The curse of the first-in-first-out queue discipline

by

Trine Tornøe Platz

and

Lars Peter Østerdal

Discussion Papers on Business and Economics
No. 10/2012

FURTHER INFORMATION
Department of Business and Economics
Faculty of Social Sciences
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark

Tel.: +45 6550 3271
Fax: +45 6550 3237
E-mail: lho@sam.sdu.dk
http://www.sdu.dk/ivoe

ISBN 978-87-91657-63-4

The curse of the first-in-first-out queue discipline*

Trine Tornøe Platz, Lars Peter Østerdal†
Department of Business and Economics
University of Southern Denmark

May 20, 2012

Abstract

We consider a congested facility where agents can line up at any time they wish after the facility opens (like airplane boarding, or drivers leaving stadium parking lots after a sports event). We show that in Nash equilibrium, within the general family of stochastic queue disciplines with no capacity waste, the focal first-in-first-out (FIFO) queue discipline is the *worst* while the last-in-first-out (LIFO) discipline is *best*.

1 Introduction

We model a congested facility that serves agents with a fixed capacity from a given point in time. Agents decide themselves when to line up for service, but they cannot line up before service begins. Situations that can be modeled in this way include passengers at the airport waiting to board a flight, a crowd exiting a theater area after a concert, and morning commute where drivers are only allowed to arrive at the bottleneck from a given point in time (due to environmental restrictions, say).

Our setting is related to the classical bottleneck model of Vickrey (1969), further analyzed and extended by Arnott et al. (1993), De Palma and Fosgerau (2009) and others, that models congestion arising from the existence of a single bottleneck in the context of morning commute and trip timing. The bottleneck is here open at all times, but agents have a preferred time for passing the bottleneck. In a somewhat different setting, Glazer and Hassin (1983) modeled equilibrium arrival patterns to a server with opening (and closing) time. Among the subsequent extensions and variations of this model, recent contributions include Jain et al. (2011) and Hassin and Kleiner (2011). Resemblances exist between this literature and the present paper, and in particular, we

*We thank workshop participants at the IRUC kick-off meeting at the Department of Economics, University of Copenhagen, November 21-22, 2011, for helpful comments. Jesper Breinbjerg has provided effective research assistance. Financial support from The Danish Council for Strategic Research is gratefully acknowledged.

†Correspondence to: Lars Peter Østerdal, Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M. E-mail: lpro@sam.sdu.dk

consider in line with Hassin and Kleiner (2011) a setup where early arrivals are not allowed.

The most commonly analyzed queue discipline in the literature on queueing with endogenous arrival times is the first-in-first-out (FIFO) queue discipline (also known as first-come-first-served). FIFO is generally considered as 'fair' and is the focal discipline in many everyday situations, such as queueing at a grocery store or bank, as well as under more serious circumstances such as in the allocation of donor organs to patients at the waiting list. However, while FIFO is intuitively fair and acceptable to most people, it may not be the best way of settling a queue. In fact, as we will show, in out setting it is the least desirable way of settling a queue.

In the context of Vickrey's bottleneck model, De Palma and Fosgerau (2009) consider risk-averse agents and a family of stochastic queue disciplines, ranging from FIFO to a completely random queue, which (to a vanishing degree) gives priority to early arrivals. They define a "no residual queue" property, which holds when there is no queue at the time the last user arrives at the queue, and prove that this property holds in equilibrium under all queueing regimes considered. Remarkably, they show that all queue disciplines within their family provide the same equilibrium utility and welfare. Thus, existing literature on queueing with endogenous arrival times has largely assumed FIFO discipline, or suggest that the queue discipline itself is irrelevant for equilibrium utility and welfare.

In this paper, we obtain a radically different conclusion for our model. We consider (pure strategy) Nash equilibria for a general familiy of stochastic queue disciplines with full capacity use. We show that with a linear cost for waiting in queue the FIFO is the *worst* discipline in terms of equilibrium utility and welfare, while the Last-In-First-Out (LIFO) queue disciplines is the *best*. Thus, these two queue disciplines provide an upper and lower bound for equilibrium utility/welfare under general stochastic queue disciplines.

The paper is organized as follows. In section 2, the model and key terms and assumptions are presented. Section 3 contains the results. Section 3.1 presents some preliminary results, and in section 3.2, the FIFO discipline is considered, while 3.3 concerns the LIFO discipline.

2 Model

2.1 Basics

Suppose that at time 0 a facility opens that can serve agents with a given fixed capacity. The capacity-use of each agent is assumed to be negligible, and the set of agents is identified with $[0, 1]$.

The *bottleneck capacity* at each unit of time is k. With full capacity use, all agents are therefore served at time $1/k$.

Let $R(t)$ be the *cumulative arrival distribution* (CAD) giving for each t the share of agents that have arrived at the bottheneck up until time t. We assume that $R(t)$ is left-differentiable where continuous, and denote the left-derivative of $R(t)$ as $R'_{-}(t)$, whenever defined. We will interpret the left-derivative as the rate (speed) at which agents arrive at any point in time where the CAD does not contain a jump. Figure 1 shows an example CAD.

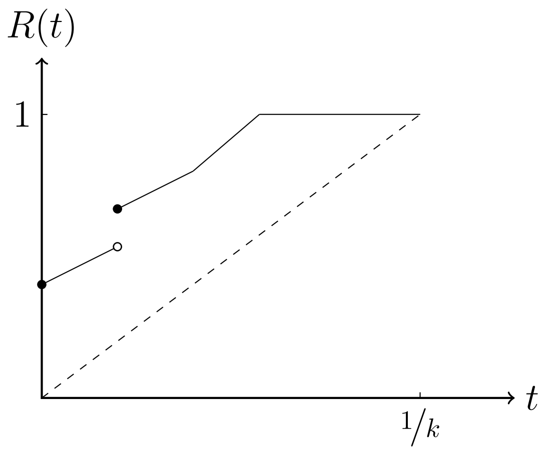

Figure 1: Example of a CAD

Let $Q(t)$ be the *backlog* at time t. We can think of $Q(t)$ as the number of agents who have joined the queue at latest at time t but remain unserved. In line with the literature, we occasionally refer to $Q(t)$ as the queue length at time t.

2.2 Agent preferences

All agents have identical preferences. Each agent wants to be served as early as possible and spend a minimum of time in the queue.

We assume that there is a *waiting cost* c for each unit of time wasted in queue. The *willingness-to-pay* for being served at time t is given by the continuous and strictly decreasing function $w(t)$.

The utility of an agent arriving at time s and being served at a time t after queuing for $t - s$ units of time is then $w(t) - c(t - s)$. We normalize $w(t)$ such that $w(1/k) = 0$. When serving time is stochastic, an agent's preferences are governed by the expected utility of the agent.

2.3 Queue discipline

The cumulative serving time distribution for an agent joining the queue at time s is denoted $T^s(t)$. For each t, $T^s(t)$ gives the cumulative probability that an agent arriving at time s has been serviced by time t. A profile of cumulative serving time distributions (one for each arrival time s) is *feasible* if the total "mass" of agents served in any interval of time does not exceed serving capacity.[1]

A *queue discipline* is a mapping that associates with a given cumulative arrival function $R(t)$ a feasible profile of cumulative serving time distributions. In other words, a queue discipline is a rule that describes when agents can expect to be served, given a specific arrival pattern.

We say that a queue discipline has *no capacity waste* if for any $R(t)$ where $R(t) \geq kt$ for all t, all agents are serviced by time $1/k$.

2.4 Equilibrium and optimality

Given a queue discipline with no capacity waste, the *expected utility* of an agent arriving at time s is denoted by $EU[T^s(t)]$. Note that the expectations exist as long as the queue discipline is terminating. Let t denote the point in time, where an agent is serviced. Then, the expected utility of an agent arriving at time s will equal:

$$EU[T^s(t)] = E[w(t) - c(t-s)]. \tag{1}$$

Let $W^s = t - s$ denote the waiting time of an agent that joins the queue at time s and is serviced at time t, and let $E[W^s]$ denote the *expected waiting time* that can be induced from the cumulative serving time distribution $T^s(t)$. We may then write:

$$EU[T^s(t)] = w(s + E[W^s]) - c(E[W^s]). \tag{2}$$

Given a queue discipline, an arrival distribution, $R(t)$, is a *(Nash) equilibrium* if no agent can unilaterally improve his expected utility by choosing another arrival time. We then say that the arrival distribution is *supported* by the queue discipline.

A queue discipline is welfare optimal if it supports an equilibrium arrival distribution that gives the highest possible expected utility of all equilibria supported by a queue discipline.

[1]Formally,

$$\sum_{\substack{t,\text{where } R(t) \text{ has jump } I_t,\\ \text{and } 0\leq t\leq y}} I_t \cdot T^t(y) + \int_0^y R'_-(t)T^t(y)dt -$$
$$\sum_{\substack{t \text{ where } R(t) \text{ has jump } I_t,\\ \text{and } 0\leq t\leq x}} I_t \cdot T^t(x) - \int_0^x R'_-(t)T^t(x)dt \leq k(y-x),$$

for any $0 \leq x < y$, Where $R'_-(t)$ denotes the left-derivative of R. Note that the specification of the cumulative serving time distribution for agents that arrive at a point in time s where arrival density is zero (i.e. where $R'_-(t) = 0$) is of no importance for feasibility. However, we may assume that they are treated the same as every other agent.

3 Results

3.1 Preliminary results

First we state some general observations about queue disciplines and equilibrium cumulative arrival distributions. Note that throughout the paper, we limit ourselves to consider queue disciplines with no capacity waste.

We are interested in comparing the welfare that arises in equilibrium under different queue disciplines. Two observations regarding comparison of welfare between equilibria under different queue disciplines are stated in the lemma below. Figure 2 provides an illustration of the type of situation considered in the second part of the lemma below.

Lemma 1. *Let $R(t) \geq kt$ be a Nash equilibrium under queue discipline 1, and let $S(t) \geq kt$ be a Nash equilibrium under queue discipline 2. Then (a): If $R(t) = S(t)$ for all t, equilibrium utility and hence welfare is the same under the two queue disciplines. (b) If $R(t) \geq S(t)$ for all t, and the inequality is strict for some time interval, then equilibrium utility is higher at $S(t)$.*

Proof. For the first part, assume that equilibrium utility is greater at $R(t)$ than $S(t)$. Then for an player arriving at s, $E[W^s]$ must be greater under $S(t)$ than $R(t)$, and this holds for all $s \in [0, \frac{1}{K}]$. However, since $R(t) = S(t)$ for all t, total waiting time is the same under both disciplines, a contradiction. For the second part: since $R(t) \geq S(t)$ for all t, total waiting time is greater at $R(t)$ than at $S(t)$. Therefore, there is an arrival time s such that the expected waiting time when arriving at s is greater for $R(t)$ than for $S(t)$. This implies that expected utility when arriving at s is lower for $R(t)$ than for $S(t)$, and equilibrium utility is therefore lower at $R(t)$. □

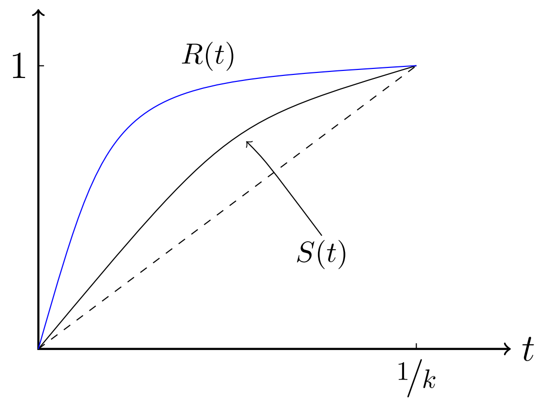

Figure 2: $R(t) \geq S(t) \geq kt$

In the following subsections we consider more specifically the FIFO and LIFO disciplines.

3.2 First-in-first-out (FIFO) queue discipline

Consider the FIFO queue discipline under which agents are served in order of arrival and as soon as capacity becomes available.

For an agent arriving at a time t where $R(t)$ has no jump, time of service is deterministic. He will be serviced as soon as every agent in the existing queue has been serviced. (Technically, the cumulative serving time distribution is a unit step function for each arrival time s). For agents arriving at a jump, time of service is uniformly distributed over an interval. Every agent already in queue must be serviced before him, and some (in expectation, half) of the agents arriving at the jump will be serviced before him as well.

Next, we turn to the analysis of Nash equilibrium in arrival patterns under FIFO.

Lemma 2. *Let $R(t)$ be a Nash equilibrium for the FIFO queue discipline. Then $R(t)$ is Lipschitz continuous.*

Proof. If $R(t)$ is supported by FIFO, the slope of $R(t)$ can not exceed k for any interval $[t_1, t_2]$, with $0 < t_1 < t_2 \leq \frac{1}{k}$. If this was the case, waiting time would be increasing with t in this interval, implying that a player with arrival time $t \in]t_1, t_2]$, could increase expected utility by arriving instead at t_1, since this would imply both a shorter waiting time and earlier service. Thus, the slope of $R(t)$ is less than k on the interval $]0, \frac{1}{k}]$, implying that $|R(t_2) - R(t_1)| \leq k|t_2 - t_1|$, for all t_1, t_2 with $0 < t_1 < t_2 \leq \frac{1}{k}$, and it follows that $R(t)$ is Lipschitz continuous. □

Note that for the FIFO discipline, although $R(t)$ has no jumps for $t > 0$, by Lemma 2, we may have $R(0) > 0$, i.e. a non-zero fraction of the agents may arrive at time 0.

Let $R'_-(t)$ denote the left-derivative of R whenever defined. Since $R(t)$ is Lipschitz continuous (and hence absolutely continuous) we have $R(t) = R(0) + \int_0^t R'_-(x)dx$.

In case not all agents arrive at time 0, the following result holds:

Lemma 3. *Suppose $R(t)$ is a Nash equilibrium for the FIFO queue discipline and $R(0) < 1$. Then the last agent that arrives is served immediately (i.e. the "no residual queue property" holds)*

Proof. Let $r = \min\{t | R(t) = 1\}$ and $R(0) < 1$, and assume that $r < \frac{1}{k}$. Then since $R(0) < 1$, and there are no jumps for $t > 0$, servicetime is deterministic and equal to $\frac{1}{k}$ for a player arriving at r. This player could therefore increase expected utility by postponing arrival until $t = \frac{1}{k}$ in which case he would be serviced at time $t = \frac{1}{k}$ while avoiding waiting time entirely. □

Thus, by no capacity waste and Lemma 3, we have $R(1/k) = 1$, and if $R(0) < 1$ we have $R(t) < 1$ for all $t < 1/k$.

Before we go any further, we would like to establish existence of an equilibrium arrival profile under FIFO.

Lemma 4. *There exists a cumulative arrival distribution supported by FIFO*

Proof. Let $R(t)$ be the CAD with $R(0) = 1$. Then, either the expected utility of arriving at time 0 is greater than the utility from arriving at $t = \frac{1}{k}$, in which case $R(t)$ is an equilibrium since no agent can profitly deviate, or the opposite is true, and $R(t)$ is not supported by FIFO.

Assume that all agents arrive at time 0 is not an equilibrium. In this case, we provide a constructive argument for the existence of an equilibrium.

Let I be the fraction of agents that would have to arrive at time 0 in order for the expected utility of these agents to equal $w(1/k)$. Note that I exists and is uniquely determined.

Note also that a player arriving immediately after time 0 will at best be serviced at time $\frac{I}{k}$. For t sufficiently small, an agent arriving in the interval from 0 to t, could therefore increase expected utility by arriving at 0 instead. This implies that the jump at 0 will be followed by a period of time where the density function vanishes, and no 'mass' of agents arrives. Next, let t^* between 0 and I/k be the point where $w(I/k) - c(I/k - t^*) = w(1/k)$, i.e., we choose t^* such that an agent obtains the same expected utility from arriving at t^* and being served at time I/k as from arriving at time 0.

For any s with $t^* \leq t \leq 1/k$, define $x(s)$ such that $s \leq s + x(s) \leq 1/k$ and

$$w(s + x(s)) - cx = w(1/k),$$

i.e., for an agent arriving at time s, $x(s)$ is the waiting time that gives the agent the same expected utility as the agents arriving at time 0. Note that $x(s)$ exists and is uniquely determined for each s. Moreover, $x(s)$ is strictly decreasing and continuous, $s + x(s)$ is strictly increasing, and $x(s) \to 0$ for $s \to 1/k$.

Now, define $R(s)$ such that $R(s) = I$ for, $0 \leq s \leq t^*$, and $R(s) = k(s + x(s))$ for $t^* < s \leq 1/k$. Then by construction $R(s)$ is an equilibrium arrival distribution function. □

In equilibrium under FIFO either every agent chooses to arrive at time 0, or some fraction of the agents arrive at time 0 followed by a period (from 0 to t^*) where the density function disappears (no arrivals) and finally a period where agents arrive smoothly until time $1/k$ where $R(t) = 1$, see Figure 3.

Figure 3 shows an example of an equilibrium arrival distribution function under FIFO. Note that for an agent arriving at time $s > t^*$, the waiting time is given by the horizontal distance between the $R(t)$ curve and the 'kt' line that shows the cumulative number of agents that has been serviced up until time t. The figure therefore also illustrates how the waiting time decreases with t (from t^*) until it reaches zero at $\frac{1}{k}$.

Next, we address the question of whether FIFO supports a unique CAD function.

Lemma 5. *Under the FIFO queue discipline there is at most one equilibrium.*

Proof. We prove this by way of contradiction. Let $R(t)$ and $S(t)$ be two distinct cumulative arrival distributions supported by FIFO, and assume that $S(0) < R(0) \leq 1$. From Lemma 3 and no capacity waste, the equilibrium utility of an agent arriving at time 0 is at least $w(\frac{1}{k})$ if $R(0) = 1$ and exactly $w(\frac{1}{k})$ if $R(0) < 1$. However, since $S(0) < R(0)$,

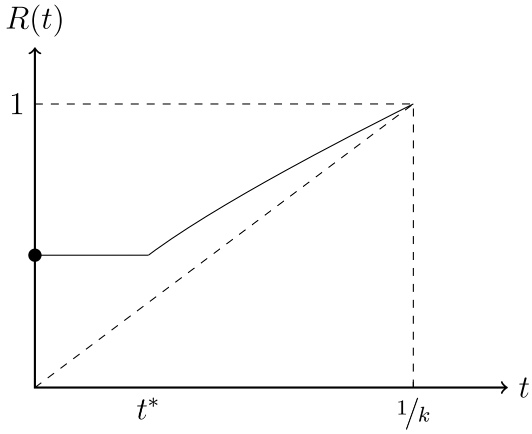

Figure 3: Equilibrium under FIFO

the expected utility of an agent arriving at 0 is greater for $S(t)$ than for $R(t)$, a contradiction. Therefore, we must have $R(0) = S(0)$. Then, given that $R(0) = S(0)$, it readily follows that t^* is the same for the two arrival profiles. Further, we know that in both cases the expected utility of every player arriving from t^* and onwards equals $w(\frac{1}{k})$, and therefore, it must be that $R(t) = S(t)$ for all t. Otherwise two players arriving at time s between t^* and $\frac{1}{k}$ would experience different waiting times, and hence, different expected utilities. $\square$

Having established that FIFO always supports a unique equilibrium CAD, we move on to state the following negative result regarding the welfare properties of the FIFO queue discipline.

Proposition 1. *Within the family of queue disciplines with no capacity waste, FIFO minimizes welfare.*

Proof. There are two cases: (a) all agent arrive at time 0, and (b) some fraction of the agents arrive at time 0 followed by a period where no-body arrives and then a period where agents arrive smoothly until time $1/k$ where the last agent arrives.

In case (a) total waiting time is the highest possible among queue disciplines with no capacity waste. Thus, no other queue-discipline can do worse, and FIFO therefore minimizes welfare among queue disciplines with no capacity waste.

In case (b), since $r = 1/k$, equilibrium utility is equal to $w(1/k)$. Under every queue discipline with no capacity waste, all players are serviced by $1/k$. Thus, a player can always choose to arrive at $t = 1/k$ and obtain utility $w(1/k)$. If, for some queue discipline $r < 1/k$, then this implies that equilibrium utility is at least $w(1/k)$, since otherwise a player arriving at r could increase expected utility by postponing arrival to $t = 1/k$. Thus, for no queue discipline can equilibrium utility be strictly lower than $w(1/k)$, and

FIFO therefore minimizes equilibrium utility among queue disciplines with no capacity waste. □

3.3 Last-in-first-out (LIFO) queue discipline

Intuitively speaking, the problem with the FIFO discipline is that the strict queue discipline gives the agents an incentive to join the queue early which in the end will hurt all agents in equilibrium. In the following, we will show that the last-in-first-out (LIFO) queue discipline is not only better than the LIFO discipline, it is in fact welfare optimal among all queue disciplines that do not waste capacity. In our setting, the LIFO queue discipline works as follows. It always gives highest priority to those agents that has arrived latest: When agents arrive continuously at a slower rate than capacity, they are all served immediately. If they arrive continuously at a higher rate than capacity, a fraction of agents are served immediately corresponding to capacity, while the rest must wait to be served until those arriving later have all been served.[2]

Before we turn to investigate equilibrium utility under LIFO, we provide some preliminary observations on equilibrium CADs and prove uniqueness and existence of an equilibrium under LIFO.

Lemma 6. *Let R be an equilibrium cumulative arrival distribution under LIFO, and let $r = \min\{t|R(t) = 1\}$. Then (a) $R(0) = 0$, (b) R is Lipschitz continuous on $[0, r[$, (c) $R'_{-}(t) > k$ for all $0 < t < r$, (d) $r < 1/k$.*

Proof. (a) In equilibrium we cannot have $R(0) > 0$, since then there is a $\varepsilon > 0$ (sufficiently small) such that an agent arriving at time 0 would be better off by arriving ε later.

(b) First, observe that R is continuous on $[0, 1/k]$, since if the CAD has a jump at time t, there is some $\varepsilon > 0$ (sufficiently small) such that an agent arriving at time t would be better off by arriving ε later. Let b be a fixed constant, where $0 < b < 1/k$, and let $\theta > 0$ such that $0 < b - \theta < b$. We show that R is Lipschitz continuous on $[0, b - \theta]$. For this, note that since R is continuous, $R'_{-}(s)$ exists for each $s \in]0, 1/k]$. It is therefore sufficient to show that $R'_{-}(s)$ is bounded. In (c) we show that $R'_{-}(t) > k$ for all $0 < t < r$, so here we focus on showing that $R'_{-}(s)$ is bounded from above, i.e., there is $K > 0$ such that $R'_{-}(s) < K$ for all $s \in [0, b - \theta]$. For this, suppose on the contrary that there is a sequence $s_1, s_2, ..$ in $[0, b - \theta]$ such that $R'_{-}(s_1) < R'_{-}(s_2) < ...$ and $R'_{-}(s_h) \to \infty$ for $h \to \infty$. Since $s_h \leq b - \theta$ an agent arriving at s_h which is not served immediately must wait for a period of time of at least θ for being served at some time after r. Since the probability of being served immediately goes to 0 as $h \to \infty$, the expected utility for an agent arriving at time s_h goes to a level below that of an agent arriving at time r (who is being served immediately with certainty). This contradicts that R is an equilibrium. Thus, $R'_{-}(s)$ is bounded from above on $[0, b - \theta]$ and the conclusion follows.

[2]Thus, in general an agent is facing a lottery over serving times with two possible outcomes, either being served immediately at arrival or later, when everyone that arrive later have been served.

(c) If $R'_{-}(t) \leq k$ for some $0 < t < r$, it means that an agent arriving at time t is served immediately with certainty and thus obtains higher expected utility than an agent arriving at time r. This contradict that R is an equilibrium CAD.

(d) Since R is Lipschitz continuous (and hence absolutely continuous) we have $R(t) = \int_0^t R'_{-}(s)ds$. By (c), the desired conclusion follows. □

Uniqueness and existence of an equilibrium is established in the following lemmas.

Lemma 7. *Under the LIFO queue discipline, there is at most one equilibrium.*

Proof. Suppose, by contradiction, that R and S are equilibrium cumulative arrival rates, $R \neq S$.

Let $r = \min\{t|R(t) = 1\}$ and $s = \min\{t|S(t) = 1\}$. We consider two cases:

(i) $r < s$

(ii) $r = s$.

(The case $r > s$ is symmetric to (i) and thus omitted).

Ad. (i): Since the agents arriving at times r and s are served immediately in the two distributions respectively, the expected utility for agents in R is greater than for the agents in S.

Let $q = \max\{t|R(t) = S(t), t < s\}$. Since $R(r) = 1 > S(r)$, and S and R are continuous functions, q is well defined. Moreover, we have $R'_{-}(q) > S'_{-}(q)$, contradicting that expected utility is higher at R, since the agents not served at time q will be served at the same later time for both R and S and the probability of being served at time q is lower in R than in S.

Ad. (ii): Since the agents arriving at time $r(= s)$ are served immediately, the expected utility for agents arriving at time $r(= s)$ is the same for both arrives profiles, and thus expected utility in equilibrium is the same for both profiles. Since $R \neq S$, $R(0) = S(0)$ and $R(r) = S(s)$ there is some t such that (a) $R(t) > S(t)$ and $R'_{-}(t) < S'_{-}(t)$ or (b) $S(t) > R(t)$ and $S'_{-}(t) < R'_{-}(t)$. If (a) then an agent arriving at t is served with higher probability at R compared to S, and if not served when arriving at t then at R the agent will be served earlier than at S since less people will arrive after. This contradicts that R and S provide the same ex ante utility. □

Lemma 8. *There exists a cumulative arrival distribution that is a Nash equilibrium under LIFO.*

Proof. Let b be a fixed constant, where $0 < b < 1/k$. Let v denote the straight line going through the points $(b, 0)$ and $(b, 1)$. Let l denote the straight line that goes through $(b, 1)$ with slope k.

Now, for each $s \in [0, 1/k]$, we define a sequence of CAD functions $Q_b^1(s), Q_b^2(s)$,... as follows.

For each $s \in [0, b]$, let $p^1(s)$ be (uniquely) determined such that the expected utility of an agent arriving at time s and being served immediately with probability $p^1(s)$ and otherwise served at time $1/k$ with probability $1-p^1(s)$ is equal to the utility of an agent arriving at time b and served immediately with certainty. Let $r^1(s) = k/p^1(s)$. (Note that if agents arrive with rate $\alpha(s) > k$, the probability of being served immediately is $k/\alpha(s)$).

Since $w(s)$ is continuous, $r^1(s)$ and $p^1(s)$ are continuous too. Note also that we have $r^1(s) \geq k$. for all s. Define $Q_b^1(s)$ on $[0, 1/k]$ such that $Q_b^1(s) = \int\limits_0^{\sigma_b(1)} r^1(t)dt$ on $[0, \sigma_b(1)]$ where $\sigma_b(1)$ is defined as the first point s where the graph of $\int\limits_0^{s} r^1(t)dt$ hits the line l, then $Q_b^1(s)$ is identified with the line l up to the point b, and $Q_b^1(s) = 1$ for $s \geq b$. By the Fundamental Theorem of Calculus $Q_b^1(s)$ is differentiable (and hence Lipschitz continuous) on $]0, \sigma_b(1)[$.

Now, we define $Q_b^h(s)$ recursively as follows.

Suppose that a CAD $Q_b^{h-1}(s)$ and a point $\sigma_b(h-1)$ has been defined such that $0 < \sigma_b(h-1) \leq b$, $Q_b^{h-1}(s)$ is differentiable on $0 \leq s < \sigma_b(h-1)$ and the derivative on this domain is greater than or equal to k, $Q_b^{h-1}(s)$ is identified with the line l for $\sigma_b(h-1) \leq s \leq b$, and $Q_b^{h-1}(s) = 1$ for $s \geq b$. For each $s \in [0, b]$, let $\beta^{h-1}(s)$ denote the point in time where the straight line from $(s, Q_b^{h-1}(s))$ with slope k meets the horisontal line connecting $(0, 1)$ and $(1/k, 1)$. Now, let $p^h(s)$ be (uniquely) determined such that the expected utility of an agent arriving at time s and being served immediately with probability $p^h(s)$ and otherwise served at time $\beta^{h-1}(s)$ with probability $1-p^h(s)$ is equal to the utility of an agent arriving at time b and served immediately with certainty. Let $r^h(s) = k/p^h(s)$. Define $Q_b^h(s)$ such that $Q_b^h(s) = \int\limits_0^{\sigma_b(h)} r^h(t)dt$, where $\sigma_b(h)$ is the first point s where the graph of $\int\limits_0^{s} r^h(t)dt$ hits the line l, and then $Q_b^h(s)$ is identified with the line l up to the point b, and $Q_b^h(s) = 1$ for $s \geq b$. Since $w(s)$ is continuous and $Q_b^{h-1}(s)$ is continuous on $[0, b[$, $r^h(s)$ and $p^h(s)$ are continuous on $[0, b]$. Note that $Q_b^h(s)$ is non-decreasing and by the Fundamental Theorem of Calculus it is differentiable (and hence Lipschitz continuous) on $]0, \sigma_b(h)[$.

Moreover, it follows from the recursive construction that $\sigma_b(1) \geq \sigma_b(2) \geq ...$, $Q_b^1(s) \leq Q_b^2(s) \leq ...$ for all $s \in [0, 1/k]$, and $r^1(s) \leq r^2(s) \leq ...$ for all $s \in [0, b]$.

We have $\lim_{h\to\infty} \sigma_b(h) > 0$, since if $\lim_{h\to\infty} \sigma_b(h) = 0$ the highest derivative of $Q_b^h(s)$ on $[0, \sigma_b(h)]$ would go to infinity as $h \to \infty$ implying that the expected utility as calculated above of a h-sequence of agents $[0, \sigma_b(h)]$ would go to zero, contradicting the construction of the sequence $Q_b^1(s), Q_b^2(s), ...$

Also, for each $0 \leq s \leq \lim_{h\to\infty} \sigma_b(h), \lim_{h\to\infty} r^h(s)$ is finite, since if $\lim_{h\to\infty} r^h(s) = \infty$ it implies that the expected utility for agents arriving at s as calculated above goes to zero, a contradiction.

In the following, let $\overline{Q}_b(s) = \lim_{h\to\infty} Q_b^h(s)$ and $\overline{\sigma}_b = \lim_{h\to\infty} \sigma_b^h(s)$

We now make the following observations:

(i) For b sufficiently close to $1/k$, $\overline{\sigma}_b < b$. This follows from observing that $Q_b^h(s)$ is increasing in b.

(ii) For b sufficiently close to 0, $\overline{\sigma}_b = b$. This follows (again) from observing that $Q_b^h(s)$ is increasing in b (since it means it decreases when b goes to zero).

(iii) $\overline{\sigma}_b$ is continuous in b. This follows since $w(s)$ is continuous and by the construction of $\overline{\sigma}_b$ and $\overline{Q}_b$.

Combining (i),(ii) and (iii), we get that there exists b such that $\overline{\sigma}_b = b$ and $\overline{Q}_b(s)$ is continuous on $[0, 1/k]$. By construction, with the CAD $\overline{Q}_b(s)$ the expected utility for an agent arriving at any time s with $\leq s < b$ is equal to the utility of an agent arriving at b (who is served immediately with certainty) and thus $\overline{Q}_b(s)$ is an equilibrium CAD under LIFO. □

Having established existence and uniqueness of LIFO, we now move on to establish the welfare properties.

Lemma 9. *Let $R(t)$ be an equilibrium arrival distribution under LIFO, and let $S(t)$ be an equilibrium under some other queue discipline that gives higher welfare. Then $s < r$, where $s = \min\{t|S(t) = 1\}$, and $r = \min\{t|R(t) = 1\}$, i.e., the latest arriving agent according to $S(t)$ arrives earlier than the latest arriving agent according to $R(t)$.*

Proof. An agent arriving at r under LIFO is served immediately. Thus, since equilibrium utility is lower under the LIFO discipline - the agent arriving at s under the alternative discipline must be served (and hence must have arrived) earlier than r, as illustrated in Figure 4 □

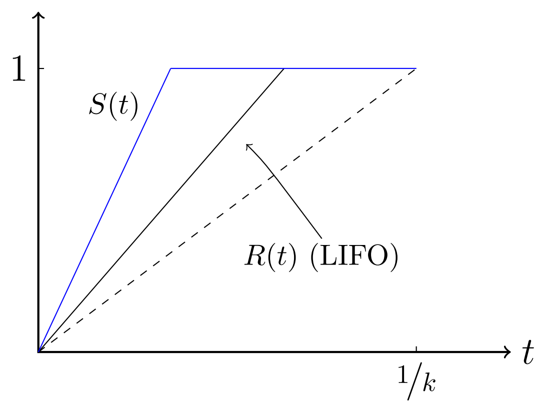

Figure 4: LIFO and possible eq. under other discipline giving higher welfare

Lemma 10. *Let $R(t)$ be an equilibrium CAD under LIFO, and let $S(t)$ be an equilibrium under some other queue discipline, where let $S(\bar{t}) = R(\bar{t})$ for some $\bar{t} < 1/k$ and $S(t) \geq R(t)$ for all $\bar{t} \leq t \leq 1/k$ with strict inequality for some interval of time. Then the equilibrium utility (and welfare) is higher under LIFO .*

Proof. Under LIFO, the agents arriving from $\bar{t}$ or later gets priority over those who arrived earlier but has not yet been served. Thus, the equilibrium utility for this group of agents must be at least as high as the equilibrium utility for the group of agents arriving from $\bar{t}$ or later under the queue priority supporting distribution $S(t)$. □

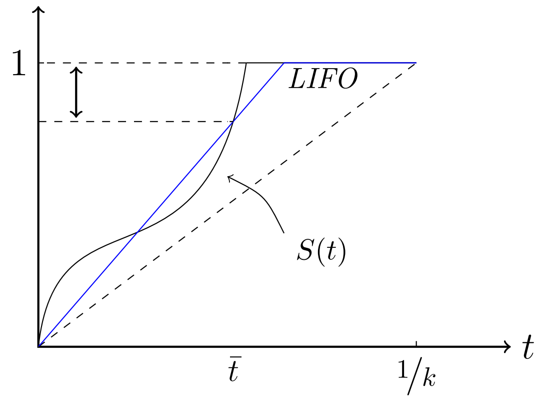

Figure 5: Higher welfare at LIFO

It follows readily from lemmas 9 and 10 that no queue discipline supports an equilibrium giving higher welfare, and we can therefore state the following proposition.

Proposition 2. *LIFO maximizes welfare.*

3.4 Welfare properties

Theorem 1. *Within the family of queue disciplines with no capacity waste, FIFO gives the lowest equilibrium utility while LIFO gives the highest equilibrium utility.*

Proof. Follows immediately from propositions 1 and 2. □

References

[1] Arnott RA, de Palma A., Lindsey R, 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *American Economic Review*, 83, 161-179.

[2] De Palma, A. and M. Fosgerau, 2009. Random queues and risk averse users. Working paper.

[3] Glazer, A. and R. Hassin (1983) ?/M/1: On the equilibrium distribution of customer arrivals. *European Journal of Operations Research*, 13, 2, 146-150.

[4] Hassin, R. and Y. Kleiner, 2011. Equilibrium and optimal arrival patters to a server with opening and closing times. IEE transactions, 43, 3, 164-175.

[5] Jain, R., Juneja, S., and N. Shimkin, 2011. The concert queueing problem: to wait or to be late. *Discrete event. Dyn. Syst.*, 21, 103-138.

[6] Vickrey, W.S, 1969. Congestion Theory and Transport Investment. *The American Economic Review*, 59, 2, 251-260.