Equilibrium balking strategies in the single server Markovian queue with catastrophes

Olga Boudali and Antonis Economou
olboudali@math.uoa.gr and aeconom@math.uoa.gr
University of Athens, Department of Mathematics
Panepistemiopolis, Athens 15784, Greece

July 13, 2011

Abstract: We consider a Markovian queue subject to Poisson generated catastrophes. Whenever a catastrophe occurs, all customers are forced to abandon the system, the server is rendered inoperative and an exponential repair time is set on. We assume that the arriving customers decide whether to join the system or balk, based on a natural reward-cost structure. We study the balking behavior of the customers and derive the corresponding Nash equilibrium strategies.

Keywords: Queueing, Catastrophes, Balking, Nash equilibrium strategies, Social optimization

1 Introduction

Queues with removals of customers before being served are often encountered in practice. One type of such a situation appears in queueing systems with reneging, where customers are impatient and as soon as their patience times expire they leave the system. Another type of such a situation occurs in systems that are subject to catastrophes/failures. Such events usually render the server(s) inoperative and in addition force the customers to leave the system. The crucial difference between these two types of abandonments is that the customers decide whether to leave the system or not according to their own desire in the case of reneging, while they are forced to abandon the system in the case of catastrophes.

During the last decades, there is an emerging tendency to study queueing systems from an economic viewpoint. More concretely, a certain reward-cost structure is imposed on the system that reflects the customers' desire for service and their unwillingness to wait. Customers are allowed to make decisions about their actions in the system, for example they may decide whether to join or balk, to wait or abandon, to retry or not etc. The customers want to maximize their benefit, taking into account that the other customers have the same objective, and so the situation can be considered as a game among the customers. In this type of studies, the main goal is to find individual and social optimal strategies. The study of queueing systems under a game-theoretic perspective was initiated by Naor (1969) who studied the $M/M/1$ model with a linear reward-cost structure. Naor (1969) assumed that an arriving customer observes the number of customers and then makes his decision whether to join or balk (observable case). His study was complemented by Edelson and Hildebrand (1975) who considered the same queueing

system but assumed that the customers make their decisions without being informed about the state of the system. Since then, there is a growing number of papers that deal with the economic analysis of the balking behavior of customers in variants of the $M/M/1$ queue, see e.g. Burnetas and Economou (2007) ($M/M/1$ queue with setup times), Economou and Kanta (2008a,b) ($M/M/1$ queue with compartmented waiting space, $M/M/1$ queue with unreliable server), Guo and Zipkin (2007) ($M/M/1$ queue with various levels of information and non-linear reward-cost structure), Hassin and Haviv (1997) ($M/M/1$ queue with priorities), Hassin (2007) ($M/M/1$ queue with various levels of information and uncertainty in the system parameters). The monographs of Hassin and Haviv (2003) and Stidham (2009) summarize the main approaches and several results in the broader area of the economic analysis of queueing systems.

The study of the equilibrium customer behavior in queueing systems with abandonments has received less attention. Hassin and Haviv (1995) identified equilibrium customer strategies regarding balking and reneging in the $M/M/1$ queue, where the reward for an individual reduces to zero if its waiting time exceeds a certain threshold time. Mandelbaum and Shimkin (2000) considered a quite general model for abandonments from a queue, due to excessive wait, assuming that waiting customers act rationally but without being able to observe the queue length. More importantly they allowed customers to be heterogeneous in their preferences and consequent behavior. Other authors have also considered the equilibrium behavior of customers in queueing systems with abandonments due to reneging. Hassin and Haviv (2003), in Chapter 5, summarize the main results for such models. However, to the best of our knowledge, studies for the equilibrium behavior of customers in queueing systems with abandonments/removals of customers due to catastrophic events do not yet exist. It is the aim of the present paper to study the equilibrium behavior of customers in the context of a simple queueing model subject to catastrophes.

More specifically, in the present paper, we investigate the equilibrium balking behavior of customers in a queue of $M/M/1$ type with complete removals at Poisson generated catastrophe epochs. A catastrophe renders the server inactive (due to either a failure or preventive check/maintenance) and a repair time is set on. During the repair time the system does not admit customers. When the repair time is completed, the system behaves as an $M/M/1$ queue till the next catastrophe and so on. We impose on the system a linear reward-cost structure as the one in Naor (1969) and Edelson and Hildebrand (1975). However, we make a modification, considering two different types of reward: the first one is the usual reward received by the customers that leave the system after service completion, while the second is a compensation received by those that are forced to abandon the system due to a catastrophe. In fact the role of this compensation is to mitigate customers' dissatisfaction. We again study the customers' behavior regarding the dilemma whether to join or balk. We consider two cases with respect to the level of information available to customers before making their decisions. More specifically, at his arrival epoch, an arbitrary customer may or may not be informed about the state of the system (observable and unobservable cases correspondingly). In each case, we characterize customer equilibrium strategies and we treat the social optimization problem. We also explore the effect of the information level on the equilibrium behavior of the customers through numerical comparisons.

The paper is organized as follows. In Section 2 we describe the dynamics of the model, the reward-cost structure and the decision framework. In Section 3 we determine equilibrium threshold strategies for the observable case, in which customers get informed about the state of

the system before making their decisions. In Section 4 we study the unobservable case, deriving mixed equilibrium balking strategies. Finally, in Section 5, we treat the social optimization problem. Moreover, we present the results from several numerical experiments that demonstrate the effect of the information level on the behavior of the customers and on the various performance measures of the system.

2   Model description

We consider a single-server queue with infinite waiting space, where customers arrive according to a Poisson process at rate λ. The service requirements of successive customers are independent and identically distributed random variables with exponential distribution with rate μ. The server serves the customers one by one. The system is subject to catastrophes/failures according to a Poisson process at rate ξ. When a catastrophe occurs all customers are forced to abandon the system prematurely, without being served. The system is rendered inoperative and a repair process is set on. The length of a repair time is exponentially distributed at rate η. During a repair time, arrivals are not accepted. We finally assume that interarrival times, service times, intercatastrophe times and repair times are mutually independent.

We represent the state of the system at time t by a pair $(Q(t), I(t))$, where $Q(t)$ records the number of customers at the system and $I(t)$ denotes the server state, with 1 describing a system in operation and 0 describing a system under repair. Note that whenever $I(t)$ is zero, $Q(t)$ should be necessarily zero too. Thus, the stochastic process $\{(Q(t), I(t)) : t \geq 0\}$ is a continuous time Markov chain with state space $S = \{(n, 1), n \geq 0\} \cup \{(0, 0)\}$ and its transition rate diagram is shown in Figure 1.
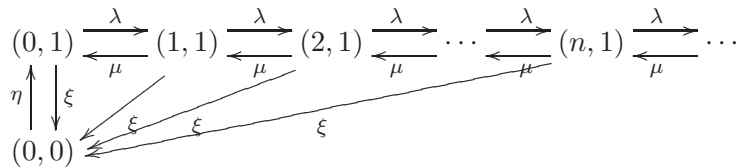

Figure 1: Transition rate diagram of $\{(Q(t), I(t))\}$.

We are interested in the behavior of customers, when they have the option to decide whether to join or balk. We model this decision framework by assuming that each customer receives either a reward of R_s units for completing service or a compensation of R_f units in case that he is forced to abandon the system due to a failure. Moreover, a customer is charged a cost of C units per time unit that he remains in the system (either in queue or in the service space). We also assume that customers are risk neutral and wish to maximize their net benefit. Finally, their decisions are assumed irrevocable, meaning that neither reneging of entering customers nor retrials of balking customers are allowed.

Since all customers are assumed indistinguishable, we can consider the situation as a symmetric game among them. Denote the common set of strategies (set of available actions) and the payoff function by S and F respectively. More concretely, let $F(s_{tagged}, s_{others})$ be the payoff for a tagged customer who follows strategy s_{tagged}, when all other customers follow s_{others}. A

3

strategy s_1 is said to dominate strategy s_2 if $F(s_1, s) \geq F(s_2, s)$, for every $s \in S$ and for at least one s the inequality is strict. A strategy s_* is said to be weakly dominant if it dominates all other strategies in S. A strategy $\tilde{s}$ is said to be a best response against a strategy s_{others}, if $F(\tilde{s}, s_{others}) \geq F(s_{tagged}, s_{others})$, for every $s_{tagged} \in S$. Finally, a strategy s_e is said to be a (symmetric) Nash equilibrium, if and only if it is a best response against itself, i.e. $F(s_e, s_e) \geq F(s, s_e)$, for every $s \in S$. The intuitive interpretation of a Nash equilibrium is that it is a stable point of the game in the sense that if all customers agree to follow it, then no one can benefit by changing it. We remark that the notion of a dominant strategy is stronger than the notion of an equilibrium. In fact, every dominant strategy is an equilibrium, but the converse is not true. Moreover, while equilibrium strategies exist in most situations, a dominant strategy rarely does.

In the next sections we obtain customer equilibrium strategies for joining/balking. We distinguish two cases with respect to the level of information available to customers at their arrival instants, before their decisions are made; the observable case (where customers observe $Q(t)$) and the unobservable case.

3 Equilibrium strategies - the observable case

In this section we study the model, under the assumption that the customers who find the server active observe the number of customers in the system, before deciding whether to enter or balk. We prove that a threshold type dominant strategy exists, which constitutes the unique equilibrium balking strategy for the customers in the system. We first give the expected net reward of a customer that observes n customers ahead of him and decides to enter. We have the following.

Proposition 3.1 *Consider the observable model of the $M/M/1$ queue with catastrophes causing complete removals of customers. The expected net benefit of a customer that observes n customers in the system upon arrival and decides to enter is given by*

$$S_{obs}(n) = R_s \left(\frac{\mu}{\mu+\xi}\right)^{n+1} + R_f \left[1 - \left(\frac{\mu}{\mu+\xi}\right)^{n+1}\right] - \frac{C}{\xi}\left[1 - \left(\frac{\mu}{\mu+\xi}\right)^{n+1}\right], \ n \geq 0. \quad (3.1)$$

Proof. Consider a tagged customer that finds the system at state $(n, 1)$ upon arrival and decides to enter. This customer may leave the system either due to its service completion or due to a catastrophe that will force him to abandon prematurely the system. For his service completion, he has to wait for a sum of $n + 1$ independent exponentially distributed times with parameter μ (note that because of the memoryless property of the exponential distribution, we can assume that the distribution of the remaining service time of the customer in service is identical to the service time distribution of the other customers). For the next catastrophe, he has to wait for an exponentially distributed time with parameter ξ. Therefore, the sojourn time of such a customer in the system is given as $Z = \min(Y_n, X)$, where Y_n follows a Gamma distribution with parameters $n + 1$, μ and X is an exponentially distributed random variable with rate ξ, independent of Y_n. Moreover, the tagged customer will be served with probability $\Pr[Y_n < X]$, while he will be forced to abandon the system due to a catastrophe with the

4

complementary probability $\Pr[Y_n \geq X]$. Therefore his net benefit will be

$$S_{obs}(n) = R_s \Pr[Y_n < X] + R_f \Pr[Y_n \geq X] - CE[Z]. \tag{3.2}$$

Note now that

$$\Pr[Y_n < X] = \int_0^\infty e^{-\xi y} \frac{\mu^{n+1}}{n!} y^n e^{-\mu y} dy = \left(\frac{\mu}{\mu + \xi}\right)^{n+1} \tag{3.3}$$

and

$$E[Z] = \int_0^\infty e^{-\xi z} \int_z^\infty \frac{\mu^{n+1}}{n!} u^n e^{-\mu u} du \, dz = \frac{1}{\xi} \left[1 - \left(\frac{\mu}{\mu + \xi}\right)^{n+1}\right]. \tag{3.4}$$

Plugging (3.3) and (3.4) in (3.2) yields (3.1). ∎

We now consider an arbitrary customer who observes upon arrival the state of the system. Since arrivals are not permitted during the repair time, if a customer observes the system at state $(0,0)$, he is not allowed to enter and thus there is no decision. So, we only consider the case where a customer observes the system at a state $(n, 1)$. Such a customer strictly prefers to enter if his net benefit is positive, is indifferent between joining and balking if it is zero and strictly prefers to balk if it is negative. In the sequel, we suppose for simplicity that customers break ties in favor of entering. We have the following.

Theorem 3.1 *In the observable model of the M/M/1 queue with catastrophes causing complete removals of customers, a unique dominant pure strategy exists (which is also the unique equilibrium strategy). There are three cases:*

Case I: $R_f < \frac{C}{\xi} - \frac{\mu R_s}{\xi}$.

Then the unique dominant strategy is always to balk.

Case II: $\frac{C}{\xi} - \frac{\mu R_s}{\xi} \leq R_f < \frac{C}{\xi}$.

Then the unique dominant strategy is the threshold strategy 'While arriving at time t and finding the system operative, observe $Q(t)$; enter if $Q(t) \leq n_e$ and balk otherwise', where n_e is given by

$$n_e = \left\lfloor \frac{\ln K}{\ln S} - 1 \right\rfloor \tag{3.5}$$

with

$$K = \frac{\frac{C}{\xi} - R_f}{R_s - R_f + \frac{C}{\xi}}, \quad S = \frac{\mu}{\mu + \xi} \tag{3.6}$$

and $\lfloor x \rfloor$ denotes the floor of x, i.e. the greatest integer which is smaller than or equal to x.

Case III: $R_f \geq \frac{C}{\xi}$.

Then the unique dominant strategy is always to enter.

Proof. Consider a tagged customer that observes the system upon arrival. If he finds the system at state $(n, 1)$ and decides to enter, then his expected net benefit is given by (3.1). The customer will prefer to enter if $S_{obs}(n) \geq 0$, which is written easily as

$$\left(R_s - R_f + \frac{C}{\xi}\right) \cdot \left(\frac{\mu}{\mu + \xi}\right)^{n+1} \geq \frac{C}{\xi} - R_f. \tag{3.7}$$

5

Since $R_s - R_f + \frac{C}{\xi} > \frac{C}{\xi} - R_f$, we have the following three cases.

Case A: $\frac{C}{\xi} - R_f > 0 \Leftrightarrow R_f < \frac{C}{\xi}$.

We can solve (3.7) with respect to n and we obtain that the tagged customer is willing to enter as long as he observes at most n_e customers in the system with n_e given by (3.5). However, it is easy to see that n_e given by (3.5) becomes negative when $R_f < \frac{C}{\xi} - \frac{\mu R_s}{\xi}$. Therefore, it is then optimal always to balk and we conclude with Case I. On the other hand, when $\frac{C}{\xi} - \frac{\mu R_s}{\xi} \le R_f < \frac{C}{\xi}$, the threshold n_e given by (3.5) is non-negative and we conclude with Case II.

Case B: $R_s - R_f + \frac{C}{\xi} > 0 \ge \frac{C}{\xi} - R_f \Leftrightarrow \frac{C}{\xi} \le R_f < \frac{C}{\xi} + R_s$.

In this case the inequality (3.7) is always true. Therefore the tagged customer is always willing to enter.

Case C: $0 \ge R_s - R_f + \frac{C}{\xi} > \frac{C}{\xi} - R_f \Leftrightarrow R_f \ge \frac{C}{\xi} + R_s$.

Solving inequality (3.7) with respect to n shows that the customer is willing to enter as long as as he observes at least n_e customers in the system with n_e given by (3.5). However, n_e is easily seen to be negative in this case, so it is always preferable for the customer to enter. Therefore, Cases B and C yield to Case III of the statement.

Note that the strategies prescribed above are dominant, since they do not depend on what the other customers do, i.e. they are best responses against any strategy of the others. ∎

Remark 3.1 *The equilibrium (dominant) strategies do not depend on the value of the repair rate η. This happens because the customers make decisions only whenever arrive at an operative system. On the contrary, the social optimal strategies do depend on η, as we will see in Section 5. Furthermore, in the limiting case where $\xi \to 0$, we can easily check, using L'Hospital rule, that the threshold n_e tends to the threshold derived by Naor (1969) for the M/M/1 system.*

4   Equilibrium strategies - the unobservable case

We now turn our interest to the unobservable case, in which the customers only know the values of the system parameters λ, μ, ξ and η and of the economic parameters R_s, R_f and C, but do not observe the state of the system upon arrival. Thus, there are only two pure strategies, 'to join' and 'to balk' and a mixed strategy is specified by the joining probability q of an arriving customer that finds the server operative. Our goal in this section is to identify the equilibrium mixed balking strategies.

Suppose that the customers follow a mixed strategy with joining probability q. Then, the system behaves as the original, but with arrival rate λq instead of λ. Its transition diagram is seen in Figure 2.
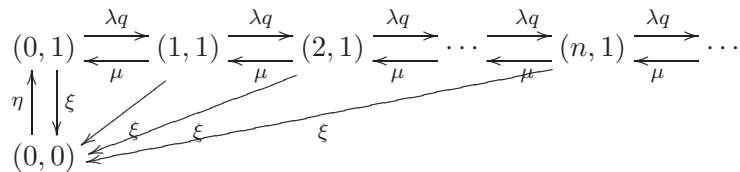

Figure 2: Transition rate diagram of $\{(Q(t), I(t))\}$ for a given mixed balking strategy q.

We have the following.

Proposition 4.1 *Consider the unobservable model of the M/M/1 queue with catastrophes causing complete removals of customers, in which the customers that find the server operative join with probability q. The stationary probabilities $p_{un}(k,i)$ of the system are given by*

$$p_{un}(0,0) = \frac{\xi}{\xi + \eta}, \tag{4.1}$$

$$p_{un}(k,1) = \frac{\eta\left(1 - x_2(q)\right) x_2(q)^k}{\xi + \eta}, \quad k \geq 0, \tag{4.2}$$

where $x_2(q)$ is given by

$$x_2(q) = \frac{(\lambda q + \mu + \xi) - \sqrt{(\lambda q + \mu + \xi)^2 - 4\lambda q\mu}}{2\mu}. \tag{4.3}$$

The expected net benefit of a customer that enters with probability q' given that the system is found operative, when the others follow a strategy q is given by

$$S_{un}(q',q) = q'\left[\left(R_s - R_f + \frac{C}{\xi}\right)\frac{\mu\left(1 - x_2(q)\right)}{\mu + \xi - \mu x_2(q)} + R_f - \frac{C}{\xi}\right]. \tag{4.4}$$

Proof. The balance equations for the stationary distribution of the Markov chain $\{(Q(t), I(t))\}$ are given as follows:

$$\eta p_{un}(0,0) = \xi \sum_{k=0}^{\infty} p_{un}(k,1), \tag{4.5}$$

$$(\lambda q + \xi)\, p_{un}(0,1) = \mu p_{un}(1,1) + \eta p_{un}(0,0), \tag{4.6}$$

$$(\lambda q + \mu + \xi)\, p_{un}(k,1) = \lambda q p_{un}(k-1,1) + \mu p_{un}(k+1,1), \quad k \geq 1. \tag{4.7}$$

Equation (4.5) and the normalization equation $p_{un}(0,0) + \sum_{k=0}^{\infty} p_{un}(k,1) = 1$ imply immediately (4.1). Equation (4.7) can be considered as a homogeneous linear difference equation of order 2 with constant coefficients and characteristic equation

$$(\lambda q + \mu + \xi)\, x = \lambda q + \mu x^2 \tag{4.8}$$

that has two roots, $x_1(q)$ and $x_2(q)$, given by

$$x_{1,2}(q) = \frac{(\lambda q + \mu + \xi) \pm \sqrt{(\lambda q + \mu + \xi)^2 - 4\lambda q\mu}}{2\mu}. \tag{4.9}$$

From the standard theory of homogeneous linear difference equations (see e.g. Elaydi (1999) Section 2.3) we conclude that $p_{un}(k,1) = c_1(q)x_1(q)^k + c_2(q)x_2(q)^k$, for $k \geq 0$, where $c_1(q)$ and $c_2(q)$ are constants to be determined. We can easily check that $x_1(q) > 1$, hence $c_1(q)$ should be necessarily 0, for $p_{un}(k,1)$, $k \geq 0$, are probabilities and so should remain bounded. The constant $c_2(q)$ can be calculated using the normalization equation and we deduce (4.2).

The expected net benefit of a customer that decides to enter when the others follow the strategy q can be computed by conditioning on the state that he observes upon arrival. The probability that an arriving customer finds n customers in the system, given that he finds the server operative (and so he can decide whether to enter or not) is

$$p_{un}^{arr(\cdot,1)}(k,1) = \frac{\lambda p_{un}(k,1)}{\sum_{i=0}^{\infty} \lambda p_{un}(i,1)} = (1 - x_2(q))x_2(q)^k, \quad k \geq 0. \tag{4.10}$$

7

Such a customer receives on the average $S_{obs}(k)$ units, given by (3.1). Therefore, the expected net benefit of a customer that decides to enter given that he has found an operative system and the others follow the strategy q is given by

$$
\begin{aligned}
S_{un}(1,q) &= \sum_{k=0}^{\infty} p_{un}^{arr(\cdot,1)}(k,1)S_{obs}(k) \\
&= \sum_{k=0}^{\infty}(1-x_2(q))x_2(q)^k \left\{ R_s \left(\frac{\mu}{\mu+\xi}\right)^{k+1} + (R_f - \frac{C}{\xi})\left[1 - \left(\frac{\mu}{\mu+\xi}\right)^{k+1}\right]\right\} \\
&= \left(R_s - R_f + \frac{C}{\xi}\right)\frac{\mu(1-x_2(q))}{\mu+\xi-\mu x_2(q)} + R_f - \frac{C}{\xi}. \tag{4.11}
\end{aligned}
$$

By the linearity of $S_{un}(q',q)$ with respect to the first argument, we have that $S_{un}(q',q) = (1-q')S_{un}(0,q) + q'S_{un}(1,q)$ and we obtain readily (4.4). ∎

We can now proceed to determine the equilibrium balking strategies of a customer in the unobservable case. We have the following.

Theorem 4.1 *In the unobservable model of the $M/M/1$ queue with catastrophes causing complete removals of customers, a unique equilibrium mixed strategy exists, with joining probability q_e given by*

$$
q_e = \begin{cases} 0 & \text{if } R_f \leq \frac{C}{\xi} - \frac{\mu R_s}{\xi} \\ \frac{(C-\xi R_f + \xi R_s)(\mu R_s + \xi R_f - C)}{\lambda(C-\xi R_f)R_s} & \text{if } \frac{C}{\xi} - \frac{\mu R_s}{\xi} < R_f < \frac{C}{\xi} - \frac{\mu R_s(1-x_2)}{\xi} \\ 1 & \text{if } R_f \geq \frac{C}{\xi} - \frac{\mu R_s(1-x_2)}{\xi}, \end{cases} \tag{4.12}
$$

where $x_2 = x_2(1)$ (using (4.3) for $q = 1$).

Proof. Suppose that customers who find the server operative enter with probability q and consider a tagged arriving customer. Then, the tagged customer prefers to enter if $S_{un}(1,q) > 0$, he is indifferent between entering and balking if $S_{un}(1,q) = 0$ and he prefers to balk if $S_{un}(1,q) < 0$. We consider the equation $S_{un}(1,q) = 0$ with $S_{un}(1,q)$ given by (4.11) and we solve for $x_2(q)$. It may be easily checked that the above equation has a unique solution given from

$$
x_{2e} = \frac{\mu R_s + \xi R_f - C}{\mu R_s} \tag{4.13}
$$

and the corresponding q_e is found by solving (4.8), for $x = x_{2e}$, with respect to q. This yields

$$
q_e = \frac{x_{2e}[\mu(1-x_{2e}) + \xi]}{\lambda(1-x_{2e})} = \frac{(C - \xi R_f + \xi R_s)(\mu R_s + \xi R_f - C)}{\lambda(C - \xi R_f)R_s}. \tag{4.14}
$$

We can now easily see that q_e given by (4.14) lies in the interval $(0,1)$ if and only if $R_f \in \left(\frac{C}{\xi} - \frac{\mu R_s}{\xi}, \frac{C}{\xi} - \frac{\mu R_s(1-x_2)}{\xi}\right)$ and we obtain the second branch of (4.12). On the other hand, we can check that $S_{un}(1,q)$ is positive for all $q \in [0,1]$, when $R_f \geq \frac{C}{\xi} - \frac{\mu R_s(1-x_2)}{\xi}$ and, as a result, a customer's best response is 1 in this case. Thus, 'enter' is the unique equilibrium strategy and we obtain the third branch of (4.12).

Finally, for $q = 0$, the system alternates between only two states, $p_{un}(0,0)$ and $p_{un}(0,1)$, and the stationary probabilities are $p_{un}(0,0) = \frac{\xi}{\xi+\eta}$ and $p_{un}(0,1) = \frac{\eta}{\xi+\eta}$. We can see that $S_{un}(1,0)$ is non-positive if and only if $R_f \leq \frac{C}{\xi} - \frac{\mu R_s}{\xi}$. Therefore, in this interval, $q = 0$ is the unique equilibrium strategy and we obtain the first branch of (4.12). ∎

8

Remark 4.1 *The equilibrium strategies do not depend on the value of the repair rate η. This happens because the customers only make decisions whenever they arrive at an operative system. However, unlike the observable case, social optimal strategies do not depend on η either, as we will see in Section 5. Furthermore, in the limiting case where $\xi \to 0$, we can easily check that the equilibrium probability q_e tends to the equilibrium probability derived by Edelson and Hildebrand (1975) for the $M/M/1$ system.*

5   Social optimal strategies - conclusions

We are now studying the problem of social optimization. We treat separately the observable and unobservable cases. We are interested in determining the optimal values of the expected net social benefit (per time unit) functions, $S_{obs}^{soc}(n)$ and $S_{un}^{soc}(q)$, and the corresponding arguments, n_{soc} and q_{soc} respectively. First we consider the observable case and we have the following Proposition 5.1.

Proposition 5.1 *Consider the observable model of the $M/M/1$ queue with catastrophes causing complete removals of customers. The expected net social benefit per time unit, given that the customers follow a threshold strategy with threshold n (i.e. arriving customers that observe at most n customers in an operative system do enter, while the rest balk without being served) is given by*

$$
\begin{aligned}
S_{obs}^{soc}(n) &= \frac{\lambda(R_s - R_f)}{(\mu + \xi)^{n+1}} \left\{ \frac{\mu d_1(n)\left[(\mu + \xi)^{n+1} - (\mu x_1)^{n+1}\right]}{\mu + \xi - \mu x_1} + \frac{\mu d_2(n)\left[(\mu + \xi)^{n+1} - (\mu x_2)^{n+1}\right]}{\mu + \xi - \mu x_2} \right\} \\
&\quad + \lambda R_f \left(\frac{\eta}{\xi + \eta} - d_1(n)x_1^{n+1} - d_2(n)x_2^{n+1} \right) \\
&\quad - \frac{C\mu^2}{\xi^2} d_1(n)x_1(1 - x_2)^2[1 - (n+2)x_1^{n+1} + (n+1)x_1^{n+2}] \\
&\quad - \frac{C\mu^2}{\xi^2} d_2(n)x_2(1 - x_1)^2[1 - (n+2)x_2^{n+1} + (n+1)x_2^{n+2}], \ n \geq 0, \quad (5.1)
\end{aligned}
$$

where $x_1 = x_1(1)$ and $x_2 = x_2(1)$ (using (4.9) for $q = 1$) and $d_1(n)$, $d_2(n)$ are given by

$$
d_1(n) = \frac{-\eta\xi[(\mu + \xi)x_2 - \lambda]x_2^n}{(\xi + \eta)\left\{(\lambda + \xi - \mu x_2)[(\mu + \xi)x_1 - \lambda]x_1^n - (\lambda + \xi - \mu x_1)[(\mu + \xi)x_2 - \lambda]x_2^n\right\}}, (5.2)
$$

$$
d_2(n) = \frac{\eta\xi[(\mu + \xi)x_1 - \lambda]x_1^n}{(\xi + \eta)\left\{(\lambda + \xi - \mu x_2)[(\mu + \xi)x_1 - \lambda]x_1^n - (\lambda + \xi - \mu x_1)[(\mu + \xi)x_2 - \lambda]x_2^n\right\}}. (5.3)
$$

Proof. The stationary distribution of the model under a threshold strategy with threshold n can be found along the same lines with the proof of Proposition 4.1 (i.e. by using the theory of linear difference equations with constant coefficients). We then obtain that

$$
p_{obs}(0,0) = \frac{\xi}{\xi + \eta}, \quad (5.4)
$$

$$
p_{obs}(k,1) = d_1(n)x_1^k + d_2(n)x_2^k, \quad 0 \leq k \leq n+1, \quad (5.5)
$$

with x_1, x_2, $d_1(n)$ and $d_2(n)$ as in the statement of the Proposition. The expected net social benefit per time unit is then found by

$$
S_{obs}^{soc}(n) = \lambda P_{obs}^{ser} R_s + \lambda P_{obs}^{cat} R_f - CE_{obs}[Q], \quad (5.6)
$$

where P_{obs}^{ser} and P_{obs}^{cat} are the fraction of customers that leave the system due to service and catastrophes respectively and $E_{obs}[Q]$ is the mean number of customers in system. Using (5.4), (5.5) and (3.3), we can compute P_{obs}^{ser}, P_{obs}^{cat} and $E_{obs}[Q]$ as

$$P_{obs}^{ser} = \sum_{k=0}^{n} p_{obs}(k,1) \left(\frac{\mu}{\mu+\xi}\right)^{k+1} = \sum_{k=0}^{n} (d_1(n)x_1^k + d_2(n)x_2^k)\left(\frac{\mu}{\mu+\xi}\right)^{k+1}, \quad (5.7)$$

$$P_{obs}^{cat} = \sum_{k=0}^{n} p_{obs}(k,1)\left[1 - \left(\frac{\mu}{\mu+\xi}\right)^{k+1}\right]$$

$$= \sum_{k=0}^{n} (d_1(n)x_1^k + d_2(n)x_2^k)\left[1 - \left(\frac{\mu}{\mu+\xi}\right)^{k+1}\right], \quad (5.8)$$

$$E_{obs}[Q] = \sum_{k=0}^{n+1} k p_{obs}(k,1) = \sum_{k=0}^{n+1} k(d_1(n)x_1^k + d_2(n)x_2^k). \quad (5.9)$$

Computing the relevant geometric sums in (5.7) - (5.9) and substituting in (5.6) yields (5.1). ∎

Unfortunately, the very involved form of (5.1) does not allow the derivation of its maximum in closed analytic form. However, it can be numerically evaluated quite easily. Thus, we turn to numerical experiments below to derive some qualitative conclusions for the behavior of the model.

In Figure 3 we consider a model with operation parameters $(\lambda, \mu, \xi, \eta) = (7, 4, 0.4, 2)$ and reward-cost parameters $(R_s, C) = (7, 3)$ and we provide a graph of the equilibrium and social optimal thresholds for the observable case as functions of the failure compensation R_f. We observe that n_e becomes infinity for large values of R_f, while n_{soc} stabilizes to a certain value for large values of R_f. Moreover, we observe that $n_{soc} \leq n_e$ for all values of R_f. These qualitative facts seem to be valid in general, as it has been verified from a large number of similar numerical experiments for other values of the parameters.

We now turn to the unobservable case and we obtain the following Proposition 5.2.

Proposition 5.2 *Consider the unobservable model of the M/M/1 queue with catastrophes causing complete removals of customers. The expected net social benefit per time unit, given that the customers follow a mixed strategy with joining probability q (i.e. arriving customers that find an operative system enter with probability q, while the rest balk without being served) is given by*

$$S_{un}^{soc}(q) = \frac{\eta x_2(q)[\mu R_s(1 - x_2(q)) + \xi R_f - C]}{(\xi + \eta)(1 - x_2(q))}, \quad (5.10)$$

with $x_2(q)$ given by (4.3).

Proof. The expected net social benefit per time unit is given by

$$S_{un}^{soc}(q) = \lambda P_{un}^{ser} R_s + \lambda P_{un}^{cat} R_f - C E_{un}[Q], \quad (5.11)$$

where P_{un}^{ser} and P_{un}^{cat} are the fractions of customers that join but leave the system due to service and catastrophes respectively and $E_{un}[Q]$ is the mean number of customers in system. Using

(4.1), (4.2) and (3.3), we can compute P_{un}^{ser}, P_{un}^{cat} and $E_{un}[Q]$ as

$$P_{un}^{ser} = \sum_{k=0}^{\infty} p_{un}(k,1)q\left(\frac{\mu}{\mu+\xi}\right)^{k+1} = \sum_{k=0}^{\infty} \frac{\eta\left(1-x_2(q)\right)x_2(q)^k}{\xi+\eta}q\left(\frac{\mu}{\mu+\xi}\right)^{k+1}, \quad (5.12)$$

$$P_{un}^{cat} = \sum_{k=0}^{\infty} p_{un}(k,1)q\left[1-\left(\frac{\mu}{\mu+\xi}\right)^{k+1}\right]$$

$$= \sum_{k=0}^{\infty} \frac{\eta\left(1-x_2(q)\right)x_2(q)^k}{\xi+\eta}q\left[1-\left(\frac{\mu}{\mu+\xi}\right)^{k+1}\right], \quad (5.13)$$

$$E_{un}[Q] = \sum_{k=0}^{\infty} kp_{un}(k,1) = \sum_{k=0}^{\infty} k\frac{\eta\left(1-x_2(q)\right)x_2(q)^k}{\xi+\eta}. \quad (5.14)$$

Computing the geometric sums in (5.12) - (5.14) and substituting in (5.11) yields (5.10). ∎

Unlike the observable case, it is possible here to obtain the optimal joining probability q_{soc} in closed form and we can also compare it with the corresponding equilibrium probability q_e. The results are summarized in the following Theorem 5.1.

Theorem 5.1 *In the unobservable model of the $M/M/1$ queue with catastrophes causing complete removals of customers, a unique social optimal strategy exists, with joining probability q_{soc} given by*

$$q_{soc} = \begin{cases} 0 & \text{if } R_f \leq \frac{C}{\xi} - \frac{\mu R_s}{\xi} \\ \frac{\sqrt{D}\left(\mu R_s - \sqrt{D}\right)\left(\xi R_s + \sqrt{D}\right)}{\lambda D R_s} & \text{if } \frac{C}{\xi} - \frac{\mu R_s}{\xi} < R_f < \frac{C}{\xi} - \frac{\mu R_s(1-x_2)^2}{\xi} \\ 1 & \text{if } R_f \geq \frac{C}{\xi} - \frac{\mu R_s(1-x_2)^2}{\xi}, \end{cases} \quad (5.15)$$

where

$$D = \mu R_s\left(C - \xi R_f\right) \quad (5.16)$$

and $x_2 = x_2(1)$ (using (4.3) for $q = 1$). Moreover, the social optimal joining probability is always smaller than the individual one, i.e.

$$q_{soc} \leq q_e. \quad (5.17)$$

Proof. We observe that the function $S_{un}^{soc}(q)$ given by (5.10) can be written as the composition of $f(x)$ and $x_2(q)$ (i.e. $S_{un}^{soc}(q) = f(x_2(q))$), with

$$f(x) = \frac{\eta x[\mu R_s(1-x) + \xi R_f - C]}{(\xi+\eta)(1-x)} \quad (5.18)$$

and $x_2(q)$ given by (4.3). Note also that the function $x_2(q)$ is strictly increasing for $q \in [0,1]$ and it takes values in $[0, x_2]$.

To proceed, we solve the equation

$$S_{un}^{soc}{}'(q) = f'(x_2(q))x_2'(q) = 0, \quad (5.19)$$

for $q \in [0,1]$. However, $x_2'(q) \neq 0$ for $q \in [0,1]$ and therefore (5.19) is reduced to $f'(x_2(q)) = 0$. So we have to solve $f'(x) = 0$, which is also written after some straightforward algebra in the form

$$\mu R_s x^2 - 2\mu R_s x + (\mu R_s + \xi R_f - C) = 0. \quad (5.20)$$

The discriminant of the quadratic polynomial in (5.20) is non-positive if and only if $R_f \geq \frac{C}{\xi}$. In that case, we conclude that $f(x)$ is increasing and consequently $S_{un}^{soc}(q)$ is also increasing. In summary we have:

11

- Case I: $R_f \geq \frac{C}{\xi}$. The social optimal joining probability is $q_{soc} = 1$.

In case where $R_f < \frac{C}{\xi}$, the equation $f'(x) = 0$ (equivalently (5.20)) has two distinct roots x_2^- and x_2^+ given by

$$x_2^- = 1 - \frac{\sqrt{D}}{\mu R_s}, \quad x_2^+ = 1 + \frac{\sqrt{D}}{\mu R_s}, \tag{5.21}$$

with D given by (5.16). Therefore, the quadratic polynomial in (5.20) is positive for $x < x_2^-$ or $x > x_2^+$ and negative for $x_2^- < x < x_2^+$. Due to the one to one correspondence between $x_2(q)$ and q and the fact that $x_2(q) \in [0, x_2]$ for $q \in [0,1]$, we have to consider several cases regarding the relative order of x_2^-, x_2^+ and 0, x_2. However, we have that $x_2 < 1 < x_2^+$ and therefore there are only 3 subcases.

- Case II-a: $R_f < \frac{C}{\xi}$ and $x_2^- \leq 0$. Then, we have necessarily $x_2^- \leq 0 < x_2 < x_2^+$. Then $f(x)$ is decreasing in $[0, x_2]$ and consequently $S_{un}^{soc}(q)$ is decreasing in $[0,1]$. The social optimal joining probability is $q_{soc} = 0$.

- Case II-b: $R_f < \frac{C}{\xi}$ and $0 < x_2^- < x_2$. Then, we have that $f'(x)$ is positive in $(0, x_2^-)$ and negative in (x_2^-, x_2) and therefore we conclude that the maximum of $S_{un}^{soc}(q)$ is attained for q such that $x_2(q) = x_2^-$. The social optimal joining probability is found by substituting x_2^- for x in (4.8) and solving for q. We obtain

$$q_{soc} = \frac{x_2^-[\mu(1 - x_2^-) + \xi]}{\lambda(1 - x_2^-)} \tag{5.22}$$

and using (5.21) we deduce that

$$q_{soc} = \frac{\sqrt{D}\left(\mu R_s - \sqrt{D}\right)\left(\xi R_s + \sqrt{D}\right)}{\lambda D R_s}. \tag{5.23}$$

- Case II-c: $R_f < \frac{C}{\xi}$ and $x_2^- \geq x_2$. Then we have that $f(x)$ is increasing in $[0, x_2]$ and consequently $S_{un}^{soc}(q)$ is increasing in $[0,1]$. The social optimal joining probability is $q_{soc} = 1$.

Using (5.21) and taking into account the common condition $R_f < \frac{C}{\xi}$ for Cases IIa-c, we can easily see that the conditions $x_2^- \leq 0$, $0 < x_2^- < x_2$ and $x_2^- \geq x_2$ can be written respectively as $R_f \leq \frac{C}{\xi} - \frac{\mu R_s}{\xi}$, $\frac{C}{\xi} - \frac{\mu R_s}{\xi} < R_f < \frac{C}{\xi} - \frac{\mu R_s(1-x_2)^2}{\xi}$ and $R_f \geq \frac{C}{\xi} - \frac{\mu R_s(1-x_2)^2}{\xi}$. By combining Cases I-II we obtain immediately (5.15).

Regarding the order between q_{soc} and q_e, formulas (4.12) and (5.15) show that $q_e = q_{soc} = 0$ for $R_f \leq \frac{C}{\xi} - \frac{\mu R_s}{\xi}$, whereas $q_e = q_{soc} = 1$ for $R_f \geq \frac{C}{\xi} - \frac{\mu R_s(1-x_2)^2}{\xi}$. Moreover, $q_e = 1$ and $q_{soc} \in (0,1)$ when $\frac{C}{\xi} - \frac{\mu R_s(1-x_2)}{\xi} \leq R_f < \frac{C}{\xi} - \frac{\mu R_s(1-x_2)^2}{\xi}$. Thus we have to check the validity of the inequality $q_{soc} \leq q_e$, only for $R_f \in (\frac{C}{\xi} - \frac{\mu R_s}{\xi}, \frac{C}{\xi} - \frac{\mu R_s(1-x_2)}{\xi})$, i.e. in the interval where both q_e and q_{soc} are strictly between 0 and 1. Using (4.14) and (5.23), we can easily see after some algebra that the inequality $q_{soc} \leq q_e$ is reduced to $\mu^2 \xi R_s^3 + D\sqrt{D} \geq 0$ which clearly holds. Thus, the inequality is also valid in this case. ∎

The inequality $n_{soc} \leq n_e$ that the numerical experiments suggest in combination with the inequality $q_{soc} \leq q_e$ that has been analytically proved shows that we have the usual situation

also encountered in the pioneering papers of Naor (1969) and Edelson and Hildebrand (1975): Individual optimization leads to longer queues than it is socially desirable (i.e. in equilibrium the customers make excessive use of the system). Indeed, a customer that decides to join the system imposes negative externalities on future arrivals.

Another topic of interest is the comparison between the observable and the unobservable cases of a given model, i.e. what is the effect of the information on customers' behavior. The value of the information has been studied in a number of papers, among them in Hassin (1986, 2007) and Guo and Zipkin (2007). In the context of the present model, we have run several numerical scenarios and have compared the expected social profit per time unit for the observable and the unobservable cases, when the customers use their individual or social optimal strategy. More concretely, we have been interested in comparing $S_{obs}^{soc}(n_e)$, $S_{obs}^{soc}(n_{soc})$, $S_{un}^{soc}(q_e)$ and $S_{un}^{soc}(q_{soc})$. The inequalities $S_{obs}^{soc}(n_e) \leq S_{obs}^{soc}(n_{soc})$ and $S_{un}^{soc}(q_e) \leq S_{un}^{soc}(q_{soc})$ are obviously valid but the other relations are not clear. For example it would be interesting to know the relationship between $S_{un}^{soc}(q_{soc})$ and $S_{obs}^{soc}(n_e)$ which corresponds to the natural question 'what is preferable for the society: to have uninformed altruistic or informed selfish agents?'. The analysis of a large number of numerical scenarios suggests that the typical ordering is $S_{un}^{soc}(q_e) \leq S_{un}^{soc}(q_{soc}) \leq S_{obs}^{soc}(n_e) \leq S_{obs}^{soc}(n_{soc})$. In this sense, it seems that in the majority of such models it is better for the society the customers to be informed and selfish than uninformed and altruistic. Such a typical case is presented in Figure 4 for $(\lambda, \mu, \xi, \eta) = (7, 2, 0.7, 1)$ and $(R_s, C) = (7, 3)$, as R_f varies in $[0, 6]$. However, there are some exceptional cases where for low values of R_f we have that $S_{un}^{soc}(q_{soc}) \leq S_{obs}^{soc}(n_e)$, whereas for high values of R_f we have the reverse inequality. For intermediate values of R_f the situation is mixed. These cases occur typically for low values of ξ. Such a numerical scenario is presented in Figure 5 for $(\lambda, \mu, \xi, \eta) = (7, 4, 0.3, 2)$ and $(R_s, C) = (4, 3)$, as R_f varies in $[0, 10]$. Note also that all graphs of $S_{obs}^{soc}(n_e)$, $S_{obs}^{soc}(n_{soc})$, $S_{un}^{soc}(q_e)$ and $S_{un}^{soc}(q_{soc})$ with respect to R_f coincide for $R_f \geq \frac{C}{\xi}$. Indeed, for values of R_f exceeding the mean waiting cost till the next catastrophe, $\frac{C}{\xi}$, it is both individually and socially optimal for the customers to join under any kind of information.

6 Bibliography

1. Burnetas, A. and Economou, A. (2007) Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Systems* **56**, 213-228.

2. Economou, A. and Kanta, S. (2008a) Optimal balking strategies and pricing for the single server Markovian queue with compartmented waiting space. *Queueing Systems* **59**, 237-269.

3. Economou, A. and Kanta, S. (2008b) Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Operations Research Letters* **36**, 696-699.

4. Edelson, N.M. and Hildebrand, K. (1975) Congestion tolls for Poisson queueing processes. *Econometrica* **43**, 81-92.

5. Elaydi, S.N. (1999) *An Introduction to Difference Equations, 2nd Edition.* Springer, New York.

6. Guo, P. and Zipkin, P. (2007) Analysis and comparison of queues with different levels of delay information. *Management Science* **53**, 962-970.

7. Hassin, R. (1986) Consumer information in markets with random products quality: The case of queues and balking. *Econometrica* **54**, 1185-1195.

8. Hassin, R. (2007) Information and uncertainty in a queuing system. *Probability in the Engineering and Informational Sciences* **21**, 361-380.

9. Hassin, R. and Haviv, M. (1995) Equilibrium strategies for queues with impatient customers. *Operations Research Letters* **17**, 41-45.

10. Hassin, R. and Haviv, M. (1997) Equilibrium threshold strategies: the case of queues with priorities. *Operations Research* **45**, 966-973.

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

12. Mandelbaum, A. and Shimkin, N. (2000) A model for rational abandonments from invisible queues. *Queueing Systems* **36**, 141-173.

13. Naor, P. (1969) The regulation of queue size by levying tolls. *Econometrica* **37**, 15-24.

14. Stidham, S. Jr. (2009) *Optimal Design of Queueing Systems.* CRC Press, Taylor and Francis Group, Boca Raton.
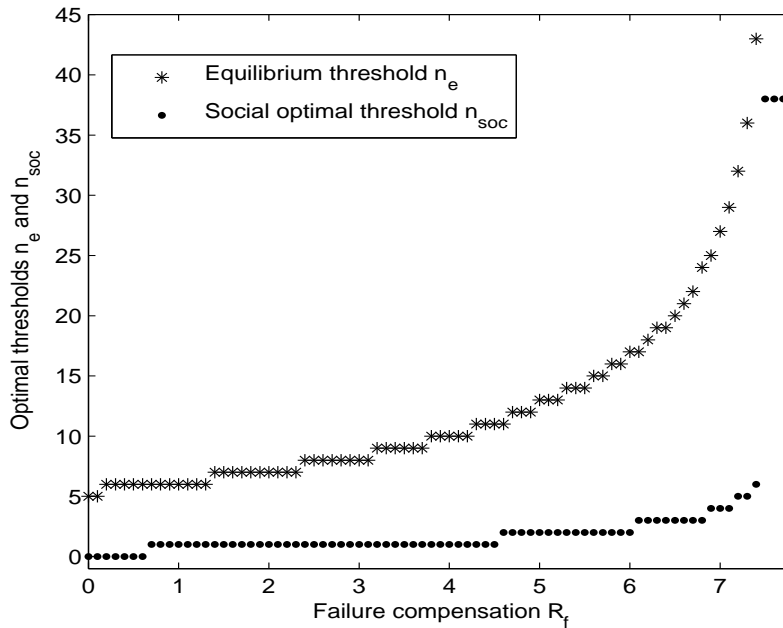
Figure 3: Equilibrium and social optimal joining thresholds with respect to R_f for the observable case with $(\lambda, \mu, \xi, \eta) = (7, 4, 0.4, 2)$ and $(R_s, C) = (7, 3)$.
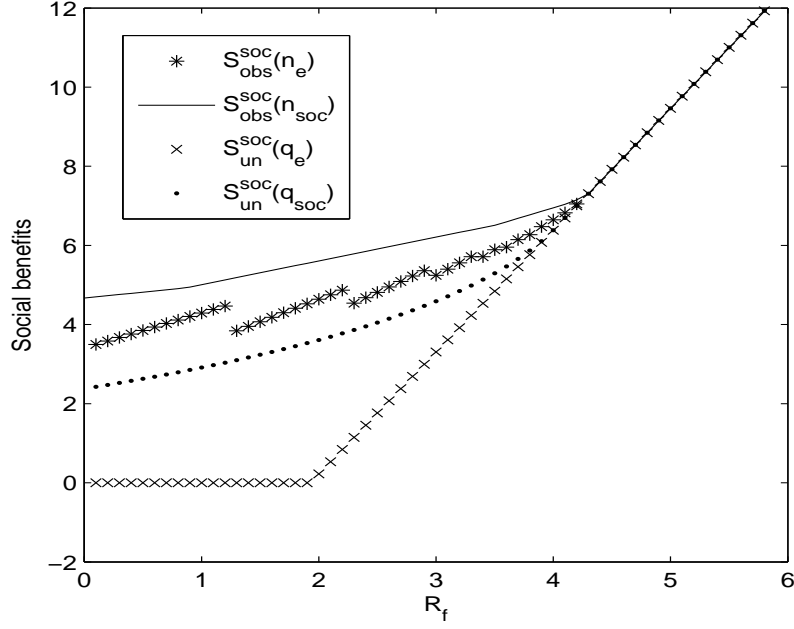
Figure 4: Social benefit per time unit with respect to R_f for a model with $(\lambda, \mu, \xi, \eta) = (7, 2, 0.7, 1)$ and $(R_s, C) = (7, 3)$.
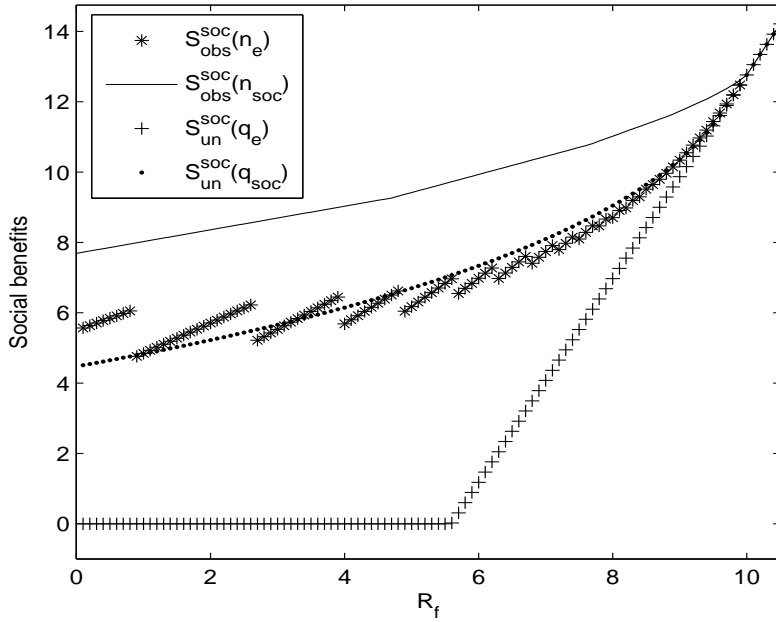

Figure 5: Social benefit per time unit with respect to R_f for a model with $(\lambda, \mu, \xi, \eta) = (7, 4, 0.3, 2)$ and $(R_s, C) = (4, 3)$.