A Partially Observed Markov Decision Process for Dynamic Pricing*

Yossi Aviv, Amit Pazgal
Olin School of Business, Washington University, St. Louis, MO 63130
aviv@wustl.edu, pazgal@wustl.edu

April, 2004

Abstract

In this paper, we develop a stylized partially observed Markov decision process (POMDP) framework, to study a dynamic pricing problem faced by sellers of fashion-like goods. We consider a retailer that plans to sell a given stock of items during a finite sales season. The objective of the retailer is to dynamically price the product in a way that maximizes expected revenues. Our model brings together various types of uncertainties about the demand, some of which are resolvable through sales observations. We develop a rigorous upper bound for the seller's optimal dynamic decision problem and use it to propose an active-learning heuristic pricing policy. We conduct a numerical study to test the performance of four different heuristic dynamic pricing policies, in order to gain insights into several important managerial questions that arise in the context of revenue management.

Keywords: Learning, Partially Observed Markov Decision Processes, Pricing, Revenue Management.

*We wish to thank the Boeing Center for Technology, Information, and Manufacturing (BCTIM), for providing funding to support this research.

1 Introduction

Retailers of seasonal, fashion-like products are faced with the complex task of dynamically setting prices in a way that maximizes their expected revenues. In this paper we focus our analysis on the revenue management process which starts at the beginning of the sales season, after the sellers already purchased the goods, and just as they are eagerly beginning to sell them to customers. The global nature of supply chains of products such as high-fashion apparels, and toys with short life cycles (e.g., those introduced simultaneously with popular movies; Johnson 2001), is that sellers often have no opportunity to replenish the inventory in their stores during a short sales season (see also Hammond and Raman 1996). In this paper, we construct a pricing model for a retailer who sells a finite stock over a short season. Our model embeds five types of uncertainty that are common in fashion-like industries: (i) the uncertainty about the *rate of arrivals* of customers to the store; (ii) the actual *number* of customers that arrive to the store during the horizon and their exact *times of arrival*; (iii) the statistical characterization of the arriving customers' *reservation prices* (the maximal prices that customers are willing to pay for the product;) (iv) the *individual* consumer's *purchase decisions* based on personal reservation prices; and finally (v) the *state* of the market, a concept we present in the next paragraph.

Demand for retail goods is usually highly dependent on changes in basic economic variables that describe either the economy as a whole, or the state of the specific industry (Song and Zipkin 1993). For fashion-like or new products, the evolution of customer perceptions of their values will determine the level of sales. To model demand processes in fluctuating business environments, researchers have suggested the use of a set of *states of the world*. Each state includes all the relevant information about the demand, and the transition between states is governed by a Markov chain (see, e.g., Sethi and Cheng 1997). In this paper, we consider a Markovian demand environment, that transitions among a finite number of *core states*, each characterized by uncertainties of the types (i)-(iv) presented above. As in most practical settings of the type we are studying, retailers do not fully observe the exact core state of the demand environment. In other words, they may experience an ongoing uncertainty about which particular demand scenario best describes the demand at each point of time (see the fifth type of uncertainty above). In such cases, it is often argued that sales figures are of critical importance in providing sellers with good means to *learn* about the state of demand; see, e.g., Fisher et al. (2000). In order to accomodate the latter type of uncertainty, we use a partially observed Markov decision process (*POMDP*) framework. A POMDP

is a sequential decision problem, pertaining to a dynamic setting, where information concerning the state of the world is incomplete (Lovejoy 1991), and is therefore applicable to our study. The potential applications of POMDP's in studying managerial problems of significant interest have been described in previous research. We refer the reader to Monahan (1982) for examples relevant to areas such as quality control, machine maintenance, internal auditing, learning, and optimal stopping. Additional applications to problems in fishery, drug therapy, and inventory, can be found in Lane (1989), Hu et al. (1996), and Treharne and Sox (2002), respectively.

The paper is organized as follows. In §1.1 we provide a review of the relevant literature. In §2 we present a *descriptive* model for the demand environment, its dynamics, and the specification of the information available to the seller. A *prescriptive* dynamic pricing problem is described in §3, along with a brief literature review on available solution methods to POMDP's. We then turn to our key section of the paper (§4), in which we develop an approximate model for studying the seller's *original* dynamic pricing problem. The key idea underlying our approach is that through an appropriate augmentation of the information structure, we can transform the highly-complex original problem into an easily solvable one. Such information structure modification not only aids tractability, but it also provides a rigorous upper bound on the optimal expected revenues. We include in ths section a description of a very simple algorithm that solves the approximate problem, as well as a heuristic pricing strategy driven by the approximate model. In §5 we report on a comprehensive numerical study we conducted in order to examine the quality of the approximate model as well as the effectiveness of the heuristic pricing policy driven by it. Additional analyses are described in §6 and the paper is concluded with a summary of our key results in §7.

Our work offers a dual contribution to the academic literature. Readers interested in revenue management can benefit from an analysis of a scenario-based model for dynamic pricing, accompanied by some managerial insights obtain through our extensive numerical study. Our models bring together not only the consideration of a variety of uncertainties that characterize demand environments for fashion-like goods, but also the scarcity of products (i.e., limited inventory). Readers with this type of interest in mind may feel comfortable skipping the technical details of §4, and rely on a general understanding of the core ideas presented there. For those interested in applications of POMDP's, we offer a methodology that may prove to be very effective in terms of providing a rigorous upper bound value approximation, fast computation times, and practically-appealing heuristics. The presentation in §4 is only loosely context-specific, and can be easily generalized.

1.1 Relevant Literature Review

Dynamic pricing for revenue maximization is a timely but not a new topic for discussion in the academic literature. We provide a brief description of the literature that is most relevant to our discussion. Gallego and van Ryzin (1994) developed a continuous-time pricing model in which customers' arrivals are governed by a Poisson process with a constant and known price-dependent arrival rate. At each point of time the seller has to balance the price appropriately. On one hand, the seller does not want to sell the product at a price that is lower than its potential. On the other hand, posting a high price risks lost sales. They find that fixed pricing policies generally do quite well. Consequently, they argue that adjusting prices in order to compensate for statistical fluctuations in demand has a very little advantage. Bitran and Mondschein (1997) study optimal pricing strategies for seasonal products, in a periodic review model similar in spirit to that of Gallego and van Ryzin (1994).

Aviv and Pazgal (2002) extend the previous work by taking into account uncertainty of type (i) listed above. Through an observation of sales during the season, the seller is able to learn about the store traffic intensity rate. They argue that dynamic pricing strategies are critical in settings with high but resolvable initial uncertainty about the demand. Yet, it is not essential to consider the impact of pricing on the learning process ("active learning"). Instead, it is reasonable to adopt a "passive learning" approach using certainty-equivalent policies, according to which the seller updates his belief about the market condition as time progresses, but at each moment, the price is set as if the market condition is known and equals to its current estimate. Other researchers have considered optimal dynamic pricing in uncertain demand environments, but when there is no limit on the level of inventory. Carvalho and Puterman (2003) develop and compare several heuristic pricing policies. For a two-period model, they construct a Taylor expansion of the reward function in order to explain the trade-off between short-term revenue maximization and future information gains. The authors use the latter result to develop a so-called one-step look ahead policy for a periodic review pricing control. Braden and Oren (1994) study a pricing problem for a firm that sells products to customers with different tastes that vary according to an index whose value is unknown. They show that it is optimal in their setting to follow a policy that maintains the separation principle (see, e.g., Bertsekas 1995): The control of prices can be done in a way that does not need to consider the evolution of information, as long as the learning about the unknown demand parameter is done periodically, prior to each decision.

Li (1988) developed a continuous time model in which both production and prices are controllable by a monopoly, and Federgruen and Heching (1999) study finite and infinite-horizon periodic-review joint inventory and pricing problems. In both cases the firm is interested in maximizing some functional form of profits generated over the sales horizon. These papers provide characterizations of optimal joint pricing and inventory replenishment strategies. Active learning about demand was studied by Balvers and Cosimano (1990). In their paper, a firm faces a demand curve that is linearly decreasing in price, but with unknown slope and intercept. Through the determination of prices, the firm controls its profit and learning about the demand. Nevertheless, inventory is not an issue, since production equals demand during every period. The literature review in Balvers and Cosimano contains references to active learning through pricing in competitive environments. See also Petruzzi and Dada (2002) for a periodic review model of perishable goods. Under their specific model features, learning about the demand is done solely through inventory decisions, and not via pricing.

2 A Descriptive Demand Model

We present below a descriptive model that characterizes three important aspects of practical demand environments: The evolution of demand in the market, the dependecy of demand on pricing decisions, and the type and level of uncertainty the seller has about the demand. The following subsections explain our modeling choices and the logic behind them.

2.1 Markov-Modulated Demand Environments

To model demand evolution during sales season of fashion-like products, one need to consider possible variations in demand that result not only due to purely random reasons, but also due to *fluctuations in the demand environment*. The latter source of variation was captured, for instance, in a Markov-modulated demand process studied by Song and Zipkin (1993). The underlying logic behind this model is that, often, there are certain factors in the marketplace that determine the demand environment. These environmental factors may change unpredictably over time, and as a consequence affect demand. In fact, Song and Zipkin use as an illustration a new product introduction setting, very relevant in our context. They argue that fluctuations in a demand environment may be particularly noticeable at the beginning and the end of the product life cycle. For instance, the timing and degree of decline at the end of the product life-cycle may be unknown and dependent on consumer tastes and innovation in the market (both dynamic in nature).

Formally, the Markov-modulated demand model includes a set of possible *core states* (this set is denoted below by the letter Ω) that represent different statistical characterizations of the demand environment. When the state of the demand during period n is $C_n = k$, we say that period n is of *type* k. Specifically, each state $k \in \Omega$ of the core process is characterized by a discrete, price- and period-dependent probability functions b_k, where $b_k(d|n,p)$ is the probability that the demand during period n of type k is d, given that a price p was set for this period. In our numerical studies, we constructed the probabilities b in the following way: For each core state $k \in \Omega$, we modeled the "potential demand" as a Poisson process with a given rate $\lambda_k > 0$. One can think of this potential demand process as customer *arrivals* to the store. Still, in order for a customer arrival to be converted into a purchase, the price of the product has to be acceptable to that individual. Let the *reservation price* be the maximal price a customer is willing to pay for a unit of the product. As long as the posted price is lower than this value, and the product is in stock, a purchase is made. We assume that the customer base is heterogeneous with individual reservation prices having an exponential distribution. Let α_k represent the mean reservation price in the market under core demand state k. Clearly, for any period n of length τ_n, the statistical distribution of the actual demand is Poisson, with a mean value of $\lambda_k e^{-p/\alpha_k}\tau_n$. With the restriction to this statistical probability structure we lose some generality, but in return we gain tractability and the ability to present our results in an easily understandable format.

The dynamics of the demand environment follows a Markov chain $C = \{C_n : n \geq 1\}$, governed by a stationary $K \times K$ transition probability matrix A, where $A_{jk} = \Pr(C_{n+1} = k | C_n = j)$. Sticking to the terminology used by Monahan (1982), we name C the *core process*. The dynamics of the core demand process during the sales horizon may take an unlimited number of forms. At one extreme, demand can be *static*, and thus reflected by the transition matrix A being equal to the $K \times K$ identity matrix. A second possible form of dynamics is *declining demand* trend, in which demand starts at a certain (possibly unknown) state, and then may shift to a state of lower demand, associated with lower α and λ values. For instance, sales of items like swimsuits, skiwear and toys typically follow this pattern (see, e.g., Hammond and Raman 1996, and Johnson 2001). In contrast, an *increasing demand* pattern environment represent cases in which, as time passes, customer arrivals will be associated with statistically higher reservation prices. For example, in sales of airline tickets or hotel bookings, we often see that the demand environment may show an increase in α values as time gets closer to the time of capacity consumption (Desiraju and Shugan, 1999). Another interesting pattern of dynamics is one that combines both trends; e.g., begins with

an increase in the demand and then shift downwards. The ability to examine the uncertainty in the timing of a decline or an increase in the demand state, seems to be important particularly in setting of the types we study. For example, Johnson (2001) describes the sales of action figure toys through licensing agreements with movie producers. The timing of introduction of the product into the market is often critical, and due to relatively long lead times, inventory replenishment opportunities are generally limited. Typically, the level of demand depends on unknown customer tastes, the success of the movie, and how saturated the market is with movie-related products. Adding to this complication is the nature of demand, that after a certain time it tends to decline quite sharply, leading retailers to offer significant markdowns for any remaining inventory.

2.2 A Partially-Observed MDP

To describe the uncertainty about the demand environment, we consider a situation in which the seller does not necessarily know the exact core state of demand, even for the most current period of the sales horizon. For instance, it may be that at the beginning of the sales season the seller has only a vague idea about the reaction of customers to the product. Such partial knowledge can be represented by a probability distribution over the set of core states Ω. We therefore define $\pi_n = (\pi_{n,1}, \ldots, \pi_{n,K}) \in \Pi$ as the *belief* of the seller about the current state of demand, generated at the beginning of period n ($\Pi = \left\{v \in \mathbb{R}^K : \sum_{k\in\Omega} v_k = 1,\ v \geq 0\right\}$). Note that π_1 is the prior belief about the state of demand during the first period, at the start of the sales horizon, typically determined on the basis of early sales information, previous experience with similar items, market research, etc. The partial observability feature of our model enables us to describe some of the crucial challenges faced, for instance, by marketers of fashion apparel, jewelry, specialty foods, books, and so on.

At the end of each period n, and just before making a pricing decision for the following period, the seller updates his knowledge about next period's core state of demand (i.e., C_{n+1}) on the basis of the previous belief (π_n), the current period's price (p_n) as well as the demand observation (d_n). In other words, we assume that the seller observes the demand process, so that at the beginning of period n, exact information about the history of demands and prices $\{(p_j, d_j) : j = 1, \ldots, n-1\}$ is available to the seller. Under these assumptions, the dynamics of the seller's belief, $(n, d_n, \pi_n, p_n) \to \pi_{n+1}$, is provided by a Bayesian transformation $\Theta : \Pi \to \Pi$:

$$\pi_{n+1,k} = \Theta_k\left(n, d_n, \pi_n, p_n\right) = \frac{\sum_{j\in\Omega} A_{j,k} \cdot b_j\left(d_n | n, p_n\right) \cdot \pi_{n,j}}{\sum_{j\in\Omega} b_j\left(d_n | n, p_n\right) \cdot \pi_{n,j}}, \quad k \in \Omega \tag{1}$$

It is instructive to note that implicitly in transformation (1) is the assumption that the only source of information about the core state gathered throughout the sales horizon is the history of demand figures and prices. Of course, in some settings, information from other sources may be available. One possible way to deal with this issue, is by enlarging the state space to accomodate additional information paths. An alternative to that is to define another process (say, $X = \{X_1, \ldots, X_N\}$) that mimics information signals obtained by the decision-maker, about the true state of the system. In fact, our paper presents a framework that is flexible enoguh to accomodate both of these approaches. The main caveat for dealing with complex information structures is the increasing complexity of the decision making process. Our paper addresses this issue as well (see §4).

3 The Dynamic Pricing Problem

The seller's goal is to follow a dynamic pricing policy scheme that brings the best expected total revenues over the sales season. This task is quite challenging, as the seller need not only balance the trade-off between current versus future revenues in setting the prices, but also take into account the way prices affect the ability of the seller to learn about the state of the demand. Suppose that q units are in stock at the beginning of period n. For a given belief π_n, and any price p, the expected number of units sold in this period is given by

$$S_n\left(q|p,\pi_n\right) = q - \sum_{j\in\Omega}\pi_{n,j}\sum_{d<q}\left(q-d\right)\cdot b_j\left(d|n,p\right)$$

We now state the dynamic optimization model for our decision process. Let $V_n^*\left(q,\pi_n\right)$ be the optimal expected total revenues generated over periods $n, n+1, \ldots, N$. Then,

$$V_n^*\left(q,\pi_n\right) = \max_p\left\{pS_n\left(q|p,\pi_n\right) + \sum_{j\in\Omega}\pi_{n,j}\sum_{d<q}V_{n+1}^*\left(q-d,\Theta\left(n,d,\pi_n,p\right)\right)\cdot b_j\left(d|n,p\right)\right\} \qquad (2)$$

We assume that no salvage value can be gained from unsold units at the end of the sales season. Hence, $V_{N+1}\equiv 0$. We also note that although the Bayesian dynamics Θ should be defined as a function of the sales and not the demand, this technical issue is of no concern since demand may not be equal to sales only if all the items were sold, in which case pricing is no longer relevant.

As reflected in (2), we used the Bayesian scheme in order to convert the partially observable Markov decision process into an equivalent, completely observable Markov decision process (this is done by including the belief vector in the state variable). Nevertheless, the price we pay for this

transformation is a significant increase of the state space dimension, as well as turning the state space into an uncountable one (see Monahan 1982). While in theory our pricing problem can be solved dynamically through (2), this approach is tedious and impractical for even small values of K. In fact, it is mainly due to the intractability of POMDPs that they have not gained high popularity in practical applications (Lovejoy 1991).

Several computational procedures exist for solving finite-horizon POMDP's. Smallwood and Sondik (1973) show that a POMDP defined on the state space Π, has an optimal value function that is piecewise linear and convex, and so the space of belief vectors can be partitioned into a finite set of convex regions separated by linear hyperplanes. For computational purposes, this property allows one to convert the state space of the dynamic problem to an appropriately selected grid. At each iteration in the dynamic algorithm, the regions have to be reconstructed and their number can grow exponentially, thus exacting high computational expense. Several papers have suggested improvement to Smallwood and Sondik's algorithm; see, e.g., White and Scherer (1989). Monahan (1982) provides a good survey on other types of algorithms. Lovejoy (1991) propose an approximation technique, applicable to both finite and infinite-horizon POMDP's, that makes use of an easily managed *fixed* grid of points in Π. In addition to providing the method for constructing such grids, Lovejoy shows how to use them in order to develop rigorous upper and lower value function bounds, and generate heuristic policies. Lovejoy (1993) blends his discretization method (Lovejoy, 1991) and the subgradient results of Sondik (1971) to create an iterative method for constructing improved (i.e., tighter) bounds on the value function of the POMDP, as well as generate heuristic policies.

4 Approximate Solutions to POMDPs via Information-Structure Modification (ISM)

In this paper, we do not make use of any of the algorithms discussed above. Instead, we propose an approximation technique that is based on *information-structure modification* (ISM) to the *original* dynamic problem (2). As we shall see, our approximate model will lend itself to an easily solvable, fully observable Markov decision process, defined on a simple discrete and finite state space. In addition to very fast computation time, the solution to the ISM model serves as a rigorous upper bound on the optimal value function of the POMDP. We use the approximate model to generate heuristic pricing policies and gauge their performance vis-a-vis the theoretical optimal pricing strategies. The approximate models also provide us with an intuition useful in the interpretation

of the results of our numerical studies.

To reduce the computational burden associated with the dynamic program (2), we take the following approach: Instead of attempting to solve the original complex dynamic program, the seller considers a similar model but with a modified information structure. Specifically, the seller will assume for the sake of mathematical tractability, that at the end of each period, he will obtain some additional information about the core process from an *hypothetical* source. We shall refer to the resulting (hypothetical) decision problem as the *ISM model*. Clearly, since in the ISM model the seller expects to obtain more information than in the original problem, the optimal solution to the approximation is higher than that of the original problem. We begin with a simple illustration.

Example 1 (Full Information) *Consider a case in which an hypothetical source provides the seller with perfect retrospective information about the state of the core process. In other words, prior to determining the price for period $n+1$, the seller finds out the exact core state in the previous period (i.e., C_n). With this modification, the seller's decision problem is simplified as follows: For a given period n_0, and q, let*

$$V_{n_0}^{FI}(q, \pi_{n_0}) = \max_{p} \left\{ \sum_{k \in \Omega} \pi_{n_0,k} \sum_{d} b_k(d|n_0, p) \left(p \cdot \min(d, q) + V_{n_0+1}^{FI}\left([q-d]^+, A_k \right) \right) \right\} \quad (3)$$

where $V_{N+1}^{FI} \equiv 0$. (The value $[x]^+$ represented the maximum between x and 0).

We call the decision problem presented in (3) the *full-information (FI) model*. It is instructive to note that the FI model is based on *retrospective*[1] information, and not on a complete knowledge of the current state of the core process. The FI model has a simple Markov decision problem form, and is easily solvable. Furthermore,

$$V_{n_0}^{*}(q, \pi_{n_0}) \leq V_{n_0}^{FI}(q, \pi_{n_0})$$

for every n_0, q, and $\pi_{n_0} \in \Pi$. Upper bound approximations of the type $V_{n_0}^{FI}(q, \pi_{n_0})$ are in fact known in the literature (e.g., Astrom 1965, White 1976, Van Hee 1978). They are attractive, since they substitute the continuous component of the state space of the Markov decision process (i.e., Π) with a discrete and finite space (Ω). This, in turn, permits an easy and quick calculation of the upper bounds. Nevertheless, we expect the upper bound values $V_{n_0}^{FI}(q, \pi_{n_0})$ to be *crude* in most interesting cases.

[1] We use the term "retrospective information" to clarify that information about the state of the core process is obtained at the *end* of a period, after the price was set for that period.

So, what value can be gained through ISM? Ideally, the seller should not base his decisions on an assumption about an hypothetical source that provides too much information about the true state of the core process. For otherwise, this might result in making poor decisions. At the same time, the decision-maker needs to smartly invent an hypothetical information structure that will make the problem easily solvable – for otherwise, we get back to the original challenge of solving a complex POMDP.

4.1 ISM: A Probabilistic Scheme

In this section, we propose a possible class of ISM schemes that do not reveal the exact current state of the core process, but rather *augment* the real information process with an hypothetical *partial* retrospective information. Formally, at the *end* of each period, on top of the known values of n, D_n, p and π_n, the decision maker obtains from the hypothetical source additional information through a (probabilistic) process $X_n\left(n, D_n, p, \pi_n, C_n\right) \in \Xi$ (no restrictions are yet imposed on the state space Ξ). Hereafter, we shall refer to the process $X = \{X_1, \ldots, X_n\}$ as the ISM. For example, the ISM underlying the upper bound $V_{n_0}^{FI}\left(q, \pi_{n_0}\right)$ is given by $X_n = C_n \in \Omega$. In order to properly define the ISM X, we use the probability functions

$$f_{C_n}\left(x|n, d, p, \pi_n\right) \doteq \Pr\left\{X_n = x|n, D_n, p, \pi_n, C_n\right\},\ x \in \Xi$$

In other words, the value of $f_{C_n}\left(x|n, d, p, \pi_n\right)$, is equal to the probability that the additional (hypothetical) information provided to the decision-maker at the end of period n, will be x, given the values $\{C_n, n, D_n = d, p, \pi_n\}$. With the process X in mind, the decision-maker is interested in solving the dynamic program

$$\begin{aligned} V_n^{ISM}\left(q, \pi_n\right) = \max_p \Big\{ & pS_n\left(q|p, \pi_n\right) \\ & + \mathsf{E}_{(C_n, D_n)}\left[\sum_{x \in \Xi} V_{n+1}^{ISM}\left(\left[q - D_n\right]^+, \Theta^X\left(n, D_n, \pi_n, p, x\right)\right) \cdot f_{C_n}\left(x|n, D_n, p, \pi_n\right)|n, p, \pi_n\right] \Big\} \quad (4) \end{aligned}$$

where Θ^X is the evolution of the belief vector in the approximate ISM model, given by the Bayesian updating mechanism

$$\pi_{n+1,k} = \Theta_k^X\left(n, d, \pi_n, p, x\right) = \frac{\sum_{j \in \Omega} \pi_{n,j} b_j\left(d|n, p\right) f_j\left(x|n, d, p, \pi_n\right) A_{j,k}}{\sum_{j \in \Omega} \pi_{n,j} b_j\left(d|n, p\right) f_j\left(x|n, d, p, \pi_n\right)},\ k \in \Omega \quad (5)$$

Generally, ISM does not necessarily result in a model that is simpler than the original POMDP (in fact, the opposite may happen). For instance, if the process X is one that randomly points to

a state, irrespective of the state of the core process (e.g., $f_j(x|n,d,p,\pi_n) = \frac{1}{K}$, where $x \in \Xi = \Omega$), the resulting model will be equivalent to the original problem. Recall that the complexity of POMDP problems arise from the fact that the initial belief vector may evolve in many possible different paths $\{\pi_1, \ldots, \pi_N\}$, even within a short time horizon N. This is particularly the case if the dynamic of the belief vector strongly depends on the actions taken in each period (i.e., prices), and the demand realizations. The ISM model discussed in Example 1 can be characterized by $\Xi = \Omega$, and $f_{C_n}(x|n,d,p,\pi_n) = \mathbf{1}\{x = C_n\}$. Applying the latter updating scheme, we get $\Theta^{FI}(n,d,\pi_n,p,x) = A_x$, the x-th row of the transition matrix A. Thus, this example serves as an illustration of a process X that simplifies the structure of the dynamic program in a way that enables us to substitute Π with a finite set of only $K+1$ belief vectors: $\{\pi_1, A_1, \ldots, A_K\}$. This ISM is therefore very appealing from a computational perspective, yet it is not necessarily a good approximation.

Clearly, since the decision-maker can always ignore the added information stream X, the solution of the ISM-based problem always serves as an upper bound for the original POMDP. We hence conclude this section with the observation that for any ISM process X,

$$V_n^*(q,\pi_n) \leq V_n^{ISM}(q,\pi_n)$$

4.2 An ISM Upper Bound

In this section we construct an ISM model that gives rise to an easily calculable upper bound on the optimal expected revenue function. We do so in two key steps: The first step is driven by the observation that a significant part of the complexity of a POMDP can be avoided if we use X-processes for which the belief evolution Θ^X is *conditionally independent* of d and p, given a value of X. We thus propose a specific class of probability forms f that yields a conditional independency. In the second step we show how one can even further simplify the dynamic program, to make it linearly hard in the horizon N, rather than exponentially hard, by further restricting the class of probabilities f. Depending on the perceived trade-off between computational simplicity and upper bound accuracy, one may either use the first step or both steps.

With the restrictions we suggest below, we do not exploit the full potential of the ISM method we outlined above. Nevertheless, the models below have a computational appeal, and they enable us to keeps the exposition of our application at a reasonable length, complexity and clarity[2].

[2] More accurate approximations through highly sophisticated ISM methods is a topic of an ongoing research. Such methods are complex mathematically, and hence a discussion of them is excluded from this work.

4.2.1 Probability Class Restriction

In the class of ISM's we consider below, the probability functions f take the special form

$$f_j(x|n,d,p,\pi_n) = \frac{g_j(x|n,\pi_n)\, h(d,x,n,p,\pi_n)}{b_j(d|n,p)} \tag{6}$$

for some general but appropriately defined[3] functions g and h. For instance, the FI model of Example 1 satisfies this property with $g_j(x|\cdot) = \mathbf{1}\{x=j\}$, and $h = b_x(d|n,p)$. The primary advantage of the probability form (6) is that in the ISM model, the decision-maker does not have to consider the subtle and complex way by which the decision variable p and the demand observation d affect the evolution of the belief vector. Observe that under the form (6), the Bayesian updating mechanism (5) reduces to:

$$\Theta_k^X(n,d,\pi_n,p,x) = \hat{\Theta}_k^X(n,\pi_n,x) = \frac{\sum_{j\in\Omega}\pi_{n,j}\cdot g_j(x|n,\pi_n)\cdot A_{j,k}}{\sum_{j\in\Omega}\pi_{n,j}\cdot g_j(x|n,\pi_n)} = \left(T_n(x,\pi_n)\,A\right)_k,\ k\in\Omega \tag{7}$$

where

$$T_n(x,\pi_n) = \left(\frac{\pi_{n,1}\cdot g_1(x|n,\pi_n)}{\sum_{j\in\Omega}\pi_{n,j}\cdot g_j(x|n,\pi_n)},\ldots,\frac{\pi_{n,K}\cdot g_K(x|n,\pi_n)}{\sum_{j\in\Omega}\pi_{n,j}\cdot g_j(x|n,\pi_n)}\right)\in\Pi$$

Therefore, the decision-maker only needs to consider the evolution of the belief vector π through the more limited number of possible X values. Clearly, the ISM model does not necessarily ignore the impact of actions on learning. While in the FI model decisions do *not* affect the values of future beliefs, the ISM in general incorporates the effects of actions on learning through the probabilities f. This is done via the functions h in dynamics of the ISM model driven by the form (6):

$$\begin{aligned} V_n^{ISM}(q,\pi_n) = \max_p \Big\{ & pS_n(q|p,\pi_n) \\ & + \sum_{j\in\Omega}\pi_{n,j}\sum_{x\in\Xi} g_j(x|n,\pi_n)\sum_{d<q} h(d,x,n,p,\pi_n)\cdot V_{n+1}^{ISM}\left(q-d,\hat{\Theta}^X(n,\pi_n,x)\right)\Big\} \end{aligned} \tag{8}$$

Model (8) gains its significant computational advantage due to the considerable reduction in the number of possible evolution paths $\{\pi_1,\ldots,\pi_N\}$ that need to be evaluated when solving the ISM problem. Particularly, when the decision variable is chosen in a continuous space, a given belief π_n may evolve, in the original model, into one out of an unlimited number of new beliefs even within a single period. In contrast, the number of possible ways a belief can evolve in model (8) over a single period is at most $|\Xi|$. Genrally, given a horizon of N periods, and an initial level of inventory q, one has to calculate at most $q\cdot|\Xi|^{N-1}$ values of the function V_N^{ISM}. Clearly, even though the

[3] All probabilities $f_j(x|n,d,p,\pi)$ have to be in the range $[0,1]$ and $\sum_{x\in\Xi} f_j(x|n,d,p,\pi) = 1$ for all possible combinations (n,j,d,p,π).

ISM model is much simpler than the original problem, the effort in solving a dynamic program of the type (8) is in general exponential in the horizon N. Therefore, for cases in which it is still hard or practically impossible to solve the latter dynamic model, we suggest that a further restriction in the probability form f should be considered.

4.2.2 Belief State Space Restriction

In order to achieve further simplification in the dynamic program (8), we take the following approach. Recall that in the FI model related of example 1, any belief vector is transformed, within a single period, into one of the K *core beliefs* $\{A_1, \ldots, A_K\}$. Therefore, the number of beliefs that need to be considered when assessing the values of V_n^{FI} $(n = 2, \ldots, N)$ is fixed and equals to K. Continuing with the functional form (6) discussed in the previous section, suppose that we select the probabilities f such that for *any* possible vector π_n, the evolution $\hat{\Theta}^X(n, \pi_n, x)$ is restricted to $K+1$ vectors, among which K are the core beliefs. In other words, we set $\Xi = \{0, 1, \ldots, K\}$, so that $\hat{\Theta}^X(n, \pi_n, x) = A_x$ for every $x \in \Omega$, but with no restriction on the belief $\hat{\Theta}^X(n, \pi_n, 0)$. Clearly, as illustrated in Figure 1 below, the maximal number of possible values the belief vector π_N can take is linear in N.

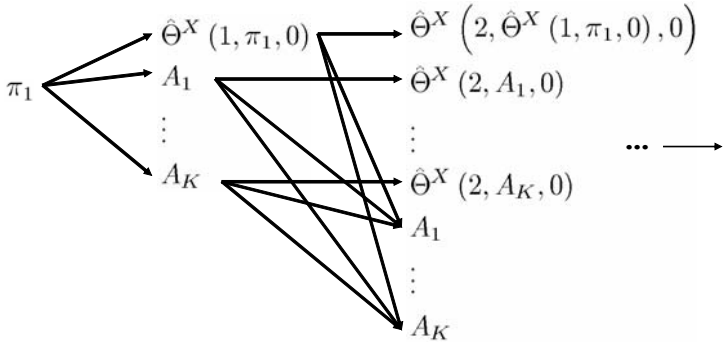

Figure 1: Linear evolution of the belief vector with state space restriction.

A feasible way to create a belief evolution of the type described above is by setting $g_j(x|n, \pi_n) = \mathbf{1}\{x = j\}$ for all $x \in \Omega$, in which case,

$$f_j(x|n, d, p, \pi_n) = \begin{cases} \frac{g_j(0|n,\pi_n)h(d,0,n,p,\pi_n)}{b_j(d|n,p)} & x = 0 \\ \mathbf{1}\{x = j\} \frac{h(d,j,n,p,\pi_n)}{b_j(d|n,p)} & x \in \Omega \end{cases}$$

The values $g_j(0|n, \pi_n)$ are then chosen in a way that for every combination of values $\{j, n, d, p, \pi_n\}$

the probabilities f are non-negative and satisfy

$$\sum_{x\in\Xi} f_j\left(x|n,d,p,\pi_n\right) = \frac{h\left(d,j,n,p,\pi_n\right)}{b_j\left(d|n,p\right)} + \frac{g_j\left(0|n,\pi_n\right)h\left(d,0,n,p,\pi_n\right)}{b_j\left(d|n,p\right)} = 1$$

The latter property imposes some important restrictions on the choice of the h functions as well as the function $g.\left(0|\cdot,\cdot\right)$. To see this, note that ideally, we would like the probability of getting the "signal" $x = 0$ to be 1. If this is possible, it means that the corresponding ISM process provides no valuable information (since it always signals the same value). Subsequently, not only that the relaxed problem would be easily solvable, but it would actually provide the *optimal* solution to the original POMDP. For example, such computationally-ideal case is attainable if the probabilities $b_j\left(d|n,p\right)$ can be written in the multiplicative form $g_j\left(0|n,\pi_n\right)h\left(d,0,n,p,\pi_n\right)$ for any feasible combination $\{j,n,d,p,\pi_n\}$. But this clearly represents a non-interesting case in which demand is independent of the core process. Intuitively, for any given set of functions $g_j\left(0|n,\pi_n\right)$, we wish to set $h\left(d,0,n,p,\pi_n\right)$ as high as possible, while maintaining the inequality $g_j\left(0|n,\pi_n\right)h\left(d,0,n,p,\pi_n\right) \leq b_j\left(d|n,p\right)$ for every $j\in\Omega$. In other words, we set $h\left(d,0,n,p,\pi_n\right) = \min_{k\in\Omega}\left\{b_k\left(d|n,p\right)/g_k\left(0|n,\pi_n\right)\right\}$ and consequently

$$f_j\left(x|n,d,p,\pi_n\right) = \begin{cases} \frac{g_j(0|n,\pi_n)}{b_j(d|n,p)}\min_{k\in\Omega}\left\{\frac{b_k(d|n,p)}{g_k(0|n,\pi_n)}\right\} & x=0 \\ 1-f_j\left(0|n,d,p,\pi_n\right) & x=j \\ 0 & \text{Otherwise} \end{cases} \tag{9}$$

(Note that $h\left(d,x,n,p,\pi_n\right)$ is simply defined as $\left(1-f_x\left(0|n,d,p,\pi_n\right)\right)\cdot b_x\left(d|n,p\right)$ for all $x\in\Omega$.) In other words, we interpret the ISM as one that at the end of each period "flips a coin" and either tells the true type of that period, or otherwise "says" 0. The specific complementary probabilities for these two outcomes are given in (9). It is instructive to note that, in general, even with the signal $X = 0$, valuable information may be passed to the decision maker. We now proceed with the presentation of the resulting ISM model.

$$\begin{aligned} V_n^{ISM}\left(q,\pi_n\right) = \max_p \Big\{ & pS_n\left(q|p,\pi_n\right) \\ & + \sum_{j\in\Omega}\pi_{n,j}\cdot\Big[\sum_{d<q}\left(b_j\left(d|n,p\right) - g_j\left(0|n,\pi_n\right)\cdot\min_{k\in\Omega}\left\{\frac{b_k\left(d|n,p\right)}{g_k\left(0|n,\pi_n\right)}\right\}\right)\cdot V_{n+1}^{ISM}\left(q-d,A_j\right) \\ & \qquad + \sum_{d<q} g_j\left(0|n,\pi_n\right)\cdot\min_{k\in\Omega}\left\{\frac{b_k\left(d|n,p\right)}{g_k\left(0|n,\pi_n\right)}\right\}\cdot V_{n+1}^{ISM}\left(q-d,T_n\left(\pi_n\right)\right)\Big]\Big\} \end{aligned} \tag{10}$$

where $T_n\left(\pi_n\right) \doteq T_n\left(0,\pi_n\right)$ is the value of the belief π_{n+1} if the ISM process X provides the value 0.

The functions $g_j\left(0|n,\pi_n\right)$ can be set with a large degree of flexibility, as long as they maintain the non-negativity constraints, and $g_j\left(0|n,\pi_n\right)>0$ for some $j\in\Omega$ for each pair (n,π_n). The simplest choice of g is to set it as a constant across $j\in\Omega$, or without loss of generality, set $g_j\left(0|n,\pi_n\right)\equiv 1$. This choice appeared to work very well for the majority of the settings we analyzed. For a subset of our instances, in which this choice has not demonstrated satisfactory results, we used an enhanced technique for setting the g-values as described below.

Consider the second component within the maximization expression in the dynamic program (10) (i.e., $\sum_{j\in\Omega}\pi_{n,j}\cdot[\ldots]$). Intuitively, we would like to "minimize" this value through the choice of the values $g_j\left(0|n,\pi_n\right)$. The rationale behind this proposition is that by minimizing the "over-optimistic" branches in the hypothetical, ISM-based decision-tree, we would gain a tighter upper bound. However, solving this problem is not an easy task, as the values of g cannot be contingent on the decision parameter p, or the demand realizations d. Moreover, the values $V_{n+1}^{ISM}\left(q-d,T_n\left(\pi_n\right)\right)$ depend on g – or in plain words – the g-values not only affect the likelihoods of the branches in the decision-tree, but also the expected rewards associated with them. We found through an intensive (heuristic) search among different functional forms, that $g_j\left(0|n,\pi_n\right)=b_j\left(0|n=1,\bar{p}\right)^{\beta}$ worked quite well. Specifically, the value of β provided us with a degree of freedom in searching for the tighter upper bound, and the value of $\bar{p}$ was set to be equal to p_1^{ISM} obtained in an initial iteration that uses $g\equiv 1$.

4.3 An Algorithm for the Upper Bound Model

In this section we outline an algorithm for solving problem (10). For simplicity of exposition, we describe the case $g_j\left(0|n,\pi_n\right)=g_j$ (i.e., constant values across all (n,π_n) combinations). Consider a given period n_0 ($n_0=1,\ldots,N$), at the beginning of which a belief π_{n_0} applies, and q_{n_0} units are available for sale. The value we are interested in calculating is $V_{n_0}^{ISM}\left(q_{n_0},\pi_{n_0}\right)$, and in order to attain this goal, we will dynamically solve (10) over the horizon $l=0,1,\ldots,N-n_0$ (the index l refers to period n_0+l). Define the vector valued function

$$\hat{\pi}\left(i\right)\doteq\begin{cases}\pi_{n_0} & i=0\\ A_i & i=1,\ldots,K\\ T\left(\hat{\pi}\left(i-K-1\right)\right)A & i=K+1,\ldots,\left(N-n_0\right)\cdot\left(K+1\right)\end{cases}$$

that can practically be stored in a matrix that includes the (only) $1+\left(K+1\right)\cdot\left(N-n_0\right)$ belief vectors used by our algorithm. Recall from Figure 1 above that by the beginning of period n_0+l,

the vector π_{n_0} evolves into one out of the $1 + l \cdot K$ belief vectors $\hat{\pi}(i)$, $i \in Z_l$, where

$$Z_l \doteq \{l \cdot (K+1), \{(K+1) \cdot (\delta - 1) + k,\, k = 1, \ldots, K;\ \delta = 1, \ldots, l\}\}$$

We thus define the functions

$$W_l(q, i) \doteq V_{n_0+l}^{ISM}(q, \hat{\pi}(i)),\ i \in Z_l$$

and note that equation (10) can be re-written as follows:

$$\begin{aligned} W_l(q,i) \;=\; & \max_p \Big\{ pS_n(q|p, \hat{\pi}(i)) \\ & + \sum_{j \in \Omega} \hat{\pi}(i)_j \cdot \Big[\sum_{d<q} \left(b_j(d|n,p) - g_j \cdot \min_{k \in \Omega} \left\{ \frac{b_k(d|n,p)}{g_k} \right\} \right) \cdot W_{l+1}(q-d, j) \\ & + \sum_{d<q} \min_{k \in \Omega} \left\{ \frac{b_k(d|n,p)}{g_k} \right\} \cdot W_{l+1}(q-d, i+K+1) \Big] \Big\} \end{aligned} \tag{11}$$

The implementation of the latter dynamic program algorithm is very simple, and can follow the next steps: For any initial values of $(n_0, \pi_{n_0}, q_{n_0})$, the matrix $\hat{\pi}$ is constructed. Then, the end-problem $W_{N-n_0}(q, i)$ is optimized for each possible combination of $q \in \{1, \ldots, q_0\}$ and $i \in Z_{N-n_0}$ (clearly, $W_l(0, \cdot) = 0$). The general dynamic programming algorithm can be applied now (i.e., calculating the values of W_l on the basis of W_{l+1}), with the values W_{l+1} discarded upon the completion of the computation of W_l. At the final step, only $W_0(q_{n_0}, 0) = V_{n_0}^{ISM}(q_{n_0}, \pi_{n_0})$ needs to be calculated. The computation of $W_0(q_{n_0}, 0)$ is usually very fast, and took no more than a couple of seconds per each instance studied in our numerical experimnetation (see §5.1).

The solution procedure for the maximization problem in each iteration of the dynamic program (11) is not necessarily simple. Under our choice of the functions g and the specific structure of the probabilities b, the expressions within the maximum operator in (11) appeared to be unimodal in the price, which lends itself to a simple search algorithm. We refer the reader to Ziya et al. (2002) for a description and comparison of various sets of sufficient assumptions that ensure unimodality.

4.4 An ISM-Based Heuristic Policy

The ISM approach can be applied not only for the purpose of generating upper bounds. We describe below an ISM-based dynamic pricing strategy, entitled the *ISM heuristic*. Suppose we are at the start of the horizon (i.e., $n_0 = 1$), having q_1 units available, and equipped with a belief π_1. In order to determine the price for the first period, the seller solves model (11) for $V_1^{ISM}(q_1, \pi_1)$. Together with this upper bound value, the seller obtains a contingent pricing scheme that is dependent on

the hypothetical ISM process X. In practice, only the first price $p_1^{ISM}(q_1, \pi_1)$ is actually used out of this pricing scheme. Then, at the beginning of the second period the seller updates his belief about demand based on the *actual* information only. In other words, the new belief would become $\pi_2 = \Theta\left(1, d_1, \pi_1, p_1^{ISM}\right)$. The seller then simply repeats the above procedure: At the beginning of each period n, the belief π_n is calculated using Bayesian scheme (1). Then, depending on the number of units on hand (q_n), the problem $V_n^{ISM}(q_n, \pi_n)$ is solved, and the initial price it prescribes, $p_n^{ISM}(q_n, \pi_n)$, is applied for that period. Let $H_1^{ISM}(q_1, \pi_1)$ be the expected revenue under the ISM heuristic policy, and note that because this heuristic is suboptimal we have

$$H_1^{ISM}(q_1, \pi_1) \leq V_1^*(q_1, \pi_1) \leq V_1^{ISM}(q_1, \pi_1)$$

We developed a software code that simulates the use of the above heuristic in a fluctuating demand environment. Using this simulation, we calculated the average total revenues collected under the heuristic, $\bar{H}_1^{ISM}(q_1, \pi_1)$, as an unbiased estimate for $H_1^{ISM}(q_1, \pi_1)$. For any given instance (N, π_1, q_1, A, K, b), the simulation was repeated many times ($10,000$ in our numerical studies) in order to generate a reasonable statistical confidence interval. Because we do not solve the optimal problem (2), the gap between the estimate $\bar{H}_1^{ISM}(q_1, \pi_1)$ and the upper bound $V_1^{ISM}(q_1, \pi_1)$,

$$\eta^{ISM} \doteq \frac{V_1^{ISM}(q_1, \pi_1)}{\bar{H}_1^{ISM}(q_1, \pi_1)} - 1$$

plays an important role. If the gap is relatively small, then we achieve two key objectives: The value $V_1^{ISM}(q_1, \pi_1)$ is a sufficiently *tight upper bound* (i.e., $V_1^{ISM}(q_1, \pi_1) \approx V_1^*(q, \pi_1)$), and in addition, the heuristic policy is *close-to-optimal* (i.e., $H_1^{ISM}(q_1, \pi_1) \approx \bar{H}_1^{ISM}(q_1, \pi_1) \approx V_1^*(q, \pi_1)$). When refering to the special ISM case of $X = C$, we will use "FI" in the superscripts.

In cotrast to the ISM heuristic, *open-loop feedback control* (OFLC) policies determine prices through a model that "freezes" the path of the belief vector. In other words, the decision-maker assumes that he would be incapable of updating his belief about the core state on the basis of prices and demands observations. For example, when setting the price for period n_0, the belief path $\left\{\bar{\pi}_{n_0+l} = \pi_{n_0} A^l : l = 0, \ldots, N - n_0\right\}$ can be used. Along the lines of Bertsekas (1995; §6.2), the seller solves the following $(N - n_0 + 1)$-period dynamic program:

$$V_n^{OFLC}(q) = \max_p \left\{ pS_n(q|p, \bar{\pi}_n) + \sum_{j \in \Omega} \bar{\pi}_{n,j} \sum_{d<q} b_j(d|n, p)\, V_{n+1}^{OFLC}(q - d) \right\}, \quad n \geq n_0 \tag{12}$$

Like the ISM heuristic, the OFLC heuristic updates the belief π_n before solving (12). Note that while the ISM method is based on an exaggeration of the content of information, OFLC policies act

as if the seller cannot access the information about historical actions and demands when it comes to learning. We shall examine this limitation in our numerical study.

4.5 Discussion

Small gaps can be expected, for instance, when demand is not significantly correlated with the core process, or when the transition probability matrix A has similar row entries. In such cases, learning does not play an important role in the decision process and so the ISM model would closely approximate the optimal value function. In fact, even the FI model would work well for these cases, both as a bound and as an embedded decision tool in the heuristic. Generally, small gap values that are based on the FI model (i.e., η^{FI}) indicate cases in which learning is not essential.

Recall that optimal pricing policies need to produce a good balance between learning about the environment and an ongoing generation of revenues. To this end, prices may be set in a way that sacrifices short term gains in lieu of information that will lead to higher expected benefits in the future. It is generally hard to clearly distinguish between these effects, since they are usually inter-related in an intricate way. Carvalho and Puterman (2003) provide an interesting treatment of this issue, by developing a Taylor series expansion of the future reward function in a two-period dynamic program. With this, they study the above trade-off and gain important insights that enabled them to construct an effective pricing strategy. Nevertheless, in their model inventory is not limited, a property that facilitates a relatively elegant treatment of the trade-off. When inventory is limited, learning might not only come at the expense of short-term gains, but also sacrifice future gains. Related research suggests that, at times, learning does not have to be actively sought after – in other words, decisions that ignore the need to learn can be very effective as long as one periodically updates the belief about the environment on the basis of most recent decision and outcomes. We refer the reader to Hu, Lovejoy and Shafer (1996), Treharne and Sox (2002), Lariviere and Porteus (1999), and Aviv and Pazgal (2002), for a discussion of this subject.

When referring to *passive learning*, one needs to be clear about the reference price setting process. Under passive learning strategies, the model used to determine the *current* action at each decision point, assumes that learning is independent of this choice. The term "passive" reflects the fact that updates to the belief vectors are made prior to each decision point, based on the history of actions and observations. Certainty-equivalent policies (see, e.g., Bertsekas 1995 for an extensive discussion of this subject) and OFLC heuristics are examples of passive learning processes. In fact, our FI heuristic falls under this category too, and we use it as a passive-learning reference process.

Indeed, the dynamic program $V_n^{FI}(q_n, \pi_n)$ solved under the FI heuristic, implicitly assumes that prices do not impact learning.

We use the title "active learning" to describe strategies in which the decisions are made through models that consider the impact of actions on the evolution of the belief vector (i.e., learning), even if this relationship is described in an imprecise way. "Limited look ahead" policies (e.g., Carvalho and Puterman 2003), as well as our general ISM heuristic (i.e., excluding special cases such as FI) fall under this category. Recall that in the dynamic program $V_n^{ISM}(q_n, \pi_n)$, prices are expected to impact learning through the process X, yet X is only hypothetical. Of course, a suboptimal active-learning strategy need not be better than a given suboptimal passive-learning strategy. In fact, it is possible that certain active-learning strategies do poorly in learning, compared to other given passive-learning strategies. The optimal policy for the original problem (2) is one that performs active learning in the best possible way.

In order to evaluate the appropriateness of our FI passive learning policy, one can consider the the gap associated with the upper bound $V_1^{ISM}(q_1, \pi_1)$ and the value of $\bar{H}_1^{FI}(q_1, \pi_1)$. Let's call this measure $\eta^{ISM/FI}$, and note that $\eta^{ISM/FI} \leq \eta^{FI}$. If the gap $\eta^{ISM/FI}$ is tight, but the gap η^{FI} is not, it would suggest that while learning is important, it is quite optimal to conduct it passively, using the FI method. If η^{ISM} is tight, but $\eta^{ISM/FI}$ is not, it would suggest that passive learning (through the FI heuristic) is insufficient, and active learning (through the ISM heuristic) is one way to perform well. We continue our discussion of this important matter after reporting on our numerical study in §5.1.

5 Performance Assessment

In this section we report on a numerical study designed to gauge the tightness of the upper bound and the quality of the heuristic policy derived from the approximate ISM-based model. We constructed a set of instances, each specified by a particular combination of system parameters

$$\left(N, q, K, A, \{\lambda_k, \alpha_k\}_{k\in\Omega}, \pi_1 \in \Pi, \{\tau_n\}_{n=1,\ldots,N}\right).$$

For each instance, we solved the problem $V_1^{ISM}(q_1, \pi_1)$, and calculated the performance estimate $\bar{H}_1^{ISM}(q_1, \pi_1)$ for our heuristic. While the computation of the upper bounds V_1^{ISM} is very quick, the simulation procedure is generally time-consuming. Therefore, we confined our analysis to a reasonably large set of $1,152$ instances, constructed by the following combinations:

$$N \in \{5, 10\}, \quad q \in \{1, 5, 10\}, \quad K \in \{2, 4, 8\}$$

and $\tau_1 = \ldots = \tau_N = 1$. (We also studied an additional sample of instances with larger values of N and q. The results were consistent with the qualitative insights we report below).

For the dynamics of the core process, we examined 4 types of matrices: (i) "static"; (ii) "absorbing"; (iii) "uniform"; and (iv) "cyclic". These matrices are given below in a respective order:

$$A \in \left\{ \mathrm{I}_{K\times K}, \begin{pmatrix} 1-\xi & \xi & 0 & \cdots & 0 \\ 0 & 1-\xi & \xi & \ddots & \vdots \\ \vdots & \ddots & 1-\xi & \xi & 0 \\ \vdots & \vdots & 0 & 1-\xi & \xi \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{K} & \cdots & \frac{1}{K} \\ \vdots & \ddots & \vdots \\ \frac{1}{K} & \cdots & \frac{1}{K} \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \vdots & \ddots & 1 \\ 1 & 0 & \cdots & 0 \end{pmatrix} \right\}$$

For the absorbing matrix, we considered the case of $\xi \in \{0.2, 0.5\}$. The uniform and cyclic matrices are used for the purpose of examining our bounds and heuristics. One can think of the uniform case as stationary but highly unpredictable demand environment. Note that in this case, learning is practically impossible from sales observations, and so even the FI model is equivalent to the original dynamic program. Under the cyclic matrix, the core demand state transitions between demand states in fixed cycles of K transitions. Like in the static environment, the seller may not necessarily know the actual state of demand during each period. In stark contrast with the uniformly random demand fluctuations, the static and cyclic environments represent cases in which knowledge of the current state of the core demand process reveals the true state of the core process for each period during the remainder of the planning horizon. For this reason, it is possible that the ISM model (10) and the heuristic policy driven by it might be bad if they reveal too much of the hidden information about the core process. By selecting these matrices, we hope to provide the reader with a conservative assessment of our upper bound and heuristic in stable (static) and unstable (cyclic) demand environments. Altogether, 5 matrices were examined in the numerical study. The value of the initial belief vector was set depending on the matrix type. We used a uniform belief $\pi_1 = \left(\frac{1}{K}, \ldots, \frac{1}{K}\right)$ for the static, uniform, and cyclic dynamics. For the absorbing state dynamics we used a perfect initial knowledge $\pi_1 = (1, 0, \ldots, 0)$.

For each instance we studied, the values of the demand parameters λ and α averaged across Ω at 2.5 and 1, respectively. We set the individual state-dependent values of these parameters by controlling three aspects. First, is the *spread*, or the level of variability of each specific parameter across the K demand scenarios. For instance, if the reservation price parameter α has a spread denoted by γ, its minimal and maximal values were set to $2/(1+\gamma)$ and $2\gamma/(1+\gamma)$, respectively. The other $K-2$ values were evenly positioned in between these two extreme values. This way

the average value of α across Ω was kept at the value 1. For the values of λ the numbers were set evenly between $5/(1+\gamma)$ and $5\gamma/(1+\gamma)$, so that they average at 2.5. We considered $\gamma \in \{2, 4\}$; i.e., a large spread in which the highest level of a parameter is 4 times larger than its lowest level, and a more modest spread in which the maximum-to-minimum ratio is 2.

The second aspect we controlled, is which parameter to vary across $k \in \Omega$. Four cases were considered: (i) only λ, with α kept constant; (ii) only α, with λ kept constant; (iii) α and λ varying with positive correlation (i.e., the largest value of α is matched with the largest value of λ, etc.); and (iv) α and λ varying with a negative correlation. The third aspect controlled was the order of the λ values – i.e., increasing or decreasing in $k \in \Omega$. Clearly, this issue is relevant only to the cases of absorbing dynamics and cyclic dynamics. For instance, a decreasing sequence of $\{\lambda_k\}$ in the absorbing dynamics case with λ's and α's positively correlated, represents a declining demand environment.

5.1 Results and Discussion

Our findings are summarized below with respect to the gap η^{ISM}. Among the 1,152 instances studied, the average gap was about 0.65%, with about 80% of the instances falling below 1%, and about 95% falling below 3%. We note that the standard error in calculating the estimates $\bar{H}_1^{ISM}$ generally fell at about 0.4% of V_1^{ISM}. The results demonstrated a consistent pattern, as follows. The largest gaps were found to be strongly dependent on the state dynamics, as reported in Table 1. The static and cyclic demand environments have the highest gaps, suggesting that settings with strong inter-temporal correlations in the core state process might not be satisfactorily addressed by the ISM approach.

Table 1: Optimality gap η^{ISM} for different types of state dynamics (the numbers in brackets represent the 0.90 and 0.95 percentiles). For the uniform case, there were minor gaps, but as expected these were only due to statistical errors in the simulation.

Static	Absorbing ($\xi = 0.5$)	Absorbing ($\xi = 0.2$)	Uniform	Cyclic
1.44%	0.3%	0.5%	0%	1.09%
(3.89%, 4.55%)	(0.89%, 1.1%)	(1.39%, 1.9%)	(0%, 0%)	(3.64%, 4.29%)

As anticipated, the gap tends to increase with the level of spread in the possible scenario values $\{\alpha_k : k \in \Omega\}$ and $\{\lambda_k : k \in \Omega\}$. For the smaller spread level of $\gamma = 2$, our method works impressively well, with a 0.99 percentile of the gap values at about 2%, and 0.95 percentile at about 1.1%. Interestingly, the gap is influenced primarily by the diversification in the values of α, whereas its dependency on the spread of λ is not significant. Among the quarter of our instances,

in which only the value of λ was varied across different core states, the average gap is 0.17%, and does not exceeds 1.6%. Among the cases in which we varied both α and λ across Ω, the correlation between these values did not have any substantial impact on the gap.

The gap tends to be slightly higher for the medium number of core states (i.e., $K = 4$). This relationship is not exact, but only represents a loose correlation pattern. Nevertheless, the results suggest that the dependency of the gap on K is not significant in general. Among all instances with a gap of 2.5% or higher, the average gap values are 3.83%, 3.98%, and 3.58% for $K = 2, 4,$ and 8, respectively. Among all $1,152$ instances, the corresponding averages are 0.54%, 0.77%, and 0.65%.

On average, the gaps are higher for the largest levels of initial inventory (q), achieving the values 0.37%, 0.8%, and 0.79% for $q = 1, 5,$ and 10, respectively. Nevertheless, among the worst-case scenarios (gap exceeding 2.5%), the average gaps are 4.39%, 3.52%, and 4.06%. The dependency on the level of initial inventory deserves a more detailed attention. As presented in §4.2.2, we solved the ISM-based model by selecting the g-values through a line search over a particular type of a functional form $b_j\left(0|n=1,\bar{p}\right)^{\beta}$. This choice of g appears to work quite well for the case $q = 1$, perhaps because the event of zero demand is the only case that has to be considered for the assessment of expected future revenues. When q is larger, it may be more appropriate to examine functional forms that take into account all the probabilities $\{b_j\left(d|n=1,\bar{p}\right) : d < q\}$. In order to gain a clearer perspective on the dependency on q, we examined all instances under the case $g \equiv 1$. The results are as follows: The average gaps are 0.91%, 1.02%, and 0.88% (statistically indistinguishable) for $q = 1, 5,$ and 10, respectively. Among the scenarios with a gap exceeding 2.5%, the averages are 4.85%, 4.17%, and 4.35% (again, statistically indistinguishable).

The number of periods (N) seems to be positively correlated with the level of the gap, but the gap values tend to grow at a diminishing rate. Figure 2 illustrates this pattern for the combination ($q = 5, K = 4, A =$ "static", $\gamma = 4$), which is among the settings with the poorest gap performance.

In order to examine the passive-learning FI heuristic, as well as the tightness of the upper bound $V_1^{FI}\left(q_1, \pi_1\right)$, we solved model (3), and simulated the FI heuristic for all $1,152$ instances. We initially calculated the gap measure η^{FI}, and reported its performance on Table 2. Comparing the corresponding values in the two tables, we note that the gap was significantly worse than η^{ISM} for setting with high inter-temporal correlation in the core state process.

We next examined the effectiveness of the FI heuristic, by measuring the gap $\eta^{ISM/FI}$ ($< \eta^{FI}$; see §4.5 above). We expected that the value $\eta^{ISM/FI}$ will be somewhere in the middle of the

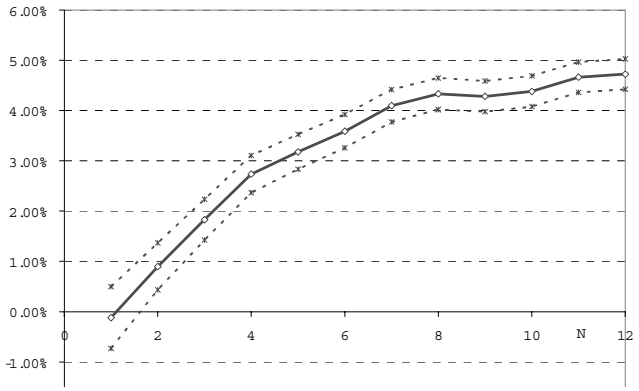

Figure 2: Gap performance as a function of the number of decision periods N. The solid line provides the gap estimate based on the simulated average revenue ($\bar{H}_1^{ISM}$). The dotted lines are based on the extreme values of the 95% confidence interval of H_1^{ISM}.

Table 2: The gap η^{FI} for different types of state dynamics (number in brackets represent the 0.90 and 0.95 percentiles).

Static	Absorbing ($\xi = 0.5$)	Absorbing ($\xi = 0.2$)	Uniform	Cyclic
4.87%	0.63%	1.23%	0%	2.45%
(12.22%, 16.11%)	(1.90%, 2.75%)	(4.01%, 5.58%)	(0%, 0%)	(5.83%, 7.84%)

range $\left[0, \eta^{FI}\right]$ for most instances. In other words, in the same way the upper bound $V_1^{FI}(q_1, \pi_1)$ is poor compared to our ISM bound $V_1^{ISM}(q_1, \pi_1)$, the FI heuristic should be poorer than our ISM heuristic. The results, however, were quite surprising. We found that the performance under the FI-based heuristic was very similar to our ISM-based heuristic, yielding approximately the same average gap across all $1,152$ instances, and having exactly the same 0.8 and 0.95 percentiles of 1% gap and 3% gap, respectively (i.e., $\eta^{ISM/FI} \approx \eta^{ISM}$). Among the instances with the poorest gap performance, one heuristic was at times a bit better than the other, but generally, we do not identify a pattern of significant advantage for the FI heuristic over the ISM heuristic, and vice versa.

Why does the passive-learning FI model works well and above our expectations? We believe that at the first part of a *long* season (relative to the number of units on hand), optimal pricing should be more aggressive in seeking short-term gains. Most likely, information about the core state would not accrue rapidly, but the season is assumed to be long enough to compensate for such shortcoming. Since high prices will most likely lead to zero demand, learning will be similar under the optimal policy, the ISM heuristic, and the FI heuristic. Specifically, when demand is zero,

no significant change to the current belief will be made, and if demand is larger than zero, a shift of belief to the most optimistic scenario (in terms of α-value) would happen. Figure 3 illustrates this rationale.

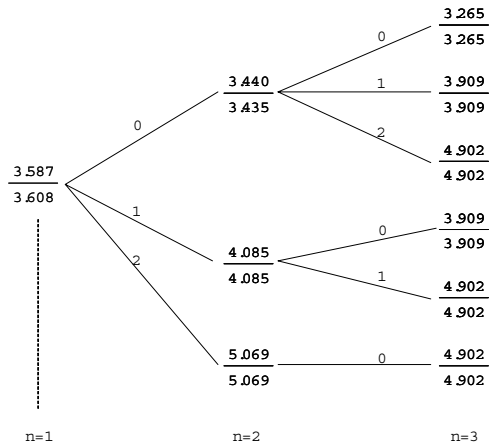

Figure 3: Price evolution over the first three periods, for the setting: $N = 10$, $q = 3$, $A =$"static", $\gamma = 4$ (only $\{\alpha_k\}$ varying across Ω). The numbers on top and bottom of a line represent those prescribed under the FI and ISM heuristics, respectively.

When the number of periods remaining to the end of the season is relatively small compared to the number of items in inventory, the seller becomes concerned about missing the opportunity to sell goods, if he bets on high prices. This concern will motivate the seller to set the prices at a medium level. We now propose an explanation similar in spirit to that provided by Lovejoy (1993) and Hu et al. (1996): When the spread among the core state parameters is large enough, medium-level prices will most likely lead to different demand levels under each core state, and hence information can be expected to accumulate quickly. If the spread is small, then information may not accrue rapidly, but the lack of information is less costly. Hence, the fact that the FI and ISM heuristics do not completely incorporate the actual learning evolution process, is not bothersome. While the ISM and FI heuristics appear to set different prices, overall, it seems to be (based on our results) that in each case the prescribed prices maintain a good balance between current vs. future revenues; i.e., in a way that is consistent with the overall goal of expected revenue optimization. For those cases that exhibit potentially unsatisfactory gap levels (say, above 3%), we believe that technical improvements within the ISM framework can yield additional reductions in the gap η^{ISM}. In fact, for a set of individual scenarios, we identified some ad-hoc structures of the g-functions, that enabled us to reduce the gap by 50% (e.g., from about 4% to 2%). Yet, the heuristics based on these

ad-hoc upper bound improvements did not yield increases in the expected revenue performance, as assessed by simulations studies with a very large number of runs per instance (400,000). This strengthens our feeling that both the FI- and the ISM-based heuristics work quite well across virtually all cases, with gaps that are possibly lower than those reported on Table 1 above.

We believe that the effectiveness of the FI and ISM policies are coming from the fact that both of them model *a* trade-off between current and future sales over the *entire* remaining horizon. This trade-off is not precise, in the sense that information is not modeled in an exact way, but as we explained above, this does not seem to be a cause of substantial sub-optimality. To further strengthen this insight, we used the simulation to examine the effectiveness of the OFLC heuristic described in §4.4. Our results demonstrate a comparable performance to our ISM and FI heuristics. Hence, a passive-learning policy happens to do well, regardless of whether the decision-maker is optimistic about learning (FI heuristic), or pessimistic (OFLC heuristic). What happens when a policy doesn't consider the whole horizon? To answer this question, we considered a one-step look-ahead OFLC policy that considers only the current and next periods in the dynamic program (12). The results confirmed our intuition, demonstrating a significant sub-performance when q is small and reasonable performance when q is large; see Table 3 below. We contend that when inventory is large enough, one may expect limited look-ahead policies to work well. But when inventory is scarce, such policies may suffer from significant subperformance, as these heuristics might not identify the potential benefits of gambling on high prices. This lesson sharpens and extends the observations in Carvalho and Puterman (2003).

Table 3: The gap measure based on the upper bound V^{ISM} and the estimated revenues under a one-step look-ahead policy.

q	Static	Absorbing ($\xi = 0.5$)	Absorbing ($\xi = 0.2$)	Uniform	Cyclic
1	28.6%	20.1%	21.1%	15.8%	17.9%
5	15.8%	10%	11.2%	7.8%	11%
10	5.7%	2.1%	3%	1.1%	4.2%

6 Analyses and Managerial Insights

In this section, we utilize our approximation technique and the ISM heuristic to gain several important managerial insights. Out of a variety of potential issues that can be handled by our models, we selected a set of relevant questions that do not lend themselves to immediate answers.

6.1 Pricing and the Length of the Sales Season

Consider the price set at the beginning of a given period n_0, $p_{n_0}^{ISM}(q_{n_0}, \pi_{n_0})$. The literature on dynamic pricing (Gallego and van Ryzin 1994) suggests that in the stationary case of a fixed and known demand function (i.e., $K = 1$), there is a clear relationship between the time left to the end of the sales season and the optimal price. As long as the number of units of the product is unchanged, prices continuously decrease over time. In other words, the further we are from the end of the season, the more affordable it is for the seller to bet on short-term gains by setting high prices. In the case of hidden state information, we need to consider the evolution of beliefs as well. For instance, Aviv and Pazgal (2002) discuss a continuous-time model in which the parameters of the demand process are constant over the sales season, but the intensity parameter (λ) is unknown to the seller. In such case, it is conceivable that the seller may want to start with low prices in order to gain better clarity about the demand state, rather than focus on short-term revenue maximization. Nevertheless, their paper shows that inspite of the fact that learning is better[4] with lower prices, optimal prices always decrease during the season, with the exception of upwards jumps when purchases occur. In our model, the price path may fluctuate even during times in which no sales occur. Yet, it is not necessarily right to assign this possible phenomenon to proactive learning. For instance, consider a cyclic demand environment in which there is a complete knowledge of the core demand state at each period. Although learning is irrelevant, prices may still fluctuate during the sales horizon even when no purchases occur.

We examine two illustrations relevant to the dependency between prices and the length of the horiozn. Both illustrations are based on the combination $q = 10$, $K = 2$, $A =$"static", $\gamma = 5$ (with both α and λ varying across the two states), and $\pi_1 = (1/2, 1/2)$. Figure 4 depicts the *price path* during a horizon of $N = 5$ periods, under the ISM and FI heuristics for a sample path in which no sales are made during the entire season. (In both cases, the heuristics yield the same expected revenues.). For each heuristic we show two paths: One in which α and λ are positively correlated ("cor+"), and one in which they are negatively correlated ("cor-").The figure shows a case in which the ISM method behaves more erraticly than the FI heuristic, when α and λ are negatively correlated. The reason for the fluctuations in the case of the negative correlation has to do mainly with the learning process. Note that when no sales are observed, it is hard to distinguish whether it is due to low average reservation prices or low store traffic intensity. If a relatively

[4]The authors define an increase in learning as a reduction in the coefficient of variation associated with the statistical distribution of the "hidden" parameter λ.

high price is set and no demand occurs, the seller's belief would shift to the low α scenario, and a smaller price would probably be set for the following period. Then, if no sale is observed, the belief would shift to the low λ scenario, pressuring for a price increase in the next period, and so forth. Intuition suggests that pricing strategies that are more aggressive in nature (price too low in order to learn, or too high in order to make short-term gains) may be susceptible to an undesirable, nervous pricing bahavior.

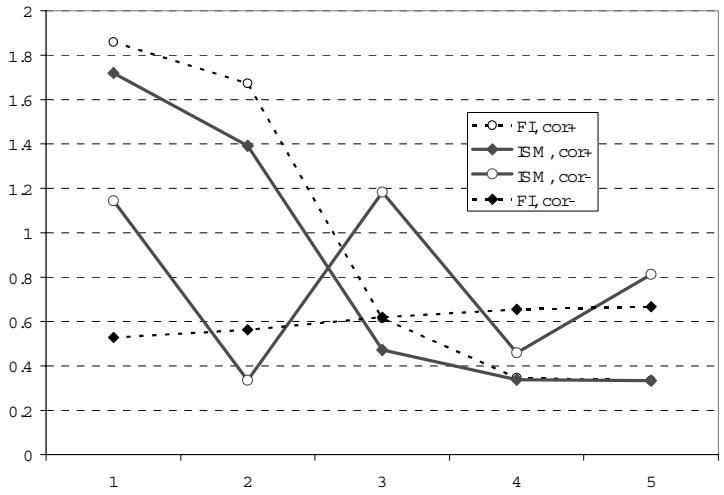

Figure 4: Price path under the ISM and FI heuristics. The evolution of prices is described for a sample path with no purchases.

Our second illustration, presented in Figure 5, depicts the *first period* prices as a function of the length of the horizon (note that this is different than a price path). Here, only the case of a negative correlation between the α and λ values is considered, in order to dramatize our message below. As expected, initial prices in the FI model are growing in N. Since model (3) ignores the aspect of active learning, the longer the season, the greater is the motivation to bet on high prices. In contrast, we see that the ISM actively pursues learning. For example, when having two periods, the seller would post a lower price than in the case of a single-period horizon, apparently in order to learn about the state of the core process. If, however, the horizon is of length 5 and above, the initial price would turn to be primarily motivated by short-term gains.

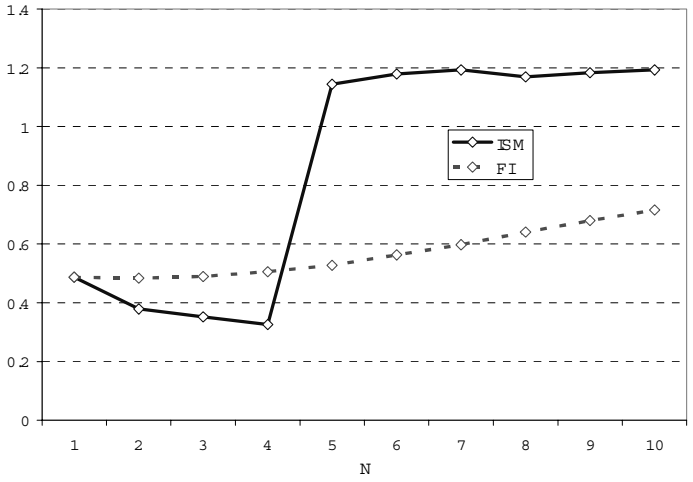

Figure 5: First period prices set under the heuristic policies FI and ISM.

6.2 Variability of the Demand Scenarios

We explored the behavior of the start-of-the-season price as a function of the dispersion (γ) among the parameters α and λ, characterizing the different demand scenarios. We found that the prices are relatively constant when there is a complete knowledge of the parameter α, but the $\{\lambda_k\}$ values vary. In such cases, a seller can comfortably assume, without a significance loss of expected revenues, that the actual value of λ in current period is equal to the *current* point estimate of its value. Yet, we emphasize that this does not mean that the seller should ignore the uncertainty about the parameter λ throughout the remainder of the season. Instead, the seller has to keep updating his belief about λ at the end of each period. Take for example the implication of this observation in cases where the value of q is large enough so that inventory is almost never depleted. Since the seller need not be concerned of selling items too quickly without exploiting the market potential, and since we observed that the seller need not worry much about the way that current prices impact learning, it would be close-to-optimal to set prices in a *myopic* way, by maximizing the current period's expected profit. Given the specific parameters examined above, the price that maximizes the first period's expected revenue $p \sum_{k\in\Omega} \pi_{1,k}\lambda_k e^{-p/\alpha}$, is $p = \alpha$.

The level of dispersion of the price reservation parameters $\{\alpha_k\}$ may have a significant impact on prices. For instance, consider the base combination $N = 2$, $K = 2$, $A =$"static", and a uniform initial belief. Figure 6 shows the initial prices for the two-period pricing problem, as a function of the spread level γ. The prices prescribed by the FI heuristic in this case are identical to the

prices that would be prescribed if $N = 1$, under both FI and ISM heuristics. In other words, none of the heuristics seems to be aggressive in short-term gain maximization. We observe that while uncertainty in the parameter λ only, has no impact on pricing decisions, prices are sensitive to the level of uncertainty about α. We suspect that the fact that the ISM prices are below the FI prices, is an indication that ISM conducts active learning.

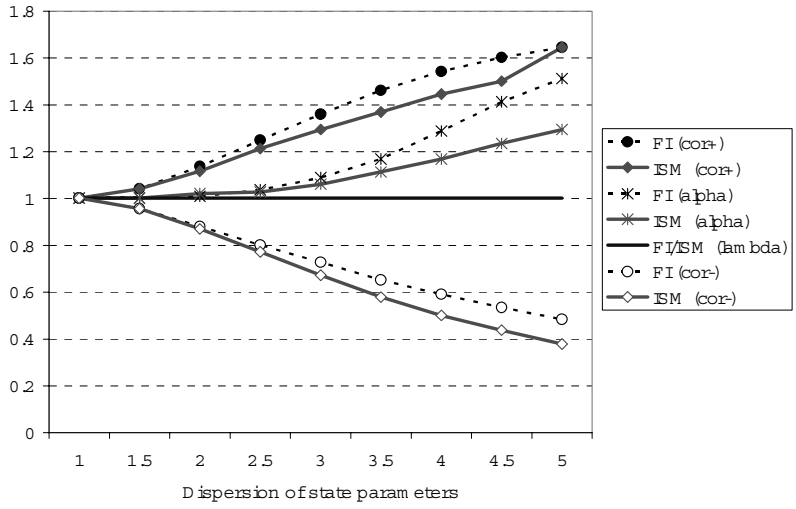

Figure 6: First-period prices under FI and ISM heuristics, as a function of the spread level γ. In brackets we specify which parameters vary. When both α and λ are varying, the correlation is noted.

6.3 Timing and Frequency of Price Changes

When price changes are costly, managers need to carefully assess their values. Our models can be utilized to study, for instance, the timing and frequency of price changes. Suppose that during a horizon of length T, there is only a single opportunity to change the price, and the timing of the change (say, τ) has to be scheduled upfront. What sorts of considerations affect the best choice of τ? To answer this question, we can study a two-period model with $\tau_1 = \tau$ as the length of the first period, and $\tau_2 = T - \tau_1$ as the length of the second period. Across a limited set of instances we examined, we found that the choice of τ depends on the level of initial inventory (q). This, as of itself, is not surprising. But we also found an additional interesting and noteworthy property (on the basis of our study of static demand patterns): If uncertainty is only about the store visit intensity λ, one does not need to worry about the dependency of the best τ on q. In fact, setting $\tau = T/2$ would be fine. But when uncertainty about the value of the average reservation prices (α)

exists, one needs to consider the timing more carefully. In the settings we studied, suboptimality due to the choice $\tau = T/2$ was about 3% in expected revenues, compared to about 0.25% when the level of α is known.

Consider now the *frequency* of the price changes, defined as the number of times a price can be set/reset during the season. Suppose that for a given frequency the pricing periods are set to be of equal length, and let the reference case represent a setting in which the price cannot be changed during the sales season. Based on a study of a limited set of instances, we observe monotonicity and concavity of the percentage increase in expected revenues, with respect to f. In other words, having more opportunities to change the price always brings benefits, however, at a diminishing rate. The expected gains due to increases in price change frequencies are felt more heavily in settings with uncertainty about α, than in settings with uncertainty about λ.

7 Conclusions

This paper discusses an important task of revenue management processes relevant to sellers of fashion-like products. We consider the pricing of goods in dynamic market environments characterized by high degree of demand uncertainties, and rapidly changing demand parameters (Song and Zipkin 1993, Johnson 2001). In such settings, it is often the case that sellers can learn about the statistical properties of the demand from actual sales figures (Fisher et al. 2000), but sales figures depend on prices. With the objective of maximizing expected revenues, the seller needs to post prices in a way that achieves an optimal balance between current and future gains. This is a complex task that has to incorporate the intricate influence of limited inventories and learning about demand, on optimal pricing. To study this problem, we develop a stylized partially observed Markov decision process (POMDP) framework. While the model provides an elegant description of the demand environment, it is prohibitively hard to solve optimally. To this end, we develop a model approximation, based on an information-structure modification (ISM) to the original problem. The optimal solution to the ISM model is relatively simple, and not less importantly, it serves as a rigorous and impressively tight upper bound on the optimal expected revenues. This bound enable us to conduct an effective numerical study in which we test four different heuristics (ISM, FI, OFLC, and one-step look ahead; see §5.1).

We found that our ISM heuristic exhibits an impressive performance in general. The sub-performance gap tends to be higher when dealing with scenarios characterized by *all* of the following

three conditions: (i) A relatively high inter-temporal correlation in the core demand state process; (ii) A high diversification of the average reservation prices across the core states Ω; and (iii) A long-enough planning horizon (N). We also investigated two types of passive-learning heuristics: The FI policy which is extremely optimistic in terms of the ability of the seller to learn about the demand, and the OFLC policy which is extremely pessimistic in that sense. Surprisingly, both heuristics performed comparably to the active-learning ISM heuristic, very close-to-optimal. This led us to make the conjecture that the most important trade-off in the pricing process has to do with the limited level of inventory. Pricing goods too high is a good betting strategy when the length of the season is long enough. In plain words, selling a product early at a low cost is a waste of limited capacity. What we assert is that learning should not interfere with this objective. To gain additional confidence for this argument, we tested a one-step look ahead policy, and found that when inventory is unlimited it works well (see also Carvalho and Puterman 2003). However, when inventory is limited, the subperformance of such policy can be as bad as 30%. Our paper also provides insights into the impact of the length of the season and demand uncertainty on pricing, as well as the timing and frequency of price-changes during a sales season.

As mentioned in the introduction, our work offers a contribution to the study of POMDP's in general. Researchers and practitioners have identified a wide variety of applications of POMDP's, ranging from inventory management, fishery and agriculture, maintenance, speach recognition, advertising, and bioinformatics. We hope that our ISM framework will offer a valuable input for continuing research in this area, and for implementation of decision-support tools in industrial applications.

Acknowledgement. We would like to thank Bill Lovejoy, the associate editor and two anonymous referees for many helpful suggestions.

References

[1] Astrom, K. 1965. Optimal control of Markov processes with incomplete state information. JMAA 10, 174-205.

[2] Aviv, Y., A. Pazgal. 2002. Pricing of Short Life-Cycle Products through Active Learning. Working paper. Washington University in St. Louis.

[3] Balvers, R.J., T.F. Cosimano. 1990. Actively learning about demand and the dynamics of price adjustments. Econom. Journal 100 882-898.

[4] Bertsekas, D.P. 1995. Dynamic programming and optimal control. Athena Scientific. Belmont, MA.

[5] Bitran, R.G., S.V. Mondschein. 1997. Periodic pricing of seasonal products in retailing. Management Science 43 64-79.

[6] Braden, D.J., S.S. Oren. 1994. Nonlinear pricing to produce information. Marketing Science 13 (3) 310-326.

[7] Carvalho, A.X., M.L. Puterman. 2003. Dynamic pricing and learning over short time horizons. Working paper. University of British Columbia.

[8] Desiraju, R., S.M. Shugan. 1999. Strategic service pricing and yield management. Journal of Marketing 63 (1) 44-56.

[9] Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. Operations Research 47 454-475.

[10] Fisher, M.L., A. Raman, A. McClelland. 2000. Rocket science retailing is almost here – Are you ready. *Harvard Business Review* (July-August).

[11] Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. Management Science. 40 999-1020.

[12] Hammond, J.H., A. Raman. 1996. Sport Obermeyer, Ltd. (Case study 9-695-022). Harvard Business School.

[13] Hu, C., W.S. Lovejoy, S.L. Shafer. 1996. Comparison of some sub-optimal control policies in medical drug therapy. Operations Research 44 (5) 696-709.

[14] Johnson, M.E., 2001. Learning from toys: Lessons in managing supply chain risk from the toy industry. California Management Review.

[15] Lane, D.E. 1989. A partially observed model for decision making by fishermen. Operations Research 37 (2) 240–254.

[16] Lariviere, M.A., E.L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. Management Science 45 (3) 346-363.

[17] Li, L. 1988. A stochastic theory of the firm. Mathematics of Operations Research 13 447-466.

[18] Lovejoy, W.S. 1991. Computationally feasible bounds for partially observed Markov decision processes. Operations Research 39 (1) 162-175.

[19] Lovejoy, W.S. 1993. Suboptimal policies, with bounds, for parameter adaptive decision processes. Operations Research 41 (3) 583-599.

[20] Monahan, G.E. 1982. A survey of partially observable Markov decision processes: Theory, models, and algorithms. Management Science 28 (1) 1-16.

[21] Petruzzi, N.C., M. Dada. 2002. Dynamic pricing and inventory control with learning. Naval Research Logistics 49 303-325.

[22] Sethi, S. P., F. Cheng. 1997. Optimality of (s,S) policies in inventory models with Markovian demand. Operations Research 45 (6) 931-939.

[23] Smallwood, R.D., E.J. Sondik. 1973. The optimal control of pratially observable Markov processes over a finite horizon. Operations Research 21, 1071-1088.

[24] Sondik, E. 1971. The optimal control of partially observable Markov processes. Ph.D. dissertation. Department of Engineering-Economic Systems, Stanford University. Stanford, CA.

[25] Song, J., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. Operations Research 41 (2) 351-370.

[26] Treharne, J.T., C.R. Sox. 2002. Adaptive inventory control for nonstationary demand and partial information. Management Science 48 (5) 607-624.

[27] Van Hee, K.M. 1978. Bayesian control of Markov chains. MAthematical Centre Tract 95, Amsterdam.

[28] White, C.C. 1976. Applications of two inequality results for concave functions to a stochastic optimization problem. JMAA 55, 347-350.

[29] White, C.C., W.T. Scherer. 1989. Solution procedures for partially observed Markov decision processes. Operations Research 37 (5) 791-797.0

[30] Ziya, S., H. Ayhan, R.D. Foley. 2002. On assumptions ensuring a strictly unimodal revenue function. To appear in *Operations Research*.