Birkbeck Working Papers in Economics & Finance

ISSN 1745-8587

Birkbeck
UNIVERSITY OF LONDON

School of Economics, Mathematics and Statistics

BWPEF 0805

Approximate Expected Delay Costs for Call and Contact Centre models under Light Traffic Regimes

S. Bhulai
VU University Amsterdam

A. C. Brooms
Birkbeck, University of London

July 2008

Approximate Expected Delay Costs for Call and Contact Centre models under Light Traffic Regimes

S.Bhulai*& A.C.Brooms†

July 10, 2008

Abstract

This paper studies the form of certain expected delay costs as a function of the arrival rate for customers who pass through a service facility that allows for reneging and retrials. We show that, under certain light traffic conditions, these costs are continuously increasing and convex functions of the arrival rate (within a finite interval). This result is first explored for the processor sharing system, in which a penalty cost is incurred for reneging from the service facility for good without ever receiving service, and then we consider a system with a more general structure governing the output processes and costs incurred per unit time, but without the penalty cost. A suggested application for these results, in which game theoretic considerations are utilized for gauging customer behaviour within a decentralized context, is briefly discussed.

KEYWORDS: CALL & CONTACT CENTRES; DELAY COST; IMPATIENCE; RETRIALS; LIGHT TRAFFIC; MONOTONICITY; JOINING RULES; NASH EQUILIBRIUM;

1 Introduction

We consider a service system consisting of a main facility Q_1, and a delay node Q_2. Customers who initially join Q_1 do so in anticipation of obtaining a certain service. Due to impatience, some customers will renege from this facility before satisfactory completion of service, either to leave the system without ever returning, at a certain cost, or in order to join the facility at a later point in time. Customers who have reneged from the main facility with the aim of trying again later can be viewed as having joined Q_2 with a single output stream leading directly to the tail of Q_1.

Facility Q_1 could represent a contact centre in which information is either solicited from or uploaded to a resource sharing facility via an internet portal according to a *processor sharing* discipline. Q_1 could also represent a call centre, in which customers

*VU University Amsterdam, Faculty of Sciences, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands.

†Corresponding Author: School of Economics, Mathematics, & Statistics, Birkbeck College, Malet Street, Bloomsbury, London WC1E 7HX, U.K.

queue up on a *first come first served* basis, in order to speak to one of the available operators.

The *expected delay cost* incurred by individual customers who join the service system is of interest in gauging the quality of service that is being experienced. The delay cost is representative of the total time spent in the system and total costs incurred as a result of a customer engaging with the system, irrespective of whether a full and satisfactory service has been obtained.

We motivate the structure of the cost function in the following way. Assume that the system is in stochastic equilibrium. Let λ be the potential arrival rate, and λs the effective arrival rate, that comes externally into the system, where $s \in (0, 1]$. Define $p_s(x, y)$ to be the probability that there are x customers in Q_1 and y customers in Q_2 just after an actual arrival epoch, and define $D_s^*(x, y)$ to be the expected delay cost of a customer who enters the system when there are x customers in Q_1 and y customers in Q_2 just after its arrival. Then the (unconditional) expected delay cost is given by

$$D_s^* = \sum_x \sum_y D_s^*(x, y) p_s(x, y).$$

It is desirable to understand the behaviour of $D_s^*(\hat{\lambda})$, as a function of the effective *external* arrival rate $\hat{\lambda} \in [0, \lambda]$, $\hat{\lambda} = \lambda s$, where λ is taken so that the system is stable whenever the state space is unbounded.

It seems intuitive that, under the right conditions, D_s^* should be an increasing function of s. However, in spite of the above decomposition being quite a natural one, it is not all that clear what behaviour would be jointly required of the $\{D_s^*(x, y)\}$ and the $\{p_s(x, y)\}$ in order to confirm this intuition. Consideration of an explicit solution for the $\{D_s^*(x, y)\}$ and the $\{p_s(x, y)\}$, even for a bounded state space, appears to be problematic: it is not clear how either the difference equations satisfied by the former, or the global balance equations satisfied by the latter, could be explicitly solved to yield closed-form solutions. Thus, a goal of this paper is to present an analytically and computationally tractable approximation to D_s^*, namely D_s, that can be shown to work well under, essentially, (i) light traffic conditions (see Daley & Rolski (1991) for discussion), and (ii) a sufficiently large, but bounded, state space (albeit the bound on the state of Q_2 should be taken to be as large as possible in view of it being a retrial queue). A main result of this paper is that, under the right conditions, D_s is strictly continuously increasing and convex in s. Another result, which is related to that of D_s, is that W_s, which measures the workload experienced by the system, is also continuously increasing and convex in s. Properties of the two quantities are obtained by working with a *relative value iteration* scheme that allows us to condition on the states of a suitably defined embedded Markov chain.

Because of its relevance to issues that arise within pricing problems for service systems, convexity of certain types of delay cost functions in queueing systems has generated some interest in the literature (see Dewan & Mendelson (1990), and Stidham (1992) for example). Although delay convexity may appear, superficially, to be intuitive in many instances, care needs to be taken that this is indeed the case, as the counterexample produced by Fridgeirsdottir & Chiu (2005) demonstrates.

The paper is organized as follows. In the next section, we present some of the notation and mathematical preliminaries that will be needed throughout the remainder of the paper. In Section 3, we derive results for a model in which Q_1 is taken to be a processor sharing system (although the actual discipline is not crucial in view of Little's Law). Section 4 considers a more general output and abandonment structure when the abondonment cost is set equal to zero. In Section 5, we explain how the expected delay costs may be computed, and present a number of these calculations for a variety of parameter settings. Although not the main focus of the paper, Section 6 discusses a potential application in which likely take-up of the service facility is assessed using game theoretic considerations. Conclusions and discussion are provided in Section 7.

2 Mathematical Preliminaries

We take $\mathbb{Z}^+ = \{1, 2, \ldots\}$, and $\mathbb{N} = \mathbb{Z}^+ \cup \{0\}$ throughout the remainder of this paper. **It will also be assumed that the system is in stochastic equilibrium**.

Define $\{(X(t), Y(t)) : t \geq 0\}$ to be a left-continuous process such that $X(t)$ and $Y(t)$ each represents the number of customers in Q_1 and Q_2 at time t, respectively.

Customers arrive at the service facility according to a Poisson process of rate λ. A customer is said to be using the joining rule $[s]$, for $s \in [0, 1]$, if upon arrival at the facility, it joins Q_1 w.p. s, and balks w.p. $1 - s$, independently of all other customers.

If each and every customer uses the joining rule $[s]$, then the customers are said to collectively adhere to the joining policy $[s]^\infty$. Thus, under $[s]^\infty$, the effective *external* arrival rate into Q_1, for a non-finite buffer, is given by λs.

Consider an arbitrary customer who, upon arrival to the service facility, joins Q_1 when there are $x - 1$ customers already present there, and y customers retrialling in Q_2, with all future customers using the policy $[s]^\infty$. Define $m_s(x, y)$ to be the event that the customer reneges and exits the system without ever returning, and $l_s(x, y)$ to be the total time that has elapsed between first entry to and last exit (departure or reneging but no retrial) from Q_1. Then define $d_s(x, y)$, the delay cost, conditioned on state (x, y), to be

$$d_s(x, y) = l_s(x, y) + \mathbb{1}\{m_s(x, y)\}R$$

and the expected delay cost $D_s^*(x, y) := E[d_s(x, y)]$.

In a similar way, define d_s^* and D_s^* to be the (unconditional) delay cost and the (unconditional) expected delay cost for an arbitrary customer.

3 The processor sharing model

Consider a service facility Q_1 in which customers arrive according to a Poisson process with rate λ. Customers at facility Q_1 are served according to a processor sharing

discipline, in which potential inter-departure times are i.i.d. exponential with mean $1/\mu$. However, customers are impatient in the sense that they renege after an exponentially distributed period with mean $1/\beta$ independently of all other customers in Q_1, whilst waiting to complete service. A customer that reneges will either abandon the system with probability $1-\psi$ or go to a retrial queue Q_2 with probability ψ. Each customer in the retrial queue independently attempts to rejoin the tail of facility Q_1 after an exponentially distributed period with mean $1/\gamma$. The retrial node will be modelled by a $\cdot/M/N$ queue. In our initial model framework, we set $N=\infty$, but eventually truncate this to obtain a finite state space: realism of the model need not be overly compromised, provided that N is taken to be large, relative to the other system parameters. The inter-arrival times, potential inter-departure times, times spent in Q_1 before reneging, times spent in Q_2 before retrial, and random variables governing exits immediately after reneging, constitute mutually independent sequences.

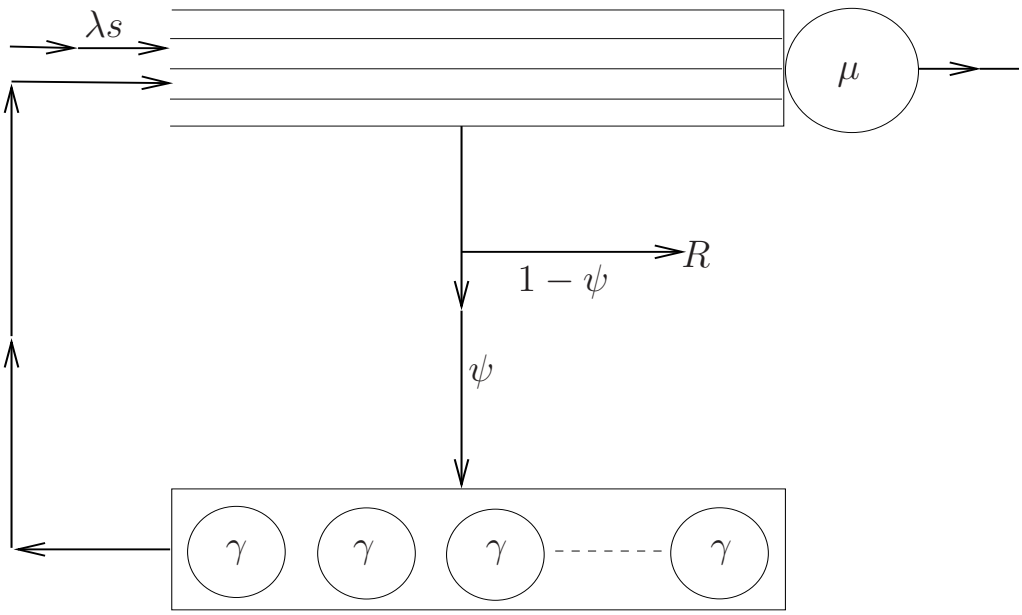

Figure 1: M/M/1/B+N processor sharing system with retrials

A customer is said to be using the joining rule $[s]$, for $s \in [0,1]$, if upon arrival at the system, the customer joins facility Q_1 with probability s, and balks with probability $1-s$, independently of all other customers. To study the expected sojourn time of an arbitrary customer as a function of s, we model the system in a Markov decision theoretic framework. For this purpose, let $\mathcal{X} = \mathbb{N} \times \mathbb{N}$ denote the state space of the system, with $(x,y) \in \mathcal{X}$ denoting that there are x customers in the facility Q_1 and y customers in the delay node Q_2. We denote the transition rate

of going from (x, y) to (x', y') by $q\big((x, y), (x', y')\big)$. Then for $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ we have

$$q\big((x,y),(x',y')\big) = \begin{cases} \lambda s & \text{if } (x',y') = (x+1,y) \\ \mu \mathbf{1}\{x>0\} + \beta x(1-\psi) & \text{if } (x',y') = (x-1,y) \\ \beta x \psi & \text{if } (x',y') = (x-1,y+1) \\ \gamma y & \text{if } (x',y') = (x+1,y-1) \\ 0 & \text{o.w.} \end{cases}$$

corresponding to arrivals, service departures and reneging customers that abandon, reneging customers that go to the retrial queue, and customers leaving the retrial queue, respectively.

For a system in equilibrium with customers adhering to the joining policy $[s]^\infty$, we define the *system workload* to be

$$W_s = \Phi L_s^{(1)} + \Psi L_s^{(2)} + \kappa$$

where $L_s^{(1)}$ and $L_s^{(2)}$ are the expected numbers of customers in Q_1 and Q_2, respectively, and where Φ, Ψ, and κ, are constants. Thus W_s can be seen to be a measure of the workload which is being borne by the facility Q_1, and the potential workload to the facility as a result of retrials from Q_2, per unit time. Without loss of generality, we consider the case in which $\Phi = 1+\beta(1-\psi)R$, $\Psi = 1$, and $\kappa = 0$. The inclusion of the "cost-like" term $\beta(1-\psi)R$ as a factor for $L_s^{(1)}$ can be seen as a way of accounting for the future potential consequences to the system of not meeting the needs of customers in a timely manner. When the system state space is unbounded, and appropriate stability conditions are satisfied, then $D_s^* = W_s(\lambda s)^{-1}$. To see this, we note that by Little's Law, the expected time spent in the system by a customer between first entry to, and final exit from, Q_1, is given by $[L_s^{(1)} + L_s^{(2)}](\lambda s)^{-1}$. Also, the external arrival rate into the system should balance with the rate at which customers exit the system for good. However, at equilibrium, the rate at which customers renege from Q_1 and then exit immediately is given by $\beta L_s^{(1)}(1-\psi)$. Hence the proportion of customers that leave the system by reneging and exiting immediately is given by $\beta L_s^{(1)}(1-\psi)(\lambda s)^{-1}$, which corresponds to the probability that an arbitrary customer exits the system by reneging rather than due to an actual service. Thus we are motivated to formally define our *approximating quantity* D_s to be

$$D_s = \frac{L_s^{(1)} + L_s^{(2)} + \beta L_s^{(1)}(1-\psi)R}{\lambda s}.$$

Conditions under which D_s serves to be a good approximation to D_s^* will be the subject matter of much of the rest of this paper.

Due to the computational complexities associated with the calculation of W_s (and hence D_s) hinted at earlier, we adopt a dynamic programming approach to obtain this quantity. Let $\tau_{x,y}$ be the expected time until the next event epoch given that the current state is (x, y). In general, $\tau_{x,y}$ is dependent on the state (x, y). In order to derive a state-independent $\tau_{x,y} = \tau$, we need to uniformize the

system. To this end, for B and N finite, truncate the state space to obtain $\mathcal{X}^{(B,N)} = \{0,\ldots,B\} \times \{0,\ldots,N\}$, and, for $(x,y) \in \mathcal{X}^{(B,N)}$, define new transition rates by

$$\widetilde{q}^{(B,N))}\big((x,y),(x',y')\big)$$

$$= \begin{cases} q\big((x,y),(x',y')\big) & (x,y) \neq (x',y') \in \mathcal{X}^{(B,N)} \\ 0 & (x',y') \notin \mathcal{X}^{(B,N)} \\ q\big((x,y),(x,y)\big) + \sum_{(m,n)\notin \mathcal{X}^{(B,N)}} q\big((x,y),(m,n)\big) & (x',y') = (x,y) \end{cases}.$$

Now, define the uniformization factor (Lippman (1975)) $\eta = \lambda + \mu + B\beta + N\gamma$. Uniformization adds dummy transitions from a state to itself, such that the rate out of each state is equal to η, and consequently $\tau = 1/\eta$.

Let $\widetilde{\pi}(x,y)$ be the probability that the state of the continuous time Markov process (CTMP) $\{(X(t),Y(t)) : t \geq 0\}$, at equilibrium, is (x,y). Then

$$W_s = L_s^{(1)} + L_s^{(2)} + \beta L_s^{(1)}(1-\psi)R = \sum_x \sum_y [x + y + \beta x(1-\psi)R]\widetilde{\pi}(x,y).$$

Also, define $\{\pi(x,y)\}$ to be the equilibrium distribution for the discrete time Markov chain (DTMC) embedded at the epochs just after potential transitions in the uniformized process. Now since $\{\widetilde{\pi}(x,y)\}$ is ergodic (that is to say, $\widetilde{\pi}(x,y)$ represents the long-run proportion of time that the CTMP spends in state (x,y)), and that for the uniformized process the expected time between potential transitions is i.i.d. exponential with finite mean τ, then it follows that

$$\widetilde{\pi}(x,y) = \pi(x,y) \text{ for all } (x,y) \in \mathcal{X}^{(B,N)}.$$

Hence

$$\frac{W_s}{\eta} = \sum_x \sum_y \frac{[x + y + \beta_0\, x(1-\psi)R_0]}{\eta}\pi(x,y)$$

where, for clarity of expression, the β and R values appearing in the expression for W_s, and in the numerator of the RHS of W_s/η, are denoted by β_0 and R_0 respectively; this is in contradistinction to the β appearing in the expression for η, whose value will alter as a result of re-scaling (see later).

By the argument presented on line 16 of p.102 in Ross (1983), the expression for W_s/η may be viewed as the long-term expected average cost of the discrete time MDP, (where the $R(x,y)$ in the notation of Ross (1983) corresponds to $c(x,y)/\eta = [x + y + \beta_0\, x(1-\psi)R_0]/\eta$ in ours, with the control action a being superfluous for our purposes), whose dynamic programming operator, T, for functions f_s defined on $\mathcal{X}^{(B,N)}$, given the value of $s \in [0,1]$, is then defined by

$$Tf_s(x,y) = \frac{c(x,y)}{\eta} + \sum_{(x',y')\in\mathcal{X}^{(B,N)}} \frac{\widetilde{q}^{(B,N)}\big((x,y),(x',y')\big)}{\eta} f_s(x',y').$$

The long-term expected average cost g_s is a solution of the optimality equation (in vector notation) $g_s + V_s = TV_s$. Another way of obtaining them is through value iteration, by recursively defining the relative value function $V_s^{n+1} = TV_s^n$, for arbitrary V_s^0. For $n \to \infty$, the vector V_s^n converges to the vector V_s that solves $g_s + V_s = TV_s$, from which g_s can then be obtained. In the sequel, we will, without loss of generality, assume that $\eta = 1$; this can also be obtained by an appropriate re-scaling. This simplifies the relative value function for $0 < x < B$ and $0 < y < N$ from

$$\begin{aligned} V_s^{n+1}(x,y) =& \lambda s/\eta\, V_s^n(x+1,y) + \mu/\eta\, V_s^n(x-1,y) \\ &+ \beta x(1-\psi)/\eta\, V_s^n(x-1,y) + \beta x\psi/\eta\, V_s^n(x-1,y+1) \\ &+ y\gamma/\eta\, V_s^n(x+1,y-1) + (\eta - \lambda s - \mu - \beta x - y\gamma)/\eta\, V_s^n(x,y) \\ &+ [(x+y) + \beta_0\, x(1-\psi)R_0]/\eta \end{aligned}$$

to the following expression in which the rates can be seen as proper transition probabilities;

$$\begin{aligned} V_s^{n+1}(x,y) =& \lambda s\, V_s^n(x+1,y) + \mu\, V_s^n(x-1,y) + \beta x(1-\psi)\,[V_s^n(x-1,y) + R] \\ &+ \beta x\psi\, V_s^n(x-1,y+1) + y\gamma\, V_s^n(x+1,y-1) \\ &+ (1 - \lambda s - \mu - \beta x - y\gamma)\, V_s^n(x,y) + (x+y) \end{aligned}$$

where $R = R_0 \times \frac{\beta_0}{\beta}$.

Before turning our attention to g_s, we first obtain some structural properties of V_s. This is done by induction on n for V_s^n, and afterwards taking the limit $V_s = \lim_{n\to\infty} V_s^n$.

Theorem 3.1. *The relative value function $V_s(x,y)$ is an increasing function in the variables x, y on $\mathcal{X}^{(B,N)}$.*

Proof. For $V_s^0(x,y) \equiv 0$, then clearly this function is increasing in both x and y. Therefore, assume that V_s^n is increasing in x and y. We first show that V_s^{n+1} is increasing in x. To this end, consider $0 < x < B-1$ and $0 < y < N$. Then,

$$\begin{aligned} V_s^{n+1}(x+1,y) - V_s^{n+1}(x,y) =& [(x+1) - x] + \lambda s\,[V_s^n(x+2,y) - V_s^n(x+1,y)] \\ &+ \mu\,[V_s^n(x,y) - V_s^n(x-1,y)] + y\gamma\,[V_s^n(x+2,y-1) - V_s^n(x+1,y-1)] \\ &+ \beta(x+1)(1-\psi)\,[V_s^n(x,y) + R] - \beta x(1-\psi)\,[V_s^n(x-1,y) + R] \\ &+ \beta(x+1)\psi\, V_s^n(x,y+1) - \beta x\psi\, V_s^n(x-1,y+1) \\ &+ (1 - \lambda s - \mu - \beta(x+1) - y\gamma)\, V_s^n(x+1,y) - (1 - \lambda s - \mu - \beta x - y\gamma)\, V_s^n(x,y) \\ \geq& \beta x(1-\psi)\,[V_s^n(x,y) - V_s^n(x-1,y)] + \beta(1-\psi)\,[V_s^n(x,y) + R] \\ &+ \beta x\psi\,[V_s^n(x,y+1) - V_s^n(x-1,y+1)] + \beta\psi\, V_s^n(x,y+1) \\ &+ (1 - \lambda s - \mu - \beta(x+1) - y\gamma)\,[V_s^n(x+1,y) - V_s^n(x,y)] - \beta\, V_s^n(x,y) \\ \geq& \beta(1-\psi)\,[V_s^n(x,y) - V_s^n(x,y) + R] + \beta\psi\,[V_s^n(x,y+1) - V_s^n(x,y)] \\ \geq& 0. \end{aligned}$$

The first two inequalities follow from applying the induction hypothesis by using increasingness in x. The first term in the third inequality is trivially non-negative; however, the second term in that inequality follows from increasingness in y. One can easily check that in a similar way the result also holds for the boundaries corresponding to $x = 0$, $x = B - 1$, $y = 0$, and $y = N$.

We now continue the proof by showing that V_s^{n+1} is increasing in y. To this end, consider $0 < x < B$ and $0 < y < N - 1$. Then,

$$\begin{aligned}
V_s^{n+1}(x, y+1) - V_s^{n+1}(x, y) &= [(y+1) - y] + \lambda s[V_s^n(x+1, y+1) - V_s^n(x+1, y)] \\
&\quad + \mu\,[V_s^n(x-1, y+1) - V_s^n(x-1, y)] \\
&\quad + \beta x(1-\psi)\,[V_s^n(x-1, y+1) + R - V_s^n(x-1, y) - R] \\
&\quad + \beta x\psi\,[V_s^n(x-1, y+2) - V_s^n(x-1, y+1)] \\
&\quad + (y+1)\gamma\, V_s^n(x+1, y) - y\gamma\, V_s^n(x+1, y-1) \\
&\quad + (1 - \lambda s - \mu - \beta x - (y+1)\gamma)\, V_s^n(x, y+1) - (1 - \lambda s - \mu - \beta x - y\gamma)\, V_s^n(x, y) \\
&\geq y\gamma\,[V_s^n(x+1, y) - V_s^n(x+1, y-1)] + \gamma\, V_s^n(x+1, y) \\
&\quad + (1 - \lambda s - \mu - \beta x - (y+1)\gamma)\,[V_s^n(x, y+1) - V_s^n(x, y)] - \gamma\, V_s^n(x, y) \\
&\geq \gamma\,[V_s^n(x+1, y) - V_s^n(x, y)] \\
&\geq 0.
\end{aligned}$$

The first two inequalities follow from applying the induction hypothesis by using increasingness in y. The last inequality follows from increasingness in x. One can easily check that in a similar way the result also holds for the boundaries corresponding to $x = 0$, $x = B$, $y = 0$, and $y = N - 1$. The proof is concluded by taking the limit as $n \to \infty$. □

Theorem 3.1 shows that h_s is increasing in both components x and y on $\mathcal{X}^{(B,N)}$. For the propagation of this result, we needed increasingness of both components simultaneously in the induction hypothesis. However, the next theorem shows that for increasingness in s we only need increasingness in x.

Theorem 3.2. *The relative value function $V_s(x, y)$ is an increasing function in s on $\mathcal{X}^{(B,N)}$.*

Proof. For $V_s^0(x, y) = 0$, clearly this function is increasing in s. Therefore, assume that V_s^n is increasing in s. We show that V_s^{n+1} is increasing in s. To this end,

consider $s' = s + \Delta \geq s$, $0 < x < B$, and $0 < y < N$. Then,

$$\begin{aligned}
V_{s'}^{n+1}(x,y) - V_s^{n+1}(x,y) &= \lambda s' \, V_{s'}^n(x+1,y) - \lambda s \, V_s^n(x+1,y) \\
&\quad + \mu \left[V_{s'}^n(x-1,y) - V_s^n(x-1,y)\right] + \beta x(1-\psi) \left[V_{s'}^n(x-1,y) - V_s^n(x-1,y)\right] \\
&\quad + \beta x \psi \left[V_{s'}^n(x-1,y+1) - V_s^n(x-1,y+1)\right] \\
&\quad + y\gamma \left[V_{s'}^n(x+1,y-1) - V_s^n(x+1,y-1)\right] \\
&\quad + (1 - \lambda s' - \mu - \beta x - y\gamma) \, V_{s'}^n(x,y) - (1 - \lambda s - \mu - \beta x - y\gamma) \, V_s^n(x,y) \\
&\geq \lambda s \left[V_{s'}^n(x+1,y) - V_s^n(x+1,y)\right] + \lambda \Delta V_{s'}^n(x+1,y) \\
&\quad + (1 - \lambda s - \mu - \beta x - y\gamma) \left[V_{s'}^n(x,y) - V_s^n(x,y)\right] - \lambda \Delta V_{s'}^n(x,y) \\
&\geq \lambda \Delta [V_{s'}^n(x+1,y) - V_{s'}^n(x,y)] \\
&\geq 0.
\end{aligned}$$

The first two inequalities follow from applying the induction hypothesis by using increasingness in s. The last inequality follows from increasingness in x (Theorem 3.1). One can easily check that in a similar way the result also holds for the boundaries corresponding to $x = 0$, $x = B$, $y = 0$, and $y = N$. The proof is concluded by taking the limit as $n \to \infty$. □

The combination of Theorem 3.1 and Theorem 3.2 show that $V_s(x,y)$ is increasing in all its components x, y, and s. In fact, since the costs $x + y$ are strictly increasing in x and y, we obtain that V_s is also strictly increasing in all its components. These are first-order properties of V_s.

Note that the results obtained so far, by construction, hold only on the state space $\mathcal{X}^{(B,N)}$. By taking B and N to be large in relation to the system parameters, it is hoped that a reasonable approximation can be obtained to the system defined on $\mathcal{X}$ with unbounded state space.

Corollary 3.3. *The relative value function $V_s(x,y)$ is a strictly increasing function in x, y, and s on $\mathcal{X}^{(B,N)}$.*

With this corollary, we are also able to derive structural properties of g_s, the expected system delay cost. Note that since the process satisfies the unichain condition, g_s does not depend on x or y. Therefore, we only derive monotonicity properties of g_s as a function of s. Recall that the relative value function $V_s(x,y)$ has the interpretation of the asymptotic difference in total costs that results from starting the process in state (x,y) instead of some reference state. Without loss of generality we take the reference state to be $(0,0)$.

The Poisson equations for state $(0,0)$ are then given by

$$g_s + \lambda s \, V_s(0,0) = \lambda s \, V_s(1,0).$$

Therefore g_s is given by

$$g_s = \lambda s \, V_s(1,0).$$

Hence, we can conclude that g_s is strictly increasing in s as well, since both λs and $V_s(1,0)$ (Corollary 3.3) are strictly increasing functions of s. Also, noting that $V_s^0(\cdot,\cdot) \equiv 0$ is (trivially) continuous in s, and assuming that $V_s^n(\cdot,\cdot)$ is continuous in s by assumption, it follows that $V_s^{n+1}(\cdot,\cdot)$ is continuous in s (since this can be expressed as the sum of continuous functions).

Hence, we have the following corollary.

Corollary 3.4. *The expected average cost g_s, and therefore also the system workload, W_s, is a strictly increasing and continuous function in s.*

Now we move on to second-order properties of V_s.

Theorem 3.5. *For all appropriate $(x,y) \in \mathcal{X}^{(B,N)}$, and λ, $\beta\psi$, and γ, sufficiently small, the relative value function $V_s(x,y)$ satisfies the following properties:*
Convex(x): $V_s(x+1,y) - 2V_s(x,y) + V_s(x-1,y) \geq 0$,
Convex(y): $V_s(x,y+1) - 2V_s(x,y) + V_s(x,y-1) \geq 0$,
Supermodular(x, y): $V_s(x+1,y+1) + V_s(x,y) - V_s(x+1,y) - V_s(x,y+1) \geq 0$.

Proof. For $V_s^0(x,y) = 0$, clearly this function satisfies Convex(x), Convex(y), and Supermodular(x, y). Therefore, assume that these properties hold for V_s^n. We first show that V_s^{n+1} satisfies Convex(x). To this end, consider $1 < x < B-1$ and $0 < y < N$. Then,

$$
\begin{aligned}
&V_s^{n+1}(x+1,y) - 2\,V_s^{n+1}(x,y) + V_s^{n+1}(x-1,y) \\
&\quad \geq \mu[V_s^n(x,y) - 2V_s^n(x-1,y) + V_s^n(x-2,y)] \\
&\qquad + \beta(1-\psi)\,[(x+1)\,V_s^n(x,y) - 2x\,V_s^n(x-1,y) + (x-1)\,V_s^n(x-2,y)] \\
&\qquad + \beta\psi\,[(x+1)\,V_s^n(x,y+1) - 2x\,V_s^n(x-1,y+1) + (x-1)\,V_s^n(x-2,y+1)] \\
&\qquad + (1-\lambda s-\mu-\beta(x+1)-y\gamma)V_s^n(x+1,y) - 2(1-\lambda s-\mu-\beta x-y\gamma)V_s^n(x,y) \\
&\qquad + (1-\lambda s-\mu-\beta(x-1)-y\gamma)V_s^n(x-1,y) \\
&\quad \geq \mu[V_s^n(x,y) - 2V_s^n(x-1,y) + V_s^n(x-2,y)] \\
&\qquad + 2\beta(1-\psi)\,[V_s^n(x,y) - V_s^n(x-1,y) - V_s^n(x,y) + V_s^n(x-1,y)] \\
&\qquad + 2\beta\psi\,[V_s^n(x,y+1) - V_s^n(x-1,y+1) - V_s^n(x,y) + V_s^n(x-1,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no x in them. The second inequality follows by rearranging terms such that the Convex($x-1$) part of the induction hypothesis can be used again. The final inequality follows from Supermodular($x-1$, y) for the last line. The result also holds trivially at the boundaries corresponding to $x = 1$, $\{(x,y) : 1 \leq x \leq B-2; y = 0\}$, $\{(x,y) : 1 \leq x \leq B-2; y = N\}$. At the boundary $x = B-1$, the result holds provided that $\lambda = 0$ and $\gamma = 0$: however, if, for example, the term with factor μ had a strictly positive lower bound (uniformly in n), then by the continuity of V_s^n in λ, and by the boundedness of V_s^n (uniformly in n), the admissible λ and γ values could

be extended to lie within contiguous (finite) sets beyond 0.

We now proceed to prove Convex(y). Consider $0 < x < B$ and $1 < y < N-1$. Then

$$
\begin{aligned}
&V_s^{n+1}(x,y+1) - 2\,V_s^{n+1}(x,y) + V_s^{n+1}(x,y-1) \\
&\quad \geq \mu[V_s^n(x-1,y+1) - 2V_s^n(x-1,y) + V_s^n(x-1,y-1)] \\
&\qquad + \gamma\,[(y+1)\,V_s^n(x+1,y) - 2y\,V_s^n(x+1,y-1) + (y-1)\,V_s^n(x+1,y-2)] \\
&\qquad + (1-\lambda s-\mu-\beta x-(y+1)\gamma)V_s^n(x,y+1) - 2(1-\lambda s-\mu-\beta x-y\gamma)V_s^n(x,y) \\
&\qquad + (1-\lambda s-\mu-\beta x-(y-1)\gamma)V_s^n(x,y-1) \\
&\quad \geq \mu[V_s^n(x-1,y+1) - 2V_s^n(x-1,y) + V_s^n(x-1,y-1)] \\
&\qquad + 2\gamma\,[V_s^n(x+1,y) - V_s^n(x+1,y-1) - V_s^n(x,y) + V_s^n(x,y-1)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no y in them. The second inequality follows by rearranging terms such that the Convex($y-1$) part of the induction hypothesis can be used again. The final inequality follows from Supermodular(x, $y-1$). The result also holds trivially at the boundaries $y = 1$, $\{(x,y) : x = 0;\ 2 \leq y \leq N-2\}$, and $\{(x,y) : x = B;\ 2 \leq y \leq N-2\}$; At the boundary $y = N-1$, the result holds for $\beta\psi = 0$: however, if, for example, the term with factor μ had a strictly positive lower bound (uniformly in n), then by the continuity of V_s^n in $\beta\psi$, the admissible $\beta\psi$ values could be extended to a contiguous (finite) set beyond 0.

We continue the proof by showing Supermodular(x, y). Consider $0 < x < B-1$ and

$0 < y < N-1$. Then

$$
\begin{aligned}
&V_s^{n+1}(x+1,y+1) + V_s^{n+1}(x,y) - V_s^{n+1}(x+1,y) - V_s^{n+1}(x,y+1) \\
&\quad \geq \mu[V_s^n(x,y+1) + V_s^n(x-1,y) - V_s^n(x,y) - V_s^n(x-1,y+1)] \\
&\qquad + \beta(1-\psi)\,[(x+1)\,V_s^n(x,y+1) + x\,V_s^n(x-1,y) - (x+1)\,V_s^n(x,y) \\
&\qquad - x\,V_s^n(x-1,y+1)] + \beta\psi\,[(x+1)\,V_s^n(x,y+2) + x\,V_s^n(x-1,y+1) \\
&\qquad - (x+1)\,V_s^n(x,y+1) - x\,V_s^n(x-1,y+2)] + \gamma\,[(y+1)\,V_s^n(x+2,y) \\
&\qquad + y\,V_s^n(x+1,y-1) - y\,V_s^n(x+2,y-1) - (y+1)\,V_s^n(x+1,y)] \\
&\qquad + (1-\lambda s-\mu-\beta(x+1)-(y+1)\gamma)\,V_s^n(x+1,y+1) \\
&\qquad + (1-\lambda s-\mu-\beta x-y\gamma)\,V_s^n(x,y) - (1-\lambda s-\mu-\beta(x+1)-y\gamma)\,V_s^n(x+1,y) \\
&\qquad - (1-\lambda s-\mu-\beta x-(y+1)\gamma)\,V_s^n(x,y+1) \\
&\quad \geq \mu[V_s^n(x,y+1) + V_s^n(x-1,y) - V_s^n(x,y) - V_s^n(x-1,y+1)] \\
&\qquad + \beta(1-\psi)\,[V_s^n(x,y+1) - V_s^n(x,y) + V_s^n(x,y) - V_s^n(x,y+1)] \\
&\qquad + \beta\psi\,[V_s^n(x,y+2) - V_s^n(x,y+1) + V_s^n(x,y) - V_s^n(x,y+1)] \\
&\qquad + \gamma\,[V_s^n(x+2,y) - V_s^n(x+1,y) + V_s^n(x,y) - V_s^n(x+1,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no x or y in them. The second inequality follows by rearranging terms such that the induction hypothesis can be used again. The final inequality follows from Convex$(y+1)$ and Convex$(x+1)$ for the last two lines. In a similar way the result also holds for the boundaries corresponding to $\{(x,y) : 0 \leq x \leq B-2;\ y = 0\}$ and $\{(x,y) : x = 0;\ 1 \leq y \leq N-2\}$. At the boundaries $x = B-1$ and $y = N-1$, the result holds for $\gamma = \beta\psi = 0$; however, if, for example, the term with factor μ had a strictly positive lower bound (uniformly in n), then by the continuity of V_s^n in both γ and $\beta\psi$, the admissible values for γ and $\beta\psi$ could be extended to contiguous (finite) sets beyond 0. Finally, the proof is concluded by taking the limit as $n \to \infty$. □

Theorem 3.5 provides the necessary ingredients for showing convexity in s. Note that, similar to the case of the first-order properties, the proof of this result only depends on the properties of the facility Q_1, i.e., convexity in x in this case. However, convexity in x depends on convexity in y for queue Q_2. For the convexity in s we also need supermodularity in s in combination with the other variables.

Theorem 3.6. *For $\Delta \geq 0$, all appropriate $(x,y) \in \mathcal{X}^{(B,N)}$, and λ, $\beta\psi$, and γ, sufficiently small, the relative value function $V_s(x,y)$ satisfies the following properties:*
Convex(s): $V_{s+\Delta}(x,y) - 2V_s(x,y) + V_{s-\Delta}(x,y) \geq 0$,
Supermodular(s, x): $V_{s+\Delta}(x+1,y) + V_s(x,y) - V_{s+\Delta}(x,y) - V_s(x+1,y) \geq 0$,
Supermodular(s, y): $V_{s+\Delta}(x,y+1) + V_s(x,y) - V_{s+\Delta}(x,y) - V_s(x,y+1) \geq 0$.

Proof. For $V_s^0(x,y) \equiv 0$, clearly this function satisfies Convex(s), Supermodular(s, x), and Supermodular(s, y). Therefore, assume that these properties hold for V_s^n.

We first show that V_s^{n+1} satisfies Convex(s). To this end, consider $0 < x < B$ and $0 < y < N$. Then,

$$
\begin{aligned}
&V_{s+\Delta}^{n+1}(x,y) - 2\,V_s^{n+1}(x,y) + V_{s-\Delta}^{n+1}(x,y) \\
&\quad \geq \lambda(s+\Delta)\,V_{s+\Delta}^{n}(x+1,y) - 2\lambda s\,V_s^{n}(x+1,y) + \lambda(s-\Delta)\,V_{s-\Delta}^{n}(x+1,y) \\
&\qquad + (1-\lambda(s+\Delta)-\mu-\beta x-y\gamma)\,V_{s+\Delta}^{n}(x,y) - 2(1-\lambda s-\mu-\beta x-y\gamma)\,V_s^{n}(x,y) \\
&\qquad + (1-\lambda(s-\Delta)-\mu-\beta x-y\gamma)\,V_{s-\Delta}^{n}(x,y) \\
&\quad \geq 2\lambda\Delta\,[V_{s+\Delta}^{n}(x+1,y) - V_s^{n}(x+1,y) - V_{s+\Delta}^{n}(x,y) + V_s^{n}(x,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no s. The second inequality follows by rearranging terms such that the induction hypothesis can be used again on terms with factor $\lambda(s-\Delta)$ and $1-\lambda(s-\Delta)-\mu-\beta x-y\gamma$. The final inequality follows from Supermodular(s, x). The result also holds at the boundaries corresponding to $x=0$, $x=B$, $y=0$, and $y=N$.

We now proceed to prove Supermodular(s, x). To this end, consider $0 < x < B-1$ and $0 < y < N$.

$$
\begin{aligned}
&V_{s+\Delta}^{n+1}(x+1,y) + V_s^{n+1}(x,y) - V_{s+\Delta}^{n+1}(x,y) - V_s^{n+1}(x+1,y) \\
&\quad \geq \lambda\,[(s+\Delta)\,V_{s+\Delta}^{n}(x+2,y) + s\,V_s^{n}(x+1,y) - (s+\Delta)\,V_{s+\Delta}^{n}(x+1,y) \\
&\qquad - s\,V_s^{n}(x+2,y)] + \beta(1-\psi)\,[(x+1)\,V_{s+\Delta}^{n}(x,y) + x\,V_s^{n}(x-1,y) \\
&\qquad - x\,V_{s+\Delta}^{n}(x-1,y) - (x+1)\,V_s^{n}(x,y)] + \beta\psi\,[(x+1)\,V_{s+\Delta}^{n}(x,y+1) \\
&\qquad + x\,V_s^{n}(x-1,y+1) - x\,V_{s+\Delta}^{n}(x-1,y+1) - (x+1)\,V_s^{n}(x,y+1)] \\
&\qquad + (1-\lambda(s+\Delta)-\mu-\beta(x+1)-y\gamma)\,V_{s+\Delta}^{n}(x+1,y) \\
&\qquad + (1-\lambda s-\mu-\beta x-y\gamma)\,V_s^{n}(x,y) - (1-\lambda(s+\Delta)-\mu-\beta x-y\gamma)\,V_{s+\Delta}^{n}(x,y) \\
&\qquad - (1-\lambda s-\mu-\beta(x+1)-y\gamma)\,V_s^{n}(x+1,y) \\
&\quad \geq \lambda\Delta\,[V_{s+\Delta}^{n}(x+2,y) - V_{s+\Delta}^{n}(x+1,y) - V_{s+\Delta}^{n}(x+1,y) + V_{s+\Delta}^{n}(x,y)] \\
&\qquad + \beta(1-\psi)\,[V_{s+\Delta}^{n}(x,y) - V_s^{n}(x,y) + V_s^{n}(x,y) - V_{s+\Delta}^{n}(x,y)] \\
&\qquad + \beta\psi\,[V_{s+\Delta}^{n}(x,y+1) - V_s^{n}(x,y+1) + V_s^{n}(x,y) - V_{s+\Delta}^{n}(x,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no s or x. The second inequality follows by rearranging terms such that the induction hypothesis can be used on terms with factors λs, $\beta(1-\psi)x$, $\beta\psi x$, or $1-\lambda s-\mu-\beta(x+1)-y\gamma$. The final inequality follows from Convex(x+1) [Theorem 3.5] for the first line, and Supermodular(s, y) for the last line. The result also holds at boundaries corresponding to $x=0$, $y=0$, and $y=N$. At $x=B-1$, the result holds

provided that $\lambda = 0$: however, if, for example, the term with factor μ had a strictly positive lower bound (uniformly in n), then by the continuity of V_s^n in λ, then the admissible λ values could be extended to lie within a contiguous (finite) set beyond 0.

We continue the proof by showing Supermodular(s, y). Consider $0 < x < B$ and $0 < y < N-1$. Then

$$\begin{aligned}
&V_{s+\Delta}^{n+1}(x,y+1) + V_s^{n+1}(x,y) - V_{s+\Delta}^{n+1}(x,y) - V_s^{n+1}(x,y+1) \\
&\quad \geq \lambda\,[(s+\Delta)\,V_{s+\Delta}^n(x+1,y+1) + s\,V_s^n(x+1,y) - (s+\Delta)\,V_{s+\Delta}^n(x+1,y) \\
&\qquad - sV_s^n(x+1,y+1)] + \gamma\,[(y+1)\,V_{s+\Delta}^n(x+1,y) + y\,V_s^n(x+1,y-1) \\
&\qquad - y\,V_{s+\Delta}^n(x+1,y-1) - (y+1)\,V_s^n(x+1,y)] \\
&\qquad + (1-\lambda(s+\Delta)-\mu-\beta x-(y+1)\gamma)\,V_{s+\Delta}^n(x,y+1) \\
&\qquad + (1-\lambda s-\mu-\beta x-y\gamma)\,V_s^n(x,y) - (1-\lambda(s+\Delta)-\mu-\beta x-y\gamma)\,V_{s+\Delta}^n(x,y) \\
&\qquad - (1-\lambda s-\mu-\beta x-(y+1)\gamma)\,V_s^n(x,y+1) \\
&\quad \geq \lambda\Delta\,[V_{s+\Delta}^n(x+1,y+1) - V_{s+\Delta}^n(x+1,y) - V_{s+\Delta}^n(x,y+1) + V_{s+\Delta}^n(x,y)] \\
&\qquad + \gamma\,[V_{s+\Delta}^n(x+1,y) - V_s^n(x+1,y) + V_s^n(x,y) - V_{s+\Delta}^n(x,y)] \\
&\quad \geq 0.
\end{aligned}$$

The first inequality follows from the induction hypothesis on the terms having factors with no s or y in them. The second inequality follows by rearranging terms such that the induction hypothesis can be used on terms with factors λs, $y\gamma$, or $1-\lambda s-\mu-\beta x-(y+1)\gamma$. The final inequality follows from Supermodular(x, y) [Theorem 3.5] and Supermodular(s, x). The result also holds at the boundaries corresponding to $x = 0$, $x = B$, $y = 0$, and $y = N-1$ (where we note that for $x = 0$, the term with factor $\beta\psi x$ disappears; and for $x > 0$, we invoke Supermodular(s,x) on this term instead). The proof is concluded by taking the limit as $n \to \infty$. □

Theorem 3.7.
(a) For W_s convex, D_s is continuously increasing on $(0,1]$.
(b) For λ, $\beta\psi$, and γ, sufficiently small and positive:
(i) D_s is continuously increasing and convex on $(0,1]$*;*
(ii) W_s is convex.

Proof. For $s' = s+\Delta$, we need to show that $D_{s'} - D_s$ is positive, i.e.

$$\frac{W_{s'}}{\lambda s'} - \frac{W_s}{\lambda s} > 0 \text{ which is equivalent to } \frac{1}{\lambda}\frac{sW_{s'} - s'W_s}{s \times s'} > 0.$$

Re-arranging yields

$$sW_{s'} - s'W_s > 0 \text{ i.e. } sW_{s'} - sW_s - \Delta W_s > 0 \text{ or } W_{s'} - W_s > \frac{\Delta}{s}W_s.$$

However, the convexity of W_s implies that for real numbers $x_1 < x_2$ in the interval $[0,1]$, and $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_1 + \alpha_2 = 1$, we have $W_{\alpha_1 x_1 + \alpha_2 x_2} < \alpha_1 W_{x_1} + \alpha_2 W_{x_2}$. Since μ and γ are positive, then the busy periods in the system are almost surely finite, thus $W_0 \equiv 0$. Setting $x_1 = 0$, $x_2 = s'$, and $\alpha_2 = s/(s+\Delta)$, establishes part (a).
Since $D_s = W_s/\lambda s = V_s(1,0)$, then (b)(i) follows trivially. On the other hand, since $W_s = \lambda s D_s$, then for valid $\Delta > 0$

$$W_{s+\Delta} - 2W_s + W_{s-\Delta} = \lambda \left\{ s[D_{s+\Delta} - 2D_s + D_{s-\Delta}] + \Delta \left[D_{s+\Delta} - D_{s-\Delta} \right] \right\}.$$

Result (b)(ii) now follows immediately from (b)(i). □

The proof of part (a) appears to suggest that in order for the delay cost D_s to be increasing in s, then not only does the *system workload* W_s have to be increasing in s, but across the intervals $[0,s]$ and $[s,s']$, the "average rate of change" over the latter interval has to be greater than that of the former. This would be guaranteed by convexity of W_s.

4 The general model

Suppose that the costs incurred at each epoch are given by $c_s(x,y)$ when the system is in state (x,y) for given s. Moreover, we assume that the rates for customers leaving the system and for reneging to the orbit are given by general functions $f(x)$ and $h(x)$, respectively, with $f(0) = h(0) \equiv 0$. Then, for $0 < x < B$ and $0 < y < N$, the relative value function is given by

$$\begin{aligned} V_s^{n+1}(x,y) = & \lambda s\, V_s^n(x+1,y) + f(x)\, V_s^n(x-1,y) + h(x)\, V_s^n(x-1,y+1) \\ & + y\gamma\, V_s^n(x+1,y-1) + (1 - \lambda s - f(x) - h(x) - y\gamma)\, V_s^n(x,y) \\ & + c_s(x,y). \end{aligned}$$

In the results that follow, we will need to assume some or all of the following conditions:

Assumptions

(I)	$c_s(x,y)$ is strictly increasing in x and y, and increasing and continuous in s
(II)	$h(x)$ is increasing in x;
(III)	$c_s(x-1,y+1) - c_s(x,y) \geq 0$;
(IV)	$f(x)$ is increasing in x;
(V)	There exists a $K \geq 0$ such that $h(x) - h(x-1) \leq K$ for $x = 1,2,\ldots,B$;
(VI)	$h(0) \equiv 0$, and $h(x)$ sufficiently small for each $x = 1,2,\ldots,B$;
(VII)	$c_s(x,y)$ is Convex(x), Convex(y), and Supermodular(x,y);
(VIII)	$f(x)$ is Concave(x), i.e., $f(x+1) - 2f(x) + f(x-1) \leq 0$;
(IX)	$h(x)$ is Convex(x), i.e., $h(x+1) - 2h(x) + h(x-1) \geq 0$;
(X)	$c_s(x,y)$ is Convex(s), Supermodular(s,x), and Supermodular(s,y).

Theorem 4.1. *Suppose that (I) and (II) hold.*
Then the relative value function $V_s(x,y)$ is strictly increasing in the variables x, y, and s, and continuous in s.

Proof. We repeat the proof of Theorem 3.1 and Theorem 3.2 given in the previous section. Since we only compare the individual terms in the relative value function, we only need to compare the terms that are different in the more general model. Thus, setting $V_s^0(x,y) = 0$, then clearly this function is increasing in x, y, and s. Therefore, assume that V_s^n is increasing in x, y, and s. We first show that V_s^{n+1} is increasing in x. To this end, consider $0 < x < B-1$ and $0 < y < N$. Then,

$$\begin{aligned}
V_s^{n+1}(x+1,y) - V_s^{n+1}(x,y) &\geq c_s(x+1,y) - c_s(x,y) \\
&\quad + f(x+1)\, V_s^n(x,y) - f(x) V_s^n(x-1,y) \\
&\quad + h(x+1)\, V_s^n(x,y+1) - h(x)\, V_s^n(x-1,y+1) \\
&\quad + (1 - \lambda s - f(x+1) - h(x+1) - y\gamma)\, V_s^n(x+1,y) \\
&\quad - (1 - \lambda s - f(x) - h(x) - y\gamma)\, V_s^n(x,y) \\
&\geq f(x)\,[V_s^n(x,y) - V_s^n(x-1,y)] + [f(x+1) - f(x)]\, V_s^n(x,y) \\
&\quad + h(x)\,[V_s^n(x,y+1) - V_s^n(x-1,y+1)] + [h(x+1) - h(x)]\, V_s^n(x,y+1) \\
&\quad + (1 - \lambda s - f(x+1) - h(x+1) - y\gamma)\,[V_s^n(x+1,y) - V_s^n(x,y)] \\
&\quad + [f(x) - f(x+1)]\, V_s^n(x,y) + [h(x) - h(x+1)]\, V_s^n(x,y) \\
&\geq [h(x+1) - h(x)]\,[V_s^n(x,y+1) - V_s^n(x,y)] \\
&\geq 0.
\end{aligned}$$

The first inequality follows from applying the induction hypothesis by using increasingness in x, and the second by the increasingness of $c_s(x,y)$ in x and a rearrangement of terms. The third inequality follows by noting that the terms with factor $f(x) - f(x+1)$ cancel, and then by applying the induction hypothesis on terms without $h(x) - h(x+1)$ as a factor. The last inequality follows from the fact that $h(x)$ is increasing in x and $V_s^n(x,y)$ is increasing in y. The above result is also easily checked at the boundaries corresponding to $x = 0$, $x = B-1$, $y = 0$ and $y = N$. In a similar way (checking boundary cases separately), it can be shown that $V_s^{n+1}(x,y+1) - V_s^{n+1}(x,y) \geq c_s(x,y+1) - c_s(x,y) \geq 0$, for $0 \leq x \leq B$, $0 \leq y \leq N-1$; and $V_{s+\Delta}^{n+1}(x,y) - V_s^{n+1}(x,y) \geq c_{s+\Delta}(x,y) - c_s(x,y) \geq 0$, for $0 \leq x \leq B$, $0 \leq y \leq N$. The proof is concluded by taking the limit as $n \to \infty$. □

Corollary 4.2. *Under the conditions of Theorem 4.1, g_s is strictly increasing and continuous in s.* □

Theorem 4.3. *Suppose that assumptions (I)-(V) hold. Then for K sufficiently small and for appropriate $(x,y) \in \mathcal{X}^{(B,N)}$, the relative value function $V_s(x,y)$ satisfies $V_s^n(x-1,y+1) - V_s^n(x,y) \geq 0$.*

Proof. For $V_s^0(x,y)=0$, clearly this function satisfies $V_s^0(x-1,y+1)-V_s^0(x,y)\geq 0$. Therefore, assume that this property holds for V_s^n. Consider $1<x<B$ and $0<y<N-1$. Then,

$$\begin{aligned}
V_s^{n+1}(x-1,y+1)-V_s^{n+1}(x,y) &\geq c_s(x-1,y+1)-c_s(x,y)\\
&\quad + f(x-1)\,V_s^n(x-2,y+1)-f(x)\,V_s^n(x-1,y)\ +h(x-1)\,V_s^n(x-2,y+2)\\
&\quad - h(x)\,V_s^n(x-1,y+1)\ +(y+1)\gamma\,V_s^n(x,y)-y\gamma\,V_s^n(x+1,y-1)\\
&\quad + (1-\lambda s-f(x-1)-h(x-1)-(y+1)\gamma)\,V_s^n(x-1,y+1)\\
&\quad - (1-\lambda s-f(x)-h(x)-y\gamma)\,V_s^n(x,y)\\
&\geq -[f(x)-f(x-1)]\,V_s^n(x-1,y)+[f(x)-f(x-1)]\,V_s^n(x,y)\\
&\quad - [h(x)-h(x-1)]\,V_s^n(x-1,y+1)+[h(x)-h(x-1)]\,V_s^n(x,y)\\
&\quad + \gamma V_s^n(x,y)-\gamma V_s^n(x,y)\\
&\geq 0.
\end{aligned}$$

The first inequality follows from the induction hypothesis on the terms having factors with no x and y in them. The second inequality follows by the subtraction and addition of the term $[1-\lambda s-f(x-1)-h(x-1)-(y+1)\gamma]V_s^n(x,y)$, rearranging terms, and invoking the induction hypothesis. The final inequality is due to the fact that $f(x)$ is increasing in x and $V_s^n(x,y)$ is increasing in x [Theorem 4.1], and the condition on $h(\cdot)$ for K sufficiently small for the second line. Indeed, it may be possible to take $K>0$ provided that, for example, the term with factor μ had a strictly positive lower bound (uniformly in n). One can easily check that in a similar way the result also holds for the boundaries corresponding to $x=1$, $x=B$, $y=0$, and $y=N-1$. The proof is concluded by taking the limit as $n\to\infty$. □

The above result says that provided conditions (I)-(IV) hold, and that the function governing the rate at which customers renege but move into the retrial orbit is sufficiently 'flat', then a reduction in system costs can be brought about by taking customers out of Q_2 and placing them in Q_1 instead.

Theorem 4.4. *Suppose that assumptions (I)-(IX) hold. Then for K, λ, and γ sufficiently small, and for appropriate $(x,y)\in\mathcal{X}^{(B,N)}$, the relative value function $V_s(x,y)$ satisfies the following properties:*
Convex(x): $V_s(x+1,y)-2V_s(x,y)+V_s(x-1,y)\geq 0$,
Convex(y): $V_s(x,y+1)-2V_s(x,y)+V_s(x,y-1)\geq 0$,
Supermodular(x, y): $V_s(x+1,y+1)+V_s(x,y)-V_s(x+1,y)-V_s(x,y+1)\geq 0$.

Proof. For $V_s^0(x,y)=0$, clearly this function satisfies Convex(x), Convex(y), and Supermodular(x, y). Therefore, assume that these properties hold for V_s^n. We first show that $V_s^{n+1}(x,y)$ satisfies Convex(x). To this end, consider $1<x<B-1$ and $0<y<N$. Then,

$$\begin{aligned}
&V_s^{n+1}(x+1,y)-2\,V_s^{n+1}(x,y)+V_s^{n+1}(x-1,y)\\
&\quad\geq c_s(x+1,y)-2\,c_s(x,y)+c_s(x-1,y)
\end{aligned}$$

$$
\begin{aligned}
&+ \lambda s[V_s^n(x+2,y) - 2V_s^n(x+1,y) + V_s^n(x,y)] \\
&+ y\gamma[V_s^n(x+2,y-1) - 2V_s^n(x+1,y-1) + V_s^n(x,y-1)] \\
&+ f(x+1)\, V_s^n(x,y) - 2f(x)\, V_s^n(x-1,y) + f(x-1)\, V_s^n(x-2,y) \\
&+ h(x+1)\, V_s^n(x,y+1) - 2h(x)\, V_s^n(x-1,y+1) + h(x-1)\, V_s^n(x-2,y+1) \\
&+ (1-\lambda s - f(x+1) - h(x+1) - y\gamma)\, V_s^n(x+1,y) \\
&- 2\,(1-\lambda s - f(x) - h(x) - y\gamma)\, V_s^n(x,y) \\
&+ (1-\lambda s - f(x-1) - h(x-1) - y\gamma)\, V_s^n(x-1,y)
\end{aligned}
$$

$$
\begin{aligned}
\geq\ & [f(x+1) - f(x-1)]\, V_s^n(x,y) - 2\,[f(x) - f(x-1)]\, V_s^n(x-1,y) \\
&- 2\,[f(x+1) - f(x)]\, V_s^n(x,y) + [f(x+1) - f(x-1)]\, V_s^n(x-1,y)] \\
&+ [h(x+1) - h(x-1)]\, V_s^n(x,y+1) - 2\,[h(x) - h(x-1)]\, V_s^n(x-1,y+1) \\
&- 2\,[h(x+1) - h(x)]\, V_s^n(x,y) + [h(x+1) - h(x-1)]\, V_s^n(x-1,y)
\end{aligned}
$$

$$
\begin{aligned}
=\ & -[f(x+1) - 2f(x) + f(x-1)]\,[V_s^n(x,y) - V_s^n(x-1,y)] \\
&+ [h(x+1) - h(x-1)]\,[V_s^n(x,y+1) - V_s^n(x-1,y+1) - V_s^n(x,y) + V_s^n(x-1,y)] \\
&+ [h(x+1) - 2h(x) + h(x-1)]\,[V_s^n(x-1,y+1) - V_s^n(x,y)]
\end{aligned}
$$

$$
\geq 0.
$$

The first inequality follows from the induction hypothesis on the terms having factors with no x in them. The second inequality follows by rearranging terms and applying the induction hypothesis on terms of the form $f(x-1)V_s^n(z,y)$, $h(x-1)v(z,y+1)$, and $[1-\lambda s - f(x+1) - h(x+1) - y\gamma]V_s^n(z,y)$ with respect to the variable z. The final inequality follows from concavity of $f(x)$ and increasingness of $V_s^n(x,y)$ in x [Theorem 4.1], Supermodular($x-1$, y) and increasingness of $h(x)$ in x, and convexity of $h(x)$ with $V_s^n(x-1,y+1) \geq V_s^n(x,y)$ [Theorem 4.3] for the three lines, respectively. The above result also holds for the boundaries corresponding to $x=1$, $\{(x,y) : 2 \leq x \leq B-2,\ y=0\}$, and $\{(x,y) : 2 \leq x \leq B-2,\ y=N\}$. It also holds at the boundary $x = B-1$ provided $\lambda = \gamma = 0$. However, if, for example, the expression with factor $f(x)$ had a strictly positive lower bound (uniformly in n), then the range of validity for λ and γ could be extended.

Similarly, for $0 < x < B$ and $1 < y < N-1$, convexity in y leads to $V_s^{n+1}(x,y+1) - 2\,V_s^{n+1}(x,y) + V_s^{n+1}(x,y-1) \geq c_s(x,y+1) - 2\,c_s(x,y) + c_s(x,y-1)$, which follows by applying the induction hypothesis on terms of the form $(y-1)\gamma V_s^n(x+1,z)$ and $[1-\lambda s - f(x) - h(x) - (y+1)\gamma]V_s^n(x,z)$ with respect to the variable z, and by exploiting Supermodular(x,$y-1$). The result also holds true for the boundaries corresponding to $\{(x,y) : x=0,\ 2 \leq y \leq N-2\}$, $\{(x,y) : x=B,\ 2 \leq y \leq N-2\}$, and $y=1$. It also holds at the boundary $y = N-1$ provided that $h(x)$ is identically equal to zero; however, if, for example, the expression with factor $f(x)$ had a strictly positive lower bound (uniformly in n), then it may be possible to take $h(x)$ to be more general.

We now proceed to prove Supermodular(x, y). To this end, consider $0 < x < B-1$ and $0 < y < N-1$. Then

$$
\begin{aligned}
&V_s^{n+1}(x+1,y+1) + V_s^{n+1}(x,y) - V_s^{n+1}(x+1,y) - V_s^{n+1}(x,y+1) \\
&\quad \geq c_s(x+1,y+1) + c_s(x,y) - c_s(x+1,y) - c_s(x,y+1) \\
&\qquad + f(x+1)\, V_s^n(x,y+1) + f(x)\, V_s^n(x-1,y) - f(x+1)\, V_s^n(x,y) \\
&\qquad - f(x)\, V_s^n(x-1,y+1) + h(x+1)\, V_s^n(x,y+2) + h(x)\, V_s^n(x-1,y+1) \\
&\qquad - h(x+1)\, V_s^n(x,y+1) - h(x)\, V_s^n(x-1,y+2) \\
&\qquad + (y+1)\gamma V_s(x+2,y) + y\gamma V_s(x+1,y-1) \\
&\qquad - y\gamma V_s(x+2,y-1) - (y+1)\gamma V_s(x+1,y) \\
&\qquad + (1-\lambda s - f(x+1) - h(x+1) - (y+1)\gamma)\, V_s^n(x+1,y+1) \\
&\qquad + (1-\lambda s - f(x) - h(x) - y\gamma)\, V_s^n(x,y) \\
&\qquad - (1-\lambda s - f(x+1) - h(x+1) - y\gamma)\, V_s^n(x+1,y) \\
&\qquad - (1-\lambda s - f(x) - h(x) - (y+1)\gamma)\, V_s^n(x,y+1) \\
&\quad \geq [f(x+1) - f(x)]\, [V_s^n(x,y+1) - V_s^n(x,y) + V_s^n(x,y) - V_s^n(x,y+1)] \\
&\qquad + [h(x+1) - h(x)]\, [V_s^n(x,y+2) - V_s^n(x,y+1) + V_s^n(x,y) - V_s^n(x,y+1)] \\
&\qquad + \gamma\, [V_s(x+2,y) - V_s(x+1,y) + V_s(x,y) - V_s(x+1,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no x and y in them. The second inequality follows by rearranging terms such that Supermodular(x,y) from the induction hypothesis can be invoked on various expressions which have either $f(x)$, $h(x)$, $[1-\lambda s - f(x+1) - h(x+1) - (y+1)\gamma]$, or $y\gamma$, as a factor. The final inequality follows from the fact that $h(x)$ is increasing in x and $V_s^n(x,y)$ satisfies Convex($y+1$) for the penultimate line, and Convex($x+1$) for the final line. The above result may also be verified for the boundaries corresponding to $\{(x,y) : x = 0,\ 0 \leq y \leq N-2\}$, and $\{(x,y) : 1 \leq x \leq B-2,\ y = 0\}$. The result also holds at the boundaries $x = B-1$ and $y = N-1$ provided that $\gamma = 0$ and $K = 0$ (constraining $h(x) \equiv 0$), respectively. However, if another expression on the RHS of these inequalities can be determined to have a strictly positive lower bound (uniformly in n), then the range of γ and the generality of $h(\cdot)$, may be extended. □

Theorem 4.5. *Suppose assumptions $(I)-(X)$ hold. Then for K, λ, and γ sufficiently small, and for appropriate $(x,y) \in \mathcal{X}^{(B,N)}$, the relative value function $V_s(x,y)$ satisfies the following properties:*
Convex(s): $V_{s+\Delta}(x,y) - 2V_s(x,y) + V_{s-\Delta}(x,y) \geq 0$,
Supermodular(s, x): $V_{s+\Delta}(x+1,y) + V_s(x,y) - V_{s+\Delta}(x,y) - V_s(x+1,y) \geq 0$,
Supermodular(s, y): $V_{s+\Delta}(x,y+1) + V_s(x,y) - V_{s+\Delta}(x,y) - V_s(x,y+1) \geq 0$.

Proof. For $V_s^0(x,y) = 0$, clearly this function satisfies Convex(s), Supermodular(s, x), and Supermodular(s, y). Therefore, assume that these properties hold for V_s^n.

We first show that V_s^{n+1} satisfies Convex(s). To this end, consider $0 < x < B$ and $0 < y < N$. Then,

$$
\begin{aligned}
&V_{s+\Delta}^{n+1}(x,y) - 2\,V_s^{n+1}(x,y) + V_{s-\Delta}^{n+1}(x,y) \\
&\quad \geq c_{s+\Delta}(x,y) - 2\,c_s(x,y) + c_{s-\Delta}(x,y) \\
&\qquad + \lambda(s+\Delta)\,V_{s+\Delta}^n(x+1,y) - 2\lambda s\,V_s^n(x+1,y) + \lambda(s-\Delta)\,V_{s-\Delta}^n(x+1,y) \\
&\qquad + (1-\lambda(s+\Delta) - f(x) - h(x) - y\gamma)\,V_{s+\Delta}^n(x,y) \\
&\qquad - 2(1-\lambda s - f(x) - h(x) - y\gamma)\,V_s^n(x,y) \\
&\qquad + (1-\lambda(s-\Delta) - f(x) - h(x) - y\gamma)\,V_{s-\Delta}^n(x,y) \\
&\quad \geq 2\lambda\Delta\,[V_{s+\Delta}^n(x+1,y) - V_s^n(x+1,y) - V_{s+\Delta}^n(x,y) + V_s^n(x,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no s. The second inequality follows by rearranging terms such that Convex(s) for the induction hypothesis can be invoked on the expressions with factor $\lambda(s-\Delta)$ and $[1-\lambda(s-\Delta)-f(x)-h(x)-y\gamma]$. The final inequality follows from Supermodular(s, x). The result may also be verified at the boundaries corresponding to $x = 0$, $x = B$, $y = 0$ and $y = N$.

We now proceed to prove Supermodular(s, x). To this end, consider $0 < x < B-1$ and $0 < y < N$. Then

$$
\begin{aligned}
&V_{s+\Delta}^{n+1}(x+1,y) + V_s^{n+1}(x,y) - V_{s+\Delta}^{n+1}(x,y) - V_s^{n+1}(x+1,y) \\
&\quad \geq c_{s+\Delta}(x+1,y) + c_s(x,y) - c_{s+\Delta}(x,y) - c_s(x+1,y) \\
&\qquad + \lambda\,[(s+\Delta)\,V_{s+\Delta}^n(x+2,y) + s\,V_s^n(x+1,y) - (s+\Delta)\,V_{s+\Delta}^n(x+1,y) \\
&\qquad - s\,V_s^n(x+2,y)] + f(x+1)\,V_{s+\Delta}^n(x,y) + f(x)\,V_s^n(x-1,y) \\
&\qquad - f(x)\,V_{s+\Delta}^n(x-1,y) - f(x+1)\,V_s^n(x,y) + h(x+1)\,V_{s+\Delta}^n(x,y+1) \\
&\qquad + h(x)\,V_s^n(x-1,y+1) - h(x)\,V_{s+\Delta}^n(x-1,y+1) - h(x+1)\,V_s^n(x,y+1)] \\
&\qquad + (1-\lambda(s+\Delta) - f(x+1) - h(x+1) - y\gamma)\,V_{s+\Delta}^n(x+1,y) \\
&\qquad + (1-\lambda s - f(x) - h(x) - y\gamma)\,V_s^n(x,y) \\
&\qquad - (1-\lambda(s+\Delta) - f(x) - h(x) - y\gamma)\,V_{s+\Delta}^n(x,y) \\
&\qquad - (1-\lambda s - f(x+1) - h(x+1) - y\gamma)\,V_s^n(x+1,y) \\
\\
&\quad \geq \lambda\Delta\,[V_{s+\Delta}^n(x+2,y) - V_{s+\Delta}^n(x+1,y) - V_{s+\Delta}^n(x+1,y) + V_{s+\Delta}^n(x,y)] \\
&\qquad + [f(x+1) - f(x)]\,[V_{s+\Delta}^n(x,y) - V_s^n(x,y) + V_s^n(x,y) - V_{s+\Delta}^n(x,y)] \\
&\qquad + [h(x+1) - h(x)]\,[V_{s+\Delta}^n(x,y+1) - V_s^n(x,y+1) + V_s^n(x,y) - V_{s+\Delta}^n(x,y)] \\
&\quad \geq 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no s or x. The second inequality follows by rearranging terms such that Supermodular(s,x) from the induction hypothesis can be invoked on a select variety of expressions that have λs, $f(x)$, $h(x)$, or $[1-\lambda s-f(x+1)-h(x+1)-y\gamma]$ as a factor. The final inequality follows from Convex($x+1$) [Theorem 4.4] for the first line, and Supermodular(s, y) for the last line. The result may also be verified at the boundaries corresponding to $x=0$, $\{(x,y):1\le x\le B-2,\ y=0\}$, and $\{(x,y):1\le x\le B-2,\ y=N\}$. It also holds at the boundary $x=B-1$ provided that $\lambda=0$; however, this could be extended in the usual way provided that an expression with (uniformly in n) positive lower bound can be identified.

We continue the proof by showing Supermodular(s, y). To this end, consider $0<x<B$ and $0<y<N-1$.

$$
\begin{aligned}
&V^{n+1}_{s+\Delta}(x,y+1)+V^{n+1}_{s}(x,y)-V^{n+1}_{s+\Delta}(x,y)-V^{n+1}_{s}(x,y+1)\\
&\quad\ge c_{s+\Delta}(x,y+1)+c_s(x,y)-c_{s+\Delta}(x,y)-c_s(x,y+1)\\
&\qquad+\lambda\left[(s+\Delta)\,V^n_{s+\Delta}(x+1,y+1)+s\,V^n_s(x+1,y)-(s+\Delta)\,V^n_{s+\Delta}(x+1,y)\right.\\
&\qquad\left.-sV^n_s(x+1,y+1)\right]+\gamma\left[(y+1)\,V^n_{s+\Delta}(x+1,y)+y\,V^n_s(x+1,y-1)\right.\\
&\qquad\left.-y\,V^n_{s+\Delta}(x+1,y-1)-(y+1)\,V^n_s(x+1,y)\right]\\
&\qquad+(1-\lambda(s+\Delta)-f(x)-h(x)-(y+1)\gamma)\,V^n_{s+\Delta}(x,y+1)\\
&\qquad+(1-\lambda s-f(x)-h(x)-y\gamma)\,V^n_s(x,y)\\
&\qquad-(1-\lambda(s+\Delta)-f(x)-h(x)-y\gamma)\,V^n_{s+\Delta}(x,y)\\
&\qquad-(1-\lambda s-f(x)-h(x)-(y+1)\gamma)\,V^n_s(x,y+1)\\
&\quad\ge\lambda\Delta\left[V^n_{s+\Delta}(x+1,y+1)-V^n_{s+\Delta}(x+1,y)-V^n_{s+\Delta}(x,y+1)+V^n_{s+\Delta}(x,y)\right]\\
&\qquad+\gamma\left[V^n_{s+\Delta}(x+1,y)-V^n_s(x+1,y)+V^n_s(x,y)-V^n_{s+\Delta}(x,y)\right]\\
&\quad\ge 0.
\end{aligned}
$$

The first inequality follows from the induction hypothesis on the terms having factors with no s or y in them. The second inequality follows by rearranging terms such that Supermodular(s,y) from the induction hypothesis can be invoked on a select variety of expressions that have either λs, $y\gamma$, or $[1-\lambda s-f(x)-h(x)-(y+1)\gamma]$. The final inequality follows from Supermodular(x, y) [Theorem 4.4] and Supermodular(s, x). The above result can also be verified at the boundaries corresponding to $x=0$, $x=B$, $y=0$ and $y=N-1$. The proof is concluded by taking the limit as $n\to\infty$. □

5 Calculations

5.1 Solution Procedure

The Bellman equations take the form

$$g + V(x,y) = TV(x,y) \qquad\qquad x = 0,1,\ldots,B; \quad y = 0,1,\ldots,N$$

where the operator T is as defined earlier, with $V(0,0) \equiv 0$ (without loss of generality).
Rather than attempt to solve this set of equations using a *relative value iteration* scheme, we instead adopt a more direct approach. Observe that the RHS can be written as $a_{x,y}+F_{x,y}V$, where $a_{x,y}$ is a scalar; $F_{x,y}$ is a row vector of length $(B+1)(N+1)$; and V is a column vector of length $(B+1)(N+1)$ also, whose $[(x+1)N+(y+1)]$-th element is equal to $V(x,y)$.
Let $\mathbf{1}$ be a $(B+1)(N+1)$ column vector of ones; let a be a $(B+1)(N+1)$ column vector whose $[(x+1)N+(y+1)]$-th element is equal to $a_{x,y}$; and let F be a $(B+1)(N+1)\times(B+1)(N+1)$ matrix, whose $[(x+1)N+(y+1)]$-th row is the vector $F_{x,y}$. Then we may express the Bellman equations in terms of the following matrix equation:

$$g\mathbf{1} + (I-F)V = a.$$

In view of the condition $V(0,0)=0$, set $\widehat{V}$ to be the vector V with the element in the 1-st position (i.e., $V(0,0)$) removed, and set M to be equal to the matrix $I-F$ with the first column removed. Then $g\mathbf{1}+M\widehat{V}=a$. Hence

$$\begin{bmatrix} g \\ - - - \\ \widehat{V} \end{bmatrix} = [\mathbf{1}|M]^{-1}\, a.$$

5.2 Computational Results

We apply the above procedure to compute the system workload, W_s, as a function of $s \in (0,1]$. By dividing W_s by λs, we can also determine the delay cost, D_s, as a function of s. The values of B and N have been taken to be 20 and 30 respectively.

Example 1

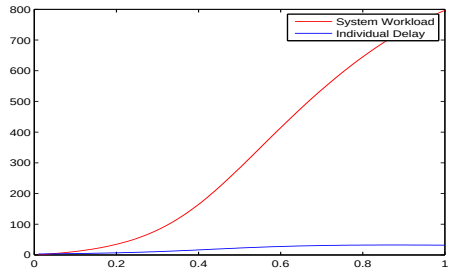

Figure 2: $\lambda = 25, \quad \mu = 10, \quad \beta = 10, \quad \gamma = 100, \quad \psi = 0.9,$ and $R = 50$

s	0.84	0.85	0.86	0.87	0.88	0.89	0.90
D_s	32.4784	32.5072	32.5254	32.5332	32.5310	32.5191	32.4981

Example 2

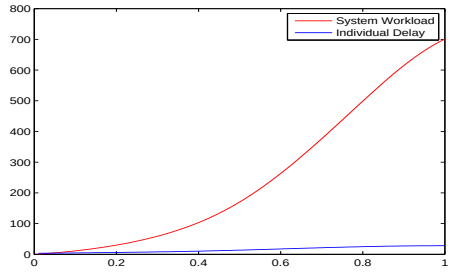

Figure 3: $\lambda = 25, \quad \mu = 10, \quad \beta = 10, \quad \gamma = 10, \quad \psi = 0.9,$ and $R = 50$

s	0.984	0.986	0.988	0.990	0.992	0.994	0.9960
D_s	28.0475	28.0492	28.0503	28.0507	28.0505	28.0495	28.0479

Example 3

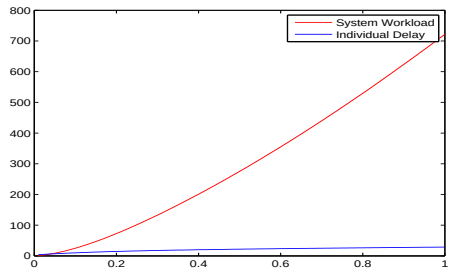

Figure 4: $\lambda = 25, \quad \mu = 10, \quad \beta = 10, \quad \gamma = 10, \quad \psi = 0.5,$ and $R = 50$

s	0.80	0.81	0.82	0.83	0.84	0.85	0.86
D_s	26.4745	26.6024	26.7294	26.8554	26.9804	27.1044	27.2275
s	0.87	0.88	0.89	0.90	0.91	0.92	0.93
D_s	27.3497	27.4710	27.5913	27.7108	27.8294	27.9471	28.0639
s	0.94	0.95	0.96	0.97	0.98	0.99	1.00
D_s	28.1800	28.2951	28.4095	28.5230	28.6357	28.7476	28.8587

Example 4

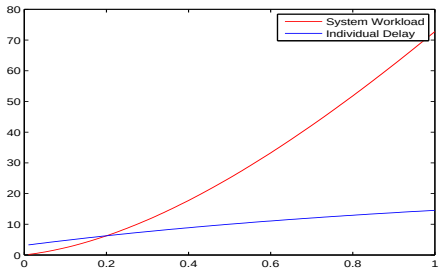

Figure 5: $\lambda = 5, \quad \mu = 10, \quad \beta = 10, \quad \gamma = 10, \quad \psi = 0.5,$ and $R = 50$

s	0.80	0.81	0.82	0.83	0.84	0.85	0.86
D_s	12.9515	13.0374	13.1226	13.2072	13.2911	13.3743	13.4570
s	0.87	0.88	0.89	0.90	0.91	0.92	0.93
D_s	13.5390	13.6204	13.7012	13.7814	13.8610	13.9400	14.0184
s	0.94	0.95	0.96	0.97	0.98	0.99	1.00
D_s	14.0962	14.1735	14.2502	14.3263	14.4019	14.4770	14.5515

Example 5

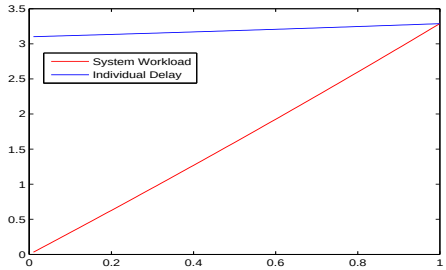

Figure 6: $\lambda = 1, \quad \mu = 10, \quad \beta = 1, \quad \gamma = 1, \quad \psi = 0.5,$ and $R = 50$

5.3 Discussion

In all five examples we see that W_s is increasing in s. However, both of the first two examples show that W_s lacks convexity. Convexity is sufficient (albeit not quite necessary) for the increasingness of D_s. Indeed, there are turning points in D_s at around $s = 0.87$ and $s = 0.990$ for examples 1 and 2 respectively. On the other hand, W_s is convex in s for examples 3 and 4, and so the corresponding D_s is increasing on the whole of $(0, 1]$ for both of these examples. Values for D_s for selected s values in the range $[0.8, 1]$ are provided to help substantiate this claim.

The examples have been chosen to explore the claims made in our results when we move from a heavy or moderate traffic regime, to parameters corresponding to a lighter traffic regime. In moving from example 1 to example 2, the parameter γ is reduced from 100 to 10; from example 2 to 3, the parameter ψ is reduced from 0.9 to 0.5; from example 3 to 4, λ is reduced from 25 to 5. These changes correspond to making γ, $\beta\psi$, and λ sufficiently small so that the convexity of W_s pertains. However, Theorem 3.7 suggests that it may be possible for W_s to be convex, but without the parameters conspiring to yield a traffic regime which is sufficiently "light" for convexity in D_s to pertain. Reducing the values for λ, β, and γ still further, as in example 5, gives rise to convex D_s (in contradistinction to examples 3 and 4); this is easily checked using, for example, the Matlab function `convexhull`.

6 Application

Anticipation of the likely demand for a service provided by either a call or contact centre is crucial for informing the allocation of resources required to meet particular performance levels. For example, given the cost of being kept waiting but perhaps not holding on long enough to get the desired service (R), and the maximum expected delay cost that the customer is prepared to accept (θ), what proportion of the potential demand (λ) will be dissuaded from trying to obtain service (represented

by $1-s$), and what proportion will at least try to obtain it (i.e., s)? One way to estimate s would be to set up a real experiment in which the resource level is set by the centre (service rate, μ, say), and then the values of R, λ, and s, are estimated through the combined use and analysis of sample surveys and collated usage statistics. Quite apart from the lack of cost-effectiveness of such a study, it is not within the gift of the experimenter to arbitrarily vary the values of R and λ (over a range of values), rather they have to be taken as they present themselves, for the given resource level. Simulations in which user decisions are based on some kind of learned information from the system data could offer more flexible experimentation over a variety of parameter values, although the computational burden involved could be significant.

Another possibility would be to work with the Nash equilibrium solution concept (see Ben-Shahar, Orda & Shimkin (2000) and references contained therein for related work in system state dependent contexts). Let the *strategy* $u(s)$ represent the probability that a particular customer, C say, joins the system (and balks with probability $1-u(s)$), when all other customers adopt a strategy of joining the system with probability s but balking with probability $1-s$. For conciseness, we represent the collection of strategies used by all customers other than C by the *policy* $[s]^\infty$.

Definition 1. *The strategy $u(s)$ is said to be optimal against the policy $[s]^\infty$ if*

$$u(s) = \begin{cases} 1 & \text{if} \quad D_s < \theta \\ 0 & \text{if} \quad D_s > \theta \\ q \in [0,1] & \text{if} \quad D_s = \theta \end{cases}.$$

□

A motivation for the above definition is provided by the following argument. Suppose that customer C adopts a strategy s against the other customers who adopt the policy $[s']^\infty$. Then the expression for the expected *overall* delay cost for C may be appropriately defined by

$$G_{s'}(s) = sD_{s'} + (1-s)\theta$$

which, subject to the constraint that $s \in [0,1]$, we seek to minimize. Clearly, this quantity is continuous in s on $[0,1]$, and differentiable on $(0,1)$, with derivative $D_{s'} - \theta$. Hence, for $D_{s'} < \theta$, $G_{s'}(s)$ is minimized at $s = 1$; for $D_{s'} > \theta$, $G_{s'}(s)$ is minimized at $s = 0$; and for $D_{s'} = \theta$, $G_{s'}(s)$ takes the same value for all $s \in [0,1]$.

With a slight abuse of terminology, in which we allow the definition of a policy to incorporate the strategy of the arbitrary customer C, we define the Nash equilibrium.

Definition 2.
We say that $[s^*]^\infty$ is a (symmetric) Nash equilibrium if s^* is optimal for any arbitrary customer C against the policy $[s^*]^\infty$ adopted by all other customers. □

To avoid trivialities, we assume $\theta > 0$: this ensures that against a policy $[s]^\infty$, it will always be optimal for an arbitrary customer to join the system for s sufficiently small.

Theorem 6.1. *Suppose that the delay cost D_s is strictly increasing in s. Then there exists a unique symmetric Nash equilibrium policy $[s^*]^\infty$. Furthermore, if $D_1 < \theta$, then $s^* = 1$, and if $D_1 \geq \theta$, then s^* is the unique solution to the equation $D_s = \theta$.*

Proof. The result is immediate from the monotonicity and continuity of D_s on $[0, 1]$. □

7 Concluding Remarks

The main result of this paper asserts that, provided that the rates associated with arrivals to the system, abandonment from the facility into the retrial orbit, and customers leaving the retrial orbit, are sufficiently small, the system workload, W_s, is convex increasing and continuous in s. In consequence, and under similar conditions, the delay cost, D_s, associated with individual customers, is continuously increasing in s. The results have been established under the proviso that the bounds on the state-space, namely B and N are taken to be finite. Finiteness of B and N is needed for the successful application of the uniformization technique in setting up the DP operator; nonetheless, taking N to be as large as possible, and setting B to match the maximum capacity of facility Q_1, our model can be calibrated to capture the realism of the situation to hand. A careful examination of many of the results in this paper will show that the proofs will tend to break down at the upper boundaries under more general assumptions for the rates. Although our results are restricted to light traffic conditions (which should be of interest in their own right), we believe that they provide some insight into what could be true for an unbounded state space ($B = N = \infty$) under more general rate assumptions. We take the view that a DP approach for the latter, and significantly more challenging, scenario would require that the DTMC be embedded at the arrival epochs. This would allow the relevant process to be (self-)uniformized, where the expected inter-epoch time is $1/\lambda$ and finite; however the more complicated structure for the transition probabilities would present a much more significant set of challenges.

References

Ben-Shahar, I., Orda, A. & Shimkin, N. (2000), 'Dynamic Service Sharing with Heterogeneous Preferences', *Queueing Systems* **35**, 83–103.

Daley, D. J. & Rolski, T. (1991), 'Light traffic approximations in queues', *Mathematics of Operations Research* **16**, 57–71.

Dewan, S. & Mendelson, H. (1990), 'User delay cost and internal pricing for a service facility', *Management Science* **36**, 1502–1517.

Fridgeirsdottir, K. & Chiu, S. (2005), 'A Note on Convexity of the Expected Delay Cost in Single-Server Queues', *Operations Research* **53**, 568–570.

Lippman, S. (1975), 'Applying a New Device in the Optimization of Exponential Queuing Systems', *Operations Research* **23**, 687–709.

Ross, S. M. (1983), *Introduction to Stochastic Dynamic Programming*, Academic Press, Inc.

Stidham, S. (1992), 'Pricing and capacity decisions for a service facility', *Management Science* **38**, 1121–1139.