OPERATIONS RESEARCH
Vol. 60, No. 6, November–December 2012, pp. 1537–1550
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.1120.1103

Blind Network Revenue Management

Omar Besbes, Assaf Zeevi
Graduate School of Business, Columbia University, New York, New York 10027
{ob2105@columbia.edu, assaf@gsb.columbia.edu}

We consider a general class of network revenue management problems, where mean demand at each point in time is determined by a vector of prices, and the objective is to dynamically adjust these prices so as to maximize expected revenues over a finite sales horizon. A salient feature of our problem is that the decision maker can only observe *realized demand* over time but does not know the underlying *demand function* that maps prices into instantaneous demand rate. We introduce a family of "blind" pricing policies that are designed to balance trade-offs between exploration (demand learning) and exploitation (pricing to optimize revenues). We derive bounds on the revenue loss incurred by said policies in comparison to the *optimal* dynamic pricing policy that *knows* the demand function a priori, and we prove that asymptotically, as the volume of sales increases, this gap shrinks to zero.

Subject classifications: revenue management; network; pricing; nonparametric estimation; minimax; learning; asymptotic optimality; curse of dimensionality.
Area of review: Stochastic Models.
History: Received December 2009; revisions received August 2010, August 2011, December 2011; accepted December 2011.

1. Introduction

1.1. Background and Overview of the Main Contributions

Background and Motivation. One of the central problems in revenue management is the so-called *tactical* pricing problem: given an initial inventory of products to be sold over a finite selling season, the objective is to devise a strategy that dynamically adjusts prices so as to maximize the expected total revenues (under the assumption that inventory levels cannot be changed after the commencement of the selling season). The recent book by Talluri and van Ryzin (2005) and survey papers by Elmaghraby and Keskinocak (2003) and Bitran and Caldentey (2003) describe numerous instances of this problem, ranging from fashion and retail, to air travel, hospitality, and leisure. In cases where there are several different product types and a set of "resources" (raw materials, flight legs or other primitive components) used to "assemble" them, the problem described above is often referred to as *network* revenue management.

Among the first papers to propose a general mathematical model for the network problem is that of Gallego and van Ryzin (1997). They formulated a finite horizon stochastic control problem where realized demand is given by a (multivariate) Poisson process whose instantaneous rate represents mean demand for each product type, and is controlled by a vector of prices chosen by the decision maker. The objective is to maximize the total expected revenues over the course of a selling season, subject to an initial inventory of primitive resources used to construct the finished products. In this setting the optimal dynamic pricing policy can be derived, at least in principle, by exploiting the Markovian structure and solving the associated Bellman equation. Roughly speaking, the resulting policy seeks to adjust prices at each point in time based on current inventory levels so as to maximize future expected profits.

The dynamic programming logic articulated above, and variants thereof, form the analytical backbone of most revenue management studies to date (cf. Talluri and van Ryzin 2005). The vast majority of these studies are predicated on the assumption of "full information," namely, that the *demand function*, i.e., the functional relationship that determines how price affects mean demand rate, is known to the decision maker at the start of the selling season. The only remaining source of uncertainty is the randomness of *realized demand*. Needless to say, this type of stipulation might be invalid in many practical settings where a priori information describing the demand function might be incomplete or lacking altogether.

The purpose of our work is to study the problem of dynamic pricing on a network à la Gallego and van Ryzin (1997), albeit in a setting where the demand function is unknown and little is assumed insofar as its properties (in particular, it need not admit a parametric representation). For this reason we refer to the class of problems studied in this paper as "blind" network revenue management, indicating the opaque nature of information available to the decision maker. The performance of any blind pricing policy will be measured relative to that of the *optimal* dynamic pricing policy that *knows* the demand function a priori. We will focus on the worst-case revenue loss over a broad class of admissible demand functions, hence ensuring the robustness of pricing policies with respect to model uncertainty.

The approach we develop for solving this class of dynamic optimization problems under model uncertainty differs significantly from almost all antecedent literature in the area of revenue management. In particular, it is not guided by dynamic programming principles. Rather, we pursue a blend of ideas from nonparametric estimation and large-scale system analysis, and the proposed pricing policies are designed to balance *exploration* (demand learning) and *exploitation* (revenue optimization) trade-offs intrinsic to the class of problems described above.

Before we explain the main contributions of our paper, it is worthwhile distinguishing between two important settings associated with the aforementioned dynamic optimization problems: one where the *action set* (feasible price set) is discrete and finite, and one where it is infinite. The former is certainly more closely related to the practice of revenue management, where pricing decisions are often constrained to a fixed and given set of values; some examples and further discussion can be found in Talluri and van Ryzin (2005, Section 5.2.1.3). In addition, because this paper will flesh out in some level of detail, this setting enjoys important theoretical properties that render it attractive from practical considerations as well. We will therefore commence our discussion and focus a significant portion of the analysis on the finite action set problem, presenting subsequently extensions to the infinite action set formulation where there is a continuum of feasible prices. This will allow us to fully elucidate the escalation in complexity in passing from the former to the latter.

Overview of the Main Contributions and Qualitative Insights. This paper advocates a simple approach to the design of blind pricing policies that hinges on a separation of estimation (demand learning) and control (pricing). In the setting where the set of feasible prices is finite and fixed in advance, we develop a simple linear programming-based policy that uses an initial learning phase to estimate demand at each price point and then determines the proportion of time each price should be used downstream to (approximately) maximize expected revenues; see Algorithm 1. We then study the revenue loss incurred by this policy, relative to the full information benchmark. As the market size (volume of sales) grows large, we establish that this relative loss in revenues shrinks to zero uniformly over the class of admissible demand functions. That is, the revenues generated by the proposed policy are close asymptotically to the best achievable revenues under full information on the demand function. Theorem 1 spells this out rigorously and characterizes the *rate* at which the relative revenue loss diminishes. The important observation here is that this rate is *dimension-independent*, an attractive property in the context of blind network revenue management problems.

The main question then is whether the magnitude of relative revenue loss associated with our proposed blind pricing policy can be significantly reduced. To this end, we derive a general lower bound on achievable performance that shows that no policy can achieve a "substantially" better performance than our proposed algorithm in the sense of significantly improving the rate of convergence of the relative revenue loss to zero; this result is formalized in Theorem 2. In addition, we show that for a slightly more restricted class of problems there do not exist policies that achieve a better convergence rate than our proposed blind pricing algorithm; see Theorem 3. These results provide a characterization of the complexity of the pricing problem in the finite action set case.

We then move on to consider the more complicated case in which the action set is a continuum of prices from which one needs to construct the optimal pricing policy. We first develop a simple policy that tests a discrete subset of prices in the exploration (demand learning) phase, and then selects the "best" price to be used in the exploitation (revenue extraction) phase; see Algorithm 2. Our analysis establishes that the policy is asymptotically optimal, but at the same time, its performance degrades significantly with the number of products being sold as a consequence of the curse of dimensionality; see Theorem 4.

We then propose a modification of the policy mentioned above that uses the demand data obtained in the price testing phase to construct a nonparametric estimate of the *entire* demand function and revenue surface. This functional estimate is then fed into a deterministic optimization problem which gives rise to the ultimate pricing policy; see Algorithm 3. The policy described above exploits prior knowledge on the smoothness of the demand function to guide both data collection (price testing) and the nonparametric curve-fitting stages. Unlike most work in the full information setting, where smoothness typically is imposed as a purely technical condition, in our context smoothness plays a much more instrumental role: it communicates important information on the unknown demand function. Roughly speaking, the smoother the demand surface, the less one suffers from dimensionality effects; this is articulated in precise mathematical terms in Theorem 5 (see also the remark following the theorem).

On a technical level, our paper contributes to the theory of revenue management, and more broadly to dynamic optimization under model uncertainty, by characterizing and formalizing mathematically some of the complexities of the blind network pricing problem. The proofs of performance bounds for our proposed policies rely on a blend of ideas: analysis of a "deterministic skeleton" problem; large deviations results that quantify fluctuations of the stochastic system relative to its fluid-model counterpart; and nonparametric estimation techniques. In terms of lower bounds on the performance that can be achieved by any pricing policy, we introduce a proof technique which is based on information theoretic arguments that help identify and formalize the "worst-case" scenarios.

On a more practical level, we would like to note a connection between the learning phase in our pricing strategies and the widespread industry practice of "price testing." A recent empirical study of 32 large U.S. retailers finds that nearly 90% of them conduct such price experiments (see Gaur and Fisher 2005), and the advent of the Internet and the direct-to-customer model have served to greatly facilitate such price testing practices and their implementation (see, e.g., Williams and Partani 2006 for further discussion and examples). Given the central role of price testing practices, there is a growing need to better understand this approach and add to its rigorous foundations. (It is worth noting that the use of "testing" ideas is not limited to prices; see Fisher and Rajaram 2000 for a study involving merchandise testing.)

Our analysis attempts to shed some light on this issue by providing simple and intuitive guidelines for selecting both the *number of prices* that should be tested, as well as the overall fraction of the selling season that should be dedicated to *experimentation*, in addition to highlighting the underlying complexity of such problems. While said guidelines are established on the basis of a theoretical analysis, they hopefully provide a basis for future development of practical and implementable pricing policies.

1.2. Related Literature

Almost all work we are aware of that incorporates model uncertainty into the dynamic pricing problem described above, has effectively been restricted to the one-dimensional case (where there is only a single product being sold). The bulk of these studies focus on a *parametric* setting where the structure of the demand function is assumed to be known up to a finite number of unknown parameters. The method of choice in the analysis of such problems has been a Bayesian formulation of dynamic programming; see Lobo and Boyd (2003), Aviv and Pazgal (2005), Araman and Caldentey (2009), and Farias and Van Roy (2010), all of which restrict attention to one or two unknown parameters. Distinct from this stream of literature is the recent work of Besbes and Zeevi (2009) that proposes a "frequentist" approach to the problem, using maximum likelihood to infer the unknown parameters and policies that hinge on a separation of estimation and control.

For any parametric approach to work well, it is crucial that the structure assumed by the policy be consistent with that of the true underlying demand function. In other words, the postulated model needs to be *well specified* with respect to the actual mechanism that determines realized demand. To remove misspecification risk, one needs to step outside the boundaries of parametric modeling assumptions. For example, one can assume that the unknown demand function satisfies some mild *nonparametric* structural conditions (e.g., that it is monotone, bounded, differentiable, etc.). Very little work has been done to date in this direction. A few recent studies consider *static* settings, which do not involve dynamic decision making over time and trade-offs between learning and pricing; see, e.g., Rusmevichientong et al. (2006) and Eren and Maglaras (2010) (see also van Ryzin and McGill 2000, Ball and Queyranne 2009 in the context of capacity allocation problems). An exception is the work of Lim and Shanthikumar (2007) that formulates a robust max–min analogue of the dynamic pricing problem of Gallego and van Ryzin (1994); see also Lim et al. (2008) for an analysis of the multiproduct case. Their work is fairly conservative insofar as an adversary (nature) is allowed to alter the *distribution* of realized demand at each point in time to counter any chosen policy, and with the exception of exceedingly simple cases, the approach is not tractable and does not lead to prescriptive solutions.

The work that is perhaps most closely related to the current paper is that of Besbes and Zeevi (2009). In terms of the problem formulation and main thrust, that paper studies an analogue of the Gallego and van Ryzin (1994) *single product* pricing problem. It focuses on the impact of parametric versus nonparametric assumptions on the structure of the unknown demand function, in particular, and quantifies the economic value of such prior information via a minimax regret formulation. The pricing policies proposed in that case hinge crucially on the one-dimensional nature of the single product problem and consider only the case where there is a continuous set of feasible prices. The network problem, which serves as the focal point of the present paper, raises a different set of issues, for example, the distinction between finite and continuous action sets, and the curse of dimensionality. In terms of contributions to methodology, the present paper shares some common threads with Besbes and Zeevi (2009), most notably the core idea of separating estimation and control. However, at a finer grain level, the policies, as well as the proof techniques needed to study their performance, differ in a significant manner and involve new ideas that draw a much stronger connection between nonparametric statistics and dynamic optimization under uncertainty. Unlike Besbes and Zeevi (2009), the present paper does not address parametric modeling. Based on the theory developed there, it is possible to establish a significant performance improvement in the network setting if one is able to restrict attention to a parametric class of demand models.

The exploration–exploitation trade-off that characterizes the blind network revenue management problem relates also to the multiarmed bandit paradigm (see, e.g., Cesa-Bianchi and Lugosi 2006 for a recent and comprehensive survey). While there is a high-level connection with this stream of work, the presence of capacity constraints in conjunction with the multidimensional aspect of the problem does not allow to establish a direct connection with existing results.

1.3. The Remainder of the Paper

The next section introduces the model and formulates the problem. Section 3 analyzes the blind network problem where the feasible price set is discrete and finite. Section 4 shifts focus to the general blind network case. All proofs are collected in three appendices: Appendix A presents the proofs of the main results, Appendix B presents the proofs of the results in §4, and Appendix C details the proofs of auxiliary lemmas. Appendix D contains some additional numerical illustrations.

2. Problem Formulation

2.1. The Model

We consider a revenue management problem in which a firm sells d different products which are generated (assembled or produced) from ℓ resources. Let $A = [a_{ij}]$ denote the *capacity consumption* matrix, whose entries $a_{ij} \geqslant 0$, $i = 1, \ldots, \ell$ and $j = 1, \ldots, d$, denote the number of units of resource i required to generate product j. It is assumed that the entries of A are integer valued and each column contains at least one nonzero entry. The selling horizon is denoted by $T > 0$, and after this time sales are discontinued, and there is no salvage value for the remaining unsold products.

Demand for products at any time $t \in [0, T]$ is given by a multivariate Poisson process with intensity $\lambda_t = (\lambda_t^1, \ldots, \lambda_t^d)$ which measures the instantaneous demand rate (in units such as number of products requested per hour, say). This intensity is determined by the price vector at time t, $p(t) = (p^1(t), \ldots, p^d(t))$ through a *demand function* λ: $\mathscr{D}_p \to \mathbb{R}_+^d$, where $\mathscr{D}_p \subseteq \mathbb{R}_{+,+}^d$ (where $\mathbb{R}_{+,+} := (0, +\infty)$) denotes the set of feasible prices which is assumed to be compact.[1] Thus the instantaneous demand rate at time t is given by $\lambda_t = \lambda(p(t))$, and the realized demand is a controlled Poisson process. More will be said on the demand function shortly.

Let $(p(t)\colon 0 \leqslant t \leqslant T)$ denote the *price process* which is assumed to have sample paths that are right continuous with left limits taking values in $\mathscr{D}_p$. Let $(N^1(\cdot), \ldots, N^d(\cdot))$ be a vector of mutually independent unit rate Poisson processes. The cumulative demand for product j up until time t is then given by $N^j(\int_0^t \lambda^j(p(s))\, ds)$.

Information Structure and the Dynamic Optimization Problem. We assume that the decision maker *does not know* the true demand function and knows only that λ belongs to the class $\mathscr{L} := \mathscr{L}(M, m, p_\infty)$ of nonnegative functions, which for finite positive constants M, m and a vector $p_\infty \in \mathscr{D}_p$ satisfies the following:

(i) Boundedness of demand: for all $\lambda \in \mathscr{L}$, $\|\lambda(p)\| < M$ for all $p \in \mathscr{D}_p$.

(ii) Minimum revenue rate: for all $\lambda \in \mathscr{L}$, $\sup\{p \cdot \lambda(p)\colon p \in \mathscr{D}_p\} > m$.

(iii) "Shut-off" price: for all $\lambda \in \mathscr{L}$, $\lambda(p_\infty) = 0$.

Here for two vectors $y, z \in \mathbb{R}^d$, $y \cdot z$ denotes the usual scalar product and $\|y\| := \max\{|y^i|\colon i = 1, \ldots, d\}$. To avoid trivialities, M, m are assumed to be such that $\mathscr{L}$ is nonempty. It is worth noting that Assumptions (i) and (ii) are quite benign and hold for many demand models used in the revenue management literature such as linear, exponential, and iso-elastic (Pareto), as long as the parameters are assumed to lie in a compact set; see, e.g., Talluri and van Ryzin (2005, §7) for further examples. The existence of a "shut-off" price in Assumption (iii) is not restrictive from a practical standpoint because in most applications there exists a finite price that yields zero demand. From a modeling perspective, this is merely a convenient way to allow for a sales denial.

While the decision maker possesses only limited information on the demand function, s/he is able to continuously observe *realized demand* at all time instants starting at time 0 and up until the end of the selling horizon T. We shall use π to denote a *pricing policy*, and its associated price process will be denoted $(p(t)\colon 0 \leqslant t \leqslant T)$. With some abuse of terminology, we will use the term *policy* to refer to the price process itself, as well as the algorithm that generates it interchangeably. For $0 \leqslant t \leqslant T$ put

$$N^{j,\pi}(t) := N^j\left(\int_0^t \lambda^j(p(s))\, ds\right), \quad \text{for } j = 1, \ldots, d, \tag{1}$$

where $N^{j,\pi}(t)$ denotes the cumulative demand, i.e., number of units requested of product j up to time t under the policy π. Let $N^\pi(t)$ denote the vector $(N^{1,\pi}(t), \ldots, N^{d,\pi}(t))$.

Let $x = (x^1, x^2, \ldots, x^\ell)$ denote the inventory level of each resource at the start of the selling season. We assume without loss of generality that $x^i > 0$, $i = 1, \ldots, \ell$. A policy π is said to be *admissible* if the induced price process is non-anticipating and satisfies

$$\int_0^T A\, dN^\pi(s) \leqslant x \quad \text{a.s.}, \tag{2}$$

$$p(s) \in \mathscr{D}_p, \quad 0 \leqslant s \leqslant T, \tag{3}$$

where A is the capacity consumption matrix defined earlier and vector inequalities are assumed to hold componentwise. The term *nonanticipating* means that at any point in time, $p(t)$ can only depend on past realized demand $(N^\pi(s)\colon 0 \leqslant s < t)$ and prices $(p(s)\colon 0 \leqslant s < t)$. (That is, the price process is adapted to the filtration generated by past values of the demand and price processes.) It is important to note that while the decision maker does not know the demand function, knowledge of p_∞ guarantees that the constraint (2) can be met. We let $\mathscr{P}$ denote the set of admissible policies, and the performance of a policy $\pi \in \mathscr{P}$ is measured in terms of cumulative expected revenues,

$$J^\pi(x, T; \lambda) := \mathbb{E}\left[\int_0^T p(s) \cdot dN^\pi(s)\right]. \tag{4}$$

It is worth noting that the decision maker is not able to compute the expectation in (4) since the *true* demand function governing customer requests is not known a priori. This lends further meaning to the terminology "blind revenue management," where one is attempting to optimize (4) in a blind manner.

The Full Information Benchmark and Main Objective. When the demand function λ is known prior to the start of the selling season, the dynamic optimization problem described above can, at least in theory, be solved; this will be referred to as the "full information" setting. This problem is precisely the one formulated in Gallego and van Ryzin (1997), who also characterize the optimal state-dependent pricing policy using dynamic programming. Suppose that we fix a demand function $\lambda \in \mathscr{L}$. Let us define

$$J^*(x, T \mid \lambda) := \sup_{\pi \in \mathscr{P}} \mathbb{E}\left[\int_0^T p(s) \cdot dN^\pi(s)\right], \tag{5}$$

where the notation reflects the fact that the optimization problem is solved "conditioned" on knowing the demand function λ at time $t = 0$.

The value of the full information optimization problem (5) serves as an upper bound on the revenues that can be achieved by any admissible policy when the demand function is known, and a fortiori for any policy in the blind setting. That is, for any demand function $\lambda \in \mathscr{L}$, we have that $J^\pi(x, T; \lambda)/J^*(x, T \mid \lambda) \leqslant 1$ for all $\pi \in \mathscr{P}$. This ratio measures the performance of any admissible policy on *relative scale*; generated revenues are expressed as a fraction of the optimal revenues in the full information setting. Our objective is to design policies that maximize this ratio *uniformly* over all demand functions in the class $\mathscr{L}$; that is, choose $\pi \in \mathscr{P}$ to maximize

$$\inf_{\lambda \in \mathscr{L}} \frac{J^\pi(x, T; \lambda)}{J^*(x, T \mid \lambda)}. \tag{6}$$

The criterion in (6) can be viewed as the result of a two-step procedure: first the decision maker selects a policy $\pi \in \mathscr{P}$, and then "nature" picks the worst possible demand function $\lambda \in \mathscr{L}$ for this particular policy. Measuring performance in this manner guarantees that "good" policies will perform well *regardless* of the true underlying demand function. The fact that admissible policies can only learn the true demand function by observing realized demand over time introduces an obvious tension between exploration (estimation/demand learning) and exploitation (optimization/pricing), and balancing these contradicting objectives is one of the main issues that will be explored in what follows.

3. Main Results

As alluded to earlier, the simplest and probably most practically relevant instance of the blind network revenue management problem occurs when the set of feasible prices is discrete and finite, say, $\mathscr{D}_p = \{p_1, \ldots, p_k, p_\infty\}$. In more generic control terminology this describes a situation where the action set is finite. In this setting, uncertainty is essentially limited to the value of the demand function at the finite collection of prices in $\mathscr{D}_p \backslash \{p_\infty\}$.

3.1. The Proposed Pricing Policy

Our proposed blind pricing policy is based on two tuning parameters $\tau \in (0, T]$ and $\delta > 0$. The first one defines the length of a learning horizon while the second one is a "fudge factor" that corrects for the remaining capacity after the learning phase. During this *exploration period* the mean demand rate at the various prices is estimated. Subsequently, these demand estimates are used as inputs for an empirical, data-driven, linear optimization problem. The solution of the latter gives rise to the pricing policy that will be used in the *exploitation period* that takes place over the interval $[\tau, T]$. In other words, the policy thus constructed, fleshed out in the pseudo-code below, "separates" estimation and control phases.

Algorithm 1 $(\pi(\tau, \delta))$

```
Step 1. Initialization: Set the learning interval to be $[0, \tau]$
    and put $\Delta = \tau/k$.
Step 2. Learning/experimentation:
    (a) While inventory is positive for all resources, price
        at $p_i$ from $t_{i-1} = (i-1)\Delta$ to $t_i = i\Delta$, $i = 1, 2, \ldots, k$
        If some resource runs out of stock, apply $p_\infty$ up
        until time $T$ and STOP.
    (b) Compute
```

$$\hat{d}_i = \frac{\text{total demand over } [t_{i-1}, t_i]}{\Delta}, \quad i = 1, \ldots, k. \tag{7}$$

```
Step 3. Optimization/exploitation: Let $\hat{t} = (\hat{t}_1, \ldots, \hat{t}_k)$ be
    the solution of the linear program
```

$$\max\left\{\sum_{i=1}^k p_i \cdot \hat{d}_i t_i \colon \sum_{i=1}^k A\hat{d}_i t_i \leqslant x - \delta e, \sum_{i=1}^k t_i \leqslant T - \tau, \\ t_i \geqslant 0,\ i = 1, \ldots, k\right\}. \tag{8}$$

```
    For $i = 1, \ldots, k$, apply $p_i$ for $\hat{t}_i$ time units on $(\tau, T]$
    until some resource is out of stock, then apply $p_\infty$ for
    the remaining time.
```

In Step 3, e denotes the vector of ones in $\mathbb{R}^d$ and A denotes the capacity-consumption matrix defined in the capacity constraint (2) of the original dynamic optimization problem (§2). In addition, it is clear that any practical implementation of the policy would not "shut off" *all* the demand once a single resource becomes unavailable but would rather do so only for those products that use the unavailable resource. The result we present in Theorem 1 is valid for policies that improve upon the above by refining Step 3 through partial and/or gradual demand "shut off."

Intuition. In Steps 1 and 2, the decision maker estimates the demand at each of the k feasible prices by testing the price on a period of time of length τ/k. To understand the logic underlying Step 3, imagine that the demand function $\lambda(\cdot)$ is revealed at the start of the selling season, and that demand is deterministic rather than governed by a Poisson process. The revenue maximization problem would then be given by the following deterministic dynamic optimization problem:

$$\max\left\{\sum_{i=1}^{k} p_i \cdot \lambda(p_i)t_i\colon \sum_{i=1}^{k} A\lambda(p_i)t_i \leqslant x,\ \sum_{i=1}^{k} t_i \leqslant T,\ t_i \geqslant 0,\ i=1,\ldots,k\right\}. \quad (9)$$

It is possible to show (see Gallego and van Ryzin 1997) that the solution of this linear program leads to near-optimal performance in the full information stochastic dynamic optimization problem. The objective of Step 3 is to get "close" to said solution, by solving a suitable *empirical version* of the deterministic problem (9). The optimal solution for this problem $\hat{t}$ is then used for the remainder of the time horizon $(\tau, T]$.

Balancing the Exploration–Exploitation Trade-Off. The choice of the key tuning parameter τ is meant to balance two contradicting objectives. As τ increases, so do the quality of the estimates of the demand function values and in turn the quality of the approximation to the deterministic skeleton problem (and its solution). However, an increase in τ also implies shorter exploitation time and higher potential revenue losses.

3.2. Theoretical Analysis

Exact analysis of the performance of the policy described in the previous section is quite difficult. We therefore introduce an asymptotic regime that facilitates an approximate analysis and that has been used in several revenue management studies to date (see, e.g., Talluri and van Ryzin 2005, §§3.6 and 5.3, and references therein). The regime is predicated on the number of initial resources and potential demand growing proportionally large. In particular, for any positive integer n the initial resource vector and the demand function are given by

$$x_n = nx, \quad \lambda_n(\cdot) = n\lambda(\cdot). \quad (10)$$

Here n, which serves as a proxy for the market size, determines both the order-of-magnitude of inventories and the rate of demand; when n is large this scaling characterizes a regime with a high volume of sales but maintains inventory constraints.[2] Such an asymptotic regime is appropriate to analyze problems where the potential demand over the sales horizon and the initial inventory are of the same order of magnitude.[3] The following notation will be useful: for real valued positive sequences $\{a_n\}$ and $\{b_n\}$ we write $a_n = O(b_n)$ if a_n/b_n is bounded from above for large enough values of n (i.e., $\limsup a_n/b_n < \infty$). If a_n/b_n is also eventually bounded away from zero (i.e., $\liminf a_n/b_n > 0$) then we write $a_n \asymp b_n$.

We will denote by $\mathscr{P}_n$ the set of admissible policies for a system with scale n, and the expected revenues under a policy $\pi_n \in \mathscr{P}_n$ will be denoted $J_n^{\pi}(x, T; \lambda)$. With some abuse of notation we will occasionally use π to denote a sequence $\{\pi_n\colon n = 1, 2, \ldots\}$ as well as any element of the sequence, omitting the subscript "n" to avoid cluttering the notation. For each $n = 1, 2, \ldots$, let $J_n^*(x, T \mid \lambda)$ denote the optimal revenues that can be achieved in the full information case, i.e., when the demand function is known a priori in a system of scale n. It follows from §2 that for all $n = 1, 2, \ldots$, we have that $J_n^{\pi}(x, T; \lambda) \leqslant J_n^*(x, T \mid \lambda)$. With this in mind, the following definition characterizes admissible policies that have "good" asymptotic properties.

Definition 1 (Asymptotic Optimality). A sequence of admissible policies $\{\pi_n\}$ is said to be asymptotically optimal if

$$\inf_{\lambda \in \mathscr{L}} \frac{J_n^{\pi}(x, T; \lambda)}{J_n^*(x, T \mid \lambda)} \to 1 \quad \text{as } n \to \infty. \quad (11)$$

Asymptotically optimal policies are those that achieve the full information upper bound on revenues as $n \to \infty$, *uniformly* over the class of admissible demand functions. For such policies it is also of interest to measure the (worst-case) magnitude of revenue loss incurred in comparison to the best achievable performance under full information. Normalized by the latter, this gives the performance loss on relative scale (say, in percentage terms), i.e., $\sup\{1 - J_n^{\pi}(x, T; \lambda)/J_n^*(x, T; \lambda)\colon \lambda \in \mathscr{L}\}$. The rate at which this shrinks to zero quantifies the rate of convergence in (11), a measure of the second-order behavior of asymptotically optimal policies.

Theorem 1. *For $\tau_n \asymp n^{-1/3}$ and $\delta_n = n(\log n)^{1/2} n^{-1/3}$, with C sufficiently large, the sequence of policies $\{\pi(\tau_n, \delta_n)\}$ defined by Algorithm* 1 *is asymptotically optimal. In particular,*

$$\sup_{\lambda \in \mathscr{L}} \left(1 - \frac{J_n^{\pi}(x, T; \lambda)}{J_n^*(x, T \mid \lambda)}\right) = O\left(\frac{(\log n)^{1/2}}{n^{1/3}}\right) \quad \textit{as } n \to \infty. \quad (12)$$

Note that learning, as measured by the length of the exploration phase τ_n, occurs on a shorter time scale than the sales horizon T. Also of significant importance is the observation that the rate of revenue loss *does not* depend on the number of products being sold, i.e., the rate of convergence above is dimension independent.

Proof Sketch. As alluded to in the discussion following Algorithm 1, there are two sources of error that impact the revenue loss relative to the maximal full information revenue benchmark, as captured by the ratio J_n^{π}/J_n^*. The gap in the performance of the prescriptions obtained from both solutions is driven by the quality of the approximation of $\lambda(p_i)$ and the time spent learning τ_n. In particular, we establish in the proof that the performance of the policy obtained from (8) will be $O(n\tau_n + n \max_{i=1,\ldots,k} |\hat{d}_i - \lambda(p_i)|)$ away from the value of (9), which serves as an upper bound to J_n^*. The first error source can be interpreted as an "*exploration bias*" that is due to experimenting with prices in the absence of information on the demand model. This results in potential revenue losses of order τ_n. The second source of error is *stochastic*, arising from the fact that only noisy observations of the demand function are available and that one uses the estimates $\hat{d}_i$ of demand to solve (8), instead of $\lambda(p_i)$ in (9). One then notes that $|\hat{d}_i - \lambda(p_i)|$ represents the deviations of the sample mean arrival rate (at a given price) from its actual mean. Since each price is held fixed for τ_n/k units of time and the demand intensity is of order n, one may use results on large deviations to establish that $|\hat{d}_i - \lambda(p_i)| \lesssim (n\tau_n/k)^{-1/2}$ with high probability, which yields that the stochastic error is of order $(n\tau_n/k)^{-1/2}$.

Hence the overall revenue loss is dictated by the sum of the two sources detailed above, namely

$$1 - J_n^\pi / J_n^* \approx C\left(\tau_n + \frac{k^{1/2}}{(n\tau_n)^{1/2}}\right). \tag{13}$$

This last expression captures mathematically the tension that must be resolved in choosing the tuning parameters associated with Algorithm 1. Roughly speaking, shortening τ_n decreases the exploration bias but increases the stochastic error because there is more "noise" at each tested price. Balancing the two error terms in (13) yields the choice of tuning parameter τ_n that minimizes the order of magnitude of the relative loss. This choice is the one reported in the theorem and gives rise to the revenue loss rate in (12). □

Remark. The proof of Theorem 1 provides an upper bound on the constant C that appears in (13), and a careful inspection of the proof reveals that the upper bound depends on the initial inventory and time horizon through $(\min\{1, \min_{i=1,\dots,\ell} x_i/T\})^{-1}$, and the number of prices, k, through $k^{4/3}$.[4] Hence, as the number of possible prices k increases, so does the constant C. The analysis we have presented focuses on cases where k is finite and small (relative to the scale of the system). On the other extreme, one has a continuum of prices; this setting is analyzed in §4. We note that the upper bound for C is fairly conservative. A possible way to evaluate the quality of the policy (and the constant C) would be to test the latter across a large range of possible demand functions and extract the corresponding "worst" constant; see, e.g., Besbes and Zeevi (2009, §6.3).

3.3. Fundamental Limits on Achievable Performance

As mentioned earlier, the rate of convergence in (11) measures the quality of asymptotically optimal policies. In what follows we will establish that the performance of the policy given by Algorithm 1, with τ_n specified as in Theorem 1, cannot be significantly improved upon in general and in certain settings is best possible.

3.3.1. A Lower Bound on Achievable Performance. For simplicity, we focus on a setting where $k = 2$, there is only a single product being sold, and there are no inventory constraints.[5]

Theorem 2. *Suppose that $\mathcal{D}_p = \{p_1, p_2, p_\infty\}$ with $p_1 < p_2$ and $p_1 M > m$, and that the inventory of primitive resources is infinite ($x = \infty$). Then, for some constant $C > 0$,*

$$\sup_{\lambda\in\mathscr{L}} \left(1 - \frac{J_n^\pi(\infty, T; \lambda)}{J_n^*(\infty, T \mid \lambda)}\right) \geqslant \frac{C}{n^{1/2}} \quad \textit{for all } n \geqslant 1, \tag{14}$$

for all admissible policies $\pi \in \mathscr{P}$.

The above result establishes a fundamental bound on the performance of *any* admissible policy: no policy can achieve a faster rate of convergence than $O(1/n^{1/2})$ over the entire class $\mathscr{L}$ when there are no inventory constraints.

Proof Sketch. The bound above is derived by restricting attention to two possible demand functions $\mu_1(\cdot)$ and $\mu_2(\cdot)$ in $\mathscr{L}$ that are "close." In particular, we focus on two demand functions that cross at p_1 and that are off by a factor of $1/n^{1/2}$ at p_2. The key underlying idea revolves around the tension that any admissible policy faces when nature is restricted to the two choices μ_1 and μ_2. On one hand, because the two demand functions coincide at p_1, one needs to price at p_2 to accumulate observations that would allow to distinguish if nature selected $\mu_1(\cdot)$ or $\mu_2(\cdot)$. At the same time, pricing at p_2 might lead to revenue losses if nature initially selected the demand function μ_1. To analyze this tension, we reduce the problem to a hypothesis test for determining if $\lambda = \mu_1$ or $\lambda = \mu_2$ and show, using information theoretic arguments, that essentially only two cases can occur. Roughly speaking, either a policy does not gather sufficient information to distinguish between the two hypotheses reliably, in which case it has a nonvanishing probability of selecting a suboptimal price, implying a worst-case revenue loss of $O(1/n^{1/2})$; or sufficient information is gathered to distinguish between the two hypotheses, but this comes at the "price" of a learning related relative revenue loss of $O(1/n^{1/2})$. It follows that for all policies $\pi \in \mathscr{P}$

$$\sup_{\lambda\in\{\mu_1(\cdot),\mu_2(\cdot)\}} \left(1 - \frac{J_n^\pi(x, T; \lambda)}{J_n^*(x, T \mid \lambda)}\right) \geqslant \frac{C}{n^{1/2}} \quad \text{for all } n \geqslant 1.$$

The result then follows. □

Discussion. The preceding discussion highlights an important feature of the policy given in Algorithm 1. While the latter is predicated on a simple separation of estimation (demand learning) and control (pricing), the rate of convergence that it achieves is "not far" from the best possible rate, which by Theorem 2 is known to be no better than order $1/n^{1/2}$. While this rate might not be achieved by said policy, its structure strikes a reasonable balance between complexity and performance. To that end, Theorem 3 below will establish that at least in the case where inventory constraints are absent, the proposed policy has performance that is best possible among all those policies that suitably constrain the number of price changes.

Remark (on the Alternative of Parametric Modeling). Careful inspection of the proof of Theorem 2 reveals that many commonly used parametric models contain demand functions that are "close" in the sense of the proof sketch following the theorem. For example, a class of linear demand models $\{\theta_1 - \theta_2 p\colon (\theta_1, \theta_2) \in [\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]\}$ such that $\underline{\theta}_1 < \bar{\theta}_1$ and $\underline{\theta}_2 < \bar{\theta}_2$ will contain the aforementioned two demand functions (for n sufficiently large). Hence, even under typical parametric modeling assumptions it is impossible to construct policies that can do better than the lower bound of order $1/n^{1/2}$ for all parameter instances, and in general performance could be much worse if the postulated model is misspecified, in which case the ratio J_n^π/J_n^* might not even converge to 1.

3.3.2. Optimality of the Proposed Policy in a Restricted Setting. We further investigate the performance of the policy presented in Algorithm 1 in settings where $k = 2$, there is a single product being sold, and there are no capacity constraints. The key observation here is that the proposed policy should select in Step 3 a single price that maximizes the estimated expected revenue rate. In other words, it need not make more than k price changes. The next result provides a lower bound on the performance of *any* admissible policy that is restricted to use *at most k* price changes.

Theorem 3. *Suppose $\mathscr{D}_p = \{p_1, p_2, p_\infty\}$ with $p_1 < p_2$ and $p_1 M > m$, and that the inventory of primitive resources is infinite ($x = \infty$). Let $\mathscr{P}' \subseteq \mathscr{P}$ denote the set of admissible policies that use at most two price changes throughout the horizon. Then, for some constant $C > 0$,*

$$\sup_{\lambda \in \mathscr{L}} \left(1 - \frac{J_n^\pi(\infty, T; \lambda)}{J_n^*(\infty, T \mid \lambda)}\right) \geqslant \frac{C}{n^{1/3}} \quad \text{for all } n \geqslant 1, \tag{15}$$

for all policies $\pi \in \mathscr{P}'$.

In the setting covered in Theorem 3, (15) establishes that the policy described in Algorithm 1, with τ specified as in Theorem 1, achieves the best possible rate of convergence among all policies in $\mathscr{P}'$ (up to a logarithmic factor).

The main intuition outlined in the proof sketch of Theorem 2 applies to Theorem 3 as well, with the key difference being that in the above it is necessary to account for the fact that the price can change only twice. This constraint implies a more subtle choice of "worst-case" demand functions and the need to track the times at which the price changes.

3.4. A Numerical Illustration

We consider an example with two products and three resources. The first, second, and third rows of the capacity consumption matrix A are given by $(1, 1)$, $(3, 1)$, and $(0, 5)$, respectively. This means that product 1 requires 1 unit of resource 1, 3 units of resource 2, and no units of resource 3, etc. We consider three different underlying demand models to test the efficacy of our proposed policy: a linear, an exponential, and a logit model.

(a) $\lambda(p_1, p_2) = (8 - 1.5p_1, 9 - 3p_2)'$,

(b) $\lambda(p_1, p_2) = (5\exp\{-0.5p_1\}, 9\exp\{-p_2\})'$,

(c) $\lambda(p_1, p_2) = 10(1 + \exp\{-p_1\} + \exp\{-p_2\})^{-1} \cdot (\exp\{-p_1\}, \exp\{-p_2\})'$.

It is important to emphasize that our policies are constructed in a blind manner, *without* knowledge of the demand function. The set of feasible prices is $\{(1, 1.5), (1, 2), (2, 3), (4, 4), (4, 6.5)\}$. In Table 1, we illustrate the performance of the policies defined by Algorithm 1 with $\tau_n = n^{-1/3}$ and $\delta_n = n(\log n)^{1/2}(n\tau_n/k)^{-1/2}$. Note that in assessing the performance ratio J_n^π/J_n^*, we use the upper bound provided by the deterministic relaxation in place of J_n^* (see Gallego and van Ryzin 1997), and hence the actual ratio J_n^π/J_n^* is *at least as high* as that reported in the table. The results are based on running 10^3 independent simulation replications from which the performance indicators were derived by averaging. The standard error was below 0.1% in all cases.

Table 1. Price restricted case.

Market "size"		$n = 10^2$	$n = 10^3$	$n = 10^4$
Tuning parameters		$\tau = 0.22$	$\tau = 0.10$	$\tau = 0.05$
		J_n^π/J_n^*	J_n^π/J_n^*	J_n^π/J_n^*
Initial	Linear	0.70	0.86	0.94
Inventory	Exponential	0.82	0.92	0.97
$x_n = (3, 5, 7) \times n$	Logit	0.85	0.93	0.98
Initial	Linear	0.77	0.83	0.92
Inventory	Exponential	0.87	0.94	0.98
$x_n = (15, 12, 30) \times n$	Logit	0.88	0.94	0.97

Notes. Results in the table give a lower bound on the performance of the policy $\pi(\tau)$ relative to the optimal performance in the full information case. Here $\tau =$ fraction of time allocated to learning.

The results in Table 1 are consistent with the asymptotic optimality statement of Theorem 1. The proposed policy generates at least 83% of the full information benchmark for $n = 10^3$, and this performance is achieved by allocating only 10% of the selling horizon to the learning phase. In addition, the performance is seen to be comparable for the various demand models tested, exhibiting the robustness asserted in Theorem 1.

4. Extensions

In this section, we deal with the case where the set $\mathscr{D}_p$ contains a continuum of feasible prices, and model uncertainty pertains to a surface $\lambda(\cdot)$ in a d-dimensional space. We henceforth assume that $\mathscr{D}_p$ is a compact convex set, and one of the key questions in the design of blind pricing policies is now concerned with the selection of "good" test prices within this feasible set.

4.1. The Blind Pricing Policy

Before describing the policy we need to define a *price grid*. Let $\mathscr{B}_p := \prod_{i=1}^d [\underline{p}^i, \bar{p}^i]$ denote the minimum volume hyperrectangle in $\mathbb{R}_+^d$, such that $\mathscr{B}_p \supseteq \mathscr{D}_p$. Given a positive integer κ, one can divide each interval $[\underline{p}^i, \bar{p}^i]$, $i = 1, \ldots, d$ into $\lfloor \kappa^{1/d} \rfloor$ intervals of equal length. Define the resulting grid of points in $\mathbb{R}_+^d$ as $\mathscr{B}_p^\kappa$. Let $e = (1, \ldots, 1) \in \mathbb{R}^\ell$. The following algorithm provides pseudo-code that defines a class of admissible learning and pricing policies that are parametrized by a triplet of tuning parameters (τ, κ, δ): $\tau \in (0, T]$ represents the length of an initial interval dedicated to learning; κ is a positive integer that defines the number of prices to "test" during the learning phase; and $\delta > 0$ is a "fudge factor" that allows for some slack in the capacity constraint (2).

Algorithm 2 $(\pi(\tau, \kappa, \delta))$

```
Step 1. Initialization:
  (a) Set the learning interval to be $[0, \tau]$, and set $\kappa$ to be
      the number of prices to experiment with.
      Put $\Delta = \tau/\kappa$.
  (b) Define $P^\kappa = \{p_1, \ldots, p_\kappa\}$ to be the prices to
      experiment with over $[0, \tau]$, where $P^\kappa \supseteq \mathscr{B}_p^\kappa \cap \mathscr{D}_p$.
Step 2. Learning/experimentation:
  (a) On the interval $[0, \tau]$ apply $p_i$ from $t_{i-1} = (i-1)\Delta$
      to $t_i = i\Delta$, $i = 1, 2, \ldots, \kappa$ as long as inventory is
      positive for all resources. If some resource runs out
      of stock, apply $p_\infty$ up until time $T$ and STOP.
  (b) Compute
```

$$\hat{d}(p_i) = \frac{\text{total demand over}[t_{i-1}, t_i)}{\Delta}, \quad i = 1, \ldots, \kappa. \tag{16}$$

```
Step 3. Optimization:
  For $i = 1, \ldots, \kappa$,
    If $A\hat{d}(p_i)T \leqslant x + \delta e$, then   [check if price is feasible]
      $\hat{r}(p_i) = p_i \cdot \hat{d}(p_i)$   [compute empirical revenue rate]
    else $\hat{r}(p_i) = 0$.
    End If
  End For
  Set $\hat{p} = \arg\max\{\hat{r}(p)\colon p \in P^\kappa\}$.
                                        [empirically optimal price]
Step 4. Pricing: On the interval $(\tau, T]$ apply $\hat{p}$ until some
  resource is out of stock, then apply $p_\infty$ for the
  remaining time.
```

Step 1 sets the first two tuning parameters: τ determines the length of interval used for learning the demand function, and κ sets the number of prices that are experimented with on that interval. In Step 2, prices in the discrete set P^κ are used to obtain an empirical approximation of the demand function.

The logic underlying Step 3 is similar to the underlying Step 3 of Algorithm 1. Here, however, one considers the following deterministic relaxation problem:

$$\max\left\{\int_0^T r(\lambda(p(s)))\,ds\colon \int_0^T A\lambda(p(s))\,ds \leqslant x,\ p(s) \in \mathscr{D}_p \quad \text{for all } s \in [0, T]\right\}, \tag{17}$$

where $r(\cdot)$ is the revenue rate. Gallego and van Ryzin (1997) show that the solution to (17) is constant over time and establish that this fixed price yields close to optimal performance in the original stochastic problem. Step 3 of the algorithm uses observed demand to form an estimate of the revenue function, and then proceeds to solve a suitable *empirical version* of the deterministic problem (17). The optimal solution for this problem $\hat{p}$ is then used for the remainder of the time horizon $(\tau, T]$. Note that in Step 3, the procedure searches for a single price, exploiting the fact that the solution to (17) is a fixed price when the price domain is a convex set. This property did not hold for the discrete price domain setting and as a result we had to solve a multidimensional linear program in Step 3 of Algorithm 1.

The choice of the tuning parameter δ allows some modest violation of the capacity constraints: the logic here is that the estimates of the demand rate are "noisy," and the δ-slack avoids restricting too drastically the search for the empirical optimal price $\hat{p}$. As opposed to Algorithm 1 which only required the specification of a learning period τ and a fudge factor δ, the continuous action set problem introduces the need for an additional tuning parameter. The choice of the learning horizon τ and number of "test prices" κ is meant to balance several contradicting objectives. As in Algorithm 1, increasing τ results in a longer time horizon over which the demand function is estimated; however, by doing so there is also a potential loss in revenues that stems from spending "too much time" on learning and exploration. Now, for every fixed choice of τ, there is also an inherent trade-off between the number of prices to experiment with, κ, and the accuracy of estimating the demand function on this price grid, which is dictated by the length $\Delta = \tau/\kappa$. In particular, using more prices translates into a better coverage of the domain of the demand surface, but it also implies that the estimates are more "noisy" because each price is used for a shorter interval. The next section explains how to balance these error sources.

4.2. Performance Analysis

In addition to the basic assumptions outlined in §2, we impose the following regularity conditions, which are standard in the revenue management literature (cf. Talluri and van Ryzin 2005).

Assumption 1. *Every demand function $\lambda(\cdot) \in \mathscr{L}$ has an inverse, denoted $\gamma(\cdot)$, the set $\mathscr{D}_\lambda := \{l\colon l = \lambda(p), p \in \mathscr{D}_p\}$ is convex, and the revenue function $r(\lambda) := \lambda \cdot \gamma(\lambda)$ is concave. In addition, $\lambda(\cdot)$ is Lipschitz continuous, i.e., $\|\lambda(p) - \lambda(p')\| \leqslant K\|p - p'\|$ for all $p, p' \in \mathscr{D}_p$, where K is a given positive constant.*

In the context of the high sales volume asymptotic regime given in (10), we provide below a result that characterizes the performance of blind policies defined by Algorithm 2.

Theorem 4. *Let Assumption* 1 *hold, and set*

$$\tau_n \asymp n^{-1/(d+3)}, \quad \kappa_n \asymp n^{d/(d+3)},$$
$$\delta_n = Cn(\log n)^{1/2} n^{-1/(d+3)}, \tag{18}$$

with $C > 0$ *sufficiently large. Then the sequence of policies* $\{\pi_n\}$ *defined by Algorithm* 2 *is asymptotically optimal. In particular,*

$$\sup_{\lambda \in \mathscr{L}} \left(1 - \frac{J_n^\pi(x, T; \lambda)}{J_n^*(x, T \mid \lambda)}\right) = O\left(\frac{(\log n)^{1/2}}{n^{1/(d+3)}}\right) \quad \textit{as } n \to \infty, \tag{19}$$

where d denotes the number of products.

Remarks.

1. As in Theorem 1, the first part of the theorem states that the value of full information diminishes for large n.

2. In contrast with the price restricted case where only finitely many actions are considered, the rate of revenue loss in the current setting degrades with the number of products (d); compare with (12). This is an obvious manifestation of the curse of dimensionality. We return to this point in §4.5, where a method is proposed to diminish this effect. It is worth noting that one could mimic the proof sketch of Theorem 1 (outlined following the statement of the result), where now, in addition to the two previous error sources, there would be an additional error source of order $\kappa_n^{-1/d}$ stemming from the discretization of the price space. In this case, the overall revenue loss would be

$$1 - J_n^\pi / J_n^* \approx C\left(\tau_n + \frac{1}{\kappa_n^{1/d}} + \frac{\kappa_n^{1/2}}{(n\tau_n)^{1/2}}\right). \tag{20}$$

4.3. A Simple State-Dependent Refinement

The learning phase in the policy described by Algorithm 2 results in an estimate of the demand function at the price vectors that are tested over the interval $[0, \tau)$. These estimates are subsequently used to solve an empirical version of the full information deterministic relaxation problem, which results in a single price that is then used for the rest of the selling season. This strategy does not make further use of the estimates of the demand function after time $t = \tau$. A simple way to refine this approach would be to *re-solve* the aforementioned optimization problem at additional points in time downstream of τ. For example, consider a policy $\pi^r(\tau, \kappa, \delta, T_r)$ that re-solves at time T_r. It proceeds as in Algorithm 2 except that Step 4 is replaced by:

```
Step 4^{(r)}. Pricing: On the interval $(\tau, T_r]$ apply $\hat{p}$. If some
    resource runs out of stock, apply $p_\infty$ up until time
    $T$ and STOP.
    Otherwise, let $I_r$ be the inventory at time $T_r$ and
    re-solve:
    For $i = 1, \ldots, \kappa$,
      If $A\hat{d}(p_i)(T - T_r) \leqslant I_r + \delta e$, then
        $\hat{r}^{(2)}(p_i) = p_i \cdot \hat{d}(p_i)$
      else $\hat{r}^{(2)}(p_i) = 0$.
      End If
    End For
    Set $\hat{p}^{(2)} = \arg\max\{\hat{r}^{(2)}(p): p \in P^\kappa\}$.
  On the interval $(T_r, T]$ apply $\hat{p}^{(2)}$ until some resource is
  out of stock, then apply $p_\infty$.
```

Intuition. While such re-solving strategies are not guaranteed to yield benefits (see, e.g., Cooper 2002 in the context of capacity allocation problems and more generally Secomandi 2008 for a detailed discussion and a review of work related to re-solving issues), the main idea here is to allow for some adaptation of the price to a given sample path of demand. As our discussion following Theorem 4 indicates, the average performance of the policy described in Algorithm 2 is essentially dictated by a law of large numbers, hence introducing re-solving points is expected to "hedge" against deviations from the average case behavior and lead to potential improvements in performance. We illustrate this in the next section.

4.4. Illustrative Numerical Examples

Note that, as in the price restricted case, $J_n^*(x, T \mid \lambda)$ is not readily computable in most cases. However, an upper bound is easy to obtain through the value of the deterministic optimization problem given in (17). This upper bound is fairly tight for moderate sized problems (see Gallego and van Ryzin 1997), and hence one can compute a "good" lower bound on the ratio $J_n^\pi(x, T; \lambda)/J_n^*(x, T \mid \lambda)$ based on this deterministic relaxation. The results depicted in Table 2 were obtained by running 10^3 independent simulation replications from which the performance indicators were derived by averaging. The standard error was below 0.5% in all cases.

The capacity consumption matrix A and true demand functions are the ones defined in §3.4. The set of feasible prices is taken to be $\mathscr{D}_p = [0.5, 5] \times [0.5, 5]$ and $T = 1$. In Table 2, we give performance results for the policy π_2 defined by Algorithm 2, and the re-solving policy π_2^r (with tuning parameters given in (18), with $C = 2$ and $T_r = 1/2$).

We observe that with inventory levels of the order of a few thousand, the expected revenues under the proposed policy π_2 exceed 65% of the optimal expected revenues in the full information case (where the demand function is known a priori). The policy utilizes approximately 18% of the time horizon T to learn the demand function, and it experiments with 16 prices. Inspecting the results, we observe that the ratio J_n^π / J_n^* approaches 1 as the market size increases, as predicted by the asymptotic optimality result in Theorem 4. We also observe that the performance of the re-solving policy π_2^r is roughly on par with that of the original policy π_2 and does not yield significant improvements. This suggests that dynamic price adjustments following the learning phase have little impact on performance and that the latter is primarily driven by the

Table 2. Results in the table give a lower bound on the ratio of the revenues extracted by the policies $\pi_2(\tau, \kappa, \delta)$ (Algorithm 2) and $\pi_2^r(\tau, \kappa, \delta, T_r)$ (Algorithm 2 with Step $4^{(r)}$) to the optimal revenues in the full information case.

Market "size"		$n = 10^2$		$n = 10^3$		$n = 10^4$	
Tuning parameters		$\kappa = 9,\ \tau = 0.29$		$\kappa = 16,\ \tau = 0.18$		$\kappa = 36,\ \tau = 0.12$	
Policy		π_2	π_2^r	π	π_2^r	π_2	π_2^r
Initial	Linear	0.57	0.60	0.65	0.67	0.76	0.74
Inventory	Exponential	0.73	0.75	0.77	0.83	0.89	0.89
$x_n = (3, 5, 7) \times n$	Logit	0.68	0.67	0.81	0.81	0.85	0.87
Initial	Linear	0.77	0.78	0.85	0.86	0.91	0.91
Inventory	Exponential	0.69	0.68	0.81	0.82	0.91	0.91
$x_n = (15, 12, 30) \times n$	Logit	0.68	0.68	0.83	0.83	0.90	0.90

Notes. Here $\kappa =$ number of prices tested by the policy, $\tau =$ fraction of time allocated to learning, and $T_r = 1/2$ is the time of re-solving and price adjustments in π_2^r.

uncertainty associated with the demand function. Comparing the results above with those in Table 1 that summarize the price-restricted case, it is evident that superior performance is observed when the set of feasible prices is finite; this is in line with intution and the performance guarantees given in Theorems 1 and 4, respectively. In particular, the performance in the finite action case does not degrade with the dimension of the set of products.

4.5. Mitigating the Curse of Dimensionality

The performance guarantee for the policies outlined in §4 degrades as the number of products d increases. The culprit here is the necessity to experiment with sufficiently many price combinations to suitably "cover" the domain of the unknown demand function. We next show how one can exploit relatively innocuous smoothness assumptions on the demand function to efficiently reconstruct the *entire* demand surface and to design "clever" pricing policies.

The form of smoothness we will be assuming is a natural strengthening of Assumption 1, which is now required to hold for the first $s - 1$ derivatives. Thus, our class of demand functions is assumed to be s-times differentiable with uniformly bounded derivatives. Note that almost all demand functions commonly used in the literature fall into this category. We state this more formally in the following assumption.

Assumption 2. *For some constant $L > 0$ and positive integer s, the demand function λ is s times differentiable, and for all $i = 1, \ldots, d$*

$$\left| \frac{\partial^{a_1, \ldots, a_d} \lambda_i(p)}{\partial p_1^{a_1} \ldots \partial p_d^{a_d}} \right| \leqslant L \tag{21}$$

for all $p \in \mathcal{D}_p$ and nonnegative integers $a_1, \ldots, a_d$ such that $a_1 + \cdots + a_d = s$.

Here d is the dimension of the set of products, and s is mnemonic for *smoothness* of the demand function. The idea now is the following: given a discrete price grid of cardinality κ, e.g., as detailed prior to the statement of Algorithm 2, and given the observed demand at each of those price vectors $y = (\hat{d}(p_1), \ldots, \hat{d}(p_\kappa))$, reconstruct an approximation $\hat{\lambda}(p; y)$ to the entire *demand surface* over the price domain, $p \in \mathcal{D}_p$. To achieve this goal, we will focus here on a standard nonparametric method based on local polynomials, which roughly works as follows: for a given price point $p \in \mathcal{D}_p$, consider a neighborhood of that point B_p that is a hypercube with edge length h; fit a polynomial of degree $s - 1$ to that neighborhood using observed demand, and approximate the value of the function $\lambda(\cdot)$ at p by that of the polynomial at the same point.

More specifically, let us focus initially on the first component of the demand function $\lambda^1(\cdot)$ and detail the development of the approximation. For $i = 1, \ldots, \kappa$ and $j = 1, \ldots, d$, let y_i^j denote the number of requests for product j when pricing at p^i in the learning phase, and let y^j denote the row vector $(y_1^j, \ldots, y_{\kappa_n}^j)$, where y_i denotes the column vector $(y_i^1, \ldots, y_i^d)^T$. Select a parameter $h > 0$ such that $h\kappa^{1/d} \geqslant s + 1$. For every $p \in \mathcal{D}_p$, we define a window $B_p = \prod_{i=1}^d B^i$, where

$$B^i = \begin{cases} [\underline{p}^i, \underline{p}^i + h] & \text{if } p^i \leqslant \underline{p}^i + h/2, \\ [\bar{p}^i - h, 1] & \text{if } p^i \geqslant \bar{p}^i - h/2, \\ [p^i - h/2, p^i + h/2] & \text{otherwise.} \end{cases}$$

The local polynomial approximation to the function $\lambda^1(\cdot)$ will be a weighted sum of the observations y_i^1. We construct a set of weights as follows. Let $\beta^1, \ldots, \beta^N$ be a basis in the space of polynomials of degree $s - 1$ of d variables. Fix a price $p \in \mathcal{D}_p$ and denote by $G = B_p \cap P^\kappa$ the set of price points in the grid that also lie in the window B_p. Let β_G^i denote the column vector whose jth component is given by the value of β^i at the jth point in G, and let $M = [\beta_G^1, \ldots, \beta_G^N]$. Given that $h\kappa^{1/d} \geqslant s + 1$, it is possible to show that M has full rank. We define a vector of weights $\omega^B(p)$ as follows:

$$\omega^B(p) = M(M^T M)^{-1} V(p), \tag{22}$$

where $V(p) = (\beta^1(p), \ldots, \beta^N(p))^T$. Note that the weights depend only on the basis of polynomials and the grid of prices. Given the weights, the approximation takes the following form:

$$\hat{\lambda}^1(p; y^1) = \sum_{i:\, p_i \in B_p} \omega_i^B(p) y_i^1. \tag{23}$$

A similar approach conducted for every component of the demand function yields the approximation

$$\hat{\lambda}(p; y) = \sum_{i:\, p_i \in B_p} \omega_i^B(p) y_i. \tag{24}$$

The proof of Theorem 5 in Appendix A contains further details on this approximation and some of its properties. See also Nemirovski (2000) for a recent reference on such approximations.

To describe a policy that uses this nonparametric regression methodology, consider now replacing Step 3 of Algorithm 2 as follows:

Algorithm 3 $(\pi(\tau, \kappa, \delta, h))$
Perform Steps 1 and 2 as in Algorithm 2.
Step 3. *Optimization*:
(a) Let $y_i = \hat{d}(p_i)$, $i = 1, \ldots, \kappa$
(b) Let $\hat{\lambda}(\cdot, y)$ be an approximation to $\lambda(\cdot)$ based on local polynomials of order $s-1$ with parameter h.
(c) Set $\hat{p} = \arg\max_{p \in \mathscr{D}_p} \{p \cdot \hat{\lambda}(p; y)\colon A\hat{\lambda}(p; y) \leqslant x + \delta e\}$
Perform Step 4 as in Algorithm 2.

The policy described by Algorithm 3 takes as input four tuning parameters $(\tau, \kappa, \delta, h)$, where h is the smoothing parameter associated with the local polynomial regression to estimate and reconstruct the demand function. In the context of the asymptotic regime given in (10), the performance of policies defined by means of Algorithm 3 is given in the following result.

Theorem 5. *Let Assumptions* 1 *and* 2 *hold. Let* π *denote the policies defined by Algorithm* 3 *where* $\tau_n \asymp n^{-1/(3+d/s)}$, $\kappa_n \asymp \lceil n^{d/(3s+d)} \rceil$, $h_n = (s+1)^{-1}\kappa_n^{-1/d}$, *and* $\delta_n \asymp C(\log n)^{1/2} n^{-1/(3+d/s)}$ *with* $C > 0$ *sufficiently large, then* $\{\pi_n\}$ *is asymptotically optimal and*

$$\sup_{\lambda \in \mathscr{L}} \left(1 - \frac{J_n^\pi(x, T; \lambda)}{J_n^*(x, T \mid \lambda)}\right) = O\left(\frac{(\log n)^{1/2}}{n^{1/(3+d/s)}}\right). \tag{25}$$

Remark (the Curse of Dimensionality). While the revenue loss relative to the full information optimal revenues given in (25) degrades as the number of products, d, increases, it is now evident that the smoother the demand function, the lesser are the curse of dimensionality effects. In particular the rate of convergence for the policy given by Algorithm 3 is $n^{-1/(3+d/s)}$ compared to $n^{-1/(3+d)}$ for the original policy described by Algorithm 2. Note that if the demand function is "very smooth" (roughly speaking infinitely continuously differentiable), then $J_n^\pi / J_n^* \approx 1 - C/n^{1/3}$, up to logarithmic terms. That is, revenue losses resulting from not knowing the demand function are *dimension-independent*, approaching the performance of the price restricted case given in Theorem 1.

Remark (Implications for Price Testing). Theorem 5 implies that the number of prices to be tested in the learning phase is reduced compared to Theorem 4. This is an important implication from a practical perspective: Algorithm 3 exploits smoothness of the demand function to extract more information per tested price. The theoretical basis for this can be found in Theorems 4 and 5, if one focuses on the magnitude of the number of price tests: $\kappa_n = n^{d/(d+3)}$ in the former; and $n^{d/(d+3s)}$ in the latter. If the function is "very smooth" (say, infinitely differentiable), then the number of prices that needs to be tested grows very slowly. In fact, it can be shown to grow roughly logarithmically with the market size n. For all practical purposes, this implies that it suffices to test a very "small" number of prices. As the remark above indicates, the performance achieved with this small number of test prices is close to being dimension-independent. Thus, with sufficient smoothness the essential complexity of the problem, in terms of both price testing and generated revenues, is "close" to that of the price restricted case (see Theorem 1). We illustrate this point in

Table 3. Exploiting smoothness.

Market "size"		$n = 10^3$		$n = 10^4$	
Tuning parameters		$\kappa = 16,\ \tau = 0.18$		$\kappa = 36,\ \tau = 0.12$	
Policy		π_2	π_3	π_2	π_3
Initial	Linear	0.65	0.76	0.76	0.83
Inventory	Exponential	0.77	0.87	0.89	0.90
$x_n = (3, 5, 7) \times n$	Logit	0.81	0.85	0.85	0.94
Initial	Linear	0.85	0.88	0.91	0.92
Inventory	Exponential	0.81	0.82	0.91	0.91
$x_n = (15, 12, 30) \times n$	Logit	0.83	0.83	0.90	0.93

Notes. Results in the table give a lower bound on the ratio of the revenues extracted by the policy π_2 (Algorithm 2) and π_3 (Algorithm 3), relative to the optimal revenues in the full information case. Here κ = number of prices tested by the policy, and τ = fraction of time allocated to learning.

a subsequent numerical example (Table 3).

As a side comment we note that the optimization problem in Step 3 of Algorithm 3 might not be concave. To this end, if the true underlying revenue function is concave in price, at least in the region where price experimentation is performed, then it is possible to obtain a tractable problem by focusing on the concave envelope of $p \cdot \hat{\lambda}(p; y)$.

Intuition. The main intuition underlying the result in Theorem 5 is as follows: as the smoothness of the underlying demand function increases, the variation of this function between any two points becomes more and more restricted. Exploiting this yields an improvement in the approximation of the demand function. In particular, one can show that with tuning parameters κ_n and h_n chosen such that $\kappa_n \asymp \lceil n^{d/(3s+d)} \rceil$ and $h_n = (s+1)^{-1}\kappa_n^{-1/d}$, one has

$$\sup_{\lambda \in \mathscr{L}} \mathbb{E}\|\hat{\lambda}_n(p; y) - \lambda(p)\|_\infty \approx (n\tau_n)^{-s/(2s+d)}. \tag{26}$$

Revisiting the remarks following Theorem 4 and in particular (20), three main error sources were highlighted: an exploration bias; a discretization error; and a stochastic error. In the current context, the key observation is that the magnitude of the discretization and stochastic errors can be reduced by exploiting smoothness. In particular, we now have for all $\lambda \in \mathscr{L}$

$$1 - \frac{J_n^\pi(x, T; \lambda)}{J_n^*(x, T; \lambda)} \approx \tau_n + (n\tau_n)^{-s/(2s+d)}, \tag{27}$$

where s is the smoothness index and d is the dimensionality. Balancing the error sources, one gets that the optimal choice of the learning horizon is $\tau_n \approx n^{-s/(3s+d)}$. Now, one has that the fraction of the optimal full information revenue extracted by π is of order $J_n^\pi / J_n^* \approx 1 - C\tau_n$. The rate τ_n degrades gracefully with the dimension d, due to the smoothness of the demand function that is exploited by the policy π.

A Numerical Example: Performance Analysis. To illustrate the performance of Algorithm 3 defined above, we consider the same setting as in §4.4. Let π denote the policy given by Algorithm 2. Let π_3 denote the policy that follows Algorithm 3 and uses local polynomials of degree 1 to approximate the demand function. In Table 3, we provide performance results when both policies use the same tuning parameters. This comparison highlights the value of "reconstructing" the demand function (as in Step 3′ of Algorithm 3) rather than restricting the search to prices that were tested in the learning phase (as in Step 3 of Algorithm 2).

We observe a general improvement when using policy π_3 compared to what is achieved by policy π_2. This improvement can be significant, at times exceeding 10% of the full information optimal revenues. In particular, for the examples considered, the performance of policy π_3 always exceeds 76% of the full information optimal revenues for market sizes of the order of 10^3. The improvements are more marginal for cases where initial inventories are large (the case $x = (15, 12, 30)$); where π already achieves a good performance. Note that the performance of policies π_2 and π_3 was illustrated using identical tuning parameters, and in particular the same number of prices were tested. Thus, one may also interpret the numerical results as follows: for any given performance level, *fewer prices* need to be tested when using π_3 as opposed to π_2.

5. Concluding Remarks

This paper formulates a class of network revenue management problems where the demand function is unknown to the decision maker. Two main settings are analyzed: one where the decision maker may select only prices from a finite set fixed a priori; and another where s/he can use any of a continuum of prices. For both settings, we develop pricing policies and characterize their theoretical performance. In particular, we show that some of these policies achieve near-optimal performance by providing lower bounds on the performance of any policy, and proving that our policies come "close" to said bounds. The paper highlights a general approach to solving dynamic optimization problems with very limited prior information, but it leaves the challenge of designing practical and implementable algorithms to future research. Another interesting direction of study would be to build on the price testing insights that have been developed in this paper and verify whether they can be translated into practical guidelines. While the paper focuses solely on the setting where the demand function is time homogeneous, it lays out foundations for studying the more complex case of time-varying demand. Here the recent work of Besbes and Zeevi (2011) might be useful as it develops an approach to online detection of changes in the market environment, and hence can be combined potentially with the ideas developed in the present paper. Developing pricing policies that can deal with limited prior information on demand as well as the possibility of temporal variation would constitute an important step toward bridging the gap between current theory and practice.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/opre.1120.1103.

Endnotes

[1] The assumption of time-homogeneity of the demand function allows to isolate learning effects from demand tracking considerations. The challenges associated with an unknown time-dependent demand function are highlighted in §5.

[2] An alternative interpretation of this regime is one where the sales rate *does not* change but the time horizon grows linearly with n: all of the results in the paper carry over with appropriate modifications to that setting.

[3] The asymptotic regime considered is not appropriate for cases when there is a fundamental mismatch between supply and demand, e.g., $x \ll \lambda(p)$.

4. This assumes that the learning phase τ_n is taken to be proportional to $k^{1/3}n^{-1/3}$.
5. We note that in Theorem 2 the assumption that $x = \infty$ is made only to simplify the argument. It is sufficient to assume unconstrained "on average," i.e., that the class of admissible demand functions $\mathscr{L}$ contains a demand function $\mu(\cdot)$ such that $\max\{\mu(p_1), \mu(p_2)\}T < x - 1$.

Acknowledgments

This research was partially supported by the National Science Foundation [Grant DMI-0447562] and the Binational Science Foundation [Grant 2006075].

References

Araman VF, Caldentey RA (2009) Dynamic pricing for nonperishable products with demand learning. *Oper. Res.* 57(5):1169–1188.

Aviv Y, Pazgal A (2005) Pricing of short life-cycle products through active learning. Working paper, Washington University in Saint Louis, Saint Louis.

Ball M, Queyranne M (2009) Toward robust revenue management: Competitive analysis of online booking. *Oper. Res.* 57(4):950–963.

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

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

Bitran G, Caldentey R (2003) An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* 5(3):203–229.

Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (Cambridge University Press, New York).

Cooper W (2002) Asymptotic behavior of an allocation policy for revenue management. *Oper. Res.* 50(4):720–727.

Elmaghraby W, Keskinocak P (2003) Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Sci.* 49(10):1287–1309.

Eren S, Maglaras C (2010) Monopoly pricing with limited demand information. *J. Revenue and Pricing Management* 9:23–48.

Farias VF, Van Roy B (2010) Dynamic pricing with a prior on market response. *Oper. Res.* 58(1):16–29.

Fisher M, Rajaram K (2000) Accurate retail testing of fashion merchandise: Methodology and application. *Marketing Sci.* 19(3):226–278.

Gallego G, van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 50(8):999–1020.

Gallego G, van Ryzin G (1997) A multiproduct dynamic pricing problem and its applications to network yield management. *Oper. Res.* 45(1):24–41.

Gaur V, Fisher ML (2005) In-store experiments to determine the impact of price on sales. *Production Oper. Management* 14(4):377–387.

Lim AEB, Shanthikumar JG (2007) Relative entropy, exponential utility, and robust dynamic pricing. *Oper. Res.* 55(2):198–214.

Lim AEB, Shanthikumar JGG, Watewai T (2008) Robust multi-product pricing. Working paper, University of California, Berkeley, Berkeley, CA.

Lobo MS, Boyd S (2003) Pricing and learning with uncertain demand. Working paper, Duke University, Durham, NC.

Nemirovski A (2000) Topics in non-parametric statistics. *Ecole d'été de Probabilités de Saint-Flour XXVIII. Lecture Notes in Mathematics* (Springer, Berlin), 1738:85–277.

Rusmevichientong P, Van Roy B, Glynn PW (2006) A nonparametric approach to multiproduct pricing. *Oper. Res.* 54(1):82–98.

Secomandi N (2008) An analysis of the control-algorithm re-solving issue in inventory and revenue management. *Manufacturing Service Oper. Management* 10(3):468–483.

Talluri KT, van Ryzin GJ (2005) *Theory and Practice of Revenue Management* (Springer-Verlag, New York).

van Ryzin G, McGill J (2000) Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Sci.* 46(6):760–775.

Williams L, Partani A (2006) Learning through price testing. *6th Annual INFORMS Revenue Management and Pricing Section Conf., June 5–6* (INFORMS, Hanover, MD).

Omar Besbes is an assistant professor in the Graduate School of Business, Columbia University. His research interests include dynamic pricing, revenue management, and service operations, with a particular emphasis on data-driven decision making.

Assaf Zeevi is a professor of business in the Graduate School of Business, Columbia University. His main research focuses on stochastic modeling and statistics and their applications in service operations, revenue management, financial engineering, and applied probability in general.