Equilibrium in Queues under Unknown Service Times and Service Value

Laurens Debo
Chicago Booth School of Business
University of Chicago
Chicago, IL 60637
Laurens.Debo@ChicagoBooth.edu

Senthil Veeraraghavan
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
senthilv@wharton.upenn.edu

July 3, 2012

Abstract

In Naor’s seminal queue-joining model, queue-joining probabilities decrease monotonously in the queue length; the longer the queue, the fewer consumers join. In practice, empirical evidence indicates that queue-joining probabilities may not always be decreasing in the queue length. For example, for restaurants, long queues may sometimes be more attractive than short queues. We rationalize non-monotonic strategies by relaxing the information assumptions in Naor’s model. Instead of assuming that the expected service time and service value are common knowledge, we assume that they are unknown to consumers, but positively correlated. Under such informational assumptions, we show that equilibria may emerge for which the joining probability increases in the queue length. We refer to these as “sputtering equilibria.” We discuss when and why such sputtering equilibria exist for discrete as well as continuously distributed priors on the expected service time (with positively correlated service value).

Keywords: Queueing games, Threshold policies, Randomization.

1 Introduction

Management of services is fundamentally influenced by how consumers consider congestion or queue lengths in making their joining decisions. Naor (1969) was the first to introduce consumers that make rational queue-joining decisions based on the queue length upon arrival. In Naor’s model, the (dominating) equilibrium queue-joining strategy is a threshold strategy: Consumers join the queue
if the length of the queue is below a threshold. Above the threshold, waiting costs dominate the service value, and hence consumers balk. Note the threshold decision is an outcome of a comparison between service value and waiting costs, with the latter increasing in queue lengths. As a result, in Naor’s seminal model (and in most of the literature), the joining rate is monotonically decreasing in the queue length, fewer consumers join a queue when the queue grows longer. This aspect of the consumer queue-joining research was extensively explored in the book by Hassin and Haviv (2003).

There is anecdotal and empirical evidence that the rates at which consumers join queues may not be monotonically decreasing in the queue length. This idea is in fact long grounded as folk wisdom in the restaurant industry: “If one walks into a restaurant and there’s nobody eating in the establishment already, a consumer is less likely to hang around, regardless of the expected food quality. Conversely, if a consumer walks in and the place is bustling, [he’s] probably going to give it a shot” (Waldman, 2009). The notion of “empty restaurant syndrome”—a restaurant parlance that describes patrons not joining a restaurant when it is empty—also supports this observation. Recently, a number of research findings support the above mentioned ideas of nonmonotonic queue-joining behavior. In an empirical study of queues at a deli, Olivares et al. (2012) find that purchase incidence is non-monotone in queue lengths. Buell and Norton (2011a, 2011b) find evidence that consumers can join longer queues with a higher affinity, a behavior that is further strengthened by the imperfect information in consumers’ minds. Giebelhausen et al. (2011), experimentally show that waiting times can indeed be a signal about quality increasing both purchase intentions and experienced satisfaction when quality is important, unknown or ambiguous.

The above observations imply that the joining rate may not always be monotonically decreasing in queue length: At shorter queue lengths, consumers may be more reluctant to join than at longer queue lengths. Such behavior cannot be explained by canonical threshold queue-joining models.

One main question in this research is to explore the antecedents of such nonmonotonic queue-joining behavior, especially those factors that are related to uncertainty about the service process characteristics. To elaborate, consumers may not exactly know a firm’s expected service time and value \textit{ex ante} (before joining). In such cases, the consumers’ decisions are influenced by uncertainties in service value and cost. We show that a simple relaxation of the information assumptions in Naor’s canonical model can result in a nonmonotonic queue-joining equilibrium structure.

To be sure, nonthreshold joining strategies have been identified in queuing games with priorities (Hassin and Haviv, 1997). In our paper, we do not consider priorities. More recent papers (Debo et al., 2011, 2012; Veeraraghavan and Debo, 2011) study the informational externalities associated with queue-joining decisions under waiting costs. In these papers, such queue-joining strategies emerge due to heterogeneity: Some consumers have better information about the service value than others. In our paper, we consider homogeneously informed consumers.

Instead, we rationalize the nonmonotonic queue-joining behavior observed in practice by relaxing Naor’s assumption that the expected service time and value are common knowledge. Thus, our
model can be considered as a simple, but theoretically and practically consequential generalization of Naor's seminal queue-joining model.

When there is uncertainty about the service value and time, dependencies between service value and service time become relevant. In many services with nonmonotonic queue-joining, service value has a positive correlation with service time. For instance, for labor-intensive services, consumers might enjoy an increased utility from longer service times as more care is given during the longer service time. Since longer waiting times in queues are a direct result of longer service times, it can also be construed that the consumers perceive an added utility in joining longer queue lengths. In fact, the nonmonotonic queue joining observed by Buell and Norton (2011a) could arise from the labor illusion—a higher value perceived from a higher labor content in the longer wait time. Anand et al. (2011) provide evidence for a variety of industries in which the offered service value is positively correlated with the wait times. Such observations of service time and value dependencies have now found stronger theoretical and empirical support in various papers. Dai et al. (2011) show that in medical services, the service quality (or value), increases in the diagnostic testing time. We refer the reader to Hopp et al. (2007), Wang et al. (2009) and Alizamir et al. (2012) for services whose value increases in the service time (discretionary services), to Kostami and Rajagopalan (2009) for dynamic tradeoffs in such services, and to Mold et al. (2010) for informational returns on clinical tests. To summarize, it appears that in many service settings, longer service times are positively correlated with increased service value.

Motivated by the above empirical and theoretical observations, we focus in this paper on the condition where both expected service time and service value are positively correlated random variables of which consumers only know the joint distribution, not the exact realization. We demonstrate that such a simple relaxation may result in queue-joining probabilities that are nonmonotonic in the queue length, as is suggested by folk wisdom and as observed empirically.

As in Naor’s model, arriving consumers must make an instantaneous decision to join or not. We consider a market in which consumers observe the queue length to learn about more than just the expected service and waiting time (and costs). Longer queues may imply longer service duration and hence higher waiting costs. However, longer service duration also implies a higher expected service value from joining. As a consequence, both the expected cost and value increase as a function of the queue length. Pure threshold joining strategies may still emerge, but not necessarily. When no pure threshold joining equilibrium exists, we demonstrate that an equilibrium exists with randomization at one (or more) queue lengths below some threshold. This creates nonmonotonicity in joining

---

1 There could be negative correlation between service value and expected service time. This assumption is consistent with traditional models. In this case, we demonstrate that Naor’s threshold equilibrium is recovered. We focus on the positive correlation case purely for the sake of brevity.

2 To make a straightforward comparison with Naor’s model, we also assume that there is no reneging. Very few papers on reneging address individual strategic consumer considerations. We refer to Assaf and Haviv (1990) and Mandelbaum and Shimkin (2000). In both these papers, either the states are unobservable, or the reneging probabilities are exogenously imposed.
behavior—a consumer joins with a probability strictly less than one at a lower queue length, but may join with probability one at a higher queue length. We label such nonthreshold joining the \textit{sputtering} equilibria.

Thus, our analysis simply illustrates that a relaxation of informational assumptions in Naor’s model about two fundamental parameters, expected service time and service value, may provide an explanation for the real-world nonmonotonic queue-joining behavior, without resorting to priority queues or heterogeneously informed consumers as was done in the previous literature.

The randomization through sputtering equilibria is relevant for service firms that provide high value (at the expense of longer service times) because it may adversely affect the firm’s throughput and revenue. Therefore, our model also allows assessing the impact of unknown service time and value on the service firm’s revenue. We show that the unobservability of the expected service time and value can either increase or decrease the throughput.

A summary of our theoretical findings follows.

1. We show that \textit{pure threshold} equilibria exist when the gap in service value between the shortest and the longest time is sufficiently small, and the prior is bivalued. When this gap is large, it is possible that no pure threshold joining strategy exists.

2. When no pure threshold joining strategy exists, we show that always a \textit{sputtering} equilibrium exists when the prior is bivalued. Under such equilibria, consumers randomize between joining and balking at a lower queue length, and strictly join at higher queue lengths.

3. Sputtering equilibria exist even when the prior beliefs are continuously distributed. We find that when the prior beliefs have a continuous distribution, there could be sputtering at multiple queue lengths. Finally, we characterize prior beliefs that could generate a specific sputtering equilibrium.

2 Model

Consumers arrive sequentially at the market according to a Poisson process with parameter $\Lambda$. If the arriving consumers cannot be immediately served, they wait and form a queue. The queue discipline is first come, first served (FCFS). All consumers incur a disutility of $c > 0$ per unit time while waiting to complete the service. Service times are exponentially distributed. The mean service time and service value are unknown to the consumers. We describe this uncertainty via a joint distribution with density

$$\chi(t, v)dt dv \text{ for } \hat{t} < t < \bar{t} \text{ and } \underline{v} < v < \overline{v}.$$
We denote the marginal density of the service rate by $X(t) = \int_\mathbb{R} \chi(t,v)dv$. We use $V(t)$ to denote the expected service value conditioned on the expected service time: $\int_\mathbb{R} \int_\mathbb{R} \chi(t,v)dtdv/\int_\mathbb{R} \chi(t,v)dv$.

We assume that $V(t)$ is linear; $V_0 + rt$, where $r > 0$. This condition implies that a service that takes more time on average is associated with a higher value. The net utility of joining at queue length $n$ is $V_0 + \{r - c(n + 1)\}t$ due to congestion costs that are linear in the expected service time. Finally, we assume that

$$t < -\frac{V_0}{r-c} < \frac{\bar{t}}{t}$$ and $r > c$. (1)

This condition accommodates negative net utilities when not enough time is spent on service, even when the queue is empty upon arrival, $V_0 + (r-c)t < 0$. When the system is empty, and the service time is known to be $\bar{t}$, the service value is less than the expected waiting cost, hence, no consumer would join. When the service time is $\bar{t}$ and the system is empty, the service is valuable enough that its value exceeds the expected waiting cost. With $r > c$, this correlation structure holds for all queue lengths: The posterior value net of the waiting costs increases in the expected service time. Let $\bar{N} = \lfloor (V_0 + r\bar{t})/(c\bar{t}) \rfloor^3$. With Assumption (1), $\bar{N} \geq 1$. $\bar{N}$ is the lowest queue length at which the expected waiting time is higher than the service value at $\bar{t}$. Hence, when the expected service time is $\bar{t}$ for sure, consumers join at all queue lengths in $\{0, 1, \ldots, \bar{N} - 1\}$ and balk at $\bar{N}$.

The model primitives are ($\Lambda, V_0, r, c, X(t)$), where $X(t)$ is defined over $[\bar{t}, \bar{t}]$.

**Game:** First, Nature determines the mean service time, $\bar{t}$, and also the corresponding value, $\bar{v} = V_0 + r\bar{t}$. Then, consumers arrive and observe the queue length, $n$, based on which they decide whether to join the queue or balk. The consumers maximize their expected net utility (i.e., service value minus the expected waiting costs). After joining, the consumers stay in the system until their service is completed. As in Naor’s model, there is no reneging.

**The consumer strategies and beliefs:** It is easy to see that $\bar{N}$ is the lowest balking threshold for all $t \in [\bar{t}, \bar{t}]$, if $t$ were perfectly known. Hence, we can restrict the relevant queue lengths to $\{0, 1, \ldots, \bar{N}\}$. We denote the consumer probability of joining the service after observing the queue length $n \in \{0, 1, \ldots, \bar{N}\}$ by $\alpha(n)$. That is $\alpha(n) : \{0, 1, \ldots, \bar{N}\} \rightarrow [0, 1]$.

The consumer’s updated density of the expected service time and value (i.e., the posterior distribution) after observing a queue of length $n$ is denoted $\gamma(t, v, n) : [\bar{t}, \bar{t}] \times [\bar{v}, \bar{v}] \times \{0, 1, \ldots, \bar{N}\} \rightarrow [0, 1]$.

For notational convenience, we denote $\gamma(t, v, n)$ simply as $\gamma(n)$. In short, the vectors $\alpha = (\alpha(0), \alpha(1), \alpha(2), \ldots, \alpha(\bar{N}))$ and $\gamma = (\gamma(0), \gamma(1), \gamma(2), \ldots, \gamma(\bar{N}))$ denote the consumer’s joining strategy and the consumer’s updated belief, respectively.

If there is one queue length (say $\hat{n}$) at which no consumer joins, none of the queue lengths that are strictly higher than $\hat{n}$ will have a strictly positive long-run probability.\(^4\) Thus, it is sufficient to

\(^3\lfloor x \rfloor\) is the smallest integer less than $x \in \mathbb{R}^+$

\(^4\)If no one joins at a queue-length, the queue does not grow any further. Even if the queue were to begin at such
focus our analysis to the joining strategies of the form $\alpha = (\alpha(0), \alpha(1), \alpha(2), \ldots, \alpha(\hat{n} - 1), 0, \ldots, 0)$, where $\hat{n} \leq \bar{N}$ (that is: balking at $\hat{n}$) where $\alpha(n) > 0$ for $0 \leq n < \hat{n}$. We need to determine the equilibrium strategies of all consumers: $\alpha^*$ and $\gamma^*$. We specify those equilibrium conditions next.

The equilibrium conditions: Consider a randomly arriving consumer. Suppose all other consumers are joining according to some strategy $\alpha$. The consumer’s expected utility of joining the queue, denoted by $u(n, \gamma)$, is a function of the queue length upon arrival, $n$ and her belief about the server’s type, $\gamma$.

Definition 1 (Equilibrium). The strategies, $\alpha^*$, and beliefs, $\gamma^*$, form an equilibrium if

(i) The consumers are rational: For each $n \in \{0, 1, \ldots, \bar{N}\}$,

$$\alpha^*(n) \in \arg \max_{\alpha' \in [0, 1]} \alpha' \times u(n, \gamma^*).$$

(ii) The consumer beliefs are consistent: The belief $\gamma^*(n)$ satisfies Bayes’ rule on all queue lengths that are reached with strictly positive probability in the long run under the strategy $\alpha^*$.5

Condition (i) of Definition 1 is referred to as the rationality condition for the consumers. Condition (ii) of Definition 1 is referred to as the consistency condition of the beliefs. Using the conditions, we can now analyze the queue-joining equilibrium.

3 Analysis of Queue-Joining Equilibria

For a given belief, $\gamma$, an arriving consumer’s expected utility after observing $n$ consumers in the system can be written

$$u(n, \gamma) = \int_{t=\bar{t}}^{\bar{t}} \int_{v=\bar{v}}^{\bar{v}} \{v - c(\nu + 1)t\} \gamma(t, v, n)dvdt.$$

With Equation (3), we write condition (i) of Definition 1 for consumers: $\alpha(n)$ for any $n$ is 1 (0) when $u(n, \gamma) > 0$ ($u(n, \gamma) < 0$) and any value in $[0, 1]$ when $u(n, \gamma) = 0$.

Condition (ii) of Definition 1 needs to be imposed on $\gamma(t, v, n)$. To that end, let $\pi(n, \alpha, t)$ be the long-run probability that $n$ consumers are in the system when the consumer strategy is $\alpha$ and the server’s mean service time is $t$. For a given mean service time and consumer joining strategy, the stochastic process that describes the queue length is a birth-and-death process (Ross, 1996), which allows us to characterize the long-run probability distribution.

higher states (queue lengths), note that those states are transient, and occur with zero probability in stationarity.

5The reader may note that when a queue length is not reached with positive probability, the belief and the action at that queue length are irrelevant. Also, if we begin with an empty queue, those states are never reached: The queue never exceeds $N$.  

6
Suppose that consumers follow the strategy profile $\alpha$. Then the stationary probability of a queue of length $n$, $\pi(n, \alpha, t)$, is

$$
\pi(n, \alpha, t) = \pi(0, \alpha, t) \prod_{j=0}^{n-1} (\alpha(j) \Delta t) \quad \text{where} \quad \pi(0, \alpha, t) = \left[ 1 + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} (\alpha(j) \Delta t) \right]^{-1}.
$$  \hfill (4)

With the PASTA property (Wolff, 1982), $\pi(n, \alpha, t)$ is also the probability that a randomly arriving consumer observes $n$ consumers in the system. Condition (ii) of Definition 1 imposes that the posterior density of the service time and value after observing a queue length of $n$ satisfies Bayes’ rule:

$$
\gamma(t, v, n)dvdt = \pi(n, \alpha, t)g(t, v)dvdt
$$  \hfill (5)

We can write the utility, $u(n, \gamma)$, when $\gamma$ of Equation (5) is substituted in Equation (3) as a function of $\alpha$, $U(n, \alpha)$:

$$
U(n, \alpha) = \frac{\int_{\frac{t}{2}}^{t} \{V(t) - c(n+1)t\} \pi(n, \alpha, t)dX(t)}{\int_{\frac{t}{2}}^{t} \pi(n, \alpha, t)dX(t)},
$$  \hfill (6)

where we used the definitions of $V(t)$, the conditional expected service value, and $X(t)$, the marginal distribution of the mean service times. We obtain an alternative condition for the expected utility at queue length $n$ to be positive:

$$
U(n, \alpha) > 0 \iff EA(n, \alpha) = \int_{\frac{t}{2}}^{t} A_n(\tau, \alpha)dX(\tau) > 0,
$$

where

$$
A_n(\tau, \alpha) = \frac{\{\tau - \tau(n)\} (\Lambda \tau)^n}{1 + \sum_{k=1}^{N} (\Lambda \tau)^k \prod_{m=0}^{k-1} \alpha_m}
$$

and $\tau(n) = V_0/(-r + c(n+1))$, $n \in \{0, \ldots, \bar{N}\}$. Note $A_n(\tau, \alpha)$ depends on service time $\tau$, which is a random variable with distribution $X(\tau)$, and the consumer joining strategy $\alpha$. $A_n$ helps us isolate the impact of the mean service time on joining utilities. When the expected value of $A_n(\tau, \alpha)$ is positive, the expected utility from joining the queue is positive. $\tau(n)$ is the mean service time above which the expected net utility of joining a queue of length $n$ becomes positive. With Assumption 1 and the definition of $\bar{N}$, it can be seen that $\tau(n)$ is increasing in queue length $n$ and lies in the interval $[\frac{t}{2}, \bar{t}]$ for $n \in \{0, \ldots, \bar{N} - 1\}$ (i.e., for all joining states). Thus, we have $\frac{t}{2} < \tau(0) < \cdots < \tau(\bar{N} - 1) < \bar{t} < \tau(\bar{N})$. 

7
We obtain an equilibrium condition expressed uniquely in terms of $\alpha^*$:

\[
\alpha^*(n) = 1(0) \quad \text{when} \quad \int_{\bar{t}}^{t} A_n(\tau, \alpha^*) dX(\tau) > (\leq) 0 \quad \text{and} \\
\alpha^*(n) \in [0, 1] \quad \text{when} \quad \int_{\bar{t}}^{t} A_n(\tau, \alpha^*) dX(\tau) = 0. 
\]

(7)

Note that the (equilibrium) conditions (i) and (ii) of Definition 1 are met by any $\alpha^*$ that satisfies Equation (7) and $\gamma^*$ that results from the long-run probability distributions characterized by Equation (4). Since the action space is compact ($[0, 1]^N$) and the best response function has a closed graph, with the Kakutani fixed-point theorem, we can conclude that at least one equilibrium in mixed strategies exists. In the next sections, we analyze the equilibrium conditions in terms of $A_n$.

4 Bivalued Priors on Service Distribution

In this section, we begin our analysis when $X(t)$ is bivalued: $t = \bar{t}$ with probability $p$ and $t = \bar{t}$ with probability $1 - p$. The model primitives are $(\Lambda, V_0, r, c, \bar{t}, \bar{t}, p)$. We can write

\[
EA(n, \alpha) = (1 - p)A_n(\bar{t}, \alpha) + pA_n(\bar{t}, \alpha).
\]

We first identify the sufficient conditions for a threshold joining equilibrium strategy (in section 4.1). We then identify nonthreshold (sputtering) equilibrium strategies in cases when threshold strategies do not exist (in section 4.2). In section 4.3, we discuss the equilibrium throughput. We generalize our analysis to include the case in which $X(t)$ has a nonnegative density over $[\bar{t}, \bar{t}]$ in section 5.

4.1 Pure threshold joining strategies

We first analyze strategies of the form $\alpha^* = (1, 1, \ldots, 1, 0, 0, \ldots, 0)$, where $\alpha^*(n) = 1$ on the $\hat{n}$ positions. (We begin the indices starting from queue length 0 to queue length $\hat{n} - 1$. Thus, the joining probability at queue length $n$ is given by the $n + 1^{th}$ component of $\alpha^*$. Therefore, a threshold joining strategy $\alpha^*$ is characterized by means of a single parameter $\hat{n}$, which is the balking threshold. With a slight abuse of notation, we write $EA(n, \alpha^*)$ as simply $EA(n, \hat{n})$.

Recall that a threshold strategy, according to Equation (7), is an equilibrium characterized by $\hat{n}$, in which the following conditions are satisfied: $EA(n, \hat{n}) \geq 0$ for $0 \leq n \leq \hat{n} - 1$ and $EA(n, \hat{n}) \leq 0$ for $n \geq \hat{n}$. It will be analytically convenient to consider $n$ and $\hat{n}$ as positive reals. We replace the integer $n$ by the real $\nu$ and the integer $\hat{n}$ by the real $\nu$. Then, $EA(\nu, n)$ can be considered as a continuous function in $(\nu, n)$. We will specify the domain of $(\nu, n)$ shortly.
Suppose that $\hat{\nu}(n)$ is the lowest root of $EA(\nu, n) = 0$ for any $n \in [0, (V_0 + rt)/(ct) - 1)$. Then $[\hat{\nu}(n)]^6$ can be interpreted as the “Best Response” of a focal consumer when all other consumers join according to a pure threshold strategy at $n \in \mathbb{N}$. The focal consumer joins according to a pure threshold strategy with his threshold at $[\hat{\nu}(n)]$. A threshold strategy of a focal consumer is an equilibrium strategy if his best response coincides with the strategy of all other consumers. Thus, the equilibrium condition for a strategy with a threshold at $\hat{n}_0$ is

$$\hat{n}_0 = [\hat{\nu}(\hat{n}_0)].$$

Typically, $\hat{\nu}(n)$ reveals further information about the consumer equilibrium. Note that for expository convenience we do not restrict the domain of $\hat{\nu}(n)$ to $[0, n]$. It is thus possible that $[\hat{\nu}(n)] > n + 1$ when $n \in \mathbb{N}$. In this case, if all consumers join with balking threshold $n$, $[\hat{\nu}(n)]$ may not be a recurrent queue length and thus cannot be an equilibrium threshold.

**Example 1:** We provide an illustrative example of an equilibrium with no threshold consumer joining. Let $V_0 = -7.55$, $r = 12.025$, $c = 1$, $p = 0.2$ and $\Lambda = 1$. In Figure 1, we plot the Best Response function for two examples. In the left panel, we have the two mean service times relatively close to each other: $t = 1$, $\bar{t} = 1.25^7$. The dashed lines indicate the locus of $EA(\nu, n) = 0$. The best response function is indicated by dotted line (with “+”). When the Best Response function intersects with the 45 degree line (i.e., when $[\hat{\nu}(\hat{n}_0)] = \hat{n}_0$), we obtain a pure strategy threshold equilibrium.

In the left panel, the best response to any threshold joining strategy, $n$, shown is to balk at a queue length of 4. The Best Response function is thus flat for the queue lengths shown in the figure. Therefore, there is a threshold equilibrium strategy in which all consumers balk at queue length $n = 4$ or higher.

In contrast, observe the right panel where we have a larger gap between the two service times ($t = 0.65$ and $\bar{t} = 1.55)^8$. The Best Response to a threshold joining strategy of $n = 4$ is to balk at a queue length of 6, as 6 is the lowest integer above the lowest root of $EA(\nu, 4) = 0$. (Recall the discussion pursuant to Equation (8).) Similarly, the Best Response to a threshold joining strategy of $n = 5$ is to balk at a queue length of 1. Hence, in this case, there cannot be a pure threshold joining strategy.

We have provided an instance of the absence of the pure threshold strategy equilibrium in Example 1. Using the illustrated example, our next step is to clarify theoretically how threshold equilibrium strategies may fail to exist. This property critically depends on the nature of the

---

6. $[x]$ is the largest integer not greater than $x \in \mathbb{R}^+$

7. As benchmark example for which a pure threshold joining strategy exists, the parameters $(c, r, t, \bar{t})$ do not satisfy Assumption 1.

8. These parameters and the parameters in all subsequent examples do satisfy Assumption 1.
function $EA(\nu, n)$ as discussed in the example.

In order to better understand the roots of $EA(\nu, n)$, we rewrite the condition $EA(\nu, n) > 0$ in terms of two separate functions $\varphi(n)$ and $\Phi(n)$. We illustrate properties of these two functions to help decipher the equilibrium structure:

\[ EA(\nu, n) > 0 \iff \left( \frac{t}{\bar{t}} \right)^{\nu} \frac{\tau(\nu) - t}{\bar{t} - \tau(\nu)} \leq \frac{p}{1 - p} \frac{1 + \sum_{k=1}^{n} (\Lambda t)^k}{1 + \sum_{k=1}^{n} (\bar{\Lambda} t)^k}, \]

(9)

With $\varphi(\nu)$ and $\Phi(n)$ as defined in Equation (9), for a given $n$, $\hat{\nu}(n)$ is the lowest value in $[0, \bar{\nu}]$ such that $\varphi(\nu) \geq p_1 - p \Phi(n)$ and $\lceil \hat{\nu}(n) \rceil = m$ is thus the first integer value, $m$, above $\hat{\nu}(n)$. With a slight abuse of notation, let $\pi(k, n, t)$ be the long-run probability that the queue length is $k$ when all consumers join according to a threshold $n$ and the expected service time is $t$. This results in an $M/M/1/n$ queue with a service rate of $1/t$. We can see that $\Phi(n) \overset{\text{def}}{=} 1 + \sum_{k=1}^{n} (\Lambda t)^k \overset{\text{def}}{=} \pi(0, n, \bar{t}) / \pi(0, n, t)$. $\Phi(n)$ can thus be described as the likelihood ratio of the server being fast to the server being slow when the queue is empty. The interpretation of $\varphi(\nu)$ is a bit more subtle. Note that the likelihood ratio of the queue being $\nu$ is given by $\overline{\pi(\nu, n, \bar{t})} / \overline{\pi(\nu, n, t)} = \left( \frac{t}{\bar{t}} \right)^{\nu} \Phi(n)$ (assuming that between queue lengths 0 and $\nu - 1$, all consumers join). Hence, when $\frac{\pi(\nu)}{\pi(\nu)} = \frac{p}{1 - p} \left( \frac{t}{\bar{t}} \right)^{\nu} \varphi(\nu)$, the consumer arriving at queue length $\nu$ is indifferent between joining and not, if the posterior likelihood ratio at the empty queue were $\varphi(\nu)$. Therefore, $\varphi(\nu)$ can be interpreted as the required likelihood ratio at the empty queue that makes an arriving consumer indifferent between joining and not at queue length $\nu$. Contrary to $\Phi$, $\varphi$ will not be monotonic, as we will show in Proposition 2. The nonmonotonic $\varphi$ will cause
the sputtering equilibrium, the nonmonotonically decreasing joining probability (as a function of the queue length).

For expositional convenience, let $\nu = (V_0 + rt)/(ct) - 1$ and $\bar{\nu} = (V_0 + r\bar{t})/(c\bar{t}) - 1$. Notice from the left-hand side of the Equation (9) that $\varphi(\nu)$ is nonnegative for $\nu \in [\underline{\nu}, \bar{\nu})$, $\varphi(\underline{\nu}) = 0$ and $\lim_{\nu \to 0^-} \varphi(\nu) = +\infty$. Also note that the set $[\underline{\nu}, \bar{\nu})$ is nonempty due to Assumption (1) and $\nu < 0 < \bar{\nu}$.

Now, we revert back to Equation (9). When the consumers expect the service time to be $\bar{t}$ with a higher prior probability (i.e., $p$ is high), the right-hand side of Equation (9) tends to infinity (as $p \to 1^-$). Therefore, $\nu = \bar{\nu}$ satisfies $\varphi(\nu) \geq \frac{p}{1-p}\Phi(n)$. As a result, $\hat{\nu}(n) = \bar{\nu}$ for any $n$, and $[\hat{\nu}(n)] = [\underline{\nu}] = \bar{N}$. It follows that $[\hat{\nu}(\bar{N})] = \bar{N}$ characterizes a pure threshold strategy equilibrium.

On the other hand, when consumers expect the service time to be slow ($t$) with a higher probability, as $p \to 0^+$, the right-hand side in Equation (9) approaches zero. Therefore, $\nu = 0$ satisfies $\varphi(\nu) \geq \frac{p}{1-p}\Phi(n)$. As a result $\hat{\nu}(n) = 0$, for any $n \geq 0$. It follows that $[\hat{\nu}(0)] = 0$ characterizes a pure strategy threshold equilibrium. These extreme cases (with the prior either very high or very low) yield intuitive equilibria in the light of Assumption (1). In general, the properties of $\Phi(n)$ and $\varphi(\nu)$ determine the Best Response to threshold joining strategy $n$. We characterize these properties in Proposition 2.

**Proposition 2.** (i) $\Phi(n)$ is monotonically decreasing in $n$ for all $n \geq 0$.
(ii) $\varphi(\nu)$ is monotonically increasing over $[\underline{\nu}, \bar{\nu})$ when

$$C \triangleq c + \frac{V_0}{4} \left( \frac{1}{\bar{t}} - \frac{1}{t} \right) \ln \left( \frac{t}{\bar{t}} \right) > 0.$$  

First, we provide intuition for $\Phi(n)$ decreasing in $n$ as observed in Proposition 2(i). Recall that $\Phi(n)$ is the likelihood ratio of the server being fast to the server being slow, at the observation of an empty queue, when the balking threshold is $n$. Suppose that balking threshold $n$ is very low. In such a situation, the recurrent state space (queue lengths) is small. At the very extreme, think of a scenario when all consumers balk at the empty queue (i.e., $n = 0$). In this extreme case, it can be observed that the queue is always empty, irrespective of the expected service time, and hence $\Phi(0) = 1$. A higher joining threshold increases the recurrent set of queue lengths. Suppose that $n$ is high. Since the queues deplete faster with a faster server, one is more likely to observe an empty queue with such a server. Therefore, $\Phi(n) < 1$. Therefore, we have illustrated that $\Phi(n)$ decreases in $n$.

Examining Proposition 2(ii), it is interesting to observe $\varphi(\nu)$ is not always monotonic. Depending on the nature of the curve, we end up with different equilibrium structures. Proposition 2(ii) provides a sufficient condition for the monotonicity of $\varphi(\nu)$, under which a unique root, $\hat{\nu}(n)$, of $EA(\nu, n) = 0$ is guaranteed. We illustrate this condition through curve $C$ in Figure 2.
Figure 2: Illustration of $C = 0$ for $V_0 = -7.55$ and $c = 1$ in the $(\bar{t}, \bar{t})$-space.

An illustration of $C$: The condition $C > 0$ is illustrated in Figure 2. The diagonal line is $\bar{t} = \bar{t}$. As $\bar{t}V(\bar{t}) > \bar{t}V(\bar{t})$, notice that $C > 0$ when $\bar{t}$ is sufficiently close to $\bar{t}$ and $c$ is high. In that case, the logarithm in the expression is close to zero. When the two service-time priors are close enough to each other, a threshold strategy equilibrium emerges (as we observed in the left panel of Figure 1 where the gap between $\bar{t}$ and $\bar{t}$ was small). Thus, the threshold equilibrium naturally emerges under little or no uncertainty in the service times, making Naor’s threshold result a special case.

When $\bar{t}$ is much higher than $\bar{t}$, $C$ becomes negative and consumers find that significantly different wait costs (and service values) associated with priors decrease their confidence in their decision. Hence, a threshold strategy equilibrium may not exist, as seen in the right panel of Figure 1. Therefore, we have

1. When $C > 0$, $\varphi(n)$ is monotonically increasing in $n$ and $\Phi(n)$ is monotonically decreasing in $n$ (as seen in Proposition 2(i–ii)). There exists always an $\hat{n}$ at which $\Phi(n)$ first drops below $\varphi(n)$. In the next subsection (in Proposition 3(i)), we will show that a pure a threshold equilibrium exists, either with balking at $\hat{n}_0$ or with mixing at $\hat{n}_0$ and balking at $\hat{n}_0 + 1$. We refer to this as the classical mixed threshold equilibrium.

2. When $C < 0$, $\varphi(n)$ may not be monotonic in $n$. No pure threshold strategy equilibrium or classical mixed threshold equilibrium can be guaranteed. In the next subsection (in Proposition 3(ii)) we will show that there always exists an equilibrium with balking at $\hat{n}_1$ and mixing at $\hat{n}_0 \leq \hat{n}_1 - 1$. When $\hat{n}_0 < \hat{n}_1 - 1$, we refer to this as the sputtering equilibrium.

Following this outline, in the next subsection, we characterize a sputtering strategy with randomization when a threshold strategy does not exist.
4.2 Sputtering Equilibrium Strategies

In this section, we analyze two cases in which there exist no pure threshold strategy equilibria. In the first case, when \( C > 0 \), \( \varphi(n) \) is monotonically increasing in \( n \) and \( \Phi(n) \) is decreasing in \( n \). In Proposition 3, for this case, we construct an equilibrium having randomization with probability \( \hat{\alpha}_0 \) at \( \hat{n}_0 \) and balking at \( \hat{n}_0 + 1 \). This corresponds to a classical joining equilibrium: pure if \( \hat{\alpha}_0 = 0 \) or mixed if \( \hat{\alpha}_0 \in (0, 1) \).

However, in the second case, \( \varphi(n) \) is nonmonotonic. In this case, there may exist an \( \hat{n} \) such that \( \hat{\nu}(\hat{n}) > \hat{n} \) and \( \hat{\nu}(\hat{n} + 1) < \hat{n} + 1 \). Then, the Best Response function has no intersection with the 45-degree line. In Proposition 3(ii), we construct an equilibrium having randomization with probability \( \hat{\alpha}_0 \) at \( \hat{n}_0 \) and balking at some \( \hat{n}_1 \geq \hat{n}_0 + 1 \), with the possibility that \( \hat{n}_1 > \hat{n}_0 + 1 \). We will denote the latter as sputtering equilibrium, because at \( \hat{n}_0 \) the queue sputters before increasing to \( \hat{n}_1 \) due to the mixing probability of \( \hat{\alpha}_0 \) at \( \hat{n}_0 \). In other words, even though consumers may balk (by randomizing) at queue length \( \hat{n}_0 \), they join with probability 1 when the queue length is \( \hat{n}_0 + 1 \) (or higher). For the next proposition, we define

\[
\hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_1) = \frac{\sum_{k=0}^{\hat{n}_0} [\Lambda\hat{t}]^k + \hat{\alpha}_0 \sum_{k=\hat{n}_0+1}^{\hat{n}_1} [\Lambda\hat{t}]^k}{\sum_{k=0}^{\hat{n}_0} [\Lambda\hat{t}]^k + \hat{\alpha}_0 \sum_{k=\hat{n}_0+1}^{\hat{n}_1} [\Lambda\hat{t}]^k},
\]

(10)

Proposition 3. (i) When \( C > 0 \), then, either there exists a pure threshold joining strategy equilibrium, or there exists a classical mixed threshold equilibrium satisfying the following conditions.

\[
\begin{align*}
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_0 + 1) &\geq \varphi(n), \quad \text{for } 0 \leq n < \hat{n}_0, \\
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_0 + 1) &\geq \varphi(n), \quad \text{for } n = \hat{n}_0, \\
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_0 + 1) &< \varphi(n), \quad \text{for } n = \hat{n}_0 + 1
\end{align*}
\]

(ii) When \( C < 0 \), either there exists a pure threshold joining strategy equilibrium, or there exists an equilibrium having randomization with probability \( \hat{\alpha}_0 \) at \( \hat{n}_0 \) and balking at \( \hat{n}_1 \), where \( \hat{n}_1 \geq \hat{n}_0 + 1 \), satisfying the following conditions.

\[
\begin{align*}
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_1) &\geq \varphi(n), \quad \text{for } 0 \leq n < \hat{n}_0 \text{ (if } 0 \leq \hat{n}_0 \text{)}, \\
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_1) &= \varphi(n), \quad \text{for } n = \hat{n}_0, \\
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_1) &\geq \varphi(n), \quad \text{for } \hat{n}_0 < n < \hat{n}_1, \\
\frac{p}{1-p} \hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_1) &< \varphi(n), \quad \text{for } n = \hat{n}_1
\end{align*}
\]

(12)

We label a sputtering equilibrium with randomization probability \( \hat{\alpha}_0 \) at queue length \( \hat{n}_0 \) and balking threshold at \( \hat{n}_1 \) when \( \hat{n}_1 > \hat{n}_0 + 1 \).

Note that \( \hat{\Phi} \) denotes the likelihood ratio at the empty queue when consumers join according to a sputtering strategy characterized by \((\hat{n}_0, \hat{\alpha}_0, \hat{n}_1)\). (We use \(^\prime\) to differentiate \( \hat{\Phi} \) from \( \Phi \), which is
Figure 3: Demonstration of the sputtering equilibrium with randomization at $n = 1$ (consumers join with probability) and a balking threshold at $n = 6$, updated service value and updated waiting cost.

the likelihood ratio at the empty queue under the classical threshold equilibrium. Proposition 3(i) provides conditions for a mixed strategy equilibrium: In this case, $\alpha^* = (1, 1, \ldots, 1, \hat{\alpha}_0, 0, 0, \ldots, 0)$, where $\hat{\alpha}_0$ is on the $\hat{n}_0 + 1$st position.

Proposition 3(ii) indicates that this classical equilibrium strategy is not always an equilibrium. Proposition 3(ii) identifies a nonthreshold equilibrium with mixing (with probability $\hat{\alpha}_0$) at some queue length ($\hat{n}_0$), joining at longer queues, and balking at queue length $\hat{n}_1$. Recall that the first component of $\alpha$ is the joining probability at the empty queue. Therefore, the equilibrium we identified is $\alpha^* = (1, 1, \ldots, 1, \hat{\alpha}_0, 1, \ldots, 1, 0, 0, \ldots, 0)$, where $\hat{\alpha}_0$ is on position $(\hat{n}_0 + 1)$ (corresponding with queue length $\hat{n}_0$) and the last 1 is at position $(\hat{n}_1 + 1)$ (corresponding with queue length $\hat{n}_1$). The equilibrium belief, $\gamma^*$, follows immediately from the long-run probability distributions.

Example 1 (contd.): Continuing with the previous example, we illustrate the sputtering equilibrium. We identify a nonthreshold equilibrium with $\hat{n}_0 = 1$ and $\hat{\alpha}_0 = 0.1695$ and $\hat{n}_1 = 6$. Under this equilibrium, all consumers always join the queue when length is strictly less than 1, they join with probability $0.1695$ (or balk with probability $0.8305$) when the queue length is 1, and they always join when the queue length is between 2 and 5, and finally, they balk from any queue that is 6 or longer. This instance is demonstrated in Figure 3, which depicts two curves as functions of $n$—dotted lines with “⋄”s representing service value, and dotted lines with “+” representing the waiting costs. It is clear that the service value matches exactly with waiting costs at the randomization queue length 1. At queue lengths $\{0, 2, 3, 4, 5\}$, the expected service value is strictly greater than the expected waiting costs. At $n = 6$ and higher, the expected waiting costs are greater than the expected service value. Hence, consumers balk. \qed
4.3 The Impact of an Unknown Service Rate and Value

Having established the conditions for sputtering equilibria in queues with value and (expected) service time uncertainty, we can now explore the effect of unknown service value in consumers’ prior distribution on the firm’s throughput. In the case when the consumers know a firm’s service value and time exactly, the Naor threshold is \( \left\lfloor \frac{V_0 + rt}{ct} \right\rfloor \), which immediately determines the joining rate, \( \alpha_n^i(t) = 1 \Leftrightarrow n < \left\lfloor \frac{V_0 + rt}{ct} \right\rfloor \). We address the throughput in the full-information case as \( TP^i(t) = \frac{1 - \pi(0, \alpha^i(t), t)}{t} \) where \( i \) denotes the service type. We can compare this full information throughput to the throughput in our case, \( TP^\ast(t) = \frac{1 - \pi(0, \alpha^\ast, t)}{t} \), when the expected service time is unknown.

Example 1 (contd.): In Table 1, we demonstrate the impact of the information about the service value, or the lack thereof, on the firm’s throughput.

<table>
<thead>
<tr>
<th>TP(t)</th>
<th>t = 1</th>
<th>t = t</th>
<th>(1 - p)TP^\omega(t) + pTP^\omega(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown t, ( \omega = * )</td>
<td>0.6981</td>
<td>0.5686</td>
<td>0.6722</td>
</tr>
<tr>
<td>Full Information t, ( \omega = i )</td>
<td>0.0000</td>
<td>0.9722</td>
<td>0.7777</td>
</tr>
</tbody>
</table>

Table 1: Throughput when the service time and value are known/unknown for numerical values in Example 1.

Notice that the throughput is zero when consumers know the mean service time and value (when \( t = \bar{t} \)) which follows from Assumption (1). Interestingly, the low-value firm (with faster service rate) cannot attract any consumers if its type is fully known. However, when there is uncertainty about the firm’s type, it free rides using the existence of a high-value firm. The high-value firm may suffer from the “empty restaurant” syndrome: Its throughput is lower when the expected service time is unobserved, compared to the case when its expected service time is observed. This observation is a direct result of the sputtering equilibrium. The randomization at short queue lengths slows down the queue joining at that state. For instance, sputtering occurs at queue length of 1 in Example 1 and fewer consumers join the queue in that state. This lower joining rate consequently inhibits the growth of long queues, leading eventually to reduced throughput. Thus, revealing the expected service times and service values ex ante, affects the throughput.

5 Generalized Sputtering Equilibria: Continuous Prior Beliefs

In section 4, we found that for bivalued discrete probability distributions of priors, an equilibrium with randomization at a single queue length (at most) always exists (as identified in Proposition 3). It is likely that we could construct a sputtering equilibrium for nonmonotonic \( \varphi(n) \). In such cases, it is natural to ask if such a result is sustained for continuous distributions.
As one might expect, a continuous distribution of prior beliefs poses additional technical constraints. For instance, under a continuous distribution of prior beliefs, we cannot readily decompose the condition $EA(\nu, n) > 0$ into two separable terms of $\varphi(n)$ and $\Phi(n)$ as in $\varphi(\nu) < \Phi(n)$ (see Equation (9)). Therefore, characterizing a sputtering equilibrium is more involved as it requires solving a fixed-point problem with a set of nonlinear equations at those queue lengths at which randomization might occur, and also set of inequalities (at the other queue lengths) involving the integrals

$$\int_{\bar{L}}^{i} A_n(\tau, \alpha)dX(\tau), \text{ for } n \in \{0, \cdots, \bar{N}\}. $$

As $A_n(\tau, \alpha)$ is a ratio of polynomials in $\tau$, no closed form of the integral exists for any general density function. Nevertheless, we can characterize some prior distributions for the existence of sputtering equilibria.

We achieve this by using a two-step approach. First, in order to derive results and insights into the structure of such equilibria, we devise a simple iterative algorithm that allows us to numerically compute an equilibrium involving randomization, arbitrarily closely. We do this in section 5.1 and in section 5.2, we discuss the impact of imperfect information on the throughput. Then, using a representative example generated through the algorithm as an illustration, we derive the analytical properties of densities (beliefs) that support a specific sputtering equilibrium structure in section 5.3.

5.1 An Algorithm for Computing an Equilibrium in Continuous Priors

In this section, we generalize our results on sputtering equilibria for bivalued distributions of the service times to the case when consumers have continuous prior distributions. To this effect, we ask the following question: Given any density, $x(\tau)$, what would be the corresponding equilibrium strategy $\alpha^*$? In order to obtain insights in the structure of the equilibrium joining strategy for any such distribution, we resort to the construction of the equilibrium through an iterative algorithm.

It is easy to see that one equilibrium exists under our prescribed conditions. Recall that our objective is more specific: We seek whether an equilibrium in which players randomize at some queue length and join with probability one at a longer queue can exist. Computational literature has shown that, in general, the problem of identifying Nash equilibria with a specific desired structure, belongs to a class of NP-hard problems (see Gilboa and Zemel, 1989). Establishing computational complexity of our specific problem structure is beyond the scope of our work. It is likely that for most instances of our problem, we can find a numerically acceptable equilibrium efficiently.

Therefore, we focus on numerically establishing an equilibrium with arbitrarily close numerical approximations on payoffs. In order to achieve this objective, we have to resort to either analyzing small problems or locating equilibria by introducing some “tolerance” on payoffs—a tolerance limit.
that is numerically dictated by the number of computational iterations of the problem instance.

To begin, we smooth the discontinuous equilibrium conditions of Equation (7) using a function $H(u, \beta) = 1/(1+\exp(-\beta u))$, where $\beta > 0$. $\beta$ is the parameter of choice. It can be seen that $H$ maps the consumer utility, $u$, into a joining probability, $H(u, \beta)$. When $\beta > 0$, large positive (negative) utilities are mapped into high (low) queue-joining probabilities. As $\lim_{\beta \to +\infty} H(u, \beta) = 1$ and $\lim_{\beta \to -\infty} H(u, \beta) = 0$. Further, when the parameter $\beta$ tends to infinity, $H(u, +\infty) = 1$ (0) for strictly positive (negative) $u$. Given any $\epsilon > 0$, there is some $\delta > 0$ such that, any utility in an interval $[-\epsilon, +\epsilon]$ can be mapped into the range $[\delta, 1-\delta]$ by selecting a large value of the parameter $\beta$.

We compute $\alpha^*$ by solving a series of fixed-point problems for different finite values of $\beta$ and computing $\alpha_\beta^*(n) = H(U(n, \alpha_\beta^*), \beta)$ for all queue lengths $n \in \{0, \ldots, \bar{N}\}$. To compute $\alpha_\beta^*$, we can use existing nonlinear optimization routines to solve $\min_{\alpha \in [0,1]} F(\alpha, \beta)$, where $F(\alpha, \beta) = \sum_{n=0}^{\bar{N}} (H(U(n, \alpha), \beta) - \alpha(n))^2$. Let $\alpha_\beta^*$ be its solution.

Then we solve a sequence of optimization problems for $\beta \in \{\beta_1, \beta_2, \ldots, \beta_T\}$ and $\beta_1 < \beta_2 < \cdots < \beta_T$, where $\beta_1$ is low and $\beta_T$ is high. Note that for $\beta = 0$, $H(u, \beta) = 1/2$ for any $\alpha$. We use $\alpha(n) = 1/2$ as the initial point of the lowest value, $\beta_1$. Next, we use $\alpha_{\beta_1}$ as the initial point for the minimization with parameter $\beta_2$. We use the solution $\alpha_{\beta_2}$ as initial point for the minimization problem with $\beta_3$ and so on. Besides rounding errors and errors of numerically computing the integrals in the objective function, $\beta_T$, the value of the objective function ($F(\alpha_{\beta_T}, \beta_T)$) and the value of $\beta_T$ determines how close the solution $\alpha_{\beta_T}$ is to $\alpha^*$. Recall that $\alpha^*$ satisfies $F(\alpha^*, +\infty) = 0$.

Hence, in selecting an upper bound, $\varepsilon_F$, on $F(\alpha_{\beta_T}, \beta_T) < \varepsilon_F$, we ensure that the best response to strategy $\alpha_{\beta_T}$ is arbitrarily close to $\alpha^*$. However, closer approximations also involve significantly increased computational burden. By selecting a high value for $\beta_T$, we can get an $\alpha_{\beta_T}$ that is arbitrarily close to $\alpha^*$. Take for example, $\varepsilon = 5/\beta_T$, we obtain $\delta = H(-5/\beta_T, \beta_T) = 1/(1+e^5) = 0.0066$ and $H(5/\beta_T, \beta_T) = 1/(1+e^{-5}) = 0.9933 \approx 1 - \delta$. Therefore, when terminating with $\beta_T = 2.000$, every strictly negative utility less than $-5/\beta_T = -0.0025$ will be mapped in a joining probability in $[0, 0.0066]$ (which can stipulated to be an arbitrarily small interval close to zero) and every strictly positive utility more than $5/\beta_T = 0.0025$ will be mapped in $[0.9933, 1]$ (arbitrarily close to one). All `close-to-zero’ utilities in $[-0.0025, +0.0025]$ are mapped into the randomization range, $[0.0066, 0.9933]$. Therefore, higher values of $\beta_T$ make $\alpha_{\beta_T}$ closer to $\alpha^*$ (at the expense of computational time).

Using an example generated by the algorithm, we examine the generalization of a sputtering equilibrium for continuous beliefs. More specifically, we will identify an equilibrium in which there is randomization of the joining-balking decision at two different queue lengths. We will use this illustrative example in section 5.3, as we derive results on beliefs that generate such equilibria.
5.2 The Impact of an Unknown Service Rate and Value

We investigate the impact of the absence of information on the true expected service time and value, as done in section 4.3. The throughput for this case, when the consumers are fully informed is

$$TP^i = \int_L^i (1 - \pi (0, \alpha^i(\tau), \tau)) \frac{1}{\tau} dX(\tau),$$

where the joining probabilities are prescribed by the Naor threshold: $\alpha^i_n(t) = 1 \Leftrightarrow n < N^i(t) = \lfloor \frac{V_0 + rt}{r} \rfloor$. We need to compare the throughput under full information with

$$TP^* = \int_L^i (1 - \pi (0, \alpha^*, \tau)) \frac{1}{\tau} dX(\tau).$$
Example 2 (contd.): We revert to Example 2 to compare the throughputs under the cases of perfect and imperfect information. When comparing the throughput with and without information on service rates and values, in Example 2, we have $TP^i = 0.6754$ and $TP^* = 0.8289$. Note how continuous priors influence throughput. The lack of information on the expected service time under continuous priors may increase the total throughput on average.

The low-value service provider thrives again by free riding on the reputation of the high-value service providers. Of course, just as discussed in section 4.3, the high-value service provider does suffer from the empty-restaurant effect. However, in this instance under continuously distributed priors, the slowing down due to the randomization at low queue lengths does not significantly hurt the high-value service provider’s throughput. The reason is that the arrival rate is relatively high compared to the mean service time (recall that $\Lambda = 2$ and the highest value service provider has a service rate of $1/\bar{t} = 0.51282$), hence, not many consumers are lost at the sputtering queue length ($n = 0$). This effect is demonstrated in Figure 5.

Now using the previous example as a specific characterization, we analytically explore what prior beliefs could sustain a sputtering equilibria, or even a generalized sputtering equilibrium, and provide further technical results.

5.3 Continuous Prior Beliefs that Support a Sputtering Equilibrium

Recall in the previous section that we addressed the question: For a given density, $x(\tau)$, what is an equilibrium strategy $\alpha^*$? In the light of the results on sputtering equilibrium in the case of a bivalued distribution on priors, it may be worth verifying if such sputtering equilibria could be sustained on continuous beliefs. Hence, it is pertinent to ask: For a given joining strategy, $\alpha^*$ which densities, $x(\tau)$, support $\alpha^*$?

In particular, we are interested in characterizing the probability distributions over the interval...
at subsets with Lebesgue measure zero) and the maximum value of
Hence, all inequalities will become ‘a.e.’ (i.e. the inequality must hold
for recurrent queue lengths for
We refer to the problem of finding the distribution with minimum maximal density over all con-
cerning, we drop \( \alpha^* \) from all arguments. (In particular, we write \( A_n(\tau, \alpha^*) \) as simply, \( A_n(\tau) \).) With this notation, we can write the equilibrium conditions of Equation (7) as
Clearly, many density functions may satisfy Equation (13) and hence, support \( \alpha^* \). We can now characterize specific density functions that allow the equilibrium constraints to hold. In order to exploit this linear structure, we identify among all density functions that satisfy Equation (13) the
min–max density can be achieved via a
characterize specific density functions that allow the equilibrium constraints to hold. In order to
infinite-dimensional linear program:

\[
x^* = \min_{x(\tau) \in L_{\infty}(\mathcal{L}]} x_0 \text{ s.t. } x_0 \geq 0, \ x_0 - x(\tau) \geq 0, \ \text{a.e. and (13)}
\]

We refer to the problem of finding the distribution with minimum maximal density over all continuous distributions in Equation (14) as the min–max density problem and its solution (i.e., the continuous distribution with the minimum maximum density on its support) as the min–max den-
sity (that supports the equilibrium structure \( \alpha^* \)). Note that this criterion spreads out the mass
over \([t, \bar{t}]\), and therefore, provides a counterpoint to perfect information on expected service time. Specifically, if we were to solve the unconstrained min–max density problem choosing from all probability distributions over the finite support \([t, \bar{t}]\), the solution results in a uniform distribution implying that the consumers are uncertain about the mean service time over \([t, \bar{t}]\).

We can provide further structure to the solution of Problem (14). To begin, we assume that \(\alpha^*\) is an equilibrium for some prior distribution to ensure that Equation (13) has at least one feasible solution. In Lemma 5 (the lemma and its proof are presented in the appendix), we obtain the dual of Problem (14):

\[
\begin{align*}
 w^* &= \max w \\
 &\quad \sum_{n \in \mathcal{N}} (-1)^{\delta^-(n)} w_n A_n(\tau) + w - w_0(\tau) \leq 0, \text{ a.e.,} \\
 &\quad \int_t^\bar{t} w_0(\tau) d\tau \leq 1, \\
 &\quad w_0(\tau) \geq 0, \text{ a.e.,} \\
 &\quad w_n \geq 0, \quad n \in \mathcal{N}^+ \cup \mathcal{N}^-, \\
\end{align*}
\]

where \(\delta^-(n) = 1\) if \(n \in \mathcal{N}^-\) and 0 otherwise. \(w \in \mathbb{R}\), the dual variable associated with the integrality constraint in Equation (13), is unrestricted in sign. \(w_n\) for \(n \in \mathcal{N}\) are the dual variables of the utility constraints at each queue length with a pure joining strategy (join or balk), and hence, need to be nonnegative for \(n \in \mathcal{N}^+ \cup \mathcal{N}^-\) and are unrestricted in sign for \(n \in \mathcal{N}^0\). Finally, the nonnegative function \(w_0(\tau) \in L^2([t, \bar{t}])\) is the dual to the constraints \(x(\tau) \leq x_0\), a.e. (We explain the choice of \(L^2([t, \bar{t}])\) for \(w_0(\tau)\) in the appendix.)

We first show that strong duality holds (i.e., \(x_0^* = w^*\)), by proving Lemma 5. Employing the complementary slackness conditions from Lemma 5, we characterize in Proposition 4 the properties of the min–max density that support the joining strategy \(\alpha^*\).

Before we state the proposition, we introduce first two subsets; \(\mathcal{N}^*\) and \(\mathcal{T}^*\).

- \(\mathcal{N}^*\) is the set of queue lengths (a subset of \(\mathcal{N}\)) for which the net utility from joining the queue is zero, under the prior that follows the min–max distribution. (Hence, by its definition, \(\mathcal{N}^0\) is a subset of \(\mathcal{N}^*\).)
- \(\mathcal{T}^*\) indicates the support of the min–max density, which is a subset of \([t, \bar{t}]\) (as follows from the last constraint in Equation (13)).

The min–max density supporting a strategy \(\alpha^*\) is characterized by the following proposition.

**Proposition 4.** The optimal solution is characterized by a subset of queue lengths, \(\mathcal{N}^*(\supseteq \mathcal{N}^0)\) and a set of expected service times, \(\mathcal{T}^* \subseteq [t, \bar{t}]\) such that

\[
\int_{\mathcal{T}^*} A_n(\tau) d\tau = 0, \quad n \in \mathcal{N}^* \quad \quad (15)
\]
and for some \( w_n^* \neq 0, \) with \( n \in N^* \),

\[
\tau \in T^* \iff \frac{1}{\int_{T^*} d\tau} + \sum_{n \in N^*} (-1)^{\delta_+^{-}(n)} w_n^* A_n(\tau) > 0
\] (16)

and \( w_n^* > 0 \) for \( n \in N^* \setminus N^0 \) and \( \int_{T^*} (-1)^{\delta_+^{-}(n)} A_n(\tau) d\tau \geq 0 \) for \( n \notin N^* \).

Then the min–max density solving Problem (14) is

\[
x^*(\tau) = \begin{cases} 
\frac{1}{\int_{T^*} d\tau} & \text{for } \tau \in T^*, \text{ a.e.,} \\
0 & \text{for } \tau \notin T^*, \text{ a.e.}
\end{cases}
\]

In addition, \( w_0^*(\tau) = \{ (\int_{T^*} d\tau)^{-1} + \sum_{n \in N^*} (-1)^{\delta_+^{-}(n)} w_n^* A_n(\tau) \}^+ \).

Proposition 4 states that the support of the min–max density fully determines the density; it is piecewise uniform. It is intuitive that the min–max density function has the same (maximal) density value over its entire support (wherever it exists), as it is a result of a linear program that minimizes the maximum of a function while keeping the area underneath equal to one. Interestingly, the Proposition also identifies a set of queue lengths, \( N^* \setminus N^0 \), at which the consumer utility is zero under the min–max density, although the consumers may join with probability 1 at these queue lengths. When \( N^* = N^0 \), we address the min–max density as preserving the utility surplus structure (of the equilibrium). That is, the expected utility at all the nonrandomization queue lengths is strictly positive or negative under the min–max density.

We apply Proposition 4 to the case of a sputtering equilibrium, \( N^0 = \{ n_0 \} \), and assume that \( \alpha^* \) is an equilibrium for some bivalued density at \( \bar{t} \) and \( \bar{\ell} \), as characterized in Proposition 3. We can now confirm through Proposition 4 whether a uniform distribution over \( [\bar{t}_1, \bar{t}_2] \subset [\bar{t}, \bar{\ell}] \) can be the solution of Problem (14) for a sputtering equilibrium while preserving the structure of the utility surplus \( N^* = N^0 \). For \( N^* = N^0 \), \( w_{n_0}^* \neq 0 \) and the condition of Equation (15) requires that \( \int_{\bar{t}_1}^{\bar{t}_2} A_{n_0}(\tau) d\tau = 0 \). Hence, the support of any distribution that satisfies Equation (15) must contain the root of \( A_{n_0}(\tau) = 0 \). Therefore, it must be that \( \tau_1^* < \tau(n_0) < \tau_2^* \) as \( A_{n_0}(\tau) \) is negative (positive) below (above) \( \tau(n_0) \); \( A_{n_0}(\bar{t}_1) < 0 < A_{n_0}(\bar{t}_2) \). But, for such \( (\tau_1^*, \tau_2^*) \), the condition from Equation (16) for \( \bar{t} < \tau_1^* < \tau_2^* < \bar{\ell} \) yields

\[
\frac{1}{\tau_2^* - \tau_1^*} + w_{n_0}^* A_{n_0}(\tau) > 0 \iff \tau \in (\tau_1^*, \tau_2^*),
\]

which requires from the continuity of \( A_{n_0}(\tau) \) that \( A_{n_0}(\bar{t}_1) = A_{n_0}(\bar{t}_2) = -\frac{1}{w_{n_0}^* (\tau_2^* - \tau_1^*)} \). However, this condition contradicts \( A_{n_0}(\tau_1^*) < 0 < A_{n_0}(\tau_2^*) \). Notice that the reasoning does not depend on the structure of \( T^* \); \( A_{n_0}(\tau) \) evaluated at any boundary of \( T^* \) of must be equal. As there must exist boundary points of \( T^* \) below and above \( \tau(n_0) \), it is impossible to make \( A_{n_0}(\tau) \) evaluated at these points equal. Hence, Equations (15) and (16) can never be satisfied simultaneously when
Example 1 (contd.): We compute the min–max density probability distribution (via the conditions in Proposition 4) among the family of distributions that support the equilibrium, \( \alpha^* \), observed in Example 1, with sputtering at a single queue length, \( \hat{n}_0 = 1 \) (with randomization \( \hat{\alpha}_0 = 0.1695 \)), balking at \( \hat{n}_1 = 6 \), and joining with probability 1 at all other queue lengths. That is, \( \mathcal{N}^0 = \{1\} \).

We obtained that \( \mathcal{N}^* = \{1, 2\} \). The support of the min–max density is \( T^* = [\bar{t}, \tau_1^*] \cup [\tau_2^*, \bar{t}] = [0.65, 0.69785] \cup [1.54073, 1.55] \), for which \( w_0^* = -376.70234 \) and \( w_2^* = 1735.21163 \). We verify that the condition of Equation (16) is satisfied.

As \( \mathcal{N}^* = \{1, 2\} \), the utility at \( n = 1 \) and \( n = 2 \) is equal to zero, even though randomization only occurs in \( \alpha^* \) at \( n = 1 \). That is, among the distributions that support an equilibrium with sputtering at a single queue length, we find one for which the expected utility at another queue length is equal to zero. Therefore, it is not surprising to find that when we use continuously distributed priors instead of the simple bivalued priors, randomization at multiple queue lengths can exist.

To illustrate this point, we use Proposition 4 to find the min–max density that supports \( \alpha''^* \), close (in the Euclidean space) to \( \alpha^* \), only, at \( n = \hat{n}_2 = 2 \), the joining probability is 0.95 instead of 1 (for \( \alpha^* \)). Thus, \( \alpha''^* = (1, 0.1695, 0.95, 1, 1, 1, 0) \) (compared with \( \alpha^* = (1, 0.1695, 1, 1, 1, 1, 0) \)). Even though \( \alpha''^* \) is close to \( \alpha^* \) in the Euclidean space, notice that it is structurally very different because now, \( \mathcal{N}^{50} = \{1, 2\} \neq \mathcal{N}^0 \). We find that the support of the min–max density supporting \( \alpha''^* \) is \( T''^* = [0.65, 0.69787] \cup [1.54100, 1.55] \), which is very close to \( T^* \) (in \( L_\infty([\bar{t}, \bar{t}]) \)), and \( \mathcal{N}''^* = \mathcal{N}^{50} \).

Therefore, \( \alpha''^* \), which involves randomization at two queue lengths, is an equilibrium under a continuous (piecewise uniform) density with support \( T''^* \).

The example above emphasizes our main observation for continuous prior distributions: When expanding the distribution space from discrete (bivalued) to continuous distributions, equilibria with randomization at more than two queue lengths can be supported. For a given sputtering equilibrium (with randomization at a single queue length), the min–max density (with support \( T^* \)) may identify another queue length, \( \hat{n}_2 \), at which the utility (under the min–max prior) is equal to zero, while the joining probability is 1. There exists a density (with support \( T''^* \) close to \( T^* \) in \( L_\infty([\bar{t}, \bar{t}]) \)) that supports an equilibrium with randomization \( \hat{\alpha}_0 \) at \( \hat{n}_0 \) and randomization \( 1 - \varepsilon' \) at \( \hat{n}_2 \), both \( \hat{n}_0 \) and \( \hat{n}_2 \) may be strictly less than the balking threshold, \( \hat{n}_1 \). We conclude that randomization at queue lengths that are strictly less than the balking threshold are not an artifice of the bivalued prior density. We obtain crisp analytical insights and intuition with such bi-valued priors. Nevertheless, with continuously distributed priors, randomization still occurs, even at multiple queue lengths. Now consider the equilibrium of Example 2.

\footnote{We report only 5 decimal digits, the internal precision was set to 20 more digits.}
Example 2 (contd.): Recall that with the Beta-distributed prior (with symmetric parameters 1.5). The equilibrium joining strategy had randomization at two queue lengths; \( N^0 = \{0, 5\} \) and balking at \( N^- = \{6\} \). At all other queue lengths, the equilibrium is to join with probability 1. We conjecture that \( N^* = \{0, 5\} \) and we compute \( T^* \) of the form \([\tau_1^*, \tau_2^*]\). We obtain \( \tau_1^* \) and \( \tau_2^* \) with Equation (15):

\[
\int_{\tau_1^*}^{\tau_2^*} A_0(\tau) d\tau = \int_{\tau_1^*}^{\tau_2^*} A_5(\tau) d\tau = 0.
\]

We solve for \( \tau_1^* = 0.4414 \) and \( \tau_2^* = 1.8451 \). Now, with Equation (16), evaluated at \( \tau_1^* \) and \( \tau_2^* \), we obtain \( w_0^* = 0.5260 \) and \( w_5^* = -0.4966 \). We verify that Equation (16)—\( \frac{1}{\tau_2^* - \tau_1^*} > w_0^* A_0(\tau) + w_5^* A_5(\tau) \)—also holds for all \( \tau \in (\tau_1^*, \tau_2^*) \). That is, a uniform distribution over \([\tau_1^*, \tau_2^*]\) also supports an equilibrium with randomization at two queue lengths \((n = 0 \text{ and } n = 5)\). Hence, \( N^* = \{1, 2\} = N^{00} \) and \( T^* = [\tau_1^*, \tau_2^*] \). In this case, the min-max density preserves the utility surplus structure.

Hence through these two illustrative examples, we have shown that continuous prior distributions can give rise to randomization—in fact, there could be sputtering randomization at multiple queue lengths. Thus, in conclusion, our results on sputtering equilibrium are robust and fundamentally expand the threshold equilibria considered in Naor (1969) for cases with informational uncertainties.

6 Conclusions

Beginning with the seminal paper by Naor (1969), queueing literature has generally focused on threshold queue-joining policies. This is perfectly understandable since threshold policies are intuitive; fewer consumers join longer queues. In this paper, we show that when both the expected service value and expected service time are unknown to the consumers, but positively correlated,
the equilibrium structure cannot be fully characterized by means of threshold policies anymore. Due to the positive correlation between service value and time, the updated value from a service in fact increases in the length of the queue. Long queues also imply long waiting times. Hence, both waiting costs and service value increase in queue lengths. A simple threshold queue-joining strategy thus may not always be an equilibrium. In this paper, we provide an equilibrium structure for such nonthreshold strategies: namely, sputtering equilibria. When the service values and time are perfectly known, or when the difference between the service times of high- and low-quality service firms is small, we recover the classical threshold strategy.

When the difference in expected service times is large enough, a sputtering equilibrium involving randomization at one or more queue lengths below a balking threshold may emerge. We note that such randomization may decrease or increase the throughput of a service firm compared to the situation in which service time and value are fully known.

We show that randomization at a single queue length arises when the prior about the service time and value is “peaked”: Either both are high, or both are low. When the prior about the expected service time and value are more uniformly distributed over some interval, randomization at multiple queue lengths may occur. Hence, irrespective of the nature of the prior distribution, pure threshold queue-joining strategies are not always an equilibrium.

Thus, we find that a focus on simple threshold policies in queuing games with uncertainty about the service time and value is restrictive. Perhaps, the overwhelming reason to employ such threshold structure is its theoretical simplicity: the joining rate is monotonically decreasing in the queue lengths. However, in practice, it appears that equilibrium queue-joining behaviors may not be monotonic (see Olivares et al. 2012). In our paper, we have provided a “classical” explanation for such nonthreshold joining equilibria, by specifying consumers as perfectly rational Bayesian agents (endowed with the ability to solve complicated equilibria using queueing theory).

There is increasing proof in the behavioral and experimental literature, that in real life, consumers may not always be rational, know queuing theory or apply Bayes’ rule correctly. Hence, our model with rational Bayesian agents is a (compelling) benchmark against which behavioral deviations of “real” human queue-joining behavior must be compared, just as Naor’s model has provided the current literature with a benchmark.

We hope that this paper will provide some scientific foundation for the recent empirical evidence suggesting that nonmonotonic queue joining may arise in practice, and spur more operations research theorists and behavioral researchers studying queue-joining behavior, analytically, empirically or in the laboratory. Our model provides predictions for laboratory queue-joining experiments or empirical tests when the expected service time and value are not fully known to the consumer base.

More analytical work may be required to incorporate behavioral deviations from rational queue-joining predictions. For example, Huang et al. (2011) analyze a model with random queue-joining
choice errors, assuming that all parameter values are known to consumers. Modeling such behavioral deviations observed in laboratory environments when all parameters are not known perfectly is also an interesting line of further research. Ultimately, we hope that understanding consumer queue-joining or queue-avoidance behavior will allow service firms to manage revenues and profits better, while meeting consumer expectations.

Proofs

Proof of Proposition 2: To save on notation, we let $\rho = \Lambda t$ and $\bar{\rho} = \bar{\Lambda} t$, where $\rho < \bar{\rho}$. Now, we consider $n$ as a continuous variable.

(i) Proof that $\Phi(n)$ is monotonically decreasing in $n$. First, we write $\Phi(n)$ as

$$\Phi(n) = \sum_{n=0}^{\infty} \frac{\rho^k}{\sum_{k=0}^{\infty} \bar{\rho}^k} = \left( \frac{1 - \rho^{n+1}}{1 - \rho} \right) \left( \frac{1 - \bar{\rho}^{n+1}}{1 - \bar{\rho}} \right).$$

and determine $d\Phi(n)/dn$:

$$d\Phi(n) = \frac{d}{dn} \left( \frac{1 - \rho^{n+1}}{1 - \rho} \right) - \frac{d}{dn} \left( \frac{1 - \bar{\rho}^{n+1}}{1 - \bar{\rho}} \right).$$

As $\frac{d(a^{n+1})}{dn} = a^{n+1} \ln a$ and $\frac{1 - \rho^{n+1}}{1 - \rho} > 0$ (the sum of $\rho^k$ is positive as $\rho > 0$), we factor out $\frac{1 - \rho^{n+1}}{1 - \rho}$, or $\Phi(n)$, to obtain $d\Phi(n)/dn$:

$$
\frac{d\Phi(n)}{dn} = \Phi(n) \left\{ \frac{-\rho^{n+1} \ln \rho}{1 - \rho^{n+1}} - \frac{-\bar{\rho}^{n+1} \ln \bar{\rho}}{1 - \bar{\rho}} \right\}.
$$

Hence, we can write the condition that $\Phi(n)$ is decreasing as

$$\frac{-\rho^{n+1} \ln \rho}{1 - \rho^{n+1}} < \frac{-\bar{\rho}^{n+1} \ln \bar{\rho}}{1 - \bar{\rho}^{n+1}}.$$

Now, we argue that $\frac{-\rho^{n+1} \ln \rho}{1 - \rho^{n+1}}$ is strictly increasing in $\rho$. Hence, the above inequality holds. Let $z = \rho^{n+1}$, and notice that

$$\rho^{n+1} \frac{-\ln \rho}{1 - \rho^{n+1}} = (n + 1) \frac{-z \ln(z)}{1 - z}$$

and, as $\frac{-z \ln(z)}{1 - z} > 0$ and is continuously increasing for all $z > 0$ and $z = \rho^{n+1}$ is monotonic in $\rho$, $\frac{-\rho^{n+1} \ln \rho}{1 - \rho^{n+1}}$ is strictly increasing in $\rho$.

(ii) Proof that $\varphi(\nu)$ is monotonic in $\nu$ when $C > 0$: We consider $\nu$ as a continuous variable and derive $\varphi(\nu)$ over an extended domain, $(\nu, \bar{\nu})$, with $\nu = (V_0 + rt)/(ct) - 1$ and $\bar{\nu} = (V_0 + r\bar{t})/(c\bar{t}) - 1$
(and, via Assumption 1, \( \nu < 0 < \bar{\nu} \)). Then, we can write \( \varphi (\nu) \) as

\[
\varphi (\nu) = \left( \frac{t}{\bar{t}} \right) ^{\nu} \frac{V_0 - \bar{t}(-r + c(\nu + 1))}{\bar{t}(r + c(\nu + 1)) - V_0}
\]

\[
= \left( \frac{t}{\bar{t}} \right) ^{\nu} \frac{V_0 + tr}{t - \frac{V_0 + tr}{ct}} - \nu - 1
\]

\[
= \left( \frac{t}{\bar{t}} \right) ^{\nu+1} \frac{\nu - \bar{\nu}}{\bar{\nu} - \nu}
\]

Notice that \( \varphi (\nu) = 0 \) and \( \lim_{\nu \to \bar{\nu}+} \varphi (\nu) = +\infty \) and \( \varphi (\nu) > 0 \) for \( \nu \in (\nu, \bar{\nu}) \). Next, we prove that over \((\nu, \bar{\nu})\), the derivative can have zero, one or two roots. We ignore the knife-edge case of one root. Now, we take the derivative:

\[
\frac{d \varphi (\nu)}{d\nu} = \left( \frac{t}{\bar{t}} \right) ^{\nu+1} \ln \left( \frac{t}{\bar{t}} \right) \frac{\nu - \bar{\nu}}{\bar{\nu} - \nu} + \left( \frac{t}{\bar{t}} \right) ^{\nu+1} \frac{(\bar{\nu} - \nu) + (\nu - \bar{\nu})}{(\bar{\nu} - \nu)^2}
\]

It is easy to see that \( \frac{d \varphi (\nu)}{d\nu} = \left( \frac{t}{\bar{t}} \right) ^{\nu+1} \frac{(\bar{\nu} - \nu)}{(\bar{\nu} - \nu)^2} > 0 \), i.e. \( \varphi (\nu) \) is increasing in the neighborhood of \( \nu \).

When

\[
\bar{\nu} - \nu > - \ln \left( \frac{t}{\bar{t}} \right) (\nu - \bar{\nu})(\bar{\nu} - \nu), \tag{17}
\]

it follows that \( \frac{d \varphi (\nu)}{d\nu} > 0 \) for all \( \nu \in (\nu, \bar{\nu}) \). In that case, \( \varphi (\nu) \) is monotonically increasing over \( \nu \in (\nu, \bar{\nu}) \). The right-hand side of Equation (17) is a quadratic equation in \( \nu \) with a maximum at \( \nu^* = \frac{\bar{\nu} + \nu}{2} \). Hence, \( \frac{d \varphi (\nu)}{d\nu} = 0 \) has zero or two (real) roots, depending on whether the maximum of the right-hand side exceeds the left-hand side or not. Substituting \( \nu^* \) in Equation (17), we obtain

\[
\bar{\nu} - \nu > - \ln \left( \frac{t}{\bar{t}} \right) \frac{(\nu - \bar{\nu})^2}{2} = \frac{d \varphi (\nu)}{d\nu} > 0, \forall \nu \in (\nu, \bar{\nu})
\]

or, after dividing by \( \bar{\nu} - \nu \) (which is strictly positive) and multiplying by \( c (> 0) \):

\[
C = c + c \frac{\nu - \bar{\nu}}{4} \ln \left( \frac{t}{\bar{t}} \right)
\]

\[
= c + \frac{1}{4} \left( \frac{V_0 + r t}{\bar{t}} - \frac{V_0 + r t}{t} \right) \ln \left( \frac{t}{\bar{t}} \right)
\]

\[
= c + \frac{V_0}{4} \left( \frac{1}{\bar{t}} - \frac{1}{t} \right) \ln \left( \frac{t}{\bar{t}} \right)
\]

27
We have thus proven that when \( C > 0 \), \( \varphi(\nu) \) is increasing for all \( \nu \in (\nu, \bar{\nu}) \).

[Note: When \( C < 0 \), there exist two roots of Equation (17), which, as \( \varphi(\nu) \) is increasing in the neighborhood of \( \nu \), correspond with a local maximum (lowest root) and minimum (highest root) respectively. We will use this property in the proof of Proposition 3.]

**Proof of Proposition 3:** Recall that we have a condition for \( U(n, \alpha) < 0 \): The sign of the expected utility of a randomly arriving consumer at queue length \( n \), when all other consumers join according to \( \alpha \) is \( EA(n, \alpha) < 0 \). At queue length \( \nu \), when \( \alpha \) is a pure joining strategy characterized by balking at queue length \( n \), we can write

\[
EA(\nu, n) < 0 \iff l_0\Phi(n) < \varphi(\nu).
\]

Recall from Proposition 2(i) that \( l_0\Phi(n) \) is monotonically decreasing in \( n \). Also, by definition of \( \nu \) and \( \bar{\nu} \), have \( \varphi(\nu) = 0 < \varphi(\bar{\nu}) = +\infty \). Recall also from its definition that \( \lceil \bar{\nu}(n) \rceil = m \) is the lowest (integer) queue length such that \( \varphi(m) \geq \Phi(n) \). Generalizing the joining strategy \( \alpha \) to one with randomization with probability \( \hat{\alpha}_0 \) at \( \hat{n}_0 \) and balking at \( \hat{n}_1 \), it is easy to see that the condition to join becomes \( l_0\hat{\Phi}(\hat{n}_0, \hat{\alpha}_0, \hat{n}_1) < \varphi(n) \) at queue length \( n \). We will use these properties below.

Consider the sequence \( \{\varphi(n) : n = [\nu], \ldots, [\bar{\nu}]\} \). For notational convenience, we set \( \varphi([\nu]) = 0 \) and \( \varphi([\bar{\nu}]) = +\infty \). We consider two cases: If \( \varphi(n) \) is monotonically increasing in \( n \), we revert to case (i). If not, we revert to case (ii).

**Case (i):** If \( \varphi(n) \) is monotonically increasing in \( n \) (when \( C > 0 \)), this means that there must exist an \( \nu' \) such that \( \varphi(\nu') = l_0\Phi(\nu') \) (\( l_0\Phi(n) \) is decreasing, \( \varphi(\nu) \) is increasing). Let \( \hat{n} = [\nu'] \), the lowest queue length above \( \nu' \) for which \( l_0\Phi(n) \) drops below \( \varphi(\nu') \): \( l_0\Phi(n) < \varphi(\hat{n}) \) and \( \varphi(\hat{n} - 1) < l_0\Phi(\hat{n} - 1) \).

Depending on \( \varphi(\hat{n} - 1) \) and \( l_0\Phi(\hat{n}) \), we consider two subcases:

**Case (i-a):** If \( \varphi(\hat{n} - 1) < l_0\Phi(\hat{n}) \), then \( \hat{n} \) characterizes a pure balking equilibrium strategy (as \( \hat{n} \) is the lowest queue length for which \( \varphi(\hat{n} - 1) < l_0\Phi(\hat{n}) < \varphi(\hat{n}) \), or \( \lceil \bar{\nu}(n) \rceil = \hat{n} \).

**Case (i-b):** If \( l_0\Phi(\hat{n}) \) \( < \varphi(\hat{n} - 1) \), then, we can always construct an equilibrium with randomization at \( \hat{n} - 1 \) and balking at \( \hat{n} \) as follows. Recall that from the definition of \( \hat{n} \), \( \varphi(\hat{n} - 1) < l_0\Phi(\hat{n} - 1) \).

If now we also have \( l_0\Phi(\hat{n}) < \varphi(\hat{n} - 1) \), combining these two inequalities, it follows that

\[
l_0\Phi(\hat{n}) < \varphi(\hat{n} - 1) < l_0\Phi(\hat{n} - 1).
\]

Evaluating \( \hat{\Phi}(\hat{n}_0, \alpha, \hat{n}_1) \) at \( \hat{n}_0 = \hat{n} - 1 \) and \( \hat{n}_1 = \hat{n} \), we obtain

\[
\hat{\Phi}(\hat{n} - 1, \alpha, \hat{n}) = \frac{\sum_{k=0}^{\hat{n} - 1} |\Lambda|^k \alpha |\Lambda|^{\hat{n}}}{\sum_{k=0}^{\hat{n} - 1} |\Lambda|^k + |\Lambda|^k},
\]

Hence, from the definition of \( \Phi(\cdot) \), we recognize: \( \hat{\Phi}(\hat{n} - 1, 0, \hat{n}) = \Phi(\hat{n} - 1) \) and \( \hat{\Phi}(\hat{n} - 1, 1, \hat{n}) = \Phi(\hat{n}) \).
so, substituting these in Equation (18), we obtain

\[ l_0 \hat{\Phi}(\hat{n} - 1, 1, \hat{n}) < \varphi(\hat{n} - 1) < l_0 \hat{\Phi}(\hat{n} - 1, 0, \hat{n}). \]

Hence, by continuity of \( \hat{\Phi}(\hat{n} - 1, \alpha, \hat{n}) \) in \( \alpha \), there exists some \( \hat{\alpha}_0 \in (0, 1) \) such that \( l_0 \hat{\Phi}(\hat{n} - 1, \hat{\alpha}_0, \hat{n}) = \varphi(\hat{n} - 1) \). As a result, we have \( l_0 \hat{\Phi}(\hat{n} - 1, \hat{\alpha}_0, \hat{n}) = \varphi(\hat{n} - 1) \), or consumers mix between joining and balking at \( \hat{n} - 1 \). As \( \varphi(n) \) is increasing, \( l_0 \hat{\Phi}(\hat{n} - 1, \hat{\alpha}_0, \hat{n}) = \varphi(\hat{n} - 1) < \varphi(\hat{n}) \), or consumers balk at \( \hat{n} \). Similarly, as \( \varphi(n) < \varphi(\hat{n} - 1) = l_0 \hat{\Phi}(\hat{n} - 1, \hat{\alpha}_0, \hat{n}) \) for \( 0 \leq n < \hat{n} - 1 \), or consumers join at \( 0 \leq n < \hat{n} - 1 \). This establishes the “classical” mixing threshold equilibrium at \( \hat{n}_0 = \hat{n} - 1 \) (with probability \( \hat{\alpha}_0 \)) and balking at \( \hat{n}_1 = \hat{n} = \hat{n}_0 + 1 \).

**Case (ii):** If \( \varphi(n) \) is not monotonically increasing in \( n \) (when \( C < 0 \)), let \( \hat{n}_0 \) be the local maximum of \( \varphi(n) \) over \( n = [\hat{\nu}, \ldots, [\hat{\nu}] \) (the lowest root, identified in the note in the proof of Proposition 2(ii)) and define \( \hat{n}_1 \) as the lowest queue length with a value that is strictly more than \( \varphi(\hat{n}_0) \):

\[ \hat{n}_1 = \min\{n \in \mathbb{N}, n > \hat{n}_0 \text{ and } \varphi(n) > \varphi(\hat{n}_0)\}. \]

Note that, by construction, \( \varphi(\hat{n}_1) > \varphi(\hat{n}_0) \). Consider the following replacement for the \( \varphi(n) \) sequence. We replace \( \{\varphi(n) : n = [\hat{\nu}, \ldots, [\hat{\nu}]\) with \( \{\varphi'(n) : n = [\hat{\nu}, \ldots, \hat{n}_0 : \varphi'(n) = \varphi(n), n = \hat{n}_0 + 1, \ldots, \hat{n}_1 - 1; \varphi'(n) = \varphi(n_0), n = \hat{n}_1, \ldots, [\hat{\nu}] : \varphi'(n) = \varphi(n)\)\}. In other words, we fill the “trough” in \( \varphi(n) \) with \( \varphi(n_0) \). \( \varphi'(n) \) is nondecreasing in \( n \). We illustrate filling the trough with a numerical example.

**Illustration of \( \varphi'(n) \) via Example 1:** For the example, \( \hat{n}_0 = 1 \) is the local maximum and \( \hat{n}_1 = 6 \) is the lowest queue length with \( \varphi(n) > \varphi(1) \). In Figure 7, the diamonds indicate \( \varphi(n) \), the circles indicate \( \varphi'(n) \). The crosses indicate \( \Phi(n) \), which is monotonically decreasing. \( \varphi'(n) \) is nondecreasing.

The sequence \( \varphi'(n) \) is nondecreasing. We use \( \varphi'(n) \) instead of \( \varphi(n) \) (as we did for case (i)). Recall that \( \Phi(n) \) is decreasing in \( n \). Similarly as in case (i), define \( \hat{n} \) be the lowest \( n \) such that \( \varphi'(n) > l_0 \Phi(n) \). If \( \hat{n} \leq \hat{n}_0 \), then, \( \hat{n} \) lies on the increasing part of \( \varphi(n) \) and we recover case (i-a). Similarly, if \( \hat{n} > \hat{n}_1 \), then we recover again case (i-a). Suppose now that \( \hat{n}_0 + 1 \leq \hat{n} \leq \hat{n}_1 \). That is, \( \hat{n} \) lies in the trough.

**Illustration of \( \hat{n} \) via Example 1:** In Figure 7, illustrating Example 1, the crosses indicate \( \Phi(n) \), which is monotonically decreasing. As \( \varphi'(n) \) is nondecreasing, observe that the lowest queue length at which \( \Phi(n) \) drops below \( \varphi'(n) \) is \( \hat{n} = 4 \) and \( \hat{n}_0 + 1 = 2 \leq \hat{n} \leq \hat{n}_1 = 6 \).

We have, when \( \hat{n}_0 + 1 \leq \hat{n} \leq \hat{n}_1 \),

\[ l_0 \Phi(\hat{n}) < \varphi'(\hat{n}), \text{ and } l_0 \Phi(\hat{n} - 1) > \varphi'(\hat{n}). \] (19)
Figure 7: Illustration of \( \varphi(n) \) and \( \varphi'(n) \) for Example 1. In this example: \( \hat{n}_0 = 1, \hat{n}_1 = 6 \). The lowest queue length at which \( l_0\Phi(n) \) drops below \( \varphi'(n) \) is \( \hat{n} = 4 \).

We have now

\[
l_0\Phi(\hat{n}_0) \geq l_0\Phi(\hat{n} - 1) \quad (\Phi(\cdot) \text{ is decreasing and } \hat{n}_0 \leq \hat{n} - 1)
\]

\[
> \varphi'(\hat{n}) \quad (\text{see Equation (19)})
\]

\[
= \varphi(\hat{n}_0) \quad (\text{by construction of } \varphi').
\]

Hence, we obtained that \( l_0\Phi(\hat{n}_0) > \varphi(\hat{n}_0) \).  \hfill (20)

We have also

\[
l_0\Phi(\hat{n}_1) \leq l_0\Phi(\hat{n}) \quad (\Phi(\cdot) \text{ is decreasing and } \hat{n} \leq \hat{n}_1)
\]

\[
< \varphi'(\hat{n}) \quad (\text{see Equation (19)})
\]

\[
= \varphi(\hat{n}_0) \quad (\text{by construction of } \varphi').
\]

Hence, we obtained that \( l_0\Phi(\hat{n}_1) < \varphi(\hat{n}_0) \).  \hfill (21)

Consider

\[
\hat{\Phi}(\hat{n}_0, \alpha, \hat{n}_1) = \frac{\sum_{k=0}^{\hat{n}_0} [\Lambda\hat{n}]^k + \alpha \sum_{k=\hat{n}_0+1}^{\hat{n}_1} [\Lambda\hat{n}]^k}{\sum_{k=0}^{\hat{n}_0} [\Lambda\hat{n}]^k + \alpha \sum_{k=\hat{n}_0+1}^{\hat{n}_1} [\Lambda\hat{n}]^k}
\]

as a continuous function of \( \alpha \). Note, that \( \hat{\Phi}(\hat{n}_0, 1, \hat{n}_1) = \Phi(\hat{n}_1) \) and \( \hat{\Phi}(\hat{n}_0, 0, \hat{n}_1) = \Phi(\hat{n}_0) \). Hence,
we have obtained via Equations (20) and (21) that

\[ l_0 \hat{\Phi}(\hat{n}_0, 1, \hat{n}_1) < \varphi(\hat{n}_0) < l_0 \hat{\Phi}(\hat{n}_0, 0, \hat{n}_1). \]

By continuity of \( l_0 \hat{\Phi}(\hat{n}_0, \alpha, \hat{n}_1) \) in \( \alpha \), there exists an \( \hat{\alpha} \in (0, 1) \) such that \( l_0 \hat{\Phi}(\hat{n}_0, \hat{\alpha}, \hat{n}_1) = \varphi(\hat{n}_0) \). Hence, the consumer arriving at queue length \( \hat{n}_0 \) is indifferent between joining and not. For \( \hat{n}_0 < n < \hat{n}_1 - 1 \), we have \( l_0 \hat{\Phi}(\hat{n}_0, \hat{\alpha}, \hat{n}_1) = \varphi(\hat{n}_0) = \varphi'(n) > \varphi(n) \) (by definition of \( \varphi' \)). Therefore, an arriving consumer always joins. Similarly, for \( 0 \leq n < \hat{n}_0 - 1 \), we have \( \varphi(n) < \varphi(\hat{n}_0) = l_0 \hat{\Phi}(\hat{n}_0, \hat{\alpha}, \hat{n}_1) \) (as \( \varphi(n) \) is increasing for \( n \leq \hat{n}_0 \)). Hence, the consumer always joins. Finally, for \( n = \hat{n}_1 \), we have \( l_0 \hat{\Phi}(\hat{n}_0, \hat{\alpha}, \hat{n}_1) = \varphi(\hat{n}_0) < \varphi(\hat{n}_1) \) (by construction of \( \hat{n}_1 \)). Therefore, an arriving consumer always balks. Thus, we have for all \( 0 \leq n < \hat{n}_0 \), an arriving consumer joins the queue; at \( n = \hat{n}_0 \) the arriving consumer is indifferent between joining and balking; for \( \hat{n}_0 < n \leq \hat{n}_1 \), an arriving consumer joins the queue, and at \( n = \hat{n}_1 \) the arriving consumer balks. Hence, we have a sputtering equilibrium as described in case (ii).

To present the proof of Proposition 4, we first introduce Lemma 5. Let \( N = \{1, \ldots, \hat{n} \} \), where \( \hat{n} \) is finite. We assume that there exist a density function, \( x^o \), that satisfies the constraints\(^\text{10}\) below and that density has a finite essential supremum (i.e. \( x^o \in L_\infty([\underline{t}, \bar{t}]) \) and \( x_0 = \esssup(x^o) \)). \( \int_{\underline{t}}^{\bar{t}} f(\tau) \, d\tau \) is the Lebesgue integral where \( f(\tau) \) is a Lebesgue integrable function over \([\underline{t}, \bar{t}]\). The primal problem is:

\[
(P) : \min x_0 \quad \begin{cases}
  x_0 - x(t) \geq 0, \text{ a.e.,} \\
  \int_{\underline{t}}^{\bar{t}} A_n(\tau)x(\tau) \, d\tau \geq 0, \quad n \in N^+, \\
  \int_{\underline{t}}^{\bar{t}} A_n(\tau)x(\tau) \, d\tau \leq 0, \quad n \in N^-, \\
  \int_{\underline{t}}^{\bar{t}} A_n(\tau)x(\tau) \, d\tau = 0, \quad n \in N^0, \\
  \int_{\underline{t}}^{\bar{t}} x(\tau) \, d\tau = 1,
\end{cases}
\]

where \( x_0 \in \mathbb{R} \), \( x \in L_\infty([\underline{t}, \bar{t}]) \) and a.e. means for all \( t \in [\underline{t}, \bar{t}] \), except for a subset of \([\underline{t}, \bar{t}]\) with Lebesgue measure zero. Now, we state:

\(^{10}\)Proving existence of a feasible point is beyond the scope of this paper. In our numerical experiments, we the feasible set is non-empty.
Lemma 5. (i) The dual of $(\mathcal{P})$, is:

$$(\mathcal{D}) : w^* = \max_{x_{\hat{n}+1}} w_{\hat{n}+1}$$

$$\sum_{n=1}^{\hat{n}} (-1)^{\delta^-(n)} w_n A_n(\tau) + w_{\hat{n}+1} \leq w_0(\tau), \ a.e.,$$

$$\int_\tau^I w_0(t) dt \leq 1,$$

$$w_n \geq 0, \ n \in \mathcal{N}^- \cup \mathcal{N}^+$$

$$w_0(\tau) \geq 0, \ a.e.$$

where $\delta^-(n) = 1$ if $n \in \mathcal{N}^-$ and $\delta^-(n) = 0$ otherwise, $w_n \in \mathbb{R}$ for $n \in \mathcal{N}$ and $w_0(\tau) \in L_2([\bar{t}, \bar{t}])$.

(ii) $x^*_n = w^*$ (strong duality).

(iii) The solutions, $(x^*_0, x^*)$ and $(w^*_0, w^*_1, ..., w^*_{\hat{n}+1})$ of Problems (22) and (23) are primal and dual feasible respectively and satisfy:

$$\begin{cases}
    x^* (\tau) (-\sum_{n=1}^{\hat{n}} (-1)^{\delta^-(n)} w_n^* A_n(\tau) - w^*_{\hat{n}+1} + w^*_0(\tau)) = 0, \ a.e., \\
    w^*_0(\tau) (x^*_0 - x^*(\tau)) = 0, \ a.e., \\
    x^*_n (1 - \int_\tau^I w^*_n(\tau) d\tau) = 0, \\
    w_n^* \int_\tau^I A_n(\tau) x^*(\tau) d\tau = 0, \ n \in \mathcal{N}^- \cup \mathcal{N}^+
\end{cases}$$

(complementary slackness).

Proof of Lemma 5. (i) The dual of Problem (22): We first introduce slack variables. Without loss of generality, we assume the following structure: $\mathcal{N}^+ = \{1, ..., m - 1\}$, $\mathcal{N}^- = \{m\}$ and $\mathcal{N}^0 = \{m + 1, ..., \hat{n}\}$. Then, we introduce $x^*(\tau)$ and $x^*_n$ for $n \in \{1, ..., m\}$, in order to obtain equality constraints:

$$(\mathcal{P}) : \min_{x_0}$$

$$-x(t) - x^*(t) + x_0 = 0, \ a.e.,$$

$$\int_\tau^I A_n(\tau) x(\tau) d\tau - x^*_n = 0, \ n \in \mathcal{N}^+, $$

$$-\int_\tau^I A_n(\tau) x(\tau) d\tau + x^*_n = 0, \ n \in \mathcal{N}^-, $$

$$\int_\tau^I A_n(\tau) x(\tau) d\tau = 0, \ n \in \mathcal{N}^0, $$

$$\int_\tau^I x(\tau) d\tau = 1, $$

$$x(t), x^*(t) \geq 0, \ a.e., $$

$$x_0, x^*_n \geq 0, \ n \in \{1, 2, ..., m\}. $$

Hence, we can, with a minor abuse of notation, write the primal variable as $x = (x, x^*, x_0, x^*) \in X$, where $x^*$ is a shortcut for $x^*_1, ..., x^*_m$. The bilinear operator $\langle x, y \rangle$, with $y = (y, y^*, y_0, y^*) \in Y$, where $y^*$ is a shortcut for $y^*_1, ..., y^*_m$, is defined as follows: $\langle y, x \rangle = \int_\tau^I x(\tau) y(\tau) d\tau + \int_\tau^I x^*(\tau) y^*(\tau) d\tau + y_0 x_0 + \sum_{n=1}^m y^*_n x^*_n$.  

32
1. We define \( c = (0, 0, 1, 0, \ldots, 0) \) with \( m \) components.

2. We define \( A \), a map from \( X \) to \( Z \). Recall that \( \delta_t(\tau) \), a Dirac-impulse at \( \tau = t \) for \( t \in [t, \bar{t}] \) yields \( \int_t^{\bar{t}} \delta_t(\tau)f(\tau)d\tau = f(t) \). We define:

\[
A = \begin{bmatrix}
-\delta_t(\tau) & -\delta_t(\tau) & 1 & 0 & \cdots & 0 & 0 \\
A_1(\tau) & 0 & 0 & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_{m-1}(\tau) & 0 & 0 & 0 & \cdots & -1 & 0 \\
-A_m(\tau) & 0 & 0 & 0 & \cdots & 0 & -1 \\
A_{m+1}(\tau) & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_{\hat{n}}(\tau) & 0 & 0 & 0 & \cdots & 0 & 0 \\
1 & 0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix}.
\]

3. We define \( b = (0, 0, 0, \ldots, 0, 1) \) with \( \hat{n} \) components.

With \( A \), \( b \) and \( c \), we can now write:

1. The objective function as \( \langle x, c \rangle \), where \( \langle x, c \rangle = x_0 \).

2. The constraints are \( Ax = b \), where \( \langle x, a_0 \rangle = b_0 \) a.e. and \( \langle x, a_n \rangle = b_n \) for \( 1 \leq n \leq \hat{n} + 1 \) where \( a_n \) is the \( n + 1 \)-st row of \( A \).

Let the function \( 0(\tau) \) over \([t, \bar{t}]\) be defined as \( 0(\tau) = 0 \) over \([t, \bar{t}]\). For convenience, we may drop the argument, \( \tau \), from \( 0(\tau) \) and write \( 0 \). \( 0 \) is the null-function in \( L_p([t, \bar{t}]) \). \( \theta \) is the null-vector in \( \mathbb{R}^m \). Let \( \theta \) be the null-vector in \( X \). The positive cone in \( X \) is \( P \) and \( P = \{ (x, x^s, x_0, x^s) \in X : x \geq 0 \) a.e., \( x^s \geq 0 \), a.e., \( x_0 \geq 0 \}, \) \( x^s \geq \theta \} \). Based on Anderson and Nash, p. 38-39, we write the primal as

\[
(\mathcal{P}) : \min \langle x, c \rangle \text{ subject to } Ax = b, \ x \geq \theta, \text{ where } x \in X, \ b \in Z, \ c \in Y.
\]

Recall that the dual space of \( L_p([t, \bar{t}]) \); \( L^*_p([t, \bar{t}]) = L_{p'}([t, \bar{t}]) \) where \( 1/p + 1/p' = 1 \) (Holder’s equality) for \( 1 \leq p < +\infty \). Following Anderson and Nash, p. 39, we introduce spaces \( (X, Y, Z, W) \): The primal variables are in \( X \), the dual variables are in \( W \). \( X (Z) \) is the dual space of \( Y (W) \):

\[
(\mathcal{P}) : \begin{bmatrix} X \\ Y \end{bmatrix} \rightarrow \begin{bmatrix} Z \\ W \end{bmatrix} : (D).
\]

We assume that there exists at least one \( (x_0^0, x^0) \) that satisfy the constraints of Problem (22). Given that the Dirac-impluse is in \( L_1([t, \bar{t}]) \), we must select \( Y = L_1([t, \bar{t}]) \times L_1([t, \bar{t}]) \times \mathbb{R}^{m+1} \). As
a consequence $X = L_\infty([\ell, \bar{t}]) \times L_\infty([\ell, \bar{t}]) \times \mathbb{R}^{m+1}$ is the primal space. For $Z$ and $W$, we choose $L_2([\ell, \bar{t}]) \times \mathbb{R}^{n+1}$. Hence, we obtain:

\[
X = L_\infty([\ell, \bar{t}]) \times L_\infty([\ell, \bar{t}]) \times \mathbb{R}^{m+1},
\]

\[
Y = L_1([\ell, \bar{t}]) \times L_1([\ell, \bar{t}]) \times \mathbb{R}^{m+1},
\]

\[
Z = L_2([\ell, \bar{t}]) \times \mathbb{R}^{n+1},
\]

\[
W = L_2([\ell, \bar{t}]) \times \mathbb{R}^{n+1}.
\]

In order to write the dual, we must introduce the dual cone $P^*$ and the adjoint matrix $A^*$ (Anderson and Nash, p. 37-38):

1. The dual cone is: $P^* \triangleq \{y \in Y : \langle x, y \rangle \geq 0 \text{ for all } x \in P\}$ or, rewritten,

\[
P^* \triangleq \left\{ y \in Y : \int_{\ell}^{\bar{t}} x(\tau) y(\tau) d\tau + \int_{\ell}^{\bar{t}} x^*(\tau) y^*(\tau) d\tau + y_0 x_0 + \sum_{n=1}^{m} y_n^\tau x_n^\tau \geq 0 \text{ for all } x \in P \right\}.
\]

2. The adjoint matrix is: $A^*$, a map from $Y$ to $W$ is such that $\langle Ax, w \rangle = \langle x, A^* w \rangle$ for each $x$ in $X$ and $w$ in $W$, or:

\[
A^* = \left[ \begin{array}{ccccccc}
-\delta_{x}(t) & A_{1}(\tau) & \cdots & A_{m-1}(\tau) & -A_{m}(\tau) & \cdots & A_{n}(\tau) & 1 \\
-\delta_{y}(t) & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & -1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & -1 & -1 & \cdots & 0 & 0 \\
\end{array} \right].
\]

**Formulation of the dual** (Anderson and Nash, p. 39): The dual is \( (D) : \max \langle b, w \rangle \) subject to $-A^* w + c \in P^*$ and $w \in W$. Note that by selecting $x_0 = 1$ and $x = 0$, every $y_0$ in $(y, y_0, y) \in P^*$ (where $y$ is a shortcut for $y_1, ..., y_m$) needs to satisfy $y_0 \geq 0$. Similarly, selecting $x_0 = 0$ and $x(\tau) = \delta_t(\tau)$, $(y_0, y, y) \in P^*$ every $y$ in $(y, y_0, y) \in P^*$ needs to satisfy $y(t) \geq 0$. Therefore, $P^* = \{(y, y_0, y) \in Y, y \geq 0, y \geq 0, \text{a.e., } y_0 \geq 0, \text{a.e.}\}$. Hence, we can rewrite the dual $(D)$ as $\langle b, w \rangle = w_{n+1}$:

\[
(D) : w^* = \max w_{n+1}
\]

\[
\sum_{n=1}^{\hat{n}} (-1)^{n} w_{n} A_{n}(\tau) + w_{n+1} \leq w_{0}(\tau), \ \text{a.e.,}
\]

\[
\frac{1}{\ell} \int_{\ell}^{\bar{t}} w_{0}(\tau) d\tau \leq 1,
\]

\[
w_{n} \geq 0, \quad n \in \mathcal{N}^- \cup \mathcal{N}^+,
\]

\[
w_{0}(\tau) \geq 0, \quad \text{a.e.}
\]
(ii) Strong duality. Now, we define the set $H$ in $Z \times \mathbb{R}$ as follows:

$$H = \{(Ax, (x, c) + r'), x \in P, r' \geq 0, r' \in \mathbb{R}\}.$$

From Theorem 3.9 in Anderson and Nash (p. 52), we have that if $H$ is closed and $(P)$ is consistent with a finite value, there is no duality gap.

1. $(P)$ is said to be consistent when if it has a feasible solution (Anderson and Nash, p. 38). By assumption that $(x^k, x^o, x^*)$ satisfies the constraints of Problem (22), $(P)$ is consistent. The value of $(P)$ is the infimum over all feasible $x$ of $(x, c)$. As $x_0^* < +\infty$, $(P)$ has a finite value.

2. Recall that $H \subseteq Z \times \mathbb{R} = L_2(\mathbb{L}, \mathbb{L}) \times \mathbb{R}^{n+1} \times \mathbb{R}$. We can write $H$ as follows.

$$H = \left\{ (x_0 - x(t) - x^s(t), \int \limits_\mathbb{L} (-1)^{\delta - (n)} A_n(\tau)x(\tau)d\tau - x_n^s, n \in \{1, \ldots, m\}, \int \limits_\mathbb{L} A_n(\tau)x(\tau)d\tau, n \in \{m+1, \ldots, n\}, \int \limits_\mathbb{L} x(\tau)d\tau, x(t), x, r', t \in [\mathbb{L}, \mathbb{L}], x \in X, r' \in \mathbb{R}, x(t) \geq 0 \text{ a.e.}, x^s(t) \geq 0 \text{ a.e.}, x_0 \geq 0, x^s \geq 0, r' \geq 0 \right\}$$

To show $H$ is closed consider a Cauchy sequence $(h^1, h^2, \ldots)$ with $h^k \in H$ and let $\lim_{k \to \infty} h^k = h$. For every $h^k$, there exist a $\pi^k = (x^k, x^s, x^0, x^k) \in P \times X$ and $r^k \in \mathbb{R}$ such that $h^k = (Ax^k, (\pi^k, c) + r^k)$. As the mappings $A\pi^k$ and $(\pi^k, c) + r^k$ are finite, $(\pi^k, r^k)$ is also a Cauchy sequence in $X \times \mathbb{R}$. As $X$ and $\mathbb{R}$ are complete, $\lim_{k \to \infty} \pi^k = \pi \in X$ and also $\lim_{k \to \infty} r^k = r' \in \mathbb{R}$. It can be shown that $h = (A\pi, (\pi, c) + r') \in Z \times \mathbb{R}$. Furthermore, as $\pi^k \in P$, also $\pi \in P$ and as $r^k \geq 0$, also $r' \geq 0$. As a consequence, there exist a $(\pi, r') \in P$ such that $h = (A\pi, (\pi, c) + r')$. Hence, $H$ is closed.

As $H$ is closed and $(P)$ has a finite value and is consistent (by assumption), we have via Anderson and Nash, Theorem 3.9, p. 52 that there is no duality gap; $w^* = z^*$.

(iii) Complementary Slackness. Due to Theorem 3.2 in Anderson and Nash (1987, p. 39), if $x$ is feasible for $(P)$ and $w$ is feasible for $(D)$ and $(x, c - A^w) = 0$ then $x$ is optimal for $(P)$ and $w$ is optimal for $(D)$. That is, $x^*$ and $w^*$ must satisfy

$$\begin{align*}
\sum_{n=1}^{n} \delta^{-}\left(\begin{array}{c}
w_n^s A_n(\tau) - w_{n+1}^s + w_n^0(\tau) \\
x_n^s(\tau)w_n^0(\tau) \\
x_n^s(1 - \int_\mathbb{L} w_n^0(\tau)d\tau)
\end{array}\right) &= 0, \text{ a.e.,} \\
x_n^*(\tau)w_n^0(\tau) &= 0, \text{ a.e.,} \\
x_n^*(1 - \int_\mathbb{L} w_n^0(\tau)d\tau) &= 0, \text{ a.e.}
\end{align*}$$

$$n \in \mathcal{N}^0 \cup \mathcal{N}^+$$

35
and primal and dual feasibility constraints:

\[
\begin{align*}
-x^*(t) - x^s(t) + x^*_0 &= 0, \text{ a.e.,} \\
\int_{\mathcal{T}} A_n(\tau)x^*(\tau)d\tau - x^*_n &= 0, \quad n \in \mathcal{N}^+ \\
-\int_{\mathcal{T}} A_n(\tau)x^s(\tau)d\tau - x^s_n &= 0, \quad n \in \mathcal{N}^- \\
\int_{\mathcal{T}} A_n(\tau)x^s(\tau)d\tau &= 0, \quad n \in \mathcal{N}^0 \\
\int_{\mathcal{T}} x^*(\tau)d\tau &= 1, \\
x^*(t), x^s(t), w^*_0(\tau) &\geq 0, \text{ a.e.,} \\
x^*_0, x^s_n &\geq 0, \quad n \in \mathcal{N}^- \cup \mathcal{N}^+
\end{align*}
\]

or, after eliminating \(x^s_n\) and \(x^*(t)\) for \(n \in \mathcal{N}^- \cup \mathcal{N}^+\), \((x^*(t), x^*_0)\) is primary feasible and \((w^*_0(\tau), w^*_n, n \in \{1, 2, \ldots, \hat{n}\})\) is dual feasible and

\[
\begin{align*}
x^*(\tau)\left(-\sum_{n=1}^{\hat{n}}(-1)^{\delta^-(n)}w^*_nA_n(\tau) - w^*_{n+1} + w^*_0(\tau)\right) &= 0, \text{ a.e.,} \\
w^*_0(\tau)(x_0^* - x^*(\tau)) &= 0, \text{ a.e.,} \\
x^*_0(1 - \int_{\mathcal{T}} w^*_0(\tau)d\tau) &= 0, \\
w^*_n \int_{\mathcal{T}} A_n(\tau)x^*(\tau)d\tau &= 0, \quad n \in \mathcal{N}^- \cup \mathcal{N}^+
\end{align*}
\]

**Proof of Proposition 4.** As the primal is feasible, \(x^*_0 > 0\), due to strong duality, \(w^* = x^*_0 > 0\). Now consider the optimal solution and define two sets, \(\mathcal{N}^*\) and \(\mathcal{T}^*\):

- Assume that \(w^*_n > 0\) for \(n \in \mathcal{N}^* \cap \{\mathcal{N}^+ \cup \mathcal{N}^-\}\) and \(w^*_n \neq 0\) for \(n \in \mathcal{N}^* \cap \mathcal{N}^0\), where \(\mathcal{N}^* \subseteq \mathcal{N}\) and \(w^*_n = 0\) for all other \(n \in \mathcal{N} \setminus \mathcal{N}^*\).

- Assume that \(x^*(\tau) > 0\) for \(\tau \in \mathcal{T}^*\) (note \(\mathcal{T}^* \neq \emptyset\)) and \(x^*(\tau) = 0\) for \(\tau \notin \mathcal{T}^*\).

Now we find a relationship between \(\mathcal{T}^*\) and \(\mathcal{N}^*\) via Equation (26) and primal and dual feasibility (see constraints in Equations (22) and (23)), strong duality \((x^*_0 = w^*)\) and the assumption that the primal is feasible \((x^*_0 > 0)\). It follows immediately from Equation (26) that

\[
\begin{align*}
\int_{\mathcal{T}^*} x^*(\tau)d\tau &= 1, \quad \text{(A)} \\
\int_{\mathcal{T}^*} w^*_0(\tau)d\tau &= 1, \quad \text{(B)} \\
\sum_{n \in \mathcal{N}^*} -w^*_n(-1)^{\delta^-(n)}A_n(\tau) + x^*_0 &= w^*_0(\tau), \quad \tau \in \mathcal{T}^*, \quad \text{(C)} \\
0 &= w^*_0(\tau), \quad \tau \notin \mathcal{T}^*, \quad \text{(D)} \\
x^*_0 &= x^*(\tau), \quad w^*_0(\tau) > 0 \quad \text{(E)} \\
\int_{\mathcal{T}^*} A_n(\tau)x^s(\tau)d\tau &= 0, \quad n \in \mathcal{N}^*, \quad \text{(F)}
\end{align*}
\]

\(w^*_n > 0, n \in \mathcal{N}^* \cap \{\mathcal{N}^+ \cup \mathcal{N}^-\}, w^*_n \neq 0, n \in \mathcal{N}^* \cap \mathcal{N}^0, w^*_0(\tau) \geq 0, \text{ a.e.} \)

Note that \(w^*_0(\tau) = 0\) for \(\tau \in \mathcal{T}^*\) can only hold for a set with measure zero as \(A_n(\tau)\) has exactly
Therefore, from Equation (27A),
\[ x \in \sum_{n} w_n \leq 0 \] for \( n \in N^* \).

We have thus obtained that the solution must satisfy the following conditions.

\[ \sum_{n \in N^*} w_n^* (\tau) = 0 \] for \( \tau \notin T^* \).

Hence, for given \( w_n^* \) and \( x_n^* \), the support of the density, \( T^* \) is determined by the sign of
\[ \sum_{n \in N^*} w_n^* (\tau) = 0 \] for \( \tau \in T^* \). That is, the optimal density is piecewise uniform.

Therefore, from Equation (27A),
\[ x_n^* = \left( \int_{T^*} d\tau \right)^{-1} \] and we obtained the following structure for \( x^*(\tau) \).

\[ x^*(\tau) = \begin{cases} (\int_{T^*} d\tau)^{-1}, & \tau \in T^* \text{ (a.e.)}, \\ 0, & \tau \notin T^* \end{cases} \]

Equation (27F) becomes \( \int_{T^*} A_n(\tau) x_n^* d\tau = 0 \), or, as \( x_n^* > 0 \), we obtain \( \int_{T^*} A_n(\tau) d\tau = 0, n \in N^* \). Equation (27B) is then always satisfied: Using Equation (27D), \( \int_{T^*} w_n^*(\tau) d\tau = \int_{T^*} x_n^* d\tau \) and

\[ \int_{T^*} w_n^*(\tau) d\tau = \int_{T^*} \left\{ \sum_{n \in N^*} w_n^* (\tau) A_n(\tau) + x_n^* \right\} d\tau \\
= \sum_{n \in N^*} w_n^* (\tau) \int_{T^*} A_n(\tau) d\tau + \int_{T^*} x_n^* d\tau. \]

In conclusion, the solution must thus satisfy the following conditions.

\[ \sum_{n \in N^*} (-1)^{\delta^-(\tau)} w_n^* A_n(\tau) + \left( \int_{T^*} d\tau \right)^{-1} > 0 \Leftrightarrow \tau \in T^* \text{ (a.e.)} \] and \( \int_{T^*} A_n(\tau) d\tau = 0, n \in N^* \), for
\[ w_n^* > 0, n \in N^* \cap \{N^+ \cup N^- \} \text{ and } w_n^* \neq 0, n \in N^* \cap N^0. \]

\[ \square \]

References


