Dynamic Pricing For Non-Perishable Products With Demand Learning

Victor Araman René Caldentey †

Abstract

A retailer is endowed with a finite inventory of a non-perishable product. Demand for this product is driven by a price-sensitive Poisson process that depends on an unknown parameter which is a proxy for the market size. The retailer has a prior belief on the value of this parameter which he updates as time and available information (prices and sales) evolves. 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 non-perishable product before a different assortment is considered. In the second case, the retailer is able to stop selling the non-perishable 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.

Keywords: Dynamic pricing, Bayesian demand learning, approximations, intensity control, non-homogeneous Poisson process, optimal stopping.

1 Introduction

This paper is concerned with dynamic pricing policies for non-perishable 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 which would cut directly into the profit margins. Second, it highlights the time value of money when assessing the

†Both authors are with the Stern School of Business, New York University, New York, NY 10012, {varaman}{rcaldent}@stern.nyu.edu

business performance. For instance, 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. 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 instance through the sales process, should be constantly performed. Such learning would shed more light on future demand and hint on 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 in order 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 Section 2, a retailer is endowed with a finite stock of a non-perishable product that he sells to a population of price sensitive customers with unknown demand characteristics. The retailer controls dynamically 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 Section 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 Section 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 where 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 Section 6, we present some extensions of the model and concluding remarks are discussed Section 7.

We believe our model contributes to the existing literature in a number of directions. First of all, 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 includes explicitly 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 non-perishable products are not necessarily decreasing functions of the level of inventory (as it 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 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 non-perishable product. Unlike the airline industry where flight departure dates are hard constraints, our modeling suits better 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 instance Shiryayev (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 instance 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 stream of research the only connection between periods occur 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 (2005) 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) study 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 (2007) 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 (2005) 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 square 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 (2005), Besbes and Zeevi (2007) 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 which 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 $\mathbb{P}_\theta$ under which $D(t)$ is Poisson process with rate θ. Note that $\mathbb{P}$ coincides with $\mathbb{P}_1$. We denote by $\mathbb{E}_\theta$ the expectation operator under $\mathbb{P}_\theta$. Also, for every adapted process f_t, non-anticipating with respect to $D(t)$, we define $I_f(t) \triangleq \int_0^t f_s \, \mathrm{d}s$.

In this probabilistic environment, we consider the following stylized retail operations. At time $t = 0$, a retailer owns N_0 identical units of a *non-perishable* 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 $\mathbb{P}_\theta$, this cumulative demand define a non-homogeneous Poisson process with intensity $\theta\,\lambda(p_t)$. The parameter $\theta > 0$ captures the magnitude of the demand intensity while 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 $p(\lambda)$ which 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 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 C1 in Appendix C describes three examples of demand models that satisfy the conditions on 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 non-perishable, 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 part 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 Section 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,...) 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 in order to maximize long-term expected cumulative profits. In particular, we consider the following intensity control problem

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

$$\text{subject to} \qquad N_t = N_0 - \int_0^t \mathrm{d}D\big(I_\lambda(s)\big), \qquad \text{(Inventory Dynamics)} \tag{2}$$

$$\tau \le \inf\{t\ge 0 \;:\; N_t = 0\}. \qquad \text{(Terminal Condition)} \tag{3}$$

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 non-perishability 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 –which 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 as 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 dumping option 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 Section 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 $\mathbb{P}_q = \mathbb{P}_\theta$. 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 inequality sign in (3) will follow directly from this analysis (see the discussion following Proposition 1).

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

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

$$\text{subject to} \qquad N_t = N_0 - \int_0^t \mathrm{d}D\big(I_\lambda(s)\big), \tag{5}$$

$$\tau = \inf\{t \geq 0 \; : \; N_t = 0\}, \tag{6}$$

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 $c^* = r\,R$.

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 in Brémaud 1980), we get the first order optimality condition for this value function in the form of the following Hamilton-Jacobi-Bellman (HJB) equation.

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

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

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

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 references, we also define the function

$$\zeta(z) \triangleq \inf \Big\{ \bar{\lambda} \in [0,\Lambda] \; : \; \bar{\lambda} = \operatorname*{argmax}_{0\leq\lambda\leq\Lambda} \{\lambda\,z + c(\lambda)\} \Big\}. \tag{9}$$

This function $\zeta(z)$ is nondecreasing and satisfies $\zeta(0) = \lambda^* \triangleq \operatorname{argmax}\{c(\lambda) \; : \; \lambda \in [0,\Lambda]\}$. Figure 1 plots ζ and the Fenchel-Legendre transforms Ψ and Φ for the case of an exponential demand rate (for further details see Section C1 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

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

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

$$\lambda^*(n;\theta) = \zeta\big(W(n-1;\theta) - W(n;\theta)\big). \tag{11}$$

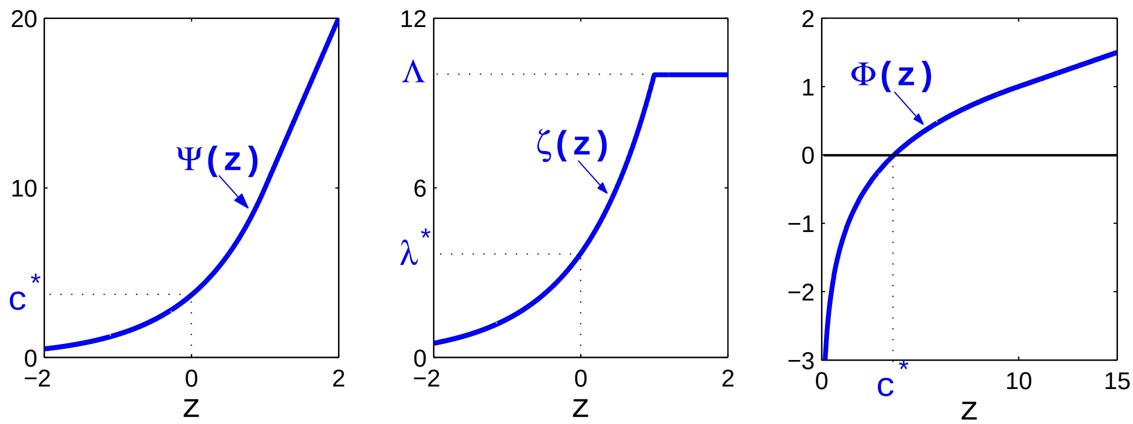

Figure 1: *Fenchel-Legendre transforms for the case of an exponential demand rate* $\lambda(p) = \Lambda\, \exp(-\alpha\, p)$ *with* $\Lambda = 10$ *and* $\alpha = 1$.

Using the monotonicity of $\Phi(z)$, the recursion in (10), and our scaling condition $c^* = r\,R$, 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) which is monotonically increasing in* θ *and satisfies* $\lim_{n\to\infty} W(n;\theta) = \theta R$. *If the scale factor* $\theta \geq 1$ *(resp.* $\theta \leq 1$*) then the value function* $W(n;\theta)$ *is increasing and concave (resp. decreasing and convex) as a function of* n.

Proof: See Section A1 in Appendix A. □

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 $\theta_L \leq 1 \leq \theta_H$.

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 $r\,R$ 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 $\max_\lambda\{\theta\, c(\lambda)\} \geq r\,R$ or equivalently $\theta\, c^* \geq r\,R$. Given the normalization $c^* = r\,R$, 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

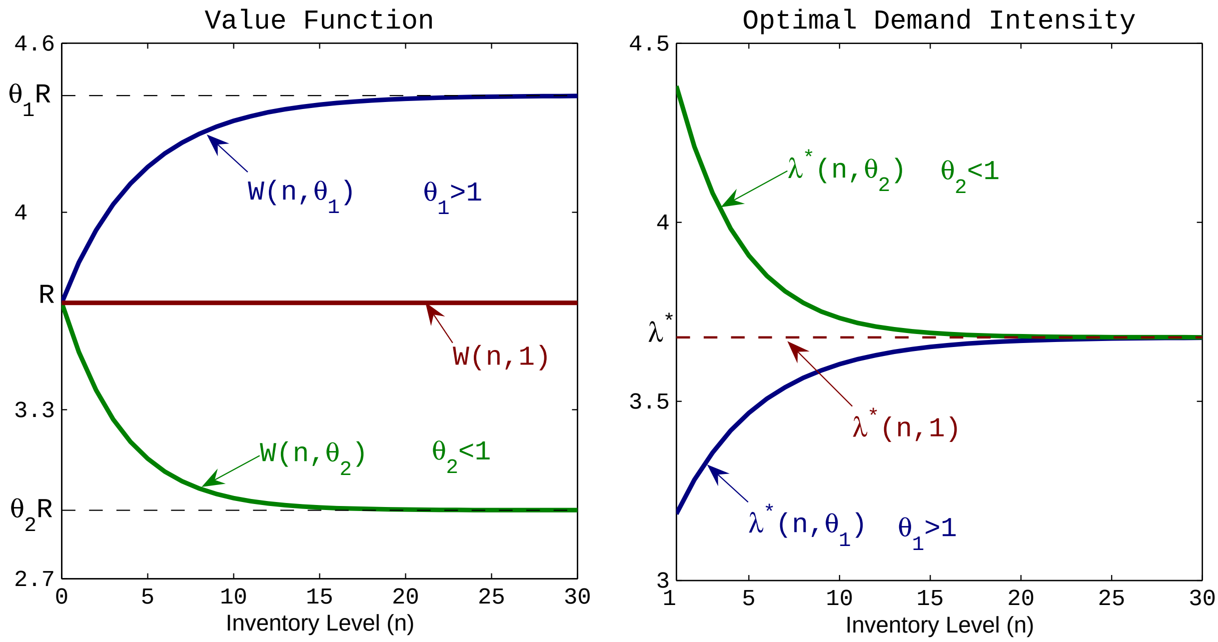

Figure 2: *Value function (left panel) and optimal demand intensity (right panel) for two values of θ under an exponential demand model $\lambda(p) = \Lambda \exp(-\alpha\, p)$. The data used is $\Lambda = 10$, $\alpha = 1$, $r = 1$, $\theta_1 = 1.2$, $\theta_2 = 0.8$, $R = \Lambda \exp(-1)/(\alpha\, r) \approx 3.68$.*

the asymptotic behavior as the inventory grows large. Specifically, we have that $W(n;\theta) \to \theta\, R$ as $n \to \infty$. 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* $\lambda^* = \text{argmax}\{c(\lambda) \,:\, \lambda \in [0, \Lambda]\}$ *be its unique maximizer. The optimal demand intensity* $\lambda^*(n;\theta)$ *is monotonically decreasing in* θ *and satisfies* $\lim_{n\to\infty} \lambda^*(n;\theta) = \lambda^*$*. If* $\theta \geq 1$ *(resp.* $\theta \leq 1$*) then* $\lambda^*(n;\theta)$ *increases (resp. decreases) with the level of inventory* n*.*

Proof: The result follows directly from concavity (resp. convexity) of $W(n;\theta)$ in Proposition 1, equation (11), and the monotonicity of $\zeta(\cdot)$. □

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 go up as well. As a result, the retailer weights less future revenues and maximizes current 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 is larger. In fact, even if low-revenue products have lower prices than high-revenue, 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 Section A2 in Appendix A. □

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 Section C1 in Appendix C. (A simple derivation of this condition translates in 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 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 $p(\lambda) = F^{-1}\left(1 - \frac{\lambda(p)}{\Lambda}\right)$. 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 $[0, \frac{\Lambda}{\alpha}]$ 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 2005).

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 to perfectly anticipate 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, way after procurement decisions are made.

With this problematic 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 consider again 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 Section 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 Section 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 in order 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 $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_q\big)$. The probability measure $\mathbb{P}_q$ satisfies (see Section 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 $\{N(t) = N_0 - \int_0^t \mathrm{d}D_s \; : t \geq 0\}$. 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

$$V(N_0, q) = \sup_{\lambda \in \mathcal{A}} \mathbb{E}_q \left[\int_0^\tau \exp(-rt)\,\theta\, c(\lambda_t) dt + \exp(-r\tau) R \right], \qquad \tau = \inf\{t \geq 0 \, : \, N_t = 0\}. \tag{13}$$

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*

$$dq_t = -\eta(q_{t-}) \Big[dD_t - (\theta_L q_{t-} + \theta_H(1-q_{t-}))\lambda_t dt\Big], \qquad \textit{where } \eta(q_t) \triangleq \frac{q_t(1-q_t)(\theta_H-\theta_L)}{\theta_L q_t + \theta_H(1-q_t)}. \tag{14}$$

Proof: See Section A3 in Appendix A. □

According to (14), the rate at which the seller beliefs change depends on the difference between the observed demand rate, $\mathrm{d}D_t$, and the expected demand rate, $(\theta_L q_{t-} + \theta_H(1-q_{t-}))\lambda_t \mathrm{d}t$, given the available information. Loosely speaking, the martingale nature of q_t follows from (14) by noticing that $\mathbb{E}_q[\mathrm{d}D_t|\mathcal{F}_t] = (\theta_L q_{t-} + \theta_H(1-q_{t-}))\lambda_t \mathrm{d}t$. 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 which 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.

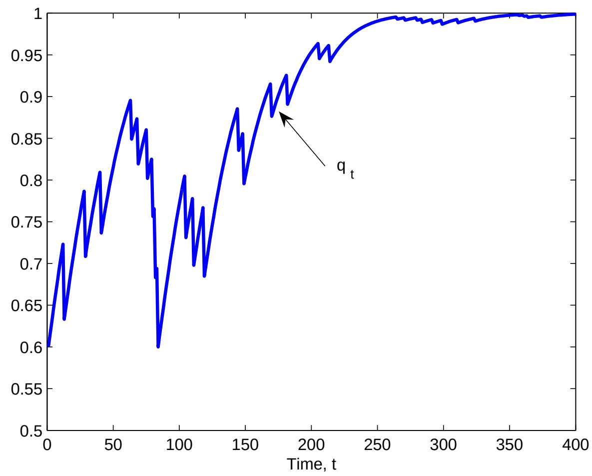

Figure 3: *Values of pathwise q_t under a constant price strategy $\lambda_t = 1$. The data used are starting inventory $n = 35$, $\theta_L = 0.8$, $\theta_H = 1.2$.*

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

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

where $\mathbb{P}_\theta|\mathcal{F}_t$ denotes the restriction of $\mathbb{P}_\theta$ 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 while 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 context, the problem's formulation can be written as follows

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

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 C2 for a derivation)

$$
rV(n, q) = \max_{0 \leq \lambda \leq \Lambda} \Big[\lambda\, \bar{\theta}(q) [V\big(n-1, q - \eta(q)\big) - V(n, q)] + \lambda\, \kappa(q) V_q(n, q) + \bar{\theta}(q)\, c(\lambda) \Big], \tag{16}
$$

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

$$
V(0, q) = R, \qquad V(n, q) + \Phi\left(\frac{r\, V(n, q)}{\bar{\theta}(q)} \right) - \eta(q)\, V_q(n, q) = V\big(n-1, q - \eta(q)\big), \ n = 1, 2, \ldots. \tag{17}
$$

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

$$
\lambda^*_V(n, q) = \zeta \circ \Phi\left(\frac{r\, V(n, q)}{\bar{\theta}(q)} \right), \tag{18}
$$

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 $V(n, q)$.

<u>Algorithm-V:</u>

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

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

$$G(q) + \Phi\left(\frac{r\,G(q)}{\bar{\theta}(q)}\right) - \eta(q)\,G'(q) = F(q), \qquad G(0) = W(n;\theta_H),\ G(1) = W(n;\theta_L). \qquad (19)$$

(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) Goto Step 2.

The main step in this algorithm is to solve the ODE in step 2. This is not a straightforward task as the border conditions at $q = 0$ and $q = 1$ are singular points for the differential equation since $\eta(0) = \eta(1) = 0$. 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. □

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* $n \to \infty$*, and*

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

Proof: See Section A4 in Appendix A. □

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 $s^*(n,q) = \bar{\theta}(q)\,\lambda_V^*(n,q)$ 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 $\bar{\theta}(q)$ even if optimal prices are increasing in $\bar{\theta}(q)$.

Proposition 6 *Suppose 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 Section A5 in Appendix A. □

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 $V(n,q)$

$$\widetilde{V}(n,q) \triangleq \mathbb{E}_q[W(n,\theta)] = q\,W(n,\theta_L) + (1-q)\,W(n,\theta_H). \tag{20}$$

In what follows, we will use the tilde (˜) 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.,*

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

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, $\left|V(n,q)/\widetilde{V}(n,q)\right| \to 1$ uniformly in q as n goes to ∞. Note also, that under perfect information, $V(n,q) = \widetilde{V}(n,q)$ 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

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

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 to use the optimality condition in (18) using $\widetilde{V}(n,q)$ instead of $V(n,q)$. It follows from the linearity of $\widetilde{V}(n,q)$ in q that the proposed approximation for $\lambda^*(n,q)$ is given by

$$\widetilde{\lambda}(n,q) = \zeta\Big((q-\eta(q))\,(\Delta W(n,\theta_H) - \Delta W(n,\theta_L)) - \Delta W(n,\theta_H)\Big), \tag{21}$$

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

Remarks.

1. Since $\zeta(z)$ 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 $\widetilde{\lambda}(n,q)$ is increasing in q.

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

$$\widetilde{\lambda}(n,q) \geq \lambda^* \qquad \text{if and only if} \qquad q-\eta(q) \geq \frac{\Delta W(n,\theta_H)}{\Delta W(n,\theta_H) - \Delta W(n,\theta_L)}.$$

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

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

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

$$\lambda^*_V(1,q) \geq \widetilde{\lambda}(1,q).$$

   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 $\widetilde{\lambda}(n,q)$. 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 to this myopic policy.

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

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 $\{(W(n,\theta_L), W(n,\theta_H) : 0 \le n \le N_0\}$ while the myopic policy is defined by N_0+1 functions $\{W(n,\bar{\theta}(q)) : 0 \le n \le N_0 \text{ and } q \in [\theta_L, \theta_H]\}$.

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

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

The last equality uses the fact that under the probability measure $\mathbb{P}_{\theta_i}$ the selling horizon τ has a Gamma distribution with parameters $(n, \lambda\,\theta_i)$, $i = L, H$. Therefore, $\mathbb{E}_{\theta_i}[e^{-r\,\tau}] = \left(\frac{\lambda\,\theta_i}{r+\lambda\,\theta_i}\right)^n$ for $i = L, H$. The corresponding demand rate associated with this single-price approximation is given by

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

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

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

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 Section 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

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

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

$$\lambda^2(n, q) = \operatorname*{argmax}_{\lambda \in [0,\Lambda]} V^2(n, q; \lambda).$$

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 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 varying uniformly 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 entire range of inventories (n) and beliefs (q) with an average error closed to 0.03%. On the contrary, the Myopic approximation performs quite poorly specially 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 closed 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 specially 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

Asymptotic Approximation

	Inventory (n)				
q	1	5	10	50	100
0.0	0.000	0.000	0.000	0.000	0.000
0.2	0.058	0.005	0.001	0.000	0.000
0.4	0.140	0.013	0.002	0.000	0.000
0.6	0.221	0.021	0.004	0.001	0.001
0.8	0.243	0.029	0.007	0.002	0.003
1.0	0.000	0.000	0.000	0.000	0.000

Myopic Approximation

	Inventory (n)				
q	1	5	10	50	100
0.0	0.000	0.000	0.000	0.000	0.000
0.2	17.533	23.853	24.281	21.777	17.270
0.4	30.897	46.410	48.784	39.314	29.458
0.6	44.113	61.565	59.753	44.493	33.122
0.8	50.472	63.913	61.237	43.911	32.513
1.0	0.000	0.000	0.000	0.000	0.000

Single-Price Policy

	Inventory (n)				
q	1	5	10	50	100
0.0	0.001	0.201	0.313	0.285	0.065
0.2	0.814	0.950	0.823	0.569	0.337
0.4	1.347	1.669	1.252	0.602	0.334
0.6	1.297	2.318	1.707	0.638	0.320
0.8	0.721	2.559	1.994	0.666	0.287
1.0	0.028	1.603	1.031	0.863	0.865

Two-Price Policy

	Inventory (n)				
q	1	5	10	50	100
0.0	0.001	0.017	0.043	0.023	0.107
0.2	0.113	0.219	0.215	0.162	0.082
0.4	0.169	0.373	0.319	0.199	0.112
0.6	0.170	0.540	0.435	0.237	0.135
0.8	0.108	0.712	0.546	0.276	0.146
1.0	0.028	0.876	0.842	0.863	0.865

Table 1: *Relative value function error* $\mathcal{E}_{\mathcal{V}}(n,q)$.

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 of all, 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 Section A6 in Appendix A.□

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 in 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 $\frac{\theta_H - 1}{\theta_H - \theta_L}$, solution itself to $\bar{\theta}(q) = 1$. However, as a result of Proposition 5 –in the limit as the inventory gets large– $\hat{q}(n)$ converges to $\frac{\theta_H - 1}{\theta_H - \theta_L}$.

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 demand rate. In this case, $\hat{q}(n)$ is indeed increasing in n. This monotonicity suggests that the seller is

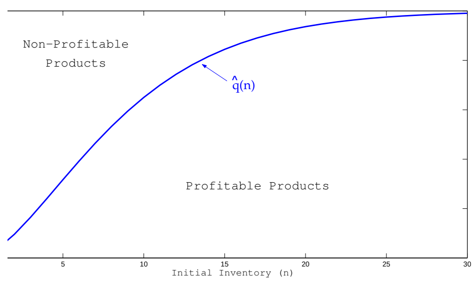

Figure 4: *Value of $\hat{q}(n)$ for an exponential demand rate $\lambda(p) = \Lambda \exp(-\alpha p)$. The data used is $\Lambda = 10$, $\alpha = 1$, $r = 1$, $\theta_H = 1.2$, $\theta_L = 0.8$, $R = c^*/r \approx 3.68$.*

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 while 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 Optimal Stopping Time Rule

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

†A US department store, famous for its basement floor where discounted items are sold.

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 Section 6.

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

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

The optimality (HJB) equation is given by

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

which can be written also as follows

$$\left\{\begin{array}{ll} U(n,q) + \Phi(\frac{r\,U(n,q)}{\bar{\theta}(q)}) - \eta(q)\,U_q(n,q) = U(n-1, q-\eta(q)) & \text{if } U(n,q) \ge R \\ U(n,q) + \Phi(\frac{r\,U(n,q)}{\bar{\theta}(q)}) - \eta(q)\,U_q(n,q) \le U(n-1, q-\eta(q)) & \text{if } U(n,q) = R. \end{array}\right. \qquad (23)$$

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 $q \ge q_n^*$. Hence, an optimal pricing strategy is only defined on the continuation region $q \in [0, q_n^*)$ and satisfies

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

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*

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

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

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

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

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

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

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

$$\lambda^*_V(n,q) \le \lambda^*_U(n,q).$$

Proof: See Section A7 in Appendix A.□

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 good 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., $V(n,q) \le U(n,q)$ for all $n \ge 0$ 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 $F(q) = U\big(n-1, q-\eta(q)\big)$ and

(i) solve for the unique solution of

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

set q^*_n to be this solution

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

$$G(q) + \Phi\left(\frac{r\,G(q)}{\bar{\theta}(q)}\right) - \eta(q)\,G'(q) = F(q), \qquad G(q_0) = R. \tag{24}$$

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

Step 3) Goto Step 2.

Again, the main step in this algorithm is solving the ODE in equation (24). The task here is simpler than in Section 4 as 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^* allows then the retailer to partition again the products in 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 $n \to \infty$. The sequence is also bounded by $\frac{\theta_H - 1}{\theta_H - \theta_L} \equiv q_1^* \leq q_n^* \leq \bar{q}_n < 1$, where for all $n \geq 1$, the upper bound $\bar{q}_n$ is the unique solution to*

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

Proof: See Section A8 in Appendix A. □

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 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 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 non-profitable 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 $\hat{q}_\infty = q_1^*$.

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 remaining of this section will be devoted first to obtain a limiting result as n gets large and second to suggest approximations to the value function that we later test numerically.

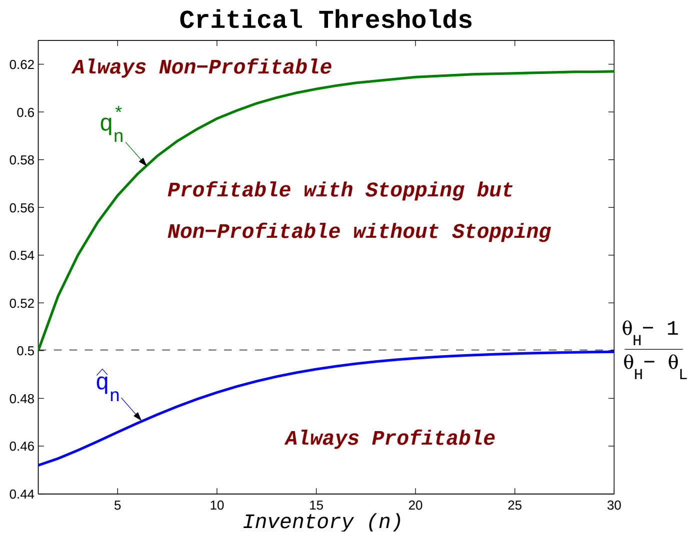

Figure 5: Thresholds q_n^* and $\hat{q}_n$ for the exponential demand model. Data: $\Lambda = 10$, $\alpha = 1$, $r = 1$, $\theta_H = 1.2$, $\theta_L = 0.8$.

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 $q \in [0, 1]$*

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

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

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

Proof: See Section A9 in Appendix A. □

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 $U^{\mathrm{L}}(n,q)$ and $U^{\mathrm{U}}(n,q)$ 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. Observe that the lower bound is just the tangent at zero until it hits R, while 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 B4 in Appendix B). Finally, $\bar{q}_n$ is solution to equation (25) which is independent of U.

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

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

The other one, is the piecewise linear

$$U^{\mathrm{H}}(n,q)=\begin{cases} V_q(n,0)\,q+W(n,\theta_H) \quad ; \quad \text{for } \ q\le q_0 \\ (V_q(n,0)\,q_0+W(n,\theta_H)-R)\,(q-\bar{q})/(q_0-\bar{q})+R \quad ; \quad \text{for } \ q\in(q_0,\bar{q}) \\ R \quad ; \quad \text{for } \ q\ge\bar{q}, \end{cases} \tag{28}$$

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, $U^{\mathrm{L}}(n,\cdot)$ for $q\le q_0$ and by a linear function linking $U^{\mathrm{L}}(n,q_0)$ at q_0 to R at $\bar{q}_n$. We put, $U^{\mathrm{H}}(n,q)=R$ for $q\ge\bar{q}_n$. Such approximation is meant to take into account the change of slope of the original value function.

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

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

We define the following performance measure,

$$\mathcal{E}^i_U(n,q)\triangleq\frac{U(n,q)-U^i_{\mathrm{Approx}}(n,q)}{U(n,q)},$$

where $U^i_{Approx}(n,q)$, 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 Sections 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 Section 4. The lower bound behaves as good as the myopic policy (this is expected as both coincide when n gets large $\lim_{n\to\infty}U^{\mathrm{M}}(n,q)-U^{\mathrm{L}}(n,q)=0$). The upper bound gives even better results. The numerical analysis is summarized in Table 2 below. We observe, that the relative error defined above range

Exponential Demand Model

Myopic: $\mathcal{E}^{\mathrm{M}}_U(n,q)$

	Inventory (n)				
q	1	5	10	25	100
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: $\mathcal{E}^{\mathrm{L}}_U(n,q)$

	Inventory (n)				
q	1	5	10	25	100
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 %

from 0 to 3.5% across the different approximations. The worse 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

Upper Bound: $\mathcal{E}_U^{\mathbf{U}}(n,q)$	Inventory (n)				
q	1	5	10	25	100
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: $\mathcal{E}_U^{\mathbf{H}}(n,q)$	Inventory (n)				
q	1	5	10	25	100
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 %

Table 2: *Relative Value function error for the exponential demand model $\lambda(p) = \Lambda\ \exp(-\alpha\, p)$, with $\Lambda = 10$ and $\alpha = 1$.*

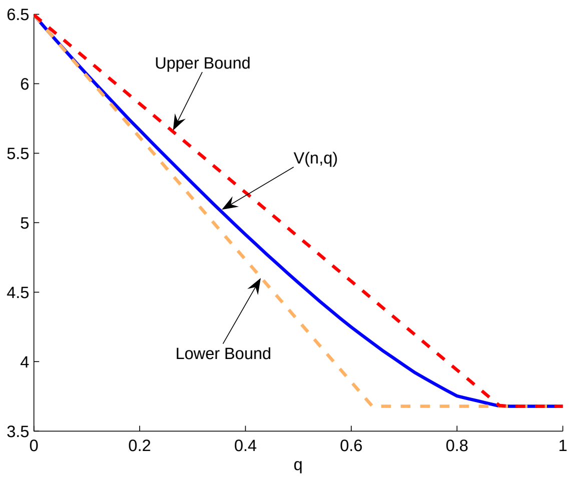

Figure 6: *Value function $V(n,q)$ (for $n = 10$) compared to its linear lower and upper bounds for the exponential demand model $\lambda(p) = \Lambda\ \exp(-\alpha\, p)$, with $\Lambda = 10$ and $\alpha = 1$.*

bound approximation is giving 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 $\{\theta_1, \ldots, \theta_d\}$ ($d \geq 2$) 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 $q(t) = (q_1(t), \ldots, q_d(t))$ where $q_i(t) = \mathbb{P}_q[\theta = \theta_i | \mathcal{F}_t]$. The optimization problem in this case becomes (we omit the derivation of the SDE for $q(t)$)

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

$$\text{subject to} \quad N_t = n - \int_0^t \mathrm{d}D\big(I_\lambda(s)\big), \tag{30}$$

$$\mathrm{d}q_i(t) = q_i(t-)\left(\frac{\bar{\theta}(q(t-)) - \theta_i}{\bar{\theta}(q(t-))}\right)(\lambda_t\,\bar{\theta}(q(t))\,\mathrm{d}t - \mathrm{d}D_t), \qquad i = 1,\ldots,d, \tag{31}$$

$$\tau \le \inf\{t \ge 0 \ :\ N_t = 0\}, \tag{32}$$

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* $\tau = \inf\{t \ge 0 \ :\ N_t = 0\}$ *(*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*

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

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 in 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 assumption and in the case where no stopping is allowed before all units are sold, we

can write the following

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

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 Section 4. All the results obtained there will hold for $\hat{V}$ when replacing $c \curvearrowright \hat{c}$ and $R \curvearrowright \hat{R}$. In particular, the approximation suggested for $V(n,q)$ under a terminal reward affine in θ, is given by

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

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 Section 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 Section 5) then we may want to include a salvage value for the unsold units. Specifically, let $R(n)$ be 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 Section 5 except for the objective function which becomes

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

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 $q \in [0,1]$

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

Most of the results of Section 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 as it follows the same line of arguments than the proofs of Propositions 10 and 11. Before stating the result, we introduce the following quantities. Similar to Section 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

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

and

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

Proposition 13 *Suppose 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*

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

We note that in this case, the threshold q_n^* is not necessarily increasing in n. Finally, we mention that other results of Section 5 hold as well; for instance, 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 non-perishable 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 which 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 includes explicitly 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 Section 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 non-perishable 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 Section 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 as 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 which is less than 1%. This is a remarkable good performance if we consider that the Myopic policy (which is also asymptotically optimal) has an average relative error closed to 30%.

In Section 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 Section 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 which, 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 Sections 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 to sell 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 of all, we can generalize our formulation by considering a non-stationary 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 as 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 Section 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 (2005) 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 that to the best of our knowledge has received little attention in the dynamic pricing literature.

APPENDIX A: Main Proofs

A1. Proof of Proposition 1

- Existence and Uniqueness: The existence and uniqueness of a solution $W(n;\theta)$ follows by noticing that recursion (10) is equivalent to $\mathcal{F}(W(n;\theta)) = W(n-1;\theta)$, where the function

$$\mathcal{F}(z) \triangleq z + \Phi\left(\frac{r\,z}{\theta}\right)$$

is continuous, strictly increasing and ranges from $[\Phi(0), \infty)$. This follows from the fact that $\Phi(z)$ is continuous and nondecreasing. Therefore, $\mathcal{F}$ admits an inverse function $\mathcal{F}^{-1}$ which is continuous strictly increasing and non-negative in the domain $[\Phi(0), \infty)$. Since $\Phi(0) \leq 0 \leq R$, it follows that $W(n;\theta)$ is uniquely determined through the recursion

$$W(0;\theta) = R, \qquad W(n;\theta) = \mathcal{F}^{-1}(W(n-1);\theta),\ n = 1, 2, \ldots.$$

- Monotonicity on θ: To prove the monotonicity of $W(n;\theta)$ on θ we use induction over n. First note that $W(1;\theta)$ solves

$$rW(1;\theta) = \theta\,\Psi(R - W(1;\theta)).$$

Since the function $h_1(z,\theta) = \theta\,\Psi(R-z)$ is increasing in θ (because Ψ is nonnegative) and decreasing in z it follows that $W(1;\theta)$ increases with θ. Let us assume that $W(n-1;\theta)$ is increasing in θ for some n. Now, $W(n;\theta$ solves

$$rW(1;\theta) = \theta\,\Psi(R - W(1;\theta)).$$

Again, the function $h_n(z,\theta) = \theta\,\Psi(R-z)$ is increasing in θ and decreasing in z. We conclude that $W(n;\theta)$ is also increasing with θ.

- Monotonicity and Concavity/Convexity on n: We now prove the monotonicity and concavity of $W(n;\theta)$ for the case $\theta \geq 1$. The proof in the case $\theta \leq 1$ uses the same line of arguments and it is left to the reader.
Suppose that $\theta \geq 1$. We proof the monotonicity of $W(n;\theta)$ in n by induction.

I) First, for $n = 1$ we have that $W(n-1;\theta) = W(0;\theta) = R$. Suppose, by contradiction, that $W(1;\theta) < W(0;\theta)$. Under this hypothesis, condition (10) implies $\Phi\left(\frac{r\,W(1;\theta)}{\theta}\right) > 0$. In addition, by construction $\Phi(z) > 0$ implies $z > c^*$ and so $rW(1;\theta) > \theta\,c^*$. But this last inequality implies that $W(1;\theta) > \theta\,R \geq R = W(0;\theta)$, since $c^* = r\,R$ and $\theta \geq 1$. Therefore, we conclude that $W(1;\theta) \geq W(0;\theta) = R$.

II) Suppose that $W(k;\theta) \geq W(k-1;\theta)$ for all $k = 1, \ldots, n-1$, some $n \geq 1$.

III) Let us prove that $W(n;\theta) \geq W(n-1;\theta)$. Again, by contradiction, let us suppose that $W(n;\theta) < W(n-1;\theta)$. Condition (10) implies $\Phi\left(\frac{r\,W(n;\theta)}{\theta}\right) > 0$ and so we must have $\theta\,R < W(n;\theta) < W(n-1;\theta)$. In addition, by condition (10) we also have that

$$W(n-1) = W(n-2) - \Phi\left(\frac{r\,W(n-1;\theta)}{\theta}\right).$$

Since $\Phi(z)$ is monotonically increasing and $W(n-1;\theta) > \theta\, R$ we conclude

$$W(n-1;\theta) < W(n-2;\theta) - \Phi(r\, R) = W(n-2;\theta),$$

which contradicts the induction step (II). We conclude that $W(n;\theta) \geq W(n-1;\theta)$.

To prove the concavity of $W(n;\theta)$ simply note that condition (10) implies

$$W(n;\theta) - W(n-1;\theta) = -\Phi\left(\frac{r\, W(n;\theta)}{\theta}\right).$$

Since both $\Phi(z)$ and $W(n;\theta)$ are monotonically increasing in their corresponding arguments, we conclude that the right hand side above is monotonically decreasing in n and so $W(n;\theta)$ is concave.

- Limiting Behavior: Finally, to prove the asymptotic behavior of $W(n;\theta)$, we first note that $W(n;\theta)$ is bounded. In fact, for the case $\theta \leq 1$ the boundedness follows since $W(n)$ is decreasing and nonnegative and so $W(n;\theta) \in [0, W(0;\theta)]$. On the other hand, for the case $\theta \geq 1$, $W(n;\theta)$ is increasing in n and so by condition (10) and the monotonicity of $\Phi(z)$ it follows that $rW(n;\theta)/\theta \leq c^*$, or equivalently, $W(n;\theta) \leq \theta\, R$. Given that $W(n;\theta)$ is bounded and monotonic (either increasing if $\theta \geq 1$ or decreasing if $\theta \leq 1$), we have that $\lim_{n\to\infty} W(n;\theta)$ exists. If we denote by $W(\infty;\theta)$ this limit, then letting $n \to \infty$ in condition (10) and using the continuity of $\Phi(z)$, we conclude that $\Phi(\frac{r\, W(\infty;\theta)}{\theta}) = 0$ or $W(\infty;\theta) = \theta\, c^*/r = \theta\, R$. □

A2. Proof of Proposition 2

Combining equations (10) and (11), it follows that

$$s^*(n;\theta) = \theta\, \zeta \circ \Phi\left(\frac{rW(n;\theta)}{\theta}\right),$$

where $\zeta \circ \Phi$ is the composition of Φ and ζ.
Our assumption that $\lambda\, p'(\lambda)$ is decreasing in λ implies that the function $\lambda^2\, p'(\lambda)$ is also decreasing in λ. Because $\lambda \in [0, \Lambda]$, we denote by $\bar{z} \triangleq \Lambda^2\, p'(\Lambda)$ its minimum value. The following lemma will be useful.

Lemma 1 *The function $\zeta \circ \Phi$ satisfies*

$$\zeta \circ \Phi(z) = \begin{cases} \lambda \text{ solution to } \lambda^2\, p'(\lambda) = -z & \text{if } 0 \leq z \leq -\bar{z} \\ \Lambda & \text{otherwise.} \end{cases}$$

It follows from Lemma 1 that if $rW(n;\theta)/\theta \geq -\bar{z}$ then $\zeta \circ \Phi(rW(n;\theta)/\theta) = \Lambda$ in which case $s^*(n\,\theta)$ is trivially (locally) increasing in θ. Let us then assume then that $rW(n;\theta)/\theta < -\bar{z}$. According to Lemma 1, the optimal demand intensity $\lambda^*(n;\theta)$ satisfies

$$(\lambda^*(n;\theta))^2\, p'(\lambda^*(n;\theta)) = -\frac{rW(n;\theta)}{\theta},$$

which implies

$$s^*(n;\theta) = -\frac{rW(n;\theta)}{\lambda^*(n;\theta)\, p'(\lambda^*(n;\theta))}.$$

To complete the proof note that(i) $W(n;\theta)$ increases with θ (by Proposition 1), $\lambda p'(\lambda)$ decreases with λ (by assumption), and (iii) $\lambda^*(n;\theta)$ decreases with θ *(by Corollary 1). □

A3. Proof of Proposition 3

We recall that $D_t = N_0 - N_t$ is the cumulative demand up to time t, which has a Poisson distribution with mean $\theta I_\lambda(t)$. Recall that $I_\lambda(t) = \int_0^t \lambda_s \, \mathrm{d}s$. The function $\lambda_t = \lambda(p_t)$ is the unscaled demand intensity at time t given the pricing policy p_t selected by the seller. Using the Poisson distribution of cumulative demand in $[0, t]$ and Bayes' rule we get that

$$
\begin{aligned}
q_t &= \mathbb{P}_q(\theta = \theta_L | \mathcal{F}_t) \\
&= \frac{q \cdot (\theta_L \, I_\lambda(t)^{D_t} \exp(-\theta_L \, I_\lambda(t))/D_t!}{q \cdot (\theta_L \, I_\lambda(t))^{D_t} \exp(-\theta_L \, I_\lambda(t))/D_t! + (1-q) \cdot (\theta_H \, I_\lambda(t))^{D_t} \exp(-\theta_H \, I_\lambda(t))/D_t!} \\
&= \frac{q}{q + (1-q)(\theta_H/\theta_L)^{D_t} \exp(-(\theta_H - \theta_L) \, I_\lambda(t))}.
\end{aligned} \tag{a1}
$$

The second equality follows from the Markov property of the demand process. We can now obtain the dynamics of the seller's belief process $(q_t : t \geq 0)$. For that, we write $q_t = f(Y_t)$, where $Y_t \triangleq \ln(\theta_H/\theta_L) D_t - (\theta_H - \theta_L) \, I_\lambda(t)$ is an $\mathcal{F}_t$-semimartingale and f is a twice differentiable and bounded function given by $f(n) \triangleq \frac{q_0}{q_0 + (1-q_0)\exp(n)}$. From Itô's lemma (*e.g.*, Ethier and Kurtz (1986)) and the fact that Y_t is a finite variation process (which follows from the fact that $D(t)$ is a pure-jump process and $I_\lambda(t)$ is non-decreasing), we get

$$
\mathrm{d}q_t = f'(Y_{t-}) \, \mathrm{d}Y_t + f(Y_t) - f(Y_{t-}) - f'(Y_{t-}) \, \Delta Y_t.
$$

Taking advantage of the pure-jump nature of D_t and the continuity of $I_\lambda(t)$, we have $\mathrm{d}D_t = \Delta D_t$, $\mathrm{d}Y_t = \Delta Y_t - (\theta_H - \theta_L) \, \mathrm{d}I_\lambda(t)$, and $f(Y_t) - f(Y_{t-}) = [f(Y_{t-} + \ln(\theta_L/\theta_H)) - f(Y_{t-})]\mathrm{d}D_t$, so that

$$
\begin{aligned}
\mathrm{d}q_t &= -f'(Y_{t-})(\theta_H - \theta_L) \, \mathrm{d}I_\lambda(t) + [f(Y_{t-} + \ln(\theta_L/\theta_H)) - f(Y_{t-})] \, \mathrm{d}D_t \\
&= (\theta_H - \theta_L) \frac{q(1-q)\exp(Y_{t-})}{(q + (1-q)\exp(Y_{t-}))^2} \, \mathrm{d}I_\lambda(t) \\
&\quad + \left[\frac{q}{q + (1-q) \, \exp(Y_{t-}) \frac{\theta_H}{\theta_L}} - \frac{q}{q + (1-q) \, \exp(Y_{t-})} \right] \mathrm{d}D_t \\
&= -\eta(q_{t-}) \left[\mathrm{d}D_t - (\theta_L q_{t-} + \theta_H(1-q_{t-}))\mathrm{d}I_\lambda(t) \right], \quad \text{where } \eta(q_t) \triangleq \frac{q_t(1-q_t)(\theta_H - \theta_L)}{\theta_L q_t + \theta_H(1-q_t)}. \; \square
\end{aligned} \tag{a2}
$$

A4. Proof of Proposition 5

The monotonicity and boundedness of $V(n, q)$ are proven in the proof of proposition 4 in Appendix B. To prove the convexity of $V(n, q)$ with respect to q, we define

$$
J_\lambda(n, \theta) \triangleq \int_0^\tau \exp(-rt) \, \theta \, c(\lambda_t) dt + \exp(-r\tau) R, \qquad \tau = \inf\{t \geq 0 \, : \, N_t = 0\},
$$

for an arbitrary policy $\lambda \in \mathcal{A}$. We consider a pair of beliefs $q_1, q_2 \in [0,1]$ and set $q = \alpha\, q_1 + (1-\alpha)\, q_2$ for some $\alpha \in [0,1]$. Then, convexity follows from

$$\begin{aligned} V(n,q) &= \sup_{\lambda\in\mathcal{A}} \big\{\mathbb{E}_q[J_\lambda(n,\theta)]\big\} = \sup_{\lambda\in\mathcal{A}} \big\{q\,\mathbb{E}_{\theta_L}[J_\lambda(n,\theta)] + (1-q)\,\mathbb{E}_{\theta_H}[J_\lambda(n,\theta)]\big\} \\ &= \sup_{\lambda\in\mathcal{A}} \big\{(\alpha\, q_1 + (1-\alpha)\, q_2)\,\mathbb{E}_{\theta_L}[J_\lambda(n,\theta)] + (1-\alpha\, q_1 - (1-\alpha)\, q_2)\,\mathbb{E}_{\theta_H}[J_\lambda(n,\theta)]\big\} \\ &= \sup_{\lambda\in\mathcal{A}} \big\{\alpha\,\mathbb{E}_{q_1}[J_\lambda(n,\theta)] + (1-\alpha)\,\mathbb{E}_{q_2}[J_\lambda(n,\theta)]\big\} \\ &\le \alpha \sup_{\lambda\in\mathcal{A}} \big\{\mathbb{E}_{q_1}[J_\lambda(n,\theta)]\big\} + (1-\alpha) \sup_{\lambda\in\mathcal{A}} \big\{\mathbb{E}_{q_2}[J_\lambda(n,\theta)]\big\} \\ &= \alpha\, V(n,q_1) + (1-\alpha)\, V(n,q_2). \end{aligned}$$

Finally, to prove the uniform convergence of $V(n,q)$, let τ_n be the time it takes to sell n units under an optimal pricing policy. Similarly, let $\tau_n(\lambda)$ be the time to deplete n units while keeping the demand rate constant at λ. Observe that

$$R\,\bar{\theta}(q) = \max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q\left[\int_0^\infty \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt\right].$$

To see this note that the Bounded Convergence Theorem allows an interchange of the expected value and the integral. It is then clear that the LHS is an upper bound of the RHS and is achieved for $\lambda_t \equiv \lambda^*$. Hence,

$$R\bar{\theta}(q) \le \max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q \int_0^{\tau_n} \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt + \max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q \int_{\tau_n}^{\infty} \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt.$$

The second term of the RHS of the previous inequality is bounded by

$$\max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q \int_{\tau_n}^{\infty} \exp(-rt)\theta_H c^* dt \le \frac{1}{r} c^* \theta_H\, \mathbb{E}_q \exp(-r\tau_n(\Lambda)).$$

Using this bound, we write

$$\begin{aligned} &|R\bar{\theta}(q) - V(n,q)| \\ &= |R\bar{\theta}(q) - \max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q\{\int_0^{\tau_n} \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt + R\exp(-r\tau_n)\}| \\ &\le |R\bar{\theta}(q) - \max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q\{\int_0^{\tau_n} \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt| \\ &\qquad + |\max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q\{\int_0^{\tau_n} \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt + R\exp(-r\tau_n)\} - \max_{0\le\lambda_t\le\Lambda} \mathbb{E}_q \int_0^{\tau_n} \exp(-rt)\bar{\theta}(q_t)c(\lambda_t)dt| \\ &\le \frac{1}{r} c^*\theta_H\, \mathbb{E}_q \exp(-r\tau_n(\Lambda)) + \max_{0\le\lambda_t\le\Lambda} R\,\mathbb{E}_q \exp(-r\tau_n) \\ &\le (\frac{1}{r} c^*\theta_H + R)\,\mathbb{E}_q \exp(-r\tau_n(\Lambda)). \end{aligned}$$

For the second inequality we use the fact that both differences on the RHS of the first inequality are non-negative. We also used implicitly that $\tau_n(\lambda)$ stochastically dominates $\tau_n(\Lambda)$ for all $0 \le \lambda \le \Lambda$. We can easily show that $\mathbb{E}_q \exp(-r\tau_n(\Lambda)) = (\Lambda/(r+\Lambda))^n$. The RHS is then independent of q and converges to 0 as $n \to \infty$. We just showed that $(V(n,\cdot) : n \in \mathbb{N})$ converge uniformly to a function

V on [0,1] as $n \to \infty$. This is in agreement with the limiting differential equation obtained from relation (17) by letting n goes to infinity

$$V(q) = V\big(q - \eta(q)\big) + \eta(q)\, V_q(q) - \Phi\left(\frac{r\, V(q)}{\bar{\theta}(q)}\right) \quad \text{with} \quad V(0) = R\theta_H. \tag{a3}$$

The linear function $R\bar{\theta}(q)$ is indeed the unique solution of this ODE. □

A5. Proof of Proposition 6

Suppose that $\lambda_V^*(n,q)$ is locally decreasing in q then it follows trivially that $s^*(n,q)$ is also locally decreasing in q. So, let us assume that $\lambda_V^*(n,q)$ is locally increasing in q. According to equation (18), the selling rate $s^*(n,q)$ satisfies

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

From here, we can use exactly the same steps as in the proof of Proposition 2 replacing $W(n;\theta)$ by $V(n,q)$. □

A6. Proof of Proposition 8

We start by studying the difference $W_c(n,\theta) = W(n,\theta) - R\theta$. We observe based on the recursion (10) and a first order Taylor expansion that

$$\begin{aligned} W_c(n-1;\theta) &= W_c(n;\theta) + \Phi\Big(\frac{rW_c(n;\theta)}{\theta} + c^*\Big) \\ &= W_c(n;\theta)\Big(1 + \frac{r}{\theta}\Phi'(c^*) + o(1)\Big). \end{aligned}$$

It is then easily seen that

$$\frac{W(n;\theta_L) - R\theta_L}{R\theta_H - W(n;\theta_H)} \sim \alpha \cdot c^{-n}, \tag{a4}$$

as $n \to \infty$; where $\alpha > 0$, $c > 1$ and $f(x) \sim g(x)$ as $x \to \infty$ means that $f(x)/g(x) \to 1$ as $x \to \infty$. Now notice that

$$\begin{aligned} \tilde{V}(n,q) - R\bar{\theta}(q) &= qW_c(n;\theta_L) - (1-q)W_c(n;\theta_H) \\ &\sim (q\alpha c^{-n} - (1-q))W_c(n;\theta_H) \end{aligned}$$

which is negative for large n. Finally, considering the linear approximation it is easy to see that $\tilde{V}(n,q) > \tilde{V}(n-1,q)$ if and only if $\Phi(\frac{r\tilde{V}(n,q)}{\bar{\theta}(q)}) < 0$ or equivalently $\tilde{V}(n,q) - R\bar{\theta}(q) < 0$ which completes the proof. □

A7. Proof of Proposition 9

– We start with (ii). The convexity proof follows exactly the same steps as in the case of $V(n,\cdot)$ in Proposition 5. Similarly, the monotonicity in q follows from the same arguments used in the lemma B2 in Appendix B restricted to $(0, q_n^*]$.

– For part (i) Observe that the first equation in (23) with the border condition $U(n,q_n^*) = R$ defines an ODE where a classical Lipschitz continuity argument proves existence and uniqueness of a continuously differentiable solution on $[0,q_n^*]$. Clearly, $U(n,q_n^*) = R$ on $[q_n^*,1]$. $U(n,\cdot)$ is continuous on q_n^* and so it remains to study the continuity of q_n^*. For that let $\epsilon > 0$, define $q_n(\epsilon) = q_n^* + \epsilon$ and $q_n'(\epsilon) = q_n^* - \epsilon$. Note that $U(n, q_n'(\epsilon)) > R$ while $U(n,q_n(\epsilon)) = R$. By taking the difference between the equations in the previous system (23) at the point q_n^*, and letting ϵ goes to zero, we obtain by continuity of the functions $U(n,\cdot)$ and $U(n-1,\cdot)$ that

$$0 \le \eta(q_n^*) U_q(n, q_n^{*-}).$$

The function U_q being non-positive, we conclude that $U_q(n,q_n^{*-}) = U_q(n,q_n^{*+}) = 0$, and $U(n,\cdot)$ is continuously differentiable on $[0,1]$. Putting $U(n,q) = R$ we get that q_n^* is the unique solution of $R + \Phi\left(\frac{r\,R}{\bar\theta(q)}\right) = U(n-1, q-\eta(q))$.

– For part (iii) Fix an initial belief q. With the option of stopping available, a retailer with $n+1$ units can follow the same policy than with n units and so $U(n+1,q) \ge U(n,q)$. As we saw before, this is not necessarily true if the option of stopping is not available. Finally, the bounds are straightforward.

– The proof of (iv) is essentially the same as in Proposition 6.

– To conclude, the prove of part (v) follows from the inequality $V(n,q) \le U(n,q)$, the identities

$$\lambda_V^*(n,q) = \zeta \circ \Phi\left(\frac{r\,V(n,q)}{\bar\theta(q)}\right) \qquad \text{and} \qquad \lambda_U^*(n,q) = \zeta \circ \Phi\left(\frac{r\,U(n,q)}{\bar\theta(q)}\right),$$

and the fact that $\zeta \circ \Phi$ is an increasing function. □

A8. Proof of Proposition 10

The monotonicity of q_n^* is a direct consequence of the monotonicity of $U(n,q)$ in n. To proof that the limiting value $q_\infty^* < 1$ note that $R + \Phi\left(\frac{r\,R}{\bar\theta(q_\infty^*)}\right) = U_\infty(q_\infty^* - \eta(q_\infty^*))$ and $R + \Phi\left(\frac{r\,R}{\bar\theta(1)}\right) > R = U_\infty(1) = U_\infty(1-\eta(1))$. Hence, we must have $q_\infty^* < 1$.

The prove that $\underline{q} := \frac{\theta_H - 1}{\theta_H - \theta_L}$ is a lower bound for q_n^* we note that $R + \Phi\left(\frac{r\,R}{\bar\theta(q)}\right) < R$ for all $q < \underline{q}$. Since by definition q_n^* satisfies $R + \Phi\left(\frac{r\,R}{\bar\theta(q_n^*)}\right) = U_\infty(q_n^* - \eta(q_n^*)) \ge R$, it follows that $q_n^* \ge \underline{q}$.

To derive the upper bound, let us first define the linear function $\mathcal{U}(n,q) := q\,R + (1-q)\,W(n,\theta_H)$. From the convexity of $U(n,q)$ as a function of q and the fact that $U(n,0) = W(n,\theta_H)$ and $U(n,1) = R$ if follows that

$$U(n,q) \le \mathcal{U}(n,q) \qquad \text{for all } q \in [0,1].$$

Since $R + \Phi\left(\frac{r\,R}{\bar\theta(q)}\right)$ is an increasing function of q, it follows that q_n^* solution of $R + \Phi\left(\frac{r\,R}{\bar\theta(q)}\right) = U(n-1, q-\eta(q))$ must be bounded above by $\bar q_n$ solution of $R + \Phi\left(\frac{r\,R}{\bar\theta(q)}\right) = \mathcal{U}(n-1, q-\eta(q))$. □

A9. Proof of Proposition 11

The bounds on $U(n, q)$ follows from its convexity and the definition of $\bar{q}_n$. The uniform convergence is due to Dini's Theorem. (Dini's Theorem states that if a monotone sequence of continuous real-valued functions converge pointwise on a compact set to a continuous function, then the convergence is uniform, see Cheney (2001).) The bounds on $U_\infty(q)$ are again due to the convexity of U_∞ preserved by the uniform convergence. $\square$

APPENDIX B: Proof of Proposition 4

In this appendix we investigate the existence and uniqueness of a solution to the following systems of ODEs.

$$V(0,q) = R, \qquad V(n,q) + \Phi\left(\frac{r\,V(n,q)}{\bar{\theta}(q)}\right) - \eta(q)\,V_q(n,q) = V\big(n-1, q-\eta(q)\big),$$

with border conditions $V(n,0) = W(n;\theta_H)$ and $V(n,1) = W(n;\theta_L)$. We approach this task recursively. That is, we will assume that we have a solution to $V(n-1,q)$ with the desired properties and use it to compute $V(n,q)$.

In what follows, we will drop the dependence on n and use the notation $F(q) = V(n-1, q-\eta(q))$, $G(q) = V(n,q)$, $G_0 = W(n,\theta_H)$ and $G_1 = W(n,\theta_H)$. Also, and due to some mathematical technicalities, we will solve the following weaker version of the problem.

Problem-L: Consider two continously differentiable functions $F(q)$ and $\Phi(q)$. $F(q)$ is decreasing and $\Phi(q)$ is increasing in $q \in (0,1]$. We are interested to find a continuously differentiable function $G(q)$ in $(0,1)$ that solves the ODE

$$G'(q) = \frac{G(q) - F(q) + \Phi\left(\frac{r\,G(q)}{\bar{\theta}(q)}\right)}{\eta(q)}, \qquad q \in (0,1) \tag{b1}$$

with boundary condition

$$\lim_{q\uparrow 1} G(q) = G_1, \qquad \text{where } G_1 \text{ solves} \qquad G_1 - F(1) + \Phi\left(\frac{r\,G_1}{\bar{\theta}(1)}\right) = 0. \tag{b2}$$

We note that we have replaced the original border conditions at $q = 0$ and $q = 1$ by (b2) which is only a limiting condition at $q = 1$. Fortunately, we will show that any solution to this weaker Problem-L satisfies the original border conditions at both $q = 0$ and $q = 1$.

For completeness, we also define G_0 to be the unique root of

$$G_0 - F(0) + \Phi\left(\frac{r\,G_0}{\bar{\theta}(0)}\right) = 0.$$

The monotonicity of the function $h(x) = x + \Phi(\frac{r\,x}{\bar{\theta}})$ guarantees that both G_0 and G_1 are uniquely defined.

To avoid confusion we will use the following terminology. We will say that $G(q)$ is a solution to the ODE if it solves (b1) in $(0,1)$. We say that $G(q)$ is a solution to Problem-L if is a solution to the ODE that satisfies the boundary condition (b2).

In what follows, we will prove that there exists a unique solution to Problem-L. This solution will also satisfy the border condition at $q = 0$. Also, because we are solving the system of ODE recursively, the solution $G(q)$ becomes the function $F(q)$ in the next iteration. Hence, we also need to show that $G(q)$ is continuously differentiable and decreasing in $(0,1]$. The following three sections address these issues of the existence, uniqueness and differentiability and monotonicity of a solution to Problem-L, respectively.

B1 Existence

Before discussing the existence of a solution to problem-L, let us first prove three lemmas.

Lemma B1 *Let $G(q)$ be a solution to the ODE. Let $\bar{q} \in (0,1)$ and $\bar{G} = G(\bar{q})$.*

i) If $\bar{G} < G_1$ then $lim_{q\uparrow 1} G(q) < G_1$.

ii) If $\bar{G} > G_0$ then $lim_{q\uparrow 0} G(q) > G_0$.

Proof: We prove only part (ii). The proof of (ii) uses the same arguments. Suppose $\bar{G} \leq G_1$ then by the definition of G_1 we get

$$\begin{aligned} G'(\bar{q}) &= \frac{\bar{G} - F(\bar{q}) + \Phi\left(\frac{r\,\bar{G}}{\bar{\theta}(\bar{q})}\right)}{\eta(\bar{q})} < \frac{G_1 - F(\bar{q}) + \Phi\left(\frac{r\,G_1}{\bar{\theta}(\bar{q})}\right)}{\eta(\bar{q})} \\ &= \frac{F(1) - F(\bar{q}) + \Phi\left(\frac{r\,G_1}{\bar{\theta}(\bar{q})}\right) - \Phi\left(\frac{r\,G_1}{\bar{\theta}(1)}\right)}{\eta(\bar{q})} \leq 0. \end{aligned}$$

The last inequality follows from the monotonicity of F and Φ and the fact that $\bar{\theta}(\bar{q}) \geq \bar{\theta}(1)$. Then, $G(q)$ is decreasing at $q = \bar{q}$ and so by its continuity we conclude that $lim_{q\uparrow 1} G(q) < G_1$. □

Lemma B2 *Let $G(q)$ be a solution to the ODE. If there is $q_0 \in (0,1)$ such that $G'(q_0) \geq 0$ then $G'(q) \geq 0$ for all $q \geq q_0$.*

Proof: The result follows from noticing that the function

$$h(q,x) := x - F(q) + \Phi\left(\frac{r\,x}{\bar{\theta}(q)}\right)$$

is increasing in both x and q. That is, $h_q(q,x) \geq 0$ and $h_x(q,x) \geq 0$ for all (q,x), where h_q and h_x are the partial derivatives of $h(q,x)$ with respect to the first and second argument respectively. Hence,

$$G'(q) = \frac{1}{\eta(q)}\,h(q,G(q)) = \frac{1}{\eta(q)}\left(\eta(q_0)\,G'(q_0) + \int_{q_0}^{q}\left[h_q(s,G(s)) + h_x(s,G(s))\,G'(s)\right]\mathrm{d}s\right) \geq 0,$$

where the inequality follows from the assumption $G'(q_0) \geq 0$. □

Lemma B3 *Let $G(q)$ be a bounded solution to the ODE. If $\lim_{q\downarrow 0}|G(q)| < \infty$ then $\lim_{q\downarrow 0} G(q) = G_0$. Similarly, if $\lim_{q\uparrow 1}|G(q)| < \infty$ then $\lim_{q\uparrow 1} G(q) = G_1$.*

Proof: We prove only the limit at $q = 0$. The argument in limit at $q = 1$ is similar and it is left to the reader. Because $G(q)$ solves the ODE it follows that

$$\eta(q)\,G'(q) = G(q) - F(q) + \Phi\left(\frac{r\,G(q)}{\bar{\theta}(q)}\right), \quad \text{for all } q \in (0,1). \tag{b3}$$

Suppose that $\lim_{q\downarrow 0} G(q) = \check{G}$ for some real $\check{G}$. We will show, by contradiction, that $\check{G} = G_0$. Let us assume that $\check{G} \neq G_0$. Because of the continuity and boundedness of $G(q)$, $F(q)$ and $\Phi(q)$, it follows from condition (b3) that there is constant $K \neq 0$ such that

$$\lim_{q\downarrow 0} \eta(q)\, G'(q) = \check{G} - F(0) + \Phi\left(\frac{r\,\check{G}}{\bar{\theta}(0)}\right) = K.$$

The fact that $K \neq 0$ follows from the definition (and uniqueness) of G_0 and the assumption $\check{G} \neq G_0$. Suppose $K > 0$ (the case $K < 0$ uses similar arguments). Because $\eta(q) \sim q$ around $q = 0$ and $K \neq 0$, the limit above implies that $G'(q) \sim q^{-1}$ or equivalently $G(q) \sim \ln(q)$ which violates the assumption $\lim_{q\downarrow 0} |G(q)| < \infty$. We conclude that $\check{G} = G_0$. $\square$

We can move to the proof of existence of a solution to Problem-L. For this, we define three families of solutions to the ODE in (b1).

$$\bar{\mathcal{G}} := \left\{ G(q) \quad \text{solution to the ODE in (b1) such that } \lim_{q\uparrow 1} G(q) > G_1 \right\}.$$

$$\mathcal{G} := \left\{ G(q) \quad \text{solution to the ODE in (b1) such that } \lim_{q\uparrow 1} G(q) = G_1 \right\}.$$

$$\underline{\mathcal{G}} := \left\{ G(q) \quad \text{solution to the ODE in (b1) such that } \lim_{q\uparrow 1} G(q) < G_1 \right\}.$$

Proving existence of a solution requires showing $\mathcal{G} \neq \emptyset$. Suppose, by contradiction that $\mathcal{G} = \emptyset$. We now define the following auxiliary functions

$$\bar{G}(q) = \inf_{G \in \bar{\mathcal{G}}} \left\{ G(q) \right\} \quad \text{and} \quad \underline{G}(q) = \sup_{G \in \underline{\mathcal{G}}} \left\{ G(q) \right\}, \quad \text{for all } q \in (0,1),$$

these are the lower and upper envelopes of the set $\bar{\mathcal{G}}$ and $\underline{\mathcal{G}}$, respectively. Note that Lemma B1 guarantees that both $\bar{\mathcal{G}}$ and $\underline{\mathcal{G}}$ are nonempty and so the infimum and supremum are well defined.

Proposition B1 *Suppose $\mathcal{G} = \emptyset$, then for any $q \in (0,1)$*

$$G_1 \leq \underline{G}(q) = \bar{G}(q) \leq G_0.$$

Define, $\tilde{G}(q) = \underline{G}(q)$, then $\tilde{G}(q)$ is a solution to the ODE and satisfies

$$\lim_{q\downarrow 0} \tilde{G}(q) = G_0 \quad \textit{and} \quad \lim_{q\uparrow 1} \tilde{G}(q) = G_1.$$

Proof: The lower bound on $\underline{G}(q)$ and upper bound on $\bar{G}(q)$ follow from Lemma B1. The equality $\underline{G}(q) = \bar{G}(q)$ follows from the assumption $\mathcal{G} = \emptyset$. To prove that $\tilde{G}(q)$ satisfies the ODE, let $q_0 \in (0,1)$ and $\hat{G}(q)$ be the solution of the ODE passing through $(q_0, \tilde{G}(q_0))$. We will show that $\tilde{G}(q) = \hat{G}(q)$ for all $q \in (0,1)$ and so $\tilde{G}(q)$ satisfies the ODE.

Because we are assuming that $\mathcal{G} = \emptyset$, we must have $\hat{G} \in \bar{\mathcal{G}}$ or $\hat{G} \in \underline{\mathcal{G}}$. We consider only the case $\hat{G} \in \bar{\mathcal{G}}$, the proof in the other case follows the same steps. Suppose $\hat{G}(q) \neq \tilde{G}(q)$ then (by the fact that $\tilde{G}(q)$ is the lower envelope of the set $\bar{\mathcal{G}}$) there exists $\check{G}(q) \in \bar{\mathcal{G}}$ such that $\check{G}(q) < \hat{G}(q)$ for all $q \in (0,1)$. But at q_0 the following holds $\tilde{G}(q_0) \leq \check{G}(q_0) < \hat{G}(q_0) = \tilde{G}(q_0)$. This contradiction implies that $\hat{G}(q) = \tilde{G}(q)$ as required.

Next, we need to show that $\lim_{q\downarrow 0} \tilde{G}(q) = G_0$ and $\lim_{q\uparrow 1} \tilde{G}(q) = G_1$. We start showing that these limits exists. For the right limit at $q = 1$, note that $\tilde{G}(q)$ is continuous in $(0,1)$; this follows from the fact that $\tilde{G}(q)$ satisfies the ODE. Now if there is $q_0 \in (0,1)$ such that $\tilde{G}'(q_0) \geq 0$ then by Lemma B2 the function $\tilde{G}(q)$ is increasing in $[q_0, 1)$. Furthermore, by the first part of this proposition, $\tilde{G}(q)$ is also bounded. Hence, the limit (as $q \uparrow 1$) of an increasing and bounded function always exists. On the other hand, if for all $q \in (0,1)$ $\tilde{G}'(q) < 0$ then $\tilde{G}(q)$ is decreasing and bounded in $(0,1)$ and, therefore, it must have a limit as $q \uparrow 1$.
For the left limit at $q = 0$, we use a similar argument. Suppose there exists a $q_0 \in (0,1)$ such that $\tilde{G}'(q_0) < 0$. Then, by Lemma B2 and the fact that $\tilde{G}(q)$ satisfies the ODE, we have that $\tilde{G}'(q) < 0$ for all $q \in (0, q_0]$. Hence by the boundedness of $\tilde{G}(q)$ we conclude that $\lim_{q\downarrow 0} \tilde{G}(q)$ exists. On the other hand, if for all $q \in (0,1)$ $\tilde{G}'(q) \geq 0$ then this monotone condition and the boundedness of $\tilde{G}(q)$ imply again that $\lim_{q\downarrow 0} \tilde{G}(q)$ exists.

Finally, the desired limits $\lim_{q\downarrow 0} \tilde{G}(q) = G_0$ and $\lim_{q\uparrow 1} \tilde{G}(q) = G_1$ follow from (i) the boundedness of $\tilde{G}(q)$, (ii) the fact that $\tilde{G}$ solves the ODE, and (iii) Lemma B3. □

The proposition shows that $\tilde{G}(q)$ is a solution to Problem-L and we must have $\mathcal{G} \neq \emptyset$.

B2 Uniqueness

From the previous section, we already know that there exists a solution $\tilde{G}(q)$ to Problem-L that it is bounded and decreasing in $(0,1)$ and satisfies $\lim_{q\downarrow 0} \tilde{G}(q) = G_0$ and $\lim_{q\uparrow 1} \tilde{G}(q) = G_1$. We can then extend the domain of $\tilde{G}(q)$ to $[0,1]$ by defining $\tilde{G}(0) = G_0$ and $\tilde{G}(1) = G_1$.

In order to prove the uniqueness of this function $\tilde{G}(q)$, we need the following result.

Lemma B4 *Let $\tilde{G}(q)$ be a solution to Problem-L. Then,* $\lim_{q\uparrow 1} \tilde{G}'(q)$ *exists and*

$$\tilde{G}'(1) = \frac{F'(1) + \left(\frac{r\,\bar{\theta}'(1)\,\tilde{G}(1)}{\bar{\theta}^2(1)}\right)\,\Phi'\left(\frac{r\,\tilde{G}(1)}{\bar{\theta}(1)}\right)}{1 + \left(\frac{r}{\bar{\theta}(1)}\right)\,\Phi'\left(\frac{r\,\tilde{G}(1)}{\bar{\theta}(1)}\right) - \eta'(1)}. \tag{b4}$$

Proof: Let us suppose, by contradiction, that $\lim_{q\uparrow 1} \tilde{G}'(q)$ does not exist. Since $\tilde{G}(q)$ is decreasing, this is equivalent to assume that $\lim_{q\uparrow 1} \tilde{G}'(q) = -\infty$.
Let us define the auxiliary function $h(q) := \eta(q)\,\tilde{G}'(q)$ and note that $h(1) = 0$ and

$$h'(q) = \tilde{G}'(q) - F'(q) + \Phi'\left(\frac{r\,\tilde{G}(q)}{\bar{\theta}(q)}\right)\left(\frac{r\,\tilde{G}'(q)\,\bar{\theta}(q) - r\,\tilde{G}(q)\,\bar{\theta}'(q)}{\bar{\theta}^2(q)}\right).$$

By assumption, $F'(q)$ is bounded and $\Phi'(x)$ is nonnegative. Furthermore, by proposition B1 $\tilde{G}(q)$ is also bounded . Hence, the assumption $\lim_{q\uparrow 1} \tilde{G}'(q) = -\infty$ implies that there exists $q_0 \in (0,1)$ such that $h'(q) < 0$ for all $q \geq q_0$.

Take $\epsilon > 0$ such that $q_0 \le 1 - \epsilon$. Then, from a first order Taylor expansion we get

$$h(q_0) = h(1-\epsilon) - \int_{q_0}^{1-\epsilon} h'(q)\,\mathrm{d}q.$$

Since this is true for any $\epsilon > 0$ and $h(1) = 0$ and $h'(q) < 0$, it follows that $h(q_0) > 0$. We conclude that

$$\tilde{G}'(q_0) = \frac{h(q_0)}{\eta(q_0)} > 0.$$

This is not possible because $\tilde{G}(q)$ is decreasing. We conclude that the assumption $\lim_{q\uparrow 1} \tilde{G}'(q) = -\infty$ cannot hold. That is, $\tilde{G}(q)$ admits a left derivative at $q = 1$. We can use L'Hôpital's rule to compute $\tilde{G}'(1)$.

$$\tilde{G}'(1) = \lim_{q\uparrow 1} \frac{h(q)}{\eta(q)} = \lim_{q\uparrow 1} \frac{h'(q)}{\eta'(q)} = \frac{1}{\eta'(1)} \left[\tilde{G}'(1) - F'(1) + \Phi'\left(\frac{r\,\tilde{G}(1)}{\bar{\theta}(1)}\right)\left(\frac{r\,\tilde{G}'(1)\,\bar{\theta}(1) - r\,\tilde{G}(1)\,\bar{\theta}'(1)}{\bar{\theta}^2(1)}\right)\right].$$

Solving for $G'(1)$ we get condition (b4). □

The lemma asserts that any solution $\tilde{G}(q)$ to Problem-L must have bounded derivative in $(0,1]$ where $\tilde{G}'(1)$ is understood to be the left derivative at $q = 1$.

Now, let us suppose that we have two bounded solutions $\tilde{G}(q)$ and $\tilde{g}(q)$ to Problem-L. Without lost of generality let us suppose that $\tilde{g}(q) \le \tilde{G}(q)$ in $(0,1)$. Otherwise, if $\tilde{g}(q_0) = \tilde{G}(q_0)$ for some $q_0 \in (0,1)$ then they must agree in the entire $(0,1)$ as they solve the same ODE in (b1).
Since both $\tilde{G}(p)$ and $\tilde{g}(p)$ satisfy the ODE it follows that for every q

$$-[\tilde{G}(q) - \tilde{g}(q)] = \int_q^1 [\tilde{G}'(x) - \tilde{g}'(x)]\,\mathrm{d}x = \int_q^1 \frac{1}{\eta(x)}\left[\tilde{G}(x) - \tilde{g}(x) + \Phi\left(\frac{r\,\tilde{G}(x)}{\bar{\theta}(x)}\right) - \Phi\left(\frac{r\,g(x)}{\bar{\theta}(x)}\right)\right]\mathrm{d}x. \tag{b5}$$

The monotonicity of $\Phi(x)$ and the boundedness of $\tilde{G}(q)$ and $\tilde{g}(q)$ imply that there exists a bounded and nonnegative function $\xi(q)$ such that

$$\Phi\left(\frac{r\,\tilde{G}(q)}{\bar{\theta}(q)}\right) - \Phi\left(\frac{r\,\tilde{g}(q)}{\bar{\theta}(q)}\right) = \xi(q)\,(\tilde{G}(q) - \tilde{g}(q)).$$

Then,

$$-[\tilde{G}(q) - \tilde{g}(q)] = \int_q^1 (\tilde{G}(x) - \tilde{g}(x))\left[\frac{1+\xi(x)}{\eta(x)}\right]\mathrm{d}x. \tag{b6}$$

By assumption, the left-hand side is nonpositive and the right-hand side is nonnegative. Hence, they must be equal to zero. We conclude then that $\tilde{G}(q) = \tilde{g}(q)$ which shows uniqueness. □

B3 Differentiability and Monotonicity

In order to solve recursively the control problem for $V(n,q)$ using Problem-L, we need to show that any solution $\tilde{G}(q)$ to this problem is continuously differentiable and decreasing in $(0,1]$.

The fact that $\tilde{G}(q)$ is continuously differentiable in $(0, 1]$ follows from proposition B1 and lemma B4. The monotonicity of $\tilde{G}(q)$ follows directly from proposition B1 and lemma B2. In fact, from proposition B1 we know that any solution $\tilde{G}(q)$ to Problem-L is a bounded solution to the ODE and satisfies $G_1 \leq G(q) \leq G_0$ and $\lim_{q\downarrow 0} G(q) = G_0$ and $\lim_{q\uparrow 1} G(q) = G_1$. By lemma B2 if $\tilde{G}(q)$ is non-decreasing at any $q_0 \in (0, 1)$ then it is non-decreasing at any $q \geq q_0$. Combining these properties of $\tilde{G}(q)$ it follows that it must be decreasing in $(0, 1]$.

APPENDIX C: Supplements

C1. Three Examples of Demand Functions

– Exponential Demand Model: Consider the case in which the demand intensity is well approximated by the following exponential demand model

$$\lambda(p) = \Lambda \exp(-\alpha\, p), \qquad p \geq 0.$$

For this demand model, the corresponding inverse demand function and corresponding revenue rates are given by

$$p(\lambda) \triangleq \frac{1}{\alpha} \ln\left(\frac{\Lambda}{\lambda}\right) \quad \text{and} \quad c(\lambda) \triangleq \lambda\, p(\lambda) = \frac{\lambda}{\alpha} \ln\left(\frac{\Lambda}{\lambda}\right), \qquad \lambda \in [0, \Lambda]$$

respectively. The demand rate that maximizes $c(\lambda)$ is $\lambda^* = \Lambda \exp(-1)$ and $c^* = \lambda^*/\alpha$. In addition, the Fenchel-Legendre transform Ψ of $c(\lambda)$ satisfies

$$\Psi(z) \triangleq \max_{0 \leq \lambda \leq \Lambda} \{\lambda\, z + c(z)\} = \begin{cases} \frac{\Lambda}{\alpha} \exp(\alpha z - 1) & \text{if } z \leq \frac{1}{\alpha} \\ \Lambda\, z & \text{if } z \geq \frac{1}{\alpha}. \end{cases}$$

The corresponding maximizer is

$$\zeta(z) \triangleq \operatorname*{argmax}_{0 \leq \lambda \leq \Lambda} \{\lambda\, z + c(z)\} = \begin{cases} \Lambda \exp(\alpha z - 1) & \text{if } z \leq \frac{1}{\alpha} \\ \Lambda & \text{if } z \geq \frac{1}{\alpha}. \end{cases}$$

The function $\Psi(z)$ is continuously differentiable, increasing and convex. The associated inverse function satisfies

$$\Phi(z) \triangleq \begin{cases} \frac{1}{\alpha}\left[1 + \ln\left(\frac{\alpha z}{\Lambda}\right)\right] & \text{if } 0 < z \leq \frac{\Lambda}{\alpha} \\ \frac{z}{\Lambda} & \text{if } z \geq \frac{\Lambda}{\alpha}. \end{cases}$$

Similarly, this function is continuously differentiable, increasing and concave. Note also that $\Phi(c^*) = 0$.

– Linear Demand Model: Consider the case in which the demand intensity is given by the following linear demand model

$$\lambda(p) = \Lambda - \alpha\, p, \qquad 0 \leq p \leq \frac{\Lambda}{\alpha}.$$

For this demand model, the corresponding inverse demand function and corresponding revenue rates are given by

$$p(\lambda) \triangleq \frac{\Lambda - \lambda}{\alpha} \quad \text{and} \quad c(\lambda) \triangleq \lambda\, p(\lambda) = \frac{\lambda(\Lambda - \lambda)}{\alpha}, \qquad \lambda \in [0, \Lambda]$$

respectively. The demand rate that maximizes $c(\lambda)$ is $\lambda^* \triangleq \frac{\Lambda}{2}$ and $c^* \triangleq (\lambda^*)^2/\alpha$. In addition, the Fenchel-Legendre transform Ψ of $c(\lambda)$ satisfies

$$\Psi(z) \triangleq \max_{0 \leq \lambda \leq \Lambda} \{\lambda\, z + c(z)\} = \begin{cases} 0 & \text{if } z \leq -\frac{\Lambda}{\alpha} \\ \frac{(\Lambda + \alpha\, z)^2}{4\,\alpha} & \text{if } -\frac{\Lambda}{\alpha} \leq z \leq \frac{\Lambda}{\alpha} \\ \Lambda\, z & \text{if } z \geq \frac{\Lambda}{\alpha}. \end{cases}$$

The corresponding maximizer is

$$\zeta(z) \triangleq \underset{0\le\lambda\le\Lambda}{\operatorname{argmax}} \{\lambda\, z + c(z)\} = \begin{cases} 0 & \text{if } z \le -\frac{\Lambda}{\alpha} \\ \frac{\alpha z+\Lambda}{2} & \text{if } -\frac{\Lambda}{\alpha} \le z \le \frac{\Lambda}{\alpha} \\ \Lambda & \text{if } z \ge \frac{\Lambda}{\alpha}. \end{cases}$$

The function $\Psi(z)$ is continuously differentiable, nondecreasing and convex. In the domain $z \ge -\frac{\Lambda}{\alpha}$, $\Psi(z)$ admits an inverse function

$$\Phi(z) \triangleq \begin{cases} \frac{\sqrt{4\alpha z}-\Lambda}{\alpha} & \text{if } 0 \le z \le \frac{\Lambda^2}{\alpha} \\ \frac{z}{\Lambda} & \text{if } z \ge \frac{\Lambda^2}{\alpha}. \end{cases}$$

Similarly, this function is continuously differentiable, increasing and concave. Note also that $\Phi(c^*) = 0$.

– Quadratic Demand Model: Consider the case in which the demand intensity is given by the following quadratic demand model

$$\lambda(p) = \sqrt{\Lambda^2 - \alpha\, p}, \qquad 0 \le p \le \frac{\Lambda^2}{\alpha}.$$

For this demand model, the corresponding inverse demand function and corresponding revenue rates are given by

$$p(\lambda) \triangleq \frac{\Lambda^2 - \lambda^2}{\alpha} \quad \text{and} \quad c(\lambda) \triangleq \lambda\, p(\lambda) = \frac{\lambda(\Lambda^2 - \lambda^2)}{\alpha}, \quad \lambda \in [0, \Lambda]$$

respectively. The demand rate that maximizes $c(\lambda)$ is $\lambda^* \triangleq \frac{\Lambda}{\sqrt{3}}$ and $c^* \triangleq 2\,(\lambda^*)^3/\alpha$. In addition, the Fenchel-Legendre transform Ψ of $c(\lambda)$ satisfies

$$\Psi(z) \triangleq \max_{0\le\lambda\le\Lambda} \{\lambda\, z + c(z)\} = \begin{cases} 0 & \text{if } z \le -\frac{\Lambda^2}{\alpha} \\ 2\,\frac{(\Lambda^2+\alpha\, z)^{\frac{3}{2}}}{3\sqrt{3}\,\alpha} & \text{if } -\frac{\Lambda^2}{\alpha} \le z \le 2\,\frac{\Lambda^2}{\alpha} \\ \Lambda\, z & \text{if } z \ge 2\,\frac{\Lambda^2}{\alpha}. \end{cases}$$

The corresponding maximizer is

$$\zeta(z) \triangleq \underset{0\le\lambda\le\Lambda}{\operatorname{argmax}} \{\lambda\, z + c(z)\} = \begin{cases} 0 & \text{if } z \le -\frac{\Lambda^2}{\alpha} \\ \sqrt{\frac{\Lambda^2+\alpha z}{3}} & \text{if } -\frac{\Lambda^2}{\alpha} \le z \le 2\,\frac{\Lambda^2}{\alpha} \\ \Lambda & \text{if } z \ge 2\,\frac{\Lambda^2}{\alpha}. \end{cases}$$

The function $\Psi(z)$ is continuously differentiable, nondecreasing and convex. In the domain $z \ge -\frac{\Lambda^2}{\alpha}$, $\Psi(z)$ admits an inverse function

$$\Phi(z) \triangleq \begin{cases} \frac{3\,(\alpha\, z/2)^{\frac{2}{3}}-\Lambda^2}{\alpha} & \text{if } 0 \le z \le 2\,\frac{\Lambda^3}{\alpha} \\ \frac{z}{\Lambda} & \text{if } z \ge 2\,\frac{\Lambda^3}{\alpha}. \end{cases}$$

Similarly, this function is continuously differentiable, increasing and concave. Note also that $\Phi(c^*) = 0$.

C2. Derivation of the HJB optimality condition

We consider the stochastic control problem

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

The dynamic programming equation for this infinite horizon is

$$
rV(n, q) = \max_{0 \leq \lambda \leq \Lambda} [-\mathcal{G}^\lambda V(n, q) + \bar{\theta}(q) c(\lambda)],
$$

where $\mathcal{G}^\lambda$ is the infinitesimal generator of (N_t, q_t) given the control λ, which following the notations of Fleming and Soner (1993) is defined by

$$
\mathcal{G}^\lambda V(n, q) \triangleq -\lim_{h \to 0} h^{-1} [\mathbb{E}_q V(n(h), q(h)) - V(n, q)].
$$

To compute this last term we apply Itô's lemma to the function V, while noticing that both processes N_t and q_t have finite variation. Using the fact that N_t is a pure-jump process and so $\mathrm{d}N_t = \Delta N_t$, we obtain

$$
\mathrm{d}V(N_t, q_t) = V_q(N_{t-}, q_{t-}) \mathrm{d}q_t + V(N_t, q_t) - V(N_{t-}, q_{t-}) - V_q(N_{t-}, q_{t-}) \Delta q_t,
$$

where the notation V_q stands for $\frac{\partial V}{\partial q}$. From the dynamics of q_t in (14), it follows that $N_t = N_0 - D_t$ and q_t have common jumps of size -1 and $-\eta(q_t)$, respectively. From this observation, it follows that

$$
V(N_t, q_t) - V(N_{t-}, q_{t-}) = -\left[V\big(N_{t-} - 1, q_{t-} - \eta(q_{t-})\big) - V(N_{t-}, q_{t-}) \right] \mathrm{d}N_t
$$

and so for a fixed control λ

$$
\mathrm{d}V(N_t, q_t) = q_t(1 - q_t)(\theta_H - \theta_L) V_q(N_t, q_t) \lambda \mathrm{d}t - \left[V\big(N_{t-} - 1, q_{t-} - \eta(q_{t-})\big) - V(N_{t-}, q_{t-}) \right] \mathrm{d}N_t.
$$

We define $\kappa(q) \triangleq q(1 - q)(\theta_H - \theta_L)$ and write

$$
\begin{aligned}
& h^{-1}\, \mathbb{E}_q [V(n(h), q(h)) - V(n, q)] \\
& \quad = h^{-1}\, \mathbb{E}_q \left[\int_0^h \kappa(q_t) V_q(N_t, q_t)\, \lambda\, \mathrm{d}t - \int_0^h \left[V\big(N_{t-} - 1, q_{t-} - \eta(q_{t-})\big) - V(N_{t-}, q_{t-}) \right] \mathrm{d}N_t \right] \\
& \quad = h^{-1}\, \mathbb{E}_q \left[\int_0^h \kappa(q_t) V_q(N_t, q_t)\, \lambda\, \mathrm{d}t + \int_0^h \bar{\theta}(q_t) \left[V\big(N_t - 1, q_t - \eta(q_t)\big) - V(N_t, q_t) \right] \lambda dt \right],
\end{aligned}
$$

where the second equality follows from the fact that for a given control λ the process $N_t + \bar{\theta}(q_t)\lambda t - N_0$ is an $\mathcal{F}_t$-martingale. Finally, letting $h \downarrow 0$ we conclude that

$$
\mathcal{G}^\lambda V(n, q) = -\lambda \Big(\kappa(q)\, V_q(n, q) + \bar{\theta}(q) [V\big(n - 1, q - \eta(q)\big) - V(n, q)] \Big).
$$

APPENDIX D: HJB Numerical Solution

In this appendix we describe the algorithm that we use in our numerical computations. The method used to solve the HJB optimality conditions is essentially the same whether the optimal stopping option is available or not. The algorithm is based on a finite-difference scheme[‡]in which the belief space $\{q \in [0,1]\}$ is partitioned using a mesh $\mathcal{M}(\Delta q) := \{q_0, \ldots, q_M\}$ such that $q_0 = 0$, $q_M = 1$ and $q_j - q_{j-1} = \Delta q$ for all $j = 1, \ldots, M$. In our computations, the size of the mesh was chosen equal to $\Delta q = 10^{-3}$.

For simplicity, in what follows we will use the index i to refer to the belief q_i and the index n to refer to the level of inventory. In addition, we will use the same notation U and V for the (numerically computed) value function for the case with and without the stopping option, respectively. For example $V(n,i)$ is the (numerically computed) value function when the inventory is n, the belief is q_i and the option to stop is not available.

D1. Algorithm Without the Optimal Stopping Option

Suppose the system is in state (n,i) and we choose a demand intensity equal to λ. Given our discrete mesh $\mathcal{M}$, we will keep this value of λ until (i) a sale occurs and the state jumps to $(n-1, i-\eta(i))$ or (ii) the belief process moves up to q_{i+1} and the state becomes $(n, i+1)$.

Let us denote by Δt_i^λ the length of the time interval during which the control λ is kept constant if there is no sale. We need to write this time as a function of the initial belief q_i and the control λ in order to ensure *local consistency* between this discrete approximation and the actual continuous time evolution of q_t. In fact, suppose that $q_t = q_i$ and the control is fixed at λ then by definition Δt_i^λ should satisfy

$$q_{t+\Delta t_i(\lambda)} = q_{i+1} \quad \text{conditional on the fact that there is no sale in the time interval } [t, t+\Delta t_i(\lambda)).$$

From equation (14) it follows that in the absence of sales and for a fixed λ, q_t evolves according to the deterministic ODE

$$\mathrm{d}q_t = \lambda\, q_t\,(1-q_t)\,(\theta_H - \theta_L)\,\mathrm{d}t.$$

After integration we can show that

$$\Delta t_i^\lambda = \frac{1}{\lambda\,(\theta_H - \theta_L)} \ln\left(\frac{q_{i+1}\,(1-q_i)}{q_i\,(1-q_{i+1})}\right).$$

Let us denote by τ_i^λ the random time at which a sale occurs if we kept the demand intensity λ fixed when the initial belief is q_i. Then, the discrete version of the HJB optimality condition for $V(n,q)$ in equation (16) takes the form

$$V(n,i) = \max_{0\le\lambda\le\Lambda} \mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\, \left(p(\lambda) + V(n-1, i-\eta(i))\right) + \mathbb{1}_{\{\tau_i^\lambda > \Delta t_i^\lambda\}}\, e^{-r\,\Delta t_i^\lambda}\, V(n, i+1)\right], \tag{d1}$$

[‡]See *Numerical Methods for Stochastic Control Problems in Continuous Time* by H. Kushner and P. Dupuis. Springer-Verlag, New York, 2001.

with border conditions $V(0,i) = R$ and $V(n,M) = W(n,\theta_L)$, for all $i = 1,\dots,m$ and all n.
It follows from the Markovian dynamics of q_t that

$$\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda > \Delta t_i^\lambda\}}\right] \equiv P(\tau_i^\lambda > \Delta t_i^\lambda) = \exp\left(-\lambda \int_0^{\Delta t_i^\lambda} \bar{\theta}(q_t)\,\mathrm{d}t\right),$$

where q_t satisfies

$$\mathrm{d}q_t = q_t\,(1-q_t)\,(\theta_H - \theta_L)\,\mathrm{d}t, \quad q_0 = q_i, \quad \text{for all } t \in [0, \Delta t_i^\lambda).$$

The following result follows after integration and we omit the details.

Lemma B5 *For all $i = 0,\dots,M-1$ and $\lambda \in [0,\Lambda]$,*

$$\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda > \Delta t_i^\lambda\}}\right] = \left(\frac{q_i}{q_{i+1}}\right)^{\frac{\theta_H}{\theta_H - \theta_L}} \left(\frac{1-q_{i+1}}{1-q_i}\right)^{\frac{\theta_L}{\theta_H-\theta_L}}. \tag{d2}$$

Computing the value of $\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\right]$ is more involved and no closed-form solution is available. However, we can rewrite this expectation as follows

$$\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\right] = \int_0^{\Delta t_i^\lambda} e^{-r\,t}\,\mathrm{d}F_i^\lambda(t) = e^{-r\,\Delta t_i^\lambda}\, F_i^\lambda(\Delta t_i^\lambda) + \int_0^{\Delta t_i^\lambda} r\, e^{-r\,t}\, F_i^\lambda(t)\,\mathrm{d}t,$$

where $F_i^\lambda(t)$ is the cumulative distribution function of τ_i^λ. For all $t \in [0, \Delta t_i^\lambda)$ we have that $\bar{\theta}(i+1) \le \bar{\theta}(q_t) \le \bar{\theta}_i$ and it follows that $1 - \exp(-\lambda\,\bar{\theta}_{i+1}\,t) \le F_i^\lambda(t) \le 1 - \exp(-\lambda\,\bar{\theta}_i\,t)$. As result, we get that the inequalities

$$\begin{aligned}
\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\right] &\le e^{-r\,\Delta t_i^\lambda}\, F_i^\lambda(\Delta t_i^\lambda) + \frac{\lambda\,\bar{\theta}_i}{r+\lambda\,\bar{\theta}_i} - \left(1 - \frac{r}{r+\lambda\,\bar{\theta}_i}\, e^{-\lambda\,\bar{\theta}_i\,\Delta t_i^\lambda}\right) e^{-r\,\Delta t_i^\lambda} \\
\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\right] &\ge e^{-r\,\Delta t_i^\lambda}\, F_i^\lambda(\Delta t_i^\lambda) + \frac{\lambda\,\bar{\theta}_{i+1}}{r+\lambda\,\bar{\theta}_{i+1}} - \left(1 - \frac{r}{r+\lambda\,\bar{\theta}_{i+1}}\, e^{-\lambda\,\bar{\theta}_{i+1}\,\Delta t_i^\lambda}\right) e^{-r\,\Delta t_i^\lambda}.
\end{aligned}$$

Note that we can write $F_i^\lambda(\Delta t_i^\lambda) = 1 - \mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda > \Delta t_i^\lambda\}}\right]$ and use Lemma B5 to get its value.
If the mesh size Δq is small then the upper and lower bound above are closed to each other. Hence, we can get an asymptotically optimal approximation of $\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\right]$ using one of the two bounds. In our computations, we use

$$\mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda \le \Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\right] \approx \frac{\lambda\,\bar{\theta}_i}{r+\lambda\,\bar{\theta}_i} - \left[\left(\frac{q_i}{q_{i+1}}\right)^{\frac{\theta_H}{\theta_H-\theta_L}} \left(\frac{1-q_{i+1}}{1-q_i}\right)^{\frac{\theta_L}{\theta_H-\theta_L}} - \frac{r}{r+\lambda\,\bar{\theta}_i}\, e^{-\lambda\,\bar{\theta}_i\,\Delta t_i^\lambda}\right] e^{-r\,\Delta t_i^\lambda}. \tag{d3}$$

After plugging the expressions in equations (d2) and (d3) back into the DP (d1) we get the recursion that we use to solve numerically the value function $V(n,i)$. A few technical remarks about this DP recursion are now in order.

- First, note that if a sale occur within the time interval Δt_i^λ then the belief process will jump backward from q_i to $q_i - \eta(q_i)$. Because of the discreteness of our mesh this value $q_i - \eta(q_i)$ might not be a member of $\mathcal{M}$. In this case, we replace $q_i - \eta(q_i)$ by the closest value $q_j \in \mathcal{M}$. Because $V(n,q)$ is continuous and uniformly bounded on q, this approximation is asymptotically exact as $\Delta q \downarrow 0$.

- Second, the value of Δt_i^λ above is equal to infinity for $i = 0$ and $i = M-1$. This is a consequence of the fact that the HJB has two singularities at $q = 0$ and $q = 1$. Because the DP recursion works backward on q, we are only concerned with the singularity at $q = 1$ that defines one of the border conditions. To bypass this technical problem, we use the (left) continuity of $V(n,q)$ at $q = 1$ and the border condition $V(n,1) = W(n,\theta_L)$ to define a new border condition at $q_{M-1} = 1 - \Delta q$ such that

$$V(n, M-1) = W(n,\theta_L) - \Delta q \, \frac{\mathrm{d}V(n,q)}{\mathrm{d}q}\Big|_{q=1^-}.$$

By Lemma B4 in Appendix B, we can compute recursively the derivative of $V(n,q)$ as follows.

$$\frac{\mathrm{d}V(n,q)}{\mathrm{d}q}\Big|_{q=1^-} := V_q(n,1) = \frac{V_q(n-1,1) - \frac{r\,(\theta_H-\theta_l)\,W(n,\theta_L)}{\theta_L\,\theta_H}\,\Phi'\left(\frac{r\,W(n,\theta_L)}{\theta_L}\right)}{1 + \frac{r}{\theta_H}\,\Phi'\left(\frac{r\,W(n,\theta_L)}{\theta_L}\right)} \quad \text{and} \quad V_q(0,1) = 0.$$

- In terms of computational complexity, we solve the optimization in (d1) using a line search.

D2. Algorithm With the Optimal Stopping Option

In this case, the only change that we need to introduce in the previous algorithm is to change the recursion in equation (d1) to

$$U(n,i) = \max\left\{R\,,\ \max_{0\le\lambda\le\Lambda} \mathbb{E}_{q_i}\left[\mathbb{1}_{\{\tau_i^\lambda\le\Delta t_i^\lambda\}}\, e^{-r\,\tau_i^\lambda}\,\left(p(\lambda) + U(n-1, i-\eta(i))\right) + \mathbb{1}_{\{\tau_i^\lambda>\Delta t_i^\lambda\}}\, e^{-r\,\Delta t_i^\lambda}\, U(n,i+1)\right]\right\}, \tag{d4}$$

with border conditions $U(0,i) = R$ and $U(n,M) = R$, for all $i = 1,\ldots,m$ and all n. Once $U(n,i)$ has been computed for all i, we can determine the threshold q_n^* to be equal to $i^*\,\Delta q$, where $i^* = \min\{i\,:\,U(n,i) = R\}$.

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., M. Queyranne. 2005. Toward robust revenue management: Competitive analysis of online booking. Working Paper, University of Maryland. Available at SSRN: http://ssrn.com/abstract=896547.

Bertsimas, D., G. Perakis. 2005. Dynamic pricing: A learning approach. Working Paper, Sloan School of Management, MIT, Cambridge. Massachusetts.

Besbes, O., A. Zeevi. 2007. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. Working Paper, Columbia University.

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, NY.

Caro, F., J. Gallien. 2005. Dynamic assortment with demand learning for seasonal consumer goods. Working Paper, Sloan School of Management, MIT, Cambridge. Massachusetts.

Carvalho, A.X., M.L. Puterman. 2004. Dynamic pricing and learning over short time horizons. Working Paper, Statistisc Department, University of British Columbia, Vancouver.

Cheney, W. 2001. *Analysis for applied mathematics*. Springer-Verlag, New York, NY.

Cope, E. 2004. Nonparametric strategies for dynamic pricing in e-commerce. Working Paper, The Sauder School of Business, University of British Columbia, Vancouver.

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.

Ethier, S.N., T.G. Kurtz. 1986. *Markov Processes: Characterization and Convergence*. Wiley, New York, NY.

Farias, V.F., B. Van Roy. 2007. Dynamic pricing with a prior on market response. Working Paper, Stanford University, California.

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, NY.

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

Keller, G., S. Rady. 1999. Optimal experimentation in a changing environment. *Review of Economic Studies* **66** 475–507.

Lariviere, M.A. 2005. A note on probability distributions with increasing generalized failure rates.

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. Sousa, S. Boyd. 2003. Pricing and learning with uncertain demand. Working Paper, Duke University.

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. Stat.* **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, NJ.

Scarf, H. 1958. Bayes solutions of the statistical inventory problems. *Ann. Math. Stat.* **30** 490–508.

Shiryayev, A.N. 1978. *Optimal Stopping Rules*. Springer-Verlag, New York, NY.

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*. Kluwer Academic Press.

Xu, X., W.J. Hopp. 2005. Dynamic pricing and inventory control: The value of demand leraning. Department of Industrial Engineering, Northwestern University.