Blind Queues: The Impact of Consumer Beliefs on Revenues and Congestion*

Shiliang Cui Senthil Veeraraghavan

{scui, senthilv}@wharton.upenn.edu

The Wharton School, University of Pennsylvania, Philadelphia PA 19104.

January 2013

Abstract

In many service settings, consumers have to join the queue without being fully aware of the parameters of the service provider (for e.g., customers at check-out counters may not know the true service rate prior to joining). In such "blind queues", consumers typically make their decisions based on the limited information about the service provider's operational parameters (from past service experiences, reviews, etc.), and the current state of the queue (the queue length). We consider a firm that serves a consumer population that may have arbitrarily misinformed beliefs about the service parameters. We show while revealing the service information to consumers improves revenues under certain consumer beliefs, it may however destroy consumer welfare or social welfare. Given a market size, the consumer welfare can be significantly reduced when a fast server announces its true service parameter. We also show that learning the service rate through sampling in blind queues leads to optimistically biased but asymptotically consistent consumer beliefs.

Keywords: Queueing Games, Service Revelation, Consumer Beliefs.

1 Introduction

Much of the literature on queues assumes that the service parameters are common knowledge and fully known to consumers when making their decisions. In reality, it is likely that only the service firm knows

*Our sincere thanks for comments and suggestions from Gerard Cachon, Noah Gans, Sergei Savin and Xuanming Su, that led to improvements in the manuscript.

its capacity and the consumers may not be fully informed of the service capacity. It is even possible that consumers could be systematically misinformed about a firm's service capacity. Hence, it is important to understand if the firm is motivated to reveal its private information on service rate to the customers, and if the firm does reveal the information, whether the information would increase consumer welfare or firm revenues. To be sure, there are papers that have focused on firms announcing (real time) delay information in terms of queue and waiting time information to its customers. However, those models also typically assume that the firm's service parameters are known to the customers.

We expect that consumers that have had limited past interactions with the service provider will not be able to accurately predict its true service rate. For instance, a customer might have visited a restaurant or an amusement park only once or twice, and it is conceivable that her best estimate of the service capacity will be based on the service times she had experienced (in the absence of other inputs). In some cases, consumers might augment their information using feedback from external acquaintances, but even such information is likely to be a smaller sample than what is needed to know the full service distribution (which is often assumed to be known accurately in the literature). In line with many real-life services, but in contrast to the existing literature, we allow for the customers to *not know* the service parameters accurately, unless they are informed about it by the firm. We term such queues as *blind queues*.

Our approach is general, i.e., individual consumers can have arbitrarily different beliefs about the service rate. It might be possible that the population is correct on average but individual customers may be misinformed in a random fashion. We also consider the possibility that the consumer population as a whole is mis-informed systematically.

In observable queues, when customers arrive with different beliefs about the true service rate, they end up with different balking behaviors based on their beliefs. As a result of his belief, a consumer may be misinformed about when to balk. For instance, a customer may join the queue, when he ought not to (if he overestimates the service rate), and an impatient misinformed customer may balk from the queue when he should not (due to underestimation). There is some recent empirical evidence (see Olivares et al. (2011) that uses queueing data from a Deli), supporting the approach that customers in observable queues may rely primarily on the length of the queue to make their purchasing/joining decisions. Thus, by understanding the impact of balking thresholds on system performance, our results could be further implemented to models where service values and waiting costs are heterogeneous.

1.1 Related Literature

The literature on queueing models with strategic customers dates back to the seminal paper of Naor (1969), who studies a single-server system with an observable queue. In Naor's model, homogeneous customers (who *know* the service parameters) observe the queue length upon arrival before making a decision to join the system. Because of homogeneity, customers have identical balking thresholds.

In our paper, consumers are not aware of the true service parameters. We allow them to have arbitrarily distributed heterogeneous beliefs over the service rate. Thus, our work is also closely related to the classical queueing papers with heterogeneous customers (with full information), in addition to those papers that deal with announcement of delay information. Following Naor (1969), queues with heterogeneous service values and time costs have been studied, as seen in the comprehensive review by Hassin and Haviv (2003).

There is a large volume of literature that examines the provision of fixed or variable *delay* information (i.e. queue length or real-time waiting time, etc.) to arriving consumers. In the context of call-centers, there are several papers that study the provision of current delay information. For instance, see Armony and Maglaras (2004a,b) and Jouini et al. (2011). We refer the reader to an excellent review by Aksin et al. (2007) of the call center literature, on the role of delay information on customers' balking behavior. Nevertheless, the service capacity and arrival information is often assumed to be known to all customers in these papers.

Hassin (1986) considers a revenue-maximizing server who may hide queue lengths to improve revenue. Whitt (1999) shows that customers are more likely to be blocked in a system where the delay information is not provided to a system where it is provided. Guo and Zipkin (2007) studies an M/M/1 queue extension with three modes of information: no information, partial information (the queue length) and full information (the exact waiting time). Economou and Kanta (2008) and Guo and Zipkin (2009) study models where some partitioned queue information (such as range of queue-lengths) is available to consumers to make their decisions. However, in all the aforementioned papers (including the no-information cases), the consumers are aware of the service rate parameter.

Thus, there are very few papers that consider server information as private information of the firm and then examine the impact of announcing this information to customers. Hassin (2007) considers an unobservable single server queuing system where the true service rate is either fast or slow, but the distribution is known to consumers. Besbes, Dooley and Gans (2011) and Debo and Veeraraghavan (2011) analyze equilibrium joining in queues with limited information on service rate. Our paper is also related to Guo et al. (2011) who studies an unobservable queue, in which partial distributional information is known to the customers, who then employ the max-entropy distribution in deciding whether to join or balk from the queues. In contrast to all of the above papers, we do not impose any distributional criterion on the consumers' knowledge on the service rate. Thus, our results do not depend on the consumer belief distribution.

Finally, our approach complements the perspective in Besbes and Maglaras (2009) and Haviv and Randhawa (2011), where the service firm does not fully know the demand (volume) information. Instead, we study a system where consumers do not know a firm's service information. However, we focus only on the decision whether the firm should reveal the unknown information to the consumers. To our knowledge, we are not aware of other papers that deal with consumer decisions when the service provider has not provided *any* information about its service parameters to the consumers. Our main theoretical contributions can be summarized as follows:

1. Consumers may have arbitrary balking thresholds due to their beliefs. We characterize those beliefs under which the firm will gain revenues from revealing its true service information.

2. We show that as consumer balking threshold beliefs become less dispersed in the population, the firm improves its revenues. If the consumer population systematically underestimates the service capacity, the service provider should always reveal its service rate information.

3. The welfare effects of information revealing are mixed. Typically, congestion (both queue lengths and wait times) increase with information revelation. Individual consumer welfare thus typically worsens with more information, especially when the server provides fast service. We find that when information is revealed, the improvement in revenues may not often overcome the consumer welfare loss, leading to reduced social welfare.

4. Our approach on blind queues is general and does not depend on the origin of the initial belief distributions, which may emerge from bounded rationality, sampling, learning from past experiences, etc. We show that sampling from finite data creates consumer optimism, but will lead to true-learning asymptotically. We show that Quantal response beliefs (bounded rational errors) can be biased, but are not consistent with learning by sampling.

The paper is structured as follows. Section 2 introduces the model and characterizes the system performance in terms of belief distributions. In Section 3, we analyze consumer populations with different beliefs for the single-server queue, and verify our results for beliefs in multi-server queues and beliefs with infinite support. In Section 4, we investigate the impact of the revelation of service information on revenues, congestion, and consumer welfare. Finally, we incorporate our belief structures into different cognitive models in the literature: quantal response errors (in §5.1), sampling (in §5.2), and limited memories (in §5.3) and conclude the paper by discussing some policy implications. All technical proofs are deferred to the appendix.

2 Model

We begin with a single-server queueing system (an extension to multi-server systems shown in §3.3 is straightforward). Consumers arrive to the queue according to a Poisson process at rate $\lambda > 0$ per unit time. The service time at the server is distributed exponentially with service rate $\mu > \lambda$. Let $\rho \triangleq \lambda/\mu$ denote the traffic intensity. Arriving customers line up at the server if the server is busy, and the queue discipline is first-come first-served (FCFS). Every arriving consumer can observe the number of the customers already waiting in the system. All consumers incur a linear waiting cost of c per unit time when they wait. The server provides a service of value v. Thus all customers are homogeneous in their valuation of the service and in their waiting costs. The firm charges a price p for its service.

Suppose customers are fully informed of the service rate μ. Upon arrival they irrevocably decide whether to join the queue based on the net value they receive from the service (i.e., $v-p$) and their expected waiting costs. For instance, a customer arriving when there are n customers in the system, joins when $v-p-(n+1)c/\mu \geq 0$, and balks otherwise. We do not consider reneging.

Model of Consumer (mis-)Information: In contrast to the existing queueing literature, we relax the assumption that customers are aware of the service time distribution or the service rate. We posit that consumers typically will not have complete information on the true distribution. For instance, consumers with their limited and idiosyncratic past interactions with the server, may have widely varying service rate beliefs.

In this paper, we use the superscript ˜ to describe the consumer beliefs about the service parameters. Consumers have heterogeneous beliefs on the service rate. (Specifically, consumer j may believe that the service rate is $\tilde{\mu}_j$, which may differ arbitrarily from the true service rate μ). We denote consumer beliefs by $\tilde{\mu} \in (0,\infty)$ with some cumulative distribution function (cdf) $G_{\tilde{\mu}}$ across the entire population. Note that every consumer has a deterministic belief. The beliefs form a random distribution because consumers with different beliefs arrive to the system randomly.

The mean of the random variable, $\tilde{\mu}$, could be the same as the true μ. In this case, we describe the population beliefs as *consistent*, i.e., the belief of the population is "correct" on average. If the mean of the random variable across the population is not the true μ, then we address the population beliefs as being biased. Specifically, if the population mean is greater (less) than μ, the beliefs are *optimistic* (*pessimistic*), i.e., the population is optimistic (pessimistic) on the service speed.

Upon arrival, consumer j with belief $\tilde{\mu}_j$ who observes n customers currently waiting in the system (including the person who is under the service, if any) makes the following decision:

$$\begin{cases} v-p \geq \frac{(n+1)c}{\tilde{\mu}_j}: & \text{customer } j \text{ joins the queue;} \\ \text{otherwise:} & \text{customer } j \text{ balks from the queue.} \end{cases}$$

We will assume $v-p \geq \frac{c}{\tilde{\mu}}$, to eliminate trivial outcomes and ensure that consumers will join an empty queue.

Balking Threshold Beliefs: For each customer j, define, $\tilde{N}_j \triangleq \lfloor \frac{\tilde{\mu}_j(v-p)}{c} \rfloor$, i.e., $\tilde{N}_j$ is an integer such that $\tilde{N}_j \leq \frac{\tilde{\mu_j}(v-p)}{c} < \tilde{N}_j+1$. Intuitively, $\tilde{N}_j$ describes the *balking threshold belief* for a consumer j: Consumer j who arrives to see n customers waiting in the system will join if $n+1 \leq \tilde{N}_j$ and balks, otherwise. Since all customers are homogeneous and differ only in their i.i.d. beliefs, we drop the subscript j, and denote the balking threshold beliefs by a random variable $\tilde{N}$. Again, note that a customer with a balking threshold belief $\tilde{N}_j$ is making a deterministic decision upon observing the queue length. The balking thresholds are random because the customers with different balking thresholds are appearing at random at the queue.

Let $F_{\tilde{N}}$ be the cdf that characterizes the random variable $\tilde{N}$. Note that customers' beliefs on the service rate are drawn from the continuous distribution $G_{\tilde{\mu}}$, whereas the balking threshold beliefs are drawn from a discrete distribution $F_{\tilde{N}}(n) = \Pr[\tilde{N} \leq n]$. Since $v - p \geq \frac{c}{\tilde{\mu}}$, we have $\tilde{N} \in \{1, 2, \ldots\}$. In essence, we translate the (uncountable) consumer beliefs on service rate to actions dictated by beliefs on balking thresholds (which is countable). Henceforth, we will focus most of our analysis on the balking threshold beliefs $\tilde{N}$. For notational convenience, we suppress the subscript $\tilde{N}$ in $F_{\tilde{N}}$ and denote $F_{\tilde{N}}$ simply as F wherever unambiguous. Our terminology on biases in beliefs (pessimism, optimism and consistency) also applies to balking threshold beliefs.

System Evolution under Threshold Beliefs: We have a population comprising of consumers who are heterogeneous in their joining behavior due to the varying individual balking threshold beliefs. Since $\tilde{N} \in \{1, 2, \ldots\}$, we have a queuing system with state-dependent arrivals - a system whose buffer size equal to the maximum balking threshold (possibly infinity, in which case we have an M/M/1 system). In contrast, note that when consumers fully know μ, we get the classical M/M/1/K system with state-independent arrivals that emerges in Naor (1969).

Let the state of system be denoted by i where i is the number of customers in the system (including the customer at the server). Since $\lambda < \mu$, this queueing system is recurrent, and long-run steady state probabilities exist. Let π_i denote the long-run probability that the system is in state i. Now consider state i: among all arrivals, only those customers who have the balking threshold greater than or equal to $i+1$ will join the queue. Thus, the effective joining probability at state i is given by $\Pr[\tilde{N} \geq i+1] = \Pr[\tilde{N} > i] = \bar{F}(i)$ (by letting $\bar{F}_{\tilde{N}}(\cdot) = 1 - F_{\tilde{N}}(\cdot)$). The effective arrival rate at any state i is $\lambda \bar{F}(i)$.

From the steady state rate balance equations, we have $\pi_{i+1} = \rho \bar{F}(i) \pi_i$ for $i \in \{0, 1, 2, ...\}$, which gives $\pi_i = \rho^i \pi_0 \prod_{n=0}^{i-1} \bar{F}(n)$ for $i \in \{1, 2, 3, ...\}$. Since $\rho < 1$, it follows from $\sum_{i=0}^{\infty} \pi_i = 1$ that

$$\pi_0 = 1 \bigg/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right). \tag{1}$$

The average number of customers in the system, L, is given by

$$L = \sum_{i=0}^{\infty} i\pi_i \quad = \quad \sum_{i=1}^{\infty} i\pi_i \quad = \pi_0 \sum_{i=1}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}(n) = \sum_{i=1}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}(n) \bigg/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right). \tag{2}$$

By convention, we set all the empty products in this paper to 1. For instance, $\prod_{n=0}^{-1} \bar{F}(n) = 1$. Then,

$$L = \sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}(n) \bigg/ \left(\sum_{i=0}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right). \tag{3}$$

Also note that, $\pi_{i+1} = \rho \bar{F}(i) \pi_i$ for $i \in \{0, 1, 2, ...\}$. Summing up over $i \geq 0$, we get,

$$\sum_{i=0}^{\infty} \pi_{i+1} = \sum_{i=0}^{\infty} \rho \bar{F}(i) \pi_i \Leftrightarrow \sum_{i=1}^{\infty} \pi_i = \rho \sum_{i=0}^{\infty} \bar{F}(i) \pi_i \Leftrightarrow (1 - \pi_0) = \rho \sum_{i=0}^{\infty} \bar{F}(i) \pi_i. \tag{4}$$

The long-run revenue rate at the server, R, is given by the product of price charged by the server and the long-run effective arrival rate at the system: $\lambda_e \triangleq \sum_{i=0}^{\infty} \pi_i \lambda \bar{F}(i)$. Thus, we have

$$R \quad = \quad p\lambda \sum_{i=0}^{\infty} \bar{F}(i)\pi_i = p\mu \cdot \rho \sum_{i=0}^{\infty} \bar{F}(i)\pi_i = p\mu(1-\pi_0) \quad \text{(from Condition (4))}. \tag{5}$$

Let W denote the average time a customer spends in the system, i.e, his waiting time in the queue plus his service time. By Little's Law,

$$W = \frac{L}{\lambda_e} = \frac{\pi_0 \sum_{i=1}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}(n)}{\mu(1-\pi_0)} \quad \text{(from conditions (2) and (5))} = \frac{1}{\mu} \frac{\pi_0}{1-\pi_0} \sum_{i=1}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}(n). \tag{6}$$

Now recall that $\pi_0 = 1 \Big/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n)\right)$, hence, $\frac{\pi_0}{1-\pi_0} = 1 \Big/ \left(\sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n)\right)$. Plugging in (6), we have

$$W = \sum_{i=1}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}(n) \Bigg/ \left(\mu \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n)\right). \tag{7}$$

Thus, as long as we can characterize $\tilde{N}$, we can immediately derive the steady state probabilities and the performance measures for the queueing system. This allows us to compare any two systems with populations that differ arbitrarily in their beliefs. To this end, in the next section, we set up a sequence of systems with consumer beliefs that are stochastically ordered in some sense. We first analyze the single-server queue in which the consumers' balking threshold beliefs are distributed over a finite interval. In other words, there is no customer in the population who has infinite patience and will always join the queue. We then consider multiple servers and beliefs with infinite support in subsequent sections.

3 Consumer Beliefs under the Lack of Service Information

When μ is fully known to consumers, the belief distribution is a one-point distribution (i.e., all consumers have identical beliefs). In contrast, the beliefs are distributed arbitrarily, when consumers are not fully informed. Recall that when there is optimism (pessimism) bias, the average balking threshold is higher (lower) than the true threshold. In §3.1, we consider belief distributions that have bias. In §3.2, we look at populations with consistent beliefs, where the average balking threshold beliefs are accurate, but there is arbitrary variability on the individual threshold joining thresholds. Our analysis in these sections assists us in pinning down the performance differences between the queueing systems which differ in consumer beliefs.

3.1 Population with Biased Beliefs

Recall that the random variable associated with balking threshold beliefs $\tilde{N}$ has a cdf F. When the threshold beliefs are finite, F takes on values over a sequence of (not necessarily consecutive) natural numbers

$\{a_1, a_2, \ldots, a_{n-1}, a_n\}$ such that $a_1 < a_2 < \ldots a_{n-1} < a_n < \infty$. We now compare a system under population beliefs $\tilde{N}$ and $\tilde{N}'$ that differ in their mean. We use first order stochastic dominance to order such random variables.

Definition 1 ***First Order Stochastic Dominance (FOSD):*** *[Quirk and Saposnik (1962)] Let $\mathcal{F}$ and $\mathcal{G}$ be the cdf's of random variables X and Y. X is said to be smaller than Y with respect to the first-order stochastic order (written $X \leq_{st} Y$) if $\mathcal{F}(t) \geq \mathcal{G}(t)$ for all real t, or equivalently, if $\bar{\mathcal{F}}(t) \leq \bar{\mathcal{G}}(t)$ for all real t.*

FOSD relation is also termed as *usual stochastic order* by Müller and Stoyan (2002), and frequently called *the stochastic order*. Variables ordered by FOSD have different means as seen in Lemma 1.

Lemma 1 ***(Theorem 1.2.9/(a) in Müller and Stoyan (2002))*** *Let $\tilde{N}$ and $\tilde{N}'$ be random variables with finite expectations. $\tilde{N} \leq_{st} \tilde{N}'$ implies $\mathbb{E}(\tilde{N}) \leq \mathbb{E}(\tilde{N}')$.*

Through Lemma 1, we can compare two threshold beliefs with respect to their 'biases'. Essentially, a threshold belief that is more pessimistic is stochastically dominated by more optimistic beliefs. We use $R_{\tilde{N}}$, $L_{\tilde{N}}$, $W_{\tilde{N}}$ ($R_{\tilde{N}'}$, $L_{\tilde{N}'}$, $W_{\tilde{N}'}$) to denote the long-run revenue (rate) at the firm, the average number of customers and the average time a customer spends in the system when the customers' balking threshold beliefs are characterized by the random variable $\tilde{N}$ (and $\tilde{N}'$ respectively). Using the results of Lemma 1, we can now compare the performance metrics, as seen in following Theorem 1.

Theorem 1 *If $\tilde{N} \leq_{st} \tilde{N}'$, then* (i) $R_{\tilde{N}} \leq R_{\tilde{N}'}$, (ii) $L_{\tilde{N}} \leq L_{\tilde{N}'}$ *and* (iii) $W_{\tilde{N}} \leq W_{\tilde{N}'}$.

When $\tilde{N} \leq_{st} \tilde{N}'$, we have $\bar{F}_{\tilde{N}} < \bar{F}_{\tilde{N}'}$, which means at any state, a smaller fraction of arrivals will join under beliefs $\tilde{N}$ (compared to $\tilde{N}'$). In other words, when $\tilde{N} \leq_{st} \tilde{N}'$ the consumers are stochastically less patient (to waiting) under $\tilde{N}$ than under $\tilde{N}'$. In the light of this observation, the conclusion on the revenue from Theorem 1/(i) becomes intuitive. Note that the revenue in the queueing system decreases in π_0. Examining equation (1), it is clear from that π_0 under $\tilde{N}$ is greater than π_0 under $\tilde{N}'$ whenever $\tilde{N} \leq_{st} \tilde{N}'$. This is because, under beliefs $\tilde{N}'$, stochastically more customers join at all states higher than 0, which reduces the visit frequency of the underlying stochastic process to state 0, and hence reduces π_0.

However, the conclusions on the average queue length (Theorem 1/(ii)) and the average waiting time (Theorem 1/(iii)) are not as immediate. For instance, examining the expression for L in equation (3), both the numerator and the denominator in the ratio for L are smaller under $\tilde{N}$ (than under $\tilde{N}'$). Since $\tilde{N}$ and $\tilde{N}'$ are arbitrarily different (except for the stochastic dominance), it is unclear when the ratio L increases or decreases. Optimism in the consumer population beliefs causes two effects. First, it leads to stochastically larger queue buffers which allow for more arrivals to join the system. Secondly, stochastically more customers join at each state, and as a consequence, the queue grows faster at each state. Thus, the expected queue length is longer. In contrast, when the threshold beliefs are pessimistic each customer expects to spend less time waiting in the system because of the shorter queues and the throughput is lower.

Note that raising the value of the service v, or lowering the charged price p, or letting the customers believe in a faster service rate ($\tilde{N}$ with higher expectation) all lead to more optimistic balking thresholds. Thus, it is intuitive that if the average customer can be made more patient, the firm would get more customers.

Finally, the results from Theorem 1 are distribution-free, i.e., the performance metrics of queueing systems can be ordered for any belief distribution, as long as the underlying belief distributions can be (first-order) stochastically ordered. Also, note that the ordering of performance metrics is invariant to the true service rate of the system.

3.2 Population with Consistent Beliefs

In this section, we consider mean-preserving spreads to examine consistent belief distributions that have the same mean (as the true belief distribution), but differ in how the individual thresholds are distributed. A specific consistent belief distribution is the "true" belief distribution which is deterministic: If all consumers knew the service rate as common knowledge, then every consumer uses the same balking threshold.

Definition 2 ***Single Mean Preserving Spread (SMPS):*** *[Rothschild and Stiglitz (1970)] Let $\mathcal{F}$ and $\mathcal{G}$ be the cdf's of two discrete random variables X and Y whose common support is a sequence of real numbers $a_1 < a_2 < ... < a_n$. Suppose the probability functions f and g describe X and Y completely:* $\Pr(X = a_i) = f_i$ *and $Pr(Y = a_i) = g_i$ where $\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} g_i = 1$. Suppose $f_i = g_i$ for all but four i, say i_1, i_2, i_3 and i_4 where $i_k < i_{k+1}$. Define $\gamma_{i_k} = g_{i_k} - f_{i_k}$. Then we say that Y differs from X by a single Mean Preserving Spread (written $\mathcal{F} \leq_{SMPS} \mathcal{G}$) if $\gamma_{i_1} = -\gamma_{i_2} \geq 0$, $\gamma_{i_4} = -\gamma_{i_3} \geq 0$ and $\sum_{k=i}^{4} a_{i_k}\gamma_{i_k} = 0$.*

The notion of mean preserving spread (MPS) is often employed to model risk ordering of two random variables that may have the same mean but different variability. If two distributions $\mathcal{F}$ and $\mathcal{G}$ describe the returns of two risky investments, and $\mathcal{F} \leq_{MPS} \mathcal{G}$, then the distribution $\mathcal{F}$ is considered less riskier. SMPS in Definition 2 is a stricter condition than MPS: $\mathcal{F} \leq_{SMPS} \mathcal{G} \Rightarrow \mathcal{F} \leq_{MPS} \mathcal{G}$.

Now consider consistent beliefs $\tilde{N}$, i.e., $\mathbb{E}[\tilde{N}]$ equals the balking threshold when the service parameters are fully known to the consumers. Since balking threshold beliefs are countable and finite, we can write the cdf F corresponding to the discrete random variable $\tilde{N}$ as follows: F has support over a finite sequence of (not necessarily consecutive) natural numbers $a_1 < a_2 < ... < a_n < \infty$, with probability mass function $f(a_i) > 0$ for all $i \in \{1, \ldots, n\}$ and $\sum_{i=1}^{n} f(a_i) = 1$.

We seek to compare the performance metrics under beliefs $\tilde{N}$ to the metrics if the true parameters of the system were known. To accomplish this, we now create a sequence of random variables that begin at the initial belief distribution under our consideration. Using a fairly general, but an intuitive construction technique, we will show that the sequence (generated using our construction) will terminate at a specific "final" distribution within finite number of steps, regardless of the starting distribution. We then characterize an ordering of performance metrics for the entire sequence. This construction allows us to not only compare the performance with the initial belief system to the canonical system with fully informed consumers, but also facilitates a comparison between any two arbitrary (consistent) belief distributions and their corresponding system performances.

Let our initial beliefs be characterized by the random variable $\tilde{N}_0$. In Construction 1, we create a sequence of random variables $\{\tilde{N}_K\}$ (term K in the sequence is distributed with the cdf F_K), and discuss the properties of the sequence. A cdf F_K corresponding to $\tilde{N}_K$ in the sequence has support over some finite sequence of natural numbers $a_{K_1} < a_{K_2} < ... < a_{K_n}$ where $K_n \geq 2$. Again, let f_K be its probability mass function such that $f_K(a_{K_i}) > 0$ for $i \in \{1, 2, ..., n\}$ and $\sum_{i=1}^{n} f_K(a_{K_i}) = 1$.

Consider the transformation of $\tilde{N}_K$ to $\tilde{N}_{K+1}$ in Construction 1. The succeeding random variable in the sequence, $\tilde{N}_{K+1}$ is constructed from the preceding random variable $\tilde{N}_K$ by taking an equal probability mass from both ends of the distribution F_K and adding those weights to the "middle" of the support.

Construction 1

$$\textit{When } a_{K_n} - 1 > a_{K_1} + 1, \quad \begin{cases} f_{K+1}(a_{K_1}) = f_K(a_{K_1}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_1}+1) = f_K(a_{K_1}+1) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(x) = f_K(x) \;\; \forall \;\; x \in \{a_{K_1}+2, a_{K_1}+3, ..., a_{K_n}-2\} \\ f_{K+1}(a_{K_n}-1) = f_K(a_{K_n}-1) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_n}) = f_K(a_{K_n}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1} = 0 \;\; \textit{otherwise.} \end{cases}$$

$$\textit{When } a_{K_n} - 1 = a_{K_1} + 1, \quad \begin{cases} f_{K+1}(a_{K_1}) = f_K(a_{K_1}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_1}+1) = f_K(a_{K_1}+1) + 2\min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_n}) = f_K(a_{K_n}) - \min\{f_K(a_{K_{K_1}}), f_K(a_{K_n})\} \\ f_{K+1} = 0 \;\; \textit{otherwise.} \end{cases}$$

Stop the sequence when $\tilde{N}_T$ *is such that* $a_{T_n} - 1 < a_{T_1} + 1$.

In Construction 1, we created the sequence $\{\tilde{N}_0, \tilde{N}_1, \ldots, \tilde{N}_T\}$ with corresponding cdf's $\{F_0, F_1, \ldots, F_T\}$ respectively. Thus, we now have a sequence of random variables that describe consumer threshold beliefs that are ordered in some sense. In the following Lemma 2, we show they are mean-preserving spreads.

Lemma 2 *Consider the sequence of random variables* $\{\tilde{N}_K\}$ *with corresponding cdf's* $\{F_K\}$. *Then,* (i) *The sequence terminates at some finite* $K = T$. (ii) $F_K \leq_{SMPS} F_{K-1}$ *for* $K \in \{1, 2, \ldots, T\}$.

Lemma 2 states that all the distributions along the sequence built through Construction 1 have the same mean (i.e., they obey the mean preserving property). As long as the first distribution is consistent, all belief distributions in Construction 1 will be consistent. Furthermore, every succeeding distribution in the sequence is dominated (under the SMPS criterion) by the preceding distribution, i.e., every distribution in the sequence is followed by a distribution that has a lower "spread" or variability.

Corollary 1 *All terms in the sequence $\{F_K\}_{K=0,1,\ldots,T}$ have the same mean $\mathbb{E}(\tilde{N}_0)$, and are ordered according to SMPS.*

Corollary 1 (proof skipped) follows immediately from Lemma 2 and the definition of SMPS. We now show in Lemma 3, that for any initial belief distribution that belongs to the family of distributions with the same mean, the sequence *always* terminates at the same distribution $\tilde{N}_T$. Depending on the parameters of the initial distribution (support etc.), the number of steps taken to reach the final distribution may differ. Thus, T depends on the distribution of $\tilde{N}_0$, but $\tilde{N}_T$ does not. In addition, we characterize this terminal distribution.

Lemma 3 *Given any $\tilde{N}_0$, the sequence $\{F_K\}$ terminates at the same F_T with the random variable $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}_0) \rfloor, \lceil \mathbb{E}(\tilde{N}_0) \rceil\}$ such that $\mathbb{E}(\tilde{N}_T) = \mathbb{E}(\tilde{N}_0)$.*

Now that we have a sequence of random variables ordered SMPS, by Construction 1, we can compare the performance metrics of the queueing system under different beliefs along the sequence. Using our notation introduced earlier, let $R_{\tilde{N}_K}$, $L_{\tilde{N}_K}$ and $W_{\tilde{N}_K}$ be the revenues, the average queue length, and the average waiting time corresponding to $\tilde{N}_K$ in the sequence of belief distributions $\{\tilde{N}_K\}$. We can now compare the performance metrics of the queues associated with the beliefs defined along the sequence in Construction 1.

Lemma 4 *Let $\{\tilde{N}_K\}$ be any sequence from Construction 1. Then* (i) *$R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$ for all ρ;* (ii) *$L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ if $\rho \leq \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1}\right)$; and* (iii) *$W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ if $\rho \leq \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}-1}{a_{K_n}-2})^2 + 4} - \frac{a_{K_n}-1}{a_{K_n}-2}\right)$.*

Recall that Construction 1 builds a sequence of consumer belief distributions with decreasing spreads, while maintaining consistency (i.e., identical means). It follows from Lemma 4/(i) that, regardless of the traffic in the system, the revenues at the firm improve as the consumers' belief distributions become less spread out (or narrower).

Note again, the revenues are in fact decreasing in idle state probability π_0. One could incorrectly surmise that the revenue ordering arises because π_0's are ordered in the sequence. However, unlike the biased beliefs case in §3.1, examining the π_0 terms along the sequence of beliefs (in Construction 1) yields no ordering. Also the result that the revenues improve when the beliefs become less spread out, is not due to Jensen's inequality.[1] Revenue improvements along the sequence emerge from the following two mechanisms: (i) The

[1]Consider belief $\tilde{N}_0$ with pdf f_0, cdf F_0 and integer $\mathbb{E}(\tilde{N}_0)$. Jensen's inequality would imply $R_{\mathbb{E}(\tilde{N}_0)} > f_0(N)R_N$ where $R_{\mathbb{E}(\tilde{N}_0)}$ and R_N are revenues when *all* consumers use the balking threshold $\mathbb{E}(\tilde{N}_0)$ and N respectively. Lemma 4 states that

consumer beliefs are gradually *altered* along the sequence in the construction which immediately changes the long-run probabilities for all states. (ii) Along the construction path, the balking threshold increases for some consumers, and decreases for some other consumers. We prove that the increased joining of the consumers with improved balking thresholds, compensates for the decreased joining of those consumers with decreased balking thresholds. This is proven for *any* belief distribution.

Using similar proof arguments, Lemma 4/(ii) and (iii) provide *distribution-free* sufficient conditions for $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$, respectively. (It is possible to derive stronger distribution-specific conditions for each inequality). Unlike revenues which always increase as beliefs become less spread out, the expected queue lengths and/or the expected waiting times can decrease. We provide numerical examples below to support this observation.

Numerical Illustration: We explore the performance metrics as the balking threshold belief $\tilde{N}_K$ is transformed into $\tilde{N}_{K+1}$ according to Construction 1. We know λ_e must increase because $R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$. Following Little's Law, it is impossible to have $L_{\tilde{N}_K} > L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ at the same time. All three other cases are possible: (i) $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$, (ii) $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} > W_{\tilde{N}_{K+1}}$, and, (iii) $L_{\tilde{N}_K} > L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} > W_{\tilde{N}_{K+1}}$. For example, consider the random variable $\tilde{N}_K \in \{3, 4, 5\}$, such that its probability mass function $f_K(x) = \{0.2, 0.6, 0.2\}$ for $x = \{3, 4, 5\}$ respectively. Following Construction 1, we have $\tilde{N}_{K+1}$ such that $Pr(\tilde{N}_{K+1} = 4) = 1$. It is also clear that $\tilde{N}_{K+1}$ is the last transformation step in Construction 1. Let $\mu = 1$ in all cases below.

Case (i): $\rho = 0.4$: $L_{\tilde{N}_K} = 0.609 < L_{\tilde{N}_{K+1}} = 0.615$ and $W_{\tilde{N}_K} = 1.551 < W_{\tilde{N}_{K+1}} = 1.562$,
Case (ii): $\rho = 0.825$: $L_{\tilde{N}_K} = 1.617 < L_{\tilde{N}_{K+1}} = 1.621$ and $W_{\tilde{N}_K} = 2.265 > W_{\tilde{N}_{K+1}} = 2.262$.
Case (iii): $\rho = 0.9$: $L_{\tilde{N}_K} = 1.793 > L_{\tilde{N}_{K+1}} = 1.790$ and $W_{\tilde{N}_K} = 2.380 > W_{\tilde{N}_{K+1}} = 2.368$. □

Having illustrated the comparative statics for the sequence of beliefs in Lemma 4, we can now compare the performance metrics of (any) initial belief with the terminal belief distribution. This is captured in Theorem 2. It turns out that when consumers beliefs become more accurate the firm always improves its revenues, while individual consumers have to wait longer if the traffic is smaller than some threshold level.

Theorem 2 *Let $\tilde{N}$ be any balking threshold belief. Let $\tilde{N}_T$ be the last term from Construction 1 initiated at $\tilde{N}_0 = \tilde{N}$. Then,* (i) $R_{\tilde{N}} < R_{\tilde{N}_T}$*;* (ii) $\exists\ \rho_L$ *s.t.* $L_{\tilde{N}} < L_{\tilde{N}_T}\ \forall\ \rho \leq \rho_L$*; and* (iii) $\exists\ \rho_W$ *s.t.* $W_{\tilde{N}} < W_{\tilde{N}_T}$ *if* $\rho \leq \rho_W$.

Theorem 2 indicates that revenue at the server always improves when beliefs become less spread-out. In the meantime, some consumers with high balking threshold thresholds become less patient, and some others with low balking thresholds become more patient. When the firm provides fast service (small ρ), the system

$R_{\mathbb{E}(\tilde{N}_0)} > R_{\tilde{N}_0} = p\mu(1-\pi_0)$ where $\pi_0 = 1 \Big/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_0(n)\right)$ (from Equation (1)) which is not equivalent to $f_0(N)R_N$.

occupancy is typically low. As a result, there is increased joining of customers with low-thresholds who became more patient in their beliefs. On the other hand, there is not much effect from the higher-threshold consumers who became less-patient as the higher states are rarely reached. Thus, when the traffic intensity ρ is small, beliefs that are *less* spread-out can increase congestion (L or W).

As beliefs become more accurate, customers wait longer and suffer higher disutility at a fast server. This result is intriguing and quite pronounced: even when the sufficient conditions do not hold, the queue-lengths (L) and waiting times (W) exhibit this property. Theorem 2 indicates that L and W can decrease only when traffic ρ is sufficiently large, as seen in the Numerical Illustration Cases (ii)-(iii).

Lower Bounds: We can use the analytical properties of the bounds along the sequence to derive *distribution-free* (w.r.t. $\tilde{N}$) bounds on ρ_L and ρ_W that hold for any arbitrary consumer belief. The lower bounds $\underline{\rho}_L$ and $\underline{\rho}_W$ are such that $\rho_L \geq \underline{\rho}_L = 0.5$ and $\rho_W \geq \underline{\rho}_W = 0.414$ respectively. We defer the details of the derivation to the appendix.

We now extend our theoretical findings to the case of a firm with multiple servers (in §3.3) and consumer beliefs that have infinite support (in §3.4).

3.3 Beliefs with Multiple Server Queues

We begin by characterizing the evolution of the queue when there are s identical servers each with service rate parameter μ ($M/M/s$ model). Assume that $s\mu > \lambda$ so the traffic $\rho = \lambda/s\mu < 1$. All other aspects of the model are the same as in the single-server setting.

Let $\tilde{N} \in \{1, 2, 3, \dots\}$, whose cdf is F, describe consumers' balking beliefs. As in the single server case, we assume that every consumer will join the system on arrival if one of the servers is idle. Thus we associate a consumer j's balking threshold belief $\tilde{N}_j \in \{1, 2, 3, \dots\}$ in the $M/M/s$ system in the following way: Consumer j with $\tilde{N}_j$, will join the $M/M/s$ system upon arrival, if and only if she observes less than $\tilde{N}_j + s - 1$ consumers already in the system.[2] This specification ensures that no-one balks when a server is idle, and is consistent with the single-server model when $s = 1$.

Let $\{0, 1, 2, \dots\}$ be the states of the $M/M/s$ system (number of consumers in the system), and $\{\pi_i : i = 0, 1, 2, \dots\}$ be the corresponding steady-state probabilities. From the rate balance equations, we have:

$$\pi_i = \begin{cases} \frac{\rho^i}{i!}\pi_0 & \text{for } i = 1, 2, \dots, s-1, s. \\ \frac{\rho^i}{s!}\prod_{n=0}^{i-s}\bar{F}(n)\pi_0 & \text{for } i = s, s+1, s+2, \dots \end{cases} \tag{8}$$

Note that when $i = s$ the two cases in Equation (8) provide the same result, i.e., $\frac{\rho^i}{i!}\pi_0 = \frac{\rho^i}{s!}\prod_{n=0}^{i-s}\bar{F}(n)\pi_0$,

[2]For example, consider consumer j with the strictest balking threshold, i.e., $\tilde{N}_j = 1$. This consumer will join the system if and only if she observes less than s ($= \tilde{N}_j + s - 1$) consumers in the system, i.e., at least one of the servers is idle.

because $\bar{F}(0) = 1$. Let $a \wedge b \triangleq \min\{a, b\}$ and let empty products, if any, be equal to 1, then from (8) we have

$$\pi_i = \frac{\rho^i}{(i \wedge s)!} \prod_{n=0}^{i-s} \bar{F}(n) \pi_0 \text{ for } i = 1, 2, 3, \ldots \tag{9}$$

From (9), we derive expressions for performance metrics for the $M/M/s$ system under the consumer beliefs $\tilde{N}$ with cdf F:

$$\pi_0 = 1 \Bigg/ \left(\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \frac{\rho^{s-1}}{s!} \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \tag{10}$$

$$R_{\tilde{N}} = p \cdot \mu [s - (s\pi_0 + (s-1)\pi_1 + \ldots + 2\pi_{s-2} + 1\pi_{s-1})] \tag{11}$$

$$L_{\tilde{N}} = \sum_{i=0}^{\infty} i\pi_i = \sum_{i=0}^{\infty} \frac{i\rho^i}{(i \wedge s)!} \prod_{n=0}^{i-s} \bar{F}(n) \Bigg/ \sum_{i=0}^{\infty} \frac{\rho^i}{(i \wedge s)!} \prod_{n=0}^{i-s} \bar{F}(n) \tag{12}$$

$$W_{\tilde{N}} = \frac{L_{\tilde{N}}}{\lambda_e} = \sum_{i=1}^{\infty} \frac{i\rho^i}{(i \wedge s)!} \prod_{n=0}^{i-s} \bar{F}(n) \Bigg/ \mu \sum_{i=1}^{\infty} (i \wedge s) \frac{\rho^i}{(i \wedge s)!} \prod_{n=0}^{i-s} \bar{F}(n) \tag{13}$$

Note that all expressions from (10) to (13) coincide with the corresponding expressions for the $M/M/1$ system when $s = 1$. We now recover the results of Theorem 1 and Theorem 2 for consumer beliefs in multi-server queues. The proofs can be found in the appendix.

Theorem 1' *Consider consumer beliefs $\tilde{N}$ and $\tilde{N}'$ at an $M/M/s$ queue. If $\tilde{N} \leq_{st} \tilde{N}'$, then* (i) $R_{\tilde{N}} \leq R_{\tilde{N}'}$, (ii) $L_{\tilde{N}} \leq L_{\tilde{N}'}$ *and* (iii) $W_{\tilde{N}} \leq W_{\tilde{N}'}$.

Theorem 2' *Let $\tilde{N}$ be any balking threshold beliefs distribution for the $M/M/s$ queue. Let $\tilde{N}_T$ be the last term from Construction 1 initiated at $\tilde{N}_0 = \tilde{N}$. Then,* (i) $R_{\tilde{N}} < R_{\tilde{N}_T}$; (ii) $\exists\, \rho_L$ *s.t.* $L_{\tilde{N}} < L_{\tilde{N}_T}$ $\forall\, \rho \leq \rho_L$; *and* (iii) $\exists\, \rho_W$ *s.t.* $W_{\tilde{N}} < W_{\tilde{N}_T}$ *if* $\rho \leq \rho_W$.

3.4 Beliefs over an Infinite Support

We further relax the assumption that the balking threshold beliefs have finite support. Recall that Theorem 1 in §3.1 states that when consumers have two sets of balking threshold beliefs $\tilde{N}$ and $\tilde{N}'$ (which have finite supports) such that $\tilde{N} \leq_{st} \tilde{N}'$, then $R_{\tilde{N}} \leq R_{\tilde{N}'}$, $L_{\tilde{N}} \leq L_{\tilde{N}'}$ and $W_{\tilde{N}} \leq W_{\tilde{N}'}$. It is clear that the approach used in the proof of Theorem 1 continues to hold when $\tilde{N}$ and $\tilde{N}'$ have infinite support. Thus we have the following theorem.

Theorem 1" *Consider consumer beliefs $\tilde{N}$ and $\tilde{N}'$ that may be distributed on an infinite support. If $\tilde{N} \leq_{st} \tilde{N}'$, then* (i) $R_{\tilde{N}} \leq R_{\tilde{N}'}$, (ii) $L_{\tilde{N}} \leq L_{\tilde{N}'}$ *and* (iii) $W_{\tilde{N}} \leq W_{\tilde{N}'}$.

However, to extend Theorem 2 to the infinite support case, we need additional preparatory groundwork. Recall that for consistent beliefs $\tilde{N}$ with finite support, Theorem 2 states that $R_{\tilde{N}} < R_{\tilde{N}_T}$ for all ρ, $L_{\tilde{N}} < L_{\tilde{N}_T}$

and $W_{\tilde{N}} < W_{\tilde{N}_T}$ for small ρ, where $\tilde{N}_T$ is the terminal distribution in the sequence formed by Construction 1 initiated from $\tilde{N}$. Lemma 3 shows that $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}) \rfloor, \lceil \mathbb{E}(\tilde{N}) \rceil\}$ such that $\mathbb{E}(\tilde{N}_T) = \mathbb{E}(\tilde{N})$.

Now we relax the assumption on the finite support and allow $\tilde{N}$ to take values from a countable set within $\{1, 2, 3, \ldots\}$ with finite mean $\mathbb{E}(\tilde{N})$. Our approach is to develop a new sequence of random variables $\tilde{N}^K$, in Construction 2, each with finite support and mean $\mathbb{E}(\tilde{N})$. We then show that the sequence $\{\tilde{N}^K\}$ converges to $\tilde{N}$ in probability, i.e., $\tilde{N}^K \xrightarrow{p} \tilde{N}$.

Construction 2 *Let* $\tilde{N} \in \{1, 2, 3, \ldots\}$ *with* $\mathbb{E}(\tilde{N}) < \infty$. *For each* K, *let* $\tilde{N}^K$ *be a random variable that takes values only in* $\{1, 2, \ldots, K-1, K\} \cup \{\lfloor \mathbb{E}[\tilde{N}|\tilde{N} > K] \rfloor, \lceil \mathbb{E}[\tilde{N}|\tilde{N} > K] \rceil\}$ *such that*

$$\text{When } \mathbb{E}[\tilde{N}|\tilde{N} > K] \text{ is an integer,} \quad \begin{cases} \Pr(\tilde{N}^K = n) = \Pr(\tilde{N} = n) \text{ for } n \in \{1, 2, \ldots, K-1, K\} \\ \Pr(\tilde{N}^K = \mathbb{E}[\tilde{N}|\tilde{N} > K]) = \Pr(\tilde{N} > K) \end{cases}$$

$$\text{Otherwise,} \quad \begin{cases} \Pr(\tilde{N}^K = n) = \Pr(\tilde{N} = n) \text{ for } n \in \{1, 2, \ldots, K-1, K\} \\ \Pr(\tilde{N}^K = \lfloor \mathbb{E}[\tilde{N}|\tilde{N} > K] \rfloor) = (\lceil \mathbb{E}[\tilde{N}|\tilde{N} > K] \rceil - \mathbb{E}[\tilde{N}|\tilde{N} > K]) \Pr(\tilde{N} > K) \\ \Pr(\tilde{N}^K = \lceil \mathbb{E}[\tilde{N}|\tilde{N} > K] \rceil) = (\mathbb{E}[\tilde{N}|\tilde{N} > K] - \lfloor \mathbb{E}[\tilde{N}|\tilde{N} > K] \rfloor) \Pr(\tilde{N} > K) \end{cases}$$

Construction 2 replaces the tail of the distribution of $\tilde{N}$ (the portion where $\tilde{N} > K$), with a single or two finite probability mass points that take on the corresponding conditional mean $\mathbb{E}[\tilde{N}|\tilde{N} > K]$. It then can be easily verified that $\mathbb{E}(\tilde{N}^K) = \mathbb{E}(\tilde{N})$. Thus Construction 2 provides a mean-preserving transformation, i.e., the consistency of beliefs is preserved.

Let $\{\tilde{N}_K\}_{K=1,2,3,\ldots}$ be built from $\tilde{N}$ using Construction 2. Note that $\Pr(\tilde{N} \neq \tilde{N}^K) \leq \Pr(\tilde{N} > K)$ and $\lim_{K \to \infty} \Pr(\tilde{N} > K) = 0$. So $\{\tilde{N}^K\}$ converges to $\tilde{N}$ in probability which also implies the convergence in distribution. It immediately follows that $\lim_{K \to \infty} R_{\tilde{N}^K} = R_{\tilde{N}}$, $\lim_{K \to \infty} L_{\tilde{N}^K} = L_{\tilde{N}}$, $\lim_{K \to \infty} W_{\tilde{N}^K} = W_{\tilde{N}}$. For each K, $\tilde{N}^K$ is a random variable with mean $\mathbb{E}(\tilde{N})$ and a finite support. By Theorem 2 (for the finite support case), we have (for each K), $R_{\tilde{N}^K} < R_{\tilde{N}_T}$ for all ρ, $L_{\tilde{N}^K} < L_{\tilde{N}_T}$ and $W_{\tilde{N}^K} < W_{\tilde{N}_T}$ for small ρ. By letting $K \to \infty$, we can extend Theorem 2 to the case when $\tilde{N}$ takes an infinite support.

Theorem 2" *Let* $\tilde{N}$ *be any balking threshold beliefs distribution that may have an infinite support. Let* $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}) \rfloor, \lceil \mathbb{E}(\tilde{N}) \rceil\}$ *such that* $\mathbb{E}(\tilde{N}_T) = \mathbb{E}(\tilde{N})$. *Then,* (i) $R_{\tilde{N}} \leq R_{\tilde{N}_T}$; (ii) $\exists\, \rho_L$ *s.t.* $L_{\tilde{N}} \leq L_{\tilde{N}_T}\ \forall\, \rho \leq \rho_L$; *and* (iii) $\exists\, \rho_W$ *s.t.* $W_{\tilde{N}} \leq W_{\tilde{N}_T}$ *if* $\rho \leq \rho_W$.

Thus, we recover the conclusions from Theorems 1 and 2, when applying our results to beliefs that have infinite supports. The lower bounds derived on $\underline{\rho}_L$ and $\underline{\rho}_W$ in §3.2 also continue to hold.

4 Impact of Revealing Service Information

So far, we discussed the revenue and congestion effects of consumer beliefs, when only the firm knows its service parameters. Now, we address whether a firm should reveal service information, i.e., its true service rate μ, and then calibrate the impact of revealing the true information. To begin with, when consumers are uninformed, they may have arbitrary beliefs over the service rate, and the balking thresholds could be distributed according to some $\tilde{N}$. When the firm chooses to inform its consumers of the true service rate, the consumers will follow identical balking thresholds, say N.[3] Since the native beliefs are arbitrary, it is unclear when a firm should reveal its service rate. We characterize this condition in the following Proposition 1.

Proposition 1 *In an $M/M/s$ queue, when $\mathbb{E}(\tilde{N}) \leq N$, the firm benefits from revealing service information ($R \uparrow$). In addition, when traffic ρ is small, the average queue length and the average waiting time for a customer both increase on announcement ($L, W \uparrow$).*

From Proposition 1, we find that when consumer balking beliefs are pessimistic, i.e., $\mathbb{E}(\tilde{N}) \leq N$, it is always in the firm's interest to reveal its service rate (the firm sees more revenue as the announcement is made). When ρ is small, the system congestion (the average queue length and the average customer waiting time) increases on consumers knowing the true information.

These results are distribution-free w.r.t. $\tilde{\mu}$ and also parameter-free w.r.t. μ. It is sufficient for the firm to only know that the beliefs are pessimistic or consistent, before the decision to reveal true information is made. Under such cases, the exact distribution of beliefs do not influence the decision to reveal information.

To intuit this result, we first consider a population with beliefs $\tilde{N}$. Let $\{\tilde{N}_0, \tilde{N}_1, \ldots, \tilde{N}_T\}$ be a sequence of balking threshold beliefs from Construction 1 starting with $\tilde{N}_0 = \tilde{N}$. (We can use a similar argument using Construction 2 for beliefs with infinite support). Theorem 2 states that $R \uparrow$ and $L, W \uparrow$ (for small ρ) when consumers adopt $\tilde{N}_T$ instead of $\tilde{N}$. Now suppose the beliefs are pessimistic or consistent, i.e., $\mathbb{E}(\tilde{N}) \leq N$, it then follows that $\tilde{N}_T \leq_{st} N$. So by Theorem 1, on revealing, we have $R \uparrow$ and $L, W \uparrow$. Thus, combining Theorems 1 and 2, when $\mathbb{E}(\tilde{N}) \leq N$, we have $R \uparrow$ and $L, W \uparrow$ (for small ρ).

Now suppose that consumers have optimism bias ($N < \mathbb{E}(\tilde{N})$). Again, we construct the sequence $\{\tilde{N}_0, \tilde{N}_1, \ldots, \tilde{N}_T\}$ using Construction 1 starting with $\tilde{N}_0 = \tilde{N}$. We have $R_{\tilde{N}} < R_{\tilde{N}_1} < \ldots < R_{\tilde{N}_T}$ by Theorem 2. On the other hand, $N < \mathbb{E}(\tilde{N})$ implies $N <_{st} \tilde{N}_T$, so by Theorem 1 we have $R_N < R_{\tilde{N}_T}$. Depending on $\tilde{N}$, we may have $R_N < R_{\tilde{N}}$ or $R_N > R_{\tilde{N}}$. Recall that $\{\tilde{N}_0, \tilde{N}_1, \ldots, \tilde{N}_T\}$ is a sequence of beliefs that have progressively lower spreads. So we conclude that when consumers population is optimistic, the firm may still reveal its service information as long as it observes high variance in consumers' balking behaviors. We provide numerical examples to support this observation in the following section.

[3] For example, in the $M/M/1$ queue, the balking threshold beliefs are given by $\tilde{N} = \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor$ and true balking threshold is given by $N = \lfloor \frac{\mu(v-p)}{c} \rfloor$. Similar conclusions hold for the $M/M/s$ queues.

4.1 Revenue and Welfare Effects of Service Revelation under Bias

We now examine specific cases where our findings will apply by studying $M/M/1$ queues under different beliefs in the population. In all cases, we set the service value $v = \$8$, price $p = \$2$, and consumer linear waiting cost at $c = \$4/\text{min}$. Consumers are not aware of the provider's true service rate, μ and their beliefs are uniformly distributed over $[2, 8]$, i.e., $\tilde{\mu} \sim U[2, 8]$ (with mean 5). As a result, consumers' balking threshold beliefs, $\tilde{N} = \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor = \lfloor 3\tilde{\mu}/2 \rfloor$, is a discrete uniform distribution taking values $\{3, 4, \ldots, 11\}$ with $\mathbb{E}(\tilde{N}) = 7$.

We examine three scenarios where the true threshold N from announcing the true service rate is (i) greater than, (ii) equal to or (iii) less than $\mathbb{E}(\tilde{N})$. These three instances correspond to pessimism, consistency and optimism in beliefs. For each case, we examine the firm's revenue, the average queue length and the consumer average waiting time, as well as consumer welfare and social welfare.

The first line in each table that follows corresponds to the situation when the firm hides the service information from its customers (customers adopt balking threshold beliefs $\tilde{N}$); the last line of each table corresponds to the situation when the firm reveals the service information to its customers (customers thus adopt balking threshold beliefs N). All rows in between the first and the last rows communicate the terms in the sequence in Construction 1. The percentage change in a parameter (compared to the original beliefs $\tilde{N}$, first line) is noted in parenthesis.

Pessimistic Beliefs: The arrival rate is $\lambda = 5/\text{min}$ and the true service rate is $\mu = 6/\text{min}$. Note that $\mu = 6 > \mathbb{E}(\tilde{\mu}) = 5$ and $N = 9 > \mathbb{E}(\tilde{N}) = 7$. Therefore, consumers have pessimistic beliefs.

Beliefs	Firm Revenue	Avg. Queue Length	Avg. Waiting Time	Consumer Welfare	Social Welfare
Uninformed $\tilde{N} = \tilde{N}_0$	8.75	1.86	0.42	18.81	27.56
Construction:$\tilde{N}_1$	8.87 (+1.41%)	1.94 (+4.32%)	0.44 (+2.87%)	18.86 (+0.26%)	27.73 (+0.62%)
$\tilde{N}_2$	9.03 (+3.23%)	2.08 (+11.94%)	0.46 (+8.44%)	18.77 (-0.21%)	27.80 (+0.88%)
$\tilde{N}_3$	9.17 (+4.88%)	2.24 (+20.47%)	0.49 (+14.86%)	18.57 (-1.28%)	27.75 (+0.68%)
$\tilde{N}_4 = \tilde{N}_T$	9.23 (+5.47%)	2.29 (+23.35%)	0.50 (+16.95%)	18.52 (-1.58%)	27.74 (+0.66%)
Informed N	9.52 (+8.83%)	2.84 (+52.78%)	0.60 (+40.38%)	17.21 (-8.52%)	26.73 (-3.01%)

In this case, revealing the true service rate increases the firm's revenue by 8.83% but also increases the average queue length by 52.78% and the average waiting time by 40.38%. The firm thus benefits from revealing its service information (in line with Proposition 1), but the increased benefit is not sufficient to overcome the loss in consumer welfare (-8.52%). As a result, the overall social welfare drops by 3.01%.

Consistent (but Noisy) Beliefs: Let $\lambda = 4/\text{min}$, $\mu = 5/\text{min}$. Note that $\mu = \mathbb{E}(\tilde{\mu}) = 5$ and $N = \mathbb{E}(\tilde{N}) = 7$. Therefore beliefs are consistent in the population. Nevertheless, the individual consumer beliefs could vary arbitrarily (uniformly distributed in this case).

Belief Type	Firm Revenue	Avg. Queue Length	Avg. Waiting Time	Consumer Welfare	Social Welfare
Uninformed $\tilde{N} = \tilde{N}_0$	7.08 (+0.00%)	1.74 (+0.00%)	0.49 (+0.00%)	14.29 (+0.00%)	21.37 (+0.00%)
Construction $\tilde{N}_1$	7.18 (+1.43%)	1.82 (+4.37%)	0.51 (+2.89%)	14.29 (+0.00%)	21.47 (+0.47%)
$\tilde{N}_2$	7.31 (+3.24%)	1.95 (+11.89%)	0.53 (+8.37%)	14.15 (-0.97%)	21.46 (+0.43%)
$\tilde{N}_3$	7.43 (+4.86%)	2.09 (+20.19%)	0.56 (+14.62%)	13.92 (-2.61%)	21.34 (-0.13%)
$\tilde{N}_4 = \tilde{N}_T$	7.47 (+5.46%)	2.14 (+23.12%)	0.57 (+16.75%)	13.84 (-3.15%)	21.31 (-0.30%)
Informed $N = \tilde{N}_T$	7.47 (+5.46%)	2.14 (+23.12%)	0.57 (+16.75%)	13.84 (-3.15%)	21.31 (-0.30%)

In the consistent beliefs case, revealing the true service rate improves revenues (by 5.46%) in line with Proposition 1. On the other hand, the average queue length and the average waiting time both increase significantly (by 23.12% and 16.75% respectively). The firm benefits from revealing the service rate, almost fully at the expense of consumer welfare (-3.15%), but the overall social welfare does not fall significantly (-0.30%) due to the increase in throughput (i.e., number of consumers served).

Optimistic Beliefs: Let $\lambda = 3$/min and $\mu = 4$/min. Note $\mu = 4 < \mathbb{E}(\tilde{\mu}) = 5$ and $N = 6 < \mathbb{E}(\tilde{N}) = 7$. Hence, population beliefs are pessimistic.

Beliefs	Firm Revenue	Avg. Queue Length	Avg. Waiting Time	Consumer Welfare	Social Welfare
Uninformed $\tilde{N} = \tilde{N}_0$	5.40 (+0.00%)	1.57 (+0.00%)	0.58 (+0.00%)	9.93 (+0.00%)	15.33 (+0.00%)
Construction: $\tilde{N}_1$	5.48 (+1.45%)	1.64 (+4.39%)	0.60 (+2.90%)	9.89 (-0.41%)	15.37 (+0.25%)
$\tilde{N}_2$	5.58 (+3.22%)	1.75 (+11.65%)	0.63 (+8.17%)	9.72 (-2.11%)	15.30 (-0.23%)
$\tilde{N}_3$	5.66 (+4.76%)	1.88 (+19.48%)	0.66 (+14.04%)	9.48 (-4.54%)	15.14 (-1.26%)
$\tilde{N}_4 = \tilde{N}_T$	5.69 (+5.34%)	1.92 (+22.37%)	0.68 (+16.16%)	9.39 (-5.43%)	15.08 (-1.64%)
Informed N	5.57 (+3.03%)	1.70 (+8.31%)	0.61 (+5.13%)	9.90 (-0.31%)	15.46 (-0.86%)

Although consumers are optimistic about the service rate, revealing the true service rate would still increase firm's revenue by 3.03%. Examining the second column of the table (firm's revenue column) reveals what we have discussed for the optimism bias case: When consumers' optimistic balking threshold beliefs are more dispersed (as in the example the original belief $\tilde{N}$), it is beneficial for the firm to reveal service rate.

In this example, we see that if consumers' balking beliefs is characterized by $\tilde{N}$ or $\tilde{N}_1$, the firm increases its revenue from revealing the true service rate. As in the previous cases, the revenue accrual comes from the expense of increased queue lengths and waiting times for consumers. On the contrary, if consumer beliefs are less dispersed, (for e.g., if the beliefs were $\tilde{N}_3$), the firm does not gain from revealing its service rate. In this case, consumers are better off in both expected queue lengths and wait times.

To summarize, while the revenues improve with more information, the welfare effects are mixed. Typically the firm benefits from revealing service information, to the detriment of consumer welfare. Often, but not always, the gains in revenues are lower than consumer welfare loss. In such case, the social welfare reduces, as a consequence of more information in the system. However, it is also possible that both the firm revenues and consumer welfare improve upon service information revelation. This can occur when the traffic is very high and consumers' prior beliefs are almost deterministic. One such example is given by Case (iii) of the numerical illustration in §3.2.

5 Applying our Findings to Specific Belief Models

While our results hold for any general belief structure, it is helpful to evaluate what our findings imply under some specific belief considerations that have been examined in the literature. In this context, it is germane to consider the following issue: If consumers arrive to a queue endowed with some pre-existing beliefs, how do these different beliefs form? In the following section, we consider some behavioral/operational antecedents to belief structures, show our analysis apply to those cases and derive conclusions from those applications.

5.1 Quantal Response Errors

Quantal response models are used to model deviations from optimal consumer decisions in the absence of full information. For instance, in queues, consumers may make "errors" in their estimate of the true service rate due to cognitive limitations following Quantal Choice Theory (Luce, 1959), which argues that decision makers do not always choose the "correct" alternative, but better alternatives (i.e., alternatives with smaller errors) are chosen with a higher probability than the alternatives that are worse. Quantal choice approach has been employed to model bounded rationality in the newsvendor contexts by Su (2008), and subsequently in queueing settings by Huang et al. (2012).

If consumer population made i.i.d. belief draws from a distribution that align with Quantal Choice Theory, a large fraction of the population will have small errors in their beliefs about the true service rate, and a diminished fraction of customers make arbitrarily large errors in their beliefs. Furthermore, the mode of such a belief distribution will coincide with the true service parameter.

For a consumer j, let the belief on the true service rate (μ) be $\tilde{\mu}_j$. We use $[|\tilde{\mu}_j - \mu| + 1]^{-1} \in (0, 1]$ to indicate the accuracy of her belief.[4] Assuming i.i.d. customers, we could use a logit model – the mostly commonly used Quantal response distribution – for $\tilde{\mu}$, to model the accuracy of consumer beliefs. Then, the pdf for the belief distribution $\tilde{\mu}$ is given by:

$$f_{\tilde{\mu}}(x) = \frac{\exp\{[\beta(|x-\mu|+1)]^{-1}\}}{\int\limits_{x=\underline{\mu}}^{x=\bar{\mu}} \exp\{[\beta(|x-\mu|+1)]^{-1}\}dx}$$

where β is a cognitive parameter that measures "distance" from perfect rationality. As $\beta \to \infty$, $\tilde{\mu} \sim U[\underline{\mu}, \bar{\mu}]$ (consumers are totally uninformed and make 'random' errors), and when $\beta \to 0$, $Pr(\tilde{\mu} = \mu) = 1$ (consumers

[4] Other measures of accuracy could be employed without altering our conclusions.

are fully informed and we recover Naor's model in this context). When the belief distribution is symmetric, i.e., $\mu = (\underline{\mu} + \bar{\mu})/2$, we note that as β decreases from ∞ to 0, the underlying consumer beliefs undergo the transformation described in Construction 1.

Quantal response choices only require that the zero-error choice is chosen with the highest probability, and therefore are *not necessarily* consistent beliefs. For example, consumer beliefs can be optimistic (if $\mu < (\underline{\mu} + \bar{\mu})/2$) or pessimistic (if $\mu > (\underline{\mu} + \bar{\mu})/2$). Quantal response beliefs are examples of beliefs where the results of our paper apply. Although Quantal response beliefs can explain some deviations from the optimal/true choice, they do not inform how these beliefs form. We examine some specific causes (e.g., past experiences) in the following sections.

5.2 Learning by Sampling Past Experiences

In many service instances, consumers have limited and infrequent interactions with the service provider. In such cases, consumers could use their past service experience as samples to learn more about the service rate. This sampling helps consumers to arrive at their beliefs and eventually make their decisions. Suppose that all consumers in a sufficiently large population, use only their past service experience to estimate the service rate. Specifically, let us examine a case in which all consumers have visited the server s times, ($s \geq 1$) or only remember the past s service time experiences. We assume that consumers are homogeneous in s in this section, but relax the assumption in §5.3.

Consider a consumer with the following service time samples $\{\tau_1, \tau_2, \ldots, \tau_s\}$. A rational consumer who knows the service distribution, but not the exact parameters, will use the observed samples to arrive at an estimate that maximizes the likelihood of observing those s samples. Simply, a rational consumer would use Maximum Likelihood Estimator (MLE) for calculating the parameters of the service distribution. Suppose the service times are i.i.d. exponential, then it is well known that the MLE for the service rate is given by

$$\hat{\mu}|(\tau_1, \tau_2, \ldots, \tau_s) = s \Bigg/ \sum_{i=1}^{s} \tau_i \ . \tag{14}$$

This is the point estimate for the service rate for the consumer with samples $\{\tau_1, \tau_2, \ldots, \tau_s\}$. Thus, consumers will have different beliefs (estimates) based on their individual samples. Define $\tilde{\mu}_s$ to be the random variable associated with the belief distribution, when all consumers use s samples to arrive at their beliefs through MLE. We note that $\sum_{i=1}^{s} \tau_i$ in equation (14) has an Erlang distribution with shape parameter s and rate parameter μ. It follows that over the population, the individual consumer beliefs $\tilde{\mu}_s$ are distributed Inverse-Gamma with shape parameter s and scale parameter $s\mu$, i.e., $\tilde{\mu}_s \sim$ *Inv-Gamma*$(s, s\mu)$. The pdf for $\tilde{\mu}_s$ is given by

$$f_{\tilde{\mu}_s}(x) = \frac{(s\mu)^s}{\Gamma(s)} x^{-s-1} \exp(-\frac{s\mu}{x}) \text{ for } x > 0 \tag{15}$$

where Γ denotes the upper incomplete gamma function. Furthermore, if the service times are independent, the estimate remains unchanged if these samples are collected on a single visit or over multiple visits.

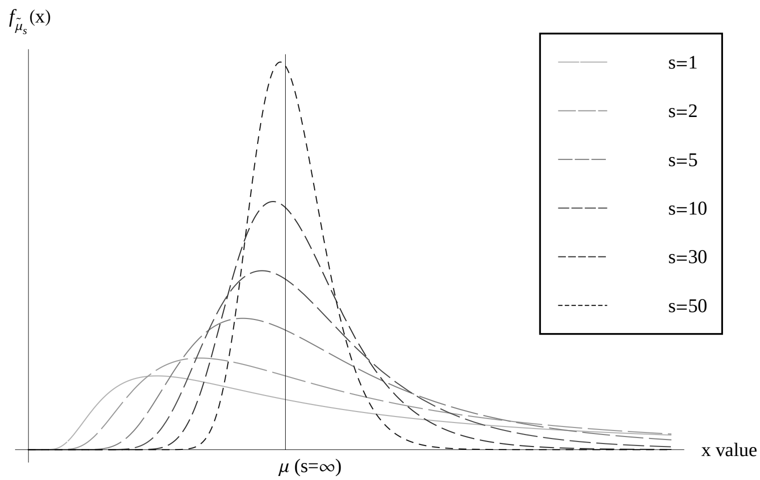

Figure 1: As consumers remember more service experiences ($s \uparrow \infty$), their estimates of the service rate become consistent with the true service rate. Specifically, we observe that (i) $\mathbb{E}(\tilde{\mu}_s) \downarrow \mu$, and (ii) $\mathrm{Var}(\tilde{\mu}_s) \downarrow 0$.

In Figure 1, we illustrate $\tilde{\mu}_s$ for different sampling sizes s. Since $\mathbb{E}(\tilde{\mu}_s) = \frac{s}{s-1}\mu$ and $\mathrm{Var}(\tilde{\mu}_s) = \frac{s^2}{(s-1)^2(s-2)}\mu$, we note that for any finite s, the population mean is higher than the true μ. Thus the population is optimistically biased. Further, as s increases, consumer beliefs get less noisy (i.e., $\mathrm{Var}(\tilde{\mu}_s) \to 0$ as $s \uparrow \infty$), as reflected in the distributions getting less spread-out in Figure 1. Eventually, as the number of samples approaches infinity in (15), the distribution of $\tilde{\mu}_s$ approaches a one-point distribution at μ, i.e., $Pr(\tilde{\mu} = \mu) = 1$. As consumers learn service rate through sampling, they remain optimistic but diminishingly so, as they collect more samples. Thus MLE derived through sampling is *biased* but *asymptotically consistent*.

Finally, for any s, the mode of the belief distribution $\tilde{\mu}_s$ does not coincide with the true service rate μ ($s = \infty$ line). Therefore, the beliefs that emerge from learning through sampling, are not Quantal choice beliefs. Hence, sampling distributions are another distinct example of our belief distribution results.

Learning through Waiting Times: Note that for a given s, as long as the service times are i.i.d., it does not matter whether a consumer's MLE is built from her own service times, or from her observations of service times for other consumers. In some cases, consumers may not be able to observe all other service times, due to limited cognitive attention to the sequence of events, or due to system environment. As an example, consider ticket queues (see Xu et al. (2007)) where a consumer learns only her own wait time and ticket number. Another similar setting occurs at emergency room queues, where consumers learn their own total wait time but not the exact service times of everyone in the queue.

Suppose that a consumer only knows her total wait time W_s and the number of customers during the wait, s (for e.g., queue length), but does not observe the individual service times $(\tau_1, \tau_2, \ldots, \tau_s)$ during the wait. Her service time belief is based on the observed waiting time, i.e., $\hat{\tau}|W_s = W_s/s$ and the corresponding service rate belief is $\hat{\mu}|W_s = [\hat{\tau}|W_s]^{-1} = s/W_s$. This estimate $\hat{\mu}|W_s$ is identical to the MLE in equation (14) for a consumer who observed individual service times which sum up to W_s. The overall belief distribution using the waiting time estimate W_s is thus identical to the MLE case with s samples.

5.3 Consumer Heterogeneity in Learning

So far, we limited consumer sampling to be homogeneous across the entire population. However, it is possible that consumers differ in s. Typically, there is some underlying distribution of s in the population, based on how consumers accumulate information or are exposed to it. For instance, consumers may consider external reviews or auxiliary information from other consumers. We denote this sampling heterogeneity by a discrete random variable S taking values in $\{1, 2, \ldots\}$ with pdf f_S. When S is a one-point distribution, we retain homogeneity in learning. With (15) in hand, we can write the continuous *unconditional* distribution of beliefs denoted by $\tilde{\mu}$ in the population, through the following pdf:

$$f_{\tilde{\mu}}(x) = \sum_{s \in S} \frac{(s\mu)^s}{\Gamma(s)} x^{-s-1} \exp(-\frac{s\mu}{x}) f_S(s) \text{ for } x > 0. \tag{16}$$

We now consider two different ways to model the learning heterogeneity, $f_S(s)$.

Sampling through Poisson Arrivals: One natural distribution of sampling in the population could be based on what consumers observe when they arrive to a queue. Applying PASTA property (Wolff, 1982), when consumers arrive according to a Poisson process to the server, their sampling distribution follows the steady state distribution of the queuing system. In an $M/M/1$ queue, a customer who arrives at the state $s-1$ observes s samples in total. With traffic intensity $\rho = \lambda/\mu$, the sampling distribution would be $f_S(s) = (1-\rho)\rho^{s-1},\ s \geq 1$. Clearly, if the queue is a system with limited buffer size, then the sampling distribution is on finite support, and therefore the population becomes more optimistic, following our previous discussion.

Limited Recall/External Reviews: Consumers may come across information (for e.g., reviews on bulletin boards or websites), but may not recall all observed information due to cognitive limitations. When subjected to an increasing amount of information, a consumer may remember only a limited amount of information, and forget an amount that is proportional to the information she is exposed to. We examine such a sampling system, using the model of limited memories due to Nelson (1974). We assume that consumers confront reviews according to a Poisson Process with rate r and they also forget a review at a rate that is directly proportional to the number of reviews a person remembers. Let a be the constant of the proportionality. We denote the state $n \in \{0, 1, 2, \ldots\}$ as the number of reviews that a particular customer remembers in the long run. The fraction of customers who remember n reviews is given by steady-state probability distribution $\{\pi_n : n = 0, 1, 2, \ldots\}$. We then have $\pi_n = (r/a)^n e^{-r/a}/n!$ for $n \in \{0, 1, 2, \ldots\}$.

Suppose that consumers form their beliefs only after reading at least one review or service experience. Then, the fraction of consumers who remember s reviews among the population is given by

$$f_S(s) = \frac{\pi_s}{1 - \pi_0} = \frac{(r/a)^n}{n!} \frac{e^{-r/a}}{1 - e^{-r/a}} \text{ for } s = 1, 2, \ldots$$

The fraction $f_S(s)$ use s reviews to arrive at their belief $\tilde{\mu}_s$. We can then use f_S to build the distribution of consumer beliefs in the queue through equation (16).

Regardless of how we model the sampling heterogeneity f_S, we conclude that consumer population will tend to be optimistic, even though all reviews/service experiences are assumed to be accurate. This is because, using MLE, each consumer class is optimistic. Thus, learning through sampling can lead to optimism bias even in heterogenous populations. We briefly examine how pessimistic beliefs can emerge.

From a behavioral point of view, consumers may have *availability bias* (Tversky and Kahneman, 1973) when processing information. Under availability bias, consumers remember *unusual* experiences saliently, in forming their beliefs. In the case of exponential servers, when consumers have finite sample sizes, short service times are much likely to be present in their sample than long service times. Thus, consumers recall longer-than-usual service times more vividly. As a result, availability bias could lead to tempered optimism or even pessimism in the population beliefs, i.e., $\mathbb{E}(\tilde{\mu}) < \mu$. Another cause of pessimistic bias, can be due to Prospect Theory (Kahneman and Tversky, 1979), where longer (worse) service times affect updating more significantly. See Gaur and Park (2007) for such asymmetric consumer learning in inventory context.

5.4 Conclusions and Implications.

Consumers often join queues with very limited information. Much of the literature has assumed that service parameters that influence the joining behavior as common knowledge. For instance, almost all queueing research in observable queues assumes that the service capacity μ is known. However, consumers cannot always fully characterize these service parameters; sometimes even the calculation of mean service time may require repeated sampling or collection of data. In such blind queues, not much is known on how revenues and welfare are impacted, when a firm reveals its service information (specifically, service capacity). Our paper seeks to fill this gap.

Our approach to solving the information problem is distribution-free. We begin with *any* belief distribution that consumers may have on the service rate in the uncountable space, and reduce it to balking threshold beliefs in the countable space. Using our general but intuitive approach, we calibrate the impact of information revelation on the performance of the queueing system, without any restrictions on the distribution of the initial consumer beliefs.

We can apply the results from our general model on specific belief structures, such as Quantal-response based bounded rationality, learning through sampling either from past experiences or reviews, and other cognitive biases to characterize their effects on revenues and consumer welfare. In fact, we find that learning from sampling can lead to optimism bias, even though sampling can be asymptotically consistent. We show that learning through sampling imposes fundamentally different beliefs structures from the bounded rational models based on Quantal response.

We find that, under consistent or pessimistic beliefs, the service provider always improve revenues by revealing its service parameters. Unlike the impact on revenues, the impact of service rate information on congestion and welfare is mixed. Even though a firm's revenues improve on announcing its service rate, the impact on the congestion levels (such as the average queue lengths, or average wait times) are typically negative. As a result, consumer welfare worsens on revelation, despite the increased market coverage. In fact, the impact on consumer welfare could be significantly negative, compared to revenue improvements at the firm. Hence, social welfare can fall even though there is increased information in the population.

Given a market size, consumer welfare likely worsens in the case when a fast service reveals its service rate, compared to the case when a slow server reveals its rate. Thus, consumers are worse off, if a *faster/better* firm with more service capacity reveals its service information.

Hence, intriguingly, with informational uncertainty, the social welfare typically improves, compared to the case in which consumers have full information. When left to their own devices, with full information, more consumers join the queue than what is socially optimal. In Naor (1969), tolls/taxes are levied to control the joining population which improves welfare. Likewise, we find that the lack of information acts as an *information tax* that deters admission, which can lead to improved welfare. Several future research directions related to queue information appear promising. For instance, we could examine the impact of unobservability of queue lengths using our approach, and measure its impact on consumer beliefs.

Our findings have several implications for queue management policies. In primary healthcare settings where the access to service providers is important, revealing the capacity information can lead to an increased consumer access to the queue (i.e., more consumers will visit the service provider). Nevertheless, consumers will observe longer queues on average, and also suffer a higher disutility in waiting time on average.

Thus when there is a significant impetus on treating admitted patients quickly, as in some emergency room settings, revealing the queue information may lead to increased crowding and worsen the average wait times. Furthermore, this effect is exacerbated for a facility that has ample service capacity. So, the decision to reveal the queue information in aforementioned settings depends critically on the tradeoffs between improved access and increased congestion.

Appendix

Proof of Theorem 1:

(i) For $j \in \{\tilde{N}, \tilde{N}'\}$, recall from (5) that $R_j = p\mu(1-\pi_0)$ where $\pi_0 = 1 \Big/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_j(n)\right)$. It is clear that the greater $\sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_j(n)$ is, the greater R_j will be. Since $\tilde{N} \leq_{st} \tilde{N}'$, we have $\bar{F}_{\tilde{N}}(n) \leq \bar{F}_{\tilde{N}'}(n)$ for every n. Thus $\sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \leq \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n)$ and $R_{\tilde{N}} \leq R_{\tilde{N}'}$.

(ii) Using equation (3) for the random variable associated with beliefs, $\tilde{N}$, and given $\tilde{N} \leq_{st} \tilde{N}'$, we would like to prove that

$$\frac{\sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n)}{\sum_{i=0}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n)} \geq \frac{\sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n)}{\sum_{i=0}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n)} \Leftrightarrow \frac{\sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n)}{\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)} \geq \frac{\sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n)}{\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n)},$$

$$\Leftrightarrow \left[\sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n)\right]\left[\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n)\right] \geq \left[\sum_{i=0}^{\infty} i\rho^i \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n)\right]\left[\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)\right],$$

$$\Leftrightarrow \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i\rho^{i+j} \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) \geq \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i\rho^{i+j} \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n),$$

$$\Leftrightarrow \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)\right) \geq 0,$$

$$\Leftrightarrow \sum_{i,j\geq 0:}\sum_{i\neq j} i\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)\right) \geq 0,$$

$$\Leftrightarrow \sum_{i,j\geq 0:}\sum_{i>j} i\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)\right)$$
$$+ \sum_{i,j\geq 0:}\sum_{i<j} i\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)\right) \geq 0,$$

$$\Leftrightarrow \sum_{i,j\geq 0:}\sum_{i>j} i\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n)\right)$$
$$+ \sum_{i,j\geq 0:}\sum_{i>j} j\rho^{i+j} \left(\prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n)\right) \geq 0.$$

Regrouping again, gives

$$\sum_{i,j\geq 0:}\sum_{i>j} \left((i-j)\rho^{i+j} \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) + (j-i)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n)\right) \geq 0,$$

$$\Leftrightarrow \sum_{i,j\geq 0:}\sum_{i>j} (i-j)\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) - \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{i-1} \bar{F}_{\tilde{N}}(n)\right) \geq 0,$$

$$\Leftrightarrow \sum_{i,j\geq 0:}\sum_{i>j} (i-j)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) \left(\prod_{n=j}^{i-1} \bar{F}_{\tilde{N}'}(n) - \prod_{n=j}^{i-1} \bar{F}_{\tilde{N}}(n)\right) \geq 0. \tag{17}$$

Finally, $N \leq_{st} N'$ implies that $\prod_{n=j}^{i-1} \bar{F}_{\tilde{N}'}(n) - \prod_{n=j}^{i-1} \bar{F}_{\tilde{N}}(n) \geq 0$. Thus $L_{N'} \geq L_N$ from (17).

(iii) Using equation (7) for the beliefs $\tilde{N}$, and comparing equation (3) to (7), we find that a similar approach used in the proof of part (ii) would be applied to show that

$$W_{\tilde{N}'} \geq W_{\tilde{N}} \Leftrightarrow \sum_{i,j\geq 1:} \sum_{i>j} (i-j)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}'}(n) \prod_{n=0}^{j-1} \bar{F}_{\tilde{N}}(n) \left(\prod_{n=j}^{i-1} \bar{F}_{\tilde{N}'}(n) - \prod_{n=j}^{i-1} \bar{F}_{\tilde{N}}(n) \right) \geq 0. \quad (18)$$

The result thus follows again from the fact that $\prod_{n=j}^{i-1} \bar{F}_{\tilde{N}'}(n) - \prod_{n=j}^{i-1} \bar{F}_{\tilde{N}}(n) \geq 0$. □

Proof of Lemma 2:

(i) By construction, the entire probability mass at one end of the distribution is transferred to the middle of the support. As a result, the range of the random variable $\tilde{N}_{K+1}$ is a strict subset of the range of $\tilde{N}_K$. Specifically, $a_{(K+1)_1} = a_{K_1} + 1$ if $f_K(a_{K_1}) \leq f_K(a_{K_n})$ and $a_{(K+1)_N} = a_{K_N} - 1$ if $f_K(a_{K_1}) \geq f_K(a_{K_n})$. The *length* of the range of $\tilde{N}_K$, $|a_{K_n} - a_{K_1}|$, is strictly decreasing in K. Within a finite number of steps, for some time $K = T$, the length will be less than 2. When $a_{T_n} - 1 < a_{T_1} + 1$, the process stops. Thus, T is finite.

(ii) We show that $F_{K+1} \leq_{SMPS} F_K$. Let a_{i_1}, a_{i_2}, a_{i_3}, a_{i_4} in Definition 2 be a_{K_1}, $a_{K_1}+1$, $a_{K_n}-1$ and a_{K_n} respectively. $f_{K+1} = f_K$ for all but these four points. Define $\gamma_{i_k} = f_K(a_{i_k}) - f_{K+1}(a_{i_k})$ for $k = 1,2,3,4$. Then, $\gamma_{i_1} = -\gamma_{i_2} = -\gamma_{i_3} = \gamma_{i_4} = \min\{f_K(a_{K_1}), f_K(a_{K_n})\} > 0$. Moreover, $\sum_{k=i}^{4} a_{i_k}\gamma_{i_k} = [a_{K_1} - (a_{K_1}+1) - (a_{K_n}-1) + a_{K_n}] \min\{f_K(a_{K_1}), f_K(a_{K_n})\} = 0 \cdot \min\{f_K(a_{K_1}), f_K(a_{K_n})\} = 0$. □

Proof of Lemma 3:

Suppose that $\tilde{N}_T$ has two elements which are not consecutive. Then, it must be that $a_{T_1} + 1 < a_{T_n}$. By Construction 1, then sequence is not completed, which contradicts the definition of T. Else, suppose that $\tilde{N}_T$ has three or more elements. Again, it must be that $a_{T_1} + 1 < a_{T_n}$, and hence, the sequence in Construction 1 is incomplete, which contradicts the definition of T. Therefore, $\tilde{N}_T$ can either take a single value or two consecutive values. Case (i): When if $\mathbb{E}(\tilde{N}_0)$ is an integer, since the transformation is mean preserving, we have $\tilde{N}_T$ is a singleton with $\tilde{N}_T = \mathbb{E}(\tilde{N}_0) = \lfloor \mathbb{E}(\tilde{N}_0) \rfloor = \lceil \mathbb{E}(\tilde{N}_0) \rceil$. Case (ii): When $\mathbb{E}(\tilde{N}_0)$ is not an integer, $\tilde{N}_T$ cannot be a singleton. Thus, $\tilde{N}_T$ takes on two consecutive values. Since the transformation in Construction 1 is mean-preserving with $\mathbb{E}(\tilde{N}_0)$, we have $\mathbb{E}(\tilde{N}_T) = \mathbb{E}(\tilde{N}_0)$. Then we must have $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}_0) \rfloor, \lceil \mathbb{E}(\tilde{N}_0) \rceil\}$, with $\Pr(\tilde{N}_T = \lfloor \mathbb{E}(\tilde{N}_0) \rfloor) \lfloor \mathbb{E}(\tilde{N}_0) \rfloor + \Pr(\tilde{N}_T = \lceil \mathbb{E}(\tilde{N}_0) \rceil) \lceil \mathbb{E}(\tilde{N}_0) \rceil = \mathbb{E}(\tilde{N}_0)$. It is also clear that the distribution of $\tilde{N}_T$ is independent of the distribution of $\tilde{N}_0$. □

Proof of Lemma 4:

(i) For $j \in \{K, K+1\}$, recall from (5) that $R_{\tilde{N}_j} = p\mu(1-\pi_0)$ where $\pi_0 = 1/\left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_j(n)\right)$. It is thus sufficient to show that

$$\sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n). \quad (19)$$

To verify (19), our strategy is to form a partition of $i \in \{0,1,2,...\}$ based on the sign of $\rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) -$

$\rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$. We specifically focus on terms that make the product $\prod_{n=0}^{i-1} \bar{F}_K(n)$, namely $\bar{F}_K(n)$. Since $\bar{F}_K(n) = f_K(n+1) + f_K(n+2) + \ldots$, applying Construction 1, we have

$$
\begin{aligned}
\bar{F}_{K+1}(0) &= \bar{F}_K(0) \quad &&= 1,\\
\bar{F}_{K+1}(1) &= \bar{F}_K(1) \quad &&= 1,\\
&\vdots\\
\bar{F}_{K+1}(a_{K_1}-2) &= \bar{F}_K(a_{K_1}-2) \quad &&= 1,\\
\bar{F}_{K+1}(a_{K_1}-1) &= \bar{F}_K(a_{K_1}-1) \quad &&= 1,\\
\bar{F}_{K+1}(a_{K_1}) &= \bar{F}_K(a_{K_1}) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \quad &&\in (0,1],\\
\bar{F}_{K+1}(a_{K_1}+1) &= \bar{F}_K(a_{K_1}+1) \quad &&\in (0,1),\\
\bar{F}_{K+1}(a_{K_1}+2) &= \bar{F}_K(a_{K_1}+2) \quad &&\in (0,1),\\
&\vdots\\
\bar{F}_{K+1}(a_{K_n}-2) &= \bar{F}_K(a_{K_n}-2) \quad &&\in (0,1),\\
\bar{F}_{K+1}(a_{K_n}-1) &= \bar{F}_K(a_{K_n}-1) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \quad &&\in [0,1),\\
\bar{F}_{K+1}(a_{K_n}) &= \bar{F}_K(a_{K_n}) \quad &&= 0,\\
\bar{F}_{K+1}(a_{K_n}+1) &= \bar{F}_K(a_{K_n}+1) \quad &&= 0,\\
&\vdots
\end{aligned}
\tag{20}
$$

Thus, using transformation of $\tilde{N}_K$ to $\tilde{N}_{K+1}$ in Construction 1, we see that $\bar{F}_K(n)$ differs from $\bar{F}_{K+1}(n)$ at only two points, specifically $n = a_{K_1}$ and $n = a_{K_n} - 1$. In order to show (19), we verify, as an intermediate step, that $\bar{F}_{K+1}(a_{K_1}) \cdot \bar{F}_{K+1}(a_{K_n} - 1) < \bar{F}_K(a_{K_1}) \cdot \bar{F}_K(a_{K_n} - 1)$. We have

$$
\begin{aligned}
&\bar{F}_{K+1}(a_{K_1}) \cdot \bar{F}_{K+1}(a_{K_n}-1) = [\bar{F}_K(a_{K_1}) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\}][\bar{F}_K(a_{K_n}-1) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\}]\\
=&\bar{F}_K(a_{K_1})\bar{F}_K(a_{K_n}-1) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\}[\bar{F}_K(a_{K_n}-1) - \bar{F}_K(a_{K_1}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\}]\\
<&\bar{F}_K(a_{K_1})\bar{F}_K(a_{K_n}-1) \text{ since } \bar{F}_K(a_{K_n}-1) \le \bar{F}_K(a_{K_1}).
\end{aligned}
\tag{21}
$$

Now we define $\mathcal{S}_1 \triangleq \{1, 2, \ldots, a_{K_1}\}$; $\mathcal{S}_2 \triangleq \{a_{K_1}+1, a_{K_1}+2, \ldots, a_{K_n}-1\}$; $\mathcal{S}_3 \triangleq \{a_{K_n}\}$; and $\mathcal{S}_4 \triangleq \{a_{K_n}+1, a_{K_n}+2, \ldots\}$. S_1, S_2, S_3 and S_4 then form a partition of the space $\{1, 2, 3, \ldots\}$. Our goal (19) is equivalent to $\sum_{i \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$.

From (20) and (21) we have $\forall i \in \mathcal{S}_1 : \prod_{n=0}^{i-1} \bar{F}_K(n) = \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) = 1$; $\forall i \in \mathcal{S}_2 : \prod_{n=0}^{i-1} \bar{F}_K(n) < \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$; $\forall i \in \mathcal{S}_3 : \prod_{n=0}^{i-1} \bar{F}_K(n) > \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$; and $\forall i \in \mathcal{S}_4 : \prod_{n=0}^{i-1} \bar{F}_K(n) = \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) = 0$.

It is clear that $\mathcal{S}_1$ and $\mathcal{S}_4$ are collection of the indices i where $\prod_{n=0}^{i-1} \bar{F}_K(n) = \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$. Hence, to prove (19), we just need to show that $\sum_{i \in \mathcal{S}_2 \bigcup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \mathcal{S}_2 \bigcup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$.

As $\rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$ for all $i \in \mathcal{S}_2$, the previous inequality will hold, if there exists some $\mathcal{S}_{2'} \subseteq \mathcal{S}_2$ such that

$$\sum_{i \in \mathcal{S}_{2'} \cup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \mathcal{S}_{2'} \cup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n). \tag{22}$$

On the other hand, the existence of $\tilde{N}_{K+1}$ guarantees that $a_{K_n} - 1 \geq a_{K_1} + 1$ and $a_{K_1} \geq 1$ so there exists at least one index in $\mathcal{S}_2$ (i.e., $i = a_{K_1} + 1$). There is only one element in $\mathcal{S}_3$ (i.e., $i = a_{K_n}$). Define $\mathcal{S}_{2'} \triangleq \{a_{K_1} + 1\}$. So $S_{2'} \cup \mathcal{S}_3 = \{a_{K_1} + 1, a_{K_n}\}$. Inequality (22) is therefore equivalent to

$$\sum_{i \in \{a_{K_1}+1, a_{K_n}\}} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \{a_{K_1}+1, a_{K_n}\}} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n),$$

$$\Leftrightarrow \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}} \bar{F}_K(n) + \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_K(n) < \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}} \bar{F}_{K+1}(n) + \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_{K+1}(n).$$

And the last condition is true because

$$\begin{aligned}
&\rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_K(n) - \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_{K+1}(n) < \rho^{a_{K_n}} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n\} \\
\leq &\rho^{a_{K_n}} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} < \rho^{a_{K_1}+1} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\
= &\rho^{a_{K_1}+1} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \prod_{n=0}^{a_{K_1}-1} \bar{F}_K(n) = \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}} \bar{F}_{K+1}(n) - \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}} \bar{F}_K(n),
\end{aligned}$$

which implies in the backward direction that inequality (22)-(19) all hold, and thus $R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$.

(ii) From equation (17) in the proof of Theorem 1, we note that $L_K < L_{K+1}$ if and only if

$$\sum_{i,j \geq 0:} \sum_{i > j} (i-j) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \left(\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n) \right) > 0.$$

Since, a_{K_n} is the largest value on the support of $\bar{F}_K$, we have $\bar{F}_{K+1}(i-1) = \bar{F}_K(i-1) = 0$ for $i \in \{a_{K_n}+1, a_{K_n}+2, \ldots\}$. Hence, those indices can be dropped, which gives us

$$\sum_{a_{K_n} \geq i >} \sum_{j \geq 0} (i-j) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \left(\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n) \right) > 0. \tag{23}$$

$$\text{Let us define } A_{K+1}(i,j) \triangleq (i-j) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \prod_{n=j}^{i-1} \bar{F}_{K+1}(n);$$

$$A_K(i,j) \triangleq (i-j) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \prod_{n=j}^{i-1} \bar{F}_K(n). \tag{24}$$

Then (23) reduces to $\sum_{a_{K_n} \geq i >} \sum_{j \geq 0} [A_{K+1}(i,j) - A_K(i,j)] > 0$.

Similar to the approach used in the proof of part (i), our strategy is to form a partition of (i,j) based on the sign of $A_{K+1}(i,j) - A_K(i,j)$. The underlying space is the 2-dimensional set $\{(i,j) : a_{K_n} \geq i > j \geq 0\}$. Since $A_{K+1}(i,j) > A_K(i,j)$ if and only if $\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) > \prod_{n=j}^{i-1} \bar{F}_K(n)$, we shall seek a partition over $\{(i,j) : a_{K_n} \geq i > j \geq 0\}$ based on the sign of $\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n)$ instead.

Define $\mathcal{G}_1 \triangleq \{a_{K_n}\} \times \{0, 1, 2, \ldots, a_{K_n}-1\}$; $\mathcal{G}_2 \triangleq \{a_{K_1}+1, a_{K_1}+2, ..., a_{K_n}-2, a_{K_n}-1\} \times \{0, 1, 2, \ldots, a_{K_1}\}$; and $\mathcal{G}_3 \triangleq \{(i,j) : a_{K_n} \geq i > j \geq 0\} - \{\mathcal{G}_1 \cup \mathcal{G}_2\}$, i.e., $\mathcal{G}_3$ contains all the elements that are not in $\mathcal{G}_1$ or $\mathcal{G}_2$.

From (20) and (21), we can verify that

$$\forall (i,j) \in \mathcal{G}_1 : \prod_{n=j}^{i-1} \bar{F}_{K+1}(n) < \prod_{n=j}^{i-1} \bar{F}_K(n) \Rightarrow A_{K+1}(i,j) - A_K(i,j) < 0;$$

Order l	$\mathcal{G}_1$ contains:	$\mathcal{G}_{2'}$ contains:
$l=1$	$(a_{K_n}, a_{K_n}-1)$	$(a_{K_1}+1, a_{K_1})$
$l=2$	$(a_{K_n}, a_{K_n}-2)$	$(a_{K_1}+1, a_{K_1}-1)$
$l=3$	$(a_{K_n}, a_{K_n}-3)$	$(a_{K_1}+1, a_{K_1}-2)$
$\vdots$	$\vdots$	$\vdots$
$l=a_{K_1}-1$	$(a_{K_n}, a_{K_n}-a_{K_1}+1)$	$(a_{K_1}+1, 2)$
$l=a_{K_1}$	$(a_{K_n}, a_{K_n}-a_{K_1})$	$(a_{K_1}+1, 1)$
$l=a_{K_1}+1$	$(a_{K_n}, a_{K_n}-a_{K_1}-1)$	$(a_{K_1}+1, 0)$
$l=a_{K_1}+2$	$(a_{K_n}, a_{K_n}-a_{K_1}-2)$	$(a_{K_1}+2, 0)$
$l=a_{K_1}+3$	$(a_{K_n}, a_{K_n}-a_{K_1}-3)$	$(a_{K_1}+3, 0)$
$\vdots$	$\vdots$	$\vdots$
$l=a_{K_n}-2$	$(a_{K_n}, 2)$	$(a_{K_n}-2, 0)$
$l=a_{K_n}-1$	$(a_{K_n}, 1)$	$(a_{K_n}-1, 0)$
$l=a_{K_n}$	$(a_{K_n}, 0)$	

Table 1: $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ used in the proof of Lemma 4/(ii)

Order l	$\mathcal{G}_1$ contains:	$\mathcal{G}_{2'}$ contains:
$l=1$	$(a_{K_n}, a_{K_n}-1)$	$(a_{K_1}+1, a_{K_1})$
$l=2$	$(a_{K_n}, a_{K_n}-2)$	$(a_{K_1}+1, a_{K_1}-1)$
$l=3$	$(a_{K_n}, a_{K_n}-3)$	$(a_{K_1}+1, a_{K_1}-2)$
$\vdots$	$\vdots$	$\vdots$
$l=a_{K_1}-1$	$(a_{K_n}, a_{K_n}-a_{K_1}+1)$	$(a_{K_1}+1, 2)$
$l=a_{K_1}$	$(a_{K_n}, a_{K_n}-a_{K_1})$	$(a_{K_1}+1, 1)$
$l=a_{K_1}+1$	$(a_{K_n}, a_{K_n}-a_{K_1}-1)$	$(a_{K_1}+2, 1)$
$l=a_{K_1}+2$	$(a_{K_n}, a_{K_n}-a_{K_1}-2)$	$(a_{K_1}+3, 1)$
$l=a_{K_1}+3$	$(a_{K_n}, a_{K_n}-a_{K_1}-3)$	$(a_{K_1}+4, 1)$
$\vdots$	$\vdots$	$\vdots$
$l=a_{K_n}-2$	$(a_{K_n}, 2)$	$(a_{K_n}-1, 1)$
$l=a_{K_n}-1$	$(a_{K_n}, 1)$	

Table 2: $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ used in the proof of Lemma 4/(iii)

$$\forall (i,j) \in \mathcal{G}_2 : \prod_{n=j}^{i-1} \bar{F}_{K+1}(n) > \prod_{n=j}^{i-1} \bar{F}_K(n) \Rightarrow A_{K+1}(i,j) - A_K(i,j) > 0;$$

$$\forall (i,j) \in \mathcal{G}_3 : \prod_{n=j}^{i-1} \bar{F}_{K+1}(n) = \prod_{n=j}^{i-1} \bar{F}_K(n) \Rightarrow A_{K+1}(i,j) - A_K(i,j) = 0.$$

Since $\mathcal{G}_3$ contains all (i,j) where $A_{K+1}(i,j)-A_K(i,j)=0$, it suffices to show that $\sum_{(i,j)\in}\sum_{\mathcal{G}_1 \bigcup \mathcal{G}_2} [A_{K+1}(i,j) - A_K(i,j)] > 0$ as a goal. Also since $A_{K+1}(i,j) - A_K(i,j) > 0$, $\forall\ (i,j) \in \mathcal{G}_2$, the previous inequality will hold if there exists a subset $\mathcal{G}_{2'} \subseteq \mathcal{G}_2$ such that

$$\sum_{(i,j)\in}\sum_{\mathcal{G}_1 \cup \mathcal{G}_{2'}} [A_{K+1}(i,j) - A_K(i,j)] > 0. \tag{25}$$

We shall prove that the sufficient condition on ρ stated in the lemma guarantees for inequality (25) to hold. To do that, we need to consider the elements of $\mathcal{G}_1$ and $\mathcal{G}_2$ in greater detail.

From the construction of the partition above, we have $|\mathcal{G}_1| = a_{K_n}$, i.e., there are a_{K_n} pairs of (i,j) in $\mathcal{G}_1$, represented by $\{(a_{K_n}, a_{K_n}-1), (a_{K_n}, a_{K_n}-2), (a_{K_n}, a_{K_n}-3), ..., (a_{K_n}, 1), (a_{K_n}, 0)\}$. On the other hand, $|\mathcal{G}_2| = (a_{K_n} - a_{K_1} - 1)(a_{K_1}+1)$. Treating $a_{K_n} - a_{K_1} - 1$ and $a_{K_1}+1$ as the base and the height of a rectangular and using the fact that a rectangular shape of fixed perimeter (a_{K_n}) contains less area ($|\mathcal{G}_2|$) when the shape is more asymmetric, we can show that $|\mathcal{G}_2| \geq a_{K_n} - 1$, with equality holds only when $a_{K_1} = a_{K_n} - 2$ (or $a_{K_1} = 0$ but is not possible).

The $a_{K_n}-1$ pairs of (i,j) that are guaranteed to reside in $\mathcal{G}_2$ can be parametrized as $\{(a_{K_1}+1, a_{K_1}), (a_{K_1}+1, a_{K_1}-1), (a_{K_1}+1, a_{K_1}-2), ..., (a_{K_1}+1, 1), (a_{K_1}+1, 0), (a_{K_1}+2, 0), (a_{K_1}+3, 0), ..., (a_{K_n}-2, 0), (a_{K_n}-1, 0)\}$. We define this set to be $\mathcal{G}_{2'}$. Note that $|\mathcal{G}_1| = a_{K_n}$ and $|\mathcal{G}_{2'}| = a_{K_n} - 1$. We now order elements of $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ in a specific way displayed in Table 1 (each element in either group is itself an (i,j) pair).

We denote $\mathcal{G}_1^l : l \in \{1, 2, \ldots, a_{K_n}\}$ and $\mathcal{G}_{2'}^l : l \in \{1, 2, \ldots, a_{K_n} - 1\}$ the l-th element in $\mathcal{G}_1$ and $\mathcal{G}_{2'}$, respectively, according to the order specified in Table 1. Furthermore, for each $\mathcal{G}_1^l$ and each $\mathcal{G}_{2'}^l$, we specific its Cartesian coordinates by subscript i and j, i.e., $\mathcal{G}_1^l = (\{\mathcal{G}_1^l\}_i, \{\mathcal{G}_1^l\}_j)$ and $\mathcal{G}_{2'}^l = (\{\mathcal{G}_{2'}^l\}_i, \{\mathcal{G}_{2'}^l\}_j)$. For example, $\{\mathcal{G}_1^{a_{K_n}}\}_i = a_{K_n}$ and $\{\mathcal{G}_{2'}^{a_{K_1}}\}_j = 1$. We note that $\forall\, l \in \{1, 2, \ldots, a_{K_n} - 1\}$,

$$\{\mathcal{G}_1^l\}_i = a_{K_n} = (a_{K_n} - 1) + 1 \geq \{\mathcal{G}_{2'}^l\}_i + 1 > \{\mathcal{G}_{2'}^l\}_i \tag{26}$$

$$\{\mathcal{G}_1^l\}_i - \{\mathcal{G}_1^l\}_j = l = \{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j \tag{27}$$

$$\{\mathcal{G}_1^l\}_i + \{\mathcal{G}_1^l\}_j = a_{K_n} + \{\mathcal{G}_1^l\}_j = 2a_{K_n} - l > 2(a_{K_n} - 1) - l \geq 2\{\mathcal{G}_{2'}^l\}_i - l \geq \{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j \tag{28}$$

Recall from (25), our goal is to find a sufficient condition such that the summation of $A_{K+1}(i,j) - A_K(i,j)$ over all (i,j) in $\mathcal{G}_1 \cup \mathcal{G}_{2'}$ is positive. We describe all the elements in $\mathcal{G}_1 \cup \mathcal{G}_{2'}$ by considering the first $(a_{K_n} - 2)$ rows of the $\mathcal{G}_1^l$, $\mathcal{G}_{2'}^l$ pair plus the last three elements at the $(a_{K_n} - 1)$-th and the a_{K_n}-th rows (namely $\mathcal{G}_1^{a_{K_n}-1}$, $\mathcal{G}_{2'}^{a_{K_n}-1}$ and $\mathcal{G}_1^{a_{K_n}}$) from Table 1 . Therefore, it is sufficient for (25) to hold when (a) $\forall\, l \in \{1, 2, \ldots, a_{K_n} - 2\}$, $\sum_{(i,j)\in\{\mathcal{G}_1^l, \mathcal{G}_{2'}^l\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$ and (b) $\sum_{(i,j)\in\{\mathcal{G}_1^{a_{K_n}-1}, \mathcal{G}_{2'}^{a_{K_n}-1}, \mathcal{G}_1^{a_{K_n}}\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$.

We first show that (a) is true for all ρ. $\forall\, l \in \{1, 2, \ldots, a_{K_n} - 2\}$. Recall that $[A_{K+1}(i,j) - A_K(i,j)]$ evaluated at $(i,j) = \mathcal{G}_1^l$ is negative, and $[A_{K+1}(i,j) - A_K(i,j)]$ evaluated at $(i,j) = \mathcal{G}_{2'}^l$ is positive. It is thus equivalent to show that $[A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=\mathcal{G}_1^l} < [A_{K+1}(i,j) - A_K(i,j)]\Big|_{(i,j)=\mathcal{G}_{2'}^l}$.

Denote $d = \min\{f_K(a_{K_1}), f_K(a_{K_n})\} > 0$, we have from (24) that

$$
\begin{aligned}
& [A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=\mathcal{G}_1^l} \\
= \;& (\{\mathcal{G}_1^l\}_i - \{\mathcal{G}_1^l\}_j)\rho^{\{\mathcal{G}_1^l\}_i + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_1^l\}_j}^{\{\mathcal{G}_1^l\}_i - 1} \bar{F}_K(n) - \prod_{n=\{\mathcal{G}_1^l\}_j}^{\{\mathcal{G}_1^l\}_i - 1} \bar{F}_{K+1}(n) \right) \\
= \;& (a_{K_n} - \{\mathcal{G}_1^l\}_j)\rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n} - 1} \bar{F}_K(n) - \prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n} - 1} \bar{F}_{K+1}(n) \right) \\
= \;& l \cdot \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n} - 1} \bar{F}_K(n) - \prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n} - 1} \bar{F}_{K+1}(n) \right) \\
< \;& l \cdot \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_K(n) \left(d \cdot \prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n} - 2} \bar{F}_K(n) \right) \\
= \;& l \cdot d \cdot \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_{K+1}(n) \prod_{n=0}^{a_{K_n} - 2} \bar{F}_K(n) \\
< \;& (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \cdot d \cdot \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j - 1} \bar{F}_{K+1}(n) \prod_{n=0}^{a_{K_n} - 2} \bar{F}_K(n) \\
& \text{because } l = \{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j \text{ see (27), } \rho < 1 \text{ and } a_{K_n} + \{\mathcal{G}_1^l\}_j > \{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j \text{ see (28)} \\
\leq \;& (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \cdot d \cdot \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=0}^{a_{K_n} - 2} \bar{F}_K(n) \\
\leq \;& (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \cdot d \cdot \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=a_{K_1}+1}^{\{\mathcal{G}_{2'}^l\}_i - 1} \bar{F}_K(n) \text{ because } a_{K_n} - 2 \geq \{\mathcal{G}_{2'}^l\}_i - 1 \text{ see (26)}
\end{aligned}
$$

$$= (\{\mathcal{G}^l_{2'}\}_i - \{\mathcal{G}^l_{2'}\}_j)\rho^{\{\mathcal{G}^l_{2'}\}_i+\{\mathcal{G}^l_{2'}\}_j} \prod_{n=0}^{\{\mathcal{G}^l_{2'}\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}^l_{2'}\}_j-1} \bar{F}_K(n) \left(d \cdot \prod_{n=a_{K_1}+1}^{\{\mathcal{G}^l_{2'}\}_i-1} \bar{F}_K(n) \right)$$

because $\prod_{n=0}^{\{\mathcal{G}^l_{2'}\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}^l_{2'}\}_j-1} \bar{F}_K(n) = 1$

$$= (\{\mathcal{G}^l_{2'}\}_i - \{\mathcal{G}^l_{2'}\}_j)\rho^{\{\mathcal{G}^l_{2'}\}_i+\{\mathcal{G}^l_{2'}\}_j} \prod_{n=0}^{\{\mathcal{G}^l_{2'}\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}^l_{2'}\}_j-1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}^l_{2'}\}_j}^{\{\mathcal{G}^l_{2'}\}_i-1} \bar{F}_{K+1}(n) - \prod_{n=\{\mathcal{G}^l_{2'}\}_j}^{\{\mathcal{G}^l_{2'}\}_i-1} \bar{F}_K(n) \right)$$

because $\prod_{n=\{\mathcal{G}^l_{2'}\}_j}^{a_{K_1}-1} \bar{F}_{K+1}(n) = \prod_{n=\{\mathcal{G}^l_{2'}\}_j}^{a_{K_1}-1} \bar{F}_K(n) = 1,\ \bar{F}_{K+1}(a_{K_1}) - \bar{F}_K(a_{K_1}) = d$

and $\bar{F}_{K+1}(x) = \bar{F}_K(x),\ \forall x \in \{a_{K_1}+1, \ldots, \{\mathcal{G}^l_{2'}\}_i - 1\}$

$$= [A_{K+1}(i,j) - A_K(i,j)]\Big|_{(i,j)=\mathcal{G}^l_{2'}}, \text{ as required.}$$

Next, for (b) to hold, i.e., $\sum_{(i,j)\in\{\mathcal{G}_1^{a_{K_n}-1}, \mathcal{G}_{2'}^{a_{K_n}}, \mathcal{G}_1^{a_{K_n}}\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$, equivalently we have:

$$[A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=\mathcal{G}_1^{a_{K_n}-1}} + [A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=\mathcal{G}_1^{a_{K_n}}} < [A_{K+1}(i,j) - A_K(i,j)]\Big|_{(i,j)=\mathcal{G}_{2'}^{a_{K_n}-1}} \Leftrightarrow$$

$$[A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=(a_{K_n},1)} + [A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=(a_{K_n},0)} < [A_{K+1}(i,j) - A_K(i,j)]\Big|_{(i,j)=(a_{K_n}-1,0)} \quad (29)$$

Note that (with any empty product being equal to $= 1$)

$$\begin{aligned}
[A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=(a_{K_n},1)} &= (a_{K_n}-1)\rho^{a_{K_n}+1} \cdot \left(\prod_{n=1}^{a_{K_n}-1} \bar{F}_{K+1}(n) - \prod_{n=1}^{a_{K_n}-1} \bar{F}_K(n) \right) \\
&< (a_{K_n}-1)\rho^{a_{K_n}+1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \\
[A_K(i,j) - A_{K+1}(i,j)]\Big|_{(i,j)=(a_{K_n},0)} &= a_{K_n}\rho^{a_{K_n}} \cdot \left(\prod_{n=0}^{a_{K_n}-1} \bar{F}_{K+1}(n) - \prod_{n=0}^{a_{K_n}-1} \bar{F}_K(n) \right) \\
&< a_{K_n}\rho^{a_{K_n}} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \\
[A_{K+1}(i,j) - A_K(i,j)]\Big|_{(i,j)=(a_{K_n}-1,0)} &= (a_{K_n}-1)\rho^{a_{K_n}-1} \cdot \left(\prod_{n=0}^{a_{K_n}-2} \bar{F}_{K+1}(n) - \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n) \right) \\
&= (a_{K_n}-1)\rho^{a_{K_n}-1} \cdot \left(\prod_{n=a_{K_1}}^{a_{K_n}-2} \bar{F}_{K+1}(n) - \prod_{n=a_{K_1}}^{a_{K_n}-2} \bar{F}_K(n) \right) \\
&= (a_{K_n}-1)\rho^{a_{K_n}-1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n).
\end{aligned}$$

Therefore, it is sufficient for (29) to hold if

$$(a_{K_n}-1)\rho^{a_{K_n}+1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) + a_{K_n}\rho^{a_{K_n}} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) \quad \leq \quad (a_{K_n}-1)\rho^{a_{K_n}-1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n),$$

$$\begin{aligned}\Leftrightarrow (a_{K_n}-1)\rho^{a_{K_n}+1}+a_{K_n}\rho^{a_{K_n}} &\leq (a_{K_n}-1)\rho^{a_{K_n}-1},\\ \Leftrightarrow \rho^2+\frac{a_{K_n}}{a_{K_n}-1}\rho &\leq 1. \end{aligned} \quad (30)$$

Solving quadratic equation (30) gives the condition

$$\frac{1}{2}\left(-\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4}-\frac{a_{K_n}}{a_{K_n}-1}\right)\leq \rho \leq \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4}-\frac{a_{K_n}}{a_{K_n}-1}\right).$$

where it is clear that $\frac{1}{2}\left(-\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4}-\frac{a_{K_n}}{a_{K_n}-1}\right)<0$ and $0<\frac{1}{2}\left(\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4}-\frac{a_{K_n}}{a_{K_n}-1}\right)<1$.

Since $\rho\in(0,1)$, we have that, when $\rho\leq\frac{1}{2}\left(\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4}-\frac{a_{K_n}}{a_{K_n}-1}\right)$, (30) (29),(25) and (23) all hold and thus $L_{\tilde{N}_K}<L_{\tilde{N}_{K+1}}$. This completes the proof of part (ii).

(iii) Recall from (18) that

$$W_{\tilde{N}'}\geq W_{\tilde{N}} \Leftrightarrow \sum_{i,j\geq 1}\sum_{i>j:}(i-j)\rho^{i+j}\prod_{n=0}^{j-1}\bar{F}_{\tilde{N}'}(n)\prod_{n=0}^{j-1}\bar{F}_{\tilde{N}}(n)\left(\prod_{n=j}^{i-1}\bar{F}_{\tilde{N}'}(n)-\prod_{n=j}^{i-1}\bar{F}_{\tilde{N}}(n)\right)\geq 0.$$

Thus, $W_{\tilde{N}'}>W_{\tilde{N}}\Leftrightarrow \sum_{a_{K_n}\geq i>}\sum_{j\geq 1}[A_{K+1}(i,j)-A_K(i,j)]>0$. The rest of the proof is then almost identical to the proof of part (ii) except now i,j cannot take on 0. Define $\mathcal{G}_1\triangleq\{a_{K_n}\}\times\{1,2,\ldots,a_{K_n}-1\}$, $\mathcal{G}_2\triangleq\{a_{K_1}+1,a_{K_1}+2,...,a_{K_n}-2,a_{K_n}-1\}\times\{1,2,\ldots,a_{K_1}\}$, and $\mathcal{G}_3\triangleq\{(i,j):a_{K_n}\geq i>j\geq 1\}-\{\mathcal{G}_1\cup\mathcal{G}_2\}$.

There are now at least $a_{K_n}-2$ elements in the set $\mathcal{G}_2$ which defines the subset $\mathcal{G}_{2'}$. The elements in $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ are ordered in a similar fashion as before and displayed in Table 2. A sufficient condition for $W_{\tilde{N}_K}<W_{\tilde{N}_{K+1}}$ is that $\sum_{(i,j)\in\mathcal{G}_1\cup\mathcal{G}_{2'}}[A_{K+1}(i,j)-A_K(i,j)]>0$. It can be shown that conditions (26)-(28) still hold, and thus $\forall\ l\in\{1,2,\ldots,a_{K_n}-3\}$ and for all ρ, $\sum_{(i,j)\in\{\mathcal{G}_1^l,\mathcal{G}_{2'}^l\}}[A_{K+1}(i,j)-A_K(i,j)]>0$. A sufficient condition for $W_{\tilde{N}_K}<W_{\tilde{N}_{K+1}}$ is then that $\sum_{(i,j)\in\{\mathcal{G}_1^{a_{K_n}-2},\mathcal{G}_{2'}^{a_{K_n}-2},\mathcal{G}_1^{a_{K_n}-1}\}}[A_{K+1}(i,j)-A_K(i,j)]>0$.

$$\begin{aligned}\text{Since } [A_K(i,j)-A_{K+1}(i,j)]\Big|_{(i,j)=(a_{K_n},2)} &< (a_{K_n}-2)\rho^{a_{K_n}+2}\cdot d\cdot\prod_{n=a_{K_1}+1}^{a_{K_n}-2}\bar{F}_K(n),\\ [A_K(i,j)-A_{K+1}(i,j)]\Big|_{(i,j)=(a_{K_n},1)} &< (a_{K_n}-1)\rho^{a_{K_n}+1}\cdot d\cdot\prod_{n=a_{K_1}+1}^{a_{K_n}-2}\bar{F}_K(n),\\ \text{and } [A_{K+1}(i,j)-A_K(i,j)]\Big|_{(i,j)=(a_{K_n}-1,1)} &= (a_{K_n}-2)\rho^{a_{K_n}}\cdot d\cdot\prod_{n=a_{K_1}+1}^{a_{K_n}-2}\bar{F}_K(n),\ \text{we have}\end{aligned}$$

$$(a_{K_n}-2)\rho^{a_{K_n}+2}\cdot d\cdot\prod_{n=a_{K_1}+1}^{a_{K_n}-2}\bar{F}_K(n)+(a_{K_n}-1)\rho^{a_{K_n}+1}\cdot d\cdot\prod_{n=a_{K_1}+1}^{a_{K_n}-2}\bar{F}_K(n)\leq(a_{K_n}-2)\rho^{a_{K_n}}\cdot d\cdot\prod_{n=a_{K_1}+1}^{a_{K_n}-2}\bar{F}_K(n),$$

$$\Leftrightarrow (a_{K_n}-2)\rho^{a_{K_n}+2}+(a_{K_n}-1)\rho^{a_{K_n}+1}\leq(a_{K_n}-2)\rho^{a_{K_n}},$$

$$\Leftrightarrow \rho^2+\frac{a_{K_n}-1}{a_{K_n}-2}\rho\leq 1.$$

The solution of the quadratic inequality on the set $\rho\in(0,1)$ is $\rho\leq\frac{1}{2}\left(\sqrt{(\frac{a_{K_n}-1}{a_{K_n}-2})^2+4}-\frac{a_{K_n}-1}{a_{K_n}-2}\right)$. □

Proof of Theorem 2:

(i) Result follows immediately from Lemma 4/(i) since $R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$ for all K.

(ii) Recall from Lemma 4/(ii) that $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ if $\rho \le \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4} - \frac{a_{K_n}}{a_{K_n}-1}\right)$. It can be easily verified that $\frac{1}{2}\left(\sqrt{(\frac{a_{K_n}}{a_{K_n}-1})^2+4} - \frac{a_{K_n}}{a_{K_n}-1}\right)$ increases in a_{K_n}. Plugging in the smallest possible value of a_{K_n} which is 3, we get $\rho = 0.5$. Therefore, when $\rho \le 0.5$, $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ for all K (regardless of the distributions of $\{\tilde{N}_K\}_{K=0,1,2,\ldots,T}$). Result thus follows.

(iii) Recall from Lemma 4/(iii) that $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ if $\rho \le \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}-1}{a_{K_n}-2})^2+4} - \frac{a_{K_n}-1}{a_{K_n}-2}\right)$. It can be verified that $\frac{1}{2}\left(\sqrt{(\frac{a_{K_n}-1}{a_{K_n}-2})^2+4} - \frac{a_{K_n}-1}{a_{K_n}-2}\right)$ increases in a_{K_n}. Plugging in the smallest possible value of a_{K_n} which is 3, we get $\rho = 0.414$. Therefore, when $\rho \le 0.414$, $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ for all K. Result thus follows. □

Proof of Theorem 1':

(i) From (11), we then have $R_{\tilde{N}} = p\cdot\mu[s-(s\pi_0+(s-1)\pi_1+\ldots+2\pi_{s-2}+1\pi_{s-1})]$ so $R_{\tilde{N}}$ is decreasing in π_0 (given that p,μ,λ,ρ,s are all fixed). Since π_0 itself is decreasing in $\sum_{i=1}^{\infty}\rho^i\prod_{n=0}^{i-1}\bar{F}(n)$ from (10), we have $R_{\tilde{N}}$ increases in $\sum_{i=1}^{\infty}\rho^i\prod_{n=0}^{i-1}\bar{F}(n)$. The rest of the proof follows the proof of Theorem 1/(i).

(ii) Proof is similar to that of Theorem 1/(ii). Result follows because $L_{\tilde{N}} \le L_{\tilde{N}'}$ if and only if

$$\sum_{i,j\ge 0:}\sum_{i>j}(i-j)\frac{\rho^{i+j}}{(i\wedge s)!(j\wedge s)!}\prod_{n=0}^{j-s}\bar{F}_{\tilde{N}'}(n)\prod_{n=0}^{j-s}\bar{F}_{\tilde{N}}(n)\left(\prod_{n=j-s+1}^{i-s}\bar{F}_{\tilde{N}'}(n)-\prod_{n=j-s+1}^{i-s}\bar{F}_{\tilde{N}}(n)\right)\ge 0.$$

(iii) Proof is similar to that of Theorem 1/(iii). We have $W_{\tilde{N}} \le W_{\tilde{N}'}$ if and only if

$$\sum_{i,j\ge 1:}\sum_{i>j}(i(j\wedge s)-j(i\wedge s))\frac{\rho^{i+j}}{(i\wedge s)!(j\wedge s)!}\prod_{n=0}^{j-s}\bar{F}_{\tilde{N}'}(n)\prod_{n=0}^{j-s}\bar{F}_{\tilde{N}}(n)\left(\prod_{n=j-s+1}^{i-s}\bar{F}_{\tilde{N}'}(n)-\prod_{n=j-s+1}^{i-s}\bar{F}_{\tilde{N}}(n)\right)\ge 0$$

and result follows because $(i(j\wedge s)-j(i\wedge s))\ge 0$ for all $\{i,j\ge 1: i>j\}$. □

Proof of Theorem 2':

We will prove the following lemma (a general version of Lemma 4 but with the $M/M/s$ queue setting) then the results of Theorem 2' immediately follow.

Let $\{\tilde{N}_K\}$ be any sequence from Construction 1 in an $M/M/s$ queue. We can show (i) $R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$ for all ρ; (ii) $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ if $\rho \le \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}+s-1}{a_{K_n}+s-2})^2+4} - \frac{a_{K_n}+s-1}{a_{K_n}+s-2}\right)$; And (iii) when $s=1$, $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ if $\rho \le \frac{1}{2}\left(\sqrt{(\frac{a_{K_n}-1}{a_{K_n}-2})^2+4} - \frac{a_{K_n}-1}{a_{K_n}-2}\right)$; when $s\ge 2$, $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ if $\rho \le \frac{1}{2}\left(\sqrt{1+4(\frac{a_{K_n}-2}{a_{K_n}-1})}-1\right)$.

(i) Recall from the proof of Theorem 1'/(i), $R_{\tilde{N}}$ increases in $\sum_{i=1}^{\infty}\rho^i\prod_{n=0}^{i-1}\bar{F}(n)$. Therefore it is sufficient to show that $\sum_{i=1}^{\infty}\rho^i\prod_{n=0}^{i-1}\bar{F}_K(n) < \sum_{i=1}^{\infty}\rho^i\prod_{n=0}^{i-1}\bar{F}_{K+1}(n)$. Result then follows the proof of Lemma 4/(i).

(ii) We can apply the same approach used in the proof of Lemma 4/(ii) here. Define

$$A_{K+1}(i,j) \triangleq (i-j)\frac{\rho^{i+j}}{(i\wedge s)!(j\wedge s)!}\prod_{n=0}^{j-s}\bar{F}_{K+1}(n)\prod_{n=0}^{j-s}\bar{F}_K(n)\prod_{n=j-s+1}^{i-s}\bar{F}_{K+1}(n)$$

$$A_K(i,j) \triangleq (i-j)\frac{\rho^{i+j}}{(i\wedge s)!(j\wedge s)!}\prod_{n=0}^{j-s}\bar{F}_{K+1}(n)\prod_{n=0}^{j-s}\bar{F}_K(n)\prod_{n=j-s+1}^{i-s}\bar{F}_K(n)$$

Order $l =$	$\mathcal{G}_1$ contains:	$\mathcal{G}_{2'}$ contains:
1	$(a_{K_n}+s-1, a_{K_n}+s-2)$	$(a_{K_1}+s, a_{K_1}+s-1)$
2	$(a_{K_n}+s-1, a_{K_n}+s-3)$	$(a_{K_1}+s, a_{K_1}+s-2)$
3	$(a_{K_n}+s-1, a_{K_n}+s-4)$	$(a_{K_1}+s, a_{K_1}+s-3)$
$\vdots$	$\vdots$	$\vdots$
$a_{K_1}-1$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1})$	$(a_{K_1}+s, 2)$
a_{K_1}	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-1)$	$(a_{K_1}+s, 1)$
$a_{K_1}+1$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-2)$	$(a_{K_1}+s, 0)$
$a_{K_1}+2$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-3)$	$(a_{K_1}+s+1, 0)$
$a_{K_1}+3$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-4)$	$(a_{K_1}+s+2, 0)$
$\vdots$	$\vdots$	$\vdots$
$a_{K_n}+s-3$	$(a_{K_n}+s-1, 2)$	$(a_{K_n}+s-3, 0)$
$a_{K_n}+s-2$	$(a_{K_n}+s-1, 1)$	$(a_{K_n}+s-2, 0)$
$a_{K_n}+s-1$	$(a_{K_n}+s-1, 0)$	

Table 3: $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ used in the proof of Theorem 2'/(ii)

Order $l =$	$\mathcal{G}_1$ contains:	$\mathcal{G}_{2'}$ contains:
1	$(a_{K_n}+s-1, a_{K_n}+s-2)$	$(a_{K_1}+s, a_{K_1}+s-1)$
2	$(a_{K_n}+s-1, a_{K_n}+s-3)$	$(a_{K_1}+s, a_{K_1}+s-2)$
3	$(a_{K_n}+s-1, a_{K_n}+s-4)$	$(a_{K_1}+s, a_{K_1}+s-3)$
$\vdots$	$\vdots$	$\vdots$
$a_{K_1}-1$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1})$	$(a_{K_1}+s, 2)$
a_{K_1}	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-1)$	$(a_{K_1}+s, 1)$
$a_{K_1}+1$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-2)$	$(a_{K_1}+s+1, 1)$
$a_{K_1}+2$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-3)$	$(a_{K_1}+s+2, 1)$
$a_{K_1}+3$	$(a_{K_n}+s-1, a_{K_n}+s-a_{K_1}-4)$	$(a_{K_1}+s+3, 1)$
$\vdots$	$\vdots$	$\vdots$
$a_{K_n}+s-3$	$(a_{K_n}+s-1, 2)$	$(a_{K_n}+s-2, 1)$
$a_{K_n}+s-2$	$(a_{K_n}+s-1, 1)$	

Table 4: $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ used in the proof of Theorem 2'/(iii)

It can be shown that $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ if $\sum_{(i,j)\in} \sum_{\mathcal{G}_1 \cup \mathcal{G}_{2'}} [A_{K+1}(i,j) - A_K(i,j)] > 0$ where the elements of $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ are listed in Table 3. It then can be verified that conditions (26)-(28) still hold, and that $\forall\, l \in \{1, 2, \ldots, a_{K_n} + s - 3\}$ and for all ρ, $\sum_{(i,j)\in\{\mathcal{G}_1^l, \mathcal{G}_{2'}^l\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$. One sufficient condition for $\sum_{(i,j)\in\{\mathcal{G}_1^{a_{K_n}+s-2}, \mathcal{G}_{2'}^{a_{K_n}+s-2}, \mathcal{G}_1^{a_{K_n}+s-1}\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$, which also makes $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$, is that $(a_{K_n} + s - 2)\rho^{a_{K_n}+s} + (a_{K_n} + s - 1)\rho^{a_{K_n}+s-1} \leq (a_{K_n} + s - 2)\rho^{a_{K_n}+s-2}$. It then follows by solving the quadratic equation that $\rho \leq \frac{1}{2}\left(\sqrt{\frac{(a_{K_n}+s-1)^2}{(a_{K_n}+s-2)^2} + 4} - \frac{a_{K_n}+s-1}{a_{K_n}+s-2}\right)$.

(iii) We can apply the same approach used in the proof of Lemma 4/(iii) here. Define

$$A_{K+1}(i,j) \triangleq (i(j \wedge s) - j(i \wedge s)) \frac{\rho^{i+j}}{(i \wedge s)!(j \wedge s)!} \prod_{n=0}^{j-s} \bar{F}_{K+1}(n) \prod_{n=0}^{j-s} \bar{F}_K(n) \prod_{n=j-s+1}^{i-s} \bar{F}_{K+1}(n)$$

$$A_K(i,j) \triangleq (i(j \wedge s) - j(i \wedge s)) \frac{\rho^{i+j}}{(i \wedge s)!(j \wedge s)!} \prod_{n=0}^{j-s} \bar{F}_{K+1}(n) \prod_{n=0}^{j-s} \bar{F}_K(n) \prod_{n=j-s+1}^{i-s} \bar{F}_K(n)$$

It can be shown that $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ if $\sum_{(i,j)\in} \sum_{\mathcal{G}_1 \cup \mathcal{G}_{2'}} [A_{K+1}(i,j) - A_K(i,j)] > 0$ where the elements of $\mathcal{G}_1$ and $\mathcal{G}_{2'}$ are listed in Table 4. It then can be verified that $\forall\, l \in \{1, 2, \ldots, a_{K_n} + s - 4\}$ and for all ρ, $\sum_{(i,j)\in\{\mathcal{G}_1^l, \mathcal{G}_{2'}^l\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$. Further, when $s \geq 2$ (the case when $s = 1$ is proved in Lemma 4/(iii)), one sufficient condition for $\sum_{(i,j)\in\{\mathcal{G}_1^{a_{K_n}+s-3}, \mathcal{G}_{2'}^{a_{K_n}+s-3}, \mathcal{G}_1^{a_{K_n}+s-2}\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$, which leads to $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$, is that $(a_{K_n} - 1)\rho^{a_{K_n}+s+1} + (a_{K_n} - 1)\rho^{a_{K_n}+s} \leq (a_{K_n} - 2)\rho^{a_{K_n}+s-1}$. It then follows by solving the quadratic equation that $\rho \leq \frac{1}{2}\left(\sqrt{1 + 4(\frac{a_{K_n}-2}{a_{K_n}-1})} - 1\right)$.

Proof of Proposition 1:

Consider the random variable $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}) \rfloor, \lceil \mathbb{E}(\tilde{N}) \rceil\}$ such that $E(\tilde{N}_T) = E(\tilde{N})$. By Theorem 2, we have $R_{\tilde{N}} \leq R_{\tilde{N}_T}$ for all ρ, and $L_{\tilde{N}} \leq L_{\tilde{N}_T}$, $W_{\tilde{N}} \leq W_{\tilde{N}_T}$ for small ρ. On the other hand, since $\mathbb{E}(\tilde{N}) \leq N$ and N is an integer, we must have $\lceil \mathbb{E}(\tilde{N}) \rceil \leq N$. It follows that $\tilde{N}_T \leq_{st} N$ so by Theorem 1, we have $R_{\tilde{N}_T} \leq R_N$ for all ρ, and $L_{\tilde{N}_T} \leq L_N$, $W_{\tilde{N}_T} \leq W_N$ for small ρ. Result thus follows. $\square$

References

Aksin, Z., M. Armony, and V. Mehrotra (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management 16*(6), 665–688.

Armony, M. and C. Maglaras (2004a). Contact centers with a call-back option and real-time delay information. *Operations Research 52*(4), 527–545.

Armony, M. and C. Maglaras (2004b). On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Operations Research 52*(2), 271–292.

Besbes, O., B. Dooley, and N. Gans (2011). Dynamic service control of a queue with congestion-sensitive customers. *Working Paper*.

Besbes, O. and C. Maglaras (2009). Revenue optimization for a make-to-order queue in an uncertain market environment. *Operations Research 57*(6), 1438–1450.

Debo, L. G. and S. Veeraraghavan (2011). Equilibrium in queues under unknown service rates and service value. *Chicago Booth Research Paper* (11-45).

Economou, A. and S. Kanta (2008). Optimal balking strategies and pricing for the single server markovian queue with compartmented waiting space. *Queueing Systems 59*(3), 237–269.

Gaur, V. and Y. Park (2007). Asymmetric consumer learning and inventory competition. *Management Science 53*(2), 227–240.

Guo, P., W. Sun, and Y. Wang (2011). Equilibrium and optimal strategies to join a queue with partial information on service times. *European Journal of Operational Research 214*(2), 284–297.

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

Guo, P. and P. Zipkin (2009). The effects of the availability of waiting-time information on a balking queue. *European Journal of Operational Research 198*(1), 199–209.

Hassin, R. (1986). Consumer information in markets with random product quality: The case of queues and balking. *Econometrica 54*(5), 1185–1195.

Hassin, R. (2007). Information and uncertainty in a queuing system. *Probability in the Engineering and Informational Sciences 21*(3), 361.

Hassin, R. and M. Haviv (2003). *To queue or not to queue: Equilibrium behavior in queueing systems*, Volume 59. Springer International Series in Operations Research & Management Science.

Haviv, M. and R. Randhawa (2011). Pricing in queues without demand information. *Working Paper*.

Huang, T., G. Allon, and A. Bassamboo (2012). Bounded rationality in service systems. *M&SOM, Forthcoming*.

Jouini, O., Z. Akşin, and Y. Dallery (2011). Call centers with delay information: Models and insights. *M&SOM 13*(4), 534–548.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

Luce, R. (1959). *Individual Choice Behavior.* Wiley, New York.

Müller, A. and D. Stoyan (2002). *Comparison methods for stochastic models and risks*, Volume 389. Wiley Series in Probability and Statistics.

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

Nelson, P. (1974). Advertising as information. *Journal of Political Economy 82*(4), 729–754.

Olivares, M., Y. Lu, A. Musalem, and A. Schilkrut (2011). Measuring the effect of queues on customer purchases. *Columbia Business School Research Paper* (12/23).

Quirk, J. P. and R. Saposnik (1962). Admissibility and measurable utility functions. *The Review of Economic Studies*, 140–146.

Rothschild, M. and J. E. Stiglitz (1970). Increasing risk: I. a definition. *J. of Econ. Theory 2*(3), 225–243.

Su, X. (2008). Bounded rationality in newsvendor models. *M&SOM 10*(4), 566–589.

Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology 5*(1), 207–233.

Whitt, W. (1999). Improving service by informing customers about anticipated delays. *Management science 45*(2), 192–207.

Wolff, R. (1982). Poisson arrivals see time averages. *Operations Research 30*(2), 223–231.

Xu, S., L. Gao, and J. Ou (2007). Service performance analysis and improvement for a ticket queue with balking customers. *Management Science 53*(6), 971–990.