mathbf{D}_{(i-1)S-1} = \mathbf{d}_{(i-1)S-1}\Big] \tag{40}$$

$$= \frac{2}{(\beta_0 + \beta_1 p_{2(i-1)})(1 - (\beta_0 + \beta_1 p_{2(i-1)}))} \tag{41}$$

$$\le \sup_{p\in\mathcal{P},\beta\in B} \frac{2}{(\beta_0 + \beta_1 p)(1 - (\beta_0 + \beta_1 p))} \tag{42}$$

$$= \frac{2}{(13/136)\cdot(1 - 15/32)} = \frac{8704}{221} < 40, \tag{43}$$

and by applying the chain rule for Fisher information (Lemma EC.5.2 of the e-companion of Broder and Rusmevichientong (2012)), it follows that

$$E\Big[(\frac{\partial}{\partial\beta_0}\log Q^{\beta_0}(\mathbf{D}_{(i-1)S}))^2\Big] \le 40(i-1)S < 80i. \tag{44}$$

We furthermore have $E[(\frac{\partial}{\partial\beta_0}\log\lambda(\beta_0))^2] = \pi^2/(1/8)^2$ and $(E[\frac{\partial}{\partial\beta_0}\pi_\beta^*(1,1)])^2 = (43/40)^2$, and thus

$$E[(p_{1+2(i-1)} - \pi_\beta^*(1,1))^2] \ge \frac{(43/40)^2}{80i + 64\pi^2}.$$

This implies

$$\sup_{\beta\in B}\text{Regret}(\psi,T)\geq\frac{10}{16}\sum_{i=2}^{T}\frac{(43/40)^2}{80i+64\pi^2}\geq\frac{10}{16}\frac{43^2}{40^2}\frac{1}{80+64\pi^2}\sum_{i=2}^{T}\frac{1}{i}\geq\frac{10}{16}\frac{43^2}{40^2}\frac{1}{80+64\pi^2}\frac{1}{2}\log(T).$$

9 Appendix: Auxiliary Lemmas

In this section we formulate and prove several auxiliary results that are used in the proofs of the main theorems of the paper.

Lemma 1 shows that the assumptions we impose on $g_{a,\beta}(p)$ do not only hold for $a\in[0,r^*]$ and $\beta=\beta^{(0)}$, but also on an open neighborhood around $[0,r^*]\times\{\beta^{(0)}\}$. This result enables us in later proofs to apply the implicit function theorem. Lemma 2 considers the optimization problem underlying (3), and shows uniqueness, differentiability, and sensitivity properties. These results are applied in Lemma 3 to conclude that $(\pi^*_\beta)_{1\leq c\leq C,1\leq s\leq S}$ is uniquely defined and continuous in β, on an open neighborhood around $\beta^{(0)}$. Lemma 4 relates price differences to the growth of $\lambda_{\min}(P_t)$ during a selling season.

Lemma 1. *There are open sets $U_a\subset\mathbb{R}$ containing $[0,r^*]$, and $U_B\subset B$ containing $\beta^{(0)}$, with*

$$\sup_{\beta\in U_B}\max_{p\in[p_l,p_h]}p\cdot h(\beta_0+\beta_1p)\in U_a,\tag{45}$$

and such that

$$g_{a,\beta}(p_l)<1,\ g_{a,\beta}(p_h)>1\ \text{and}\ g_{a,\beta}(p)\ \text{strictly increasing in}\ p,\tag{46}$$

holds for all $(a,\beta)\in U_a\times U_B$.

Lemma 2. *Let U_a and U_B be as in Lemma 1, and for all $(a,\beta)\in U_a\times U_B$ define the function $f_{a,\beta}(p)=(p-a)h(\beta_0+\beta_1p)$. 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 $p^*_{a,\beta}=\underset{p\in[p_l,p_h]}{arg\,max}\,f_{a,\beta}(p)$. Then:*

(i) $p^*_{a,\beta}$ is the unique solution to $\dot f_{a,\beta}(p)=0$, lies in (p_l,p_h), and in addition satisfies $\ddot f_{a,\beta}(p^*_{a,\beta})<0$.

(ii) $p^*_{a,\beta}$ is continuously differentiable in a and β, strictly increasing in a, and $f_{a,\beta}(p^*_{a,\beta})$ is strictly decreasing in a.

(iii) There is a $K_0>0$ such that for all $(a,\beta)\in U_a\times U_B$ and $p\in[p_l,p_h]$,

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

Lemma 3. *Let U_B be as in Lemma 1. For each $\beta\in U_B$ and $(c,s)\in\mathcal{X}$ with $c>0$ we have $\Delta V_\beta(c,s)\in U_a$. Furthermore, $\pi^*_\beta(c,s)$ is uniquely defined and continuous in β.*

Lemma 4. *Let $k\in\mathbb{N}$. If there are $s,s'\in\{1,\dots,S\}$ such that $|p_{s+(k-1)S}-p_{s'+(k-1)S}|\geq\delta$, then $\lambda_{min}(P_{kS})\geq\lambda_{min}(P_{(k-1)S})+\frac{1}{2}\delta^2(1+p_h^2)^{-1}$.*

Proof of Lemma 1

The lemma follows directly from the continuity assumptions on h.

Proof of Lemma 2

Since

$$\dot{f}_{a,\beta}(p) = h(\beta_0 + \beta_1 p)\left[1 + (p-a)\beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)}\right] = h(\beta_0 + \beta_1 p)\left[1 - g_{a,\beta}(p)\right],$$

and $h(\beta_0 + \beta_1 p) > 0$ for all $\beta \in U_B$, $p \in [p_l, p_h]$, we have $\dot{f}_{a,\beta}(p) = 0$ if and only if $g_{a,\beta}(p) = 1$. By Lemma 1, for all $(a, \beta) \in U_a \times U_B$ there is a unique $p^*_{a,\beta} \in (p_l, p_h)$ such that $g_{a,\beta}(p^*_{a,\beta}) = 1$. From

$$\begin{aligned}\ddot{f}_{a,\beta}(p) &= \frac{\partial}{\partial p}\Big[h(\beta_0 + \beta_1 p)(1 - g_{a,\beta}(p))\Big] \\ &= \beta_1 \dot{h}(\beta_0 + \beta_1 p)(1 - g_{a,\beta}(p)) - h(\beta_0 + \beta_1 p)\frac{\partial}{\partial p} g_{a,\beta}(p)\end{aligned}$$

follows

$$\ddot{f}_{a,\beta}(p^*_{a,\beta}) = -h(\beta_0 + \beta_1 p^*_{a,\beta})\frac{\partial}{\partial p} g_{a,\beta}(p^*_{a,\beta}) < 0,$$

since by Lemma 1, $g_{a,\beta}$ is strictly increasing in p. This proves (i).

For all $(a, \beta) \in U_a \times U_B$, $p^*_{a,\beta}$ is the unique solution in (p_l, p_h) to $g_{a,\beta}(p) - 1 = 0$, and

$$\frac{\partial g_{a,\beta}(p)}{\partial p}\Big|_{p = p^*_{a,\beta}} > 0.$$

The implicit function theorem (see e.g. Duistermaat and Kolk, 2004) then implies that $p^*_{a,\beta}$ is continuously differentiable at every $(a, \beta) \in U_a \times U_B$.

Furthermore, for all $(a, \beta) \in U_a \times U_B$ and $p \in [p_l, p_h]$ we have

$$\frac{\partial g_{a,\beta}(p)}{\partial a} = \beta_1 \frac{\dot{h}(\beta_0 + \beta_1 p)}{h(\beta_0 + \beta_1 p)} < 0.$$

This implies that for all $a \in U_a$, $a' \in U_a$, with $a < a'$, and all $p \in [p_l, p_h]$ with $p \leq p^*_{a,\beta}$, we have $g_{a',\beta}(p) \leq g_{a,\beta}(p) \leq 1$. Therefore $p^*_{a',\beta} > p^*_{a,\beta}$ for all $a < a'$, and thus $p^*_{a,\beta}$ is strictly monotone increasing in a.

Using $g_{a,\beta}(p^*_{a,\beta}) = 1$ and thus $(p^*_{a,\beta} - a) = (-\beta_1^{-1})\frac{h(\beta_0 + \beta_1 p^*_{a,\beta})}{\dot{h}(\beta_0 + \beta_1 p^*_{a,\beta})}$, we have

$$f_{a,\beta}(p^*_{a,\beta}) = (p^*_{a,\beta} - a)h(\beta_0 + \beta_1 p^*_{a,\beta}) = (-\beta_1^{-1})\frac{h(\beta_0 + \beta_1 p^*_{a,\beta})^2}{\dot{h}(\beta_0 + \beta_1 p^*_{a,\beta})},$$

and thus

$$\frac{\partial}{\partial a} f_{a,\beta}(p^*_{a,\beta})) = (-\beta_1^{-1}) \left(\frac{\partial}{\partial z} \frac{h(z)^2}{\dot{h}(z)} \Big|_{z=\beta_0+\beta_1 p^*_{a,\beta}} \right) \beta_1 \frac{\partial}{\partial a} p^*_{a,\beta}. \tag{47}$$

Log-concavity of h implies $\frac{\partial^2 \log(h(z))}{\partial z^2} = \frac{h(z)\ddot{h}(z) - \dot{h}(z)^2}{h(z)^2} \leq 0$, and thus

$$\begin{aligned}
\frac{\partial}{\partial z} \frac{h(z)^2}{\dot{h}(z)} &= \frac{2h(z)\dot{h}(z)^2 - h(z)^2\ddot{h}(z)}{\dot{h}(z)^2} = h(z) \left[2 - \frac{h(z)\ddot{h}(z)}{h(z)^2} \frac{h(z)^2}{\dot{h}(z)^2} \right] \\
&\geq h(z) \left[2 - \frac{\dot{h}(z)^2}{h(z)^2} \frac{h(z)^2}{\dot{h}(z)^2} \right] = h(z).
\end{aligned}$$

Since $\frac{\partial}{\partial a} p^*_{a,\beta} > 0$, it follows that $f_{a,\beta}(p^*_{a,\beta})$ is strictly decreasing in a. This completes the proof of (ii).

Let $K_0 = \sup_{(a,\beta,p)\in U_a \times U_B \times [p_l,p_h]} -\ddot{f}_{a,\beta}(p)/2$. Since $(a,\beta,p) \mapsto f_{a,\beta}(p)$ is twice continuously differentiable on $\mathbb{R} \times B \times [p_l, p_h]$ and $\ddot{f}_{a,\beta}(p^*_{a,\beta}) < 0$, it follows that $0 < K_0 < \infty$. By a Taylor expansion, there is a $\tilde{p}_{a,\beta}$ on the line segment between p and $p^*_{a,\beta}$, such that

$$\begin{aligned}
f_{a,\beta}(p) &= f_{a,\beta}(p^*_{a,\beta}) + \dot{f}_{a,\beta}(p^*_{a,\beta})(p - p^*_{a,\beta}) + \frac{1}{2}\ddot{f}_{a,\beta}(\tilde{p}_{a,\beta})(p - p^*_{a,\beta})^2 \\
&\geq f_{a,\beta}(p^*_{a,\beta}) - K_0(p - p^*_{a,\beta})^2,
\end{aligned}$$

using $\dot{f}_{a,\beta}(p^*_{a,\beta}) = 0$. This proves (iii).

Proof of Lemma 3

Let $\beta \in U_B$. We show $0 \leq \Delta V_\beta(c,s) \leq \max_{p\in[p_l,p_h]} ph(\beta_0 + \beta_1 p)$, for all $(c,s) \in \mathcal{X}$. By (45), this implies $\Delta V_\beta(c,s) \in U_a$. In view of (3), uniqueness and continuity of π^*_β then follow from repeated application of Lemma 2(i, ii), for each $(c,s) \in \mathcal{X}$.

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

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

and

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

Proof of Lemma 4

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

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

Acknowledgment

We thank Sandjai Bhulai for useful discussions and providing literature references. We also thank the referees and the associate editor, whose comments have greatly improved the paper. Part of this research was done while the first author was with CWI, Eindhoven University of Technology, and University of Amsterdam. The research of Bert Zwart is partly supported by an NWO VIDI grant and an IBM faculty award.

References

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

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

V. F. Araman and R. Caldentey. Revenue management with incomplete demand information. In J. J. Cochran, editor, *Encyclopedia of Operations Research*. Wiley, 2011.

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

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

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

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

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

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

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

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

A. V. den Boer. Dynamic pricing with multiple products and partially specified demand distribution. Submitted for publication, 2011.

A. V. den Boer and B. Zwart. Simultaneously learning and optimizing using controlled variance pricing. *Management Science*, Accepted for publication, 2010.

A. V. den Boer and B. Zwart. Mean square convergence rates for maximum quasi-likelihood estimators. Submitted for publication, 2011.

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

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

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

A. Goldenshluger and A. Zeevi. Woodroofe's one-armed bandit problem revisited. *The Annals of Applied Probability*, 19(4):1603–1633, 2009.

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

J. M. Harrison, N. B. Keskin, and A. Zeevi. Dynamic pricing with an unknown linear demand model: asymptotically optimal semi-myopic policies. Working paper, http://faculty-gsb.stanford.edu/harrison/Documents/hkz-2.pdf, 2011.

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

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

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

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

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

K. Kalyanam, R. Lal, and G. Wolfram. Future store technologies and their impact on grocery retailing. In M. Krafft and M. K. Mantrala, editors, *Retailing in the 21st Century*, chapter 7, pages 95–112. Springer, Berlin, Heidelberg, 2006.

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

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

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

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

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

M. A. Lariviere. A note on probability distributions with increasing generalized failure rates. *Operations Research*, 54(3):602–604, 2006.

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

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

M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York, first edition, 1994.

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

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

Z. Wang, S. Deng, and Y. Ye. Close the gaps: a learning-while-doing algorithm for a class of single-product revenue management problems. Working paper, 2011.

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