Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms*

Omar Besbes[†]

University of Pennsylvania

Assaf Zeevi[‡]

Columbia University

Submitted: November 10, 2006   First revision submitted: June 8, 2007

Last revised: December 16, 2007

Abstract

We consider a single product revenue management problem where, given an initial inventory, the objective is to dynamically adjust prices over a finite sales horizon to maximize expected revenues. Realized demand is observed over time, but the underlying functional relationship between price and mean demand rate that governs these observations (otherwise known as the demand function or demand curve), is not known. We consider two instances of this problem: i.) a setting where the demand function is assumed to belong to a known parametric family with unknown parameter values; and ii.) a setting where the demand function is assumed to belong to a broad class of functions that need not admit any parametric representation. In each case we develop policies that learn the demand function "on the fly," and optimize prices based on that. The performance of these algorithms is measured in terms of the *regret*: the revenue loss relative to the maximal revenues that can be extracted when the demand function is known prior to the start of the selling season. We derive lower bounds on the regret that hold for any admissible pricing policy, and then show that our proposed algorithms achieve a regret that is "close" to this lower bound. The magnitude of the regret can be interpreted as the economic value of prior knowledge on the demand function; manifested as the revenue loss due to model uncertainty.

Keywords: Revenue management, pricing, estimation, learning, exploration-exploitation, value of information, asymptotic analysis.

1

1 Introduction

Since the deregulation of the airline industry in the 1970's, revenue management practices that deal with complex pricing and demand management decisions have become increasingly prevalent in a variety of industries; see Bitran and Caldentey (2003), Elmaghraby and Keskinocak (2003) and Talluri and van Ryzin (2005) for a broad overview of the field and recent developments. (The literature review in Section 2 provides a detailed survey of work directly related to the current paper.) A critical assumption made in most academic studies of revenue management problems is that the functional relationship between the mean demand rate and price, often referred to as the *demand function* or *demand curve*, is known to the decision maker. This makes the underlying problems more tractable and allows one to extract structural insights. At the same time, this assumption of "full information" endows the decision maker with knowledge that s/he does not typically possess in practice.

Lack of information concerning the demand model raises several fundamental questions. First, and foremost, is it possible to quantify the "value" of full information (for example, by measuring the revenue loss due to imperfect information)? Second, is it possible to achieve anything close to the maximal revenues in the full information setting by judiciously combining real-time demand learning and pricing strategies? Finally, how would such strategies exploit prior information, if any, on the structure of the demand function?

The main objective of this paper is to shed some light on the aforementioned questions. Our departure point will be a prototypical single product revenue management problem first introduced and formalized by Gallego and van Ryzin (1994). This formulation models realized demand as a Poisson process whose intensity at each point in time is determined by a price set by the decision maker. Given an initial inventory, the objective is to dynamically price the product so as to maximize expected revenues over a finite selling horizon. In the dynamic optimization problem considered in Gallego and van Ryzin (1994), the decision maker knows the demand function prior to the start of the selling season and designs optimal policies based on this information. In the setting we pursue in this paper it is only possible to observe *realized demand* over time, and the demand function itself is *not known*. To that end, we consider two levels of uncertainty with regard to the demand model: i.) a *nonparametric setting* where the demand function is only assumed to belong to a broad functional class satisfying mild regularity conditions; and ii.) a *parametric setting* in which the demand function admits a given parametric structure but the parameter values are not known.

The absence of perfect prior information concerning the demand model introduces an important new component into the above dynamic optimization problem, namely, tension between exploration (demand learning) and exploitation (pricing). The longer one spends learning the demand characteristics, the less time remains to exploit that knowledge and optimize profits. On the other hand,

less time spent on demand learning leaves more residual uncertainty that could hamper any pricing strategy in the exploitation phase. One of the main contributions of this paper is to formulate this dynamic pricing problem under incomplete information, and to pursue an analysis that highlights the key tradeoffs discussed above, and articulate them in a precise mathematical manner.

To address uncertainty with regard to the demand model, we introduce a family of pricing policies that learn the demand function "on the fly." Their performance will be measured in terms of the revenue loss relative to a full information benchmark that assumes knowledge of the demand function. We refer to this loss as the *regret* associated with not knowing the demand function a priori; the magnitude of the regret quantifies the *economic value* of prior model information. The policies we consider are designed with the objective of achieving a "small" regret *uniformly* over the relevant class of demand functions (either parametric or nonparametric). This adversarial setting, where nature is allowed to counter a chosen policy with the "worst" demand function, ensures that policies exhibit "good" performance irrespective of the true demand model.

The complexity of the problem described above makes it difficult to evaluate the performance of any reasonable policy, except via numerical experiments. To address this issue, we consider an asymptotic regime which is characterized by a high volume of sales. More specifically, the initial level of inventory and the magnitude of demand ("market size") grow large in proportion to each other; see Gallego and van Ryzin (1994) and Talluri and van Ryzin (2005) for further examples in the revenue management literature that adopt this framework. This regime allows us to bound the magnitude of the regret, and to establish a rather surprising result with regard to our proposed policies: as the sales volume grows large, the regret eventually shrinks to zero. That is, these policies achieve (asymptotically) the maximal full information revenues, despite the absence of prior information regarding the demand function; in that sense, they are *asymptotically optimal*.

In more detail, the main contributions of this paper are summarized as follows.

i.) We introduce a nonparametric pricing policy (see Algorithm 1) that requires almost no prior information on the demand function. In settings where the structure of the demand function is known up to the value/s of certain parameter/s, we develop a parametric pricing policy based on Maximum Likelihood estimation (see Algorithm 2 and Algorithm 3).

ii.) We establish lower bounds on the regret that hold for *any* admissible learning and pricing policy (see Propositions 2 and 4).

iii.) We derive upper bounds on the performance of our nonparametric and parametric pricing policies (see Propositions 1 and 3). In all cases the proposed policies achieve a regret that is "not far" from the lower bound described above. In the parametric setting when only one parameter is unknown, we prove that essentially no admissible pricing policy can achieve a smaller regret than our proposed method (see Proposition 5).

iv.) Building on ideas from stochastic approximations, we indicate how one can develop more refined sequential policies, and illustrate this in the nonparametric setting (see Algorithm 4).

Returning to the questions raised earlier in this section, our results shed light on the following issues. First, despite having only limited (or almost no) prior information, it is possible to construct joint learning and pricing policies that generate revenues which are "close" to the best achievable performance with full information. Our results highlight an interesting observation. In the full information setting, Gallego and van Ryzin (1994) prove that fixed price heuristics lead to near-optimal revenues, hence the value of dynamic price changes is (at least asymptotically) limited. In a setting with incomplete information, price changes play a much more pivotal role, as they are relied upon to resolve uncertainty with regard to the demand function.

The regret bounds described above rigorously quantify the economic value of a priori information on the demand model. Alternatively, the lost revenues can be viewed as quantifying the "price" paid for model uncertainty. Finally, our work highlights an important issue related to model misspecification risk. In particular, if an algorithm is designed under parametric assumptions, it is prone to such risk as the true demand function may not (and in many cases will not) belong to the assumed parametric family. Our regret bounds provide a means for quantifying the "price" that one pays for eliminating this risk via nonparametric approaches; see also the numerical illustration in Section 6.

The remainder of the paper. The next section reviews related literature. Section 3 introduces the model and formulates the problem. Section 4 studies the nonparametric setting and Section 5 focuses on cases where the demand function possesses a parametric structure. Section 6 presents numerical results and discusses some qualitative insights. Section 7 formulates adaptive versions of the nonparametric algorithm and illustrates their performance. All proofs are collected in two appendices: Appendix A contains the proofs of the main results; and Appendix B contains proofs of auxiliary lemmas.

2   Related Literature

Parametric approaches. The majority of revenue management studies that address demand function uncertainty do so by assuming that one or more parameters characterizing this function are unknown. The typical approach here follows a dynamic programming formulation with Bayesian updating, where a prior on the distribution of the unknown parameters is initially postulated. Recent examples include Aviv and Pazgal (2005), Araman and Caldentey (2005) and Farias and Van Roy (2006), all of which assume a single parameter is unknown. (See also Lobo and Boyd (2003) and Carvalho and Puterman (2005).) Scarf (1959) was one of the first papers to use this Bayesian formulation, though in the context of inventory management.

While the Bayesian approach provides for an attractive stylized analysis of the joint learning and pricing problem, it suffers from significant shortcomings. Most notably, the objective of the dynamic optimization problem involves an expectation that is taken relative to a prior distribution over the unknown parameters. Hence any notion of optimality associated with a Bayesian-based policy is with respect to that prior. Moreover, the specification of this prior distribution is typically constrained to so-called conjugate families and is not driven by "real" prior information; the hindering element here is the computation of the posterior via Bayes rule. The above factors introduce significant restrictions on the models that are amenable to analysis via Bayesian dynamic programming. Bertsimas and Perakis (2003) have considered an alternative to this formulation using a least squares approach in the context of a linear demand model. Our work in the parametric setting is based on maximum likelihood estimation, and hence is applicable to a wide class of parametric models; as such, it offers an alternative to Bayesian approaches that circumvents some of their deficiencies.

Nonparametric approaches. The main difficulty facing nonparametric approaches is loss of tractability. Most work here has been pursued in relatively simple static settings that do not allow for learning of the demand function; see, e.g., Ball and Queyranne (2006) and Eren and Maglaras (2007) for a competitive ratio formulation, and Perakis and Roels (2006) for a minimax regret formulation. These studies focus almost exclusively on structural insights. The recent paper by Rusmevichientong et al. (2006) develops a nonparametric approach to a multiproduct static pricing problem, based on historical data. The formulation does not incorporate inventory constraints or finite sales horizon considerations. In the context of optimizing seat allocation policies for a single flight multi-class problem, van Ryzin and McGill (2000) show that one can use stochastic approximation methods to reach near-optimal capacity protection levels in the long run. The absence of parametric assumptions in this case is with respect to the distributions of customers' requests for each class. (See also Huh and Rusmevichientong (2006) for a related study.)

Perhaps the most closely related paper to our current work is that of Lim and Shanthikumar (2006) who formulate a robust counterpart to the single product revenue management problem of Gallego and van Ryzin (1994). In that paper the uncertainty arises at the level of the point process distribution characterizing realized demand, and the authors use a max-min formulation where nature is adversarial at *every point in time*. This type of conservative setting effectively precludes any real-time learning, and moreover does not lend itself to prescriptive solutions.

Related work in other disciplines. The general problem of dynamic optimization with limited or no information about a response function has also attracted attention in other fields. In economics, a line of work that traces back to Hannan (1957) studies settings where the decision maker faces an oblivious opponent. The objective is to minimize the difference between the rewards accumulated by a given policy, and the rewards accumulated by the best possible single action had

5

the decision maker known in advance the actions of the adversary; see Foster and Vohra (1999) for a review of this line of work and its relation to developments in other fields.

A classical formulation of sequential optimization under uncertainty that captures the essence of the exploration-exploitation tradeoff, is the multiarmed bandit paradigm that dates back to the work of Robbins (1952). This was originally introduced as a model of clinical trials in the statistics literature, but has since been used in many other settings; see, e.g., Lai and Robbins (1985) and references therein. Related studies in the computer science literature include Auer et al. (2002) who study an adversarial version of a multi-armed bandit problem, and Kleinberg and Leighton (2003) who provide an analysis of an on-line posted-price auction using these tools. (See Cesa-Bianchi and Lugosi (2006) for a recent and comprehensive survey). Our work shares an important common theme with the streams of literature survey above, insofar as it too highlights exploration-exploitation tradeoffs. On the other hand, our work represents a significant departure from antecedent literature along three important dimensions that are characteristic of our dynamic pricing problem: we deal with a *constrained* dynamic optimization problem (the constraint arising from the initial inventory level); the action space of the decision maker, namely the feasible price set, is uncountable; and the action space of the adversary (nature), namely, the class of admissible demand functions, is also uncountable.

3 Problem Formulation

Model primitives and basic assumptions. We consider a revenue management problem in which a monopolist sells a single product. 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 the product at any time $t \in [0, T]$ is given by a Poisson process with intensity λ_t which measures the instantaneous demand rate (in units such as number of products requested per hour, say): Letting $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ denote the *demand function*, then if the price at time t is $p(t)$, the instantaneous demand rate at time t is given by $\lambda_t = \lambda(p(t))$, and realized demand is a controlled Poisson process with this intensity.

We assume that the set of feasible prices is $[\underline{p}, \overline{p}] \cup p_\infty$, where $0 < \underline{p} < \overline{p} < \infty$ and $p_\infty > 0$ is a price that "turns off" demand (and revenue rate), i.e., $\lambda(p_\infty) = 0$[1]. With regard to the demand function, we assume that $\lambda(\cdot)$ is non-increasing in the price p, has an inverse denoted by $\gamma(\cdot)$, and the revenue rate $r(\lambda) := \lambda\gamma(\lambda)$ is concave. These assumptions are quite standard in the revenue management literature resulting in the term *regular* affixed to demand functions satisfying these conditions; see, e.g., Talluri and van Ryzin (2005, §7).

Let $(p(t) : 0 \le t \le T)$ denote the *price process* which is assumed to have sample paths that are

---

[1]The case $p_\infty = \infty$ can be incorporated by assuming $r(\lambda(p_\infty)) := \lim_{p \to p_\infty} r(\lambda(p)) = 0$ which implies that $\lim_{p \to p_\infty} \lambda(p) = 0$.

right continuous with left limits taking values in $[\underline{p}, \overline{p}] \cup p_\infty$. Let $N(\cdot)$ be a unit rate Poisson process. The cumulative demand for the product up until time t is then given by $D(t) := N(\int_0^t \lambda(p(s))ds)$. We say that $(p(t) : 0 \le t \le T)$ is *non anticipating* if the value of $p(t)$ at each time $t \in [0, T]$ is only allowed to depend on past prices $\{p(s) : s \in [0, t)\}$ and demand values $\{(D(s)) : s \in [0, t)\}$. (More formally, the price process is adapted to the filtration generated by the past values of the demand and price processes.)

Information structure and the economic optimization problem. We assume that the decision maker *does not know* the true demand function λ, but 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. The only information available regarding λ is that it belongs to a class of admissible demand functions, $\mathcal{L}$; in Section 4, $\mathcal{L}$ will be taken to be a *nonparametric* class of functions, and in Section 5 it will be restricted to a *parametric* class. Thus, the makeup of the class $\mathcal{L}$ summarizes prior information on the demand model.

We shall use π to denote a *pricing policy*, which, roughly speaking, maps the above information structure to a non anticipating price process $(p(t) : 0 \le t \le T)$. With some abuse of terminology, we will use the term "policy" to refer to the price process itself and the algorithm that generates it, interchangeably. Put

$$N^\pi(t) := N\left(\int_0^t \lambda(p(s))ds\right) \qquad \text{for } 0 \le t \le T, \tag{1}$$

where $N^\pi(t)$ denotes the cumulative demand up to time t under the policy π.

Let $x > 0$ denote the inventory level (number of products) at the start of the selling season. A pricing policy π is said to be *admissible* if the induced price process satisfies

$$\int_0^T dN^\pi(s) \le x \qquad \text{a.s.}, \tag{2}$$

$$p(s) \in [\underline{p}, \overline{p}] \cup p_\infty, \ 0 \le s \le T. \tag{3}$$

It is important to note that while the decision maker does not know the demand function, knowledge that $\lambda(p_\infty) = 0$ guarantees that the constraint (2) can be met. Let $\mathcal{P}$ denote the set of admissible pricing policies.

The dynamic optimization problem faced by the decision maker under the information structure described above is: choose $\pi \in \mathcal{P}$ to maximize the total expected revenues

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

The dependence on λ in the left-hand-side is indicative of the fact that the expectation on the right-hand-side is taken with respect to the *true* demand distribution. Since the decision maker cannot compute the expectation in (4) without knowing the underlying demand function, the above

optimization problem does not seem to be well posed. In a sense one can view the solution of (4) as being made possible only with the aid of an "oracle" which can compute the quantity $J^{\pi}(x, T; \lambda)$ for any given policy. We will now redefine the decision maker's objective in a more suitable manner, using the notion of a full information benchmark.

A full information benchmark. Let us first explain how the analysis of the dynamic optimization problem described in (4) proceeds when one removes two significant obstacles: lack of knowledge of the demand function λ prior to the start of the selling season; and stochastic variability in realized demand. In particular, consider the following full information deterministic optimization problem, in which the function λ is assumed to be known at time $t = 0$:

$$\sup \int_0^T r(\lambda(p(s)))ds, \tag{5}$$

$$\text{s.t.} \int_0^T \lambda(p(s))ds \leq x,$$

$$p(s) \in [\underline{p}, \overline{p}] \cup p_\infty \quad \text{for all } s \in [0, T].$$

This problem is obtained from (4), and the admissibility conditions (2)-(3), by replacing the random process characterizing customer purchase requests by its mean rate. For example, if one focuses on the objective (4), then the deterministic objective in (5) is obtained by substituting "$\lambda(p(s))ds$" for "$dN^{\pi}(s)$" since $r(\lambda(p(s))) = p(s)\lambda(p(s))$. The same parallel can be drawn between the first constraint of the deterministic problem and (2). Consequently, it is reasonable to refer to (5) as a *full information deterministic relaxation* of the original dynamic pricing problem (4).

Let us denote the value of (5) as $J^D(x, T|\lambda)$ where 'D' is mnemonic for deterministic and the choice of notation with respect to λ reflects the fact that the optimization problem is solved "conditioned" on knowing the true underlying demand function. The value of the full information deterministic relaxation provides, as one would anticipate, an upper bound on expected revenues generated by any pricing policy $\pi \in \mathcal{P}$, that is, $J^{\pi}(x, T; \lambda) \leq J^D(x, T|\lambda)$ for all $\lambda \in \mathcal{L}$; this rather intuitive observation is formalized in Lemma 1 in Appendix A which essentially generalizes Gallego and van Ryzin (1994, Proposition 2).

The minimax regret objective. As indicated above, for any demand function $\lambda \in \mathcal{L}$, we have that $J^{\pi}(x, T; \lambda) \leq J^D(x, T|\lambda)$ for all admissible policies $\pi \in \mathcal{P}$. With this in mind, we define the *regret* $\mathcal{R}^{\pi}(x, T; \lambda)$, for any given function $\lambda \in \mathcal{L}$ and policy $\pi \in \mathcal{P}$, to be

$$\mathcal{R}^{\pi}(x, T; \lambda) = 1 - \frac{J^{\pi}(x, T; \lambda)}{J^D(x, T|\lambda)}. \tag{6}$$

The regret measures the percentage loss in performance of any policy π in relation to the benchmark $J^D(x, T|\lambda)$. By definition, the value of the regret always lies in the interval $[0, 1]$, and the smaller the regret, the better the performance of a policy π; in the extreme case when the regret is zero, then the policy π is guaranteed to extract the maximum full information revenues. Since the

decision maker does not know which demand function s/he will face in the class $\mathcal{L}$, it is attractive to design pricing policies that perform well irrespective of the actual underlying demand function. In particular, if the decision maker uses a policy $\pi \in \mathcal{P}$, and nature then "picks" the worst possible demand function for that policy, then the resulting regret would be

$$\sup_{\lambda \in \mathcal{L}} \mathcal{R}^{\pi}(x, T; \lambda). \tag{7}$$

In this game theoretic setting it is now possible to restate the decision maker's objective, initially given in (4), as follows: pick $\pi \in \mathcal{P}$ to minimize (7). The advantage of this formulation is that the decision maker's problem is now well posed: for any $\pi \in \mathcal{P}$ and fixed $\lambda \in \mathcal{L}$ it is possible, at least in theory, to compute the numerator on the right hand side in (6), and hence (7). Roughly speaking, one can attach a worst case $\lambda \in \mathcal{L}$ to "each" policy $\pi \in \mathcal{P}$, and subsequently one can try to "optimize" this by searching for the policy with the best worst-case performance. In other words, we are interested in characterizing the *minimax regret*

$$\inf_{\pi \in \mathcal{P}} \sup_{\lambda \in \mathcal{L}} \mathcal{R}^{\pi}(x, T; \lambda). \tag{8}$$

This quantity has an obvious physical interpretation: it measures the monetary value (in normalized currency units) of knowing the demand function a priori. The issue of course is that, barring exceedingly simple cases, it is not possible to compute the minimax regret. Our objective in what follows will be to characterize this quantity by deriving suitable bounds on (8) in cases where the class $\mathcal{L}$ is nonparametric or is restricted to a suitable parametric class of demand functions.

4 Main Results: The Nonparametric Case

4.1 A nonparametric pricing algorithm

We introduce below a learning and pricing policy defined through two tuning parameters (κ, τ): κ is a positive integer and $\tau \in (0, T]$. The general structure, which is summarized for convenience in algorithmic form, is divided into two main stages. A "learning" phase (exploration) of length τ is first used, in which κ prices are tested. Then a "pricing" phase (exploitation) fixes a "good" price based on demand observations in the first phase. The intuition underlying the method is discussed immediately following the description of the method.

Algorithm 1 : $\pi(\tau, \kappa)$

Step 1. Initialization:

    (a) Set the learning interval to be $[0, \tau]$, and the number of prices to experiment with to be κ. Put $\Delta = \tau / \kappa$.

(b) Divide $[\underline{p}, \overline{p}]$ into κ equally spaced intervals and let $\{p_i, i = 1, ..., \kappa\}$ be the left endpoints of these intervals.

Step 2. Learning/experimentation:

(a) On the interval $[0, \tau]$ apply p_i from $t_{i-1} = (i-1)\Delta$ to $t_i = i\Delta$, $i = 1, 2, ..., \kappa$, as long as inventory is positive. If no more units are in stock, apply p_∞ up until time T and STOP.

(b) Compute

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

Step 3. Optimization:

Compute
$$\hat{p}^u = \arg\max_{1 \le i \le \kappa}\{p_i \hat{d}(p_i)\}, \qquad \hat{p}^c = \arg\min_{1 \le i \le \kappa}\left|\hat{d}(p_i) - x/T\right|, \tag{9}$$

and set
$$\hat{p} = \max\{\hat{p}^c, \hat{p}^u\}. \tag{10}$$

Step 4. Pricing:

On the interval $(\tau, T]$ apply $\hat{p}$ as long as inventory is positive, then apply p_∞ for the remaining time.

Intuition and key underlying ideas. At first, a nonparametric empirical estimate of the demand function is obtained based on a learning phase of length τ described in Steps 1 and 2. The intuition underlying Steps 3 and 4 is based on the analysis of the deterministic relaxation (5) whose solution (see Lemma 1 in Appendix A) is given by $p(s) = p^D := \max\{p^u, p^c\}$ for $s \in [0, T']$ and $p(s) = p_\infty$ for $s > T'$, where

$$p^u = \arg\max_{p \in [\underline{p}, \overline{p}]}\{r(\lambda(p))\}, \qquad p^c = \arg\min_{p \in [\underline{p}, \overline{p}]} |\lambda(p) - x/T|,$$

and $T' = \min\{T, x/\lambda(p^D)\}$. Here, the superscripts "u" and "c" stand for *unconstrained* and *constrained*, respectively, in reference to whether the inventory constraint in (5) is binding or not. In particular, the deterministic problem (5) can be solved by restricting attention to the two prices p^u and p^c. Algorithm 1 hinges on this observation.

In Step 3 the objective is to obtain an accurate estimate of p^D based on the observations during the "exploration" phase, while at the same time keeping τ "small" in order to limit the revenue loss over this learning phase. The algorithm then applies this price on $(\tau, T]$. With the exception of the "short" initial phase $[0, \tau]$, the expected revenues (in the real system) will be close to those achieved by p^D over $[0, T]$. The analysis in Gallego and van Ryzin (1994) establishes that those revenues would be close to $J^D(x, T|\lambda)$, and as a result, the regret $\mathcal{R}^\pi(x, T; \lambda)$ should be small.

To estimate p^D, the algorithm dedicates an initial portion $[0, \tau]$ of the total selling interval $[0, T]$ to an exploration of the price domain. On this initial interval, the algorithm experiments with κ prices where each is kept fixed for τ/κ units of time. This structure leads to three main sources of error in the search for p^D. First, during the learning phase one incurs an *exploration bias* since the prices being tested there are not close to p^D (or close to the optimal fixed price for that matter). This incurs losses of order τ. Second, experimenting with only a finite number of prices κ in the search for p^D results in a *deterministic error* of order $1/\kappa$ in Step 3. Finally, only "noisy" demand observations are available at each of the κ price points, and the longer a price is held fixed, the more accurate the estimate of the mean demand rate at that price. This introduces a *stochastic error* of order $(\tau/\kappa)^{-1/2}$ stemming from the nature of the Poisson process. The crux of the matter is to balance these three error sources by a suitable choice of the tuning parameters τ and κ.

4.2 Model uncertainty: the class of demand functions

The nonparametric class of functions we consider consists of regular demand function (satisfying the standard conditions laid out in Section 3) which in addition satisfy the following.

Assumption 1 For some finite positive constants $M, \underline{K}, \overline{K}, m$, with $\underline{K} \leq \overline{K}$:

(i.) Boundedness: $|\lambda(p)| \leq M$ for all $p \in [\underline{p}, \overline{p}]$.

(ii.) Lipschitz continuity: $|\lambda(p) - \lambda(p')| \leq \overline{K}|p - p'|$ for all $p, p' \in [\underline{p}, \overline{p}]$ and $|\gamma(l) - \gamma(l')| \leq \underline{K}^{-1}|l - l'|$ for all $l, l' \in [\lambda(\overline{p}), \lambda(\underline{p})]$.

(iii.) Minimum revenue rate: $\max\{p\lambda(p) : p \in [\underline{p}, \overline{p}]\} \geq m$.

Let $\mathcal{L} := \mathcal{L}(M, \underline{K}, \overline{K}, m)$ denote this class of demand functions. Assumption 1(i.) and 1(ii.) are quite benign, only requiring minimal smoothness of the demand function; (ii.) also ensures that when $\lambda(\cdot)$ is positive, it does not have "flat regions." Assumption 1(iii.) states that a minimal revenue rate exists, hence avoiding trivialities.

Note that assumptions (i.)-(iii.) above hold for many models of the demand function used in the revenue management and economics literature (e.g., linear, exponential and iso-elastic/Pareto with parameters lying in a compact set; cf. Talluri and van Ryzin (2005, §7) for further examples).

4.3 Performance analysis

Since minimax regret is hardly a tractable quantity, we introduce in this section an asymptotic regime characterized by a "high volume of sales," which will be used to analyze the performance of Algorithm 1. We consider a regime in which both the size of the initial inventory as well as potential demand grow proportionally large. In particular for a market of "size" n, where n is a

positive integer, the initial inventory and the demand function are now assumed to be given by

$$x_n = nx \qquad \text{and} \qquad \lambda_n(\cdot) = n\lambda(\cdot). \qquad (11)$$

Thus, the index n determines the order of magnitude of both inventory and rate of demand. We will denote by $\mathcal{P}_n$ the set of admissible policies for a market of scale n, and the expected revenues under a policy $\pi_n \in \mathcal{P}_n$ will be denoted $J_n^\pi(x, T; \lambda)$. With some abuse of notation, we will occasionally use π to denote a sequence of policies $\{\pi_n, n = 1, 2, \ldots\}$ as well as any element of that sequence, omitting the subscript "n" to avoid cluttering the notation. For each $n = 1, 2, \ldots$, we denote by $J_n^D(x, T | \lambda)$ the value of the deterministic relaxation given in (5) with the scaling given in (11); it is straightforward to verify that $J_n^D(x, T | \lambda) = nJ^D(x, T | \lambda)$. Finally let the regret be denoted as $\mathcal{R}_n^\pi(x, T; \lambda) := 1 - J_n^\pi(x, T; \lambda)/J_n^D(x, T | \lambda)$. The following definition characterizes admissible policies that have "good" asymptotic properties.

Definition 1 (Asymptotic optimality) *A sequence of admissible policies $\pi_n \in \mathcal{P}_n$ is said to be asymptotically optimal if*

$$\sup_{\lambda \in \mathcal{L}} \mathcal{R}_n^\pi(x, T; \lambda) \to 0 \qquad \text{as } n \to \infty. \qquad (12)$$

In other words, asymptotically optimal policies achieve the full information upper bound on revenues as $n \to \infty$, *uniformly* over the class of admissible demand functions.

For the purpose of asymptotic analysis we use the following notation: 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 by a constant, and if a_n/b_n is also bounded from below then we write $a_n \asymp b_n$. We now analyze the performance of policies associated with Algorithm 1 in the asymptotic regime described above.

Proposition 1 *Let Assumption 1 hold. Set $\tau_n \asymp n^{-1/4}$, $\kappa_n \asymp n^{1/4}$ and let $\pi_n := \pi(\tau_n, \kappa_n)$ be given by Algorithm 1. Then, the sequence $\{\pi_n\}$ is asymptotically optimal, and for all $n \geq 1$*

$$\sup_{\lambda \in \mathcal{L}} \mathcal{R}_n^\pi(x, T; \lambda) \leq \frac{C(\log n)^{1/2}}{n^{1/4}}, \qquad (13)$$

for some finite positive constant C.

The constant C above depends only on the parameters characterizing the class $\mathcal{L}$, the initial inventory x, and the time horizon T. The exact dependence is somewhat complex and is omitted, however, we note that it is fully consistent with basic intuition: as one expands the class $\mathcal{L}$ by suitably increasing or decreasing the value of the parameters in Assumption 1, the magnitude of C grows, and vice versa. (This can also be inferred by carefully inspecting the proof of the proposition.)

We next present a lower bound on the minimax regret which establishes a fundamental limit on the performance of *any* admissible pricing policy. Roughly speaking, the main idea behind this result is to construct a "worst case" demand function such that the regret is large for any policy.

Proposition 2 *Let Assumption 1 hold with $M, \overline{K}$ satisfying $M \geq \max\{2\overline{K}\overline{p}, \overline{K}\overline{p} + x/T\}$. Then there exists a finite positive constant C' such that for any sequence of admissible policies $\{\pi_n\}$ and for all $n \geq 1$*

$$\sup_{\lambda \in \mathcal{L}} \mathcal{R}_n^\pi(x, T; \lambda) \geq \frac{C'}{n^{1/2}}. \tag{14}$$

Combining Propositions 1 and 2, one can characterize the magnitude of the minimax regret as follows:

$$\frac{C'}{n^{1/2}} \leq \inf_{\pi \in \mathcal{P}_n} \sup_{\lambda \in \mathcal{L}} \mathcal{R}_n^\pi(x, T; \lambda) \leq \frac{C(\log n)^{1/2}}{n^{1/4}}, \tag{15}$$

and hence the performance of Algorithm 1 is "not far" from being minimax optimal (i.e., achieving the lower bound). A question that remains open is whether one can close this gap by further refining the algorithm. We revisit this point in Sections 5 and 7.

5 Main Results: The Parametric Case

In this section we assume the demand function is known to have a parametric form. Our goal is to develop pricing policies that exploit this information and that will work well for a large class of admissible parametric demand functions. One of the main questions of interest here is whether these policies achieve a smaller regret relative to the nonparametric case, and if there exist policies whose performance cannot be improved upon.

Preliminaries. Let k denote a positive integer and $\mathcal{L}(\Theta) = \{\lambda(\cdot; \theta) : \theta \in \Theta\}$ be a parametric family of demand functions, where $\Theta \subseteq \mathbb{R}^k$ is assumed to be a convex compact set, $\theta \in \Theta$ is a parameter vector, and $\lambda : \mathbb{R}_+ \times \Theta \to \mathbb{R}_+$. We consider all parametric families that are subsets of the class of admissible regular demand functions defined in Section 4, i.e., $\mathcal{L}(\Theta) \subset \mathcal{L} := \mathcal{L}(M, \underline{K}, \overline{K}, m)$.

In the current setting, we will assume that the decision maker knows that $\lambda \in \mathcal{L}(\Theta)$, i.e., s/he knows the parametric structure of the demand function, but *does not know* the value of the parameter vector θ. We continue to denote by $\mathcal{P}$ the set of admissible pricing policies, i.e., policies that satisfy (2)-(3). For any policy $\pi \in \mathcal{P}$, let $J^\pi(x, T; \theta)$ denote the expected revenues under π, and let $J^D(x, T|\theta)$ denote the value of the deterministic relaxation (5) when the value of the unknown parameter vector is revealed to the decision maker prior to the start of the selling season. Let $\mathcal{R}^\pi(x, T; \theta)$ denote the regret under a policy π, namely

$$\mathcal{R}^\pi(x, T; \theta) = 1 - \frac{J^\pi(x, T; \theta)}{J^D(x, T|\theta)}. \tag{16}$$

5.1 The proposed method

We consider a simple modification of the approach taken in Algorithm 1 that now exploits the assumed parametric structure of the demand function.

Algorithm 2: $\pi(\tau)$

Step 1. Initialization:

    (a) Set the learning interval to be $[0, \tau]$. Put $\Delta = \tau/k$, where k is the dimension of the parameter space Θ.

    (b) Choose a set of k prices, $\{p_i, i = 1, ..., k\}$.

Step 2. Learning/experimentation:

    (a) On the interval $[0, \tau]$ apply p_i from $t_{i-1} = (i-1)\Delta$ to $t_i = i\Delta$, $i = 1, ..., k$, as long as the inventory is positive. If no more units are in stock, apply p_∞ up until time T and STOP.

    (b) Compute
$$\hat{d}_i = \frac{\text{total demand over } [t_{i-1}, t_i)}{\Delta}, \qquad i = 1, ..., k.$$

    (c) Let $\hat{\theta}$ be a solution of $\{\lambda(p_i; \theta) = \hat{d}_i, \ i = 1, \ldots, k\}$

Step 3. Optimization:

$$
\begin{aligned}
p^u(\hat{\theta}) &= \arg\max\{p\lambda(p; \hat{\theta}) : p \in [\underline{p}, \overline{p}]\}, \\
p^c(\hat{\theta}) &= \arg\min\{|\lambda(p; \hat{\theta}) - x/T| : p \in [\underline{p}, \overline{p}]\}, \\
\text{set} \qquad \hat{p} &= \max\{p^u(\hat{\theta}), p^c(\hat{\theta})\}.
\end{aligned}
$$

Step 4. Pricing:
On the interval $(\tau, T]$ apply $\hat{p}$ as long as inventory is positive, then apply p_∞ for the remaining time.

Note that Step 1(b) requires one to define the prices $\{p_1, ..., p_k\}$ and Step 2(c) implicitly assumes that the system of equations admits a solution. (We define the prices and state this assumption more formally when analyzing the performance of Algorithm 2.) The intuition behind this algorithm is similar to the one that underlies the construction of Algorithm 1, the only difference being that the parametric structure allows one to infer accurate information about the demand function using only a "small" number of "test" prices (k). In particular, recalling the discussion following Algorithm 1, a key difference is that the deterministic error source associated with price granularity does not affect the performance of Algorithm 2.

5.2 The parametric class of demand functions

First note that the inclusion $\mathcal{L}(\Theta) \subset \mathcal{L}$ implies that $\lambda(\cdot; \theta)$ has an inverse for all $\theta \in \Theta$; this inverse will be denoted $\gamma(\cdot; \theta)$. (Note also that conditions (i.)-(iii.) in Assumption 1 hold.) For any parameter vector $\theta \in \Theta$ we denote the revenue function by $r(l; \theta) := \gamma(l; \theta)l$. Let $\{P_\theta : \theta \in \Theta\}$ be the family of demand distributions corresponding to Poisson processes with controlled intensities $\lambda(\cdot; \theta)$, $\theta \in \Theta$. We denote by $f_{\lambda(p;\theta)}(\cdot)$ the probability mass function of a Poisson random variable with intensity $\lambda(p; \theta)$.

The following technical conditions articulate standard regularity assumptions in the context of Maximum Likelihood estimation (cf. Borovkov (1998)), and are used to define the admissible class of parametric demand functions.

Assumption 2

(i.) There exists a vector of distinct prices $\vec{p} = (p_1, ..., p_k) \in [\underline{p}, \overline{p}]$ such that:

    a) For some $l_0 > 0$, $\min_{1 \leq i \leq k} \inf_{\theta \in \Theta} \lambda(p_i; \theta) > l_0$.

    b) For any vector $\vec{d} = (d_1, \ldots, d_k)$, the system of equations $\{\lambda(p_i; \theta) = d_i, \, i = 1, \ldots, k\}$ has a unique solution in θ. Let $g(\vec{p}, \vec{d})$ denote this solution. We assume in addition that $g(\vec{p}, \cdot)$ is Lipschitz continuous with constant $\alpha > 0$.

    c) For $i = 1, ..., k$, $\sqrt{\lambda(p_i; \theta)}$ is differentiable on Θ.

(ii.) For some $\overline{K}_2 > 0$, $|\lambda(p; \theta) - \lambda(p; \theta')| \leq \overline{K}_2 \|\theta - \theta'\|_\infty$ for all $p \in [\underline{p}, \overline{p}]$ and $\theta, \theta' \in \Theta$.

Condition (i.) ensures that the parametric model is identifiable based on a sufficient set of observations. Condition (ii.) is a mild regularity assumption on the parametric class, controlling for changes in the demand curve as parameters vary. We provide below an example of a parametric class that satisfies Assumption 2.

Example 1 (Linear demand function) Let $\lambda(p; \theta) = \theta_1 - \theta_2 p$, and set $[\underline{p}, \overline{p}] = [1, 2]$. Let $\mathcal{L}(\Theta) = \{\lambda(\cdot; \theta) : \theta \in [10, 20] \times [1, 4]\}$. If one sets $(p_1, p_2) = (1, 2)$, it is straightforward to verify conditions (i.)a), (i.)c) and that (ii.) is satisfied with $\overline{K}_2 = 1$. It is also easy to see that the unique solution associated with condition (i.)b) is given by

$$g(p_1, p_2, d_1, d_2) = \left(d_1 + \frac{p_1}{p_2 - p_1} [d_1 - d_2] \, , \, \frac{1}{p_2 - p_1} [d_1 - d_2] \right)$$

and $g(p_1, p_2, d_1, d_2)$ is clearly Lipschitz continuous with respect to (d_1, d_2). ∎

5.3 Performance analysis

Suppose that the initial inventory level and the demand function are scaled according to (11), and denote the regret by $\mathcal{R}_n^\pi(x, T; \theta) = 1 - J_n^\pi(x, T; \theta) / J_n^D(x, T | \theta)$.

Proposition 3 *Let Assumptions 1 and 2 hold and let $\{p_1, ..., p_k\}$ be as in Assumption 2(i.). Set $\tau_n \asymp n^{-1/3}$ and let $\pi_n := \pi(\tau_n)$ be defined by Algorithm 2. Then the sequence of policies $\{\pi_n\}$ is asymptotically optimal and satisfies*

$$\sup_{\theta \in \Theta} \mathcal{R}_n^\pi(x, T; \theta) \leq \frac{C(\log n)^{1/2}}{n^{1/3}}, \tag{17}$$

for all $n \geq 1$ and some finite positive constant C.

Contrasting the above with Proposition 1, we observe that the parametric structure of the demand function translates into improved performance bounds for the learning and pricing algorithm that is designed with this knowledge in mind. In particular, the regret is now of order $\mathcal{R}_n = O((\log n)^{1/2} n^{-1/3})$ as opposed to $\mathcal{R}_n = O((\log n)^{1/2} n^{-1/4})$ in the nonparametric case. In addition, note that the upper bound above holds for *all* admissible parametric families. The improvement in terms of generated revenues, as quantified by the smaller magnitude of the regret, spells out the advantages of using a parametric approach versus a nonparametric one. The downside, namely, model misspecification and its consequences, is illustrated and discussed in Section 6.

At an intuitive level, classical estimation theory tells us that parameter uncertainty cannot be resolved faster than rate $n^{-1/2}$ with n observations. This suggests that for the asymptotic regime we consider in this paper, no admissible pricing policy would be able to achieve a convergence rate faster than $n^{-1/2}$. This intuition is made rigorous in the next proposition.

Proposition 4 *Let Assumption 1 hold with $M, \overline{K}$ satisfying $M \geq \max\{2\overline{K}\overline{p}, \overline{K}\overline{p} + x/T\}$. Then, there exists a parametric family $\mathcal{L}(\Theta) \subset \mathcal{L}(M, \underline{K}, \overline{K}, m)$ satisfying Assumption 2 such that for some positive constant C' and for all admissible policies $\{\pi_n\}$*

$$\sup_{\theta \in \Theta} \mathcal{R}_n^\pi(x, T; \theta) \geq \frac{C'}{n^{1/2}}, \tag{18}$$

for all $n \geq 1$.

This result follows in a relatively straightforward manner from Proposition 2.

5.4 An optimal algorithm when a single parameter is unknown

Given the lower bound in (18), the natural question that arises is whether it is possible to close the remaining gap with respect to the upper bound in Proposition 3. We study this question in the context of a single unknown parameter, i.e., $k = 1$ and hence $\Theta \subseteq \mathbb{R}$. Consider the following ℓ-step policy $\pi(\ell, \Delta^{(1)}, \ldots, \Delta^{(\ell)})$.

Algorithm 3: $\pi(\ell, \Delta^{(1)}, \ldots, \Delta^{(\ell)})$

Step 1. Initialization:

    (a) Set the number of steps to be ℓ and define $\Delta^{(i)}$, $i = 1, \ldots, \ell$ so that $\Delta^{(1)} + \ldots + \Delta^{(\ell)} = T$.

    (b) Choose a price $\hat{p}_1 \in [\underline{p}, \overline{p}]$ as in Assumption 2(i.).

Step 2. Learning/Optimization/Pricing:
Set $t_1 = 0$.

For $i = 1, \ldots, \ell$,

    (a) **Learning/Pricing:**

        i.) Apply $\hat{p}_i$ on the interval $[t_i, t_i + \Delta^{(i)})$ as long as the inventory is positive. If no more units are in stock, apply p_∞ up until time T and STOP.

        ii.) Compute

$$\hat{d}_i = \frac{\text{Total demand over } [t_i, t_i + \Delta^{(i)})}{\Delta^{(i)}}.$$

        iii.) Set $t_{i+1} = t_i + \Delta^{(i)}$.

        iv.) Let $\hat{\theta}^i$ be the unique solution of $\lambda(p_i, \theta) = \hat{d}_i$.

    (b) **Optimization:**

$$\begin{aligned}
p^u(\hat{\theta}^i) &= \arg\max\{p\lambda(p; \hat{\theta}^i) : p \in [\underline{p}, \overline{p}]\}, \\
p^c(\hat{\theta}^i) &= \arg\min\{|\lambda(p; \hat{\theta}^i) - x/t| : p \in [\underline{p}, \overline{p}]\}, \\
\hat{p}_{i+1} &= \max\{p^u(\hat{\theta}^i), p^c(\hat{\theta}^i)\}.
\end{aligned}$$

End For

The intuition underlying Algorithm 3 is as follows. When a single parameter is unknown, one can infer information about the parameter from observations of demand at a single price. Given this, the idea of Algorithm 3 is to price "close" to p^D after the first stage, but to *continue* learning. In particular, the estimate of p^D is improved from stage to stage using the demand observations from the previous stage. As a result, losses are mitigated by two effects: i.) after stage 1, the price is always close to p^D; and ii.) the estimate of p^D becomes more precise. For what follows, we slightly strengthen Assumption 2(i.)

Assumption 3 $\inf_{p \in [\underline{p}, \overline{p}]} \inf_{\theta \in \Theta} \lambda(p; \theta) > l_0$ *and for any price $p \in [\underline{p}, \overline{p}]$ and any $d \geq 0$, the equation $\lambda(p; \cdot) = d$ has a unique solution. If $g(p, d)$ denotes this solution, then $g(p, \cdot)$ is Lipschitz continuous with constant $\alpha > 0$.*

We now analyze the performance of policies associated with Algorithm 3. In particular suppose that the initial inventory level and the demand function are scaled according to (11), and define the sequence of tuning parameters $\{\ell_n, \Delta_n^{(1)}, \ldots, \Delta_n^{(\ell_n)}\}$ as follows:

$$\ell_n = (\log 2)^{-1} \log \log n \tag{19}$$

$$\Delta_n^{(m)} = \beta_n n^{(a_{\ell_n}/a_m)-1}, \qquad m = 1, ..., \ell_n, \tag{20}$$

where $a_m = 2^{m-1}/(2^m - 1)$ for $m \geq 1$, and $\beta_n > 0$ is a normalizing constant chosen so $\Delta_n^{(1)} + ... + \Delta_n^{(\ell_n)} = T$. We then have the following result.

Proposition 5 *Let Assumptions 1,2 and 3 hold. Let $\{\ell_n, \Delta_n^{(1)}, \ldots, \Delta_n^{(\ell_n)}\}$ be defined as in (19)-(20) and put $\pi_n := \pi(\ell_n, \Delta_n^{(1)}, \ldots, \Delta_n^{(\ell_n)})$, defined by Algorithm 3. Then the sequence of policies $\{\pi_n\}$ is asymptotically optimal and satisfies*

$$\sup_{\theta \in \Theta} \mathcal{R}_n^\pi(x, T; \theta) = O\Big(\frac{(\log \log n)(\log n)^{1/2}}{n^{1/2}}\Big). \tag{21}$$

Note that the regret of the sequence for policies introduced in Proposition 5 achieves the lower bound spelled out in Proposition 4 (up to logarithmic terms). In that sense, these policies cannot be improved upon. On the other hand, Algorithm 3 is restricted to the case where only one parameter is unknown, and exploits the fact that in this setting it is possible to learn the single unknown parameter by conducting price experiments in the neighborhood of the near-optimal price p^D. Designing policies that achieve the lower bound in Proposition 4 in the multi-parameter case remains an open question.

6 Numerical Results and Qualitative Insights

6.1 Performance of the parametric/nonparametric policies and the "price" of uncertainty

We examine the performance of three polices developed in the previous sections: i.) The nonparametric policy defined in Algorithm 1; ii.) The parametric policy defined by means of Algorithm 2 and designed for are a finite number of unknown parameters; and iii.) The parametric policy defined in Algorithm 3 designed for cases with a single unknown parameter. (The tuning parameters are taken as in Propositions 1, 3 and 5, respectively.)

The performance of these policies are measured by the magnitude of the regret. Note that $\mathcal{R}_n^\pi(x, T; \lambda) \approx C/n^\gamma$ implies that $\log \mathcal{R}_n^\pi(x, T; \lambda)$ should be approximately linear in $\log n$ with slope $-\gamma$. In Figure 1, we depict $\mathcal{R}_n^\pi(x, T; \lambda)$ as a function of n in a log-log plot for large values of n, and compute the best (least squares) linear fit for these values. The results depicted are based on running 10^3 independent simulation replications from which the performance indicators were derived by averaging. The standard error for $\mathcal{R}_n^\pi(x, T; \lambda)$ was below 0.05% in all cases.

18

Figure 1(a) summarizes results for an underlying exponential demand model $\lambda(p) = \theta \exp(-.5p)$, where $\theta = 10 \exp(1)$, and Figure 1(b) presents results for an underlying linear demand model $\lambda(p) = 30 - \theta p$ where $\theta = 3$. The nonparametric algorithm does not make any assumptions with regard to the structure of the demand function (beyond those spelled out in Section 4), while the parametric algorithms are assumed to know the parametric structure but the true value of θ is not revealed to them. In both cases, the initial normalized inventory level was $x = 20$, the selling horizon was $T = 1$ and the feasible price set was $[\underline{p}, \overline{p}] = [0.1, 10]$.
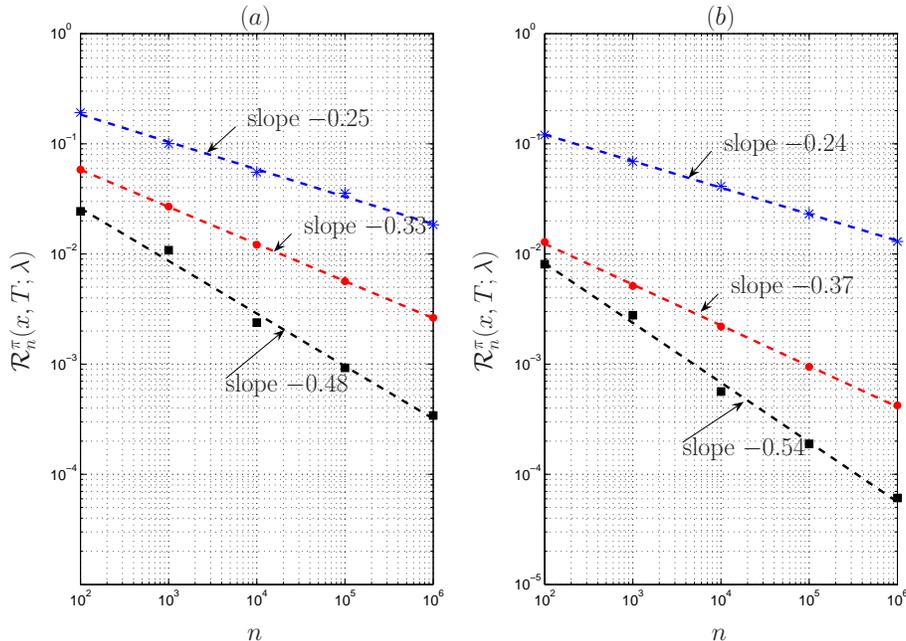
Figure 1: Performance of pricing policies as a function of the market size (n). Stars show the performance of the nonparametric policy defined in Algorithm 1; dots represent the performance of the parametric policy defined Algorithm in 2; and squares depict the performance of the parametric policy defined in Algorithm 3. The dashed lines represent the best linear fit to each set of points; in panel (a), the demand function is exponential and in panel (b) linear.

Discussion. The slopes of the best linear fit in Figures 1(a) and 1(b) are very close to $\gamma = -1/4$, $\gamma = -1/3$ and $\gamma = -1/2$, predicted by the upper bounds in (13), (17) and (21), respectively. These results provide a "picture proof" of the Propositions 1, 3 and 5. As is evident, the less structure is assumed a priori, the higher the profit loss relative to the full information benchmark: informally speaking, this is the "price" paid due to increasing uncertainty with regard to the demand model.

An additional fundamental difference between the various policies concerns the degree to which the price domain is explored. The nonparametric policy (Algorithm 1) essentially needs to explore the "entire" price domain; the general parametric policy (Algorithm 2) needs to test a number of prices equal to the number of unknown parameters; and the parametric policy designed for the case

19

of a single unknown parameter (Algorithm 3) explores only prices in the vicinity of the price p^D. In other words, as uncertainty "decreases," the price exploration region shrinks as well.

6.2 The "price" of hedging misspecification risk

We have seen in the previous section that more refined information regarding the demand model yields higher revenues. Thus, it becomes tempting to postulate parametric structure, which in turn may be incorrect relative to the *true* underlying demand model. This misspecification risk can be eliminated via nonparametric pricing policies, but at the price of settling for more modest performance revenue-wise. We now provide an illustration of this trade-off.

We fix the time horizon to be $T = 1$ and the set of feasible prices to be $[\underline{p}, \overline{p}] = [0.1, 10]$. Figure 2 depicts the regret $\mathcal{R}_n^\pi(x, T; \lambda)$ for two underlying demand models, various values of n, and three policies. The first policy assumes an exponential parametric structure for the demand function, $\lambda(p; \theta) = \theta_1 \exp(-\theta_2 p)$ with $\theta = (\theta_1, \theta_2)$, and applies Algorithm 2 with $\tau_n = n^{-1/3}$. The second policy assumes a linear parametric structure $\lambda(p; \theta) = (\theta_1 - \theta_2 p)^+$ and applies Algorithm 2 with $\tau_n = n^{-1/3}$. Finally, the third policy is the one given by the nonparametric method described in Algorithm 1 (with tuning parameters as in Proposition 1). The point of this comparison is to illustrate that the parametric algorithm outperforms its nonparametric counterpart when the assumed parametric model is consistent with the true demand function, otherwise the parametric algorithm leads to a regret that does not converge to zero due to a model misspecification error.

In Figures 2(a),(b) the underlying demand model is given by $\lambda(p) = a \exp(-\alpha p)$ with $a = 10 \exp(1)$ and $\alpha = 1$, and hence the policy that assumes the exponential structure (squares) is *well specified*, while the policy that assumes a linear demand function (crosses) suffers from *model misspecification*.

In Figures 2(c),(d) the underlying demand model is given by $\lambda(p) = (a - \alpha p)^+$ with $a = 30$ and $\alpha = 3$. Now the policy that assumes the exponential structure (squares) corresponds to a *misspecified* case, while the policy that assumes a linear demand function (crosses) corresponds to a *well specified* case.

In Figures 2(a),(c) the normalized inventory is $x = 8$ and in Figures 2(b),(d) it is taken to be $x = 20$. The results depicted in the figures are based on running 10^3 independent simulation replications from which the performance indicators were derived by averaging. The standard error was below 0.9% in all cases.

Discussion. We focus on Figures 2(a),(b) as similar remarks apply to Figures 2(c),(d). First, note that the parametric algorithm is able to achieve close to 90% of the full information revenues when the parametric model is well specified, when the market size is $n = 100$ or more. In addition, the convergence of the regret for Algorithm 2 is faster for small n under the well specified parametric assumption (squares) in comparison to the nonparametric algorithm (stars). [We note that the non-
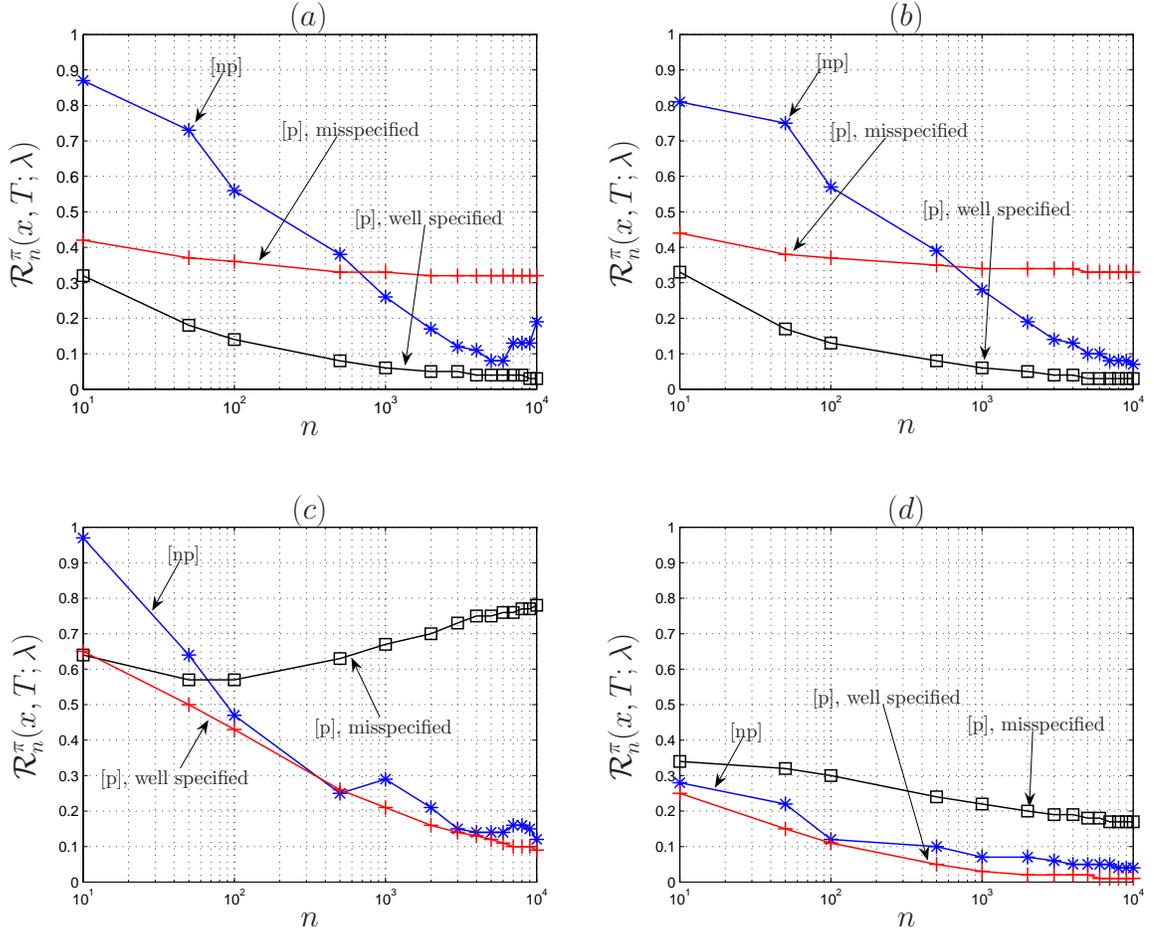
Figure 2: Performance of the parametric (Algorithm 2) and nonparametric policies (Algorithm 1) as a function of the market size: the parametric algorithm that assumes a linear model (crosses) and exponential model (squares); and the nonparametric algorithm (stars). The underlying demand model is exponential in Figures (a) and (b) and linear in Figures (c) and (d). Arrows indicate if the approach used is nonparametric [np] or parametric [p], and whether the model is well/misspecified in the case of the parametric algorithm.

monotonic performance of the nonparametric algorithm (stars) stems from the fact that the price grid used by this algorithm changes with n, and consequently the minimal distance of the sought fixed price to any price on the grid need not be monotonic with respect to n.]

In contrast, if the model is misspecified, the regret $\mathcal{R}_n^\pi(x, T; \lambda)$ converges to a strictly positive value (crosses), as expected, and hence the performance of the parametric algorithm fails to asymptotically achieve the full information revenues. The takeaway message here is that a nonparametric approach eliminates the risk stemming from model misspecification but at a price of extracting lower

revenues than its parametric counterpart. The regret bounds derived in earlier sections precisely spell out the magnitude of this "price" of misspecification.

6.3 Numerical analysis of the bounds in Propositions 1 and 3

The results of Propositions 1 and 3 derive upper bounds on the regret that involve constants whose values have not been specified. Ideally, one would like to characterize the smallest possible constants for which the upper bounds hold, in order to determine the non-asymptotic behavior of the policies. While an exact analysis along these lines is essentially intractable, it is possible to investigate this point numerically and assess the "typical" magnitude of these constants.

Fix $T = 1$ and $[\underline{p}, \overline{p}] = [5, 10]$, and consider two families of demand functions: the first corresponds to an exponential class $\mathcal{L}_1 = \{a \exp\{-\alpha p\} : a \in [\underline{a}, \overline{a}], \alpha \in [\underline{\alpha}, \overline{\alpha}]\}$ where $[\underline{a}, \overline{a}] = [5, 10]$ and $[\underline{\alpha}, \overline{\alpha}] = [0.1, 0.2]$; the second corresponds to a linear class $\mathcal{L}_2 = \{(b - \beta p)^+ : b \in [\underline{b}, \overline{b}], \beta \in [\underline{\beta}, \overline{\beta}]\}$ where $[\underline{b}, \overline{b}] = [10, 20]$ and $[\underline{\beta}, \overline{\beta}] = [0.2, 1]$. Note that all of the functions in the union of $\mathcal{L}_1$ and $\mathcal{L}_2$ are bounded by $M = \max\{\overline{a} \exp\{-\underline{\alpha}\underline{p}\}, \overline{b} - \underline{\beta}\underline{p}\} = 19$, are $\overline{K}$-Lipschitz with $\overline{K} = \max\{\overline{a}\overline{\alpha} \exp\{-\underline{\alpha}\underline{p}\}, \overline{\beta}\} = 1.21$, and have a first derivative bounded below by $\underline{K} = \min\{\underline{a}\underline{\alpha} \exp\{-\overline{\alpha}\overline{p}\}, \underline{\beta}\} = 0.067$. Similarly, the maximum revenue rate is bounded below by $m = \min\{\underline{p}\underline{a} \exp\{-\overline{\alpha}\overline{p}\}, \underline{p}(\underline{b} - \overline{\beta}\underline{p})\} = 3.38$.

To estimate the constants characterizing the revenue losses of the parametric and nonparametric policies we proceed as follows. We draw 100 demand functions from each class of demand models above, using a uniform distribution over each interval defining the space of parameters. For each function, we estimate the regret associated with the proposed nonparametric (Algorithm 1) and parametric (Algorithm 2) policies for market sizes of $n = 100$, $n = 10^3$ and $n = 10^4$, and for two initial (normalized) inventory levels ($x = 5, x = 10$). This step is executed as in the previous subsections. For each draw and each market size, we compute the value of $\mathcal{R}_n^\pi(x, T; \lambda)n^\gamma$, where $\gamma = 1/4$ for the nonparametric policy, and $\gamma = 1/3$ for the parametric one. Taking the maximum of these values over all 100 draws, we arrive at an estimate, C_{max}, for each market size, which corresponds to the worst case constant observed for all models. This value serves as a proxy for the magnitude of the constants characterizing the performance of the proposed algorithms. Results are summarized in Table 1.

The reasonable magnitude of the constants (all fall in the range of 0.5-1.3) suggests that our policies should perform well in practical settings. In particular, based on these results, we expect the nonparametric policy to achieve at most a regret of 35% for $n = 100$, 23% for $n = 10^3$ and 14% for $n = 10^4$, while the parametric policy should achieve a regret of 24% for $n = 100$, 12% for

demand functions	$\mathcal{L}_1$			$\mathcal{L}_2$		
	C_{max}			C_{max}		
market size (n)	10^2	10^3	10^4	10^2	10^3	10^4
$x = 5,$ $\quad \pi_{np}$	0.92	1.04	1.07	1.09	1.25	1.31
π_p	0.79	0.79	0.79	1.11	1.11	1.11
$x = 10,$ $\quad \pi_{np}$	0.66	0.75	0.77	0.74	0.85	0.87
π_p	0.61	0.53	0.39	0.58	0.60	0.54

Table 1: Estimates of the magnitude of the constants characterizing the performance of the algorithms: π_{np} corresponds to the nonparametric policy; and π_p to the parametric one.

$n = 10^3$ and 6% for $n = 10^4$.

7 Sequential Learning Strategies in the Nonparametric Setting

The performance guarantees associated with Algorithm 1 rely heavily on how well one is able to estimate the two prices p^u and p^c arising from the solution of the deterministic relaxation (5). In Algorithm 1, this estimation is executed in a single stage over the interval $[0, \tau]$. A natural extension would be to consider policies in which the learning phase searches for the optimal price in a sequential and dynamic manner.

Recall the definitions of p^u and p^c, the unconstrained and constrained solutions of the deterministic relaxation: $p^u = \arg\max_{p \in [\underline{p}, \overline{p}]} \{r(\lambda(p))\}$, $p^c = \arg\min_{p \in [\underline{p}, \overline{p}]} |\lambda(p) - x/T|$. Suppose for simplicity that the two critical prices belong to the interior of the feasible price set $[\underline{p}, \overline{p}]$. Then p^u is the unique maximizer of the revenue function and p^c is the x/T-crossing of the demand function, which is assumed to be decreasing. Using this interpretation, one could use stochastic approximation schemes to search for the two prices rather than via the estimation of the demand function over the whole price domain (as in Algorithm 1). We detail below how such schemes could be designed.

Background on stochastic approximations. Suppose one wants to estimate the maximizer z^* of a real valued function $M(z)$ on a domain $[\underline{z}, \overline{z}]$, and that the available observations are in the form $H(z) = M(z) + \epsilon(z)$, where $\epsilon(z)$ is a stochastic noise term with $\mathbb{E}[\epsilon(z)] = 0$ and $\mathbb{E}[\epsilon^2(z)] < \infty$ for all $z \in [\underline{z}, \overline{z}]$. The basic Kiefer-Wolfowitz (KW) scheme estimates z^* using the recursion

$$z_{k+1} = z_k + a_k \frac{y_{2k} - y_{2k-1}}{c_k}, \qquad k = 1, 2, \ldots, \tag{22}$$

where $y_{2k} = H(z_k + c_k)$ and $y_{2k-1} = H(z_k - c_k)$, and $\{a_k\}$ and $\{c_k\}$ are predefined deterministic real-valued positive sequences. Under suitable technical conditions on the function $M(\cdot)$ and the sequences $\{a_k\}$ and $\{c_k\}$, Kiefer and Wolfowitz (1952) showed that the sequence $z_k \to z^*$ in probability as $k \to \infty$. Similar ideas lead to a sequence $\{z_k\}$ that converges to an α-crossing $\{z : M(z) = \alpha\}$ via the Robbins-Monro (RM) procedure (see Robbins and Monro (1951)), viz.

$$z_{k+1} = z_k + a_k(\alpha - y_k), \tag{23}$$

where $y_k = H(z_k)$. For further details and other variants the reader is referred to Benveniste, Métivier, and Priouret (1990).

A stochastic approximation-based algorithm. A straightforward way to blend adaptive search procedures into Algorithm 1 is given below. Here $h(\cdot)$ denotes the projection operator $h(y) := \min\{\max\{y, \underline{p}\}, \bar{p}\}$.

Algorithm 4 : $\quad \pi^{SA}(\tau, \Delta, \{a_k\}, \{c_k\})$

Step 1. Initialization:

Set the learning interval to be $[0, \tau]$, and set Δ to be the holding time for each price during that period.

Step 2. Learning/experimentation:

(a) **KW-type scheme:**

Set $\kappa_1 = \left\lfloor \tau/(4\Delta) \right\rfloor$, $t_1 = 0$, $z_1 = \underline{p}$ and $t_i = t_1 + (i-1)\Delta, \quad i = 2, \ldots$

For $k = 1, \ldots, \kappa_1$

i.) Apply $p_k = \min\{z_k + c_k, \bar{p}\}$ on $[t_{2k-2}, t_{2k-1})$ and compute

$$y_{2k} = \left(\text{total demand over } [t_{2k-2}, t_{2k-1}) \right)/\Delta$$

ii.) Apply $p_k = \max\{z_k - c_k, \underline{p}\}$ on $[t_{2k-1}, t_{2k})$ and compute

$$y_{2k-1} = \left(\text{total demand over } [t_{2k-1}, t_{2k}) \right)/\Delta$$

iii.) Update: $z_{k+1} = h\left(z_k + (a_k/c_k)(y_{2k} - y_{2k-1}) \right)$

end For

Set $\hat{p}^u = z_{\kappa_1 + 1}$

(b) **RM-type scheme:**

Set $\kappa_2 = \left\lfloor \tau/(2\Delta) \right\rfloor$, $t_1 = 2\kappa_1 \Delta$, $z_1 = \underline{p}$ and $t_i = t_1 + (i-1)\Delta, \quad i = 2, \ldots$

For $k = 1, \ldots, \kappa_2$

i.) Apply $p_k = \min\{z_k, \bar{p}\}$ on $[t_k, t_{k+1})$ and compute

$$y_k = \left(\text{total demand over } [t_k, t_{k+1}) \right)/\Delta$$

ii.) Update: $z_{k+1} = h(z_k + a_k(x/T - y_k))$

end For

Set $\hat{p}^c = z_{\kappa_2 + 1}$

Step 3. Optimization: Set $\hat{p} = \max\{\hat{p}^u, \hat{p}^c\}$.

Step 4. Pricing: Set $\tau' = (2\kappa_1 + \kappa_2)\Delta$. On the interval $(\tau', T]$ apply $\hat{p}$ as long as inventory is positive, then apply p_∞ for the remaining time.

Comments on the structure of the algorithm. The algorithm exploits the structure of the single product problem, adaptively estimating each of the two prices p^u and p^c based on its characterization as either a point of maximum or a level crossing; the former is executed in Step 2(a) and the latter in the step 2(b). This also requires two consecutive price testing phases to be carried out. Note that in Step 2 above, the number of iterations is different for the KW and RM procedures. This is due to the fact that a KW-based procedure uses two prices at each iteration in order to estimate an improvement direction. (For brevity, in Step 2 it was left implicit that as soon as inventory is depleted, the shut-off price p_∞ is applied.)

Numerical results. We present a comparison between the performance of the policy defined by Algorithm 1 with tuning parameters as in Proposition 1, and the policy given by Algorithm 4 that uses stochastic approximations. We use π_1^{SA} to denote the policy which uses Algorithm 4 taking $\tau_n = n^{-1/4}$ and $\Delta_n = n^{-1}$; this corresponds to the KW-scheme using $\left\lfloor (1/4)n^{3/4} \right\rfloor$ steps. Similarly, we let π_2^{SA} denote the policy which corresponds to taking $\tau_n = n^{-1/4}$ and $\Delta_n = n^{-1/2}$. (The choice of τ_n is driven by our previous asymptotic considerations and is meant only for illustrative purposes.) The key difference between π_1^{SA} and π_2^{SA} is that in the latter there are fewer test prices being used, and the observations for a given price are less noisy since each price is held for a longer time interval.

Table 2 presents results for the three policies discussed above and for two normalized initial inventory levels (x); if the problem size is n and the total inventory is y then $x = y/n$. For each policy we record the regret $\mathcal{R}_n^\pi(x, T; \lambda)$, and the number of prices κ that are used during the interval $[0, T]$. To generate the results in Table 2 we take the underlying demand function to be linear, $\lambda(p) = (10 - 2p)^+$ and the set of feasible prices is taken to be $[\underline{p}, \overline{p}] = [0.1, 4.5]$. In the first case where $x = 3$, and hence $p^c = 3.5$ and $p^u = 2.5$, the constrained price p^c is optimal in the deterministic relaxation (5). In the second case, there is a larger initial normalized inventory $x = 8$ and the prices of interest are given by $p^c = 1$ and $p^u = 2.5$, hence the unconstrained price is optimal in the deterministic relaxation.

The results depicted in the table are based on running 500 independent simulation replications from which the performance indicators were derived by averaging. The standard error for $\mathcal{R}_n^\pi(x, T; \lambda)$ was below 0.8% in all cases. (We note that numerical experiments with other simple demand models give rise to similar results and therefore, for space considerations, we only report on the

Problem "size"	$n = 10^2$		$n = 10^3$		$n = 10^4$	
	$\mathcal{R}_n^\pi(x,T;\lambda)$	κ	$\mathcal{R}_n^\pi(x,T;\lambda)$	κ	$\mathcal{R}_n^\pi(x,T;\lambda)$	κ
$x = 3$, π	.44	5	.19	7	.12	12
π_1^{SA}	.25	33	.12	180	.05	1002
π_2^{SA}	.27	12	.12	18	.05	33
$x = 8$, π	.86	5	.08	7	.04	12
π_1^{SA}	.28	33	.17	180	.10	1002
π_2^{SA}	.22	12	.08	18	.01	33

Table 2: Performance of stochastic approximation based schemes. $\mathcal{R}_n^\pi(x,T;\lambda)$ represents the regret associated with a given policy, and x is the normalized inventory level. Here π is the basic non-parametric pricing policy defined by Algorithm 1, π_1^{SA} and π_2^{SA} correspond to two implementations of Algorithm 4, and $\kappa =$ number of prices used by a given policy.

linear demand example.)

Discussion. A quick inspection of the results in Table 2 leads to the following observations. First, all three policies are seen to have comparable performance achieving a regret which is smaller than 20% for problems where the inventory is of the order of a thousand; this translates to excess of 80% of the optimal full information revenues. It is also interesting to contrast the structure of the two adaptive policies π_1^{SA} and π_2^{SA}. We observe that the former uses significantly more prices than the latter (180 prices for a problem with $n = 10^3$, compared with 18 for π_2^{SA}) and hence is not very practical to implement. Considering that the performance of π_1^{SA} and π_2^{SA} is comaprable, it seems that choosing the length Δ so that observations are less noisy is potentially pereferable, even though less price changes are allowed. The nonparametric policy π uses less than half of the price changes that π_2^{SA} utilizes, and is overall the simplest to implement.

A detailed performance analysis of Algorithm 4 is beyond the scope of the paper. However, based on convergence rate results associated with stochastic approximation schemes (see Polyak and Tsybakov (1990) and Polyak and Juditsky (1990)), we expect that the policies associated with Algorithm 4 would achieve at best a regret of order $\mathcal{R}_n \approx n^{-1/3}$.

In Algorithm 4, we use a KW scheme that seeks an estimate of p^u (the unconstrained maximizer of $r(\cdot)$) and an RM-based scheme that seeks an estimate of p^c (the inventory-constrained "run out" price). Only after these estimates are obtained, by running each stochastic approximation procedure separately, can one attempt to deduce which of the two prices is "near-optimal" for the given problem. An interesting direction is to investigate methods for finding an estimate of the optimal price of the deterministic relaxation using a single combined procedure. Stochastic approximations for constrained problems exist (see, e.g., Kushner and Sanvicente (1975)), however, to the best of our knowledge there are still open questions with regard to their performance and rates of convergence.

A   Proofs of Main Results

Preliminaries and notation. For any real number x, x^+ will denote $\max\{x, 0\}$. $C_1, C_2, \ldots$ will be used to denote positive constants which are independent of a given demand function, but may depend on the parameters of the class $\mathcal{L}$ of admissible demand functions and on x and T. A sequence $\{a_n\}$ of real numbers is said to increase to infinity at a *polynomial rate* if there exist a constant $\beta > 0$ such that $\liminf_{n\to\infty} a_n/n^\beta > 0$. To lighten the notation, we will occasionally omit the arguments of $J_n^\pi(x, T; \lambda)$ and $J_n^D(x, T|\lambda)$. In what follows we will make use of the following facts.

Fact 1. Recall the definition of $J^D(x, T|\lambda)$, the optimal value of the deterministic relaxation (5). First note that $J_n^D = nJ^D$. We will also use the fact that

$$\inf_{\lambda \in \mathcal{L}} J^D(x, T|\lambda) \geq m^D,$$

where $m^D = mT' > 0$ and $T' = \min\{T, x/M\}$. Indeed, for any $\lambda \in \mathcal{L}$, there is a price $q \in [\underline{p}, \overline{p}]$ such that $r(q) \geq m$. Consider the policy that applies q on $[0, T']$ and then applies p_∞ up until T. This solution is feasible since $\lambda(q)T' \leq MT' \leq x$. In addition the revenues generated from this policy are given by mT'.

We provide below a result that generalizes Gallego and van Ryzin (1994, Proposition 2) and that is used throughout the proofs of the main results. Its proof can be found in Appendix B.

Lemma 1 *The solution to problem (5) is given by $p(s) = p^D := \max\{p^u, p^c\}$ for $s \in [0, T']$ and $p(s) = p_\infty$ for $s > T'$, where $p^u = \arg\max_{p \in [\underline{p}, \overline{p}]}\{r(\lambda(p))\}$, $p^c = \arg\min_{p \in [\underline{p}, \overline{p}]}|\lambda(p) - x/T|$ and $T' = \min\{T, x/\lambda(p^D)\}$. In addition, for all $\lambda \in \mathcal{L}$, the optimal value of (5), $J^D(x, T|\lambda)$, serves as an upper bound on $J^\pi(x, T; \lambda)$ for all $\pi \in \mathcal{P}$.*

Before proceeding, we state the following lemma which is needed throughout the analysis and whose proof is deferred to Appendix B.

Lemma 2 *Suppose that $\mu \in [0, M]$ and $r_n \geq n^\beta$ with $\beta > 0$. If $\epsilon_n = 2\eta^{1/2}M^{1/2}(\log n)^{1/2}r_n^{-1/2}$, then for all $n \geq 1$*

$$\mathbb{P}\Big(N(\mu r_n) - \mu r_n > r_n\epsilon_n\Big) \leq \frac{C}{n^\eta},$$

$$\mathbb{P}\Big(N(\mu r_n) - \mu r_n < -r_n\epsilon_n\Big) \leq \frac{C}{n^\eta},$$

for some suitably chosen constant $C > 0$.

This lemma bounds the deviations of a Poisson process from its mean. It will be used to control for the estimates of the demand function evaluated at given prices.

A.1 Proofs of the results in Section 4

Proof of Proposition 1. Fix $\lambda \in \mathcal{L}$, $\eta = 2$. We consider the sequence of policies $\pi_n := \pi(\tau_n, \kappa_n)$, $n = 1, 2, \ldots$ defined by means of Algorithm 1. The proof is organized in four steps. The first step develops an expression for a lower bound on the expected revenues achieved by the proposed policy. This lower bound highlights the terms that need to be analyzed. The second step provides probabilistic bounds on the estimate $\hat{p}$ of the price p^D and the associated revenue rate. This is done by controlling the deviations of a Poisson process from its mean (see Lemma 3). The third step finalizes the analysis of the lower bound. For this step, two cases have to be considered separately depending on the starting level of inventory. The key issue here is to control for stochastic fluctuations of customer requests at the estimated price $\hat{p}$. In the last step, we conclude the proof by combing the results from the two cases and plugging in the tuning parameters.

Let τ_n be such that $\tau_n \to 0$ and $n\tau_n \to \infty$ at a polynomial rate as $n \to \infty$. Let κ_n be a sequence of integers such that $\kappa_n \to \infty$ and $n\Delta_n := n\tau_n/\kappa_n \to \infty$ at a polynomial rate. We divide the interval $[\underline{p}, \overline{p}]$ into κ_n equal length intervals and we let $P_n = \{p_i, i = 1, ..., \kappa_n\}$ be the left endpoints of these intervals. Now partition $[0, \tau_n]$ into κ_n intervals of length Δ_n and apply the price p_i on the ith interval. Define

$$\widehat{\lambda}(p_i) = \frac{N\left(\sum_{j=1}^{i} n\lambda(p_j)\Delta_n\right) - N\left(\sum_{j=1}^{i-1} n\lambda(p_j)\Delta_n\right)}{n\Delta_n}, \qquad i = 1, ..., \kappa_n, \tag{A-1}$$

where $N(\cdot)$ is a unit rate Poisson process. Thus $\widehat{\lambda}(p_i)$ denotes the number of product requests over successive intervals of length Δ_n, normalized by $n\Delta_n$.

Step 1. Here, we derive a lower bound on the expected revenues under the policy π_n. Let $X_n^{(L)} = \sum_{i=1}^{\kappa_n} \lambda(p_i)n\Delta_n$, $X_n^{(P)} = \lambda(\hat{p})n(T - \tau_n)$ and put $Y_n = N(X_n^{(L)} + X_n^{(P)})$, $Y_n^{(L)} = N(X_n^{(L)})$ and $Y_n^{(P)} = Y_n - Y_n^{(L)}$. $Y_n^{(L)}$ represents the maximum number of requests during the learning phase if the system would not run out of resources, $Y_n^{(P)}$ the maximum number of requests during the pricing phase and Y_n the maximum total number of requests throughout the sales horizon. Now note that one can lower bound the revenues achieved by π_n by those accumulated during the pricing phase and during the latter, the maximum number of units that can be sold is exactly $\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\}$. We deduce that

$$J_n^\pi \geq \mathbb{E}\left[\hat{p}\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\}\right] \tag{A-2}$$

Next, we analyze the lower bound above by getting a handle on $\hat{p}$, the estimate of p^D and the quantities $Y_n^{(P)}$ and $Y_n^{(L)}$.

Step 2. Here, we analyze the estimate $\hat{p}$ through the estimates $\hat{p}^u$ and $\hat{p}^c$. Recall the definition of

$$\hat{p}^u = \arg\max_{1 \leq i \leq \kappa_n}\{p_i\widehat{\lambda}(p_i)\}, \quad \hat{p}^c = \arg\min_{1 \leq i \leq \kappa_n}|\widehat{\lambda}(p_i) - x/T|,$$

28

which are estimates of p^u and p^c, respectively. We define the following quantity

$$u_n = (\log n)^{1/2} \max\left\{1/\kappa_n, 1/(n\Delta_n)^{1/2}\right\}, \tag{A-3}$$

that will be used throughout the analysis to quantify deviations of various quantities associated with the proposed policy from their full information counterparts. We provide below a result that characterizes the revenue rate at $\hat{p} = \max\{\hat{p}^c, \hat{p}^u\}$ as well as $\hat{p}^c$. The proof of this lemma can be found in Appendix B.

Lemma 3 *For some positive constants C_1, C_2, C_3, for and all $n \geq 1$,*

$$\mathbb{P}\left\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_1 u_n\right\} \leq \frac{C_3}{n^{\eta-1}}, \tag{A-4}$$

$$\mathbb{P}\left\{|\hat{p}^c - p^c| > C_2 u_n\right\} \leq \frac{C_3}{n^{\eta-1}}. \tag{A-5}$$

Step 3. In what follows we will separate two cases: $\lambda(\bar{p}) \leq x/T$ and $\lambda(\bar{p}) > x/T$. The latter case is one where the decision-maker would on average run out of inventory during the sales horizon and has to be analyzed separately.

Case 1. Suppose first that $\lambda(\bar{p}) \leq x/T$. Note that $\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\} = Y_n^{(P)} - (Y_n - nx)^+$. Hence, using (A-2) and the fact that $\hat{p} \leq \bar{p}$, we have

$$\begin{aligned} J_n^{\pi} &\geq \mathbb{E}\left[\hat{p}Y_n^{(P)}\right] - \bar{p}\mathbb{E}\left[(Y_n - nx)^+\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\hat{p}N\left(\lambda(\hat{p})n(T - \tau_n)\right) \Big| \hat{p}\right]\right] - \bar{p}\mathbb{E}\left[(Y_n - nx)^+\right] \\ &= \mathbb{E}[r(\lambda(\hat{p}))]n(T - \tau_n) - \bar{p}\mathbb{E}\left[(Y_n - nx)^+\right]. \end{aligned} \tag{A-6}$$

Now,

$$\begin{aligned} \mathbb{E}[r(\lambda(\hat{p}))] &= r(\lambda(p^D)) \\ &\quad + \mathbb{E}\left[r(\lambda(\hat{p})) - r(\lambda(p^D)) \,\Big|\, r(\lambda(p^D)) - r(\lambda(\hat{p})) \leq C_1 u_n\right] \mathbb{P}\left\{r(\lambda(p^D)) - r(\lambda(\hat{p})) \leq C_1 u_n\right\} \\ &\quad + \mathbb{E}\left[r(\lambda(\hat{p})) - r(\lambda(p^D)) \,\Big|\, r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_1 u_n\right] \mathbb{P}\left\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_1 u_n\right\} \\ &\overset{(a)}{\geq} r(\lambda(p^D)) - C_1 u_n - \frac{C_4}{n^{\eta-1}}, \end{aligned} \tag{A-7}$$

where (a) follows from Lemma 3 and one can take $C_4 = 2\bar{p}MC_3$ (recall that $\lambda(\cdot)$ is bounded above by M under Assumption 1 (i) and hence $r(\cdot)$ is bounded above by $\bar{p}M$.) We now turn to analyze the second term on the RHS of (A-6). The following result gives a probabilistic bound on the deviations of the random variable Y_n from nx and its proof is deferred to Appendix B.

Lemma 4 *For some positive constants C_5, C_6, for and all $n \geq 1$,*

$$\mathbb{P}\left(Y_n - nx > C_5 n u_n\right) \leq \frac{C_6}{n^{\eta-1}} \tag{A-8}$$

Using (A-8), we have

$$
\begin{aligned}
\mathbb{E}\Big[(Y_n - nx)^+\Big] &\leq C_5 n u_n + \mathbb{E}\Big[(Y_n - nx)^+ \; ; \; Y_n - nx \geq C_5 n u_n\Big] \\
&\stackrel{(a)}{\leq} C_5 n u_n + (nx + C_5 n u_n + 1 + nM)\frac{C_6}{n^{\eta-1}} \\
&\leq C_7 n u_n, \tag{A-9}
\end{aligned}
$$

where C_7 is a suitably large positive constant, (a) follows from the fact that for a Poisson random variable Z with mean μ, $\mathbb{E}[Z \mid Z > a] \leq a + 1 + \mu$. We thus conclude from (A-7) and (A-9) that

$$
\begin{aligned}
J_n^\pi &\geq \Big[r(\lambda(p^D)) - C_1 u_n - \frac{C_3}{n^\eta}\Big]n(T - \tau_n) - \bar{p}C_7 n u_n \\
&\geq nr(\lambda(p^D))T - C_8 n(u_n + \tau_n), \tag{A-10}
\end{aligned}
$$

where C_8 is a suitably chosen constant. Now, from (A-10) and the definition of m^D in Fact 1 in the preamble of the appendix, we have that

$$
\frac{J_n^\pi}{J_n^D} = \frac{J_n^\pi}{nJ^D} \geq 1 - \frac{C_8}{m^D}(u_n + \tau_n). \tag{A-11}
$$

Case 2. Now suppose $\lambda(\bar{p}) > x/T$. In this case, note that $p^c = p^D = \bar{p}$. Recalling the lower bound expression in (A-2), we seek to get a handle on $\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\}$. The following result, whose proof can be found in Appendix B, says that the previous quantity will be "close" to the total units in inventory, nx, with very high probability.

Lemma 5 *Define $\mathcal{A} := \{\omega : \min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\} \geq nx - C_9 u_n, \; |\hat{p} - p^D| \leq C_2 u_n\}$. Then for some positive constants C_9, C_{10}, for and all $n \geq 1$,*

$$
\mathbb{P}(\mathcal{A}) \geq 1 - \frac{C_{10}}{n^{\eta-1}} \tag{A-12}
$$

The revenues generated by π_n can be bounded below as follows

$$
\begin{aligned}
J_n^\pi &\geq \mathbb{E}\Big[\hat{p}\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\}\Big] \\
&\geq \mathbb{E}\Big[\hat{p}\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\} \,\Big|\, \mathcal{A}\Big]\mathbb{P}(\mathcal{A}) \\
&\stackrel{(a)}{\geq} \mathbb{E}\Big[(p^D - C_2 u_n)\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\} \,\Big|\, \mathcal{A}\Big]\mathbb{P}(\mathcal{A}) \\
&\stackrel{(b)}{\geq} (p^D - C_2 u_n)(nx - C_9 n u_n)\Big(1 - \frac{C_{10}}{n^{\eta-1}}\Big) \tag{A-13} \\
&\geq p^D nx - nC_{11}u_n \tag{A-14}
\end{aligned}
$$

where both (a) and (b) follow from the definition of $\mathcal{A}$ and (A-12) and where $C_{10} > 0$ is suitably large. Now, recall that since $\lambda(\bar{p}) > x/T$, $p^D = \bar{p}$ and $J_n^D = nx\bar{p}$. Hence

$$
\frac{J_n^\pi}{J_n^D} = \frac{J_n^\pi}{nx\bar{p}} \geq 1 - \frac{C_{11}}{\bar{p}x}u_n. \tag{A-15}
$$

Step 4. Let $C_{12} = \max\{C_8/(m^D), C_{11}/(\bar{p}x)\}$. Combining (A-11) and (A-15), we have for all $\lambda \in \mathcal{L}$, $\mathcal{R}_n^{\pi}(x, T; \lambda) \leq C_{12}(u_n + \tau_n)$, and note that C_{12} does not depend on the specific function $\lambda \in \mathcal{L}$. Note that the choice of tuning parameters that minimizes the order of the the upper bound on the regret is exactly $\tau_n \asymp n^{-1/4}$ and $\kappa_n \asymp n^{1/4}$. Plugging in, we get for some $C_{13} > 0$

$$\sup_{\lambda \in \mathcal{L}} \mathcal{R}_n^{\pi}(x, T; \lambda) \leq \frac{C_{13}(\log n)^{1/2}}{n^{1/4}}.$$

The proof is complete. ∎

Proof of Proposition 2. Since $J_n^{\pi}(x, T; \lambda) \leq J_n^*(x, T|\lambda)$, the result will be established if we can prove that for some $C > 0$

$$\sup_{\lambda \in \mathcal{L}} \left(1 - \frac{J_n^*(x, T|\lambda)}{J_n^D(x, T|\lambda)}\right) > \frac{C}{n^{1/2}}. \tag{A-16}$$

Consider the demand function $\lambda(p) = (a - bp)^+$ with $b \in [\underline{K}, \overline{K}]$ and $a = \max\{2\overline{K}\bar{p}, \overline{K}\bar{p} + x/T\}$. Note in particular that $\lambda \in \mathcal{L}$ since $a \leq M$. In addition, for this demand function we have $p^u = p^c = \bar{p}$. By Proposition 1 in Gallego and van Ryzin (1994), any dynamic pricing policy will never price below p^u, hence the optimal dynamic pricing policy is to price at p^u until T or the time when inventory is depleted. We deduce that

$$
\begin{aligned}
J_n^*(x, T|\lambda) &= p^u \mathbb{E}[N(n\lambda(p^u)T)] - p^u \mathbb{E}[(N(n\lambda(p^u)T) - nx)^+] \\
&= nr(p^u)T - p^u \mathbb{E}[(N(nx) - nx)^+] \\
&= J_n^D(x, T|\lambda) - \bar{p}\mathbb{E}[(N(nx) - nx)^+] \\
&= J_n^D(x, T|\lambda) - \bar{p}nxe^{-nx}\frac{(nx)^n x}{(nx)!}.
\end{aligned}
$$

Hence,

$$\frac{J_n^D(x, T|\lambda) - J_n^*(x, T|\lambda)}{J_n^D(x, T|\lambda)} = \frac{1}{n\bar{p}\lambda(\bar{p})T}\bar{p}nxe^{-nx}\frac{(nx)^n x}{(nx)!}.$$

Using Sterling's approximation, $n^{1/2}(1 - J_n^*(x, T|\lambda)/J_n^D(x, T|\lambda)) \to C_1 > 0$ as $n \to \infty$ and hence (A-16) is established. This completes the proof. ∎

A.2 Proofs of the results in Section 5

In this section, we define for any $\theta \in \Theta$, $p^D(\theta) := \max\{p^u(\theta), p^c(\theta)\}$, where $p^u(\theta) = \arg\max\{p\lambda(p; \theta) : p \in [\underline{p}, \bar{p}]\}$, $p^c(\theta) = \arg\min\{|\lambda(p; \theta) - x/T| : p \in [\underline{p}, \bar{p}]\}$. In addition, we define $\mathcal{D}_\lambda(\theta) = \{l : l = \lambda(p; \theta)$ for some $p \in [\underline{p}, \bar{p}]\}$.

Proof of Proposition 3. In what follows, we let θ^* denote the true parameter value. The proof follows the structure of the proof of Proposition 1. The first step provides a lower on the revenues achieved by the proposed policy. The second step bounds the difference between the estimated parameter vector and its true value and analyzes the revenues at $p^u(\hat{\theta})$, $p^c(\hat{\theta})$ and $\hat{p} =$

31

$\max\{p^u(\hat{\theta}), p^c(\hat{\theta})\}$. Step 3 finalizes the analysis of the lower bound. The proof concludes with the resulting upper bound obtained for the regret.

Step 1. Here, we derive a lower bound on the expected revenues under the policy π_n. Let $X_n^{(L)} = \sum_{i=1}^k \lambda(p_i) n \Delta_n$, $X_n^{(P)} = \lambda(\hat{p}) n(T - \tau_n)$ and put $Y_n = N(X_n^{(L)} + X_n^{(P)})$, $Y_n^{(L)} = N(X_n^{(L)})$ and $Y_n^{(P)} = Y_n - Y_n^{(L)}$. As in the proof of Proposition 1, we have

$$J_n^\pi \geq \mathbb{E}\left[\hat{p}\min\{Y_n^{(P)}, (nx - Y_n^{(L)})^+\}\right] \tag{A-17}$$

Next, we analyze the lower bound above by getting a handle on $\hat{p}$, the estimate of p^D and the quantities $Y_n^{(P)}$ and $Y_n^{(L)}$.

Step 2. Here, we analyze the estimate $\hat{p}$.

We first analyze $p^u(\hat{\theta})$. Note that for any $p \in [\underline{p}, \overline{p}]$ and $\theta, \theta' \in \Theta$, we have

$$\begin{aligned}
|r(\lambda(p; \theta); \theta) - r(\lambda(p; \theta'); \theta')| &= |p\lambda(p; \theta) - p\lambda(p; \theta')| \\
&\leq \overline{p}|\lambda(p; \theta) - \lambda(p; \theta')| \\
&\overset{(a)}{\leq} \overline{p}\overline{K}_2\|\theta - \theta'\|_\infty,
\end{aligned} \tag{A-18}$$

where (a) follows from Assumption 2(ii.). Now

$$\begin{aligned}
0 \leq r(\lambda(p^u(\theta^*); \theta^*); \theta^*) &- r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*) \\
= \quad &\left[r(\lambda(p^u(\theta^*); \theta^*); \theta^*) - r(\lambda(p^u(\theta^*); \hat{\theta}); \hat{\theta})\right] + \left[r(\lambda(p^u(\theta^*); \hat{\theta}); \hat{\theta}) - r(\lambda(p^u(\hat{\theta}); \hat{\theta}); \hat{\theta})\right] \\
&+ \left[r(\lambda(p^u(\hat{\theta}); \hat{\theta}); \hat{\theta}) - r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*)\right] \\
\overset{(a)}{\leq} \quad &\overline{K}_2\overline{p}\|\hat{\theta} - \theta^*\|_\infty + 0 + \overline{K}_2\overline{p}\|\hat{\theta} - \theta^*\|_\infty,
\end{aligned}$$

where (a) follows from (A-18) and the fact that $p^u(\hat{\theta})$ maximizes $r(\lambda(\cdot; \hat{\theta}); \hat{\theta})$ on $[\underline{p}, \overline{p}]$. Hence, we have established that

$$0 \leq r(\lambda(p^u(\theta^*); \theta^*); \theta^*) - r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*) \leq 2\overline{K}_2\overline{p}\|\hat{\theta} - \theta^*\|_\infty. \tag{A-19}$$

We now turn to analyze $p^c(\hat{\theta})$. Note that under Assumption 1(i.),(ii.), $r(\cdot, \theta)$ is Lipschitz with constant $\overline{p} + M\underline{K}^{-1}$. Using this,

$$|r(\lambda(p^c(\theta^*); \theta^*); \theta^*) - r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*)| \leq (\overline{p} + M\underline{K}^{-1})|\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \theta^*)|. \tag{A-20}$$

Now note that

$$\begin{aligned}
|\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \theta^*)| &\leq |\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \hat{\theta})| + |\lambda(p^c(\hat{\theta}); \hat{\theta}) - \lambda(p^c(\hat{\theta}); \theta^*)| \\
&\overset{(a)}{\leq} |\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \hat{\theta})| + \overline{K}_2\|\hat{\theta} - \theta^*\|_\infty,
\end{aligned}$$

where (a) follows from Assumption 2(ii.). Let $h_2(l; \theta)$ denote the projection on $\mathcal{D}_\lambda(\theta)$, i.e., $h_2(l; \theta) = \max\{\min\{y, \lambda(\underline{p}; \theta)\}, \lambda(\overline{p}; \theta)\}$ and note that $\lambda(p^c(\theta); \theta) = h_2(x/T; \theta)$. It is easy to show that for all

y, $h_2(y, \cdot)$ is Lipschitz with constant $\overline{K}_2$ where the latter was defined in Assumption 2(ii.). Using this, we have $|\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \hat{\theta})| = |h_2(x/T; \theta^*) - h_2(x/T; \hat{\theta})|$, and we deduce that

$$|\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \theta^*)| \leq 2\overline{K}_2 \|\hat{\theta} - \theta^*\|_\infty. \tag{A-21}$$

Combining (A-20) and (A-21), we get

$$|r(\lambda(p^c(\theta^*); \theta^*); \theta^*) - r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*)| \leq 2(\overline{p} + M\underline{K}^{-1})\overline{K}_2 \|\hat{\theta} - \theta^*\|_\infty. \tag{A-22}$$

We now turn to analyze the revenues achieved by $\hat{p} = p^D(\hat{\theta})$. For this purpose, we divide the analysis into two cases:

Case 1. Suppose first that $p^u(\theta^*) \geq p^c(\theta^*)$, i.e., $p^D(\theta^*) = p^u(\theta^*)$.

i.) If $p^u(\hat{\theta}) \geq p^c(\hat{\theta})$, then

$$|r(\lambda(p^D(\theta^*); \theta^*); \theta^*) - r(\lambda(\hat{p}; \theta^*); \theta^*)| = |r(\lambda(p^u(\theta^*); \theta^*); \theta^*) - r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*)| \leq 2\overline{K}_2\overline{p}\|\hat{\theta} - \theta^*\|_\infty.$$

ii.) If $p^u(\hat{\theta}) < p^c(\hat{\theta}) \leq p^u(\theta^*)$, then the fact that $r(\lambda(\cdot; \theta^*), \theta^*)$ is nondecreasing to the left of $p^u(\theta^*)$ implies that

$$\begin{aligned}
|r(\lambda(p^D(\theta^*); \theta^*); \theta^*) - r(\lambda(\hat{p}; \theta^*); \theta^*)| &= |r(\lambda(p^u(\theta^*); \theta^*); \theta^*) - r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*)| \\
&\leq |r(\lambda(p^u(\theta^*); \theta^*); \theta^*) - r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*)| \\
&\leq 2\overline{K}_2\overline{p}\|\hat{\theta} - \theta^*\|_\infty.
\end{aligned}$$

iii.) If $p^u(\hat{\theta}) < p^c(\hat{\theta})$ and $p^c(\hat{\theta}) > p^u(\theta^*)$, then

$$\begin{aligned}
|r(\lambda(p^D(\theta^*); \theta^*); \theta^*) - r(\lambda(\hat{p}; \theta^*); \theta^*)| &= |r(\lambda(p^u(\theta^*); \theta^*); \theta^*) - r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*)| \\
&\leq (\overline{p} + M\underline{K}^{-1})|\lambda(p^u(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \theta^*)| \\
&\overset{(a)}{\leq} (\overline{p} + M\underline{K}^{-1})|\lambda(p^c(\theta^*); \theta^*) - \lambda(p^c(\hat{\theta}); \theta^*)| \\
&\overset{(b)}{\leq} 2(\overline{p} + M\underline{K}^{-1})\overline{K}_2 \|\hat{\theta} - \theta^*\|_\infty.
\end{aligned}$$

where (a) follows from the fact that $p^c(\theta^*) \leq p^u(\theta^*) < p^c(\hat{\theta})$ and that $\lambda(\cdot; \theta^*)$ is nonincreasing and (b) follows from (A-21).

Case 2. Suppose now that $p^u(\theta^*) < p^c(\theta^*)$, implying that $p^D(\theta^*) = p^c(\theta^*)$.

i.) If $p^c(\hat{\theta}) \geq p^u(\hat{\theta})$, then

$$|r(\lambda(p^D(\theta^*); \theta^*); \theta^*) - r(\lambda(\hat{p}; \theta^*); \theta^*)| = |r(\lambda(p^c(\theta^*); \theta^*); \theta^*) - r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*)| \leq 2(\overline{p} + M\underline{K}^{-1})\overline{K}_2 \|\hat{\theta} - \theta^*\|_\infty.$$

ii.) If $p^c(\hat{\theta}) < p^u(\hat{\theta})$, then

$$\begin{aligned}
r(\lambda(p^D(\theta^*); \theta^*); \theta^*) - r(\lambda(\hat{p}; \theta^*); \theta^*) &= r(\lambda(p^c(\theta^*); \theta^*); \theta^*) - r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*) \\
&= r(\lambda(p^c(\theta^*); \theta^*); \theta^*) - r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*) \\
&\quad + r(\lambda(p^c(\hat{\theta}); \theta^*); \theta^*) - r(\lambda(p^u(\theta^*); \theta^*); \theta^*) \\
&\quad + r(\lambda(p^c(\theta^*); \theta^*); \theta^*) - r(\lambda(p^u(\hat{\theta}); \theta^*); \theta^*) \\
&\overset{(a)}{\leq} \left(2(\overline{p} + M\underline{K}^{-1})\overline{K}_2 + 2\overline{K}_2\overline{p}\right)\|\hat{\theta} - \theta^*\|_\infty.
\end{aligned}$$

33

where (a) follows from the fact that $p^u(\theta^*)$ maximizes $r(\lambda(\cdot;\theta^*);\theta^*)$ on $[\underline{p},\overline{p}]$ and the inequalities (A-19) and (A-22). Combining the results from cases 1 and 2, we have established that

$$r(\lambda(p^D(\theta^*);\theta^*);\theta^*) - r(\lambda(\hat{p};\theta^*);\theta^*) \leq C_1 \|\hat{\theta} - \theta^*\|_\infty, \tag{A-23}$$

where $C_1 = \left(2(\overline{p} + M\underline{K}^{-1})\overline{K}_2 + 2\overline{K}_2\overline{p}\right)$. We are now left with controlling $\hat{\theta}$. This is done in Lemma 6, whose proof is deferred to Appendix B.

Lemma 6 *Under Assumption 2, for some $C_2 > 0$,*

$$\mathbb{E}\|\hat{\theta} - \theta^*\|_\infty \leq \frac{C_2}{(n\tau_n)^{1/2}} \qquad \text{for all } n \geq 1. \tag{A-24}$$

Using the result above and (A-23), we have

$$
\begin{aligned}
\mathbb{E}\left[r(\lambda(\hat{p};\theta^*);\theta^*)\right] &= r(\lambda(p^D(\theta^*);\theta^*);\theta^*) + \mathbb{E}\left[r(\lambda(\hat{p};\theta^*);\theta^*) - r(\lambda(p^D(\theta^*);\theta^*));\theta^*)\right] \\
&\geq r(\lambda(p^D(\theta^*);\theta^*);\theta^*) - \frac{C_1 C_2}{(n\tau_n)^{1/2}}.
\end{aligned}
$$

Step 3. Finalizing the analysis of the lower bound can be conducted in a similar manner as in the proof of Proposition 1 (Step 3) by letting in this case $u_n = (\log n)^{1/2}(n\tau_n)^{-1/2}$.

Step 4. As in the proof of Proposition 1 (Step 4), we get that for some $C_3 > 0$,

$$\left(1 - \frac{J_n^\pi}{J_n^D}\right) \leq \frac{C_3}{m^D}\left(\tau_n + (\log n)^{1/2}(n\tau_n)^{-1/2}\right), \tag{A-25}$$

and the result follows by plugging in $\tau_n \asymp n^{-1/3}$. This completes the proof. $\blacksquare$

Proof of Proposition 5. Throughout the proof, we let θ^* denote the true underlying parameter value. Also, to simplify notation, we let $p^D := p^D(\theta^*)$. The proof is organized as follows. We first lower bound the expected revenues achieved by the proposed policy. The key issue here is to account for the performance losses throughout the various stages of the algorithm. This is done by analyzing the estimates $\hat{\theta}^i$ and the prices $\hat{p}^i$ at each stage and controlling for the amount of inventory depleted as well as the revenues generated.

For $i = 1, ..., \ell$, let $X_n^i = \sum_{j=1}^i n\lambda(\hat{p}_j;\theta^*)\Delta_n^{(j)}$, and put $Y_n^i = N(X_n^i) - N(X_n^{i-1})$. Finally, let $Y_n = \sum_{i=1}^\ell Y_n^i$ and recall that $p^D(\theta) = \max(p^u(\theta), p^c(\theta))$. We restrict attention to the case $\lambda(\overline{p};\theta^*) \leq x/T$ (the case of $\lambda(\overline{p};\theta^*) > x/T$ can be handled by following arguments similar to those in the second case of Step 3 in the proof of Proposition 1). The total revenues over the selling

horizon under the policy π_n can be bounded below as follows

$$
\begin{aligned}
J_n^\pi \;&\geq\; \mathbb{E}\Big[\hat{p}_1 Y_n^1 + \sum_{i=2}^{\ell_n}\hat{p}_i Y_n^i\Big] - \overline{p}\,\mathbb{E}\Big[(Y_n - nx)^+\Big]\\[4pt]
&=\; n\hat{p}_1 \lambda(\hat{p}_1;\theta^*)\Delta_n^{(1)} + \sum_{i=2}^{\ell_n}\mathbb{E}\Big[\mathbb{E}\Big[\hat{p}_i N(n\lambda(\hat{p}_i;\theta^*)\Delta_n^{(i)})\,\Big|\,\hat{p}_i\Big]\Big] - \overline{p}\,\mathbb{E}\Big[(Y_n - nx)^+\Big]\\[4pt]
&=\; nr(\lambda(\hat{p}_1;\theta^*);\theta^*)\Delta_n^{(1)} + n\sum_{i=2}^{\ell_n}\mathbb{E}\Big[\hat{p}_i\lambda(\hat{p}_i;\theta^*)\Big]\Delta_n^{(i)} - \overline{p}\,\mathbb{E}\Big[(Y_n - nx)^+\Big]\\[4pt]
&=\; nr(\lambda(p^D;\theta^*);\theta^*)\sum_{i=1}^{\ell_n}\Delta_n^{(i)} - n\big[r(\lambda(p^D;\theta^*);\theta^*) - r(\lambda(\hat{p}_1;\theta^*);\theta^*)\big]\Delta_n^{(1)}\\[4pt]
&\quad -n\sum_{i=2}^{\ell_n}\Big(r(\lambda(p^D;\theta^*);\theta^*) - \mathbb{E}[r(\lambda(\hat{p}_i;\theta^*);\theta^*)]\Big)\Delta_n^{(i)} - \overline{p}\,\mathbb{E}\Big[(Y_n - nx)^+\Big] \qquad\text{(A-26)}\\[4pt]
&=\; r(\lambda(p^D;\theta^*);\theta^*)nT - n\big[r(\lambda(p^D;\theta^*);\theta^*) - r(\lambda(\hat{p}_1;\theta^*);\theta^*)\big]\Delta_n^{(1)}\\[4pt]
&\quad -n\sum_{i=2}^{\ell_n}\Big(r(\lambda(p^D;\theta^*);\theta^*) - \mathbb{E}[r(\lambda(\hat{p}_i;\theta^*);\theta^*)]\Big)\Delta_n^{(i)} - \overline{p}\,\mathbb{E}\Big[(Y_n - nx)^+\Big], \qquad\text{(A-27)}
\end{aligned}
$$

where $\hat{p}_i$, $i = 2, ..., \ell_n$ were defined in Algorithm 3 (Step 2(b)).

Note that $(r(\lambda(p^D;\theta^*);\theta^*) - r(\lambda(\hat{p}_1;\theta^*);\theta^*))\Delta_n^{(1)} \leq r(\lambda(p^D;\theta^*);\theta^*)\Delta_n^{(1)} \leq C_1 n^{a_\ell - 1}$, where $C_1/2 = M\overline{p}$ bounds the revenue rate. The main task is to derive bounds on $r(\lambda(p^D;\theta^*);\theta^*) - \mathbb{E}[r(\lambda(\hat{p}_i;\theta^*);\theta^*)]$. Using Lemma 6 and a parallel reasoning to the one in Step 2 of the proof of Proposition 3, we have that for some $C_2 > 0$

$$
r(\lambda(p^D;\theta^*);\theta^*) - \mathbb{E}[r(\lambda(\hat{p}_i;\theta^*);\theta^*)] \leq \frac{C_2}{(n\Delta_n^{(i-1)})^{1/2}},
$$

and hence,

$$
\begin{aligned}
\sum_{i=2}^{\ell_n}\Big(r(\lambda(p^D;\theta^*);\theta^*) - \mathbb{E}[r(\lambda(\hat{p}_i;\theta^*);\theta^*)]\Big)\Delta_n^{(i)} \;&\leq\; C_1\sum_{i=2}^{\ell_n}\frac{\Delta_n^{(i)}}{(n\Delta_n^{(i-1)})^{1/2}}\\[4pt]
&=\; C_2\beta^{1/2}\sum_{i=2}^{\ell_n}n^{a_\ell - 1} \leq C_2 T^{1/2}\ell_n n^{a_\ell - 1},
\end{aligned}
$$

where the last inequality follows since $\Delta_n^{(\ell)} = \beta$ and hence $\beta \leq T$. We deduce that

$$
\begin{aligned}
n\big[r(\lambda(p^D;\theta^*);\theta^*) &- r(\lambda(\hat{p}_1;\theta^*);\theta^*)\big]\Delta_n^{(1)}\\[4pt]
&+n\sum_{i=2}^{\ell_n}\Big(r(\lambda(p^D;\theta^*);\theta^*) - \mathbb{E}[r(\lambda(\hat{p}_i;\theta^*);\theta^*)]\Big)\Delta_n^{(i)} + \overline{p}\,\mathbb{E}[(Y_n - nx)^+]\\[4pt]
&\leq\; n\max\{C_2 T^{1/2}, C_1\}\ell_n n^{a_\ell - 1} + \overline{p}\,\mathbb{E}[(Y_n - nx)^+]. \qquad\text{(A-28)}
\end{aligned}
$$

Let ϵ_n be a sequence of positive numbers to be defined later. We have

$$
\begin{aligned}
\mathbb{E}\Big[(Y_n - nx)^+\Big] &= \mathbb{E}\Big[(Y_n - nx)^+; Y_n - nx \le n\epsilon_n\Big] + \mathbb{E}\Big[(Y_n - nx)^+; Y_n - nx > n\epsilon_n\Big] \\
&\le n\epsilon_n + \mathbb{E}\Big[(Y_n - nx)^+ \mid Y_n - nx > n\epsilon_n\Big]\mathbb{P}(Y_n - nx > n\epsilon_n) \\
&\overset{(a)}{\le} n\epsilon_n + (n\epsilon_n + nM + 1)\mathbb{P}(Y_n - nx > n\epsilon_n),
\end{aligned}
$$

where (a) follows from the fact that for a Poisson random variable Z with mean μ, $\mathbb{E}[Z \mid Z > a] \le a + 1 + \mu$. Now note that $nx = n(x/T)T \ge n\lambda(p^c; \theta^*)\sum_{i=1}^{\ell_n}\Delta_n^{(i)}$. (Note that under the assumption that $\lambda(\overline{p}; \theta^*) \le x/T$ and since $\lambda(\cdot; \theta^*)$ is continuous and decreasing, either $\lambda(\underline{p}; \theta^*) \ge x/T$ in which case $\lambda(p^c; \theta^*) = x/T$ or $\lambda(\underline{p}; \theta^*) < x/T$ in which case $\lambda(p^c) < x/T$.) Hence we have

$$
\begin{aligned}
\mathbb{P}(Y_n - nx > \ell_n n\epsilon_n) &\le \mathbb{P}\Big\{N(n\lambda(\hat{p}_1; \theta^*)\Delta_n^{(1)}) > n\lambda(p^c; \theta^*)\Delta_n^{(1)} + n\epsilon_n\Big\} \\
&\quad + \sum_{i=2}^{\ell_n}\mathbb{P}\Big\{N(n\lambda(\hat{p}_i; \theta^*)\Delta_n^{(i)}) > n\lambda(p^c; \theta^*)\Delta_n^{(i)} + n\epsilon_n\Big\}. \quad \text{(A-29)}
\end{aligned}
$$

Choose $\epsilon_n = C_3(\log n)^{1/2}\Delta_n^{(1)} = C_3(\log n)^{1/2}n^{a_{\ell_n}-1}$, where $C_3 > 0$ is suitably large. Note that $\Delta_n^{(1)}/\epsilon_n \to 0$ as $n \to \infty$, hence by Lemma 2 the first term on the RHS of (A-29) is bounded from above by C_4/n^η for some $C_4 > 0$ and some $\eta \ge 2$. Consider any term in the sum which constitutes the second term on the RHS of (A-29). Using the same reasoning as the one leading to (B-10), we have for some $C_5, C_6 > 0$

$$
\mathbb{P}\left\{\lambda(\hat{p}_i; \theta^*) - \lambda(p^c; \theta^*) > \frac{C_5}{(\Delta_n^{(i-1)})^{1/2}}\right\} \le \frac{C_6}{n^{\eta-1}}. \quad \text{(A-30)}
$$

Now, for $i = 2, ..., \ell$

$$
\begin{aligned}
&\mathbb{P}\Big\{N(n\lambda(\hat{p}_i; \theta^*)\Delta_n^{(i)}) > n\lambda(p^c; \theta^*)\Delta_n^{(i)} + n\epsilon_n\Big\} \\
&\le \mathbb{P}\left\{N(n\lambda(\hat{p}_i; \theta^*)\Delta_n^{(i)}) > n\lambda(p^c; \theta^*)\Delta_n^{(i)} + n\epsilon_n \,,\, \lambda(\hat{p}_i; \theta^*) \le \lambda(p^c; \theta^*) + \frac{C_6}{(n\Delta_n^{(i-1)})^{1/2}}\right\} \\
&\quad + \mathbb{P}\left\{\lambda(\hat{p}_i; \theta^*) - \lambda(p^c; \theta^*) > \frac{C_6}{(n\Delta_n^{(i-1)})^{1/2}}\right\} \\
&\overset{(a)}{\le} \mathbb{P}\left\{N(n\lambda(p^c; \theta^*) + nC_6\Delta_n^{(i)}/(\Delta_n^{(i-1)})^{1/2}) > n\lambda(p^c; \theta^*)\Delta_n^{(i)} + n\epsilon_n\right\} + \frac{C_6}{n^{\eta-1}} \\
&\le \mathbb{P}\left\{N(n\lambda(p^c; \theta^*) + C_6 n^{a_\ell}) - (n\lambda(p^c; \theta^*)\Delta_n^{(i)} + C_6 n^{a_\ell}) > -nC_6 n^{a_\ell - 1} + n\epsilon_n\right\} + \frac{C_6}{n^{\eta-1}} \\
&\overset{(b)}{\le} \mathbb{P}\left\{N(n\lambda(p^c; \theta^*) + C_6 n^{a_\ell}) - (n\lambda(p^c; \theta^*)\Delta_n^{(i)} + C_6 n^{a_\ell}) > \frac{1}{2}n\epsilon_n\right\} + \frac{C_6}{n^{\eta-1}},
\end{aligned}
$$

where (a) follows from (A-30) and (b) follows from the definition of the sequence $\Delta_n^{(i)}, i = 1, ..., \ell$. Appealing to Lemma 2 yields that the first term on the RHS above can be bounded by C_7/n^η for some $C_7 > 0$. We conclude that for some $C_8 > 0$,

$$
\mathbb{E}\Big[(Y_n - nx)^+\Big] \le nC_8(\log n)^{1/2}n^{a_{\ell_n}-1},
$$

36

and this in turn, in conjunction with (A-27) and (A-28), implies that for some $C_9 > 0$

$$\frac{J_n^\pi}{J_n^D} \geq 1 - \frac{C_9}{m^D}\Big[\ell_n n^{a_{\ell_n}-1} + (\log n)^{1/2} n^{a_{\ell_n}-1}\Big]. \tag{A-31}$$

By taking $a_{\ell_n} = 2^{\ell_n-1}/(2^{\ell_n}-1)$ with $\ell_n = (\log 2)^{-1}\log\log n$, we get that $n^{a_{\ell_n}-1} \leq \exp(1)n^{-1/2}$. Plugging in (A-31), we get the final assertion of the proposition. A similar result holds when $\lambda(\bar{p},\theta^*) > x/T$ and the the proof is thus complete. ∎

B Proofs of Auxiliary Results

In what follows, C_i', $i = 1, 2, \ldots$ will denote positive constants that depend only on x, T and the class $\mathcal{L}$ (but not on a specific function $\lambda \in \mathcal{L}$).

Proof of Lemma 1. Fix $\lambda \in \mathcal{L}$ and let $\mathcal{P}_\lambda$ denote the class of policies that "know" the demand function prior the the start of the selling season, and whose corresponding price process is non anticipating and satisfies the admissibility conditions (2)-(3). Put

$$J^*(x,T|\lambda) := \sup_{\pi\in\mathcal{P}_\lambda} \mathbb{E}\Big[\int_0^T p(s)dN^\pi(s)\Big], \tag{B-1}$$

Problem (B-1) would be the one that the decision maker would solve if s/he would know the demand function λ prior to the start of the selling season. Clearly $J^\pi(x,T;\lambda) \leq J^*(x,T|\lambda)$ for all $\pi \in \mathcal{P}$. Hence, to establish the bound of the lemma, it is sufficient to show that $J^*(x,T|\lambda) \leq J^D(x,T|\lambda)$. Consider the optimization problem given in (5) and its equivalent formulation in terms of rates:

$$J^D(x,T|\lambda) = \max\Big\{\int_0^T r(\lambda(s))ds \,:\, \int_0^T \lambda(s)ds \leq x \,, \, \lambda(s) \in [\underline{l},\bar{l}] \cup \{0\} \text{ for all } 0 \leq s \leq T\Big\}. \tag{B-2}$$

If $\underline{l} = 0$, then the set of feasible rates is convex and Gallego and van Ryzin (1994, Theorem 2) yields the result. Suppose now that $\underline{l} > 0$ and let

$$\lambda^c = \arg\min_{l\in[\underline{l},\bar{l}]} |l - x/T|\,, \quad \lambda^u = \arg\max_{l\in[\underline{l},\bar{l}]} r(l)\,, \quad \lambda^D = \min\{\lambda^u, \lambda^c\}.$$

Claim 1: If $x/T \geq \underline{l}$, then setting $\lambda(s) = \lambda^D$ for $0 \leq s \leq T$ is an optimal solution to the deterministic problem (B-2). In addition the value of the deterministic problem serves as an upper bound on the original stochastic problem.

To verify the claim we consider two cases. Suppose first that $x/T \in [\underline{l},\bar{l}]$, in which case $\lambda^c = x/T$. If $\lambda^u \leq x/T$, then applying λ^u throughout is feasible and is hence optimal. In addition, the optimal value of the deterministic problem is given by $r(\lambda^u)T$ which is an upper bound on the stochastic problem (since $r(l) \leq r(\lambda^u)$ for all $l \in [\underline{l},\bar{l}]$) . If $\lambda^u > x/T$, then $\lambda^u > \underline{l}$ and hence $r(\cdot)$ is necessarily non-decreasing on $[0,\lambda^u]$ (by the concavity assumption). Thus if we relax the domain of feasible rates to the convex set $[0,\bar{l}]$, applying λ^D throughout is an optimal solution (see Gallego and van

Ryzin (1994)). This is a feasible solution for the constrained problem and is hence optimal for that problem as well. Let $J^D[0,\bar{l}]$ be the value of the deterministic problem with feasible rates in $[0,\bar{l}]$ and let $J^*[0,\bar{l}]$ be the value of the stochastic problem. We clearly have that

$$J^D[\underline{l},\bar{l}] \overset{(a)}{=} J^D[0,\bar{l}] \overset{(b)}{\geq} J^*[0,\bar{l}] \overset{(c)}{\geq} J^*[\underline{l},\bar{l}],$$

where: (a) follows from the result we just established; (b) follows from Gallego and van Ryzin (1994, Theorem 2); and (c) is a clear consequence of constraint relaxation.

Suppose now that $x/T > \bar{l}$. In this case, $\lambda^u \leq \bar{l} < x/T$. Applying λ^u throughout is feasible and hence optimal. The optimal value of the deterministic problem is hence given by $r(\lambda^u)T$ which is an upper bound on the stochastic problem. The assertion of the claim follows.

Claim 2: If $x/T < \underline{l}$, then $\lambda(s) = \lambda^D = \underline{l}$ for $s \in [0,T']$ and $\lambda(s) = 0$ for $s \in (T',T]$, where $T' = x/\underline{l}$, is an optimal solution to the deterministic relaxation.

To verify the claim, note that all units are sold at the maximum allowable price with the proposed solution and hence the revenues achieved are $x\bar{p}$, which is a clear upper bound on the performance of any solution of the deterministic or the stochastic problem. Thus the assertion of the claim follows.

Combining the two claims yields the result of the lemma in the case $\underline{l} > 0$. ∎

Proof of Lemma 2. If $\mu = 0$, then result is immediate. Now suppose $\mu \in (0,M]$. Using a Chernoff bound, we have for any $\theta > 0$

$$\mathbb{P}\Big(N(\mu r_n) - \mu r_n > r_n \epsilon_n\Big) \leq \exp\Big\{-\theta r_n(\mu + \epsilon_n) + (\exp\{\theta\} - 1)\mu r_n\Big\}. \tag{B-3}$$

The expression in each of the exponents is minimized by setting $\theta = \log\big(1 + \epsilon_n/\mu\big)$. Plugging this into (B-3) yields

$$\mathbb{P}\Big(N(\mu r_n) - \mu r_n > r_n \epsilon_n\Big) \leq \exp\Big\{-\log\Big(1 + \frac{\epsilon_n}{\mu}\Big)(\mu + \epsilon_n) + \epsilon_n\Big\}$$
$$\leq \exp\Big\{r_n\Big(-\log\Big(1 + \frac{\epsilon_n}{M}\Big)(M + \epsilon_n) + \epsilon_n\Big)\Big\}. \tag{B-4}$$

For the last inequality, note that the derivative of the term in the exponent with respect to μ is given by $-\log(1 + \epsilon_n/\mu) + \epsilon_n/\mu$, which is always positive for $\epsilon_n > 0$. Now, using a Taylor expansion we get that for some $\xi \in [0,\epsilon_n]$,

$$-\log\Big(1 + \frac{\epsilon_n}{M}\Big)(M + \epsilon_n) + \epsilon_n = -\frac{1}{2}\frac{1}{1+\xi}\frac{\epsilon_n^2}{M}.$$

If $\epsilon_n \leq 1$, we have

$$-\frac{1}{2}\frac{1}{1+\xi}\frac{\epsilon_n^2}{M} \leq -\frac{\epsilon_n^2}{4M}.$$

If $\epsilon_n > 1$, then

$$-\frac{1}{2}\frac{1}{1+\xi}\frac{\epsilon_n^2}{M} \leq -\frac{1}{2M}\frac{\epsilon_n}{1+\epsilon_n} \leq -\frac{1}{4M}.$$

38

Letting $C(\eta) = 2\eta^{1/2}M^{1/2}$ and substituting for ϵ_n, we get

$$\mathbb{P}\Big(N(\mu r_n) - \mu r_n > r_n \epsilon_n\Big) \;\leq\; \exp\Big\{-r_n \min\Big\{\frac{1}{4M}, \frac{C^2(\eta)\log n}{4Mr_n}\Big\}\Big\} = \max\Big\{\exp\Big\{\frac{-r_n}{4M}\Big\}, \frac{1}{n^\eta}\Big\},$$

Now it is easy to check that $\exp\{-r_n/4M\} \leq (4M\eta/\beta)^{1/\beta}\exp\{-\eta/\beta\}1/n^\eta$. Hence the first result follows with $C_1 = \max\{1, (4M\eta/\beta)^{1/\beta}\exp\{-\eta/\beta\}\}$. The other inequality is established using identical arguments. This completes the proof. ∎

Proof of Lemma 3. The proof is divided into three main steps. The first step analyzes key properties of the estimate $\hat{p}^c$ and establishes (A-5). The second step focuses on $\hat{p}^u$, and the third step uses the results from the previous two steps to derive properties of $\hat{p} = \max\{\hat{p}^u, \hat{p}^c\}$ and establishes (A-4).

    Step 1. Here, we focus on $\hat{p}^c$ and show the following. Put $C_1' = (M + \overline{K}\overline{p})C_2'$ where $C_2' = \underline{K}^{-1}\max\{8\eta^{1/2}M^{1/2}, 2\overline{K}(\overline{p} - \underline{p})\}$, then for some $C_3' > 0$ and for all $n \geq 1$

$$\mathbb{P}\{|r(\lambda(p^c)) - r(\lambda(\hat{p}^c))| > C_1'u_n\} \leq \frac{C_3'}{n^{\eta-1}}, \tag{B-5}$$

$$\mathbb{P}\{|\hat{p}^c - p^c| > C_2'u_n\} \leq \frac{C_3'}{n^{\eta-1}}. \tag{B-6}$$

For each n, let $[p_i, p_{i+1}]$ be the interval that contains p^c. Here, we drop the dependence on n to avoid cluttering the notation. Now

$$
\begin{aligned}
|\lambda(\hat{p}^c) - \lambda(p^c)| &\leq |\lambda(\hat{p}^c) - \widehat{\lambda}(\hat{p}^c)| + |\widehat{\lambda}(\hat{p}^c) - \lambda(p^c)| \\
&\overset{(a)}{\leq} |\lambda(\hat{p}^c) - \widehat{\lambda}(\hat{p}^c)| + |\widehat{\lambda}(p_i) - \lambda(p^c)| \\
&\overset{(b)}{\leq} 2\max_{1\leq k\leq \kappa_n}|\widehat{\lambda}(p_k) - \lambda(p_k)| + |\lambda(p_i) - \lambda(p^c)| \\
&\overset{(c)}{\leq} 2\max_{1\leq k\leq \kappa_n}|\widehat{\lambda}(p_k) - \lambda(p_k)| + \frac{\overline{K}(\overline{p} - \underline{p})}{\kappa_n},
\end{aligned}
$$

where (a) follows from the definition of $\hat{p}^c = \arg\min_{1\leq j\leq \kappa_n}|\widehat{\lambda}(p_j) - \lambda(p^c)|$, (b) follows from the triangle inequality, and (c) follows from the fact that $\lambda(\cdot)$ is $\overline{K}$-Lipschitz (Assumption 1(ii.)). We have

$$
\begin{aligned}
\mathbb{P}\{|\hat{p}^c - p^c| > C_2'u_n\} &\overset{(a)}{\leq} \mathbb{P}\{|\lambda(\hat{p}^c) - \lambda(p^c)| > C_2'\underline{K}u_n\} \\
&\leq \mathbb{P}\Big\{\max_{1\leq k\leq \kappa_n}|\widehat{\lambda}(p_k) - \lambda(p_k)| > \frac{(C_2'\underline{K})u_n}{2} - \frac{\overline{K}(\overline{p} - \underline{p})}{2\kappa_n}\Big\} \\
&\overset{(b)}{\leq} \sum_{k=1}^{\kappa_n}\mathbb{P}\Big\{|\widehat{\lambda}(p_k) - \lambda(p_k)| > \frac{(C_2'\underline{K})u_n}{4}\Big\} \\
&\overset{(c)}{\leq} \frac{C_3'}{n^{\eta-1}},
\end{aligned}
$$

where (a) follows from the fact that $\gamma(\cdot)$ is $\underline{K}^{-1}$-Lipschitz (Assumption 1(ii.)), (b) follows from a union bound and the fact that $C_2' \geq 2\overline{K}(\overline{p}-\underline{p})\underline{K}^{-1}$ and and (c) follows from Lemma 2 in conjunction with the inequality $C_2'/(4\underline{K}^{-1}) \geq 2\eta^{1/2}M^{1/2}$. Note that (A-5) has now been established by setting $C_3 = C_2'$. The inequality for revenues now follows. Indeed, note that under Assumption 1, $r(\lambda(\cdot))$ is Lipschitz with constant $M + \overline{p}\overline{K}$. Hence, $\left|r(\lambda(p^c)) - r(\lambda(\hat{p}^c))\right| \leq (M + \overline{K}\overline{p})\left|p^c - \hat{p}^c\right|$, and

$$
\begin{aligned}
\mathbb{P}\left\{\left|r(\lambda(p^c)) - r(\lambda(\hat{p}^c))\right| > C_1'u_n\right\} &\leq \mathbb{P}\left\{\left|\hat{p}^c - p^c\right| > C_1'u_n/(M + \overline{K}\overline{p})\right\} \\
&= \mathbb{P}\left\{\left|\hat{p}^c - p^c\right| > C_2'u_n\right\} \\
&\leq \frac{C_3'}{n^{\eta-1}}.
\end{aligned}
$$

This concludes the proof of (B-6) and (B-5). In Step 2, we derive properties of $\hat{p}^u$ which are subsequently used in Step 3 to prove (A-4).

Step 2. Here, we focus on $\hat{p}^u$ and show the following. Put $C_4' = \max\{8\eta^{1/2}M^{1/2}\overline{p}, 2M + \overline{p}\overline{K}\}$, then for some $C_5' > 0$ and all $n \geq 1$

$$
\mathbb{P}\left\{r(\lambda(p^u)) - r(\lambda(\hat{p}^u)) > C_4'u_n\right\} \leq \frac{C_5'}{n^{\eta-1}}, \tag{B-7}
$$

First, for each n, let j be the index such that $p^u \in [p_j, p_{j+1}]$. Recall that under Assumption 1, $r(\lambda(\cdot))$ is Lipschitz with constant $M + \overline{p}\overline{K}$. This, in conjunction with the fact that $|p^u - p_{j+1}| \leq (\overline{p} - \underline{p})/\kappa_n$ yields

$$
\left|r(\lambda(p^u)) - r(\lambda(p_{j+1}))\right| \leq \frac{(M + \overline{p}\overline{K})(\overline{p} - \underline{p})}{\kappa_n}. \tag{B-8}
$$

We now establish an upper bound on the difference in revenues $r(\lambda(p^u)) - r(\lambda(\hat{p}^u))$ as follows

$$
\begin{aligned}
r(\lambda(p^u)) - r(\lambda(\hat{p}^u)) &= r(\lambda(p^u)) - p_{j+1}\widehat{\lambda}(p_{j+1}) + p_{j+1}\widehat{\lambda}(p_{j+1}^n) - \hat{p}^u\widehat{\lambda}(\hat{p}^u) + \hat{p}^u\widehat{\lambda}(\hat{p}^u) - \hat{p}^u\lambda(\hat{p}^u) \\
&\overset{(a)}{\leq} r(\lambda(p^u)) - p_{j+1}\widehat{\lambda}(p_{j+1}) + \hat{p}^u\widehat{\lambda}(\hat{p}^u) - \hat{p}^u\lambda(\hat{p}^u) \\
&\leq \left|r(\lambda(p^u)) - r(\lambda(p_{j+1}))\right| + 2\max_{1\leq k\leq\kappa_n}\left|p_k\lambda(p_k) - p_k\widehat{\lambda}(p_k)\right| \\
&\overset{(b)}{\leq} \frac{(M + \overline{p}\overline{K})(\overline{p} - \underline{p})}{\kappa_n} + 2\max_{1\leq k\leq\kappa_n}\left|p_k\lambda(p_k) - p_k\widehat{\lambda}(p_k)\right|,
\end{aligned}
$$

where (a) follows from the definition of $\hat{p}^u$ given in (10), and (b) follows from (B-8). Now

$$
\begin{aligned}
\mathbb{P}\left\{r(\lambda(p^u)) - r(\lambda(\hat{p}^u)) > C_4'u_n\right\} &\leq \mathbb{P}\left\{\max_{1\leq k\leq\kappa_n}\left|p_k\lambda(p_k) - p_k\widehat{\lambda}(p_k)\right| > \frac{u_n}{2} - \frac{(M + \overline{p}\overline{K})(\overline{p} - \underline{p})}{2\kappa_n}\right\} \\
&\overset{(a)}{\leq} \mathbb{P}\left\{\max_{1\leq k\leq\kappa_n}\left|p_k\lambda(p_k) - p_k\widehat{\lambda}(p_k)\right| > \frac{C_4'u_n}{4}\right\} \\
&\overset{(b)}{\leq} \sum_{k=1}^{\kappa_n}\mathbb{P}\left\{\left|\lambda(p_k) - \widehat{\lambda}(p_k)\right| > \frac{C_4'u_n}{4\overline{p}}\right\} \\
&\overset{(c)}{\leq} \frac{\kappa_n C_5'}{n^\eta} \\
&\leq \frac{C_5'}{n^{\eta-1}},
\end{aligned}
$$

40

where (a) follows from the fact that $(C_4'/2)u_n \geq C_3'(\bar{p} - \underline{p})/\kappa_n$, (b) follows from a union bound and (c) follows from the fact that $C_4'/(4\bar{p}) \geq 2\eta^{1/2}M^{1/2}$ and Lemma 2. We have now established (B-7).

Step 3. In this last step, we now turn to analyze the revenue function evaluated at $\hat{p} = \max\{\hat{p}^u, \hat{p}^c\}$, which is an estimate of $p^D = \max\{p^u, p^c\}$. Define $C_7' = \max\{2C_1', 2C_4', C_2'(M + \overline{K}\bar{p})\}$. We divide the analysis in two cases: $p^u \geq p^c$ and $p^u < p^c$.

Case 1. Suppose first that $p^u \geq p^c$, i.e., $p^D = p^u$.

i.) If $\hat{p}^u \geq \hat{p}^c$, then $r(\lambda(p^D)) - r(\lambda(\hat{p})) = r(\lambda(p^u)) - r(\lambda(\hat{p}^u))$.

ii.) If $\hat{p}^u < \hat{p}^c \leq p^u$, then $r(\lambda(p^D)) - r(\lambda(\hat{p})) = r(\lambda(p^u)) - r(\lambda(\hat{p}^c)) \leq r(\lambda(p^u)) - r(\lambda(\hat{p}^u))$, where the last inequality follows from the fact that $r(\lambda(\cdot))$ is non-decreasing on $[\hat{p}^u, p^u]$. Since, $C_7' \geq C_4'$, in both i.) and ii.), we have by (B-7) that $\mathbb{P}\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_7'u_n\} \leq C_5'/n^{\eta-1}$.

iii.) The last subcase we analyze is when $\hat{p}^u < \hat{p}^c$ and $\hat{p}^c > p^u$. In this case, note that

$$
\begin{aligned}
\mathbb{P}\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_7'u_n\} &= \mathbb{P}\{r(\lambda(p^u)) - r(\lambda(\hat{p}^c)) > C_4'u_n\} \\
&\leq \mathbb{P}\{r(\lambda(p^u)) - r(\lambda(\hat{p}^c)) > C_7'u_n \; ; \; |\hat{p}^c - p^c| \leq C_2'u_n\} \\
&\quad + \mathbb{P}\{|\hat{p}^c - p^c| > C_2'u_n\} \\
&\overset{(a)}{\leq} \mathbb{P}\{(M + \overline{K}\bar{p})|p^u - \hat{p}^c| > C_7'u_n \; ; \; |\hat{p}^c - p^c| \leq C_2'u_n\} + \frac{C_3'}{n^{\eta-1}} \\
&\overset{(b)}{\leq} \mathbb{P}\{(M + \overline{K}\bar{p})|p^c - \hat{p}^c| > C_7'u_n \; ; \; |\hat{p}^c - p^c| \leq C_2'u_n\} + \frac{C_3'}{n^{\eta-1}} \\
&\overset{(c)}{=} \frac{C_3'}{n^{\eta-1}},
\end{aligned}
$$

where (a) follows the fact that $r(\lambda(\cdot))$ is $(M + \overline{K}\bar{p})$-Lipschitz under Assumption 1, and (B-6). Note that (b) stems from the fact that in case iii.), $|p^u - \hat{p}^c| \leq |p^c - \hat{p}^c|$ since $\hat{p}^c > p^u \geq p^c$ and (c) follows since $C_7'/(M + \overline{K}\bar{p}) \geq C_2'$. In all subcases, we have

$$
\mathbb{P}\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_7'u_n\} \leq \frac{\max\{C_5', C_3'\}}{n^{\eta-1}}.
$$

Case 2. We now turn to the case where $p^u < p^c$, implying that $p^D = p^c$.

i.) If $\hat{p}^c \geq \hat{p}^u$, then $r(\lambda(p^D)) - r(\lambda(\hat{p})) = r(\lambda(p^c)) - r(\lambda(\hat{p}^c))$. Hence,

$$
\mathbb{P}\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_7'u_n\} = \mathbb{P}\{r(\lambda(p^c)) - r(\lambda(\hat{p}^c))u)) > C_7'u_n\} \overset{(a)}{\leq} \frac{C_3'}{n^{\eta-1}},
$$

where (a) follows from (B-5) and the fact that $C_7' \geq C_1'$.

ii.) If $\hat{p}^c < \hat{p}^u$, then $r(\lambda(p^D)) - r(\lambda(\hat{p})) = r(\lambda(p^c)) - r(\lambda(\hat{p}^u)) = r(\lambda(p^c)) - r(\lambda(\hat{p}^c)) + r(\lambda(\hat{p}^c)) - r(\lambda(p^u)) + r(\lambda(p^u)) - r(\lambda(\hat{p}^u)) \leq r(\lambda(p^c)) - r(\lambda(\hat{p}^c)) + r(\lambda(p^u)) - r(\lambda(\hat{p}^u))$, where the inequality follows from the definition of p^u. We deduce that

$$
\begin{aligned}
\mathbb{P}\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_7'u_n\} &\leq \mathbb{P}\{r(\lambda(p^c)) - r(\lambda(\hat{p}^c)) + r(\lambda(p^u)) - r(\lambda(\hat{p}^u)) > C_7'u_n\} \\
&\leq \mathbb{P}\{r(\lambda(p^c)) - r(\lambda(\hat{p}^c)) > C_7'u_n/2\} + \mathbb{P}\{r(\lambda(p^u)) - r(\lambda(\hat{p}^u)) > C_7'u_n/2\} \\
&\overset{(a)}{\leq} \frac{C_5' + C_3'}{n^{\eta-1}},
\end{aligned}
$$

41

where (a) follows from (B-5) and (B-7) and the fact that $C_7' \geq 2\max\{C_1', C_4'\}$.

Putting together the results from cases 1 and 2, we have shown that

$$\mathbb{P}\big\{r(\lambda(p^D)) - r(\lambda(\hat{p})) > C_7' u_n\big\} \quad \leq \quad \frac{C_5' + C_3'}{n^{\eta-1}},$$

which establishes (A-4). The proof is now complete. ∎

Proof of Lemma 4. Put $C_1' = 4\max\{\overline{K}TC_2, M, 2\eta^{1/2}M^{1/2}\}$ where C_2 was defined in Lemma 3. First, using the fact that $Y_n = Y_n^{(L)} + Y_n^{(P)}$, note that

$$\mathbb{P}\big(Y_n - nx > C_1' n u_n\big) \quad \leq \quad \mathbb{P}\big(Y_n^{(L)} > C_1' n u_n/2\big) + \mathbb{P}\big(Y_n^{(P)} > nx + C_1' n u_n/2\big). \qquad \text{(B-9)}$$

Prior to bounding the second term on the RHS above, note that since $p^c - \hat{p} = p^c - \max\{\hat{p}^u, \hat{p}^c\} \leq p^c - \hat{p}^c$ and hence

$$\mathbb{P}\big\{p^c - \hat{p} > C_2 u_n\big\} \quad \leq \quad \mathbb{P}\big\{p^c - \hat{p}^c > C_2 u_n\big\} \leq \frac{C_3}{n^{\eta-1}},$$

where the last inequality follows from Lemma 3. Let $C_2' = \overline{K}C_2$. Recalling that $\lambda(\cdot)$ is monotone decreasing and $\overline{K}$-Lipschitz, we have that $\lambda(\hat{p}) - \lambda(p^c) > C_2' u_n$ implies that $p^c - \hat{p} > C_2 u_n$. Hence

$$\mathbb{P}\Big\{\lambda(\hat{p}) > \lambda(p^c) + C_2' u_n\Big\} \leq \mathbb{P}\big\{p^c - \hat{p} > C_2 u_n\big\} \leq \frac{C_3}{n^{\eta-1}}. \qquad \text{(B-10)}$$

Coming back to the second term on the RHS of (B-9), we have

$$\mathbb{P}\Big(Y_n^{(P)} > nx + C_1' n u_n/2\Big)$$

$$\leq \mathbb{P}\Big(N\big(\lambda(\hat{p})n(T - \tau_n)\big) > nx + C_1' n u_n/2 \,,\, \lambda(\hat{p}) \leq \lambda(p^c) + C_2' u_n\Big) + \mathbb{P}\Big(\lambda(\hat{p}) > \lambda(p^c) + C_2' u_n\Big)$$

$$\overset{(a)}{\leq} \mathbb{P}\Big(N\big((\lambda(p^c) + C_2' u_n)nT\big) > nx + C_1' n u_n/2\Big) + \frac{C_3}{n^{\eta-1}},$$

$$\overset{(b)}{\leq} \mathbb{P}\Big(N\big((x/T + C_2' u_n)nT\big) > nx + C_1' n u_n/2\Big) + \frac{C_3}{n^{\eta-1}},$$

$$= \mathbb{P}\Big(N\big(x(1 + C_2'Tu_n/x)n\big) - x(1 + C_2'Tu_n/x)n > -C_2'Tnu_n + C_1' n u_n/2\Big) + \frac{C_3}{n^{\eta-1}},$$

where (a) follows from (B-10), and (b) follows from the fact that $\lambda(p^c) \leq x/T$. For the latter, note that under the assumption that $\lambda(\overline{p}) \leq x/T$ and since $\lambda(\cdot)$ is continuous and decreasing, either $\lambda(\underline{p}) \geq x/T$ in which case $\lambda(p^c) = x/T$ or $\lambda(\underline{p}) < x/T$ in which case $\lambda(p^c) < x/T$. Now, using the fact that $C_1' - C_2'T \geq C_2'T$, we have

$$\mathbb{P}\Big(N\big((\lambda(p^c) + u_n)nT\big) > nx + C_1' n u_n/2\Big) \quad \leq \quad \mathbb{P}\Big(N\big(x(1 + C_2'Tu_n/x)n\big) - x(1 + C_2'Tu_n/x)n > C_2'Tnu_n\Big)$$

$$\leq \quad \frac{C_3'}{n^{\eta}}, \qquad \text{(B-11)}$$

42

where C_3' is suitably large and we have used Lemma 2 for the last inequality (since $u_n/((\log n)^{1/2}n^{-1/2})$ as $n \to \infty$). The first term on the RHS of (B-9) can be bounded using Lemma 2.

$$
\begin{aligned}
\mathbb{P}\Big(Y_n^{(L)} > C_1' n u_n/2\Big) &= \mathbb{P}\Big(N\Big(\sum_{i=1}^{\kappa_n} \lambda(p_i)n\Delta_n\Big) > C_1' n u_n/2\Big) \\
&\leq \sum_{i=1}^{\kappa_n} \mathbb{P}\Big(N(\lambda(p_i)n\Delta_n) > \frac{C_1' n u_n}{2\kappa_n}\Big) \\
&\overset{(a)}{\leq} \kappa_n \mathbb{P}\Big(N(Mn\Delta_n) - Mn\Delta_n > \frac{C_1' n u_n}{4\kappa_n}\Big) \\
&\overset{(b)}{\leq} \kappa_n \frac{C_3'}{n^\eta} \overset{(c)}{\leq} \frac{C_4'}{n^{\eta-1}},
\end{aligned}
$$

where $C_4' > 0$ is suitably large, (a) follows from the fact that $C_1' n u_n/(4\kappa_n) \geq Mn\Delta_n$, (b) follows from Lemma 2 and the fact that $C_1' n u_n/(4\kappa_n) \geq 2\eta^{1/2}M^{1/2}(\log n)^{1/2}(n\Delta_n)^{1/2}$ and (c) follows from the fact that $\kappa_n = o(n)$. Now, combining the above, we have

$$
\mathbb{P}\big(Y_n - nx > C_1' n u_n\big) \leq \frac{C_3 + C_3' + C_4'}{n^{\eta-1}}. \tag{B-12}
$$

The result is now established. ∎

Proof of Lemma 5. Note that $\mathcal{A} = \{\omega : Y_n^{(P)} \geq nx - C_9 n u_n, \ Y_n^{(L)} \leq C_9 n u_n, \ |\hat{p} - p^D| \leq C_2 u_n\}$. Now,

$$
\begin{aligned}
\mathbb{P}(\mathcal{A}^c) &\overset{(a)}{\leq} \mathbb{P}(Y_n^{(P)} < nx - C_9 n u_n) + \mathbb{P}(Y_n^{(L)} > C_9 n u_n) + \mathbb{P}(|\hat{p} - p^D| > C_2 u_n) \\
&\overset{(b)}{\leq} \mathbb{P}(Y_n^{(P)} < nx - C_9 n u_n) + \mathbb{P}(Y_n^{(L)} > C_9 n u_n) + \frac{C_3}{n^{\eta-1}}, \tag{B-13}
\end{aligned}
$$

where (a) follows from a union bound and (b) follows from Lemma 3 (since $|\hat{p} - p^D| = |\hat{p} - \bar{p}| \leq |\hat{p}^c - \bar{p}| = |\hat{p}^c - p^c|$). We now focus on the first term on the RHS of (B-13).

$$
\begin{aligned}
\mathbb{P}(Y_n^{(P)} < nx - C_9 n u_n) &= \mathbb{P}(N(\lambda(\hat{p})n(T - \tau_n)) < nx - C_9 n u_n) \\
&\leq \mathbb{P}(N(\lambda(\bar{p})n(T - \tau_n)) < n\lambda(\bar{p})T - n\epsilon_n) \\
&\leq \mathbb{P}(N(\lambda(\bar{p})n(T - \tau_n)) - \lambda(\bar{p})n(T - \tau_n) < n(M\tau_n - C_9 u_n)) \\
&\overset{(a)}{\leq} \frac{C_1'}{n^{\eta-1}},
\end{aligned}
$$

where (a) follows from Lemma 2 as long C_9 is chosen large enough and $C_1' > 0$ is suitably large. We now turn to the second term on the RHS of (B-13).

$$
\begin{aligned}
\mathbb{P}(Y_n^{(L)} > C_9 n u_n) &\leq \sum_{i=1}^{\kappa_n} \mathbb{P}(\widehat{\lambda}_i n\Delta_n > C_9 n/\kappa_n u_n) \\
&\leq \sum_{i=1}^{\kappa_n} \mathbb{P}(\widehat{\lambda}_i n\Delta_n - \lambda(p_i)n\Delta_n > C_9 n/\kappa_n u_n - Mn\Delta_n) \\
&\overset{(a)}{\leq} \kappa_n \frac{C_2'}{n^{\eta-1}} \leq \frac{C_2'}{n^{\eta-1}},
\end{aligned}
$$

43

where (a) follows from Lemma 2 as long as C_9 is chosen large enough and $C_2' > 0$ is suitably chosen. Coming back to (B-13), we get

$$\mathbb{P}(\mathcal{A}^c) \ \leq \ \frac{C_1' + C_2' + C_3}{n^{\eta-1}}, \tag{B-14}$$

and the result follows. ∎

Proof of Lemma 6. Consider each interval $[(i-1)\Delta_n, i\Delta_n]$ for $i = 1, ..., k$. Suppose each is divided into $\lfloor n\Delta_n \rfloor$ intervals of length $1/n$ each and measurements are taken every $1/n$ units of time. Let $X_n^i = (x_1^i, ..., x_{\lfloor n\Delta_n \rfloor}^i)$ denote the vector of total demand observed in each interval. Focusing on the ith interval where price p_i is applied, and letting $\mu_i = \lambda(p_i, \theta)$, the log-likelihood function can be written as

$$L_n^i(X_n^i, \mu_i) \ = \ \sum_{j=1}^{\lfloor n\Delta_n \rfloor} \log f_{\mu_i}(x_j^i).$$

This expression is maximized by $\hat{\mu}_i = \sum_{j=1}^{\lfloor n\Delta_n \rfloor} x_j^i/\Delta_n$. This follows since for a Poisson process with unknown constant intensity μ, given n observations of the number of events in non-overlapping intervals of unit length $(y_1, ..., y_n)$, the ML estimate $\hat{\mu}$ is given by $\hat{\mu} = (y_1 + ... + y_n)/n$.

Now note that for a Poisson process with unknown intensity $\mu \in (l_0, M]$, the Fisher information I is bounded and positive, where I is given by $I = \mathbb{E}_\mu \big[\partial \log f_\mu(x)/\partial \mu \big]^2$. In addition, note that for any real number $s \geq 2$, we have $\eta := \mathbb{E}_\mu |\partial \log f_\mu(x)/\partial \mu|^s = \sup_{\mu \in (l_0, M]} \mu^{-s} \mathbb{E}_\mu |x - \mu|^s < \infty$. Hence, under Assumption 2, the conditions of Theorem 3 in Borovkov (1998, II.36) are satisfied and we have that there exists values $0 < c < \infty$ and $\delta > 0$ such that for all v and $n \geq 1$,

$$\mathbb{P}\Big((n\Delta_n)^{1/2} |\hat{\mu}_i - \mu_i^*| \geq v \Big) \leq c e^{-\delta v^2}. \tag{B-15}$$

Setting $\mu_i = \lambda(p_i, \theta^*)$ where θ^* is the true parameter and using condition (i.)(b) in Assumption 2, we deduce that

$$\mathbb{P}\Big((n\Delta_n)^{1/2} \|\hat{\theta} - \theta^*\|_\infty \geq \alpha v \Big) \leq \sum_{i=1}^k \mathbb{P}\Big((n\Delta_n)^{1/2} |\hat{\mu}_i - \mu_i^*| \geq v \Big) \leq ck e^{-\delta v^2}, \tag{B-16}$$

and in turn,

$$\begin{aligned} \mathbb{E}\|\hat{\theta} - \theta^*\|_\infty \ &= \ \int_0^\infty \mathbb{P}\{(n\Delta_n)^{1/2}\|\hat{\theta} - \theta^*\|_\infty > (n\Delta_n)^{1/2}s\}ds \\ &\leq \ a + \int_a^\infty ck e^{-\delta n \Delta_n s^2/\alpha^2} ds \\ &\leq \ a + \frac{ck\alpha^2}{2\delta n \Delta_n a} e^{-\delta n \Delta_n a^2/\alpha^2}. \end{aligned}$$

Taking $a = (n\Delta_n)^{-1/2}$, one gets

$$\mathbb{E}\|\hat{\theta} - \theta^*\|_\infty \leq \frac{C_1'}{(n\tau_n)^{1/2}}, \tag{B-17}$$

where $C_1' = 1 + ck\alpha^2/(2\delta)e^{-\delta/\alpha^2}$. This concludes the proof. ∎

References

Araman, V. F. and Caldentey, R. A. (2005), 'Dynamic pricing for non-perishable products with demand learning', *working paper, New York University* .

Auer, P., Cesa-Bianchi, N., Freund, Y. and Schapire, R. E. (2002), 'The nonstochastic multiarmed bandit problem', *SIAM Journal of Computing* **32**, 48–77.

Aviv, Y. and Pazgal, A. (2005), 'Pricing of short life-cycle products through active learning', *working paper, Washington University* .

Ball, M. and Queyranne, M. (2006), 'Toward robust revenue management: Competitive analysis of online booking', *working paper, University of Maryland* .

Benveniste, A., Métivier, M., and Priouret, P. (1990), *Adaptive algorithms and stochastic approximations*, Springer Verlag.

Bertsimas, D. and Perakis, G. (2003), 'Dynamic pricing: A learning approach', *working paper, Massachusetts Institute of Technology* .

Bitran, G. and Caldentey, R. (2003), 'An overview of pricing models for revenue management', *Manufacturing & Service Operations Management* **5**, 203–229.

Borovkov, A. (1998), *Mathematical Statistics*, Gordon and Breach Science Publishers.

Carvalho, A. X. and Puterman, M. L. (2005), 'Dynamic pricing and learning over short time horizons', *working paper, University of British Columbia* .

Cesa-Bianchi, N. and Lugosi, G. (2006), *Prediction, learning, and games*, Cambridge University Press.

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

Eren, S. and Maglaras, C. (2007), 'Pricing without market information', *working paper, Columbia University* .

Farias, V. F. and Van Roy, B. (2006), 'Dynamic pricing with a prior on market response', *working paper, Stanford University* .

Foster, D. P. and Vohra, R. (1999), 'Regret in the on-line decision problem', *Games and Economic Behavior* **29**, 7–35.

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

Hannan, J. (1957), 'Approximation to Bayes risk in repeated play', *Contributions to the Theory of Games, Princeton University Press* **III**, 97–139.

Huh, W. T. and Rusmevichientong, P. (2006), 'An adaptive algorithm for multiple-fare-class capacity control problems', *working paper, Columbia University* .

Kiefer, J. and Wolfowitz, J. (1952), 'Stochastic estimation of the maximum of a regression function', *Annals of Mathematical Statistics* **23**, 462–466.

Kleinberg, R. and Leighton, F. T. (2003), 'The value of knowing a demand curve: Bounds on regret for online posted-price auctions', *Proc. of the 44th Annual IEEE Symposium on Foundations of Computer Science* .

Kushner, H. and Sanvicente, E. (1975), 'Stochastic approximation of constrained systems with system and constraint noise', *Automatica* **11**, 375–380.

Lai, T. L. and Robbins, H. (1985), 'Asymptotically efficient adaptive allocation rules', *Advances in Applied Mathematics* **6**, 4–22.

Lim, A. and Shanthikumar, J. (2006), 'Relative entropy, exponential utility, and robust dynamic pricing', *Forthcoming Operations Research* .

Lobo, M. S. and Boyd, S. (2003), 'Pricing and learning with uncertain demand', *working paper, Duke University* .

Perakis, G. and Roels, G. (2006), 'Distribution-free revenue management', *6th Informs Revenue Management Conference, June 2006, http://www.demingcenter.com* .

Polyak, B. T. and Juditsky, A. B. (1990), 'Acceleration of stochastic approximation by averaging', *SIAM J. Control and Optimization* **30**, 838–855.

Polyak, B. T. and Tsybakov, A. (1990), 'Optimal order of accuracy of search algorithms in stochastic optimization', *Problems of Information Transmission* **26**, 126–133.

Robbins, H. (1952), 'Some aspects of the sequential design of experiments', *Bull. Amer. Math. Soc.* **58**, 527–535.

Robbins, H. and Monro, S. (1951), 'A stochastic approximation method', *Annals of Mathematical Statistics* **22**, 400–407.

Rusmevichientong, P., Van Roy, B. and Glynn, P. W. (2006), 'A non-parametric approach to multi-product pricing', *Operations Research* **54**, 82–98.

Scarf, H. (1959), 'Bayes solutions of the statistical inventory problem', *Annals of Mathematical Statistics* **30**, 490–508.

Talluri, K. T. and van Ryzin, G. J. (2005), *Theory and Practice of Revenue Management*, Springer-Verlag.

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