Equilibrium customers choice between FCFS and Random servers

Refael Hassin*

July 22, 2009

Dedicated to the memory of my mother Zipora (Fella) Hassin (1924-2009)

Abstract

Consider two servers of equal service capacity, one serving in a first-come first-served order (FCFS), and the other serving its queue in random order. Customers arrive as a Poisson process and each arriving customer observes the length of the two queues and then chooses to join the queue that minimizes its expected queueing time. Assuming exponentially distributed service times, we numerically compute a Nash equilibrium in this system, and investigate the question of which server attracts the greater share of customers. If customers who arrive to find both queues empty independently choose to join each queue with probability 0.5, then we show that the server with FCFS discipline obtains a slightly greater share of the market. However, if such customers always join the same queue (say of the server with FCFS discipline) then that server attracts the greater share of customers.

1 Introduction

The subject of this paper is the analysis of a competition between two servers facing a homogeneous population of customers. The only difference between the servers is that one maintains a first-come first-served (FCFS) queue whereas the other maintains random order, meaning that the next customer to be served is randomly and uniformly selected from those present in the queue. We assume that the queues are *observable*, meaning that an arriving customer knows the length of the queue when choosing a

*School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel. hassin@post.tau.ac.il; http://www.math.tau.ac.il/~hassin/. This research was supported by the Israel Science Foundation grant no. 526/08.

server. We are interested in the way the market splits between the two servers. For simplicity we assume that no admission fee is charged by either of the servers so that customers choice between the servers is solely aimed at minimizing their expected waiting time.

The following deterministic example is illustrative: Consider N customers who need service. There are two servers that are identical in all but their service regime. One serves its customers in FCFS order, and the other server serves them in random order. The number of customers, N, is common knowledge. Both servers start operating at time 0. Before this instant, customers choose servers according to an arbitrary order. This choice is irrevocable. No new customers are expected to arrive, and therefore each server operates till its queue is empty. How will the N customers split between the two queues in a Nash equilibrium? The answer is that two thirds of the customers will join the random queue. (If $N = 3k + 1$ where k is an integer then there are two Nash equilibrium solutions, where the random server gets $2k + 1$ or $2k$ customers.) This way, the expected wait of an additional customer in either of the queues exceeds his expected wait in the queue he had chosen.

Assuming that the servers make profit from each served customer, we see that the server using the random discipline obtains a higher profit. This is because the server using the FCFS discipline is unable to charge a higher price to those who get a better position in the queue (whereas the expected welfare of every customer in the random queue is identical).

In this paper, we consider a structured model of a congested system - that of a Markovian queue managed by competing servers who aim to maximize profits. We consider a queueing system with a Poisson arrival process risk neutral customers who are identical in all but their arrival time. We denote the rate of arrival by λ. There are two servers with exponential service of rate μ each. The queue length is observable to the customers while making their decision of whether or not to join it. The two servers are identical except for that one server conducts a FCFS queue (denoted Q_F) while in the other queue (denote Q_R) the service order is random. Our main question is which of these two servers is expected to get a higher share of the demand. We note that by our assumptions, joining the shorter line (with an arbitrary tie breaking rule) is socially optimal. However, we assume self-interested customers who independently seek to maximize their own welfare.

Our model can be viewed as a noncooperative game where the players are the customers and a player's objective is to minimize his payoff defined as the time until the end of his service. (The payoff may be infinity, if the joining rate to the random queue exceeds its capacity.)

Existence of an equilibrium in this model follows from a standard argument using a fixed-point theorem. There are countably many players. A pure strategy is a function that assigns, to every possible state of the two queues, whether the player joins the ordered queue or the random queue. Since there are countably many possible states for the two queues, the space of mixed strategies is compact in the product topology.

The player's expected payoff if he joins the ordered queue depends only on the state of the ordered queue, while his payoff if he joins the random queue depends on the states of the two queues, as well as on the strategies of the other players. However, this function is continuous in the strategies of the players (when we consider the one point compactification of the set of real numbers). A standard use of a fixed-point theorem shows that a symmetric equilibrium exists. However, an analytic construction of an equilibrium solution or even a proof of uniqueness of the solution is out of reach. Therefore we concentrate on a numerical solution of the model.

Uniqueness of the threshold strategy is fundamental to enable us to draw meaningful conclusions from our numerical study. We are not able to give an analytical proof of uniqueness, and leave it as an open problem. However, we observe in our numerical study that the search for a solution converged to the same solution while starting it from different initial solutions, thus providing numerical support to the claim that the equilibrium is indeed unique.

2 Related literature

The choice of an appropriate service regime in a queueing system has been the subject of extensive research. In practice, it seems that the FCFS discipline is the natural regime when the population of potential customers is homogeneous. The main argument in favor of this regime is that it represents some sort of *fairness* (see, for example, [6, 5]). Giving priorities to certain customers over others is natural when customers differ in their service requirements or cost parameters (see [16] for a survey). Last-Come First-Served (LCFS) has been shown to have surprising advantages over FCFS when customers can observe the queue before making a decision of whether or not to join it. The reason lies in the sub-optimality of a customer's decision as first analyzed by Naor [21], caused by negative externalities associated with joining. As observed by Hassin [12] (see also [1]) these externalities are not present under LCFS.

Information, if properly used, may be useful to both customers and servers. However, as follows from the work of Naor [21] and Edelson and Hildebrand [8], customers and servers may exploit information in a way that eventually reduces their welfare. This raises the question of whether customers should be informed about the queue size or other parameters that are relevant to their decisions (such as the quality of service, and conditions of waiting). Hassin [11] shows that depending on the input parameters, it may or it may not be socially desirable to reveal the length of the queue in front of a profit maximizing server to customers who consider whether or not to join it.

Guo and Zipkin [10] consider customers whose waiting time costs per unit of time are uniformly distributed on [0,1], and three levels of delay information: no information, length of queue, and exact waiting time. They show how to compute the performance measures in the three systems. Their model has no entry pricing, and welfare increases when more accurate delay information is available.

Debo, Parlour, and Rajan [7] consider a firm that knows the quality of the service it provides but cannot credibly communicate it to its potential customers. Customers possess private imperfect information on the quality of service. There are two types of customers, differing by the quality of their private information, as expressed by the probability that it provides the correct answer. Each customer updates its beliefs after observing the length of the queue and decides whether or not to join or balk. The authors show that in general, when waiting in the queue is costly, the equilibrium behavior is not of the threshold type. Other results on the value of information in queueing systems are discussed in [2, 13, 15, 19].

Arnott, de Palma, and Lindsey [3, 4] analyze effects of information on participation and time-of-use decisions in congestible systems, when capacity and demand fluctuate. Some other papers dealing with related questions are mentioned in these references. They conclude that with appropriate price regulation of the system, added information can improve the efficiency of the system. Otherwise, the effect of information may be negative.

The appropriate regime may also be a decision variable for servers under competition. Hassin [13] shows that a server who reveals the queue length has an advantage over his competitor who hides this information (see also [2]).

We study equilibrium in threshold strategies. For a theoretical analysis of such equilibria one usually assumes some monotonicity conditions on the equilibrium payoff, and then uses the dynamic programming principle to show that if the continuation payoff satisfies these conditions then so does the solution of the dynamic programming equation. When one studies one-player problems (like bandit problems), it is possible to prove uniqueness. When one studies multi-player games, the equilibrium is usually a fixed-point of the dynamic programming equation, and so its existence is guaranteed by a fixed-point argument, that does not yield uniqueness. Uniqueness of threshold strategies has been proved for several models, some of which are described below. Note that the game associated with our queueing system is an undiscounted stochastic game with perfect information and infinite number of players and states, in which each player plays only once. Therefore, none of the results described below, which deal with discounted games, applies to our model.

Gittins and Jones [9] proved the existence of a unique threshold strategy in discounted bandit problems (one-player decision-making problems with incomplete information). Rosenberg, Solan and Vieille [23] proved that in two player bandit games, in which each player operates a two-arm bandit machine, one arm is safe and the other is risky, the risky arm has two possible types, and the types of the risky arms of the two machines are the same (and unknown), all equilibria are in cut-off strategies. In general equilibrium models, that are one-shot games, equilibrium uniqueness can be proved under proper conditions (e.g., Kehoe [17]) and tests have been devised to check whether such a model has a unique equilibrium (e.g., Kehoe and Whalley [18]). One economic model (one shot) in which uniqueness of equilibrium was proved is a model of self-fulfilling currency attacks (Morris and Shin [20]). In discounted stochastic games, uniqueness was proved in specific models. One is Pauzner and Frankel [22], where the key assumption is strategic complementarity, which does not hold in our model.

3 Equilibrium

The expected waiting time of a customer who joins the FCFS queue only depends on the length of the queue. However, for a customer who joins the random queue, the expected waiting time depends not only on the queue length there but also on the behavior of future arrivals (the more they tend to join the random queue the higher is the expected wait there). Thus, we are looking for a (Nash) equilibrium, such that each customer adopts a rule for deciding which queue to join (as a function of the queue lengths it sees upon arrival), and no one customer can become better off by changing its decision rule. We are interested here in a natural type of equilibrium, that of equilibrium threshold strategies. Note that we do not exclude the existence of equilibrium solutions of different types. For example in [14], in a model where it natural to have threshold strategies in equilibrium, it is shown that for specific data it is possible to have an non-threshold equilibrium strategy. This construction assumes specific that a choice is made when customers are indifferent to two alternative acts.

A customer who observes upon arrival that both servers are idle, is indifferent to joining either of the queues. We make the natural assumption that in this case, the customer randomly selects one of the queues with equal probabilities.

The principles of equilibrium behavior in queues are well investigated in a rich literature, and surveyed in the book [16].

In the context of our model, it is natural to consider strategies in which, given the length of Q_F, customers enter Q_R if and only if its length is less than a given constant called a *threshold*. However, it is possible that if everyone in the population has a threshold n then a deviant whose threshold is $n+1$ has a smaller expected wait, while if everyone in the population adopts a threshold of $n+1$ then deviating to a threshold of n reduces the expected wait. In such a case no pure strategy of the threshold type defines an equilibrium. Consequently, we follow [15] (see also [13]) and extend the definition of a threshold as follows. Denote by (f, r) the it state of the system, where f and r respectively denote the number of customers at the FCFS and random systems, including customers in service. Consider a customer who observes state (f, r) upon arrival.

> *Under a threshold strategy with a threshold vector* $x_f = n_f + p_f$, $f = 0, 1, \ldots$, $n_f \in I\!N$, $p_f \in [0, 1)$, *a customer always joins* Q_R *if* $r < n_f$, *always joins* Q_F *if* $r > n_f$, *and when* $r = n_f$, *the customer joins* Q_R *with probability* p_f *and* Q_F *with probability* $1 - p_f$.

If x_f are integers ($r_f = 0$ for every f), the strategy is *pure*. Otherwise, it is *mixed*.

By assumption, when $r = f = 0$, a new arrival selects each one of the servers with probability 0.5. Clearly, if $f = 0$ and $r > 0$ the new arrival joins Q_F. This means that

$x_0 = 0.5$. We also observe that if $f = 1$ then a new arrival joins Q_R if and only if $r = 0$. This means that $x_1 = 1$.

Definition 3.1 $E_q(f, r)$ - the expected *queueing time* (that is, excluding service) of a customer at Q_R, given that there are r customers in Q_R and f customers in Q_F (f and r include customers in service). This variable is defined for $r \geq 2$ and $f \geq 0$ since we assume that Q_R has at least one customer in service and another one in the queue.

The following conditions, for every $f > 0$, are necessary and sufficient for an equilibrium:

1. If $p_f > 0$ then $E_q(f, n_f + 1) = \frac{f}{\mu}$.
2. Otherwise, $E_q(f, n_f) \leq \frac{f}{\mu} \leq E_q(f, n_f + 1)$.

The first condition means that at state (f, n_f) a customer is indifferent to joining Q_F and waiting during f service periods before starting service, or joining Q_R and having expected queueing time of $E_q(f, n_f + 1)$. The second condition states that at state $(f, n_f - 1)$ one prefers to join Q_R and have expected queueing time of $E_q(f, n_f)$ rather than wait for f service completions. However, at state (f, n_f) one should join Q_F and having expected queueing time of $\frac{f}{\mu}$ rather than joining Q_R and have expected queueing time of $E_q(f, n_f + 1)$.

We now define a set of equations satisfied by the conditional expected queueing times for a given threshold strategy. We find it convenient to define $E_q(f, 1) = 0$ for every $f \geq 0$.

For $r \geq 2$:

$$(\lambda + \mu)E_q(0, r) = 1 + \mu\frac{r-2}{r-1}E_q(0, r-1) + \lambda E_q(1, r).$$

For $f > 0$ and $r \geq 2$:

$$\begin{aligned}(\lambda + 2\mu)E_q(f, r) &= 1 + \mu\left(E_q(f-1, r) + \frac{r-2}{r-1}E_q(f, r-1)\right) + \lambda\left(E_q(f+1, r)\mathbf{I}_{r>n_f}\right.\\ &+ \left. E_q(f, r+1)\mathbf{I}_{r<n_f} + \left[p_f E_q(f, r+1) + (1-p_f)E_q(f+1, r)\right]\mathbf{I}_{r=n_f}\right),\end{aligned}$$

where $\mathbf{I}$ denotes the indicator function. These equations use the fact that in state (f, r) there are $r - 1$ customers in the queue Q_R, and when service is completed, there is a probability of $\frac{1}{r-1}$ for each of them to obtain the next service.

4 Steady-state probabilities and arrival rates

Consider a given threshold strategy. Let $\pi_{f,r}$ denote the steady state probability of state (f, r). Then:

$$
\begin{aligned}
\lambda\pi_{0,0} &= \mu[\pi_{1,0} + \pi_{0,1}] \\
(\lambda+\mu)\pi_{1,0} &= \mu[\pi_{1,1} + \pi_{2,0}] + 0.5\lambda\pi_{0,0}. \\
(\lambda+\mu)\pi_{0,1} &= \mu[\pi_{1,1} + \pi_{0,2}] + 0.5\lambda\pi_{0,0}. \\
(\lambda+\mu)\pi_{f,0} &= \mu[\pi_{f,1} + \pi_{f+1,0}], \qquad f \geq 2 \\
(\lambda+\mu)\pi_{0,r} &= \mu[\pi_{1,r} + \pi_{0,r+1}] \qquad r \geq 2 \\
(\lambda+2\mu)\pi_{f,r} &= \mu[\pi_{f+1,r} + \pi_{f,r+1}] \\
&\quad +\lambda\Big(\pi_{f-1,r}\mathbf{I}_{r>n_{f-1}} + (1-p_{f-1})\pi_{f-1,r}\mathbf{I}_{r=n_{f-1}} \\
&\quad + p_f\pi_{f,r-1}\mathbf{I}_{r-1=n_f} + \pi_{f,r-1}\mathbf{I}_{r-1<n_f}\Big). \qquad f, r \geq 1 \\
\sum_{f\geq 0}\sum_{r\geq 0}\pi_{f,r} &= 1.
\end{aligned}
$$

Given an equilibrium solution, the arrival rate to Q_R is

$$
\lambda_R = \lambda \sum_{f\geq 0}\left(\sum_{r=0}^{n_f-1}\pi_{f,r} + \pi_{f,n_f}p_f\right),
$$

and the arrival rate to Q_F is $\lambda_F = \lambda - \lambda_R$. The *market share* of Q_R is defined as $\frac{\lambda_R}{\lambda}$, and similarly, the market share of Q_F is $\frac{\lambda_F}{\lambda}$.

5 Method

In our numerical study, we approximate the solution as follows. We consider a queue with a finite waiting room for each of of the servers. Thus, for a parameter K, we assume $r, f \leq K$. Our choice of $K = 120$ gave consistent results for $\rho = \frac{\lambda}{2\mu} \leq 0.96$. This value can also be justified by the following argument. Approximate the system by an $M/M/1$ queue with the same utilization factor ρ. The probability that the total number of customers exceeds $2K$ is then ρ^{2K}. Choosing $K = 120$ is sufficient to ensure that the boundary is reached with a very small probability for $\rho \leq 0.96$, since $0.96^{240} \approx 0.00005$.

We assume that arrivals who want to join a queue, but this queue is at its maximal length, are lost. Thus we obtain a finite-state system.

Starting with an arbitrary initial threshold strategy, we recursively solve approximate values for $E_q(f, r)$ for $f, r = 0, \ldots, K$. Given these values, we compute in each state the preferred action between joining Q_F and Q_R obtaining best response pure (in general) threshold strategies. Let $q(f)$ be the maximum value of r for which a customer who arrives to state (f, r) strictly prefers joining the random queue. Note that $E_q(f, q+2) > \frac{f}{\mu}$.

We define a tolerance parameter $\epsilon = 0.0001$. We assume that an equilibrium has been reached if for every given f, $q(f)+1 \le x(f) < q(f)+2$, and if $x(f) > q(f)+1+\epsilon$ then we also require $E_q(f, q+2) \le \frac{f}{\mu}+\epsilon$. If $x(f)$ does not satisfy the above requirements then we slightly modify its value towards $q(f)+1$ or $q(f)+2$, according to the case.

We repeat the process until the equilibrium requirements are satisfied for all $f = 1, \ldots, K$. We note that the threshold values $x_0 = 0.5$ and $x_1 = 1$ are maintained throughout by this process. Convergence was obtained quite rapidly.

For the equilibrium threshold strategy, we solve a finite system of $(K+1)^2$ linear equations with the same number of variables. (Note that we have given $(K+1)^2+1$ equations, including the summation to 1, but these equations are linearly dependent and one of them is removed). Its solution gives the steady state probabilities and enable computing the equilibrium arrival rates at each of the servers.

Altogether we solved the system for about 110 different values of ρ, evenly spaced between 0 and 0.96 (with higher concentration between 0.9 and 0.96). To verify uniqueness of the solution we obtained ten solutions for every given value of ρ, each time starting the computations from a different random initial solution. Specifically, we initially chose $x(f)$, for $f = 2, \ldots, K$, as a uniform random variable over $[0, f]$. In all cases the computations converged to the same solution up to five decimal points.

6 Results

Consider the static model, with no more future arrivals. Consider a customer's choice between joining Q_F given that there are already f customers there, or Q_R with r customers. In the former case he will queue for f service periods, and in the latter his expected queueing time is $1 + \frac{r-1}{2}$ service periods. The point of indifference is when $r = 2f - 1$.

Now consider the dynamic case. Recall our assumption that a customer who faces state $(0, 0)$ upon arrival randomly selects one of the queues with equal probability. Let $\rho = \frac{\lambda}{2\mu}$ denote the *utilization factor* of the system.

Consider ρ which is very small but still positive. A customer who faces state $(1, 1)$, will be the first in the queue which he chooses to join, and since the arrival rate is positive, a smaller expected waiting time is guaranteed at Q_F. For sufficiently small arrival rate, a dominant strategy is to join Q_R only if $r < 2f - 1$, meaning that $x_f = 2f - 1$. Note that given $r > 1$ customers in a random queue with arrival rate λ, the expected queueing time of any of the $r - 1$ customers in the queue is $\frac{r}{2\mu-\lambda}$. In our model, this means that if $\frac{f}{\mu} \ge \frac{r+1}{2\mu-\lambda}$ then an arriving customer prefers Q_F. Consequently, $x_f \ge 2(1-\rho)f - 1$. Moreover, since $x_f \le 2f - 1$, it follows that when $\rho < \frac{1}{2f}$, $x_f = 2f - 1$. As we can see from Figure 1 these bounds are not tight. Figure 1 contains the equilibrium thresholds for two extremes. With

$\rho = 0.001$ the solution resembles that of the static model, where $x_f = 2f - 1$ for $f \geq 1$. By assumption, customers randomize with equal probability when both servers are idle, giving $x_0 = 0.5$. With $\rho = 0.95$ we obtain different behavior, and for the initial values $f \leq x_f \leq f + 1$. (The initial threshold values for $\rho = 0.95$ are $(0.5, 1, 2, 3.072, 4.0822, 5.7707, 7, 8, 9, 10, 11, 12 \ldots)$.) Generally, $x_f = f$ for an interval of initial f values which becomes longer as ρ increases. We observe that when $\rho \to 1$ $x_f = f$ for every f, so that an arriving customer joins the short queue, and when the queue lengths are identical then he joins Q_F. Figure 1 gives the threshold functions for some other values of ρ. As expected, the thresholds are nonincreasing in ρ. It is interesting to observe that the number and location of states for which the threshold is not an integer looks arbitrary. For example, with $\rho = 0.775$ the values of $x_1, \ldots, x_{20}$ are $(1, 2.476, 4, 5, 6, 8, 9, 10, 11.44, 13, 14, 15, 16.678, 18, 19, 20, 21.761, 23, 24, 25)$.

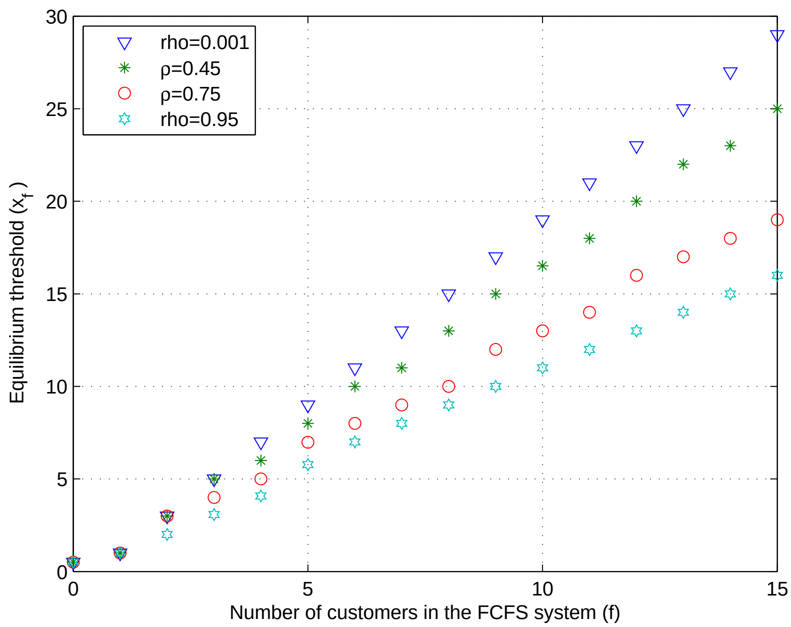

Figure 1: Equilibrium thresholds

Figure 2 shows the market share of the random server as a function of ρ. We see that the market share of the random server is always between 0.48 and 0.5, thus it is smaller, but only by a small amount, than the market share of the FCFS server.

When seeing a single customer in each queue (state $(1, 1)$) a customer in the static model is indifferent to joining each of the queues. But in the dynamic model, even with very small arrival rate, he still prefers joining Q_F. This means in particular that when ρ is sufficiently small, the market share of Q_F is strictly greater than 0.5. It is hard to tell intuitively whether the same result is expected when ρ is not very small. Our numerical results indicate that for every value of $\rho > 0$, the market share of Q_F is strictly greater than 0.5. When ρ increases from 0, the probability of state $(1, 1)$ increases, and this may explain the decrease in the market share of the random server. However, at some point, a further increase of ρ causes a decrease in $\pi_{1,1}$ and the market share increases as a function of ρ. When $\rho \to \infty$ both servers serve at their full capacity, μ customers per time unit, and, as expected, they split the market equally. The function depicted in Figure 2 is not unimodal, and the reason for this can be found in the following finding.

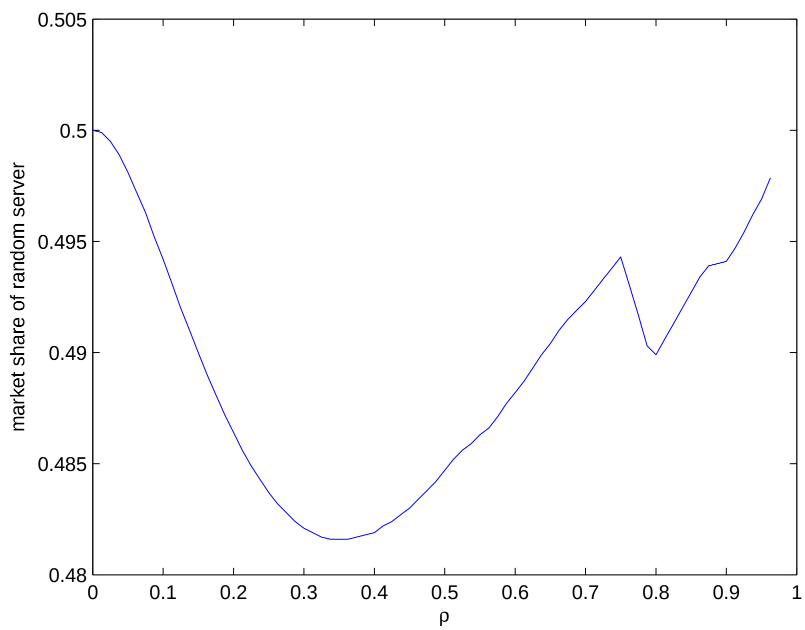

Figure 2: Equilibrium random server's market share

Figure 3 gives a detailed view of the thresholds x_2 and x_3 as functions of ρ. We see that up to approximately 0.7502, $x_2 = 3$. Then it decreases almost linearly until we again get an integer threshold, of value 2. Similar behavior is demonstrated by x_3. Comparing the data with that of the market share, we see that the interval in which x_2 decreases is exactly the interval in which the market share of the random server falls, approximately $[0.7502, 0.795]$. Similarly, the two intervals in which x_3 decrease can be observed in Figure 3 where the increase in the market share of the random server slows down.

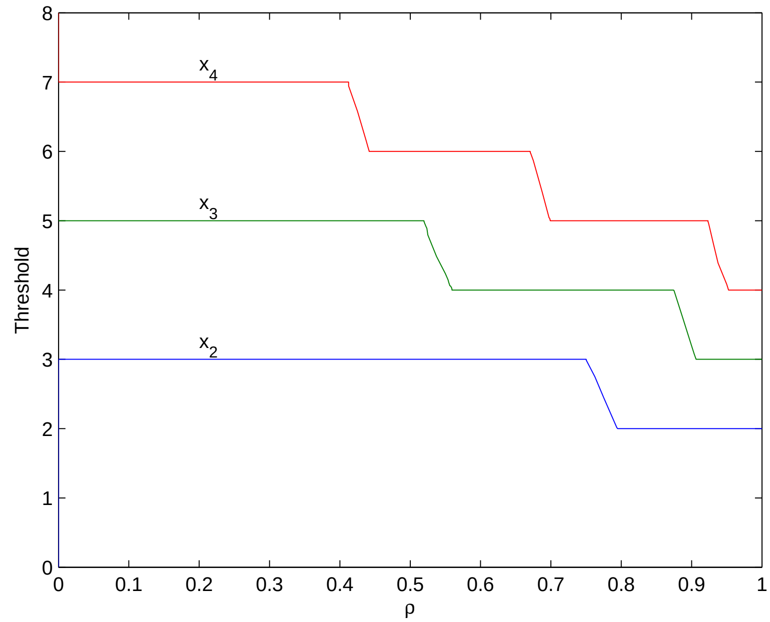

Figure 3: Thresholds x_2, x_3, x_4

We note that for $\rho < 0.7502$, the threshold values satisfy $x_i \geq i + 1$ for every $i > 1$. Thus the random server has an advantage over the FCFS server except for in the state $(1, 1)$ where customers prefer the latter. However, since the the probabilities for long queues are small for such values of ρ, the advantage at $(1, 1)$ suffices to guarantee the FCFS server more than half of the market share. When ρ increases, the FCFS server gradually gains preference over the random server for higher queue lengths as well, as demonstrated by Figure 3, and similar curves for higher queue lengths.

7 Concluding remarks

This paper opened with an example of a static model where there is a number of customers to be served and no future arrivals are expected. In this situation, the random servers gets a significantly higher share of the demand. We then concentrated on a dynamic model in which there is a stochastic process of arrivals. We see that the equilibrium strategy of customers resembles that of the static model only if the system's utilization is low. However, in this case, the probability of having customers in the queue is low, and the first person to queue prefers the FCFS, thus leading to a higher market share for the FCFS server. When the system's utilization gets higher, the equilibrium strategy changes and the server with the ordered queue is always better off. Our numerical results indicate however that the difference between the market shares is never very high, being maximized at $\rho \approx 0.354$ where the random server obtains just about 0.482 of the demand.

While considering self-interested customers who only care about their own welfare, a customer who arrives while both servers are idle is indifferent with respect to which of the queues to join. Therefore, we made the natural assumption that such a customer chooses each of the queues with equal probability. Due to the small difference in market shares under equilibrium, this assumption turns out to be crucial for our conclusion that the ordered queue obtains a larger market share. Table 1 gives the market share of the random server for some values of ρ under three models where an arrival at an empty system joins the random queue with probability 0,0.5, and 1. We see that in the latter case the market share of random server obtains more than half of the market. Note that x_f, $f \geq 1$, is independent of x_0. This is because the system will not be empty before the arriving customer completes service.

ρ	$x_0 = 0$	$x_0 = 0.5$	$x_0 = 1$
.5	.3158	.4847	.6426
.625	.3841	.4893	.5945
.75	.4317	.4943	.5556
.875	.4674	.4939	.5217
.9	.4742	.4946	.5167
.925	.4813	.4954	.5122
.95	.4885	.4967	.5087

Table 1: Market share of random server when $x_0 \in \{0, 0.5, 1\}$

As noted in the introduction, if social welfare is additive and does not discriminate among customers (i.e., only the average waiting matters, not who waits), then it is socially optimal that customers join the shorter queue, with ties arbitrarily broken. This outcome can be obtained if both servers apply FCFS service order. Moreover, since our conclusion is that the FCFS server has an advantage over the random server, obtaining a larger portion of the market share, the random server is motivated to switch to FCFS as well. We conclude that under the assumptions of our model, both servers prefer to apply FCFS, and this outcome also conforms with social welfare optimization.

Acknowledgment

The author is indebted to an anonymous referee for critical reading and constructive suggestions, and to Eilon Solan for his assistance regarding state-of-the-art results on existence and uniqueness of equilibrium solutions.

References

[1] H. Alperstein, "Optimal pricing for the service facility offering a set of priority prices," *Management Science* **34** (1988) 666-671.

[2] E. Altman, T. Jiménez, R. Núñez-Queija, and U. Yechiali (2004) "Optimal routing among $\cdot/M/1$ queues with partial information," *Stochastic Models* **20** (2004) 149-171.

[3] R. Arnott, A. de Palma, and R. Lindsey, "Information and usage of free-access congestible facilities with stochastic capacity and demand," *International Economic Review* **37** (1996) 181-203.

[4] R. Arnott, A. de Palma, and R. Lindsey, "Information and time-of-usage decisions in the bottleneck model with stochastic capacity and demand," *European Economic Review* **43** (1999) 525-548.

[5] B. Avi-Itzhak, H. Levy, and D. Raz, "Quantifying fairness in queueing systems: principles, approaches and applicability," *Probability in the Informational Sciences*, to appear.

[6] B. Avi-Itzhak, H. Levy, and D. Raz, "A Resource Allocation Queueing Fairness Measure: Properties and Bounds," *Queueing Systems Theory and Application* **56** (2007) 65-71.

[7] L. G. Debo, C. A. Parlour, and U. Rajan , "Herding and Congestion," 2005.

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

[9] J. C. Gittins and D. M. Jones, "A dynamic allocation index for the discounted multi-armed bandit problem" *Biometrika* **66** (1979) 561-565.

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

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

[12] R. Hassin, "On the optimality of first-come last-served queues," *Econometrica* **53** (1985), 201-202.

[13] R. Hassin, "On the advantage of being the first server," *Management Science* **42** (1996) 618-623.

[14] R. Hassin and M. Haviv, "Equilibrium threshold strategies: The case of queues with priorities," *Operations Research* **45** (1997) 966-973.

[15] R. Hassin and M. Haviv "Equilibrium strategies and the value of information in a two line queueing system with threshold jockeying," *Communications in Statistics - Stochastic Models* **10** (1994) 415-436.

[16] R. Hassin and M. Haviv *To Queue or not to Queue: Equilibrium Behavior in Queueing Systems,* Kluwer International Series, 2003.

[17] T. J. Kehoe, "An index theorem for general equilibrium models with production," *Econometrica* **48** (1980) 1211-1232,

[18] T.J. Kehoe and J. Whalley, "Uniqueness of equilibrium in large-scale numerical general equilibrium models," *Journal of Public Economics* **28** (1985) 247-254.

[19] C. Larsen, "Investigating sensitivity and the impact of information on pricing decisions in an $M/M/1/\infty$ queueing model," *International Journal of Production Economics* **56-57** (1998) 365-377.

[20] S. Morris and H. S. Shin "Unique equilibrium in a model of self-fulfilling currency attacks," *The American Economic Review* **88**, (1988) 587-597.

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

[22] A. Pauzner and D. Frankel, "Resolving Indeterminacy in Dynamic Settings: The Role of Shocks," *Quarterly Journal of Economics* **115** (2000) 283-304.

[23] D. Rosenberg, E. Solan and N. Vieille, "Social Learning in One-Arm Bandit Problems," *Econometrica* **75** (2007) 1591-1611.