

Dynamic Pricing for Nonperishable Products with Demand Learning

Victor F. Araman

Stern School of Business, New York University, New York, New York 10012, and Olayan School of Business, American University of Beirut, Beirut, Lebanon, va03@aub.edu.lb

René Caldentey

Stern School of Business, New York University, New York, New York 10012, rcaldent@stern.nyu.edu

A retailer is endowed with a finite inventory of a nonperishable product. Demand for this product is driven by a price-sensitive Poisson process that depends on an unknown parameter that is a proxy for the market size. The retailer has a prior belief on the value of this parameter that he updates as time and available information (prices and sales) evolve. The retailer's objective is to maximize the discounted long-term average profits of his operation using dynamic pricing policies. We consider two cases. In the first case, the retailer is constrained to sell the entire initial stock of the nonperishable product before a different assortment is considered. In the second case, the retailer is able to stop selling the nonperishable product at any time and switch to a different menu of products. For both cases, we formulate the retailer's problem as a (Poisson) intensity control problem and derive structural properties of an optimal solution, and suggest a simple and efficient approximated solution. We use numerical computations, together with asymptotic analysis, to evaluate the performance of our proposed policy.

Subject classifications: dynamic pricing; Bayesian demand learning; approximations; intensity control; nonhomogeneous Poisson process; optimal stopping.

Area of review: Manufacturing, Service, and Supply Chain Operations.

History: Received December 2005; revisions received August 2008, January 2009; accepted April 2009.

1. Introduction

This paper is concerned with dynamic pricing policies for nonperishable products in the context of a retail operation with uncertain demand. In particular, we investigate the interplay between demand learning and pricing decisions and their impact on the long-term performance of the business.

Effective retail management is about managing a limited available capacity to procure and sell the right assortment of products while considering present and future market developments. This point of view is captured by one of the most popular measures in the retail industry, namely, average sales per square foot per unit time. Indeed, this measure highlights two fundamental aspects of a retail operation. First, it emphasizes the fact that capacity, measured by store or shelf space, is one of the retailer's key assets, and thus must be managed as such; the challenge resides in choosing the best possible menu of products; failure to do so results in opportunity costs that would cut directly into the profit margins. Second, it highlights the time value of money when assessing the business performance. For example, a retailer might prefer to sell a product with a 5% margin over another one with a 10% margin if the former sells much faster than the latter (see Gaur et al. 2005 for a comprehensive empirical analysis of this “earn versus

turns” trade-off). Thus, in optimizing this measure, retailers must balance the short-term benefits obtained by selling a given menu of products and the long-term opportunity costs incurred by allocating their resources (shelf space, time, capital, etc.) to these products instead of a different assortment.

In addition to such critical trade-offs, retailers cannot overlook the market conditions in which they compete. Customers' preferences, competitors' actions, new product introduction, regulations, and so on, are often unknown to the retailer and need to be factored into the business strategy. As a result, learning about these market factors—induced, for example, through the sales process—should be constantly performed. Such learning would shed more light on future demand and hint at the current strategy to adopt. The same product that sells well today might get stocked on the shelves tomorrow, wasting valuable space that could be used to sell a more profitable alternative. To prevent such a highly undesirable situation, a retailer must continuously monitor the products sales to infer customers' preferences, identify early-on the selling pattern of each product, and adopt the appropriate strategy. Low-selling items must be removed either by shipping them to a secondary market (e.g., an outlet) or by liquidating their inventory through active price markdowns. It is precisely this relationships between demand learning, pricing policies, and inventory

turns that we study in this paper using a stylized retail operation.

In our model, which is described in detail in §2, a retailer is endowed with a finite stock of a nonperishable product that he sells to a population of price-sensitive customers with unknown demand characteristics. The retailer dynamically controls the price of the product and uses all available information (i.e., price and sales history) to learn demand attributes over time. The problem faced by this retailer is based on the so-called *exploration* versus *exploitation* trade-off. On the one hand, pricing policies affect immediate revenues (exploitation). On the other hand, the selling pattern they induce impacts the retailer's ability to learn demand (exploration); a knowledge that can be used to increase future profits. We tackle this problem using a sequence of models with increasing degree of complexity. First, in §3, we study the perfect information case in which the retailer knows all demand parameters with certainty. In this setting, we derive an optimal pricing policy and characterize the retailer's long-term profit as a function of the inventory level. From a practical standpoint, we view this full-information case as a good approximation for an experienced retailer that sells in a mature market. In §4, we relax the perfect information assumption and consider the case in which the demand intensity depends on an unknown parameter, θ ; a proxy for the size of the market. The retailer has a prior belief with regard to the value of θ that he dynamically updates over time. We also assume that the retailer must deplete the entire stock of the product before a different assortment can be sold. This condition is satisfied in many practical situations in which the retailer has no secondary market to which to ship unpopular items or the cost of this shipment is excessive. Section 5 discusses a more general case in which the retailer has imperfect information about the value of θ while having the option to stop selling the product at any time and switch to a more profitable alternative. In §6, we present some extensions of the model, and concluding remarks are discussed §7.

We believe our model contributes to the existing literature in a number of directions. First, we propose a parsimonious continuous-time formulation to model the problem of a retailer selling nonperishable products with uncertain demand characteristics. We use dynamic programming methods to formulate the problem and propose a set of simple algorithms to efficiently solve it. A distinguishing feature of our formulation is that it explicitly includes a terminal reward that captures the opportunity cost of the seller's operations. This opportunity cost can induce the retailer to stop selling the product at any time discarding unsold units, a feature that is not captured by traditional revenue management models. We also derive simple managerial guidelines that reflect the essential characteristics of an optimal policy (pricing and stopping) and use numerical experiments and asymptotic analysis to evaluate their performance.

In summary, some of the main managerial insights that we draw from this paper are the following. The optimal pricing policies in the context of nonperishable products are not necessarily decreasing functions of the level of inventory (as is the case in the traditional revenue management setting). Indeed, whether prices are decreasing or increasing with inventory depends on the size of the market, that is, on the value of θ . We show that if θ (or the belief that θ is large) is large (respectively, small), then prices do decrease (increase) with inventory. Furthermore, we show that for a given inventory level, the optimal price is monotonically increasing with θ ; but, despite this monotonicity, we show that the resulting selling rate is also increasing with θ . In other words, under an optimal pricing policy, popular products have both higher prices and higher inventory turns compared to less-popular products. Moreover, we show that optimal prices are lower when the retailer has the option to stop selling a product and switch to a different assortment than when he does not have such a stopping option. This result is consistent with the intuition that learning is more valuable when the option to stop is available and therefore the retailer is willing to sacrifice immediate revenues—by setting lower prices—and induce demand learning. On the procurement side, our analysis reveals that for any given level of uncertainty about θ the retailer prefers larger batches to smaller batches. In general, large batches give the retailer more time to learn about the true value of θ . Hence, this result suggests that the upside reward of learning good news (i.e., that θ is high) dominates the downside cost of learning bad news (i.e., that θ is low) when the inventory position is large.

We conclude this introduction by attempting to position our paper within the vast literature on dynamic pricing and demand learning in operations management. Our pricing formulation is closely related to the continuous-time (Poisson) intensity control problem studied by Gallego and van Ryzin (1994) (see also Bitran and Caldentey 2003, Elmaghraby and Keskinocak 2003, Talluri and van Ryzin 2004 for related references), but with some noticeable differences. First, we depart from the revenue management setting by considering a nonperishable product. Unlike the airline industry, where flight departure dates are hard constraints, our modeling better suits a retail operation in which the seller has the flexibility to adjust the duration of the selling season based on market contingencies. As a consequence, we look at the retailer's infinite-horizon operations and use a discounted long-term average profit objective function.

As we mentioned before, a distinguishing feature of our demand model is that it depends on an unknown parameter. Practically, we apply dynamic pricing to maximize revenues, which creates an incentive for (*Bayesian*) learning. The underlying process is then a nonhomogeneous Poisson process parameterized by the unknown parameter θ . From a mathematical point of view, the learning side of our paper

resembles the sequential testing hypotheses problem studied broadly in statistics; see, for example, Shiryaev (1978) or, more recently, Peskir and Shiryaev (2000). The latter study the problem of observing the output of a homogeneous Poisson process with unknown rate (either high or low) up to a time that needs to be optimally chosen based on cost considerations.

The economics literature borrows some of these ideas. Indeed, learning and experimentation through Bayesian updates in an infinite-horizon setting has been extensively studied. Some of the most fundamental questions that these types of studies try to answer relate to the value of learning and whether, for example, optimal strategies eventually converge to the true state of the system or not (see Bolton and Harris 1999, Keller and Rady 1999, and references therein). Often in such a stream of research the only connection between periods occurs through the belief process, as opposed to operations in general, and our paper in particular, where other state variables such as manufacturing capacity or inventory levels are included.

Bayesian learning in the scope of a periodic inventory control problem has been pioneered by Scarf (1958); see also Azoury (1985), Lovejoy (1990), Eppen and Iyer (1997), Lariviere and Porteus (1999), and references therein. This literature is mainly concerned with determining optimal inventory decisions under various modes of procurement such as periodic replenishment or newsvendor-type models. The problem of optimal assortment in a multiproduct setting has also received some attention. For example, Caro and Gallien (2007) study a discrete-time finite-horizon problem using a multiarmed bandit formulation and Bayesian learning. At each time period, the seller must decide the subset of products to offer based on historical sales data. The authors propose a simple index policy based on a relaxation of the dynamic program. In most of this inventory-related research, however, pricing policies and their impact on revenues and demand learning are not investigated.

More recently, there has been an increased interest in demand learning models in the context of dynamic pricing. Most of this literature focuses on the finite-horizon setting. Petruzzi and Dada (2002) analyze the problem of learning while controlling inventory and prices in a discrete time setting. Demand in every period is a deterministic function of price perturbed by an unknown parameter and its probability distribution is updated using successive censored sale data. In this setting, the authors characterize the structure of an optimal policy. Recently, Carvalho and Puterman (2004) studied dynamic pricing of an uncapacitated system under an exponential demand function (perturbed by a Gaussian noise) with unknown parameters, estimated through a Kalman filter. Similarly, Lobo and Boyd (2003) consider a linear price demand function and obtain approximate solutions using convex programming methods.

In the context of revenue management, Aviv and Pazgal (2002) introduce Bayesian learning within the dynamic pricing model of Gallego and van Ryzin (1994), but with unknown demand intensity. The prior distribution of this intensity is assumed to be Gamma, which is a conjugate distribution for the Poisson demand process. In a similar setting, Aviv and Pazgal (2005) propose a partially observed Markov decision process framework to compute an upper bound on the seller's revenue and derive some heuristics to approximate the optimal pricing policy. Similar to our infinite-horizon model, Farias and Van Roy (2009) propose a special heuristic (*decay balancing*) that shows a good numerical performance for the case in which the unknown demand intensity has a Gamma distribution (as in Aviv and Pazgal 2002). Xu and Hopp (2005) propose a piecewise-linear demand model with unknown parameters and use Bayes updating to investigate some martingale properties of the optimal price process. Bertsimas and Perakis (2006) consider a discrete-time model in which demand is a linear function of the price with unknown coefficients and perturbed by a white noise. Both the monopolistic and oligopolistic cases are studied. Instead of Bayesian learning, the authors use a least-squares estimation embedded in a dynamic program with incomplete state information. Some approximations and heuristics are proposed to reduce the dimensionality of the problem.

Finally, there is a growing stream of literature that discusses revenue management policies under unknown demand characteristics using a nonparametric approach. A few representative examples of this stream are Cope (2004), Lim and Shanthikumar (2006), Ball and Queyranne (2009), Besbes and Zeevi (2009), and Eren and Maglaras (2006). In most of these papers, demand uncertainty, or more precisely model ambiguity, is represented by an *uncertainty set*, that is, the set of all demand models that could potentially be the real one. This ambiguity is handled using a robust formulation that identifies operating policies that will guarantee the best-possible level of performance (in a min-max, competitive ratio, or minimum regret criteria, among others) for a given uncertainty set.

2. Model Description

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space endowed with a standard (rate 1) Poisson process $D = (D(t): t \geq 0)$ and let $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ be the usual filtration generated by D . For a given $\theta > 0$, we define the probability measure \mathbb{P}_θ under which $D(t)$ is a Poisson process with rate θ . Note that \mathbb{P} coincides with \mathbb{P}_1 . We denote by \mathbb{E}_θ the expectation operator under \mathbb{P}_θ . Also, for every adapted process f_t , nonanticipating with respect to $D(t)$, we define $I_f(t) \triangleq \int_0^t f_s ds$.

In this probabilistic environment, we consider the following stylized retail operations. At time $t = 0$, a retailer owns N_0 identical units of a *nonperishable* product that he can sell to a stochastically arriving stream of buyers. These buyers are price sensitive and their purchasing behavior

is modulated by an \mathcal{F}_t -adapted price process $\{p_t; t \geq 0\}$ selected by the retailer. In particular, any given price p affects instantaneously the demand rate, which we denote by $\lambda(p)$. We let $D(I_\lambda(t))$ be the corresponding cumulative demand process up to time t . Under \mathbb{P}_θ , this cumulative demand defines a nonhomogeneous Poisson process with intensity $\theta\lambda(p_t)$. The parameter $\theta > 0$ captures the magnitude of the demand intensity, whereas the quantity $\lambda(p)$ models buyers' sensitivity to price. We refer to θ as the (demand) scale factor and $\lambda(p)$ as the unscaled demand intensity.

Consistent with standard economic theory, we assume that the mapping $p \mapsto \lambda(p)$ is a continuous, nonnegative, and strictly decreasing function of the price p . Furthermore, to avoid unrealistic unbounded optimal pricing strategies, we impose the additional condition that there exists a price p_∞ (possibly infinite) such that $\lim p\lambda(p) = 0$ as $p \uparrow p_\infty$. These assumptions guarantee the existence of an inverse demand function $p(\lambda)$ that is well defined and continuous in the domain $[0, \Lambda]$, where $\Lambda \triangleq \lambda(0)$. Based on this one-to-one correspondence between prices and demand intensities, we find it convenient to let the seller control demand intensities rather than prices. This is a recurrent modeling approach in the revenue management literature that has proven to be calligraphically efficient (e.g., Gallego and van Ryzin 1994). Under this change of control variable, we define an *admissible* selling strategy as an adapted mapping $\lambda: t \mapsto \lambda_t$, where for each time $t \geq 0$, $\lambda_t \in [0, \Lambda]$. We denote the set of such admissible strategies by \mathcal{A} .

Section C.1 in Appendix C describes three examples of demand models that satisfy the conditions in the previous paragraph: the *exponential* demand model with $\lambda(p) = \Lambda \exp(-\alpha p)$ (e.g., Smith and Achabal 1998), the *linear* demand model with $\lambda(p) = \Lambda - \alpha p$, and the *quadratic* demand model with $\lambda(p) = \sqrt{\Lambda^2 - \alpha p}$. In these cases, Λ is the maximum unscaled demand intensity and α captures customers' sensitivity to price. We will use these models in our computational experiments throughout the paper.

The products we consider in this setting are nonperishable in the sense that there is no predetermined end of season. Basically, the season will end either when all units have been sold or before if the retailer decides to stop before this depletion time. He can choose to do so at any random stopping time. We denote by \mathcal{T} the set of stopping times with respect to \mathbb{F} .

There are two sources of demand uncertainty in our model. First, as described above, we use a Poisson process to model the arriving pattern of customers. Our choice of a price-sensitive Poisson process provides mathematical tractability to our model and is a recurrent assumption within the dynamic pricing literature in operations; see Bitran and Caldentey (2003) for more details. Second, we assume that the retailer has only partial information about the value of the scale factor θ . In particular, θ is a random variable taking values on a discrete set Θ . For most of the paper we restrict the analysis to the case in which

$\Theta = \{\theta_L, \theta_H\}$ with $\theta_L \leq \theta_H$, where the subscripts L and H stand for low and high market size, respectively. In §6, we show how to extend our results to the case in which Θ is a general finite set.

We note that by modeling θ as a fixed random variable we are implicitly assuming that market conditions (e.g., customers' preferences, competition, etc.) are not changing over time. Otherwise, it would be more appropriate to model θ as a Θ -valued stochastic process. In this respect, our model with a fixed θ is well suited for products with a short life cycle (such as seasonal, perishable, or fashionable items) with only a few months of selling horizon and for which market conditions tend to be relatively stable.

The retailer starts the selling season with a prior belief q that $\theta = \theta_L$. As time goes by, and demand data is collected, the retailer is able to update his estimate on the true value of θ . For a given prior $q \in [0, 1]$, we use a slight abuse of notation and define the probability measure $\mathbb{P}_q \triangleq q\mathbb{P}_{\theta_L} + (1 - q)\mathbb{P}_{\theta_H}$, with expectation operator \mathbb{E}_q .

The seller's problem is to dynamically adjust the demand intensity λ_t to maximize long-term expected cumulative profits. In particular, we consider the following intensity control problem:

$$\sup_{\lambda \in \mathcal{A}, \tau \in \mathcal{T}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt) p(\lambda_t) dD(I_\lambda(t)) + \exp(-r\tau)R \right] \quad (1)$$

$$\text{subject to } N_t = N_0 - \int_0^t dD(I_\lambda(s)) \quad (\text{inventory dynamics}), \quad (2)$$

$$\tau \leq \inf\{t \geq 0: N_t = 0\} \quad (\text{terminal condition}). \quad (3)$$

A few remarks about this control problem are in order. Our modeling differs from the more traditional revenue management literature (e.g., Talluri and van Ryzin 2004, Bitran and Caldentey 2003, Elmaghraby and Keskinocak 2003) in a couple of dimensions. Because of the nonperishability of the product, our model does not consider a fixed finite horizon, but rather an infinite-horizon stopping time problem. Note that the stopping time τ allows the retailer to stop selling the product at any time satisfying constraint (3), and so backorders are not allowed. Another difference—that is consistent with our infinite-horizon view of the retailer's operation—is the use of discount rate $r > 0$ that penalizes future cash flows. Finally, a distinguishing aspect of our model is the terminal value R , which captures the opportunity cost of the retailer's operation. We interpret R as the expected discounted cash flows that the seller can get from his retail business after he stops selling the current product. In practice, estimating the "correct" value of R is a difficult task. A commonly used rule-of-thumb is to consider the historical returns of the operation. (Other interpretations based on operational costs or property values

are also possible.) However, this measure fails to take into account new information about markets and products. We do not model the problem of computing this opportunity cost because it lies beyond the scope of this paper. Instead, we assume that the retailer has been able to get a good estimate of the value of R . It is possible that in some cases the reward R is a function of the terminal inventory N_τ (similar to the option to “dump” inventory in Eppen and Iyer 1997) or even a function of the seller’s updated beliefs on θ at time τ (in case of demand correlation between two consecutive assortments). We postpone the discussion of these and other extensions to §6.

In the following sections, we study problem (1)–(3) under different degrees of complexity. We start by looking at the simplest (full-information) case, in which the retailer knows the value of θ at $t = 0$, and then move to the case where θ is unknown.

3. Dynamic Pricing with Perfect Demand Information

In this section, we solve the retailer’s optimization problem and derive structural properties of its solution, assuming that θ is fully known so that $\mathbb{P}_\theta = \mathbb{P}_\theta$. Also, to ease the exposition, we first solve problem (1)–(3), replacing the inequality sign in (3) by an equality sign. That is, we assume that all units must be sold before the retailer can start selling a different assortment. The solution for the case with the inequality sign in (3) will follow directly from this analysis (see the discussion following Proposition 1).

Under some minor technical conditions on λ (see §III.3 in Brémaud 1980), we can rewrite the seller’s optimization problem as follows:

$$W(N_0; \theta) = \sup_{\lambda_t \in \mathcal{A}} \mathbb{E}_\theta \left[\int_0^\tau \exp(-rt) \theta c(\lambda_t) dt + \exp(-r\tau) R \right] \quad (4)$$

$$\text{subject to } N_t = N_0 - \int_0^t dD(I_\lambda(s)), \quad (5)$$

$$\tau = \inf\{t \geq 0: N_t = 0\}, \quad (6)$$

where $c(\lambda) \triangleq \lambda p(\lambda)$ is the unscaled revenue rate function. We denote by $c^* \triangleq \max\{c(\lambda): \lambda \in [0, \Lambda]\}$ the maximum unscaled revenue rate, which is guaranteed to exist given the continuity of $c(\lambda)$ in $[0, \Lambda]$. Without loss of generality, and for the rest of the paper, we normalize the unscaled revenue rate function (by adequately adjusting the scale factor θ) so that $c^* = rR$.

We interpret $W(n; \theta)$ as the value function for the associated dynamic programming formulation, which measures the expected discounted cumulative revenue when the current inventory level is n and the demand scale factor is θ . Observe that W includes revenues from both the current product and future ones (captured by R).

Invoking standard stochastic control arguments (Chapter VII, Brémaud 1980), we get the first-order optimality

condition for this value function in the form of the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\max_{0 \leq \lambda \leq \Lambda} \{-\lambda \theta (W(n; \theta) - W(n-1; \theta)) - rW(n; \theta) + \theta c(\lambda)\} = 0, \quad W(0; \theta) = R. \quad (7)$$

To solve this HJB equation, we find it convenient to rewrite it as follows:

$$\frac{rW(n; \theta)}{\theta} = \Psi(W(n-1; \theta) - W(n; \theta)) \quad \text{and} \quad W(0; \theta) = R, \quad \text{where } \Psi(z) \triangleq \max_{0 \leq \lambda \leq \Lambda} \{\lambda z + c(\lambda)\}. \quad (8)$$

The function $\Psi(\cdot)$ defined on the real line is nonnegative and monotonically increasing. It admits an inverse function given by $\Phi(z) \triangleq \Psi^{-1}(z)$ ($z \in \mathbb{R}_+$). The function $\Psi(\cdot)$ is known as the Fenchel-Legendre transform of $c(\cdot)$ and has been extensively studied in the context of convex analysis (see Rockafellar 1997). For future reference, we also define the function

$$\zeta(z) \triangleq \inf \left\{ \bar{\lambda} \in [0, \Lambda]: \bar{\lambda} = \arg \max_{0 \leq \lambda \leq \Lambda} \{\lambda z + c(\lambda)\} \right\}. \quad (9)$$

This function $\zeta(z)$ is nondecreasing and satisfies $\zeta(0) = \lambda^* \triangleq \arg \max\{c(\lambda): \lambda \in [0, \Lambda]\}$. Figure 1 plots ζ , and the Fenchel-Legendre transforms Ψ and Φ for the case of an exponential demand rate (for further details see §C.1 in Appendix C). We note that $\Psi(0) = c^*$ and $\Phi(c^*) = 0$. Based on Equation (8), we can compute the value function iteratively through the recursion

$$W(0; \theta) = R \quad \text{and} \quad W(n; \theta) + \Phi\left(\frac{rW(n; \theta)}{\theta}\right) = W(n-1; \theta), \quad n = 1, 2, \dots \quad (10)$$

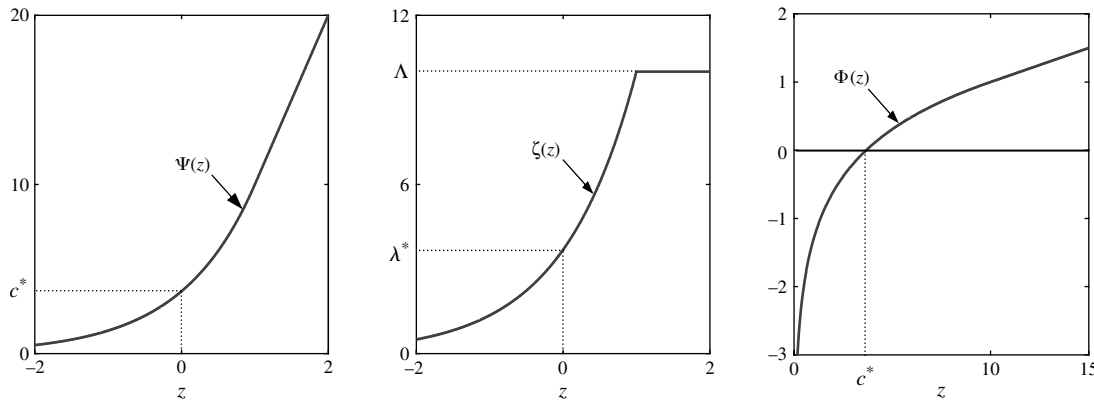
To complete the characterization of the optimal solution, the optimal demand intensity $\lambda^*(n; \theta)$ for an inventory of n is given by

$$\lambda^*(n; \theta) = \zeta(W(n-1; \theta) - W(n; \theta)). \quad (11)$$

Using the monotonicity of $\Phi(z)$, the recursion in (10), and our scaling condition $c^* = rR$, we obtain the following result.

PROPOSITION 1. *For every $\theta > 0$ and $R \geq 0$ there is a unique solution $W(n; \theta)$ to the recursion (10) that is monotonically increasing in θ and satisfies $\lim_{n \rightarrow \infty} W(n; \theta) = \theta R$. If the scale factor $\theta \geq 1$ (respectively $\theta \leq 1$), then the value function $W(n; \theta)$ is increasing and concave (respectively decreasing and convex) as a function of n .*

Figure 1. Fenchel-Legendre transforms for the case of an exponential demand rate $\lambda(p) = \Lambda \exp(-\alpha p)$ with $\Lambda = 10$ and $\alpha = 1$.



PROOF. See §A.1 in Appendix A. \square

Proposition 1 highlights the effect of the scale parameter θ on revenues. For $\theta \geq 1$, the revenue function $W(n; \theta)$ is always larger than R , and increases with the inventory level. The opposite conclusion holds for $\theta \leq 1$. Based on this distinction, we say that a product is *high revenue* (or profitable) if $\theta \geq 1$, and we say that a product is *low revenue* (or unprofitable) if $\theta \leq 1$. From now on, we assume that $\theta_L \leq 1 \leq \theta_H$.

The difference between high-revenue and low-revenue products comes from the underlying trade-off that the seller experiences in terms of present and future revenues. In our model, the quantity R captures the future value of the seller's operations after the current product has been depleted. Therefore, for a given discount rate r , the term rR represents the seller's average revenue rate from future businesses. On the other hand, the revenue rate generated by the current product is $\theta c(\lambda)$, for a demand intensity $\theta\lambda$. Thus, the seller considers the current operations to be more profitable than the average future business if $\max_{\lambda} \{\theta c(\lambda)\} \geq rR$ or, equivalently, $\theta c^* \geq rR$. Given the normalization $c^* = rR$, this condition reduces to $\theta \geq 1$. In other words, for $\theta \geq 1$ the current product offers higher returns than the average product that the seller usually sells, and so the value function increases with n ; in this case, the retailer will always choose to sell this product until no more units are available. On the other hand, if $\theta \leq 1$, then the seller would like to switch as soon as possible from the current product to a new (more profitable) alternative. If the seller has to deplete all units before switching to another product, then the corresponding value function is a decreasing function of the inventory. In other words, the more units of this low-revenue product the seller has, the longer it is going to take to sell them all and move to a better product. However, if the retailer can stop selling the product at any time, then for $\theta < 1$ he chooses to stop immediately, i.e., $\tau = 0$.

An example of the results in Proposition 1 is depicted on the left panel in Figure 2. The right panel shows the

corresponding optimal demand intensity $\lambda^*(n, \theta)$ that we discuss in Corollary 1 below. Besides the monotonicity and convexity properties of the value function, Figure 2 also confirms the asymptotic behavior as the inventory grows large. Specifically, we have that $W(n; \theta) \rightarrow \theta R$ as $n \rightarrow \infty$. Interestingly, Proposition 1 holds true without requiring any specific condition (such as concavity) on the revenue rate function $c(\lambda)$. This is a distinguishing feature of our representation of the value function in (8) in terms of the Fenchel-Legendre transform Ψ and its inverse Φ . Indeed, it is well known that Ψ is unaffected if we replace $c(\lambda)$ by its *concave hull* in Equation (8). The following corollary follows from Proposition 1.

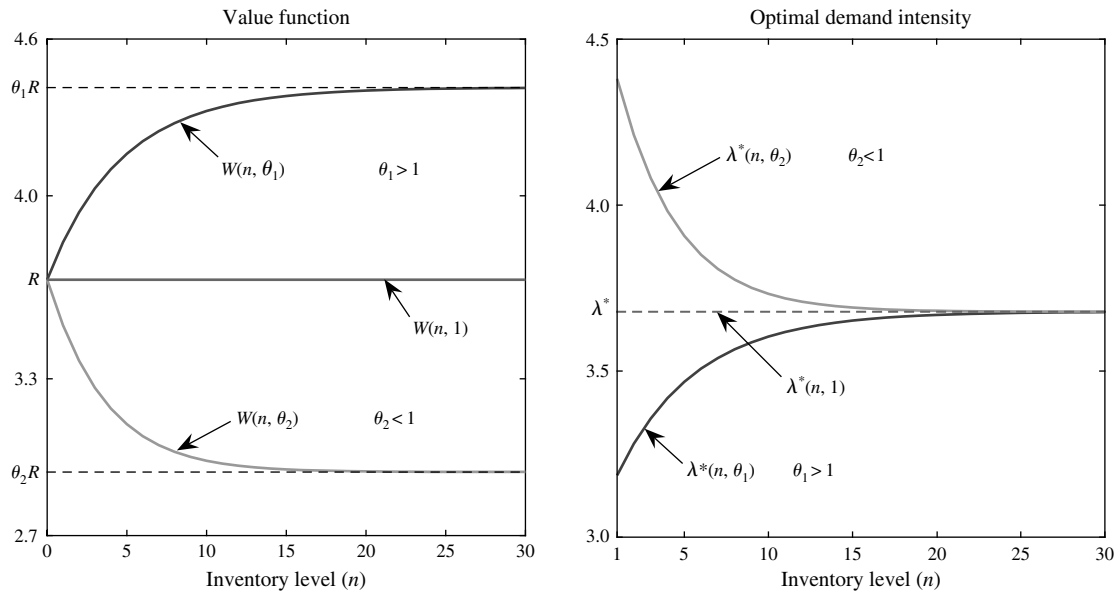
COROLLARY 1. Let $\lambda^* = \arg \max \{c(\lambda) : \lambda \in [0, \Lambda]\}$ be its unique maximizer. The optimal demand intensity $\lambda^*(n; \theta)$ is monotonically decreasing in θ and satisfies $\lim_{n \rightarrow \infty} \lambda^*(n; \theta) = \lambda^*$. If $\theta \geq 1$ (respectively, $\theta \leq 1$), then $\lambda^*(n; \theta)$ increases (respectively, decreases) with the level of inventory n .

PROOF. The result follows directly from concavity (respectively, convexity) of $W(n; \theta)$ in Proposition 1, Equation (11), and the monotonicity of $\zeta(\cdot)$. \square

From a pricing perspective, we note that for a low-revenue product the price increases with the available stock. This is in contrast to most of the dynamic pricing literature (e.g., Gallego and van Ryzin 1994), which is more in synch with our high-revenue product where optimal prices decrease with the inventory level. This apparently counterintuitive result relies on a simple observation. In our setting, the retailer's trade-off is current versus future revenues. As the initial stock increases, the time required to deplete these units goes up as well. As a result, the retailer is better off maximizing current revenues and minimizing future revenues by increasing the price. In contrast, for high-revenue products the price decreases with inventory.

The different pricing behavior between low- and high-revenue products raises an important issue regarding depletion time, specifically, whether we are selling faster when θ

Figure 2. Value function (left panel) and optimal demand intensity (right panel) for two values of θ under an exponential demand model $\lambda(p) = \Lambda \exp(-\alpha p)$.



Note. The data used is $\Lambda = 10$, $\alpha = 1$, $r = 1$, $\theta_1 = 1.2$, $\theta_2 = 0.8$, $R = \Lambda \exp(-1)/(\alpha r) \approx 3.68$.

is larger. In fact, even if low-revenue products have lower prices than high-revenue products, their demand scale factor, θ , is smaller. Hence, the net effect on the net demand rate $\theta\lambda(p)$ is unclear. According to Corollary 1, for n sufficiently large the pricing policies for both low- and high-revenue products are almost identical, and so the effective rate of sales increases with θ . The following proposition shows that under mild conditions on the demand model (condition (12) below), this conclusion holds for all inventory levels.

PROPOSITION 2. Let $s^*(n; \theta) \triangleq \theta\lambda^*(n; \theta)$ be the optimal rate of sales for a given θ and inventory level n . If

$$\frac{d}{d\lambda}(\lambda p'(\lambda)) \leq 0, \tag{12}$$

then the sales rate $s^*(n; \theta)$ increases with θ for all n .

PROOF. See §A.2 in Appendix A. \square

Condition (12) on the pricing function $p(\cdot)$ is not particularly restrictive, and it is satisfied by the three demand models (exponential, linear, and quadratic) that we describe in §C.1 in Appendix C. (A simple derivation of this condition translates to a slightly stronger requirement on $c(\cdot)$ than just concavity.) Interestingly, according to this proposition, even if prices increase with θ the net demand rate, $\theta\lambda^*$, still increases with θ . In other words, the inventory turns of high-revenue products are higher than those of low-revenue products even though the former are sold at a higher price than the latter.

As a side remark, we can get an alternative interpretation of condition (12) using the notion of *reservation price* (e.g.,

Bitran and Mondschein 1997). Suppose that every arriving buyer has a maximum price that he is willing to pay for the product. The seller is unable to observe this reservation price, but only knows its probability distribution (F) among the population of buyers. In this setting, if the seller charges a price p the resulting demand intensity equals $\lambda(p) = \Lambda(1 - F(p))$ with corresponding inverse demand function $p(\lambda) = F^{-1}(1 - \lambda/\Lambda)$. For example, if the reservation price is exponentially distributed with parameter α , then we recover the exponential demand model $\lambda(p) = \Lambda \exp(-\alpha p)$, and if the reservation price is uniformly distributed in $[0, \Lambda/\alpha]$, then we recover the linear demand model $\lambda(p) = \Lambda - \alpha p$. With this interpretation of the demand process, it is a matter of simple calculations to show that condition (12) is equivalent to the reservation price distribution (F) having weakly increasing failure rate (IFR) (e.g., Lariviere 2006).

We conclude this section with a brief summary of our findings under full information. According to our model, the seller can partition the set of products that he sells in two groups: (i) *high-revenue products* for which $\theta \geq 1$, and (ii) *low-revenue products* for which $\theta \leq 1$. High-revenue products sell faster (higher inventory turns) and at a higher price than their low-revenue counterparts. Hence, if the seller were able to identify which products offer high revenues and which do not, then he would never engage in procuring and selling low-revenue products. In practice, however, the seller is rarely capable of perfectly anticipating the selling pattern of a given product. This pattern, which depends on customers' preferences and market competition, only reveals itself as the selling season progresses, long after procurement decisions are made.

With this problem in mind, we study in the following section optimal pricing strategies for the case in which the seller has imperfect knowledge about customers’ preferences, or, in our case, the value of the scale factor θ .

4. Dynamic Pricing with Incomplete Demand Information

In this section, we consider the case in which the retailer starts the selling horizon having only partial information about the demand scale factor θ . We again consider the case in which θ can take only two values $\{\theta_L, \theta_H\}$, with $\theta_L \leq 1 \leq \theta_H$. (A generalization to the case of a multidimensional vector θ is discussed in §6.) This is the most interesting case in the sense that the retailer cannot tell whether the product being sold is a high-revenue ($\theta = \theta_H > 1$) or a low-revenue ($\theta = \theta_L < 1$) product. The retailer starts the selling season with a prior belief q that $\theta = \theta_L$. We also assume in this section that all initial N_0 units must be depleted before a different product can be offered. This final assumption is relaxed in §5.

The setting here describes, for example, those situations where the retailer is bringing a new product into the market and has uncertain information about how well this product will sell. As the selling period progresses and the demand process materializes, the retailer updates his information and adjusts the price accordingly to maximize cumulative discounted profits. This *active learning* process is essentially a Bayes update on the distribution of θ while the retailer is only observing the sales process over time. It is active in the sense that the optimal price is not only a result of the current belief, but also on how it will evolve in the future.

In formal terms, we embed the model in this section in a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}_q)$. The probability measure \mathbb{P}_q satisfies (see §2 for notation)

$$\mathbb{P}_q = q\mathbb{P}_{\theta_L} + (1 - q)\mathbb{P}_{\theta_H}.$$

Given the retailer’s initial beliefs q , the random variable θ satisfies $\mathbb{P}_q(\theta = \theta_L) = 1 - \mathbb{P}_q(\theta = \theta_H) = q$. We let $q_t \triangleq \mathbb{P}_q(\theta = \theta_L | \mathcal{F}_t)$ be the retailer’s belief about the value of θ at time t conditional on the available information \mathcal{F}_t . Recall that $(\mathcal{F}_t; t \geq 0)$ is the filtration generated by the inventory (or equivalently sales) process $\{N(t) = N_0 - \int_0^t dD_s; t \geq 0\}$. Note also that the process $\{(q_t, \mathcal{F}_t); t \geq 0\}$ is by definition a \mathbb{P}_q -martingale.

In this setting, the retailer problem becomes

$$V(N_0, q) = \sup_{\lambda \in \mathcal{A}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt) \theta c(\lambda_t) dt + \exp(-r\tau) R \right],$$

$$\tau = \inf\{t \geq 0: N_t = 0\}. \quad (13)$$

We will tackle a solution to (13) using dynamic programming. For this, we will first derive the specific dynamics of q_t using Bayes’s rule and Itô’s lemma.

PROPOSITION 3. *The \mathbb{P}_q -martingale (belief) process $\{(q_t, \mathcal{F}_t); t \geq 0\}$ satisfies the stochastic differential equation*

$$dq_t = -\eta(q_{t-})[dD_t - (\theta_L q_{t-} + \theta_H(1 - q_{t-}))\lambda_t dt],$$

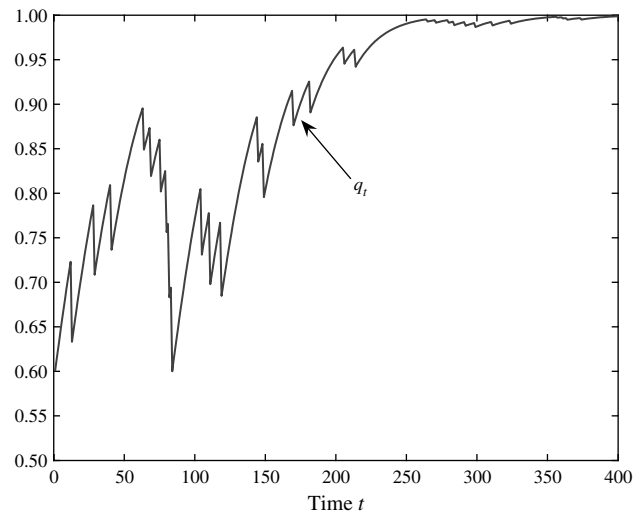
$$\text{where } \eta(q_t) \triangleq \frac{q_t(1 - q_t)(\theta_H - \theta_L)}{\theta_L q_t + \theta_H(1 - q_t)}. \quad (14)$$

PROOF. See §A.3 in Appendix A. \square

According to (14), the rate at which the seller’s beliefs change depends on the difference between the observed demand rate, dD_t , and the expected demand rate, $(\theta_L q_{t-} + \theta_H(1 - q_{t-}))\lambda_t dt$, given the available information. Loosely speaking, the martingale nature of q_t follows from (14) by noticing that $\mathbb{E}_q[dD_t | \mathcal{F}_t] = (\theta_L q_{t-} + \theta_H(1 - q_{t-}))\lambda_t dt$. Observe as well that q_t is a jump process driven by the Poisson demand, and as a bounded martingale will converge to q_∞ \mathbb{P}_q -a.s. (q_∞ is a random variable that takes under \mathbb{P}_q the value 1 with probability q and 0 with probability $1 - q$.) As long as no sales occur, q_t increases deterministically towards one; the process jumps downward by a factor of $\eta(q_{t-})$ when a sale occurs. These jumps depend on the value of the belief and tend to zero as q approaches either zero or one (see Figure 3, which depicts a pathwise evolution of the belief process under a constant price policy.) As we should expect, Equation (14) also reveals that in all three cases— $q = 0$, $q = 1$, and $\theta_L = \theta_H$ —the beliefs of the seller are actually constant over time, which brings us back to the model of the previous section with perfect demand information.

Another important feature of the belief process in (14) is that it implies that learning is maximized when the demand rate is set to its maximum value $\lambda_t = \Lambda$ for all t . Indeed, for a given a pricing strategy $\lambda = (\lambda_t)_{t \geq 0}$ the corresponding

Figure 3. Values of pathwise q_t under a constant price strategy $\lambda_t = 1$.



Note. The data used are starting inventory $n = 35$, $\theta_L = 0.8$, $\theta_H = 1.2$.

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

likelihood ratio process associated with the simple hypotheses $H_H = \{\theta = \theta_H\}$ and $H_L = \{\theta = \theta_L\}$ is equal to (see Brémaud 1980)

$$\mathcal{L}_t \triangleq \frac{d(\mathbb{P}_{\theta_L} | \mathcal{F}_t)}{d(\mathbb{P}_{\theta_H} | \mathcal{F}_t)} = \left(\frac{\theta_L}{\theta_H}\right)^{D_t} \exp((\theta_H - \theta_L)I_\lambda(t)), \quad (15)$$

where $\mathbb{P}_\theta | \mathcal{F}_t$ denotes the restriction of \mathbb{P}_θ to \mathcal{F}_t . Hence, for any history \mathcal{F}_t the likelihood-ratio process is maximized if we choose a pricing strategy λ that maximizes $I_\lambda(t)$, that is, setting $\lambda_t = \Lambda$ for all t .

Of course, we are not interested in choosing a pricing strategy that maximizes the seller's learning, but one that maximizes the discounted expected payoff in Equation (13). Note that the retailer controls the unscaled demand rate λ_t , whereas the actual rate realized is in fact $\theta\lambda_t$, which in turns induces a revenue rate of $\theta c(\lambda_t)$. The revenue rate function $c(\cdot)$ satisfies the same set of assumptions as in the previous section. Therefore, in such a context, the problem's formulation can be written as follows:

$$\begin{aligned} & V(N_0, q) \\ &= \sup_{\lambda \in \mathcal{A}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt) \theta c(\lambda_t) dt + \exp(-r\tau) R \right] \\ & \text{subject to } N_t = N_0 - \int_0^t dD(I_\lambda(s)), \\ & \quad dq_t = \eta(q_{t-}) [dD_t - (\theta_L q_{t-} + \theta_H(1 - q_{t-})) \lambda_t dt], \\ & \quad q_0 = q, \\ & \quad \tau = \inf\{t \geq 0: N_t = 0\}. \end{aligned}$$

The value function associated with the dynamic programming formulation is now $V(n, q)$, where the state variables are the level of inventory n and the seller's beliefs q . We define $\bar{\theta}(q_t) \triangleq \mathbb{E}_q[\theta | \mathcal{F}_t] = \theta_L q_t + \theta_H(1 - q_t)$ to be the expected demand scale factor given the available information at time t .

The HJB equation is then given by (see Appendix C.2 for a derivation)

$$\begin{aligned} rV(n, q) = \max_{0 \leq \lambda \leq \Lambda} & [\lambda \bar{\theta}(q) [V(n-1, q - \eta(q)) - V(n, q)] \\ & + \lambda \kappa(q) V_q(n, q) + \bar{\theta}(q) c(\lambda)], \quad (16) \end{aligned}$$

with $\kappa(q) \triangleq q(1 - q)(\theta_H - \theta_L)$ and boundary conditions $V(0, q) = R$, $V(n, 0) = W(n; \theta_H)$, and $V(n, 1) = W(n; \theta_L)$. Recall that the function $W(n; \theta)$ is the value function when there is no uncertainty about the true value of θ and is computed using the recursion in Equation (10).

As in Equation (10), we can rewrite the HJB condition using the Fenchel-Legendre transform of $c(\lambda)$ in the following convenient form:

$$\begin{aligned} & V(0, q) = R, \\ & V(n, q) + \Phi\left(\frac{rV(n, q)}{\bar{\theta}(q)}\right) - \eta(q) V_q(n, q) \\ &= V(n-1, q - \eta(q)), \quad n = 1, 2, \dots \end{aligned} \quad (17)$$

It also follows from Equations (16) and (17) that the optimal demand intensity $\lambda_V^*(n, q)$ satisfies

$$\lambda_V^*(n, q) = \zeta \circ \Phi\left(\frac{rV(n, q)}{\bar{\theta}(q)}\right), \quad (18)$$

where $\zeta(\cdot)$ is defined in (9) and “ \circ ” stands for the composition operator.

In general, we have not been able to solve explicitly the difference-differential equation in (17) to derive the corresponding optimal pricing policy. However, its recursive structure suggests the following algorithm to compute $V(n, q)$.

Algorithm -V

Step 1. Initialization: Set $V(0, q) = R$ for all $q \in [0, 1]$ and $n = 1$.

Step 2. Iteration: Set $F(q) = V(n-1, q - \eta(q))$ and solve the following ordinary differential equation (ODE) for $G(q)$ in the domain $q \in [0, 1]$:

$$\begin{aligned} & G(q) + \Phi\left(\frac{rG(q)}{\bar{\theta}(q)}\right) - \eta(q)G'(q) = F(q), \quad (19) \\ & G(0) = W(n; \theta_H), \quad G(1) = W(n; \theta_L). \end{aligned}$$

(Appendix D describes a finite-difference method that we use to solve this ODE.)

Set $V(n, q) = G(q)$ and $n = n + 1$.

Step 3. Go to Step 2.

The main step in this algorithm is to solve the ODE in Step 2. This is not a straightforward task because the border conditions at $q = 0$ and $q = 1$ are singular points for the differential equation because $\eta(0) = \eta(1) = 0$. Hence, even the existence of a solution to (19) is a subtle issue. Fortunately, the following proposition takes care of this problem.

PROPOSITION 4. *There exists a unique sequence of functions, $\{V(n, \cdot): n \geq 1\}$, defined on $[0, 1]$ and satisfying the system of Equations (17) with border conditions $V(n, 0) = W(n; \theta_H)$ and $V(n, 1) = W(n; \theta_L)$.*

PROOF. The proof of this proposition requires a number of intermediate steps and can be found in Appendix B. \square

Despite the fact that we do not have an analytical solution to (17), this optimality condition provides enough information to derive some useful properties that we use to approximate the value function and the corresponding pricing strategy.

PROPOSITION 5. (i) *The value function $V(n, \cdot)$ is monotonically decreasing and convex in q . It is also bounded by the perfect information values for all $n \geq 1$ and $q \in [0, 1]$:*

$$W(n; \theta_L) \leq V(n, q) \leq W(n; \theta_H).$$

(ii) *Furthermore, $V(n, \cdot)$ converges uniformly to the linear function $R\bar{\theta}(\cdot)$ as $n \rightarrow \infty$, and*

$$\lim_{n \rightarrow \infty} \lambda_V^*(n, q) = \lambda^*.$$

PROOF. See §A.4 in Appendix A. \square

As expected, part (i) of the proposition shows that the value function decreases with q and it is bounded by the value function in the full-information case when $\theta = \theta_L$ and $\theta = \theta_H$. The asymptotic result in part (ii) shows that the optimal demand intensity converges to λ^* , which maximizes the instantaneous revenue rate. Hence, as n gets large the retailer favors revenue exploitation over demand exploration when selecting the optimal selling rate. The asymptotic result also shows that the value function converges to the linear function $R(\theta_L q + \theta_H(1 - q))$ as the number of initial units grows to infinity. This limiting behavior suggests a simple method to approximate the value function, which we undertake in the following subsection.

Before jumping into this asymptotic analysis, let us use the result in Proposition 5 to extend the result in Proposition 2 to this case with an unknown θ . For this, we define $s^*(n, q) = \theta(q)\lambda_V^*(n, q)$ to be the expected selling rate when the inventory is n and the belief process is equal to q . As in the full-information case, the following proposition reveals that $s^*(n, q)$ increases with the (expected) market size $\theta(q)$ even if optimal prices are increasing in $\theta(q)$.

PROPOSITION 6. *Suppose that the demand function satisfies*

$$\frac{d}{d\lambda}(\lambda p'(\lambda)) \leq 0.$$

Then, the sales rate $s^(n, q)$ decreases with q for all n .*

PROOF. See §A.5 in Appendix A. \square

4.1. Asymptotic Approximation

Based on Proposition 5, it seems that (for a fixed inventory level n) $V(n, q)$ is well approximated by a linear function of q . In particular, we consider for each $n \geq 1$ and $q \in [0, 1]$, the following approximation of $V(n, q)$:

$$\begin{aligned} \tilde{V}(n, q) &\triangleq \mathbb{E}_q[W(n, \theta)] \\ &= qW(n, \theta_L) + (1 - q)W(n, \theta_H). \end{aligned} \quad (20)$$

In what follows, we will use the tilde ($\tilde{\cdot}$) notation to denote the asymptotic approximation of quantities such as the value function in (20) or the demand intensity in (21).

The next result shows that the linear approximation is not only suggested by the limiting result on $V(n, \cdot)$, but it also represents an upper bound for the value function. More importantly, it approaches the value function in a strong sense, i.e., their ratio goes to one uniformly in q . Combining Propositions 1 and 5, we obtain that:

PROPOSITION 7. *The approximation in (20) defines an upper bound of the value function, i.e.,*

$$V(n, q) \leq \tilde{V}(n, q),$$

for all $q \in [0, 1]$ and for all n . Furthermore, the approximation is asymptotically and uniformly (in q) exact, as n goes to infinity. That is, $|V(n, q)/\tilde{V}(n, q)| \rightarrow 1$ uniformly in q as n goes to ∞ . Note also that under perfect information $V(n, q) = \tilde{V}(n, q)$ for $q \in \{0, 1\}$ or $\theta_L = \theta_H$.

PROOF. The upper bound is due to the convexity of V in q . Because of the boundedness of V , the uniform convergence of the ratio is guaranteed if the difference converges uniformly to zero. Using triangle inequality, we write

$$|V(n, q) - \tilde{V}(n, q)| \leq |V(n, q) - R\bar{\theta}(q)| + |R\bar{\theta}(q) - \tilde{V}(n, q)|.$$

Both terms on the right converge to zero uniformly in q —the first one through Proposition 5. The second term is smaller than $R\theta_H - W(n; \theta_H) + W(n; \theta_L) - R\theta_L$, which is independent of q and converges to zero. \square

Let us turn to the pricing strategy. The asymptotic approximation in (20) works directly with the value function, and thus it is unclear how to estimate the optimal demand rate $\lambda_V^*(n, q)$. To fill this gap, we propose using the optimality condition in (18), using $\tilde{V}(n, q)$ instead of $V(n, q)$. It follows from the linearity of $\tilde{V}(n, q)$ in q that the proposed approximation for $\lambda^*(n, q)$ is given by

$$\begin{aligned} \tilde{\lambda}(n, q) &= \zeta\left((q - \eta(q))(\Delta W(n, \theta_H) - \Delta W(n, \theta_L)) \right. \\ &\quad \left. - \Delta W(n, \theta_H)\right), \end{aligned} \quad (21)$$

where $\Delta W(n, \theta) = W(n, \theta) - W(n - 1, \theta)$.

REMARKS. 1. Because $\zeta(z)$ increases with z , $q - \eta(q)$ increases with q , and $\Delta W(n, \theta_H) \geq 0 \geq \Delta W(n, \theta_L)$, it follows that $\tilde{\lambda}(n, q)$ is increasing in q .

2. Furthermore, because $\zeta(0) = \lambda^*$, we have that

$$\begin{aligned} \tilde{\lambda}(n, q) &\geq \lambda^* \quad \text{if and only if} \\ q - \eta(q) &\geq \frac{\Delta W(n, \theta_H)}{\Delta W(n, \theta_H) - \Delta W(n, \theta_L)}. \end{aligned}$$

3. Using the convexity of V and the fact that \tilde{V} is an upper bound of V , we get that

$$\begin{aligned} R - V(1, q) + \eta(q)V_q(1, q) &\geq R - V(1, q - \eta(q)) \\ &\geq R - \tilde{V}(1, q - \eta(q)). \end{aligned}$$

If we apply ζ (which is an increasing function) to both sides, we conclude that

$$\lambda_V^*(1, q) \geq \tilde{\lambda}(1, q).$$

That is, the asymptotic approximation overprices the optimal solution for $n = 1$. Unfortunately, for $n \geq 2$ we have not been able to prove (or disprove) a similar claim.

Let us now assess the performance of the asymptotic approximation by comparing the optimal expected discounted revenue $V(n, q)$ to the one obtained using the demand rate $\tilde{\lambda}(n, q)$. Also, to measure the performance of our approximation with respect to other alternative policies, we consider the following three heuristics.

1. MYOPIC POLICY. The popular myopic (or certainty equivalent) approximation of the value function is defined as

$$V^0(n, q) \triangleq W(n, \mathbb{E}_q(\theta)) = W(n, \bar{\theta}(q)).$$

We note that this policy is asymptotically optimal in the sense that $V^0(n, q)$ converges to $R\bar{\theta}(q)$ as n goes to infinity. We call this approximation myopic because it models the discounted profit that a retailer would expect to get if he myopically considers the expected value $\bar{\theta}(q)$ to be the true value of the scale factor θ . As opposed to our original active learning strategy, such strategy falls into the category of passive learning. Like our asymptotic policy, this myopic policy does not generate a pricing policy directly. It rather proposes an approximation for the value function that we need to translate into an implementable pricing strategy. Again, we can use the optimality condition (18) to get a demand rate associated with this myopic policy:

$$\lambda^0(n, q) = \zeta(V^0(n-1, q-\eta(q)) - V^0(n, q) + \eta(q)V_q^0(n, q)).$$

We note that despite its simplicity, the computational effort required to compute the myopic policy is substantially higher than the one needed for the asymptotic policy. Indeed, our asymptotic approximation is fully characterized by $2(N_0 + 1)$ values $\{(W(n, \theta_L), W(n, \theta_H)): 0 \leq n \leq N_0\}$, whereas the myopic policy is defined by $N_0 + 1$ functions $\{W(n, \bar{\theta}(q)): 0 \leq n \leq N_0 \text{ and } q \in [\theta_L, \theta_H]\}$.

2. SINGLE-PRICE POLICY. Another popular approximation in the revenue management literature is the single-price policy. Under this approximation, the price is kept fixed for the entire planning horizon. The popularity of this approximation comes from (i) its simplicity from an implementation point of view, and (ii) its asymptotic optimality in certain settings with large initial inventory and large demand rate (e.g., Gallego and van Ryzin 1994 or Bitran and Caldentey 2003). Let us denote by $V^1(n, q; \lambda)$ the retailer's expected discounted payoff, starting with n units of inventory and a belief of q if the fixed-price policy $\lambda_t = \lambda$ is used. It follows that

$$\begin{aligned} V^1(n, q; \lambda) &= \mathbb{E}_q \left[\int_0^\tau e^{-rt} \theta c(\lambda) dt + e^{-r\tau} R \right] \\ &= \mathbb{E}_q \left[\left(\frac{\theta c(\lambda)}{r} \right) (1 - e^{-r\tau}) + e^{-r\tau} R \right] \\ &= \frac{\bar{\theta}(q)c(\lambda)}{r} + \mathbb{E}_q \left[\left(R - \frac{\theta c(\lambda)}{r} \right) e^{-r\tau} \right] \\ &= \frac{\bar{\theta}(q)c(\lambda)}{r} + q \left(R - \frac{\theta_L c(\lambda)}{r} \right) \left(\frac{\lambda \theta_L}{r + \lambda \theta_L} \right)^n \\ &\quad + (1-q) \left(R - \frac{\theta_H c(\lambda)}{r} \right) \left(\frac{\lambda \theta_H}{r + \lambda \theta_H} \right)^n. \end{aligned}$$

The last equality uses the fact that under the probability measure \mathbb{P}_{θ_i} the selling horizon τ has a Gamma

distribution with parameters $(n, \lambda \theta_i)$, $i = L, H$. Therefore, $\mathbb{E}_{\theta_i}[e^{-r\tau}] = (\lambda \theta_i / (r + \lambda \theta_i))^n$ for $i = L, H$. The corresponding demand rate associated with this single-price approximation is given by

$$\lambda^1(n, q) = \arg \max_{\lambda \in [0, \Lambda]} V^1(n, q; \lambda).$$

It is worth noticing that this single-price policy is also asymptotically optimal in the sense that

$$\begin{aligned} \sup_{\lambda \in [0, \Lambda]} \lim_{n \rightarrow \infty} V^1(n, q; \lambda) &= \bar{\theta}(q)c(\lambda^*)/r = R\bar{\theta}(q) \\ &= \lim_{n \rightarrow \infty} V(n, q). \end{aligned}$$

3. TWO-PRICE POLICY. An important limitation of the previous approximation is its inability to adjust the price based on the realized demand. This is particularly serious in our setting, where the demand distribution is unknown. To partially address this limitation, and at the same time preserve the operational simplicity of the single-price policy, we consider a two-price policy in which the retailer is able to change the price only once. (Feng and Gallego 1995 provide structural properties of this type of policies under full-demand information in a finite-horizon setting.) A major difficulty for determining the optimal two-price policy is that it requires solving an optimal stopping time problem. From a computational standpoint, this is at least as demanding as computing the optimal value function. For this reason, we only consider a suboptimal version that makes a single price change right after the first unit is sold. The discussion of optimal pricing policies based on stopping time rules is postponed to §5. Under this restriction, let us denote by $V^2(n, q; \lambda)$ the retailer's expected discounted payoff starting with n units of inventory and a belief q when the initial demand intensity is set to $\lambda_t = \lambda$. It follows that

$$V^2(n, q; \lambda) = \mathbb{E}[e^{-r\tau_\lambda}[p(\lambda) + V^1(n-1, q_{\tau_\lambda})]],$$

where τ_λ is the (random) time at which the first unit is sold if the seller uses a fixed strategy $\lambda_t = \lambda$, $t \in [0, \tau_\lambda]$. The corresponding demand rate associated with this two-price approximation is given by

$$\lambda^2(n, q) = \arg \max_{\lambda \in [0, \Lambda]} V^2(n, q; \lambda).$$

Let us now compare the performance of the asymptotic approximation and the other three heuristics in terms of their relative error with respect to the optimal solution. If we let $\mathcal{V}(n, q)$ be the expected discounted payoff generated by any of these approximations (using the corresponding pricing policy), then the relative error is defined by

$$\mathcal{E}_{\mathcal{V}}(n, q) \triangleq \frac{V(n, q) - \mathcal{V}(n, q)}{V(n, q)} \times 100.$$

Table 1. Percentage relative error $\mathcal{E}_{\text{qr}}(n, q)$.

q	Inventory (n) (%)				
	1	5	10	50	100
Asymptotic approximation					
0.0	0.00	0.00	0.00	0.00	0.00
0.2	0.06	0.01	0.00	0.00	0.00
0.4	0.14	0.01	0.00	0.00	0.00
0.6	0.22	0.02	0.00	0.00	0.00
0.8	0.24	0.03	0.01	0.00	0.00
1.0	0.00	0.00	0.00	0.00	0.00
Myopic approximation					
0.0	0.00	0.00	0.00	0.00	0.00
0.2	17.53	23.85	24.28	21.78	17.27
0.4	30.90	46.41	48.78	39.31	29.46
0.6	44.11	61.57	59.75	44.49	33.12
0.8	50.47	63.91	61.24	43.91	32.51
1.0	0.00	0.00	0.00	0.00	0.00
Single-price policy					
0.0	0.00	0.20	0.31	0.29	0.07
0.2	0.81	0.95	0.82	0.57	0.34
0.4	1.35	1.67	1.25	0.60	0.33
0.6	1.30	2.32	1.71	0.64	0.32
0.8	0.72	2.56	1.99	0.67	0.29
1.0	0.03	1.60	1.03	0.863	0.87
Two-price policy					
0.0	0.00	0.02	0.04	0.02	0.11
0.2	0.11	0.22	0.22	0.16	0.08
0.4	0.17	0.37	0.32	0.20	0.11
0.6	0.17	0.54	0.44	0.24	0.14
0.8	0.11	0.71	0.55	0.28	0.15
1.0	0.03	0.88	0.84	0.86	0.87

Table 1 shows the average relative error for the four approximations. We compute this average over the three demand models (exponential, linear, and quadratic) described in Appendix C, uniformly varying the parameters Λ , θ_H , and θ_L in the ranges $[1, 20]$, $[1.1, 8]$, and $[0.1, 0.9]$, respectively, for a total of 225 different instances.

As we can see from Table 1, the asymptotic policy performs extremely well for the entire range of inventories (n) and beliefs (q) with an average error close to 0.03%. On the contrary, the myopic approximation performs quite poorly, especially for intermediate values of the inventory and belief; this is despite the fact that it is optimal for $q \in \{0, 1\}$. The average error of this myopic policy is close to 26.5%. The single-price and two-price policies offer a reasonably good performance across the board, with an average error of 0.9% and 0.3%, respectively (although an order of magnitude higher than the asymptotic policy). Thus, limiting the number of price changes can lead to good results especially for small values of the inventory. In conclusion, our proposed asymptotic policy is simple to compute (a linear function of q) and performs very well for the entire range of inventory.

In terms of implementation, we note that the performance of the asymptotic policy tends to degrade for small values of inventory. Hence, it seems reasonable to

implement a hybrid solution method that uses the asymptotic policy for large values of n and then switches to the computation of the optimal solution using Algorithm-V for small values of n .

We conclude this section with a brief discussion of the seller's preferences over different states (n, q). First, we note that (similarly to the full-information case) even though the cost of the initial units is sunk, it is not necessarily true that the value function is increasing in n , i.e., the retailer is not always better off with more units. Specifically, in the case where the initial belief q is near one, more units will delay the retailer from liquidating this low-revenue product. On the other hand, more units gives the retailer more time for learning. In the next proposition, we study the monotonicity of V through its approximation \tilde{V} and show that, for all $q < 1$, there exists an inventory threshold after which $\tilde{V}(\cdot, q)$ becomes increasing in n .

PROPOSITION 8. *For any fixed level of the prior $q < 1$, there exists a level of inventory, $n_0(q)$, such that the approximated value function $\tilde{V}(\cdot, q)$ is increasing in n for all $n \geq n_0$.*

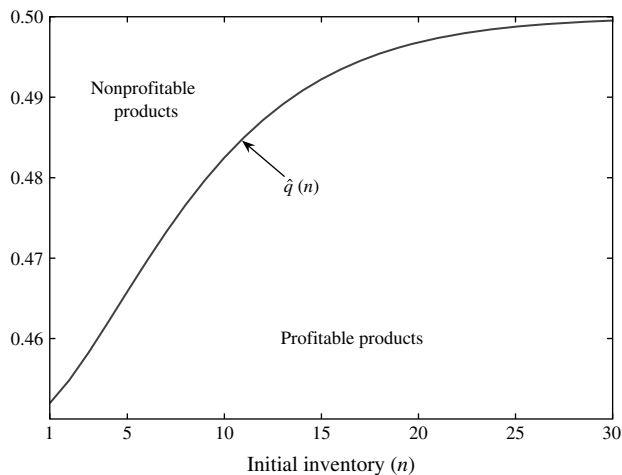
PROOF. See §A.6 in Appendix A. \square

The proof of the previous proposition is based on the following behavior, where simple calculations show that $W(n, \theta_L)$ converges faster to $R\theta_L$ than does $W(n, \theta_H)$ to $R\theta_H$. Recall that $\tilde{V}(n, q)$ is a linear combination of $W(n, \theta_L)$ (decreasing in n) and $W(n, \theta_H)$ (increasing in n). Hence, for n large enough, $\tilde{V}(n, q) \approx qR\theta_L + (1 - q)W(n, \theta_H)$, which is increasing in n . From a pricing perspective, Equation (18) implies that the optimal prices decrease eventually with the inventory level n , which implies a higher learning rate. We conclude that the monotonicity of the value function is the result of an increasing value of learning that is achieved for a large inventory level.

In the full-information case, the retailer is able to partition the products into two categories (high-revenue and low-revenue products), based on the value of θ compared to 1. In the incomplete information case, such partition depends on the initial belief and the inventory level. Prior to accepting a batch of n units of a product, the seller would like to compare his prior q , to the value $\hat{q}(n)$ solution to $V(n, q) = R$. This quantity defines the belief threshold between high-revenue and low-revenue products as a function of the initial stock. The monotonicity in q of the value function implies that (i) $\hat{q}(n)$ is unique, and (ii) in expectation, the retailer would be better off discarding the product if $q > \hat{q}(n)$. Observe as well that, in general, $\hat{q}(n)$ is different from $(\theta_H - 1)/(\theta_H - \theta_L)$, solution to $\tilde{\theta}(q) = 1$. However, as a result of Proposition 5—in the limit as the inventory gets large— $\hat{q}(n)$ converges to $(\theta_H - 1)/(\theta_H - \theta_L)$.

One can show, similarly to the proof of Proposition 8, that the solution to the equation $\tilde{V}(n, q) = R$ is monotone in n for n larger than a certain threshold. Therefore, one expects a similar behavior for $\hat{q}(n)$. Figure 4 plots the values of $\hat{q}(n)$ as a function of n for the case of an exponential

Figure 4. Value of $\hat{q}(n)$ for an exponential demand rate $\lambda(p) = \Lambda \exp(-\alpha p)$.



Note. The data used is $\Lambda = 10$, $\alpha = 1$, $r = 1$, $\theta_H = 1.2$, $\theta_L = 0.8$, $R = c^*/r \approx 3.68$.

demand rate. In this case, $\hat{q}(n)$ is indeed increasing in n . This monotonicity suggests that the seller is willing to take more risk (measured by an increase in q) for larger orders (measured by an increase in n). For example, in Figure 4, if the seller's initial belief is $q = 0.48$, then an order of $n = 5$ units is not attractive, whereas an order of $n = 15$ units becomes attractive. Intuitively, for small orders, the event that the product is high revenue (i.e., $\theta = \theta_H$) has a small impact on the cumulative discounted profit with respect to the opportunity cost R . In other words, the potential value for demand learning increases with the size of the order.

5. Dynamic Pricing Under the Optimal Stopping Time Rule

In many settings, a retailer that has acquired a certain number of units of a nonperishable product will carry on selling those units until they are sold out. However, in some cases the seller has the option of discontinuing the current sales at any random time. This can occur, for example, by moving the current product to a secondary market (or simply to another floor, like Filene's¹ basement).

In this section, we consider a similar setting to the one discussed in the previous section, but allowing the seller to stop the current sales and achieve the terminal value R at any point in time. We restrict ourselves to times that depend on the current history (i.e., stopping times). In the full-information case the seller chooses at time zero either to acquire the units or not. When learning is taken into account, the seller will pursue the business as long as the value function is greater than R and will decide to drop it as soon as the value function hits R . The case in which this terminal reward R depends on the number of unsold units at the time of stopping is discussed in §6.

The formulation of this problem in this case can be written as follows:

$$\begin{aligned}
 &U(N_0, q) \\
 &= \sup_{\lambda \in \mathcal{A}, \tau \in \mathcal{T}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt) \bar{\theta}(q_t) c(\lambda_t) dt + \exp(-r\tau) R \right] \\
 &\text{subject to } N_t = N_0 - \int_0^t dD(I_\lambda(s)), \quad (22) \\
 &\quad dq_t = \eta(q_{t-}) [dD_t - (\theta_L q_{t-} + \theta_H (1 - q_{t-})) dI_\lambda(t)], \\
 &\quad q_0 = q, \\
 &\quad \tau \leq \inf\{t \geq 0: N_t = 0\}.
 \end{aligned}$$

The optimality (HJB) equation is given by

$$\begin{aligned}
 \max \left\{ R - U(n, q), \Psi(U(n-1, q - \eta(q)) - U(n, q)) \right. \\
 \left. + \eta(q) U_q(n, q) - \frac{rU(n, q)}{\bar{\theta}(q)} \right\} = 0,
 \end{aligned}$$

which can also be written as follows:

$$\begin{cases} U(n, q) + \Phi\left(\frac{rU(n, q)}{\bar{\theta}(q)}\right) - \eta(q)U_q(n, q) \\ = U(n-1, q - \eta(q)) \quad \text{if } U(n, q) \geq R, \\ U(n, q) + \Phi\left(\frac{rU(n, q)}{\bar{\theta}(q)}\right) - \eta(q)U_q(n, q) \\ \leq U(n-1, q - \eta(q)) \quad \text{if } U(n, q) = R. \end{cases} \quad (23)$$

We denote by q_n^* the smallest value of q for which $U(n, q) = R$. The following proposition shows, among other things, that $U(n, q) = R$ for all $q \geq q_n^*$. Hence, an optimal pricing strategy is only defined on the continuation region $q \in [0, q_n^*)$ and satisfies

$$\lambda_U^*(n, q) = \zeta \circ \Phi\left(\frac{rU(n, q)}{\bar{\theta}(q)}\right).$$

PROPOSITION 9. (i) *The system of equations given by (23) admits a unique continuously differentiable solution $U(n, \cdot)$ defined on $[0, 1]$ such that $U(n, q) > R$ on $[0, q_n^*)$ and $U(n, q) = R$ on $[q_n^*, 1]$, where q_n^* is the unique solution of the smooth pasting condition*

$$R + \Phi\left(\frac{rR}{\bar{\theta}(q)}\right) = U(n-1, q - \eta(q)).$$

(ii) *The value function $U(n, \cdot)$ is decreasing and convex in q on $[0, 1]$.*

(iii) *The sequence $(U(n, \cdot): n \geq 1)$ is increasing in n and satisfies for all $n \geq 1$ and $q \in [0, 1]$,*

$$R \leq U(n, q) \leq W(n, \theta_H).$$

(iv) Let $s^*(n, q) = \bar{\theta}(q)\lambda_U^*(n, q)$ be the expected selling rate. Then, if the demand function satisfies

$$\frac{d}{d\lambda}(\lambda p'(\lambda)) \leq 0,$$

then the sales rate $s^*(n, q)$ decreases with q for all n .

(v) Let $\lambda_U^*(n, q)$ and $\lambda_V^*(n, q)$ be the optimal demand rate for the cases where the option to stop is and is not available, respectively. Then, for all n and q ,

$$\lambda_V^*(n, q) \leq \lambda_U^*(n, q).$$

PROOF. See §A.7 in Appendix A. \square

The previous proposition shows that most properties of the value function are maintained when the option of stopping is permitted. A fundamental difference, however, is that U is increasing in n for all q , as opposed to V that might be decreasing in n , for some values of n and large values of q . Indeed, with the option of stopping available, one can do at least as well with $n + 1$ units than with n (under the assumption that the cost of the initial inventory is sunk). It should also be clear that the value function when the option of stopping is not allowed represents a lower bound for U , i.e., $V(n, q) \leq U(n, q)$ for all $n \geq 0$ and $q \in (0, 1)$. Part (v) in the proposition follows directly from this inequality. Intuitively, this result follows from the fact that the value of demand learning is higher when the option to stop is available, which gives the retailer more incentives to set lower prices to learn faster (see Equation (15) and the discussion that follows it).

Now, we suggest the following algorithm to compute the value function.

Algorithm -U

Step 1. Initialization: Set $U(0, q) = R$ for all $q \in [0, 1]$ and $n = 1$.

Step 2. Iteration: Set $F(q) = U(n - 1, q - \eta(q))$ and
 (i) solve for the unique solution of

$$R + \Phi\left(\frac{rR}{\bar{\theta}(q)}\right) = F(q),$$

set q_n^* to be this solution.

(ii) Solve the following ordinary differential equation (ODE) for $G(q)$ in the domain $q \in [0, q_n^*]$:

$$G(q) + \Phi\left(\frac{rG(q)}{\bar{\theta}(q)}\right) - \eta(q)G'(q) = F(q), \tag{24}$$

$$G(q_0) = R.$$

(iii) Set $U(n, q) = G(q)$ for $q \leq q_n^*$ and $U(n, q) = R$ otherwise. Set $n = n + 1$.

Step 3. Go to Step 2.

Again, the main step in this algorithm is solving the ODE in Equation (24). The task here is simpler than in §4 because the border condition is well defined; that is, the ODE does not have a singularity at q_n^* and can be solved using standard methods (e.g., Picard iteration). Appendix D describes a finite-difference scheme that can be used to solve this ODE.

We now discuss some properties of q_n^* , which is the threshold value of the belief (when the current stock is n units) at which the retailer will choose to stop selling the current product and move to the next one. The quantity q_n^* then allows the retailer to again partition the products into two categories of high-revenue and low-revenue ones.

PROPOSITION 10. *The sequence $(q_n^*; n \geq 1)$ is increasing in n and converges to $q_\infty^* < 1$ as $n \rightarrow \infty$. The sequence is also bounded by $(\theta_H - 1)/(\theta_H - \theta_L) \equiv q_1^* \leq q_n^* \leq \bar{q}_n < 1$, where for all $n \geq 1$, the upper bound \bar{q}_n is the unique solution to*

$$R + \Phi\left(\frac{rR}{\bar{\theta}(q)}\right) = (q - \eta(q))R + (1 - q + \eta(q))W(n - 1, \theta_H). \tag{25}$$

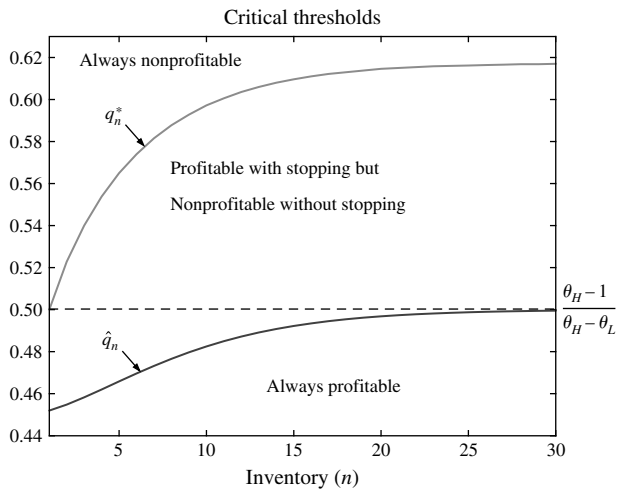
PROOF. See §A.8 in Appendix A. \square

In the setting where stopping is allowed, we have showed that the value function U is always increasing in the current inventory n . Hence, the threshold q_n^* (solution to $U(n, q) = R$) is increasing as well in n . This monotonicity suggests that the seller is willing to take more risks (i.e., measured by larger values of q) for larger initial inventory n . Indeed, higher initial inventory levels offer a greater opportunity for learning, which make them more attractive to the seller. Observe, however, that the upper bound q_∞^* is strictly less than 1, and so the willingness to take risk is limited; if q is greater than q_∞^* , then independently of the order size the seller always rejects such a product.

We recall here that for a particular value of inventory and belief, the value function in the case where stopping is allowed is always larger than the value function when such an option is not available ($V(n, q) \leq U(n, q)$). Therefore, the threshold q_n^* is always larger than \hat{q}_n (solution to $V(n, q) = R$). This inequality implies that the values of the belief for which the product is assumed to be a high-revenue one is larger in the case where stopping is allowed compared to the case where it is not. Put differently, consider n units of a product that the retailer is contemplating selling. If the product's prior q is such that $\hat{q}_n < q < q_n^*$, then the product is considered a nonprofitable one (low-revenue) in the case where stopping is not allowed and a profitable product (high-revenue) in the case where stopping is allowed. Figure 5 depicts the behavior of q_n^* and \hat{q}_n as a function of n . As we can see, the option of stopping has a significant effect on the seller's segmentation of profitable and non-profitable products. Indeed, it is worth noticing that $\hat{q}_\infty = q_1^*$.

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Figure 5. Thresholds q_n^* and \hat{q}_n for the exponential demand model.



Note. Data: $\Lambda = 10$, $\alpha = 1$, $r = 1$, $\theta_H = 1.2$, $\theta_L = 0.8$.

5.1. Bounds and Approximations

We suggested above an algorithm to solve numerically for the value function U ; however, it is impossible in general to obtain a closed-form expression for it. The remainder of this section will be devoted first to obtain a limiting result as n gets large, and second to suggesting approximations to the value function that we later test numerically.

PROPOSITION 11. For all $n \geq 1$, the value function $U(n, \cdot)$ is bounded above and below by piecewise-linear functions such that for all $q \in [0, 1]$,

$$\begin{aligned} & \max\{V_q(n, 0)q + W(n; \theta_H), R\} \\ & \leq U(n, q) \\ & \leq \max\left\{W(n; \theta_H) - (W(n; \theta_H) - R)\frac{q}{\bar{q}_n}, R\right\}. \end{aligned} \quad (26)$$

Furthermore, the sequence of value functions $(U(n, \cdot); n \geq 1)$ converges uniformly on $[0, 1]$ to a continuously differentiable function, $U_\infty(\cdot)$, as $n \rightarrow \infty$, such that for all $q \in [0, 1]$,

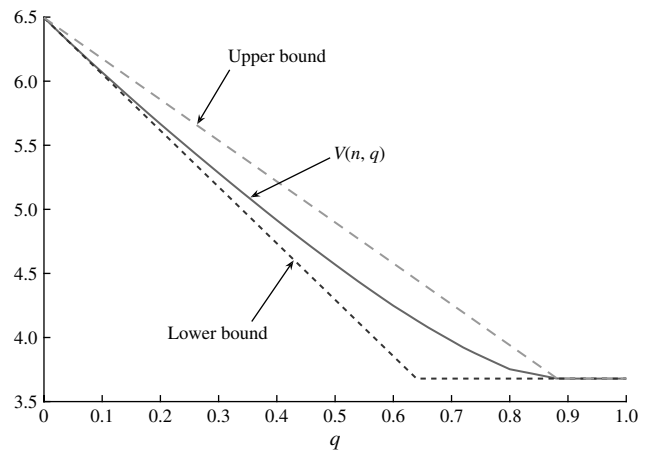
$$R \max\{\bar{\theta}(q), 1\} \leq U_\infty(q) \leq R \max\left\{\theta_H - \frac{\theta_H - 1}{\bar{q}_\infty}q, 1\right\}. \quad (27)$$

PROOF. See §A.9 in Appendix A. \square

The previous proposition shows that the value function converges as the number of the initial inventory becomes large. It also gives a lower and an upper bound, both linear in q and truncated at R .

We denote by $U^l(n, q)$ and $U^u(n, q)$ the lower and upper bounds of $U(n, q)$ appearing in (26) (see Figure 6). Lacking a simple limiting result, we suggest these bounds as possible approximations of the optimal value function.

Figure 6. Value function $V(n, q)$ (for $n = 10$) compared to its linear lower and upper bounds for the exponential demand model $\lambda(p) = \Lambda \exp(-\alpha p)$, with $\Lambda = 10$ and $\alpha = 1$.



Observe that the lower bound is just the tangent at zero until it hits R , whereas the upper bound is a straight line linking the value $U(n, 0)$ at zero, to R at \bar{q}_n (recall that \bar{q}_n is an upper bound of q_n^*). All the parameters of these segments are known without having to solve for U . Indeed, the value function at zero is $W(n, \theta_H)$, the tangent at zero is equal to the known value $V_q(n, 0)$ (we can prove this by induction using similar arguments as those in Lemma B.4 in Appendix B). Finally, \bar{q}_n is a solution to Equation (25), which is independent of U .

To assess the performance of the bound-based approximations, we follow a similar approach to the one used in §4. We first introduce two additional approximations. The first one is an approximation based on the myopic policy discussed in §4,

$$U^M(n, q) = \max\{W(n, \bar{\theta}(q)), R\}.$$

The other one, is the piecewise-linear,

$$U^H(n, q) = \begin{cases} V_q(n, 0)q + W(n, \theta_H) & \text{for } q \leq q_0, \\ (V_q(n, 0)q_0 + W(n, \theta_H) - R) \cdot (q - \bar{q}) / (q_0 - \bar{q}) + R & \text{for } q \in (q_0, \bar{q}), \\ R & \text{for } q \geq \bar{q}, \end{cases} \quad (28)$$

where q_0 is such that each “piece” of U^H covers half of the range of U , i.e., $V_q(n, 0)q_0 + W(n, \theta_H) = (W(n, \theta_H) - R)/2$; see Figure 6. The function U^H is a hybrid function, approximating the value function by the lower bound, $U^l(n, \cdot)$ for $q \leq q_0$ and by a linear function linking $U^l(n, q_0)$ at q_0 to R at \bar{q}_n . We put, $U^H(n, q) = R$ for $q \geq \bar{q}_n$.

Such approximation is meant to take into account the change of slope of the original value function.

To compare these four different approximations, we compute for each of them a corresponding pricing strategy given for $i \in \{U, L, M, H\}$ by the following equation

$$\lambda^i(n, q) = \zeta(U^i(n - 1, q - \eta(q)) - U^i(n, q) + \eta(q)U_q^i(n, q)).$$

We define the following performance measure,

$$\mathcal{E}_U^i(n, q) \triangleq \frac{U(n, q) - U_{\text{Approx}}^i(n, q)}{U(n, q)},$$

where $U_{\text{Approx}}^i(n, q)$, is the seller’s discounted profit under a particular approximating pricing policy.

We cannot expect these approximations to perform as well as the one suggested in §4. Indeed, none of them become asymptotically close to the optimal value function. Figure 6 depicts the gap between the bounds and the optimal value function. This gap will not improve much as n gets large. These approximations, however, have the advantage of being simple (linear or piecewise-linear functions of q), consistent with our previous approximation in §4. The lower bound behaves as well as the myopic policy (this is expected because both coincide when n gets large, $\lim_{n \rightarrow \infty} U^M(n, q) - U^L(n, q) = 0$). The upper bound gives even better results. The numerical analysis is summarized in Table 2. We observe that the relative error defined above ranges from 0 to 3.5% across the different approximations. The worst cases belong to the values of q between 0.4 and 0.6. The error is much smaller for higher values of q . It seems that the upper-bound approximation gives the best results, with a relative error strictly less than 1% for all values of the belief q and inventory n . Finally, as argued above, the value function corresponding to the lower bound behaves numerically similarly to the myopic strategy.

6. Extensions

6.1. Multidimensional Scale Factor θ

The models discussed so far assume that the unknown scale factor θ can take only two values θ_H and θ_L . In many practical situations, the seller may want to enlarge this set of possible values to $\{\theta_1, \dots, \theta_d\}$ ($d \geq 2$) to enrich the modeling of the demand process. Naturally, the choice of d trades off the accuracy of the demand model and the computational effort needed to solve the corresponding multidimensional dynamic program.

From a mathematical standpoint, expanding the support of θ is equivalent to expanding the belief process to a multidimensional vector $q(t) = (q_1(t), \dots, q_d(t))$, where $q_i(t) = \mathbb{P}_q[\theta = \theta_i | \mathcal{F}_t]$. The optimization problem

Table 2. Percentage relative error for the exponential demand model $\lambda(p) = \Lambda \exp(-\alpha p)$, with $\Lambda = 10$ and $\alpha = 1$.

Exponential demand model					
Inventory (n) (%)					
q	1	5	10	25	100
Myopic: $\mathcal{E}_U^M(n, q)$					
0.0	0.00	0.00	0.00	0.00	0.00
0.2	0.22	0.55	1.05	0.85	0.78
0.4	0.48	1.01	1.62	1.46	1.38
0.6	0.26	1.76	2.58	2.69	2.62
0.8	0.00	0.42	0.66	1.37	1.41
1.0	0.00	0.00	0.00	0.00	0.00
Lower bound: $\mathcal{E}_U^L(n, q)$					
0.0	0.00	0.00	0.00	0.00	0.00
0.2	0.15	0.75	1.15	0.84	0.77
0.4	0.32	1.96	1.93	1.44	1.33
0.6	0.05	1.73	3.50	2.70	2.53
0.8	0.00	0.48	0.73	1.34	1.40
1.0	0.00	0.00	0.00	0.00	0.00
Upper bound: $\mathcal{E}_U^U(n, q)$					
0.0	0.00	0.00	0.00	0.00	0.00
0.2	0.19	0.28	0.72	0.58	0.52
0.4	0.42	0.32	0.60	0.60	0.56
0.6	0.24	0.51	0.54	0.79	0.77
0.8	0.00	0.45	0.39	0.85	0.85
1.0	0.00	0.00	0.00	0.00	0.00
Hybrid: $\mathcal{E}_U^H(n, q)$					
0.0	0.00	0.00	0.00	0.00	0.00
0.2	7.52	4.71	1.36	0.95	0.90
0.4	3.02	3.69	1.47	1.21	1.19
0.6	0.28	2.34	1.23	1.15	1.15
0.8	0.00	0.50	0.62	0.97	0.99
1.0	0.00	0.00	0.00	0.00	0.00

in this case becomes (we omit the derivation of the SDE for $q(t)$)

$$V(n, q) = \sup_{\lambda \in \mathcal{A}, \tau \in \mathcal{T}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt) \bar{\theta}(q(t)) c(\lambda_t) dt + \exp(-r\tau) R \right] \quad (29)$$

$$\text{subject to } N_i = n - \int_0^t dD(I_\lambda(s)), \quad (30)$$

$$dq_i(t) = q_i(t-) \left(\frac{\bar{\theta}(q(t-)) - \theta_i}{\bar{\theta}(q(t-))} \right) \cdot (\lambda_t \bar{\theta}(q(t)) dt - dD_t), \quad i = 1, \dots, d, \quad (31)$$

$$\tau \leq \inf\{t \geq 0 : N_t = 0\}, \quad (32)$$

where $\bar{\theta}(q(t)) := \mathbb{E}_q[\theta | \mathcal{F}_t] = \sum_{i=1}^d q_i(t) \theta_i$ is the expected value of θ given the belief $q(t)$.

Based on the results in the previous sections, we know that even for the simplest case $d = 2$, the corresponding HJB optimality condition does not admit a tractable analytical solution. For this reason, we will not analyze this model in full detail but simply present the following extension of the asymptotic approximation in Proposition 5 to this multidimensional case.

PROPOSITION 12. *Consider the seller's optimization problem (29)–(32) with constraint (32) replaced by $\tau = \inf\{t \geq 0: N_t = 0\}$ (i.e., the stopping time option is not available). Then, the corresponding value function $V(n, q)$ is convex in q and converges (uniformly in q) as n goes to infinite*

$$\lim_{n \rightarrow \infty} V(n, q) = \bar{\theta}(q)R.$$

The proof of this result mimics the proof of Proposition 5, and it is omitted. Based on this result, we propose the following approximation for $V(n, q)$ if the stopping option is not available:

$$V(n, q) \approx \sum_{i=1}^d q_i W(n, \theta_i).$$

Each of the $W(n, \theta_i)$ is computed using the recursion in (10).

6.2. Final Reward Function of the Market Proxy θ

The final reward, R , is a critical factor of our model and represents the expected discounted future cashflows of the retailer's operations (otherwise, it could represent the opportunity cost of the space devoted for one product). One contribution of this paper is measuring the effect of this constant on the optimal pricing strategy. This value represents a reference compared to which one continuously tries to guess whether the product is of low revenue or of high revenue. It is likely in certain cases that the future value of a business is affected by the performance of the current product that is being sold. In our case, this translates to R being a function of θ ; the final reward, $R(\theta)$, is henceforth uncertain and revealed only in the long run. We consider here an extension of the model studied in the previous sections by adopting a linear model where $R(\theta) = R_1 + R_2\theta$.

Under such an assumption and in the case where no stopping is allowed before all units are sold, we can write the following:

$$\begin{aligned} V(n, q) &= \sup_{\lambda \in \mathcal{A}, \tau \in \mathcal{T}} \mathbb{E}_q \left\{ \int_0^\tau \exp(-rt) \theta [c(\lambda_t) - rR_2] dt \right. \\ &\quad \left. + \exp(-r\tau)(R_1 + R_2\theta) \right\} \\ &= \sup_{\lambda \in \mathcal{A}, \tau \in \mathcal{T}} \mathbb{E}_q \left\{ \int_0^\tau \exp(-rt) \theta \hat{c}(\lambda_t) dt + \exp(-r\tau) \hat{R} \right\} \\ &\quad + R_2 \bar{\theta}(q), \end{aligned}$$

where $\hat{c}(\lambda) = c(\lambda) - rR_2$ and $\hat{R} = R_1$. The second equality is obtained by writing that $\exp(-r\tau) = 1 - \int_0^\tau r \exp(-rt)$. We let $\hat{V}(n, q) = V(n, q) - R_2 \bar{\theta}(q)$, which brings us back to our original problem of §4. All the results obtained there will hold for \hat{V} when replacing $c \rightsquigarrow \hat{c}$ and $R \rightsquigarrow \hat{R}$. In particular, the approximation suggested for $V(n, q)$ under a terminal reward affine in θ is given by

$$\tilde{V}(n, q) = q(\hat{W}(n, \theta_L) + R_2\theta_L) + (1-q)(\hat{W}(n, \theta_H) + R_2\theta_H),$$

where \hat{W} is the value function corresponding to \hat{V} under perfect information. It is interesting to note that based on (16), the pricing strategy is not affected by this additional linear term in θ . In a similar fashion, we can also generalize the optimal stopping problem of §5 to the case where, again, the terminal reward $R(\theta)$ is an affine function of θ .

6.3. Final Reward with Salvage Value

A potential improvement of our model is to make the opportunity cost R a function of the terminal level of inventory. Obviously, this extension is irrelevant if the option to stop is not available; in this case the final inventory is always zero. However, if the retailer can stop selling the product at any time (as in §5), then we may want to include a salvage value for the unsold units. Specifically, let $R(n)$ be the seller's opportunity cost when there are n units of inventory.

The formulation of the problem remains almost the same than in the case of §5 except for the objective function, which becomes

$$\begin{aligned} U(N_0, q) &= \sup_{\lambda \in \mathcal{A}, \tau \in \mathcal{T}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt) \bar{\theta}(q_t) c(\lambda_t) dt \right. \\ &\quad \left. + \exp(-r\tau) R(N_0 - N_\tau) \right]. \end{aligned} \quad (33)$$

The only difference between this formulation and (22) is the value of the terminal reward $R(n)$. Such modification will result in the following optimality (HJB) equation for each fixed n and $q \in [0, 1]$:

$$\begin{cases} U(n, q) + \Phi\left(\frac{rU(n, q)}{\bar{\theta}(q)}\right) - \eta(q)U_q(n, q) \\ \quad = U(n-1, q - \eta(q)) \quad \text{if } U(n, q) \geq R(n), \\ U(n, q) + \Phi\left(\frac{rU(n, q)}{\bar{\theta}(q)}\right) - \eta(q)U_q(n, q) \\ \quad \leq U(n-1, q - \eta(q)) \quad \text{if } U(n, q) = R(n). \end{cases} \quad (34)$$

Most of the results of §5 extend directly to this case if we impose the following two conditions on $R(n)$.

- (A1) The function $R(n)$ is increasing in n .
- (A2) $R(n) < W(n, \theta_H)$ for all n .

The monotonicity of $R(n)$ is a natural requirement consistent with the notion of salvage value. The second condition guarantees that for every inventory level n the retailer always prefers to sell a high-revenue product ($\theta = \theta_H$) than to liquidate it and collect the terminal reward $R(n)$.

The following proposition summarizes the main properties of $U(n, q)$. The proof is omitted because it follows the same line of arguments as the proofs of Propositions 10 and 11. Before stating the result, we introduce the following quantities. Similar to §5, let q_n^* be the solution to $U(n, q) = R(n)$. Denote by \underline{q}_n and \bar{q}_n , respectively, the solutions to

$$\Phi\left(\frac{rR(n)}{\bar{\theta}(q)}\right) = R(n-1) - R(n)$$

and

$$R(n) + \Phi\left(\frac{rR(n)}{\bar{\theta}(q)}\right) = (q - \eta(q))R(n-1) \\ + (1 - q + \eta(q))W(n-1, \theta_H).$$

PROPOSITION 13. *Suppose that conditions (A1) and (A2) hold. Then, the value function $U(n, q)$ is increasing in n and decreasing and convex in q . Furthermore, for every n , the threshold $q_n^* \in [\underline{q}_n, \bar{q}_n]$, and the value function is bounded by*

$$\max\left\{W(n, \theta_H) - (W(n, \theta_H) - R(n))\frac{q}{\underline{q}_n}, R(n)\right\} \\ \leq U(n, q) \\ \leq \max\left\{W(n, \theta_H) - (W(n, \theta_H) - R(n))\frac{q}{\bar{q}_n}, R(n)\right\}.$$

We note that in this case, the threshold q_n^* is not necessarily increasing in n . Finally, we mention that other results of §5 hold as well; for example, we can similarly use the upper and lower bounds on $U(n, q)$ as approximations of the true value function to estimate the optimal pricing strategy.

7. Concluding Remarks

In this paper, we have studied the problem faced by a retailer that sells nonperishable products to a Poisson arrival stream of price-sensitive customers with unknown demand intensity. The uncertainty in the demand rate is modeled by a single factor θ that is used as a proxy to capture the unknown size of the market. The retailer is initially endowed with a finite inventory of the product and a prior belief about the value of θ . In this setting, the retailer's problem is to maximize the expected discounted cumulative revenue, adjusting dynamically the price of the product and using Bayesian learning to update the distribution of θ . Besides the uncertainty with regard to the demand intensity, the model differs from the traditional revenue management

problem in two important aspects. First, because the product is nonperishable, the selling horizon is not (a priori) bounded. Second, the model explicitly includes an opportunity cost that the retailer incurs when he decides to sell a particular nonperishable product instead of a different assortment.

The analysis of the retailer's problem was divided in three parts. In §3, we considered the case in which θ is known with certainty at time 0. In this perfect information case, the problem admits a tractable dynamic programming formulation that we showed how to solve efficiently. The main insight in this case is that the retailer can partition the set of nonperishable products in two categories depending on the value of θ . If θ is larger than a fixed threshold (that we normalized to 1) the product offers high returns compared to the retailer's average revenue (captured by the opportunity cost R). On the other hand, if θ is low (less than 1 in our normalized system), the product generates lower-than-average revenues. Hence, if the retailer were able to observe in advance the value of θ , he would only engage in selling high-revenue products. An interesting feature of the solution is that even though optimal prices increase with θ , the resulting optimal selling rates also increase with θ . That is, high-revenue products are sold at a higher price and have a higher inventory turnover than low-revenue products.

In §4, we relaxed the perfect information assumption and considered the case in which θ is unknown. We also assumed that the retailer must sell the initial inventory completely before a different assortment can be offered. The analysis of this model is more involved because the state description requires a new state variable to capture the retailer's beliefs about the value of θ . As a result, the resulting dynamic program does not admit a simple analytical solution. Nevertheless, we propose a recursive algorithm to solve the corresponding HJB that requires solving a one-dimensional ODE in each iteration. Because of this lack of tractability, we propose a simple approximation to compute the value function and associated optimal pricing strategy. The proposed policy is based on the fact that as the inventory gets large, the retailer's discounted revenue (as a function of the initial belief) converges uniformly to a straight line that we can characterize in closed form. This asymptotic property is used to develop a simple approximation that showed a good performance when compared numerically to the optimal solution. Our computational experiments, summarized in Table 1, reveal that the asymptotic approximation has on average a relative error that is less than 1%. This is a remarkably good performance if we consider that the myopic policy (which is also asymptotically optimal) has an average relative error close to 30%.

In §5, we considered the case in which the retailer can stop selling the product at any time and move to a different assortment. This stopping decision depends on the inventory level, the retailer's beliefs about the true value

of θ , and the opportunity cost. The HJB optimality condition in this case resembles the one encountered in §4, but includes an extra degree of complexity. The stopping time option creates a free boundary condition that complicates the analysis and solution techniques. In particular, the asymptotic analysis that proved so effective in the model of §4 does not produce a similar result in the case where stopping is allowed. Moreover, we were not able to characterize in simple terms the asymptotic limit of the value function (as inventory gets large), and therefore could not derive an asymptotic approximation as we did in the case without the stopping option. Instead, we derive piecewise-linear upper and lower bounds for the optimal value function that, together with the HJB condition, produce a simple procedure to estimate the optimal pricing strategy. The numerical experiments in Table 2 suggest that the upper-bound approximation performs better than the other approximations with an average relative error of less than 1%.

A distinguishing feature of our model with uncertain demand intensity is that the retailer must consider the trade-offs between exploration and exploitation. That is, by adjusting the price the retailer can influence both the rate at which new information is gathered and the rate at which revenues are collected. Our results in §§4 and 5 suggest that the retailer is willing to take more risk—measured by an increase in the probability that the product is low revenue—for larger orders. Furthermore, when the stopping time option is available the retailer might accept selling a large batch even if his initial belief of θ is strictly less than one. This behavior can be explained by the fact that larger batches offer a larger exploration opportunity. That is, with larger batches the retailer has more time to learn, and hence is willing to take more risk for this option to learn.

There are a number of possible future research directions. First, we can generalize our formulation by considering a nonstationary demand process including, for example, a time component in the unscaled intensity, $\lambda(t, p)$, and in the terminal reward, $R(t)$. This is an important extension in our nonperishable product setting because it captures the evolution of the product life cycle as well as the fact that changing from one assortment to another is an option that is typically not equally available over time. Another extension to this model would be to expand our analysis in §6.2 to consider an arbitrary dependence of R on θ . This will cover situations where learning not only informs the retailer about the current product's demand, but also helps him predict demand in the future (capturing possible correlation among successive products and economic business cycles). A special case of such a setting occurs when the seller's inventory decisions are made contingent upon his knowledge of the market captured by θ .

In revenue management problems in general and in ours in particular, one assumes the cost of initial units to be sunk and no replenishment permitted. However, in some retail businesses, replenishment is certainly an option. An interesting research project would be to generalize our dynamic

pricing with learning to the case when the retailer can choose either to continue with the current product by ordering a new amount or moving to a different product and basically making R . The cost component needs to be introduced in this case. The problem becomes even more complicated, but seems an interesting and a natural continuation of this paper.

Another possible extension would be to consider the operation of a retailer that sells simultaneously a menu of substitute and complementary products. It would be interesting to embed our modeling framework with unknown demand intensity in this case. Some preliminary results in this direction are presented in Caro and Gallien (2007) using a finite-horizon setting with no opportunity costs. Finally, another interesting extension is to consider a different learning approach. Our Bayesian method assumes that the retailer has a prior belief about the value of θ . This prior preconditions the retailer's learning process and, therefore, the resulting pricing strategy. It would be interesting to study a related problem in which the retailer does not have a prior, but instead uses a maximum-likelihood approach. We are currently exploring this variant, which to the best of our knowledge has received little attention in the dynamic pricing literature.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

Endnote

1. A U.S. department store, famous for its basement floor where discounted items are sold.

References

- Aviv, Y., A. Pazgal. 2002. Pricing of short life-cycle products through active learning. Working paper, Olin School of Business, Washington University, St. Louis.
- Aviv, Y., A. Pazgal. 2005. A partially observed Markov decision process for dynamic pricing. *Management Sci.* **51**(9) 1400–1416.
- Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Sci.* **31**(9) 1150–1160.
- Ball, M. O., M. Queyranne. 2009. Toward robust revenue management: Competitive analysis of online booking. *Oper. Res.* **57**(4) 950–963.
- Bertsimas, D., G. Perakis. 2006. Dynamic pricing: A learning approach. D. Hearn, S. Lawphongpanich, eds. *Mathematical and Computational Models for Congestion Charging*. Applied Optimization, Vol. 101. Springer, New York, 45–79.
- Besbes, O., A. Zeevi. 2009. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Oper. Res.*, ePub ahead of print April 24, <http://or.journal.informs.org/cgi/content/abstract/opre.1080.0640v1>.
- Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* **5** 203–229.
- Bitran, G. R., S. V. Mondschein. 1997. Periodic pricing of seasonal products in retailing. *Management Sci.* **43**(1) 64–79.

- Bolton, P., C. Harris. 1999. Strategic experimentation. *Econometrica* **67**(2) 349–374.
- Brémaud, P. 1980. *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag, New York.
- Caro, F., J. Gallien. 2007. Dynamic assortment with demand learning for seasonal consumer goods. *Management Sci.* **53**(2) 276–292.
- Carvalho, A. X., M. L. Puterman. 2004. Dynamic pricing and learning over short time horizons. Working paper, Statistics Department, University of British Columbia, Vancouver, British Columbia, Canada.
- Cheney, W. 2001. *Analysis for Applied Mathematics*. Springer-Verlag, New York.
- Cope, E. 2004. Nonparametric strategies for dynamic pricing in e-commerce. Working paper, The Sauder School of Business, University of British Columbia, Vancouver, British Columbia, Canada.
- Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Sci.* **49**(10) 1287–1309.
- Eppen, G. D., V. Iyer. 1997. Improved fashion buying with Bayesian updates. *Oper. Res.* **45**(6) 805–819.
- Eren, S., C. Maglaras. 2006. Pricing without market information. Working paper, Columbia University, New York.
- Ethier, S. N., T. G. Kurtz. 1986. *Markov Processes: Characterization and Convergence*. Wiley, New York.
- Farias, V. F., B. Van Roy. 2009. Dynamic pricing with a prior on market response. *Oper. Res.* Forthcoming.
- Feng, Y., G. Gallego. 1995. Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Sci.* **41** 1371–1391.
- Fleming, W. H., H. M. Soner. 1993. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York.
- Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* **40** 999–1020.
- Gaur, V., M. L. Fischer, A. Raman. 2005. An econometric analysis of inventory turnover performance in retail services. *Management Sci.* **51** 181–194.
- Keller, G., S. Rady. 1999. Optimal experimentation in a changing environment. *Rev. Econom. Stud.* **66** 475–507.
- Lariviere, M. A. 2006. A note on probability distributions with increasing generalized failure rates. *Oper. Res.* **54**(3) 602–604.
- Lariviere, M. A., E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.* **4**(3) 346–363.
- Lim, A. E. B., J. G. Shanthikumar. 2006. Relative entropy, exponential utility, and robust dynamic pricing. *Oper. Res.* **55**(2) 198–214.
- Lobo, M. S., S. Boyd. 2003. Pricing and learning with uncertain demand. Working paper, Duke University, Durham, NC.
- Lovejoy, W. S. 1990. Myopic policies for some inventory models with uncertain demand distribution. *Management Sci.* **36**(6) 724–738.
- Peskir, G., A. N. Shiryaev. 2000. Sequential testing problems for Poisson processes. *Ann. Statist.* **28** 837–859.
- Petruzzi, N. C., M. Dada. 2002. Dynamic pricing and inventory control with learning. *Naval Res. Logist.* **49** 303–325.
- Rockafellar, R. T. 1997. *Convex Analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ.
- Scarf, H. 1958. Bayes solutions of the statistical inventory problems. *Ann. Math. Statist.* **30** 490–508.
- Shiryaev, A. N. 1978. *Optimal Stopping Rules*. Springer-Verlag, New York.
- Smith, S. A., D. Achabal. 1998. Clearance pricing and inventory policies for retail chains. *Management Sci.* **44** 285–300.
- Talluri, K., G. van Ryzin. 2004. *The Theory and Practice of Revenue Management*. Springer Science, New York.
- Xu, X., W. J. Hopp. 2005. *Dynamic Pricing and Inventory Control: The Value of Demand Learning*. Department of Industrial Engineering, Northwestern University, Evanston, IL.