Manufacturing & Service Operations Management

Bayesian Dynamic Pricing in Queueing Systems with Unknown Delay Cost Characteristics

Philipp Afèche, Barış Ata,

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 15, No. 2, Spring 2013, pp. 292–304
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)

http://dx.doi.org/10.1287/msom.1120.0418

Bayesian Dynamic Pricing in Queueing Systems with Unknown Delay Cost Characteristics

Philipp Afèche
Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada,
afeche@rotman.utoronto.ca

Barış Ata
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
b-ata@kellogg.northwestern.edu

The revenue management literature for queues typically assumes that providers know the distribution of customer demand attributes. We study an observable $M/M/1$ queue that serves an *unknown* proportion of patient and impatient customers. The provider has a Bernoulli prior on this proportion, corresponding to an optimistic or pessimistic scenario. For every queue length, she chooses a low or a high price, or turns customers away. Only the high price is informative. The optimal Bayesian price for a queue state is belief-dependent if the optimal policies for the underlying scenarios disagree at that queue state; in this case the policy has a belief-threshold structure. The optimal Bayesian pricing policy as a function of queue length has a zone (or, nested-threshold) structure. Moreover, the price convergence under the optimal Bayesian policy is sensitive to the system size, i.e., the maximum queue length. We identify two cases: prices converge (1) almost surely to the optimal prices in either scenario or (2) with positive probability to suboptimal prices. Only Case 2 is consistent with the typical *incomplete learning* outcome observed in the literature.

Key words: Bayesian learning; delay; dynamic pricing; revenue management; queueing
History: Received: November 7, 2005; accepted: September 16, 2012. Published online in *Articles in Advance* March 1, 2013.

1. Introduction

Virtually the entire revenue management literature for queues assumes that providers have perfect information on the distribution of customer demand attributes (see Hassin and Haviv 2003). This paper relaxes this standard assumption and studies optimal dynamic pricing under Bayesian updating in the case of uncertainty with respect to the distribution of customers' delay cost parameters. We address two fundamental questions that arise in such settings. First, how should pricing decisions be made to optimally trade off revenue generation and demand learning? Second, what is the probability that the prices under the optimal Bayesian policy converge to the prices that are optimal for the underlying demand scenario?

We study these questions in the context of an $M/M/1$ queueing system that serves an unknown proportion of two customer types, patient versus impatient, who differ in their costs of waiting for service. The provider cannot tell apart patient types from impatient ones and, therefore, cannot price discriminate based on customer identity. She has a Bernoulli prior on the proportion of patient versus impatient customers that corresponds to an optimistic or pessimistic demand scenario. She updates this prior depending on whether customers buy or not at the posted prices. Customers observe the queue length upon arrival. For every queue length there is a low price that all customers are willing to pay and a high price that only patient customers are willing to pay. Consequently, only the high price is informative and allows learning; we assume that lost sales are observed. The provider's decision is to choose for every queue length between the corresponding low price, the high price, and rejecting customers. She seeks to maximize expected discounted revenues over an infinite horizon by varying the price depending on the queue length and her updated prior, so as to optimally trade off the revenue impact against the value of learning.

This paper provides what appears to be the first analysis of the learning-and-earning problem for a capacity-constrained operation that serves heterogeneous time-sensitive customers whose delay costs are drawn from an unknown distribution. As such, we view the paper's primary contributions in the novelty of the problem formulation and the insights it generates on the structure and price convergence of the optimal Bayesian pricing policy.

2. Literature Review

Our model is closely related to that in the seminal paper of Naor (1969). He studies *static* pricing for an observable $M/M/1$ queue, assuming *identical* customers with a (known) linear delay cost, and shows that the revenue-maximizing price (weakly) exceeds the socially optimal price. Chen and Frank (2001) show that under dynamic pricing, the revenue-maximizing and socially optimal policies agree in Naor's model. They also show numerical results for the case with a *known* proportion of two customer types with identical delay costs but different valuations for service.

The queueing literature on pricing problems under imperfect aggregate demand information is sparse. Masuda and Whang (1999) consider social optimization for a network provider who knows customers' (identical) linear delay costs but lacks full knowledge of the relationship between arrival rates and service valuations. They assume that customers do not observe the queues and study adaptive pricing heuristics and delay expectation models. Besbes and Maglaras (2009) consider a setting where the market size is unobservable but varying slowly. By exploiting the separation of time scales, they develop good policies based on fluid approximations. Haviv and Randhawa (2012) consider uninformed *static* pricing without using demand rate information. They show that the optimal uninformed price can have excellent revenue performance. In both of these studies, unlike in ours, the provider is perfectly informed about the distribution of customer preferences.

Outside the queueing literature, there are many papers on decision making and demand learning. We distinguish two main streams, on *dynamic pricing without supply constraints*, which originated in the economics literature, and on *inventory control and/or pricing with finite inventories*, which originated in the operations research and management science (OR/MS) literature. In problems without supply constraints, the only connection between time periods occurs through beliefs. Our setup gives rise to two state variables, the belief and the queue length, and a sale may affect both. Papers on pricing with finite inventories also deal with two state variables, the belief and the inventory level, and a sale may affect both variables. However, in these papers, the consumer's response to a given price does *not* depend on the inventory level (so long as there is any inventory), whereas in our setup, the consumer's response to a given price *does* depend on the queue length.

Dynamic pricing without supply constraints. There are many economics papers on maximizing the payoff earned while simultaneously learning, which is often referred to as the learning-and-earning problem, first studied by Rothschild (1974). He considers a seller who chooses among two prices and shows that the Bayesian optimal decision may not eventually coincide with the true optimal decision when the underlying demand scenario is known. Moreover, a seller who follows the ex ante optimal policy may never learn the true demand curve. This result is referred to as *incomplete learning* and applies to multiarmed bandit formulations more generally (see Banks and Sundaram 1992, Brezzi and Lai 2002). Keller and Rady (1999) consider a firm facing an unobserved demand curve that switches between two alternatives. They identify two optimal experimentation regimes, one where the firm switches between the optimal actions corresponding to the two demand curves, and another where it is trapped in an uninformative set of actions.

The incomplete learning outcome is common in the literature. Two notable exceptions are Mersereau et al. (2009) and Easley and Kiefer (1988). Mersereau et al. (2009) consider a multiarmed bandit problem with correlated arms (so that every action is informative) and establish *complete learning*, i.e., learning with probability one. Easley and Kiefer (1988) establish complete learning in the absence of confounding beliefs (which is a slight generalization of every action being informative). (See also Aghion et al. 1991, who show that learning happens with probability one when there is no noise or no discounting.)

In our setting, learning can be complete or incomplete, depending on the system size, i.e., the maximum queue length under the optimal Bayesian pricing policy. Moreover, our problem is fundamentally different from those in the learning papers discussed above. In these papers the optimal policy for each known parameter value consists of a *single* decision. In our setup, the optimal policy for each known scenario comprises *multiple* prices, one for each queue length.

The learning-and-earning problem without supply constraints has also been studied in the OR/MS literature. Lobo and Boyd (2003) derive approximate solutions using convex programming and simulations. Harrison et al. (2012a) study a dynamic pricing problem with model uncertainty, where a seller has a binary prior on two demand curves. They show that the myopic Bayesian policy fails to learn the true demand curve with positive probability (i.e., the incomplete learning phenomenon occurs). They further show that this can be avoided by enhancing their myopic policy with a constraint on the minimum price deviations, and that regret associated with this policy is bounded by a constant as the planning horizon gets large. The framework of den Boer and Zwart (2011) involves a demand model with two unknown parameters. The authors analyze a simple policy and derive an upper bound on the asymptotic regret of the analyzed policy. Broder and Rusmevichientong (2012) consider a general parametric demand model to evaluate maximum-likelihood-based policies and develop bounds on the regret under their proposed pricing policies. Harrison et al. (2012b) provide a unified set of conditions to achieve near optimal performance in dynamic pricing problems with demand model uncertainty.

Inventory control and/or pricing with finite inventories. The study of Bayesian learning models of inventory management started with Scarf (1959), who considers a nonperishable item with unknown demand distribution but observable customer demand (see also Azoury 1985, Lovejoy 1990). These papers focus on reducing the computational complexity of the problem. Harpaz et al. (1982) consider a perishable item with unobserved lost sales and recognize that one should increase the inventory level to learn the demand distribution. Chen and Plambeck (2008) show that this "stock more" result can be reversed, in the case of nonperishable inventory with unobserved lost sales, and if lost sales are observable and one wishes to learn about demand substitution rates.

The last decade has seen a growing interest in the OR/MS literature in Bayesian dynamic pricing and learning models in the presence of inventory constraints. Petruzzi and Dada (2002) consider an optimal pricing, demand learning, and inventory control problem in discrete time. Aviv and Pazgal (2005) use a partially observed Markov decision process framework to study the dynamic pricing problem of a seller who owns a finite stock of goods and sells them without replenishment over a finite horizon. Farias and van Roy (2010) and Araman and Caldentey (2009) consider infinite horizon versions of this problem. Besbes and Zeevi (2009) study a similar problem via a non-Bayesian approach and minimize regret in an asymptotic framework.

3. Model

We consider a capacity-constrained operation, modeled as a first-in-first-out $M/M/1$ system with arrival rate λ and service rate μ. Let n denote the queue length, including the customer in service. All customers get an identical reward R that equals their willingness to pay for instant service without waiting. The customer population comprises *patient* and *impatient* customers. Patient (impatient) customers incur a delay cost c_L (c_H) per unit of time waiting for (but not during) service, where $c_L < c_H$. We assume that $c_H < R\mu$ to rule out trivial cases. The proportion of patient customers is $q \in (0, 1)$. A given q specifies what we call a *demand scenario*.

The provider knows the parameters λ, μ, R, c_L, and c_H, and observes all customer arrivals, including lost sales. We consider both the benchmark case where the provider knows the demand scenario q, and the case where she does not know the true scenario. In the latter setting, we consider Bayesian updating of a prior belief with a Bernoulli distribution over two conceivable demand scenarios, *pessimistic* and *optimistic* with fractions of patient customers q_p and q_o, respectively, where $q_p < q_o$. Under both known and unknown scenarios, the provider cannot tell apart patient customers from impatient ones and, therefore, cannot price discriminate based on customer identity. The provider seeks to maximize her expected discounted revenues over an infinite horizon by choosing the price p depending on the system state. Let $\beta > 0$ denote the discount rate.

Customers know the service rate μ and observe the price p and the queue length n. They buy the service (and join the queue) if and only if their expected net value from service, i.e., $R - n(c_H/\mu)$ or $R - n(c_L/\mu)$, weakly exceeds the price. Customers do not renege and we assume no retrials.

The Bayesian pricing problem does not seem analytically tractable for a system without buffer limit, which we call a *general buffer system*. Therefore, we first study a small buffer system analytically and generate various structural insights in §4. We first study the known scenario benchmark in that case (§4.1) and build on the insights gleaned for the analysis of the unknown scenario (§4.2). In §5, we consider a general buffer system. We analytically characterize the optimal policy for the known scenario benchmark as a nested threshold policy (§5.1). Then, combining the insights from this benchmark and a numerical study, we identify and discuss the properties of the optimal Bayesian pricing policy in §5.2.

4. Analysis of a Small Buffer System

The small buffer system has a maximum queue length of two. We refer to the three states of the queue length n as *empty* $(n = 0)$, *congested* $(n = 1)$ when there is a job in service and none is waiting, and *full* $(n = 2)$ when one job is being served and one is waiting for service. Any customer who arrives to an empty system is willing to pay up to R. A customer arriving to a congested system is willing to pay up to $P_H := R - (c_L/\mu)$ if he is patient and up to $P_L := R - (c_H/\mu)$ if he is impatient, where $P_H > P_L$. The only rational prices at which to admit a customer when the system is congested are P_H and P_L. Similarly, R is the only rational price to charge when the system is empty. When the system is full, all arriving customers are turned away.

4.1. Optimal Dynamic Pricing Under Known Demand Scenario

Under known demand scenario, the system state is the queue length n and the state space is $\mathcal{N} = \{0, 1, 2\}$. Without loss of generality we restrict attention to stationary pricing policies. It is optimal to charge $p = R$ when the system is empty. The provider is required to turn customers away when the system is full; we define the reject price $P_R := R$. Therefore, the provider's policy is fully specified by the price $p \in \{P_L, P_H, P_R\}$ that she charges in the congested system. We denote a policy by $\pi \in \{l, h, r\}$, where l specifies to *price low* ($p = P_L$), h to *price high* ($p = P_H$), and r to *reject* customers ($p = P_R$) when the system is congested.

Let v_n be the infinite horizon expected net present revenue under the optimal pricing policy starting in state $n \in \mathcal{N}$. Following the standard solution approach, one obtains the following Bellman equation for the uniformized system (see Bertsekas 1995):

$$(\beta+\lambda)v_0 = \lambda[R+v_1], \tag{1}$$

$$\begin{aligned}(\beta+\lambda+\mu)v_1 = \max\{&\lambda[P_L+v_2]+\mu v_0,\ \lambda q[P_H+v_2]\\ &+\lambda(1-q)v_1+\mu v_0,\ \lambda v_1+\mu v_0\},\end{aligned} \tag{2}$$

$$(\beta+\mu)v_2 = \mu v_1. \tag{3}$$

As a stepping stone for the analysis of dynamic pricing under unknown demand scenario, Proposition 1 explicitly characterizes the solution of (1)–(3) and the corresponding optimal policy as a function of the problem parameters. Let v_n^π denote the expected net present revenue starting in state $n \in \mathcal{N}$ under policy $\pi \in \{l, h, r\}$. For each policy $\pi \in \{l, h, r\}$, v_n^π satisfies a set of equations similar to (1)–(3) with the right-hand side of (2) replaced by the appropriate term.

Proposition 1. *The expected net present revenue of policy $\pi \in \{l,\ h,\ r\}$ satisfies $v_0^\pi = \frac{\lambda}{\beta+\lambda}v_1^\pi + \frac{\lambda}{\beta+\lambda}R$ and $v_2^\pi = \frac{\mu}{\beta+\mu}v_1^\pi$, where*

$$v_1^r = \frac{\frac{\mu}{\beta+\mu}\frac{\lambda}{\beta+\lambda}R}{1-\frac{1}{\beta+\mu}\frac{\lambda\mu}{\beta+\lambda}}, \tag{4}$$

$$v_1^h(q) = \frac{\frac{1}{\beta+\lambda q+\mu}\big[\lambda q P_H + \mu\frac{\lambda}{\beta+\lambda}R\big]}{1-\frac{1}{\beta+\lambda q+\mu}\big[\lambda q\frac{\mu}{\beta+\mu}+\mu\frac{\lambda}{\beta+\lambda}\big]}, \tag{5}$$

$$v_1^l = \frac{\frac{1}{\beta+\lambda+\mu}\big[\lambda P_L + \mu\frac{\lambda}{\beta+\lambda}R\big]}{1-\frac{1}{\beta+\lambda+\mu}\big[\lambda\frac{\mu}{\beta+\mu}+\mu\frac{\lambda}{\beta+\lambda}\big]}. \tag{6}$$

The expected net present revenue of the optimal policy in the congested state ($n = 1$) satisfies $v_1 = \max\{v_1^l,\ v_1^h(q), v_1^r\}$. The optimal action in the congested state is as follows:

$$\textit{if } R \le R_L,\ \textit{then reject customers for all } q \in (0,1), \tag{7}$$

$$\textit{if } R \in (R_L, R_H],\ \textit{then price high for all } q \in (0,1), \tag{8}$$

$$\begin{aligned}&\textit{if } R > R_H,\ \textit{then price low if } q \le \underline{q}\\ &\qquad\qquad \textit{and price high if } q > \underline{q},\end{aligned} \tag{9}$$

where

$$R_L = \frac{c_L/\mu}{1-\frac{\lambda}{\lambda+\mu+\beta}\frac{\mu}{\mu+\beta}}, \tag{10}$$

$$R_H = \frac{c_H/\mu}{1-\frac{\lambda}{\lambda+\mu+\beta}\frac{\mu}{\mu+\beta}}, \tag{11}$$

$$\underline{q} = \frac{(R-R_H)\big(1-\frac{\lambda}{\beta+\lambda+\mu}\frac{\mu}{\beta+\mu}\big)}{(R-R_L)\big(1-\frac{\lambda}{\beta+\lambda+\mu}\frac{\mu}{\beta+\mu}\big)\big(\frac{c_H}{\mu}-\frac{c_L}{\mu}\big)\frac{\lambda}{\beta+\lambda+\mu}\frac{\beta+\lambda}{\beta+\mu}}. \tag{12}$$

The expected net present revenues under the three policies starting in the congested state $n = 1$ are given by (4)–(6). The numerators capture the respective expected net present revenues during the first cycle, starting in the congested state until return to that state. Naturally, the expected net present revenues from pricing low or rejecting customers do not depend on the scenario q because all (no) customers are willing to buy under the former (latter) policy. However, $v_1^h(q)$, the expected discounted revenue from pricing high in the congested state, is a function of q. Observe that $v_1^h(0) = v_1^r$ because pricing high is tantamount to rejecting all customers if none are patient and that $v_1^h(1) > v_1^l$ because pricing high yields the same number of sales as pricing low when all customers are patient and therefore generates more revenue.

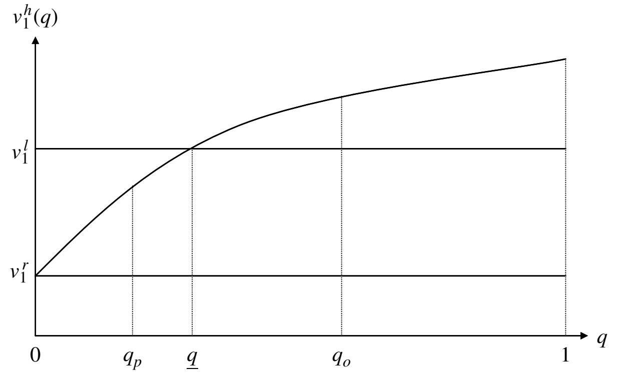
The optimal policy specified by (7)–(12) has the following intuition. In choosing the price for the congested state, the provider must appropriately trade off current and future revenues, i.e., selling now at a price $R - c_L/\mu$ or $R - c_H/\mu$ versus later at a higher price R when the system is empty. If the reward R is below the threshold R_L, the expected delay cost of both customer types is so high that it is worth delaying a sale until the system empties; i.e., the optimal policy is to turn away both types in the congested state. If the reward R lies between the thresholds R_L and R_H, then it is only worth selling to the patient customers (at high price) in the congested state, whereas the impatient customers' delay cost c_H is so high that the provider is better off selling to them only at a higher price when the system is empty. By (8) and (10)–(11), the larger the difference between the impatient and patient customers' delay costs, the larger the range of rewards for which this policy is optimal. Only if R exceeds R_H can it be optimal to sell to both types by pricing low in the congested state. Whether doing so is optimal depends on the scenario q: price high if the proportion of patient customers is sufficiently high ($q > \underline{q}$) and price low otherwise.

The only parameter combinations of interest for the Bayesian dynamic pricing problem are those that yield a different optimal policy for the pessimistic and optimistic scenarios. Based on Proposition 1, these parameters satisfy

$$R > R_H \quad \text{and} \quad q_p < \underline{q} < q_o; \tag{13}$$

i.e., the optimal price is low in the pessimistic scenario but high in the optimistic scenario. Figure 1

Figure 1 Parameters of Interest for Dynamic Pricing Under Unknown Scenario Yield $v_1^h(q_p) < v_1^l < v_1^h(q_o)$



Note. Optimal policy is to price low or high depending on whether the scenario is pessimistic or optimistic.

illustrates this case of interest that we focus on in the next section.

4.2. Optimal Dynamic Pricing Under Unknown Demand Scenario

The provider knows that the scenario is either optimistic with q_o or pessimistic with q_p, where $q_p < \underline{q} < q_o$, but is uncertain about which one it is. She has a Bernoulli prior on $\{q_p, q_o\}$; that is, she assigns a prior probability α to the optimistic scenario and $1-\alpha$ to the pessimistic scenario. We allow $\alpha \in [0,1]$ to provide a unified treatment of the known and unknown scenario cases.

The provider only needs to choose a price when the system is congested because it is always optimal to sell at a price $p = R$ when the system is empty and customers must be turned away when the system is full. Without loss of generality we restrict attention to stationary Markov pricing policies (see Ritt and Sennott 1992). Let $\mathcal{S} = \mathcal{N} \times [0,1]$ denote the state space and $(n, \alpha) \in \mathcal{S}$ describe the state of the system. The provider's pricing policy may depend on her prior α. Based on the analysis for known demand scenario, when $q_p < \underline{q} < q_o$ and $R > R_H$ the optimal policy is to price low or high in the congested state depending on whether the scenario is pessimistic or optimistic, respectively, whereas turning customers away is suboptimal in either scenario. This section focuses on this case. We denote a stationary pricing policy by a function $p(\cdot)\colon [0,1] \to \{P_L, P_H\}$; that is, $p(\alpha)$ specifies the high or low price when the system is congested $(n=1)$ and the prior is α.

We show that the optimal dynamic pricing policy p^* has a simple threshold structure, which is to price high $(p = P_H)$ and sell only to patient customers if the provider's prior α is greater than a strictly positive threshold α^*, and to price low $(p = P_L)$ and accept all customers otherwise.

Let $v_n(\alpha)$ be the infinite horizon expected net present revenue under the optimal pricing policy starting in state $(n, \alpha) \in \mathcal{S}$. Denote by $\bar{\alpha}$ and $\underline{\alpha}$, respectively, the Bayesian posteriors corresponding to a prior α depending on whether a customer buys or not at the high price when the system is congested. They satisfy

$$\bar{\alpha} = \frac{\alpha}{\alpha + (1-\alpha)(q_p/q_o)} \quad \text{and} \quad \underline{\alpha} = \frac{\alpha}{\alpha + (1-\alpha)((1-q_p)/(1-q_o))}. \tag{14}$$

Following the standard solution approach for infinite horizon continuous time problems with a discounted cost criterion, we arrive at the following Bellman equation for the uniformized system:

$$(\beta + \lambda) v_0(\alpha) = \lambda \cdot [R + v_1(\alpha)], \tag{15}$$

$$\begin{aligned}(\beta + \mu + \lambda) v_1(\alpha) &= \max\{\lambda \cdot [P_L + v_2(\alpha)] + \mu \cdot v_0(\alpha), \\ &\quad \lambda(\alpha q_0 + (1-\alpha) q_p) \cdot [P_H + v_2(\bar{\alpha})] \\ &\quad + \lambda(\alpha(1-q_o) + (1-\alpha)(1-q_p)) \cdot v_1(\underline{\alpha}) \\ &\quad + \mu \cdot v_0(\alpha)\},\end{aligned} \tag{16}$$

$$(\beta + \mu) v_2(\alpha) = \mu v_1(\alpha), \tag{17}$$

for $\alpha \in [0,1]$. (We omit the term corresponding to rejecting customers when $n = 1$ because this action is suboptimal for the parameter regime of interest.) For notational efficiency define $q(\alpha) := \alpha q_o + (1-\alpha) q_p$ to be the expected fraction of patient customers when the provider's updated prior is α. Using (15) and (17) to substitute for $v_o(\alpha)$ and $v_2(\alpha)$ into (16) yields

$$\begin{aligned}(\beta + \mu + \lambda) v_1(\alpha) &= \max\Big\{\lambda\Big[P_L + \tfrac{\mu}{\beta+\mu} v_1(\alpha)\Big] + \mu \tfrac{\lambda}{\beta+\lambda}[R + v_1(\alpha)], \\ &\quad \lambda q(\alpha)\Big[P_H + \tfrac{\mu}{\beta+\mu} v_1(\bar{\alpha})\Big] + \lambda(1 - q(\alpha)) v_1(\underline{\alpha}) \\ &\quad + \mu \tfrac{\lambda}{\beta+\lambda}[R + v_1(\alpha)]\Big\}\end{aligned} \tag{18}$$

for $\alpha \in [0,1]$. In deciding whether to price low $(p = P_L)$ or high $(p = P_H)$ when the system is congested $(n = 1)$, the provider faces the following trade-off between revenue generation and the value of learning. Starting in the congested state, the next sale occurs either in that state, at a price P_L or P_H, or in the empty system at the higher price R. Compared with pricing low, a high price may delay the time until the next sale and reduce the relative probability that this sale occurs in the congested system because only a fraction $q(\alpha)$ of customers are willing to pay P_H. From a revenue perspective, pricing high is, therefore, only beneficial if the fraction of patient customers $q(\alpha)$ is sufficiently high. However, only a high price generates information about the true fraction of patient customers, whereas all customers behave the same under low pricing. Therefore, the cost of losing a current sale as a result of pricing high in the congested system may be offset by the information value of this lost sale: by updating her prior about the demand scenario, and ultimately her pricing, the provider may be able to increase future revenues.

To simplify (18), observe that if pricing low ($p = P_L$) is optimal for some α, then $v_1(\alpha) = v_1^l$, as given by (6): revenues under the low price do not depend on α because all customers purchase at that price. Setting $J(\alpha) := v_1(\alpha)$ for all α to economize on notation yields

$$\begin{aligned} J(\alpha) = \max\big\{v_1^l,\, \big(\lambda q(\alpha)\big[P_H + \tfrac{\mu}{\beta+\mu} J(\bar{\alpha})\big] \\ + \lambda(1-q(\alpha))J(\underline{\alpha}) + \mu \tfrac{\lambda}{\beta+\lambda}[R + J(\alpha)]\big) \\ \cdot \tfrac{1}{\beta+\lambda+\mu}\big\}, \quad \alpha \in [0,1]. \end{aligned} \tag{19}$$

Proposition 2 characterizes the structure of the value function J.

Proposition 2. *The Bellman equation* (19) *has a unique continuous solution J. It is nondecreasing and satisfies $J(0) = v_1^l$, $J(1) = v_1^h(q_o)$, and*

$$J(\alpha) \le P_H \frac{\beta+\mu}{\beta} \quad \textit{for } 0 \le \alpha \le 1. \tag{20}$$

Proposition 3 establishes that a threshold pricing policy is optimal.

Proposition 3. *If $R > R_H$ and $q_p < q < q_o$, then the optimal pricing policy $p^*(\cdot)$ has a threshold structure. It satisfies*

$$p^*(\alpha) = \begin{cases} P_H & \textit{if } \alpha > \alpha^*, \\ P_L & \textit{if } \alpha \le \alpha^*, \end{cases} \tag{21}$$

where $\alpha^* := \sup\{0 \le \alpha \le 1\colon J(\alpha) = v_1^l\}$ and $0 < \alpha^* < 1$.

Considering that pricing high allows the provider to update her beliefs, whereas pricing low does not, setting a high price can be interpreted as experimenting or learning. Because the threshold α^* is strictly positive, the provider will stop experimenting with positive probability at some finite point in time and subsequently will charge the low price forever, even if the underlying scenario is optimistic. This phenomenon is referred to in the literature as *incomplete learning*, i.e., the provider never learns the true demand scenario (see Rothschild 1974). We study this further in §6.

The pricing problem considered in this section can be viewed as a generalization of a two-armed bandit problem, where one arm corresponds to pricing high, and the other to pricing low. However, our setup differs from the standard multiarmed bandit formulation in that the time until the provider can pull an arm again is random and depends on the outcome of the previous decision. An alternate solution approach involves modifying the existing theory (e.g., Tsitsiklis 1994) to accommodate an infinite state space, characterizing the corresponding Gittins indices, and showing their equivalence to a threshold rule on the prior. We chose to study the problem from first principles because this approach is more direct. More importantly, because of the underlying queueing dynamics, there is no straightforward way to map the Bayesian pricing problem for a general buffer system to a multiarmed bandit problem by appropriately defining arms. In contrast, the dynamic programming formulation (15)–(17) readily extends to a general buffer system; see §5.2.

5. General Buffer System

In §5.1, we characterize the optimal policy for the known scenario benchmark as a nested threshold policy. It appears that this result is new and may be of interest in its own right. In §5.2, we discuss the Bayesian pricing problem under unknown scenario, which does not seem analytically tractable for a general buffer system. However, our result for the known scenario benchmark, combined with a numerical study, allow us to clearly identify the structural properties of the optimal Bayesian pricing policy.

5.1. Optimal Dynamic Pricing Under Known Demand Scenario

The system state is the queue length $n \in \mathbb{N}$. We assume that the system is initially empty. As done earlier, we restrict attention to stationary pricing policies. In each state $n \ge 1$, the provider chooses among pricing low, pricing high, or rejecting customers by charging prices $P_L(n) := R - n(c_H/\mu)$, $P_H(n) := R - n(c_L/\mu)$, and $P_R = R$, respectively. Pricing low corresponds to admitting all customers, and pricing high corresponds to admitting only the patient customers.

Let v_n denote the infinite horizon expected net present revenue under the optimal pricing policy, starting with n customers. We have the following Bellman equation for the uniformized system:

$$(\beta+\lambda)v_0 = \lambda[R + v_1], \tag{22}$$

$$\begin{aligned} &(\beta+\lambda+\mu)v_n \\ &\quad = \max\{\mu v_{n-1} + \lambda[P_L(n) + v_{n+1}], \\ &\qquad \mu v_{n-1} + \lambda q[P_H(n) + v_{n+1}] + \lambda(1-q)v_n, \\ &\qquad \mu v_{n-1} + \lambda v_n\}, \qquad n \ge 1. \end{aligned} \tag{23}$$

Defining $\Delta_n := v_{n+1} - v_n$ for $n \ge 0$, and rearranging terms leads to the following simplified equation:

$$(\beta+\lambda)v_0 = \lambda v_0 + \lambda[R + \Delta_0], \tag{24}$$

$$(\beta+\lambda+\mu)v_n=\mu v_{n-1}+\lambda v_n +\lambda\max\{P_L(n)+\Delta_n,\, q[P_H(n)+\Delta_n],0\},\quad n\ge 1. \tag{25}$$

Define $\bar{n} := \lceil R\mu/c_L\rceil$, which is the shortest queue length at which no customer is willing to pay a strictly positive price, i.e., $P_L(n) < P_H(n) \le 0$ for $n \ge \bar{n}$. Furthermore, let

$$f(n) := \frac{qP_H(n)-P_L(n)}{1-q}, \quad n\ge 1. \tag{26}$$

The following condition will be used to prove the main result of this section:

$$\frac{1-q}{q}\,\frac{\lambda-\beta}{\beta} \le \frac{c_H-c_L}{c_H}, \tag{27}$$

which is only a sufficient condition. Numerical examples suggest that Propositions 4 and 5, which characterize the optimal pricing policy under known scenario, also hold when (27) is violated.

Proposition 4. *For $n\ge 0$, we have that $\Delta_n \le 0$. If* (27) *holds, then $\Delta_n - f(n)$ is decreasing.*

To state the main result of this section, define the queue length thresholds

$$n_h := \inf\{n\ge 1\colon\, \Delta_n - f(n)\le 0\}, \tag{28}$$

where $\inf\varnothing=\infty$, and

$$n_r := \inf\{n\ge n_h\colon\, \Delta_n + P_H(n)\le 0\}. \tag{29}$$

Proposition 5. *Assume* (27) *holds. We have that $0 < n_h \le n_r \le \bar{n}$, and the optimal pricing policy p^* is given by*

$$p^*(n)=\begin{cases} P_L(n) & \text{for } n<n_h,\\ P_H(n) & \text{for } n_h\le n<n_r,\\ P_R & \text{for } n\ge n_r. \end{cases} \tag{30}$$

This nested threshold structure also holds in the small buffer system; see (7)–(9) in Proposition 1. The intuition for this result is that, as the queue length increases, the profitability of both customer types decreases and impatient types become relatively less profitable than patient ones. Similar policies have been shown to be optimal in the queueing control literature; for example, Ata (2006) proves for a multiclass admission control problem without delay costs the (asymptotic) optimality of a policy that rejects successively more expensive classes as the queue length increases. However, unlike in that problem, the value function is not concave in ours. The proof of Proposition 5, therefore, involves additional challenges, and it does not seem to be covered by existing results.

5.2. Optimal Dynamic Pricing Under Unknown Demand Scenario

Letting $v_n(\alpha)$ denote the expected net present revenue under the optimal pricing policy given queue length n and prior $\alpha\in[0,1]$, we obtain the following Bellman equation for the uniformized system:

$$(\beta+\lambda)v_0(\alpha)=\lambda[R+v_1(\alpha)], \tag{31}$$

$$\begin{aligned}(\beta+\mu+\lambda)v_n(\alpha) &=\max\{\mu v_{n-1}(\alpha)+\lambda[P_L(n)+v_{n+1}(\alpha)],\\ &\quad \mu v_{n-1}(\alpha)+\lambda q(\alpha)[P_H(n)+v_{n+1}(\bar{\alpha})]\\ &\qquad +\lambda(1-q(\alpha))v_n(\underline{\alpha}),\\ &\quad \mu v_{n-1}(\alpha)+\lambda v_n(\alpha)\},\end{aligned} \tag{32}$$

for $\alpha\in[0,1]$ and $n\ge1$, where $q(\alpha)=\alpha q_o+(1-\alpha)q_p$, and $\underline{\alpha}$ and $\bar{\alpha}$ are the updated priors given by (14). For $\alpha=0$, (31)–(32) specialize to (22)–(23) with $q=q_p$, and similarly for $\alpha=1$ and $q=q_o$.

The Bellman equation appears to be analytically intractable. We, therefore, compute the optimal Bayesian policy (using linear programming) for a set of test problems and discuss its properties in relation to the optimal pricing policies for the underlying known demand scenarios.

5.2.1. Known Scenario: Sensitivity of the Optimal Policy to the Fraction of Patient Types q.

By Proposition 5, the two queue length thresholds $n_h\le n_r$ fully characterize the optimal pricing policy for a given known scenario. We, henceforth, write $n_h(q)$, $n_r(q)$, and $p^*(n;q)$ to emphasize their dependence on the scenario q. To determine how these thresholds depend on q, we solve the dynamic program (24)–(25) for a number of test problems covering a wide range of parameter combinations. In each case, we fix the parameters $R, c_L, c_H, \beta, \lambda$, and μ and compute the solution for $q\in[0,1]$. The following observation summarizes the results of this numerical study.

Observation 1. *Fix all system parameters except for q.*

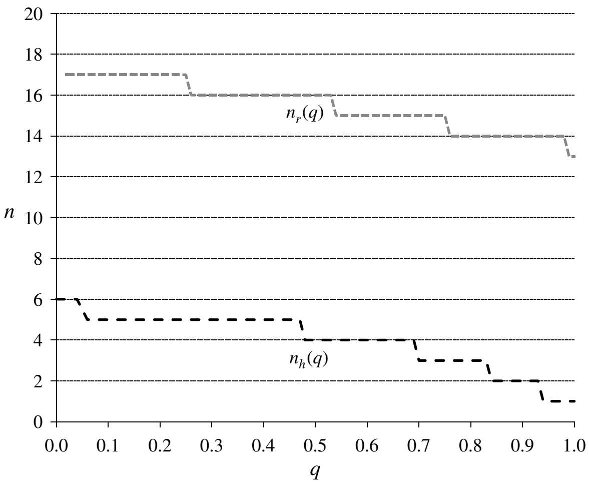
1. *The optimal price-high threshold $n_h(q)$ and the optimal reject threshold $n_r(q)$ satisfy*

$$n_h(1)=1\le n_h(q)\le n_h(0)=n_r(0)\le n_r(q)\quad \textit{for all } q.$$

2. *The optimal price-high threshold $n_h(q)$ is nonincreasing in q, so $n_h(q_o)\le n_h(q_p)$.*
3. *The optimal reject threshold $n_r(q)$ is nonincreasing in $q>0$, so $n_r(q_o)\le n_r(q_p)$ for $q_p>0$.*

Observation 1 is consistent with Proposition 1 for the small buffer system. Figure 2 illustrates Observation 1 for $R=100$, $\lambda=\mu=1$, $\beta=0.1$, $c_L=5$, and $c_H=10$. This example is representative of a wide range of cases in which the optimal reject and price-high thresholds depend on q. The reject threshold jumps up at $q=0$, from $n_r(0)=6$ to $n_r(0.1)=17$. It is nonincreasing thereafter.

Figure 2 **Known Demand Scenario**



Notes. Optimal price high threshold $n_h(q)$ and reject threshold $n_r(q)$ as a function of the fraction of patient customers q. ($R = 100$, $\lambda = \mu = 1$, $\beta = 0.1$, $c_L = 5$ and $c_H = 10$.)

5.2.2. Unknown Scenario: Properties of the Optimal Bayesian Pricing Policy.

The structural insights from our numerical study for this case are summarized in the following observation.

Observation 2. *Fix R, c_L, c_H, λ, μ, and β, and scenarios $0 < q_p < q_o < 1$. Let $p^*(n, \alpha)$ denote the price under the optimal Bayesian policy for queue length n and prior α.*

1. *Fix a queue length n. The optimal Bayesian price is independent of α if the optimal policies for known optimistic and pessimistic scenarios charge the same price at queue length n. In this case, $p^*(n, \alpha) = p^*(n; q_p) = p^*(n; q_o)$. Otherwise, there is a unique $\alpha^*(n) \in (0, 1)$, such that*

$$p^*(n,\alpha) = \begin{cases} p^*(n; q_p) & \text{if } \alpha \le \alpha^*(n), \\ p^*(n; q_o) & \text{if } \alpha > \alpha^*(n). \end{cases}$$

2. *The optimal Bayesian pricing policy partitions the set of queue lengths into at most five zones:*

$$p^*(n,\alpha) = \begin{cases} P_L(n) & n < n_h(q_o), \\ P_L(n) \text{ if } \alpha \le \alpha^*(n) \text{ and} & \\ \quad P_H(n) \text{ if } \alpha > \alpha^*(n) & n_h(q_o) \le n < n_h(q_p), \\ P_H(n) & n_h(q_p) \le n < n_r(q_o), \\ P_H(n) \text{ if } \alpha \le \alpha^*(n) \text{ and} & \\ \quad P_R \text{ if } \alpha > \alpha^*(n) & n_r(q_o) \le n < n_r(q_p), \\ P_R & n_r(q_p) \le n. \end{cases} \quad (33)$$

By part 1 of Observation 2, the optimal Bayesian policy for a general buffer system preserves two key properties that hold for the small buffer system (see Proposition 3): (i) The optimal price at a queue length n depends on the prior α if and only if the optimal policies under known pessimistic and optimistic scenarios disagree at that queue length; (ii) for such n, the optimal Bayesian policy has a belief-threshold structure: charge the price that is optimal under known pessimistic scenario if $\alpha \le \alpha^*(n)$ and the one that is optimal under known optimistic scenario if $\alpha > \alpha^*(n)$. These properties of the optimal Bayesian policy are quite intuitive. The lower α, the more confident the provider about the demand scenario being pessimistic, and $q(\alpha) \to q_p$ as $\alpha \to 0$. Conversely, the higher α, the more likely the optimistic scenario, and $q(\alpha) \to q_o$ as $\alpha \to 1$. If the optimal policies for the underlying known scenarios q_p and q_o disagree at a given queue length n, it is, therefore, natural for the provider to price as in the pessimistic scenario if α is sufficiently low, and as in the optimistic scenario if α is sufficiently high.

The zone structure for the optimal Bayesian policy specified in (33) follows from Observation 1 and part 1 of Observation 2. Observation 1 implies that the optimal price-high and reject thresholds for the known pessimistic and optimistic scenarios satisfy the following ranking:

$$n_h(q_o) \le n_h(q_p) \le n_r(q_o) \le n_r(q_p). \quad (34)$$

Combined with part 1 of Observation 2, this ranking implies the optimal Bayesian pricing policy in (33). The resulting queue length partition consists of at most five zones; the pricing rule is the same within a zone and differs across zones. This partition has two key features.

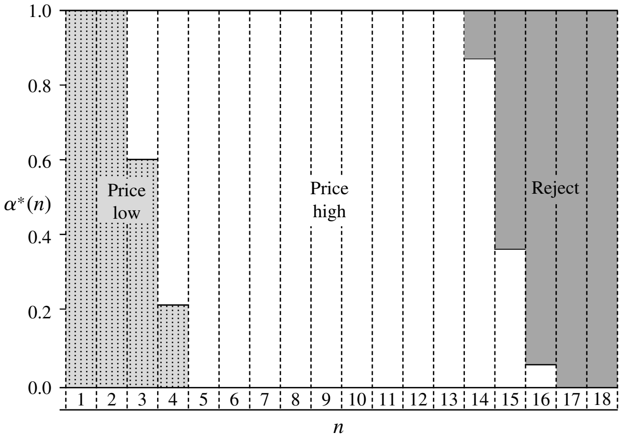
First, there is at least one queue length with optimal *α-contingent high pricing*, if the known pessimistic and optimistic demand scenarios yield different optimal policies. Moreover, at every queue length with optimal α-contingent pricing, the choice is between pricing high and either pricing low or rejecting. For no queue length is it optimal to choose between pricing low and rejecting.

Second, there is at least one queue length with optimal *α-independent high pricing*, if $n_h(q_p) < n_r(q_o)$. This property holds for any pair of scenarios if the optimal reject threshold in a system with only impatient customers is strictly smaller than in one with only patient customers, i.e., $n_r(0) = n_h(0) < n_r(1)$, because $n_h(q)$ and $n_r(q)$ are nonincreasing by Observation 1.

We classify the possible zone structures into the following two main cases.

Case 1: There is a queue length with optimal α-independent high pricing. In this case the provider charges the high price infinitely often. Because she updates her prior when pricing high, this case implies *complete learning;* i.e., the prices under the optimal Bayesian policy converge almost surely to the optimal prices for the underlying demand scenario. For illustration, take $R = 100$, $\lambda = \mu = 1$, $\beta = 0.1$, $c_L = 5$, and $c_H = 10$, as in Figure 2. Consider the optimal Bayesian

Figure 3 Case 1. Optimal Bayesian Pricing Policy with α-Independent High Pricing for Several Queue Lengths



Note. $R = 100$, $\lambda = \mu = 1$, $\beta = 0.1$, $c_L = 5$, $c_H = 10$, $q_p = 0.2$ and $q_o = 0.8$.

pricing policy for pessimistic scenario $q_p = 0.2$ and optimistic scenario $q_o = 0.8$. From Figure 2, the optimal queue length thresholds satisfy $n_h(q_o) = 3 < n_h(q_p) = 5 < n_r(q_o) = 14 < n_r(q_p) = 17$. By (33), the optimal Bayesian policy partitions the queue lengths into three zones with α-independent pricing and two zones with α-contingent high pricing:

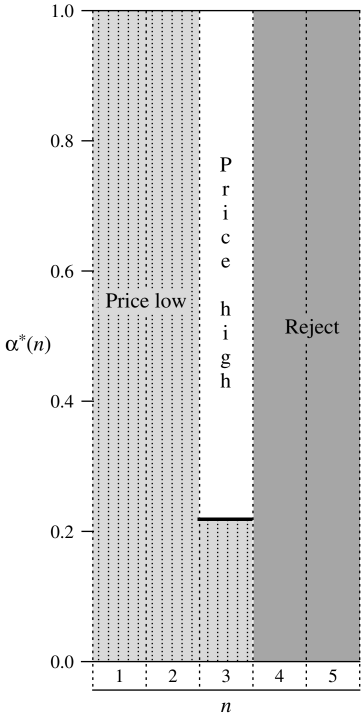
$$p^*(n,\alpha) = \begin{cases} P_L(n) & n < 3, \\ P_L(n) \text{ if } \alpha \le \alpha^*(n) \text{ and} & \\ \quad P_H(n) \text{ if } \alpha > \alpha^*(n) & 3 \le n < 5, \\ P_H(n) & 5 \le n < 14, \\ P_H(n) \text{ if } \alpha \le \alpha^*(n) \text{ and} & \\ \quad P_R \text{ if } \alpha > \alpha^*(n) & 14 \le n < 17, \\ P_R & 17 \le n. \end{cases}$$

The belief thresholds $\alpha^*(n)$ are shown in Figure 3, which illustrates this policy.

Case 2: *α-independent high pricing is not optimal at any queue length*. In this case the provider may stop charging the high price forever after a finite amount of time, which is suboptimal if the scenario is optimistic. This case, therefore, implies *potentially incomplete learning*; i.e., prices may fail to converge to optimal levels with positive probability. For example, set $R = 100$, $\lambda = \mu = 1$, $\beta = 0.1$, $c_L = 14$, and $c_H = 16$. Consider the optimal Bayesian pricing policy for scenarios $q_p = 0.1$ and $q_o = 0.3$. The optimal queue length thresholds satisfy $n_h(q_o) = 3 < n_h(q_p) = n_r(q_o) = n_r(q_p) = 4$. The optimal Bayesian policy has a three-zone structure as specified in (33); see Figure 4. Pricing high is optimal if and only if the queue length $n = 3$ and the prior $\alpha > \alpha^*(3) = 0.21$. If the provider's updated prior drops below $\alpha^*(3) = 0.21$, she stops updating her prior and charges the low price for queue length $n = 3$, which is suboptimal if the scenario is optimistic.

In the small buffer system, only Case 2 arises (see Proposition 3), whereas Case 1 corresponds to the degenerate case where the underlying known scenarios yield identical optimal policies. Case 2 of the general buffer system also has a small system size, i.e., maximum queue length. However, it arises endogenously as a result of the optimal Bayesian pricing policy, unlike in the small buffer system, where the system size is exogenously given. In particular, in Case 2 of the general buffer system, delay costs of both customer types are quite high (and higher compared with Case 1). Consequently, the reject thresholds for both underlying scenarios are very low, e.g., $n_r(q_0) = n_r(q_p) = 4$.

Figure 4 Case 2. Optimal Bayesian Pricing Policy Without α-Independent High Pricing



Note. $R = 100$, $\lambda = \mu = 1$, $\beta = 0.1$, $c_L = 14$, $c_H = 16$, $q_p = 0.1$ and $q_o = 0.3$.

6. Price Convergence Under Optimal Bayesian Policy

In this section, we focus primarily on the small buffer system studied in §4. Denote the provider's prior at time t by $\alpha(t)$ and define $T := \inf\{t \ge 0\colon \alpha(t) \le \alpha^*\}$. It follows from Proposition 3 that the provider prices high until T and prices low thereafter. We are interested in the probability that under the optimal Bayesian policy, the price in the congested state converges to the price that is optimal for the true underlying scenario. When the true scenario is pessimistic, this probability equals $\mathbb{P}(T < \infty)$. When the true scenario is optimistic, it equals $1 - \mathbb{P}(T < \infty)$.

To characterize $\mathbb{P}(T < \infty)$, consider the evolution of Bayesian updates to the prior α. Let $\alpha_k(\alpha)$ be the provider's posterior after the first $k \geq 1$ updates of α and set $\alpha_0(\alpha) = \alpha$. To characterize the evolution of $\alpha_k(\alpha)$, define the random variable X_k as the *net* sale (sale minus the lost sale) associated with the kth customer who arrives to the congested system when the price is high:

$$X_k := \begin{cases} +1 & \text{if the } k\text{th customer joins the system,} \\ -1 & \text{if the } k\text{th customer does not} \\ & \quad \text{join the system.} \end{cases}$$

Define $Z_k := \sum_{i=1}^{k} X_i$ for $k \geq 1$ as the cumulative number of net sales from the first k arrivals to the congested system. The following lemma characterizes the prior after the first k updates.

Lemma 1. *The provider's prior after the first k updates is given by*

$$\alpha_k(\alpha) = \alpha \cdot \left(\alpha + (1-\alpha)\left(\frac{q_p}{q_o}\cdot\frac{1-q_p}{1-q_o}\right)^{k/2} \cdot \left(\frac{q_p}{q_o}\cdot\frac{1-q_o}{1-q_p}\right)^{Z_k/2}\right)^{-1}. \tag{35}$$

Define the stopping time K to be the number of updates before the prior falls below α^*:

$$\begin{aligned} K &:= \inf\{k \geq 0\colon \alpha_k(\alpha) \leq \alpha^*\} \\ &= \inf\{k \geq 0\colon Z_k - d\cdot k \leq -L\}, \end{aligned} \tag{36}$$

where the equality follows from Lemma 1 and the constants d and L are defined as follows:

$$d := \frac{\log\left(\frac{q_p}{q_o}\cdot\frac{1-q_p}{1-q_o}\right)}{\log\left(\frac{q_o}{q_p}\cdot\frac{1-q_p}{1-q_o}\right)} \quad \text{and} \quad L := \frac{\log\left(\frac{\alpha}{\alpha^*}\cdot\frac{1-\alpha^*}{1-\alpha}\right)}{\frac{1}{2}\log\left(\frac{q_o}{q_p}\cdot\frac{1-q_p}{1-q_o}\right)}. \tag{37}$$

The constant $-L < 0$ is the normalized hitting barrier for the random walk of sales and lost sales in the congested system. The constant d can be interpreted as a normalized drift term that captures the impact of zero net sales ($Z_k = 0$) on the prior. In particular, $d < 0$ if $q_p < 1 - q_o$, $d = 0$ if $q_p = 1 - q_o$, and $d > 0$ if $q_p > 1 - q_o$. For convenience, define the random variable

$$Y_k := X_k - d \quad \text{for } k \geq 1,$$

which is the net sale for the kth arrival, normalized by the drift term d. Denote by $\mathbb{E}_p[Y_1]$ and $\mathbb{E}_o[Y_1]$ the conditional expectation of Y_1 under the pessimistic and optimistic scenarios, respectively. Define the conditional cumulant-generating function of Y_1 under the pessimistic scenario

$$\begin{aligned} \psi_p(\theta) &:= \log \mathbb{E}_p[e^{\theta Y_1}] \\ &= \log(q_p e^{\theta} + (1-q_p)e^{-\theta}) - \theta d \quad \text{for } -\infty < \theta < \infty, \end{aligned}$$

and similarly define $\psi_o(\theta) := \log \mathbb{E}_o[e^{\theta Y_1}]$.

Define the random walk $S_k := \sum_{i=1}^{k} Y_i$ for $k \geq 1$, where (36) implies $K = \inf\{k \geq 0\colon S_k \leq -L\}$. It follows from standard properties that $\mathbb{E}_o[Y_1] > 0 > \mathbb{E}_p[Y_1]$. Therefore, the random walk S_k has negative drift if the true scenario is pessimistic and positive drift if the true scenario is optimistic (see also Lemma 2 in the online supplement, available at http://dx.doi.org/10.1287/msom.1120.0418).

If the true scenario is pessimistic, the probability that the provider eventually charges the optimal (low) price in the congested state equals $\mathbb{P}(T < \infty)$. Clearly, $\mathbb{P}(T < \infty) = \mathbb{P}(K < \infty)$. Because S_k is a random walk with negative drift and $K = \inf\{k \geq 0\colon S_k \leq -L\}$, it follows immediately that $\mathbb{P}(K < \infty) = 1$; i.e., the provider eventually charges the optimal low price almost surely.

In contrast, if the true scenario is optimistic, the following proposition shows that there is a positive probability that the provider experiments for a finite period of time by pricing high in the congested state and then charges the suboptimal low price forever, leading to incomplete learning.

Proposition 6. *Under the optimal Bayesian policy, if the true scenario is optimistic, the probability that the price in the congested state converges to the suboptimal low price for this scenario equals*

$$\mathbb{P}(T < \infty) = \frac{1}{\mathbb{E}_o[e^{r_o S_K} \mid S_K \leq -L]}, \tag{38}$$

where $r_o = \min\{r\colon \psi_o(r) = 0\} < 0$. *This yields*

$$e^{r_o(L+1+d)} \leq \mathbb{P}(T < \infty) \leq e^{r_o L} < 1. \tag{39}$$

As discussed in §5.2, in the general buffer system incomplete learning may arise only if the optimal Bayesian policy prescribes no α-independent high pricing (Case 2). In this case the preceding analysis goes through with minor modifications to the stopping times T and K defined with respect to the appropriate (belief) thresholds; and the formula for $\mathbb{P}(T < \infty)$ in Proposition 6 holds with that modification.

7. Concluding Remarks

This paper provides two sets of results on the learning-and-earning problem for a queueing system. First, it characterizes the optimal dynamic pricing policies both in the known scenario benchmark and under Bayesian updating. Second, it characterizes the convergence behavior of the optimal Bayesian prices.

In the known scenario benchmark, the optimal dynamic pricing policy has a nested threshold structure. It appears that this result is new and may be of interest in its own right.

The optimal Bayesian dynamic pricing policy has two key features. First, the optimal price for a given queue state is belief-dependent if and only if the optimal policies for the known pessimistic and optimistic scenarios disagree at that queue state. Second, for such queue states the optimal Bayesian policy has a belief-threshold structure: the provider charges the price that is optimal for the known optimistic scenario, if her belief for that scenario exceeds the threshold, and otherwise the price that is optimal for the known pessimistic scenario. We show these properties analytically for a small buffer system and verify them for a general buffer system through a combination of analytical and numerical results. These results suggest that the key structural properties of the optimal policies for the small buffer system continue to hold for a general buffer system.

The price convergence behavior under the optimal Bayesian policy is sensitive to the system size. In the small buffer system, learning is potentially incomplete: there is a positive probability that the provider ends up charging the suboptimal low price forever if the scenario is optimistic. For the general buffer system, we identify two cases. Case 1, complete learning, in which case the optimal Bayesian prices converge almost surely to the optimal prices in either scenario. Case 2, potentially incomplete learning, in which case the optimal Bayesian prices converge to the optimal prices if the scenario is pessimistic; however, they converge to suboptimal prices with positive probability if the scenario is optimistic. Case 2 corresponds to the price convergence behavior in the small buffer system, which is consistent with the typical incomplete learning outcome in the literature (see Rothschild 1974).

The superior price convergence performance in Case 1 of the general buffer system points to the value of a larger system for demand learning. In particular, queue states with belief-independent high pricing emerge naturally under the optimal Bayesian policy if the optimal pricing policy for each underlying scenario comprises *multiple* queue states with congestion-dependent prices, and the policies for pessimistic and optimistic scenarios agree on the informative high price in at least one queue state. Such a system arises if the patient and impatient types differ sufficiently in their delay cost, or if both types are sufficiently time insensitive. However, if the system is small, either by design (as in the small buffer system) or as a result of the optimal Bayesian policy (Case 2), the optimal policies for the underlying known scenarios consist of far shorter price vectors, and they disagree on whether to charge the informative high price. In such settings, belief-independent high pricing is not optimal at any queue length under the optimal Bayesian policy, and the provider is prone to choosing uninformative and suboptimal prices eventually.

These price convergence results also point to the role of service speed and queueing for demand learning. For given demand attributes, the system size under the optimal Bayesian policy increases in the service speed (i.e., the service rate μ). If service is slow, relative to delay costs, the system is relatively small under the optimal Bayesian policy, the provider has only a few queue states available for demand learning, and prices may converge to suboptimal levels (Case 2). However, if service is sufficiently fast, the optimal Bayesian policy gives rise to a larger system, with belief-independent high pricing in at least one queue state, which ensures complete learning (Case 1).

Our price convergence results are robust to the assumption of a Bernoulli or binary prior, i.e., that $q \in \{q_p, q_o\}$. It is not the prior distribution per se that drives these results. Rather, what matters is whether there exists a queue state with belief-independent informative high pricing. This, in turn, depends primarily on the structure of the underlying value-delay cost structure and the resulting system size. In our model, belief-independent high pricing would still emerge in every general buffer system where the optimal reject threshold with only impatient customers is smaller than with only patient customers (i.e., $n_r(0) < n_r(1)$), and it would not emerge in any nontrivial small buffer system. If there is no belief-independent high pricing, then prices can converge to suboptimal levels for any prior distribution. Ultimately, what matters from a learning perspective is a sufficient statistic of the fraction q of patient customers. For example, the number of times the provider prices high and the number of resulting sales provide a sufficient statistic. The key insight is that if the provider sees a sequence of discouraging outcomes (i.e., no sales when pricing high), then she may stop experimenting with the informative high price and stop learning as a result. This insight is independent of whether the prior is binary or has a continuous support. Consistent with this intuition, Rothschild (1974) considers a continuous prior on the unknown parameters and illustrates that the incomplete learning phenomenon can arise in that setting. Although he ignores capacity constraints, our small buffer system resembles his setting, in which the optimal pricing policy for each known parameter value consists of a single price. In a general buffer system with a general prior distribution, if the optimal Bayesian policy prescribes belief-independent high pricing for at least one queue state, then the prices will converge almost surely to the optimal levels corresponding to the true underlying scenario, just like in our model with a binary prior.

We also assume that the delay cost distribution is binary. We expect that relaxing this assumption can promote convergence to optimal prices, particularly in small buffer systems. To see this, suppose that the delay costs are drawn from an arbitrary distribution. Allowing for multiple delay cost levels potentially yields a larger set of optimal prices for each queue state, and multiple prices in this set are informative, namely, all prices except the lowest price at which all customers are willing to buy and the highest one where none are willing to buy.

Another key assumption we made is that both types have the same constant reward/valuation. If the types also differ in their valuations, say V_H and V_L, we expect the results to depend on the relative magnitude of V_H and V_L. If $V_H < V_L$, then the main effect would be that even in the small buffer system there could be learning in two queue states, $n = 0$ and $n = 1$, which would promote price convergence. If $V_H > V_L$ and the net value functions do not cross, i.e., $V_H - n(c_H/\mu) > V_L - n(c_L/\mu)$ for all $n \le V_L(\mu/c_L)$, then the main price convergence insights would likely remain the same, with the necessary modifications to reflect the reversed ranking of impatient versus patient types: only impatient customers would buy at the high price. In the small buffer system we expect it to be optimal in the congested state to price high as long as the posterior belief is *below* a threshold (enough impatient customers) and to price low otherwise. This will result in a positive probability of incomplete learning if the scenario is *pessimistic*: this occurs if the posterior belief crosses the belief threshold from below, whereupon the provider charges the suboptimal low price forever. In the case where $V_H > V_L$ and the net value functions do cross, i.e., $V_H - n(c_H/\mu) = V_L - n(c_L/\mu)$ for some n, the result may differ significantly from ours. In terms of price convergence, the performance advantage of the large buffer system from charging multiple congestion-sensitive prices, could be reduced because pricing may be uninformative in some intermediate queue state(s) where both types have the same net willingness-to-pay due to the positive correlation between valuations and delay costs. As a result, we expect to see in the large buffer system fewer queue states with belief-independent high pricing under the optimal Bayesian policy. Generally speaking, we expect multiple valuation levels to have two effects for price convergence and learning: First, they yield more price levels that are informative when queues are short (valuation differences dominate) or long (delay cost differences dominate), but they may result in less informative pricing at medium-length queues if valuations and delay costs are positively correlated.

A potential future research direction is to use frequentist methods for learning such as least squares, maximum likelihood, or even nonparametric estimation methods. Another potentially fruitful direction is to somewhat lower aspirations and focus on asymptotic (versus exact) optimality, which may facilitate obtaining further analytical results in the general buffer case. Finally, one can consider simple, easy-to-implement policies and evaluate their performance.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/msom.1120.0418.

Acknowledgments

The authors are grateful to Fikri Karaesmen, N. Bora Keskin, Refael Hassin, Ignatius J. Horstmann, the associate editor, and the referees for constructive comments that helped improve the paper significantly. Research assistance with the numerical study by Seyed M. Emadi is gratefully acknowledged.

References

Aghion P, Bolton P, Harris C, Julien B (1991) Optimal learning by experimentation. *Rev. Econom. Stud.* 58(4):621–654.

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

Ata B (2006) Dynamic control of a multiclass queue with thin arrival streams. *Oper. Res.* 54(5):876–892.

Aviv Y, Pazgal A (2005) A partially observed Markov decision process for dynamic pricing. *Management Sci.* 51(9):1400–1416.

Azoury KS (1985) Bayes solution to dynamic inventory models under unknown demand distribution. *Management Sci.* 31(9):1150–1160.

Banks JS, Sundaram RK (1992) Denumarable-armed bandits. *Econometrica* 60(5):1071–1096.

Bertsekas D (1995) *Dynamic Programming and Optimal Control*, Vol. 2 (Athena Scientific, Nashua, NH).

Besbes O, Maglaras C (2009) Revenue optimization of a make-to-order queue in an uncertain market environment. *Oper. Res.* 57(6):1438–1450.

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

Brezzi M, Lai TL (2002) Optimal learning and experimentation in bandit problems. *J. Econom. Dynam. Control* 27(1):87–108.

Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Oper. Res.* 60(4):965–980.

Chen H, Frank M (2001) State dependent pricing with a queue. *IIE Trans.* 33(10):847–860.

Chen L, Plambeck EL (2008) Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing Service Oper. Management* 10(2):236–256.

den Boer AV, Zwart B (2011) Simultaneously learning and optimizing using controlled variance pricing. Working paper, Centrum Wiskunde and Informatica, Amsterdam.

Easley D, Kiefer NM (1988) Controlling a stochastic process with unknown parameters. *Econometrica* 56(5):1045–1064.

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

Harpaz G, Lee W, Winkler R (1982) Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Sci.* 28(6):589–603.

Harrison JM, Keskin NB, Zeevi A (2012a) Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Sci.* 58(3):570–586.

Harrison JM, Keskin NB, Zeevi A (2012b) Dynamic pricing with an unknown linear demand model: Asymptotically optimal semi-myopic policies. Working paper, Stanford University, Stanford, CA.

Hassin R, Haviv M (2003) *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems* (Kluwer, Boston).

Haviv M, Randhawa R (2012) Pricing queues without demand information. Working paper, University of Southern California, Los Angeles.

Karlin S, Taylor H (1975) *A First Course in Stochastic Processes*, 2nd ed. (Academic Press).

Keller G, Rady S (1999) Optimal experimentation in a changing environment. *Rev. Econom. Stud.* 66(3):475–507.

Lobo MS, Boyd S (2003) Pricing and learning with uncertain demand. Working paper, Stanford University, Stanford, CA.

Lovejoy W (1990) Myopic policies for some inventory models with uncertain demand distribution. *Management Sci.* 36(6): 724–738.

Masuda Y, Whang S (1999) Dynamic pricing for network service: Equilibrium and stability. *Management Sci.* 45(6):857–869.

Mersereau AJ, Rusmevichientong P, Tsitsiklis JN (2009) A structured multiarmed bandit problem and the greedy policy. *IEEE Trans. Automatic Control* 54(12):2787–2802.

Naor P (1969) On the regulation of queue size by levying tolls. *Econometrica* 37(1):15–24.

Petruzzi NC, Dada M (2002) Dynamic pricing and inventory control with learning. *Naval Res. Logist.* 49(3):303–325.

Ritt RK, Sennott LI (1992) Optimal stationary policies in general state space Markov decision chains with finite action sets. *Math. Oper. Res.* 17(4):901–909.

Rothschild M (1974) A two-armed bandit theory of market pricing. *J. Econom. Theory* 9(2):185–202.

Scarf H (1959) Bayes solution of the statistical inventory problem. *Ann. Math. Statist.* 17(4):901–909.

Tsitsiklis JN (1994) A short proof of the Gittins index theorem. *Ann. App. Probab.* 4(1):194–199.