

CONGESTION-DEPENDENT PRICING IN A STOCHASTIC SERVICE SYSTEM

IDRISS MAOUI,*
HAYRIYE AYHAN^{†** ****} AND
ROBERT D. FOLEY,^{** ****} *Georgia Institute of Technology*

Abstract

We study a service facility modeled as a queueing system with finite or infinite capacity. Arriving customers enter if there is room in the facility and if they are willing to pay the price posted by the service provider. Customers belong to one of a finite number of classes that have different willingnesses-to-pay. Moreover, there is a penalty for congestion in the facility in the form of state-dependent holding costs. The service provider may advertise class-specific prices that may fluctuate over time. We show the existence of a unique optimal stationary pricing policy in a continuous and unbounded action space that maximizes the long-run average profit per unit time. We determine an expression for this policy under certain conditions. We also analyze the structure and the properties of this policy.

Keywords: Dynamic pricing; queueing; Markov decision process; revenue management

2000 Mathematics Subject Classification: Primary 90B22
Secondary 90C40

1. Introduction

A decision maker wishes to quote prices at the most profitable level. When a customer arrives, the customer decides whether to pay the quoted price and enter the service system or to depart without obtaining service. Rather than being restricted to a single price that is offered to all customers, the decision maker has a great deal of flexibility in setting prices. The decision maker is allowed to use two pieces of information when making a quotation. The decision maker knows the number of customers currently in the service system, which is a measure of the congestion in the system. Allowing the price to depend upon the level of congestion will be called *congestion-dependent pricing*. In addition, the decision maker is able to classify the customers into different types, and the decision maker knows the probability of a customer of a particular type accepting a particular price. Allowing the price to depend upon information about a customer will be referred to as *precision pricing*. Thus, the decision maker can use a congestion dependent, precision pricing strategy. If all customers are classified as the same type then the decision maker uses only congestion-dependent pricing. If the decision maker is not allowed to use information about the current level of congestion when setting prices then the decision maker will be using *static pricing*.

Received 23 May 2006; revision received 12 November 2007.

* Current address: Lehman Brothers, New York, NY 10019, USA. Email address: imaoui@lehman.com

** Postal address: H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA.

*** Email address: hayhan@isye.gatech.edu

**** Email address: rfoley@isye.gatech.edu

We assume that the probability of a particular type of customer accepting a price does not increase as the price increases, and we refer to this conditional distribution function as the *willingness-to-pay distribution*. Already the decision maker faces a trade-off. If prices are high, each customer pays a lot, but few customers pay; if prices are low, each customer pays a little, but many pay.

If there is no limit on the number of customers that can be in the service system simultaneously and the decision maker has no reason to keep the number of customers in the system at a low level, then the most profitable prices could easily result in a large average number of customers in the system. In most applications this would be unacceptable. To give the decision maker an incentive to reduce congestion, we assume that the decision maker incurs a cost at rate h_s when there are s customers in the service system. These costs will be nonnegative and nondecreasing in s .

The service system will be modeled as a queueing system. The maximum number of customers that can simultaneously be in the service system is $N \leq \infty$. The particular assumptions on the queueing system will be given in Section 3, but they will be designed so that the number of customers in the service system is Markovian. By ‘most profitable’ we mean maximizing the long-run average profit per unit time.

Under certain assumptions, we will be able to show that there is a unique, optimal pricing strategy, and we characterize the optimal prices. In addition, we will be able to determine structural and ordering properties of the optimal prices. For example, under some mild conditions, we show that the optimal price is nondecreasing in the congestion level. Under other conditions, we show that the optimal static price is a compromise lying between the highest and lowest congestion-dependent prices.

Several technical aspects of the paper may be of interest. First, we have an unbounded, continuous action space. When $N < \infty$, we are able to use the results of [8, Section 5.2] to analyze our problem. Our problem is more complicated when $N = \infty$ since the state space is uncountable. Under certain conditions, we can extend our results for $N < \infty$ to $N = \infty$ through the use of a fixed point argument.

In the following section we briefly review the literature related to our paper. In Section 3 we describe our model. Then we decompose our analysis into two parts. Section 4 focuses on finite capacity systems, and Section 5 focuses on infinite capacity systems. In Section 6 we conclude our work.

2. Literature review

Although our work directly considers the issue of dynamic pricing in queueing systems, it is inspired by a series of papers on the more general topic of congestion control in queueing systems. We can group them into two categories, depending on whether the control is static or dynamic.

The paper by Naor [13] is considered as the precursor in combining the issues of pricing and congestion control in queues. Naor’s work and many papers extending it (such as [6] and [18]) analyze systems where customers make a decision to enter a service facility based on its current queue length. Entering customers obtain a fixed reward and are charged a holding cost function of their time spent in the system. In order to maximize their utility, they decide to join or balk (join-balk rule). The service provider then imposes an entrance fee to induce an optimal customer admission rate. Larsen [7] and Hassin [5] evaluated the effect of releasing the expected queue length to potential customers as opposed to the current queue length. Mendelson and Whang [12] considered customers who make their decision to enter the system based both on

price and delay. Mendelson and Whang [12] also included different customer classes that have different demand functions and delay costs. Prices were then used by the decision maker as an incentive to induce optimal customer arrival rates and execution priorities. In all the papers mentioned above, the system controls are static; that is, the controls are independent of the congestion level.

In the second set of papers, controls are allowed to depend on the congestion level (dynamic control). Stidham [16] developed a dynamic admission control model to optimize an infinite-horizon discounted reward with convex holding costs for single server queues. Stidham's decision variable is defined by whether to accept or reject an incoming job. Each accepted job yields a fixed deterministic reward. He showed the existence of a monotonic optimal stationary policy. He also extended his results to simple networks of queues. Conversely, George and Harrison [4] allowed the service provider to dynamically control the service rate instead of the arrival rate. There was a penalty that depended on the chosen service rate and the objective was to minimize the long-run average cost in systems with holding costs.

Combining the problems of setting admission rates and service rates, Ata and Shneerson [2] considered a dynamic control model, where the service provider sets state-dependent admission rates and service rates in an M/M/1 queue with holding costs. There was a reward associated with the chosen admission rates and a cost corresponding to the chosen service rates. After explicitly solving this problem, they analyzed a decentralized model, where only service rates and prices were decision variables. The service provider must set them so that the optimal admission rates are induced by customers maximizing their own utility.

Low [9], [10] considered dynamic pricing in M/M/s queues with finite or infinite capacity, but with a finite action space. Low did not use a willingness-to-pay distribution, but each price in the action space corresponded to a given positive arrival rate. Low also considered state-dependent holding costs incurred as a lump sum as a customer arrives. He made the extra assumption that holding costs are bounded and that the facility has multiple identical servers. He showed that optimal prices are nondecreasing as the system becomes congested and developed an algorithm to solve the Markov decision process formulation of the problem. Aktaran and Ayhan [1], as well as Çil *et al.* [3], further investigated the sensitivity of the optimal prices to system parameters. Paschalidis and Tsitsiklis [14] focused on models with multiple classes of customers that have different resource requirements without holding costs.

In this paper we extend Low's work by introducing general state-dependent holding costs and service rates, and by considering a continuous unbounded action space and multiple customer classes. Note that we analyze the same model in a related paper [11], where the service provider does not have the flexibility to adjust prices in time and must implement a static pricing policy.

3. Model description

We model the service facility as a queueing system of capacity $N \leq \infty$; that is, no more than N customers are allowed in the system at any time. There are I classes of customers and customers from class $i = 1, \dots, I$ arrive according to a Poisson process with parameter $\Lambda_i > 0$. The arrival processes from customer classes are independent of each other. Note that this formulation is equivalent to having arriving customers randomly assigned to a specific class, independent of everything else.

The maximum amounts that successive class $i = 1, \dots, I$ customers are willing to pay are independent, identically distributed random variables with distribution F_i . The amount a class $i = 1, \dots, I$ customer is willing to pay is independent of the amount a class $j = 1, \dots, I$ customer is willing to pay for $i \neq j$. For all $i = 1, \dots, I$, we assume that the cumulative

distribution function $F_i(\cdot)$ is absolutely continuous with density $f_i(\cdot)$, support (α_i, β_i) , and finite mean. Let $r_i(\cdot)$ denote the hazard rate function of $F_i(\cdot)$; that is,

$$r_i(z) = \frac{f_i(z)}{1 - F_i(z)} \quad \text{for } \alpha_i < z < \beta_i.$$

In the following we assume that F_i has IGHR (increasing generalized hazard rate); that is, $zr_i(z)$ is strictly increasing for all z in $[\alpha_i, \beta_i]$. Note that we can interpret $-zr_i(z)$ as the price elasticity of the demand function for class- i customers. The service provider can advertise different prices to different classes. Without loss of generality, only prices in $[\alpha_i, \beta_i]$ can be advertised to class- i customers.

We define the state of the system, $X(t)$, as the number of customers in the system at time t . Let $\mathbf{z} \in [\alpha_1, \beta_1]^N \times \dots \times [\alpha_I, \beta_I]^N$ be a pricing (decision) rule, where price $z_{i,s}$ is advertised to class- i customers when the system is in state s . Since there is a one-to-one relationship between decision rules and stationary policies, in an abuse of notation, we also denote by \mathbf{z} the stationary pricing policy corresponding to the pricing rule \mathbf{z} ; that is, \mathbf{z} also denotes the policy of using pricing rule \mathbf{z} at every decision epoch (see [15, p. 20] for further details). Customers enter the system if it is not full and if they are willing to pay the price posted by the service provider upon arrival. Hence, the customer admission process under the stationary pricing policy \mathbf{z} is a conditional (doubly stochastic) Poisson process with rate $\mathbf{1}_{X(t) < N} \sum_{i=1}^I \lambda_i(z_{i,X(t)})$, where $\lambda_i(z) = \Lambda_i(1 - F_i(z))$. In the same fashion, the service process is a conditional Poisson process with rate $\mu_{X(t)} \mathbf{1}_{X(t) > 0}$, such that $\{\mu_s\}$ are positive real numbers that are nondecreasing in s . Under the stationary policy \mathbf{z} , the queueing system behaves as a Markovian birth–death process with birth rates $\sum_{i=1}^I \lambda_i(z_{i,s})$ and death rates μ_s .

Furthermore, the service provider must pay a holding cost h_s per unit time spent in state s , where $0 = h_0 \leq h_1 \leq \dots \leq h_N$ as it becomes more expensive to accommodate a larger number of customers. We assume that $h_1/\mu_1 < \max \beta_i$, so that we have an attainable positive reward.

We let g_N^* denote the optimal dynamic average profit per unit time under capacity N over the set of all history-dependent randomized policies; see [15, pp. 35–36]. Under the stationary pricing policy \mathbf{z} , we denote the objective function by $R(\mathbf{z})$ and the stationary probability distribution by $\{\pi_s(\mathbf{z})\}$. We show that there exists a unique optimal stationary pricing policy \mathbf{z}^* that maximizes the long-run average profit per unit time; that is, $R(\mathbf{z}) = g_N^*$ if and only if $\mathbf{z} = \mathbf{z}^*$.

We separate systems with finite capacity from systems with infinite capacity in our work. Indeed, we make extra assumptions, and we use results from finite capacity systems in order to analyze systems of infinite capacity. In the following section we focus on service facilities with finite capacity.

4. Finite capacity queues

4.1. Characterization of an optimal stationary policy

In this section we restrict ourselves to systems with finite capacity. We use a Markov decision process (MDP) approach to exhibit an optimal stationary policy. Note that the MDP associated with our system behaves as a birth–death process. Since the death rates are strictly positive, the MDP is unichain for any stationary policy. We set up the system of average-cost optimality

equations (ACOE) as detailed in Theorem 5.2.2 of [8]:

$$\begin{aligned}
 l(-1) &= 0, \\
 l(s) &= \sup_{z_0, \dots, z_I} \left\{ \frac{\sum_{i=1}^I \lambda_i(z_i)(z_i + l(s + 1)) + \mu_s l(s - 1) - g - h_s}{\sum_{i=1}^I \lambda_i(z_i) + \mu_s} \right\}, \quad 0 \leq s \leq N - 1, \\
 l(N) &= l(N - 1) - \frac{g + h_N}{\mu_N},
 \end{aligned}$$

where g is the gain and $l(\cdot)$ is the bias vector. Since the value of μ_0 does not matter as long as it is positive, we will consider $\mu_0 = \mu_1$ without loss of generality. In this system we are solving for g and $l(\cdot)$. We can transform these equations into a simpler equivalent form by letting $G(-1) = 0$ and $G(s) = l(s) - l(s + 1)$ for $s = 0, \dots, N - 1$. Then

$$G(-1) = 0, \tag{4.1}$$

$$g + h_s - \mu_s G(s - 1) = \sum_{i=1}^I \sup_z \{(z - G(s))\lambda_i(z)\} \quad \text{if } s = 0, \dots, N - 1, \tag{4.2}$$

$$G(N - 1) = \frac{g + h_N}{\mu_N}. \tag{4.3}$$

If a solution $(g, G(\cdot), z)$ to the system of ACOE exists, we call it a *canonical triplet*, where z are prices that achieve the suprema in (4.2). Precisely, for $s = 0, \dots, N - 1$ and $i = 1, \dots, I$, the component $z_{i,s}$ of z satisfies $z_{i,s} = \arg \sup\{(z - G(s))\lambda_i(z)\}$.

In the following theorem we explicitly characterize a unique optimal stationary policy.

Theorem 4.1. *There exists a canonical triplet $(g, G(\cdot), z)$ for the system of ACOE (4.1)–(4.3). Moreover, the optimal long-run average reward is $g_N^* = g$, and $z^* = z$ is a unique optimal stationary policy, where, for $s = 0, \dots, N - 1$ and $i = 1, \dots, I$,*

$$z_{i,s}^* = \inf\{z : r_i(z)(z - G(s)) \geq 1\}.$$

Before proving this theorem, we need the following two lemmas. Let $G(s, g)$ be the solution of (4.2) and (4.3) for an arbitrary $g \geq 0$.

Lemma 4.1. *For all $s = -1, \dots, N - 1$, $G(s, \cdot)$ is nondecreasing and continuous. Moreover, there exists $g \geq 0$ such that $G(-1, g) = 0$.*

Proof. Note that $G(N - 1, g) = (g + h_N)/\mu_N$ is continuous and nondecreasing in g . Suppose that $G(s, g)$ is nondecreasing and continuous in g for some state s between 0 and $N - 1$. As $\sup\{\lambda_i(z)(z - G(s, g))\}$ is the supremum of a bounded continuous function of z , we can claim that

$$\mu_s G(s - 1, g) = g - \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s, g))\}$$

is continuous and nondecreasing in g . By induction, for all $s = -1, \dots, N - 1$, $G(s, \cdot)$ is nondecreasing and continuous.

To complete the proof, we will show that $G(-1, 0) \leq 0$ and that there exists $g_b > 0$ such that $G(-1, g_b) \geq 0$. Hence, by continuity, there exists $g \in [0, g_b]$ such that $G(-1, g) = 0$.

We know that $-\mu_0 G(-1, 0) = \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(0, 0))\} \geq 0$. Therefore, $G(-1, 0) \leq 0$. Now consider $g_b = \sum_{i=1}^I \sup\{\lambda_i(z)z\}$. Note that $G(N-1, g_b) = (g_b + h_N)/\mu_N \geq 0$. Suppose that $G(s, g_b) \geq 0$ for some $s = 0, \dots, N-1$, then

$$g_b + h_s - \mu_s G(s-1, g_b) = \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s, g_b))\} \leq \sum_{i=1}^I \sup\{\lambda_i(z)z\}.$$

Therefore, $\mu_s G(s-1, g_b) \geq g_b + h_s - \sum_{i=1}^I \sup\{\lambda_i(z)z\} \geq 0$. By induction, $G(-1, g_b) \geq 0$.

Lemma 4.2. *Let $(g, G(\cdot), \mathbf{z})$ be a canonical triplet. Then, for all $s = -1, \dots, N-1$, $0 \leq G(s) \leq (g + h_{s+1})/\mu_{s+1}$.*

Proof. For all $s = 0, \dots, N-1$, $\sup\{(z - G(s))\lambda_i(z)\} \geq 0$. Therefore, we have $G(s-1) \leq (g + h_s)/\mu_s$ from (4.2). Using (4.1) and (4.3) as well, $G(s) \leq (g + h_{s+1})/\mu_{s+1}$ for all $s = -1, \dots, N-1$.

Now suppose that there exists $s = 0, \dots, N-1$ such that $G(s) < 0$. Since $G(-1) \geq 0$, there exists s such that $G(s) < 0$ and $G(s-1) \geq 0$. Hence, $\mu_{s+1}G(s) < \mu_s G(s-1)$. But, we have

$$\begin{aligned} \sum_{i=1}^I \lambda_i(z_{i,s+1})G(s+1) &= \sum_{i=1}^I \lambda_i(z_{i,s+1})z_{i,s+1} - g - h_{s+1} + \mu_{s+1}G(s) \\ &= \sum_{i=1}^I \lambda_i(z_{i,s+1})(z_{i,s+1} - G(s)) + \lambda_i(z_{i,s+1})G(s) \\ &\quad + \mu_{s+1}G(s) - g - h_{s+1} \\ &< \sum_{i=1}^I \lambda_i(z_{i,s+1})(z_{i,s+1} - G(s)) + \lambda_i(z_{i,s+1})G(s) \\ &\quad + \mu_s G(s-1) - g - h_s \\ &< \sum_{i=1}^I \lambda_i(z_{i,s+1})G(s). \end{aligned}$$

If $\sum_{i=1}^I \lambda_i(z_{i,s+1}) = 0$ then $G(s) = (g + h_{s+1})/\mu_{s+1} \geq 0$, which is impossible. Therefore, $G(s+1) < G(s) < 0$. Since $\mu_{s+2} \geq \mu_{s+1}$, we have $\mu_{s+2}G(s+1) < \mu_{s+1}G(s) < 0$. Consequently, we can repeat the argument above until we reach state $N-1$, for which $G(N-1) < 0$. But $G(N-1) = (g + h_N)/\mu_N \geq 0$, which yields a contradiction. Therefore, for all $s = -1, \dots, N-1$, $0 \leq G(s) \leq (g + h_{s+1})/\mu_{s+1}$.

Proof of Theorem 4.1. The existence of a canonical triplet $(g, G(\cdot), \mathbf{z})$ to (4.1)–(4.3) is a direct consequence of Lemma 4.1. Since the state space is finite, we can refer to Equation (5.2.12) of [8] to prove that the canonical triplet $(g, G(\cdot), \mathbf{z})$ is an optimal solution. Therefore, $g_N^* = g$ and $\mathbf{z}^* = \mathbf{z}$.

It remains to show that $z_{i,s}^* = \inf\{z : r_i(z)(z - G(s)) \geq 1\}$ and that it is the unique optimal stationary policy. For $s = 0, \dots, N-1$ and $i = 1, \dots, I$, let

$$\begin{aligned} v_{i,s}(z) &= \lambda_i(z)(z - G(s)), \\ v'_{i,s}(z) &= (1 - F_i(z)) - f_i(z)(z - G(s)) \quad \text{almost everywhere (a.e.) on } [\alpha_i, \beta_i]. \end{aligned}$$

Also note that $v'_{i,s}(z) > 0$ is equivalent to $r_i(z)(z - G(s)) < 1$ and $v'_{i,s}(z) < 0$ is equivalent to $r_i(z)(z - G(s)) > 1$. The IGHR assumption implies that $r_i(z)(z - G(s)) \geq 1$ a.e. on $(\inf\{z: r_i(z)(z - G(s)) \geq 1\}, \beta_i)$. Therefore, $v'_{i,s}(\cdot) > 0$ a.e. on $(\alpha_i, \inf\{z: r_i(z)(z - G(s)) \geq 1\})$ and $v'_{i,s}(\cdot) < 0$ a.e. on $(\inf\{z: r_i(z)(z - G(s)) \geq 1\}, \beta_i)$. Thus, $v_{i,s}(\cdot)$ is strictly unimodal and $z^*_{i,s} = \inf\{z: r_i(z)(z - G(s)) \geq 1\}$ is its unique maximizer on $[\alpha_i, \beta_i]$.

We still need to show that z^* is the unique optimal stationary policy. Under IGHR, $z^*_{i,s}$ is the unique maximizer of $\sup\{\lambda_i(z)(z - G(s))\}$. So,

$$g_N^* > \sum_{i=1}^I \lambda_i(z_i)(z_i - G(s)) + \mu_s G(s - 1) - h_s \quad \text{for all } z_i \neq z^*_{i,s}.$$

Since we have a unichain model, we can refer to Proposition 8.5.10 of [15] to prove the uniqueness of the optimal stationary policy, z^* .

We are now able to characterize an optimal stationary policy explicitly. Note that it might be possible that $z^*_{i,s} = \beta_i$ for some state s . In this case it is optimal for the service provider not to accept customers of class i when in state s . However, this can only occur if the class- i customers' willingness-to-pay distribution has finite support. Indeed, if F_i has infinite support then, for all $s = 0, \dots, N - 1$, $\sup\{\lambda_i(z)(z - G(s))\} > 0$ and $z^*_{i,s} < \infty = \beta_i$. Moreover, note that

$$z^*_{i,s} = \inf\{z: (z - G(s))r_i(z) \geq 1\} \geq \inf\{z: zr_i(z) \geq 1\}.$$

Since $\inf\{z: zr_i(z) \geq 1\}$ is the optimal price to charge when the demand function is $1 - F_i(z)$, we observe that holding costs and capacity limitations force the service provider to charge higher prices than she normally would if she had no constraints. Therefore, one can understand $G(s)$ as a price premium charged by the service provider to account for the additional congestion created by a customer's admission into state s .

4.2. Structure of optimal policies

We will now exhibit structural properties of the derived optimal stationary policy. More specifically, we are interested in the monotonicity of the optimal stationary policy. In Proposition 4.1, below, we demonstrate that the optimal prices to be charged are nondecreasing in the state index.

Proposition 4.1. *Suppose that $\{\mu_s\}_{s=0}^N$ and $\{h_s\}_{s=0}^N$ are such that there exists an integer q between 0 and N , where $\mu_0 \leq \mu_1 \leq \dots \leq \mu_q = \mu_{q+1} = \dots = \mu_N$ and $0 = h_0 = h_1 = \dots = h_q \leq h_{q+1} \leq \dots \leq h_N$. Then $z^*_{i,s}$ is nondecreasing in s .*

To prove this result, we need the following lemma.

Lemma 4.3. *Suppose that $\{\mu_s\}_{s=0}^N$ and $\{h_s\}_{s=0}^N$ are such that there exists an integer q between 0 and N , where $\mu_0 \leq \mu_1 \leq \dots \leq \mu_q = \mu_{q+1} = \dots = \mu_N$ and $0 = h_0 = h_1 = \dots = h_q \leq h_{q+1} \leq \dots \leq h_N$. Then $G(\cdot)$ is nondecreasing.*

Proof. We decompose our proof into two parts. We will first prove, by induction, that $G(s)$ is nondecreasing for states $s = 0, \dots, q - 1$. Then we will show the same for states $s = q - 1, \dots, N - 1$.

First, note that $G(-1) \leq G(0)$. Now suppose that $G(s - 1) \leq G(s)$ for some state $s \in \{0, \dots, q - 2\}$. Then $\mu_s G(s - 1) \leq \mu_{s+1} G(s)$ and

$$\begin{aligned} \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s + 1))\} &= g_N^* - \mu_{s+1} G(s) \\ &\leq g_N^* - \mu_s G(s - 1) \\ &\leq \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s))\}. \end{aligned}$$

Hence, $\sum_{i=1}^I \lambda_i(z_{i,s}^*) G(s + 1) \geq \sum_{i=1}^I \lambda_i(z_{i,s}^*) G(s)$. Therefore, we either have

$$\sum_{i=1}^I \lambda_i(z_{i,s}^*) = 0 \quad \text{or} \quad G(s) \leq G(s + 1).$$

We show that $\sum_{i=1}^I \lambda_i(z_{i,s}^*) = 0$ is impossible.

Assume that $\sum_{i=1}^I \lambda_i(z_{i,s}^*) = 0$. It implies that

$$\begin{aligned} G(s - 1) = G(s) &= \frac{g_N^*}{\mu_s} = \frac{g_N^*}{\mu_{s+1}} \quad \text{and} \\ \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s - 1))\} &= \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s))\} = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} G(s - 2) = \frac{g_N^*}{\mu_{s-1}} &\geq \frac{g_N^*}{\mu_s} = G(s - 1) \quad \text{and} \\ \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s - 2))\} &\leq \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s - 1))\} = 0. \end{aligned}$$

We can repeat this argument until we reach the contradiction $G(-1) \geq g_N^*/\mu_1$. Therefore, $G(s) \leq G(s + 1)$. By induction, $G(s - 1) \leq G(s)$ holds for all $s = 0, \dots, q - 1$.

If $q = N$, the proof is complete. Otherwise, it remains to show, by induction, that $G(s - 1) \leq G(s)$ for states $s = q, \dots, N - 1$. Recall that $G(N - 1) = (g_N^* + h_N)/\mu_N$ and $G(N - 2) \leq (g_N^* + h_{N-1})/\mu_{N-1}$. Therefore,

$$G(N - 2) \leq \frac{g_N^* + h_N}{\mu_N} = G(N - 1).$$

Now suppose that $G(s) \geq G(s - 1)$ for some state $s \in \{q + 1, \dots, N - 1\}$. Then

$$\begin{aligned} g_N^* + h_s - \mu_N G(s - 1) &= \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s))\} \\ &\leq \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s - 1))\} \\ &\leq g_N^* + h_{s-1} - \mu_N G(s - 2). \end{aligned}$$

Hence, $\mu_N(G(s - 1) - G(s - 2)) \geq h_s - h_{s-1} \geq 0$. Therefore, $G(s - 1) \geq G(s - 2)$. By induction, $G(s) \geq G(s - 1)$ for $s = q, \dots, N - 1$ and the proof is complete.

Proof of Proposition 4.1. Recall that $z_{i,s}^* = \inf\{z : r_i(z)(z - G(s)) \geq 1\}$. Since $G(s)$ is nondecreasing in s , so is $z_{i,s}^*$.

Therefore, in queues with the holding cost and service rate structure described above (such as multiple-server systems), the service provider charges more as the system becomes congested. As a consequence, the admission rates are nonincreasing with respect to the number of people in the system. Hence, the optimal policy performs a congestion control that prevents high holding costs. Moreover, in the proof of Lemma 4.3 we showed that $\sum_{i=1}^I \lambda_i(z_{i,s}^*) > 0$ for all $s < q - 1$. This property is quite intuitive since states 0 through q do not incur any holding cost, so that it is not profitable for the service provider to refuse entrance to customers in those states.

We are now interested in how the system reacts to an increase in capacity. A larger buffer size affects the holding costs as well as the revenue by welcoming more customers. As capacity increases, we show that the optimal prices decrease state by state whereas the optimal reward increases. In the following, subscripts 1 and 2 identify parameters for systems 1 and 2, respectively.

Proposition 4.2. *Consider two systems, 1 and 2, where system 1 has capacity N and system 2 has capacity $N + 1$. Then $g_{N+1}^* \geq g_N^*$. If, for all $s = 0, \dots, N - 1$, there exists $i = 1, \dots, I$ such that $z_{i,s,1}^* < \beta_i$, then $z_{i,s,2}^* \leq z_{i,s,1}^*$.*

Proof. It is straightforward to show that $g_{N+1}^* \geq g_N^*$, since the action space for system 2 includes the action space for system 1.

Suppose that $G_2(s) > G_1(s)$ for some state $s = 0, \dots, N - 1$. Therefore,

$$\sum_{i=1}^I \sup\{\lambda_i(z)(z - G_2(s))\} < \sum_{i=1}^I \sup\{\lambda_i(z)(z - G_1(s))\},$$

or

$$\sum_{i=1}^I \sup\{\lambda_i(z)(z - G_2(s))\} = \sum_{i=1}^I \sup\{\lambda_i(z)(z - G_1(s))\} = 0.$$

The latter case is impossible since it implies that $z_{i,s,1}^* = z_{i,s,2}^* = \beta_i$ for all i . So,

$$\begin{aligned} \mu_s G_2(s - 1) &= g_{N+1}^* - \sum_{i=1}^I \sup\{\lambda_i(z)(z - G_2(s))\} \\ &> g_N^* - \sum_{i=1}^I \sup\{\lambda_i(z)(z - G_1(s))\} \\ &> \mu_s G_1(s - 1). \end{aligned}$$

By induction, $0 = G_2(-1) > G_1(-1) = 0$, which yields a contradiction. Therefore, for all $s = 0, \dots, N - 1$, $G_2(s) \leq G_1(s)$ and, consequently, $z_{i,s,2}^* \leq z_{i,s,1}^*$.

We analyze how the optimal reward varies as other parameters change. As earlier, subscripts 1 and 2 refer to systems 1 and 2, respectively. In Proposition 4.3, below, we characterize the

sensitivity of the optimal reward to the willingness-to-pay distribution as well as other system parameters.

Proposition 4.3. *Consider two systems, 1 and 2, that satisfy all of the following:*

1. $\Lambda_{i,1} \geq \Lambda_{i,2}$ for $i = 1, \dots, I$,
2. $F_{i,1} \leq F_{i,2}$ for $i = 1, \dots, I$,
3. $h_{s,1} \leq h_{s,2}$ for $s = 0, \dots, N$,
4. $\mu_{s,1} \geq \mu_{s,2}$ for $s = 1, \dots, N$.

Then $g_{N,1}^* \geq g_{N,2}^*$.

Proof. We will prove the result by contradiction. Suppose that conditions 1, 2, 3, and 4 hold, and $g_{N,1}^* < g_{N,2}^*$. Therefore,

$$G_1(N - 1) = \frac{g_{N,1}^* + h_{N,1}}{\mu_{N,1}} < \frac{g_{N,2}^* + h_{N,2}}{\mu_{N,2}} = G_2(N - 1).$$

Suppose that $G_1(s) < G_2(s)$ for some $s = 0, \dots, N - 1$. Then

$$\begin{aligned} g_{N,2}^* + h_{s,2} - \mu_{s,2}G_2(s - 1) &= \sum_{i=1}^I \sup\{\lambda_{i,2}(z)(z - G_2(s))\} \\ &\leq \sum_{i=1}^I \sup\{\lambda_{i,1}(z)(z - G_1(s))\} \\ &\leq g_{N,1}^* + h_{s,1} - \mu_{s,1}G_1(s - 1). \end{aligned}$$

So, $G_1(s - 1) < G_2(s - 1)$. By induction, $0 = G_1(-1) < G_2(-1) = 0$, exhibiting a contradiction. Thus, $g_{N,1}^* \geq g_{N,2}^*$.

We now compare our optimal policy with the optimal static price derived in [11]. Now suppose that $\mu_s = \mu$ and $h_s = \kappa s$ for some constant $\kappa \geq 0$ and for all $s = 0, \dots, N$. Also suppose that $I = 1$, as in [11]. It is clear that dynamic pricing achieves a better optimal profit. Moreover, there is an ordering relationship between our optimal static price and our optimal stationary policy. Since $I = 1$, we will omit the class subscript in the remainder of this section. For instance, we will write $r(\cdot)$, $\lambda(y)$, and z_s^* instead of $r_1(\cdot)$, $\lambda_1(y)$, and $z_{1,s}^*$. As in [11], let $\pi_n(\rho, N)$ denote the stationary probability of n customers in the system under static pricing when traffic intensity is ρ . In the same fashion, let $L(\rho, N)$ denote the long-run expected number of customers in the system under static pricing and traffic intensity ρ . In the following proofs we also use the quantity $\rho(y)$ that is defined as $\rho(y) = \lambda(y)/\mu$, which is simply the traffic intensity under static price y .

Proposition 4.4. *Let $I = 1$, and let y_N^* denote the optimal static price for a system of capacity N . Then, for all $s = 1, \dots, N$, $z_0^* \leq y_N^* \leq z_{N-1}^*$.*

In order to prove Proposition 4.4, we need the following lemmas, whose proofs are given in Appendix A.

Lemma 4.4. Let $I = 1$. For $s = 0, \dots, N - 1$, if $g_N^* + h_s > \mu_s G(s - 1)$ then

$$z_s^* = \sup\{z : (g_N^* + h_s - \mu_s G(s - 1))r(z) \leq \lambda(z)\},$$

otherwise, $z_s^* = \beta$.

Lemma 4.5. Consider an $M/M/1/N$ queue under traffic intensity $\rho \geq 0$. Then,

1. $\gamma_N(\rho) \geq \pi_0(\rho, N)$,
2. $1 - \pi_N(\rho, N) \geq \gamma_N(\rho)$,
3. $N - L(\rho, N) \geq \pi_0(\rho, N)\varphi_N(\rho)$,
4. $\rho\varphi_N(\rho)(1 - \pi_N(\rho, N)) \geq L(\rho, N)$,

where

$$\begin{aligned} \varphi_N(\rho) &= -\frac{\partial L(\rho, N)/\partial \rho}{\partial \pi_0(\rho, N)/\partial \rho} \\ &= \begin{cases} \frac{1 - (N + 1)^2 \rho^N (1 + \rho^2) + 2N(N + 2)\rho^{N+1} + \rho^{2N+2}}{(1 - \rho)^2(1 - (N + 1)\rho^N + N\rho^{N+1})} & \text{if } \rho \neq 1, \\ \frac{1}{6}N^2 + \frac{1}{2}N + \frac{1}{3} & \text{if } \rho = 1, \end{cases} \end{aligned}$$

and

$$\gamma_N(\rho) = \begin{cases} \frac{1 + N\rho^{N+1} - (N + 1)\rho^N}{(1 - \rho^{N+1})(1 - \rho^N)} & \text{if } \rho \neq 1, \\ \frac{1}{2} & \text{if } \rho = 1. \end{cases}$$

Proof of Proposition 4.4. From [11] recall that

$$\begin{aligned} z_{N-1}^* &= \inf\left\{y : r(y)\left(y - \frac{g_N^* + N\kappa}{\mu}\right) \geq 1\right\} \quad \text{and} \\ y_N^* &= \inf\left\{y : r(y)\gamma_N(\rho(y))\left(y - \frac{\kappa}{\mu}\varphi_N(\rho(y))\right) \geq 1\right\}. \end{aligned}$$

To prove that $y_N^* \leq z_{N-1}^*$, we will show that, for all $y \geq 0$ such that $r(y)(y - (g_N^* + N\kappa)/\mu) \geq 1$, we have $r(y)\gamma_N(\rho(y))(y - (\kappa/\mu)\varphi_N(\rho(y))) \geq 1$. Consider $y \geq 0$ such that $r(y)(y - (g_N^* + N\kappa)/\mu) \geq 1$. Since $g_N^* \geq \mu y(1 - \pi_0(\rho(y), N)) - h\kappa L(\rho(y), N)$, we can claim, by Lemma 4.5, that

$$\begin{aligned} y - \frac{g_N^* + N\kappa}{\mu} &\leq y\pi_0(\rho(y), N) - \frac{\kappa}{\mu}(N - L(\rho(y), N)) \\ &\leq \pi_0(\rho(y), N)\left(y - \frac{\kappa}{\mu}\varphi_N(\rho(y))\right) \\ &\leq \gamma_N(\rho(y))\left(y - \frac{\kappa}{\mu}\varphi_N(\rho(y))\right). \end{aligned}$$

Hence, we have $r(y)\gamma_N(\rho(y))(y - (\kappa/\mu)\varphi_N(\rho(y))) \geq 1$, proving that $y_N^* \leq z_{N-1}^*$.

To prove that $z_0^* \leq y_N^*$, we proceed by contradiction. Recall the alternate expression of z_0^* from Lemma 4.4, which is $z_0^* = \sup\{y : g_N^* r(y) \leq \lambda(y)\}$. If $z_0^* = \beta$ then $g_N^* = 0$, which is impossible. Suppose that $\beta > z_0^* > y_N^*$, so that there exists y in (y_N^*, z_0^*) such that $g_N^* r(y) \leq \lambda(y)$ and $r(y)\gamma_N(\rho(y))(y - (\kappa/\mu)\varphi_N(\rho(y))) > 1$. Using inequalities 2 and 4 from Lemma 4.5, we have

$$\begin{aligned} \lambda(y)\gamma_N(\rho(y))\left(y - \frac{\kappa}{\mu}\varphi_N(\rho(y))\right) &\leq \lambda(y)(1 - \pi_N(\rho(y), N))\left(y - \frac{\kappa}{\mu}\varphi_N(\rho(y))\right) \\ &\leq \lambda(y)y(1 - \pi_N(\rho(y), N)) \\ &\quad - \kappa\rho(y)\varphi_N(\rho(y))(1 - \pi_N(\rho(y), N)) \\ &\leq \lambda(y)y(1 - \pi_N(\rho(y), N)) - \kappa L(\rho(y), N) \\ &\leq g_N^*. \end{aligned}$$

Therefore, $g_N^* r(y) \geq \lambda(y)r(y)\gamma_N(\rho(y))(y - (\kappa/\mu)\varphi_N(\rho(y))) > \lambda(y)$, exhibiting a contradiction. Thus, we have proved that $z_0^* \leq y_N^*$.

Therefore, y_N^* is a convex combination of z_0^* and z_{N-1}^* . We can interpret the optimal static price as a ‘compromise’ between z_0^* and z_{N-1}^* . On the one hand, when the system is empty, the service provider is willing to discount prices to attract customers. On the other hand, when the system is almost full, the service provider charges a premium for high congestion costs. Under a static pricing scheme, the service provider does not have the possibility to differentiate states when pricing service. Hence, it is intuitive that the optimal price to be charged in this case lies between the optimal dynamic prices charged in extremal states.

5. Infinite capacity queues

In this section we impose no limitation on the system capacity. This introduces some difficulties since the system of ACOE now has infinitely many equations and solution triplets. Moreover, unlike in the finite capacity case, a canonical triplet does not always translate into an optimal stationary policy. Nevertheless, under certain holding cost and service rate structures, we are able to find an optimal stationary policy that maximizes the long-term average profit per unit time.

5.1. Uniform asymptotic holding cost and service rate

Let us assume that h_s and μ_s are constant past a certain state N , that is, $\mu_s = \mu_N$ and $h_s = h_N$ for all $s \geq N$. We call this assumption the uniform asymptotic holding cost and service rate. In this case we can reduce the ACOE to only a finite number of equations and prove that the corresponding canonical triplet is optimal. To ensure stability, we also assume that $\sum_{i=1}^I \Lambda_i < \mu_N$.

Consider the mapping $\Psi : \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ defined as

$$\Psi(V, g) = \frac{g + h_N - \sum_{i=1}^I \sup\{\lambda_i(z)(z - V)\}}{\mu_N}.$$

Some properties of Ψ are stated in the following lemma, whose proof is given in Appendix A.

Lemma 5.1. *For all $g > 0$, $\Psi(\cdot, g)$ is a continuous, nondecreasing contraction mapping from \mathbb{R} to $(-\infty, (g + h_N)/\mu_N]$ and has a unique fixed point denoted by $FP(g)$. Moreover, $FP(\cdot)$ is increasing and continuous on $[0, \infty)$.*

If we can find a triplet $(g, G(\cdot), \mathbf{z})$ satisfying

$$G(-1) = 0, \tag{5.1}$$

$$g + h_s - \mu_s G(s - 1) = \sum_{i=1}^I \sup\{\lambda_i(z)(z - G(s))\}, \quad s = 0, \dots, N - 1, \tag{5.2}$$

$$G(N - 1) = \text{FP}(g), \tag{5.3}$$

we have exhibited a canonical triplet that satisfies the infinitely many ACOE. Indeed, the equation corresponding to state $s \geq N$ is $g + h_N - \mu_N G(s - 1) = \sum_{i=1}^I \sup\{(z - G(s))\lambda_i(z)\}$, which is $G(s - 1) = \Psi(G(s), g)$. It is clear that it is satisfied by $G(s) = G(s - 1) = \text{FP}(g)$. We now show that such a canonical triplet exists and that it corresponds to an optimal stationary policy.

Theorem 5.1. *There exists a canonical triplet $(g, G(\cdot), \mathbf{z})$ for the system of ACOE (5.1)–(5.3). Moreover, $\mathbf{z} = \mathbf{z}^*$ and $g = g_\infty^*$, where $z_{s,i}^* = \inf\{z : r_i(z)(z - G(s)) \geq 1\}$ for all $s \geq 0$.*

Proof. First, we prove the existence of a canonical triplet $(g, G(\cdot), \mathbf{z})$ to the ACOE (5.1)–(5.3). Note that

$$\Psi\left(0, \sum_{i=1}^I \sup\{\lambda_i(z)z\}\right) = \frac{h_N}{\mu_N} \geq 0,$$

and, consequently, $\text{FP}(\sum_{i=1}^I \sup\{\lambda_i(z)z\}) \geq 0$. Therefore, the proof is exactly the same as the proof of Lemma 4.1, except that $G(N - 1, g) = \text{FP}(g)$ is the starting point of the induction.

We now show that $(g, G(\cdot), \mathbf{z})$ corresponds to an optimal stationary policy. According to Equation (5.2.12) of [8], we need to show that

$$\lim_{t \rightarrow \infty} \inf_{d \in \Pi^{RH}} \frac{E_{s_0}^d[l(X(t))]}{t} = 0,$$

where Π^{RH} is the set of all history-dependent randomized policies and s_0 is the starting state at time $t = 0$. If this is the case, the canonical triplet corresponds to an optimal stationary policy. Recall that $l(s) - l(s + 1) = G(s)$ for all $s = 0, \dots, N - 1$. Therefore, $l(s) = l(N) - (s - N)\text{FP}(g)$ for all $s \geq N$. Hence,

$$E_{s_0}^d[l(X(t))] = E_{s_0}^d[l(X(t) \mid X(t) < N)]P(X(t) < N) + (l(N) - (E_{s_0}^d[X(t) \mid X(t) \geq N] - N)\text{FP}(g))P(X(t) \geq N).$$

Note that, for all $d \in \Pi^{RH}$,

$$|E_{s_0}^d[l(X(t) \mid X(t) < N)]P(X(t) < N)| \leq \max_{s \leq N-1} |l(s)|.$$

Therefore, we have

$$\lim_{t \rightarrow \infty} \inf_{d \in \Pi^{RH}} \frac{|E_{s_0}^d[l(X(t) \mid X(t) < N)]P(X(t) < N)|}{t} = 0.$$

It remains to show that

$$\lim_{t \rightarrow \infty} \sup_{d \in \Pi^{RH}} \frac{E_{s_0}^d[X(t) \mid X(t) \geq N]P(X(t) \geq N)}{t} = 0.$$

First, note that

$$\sup_{d \in \Pi^{RH}} E_{s_0}^d [X(t) \mid X(t) \geq N] P(X(t) \geq N) \leq \sup_{d \in \Pi^{RH}} E_{s_0}^d [X(t)]$$

and that the supremum, $\sup_{d \in \Pi^{RH}} E_{s_0}^d [X(t)]$, is attained for policy \hat{z} , where \hat{z} denotes the stationary policy of charging price 0 to all customers in all states ($\hat{z}_{i,s} = 0$ for $i = 1, \dots, I$ and $s = 0, \dots, \infty$). We have

$$\lim_{t \rightarrow \infty} \sup_{d \in \Pi^{RH}} E_{s_0}^d [X(t)] \leq \lim_{t \rightarrow \infty} E_{s_0}^{\hat{z}} [X(t)] \leq \frac{\sum_{i=1}^I \Lambda_i}{\mu_N - \sum_{i=1}^I \Lambda_i}.$$

Therefore,

$$\lim_{t \rightarrow \infty} \inf_{d \in \Pi^{RH}} \frac{E_{s_0}^d [I(X(t))]}{t} = 0,$$

and we can apply Theorem 5.2.4 of [8] and claim that $z = z^*$ and $g = g_\infty^*$.

This theorem enables us to explicitly characterize an optimal stationary policy. Note that the service provider charges the same price $z_{i,N-1}^* = \inf \{z : r_i(z)(z - FP(g_\infty^*)) \geq 1\}$ to class- i customers for all states $s \geq N - 1$. This property is quite surprising since there is no apparent symmetry in the transition structure to justify it.

We now derive structural properties of our solution in states $s < N$. In Proposition 5.1, below, we extend Proposition 4.1 to infinite capacity systems under certain conditions on the system parameters. We demonstrate that the optimal prices to be charged in states $s = 0, \dots, N - 1$ are nondecreasing in s .

Proposition 5.1. *Suppose that $\{\mu_s\}_{s=0}^N$ and $\{h_s\}_{s=0}^N$ are such that there exists an integer q between 0 and N , where $\mu_0 \leq \mu_1 \leq \dots \leq \mu_q = \mu_{q+1} = \dots = \mu_N$ and $0 = h_0 = h_1 = \dots = h_q \leq h_{q+1} \leq \dots \leq h_N$. Then $z_{i,s}^*$ is nondecreasing in s .*

To prove this result, we need the following lemma.

Lemma 5.2. *Suppose that $\{\mu_s\}_{s=0}^N$ and $\{h_s\}_{s=0}^N$ are such that there exists an integer q between 0 and N , where $\mu_0 \leq \mu_1 \leq \dots \leq \mu_q = \mu_{q+1} = \dots = \mu_N$ and $0 = h_0 = h_1 = \dots = h_q \leq h_{q+1} \leq \dots \leq h_N$. Then $G(s)$ is a nondecreasing function of s .*

Proof. Proving that the result holds for states $s = 0, \dots, q$ is the same as in the proof of Lemma 4.3.

If $q = N$, the proof is complete. Otherwise, recall that $G(N - 1) = FP(g_\infty^*)$. Therefore,

$$\begin{aligned} G(N - 2) &= \frac{g_\infty^* + h_{N-1} - \sum_{i=1}^I \sup\{\lambda_i(z)(z - FP(g_\infty^*))\}}{\mu_N} \\ &\leq \frac{g_\infty^* + h_N - \sum_{i=1}^I \sup\{\lambda_i(z)(z - FP(g_\infty^*))\}}{\mu_N} \\ &= FP(g_\infty^*) \\ &= G(N - 1). \end{aligned}$$

We prove, by induction, that $G(s) \geq G(s - 1)$ holds for states $s = q, \dots, N - 1$ in the same fashion as in the proof of Lemma 4.3.

Proof of Proposition 5.1. Recall that $z_{i,s}^* = \inf\{z: r_i(z)(z - G(s)) \geq 1\}$. Since $G(s)$ is nondecreasing with respect to s , so is $z_{i,s}^*$.

As in the finite capacity case, the service provider charges more when the system is congested, keeping higher states less attractive to customers in order to control holding costs. Let us now consider the particular case where there is no holding cost ($h = 0$).

Proposition 5.2. *If $h_s = 0$ and $\mu_s = \mu$ for all $s = 1, \dots, \infty$ then the optimal dynamic policy is the same as the optimal static policy.*

Proof. We just need to check that

$$g_\infty^* = \sum_{i=1}^I \sup\{\lambda_i(z)z\}, \quad G(s) = 0,$$

and

$$z_{i,s}^* = \inf\{z: zr_i(z) \geq 1\},$$

for $s = 0, \dots, \infty$ satisfy the ACOE, where $FP(g_\infty^*) = 0$. This is easily verified and the rest of the proof is omitted.

In this case the service manager treats all states as identical. This is quite intuitive since the absence of holding costs enables us to consider only the arrival process to optimize revenue. The queueing process becomes irrelevant in computing the long-term average reward since no customer ever leaves due to capacity restrictions and no holding cost is incurred.

5.2. Linear holding cost

Under uniform asymptotic holding cost and service rate, we are able to derive an optimal stationary pricing policy explicitly. The assumption that $\mu_s = \mu_N$ for $s \geq N$ for some N is often encountered as servers become saturated. However, we might have each customer incur the same holding cost κ while in the system. Then $h_s = \kappa s$ and the uniform asymptotic holding cost cannot hold simultaneously. In this section we investigate systems where holding costs and service rates are of the form $h_s = \kappa s$ and $\mu_{s+1} = \mu$ for $s \geq 0$. To ensure the stability of the system, we impose the condition $\sum_{i=1}^I \Lambda_i < \mu$. Although we are only considering linear holding costs and a single server here, Proposition 5.3, Proposition 5.4, and Theorem 5.2 hold in a stable q -server setting with h_s monotonically increasing to ∞ and integrable with respect to $\{(\sum_{i=1}^I \Lambda_i)^s / \mu(1 \wedge q) \cdots \mu(s \wedge q)\}$. We show the existence of an optimal stationary policy by approximating the system with a finite capacity system. First, let us consider willingness-to-pay distributions with finite support ($\beta_i < \infty$ for all $i = 1, \dots, I$). In this case we demonstrate, in the next proposition, that we can restrict our analysis to finite capacity systems when $\beta_i < \infty$ for all $i = 1, \dots, I$.

Proposition 5.3. *If $\beta_i < \infty$ for all $i = 1, \dots, I$ then $g_\infty^* = g_M^*$, where $M = \lfloor \max \beta_i \mu / \kappa \rfloor$.*

Proof. To prove this proposition, we show that for any stationary policy of the infinite capacity system, we can find a stationary policy of the truncated M -capacity system that performs as well. Let \mathbf{z} be a stationary pricing policy for the infinite capacity system such that $R(\mathbf{z}) > 0$. This policy exists since the assumption that $\max \beta_i > \kappa / \mu$ ensures the existence of a positive reward. Now consider the M -capacity stationary pricing policy \mathbf{z}^M , which is defined as the truncation of \mathbf{z} up to state $M - 1$ included. More precisely, $z_{i,s}^M = z_{i,s}$ for all $s < M$ and

$i = 1, \dots, I$. We have

$$\begin{aligned} R(\mathbf{z}) &= \sum_{s=0}^{\infty} \left(\sum_{i=1}^I \lambda_i(z_{i,s}) z_{i,s} - \kappa s \right) \pi_s(\mathbf{z}) \\ &= \sum_{s=0}^{\infty} \left(\sum_{i=1}^I \lambda_i(z_{i,s}) \left(z_{i,s} - \frac{\kappa}{\mu}(s+1) \right) \right) \pi_s(\mathbf{z}) \\ &= \sum_{s=0}^{\infty} a_s \pi_s(\mathbf{z}) \end{aligned}$$

and

$$R(\mathbf{z}^{\mathbf{M}}) = \sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}^{\mathbf{M}}),$$

where $a_s = \sum_{i=1}^I \lambda_i(z_{i,s})(z_{i,s} - (\kappa/\mu)(s+1))$. Clearly, the definition of M as $M = \lfloor \max \beta_i \mu / \kappa \rfloor$ implies that $a_s \leq 0$ for $s \geq M$. It is straightforward to show that, for all $s \leq M$,

$$\pi_s(\mathbf{z}^{\mathbf{M}}) = \frac{\pi_s(\mathbf{z})}{\sum_{s=0}^M \pi_s(\mathbf{z})}.$$

Hence, $R(\mathbf{z}^{\mathbf{M}}) - R(\mathbf{z})$ has the same sign as

$$\sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) - \sum_{s=0}^{\infty} a_s \pi_s(\mathbf{z}) \sum_{s=0}^M \pi_s(\mathbf{z}),$$

and we have

$$\begin{aligned} \sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) - \sum_{s=0}^{\infty} a_s \pi_s(\mathbf{z}) \sum_{s=0}^M \pi_s(\mathbf{z}) &\geq \sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) - \sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) \sum_{s=0}^M \pi_s(\mathbf{z}) \\ &\geq \sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) \left(1 - \sum_{s=0}^M \pi_s(\mathbf{z}) \right). \end{aligned}$$

Recall that $R(\mathbf{z}) > 0$ and

$$\sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) \geq \sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) + \sum_{s=M}^{\infty} a_s \pi_s(\mathbf{z}) \geq R(\mathbf{z}) > 0.$$

Therefore,

$$\sum_{s=0}^{M-1} a_s \pi_s(\mathbf{z}) - \sum_{s=0}^{\infty} a_s \pi_s(\mathbf{z}) \sum_{s=0}^M \pi_s(\mathbf{z}) \geq 0$$

and $R(\mathbf{z}^{\mathbf{M}}) \geq R(\mathbf{z})$, proving the result.

Without loss of generality, we now assume that at least one of the willingness-to-pay distributions F_1, \dots, F_I has infinite support in the rest of this section. To prove the existence of an optimal stationary policy, we approximate the infinite capacity system by a finite capacity model of large size and show the following two convergence results. Note that Weber and Stidham [17] provided an existence proof in the case of a compact action space.

Proposition 5.4. *If $\sum_{i=1}^I \Lambda_i < \mu$ then g_N^* converges to g_∞^* as N goes to ∞ .*

Theorem 5.2. *Let $(g_N^*, G(\cdot), z^N)$ be the canonical triplet associated with a truncated system with capacity N . Then, under the stability condition $\sum_{i=1}^I \Lambda_i < \mu$, there exists z such that $z_{i,s}^N \downarrow z_{i,s}$ for all i, s as $N \rightarrow \infty$. Moreover, $z = z^*$ is optimal for the infinite capacity model.*

We need the following lemma to prove Proposition 5.4 and Theorem 5.2.

Lemma 5.3. *Let z be a stationary pricing policy and $z^{|N}$ be the truncation of z up to state $N - 1$. Under the stability condition $\sum_{i=1}^I \Lambda_i < \mu$, $R(z^{|N}) \rightarrow R(z)$ as N goes to ∞ .*

Proof. Consider

$$\begin{aligned} R(z) &= \sum_{s=0}^{\infty} \pi_s(z) \sum_{i=1}^I \lambda_i(z_{i,s}) z_{i,s} - \kappa(s + 1) \pi_{s+1}(z) \\ &= \pi_0(z) \sum_{s=0}^{\infty} \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s-1})}{\mu^s} \left(\sum_{i=1}^I \lambda_i(z_{i,s}) \left(z_{i,s} - \frac{\kappa(s + 1)}{\mu} \right) \right). \end{aligned}$$

Moreover,

$$\begin{aligned} R(z^{|N}) &= \sum_{s=0}^{N-1} \pi_s(z^{|N}) \sum_{i=1}^I \lambda_i(z_{i,s}) z_{i,s} - \kappa(s + 1) \pi_{s+1}(z^{|N}) \\ &= \pi_0(z^{|N}) \sum_{s=0}^{N-1} \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s-1})}{\mu^s} \left(\sum_{i=1}^I \lambda_i(z_{i,s}) \left(z_{i,s} - \frac{\kappa(s + 1)}{\mu} \right) \right). \end{aligned}$$

Since

$$(\pi_0(z^{|N}))^{-1} = 1 + \sum_{s=0}^{N-1} \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s})}{\mu^{s+1}}$$

as N goes to ∞ , $\pi_0(z^{|N}) \rightarrow \pi_0(z)$. Therefore, $R(z^{|N}) \rightarrow R(z)$, which proves the desired result.

Proof of Proposition 5.4. We will prove this proposition by contradiction. From Proposition 4.2 we know that g_N^* is nondecreasing in N and that $\lim_N g_N^*$ exists and is less than or equal to g_∞^* . Now suppose that $\lim_N g_N^* < g_\infty^*$. Then, according to Lemma 5.3, there exists an N -capacity stationary policy z^N such that $\lim_N g_N^* < R(z^N) < g_\infty^*$. As $\lim_N g_N^* \geq g_N^*$, we have a contradiction and the proof is complete.

Proof of Theorem 5.2. We already proved, in Propositions 4.2 and 5.4, that g_N^* converges to g_∞^* and that $z_{i,s}^N$ is a nonincreasing sequence in N . Therefore, $\lim_N z_{i,s}^N = z_{i,s}$ exists.

We will show that $|R(z) - g_N^*| \rightarrow 0$. Since $g_N^* \rightarrow g_\infty^*$, it will imply that $R(z) = g_\infty^*$ and that z is optimal.

To do so, we will first prove that, for any s , $\pi_s(z^N) \rightarrow \pi_s(z)$ as N goes to ∞ . Since

$$\pi_s(z^N) = \pi_0(z^N) \frac{\sum_i \lambda_i(z_{i,0}^N) \cdots \sum_i \lambda_i(z_{i,s-1}^N)}{\mu^s},$$

we only need to prove that $\pi_0(\mathbf{z}^N) \rightarrow \pi_0(\mathbf{z})$. We have

$$\begin{aligned} \pi_0(\mathbf{z}^N)^{-1} - \pi_0(\mathbf{z})^{-1} &= \sum_{s=1}^N \frac{\sum_i \lambda_i(z_{i,0}^N) \cdots \sum_i \lambda_i(z_{i,s-1}^N)}{\mu^s} \\ &\quad - \sum_{s=1}^{\infty} \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s-1})}{\mu^s}. \end{aligned}$$

Let M be an arbitrary integer smaller than N ,

$$\begin{aligned} \pi_0(\mathbf{z}^N)^{-1} - \pi_0(\mathbf{z})^{-1} &= \sum_{s=1}^M \frac{\sum_i \lambda_i(z_{i,0}^N) \cdots \sum_i \lambda_i(z_{i,s-1}^N)}{\mu^s} - \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s-1})}{\mu^s} \\ &\quad + \sum_{s=M}^N \frac{\sum_i \lambda_i(z_{i,0}^N) \cdots \sum_i \lambda_i(z_{i,s-1}^N)}{\mu^s} \\ &\quad - \sum_{s=M}^{\infty} \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s-1})}{\mu^s}. \end{aligned}$$

Hence,

$$\begin{aligned} |\pi_0(\mathbf{z}^N)^{-1} - \pi_0(\mathbf{z})^{-1}| &\leq \sum_{s=1}^M \left| \frac{\sum_i \lambda_i(z_{i,0}^N) \cdots \sum_i \lambda_i(z_{i,s-1}^N)}{\mu^s} - \frac{\sum_i \lambda_i(z_{i,0}) \cdots \sum_i \lambda_i(z_{i,s-1})}{\mu^s} \right| \\ &\quad + 2 \sum_{s=M}^{\infty} \frac{(\sum_i \Lambda_i)^s}{\mu^s}. \end{aligned}$$

First, let N go to ∞ and then let M go to ∞ . We have $\pi_s(\mathbf{z}^N) \rightarrow \pi_s(\mathbf{z})$ for all $s \geq 0$ as N goes to ∞ .

Now consider

$$\begin{aligned} R(\mathbf{z}) - g_N^* &= \sum_{s=0}^{\infty} \sum_{i=1}^I z_{i,s} \lambda_i(z_{i,s}) \pi_s(\mathbf{z}) - \sum_{s=0}^{N-1} \sum_{i=1}^I z_{i,s}^N \lambda_i(z_{i,s}^N) \pi_s(\mathbf{z}^N) \\ &\quad - \kappa \sum_{s=1}^{\infty} s \pi_s(\mathbf{z}) + \kappa \sum_{s=1}^N s \pi_s(\mathbf{z}^N). \end{aligned}$$

So,

$$\begin{aligned} |R(\mathbf{z}) - g_N^*| &\leq \sum_{s=0}^M \left| \sum_{i=1}^I z_{i,s} \lambda_i(z_{i,s}) \pi_s(\mathbf{z}) - z_{i,s}^N \lambda_i(z_{i,s}^N) \pi_s(\mathbf{z}^N) \right| + \kappa \sum_{s=1}^M s |\pi_s(\mathbf{z}) - \pi_s(\mathbf{z}^N)| \\ &\quad + 2 \sup_{i,z} \{z \lambda_i(z)\} \sum_{s=M}^{\infty} \frac{(\max \Lambda_i)^s}{\mu^s} + 2\kappa \sum_{s=M}^{\infty} s \frac{(\max \Lambda_i)^s}{\mu^s}. \end{aligned}$$

Letting N go to ∞ yields

$$\lim_{N \rightarrow \infty} |R(\mathbf{z}) - g_N^*| \leq 2 \sup_{i,z} \{z \lambda_i(z)\} \sum_{s=M}^{\infty} \frac{(\max \Lambda_i)^s}{\mu^s} + 2\kappa \sum_{s=M}^{\infty} s \frac{(\max \Lambda_i)^s}{\mu^s},$$

and letting M go to ∞ implies that $\lim_N g_N^* = R(z)$. Therefore, $z = z^*$, and the proof is complete.

The key element of this result is that the stationary probability of being in higher states is negligible under any stationary policy. Thus, an infinite capacity system can be approximated by a system of large finite capacity. The optimal stationary policy exists and is the limit of finite-capacity optimal policies, which enables us to state the following proposition.

Proposition 5.5. *Under the stability condition $\sum_{i=1}^I \Lambda_i < \mu$, $z_{i,s}^*$ is nondecreasing in s .*

Proof. From Theorem 5.2 we know that $z_{i,s}^* = \lim_N z_{i,s}^N$, where $z_{i,s}^N$ is the optimal price at state s for the truncated N -capacity system. By Proposition 4.1, $z_{i,s}^N$ is nondecreasing in s . Hence, the same holds for $z_{i,s}^*$.

Although we do not characterize the optimal stationary policy explicitly in this case, we can still derive some insights. Not surprisingly, the structures of the infinite and finite capacity optimal policies are the same. High prices are posted for congested states in order to minimize holding costs.

Moreover, it can be shown that $G(s) > s(\kappa/\mu)$ and $z_{i,s}^* \geq G(s)$. Therefore, $z_{i,s}^* > s(\kappa/\mu)$. This inequality is quite intuitive since the service provider has to charge at least enough to cover the expected holding cost of a customer admitted in state s , which is greater than κ times the expected service time of the s customers in front of her. A direct consequence of this result is that when $N = \infty$, $z_{i,s}^* \rightarrow \infty$ as s goes to ∞ . Since $z\lambda_i(z)$ converges to 0 as z goes to ∞ , very little profit will be generated from congested states. Thus, profit will remain very close to optimal if the service provider decides to limit the system size to a large capacity.

6. Concluding remarks

We characterized optimal pricing policies that maximize the long-run average profit per unit time. In systems with finite capacity and in systems with infinite capacity under uniform asymptotic holding cost and service rate, we found an exact solution to the ACOE that corresponds to an optimal stationary pricing policy. In systems with infinite capacity and more general holding cost and service rate structure, we showed that an optimal stationary pricing policy exists as the limiting pricing solution for finite capacity systems whose size grows to ∞ .

Moreover, we proved that the optimal stationary prices are nondecreasing with the state index and perform a congestion control that prevents high holding costs in congested states.

In all the above, we did not consider capacity N as a decision variable. However, the service provider has indirect control on N . If the ACOE canonical triplet yields $z_{i,s}^* = \beta_i$ for some state s , then it is optimal not to let anyone from class i enter the facility in state s . The service manager limits the size of the buffer to $\min\{s : z_{i,s}^* = \beta_i \text{ for all } i\}$. Note that when the willingness-to-pay distribution F_i has infinite support, $z_{i,s}^* < \infty$ for all states s , $s = 0, \dots, N - 1$, so that the service provider never denies entry to customers at optimality.

Appendix A.

Proof of Lemma 4.4. Since $I = 1$, we omit the class subscript in this proof. All derivatives in this proof are with respect to ρ . In Theorem 4.1 we proved that $z_s^* = \inf\{z : r(z)(z - G(s)) \geq 1\}$. Now suppose that $g_N^* + h_s > \mu_s G(s - 1)$. Therefore, $z_s^* < \beta$. It is straightforward to show

that $\sup\{\lambda(z)(z - G(s)) - g_N^* - h_s + \mu_s G(s - 1)\} = 0$ is equivalent to

$$\sup\left\{z - G(s) - \frac{g_N^* + h_s - \mu_s G(s - 1)}{\lambda(z)}\right\} = 0,$$

where z_s^* is the unique price that attains the supremum. Let

$$t_s(z) = z - G(s) - \frac{g_N^* + h_s - \mu_s G(s - 1)}{\lambda(z)},$$

$$t'_s(z) = 1 - (g_N^* + h_s - \mu_s G(s - 1)) \frac{r'(z)}{\lambda(z)} \quad \text{a.e. on } [\alpha, \beta].$$

Under IGHR, Proposition 5.1 of [19] shows that $t'_s(\cdot)$ is strictly decreasing a.e on $(\inf\{z : zr(z) \geq 1\}, \beta)$, which includes (z_s^*, β) . Hence, $t'_s(\cdot) < 0$ a.e. on (z_s^*, β) . Therefore,

$$z_s^* = \sup\{z : t'_s(z) \geq 0\} = \sup\{z : (g_N^* + h_s - \mu_s G(s - 1))r(z) \leq \lambda(z)\}.$$

Suppose that $g_N^* + h_s \geq \mu_s G(s - 1)$. Since we always have $g_N^* + h_s \leq \mu_s G(s - 1)$, we can claim that $g_N^* + h_s = \mu_s G(s - 1)$ and $\sup\{(z - G(s))\lambda(z)\} = 0$. We have two possibilities: $z_s^* = \beta$ or $z_s^* = G(s) < \beta$. The latter is impossible since there must exist $\varepsilon > 0$ such that $G(s) + \varepsilon < \beta$ and $\varepsilon\lambda(G(s) + \varepsilon) > 0$. Therefore, $z_s^* = \beta$.

Proof of Lemma 4.5. For clarity, we will omit the arguments of the quantities we use in this proof. For instance, we will write π_0 instead of $\pi_0(\rho, N)$.

1. We prove first that $\gamma_N \geq \pi_0$. When $\rho = 1$, $\pi_0 = 1/(N + 1)$ and $\gamma_N = \frac{1}{2}$, which agrees with our claim. Otherwise, recall that $\pi_0 = (1 - \rho)/(1 - \rho^{N+1})$. Therefore,

$$\begin{aligned} \frac{\pi_0}{\gamma_N} &= \frac{(1 - \rho)(1 - \rho^N)}{1 + N\rho^{N+1} - (N + 1)\rho^N} \\ &= \frac{(1 - \rho)^2 \sum_{k=0}^{N-1} \rho^k}{(1 - \rho)(\sum_{k=0}^{N-1} \rho^k - N\rho^N)} \\ &= \frac{(1 - \rho) \sum_{k=0}^{N-1} \rho^k}{-\sum_{k=0}^{N-1} (1 - \rho^k) + N(1 - \rho^N)} \\ &= \frac{\sum_{k=0}^{N-1} \rho^k}{-\sum_{k=0}^{N-1} \sum_{s=0}^{k-1} \rho^s + N \sum_{s=0}^{N-1} \rho^s} \\ &= \frac{\sum_{k=0}^{N-1} \rho^k}{-\sum_{s=0}^{N-2} \rho^s (N - s - 1) + N \sum_{s=0}^{N-1} \rho^s} \\ &= \frac{\sum_{k=0}^{N-1} \rho^k}{\sum_{s=0}^{N-2} \rho^s (s + 1) + N\rho^{N-1}} \\ &= \frac{\sum_{k=0}^{N-1} \rho^k}{\sum_{s=0}^{N-1} \rho^s + \sum_{s=0}^{N-2} s\rho^s + (N - 1)\rho^{N-1}} \\ &\leq 1. \end{aligned}$$

2. Let us prove now that $1 - \pi_N \geq \gamma_N$, which is easily verified when $\rho = 1$. Now suppose that $\rho \neq 1$. Since $1 - \pi_N = (1 - \rho^N)/(1 - \rho^{N+1})$, we have

$$\begin{aligned} \frac{1 - \pi_N}{\gamma_N} - 1 &= \frac{(1 - \rho^N)^2}{1 + N\rho^{N+1} - (N + 1)\rho^N} - 1 \\ &= \frac{1 - 2\rho^N + \rho^{2N} - 1 - N\rho^{N+1} + (N + 1)\rho^N}{1 + N\rho^{N+1} - (N + 1)\rho^N} \\ &= \frac{\rho^N(\rho^N - N\rho + (N - 1))}{1 + N\rho^{N+1} - (N + 1)\rho^N}. \end{aligned}$$

We know, from [20], that

$$1 + N\rho^{N+1} - (N + 1)\rho^N \geq 0$$

and

$$\begin{aligned} \rho^N - N\rho + (N - 1) &= (\rho^N - 1) - N(\rho - 1) \\ &= (\rho - 1) \sum_{k=0}^{N-1} (\rho^k - 1) \\ &\geq 0. \end{aligned}$$

Therefore, $1 - \pi_N \geq \gamma_N$.

3. We need to show that $N - L \geq \pi_0\varphi_N$. When $\rho = 1$, we have

$$N - L = \frac{N}{2}, \quad \pi_0 = \frac{1}{N + 1}, \quad \text{and} \quad \varphi_N = \frac{1}{6}N^2 + \frac{1}{2}N + \frac{1}{3}.$$

So,

$$\frac{N - L}{\pi_0} = \frac{1}{2}N^2 + \frac{1}{2}N \geq \varphi_N.$$

Now suppose that $\rho \neq 1$. We have $L = \pi_0 \sum_{n=0}^N n\rho^n = \pi_0 F$, where $F = \sum_{n=0}^N n\rho^n$. Recall, from [11], that $\varphi_N = -L'/\pi'_0$. Using the fact that $\pi'_0 = -\pi_0^2(F/\rho)$,

$$\begin{aligned} \pi_0\varphi_N &= \pi_0 \frac{\pi'_0 F + \pi_0 F'}{-\pi'_0} \\ &= -L - \frac{\pi_0^2}{\pi'_0} F' \\ &= -L + \frac{F'\rho}{F} \\ &= -L + \frac{\sum_{k=0}^N k^2 \rho^k}{F} \\ &\leq N - L. \end{aligned}$$

4. It remains to prove that $\rho\varphi_N(1 - \pi_N) \geq L$. When $\rho = 1$, we have

$$\rho\varphi_N(1 - \pi_N) = \frac{N}{N + 1} \left(\frac{1}{6}N^2 + \frac{1}{2}N + \frac{1}{3} \right) \geq \frac{N}{2} = L.$$

Now suppose that $\rho \neq 1$. In the same fashion, $\varphi_N = -F - (\pi_0/\pi'_0)F'$. Moreover, note that $\rho(1 - \pi_N) = 1 - \pi_0$. So,

$$\begin{aligned} \rho\varphi_N(1 - \pi_N) - L &= \varphi_N(1 - \pi_0) - L, \\ &= \left(-F - \frac{\pi_0}{\pi'_0}F'\right)(1 - \pi_0) - L \\ &= -F - \frac{\pi_0}{\pi'_0}F'(1 - \pi_0) \\ &= -F + \rho^2 \frac{F'}{F} \frac{1 - \rho^N}{1 - \rho} \\ &= -F + \rho^2 \frac{F'}{F} \sum_{k=0}^{N-1} \rho^k. \end{aligned}$$

Note that this quantity has the same sign as $-F^2 + \rho^2 F' \sum_{k=0}^{N-1} \rho^k$ and that

$$\begin{aligned} -F^2 + \rho^2 F' \sum_{k=0}^{N-1} \rho^k &= -\left(\sum_{k=1}^N k\rho^k\right)^2 + \sum_{k=1}^N k^2 \rho^k \sum_{k=1}^N \rho^k \\ &= \sum_{k=2}^{2N} \rho^k \sum_{n=\max(0,k-N)}^{\min(N,k)} n(n-k) + n^2 \\ &= \sum_{k=2}^{2N} \rho^k \sum_{n=\max(0,k-N)}^{\min(N,k)} n(2n-k) \\ &= \sum_{k=2}^{2N} \rho^k \sum_{n=k/2-\min(N,k)}^{\min(N,k)-k/2} \left(\frac{k}{2} + n\right)2n. \end{aligned}$$

Let $b_{n,k} = (k/2 + n)2n$ for $k/2 - \min(N, k) \leq n \leq \min(N, k) - k/2$. For $0 \leq n \leq \min(N, k) - k/2$, we have

$$b_{n,k} = \left(\frac{k}{2} + n\right)2n \geq \left(\frac{k}{2} - n\right)2n = -b_{-n,k}.$$

Therefore,

$$\sum_{n=k/2-\min(N,k)}^{\min(N,k)-k/2} b_{n,k} = \sum_{n=k/2-\min(N,k)}^{\min(N,k)-k/2} \left(\frac{k}{2} + n\right)2n \geq 0.$$

Hence, we can claim that $\rho\varphi_N(1 - \pi_N) \geq L$ and the proof is complete.

Proof of Lemma 5.1. It is clear that $\Psi(V, g)$ is nondecreasing in V and continuity can be proven as $\Psi(V, g)$ depends on V through the supremum of a bounded continuous function of z .

We will now prove that $\Psi(\cdot, g)$ is a contraction. Suppose that $V_1 \leq V_2$. Then,

$$\begin{aligned} \mu_N(\Psi(V_2, g) - \Psi(V_1, g)) &= \sum_{i=1}^I \sup\{\lambda_i(z)(z - V_1)\} - \sum_{i=1}^I \sup\{\lambda_i(z)(z - V_2)\} \\ &\leq \sum_{i=1}^I \lambda_i(z_i(V_1))(z_i(V_1) - V_1) - \lambda_i(z_i(V_1))(z_i(V_1) - V_2), \\ &\leq (V_2 - V_1) \sum_{i=1}^I \Lambda_i, \end{aligned}$$

where $z_i(V)$ is the unique maximizer in $[\alpha_i, \beta_i]$ of $\lambda_i(z)(z - V)$. Therefore,

$$\Psi(V_2, g) - \Psi(V_1, g) \leq \frac{\sum_{i=1}^I \Lambda_i}{\mu_N} (V_2 - V_1),$$

which proves that $\Psi(\cdot, g)$ is a contraction mapping and has a unique fixed point.

It remains to show that $\text{FP}(\cdot)$ is increasing and continuous. Let $0 \leq g_1 < g_2$. Therefore, $\Psi(\text{FP}(g_2), g_1) < \Psi(\text{FP}(g_2), g_2) = \text{FP}(g_2)$, so $\text{FP}(g_1) < \text{FP}(g_2)$, proving that $\text{FP}(\cdot)$ is increasing. As $\Psi(\cdot, g) \leq (g + h_N)/\mu_N$, we also have $\text{FP}(g) \leq (g + h_N)/\mu_N$.

To prove that $\text{FP}(\cdot)$ is continuous, we show, by contradiction, that it is both left-continuous and right-continuous. Let $g_n \uparrow g$ such that $g_n \geq 0$ for all n . Therefore, $\text{FP}(g_n)$ has a limit $\lim_n \text{FP}(g_n) \leq \text{FP}(g)$. Suppose that $\lim_n \text{FP}(g_n) < \text{FP}(g)$. Hence, $\lim_n \text{FP}(g_n) < \Psi(\lim_n \text{FP}(g_n), g)$. But $\lim_n \text{FP}(g_n) \geq \text{FP}(g_m)$ for all $m \geq 0$, so $\lim_n \text{FP}(g_n) > \Psi(\lim_n \text{FP}(g_n), g_m)$. As m goes to ∞ , we have $\lim_n \text{FP}(g_n) \geq \Psi(\lim_n \text{FP}(g_n), g)$, yielding a contradiction. Hence, $\text{FP}(\cdot)$ is left-continuous.

In the same fashion, let $g_n \downarrow g$, such that $g_n \geq 0$ for all n . Therefore, $\text{FP}(g_n)$ has a limit $\lim_n \text{FP}(g_n) \geq \text{FP}(g)$. Suppose that $\lim_n \text{FP}(g_n) > \text{FP}(g)$. Hence, $\lim_n \text{FP}(g_n) > \Psi(\lim_n \text{FP}(g_n), g)$. However, we note that $\lim_n \text{FP}(g_n) \leq \text{FP}(g_m)$ for all $m \geq 0$, which implies that $\lim_n \text{FP}(g_n) < \Psi(\lim_n \text{FP}(g_n), g_m)$. Letting m go to ∞ , $\lim_n \text{FP}(g_n) \leq \Psi(\lim_n \text{FP}(g_n), g)$, yielding a contradiction. Therefore, $\text{FP}(\cdot)$ is continuous on $[0, \infty)$.

References

- [1] AKTARAN, T. AND AYHAN, H. (2007). Sensitivity of optimal prices to system parameters in a service facility. Working paper.
- [2] ATA, B. AND SHNEORSON, S. (2006). Dynamic control of an M/M/1 service system with adjustable arrival and service rates. *Manag. Sci.* **52**, 1778–1791.
- [3] ÇİL, E. B., KARAESMEN, F. AND ÖRMECI, E. L. (2007). Sensitivity analysis on a dynamic pricing problem of an M/M/c queuing system. Submitted.
- [4] GEORGE, J. M. AND HARRISON, J. M. (2001). Dynamic control of a queue with adjustable service rate. *Operat. Res.* **49**, 720–731.
- [5] HASSIN, R. (1986). Consumer information in markets with random product quality: the case of queues and balking. *Econometrica* **54**, 1185–1195.
- [6] KNUDSEN, N. C. (1972). Individual and social optimization in a multiserver queue with a general cost-benefit structure. *Econometrica* **40**, 515–528.
- [7] LARSEN, C. (1998). Investigating sensitivity and the impact of information on pricing decisions in an M/M/1/∞ queuing model. *Internat. J. Production Econom.* **56–57**, 365–377.
- [8] LASSERRE, J.-B. AND HERNÁNDEZ-LERMA, O. (1996). *Discrete-Time Markov Control Processes. Basic Optimality Criteria*. Springer, New York.
- [9] LOW, D. W. (1974). Optimal dynamic pricing policies for an M/M/s queue. *Operat. Res.* **22**, 545–561.
- [10] LOW, D. W. (1974). Optimal pricing for an unbounded queue. *IBM J. Res. Development* **18**, 290–302.

- [11] MAOUI, I., AYHAN, H. AND FOLEY, R. D. (2007). Optimal static pricing for a service facility with holding costs. To appear in *Europ. J. Operat. Res.*
- [12] MENDELSON, H. AND WHANG, S. (1990). Optimal incentive-compatible priority pricing for the M/M/1 queue. *Operat. Res.* **38**, 870–883.
- [13] NAOR, P. (1969). The regulation of queue size by levying tolls. *Econometrica* **37**, 15–24.
- [14] PASCHALIDIS, I. C. AND TSITSIKLIS, J. N. (2000). Congestion-dependent pricing of network services. *IEEE/ACM Trans. Networking* **8**, 171–184.
- [15] PUTERMAN, M. (1994). *Markov Decision Processes*. John Wiley, New York.
- [16] STIDHAM, S. (1985). Optimal control of admission to a queueing system. *IEEE Trans. Automatic Control* **30**, 705–713.
- [17] WEBER, R. R. AND STIDHAM, S. (1987). Optimal control of service rates in network queues. *Adv. Appl. Prob.* **19**, 202–218.
- [18] YECHIALI, U. (1971). On optimal balking rules and toll charges in the G/M/1/s queue. *Operat. Res.* **19**, 349–370.
- [19] ZIYA, S., AYHAN, H. AND FOLEY, R. D. (2004). Relationships among three assumptions in revenue management. *Operat. Res.* **52**, 804–809.
- [20] ZIYA, S., AYHAN, H. AND FOLEY, R. D. (2006). Optimal prices for finite capacity queueing systems. *Operat. Res. Lett.* **34**, 214–218.