When Fixed Price Meets Priority Auctions: Competing Firms with Different Pricing and Service Rules*

Jiayang Gao, Krishnamurthy Iyer, and Huseyin Topaloglu

School of Operations Research and Information Engineering
Cornell University, Ithaca, NY 14853, USA
email: {jg838,kriyer,ht88}@cornell.edu

September 19, 2018

Abstract

We consider a service system with two competing firms offering service via two different pricing and service rules. With the fixed-price firm, customers obtain service at a fixed price, and have homogeneous expected waiting times. With the bid-based firm, customers submit a bid to obtain service, and incur expected waiting times decreasing in their bids, and make payments equal to their bids. We assume the customers are heterogeneous, with different waiting costs, and choose the firm from which to obtain service strategically on arrival. We establish the existence and uniqueness of a symmetric equilibrium for the customers' decision problem for any given fixed and reserve price set by the two firms, and characterize the structure of the resulting equilibrium strategy. In particular, we show that the customers' equilibrium strategy has a simple threshold structure, where customers with either high or low waiting cost choose to obtain service from the bid-based firm, whereas those with moderate waiting cost choose to obtain service from the fixed-price firm. We use this characterization of the equilibrium strategy to study the price competition between the two firms, and show that there exists a mixed Nash equilibrium. Moreover, we analyze price and capacity competition between the firms in a limiting regime where the customer arrival rate and the service capacities increase proportionally, and show that this competition inherently favors the bid-based firm.

*Keywords: priority auctions, bidding, competition, pricing, game theory, priority queues

1 Introduction

Competing firms often offer their products and services through various modes differing in their pricing structure and service quality. One main motivation behind this service differentiation among competing firms is to target heterogeneous customers, differing in their preferences over the quality and urgency of service, their tolerance for uncertainty, etc. For example, in the cloud computing service Amazon EC2, the customers can choose to obtain service by bidding for a computing resource in a quasi-auction market ("spot instances", see Amazon (2015)), whereas in the competing service Microsoft Azure, customers pay for a fixed-price per hourly use (Microsoft 2017). Similarly, in the local transportation industry, a regular taxi service company offers a fixed price for a specific origin-destination pair, while Didi Dache, a major competitor of the taxi companies in China, gives customers the option to offer a customized amount of tip to the driver when they request for a car, and the drivers are more likely to accept a request as the amount of tip increases. (For more information about Didi Dache and tipping, see Chohan (2014) and Palmer (2015).) In this case, the Didi customers are essentially bidding in a priority queue.

With an assortment of options to obtain services, customers have to make strategic trade-offs among the cost, quality and priority of service. This choice is further complicated by the fact that, owing to the presence of resource constraints on the firms' part, each customer's choice among the different firms influences and is influenced by how other customers make the same decision. This suggests that the characteristics of the set of customers availing service from a particular firm is determined endogenously, and an equilibrium analysis is needed to understand the customers' decision. Further, each firm has to model the customers' equilibrium response in order to evaluate and optimize the design of their service in order to target customers with a specific set of characteristics.

To understand the equilibrium behavior of customers and the resulting competition among firms offering different pricing and service rules, in this paper, we consider a setting consisting of customers with heterogeneous delay-sensitivity, and two competing service providers who offer service via two different pricing and service rules. In the fixed-price firm, the service is offered at a fixed price, and the service quality (waiting time) is the same for everyone who chooses to obtain service from this firm. On the other hand, in the bid-based firm, the service is provided according to a first-price priority auction, where each customer on arrival submits a bid equal to the price they will pay for service, and the expected waiting time is decreasing in the bids. Each customer chooses the firm (and a bid, if she chooses the bid-based firm) that maximizes her total expected utility, given by the difference between the utility of receiving service, and her expected cost. This cost is comprised of the customer's payment for service and the delay cost she incurs by waiting until service completion.

A customer's strategy in this context is a function of her unit waiting cost. The strategy consists of her choice of the firm to obtain service from (or to balk without obtaining service), and her bid in the bid-based firm if she chooses to obtain service there. We analyze a symmetric equilibrium of the preceding system, where all customers adopt the same strategy, and where each customer is making a best response to others' strategies.

For given fixed price in the fixed-price firm, we show by explicit construction that there exists a *unique* symmetric equilibrium. More importantly, we characterize the structure of the equilibrium strategy, and show that it has a multi-threshold structure: the bid-based firm is visited by customers with very high or very low delay-sensitivity, whereas the fixed-price firm is visited by customers with moderate delay-sensitivity. This result can be extended to the case where the bid-based firm charges a reserve price.

This structural characterization of the symmetric equilibrium of the customers' game provides the main insight of our paper, namely that the bid-based firm simultaneously enables two sets of customers, with markedly different characteristics, to make optimal trade-offs between the delay costs and the cost of obtaining service. On one hand, for those customers with very low unit waiting cost, it acts as a means to reduce their total cost by submitting low bids and waiting longer for service. On the other hand, for customers with very high unit waiting costs, it acts as a venue to demand high priority (and lower waiting time) for service. This insight coincides with what we see in the China's local transportation service industry, in which Didi Dache has a reserve price that is much lower than the price charged by the taxi company (usually about 30% lower, see Tian (2015)), but some delay-sensitive customers choose to add a considerable amount of tip to avoid waiting (see Springwise (2013)).

Finally, using the uniqueness of the customers' equilibrium strategy, we study the price and capacity competition game between the two firms, where the fixed-price firm sets the fixed-price and the bid-based firm sets the reserve price, in a limiting regime where the arrival rate and the firms' service capacity increase proportionally. We summarize our main contributions as follows:

(1) Characterization of structure of customers' equilibrium strategy: We show that, in any symmetric equilibrium, the customers' strategy has a simple multi-threshold structure: among all customers who choose to obtain service from the system, the customers with relatively high per unit time waiting cost and those with relatively low per unit time waiting costs choose to obtain service from the bid-based firm. On the other hand, the customers who obtain service from the fixed-price firm have relatively moderate per unit time waiting cost. Consequently, in this market, the fixed-price firm only attracts customers with moderate waiting costs.

(2) Existence and uniqueness of customers' equilibrium: Using the structure of the equilibrium strategy, we show the existence and uniqueness of a symmetric equilibrium of the customers' game. Our proof proceeds by obtaining necessary and sufficient conditions on a threshold strategy to constitute a symmetric equilibrium. These conditions are obtained by imposing the continuity of the expected waiting times and total cost of the customers with unit waiting costs at the thresholds. Finally, we show the existence (and uniqueness) by explicitly constructing a solution satisfying these conditions.

(3) Competition between firms: Finally, we investigate how competing firms set their prices in order to maximize their revenue in equilibrium. Using the uniqueness of the customers' equilibrium strategy, we study the resulting price competition between the two firms, where the fixed-price firm sets the fixed-price and the bid-based firm sets the reserve price, under a limiting regime where the arrival rate and the firms' service capacity increase proportionally. Given the equilibrium for this price competition, we then study the capacity competition between the two firms facing capacity costs. We observe that the bid-based firm has an inherent advantage in this competition: with equal capacity costs, there is never an equilibrium where the fixed-price firm is a monopolist, whereas there exists values of the capacity cost for which in equilibrium the bid-based firm is a monopolist.

1.1 Literature review

There are multiple strands of literature from game theory and auctions, revenue management, and queueing theory that are related to our work.

Our work builds on existing work on revenue maximizing pricing policies of a monopolist, in the presence of strategic customers. Mendelson and Whang (1990) consider a queueing system with a single server and finite number of customer classes each with a different per unit time waiting cost and service time distribution. The authors devise an incentive compatible pricing policy, such that the social welfare is maximized when each customer class endogenously choose their arrival rate and priority. Similar models have been studied in the context of a revenue maximizing service provider (Afèche 2013, Yahalom et al. 2006, Gavirneni and Kulkarni 2016).

Although we do not specify a queueing model in our paper, our results are applicable to many commonly used queueing models (e.g., $G/M/k$ priority queues), since the expected waiting time function under those models satisfy our assumptions. Thus, the analysis of customers' behavior in the bid-based firm is closely related to that under priority queues. One of the earliest work in priority queues is by Kleinrock (1967), who considers a model of an $M/M/1$ queue where service is provided in the decreasing order of the customers' bids, and where the customers are non-strategic with an exogenously specified bid distribution. In this setting, he obtains expressions for the expected waiting time as a function of the bid in both preemptive and non-preemptive settings. Balachandran (1972), Lui (1985), Glazer and Hassin (1986), Hassin (1995), Kittsteiner and Moldovanu (2005) build on this work by considering strategic customers that determine their priorities endogenously through their payments. Our analysis of the bid-based queue further makes use of results from auction theory (Myerson 1981, Krishna 2009). More broadly, our work contributes to the literature at the intersection of game theory and queueing theory; see Hassin and Haviv (2003) for a comprehensive survey of various models of queueing systems with strategic customers and servers.

There are a number of papers that study a market with firms offering a multitude of price-quality combinations (offered by either a monopolist or competing firms) for customers to choose from. Mussa and Rosen (1978) consider a monopolist choosing a price-quality schedule to maximize its revenue. Afèche and Pavlin (2016) consider revenue maximizing price/lead-time menu for a single server system where customers' value for service completion is a monotone continuous function of their per unit time waiting cost. Nazerzadeh and Randhawa (2014) study a similar model where the per unit time waiting costs are a linear or sublinear function of the value of service, and they consider the asymptotic regime with large arrival rates and service capacity. See also Van Mieghem (2000) and Maglaras et al. (2013) for related models studying pricing and scheduling policy in the asymptotic large system regime. Kilcioglu and Rao (2016) study price and quality competition in the cloud computing market. More closely related is the paper by Kilcioglu and Maglaras (2015), who consider a revenue maximizing service provider with infinite capacity offering two service options, one with guaranteed service availability, and the other where users submit bids and the winning bids get served. The authors show that even with infinite capacity, the service provider may find it optimal to make the latter service option stochastically unavailable. In these papers, since the service provider guarantees each price-quality combination in a menu, a customer's expected utility is not affected by the actions of other customers. Thus, the decision problem of each customer gets decoupled, and the customers optimal price-quality choice can be obtained by solving an optimization problem. (We note that, in these papers, the choice of the optimal price-quality menu is a result of a Stackelberg game between the service provider and the customers.) In contrast, in our model, a customer's quality of service, as measured by her expected waiting time, is a function of both her own action, as well as those of others. Consequently, the customer behavior arises as the equilibrium outcome of a game among the customers.

There are some papers which consider the game among customers in a market with multiple pricing and service policies (Etzion et al. 2006, Caldentey and Vulcano 2007, Wang et al. 2012, Abhishek et al. 2012). Afèche and Mendelson (2004) analyze a queueing system under different pricing and service policies, where customers are strategic and have a value for service that depends multiplicatively on their delay cost. Abhishek et al. (2012) model the Amazon EC2 cloud computing service as a hybrid system where customers can choose to enter a bid-based priority queue, or to obtain service from a fixed-price queue with infinite capacity. The authors show that any equilibrium has a single threshold structure, with all customers below the threshold entering the bid-based priority queue, and the rest the fixed-price queue. Caldentey and Vulcano (2007) prove the existence of similar single threshold strategy equilibrium in the context of a seller adopting a fixed-price or an auction mechanism to sell their products. In contrast, we model a setting where the customers obtaining service from the fixed-price firm encounter congestion and have to wait for service. This waiting for service in the fixed-price firm induces customers with high unit waiting cost to choose the bid-based firm in equilibrium, resulting in an equilibrium strategy with a multi-threshold structure.

In our model, the customers with intermediate types (defined by their unit time waiting costs) choose the fixed-price firm, while the customers on the two extremes choose the bid-based firm. Similar intermediate-versus-extremes equilibrium structure was recognized in several other papers under different circumstances. Yang et al. (2015) consider trading position on a FIFO queue with an intermediary at a fee, when customers have different unit time waiting costs. They show that in equilibrium, customers with intermediate waiting costs do not participate in the trade and remain in their FIFO position, while customers with lower or higher waiting costs trade their priorities. Afèche and Pavlin (2016) study the design of a price/lead-time menu in order to maximize revenue, when customers differ in patience levels. They conclude that pricing out the customers with intermediate patience levels while serving the most patient and impatient ones may increase revenue.

The rest of this paper is organized as follows. In Section 2, we describe our model of a market with a fixed-price firm and a bid-based firm. In Section 3, we study the customers' game, and characterize the structure of the strategy in any symmetric equilibrium. Using this structure, in Section 4 we obtain necessary and sufficient conditions for a strategy to constitute an equilibrium, and prove our main result, namely the existence and uniqueness of a symmetric equilibrium of the customers' game. Finally, in Section 5, we study the firms' game, and prove the existence of mixed Nash equilibrium. We also study the price of stability of the firms' game in a limiting regime, and characterize the conditions under which it is small.

2 Model

Consider a setting with two competing firms, a fixed-price firm and a bid-based firm, offering service to a Poisson stream of customers with rate $\lambda > 0$. The fixed-price firm charges a fixed price $P > 0$ to offer a service where all customers incur identical expected waiting times until service completion. This identical expected waiting time increases with the arrival rate of customers. We assume that the fixed-price firm has a service capacity of n, such that if the arrival rate of customers to the firm is greater than or equal to n, then the firm can no longer ensure finite expected waiting times. For example, the fixed-price firm may use a $M/G/n$ FIFO queue to provide service to its customers.

On the other hand, the bid-based firm provides service via a first-price auction (Krishna 2009) with reserve price r. More precisely, customers arriving to the bid-based firm submit a bid no less than r on arrival that denotes the price they are willing to pay for service. Customers are then served in the decreasing order of their bids (with ties broken uniformly at random), and are charged their bid upon service completion. We assume that the bid-based firm has a service capacity of

k, implying that if the total arrival rate of customers to the firm is greater than or equal to k, then the firm cannot ensure finite expected waiting time to all of its customers. For most of our analysis, we assume that a customer of the bid-based firm experiences an expected waiting time that depends only on the arrival rates of customers with higher bids. For example, the bid-based firm may provide service to its customers using a $G/M/k$ preemptive priority queue. (Note that for a non-preemptive priority queue, a customer's expected waiting time also depends on the arrival rate of customers with lower bids. Our analysis in Section 3 and Section 4 can be reproduced for non-preemptive queues; see Appendix G for details.)

Each arriving customer is characterized by three features: their service requirement, their value for service completion and their cost for waiting until service completion. We address each one separately below.

We assume that each customer's service requirement is exponentially (and independently) distributed with mean 1 (equal to the service rate $\mu = 1$). The homogeneity of service requirement is a restrictive assumption, and a more general model will allow for non-homogeneous distribution for the service requirement. On the other hand, our model can easily accommodate more general distributions for service requirement, and the independence assumption can be expected to hold in many service systems where the customer base is fairly large and diverse.

Next, each customer obtains a deterministic value $V > 0$ upon service completion. As mentioned earlier, we assume that this value is identical across customers. Although this assumption may seem restrictive, one justification arises out of the interpretation of the value V as the opportunity cost faced by the customer. For example, one may consider V to be fixed price charged by a competitor to the two firms who offers service with negligible waiting times. In such settings, if each customer has access to this competitor, then our assumption of identical value of service completion holds. However, if not all customers can access this competitor, then a more appropriate model would require different values of service completion for different (classes of) customers.

Finally, we assume that the customers incur a heterogeneous cost for waiting until service completion. More precisely, each customer incurs a disutility proportional to the total time she spends in the system until service completion, and we refer to the proportionality constant as the customer's unit waiting cost. We assume that each customer's unit waiting cost c is drawn independently and identically from a continuously differentiable bounded distribution F. For ease of notation, we assume that this distribution F is the uniform distribution on $[0, 1]$ in this paper. All of our analytical results extend directly to the general case, as we discuss in Appendix A.

We assume that the arrival rate λ, the price P in the fixed-price firm, the reserve price r in the bid-based firm, the distribution of the service requirement, the value of service completion V, the service capacities k and n, and the distribution F of customers' unit waiting cost are common knowledge among the customers and the two service providers.

On arrival to the system, each customer decides based on her unit waiting cost, whether to obtain service and if so, from which firm. If she decides to obtain service from the bid-based firm, then she further chooses a bid to submit. A customer choosing not to obtain service leaves the system never to return, and obtains zero utility. A customer with unit waiting cost c waiting for a time W until service completion, and making a payment m obtains a utility equal to $V - c \cdot W - m$. (We refer to the quantity $c \cdot W + m$ as the total cost incurred by the customer.) We assume that the customers are strategic and seek to maximize their total expected utility, where the expectation is with respect to the steady state distribution of the system. (Since the customers are free to balk upon arrival, the two service systems are stable and a steady-state distribution exists in any equilibrium.) Implicitly, this steady state expectation entails assuming that the customers cannot observe the state of the system, such as the queue lengths in each firm or the existing bids in the bid-based firm, before making their decision. This assumption is valid in many settings, especially when the queue is not physical, e.g., in call centers, online service industry, etc.

Consequently, we represent a customer's strategy by a pair of functions $x(\cdot)$ and $\varphi(\cdot)$ of her unit waiting cost c, where $x(c) \in \{\mathsf{LEAVE}, \mathsf{FIX}, \mathsf{BID}\}$ denotes her decision about whether to obtain service and if so, from which firm, and $\varphi(c) \geq 0$ denotes her bid upon joining the bid-based firm. We refer to the function $x(\cdot)$ as the customer's service decision and the function $\varphi(\cdot)$ as her bid function. For the sake of completeness, we define $\varphi(c) = P$ if $x(c) = \mathsf{FIX}$ and $\varphi(c) = V$ if $x(c) = \mathsf{LEAVE}$.

We focus on the symmetric setting, where all customers follow the same strategy (x, φ). In this scenario, we let $w_F(x, \varphi)$ denote the expected waiting time in the fixed-price firm in steady state. For a customer with unit waiting cost c, the expected total cost on receiving service from the fixed-price firm is then given by $cw_F(x, \varphi) + P$. Similarly, we let $w_B(b'|x, \varphi)$ denote the expected waiting time in steady state for a customer joining the bid-based firm and making a bid b'. The total expected cost for such a customer with unit waiting cost c is then given by $cw_B(b'|x, \varphi) + b'$.

We say a strategy (x, φ) forms a symmetric equilibrium if, assuming all other customers act according to the strategy (x, φ), each customer's expected utility is maximized by following the same strategy (x, φ). Formally, we require that (x, φ) satisfy the following conditions:

$$x(c) = \begin{cases} \textsf{LEAVE}, & \text{only if } V \leq \min\{\min_{b'}\{cw_B(b'|x,\varphi) + b'\}, cw_F(x,\varphi) + P\}; \\ \textsf{FIX}, & \text{only if } cw_F(x,\varphi) + P \leq \min\{V, \min_{b'}\{cw_B(b'|x,\varphi) + b'\}\}; \\ \textsf{BID}, & \text{only if } \min_{b'}\{cw_B(b'|x,\varphi) + b'\} \leq \min\{cw_F(x,\varphi) + P, V\}, \end{cases}$$

and

$$\varphi(c) \in \arg\min_{b'} \{cw_B(b'|x,\varphi) + b'\}, \quad \text{if } x(c) = \textsf{BID}.$$

Here, we break ties arbitrarily. The first condition specifies that the customer will choose to obtain the service only if the total expected cost is less than or equal to the value of service completion. In this case, the customer will choose the fixed-price firm if the total expected cost therein is no more than that in the bid-based firm under the best possible bid. Otherwise, the customer will choose the bid-based firm. The second condition requires that upon choosing to obtain service from the bid-based firm, the customer will enter a bid that minimizes her total expected cost.

3 Equilibrium structure of the customers' game

In this section, we characterize the structure of the equilibrium strategy of the customers' game. In particular, in Section 3.1, we show that in any symmetric equilibrium, the bidding function is completely determined by the customers' service decision. Furthermore, in Section 3.2, we show that in any symmetric equilibrium, the customers' service decision has a simple multi-threshold structure. We use this structural characterization of the symmetric equilibrium later in Section 4 to prove the existence (and uniqueness) of the customers' equilibrium. For ease of notation, we assume in Sections 3 and 4 that $r = 0$. We extend all results to the case where $r > 0$ in Section 4.3.

3.1 Structure of customers' equilibrium bidding function

We begin our analysis of the symmetric equilibrium by focusing on the equilibrium bidding function. We show that in a symmetric equilibrium (x, φ), the bidding function $\varphi(\cdot)$ is completely specified once the service decision $x(\cdot)$ is known. Towards this, for $c \in [0,1]$, define $B(c|(x,\varphi))$ as the proportion of arrivals to the bid-based firm with unit waiting cost below c: $B(c|(x,\varphi)) \triangleq \mathbf{P}(\widehat{C} \leq c \mid x(\widehat{C}) = \mathsf{BID})$, where $\widehat{C}$ is a random variable distributed as F. We have the following lemma showing the monotonicity of the expected waiting time, payment, and the total cost in any symmetric equilibrium. The proof relies on the fact that a customer does not gain by unilaterally deviating from the equilibrium strategy. We provide the details in Appendix B.

Lemma 1. *In any symmetric equilibrium (x, φ) of the customers' game, for all customers that choose to obtain service, the expected waiting time is non-increasing, the expected payment is non-decreasing, and the total expected cost is strictly increasing in the unit waiting cost. Moreover, the bidding function is strictly increasing at all unit waiting cost c where $B(c|(x,\varphi))$ is strictly increasing.*

The implication of the lemma is as follows. In a symmetric equilibrium, the bidding function $\varphi(c)$ is strictly increasing whenever $B(c|(x,\varphi))$, the proportion of arrivals to the bid-based firm with unit waiting cost lower than c, is strictly increasing. (Note that $B(c|(x,\varphi))$ is always non-decreasing in c.) And since the service in the bid-based firm is in the decreasing order of bids, this implies that in equilibrium, the customers in the bid-based firm are served in decreasing order of the unit waiting cost. Thus, the expected waiting time of a customer in the bid-based firm is solely a function of the service decision $x(\cdot)$ and her unit waiting cost c. We use $w(c|x)$ to denote the expected waiting time of a customer with unit waiting cost c when everyone uses the service decision function $x(\cdot)$. Using this result, we obtain the following characterization of the expected waiting time function and the bidding function in terms of the service decision in equilibrium.

Lemma 2. *In any symmetric equilibrium (x, φ), the expected waiting time function $w(\cdot|x)$ and the bidding function $\varphi(\cdot)$ are completely determined by the customers' service decision $x(\cdot)$. In* *particular, the bidding function satisfies*

$$\varphi(c) = \int_0^c w(t|x)dt - cw(c|x). \tag{1}$$

for all c such that $x(c) \neq \mathsf{LEAVE}$.

Proof sketch. From Lemma 1, we obtain that in equilibrium, a customer's expected waiting time in the bid-based firm depends only on the proportion of arrivals to the bid-based firm with higher unit waiting costs, which in turn depends only on the service decision $x(\cdot)$ and the customer's unit waiting cost. Since the expected waiting time in the fixed-price firm depends only on the proportion of arrivals to the fixed-price firm, this holds true also for a customer in the fixed-price firm. Together, this implies that, in a symmetric equilibrium (x, φ), the expected waiting time of a customer is given by a function $w(c|x)$ only of their unit waiting cost c and the service decision $x(\cdot)$.

To obtain (1), note that in equilibrium, for any customer with unit waiting cost c, the marginal decrease in the expected waiting cost resulting from a marginal increase in the bid, must equal the marginal increase in the resulting payment. Assuming differentiability, this implies $\varphi'(c) = -cw'(t|x)$, which on integrating yields (1). See Appendix B for a rigorous argument. □

As a consequence of Lemma 2, in order to find a symmetric equilibrium, it suffices to focus only on the customers' service decision $x(\cdot)$, and use (1) to obtain the bidding function $\varphi(\cdot)$. Moreover, for any given service decision $x(\cdot)$, if customers bid according to the bidding function given by (1) in the bid-based firm, we observe that the total expected cost of a customer is given by

$$TC(c|x) \triangleq cw(c|x) + \varphi(c) = \int_0^c w(t|x)dt, \tag{2}$$

for all $x(c) \neq \mathsf{LEAVE}$.

3.2 Structure of the customers' equilibrium service decision

Having characterized the bidding function in a symmetric equilibrium of the customers' game, we now focus on the service decision $x(\cdot)$ in a symmetric equilibrium. Before we proceed, we note that starting from an equilibrium, if we alter the actions of a measure zero set of customers from their current action to a different best response action, the resulting service decision continues to be a symmetric equilibrium (without a specific tie-breaking rule). Thus, in order to avoid unnecessary technicalities, in the rest of the paper, we focus only on those symmetric equilibria where each action is either employed by a set of customers of positive measure, or never adopted by any customers. (In particular, we ignore equilibria where one of the firms services a set of customers with measure zero.) Then, the following theorem, the main result of this section, states that the service decision $x(\cdot)$ in any symmetric equilibrium has a simple multi-threshold structure.

Theorem 1. *Let $V > 0$ and $P > 0$. In any symmetric equilibrium (x, φ), the service decision function $x(\cdot)$ has multi-threshold structure. Specifically, there exists thresholds $0 < c_1 \leq c_2 \leq c_\ell \leq 1$, such that*

$$x(c) = \begin{cases} \mathsf{LEAVE}, & \text{if } c \in (c_\ell, 1]; \\ \mathsf{FIX}, & \text{if } c \in (c_1, c_2); \\ \mathsf{BID}, & \text{if } c \in [0, c_1] \cup [c_2, c_\ell]. \end{cases}$$

For a symmetric equilibrium with no customer obtaining service from the fixed-price firm, the thresholds satisfy $c_1 = c_2 = c_\ell \in (0, 1]$. For a symmetric equilibrium with some customers obtaining service from the fixed-price firm, the thresholds satisfy $0 < c_1 < c_2 < c_\ell \leq 1$.

Proof. Fix a symmetric equilibrium (x, φ). Consider a customer with unit waiting cost c, who obtains service in equilibrium. The total expected cost of such a customer in equilibrium is no more than V, i.e., $cw(c|x) + \varphi(c) \leq V$. Since $w(c|x) \geq 0$, the total expected cost of any customer with unit waiting cost $c' < c$ using action $(x(c), \varphi(c))$ is less than V. Consequently, the expected total cost of such a customer using the optimal action $(x(c'), \varphi(c'))$ is also less than V, and such a customer also chooses to obtain service. Thus, there exists a threshold $c_\ell \in [0, 1]$ such that all customers with unit waiting cost below c_ℓ choose to obtain service (i.e, $x(c) \in \{\mathsf{FIX}, \mathsf{BID}\}$), and the rest choose to leave the system without obtaining service (i.e., $x(c) = \mathsf{LEAVE}$). Observe that for $V > 0$, we have $c_\ell > 0$, for if no other customers obtain service from the system, then it is optimal for small enough $\epsilon > 0$, a customer with unit waiting cost ϵ strictly prefers to obtain service from the bid-based firm by making a small enough bid. (Note that it may be the case that $c_\ell = 1$, implying that no customers choose to leave the system without obtaining service.)

Now, we consider how customers with unit waiting cost below c_ℓ choose between the fixed-price firm and the bid-based firm. Let C_F denote the set of unit waiting costs for which the customer's equilibrium choice is to obtain service from the fixed-price firm. We begin by showing that the set C_F is convex.

Let $c, \hat{c} \in C_F$ with $\hat{c} < c$. Since $x(c) = x(\hat{c}) = \mathsf{FIX}$, we obtain in equilibrium,

$$
\begin{aligned}
cw_F(x) + P &\leq \min_{b'}\{cw_B(b'|x) + b'\}, & cw_F(x) + P &\leq V, \\
\hat{c}w_F(x) + P &\leq \min_{b'}\{\hat{c}w_B(b'|x) + b'\}, & \hat{c}w_F(x) + P &\leq V.
\end{aligned}
$$

Let $\beta \in (0, 1)$, and $\tilde{c} = \beta c + (1 - \beta)\hat{c}$. Taking convex combination of the two inequalities on the right side above, we obtain $\tilde{c}w_F(x) + P \leq V$, implying the customer with unit waiting cost $\tilde{c}$ would prefer obtaining service from the fixed-price firm over leaving without obtaining service. Similarly, we obtain

$$
\begin{aligned}
\tilde{c}w_F(x) + P &\leq \beta \min_{b'}\{cw_B(b'|x) + b'\} + (1 - \beta) \min_{b'}\{\hat{c}w_B(b'|x) + b'\} \\
&< \min_{b'}\{\tilde{c}w_B(b'|x) + b'\}.
\end{aligned}
$$

Thus, the customer with unit waiting cost $\tilde{c}$ would strictly prefer obtaining service from the fixed-price firm over obtaining service from the bid-based firm. Taken together, this implies $x(\tilde{c}) = \mathsf{FIX}$, and hence C_F is convex.

If C_F is empty, then all customers with unit waiting cost below c_ℓ choose to obtain service from the bid-based firm, and the service decision function is given by $x(c) = \mathsf{BID}$ for $c \leq c_\ell$, and $x(c) = \mathsf{LEAVE}$ for $c > c_\ell$. This fits the representation of the service decision function in the theorem with $c_1 \triangleq c_\ell$ and $c_2 \triangleq c_\ell$.

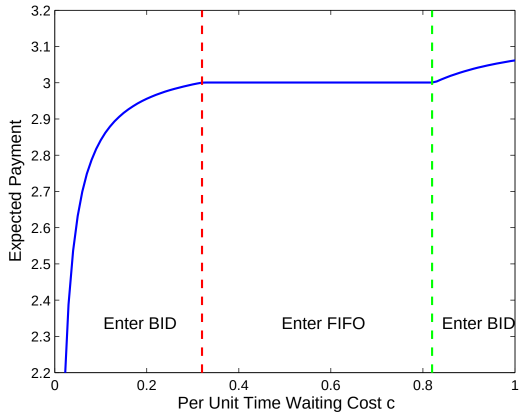
Hence, for the rest of the proof, suppose C_F is non-empty. Define $c_1 \triangleq \inf C_F$ and $c_2 \triangleq \sup C_F$, such that $c_1 < c_2$. We now show that $0 < c_1$ and $c_2 < c_\ell$.

Suppose $c_1 = 0$. By convexity of C_F, this implies that there exists an $\bar{\epsilon} > 0$, such that $x(c) = \mathsf{FIX}$ for all $c < \bar{\epsilon}$. For a customer with unit waiting cost $\epsilon < \bar{\epsilon}$, his total expected cost is $\epsilon w_F(x) + P$. Suppose instead the customer chooses to obtain service from the bid-based firm with a zero bid. This ensures that his expected waiting time is equal to that of a customer with the lowest priority in the bid-based firm. By Lemma 1, the expected waiting time in equilibrium is non-increasing in unit waiting cost, and hence the expected waiting time of a customer with the lowest priority in the bid-based firm is at most $w(\epsilon|x) = w_F(x)$. Thus, the total expected cost of the customer with unit waiting cost ϵ on choosing to obtain service from the bid-based firm with a zero bid is at most $\epsilon w_F(x) < \epsilon w_F(x) + P$. This contradicts the assumption that (x, φ) is an equilibrium. Hence, we obtain that in any symmetric equilibrium, $c_1 > 0$.

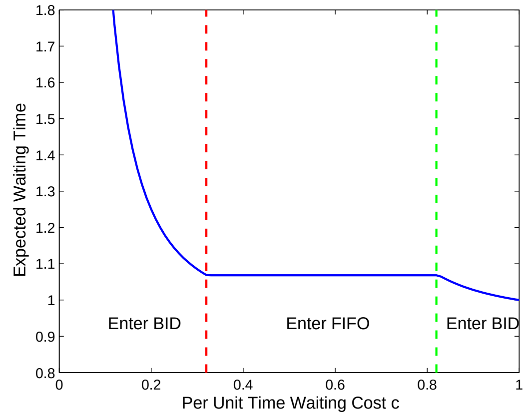
Finally, suppose $c_2 = c_\ell$. Since C_F is convex (and non-empty), this implies that all customers with unit waiting cost below c_1 obtain service from the bid-based firm, all customers with unit waiting cost between c_1 and c_ℓ obtain service from the fixed-price firm, and the rest choose to leave without obtaining service. Consequently, a customer with unit waiting cost c_1 has the highest priority in the bid-based firm, and has an expected waiting time of 1. From Lemma 1, we know that the expected waiting time in an equilibrium is non-increasing in the unit waiting cost, implying that the expected waiting time of a customer in the fixed-price firm, $w_F(x)$, is less than or equal 1. However, since C_F is non-empty, the expected waiting time in the fixed-price firm has to be strictly greater than the service completion time 1. Thus we obtain a contradiction.

Summing up, we obtain that any symmetric equilibrium (x, φ) where at least some customers choose to obtain service from the fixed-price firm, there exists thresholds $0 < c_1 < c_2 < c_\ell \leq 1$, such that for all $c \in [0, c_1] \cup [c_2, c_\ell]$, we have $x(c) = \mathsf{BID}$, for all $c \in (c_1, c_2)$, we have $x(c) = \mathsf{FIX}$, and for all $c \in (c_\ell, 1]$, we have $x(c) = \mathsf{LEAVE}$. □

The preceding theorem states that in a symmetric equilibrium, a customer's decision about which firm to obtain service from has a simple multi-threshold structure: customers with very low and very high unit waiting cost choose to obtain service from the bid-based firm, and those with intermediate unit waiting cost prefer to obtain service via the fixed-price firm. This structure suggests that the bid-based firm serves two different functions in the system. For those customers with very high unit waiting cost, the bid-based firm provides means to obtain high priority and get service quickly. On the other hand, for customers with very low unit waiting cost, the bid-based firm allows them to obtain service at low costs, albeit after longer waiting times. We reiterate that the multi-threshold structure arises mainly due to the modeling of congestion at the fixed-price firm: if the fixed-price firm has infinite capacity and the fixed-price is not exorbitant, then no customer with very high unit waiting cost would obtain service from the bid-based firm.



(a) Expected payment



(b) Expected waiting time

Figure 1: Expected payment and waiting time in equilibrium of the customers' game

In Fig. 1a and Fig. 1b we illustrate the equilibrium strategy, the expected waiting time in equilibrium and the equilibrium payment for the following parameter values: $\lambda = 40$, $P = 3$, and V is sufficiently high that every customer obtains service in equilibrium. The bid-based firm is an $M/M/20$ preemptive priority queue, and the fixed-price firm is an $M/M/30$ FIFO queue. From these plots, we observe the threshold structure of equilibrium, and the monotonicity of the expected waiting time and the payment with respect to unit time waiting cost. Note also that the low unit waiting cost customers in the bid-based firm pay less than the price in the fixed-price firm, and wait longer for service completion, while the high unit waiting cost customers in the bid-based firm pay more than the price in the fixed-price firm, and have lower waiting times.

In Figure 2, where we plot equilibrium thresholds c_1, c_2, c_ℓ and $\alpha = c_2 - c_1$ against the fixed-price P, when the fixed-price firm is an $M/M/49$ FIFO queue and the bid-based firm is an $M/M/51$ priority queue. We let arrival rate $\lambda = 95$ and $V = 4$. Note that, as long as $c_\ell = 1$, the thresholds c_1, c_2 are increasing in P and α is decreasing in P. However, this relationship may not hold if c_ℓ decreases below 1. Also, as P increases, the three thresholds c_1, c_2 and c_ℓ converge to the same value, matching with the intuition that as P becomes too large, no customer can afford to join the fixed-price firm. (In Section 4.1, we analytically derive an expression for the value of P beyond which no customer joins the fixed-price firm in equilibrium.)

From the perspective of equilibrium analysis, the preceding theorem is important. From Lemma 2, in any symmetric equilibrium (x, φ), the customers' bidding function $\varphi(\cdot)$ is fully specified from

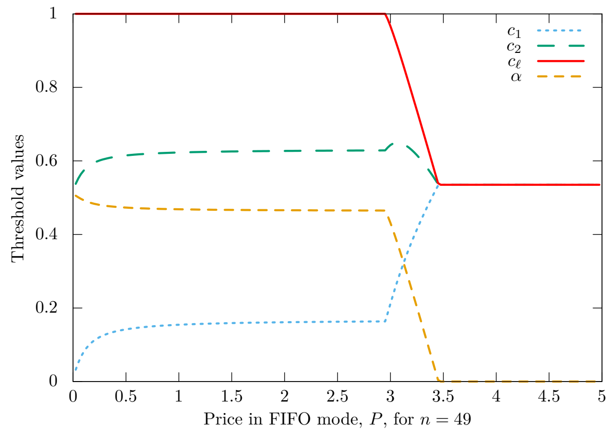


Figure 2: Equilibrium thresholds

the service decision $x(\cdot)$. Combining this result with the preceding theorem, we obtain that any symmetric equilibrium (x, φ) can be characterized by three thresholds $c_1, c_2, c_\ell \in [0, 1]$. Thus from this point on, we only need to consider symmetric strategies characterized by c_1, c_2, c_ℓ. This greatly simplifies the analysis for showing existence of the equilibrium, as we see in the following section.

4 Existence and uniqueness of equilibrium in customers' game

Having determined the structure of a symmetric equilibrium of customers' game, we are now ready to present our main results regarding the existence and uniqueness of customers' symmetric equilibrium. We present our results in two steps: first, in Section 4.1, we consider a system where obtaining service is mandatory for all customers, denoted by $\mathsf{SYS}_{\mathsf{man}}$. Although this restriction is impracticable, we use the results for this system to show existence (and uniqueness) in the original system where customers have the option to leave the system without obtaining service, which we denote by $\mathsf{SYS}_{\mathsf{op}}$. This is achieved by carefully relating an equilibrium of the $\mathsf{SYS}_{\mathsf{man}}$ system with a corresponding equilibrium of the $\mathsf{SYS}_{\mathsf{op}}$ system. We provide the details in Section 4.2. We then extend the results to systems with positive reserve price in Section 4.3.

To state our results, we need two model primitives. Let $\Gamma(x)$ denote the expected waiting time of a customer in the bid-based firm, if the arrival rate of customers to that firm with higher priority is equal to $x \in [0, k)$. Similarly, let $\Phi(x)$ denote the expected waiting time of a customer in the fixed-price firm, if the arrival rate of the customers to that firm is $x \in [0, n)$. We make the following assumptions on $\Gamma(x)$ and $\Phi(x)$.

Assumption 1. *1. The function* Γ *is finite, strictly increasing and continuous in* $[0, k)$ *with* $\lim_{x \to k} \Gamma(x) = \infty$*, and* $\Gamma(x) = \infty$ *for* $x > k$*.*

2. *The function* Φ *is finite, strictly increasing and continuous in* $[0, n)$ *with* $\lim_{x \to n} \Phi(x) = \infty$*, and* $\Phi(x) = \infty$ *for* $x > n$*.*

3. $\Gamma(0) = \Phi(0) = 1$*.*

4. $\int_0^y \Gamma(t)dt \to \infty,$ *as* $y \to k$*.*

We briefly note that in the preceding assumption, we allow for n and k to be infinite, implying that the corresponding firm has (potentially) infinite capacity. However, the first two conditions require that as the arrival rate of customers to a firm increases, the waiting time at that firm strictly increases, leading to a deteriorating quality of service. Thus, these conditions imply congestion effects at the two firms. In particular, the preceding assumption precludes settings such as an $M/M/\infty$ queue, where the service quality is constant and independent of the arrival rates.

The following lemma shows that Assumption 1 holds under many commonly studied queueing systems. The proof is given in Appendix C.

Lemma 3. *Suppose, for* $n, k < \infty$*, the fixed-price firm operates a* $G/G/n$ *queue with service rate* 1*, and the bid-based firm operates a* $G/M/k$ *preemptive queue with service rate* 1*. Then the system satisfies Assumption 1.*

The preceding lemma uses the fact that the distribution of the waiting cost is continuously differentiable, and thus has no point mass. In particular, part 3 of Assumption 1 may not hold if there is a positive mass of customer types with the highest priority in the bid-based firm. Also, we have implicitly used the fact that the queue in the bid-based firm is preemptive. Without this assumption, the expected waiting time of a customer will be a function not only of the arrival rate of the customers with higher priority, but also of the total arrival rate to the queue. However all our analysis in Sections 3 and 4 can be reproduced for non-preemptive queues. Please refer to Appendix G for details.

Following Theorem 1, we represent a customer's strategy by a vector of thresholds $\bar{c} = (c_1, c_2, c_\ell)$. Assuming all customers follow the strategy $\bar{c}$, let $w_F(\bar{c})$ be the expected waiting time in the fixed-price firm. Similarly, let $w_B(c|\bar{c})$ be the expected waiting time of a customer of unit waiting cost c if she chooses to obtain service from the bid-based firm and subsequently makes the optimal bid from (1). Letting $\alpha = c_2 - c_1$ for simplicity of notation, we have the following expressions:

$$w_F(\bar{c}) = \Phi(\lambda\alpha), \tag{3}$$

$$w_B(t|\bar{c}) = \begin{cases} \Gamma(\lambda c_\ell - \lambda t) & \text{if } t \in [c_2, c_\ell]; \\ \Gamma(\lambda c_\ell - \lambda\alpha - \lambda t) & \text{if } t \in [0, c_1]. \end{cases} \tag{4}$$

The first equation follows from the fact that, under strategy $\bar{c}$, the arrival rate of customers to the fixed-price firm equals $\lambda(c_2 - c_1) = \lambda\alpha$. The second equation follows from the fact that for $t \in [c_2, c_\ell]$, the arrival rate of customers to the bid-based firm with higher priority than the customer with unit waiting cost t is equal to $\lambda(c_\ell - t)$, whereas it equals $\lambda(c_\ell - \alpha - t)$ for $t \in [0, c_1]$, since all customers with unit waiting cost in $[c_1, c_2]$ obtain service from the fixed-price firm. (Note that these expressions make use of our assumption that the unit waiting costs are uniformly distributed. We briefly describe in Appendix A how these expressions differ for non-uniform distributions.)

4.1 System with mandatory service requirement

We first consider the setting where customers do not have the option to leave the system without obtaining service. In other words, we assume that the customers' value for service completion V is sufficiently high to render obtaining service mandatory for all customers. Denote such a system by $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ if the arrival rate is λ and the fixed-price price is P. In such a setting, customers arrive to the system and make the choice between obtaining service from the fixed-price firm or the bid-based firm, and a symmetric equilibrium for this system can be represented as $\bar{c} = (c_1, c_2, 1)$. In the following, we show that for all $\lambda \in (0, n+k)$ and for all $P \geq 0$, there exists a unique symmetric equilibrium. Towards that goal, define $P_{\max}(\lambda)$ as follows:

$$P_{\max}(\lambda) \triangleq \begin{cases} \frac{1}{\lambda}\int_0^\lambda \Gamma(t)dt - 1 & \text{if } \lambda < k; \\ \infty & \text{otherwise.} \end{cases} \tag{5}$$

For each arrival rate λ, $P_{\max}(\lambda)$ acts as a bound on the price P of service at the fixed-price firm. This is because, as we show below, for $P \geq P_{\max}(\lambda)$, no customer obtains service from the fixed-price firm in a symmetric equilibrium of the customers' game. Thus, for large values of k, i.e., when the bid-based firm has ample capacity and can offer low waiting times, we have $P_{\max}(\lambda) < \infty$, implying the fixed-price firm's choice of the price P is constrained. On the other hand, for large enough arrival rate λ, i.e, when there is sufficient demand, we have $P_{\max}(\lambda) = \infty$, implying the fixed-price firm's choice of the price P is unconstrained.

Suppose the strategy $\bar{c} = (1, 1, 1)$ constitutes a symmetric equilibrium, where no customers choose to obtain service from the fixed-price firm. In such an equilibrium, if the customer with unit waiting cost equal to one obtains service from the fixed-price firm, her expected waiting time to service completion is 1 and her expected payment is P. Since in equilibrium such a customer prefers obtaining service from the bid-based firm, it must be the case that

$$\int_0^1 w_B(t|\bar{c})\ dt \leq 1 + P. \tag{6}$$

Here, the left hand side denotes the total expected cost of the customer with unit waiting cost c_ℓ, as per (2). The necessary condition (6) requires that this be less than that of obtaining service from the fixed-price firm. Note that, using (4), it is straightforward to show that this implies $P \geq P_{\max}(\lambda)$. Thus, for $\bar{c} = (1, 1, 1)$ to be a symmetric equilibrium for $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, a necessary condition is $P \geq P_{\max}(\lambda)$. The following theorem shows that it is also sufficient. The proof of this theorem is given in Appendix D.

Theorem 2. *There exists a unique symmetric customers' equilibrium of the system* $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ *of the form* $\bar{c} = (1, 1, 1)$ *if and only if* $P \geq P_{\max}(\lambda)$.

The preceding theorem states that for any $\lambda \in (0, n + k)$, $P_{\max}(\lambda)$ is the highest price in the fixed-price firm for which one may expect customers to choose that firm for obtaining service. For values of P greater than $P_{\max}(\lambda)$, all customers prefer the bid-based firm over the fixed-price firm. The following theorem, our main result, shows that for values of P less than $P_{\max}(\lambda)$, there exists a unique symmetric equilibrium, in which there is a positive arrival rate of customers to the fixed-price firm.

Theorem 3. *For all* $0 < P < P_{\max}(\lambda)$*, there exists a unique symmetric customers' equilibrium* $\bar{c} = (c_1, c_2, 1)$*, with* $0 < c_1 < c_2 < 1$*. In other words, in equilibrium, the arrival rate of customers to* *the fixed-price firm is positive.*

Our proof proceeds by first identifying a set of necessary and sufficient conditions for a strategy $\bar{c} = (c_1, c_2, 1)$ to be a symmetric equilibrium for $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, and second showing that these conditions have a unique solution by explicit construction. In the following, we provide the intuition behind these necessary and sufficient conditions.

Suppose $0 < P < P_{\max}(\lambda)$, and consider a symmetric equilibrium $\bar{c} = (c_1, c_2, 1)$ with $0 < c_1 < c_2 < 1$, where the arrival rate of customers to the fixed-price firm is positive. In this equilibrium, a customer with unit waiting cost c_1 must be indifferent between obtaining service in the bid-based firm and the fixed-price firm. For otherwise, a customer with unit waiting costs slightly higher than c_1 would strictly prefer to mimic the behavior customer with unit waiting cost c_1. Equating the customer's total expected cost, we have

$$\int_0^{c_1} w_B(t|\bar{c})dt = c_1 w_F(\bar{c}) + P. \tag{7}$$

Note that since all customers with unit waiting cost below c_1 obtain service from the bid-based firm, we have $w(t|\bar{c}) = w_B(t|\bar{c})$ and consequently, as per (2), the left hand side denotes the total expected cost of a customer with unit waiting cost c_1. On the other hand, the right hand side denotes the total expected cost for such a customer on obtaining service from the fixed-price firm.

Similarly, consider a customer with unit waiting cost c_2. Since all customers with unit waiting costs between c_1 and c_2 obtain service from the fixed-price firm, the expected waiting time of the customer with unit waiting cost c_2 must equal that of a customer with unit waiting cost c_1. Since from Lemma 1, we know that the expected waiting time in equilibrium is non-increasing in the unit waiting cost, and since all customers in the fixed-price firm have waiting times $w_F(\bar{c})$, it follows that

$$w_B(c_2|\bar{c}) = w_B(c_1|\bar{c}) = w_F(\bar{c}). \tag{8}$$

The necessary condition (8) thus states that the expected waiting time in equilibrium must be continuous at c_1.

The following proposition formalizes the preceding discussion, and shows that the aforementioned necessary conditions are also sufficient for a strategy to be an equilibrium. We provide the proof in

Appendix C.

Proposition 1. *A strategy* $\bar{c} = (c_1, c_2, 1)$*, with* $0 < c_1 < c_2 < 1$*, is a symmetric customers' equilibrium for the system* $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ *with* $0 < P < P_{\max}(\lambda)$ *if and only if the conditions* (7) *and* (8) *hold.*

Using (3) and (4), we can summarize the conditions (7) and (8) as follows:

$$\int_0^{c_1} \Gamma(\lambda(1-\alpha-t))dt = c_1\Phi(\lambda\alpha) + P$$
$$\Gamma(\lambda(1-\alpha-c_1)) = \Phi(\lambda\alpha). \tag{9}$$

Thus, following Proposition 1, showing the existence (and uniqueness) of a symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ with $P \in (0, P_{\max}(\lambda))$ requires showing that there exists a (unique) $c_1 \in [0, 1)$ and $\alpha \in (0, 1 - c_1]$ that satisfy the set of equations (9). We show that this is indeed the case in the proof of Theorem 3. We provide the details of the proof in Appendix D.

4.2 System with optional service requirement

In this section, we extend our existence and uniqueness result to systems where customers may choose not to obtain service. We denote the system with optional service requirement with arrival rate λ, fixed price P, and the value of service completion V by $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$.

For this system, consider a symmetric equilibrium $\bar{c} = (c_1, c_2, c_\ell)$. In equilibrium, each customer choosing to obtain service must have a total expected cost that is less than or equal to the value of service completion V. In particular, this holds for a customer with unit waiting cost c_ℓ. Moreover, if c_ℓ is strictly less than one, the total expected cost of such a customer must exactly equal to V. For if this were not true, a customer with unit waiting cost slightly greater than c_ℓ would find it preferable to obtain service from the system. Thus, we obtain the following necessary condition on an equilibrium:

$$c_\ell < 1, \quad \int_0^{c_\ell} w(t|\bar{c})dt = V, \qquad \text{OR} \qquad c_\ell = 1, \quad \int_0^{c_\ell} w(t|\bar{c})dt \leq V. \tag{10}$$

Here, as per (2), the integral denotes the total expected cost of a customer with unit waiting cost c_ℓ. Note that we have $w(t|\bar{c}) = w_F(\bar{c})$ for $t \in (c_1, c_2)$ and $w(t|\bar{c}) = w_B(t|\bar{c})$ for $t \in [0, c_1] \cup [c_2, c_\ell]$. Using

(10), we can now relate the symmetric equilibria of the system $\mathsf{SYS_{op}}$ with those of the system $\mathsf{SYS_{man}}$. The following lemmas formalize this argument. The proof is provided in Appendix C.

Lemma 4. *If the strategy $(c_1, c_2, 1)$ is a symmetric equilibrium for the system $\mathsf{SYS_{man}}(\lambda, P)$ with $\lambda \in (0, n+k)$, then the strategy $\bar{c}(u) = (c_1 u, c_2 u, u)$ for $u \in (0, 1]$ is a symmetric equilibrium for the system $\mathsf{SYS_{op}}(\frac{\lambda}{u}, Pu, V)$ if and only if the condition* (10) *holds for the strategy $\bar{c}(u)$.*

The preceding lemma states that a symmetric equilibrium of the system $\mathsf{SYS_{man}}$ can be used to construct a symmetric equilibrium of a related $\mathsf{SYS_{op}}$ system, as long as one can ensure that the condition (10) is satisfied. Conversely, the following lemma constructs a symmetric equilibrium for a $\mathsf{SYS_{man}}$ system using the symmetric equilibrium of a related $\mathsf{SYS_{op}}$ system.

Lemma 5. *If the strategy (c_1, c_2, u) is a symmetric equilibrium of the system $\mathsf{SYS_{op}}(\lambda, P, V)$ then the strategy $(\frac{c_1}{u}, \frac{c_2}{u}, 1)$ is a symmetric equilibrium of the system $\mathsf{SYS_{man}}\left(\lambda u, \frac{P}{u}\right)$.*

Let $U_\lambda = (0, 1] \cap (0, \frac{n+k}{\lambda})$. The preceding lemmas, together with Theorem 3, imply that there exist functions $\mathcal{C}_i : U_\lambda \to [0, 1]$ for $i = 1, 2$, such that for each $u \in U_\lambda$, we have $\mathcal{C}_1(u) \leq \mathcal{C}_2(u) \leq u$, and the strategy $\left(\frac{\mathcal{C}_1(u)}{u}, \frac{\mathcal{C}_2(u)}{u}, 1\right)$ is the unique symmetric equilibrium of the system $\mathsf{SYS_{man}}(\lambda u, \frac{P}{u})$. Using this, we now state the existence (and uniqueness) result for the system $\mathsf{SYS_{op}}(\lambda, P, V)$.

Theorem 4. *For each $\lambda > 0$, $P > 0$ and $V > 0$, there exists $u = u(\lambda, P, V) \in U_\lambda$ such that the strategy $(\mathcal{C}_1(u), \mathcal{C}_2(u), u)$ constitutes the unique symmetric customers' equilibrium for the system $\mathsf{SYS_{op}}(\lambda, P, V)$. Further, for each $\lambda > 0$ and $V > 0$, there exists a threshold $P(\lambda, V)$ such that for all $P \geq P(\lambda, V)$, in the symmetric equilibrium, the arrival rate of customers to the fixed-price firm is zero, whereas for all $P \in (0, P(\lambda, V))$, the arrival rate of customers to the fixed-price firm is positive.*

To see the intuition behind the result, observe that, by using the expressions (4) and (3), we can write the expected waiting time function for the strategy as $\bar{\mathcal{C}}(u) = (\mathcal{C}_1(u), \mathcal{C}_2(u), u)$,

$$w(t|\bar{\mathcal{C}}(u)) = \begin{cases} \Gamma(\lambda u - \lambda t) & \text{if } t \in [\mathcal{C}_2(u), u]; \\ \Phi(\lambda \mathcal{A}(u)) & \text{if } t \in (\mathcal{C}_1(u), \mathcal{C}_2(u)); \\ \Gamma(\lambda u - \lambda \mathcal{A}(u) - \lambda t) & \text{if } t \in [0, \mathcal{C}_1(u)], \end{cases}$$

where $\mathcal{A}(u) = \mathcal{C}_2(u) - \mathcal{C}_1(u)$. By straightforward algebra, the condition (10) then reduces to

$$\int_0^{u-\mathcal{A}(u)} \Gamma(\lambda t)dt + \mathcal{A}(u)\Phi(\lambda\mathcal{A}(u)) \leq V, \quad \text{if } u < 1,$$
$$\int_0^{u-\mathcal{A}(u)} \Gamma(\lambda t)dt + \mathcal{A}(u)\Phi(\lambda\mathcal{A}(u)) = V, \quad \text{if } u = 1. \tag{11}$$

Thus, showing the existence (and uniqueness) of a symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$ requires showing that the preceding equation has a unique solution $u = u(\lambda, P, V)$. This is obtained by showing that both $\mathcal{A}(u)$ and $u - \mathcal{A}(u)$ are continuous and non-decreasing in $u \in U_\lambda$. We provide the details in Appendix E.

4.3 System with positive reserve price

The existence and uniqueness of customers' equilibrium can be easily extended to the setting where reserve price in the bid-based firm r is positive. When $P > r$, there is an straightforward one-to-one mapping between the systems with fixed price P, reserve price r, value V, and the system with fixed price $P - r$, reserve price 0, value $V - r$. In particular, the two systems have the same equilibrium thresholds, and any customer who receives service in equilibrium pays an additional r in the former system over her payment in the latter system. To see this, note that in the former system, all customers who receive service pays at least r, thus subtracting r from both the prices and the value does not change the customers' service decision. As a result, we can obtain the service decisions in the former system with reserve price, by calculating that in the latter system, which has zero reserve price. Customers' bids in the former system is then given by r plus their bids in the latter system.

When $P \leq r$, the system cannot be mapped directly to any system that we have studied in the earlier parts of the paper. However, we can perform a similar analysis to show that there exists a unique customers' equilibrium $\bar{c} = (c_1, c_2, c_\ell)$. In this equilibrium, we have $c_1 = 0$, whereas c_2 and c_ℓ satisfy a condition analogous to (10), as well as a condition which implies that if $c_2 < c_\ell$, a customer with unit waiting cost c_2 is indifferent between the two firms, and if $c_2 = c_\ell$, a customer with unit waiting cost c_2 prefers fixed-price firm over bid-based firm. Namely, we have the following two conditions:

$$c_\ell < 1, \quad P + \int_0^{c_\ell} w(t|\bar{c})dt = V, \qquad \text{OR} \qquad c_\ell = 1, \quad P + \int_0^{c_\ell} w(t|\bar{c})dt \leq V,$$

and

$$c_2 < c_\ell, \quad P + c_2 w_F(\bar{c}) = r + c_2 w_B(c_2|\bar{c}), \qquad \text{OR} \qquad c_2 = c_\ell, \quad P + c_2 w_F(\bar{c}) \leq r + c_2 w_B(c_2|\bar{c}).$$

The following theorem summarizes the discussions above.

Theorem 5. *For each $\lambda > 0$, $P > 0$, $r > 0$ and $V > 0$, there exists a strategy characterized by thresholds $\bar{c} = (c_1, c_2, c_\ell)$ that constitutes the unique symmetric customers' equilibrium for the system with optional service requirements. In particular, when $r \geq P$, we have $c_1 = 0$ and the equilibrium strategy is a two-thresholds strategy.*

5 Price and capacity competition

Having characterized the unique equilibrium among the customers for any given fixed price P and reserve price r, we are now ready to study the competition between the two firms. We are interested in the setting where each firm seeks to maximize its expected revenue, given the other firms' choices. We start by analyzing the price competition between the two firms, where the fixed-price firm sets its fixed-price P and the bid-based firm sets its reserve price simultaneously. We later visit the capacity decisions of the two firms in Section 5.3.

Without loss of generality, we assume that the fixed-price firm chooses $P \in [0, V]$ and the bid-based firm chooses $r \in [0, V]$, since otherwise no customer would obtain service from the respective firm. Furthermore, we assume that $V \geq 1$. This is because, if $V < 1$, some customers would never choose to obtain service from either firm, even if they were to be served immediately upon arrival, and such customers can be ignored for the sake of analysis. This resulting system with $V < 1$ can then be mapped to a system with $V \geq 1$ and lower λ, using techniques as in the proof of Lemma 4.

The following result follows from the continuity of the equilibrium thresholds (and hence the expected revenue of the two firms) in the prices P and r, and the fact that the prices take values in a compact set. The details are given in Appendix F.

Theorem 6. *For $V > 1$, the game between the two firms has a mixed Nash equilibrium.*

The preceding result does not allow us to conclude the existence of a pure Nash equilibrium; for this, we need the expected revenues of the two firms to be concave in their respective prices. However, since these revenues are determined as the outcome of an equilibrium between the customers, concavity properties do not arise in general. Given the implausibility of mixed equilibrium in practice, and to obtain meaningful insight into the competition, we analyze the system under specific assumptions on the expected waiting times at the two firms (discussed in Section 5.1), and under a particular parameter regime. Specifically, in Section 5.2, we study the system in the limit where the arrival rate and the service capacities both increase proportionally, and characterize the set of *pure* equilibria in the limiting game. Using standard arguments, we show in Theorem 7 that these pure equilibria are approximate equilibria for the system with finite, but large enough, arrival rate. We then use this characterization of the pure equilibria in the limiting game to study the capacity competition between the two firms in Section 5.3.

5.1 Waiting time expressions

To obtain analytical expressions for the equilibrium thresholds, we focus on a setting with explicit expected waiting time expressions $w_F(\cdot)$ and $w_B(\cdot|\bar{c})$, motivated by $M/M/1$ FIFO and (preemptive) priority queues. In particular, suppose the strategy adopted by the customers is given by $\bar{c} = (c_1, c_2, c_\ell)$. Let $\rho_F = \lambda(c_2 - c_1)/n$ denote the traffic intensity (see Gautam (2012)) at the fixed-price firm, and let $\rho_B = \lambda(c_\ell - c_2 + c_1)/k$ denote the total traffic intensity (across all priorities) at the bid-based firm.

We assume that the expected waiting time at the fixed-price firm is given explicitly by

$$w_F(\bar{c}) = \frac{\rho_F}{1 - \rho_F}\frac{1}{n} + 1.$$

Essentially, making this assumption implies that when a customer is waiting for service in the fixed-price firm, the service requirement ahead of her decreases at rate n. The first term then denotes the waiting time until the beginning of service in an $M/M/1$ FIFO queue with service rate n and arrival rate $\lambda(c_2 - c_1)$. Once the customer starts her service, her service is processed at rate 1, and incurs an additional waiting time of one, as in the second term. As an example of an instance where this waiting time expression holds, consider a airport taxi service, where customers are riders waiting for taxi, each of which arrive at rate n. Thus, the time waiting for the arrival of a taxi is given by the first term in the expression. Once a taxi arrives, the time it takes to reach the destination is fixed and independent of the service rate n.

Similarly, for the waiting time $w_B(\cdot|\bar{c})$, we again make the assumption that when a customer is waiting for service in the bid-based firm, the service requirement ahead of her decreases at rate k. With this assumption, the expected waiting time of a customer with unit waiting cost c in the bid-based firm is given by

$$w_B(c|\bar{c}) = \frac{1}{k(1 - \rho_B + \rho_B B(c;\bar{c}))^2} - \frac{1}{k} + 1.$$

Here $B(c;\bar{c})$ is the fraction of customers that join the bid-based firm with unit cost less than c, defined as

$$B(x;\bar{c}) = \begin{cases} \frac{x}{c_\ell - c_2 + c_1} & \text{if } x \leq c_1; \\ \frac{x + c_1 - c_2}{c_\ell - c_2 + c_1} & \text{if } x \geq c_2. \end{cases}$$

The first two terms in the waiting time expression denote the waiting time until the beginning of service in a $M/M/1$ priority queue with service rate k and total arrival rate $\lambda(c_\ell - c_2 + c_1)$, where we have used the waiting time expression for priority queues as in Kleinrock (1967). The last term then denotes the time in service.

5.2 High arrival rate and high capacity regime

Next, we introduce the limiting regime under which we perform our analysis. In particular, we consider the limiting regime where the arrival rate λ diverges to ∞, with the capacities at the two firms increasing proportionally as $n = q_F\lambda$ and $k = q_B\lambda$, for some $q_F > 0$ and $q_B > 0$. We refer to q_B and q_F as the capacity ratios.

To model the customer behavior in this regime, we first identify the limit of the thresholds in the unique customer equilibrium, as $\lambda \to \infty$. For fixed (P, r) and for any fixed $\lambda > 0$, let $(c_1^\lambda, c_2^\lambda, c_\ell^\lambda)$ denote the equilibrium thresholds, and let $\alpha^\lambda = c_2^\lambda - c_1^\lambda$. Let $R_F^\lambda(P, r)$ denote the expected revenue per arrival of the fixed-price firm in equilibrium, and let $R_B^\lambda(P, r)$ denote that of the bid-based firm. Substituting the waiting time expressions introduced in Section 5.1 in equations (7), (8) and (10), and letting λ approach infinity, we obtain expressions for the limiting thresholds and the expected revenue per arrival. (We denote these limits with superscript ∞ in place of λ.) For different fixed values of P and r, these expressions are summarized in the following lemma. The proof of the lemma is given in Appendix F.

Lemma 6. *Let $V \geq 1$. As λ approaches infinity, the limiting equilibrium thresholds are as follows:*

1. *If $r < P \leq V - \min\{q_B, 1\}$, we have $c_\ell^\infty(P, r) = \min\{V - P, 1, q_F + q_B\}$, $\alpha^\infty(P, r) = c_\ell^\infty(P, r) - \min\{q_B, 1\}$, and*

$$c_1^\infty(P, r) = \begin{cases} \sqrt{\dfrac{q_B}{\left(\frac{1}{q_F + q_B - c_\ell^\infty(P, r)} - \frac{1}{q_F} + \frac{1}{q_B}\right)}} & \text{if } q_B < 1; \\ 0 & \text{otherwise.} \end{cases}$$

2. *If $P \leq r \leq V - \min\{q_F, 1\}$, we have $c_\ell^\infty(P, r) = \min\{V - r, 1, q_F + q_B\}$, $c_1^\infty(P, r) = 0$, and $\alpha^\infty(P, r) = \min\{q_F, 1\}$.*

3. *If $P > r$ and $P > V - \min\{q_B, 1\}$, then $c_1^\infty(P, r) = c_2^\infty(P, r) = c_\ell^\infty(P, r) = \min\{V - r, q_B, 1\}$.*

4. *If $r \geq P$ and $r > V - \min\{q_F, 1\}$, we have $c_1^\infty(P, r) = 0$, whereas $c_2^\infty(P, r) = c_\ell^\infty(P, r) = \min\{V - P, q_F, 1\}$.*

Furthermore, the limiting expected revenue per arrival of the two firms satisfy:

$$R_F^\infty(P,r) = \begin{cases} \left(c_\ell^\infty(P,r) - \min\{q_B,1\}\right)P, & \textit{if } r < P \le V - \min\{q_B,1\}; \\ \min\{V-P, q_F, 1\}P, & \textit{if } P \le r; \\ 0, & \textit{if } P > r \textit{ and } P > V - \min\{q_B,1\}, \end{cases}$$

$$R_B^\infty(P,r) = \begin{cases} \min\{q_B,1\}P, & \textit{if } r < P \le V - \min\{q_B,1\}; \\ \left(c_\ell^\infty(P,r) - \min\{q_F,1\}\right)r, & \textit{if } P \le r \le V - \min\{q_F,1\}; \\ \min\{V-r, q_B, 1\}r, & \textit{if } P > r \textit{ and } P > V - \min\{q_B,1\}; \\ 0, & \textit{if } r \ge P \textit{ and } r > V - \min\{q_F,1\}. \end{cases}$$

The preceding lemma suggests that the limiting equilibrium has one of four possible structures, depending on the values of P and r. For values of P and r not too high, as in the first two cases of the lemma, if the capacities q_B and q_F of the firms are small, the demand is split between the two firms and we have a duopoly. In particular, if $q_F + q_B < 1$, then, for $r < P$, the bid-based firm operates at its capacity (with demand equal to λq_B) and the remaining customers go to the fixed-price firm. On the other hand, for $P \le r$, the fixed-price firm operates at capacity, with demand equal to λq_F, and the remaining customers go to the bid-based firm. However, when one of the two firms sets a sufficiently high price (fixed-price firm in the third case, and bid-based firm in the fourth case), then the demand is fully captured by the other firm, and the limiting equilibrium is a monopoly.

From Lemma 6, we obtain that the firms' limiting expected revenue per arrival has a simple closed form. Using these expressions, we study a "limiting game", where, given the capacity ratios q_B and q_F, the fixed-price and bid-based firms choose P and r respectively and simultaneously to maximize their limiting revenues per arrival $R_F^\infty(P,r)$ and $R_B^\infty(P,r)$. We show that this game has *pure* Nash equilibria, which can be explicity identified. As a justification for studying these equilibria, we present the following theorem, which states that any Nash equilibrium of the limiting game is an approximate equilibrium of the finite game (with appropriate capacities). More formally, for $\epsilon > 0$, we say a price profile (P,r) is an ϵ-equilibrium, if no firm can unilaterally deviate and increase its payoff by more than ϵ. We have the following theorem, which follows from the continuity of the equilibrium thresholds in λ. The details are given in Appendix F.

Theorem 7. *For every $\epsilon > 0$, there exists a $\lambda_0(\epsilon)$, such that for any $\lambda > \lambda_0(\epsilon)$, every equilibrium (P, r) of the limiting game with capacity ratios q_B and q_F is an ϵ-equilibrium of the finite game with arrival rate λ, and capacities $n = q_F\lambda$ and $k = q_B\lambda$.*

With this theorem in place, we now characterize the pure Nash equilibria of the limiting game. For any (q_B, q_F) with q_B and q_F both positive, let $\mathcal{S}(q_B, q_F)$ denote the set of pure Nash equilibria of the limiting game. We have the following lemma, whose proof is provided in Appendix F.

Lemma 7. *Suppose $q_F, q_B > 0$ and $V \geq 1$. Then, the pure Nash equilibria of the limiting game are as follows:*

1. *If $V \geq 2 - q_B$ and $q_F + q_B > 1$, then*

$$\mathcal{S}(q_B, q_F) = \left\{(P, r) : P = V - 1, \quad r \leq \frac{1 - \min\{q_B, 1\}}{\min\{q_F, 1\}}(V - 1)\right\}.$$

For all $(P, r) \in \mathcal{S}(q_B, q_F)$, we have $R_B^\infty(P, r) = \min\{q_B, 1\}(V - 1)$ and $R_F^\infty(P, r) = (1 - \min\{q_B, 1\})(V - 1)$.

2. *If $V < 2 - q_B$ and $q_F + q_B > 1$, then*

$$\mathcal{S}(q_B, q_F) = \left\{(P, r) : P = \frac{V - q_B}{2}, \quad r \leq V - 1\right\}.$$

For all $(P, r) \in \mathcal{S}(q_B, q_F)$, we have $R_B^\infty(P, r) = q_B\frac{V - q_B}{2}$ and $R_F^\infty(P, r) = \left(\frac{V - q_B}{2}\right)^2$.

3. *If $V \geq 2q_F + q_B$ and $q_F + q_B \leq 1$, then*

$$\mathcal{S}(q_B, q_F) = \{(P, r) : P = V - q_B - q_F, \quad r < P\}.$$

For all $(P, r) \in \mathcal{S}(q_B, q_F)$, we have $R_B^\infty(P, r) = q_B(V - q_B - q_F)$ and $R_F^\infty(P, r) = q_F(V - q_B - q_F)$.

4. *If $V < 2q_F + q_B$ and $q_F + q_B \leq 1$, then*

$$\mathcal{S}(q_B, q_F) = \left\{(P, r) : P = \frac{V - q_B}{2}, \quad r < V - q_B - q_F\right\}.$$

For all $(P, r) \in \mathcal{S}(q_B, q_F)$*, we have* $R_B^\infty(P, r) = q_B \frac{V - q_B}{2}$ *and* $R_F^\infty(P, r) = \left(\frac{V - q_B}{2}\right)^2$.

We observe that the limiting game has multiple equilbria; however, each firm's limiting revenue per arrival is the same across all equilibria. This is explained by observing that for every equilibrium among the firms of the limiting game, the customer equilibrium falls in the first case of Lemma 6. That is, the condition $r < P \leq V - \min\{q_B, 1\}$ holds in each equilibrium, implying that the bid-based firm is operating at capacity (or serving the entire demand), and receiving price P per customer served in expectation. In particular, the fixed-price firm serves the "leftover" demand. Under the condition $r < P \leq V - \min\{q_B, 1\}$, it is straightforward to verify that the limiting thresholds and the firms' revenues are functions of just the fixed-price P, and not the bid-based firm's reserve price r. Since for each case of Lemma 7, the set of pure Nash equilibria share the same value of P, we obtain that the firms' revenue per arrival is the same across all equilibria. Given this result, we abuse notation slightly, and let $R_F^\infty(q_B, q_F)$ and $R_B^\infty(q_B, q_F)$ denote the firms' expected revenue per arrival for any $(P, r) \in \mathcal{S}(q_B, q_F)$.

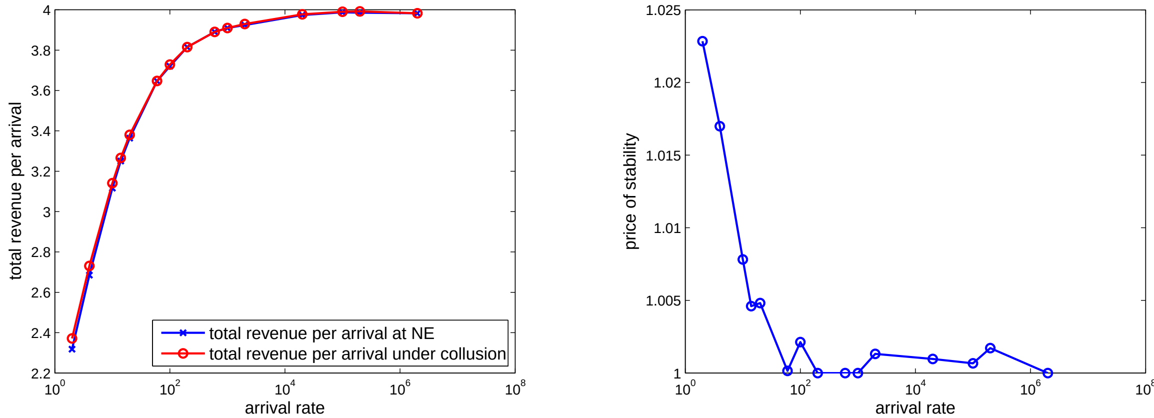
The preceding lemma also leads to an interesting implication: as long as $q_F > 0$, setting reserve price equal to zero is always a best response for the bid-based firm. In the presence of the fixed-price firm, the equilibrium is completely driven by the fixed price P, and the bid-based firm cannot improve its revenue by setting positive reserve price. This is in sharp contrast to the monopoly setting[1] where the fixed-price firm does not exist (i.e., $q_F = 0$), in which case, the bid-based firm can improve its revenue by setting a positive reserve price.

We conclude this discussion of the price competition between the firms studying the *price of stability* (PoS) of the limiting game. The price of stability is defined as the ratio of the maximum revenue achievable under collusion (i.e., when the two service modes are operated by a common firm which sets P and r to maximize its revenue) to the total revenue of the two firms in equilibrium (see Nisan et al. (2007)) . We have the following result, whose proof is provided in Appendix F:

Theorem 8. *For* $V \geq 2$*, we have* $\mathsf{PoS} = 1$*. For* $1 \leq V < 2$*,* PoS *is strictly decreasing in* V*, with* $\mathsf{PoS} \leq \frac{9}{8}$ *if* $1.5 \leq V < 2$ *and* $\mathsf{PoS} \leq \frac{1}{1 - \min\{q_B, 1\}^2}$ *if* $1 \leq V < 1.5$*.*

Thus, for small q_B or large V, the price of stability is relatively low. Moreover, for sufficiently large V, the total payoff under Nash equilibria is the same as the total payoff under the collusion.

[1] See Lemma 15 in Appendix F for optimal reserve price when the bid-based firm is a monopolist.



(a) Total revenue per arrival (b) Ratio between the total revenues per arrival

Figure 3: Total revenue per arrival in Nash equilibrium and under collusion and their ratio, when $V = 5$

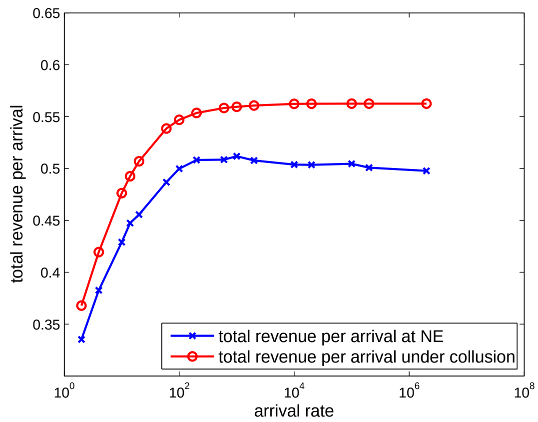
This suggests that in the limiting regime, the competition between the two firms does not significantly affect the total revenue, but only affects how it is shared between the two firms.

We illustrate our discussion above through two numerical examples. Figure 3a shows the total revenue per arrival in a Nash equilibrium and under collusion, as k, n and λ increase proportionally, assuming $q_B = q_F = 0.5$, and $V = 5$. Figure 3b shows the price of stability[2] for the same set of parameters. Observe that the ratio is very close to 1 for large λ, which coincides with the first case of Theorem 8. Similarly, Figure 4 shows the revenues and price of stability as k, n and λ increase proportionally, assuming $q_B = q_F = 0.5$, and $V = 1.5$. The price of stability fluctuates around 9/8 for large λ, which coincides with the second case of Theorem 8.

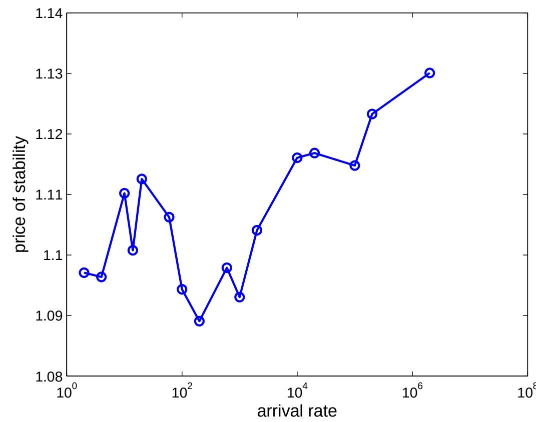
5.3 Capacity competition between firms

We end this section with a brief analysis of capacity decisions of the two firms. We model the capacity choices of the firms as long-run decisions, while the prices (fixed price P and the reserve price r), being readily adjustable, form short-run decisions. This is motivated by the fact that in practice firms must usually make staffing decisions in advance, e.g., for training purposes. Thus, in this model, capacity and pricing decisions form different layers of competition. As a consequence of this modeling assumption, once the capacity decisions are made, the prices must arise as equilibria

[2]Note that we do not have any monotonicity results for the price of stability as arrival rate increases. Thus we have zig zag patterns for price of stability.



(a) Total revenue per arrival



(b) Ratio between the total revenues per arrival

Figure 4: Total revenue per arrival in Nash equilibrium and under collusion and their ratio, when $V = 1.5$

of the limiting game as analyzed in Lemma 7. The revenue per arrival of a firm in this resulting equilibrium, minus its capacity costs, define the firm's profit, and each firm sets its capacities to maximize this profit. (Note that since all equilibria in the limiting game lead to the same revenue for the firms, the notion of a firm's profit is well-defined.)

We assume that the firms incur linear capacity costs. In particular, the fixed-price firm faces a cost of $\beta_F > 0$ per unit of capacity, whereas the bid-based firm's cost per unit of capacity is given by β_B. This assumption implies that per arrival, the capacity cost of the fixed-price firm equals $\beta_F q_F$, and that of the bid-based firm equals $\beta_B q_B$. Thus, for $q_B, q_F > 0$, the firms' profits are given by

$$\Pi_B(q_B, q_F) = R_B^\infty(q_B, q_F) - q_B\beta_B,$$
$$\Pi_F(q_B, q_F) = R_F^\infty(q_B, q_F) - q_F\beta_F.$$

The preceding expressions define the firms' profits when q_B and q_F are both positive. Since a firm may find it optimal to set zero capacity, (if capacity costs are too high), a complete description requires us to specify the firms' profits when either q_B or q_F is zero. These expressions can be obtained by a straightforward analysis, which we omit for reasons of brevity. (In particular, when q_F is zero, i.e, the fixed-price firm is not present, the bid-based firm will typically find setting positive reserve price better than zero reserve.)

Without loss of generality, we assume that $0 \le \beta_B < V$ and $0 \le \beta_F < V$. This is because V

is the maximum revenue that a firm can get from a customer, so a firm will go out of market if its capacity cost per arrival is greater than V. Furthermore, we restrict our attention to $V > 2$; from Theorem 8, we have that for $V > 2$, the price of stability in the price competition is one, thus allowing us to disentangle the effect of price and capacity competition on the equilibrium outcome. For these values of the capacity costs, the following theorem characterizes the equilibrium market structure.

Theorem 9. *Let $V > 2$ and $\beta_B, \beta_F \leq V$. Then, the equilibrium in the capacity competition is as follows:*

1. *There exists an equilibrium in which both firms operate in the market if and only if $|\beta_B - \beta_F| \leq 1$ and $\max\{2\beta_F - \beta_B, 2\beta_B - \beta_F\} < V \leq 2\beta_B - \beta_F + \frac{9}{2} - \frac{3}{2}\sqrt{5 + 4(\beta_B - \beta_F)}$. In such duopoly equilibrium, the capacities are $q_B^d = (V - 2\beta_B + \beta_F)/3$ and $q_F^d = (V - 2\beta_F + \beta_B)/3$. The set of equilibria of the resulting pricing game is $\{(P^d, r^d) : P^d = (V + \beta_B + \beta_F)/3, r^d < P^d\}$. The equilibrium payoffs are given by $\Pi_B(q_B^d, q_F^d) = (V - 2\beta_B + \beta_F)^2/9$ and $\Pi_F(q_B^d, q_F^d) = (V - 2\beta_F + \beta_B)^2/9$.*

2. *There exists an equilibrium in which the bid-based firm is a monopolist in the market, if and only if one of the following two conditions is satisfied.*

   (a) *The parameters satisfy $V \leq \min\{2\beta_F - \beta_B, \beta_B + 2, \beta_F + 1\}$. In the corresponding monopoly equilibrium, the capacity of the bid-based firm is $q_B^m = (V - \beta_B)/2$. The reserve price is $r^m = (V + \beta_B)/2$, and the bid-based firm's payoff is $\Pi_B(q_B^m, 0) = (V - \beta_B)^2/4$.*

   (b) *The parameters satisfy $\beta_B + 2 \leq V$. In the corresponding monopoly equilibrium, the capacity of the bid-based firm is $q_B^m = 1$. The reserve price is $r^m = V - 1$, and the bid-based firm's payoff is $\Pi_B(q_B^m, 0) = V - 1$.*

3. *There exists an equilibrium in which the fixed-price firm is a monopolist in the market, if and only if one of the following two conditions hold*

   (a) *The parameters satisfy $V \leq \min\{2\beta_B - \beta_F, \beta_F + 2, \beta_B + 1\}$. In such monopoly equilibrium, the capacity of the fixed-price firm is $q_F^m = V - \beta_F)/2$. The fixed price is $P^m = (V + \beta_F)/2$, and the fixed-price firm's payoff is $\Pi_F(0, q_F^m) = (V - \beta_F)^2/4$.*

(b) The parameters satisfy $2 + \beta_F \leq V \leq 1 + \beta_B$. In the corresponding equilibrium, the capacity of the fixed-price firm is given by $q_F^m = 1$. The fixed price is $P^m = V - 1$, and the fixed-price firm's payoff is $\Pi_F(0, q_F^m) = V - 1$.

To obtain meaningful insights into the capacity competition, we consider the special case where the firms' capacity costs are equal. The following corollary summarizes the equilibrium structure.

Corollary 1. *Suppose $\beta_B = \beta_F = \beta \leq V$, and $V > 2$. Then, there exists an equilibrium in which both firms operate in the market if and only if $\beta < V \leq \beta + \frac{9}{2} - \frac{3\sqrt{5}}{2}$. Furthermore, for $\beta + 2 \leq V$, there is an equilibrium where the bid-based firm is a monopolist. Finally, there is never an equilibrium in which the fixed-price firm is a monopolist.*

Thus, under equal capacity costs, if the fixed-price firm operates in the market in equilibrium, then so does the bid-based firm. On the other hand, there are values of the capacity cost for which in equilibrium, the fixed-price firm finds it optimal to not operate in the market. This suggests that the capacity competition intrinsically favors the bid-based firm.

6 Conclusion

In this paper, we analyze a model of two competing service providers offering service which differ in pricing rules and priority of service. We show under general settings the existence and uniqueness of a symmetric equilibrium of the customers' game, where the customers' strategy has a multi-threshold structure. In particular, we show the customers endogenously segregate between the two firms, with customer with very low and high waiting costs choosing to the obtain service from the bid-based firm, whereas customers with moderate waiting costs choosing to obtain service from the fixed-price firm. With the characterization of the unique equilibrium among the customers, we study the price and capacity competition between the two firms, under high arrival rate and explicit waiting time assumptions, and show that in this competition, the bid-based firm has an inherent advantage.

There are many avenues for future research. In our analysis, we have assumed that the bid-based firm uses a priority queue where customers are served in the decreasing order of their bids. One justification for this is practical: such bidding (priority) mechanisms are prevalent, for example in cloud computing services such as Amazon EC2, partly due to the ease of implementing such mechanisms. A more theoretical justification is that if bid-based firm were the only firm in the market, and the customer's waiting cost distribution were *regular*, then one can show using analysis similar to (Afèche and Pavlin 2016, Doroudi et al. 2013, Nazerzadeh and Randhawa 2014), that the revenue-optimal mechanism can be implemented as a first-price auction (with a reserve).

One natural direction for future research is to understand whether such bidding mechanisms are indeed optimal in the presence of a competing fixed-price firm. Formulating this question as a mechanism design problem leads to an endogenous participation constraint for the customers, which requires that the customers who join the bid-based firm obtain higher utility than they would have obtained if they joined the fixed price firm (or left without service). Unlike the usual mechanism design setting, the value of this outside option (the opportunity cost) depends on how many other customers choose this outside option, which is determined endogenously in equilibrium from the congestion at the fixed-price firm. Solving such a mechanism design problem with endogenous constraint is challenging.

Finally, in many practical applications, one needs to consider service abandonments and dynamic pricing of the firms in response to the real-time state of the system. Incorporating these practical considerations into our model is another interesting area for future research.

7 Acknowledgment

This work was supported by the National Science Foundation under grant CMMI-1462592. The authors thank Sven Seuken and Ludwig Dierks for pointing out an issue in an earlier version of this paper.

8 Biographies

Jiayang Gao received her Ph.D degree from the School of Operations Research and Information Engineering at Cornell University. Her thesis is on dynamic pricing and assortment personalization in strategic settings.

Krishnamurthy Iyer is an assistant professor in the School of Operations Research and Information Engineering at Cornell University. His research interests include game theory, mechanism design, and applied probability.

Huseyin Topaloglu is a professor in the School of Operations Research and Information Engineering at Cornell Tech. His research interests include revenue management, pricing and transportation logistics.

A Extensions to general distribution F

All our results in Sections 3–4, except uniqueness, can be extended to the case where the distribution of the unit waiting cost is any bounded continuously differentiable distribution F. Since we use the intermediate value theorem and monotonicity with respect to c for the proof of existence of equilibrium, our proof can be extended in a straightforward manner to general distribution F, on replacing c_1, c_2 with $F(c_1)$, $F(c_2)$ respectively, and replacing α with $F(c_2) - F(c_1)$. This argument would show the existence of equilibrium strategies with thresholds in the values of $F(c)$, which translates to the thresholds in values of c. For uniqueness, we need the additional requirement that the density of F is strictly positive on its support. This is a reasonable requirement, since otherwise there may exist multiple values of thresholds for which effectively the same set of customers obtain service from each firm.

B Proofs of the results in Section 3

Proof of Lemma 1. Fix a symmetric equilibrium (x, φ). Consider two customers, with unit waiting costs c and $c' > c$, such that both choose to obtain service. The total expected cost of the customer with unit waiting cost c is given by $\varphi(c) + cw(c|x, \varphi)$. (Here, we adopt the convention that if $x(c) = \mathsf{FIX}$, then $\varphi(c) \triangleq P$.) If this customer deviates, and chooses the service decision $x(c')$ with bid $\varphi(c')$, her total expected cost would be $\varphi(c') + cw(c'|x, \varphi)$. As (x, φ) is an equilibrium, we must have that the customer's total expected cost on following her equilibrium action should be at most as that from deviating to the action $(x(c'), \varphi(c'))$. This implies that

$$\varphi(c) + cw(c|x, \varphi) \leq \varphi(c') + cw(c'|x, \varphi). \tag{12}$$

Similarly for customer with unit waiting cost c' cannot become better off from deviating to the action $(x(c), \varphi(c))$. This implies that

$$\varphi(c') + c'w(c'|x, \varphi) \leq \varphi(c) + c'w(c|x, \varphi). \tag{13}$$

Adding the two inequalities above, we obtain that $(c' - c)\left(w(c'|x, \varphi) - w(c|x, \varphi)\right) \leq 0$. Since $c' > c$, we have $w(c|x, \varphi) \geq w(c'|x, \varphi)$, thereby proving that the expected waiting cost in equilibrium is non-increasing in the unit waiting cost.

Using the fact that $w(c|x, \varphi) \geq w(c'|x, \varphi)$ in (12), we obtain that $\varphi(c) \leq \varphi(c')$, thus proving that the expected payment in equilibrium is non-decreasing in the unit waiting cost.

Moreover, using $c' > c$ and $w(c'|x, \varphi) > 0$, we get $\varphi(c') + cw(c'|x, \varphi) < \varphi(c') + c'w(c'|x, \varphi)$. Combining this with equation (12), we obtain that $\varphi(c) + cw(c|x, \varphi) < \varphi(c') + c'w(c'|x, \varphi)$. Thus, the total expected cost is strictly increasing in the unit waiting cost.

Now we only need to show that the bids in bid-based firm $\varphi(c)$ is uniquely determined by and strictly increasing in the fraction of customers in the bid-based firm with unit waiting cost less than c, which we denote by $B(c|(x, \varphi)) = \mathbf{P}(\widehat{C} \leq c \mid x(\widehat{C} = \mathsf{BID})$, where $\widehat{C}$ is a random variable with distribution F. (In the following, we drop (x, φ) from the notation for $B(\cdot)$.) We first show that for any $c' > c$, $B(c') = B(c)$ implies $\varphi(c') = \varphi(c)$. Earlier in the proof, we show that the bids are non-decreasing in unit waiting cost, so $\varphi(c') \geq \varphi(c)$. Suppose $\varphi(c') > \varphi(c)$, then if the customer with unit waiting cost c' decreases her bid from $\varphi(c')$ to $\varphi(c)$, her priority decreases. And the set of unit waiting costs with higher priority than the customer when she bids $\varphi(c)$ while having lower priority when she bids $\varphi(c')$ is a subset of $\{\hat{c} : B(\hat{c}) = B(c')\}$. As a result, the expected waiting time increases by at most the expected time waiting for customers with unit waiting cost in the set $\{\hat{c} : B(\hat{c}) = B(c')\}$. Since the set has measure zero, the increase in expected waiting time is zero. So the customer with unit waiting cost c' is strictly better off submitting a bid $\varphi(c)$, which contradicts the fact that (x, φ) is a symmetric equilibrium. Thus, $B(c') = B(c)$ implies $\varphi(c') = \varphi(c)$, and the bidding function is uniquely determined by $B(\cdot)$.

Now we want to show that for any $c' > c$, $B(c') > B(c)$ implies $\varphi(c') > \varphi(c)$. Since bids are non-decreasing in unit waiting cost, we have $\varphi(c') \geq \varphi(c)$. Suppose $\varphi(c') = \varphi(c)$, then every customer with unit waiting cost between c and c' has the same bid. If the customer with unit waiting cost c' increases her bid to $\varphi(c') + \epsilon$, then she can get service before anyone in the set $\{\hat{c} : B(c) \leq B(\hat{c}) < B(c'), x(\hat{c}) = \mathsf{BID}\}$, and her expected waiting time decreases by at least the expected time she had to spend waiting for customers in the set $\{\hat{c} : B(c) \leq B(\hat{c}) < B(c'), x(\hat{c}) = \mathsf{BID}\}$ when her bid was $\varphi(c')$. Since the set has measure $B(c') - B(c) > 0$ and we break the ties among the customers in this set uniformly at random, the expected time one has to spend waiting for these customers to complete their service is positive (and bounded below). Thus for small enough ϵ, the customer with unit waiting cost c' is better off using bid $\varphi(c') + \epsilon$, which again contradicts the fact that (x, φ) is an equilibrium. So in equilibrium, $B(c') > B(c)$ implies $\varphi(c') > \varphi(c)$, i.e., the bidding function is strictly increasing in $B(\cdot)$. □

Proof of Lemma 2. Fix a symmetric equilibrium (x, φ). For customers with unit cost c such that $x(c) = \mathsf{LEAVE}$, the expected payment and the expected waiting time are equal to zero, and the bid is by definition equal to V. Thus, for such customers, these quantities are completely determined by the service decision x. Hence, we focus only on those customers that choose to obtain service, i.e., for values of c such that $x(c) \neq \mathsf{LEAVE}$.

For a customer choosing to obtain service in the bid-based firm, by Lemma 1, the bid is strictly increasing in $B(c)$, the fraction of arrivals to the bid-based firm with unit waiting cost less than c. Since the priority of service is determined in the descending order of the bids (with ties broken uniformly at random), service is provided in the decreasing order of $B(\cdot)$ and thus, the expected waiting time in the bid-based firm depends only on $B(\cdot)$. Similarly, since the service discipline in the fixed-price firm is first-in-first-out, the expected waiting time in the fixed-price firm depends only on the arrival rate of the customers into the fixed-price firm, which is again determined by the service decision $x(\cdot)$. This argument proves that the expected waiting cost in equilibrium is completely specified by the service decision $x(\cdot)$. Henceforth, we denote the expected waiting time of a customer with unit waiting cost c in the symmetric equilibrium (x, c) by $w(c|x)$.

Next, we show that the bidding function is uniquely determined by service decision in equilibrium. Fix a symmetric equilibrium (x, φ), and consider a customer with unit waiting cost c with $x(c) \neq \mathsf{LEAVE}$. Suppose, for some $\hat{c}$, the customer deviates to the service decision $x(\hat{c})$, with the corresponding bid $\varphi(\hat{c})$. If $\hat{c}$ is such that $x(\hat{c}) \neq \mathsf{LEAVE}$, her total expected utility is given by $\pi(\hat{c}, c) = V - cw(\hat{c}|x) - \varphi(\hat{c})$. (Here, recall our convention that $\varphi(\hat{c}) = P$ if $x(\hat{c}) = \mathsf{FIX}$.) On the other hand, if $x(\hat{c}) = \mathsf{LEAVE}$, then $\pi(\hat{c}, c) = 0$. By the fact that (x, φ) is a symmetric equilibrium, we obtain

$$\pi(c, c) = \max_{\hat{c} \in [0,1]} \pi(\hat{c}, c),$$

for all $c \in [0, 1]$, with the maximum being attained at $\hat{c} = c$. We use the Mirrlees trick (Mirrlees 1971) to compute $\pi(c, c)$ in equilibrium. Observe that for a fixed $\hat{c}$, the function $\pi(\hat{c}, c)$ is linear in c. Thus, the preceding equation implies that $\pi(c, c)$, as a function of c, is a maximum of linear functions, and hence is convex. This implies that $\pi(c, c)$ is differentiable almost everywhere (Rockafellar 1997). Then, by the envelope theorem, we obtain

$$\frac{d\pi(c,c)}{dc} = \left.\frac{\partial\pi(\hat{c},c)}{\partial c}\right|_{\hat{c}=c} = -w(c|x),$$

for almost all c such that $x(c) \neq \mathsf{LEAVE}$. (Here, the last equality follows from the assumption that $x(\hat{c}) \neq \mathsf{LEAVE}$ when $\hat{c} = c$, since otherwise, the partial derivative above is equal to zero). Now, note that if $x(c) \neq \mathsf{LEAVE}$, then from Lemma 1, we obtain that $x(c') \neq \mathsf{LEAVE}$ for all $c' \leq c$. This, combined with the fact that $\pi(0,0) = V$, yields on integrating

$$\pi(c,c) = V - \int_0^c w(t|x)dt.$$

Finally, since $x(c) \neq \mathsf{LEAVE}$, we have $\pi(c,c) = V - cw(c|x) - \varphi(c)$. Comparing the two expressions, we get

$$\varphi(c) = \int_0^c w(t|x)dt - cw(c|x),$$

for almost all c such that $x(c) \neq \mathsf{LEAVE}$. Thus, the bidding function is completely determined from the expected waiting time $w(\cdot|x)$, which in turn depends only on the service decision x. ☐

C Proofs of auxiliary results in Section 4

Proof of Lemma 3. The first two statements follow directly from a straightforward stochastic coupling argument. We provide a brief sketch of the continuity and monotonicity of Γ for completeness. To show that $\Gamma(a) < \Gamma(b)$ for $0 \leq a < b < k$, consider two coupled copies Q_a, Q_b of the preemptive bid-based priority queue with arrival rate $\lambda \in (b,k)$. Suppose two customers, one per queue, arrive at time 0 with identical service requirement, with the only difference being that the customer to the copy Q_i has a priority level such that the arrival rate of customers to Q_i with higher priority is exactly equal to i, for each $i \in \{a,b\}$. Through this coupling, it follows directly that the waiting time of the customer to the queue Q_a is almost surely less than the waiting time of the customer to the queue Q_b. Taking expectations, we obtain from the definition of Γ that $\Gamma(a) \leq \Gamma(b)$. The strict inequality and continuity follow from the fact that the arrival rate of customers with intermediate priority between a and b is positive, and decreases to zero as a approaches b.

The third statement follows directly from the fact that the expected service time is one. Finally, since $\int_0^y \Gamma(t)dt/y$ is the average waiting time in the system when there are k servers and the arrival rate is y, it is no less than the average waiting time in a work conserving system with the same number of servers and the same arrival rate. Since the latter tends to infinity as y approaches k, we have $\int_0^y \Gamma(t)dt \to \infty$ as $y \to k$. ☐

Proof of Proposition 1. We prove a more general statement, which will be useful later when we prove Lemma 4 and Lemma 5. The statement we prove here is as follows: consider a setting where for some $c_\ell \leq 1$, service is mandatory for anyone with cost $c \leq c_\ell$, and service is forbidden for anyone with cost $c > c_\ell$. In other words, the action LEAVE is not available to customers with unit costs $c \leq c_\ell$, and the actions BID and FIX are not available to those with unit costs $c > c_\ell$. Then, in such a system, a strategy $\bar{c} = (c_1, c_2, c_\ell)$, with $0 < c_1 < c_2 < c_\ell \leq 1$ is a symmetric equilibrium if and only if (7) and (8) are satisfied. The statement of the proposition then corresponds to the special case of $c_\ell = 1$.

We split the proof into two steps showing first the necessity and then the sufficiency of the conditions for equilibrium.

Necessity of (7) **and** (8)**:** Suppose $\bar{c} = (c_1, c_2, c_\ell)$, with $0 < c_1 < c_2 < c_\ell \leq 1$ is a symmetric equilibrium. We begin by showing that the condition (7) holds.

Consider a customer with unit waiting cost $c_1 + \epsilon$ for some $\epsilon \in (0, c_2 - c_1)$, who obtains service from the fixed-price firm, and incurs a total expected cost equal to $(c_1 + \epsilon)w_F(\bar{c}) + P$. If such a customer decides instead to obtain service from the bid-based firm and submit the same bid as a customer with unit waiting cost c_1, then her total expected cost is given by

$$(c_1 + \epsilon)w(c_1|\bar{c}) + \varphi(c_1|\bar{c}) = \int_0^{c_1} w(t|\bar{c})dt + \epsilon w(c_1|\bar{c}),$$

where the right hand side follows from (2). In equilibrium such a unilateral deviation must be non-preferable. Since this is true for any $\epsilon \in (0, c_2 - c_1)$, we obtain $\int_0^{c_1} w(t|\bar{c})dt \geq c_1 w_F(\bar{c}) + P$. On the other hand, since a customer with unit waiting cost c_1 prefers to obtain service from the bid-based firm instead of the fixed-price firm, we have $\int_0^{c_1} w(t|\bar{c})dt \leq c_1 w_F(\bar{c}) + P$. Together, these inequalities yield (7).

We now show that condition (8) holds. As the expected waiting time is non-increasing in c, we have $w_B(c_1|\bar{c}) \geq w_F(\bar{c}) \geq w_B(c_2|\bar{c})$. Since customers in the bid-based firm are served in the decreasing order of their waiting costs, by the fact that F is continuous and hence has no atom, it follows that the expected waiting time for a customer with unit waiting cost c_1 and c_2 must be equal. From this, we obtain (8).

Sufficiency of (7) **and** (8)**:** Suppose all customers adopt a strategy $\bar{c} = (c_1, c_2, c_\ell)$ with $0 < c_1 < c_2 < c_\ell \leq 1$ satisfying the conditions (7) and (8). We begin by obtaining the expression for the expected waiting time and the expected total cost under this strategy profile.

Observe that since the expected waiting times in the bid-based firm and the fixed-price firm satisfy the condition (8), the expected waiting time function defined by

$$w(t|\bar{c}) = \begin{cases} w_B(t|\bar{c}), & \text{for } t \in [0, c_1] \cup [c_2, c_\ell]; \\ w_F(\bar{c}), & \text{for } t \in (c_1, c_2) \end{cases} \tag{14}$$

is continuous and non-increasing over $t \in [0, c_\ell]$.

Next, note that for a customer with unit waiting cost c obtaining service in the bid-based firm, by (2), her total expected cost is given by $\int_0^c w(t|\bar{c})dt$. We show that this expression holds for all customers, even those obtaining service from the fixed-price firm. To see this, note that for such a customer with unit waiting cost $c \in (c_1, c_2)$, the total expected cost is given by

$$cw_F(\bar{c}) + P = c_1 w_F(\bar{c}) + P + (c - c_1)w_F(\bar{c}) = \int_0^{c_1} w_B(t|\bar{c})dt + (c - c_1)w_F(\bar{c}) = \int_0^c w(t|\bar{c})dt,$$

where the second equality follows (7), and the third from (14).

Next, consider a customer with unit waiting cost $c \leq 1$. If she adopts the actions of a customer with unit waiting cost $c' \neq c$, then her total expected cost, using (2), is given by

$$\begin{aligned} cw(c'|\bar{c}) + \varphi(c'|\bar{c}) &= \int_0^{c'} w(t|\bar{c})dt + (c-c')w(c'|\bar{c}) \\ &= \int_0^{c} w(t|\bar{c})dt + \int_c^{c'} w(t|\bar{c})dt + (c-c')w(c'|\bar{c}) \\ &\geq \int_0^{c} w(t|\bar{c})dt + (c'-c)w(c'|\bar{c}) + (c-c')w(c'|\bar{c}) \\ &= \int_0^{c} w(t|\bar{c})dt. \end{aligned}$$

Here, the inequality follows from the fact that the expected waiting time $w(\cdot|\bar{c})$ is non-increasing. Since the right hand side denotes the total expected cost for the customer under strategy $\bar{c}$, this implies that a best response of a customer with unit waiting cost c is the action suggested by the strategy $\bar{c}$.

Taken together, this implies that the strategy $\bar{c}$ is a best response, assuming all others follow $\bar{c}$, and hence it is a symmetric equilibrium. □

Proof of Lemma 4. We have already shown the necessity of condition (10) in the beginning of Subsection 4.2. It is left to show the sufficiency of condition (10). Suppose condition (10) holds, we show that $\bar{c}(u)$ is an equilibrium for $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$.

First consider the case when $0 < c_1 < c_2 < 1$. By Proposition 1, since $\bar{c} = (c_1, c_2, 1)$ is a symmetric equilibrium for $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, it satisfies (7) and (8). First, we prove that $\bar{c}(u)$ satisfies conditions (7) and (8) in $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$. To prove this observe that

$$\begin{aligned} \int_0^{c_1 u} w_B(t|\bar{c}(u))dt = \int_0^{c_1 u} \Gamma\left(\frac{\lambda}{u}u - \frac{\lambda}{u}(c_2 u - c_1 u) - \frac{\lambda}{u}t\right)dt &= u\int_0^{c_1} \Gamma(\lambda - \lambda(c_2 - c_1) - \lambda t)dt \\ &= u\int_0^{c_1} w_B(t|\bar{c})dt. \end{aligned}$$

Here, the first equality follows from the definition of $w_B(t|\bar{c}(u))$, the second equality follows from a change of variable and the third follows from the definition of $w_B(t|\bar{c})$. Now, since $\bar{c}$ satisfies (7) in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, we obtain

$$\begin{aligned} u\int_0^{c_1} w_B(t|\bar{c})dt = u\left(c_1 w_F(\bar{c}) + P\right) = u\left(c_1 \Phi(\lambda(c_2 - c_1)) + P\right) &= c_1 u \Phi\left(\frac{\lambda}{u}(c_2 u - c_1 u)\right) + Pu \\ &= (c_1 u) w_F(\bar{c}(u)) + Pu. \end{aligned}$$

Here, the second and the fourth equalities follow from the definition of $w_F(\bar{c})$ and $w_F(\bar{c}(u))$ respectively. From this, we obtain that condition (7) holds for $\bar{c}(u)$ in $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$. Through a similar argument, we can show that condition (8) holds for $\bar{c}(u)$ in $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$.

As we show in the proof of Proposition 1, since $\bar{c}(u)$ satisfies conditions (7) and (8) in $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$, the profile $\bar{c}(u)$ is a symmetric equilibrium in a system where customers with unit costs $c \leq u$ cannot choose LEAVE, and those with unit costs $c > u$ must choose LEAVE. Thus, it suffices to show that in the system $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$, the action LEAVE is not optimal for customers with unit costs $c \leq u$, and it is optimal for those with $c > u$.

We start by showing that LEAVE is not optimal for any customer with unit cost $c \leq u$. Note that the expected total cost for such a customer is $\int_0^c w(t|\bar{c}(u))dt$. Hence, using condition (10), we obtain

$\int_0^c w(t|\bar{c}(u))dt \leq \int_0^u w(t|\bar{c}(u))dt \leq V$, implying that the expected total cost for customer with unit cost $c \leq u$ is no more than the value of service completion V. Hence, such a customer would prefer to obtain service over leaving without service, implying that LEAVE is not optimal for such a customer.

Finally we show that LEAVE is optimal for customer with $c > u$, when $u < 1$. Suppose, for the sake of arriving at a contradiction, there exists a $c > u$ such that a customer with unit cost c has an action $a \in \{\mathsf{FIX}, \mathsf{BID}\}$ that strictly dominates LEAVE. (Here, in the case where $a = \mathsf{BID}$, the action also includes a corresponding bid.) This implies that the expected total cost of the action a for the customer is strictly less than the value V of service completion. Now, consider a customer with unit cost u who takes the action a. For this customer, the waiting costs are lower than that of a customer with unit cost $c > u$, and hence her expected total cost must also be strictly lower than V. However, this contradicts with (10), which states that when $u < 1$, a customer with a unit cost u has expected total cost equal to V. Thus, LEAVE is optimal for any customer with $c > u$.

The preceding two paragraphs together prove that when $0 < c_1 < c_2 < 1$, no customer wants to deviate from the action suggested by $\bar{c}(u)$. Thus $\bar{c}(u)$ is an equilibrium for $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$.

On the other hand, suppose $c_1 = c_2 = 1$. Repeating the same argument as in the proof of Theorem 2, it follows that in the system $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$, the action suggested by $\bar{c}(u)$ is optimal for customers with unit cost $c \leq u$, if such customers cannot choose LEAVE. Thus, to show that $\bar{c}(u)$ is an equilibrium for the system $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$, it suffices to show that LEAVE is not optimal for customers with unit costs $c \leq u$, whereas it is optimal for customers with unit costs $c > u$. This latter statement follows using the same argument as for the case with $0 < c_1 < c_2 < 1$. Thus it follows that $\bar{c}(u)$ is an equilibrium for $\mathsf{SYS}_{\mathsf{op}}(\frac{\lambda}{u}, Pu, V)$ in this case too. □

Proof of Lemma 5. Let (c_1, c_2, u) be a symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$. If $0 < c_1 < c_2 < u \leq 1$, by the argument in the proof of Proposition 1, the strategy (c_1, c_2, u) satisfies conditions (7) and (8) in $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$, when customers with unit costs $c \leq u$ cannot choose LEAVE, whereas those with unit costs $c > u$ must choose LEAVE. From this observation, using the same argument as in the proof of Lemma 4, it follows that the strategy $(\frac{c_1}{u}, \frac{c_2}{u}, 1)$ satisfies conditions (7) and (8) in $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$. Thus, using Proposition 1, we obtain that $(\frac{c_1}{u}, \frac{c_2}{u}, 1)$ is an equilibrium for $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$.

If $c_1 = c_2 = u$, by the same argument as used in the proof of Theorem 2, we obtain that the $\bar{c} = (c_1, c_2, u)$ satisfies a condition analogous to (6) in $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$, namely

$$\int_0^u w_B(t|\bar{c})dt \leq u + P,$$

where $w_B(t|\bar{c}) = \Gamma(\lambda u - \lambda t)$. After a change of variables, the preceding condition can be shown to be equivalent to $\frac{P}{u} \geq P_{\max}(\lambda u)$. Theorem 2 then implies that $(1, 1, 1) = (\frac{c_1}{u}, \frac{c_2}{u}, 1)$ is the unique symmetric equilibrium of the system $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$. □

D Proofs of Theorem 2 and Theorem 3

The main stepping stone to the proving Theorems 2 and 3 is Theorem 10, where we show that symmetric equilibria of the system $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ are in one-to-one correspondence with fixed points of a function $\Psi(\cdot, P)$. This function Ψ is obtained by combining the two necessary and sufficient conditions in Proposition 1, as listed in (9), and reproduced below:

$$\int_0^{c_1} \Gamma(\lambda(1-\alpha-t))dt = c_1\Phi(\lambda\alpha) + P$$
$$\Gamma(\lambda(1-\alpha-c_1)) = \Phi(\lambda\alpha).$$

Informally, for any $\alpha \in [0,1]$, we first determine the value of c_1 that satisfies the first equation. Then, using this value of c_1, we define $\Psi(\lambda\alpha, P)$ to equal $\Phi^{-1}(\Gamma(\lambda - \lambda\alpha - \lambda c_1))$. By the definition of $\Psi(\cdot, P)$, it follows that if $\Psi(\lambda\alpha, P) = \lambda\alpha$, then the second equation also holds. Thus, to show existence and uniqueness, one must show that the function $\Psi(\cdot, P)$ has a unique fixed point. We exhibit such a fixed point by explicit construction in (15). The technical challenge in the proof, which is addressed in the following lemmas, is to show these implicit definitions are sound, i.e., the quantities are uniquely defined, and that the function Ψ is continuous and strictly decreasing. We develop this informal argument more formally below.

Define $\nu = \min\{n, \lambda\}$ and $\kappa = \min\{k, \lambda\}$. Let ξ be defined as

$$\xi \triangleq \sup\{\lambda - \kappa < x < \nu : \Gamma(\lambda - x) > \Phi(x)\}.$$

As $x \downarrow \lambda - \kappa$, we have $\Gamma(\lambda - x) \to \Gamma(\kappa) > 1$ if $\kappa < k$ and $\Gamma(\lambda - x) \to \infty$ if $\kappa = k$. On the other hand, $\Phi(\lambda - \kappa) = \Phi(0) = 1$ if $\kappa < k$, and $\Phi(\lambda - \kappa) < \Phi(n) = \infty$ if $\kappa = k$. Hence, $\xi \in (\lambda - \kappa, \nu]$. By continuity and strict monotonicity of Γ and Φ, we obtain that for all $x \in [\lambda - \kappa, \xi)$, $\Gamma(\lambda - x) > \Phi(x)$, and $\Gamma(\xi) \geq \Phi(\xi)$, with equality if $\xi < \nu$.

Define the function s as follows: for $x \in (\lambda - \kappa, \nu)$, and $z \in [0, \lambda - x]$,

$$s(z, x) \triangleq \int_x^{z+x} \Gamma(\lambda - t)dt - z\Phi(x).$$

First observe that from Assumption 1, we obtain that $s(z, x)$ is twice-differentiable in z and x (almost everywhere). Note that we have $\frac{\partial^2 s(z,x)}{\partial z^2} = -\Gamma'(\lambda - z - x) < 0$. Thus $s(z, x)$ is strictly concave in z for each $x \in (\lambda - \kappa, \nu)$. Thus, there exists a unique maximizer of $s(z, x)$ over $z \in [0, \lambda - x]$ for all $x \in (\lambda - \kappa, \nu)$. Define $z_{\max}(x)$ to be this unique maximizer for $x \in (\lambda - \kappa, \nu)$:

$$z_{\max}(x) = \arg\max_{z \in [0, \lambda - x]} s(z, x).$$

We have the following properties of $z_{\max}(x)$.

Lemma 8. *The function $z_{\max}(x)$ is continuous over $x \in (\lambda - \kappa, \nu)$. Further, $z_{\max}(\cdot)$ is strictly decreasing over $(\lambda - \kappa, \xi)$ and $z_{\max}(x) = 0$ for all $x \in [\xi, \nu)$. Moreover, $z_{\max}(x) = \lambda - x - \Gamma^{-1}(\Phi(x)) \in (0, \lambda - x)$ for $x \in (\lambda - \kappa, \xi)$, and if $\kappa = \lambda < k$, then $z_{\max}(0) \triangleq \lim_{x \downarrow 0} z_{\max}(x) = \lambda$.*

Proof. The continuity of $z_{\max}(x)$ follows from the application of Berge's maximum theorem (Aliprantis and Border 2006) to the function $s(z, x)$, which is strictly concave in z for each $x \in (\lambda - \kappa, \nu)$.

Next, observe that for all $x \in (\lambda - \kappa, \nu)$, we have

$$\left.\frac{\partial s(z, x)}{\partial z}\right|_{z=0} = \Gamma(\lambda - x) - \Phi(x),$$

which is positive for all $x \in (\lambda - \kappa, \xi)$, equals zero for $x = \xi$ if $\xi < \nu$, and is negative for $x \in (\xi, \nu)$. Moreover, we have

$$\left.\frac{\partial s(z,x)}{\partial z}\right|_{z=\lambda-x} = \Gamma(0) - \Phi(x),$$

which is negative for all $x \in (\lambda - \kappa, \nu)$. From this, it follows that the unique maximizer $z_{\max}(x)$ of $s(z,x)$ over $z \in [0, \lambda - x]$ must lie in $(0, \lambda - x)$ for all $x \in (\lambda - \kappa, \xi)$, and equals zero for $x \in [\xi, \nu)$.

Since $z_{\max}(x) \in (0, \lambda - x)$ for all $x \in (\lambda - \kappa, \xi)$, we have by first order necessary conditions,

$$\left.\frac{\partial s(z,x)}{\partial z}\right|_{z=z_{\max}(x)} = \Gamma(\lambda - z_{\max}(x) - x) - \Phi(x) = 0, \quad \text{for all } x \in (\lambda - \kappa, \xi).$$

Thus, we obtain $z_{\max}(x) = \lambda - x - \Gamma^{-1}(\Phi(x))$ for all $x \in (\lambda - \kappa, \xi)$. Since Γ and Φ are strictly increasing, this implies that $z_{\max}(x)$ is strictly decreasing over $x \in (\lambda - \kappa, \xi)$. Finally, suppose $\kappa = \lambda < k$. Then, for all small enough $\epsilon > 0$, we have $z_{\max}(\epsilon) = \lambda - \epsilon - \Gamma^{-1}(\Phi(\epsilon))$. Taking limits as $\epsilon \to 0$, and observing that $\Gamma(0) = \Phi(0) = 1$, we obtain that $z_{\max}(0) = \lambda$. □

Lemma 9. *The function $s(z_{\max}(x), x)$ is continuous over $(\lambda - \kappa, \nu)$ and strictly decreasing over $(\lambda - \kappa, \xi)$. Further, we have $s(z_{\max}(x), x) = 0$ for all $x \in [\xi, \nu)$.*

Proof. The continuity follows trivially from the continuity of $s(z,x)$ and Lemma 8. For $x \in [\xi, \nu)$, the result follows directly from $z_{\max}(x) = 0$ and the definition of $s(z,x)$. For $x \in (\lambda - k, \xi)$, we have, by the envelope theorem,

$$\frac{ds(z_{\max}(x), x)}{dx} = \Gamma(\lambda - z_{\max}(x) - x) - \Gamma(\lambda - x) - z_{\max}(x)\Phi'(x) < 0,$$

where the last inequality follows from the fact that Γ is strictly increasing, Φ is strictly increasing, and $z_{\max}(x) > 0$. □

Lemma 10. *We have*

$$\lim_{x \downarrow \lambda - \kappa} s(z_{\max}(x), x) = \lambda P_{\max}(\lambda) = \begin{cases} \int_0^\lambda \Gamma(t)dt - \lambda & \textit{if } \kappa = \lambda < k; \\ \infty & \textit{if } \kappa = k. \end{cases}$$

Proof. If $\kappa = \lambda < k$, then $\lambda - \kappa = 0$. Then, by continuity of $s(z,x)$ and $z_{\max}(x)$, we obtain

$$\begin{aligned} \lim_{x \downarrow \lambda - \kappa} s(z_{\max}(x), x) &= s(z_{\max}(0), 0) \\ &= s(\lambda, 0) \\ &= \int_0^\lambda \Gamma(\lambda - t)dt - \lambda\Phi(0) \\ &= \int_0^\lambda \Gamma(t)dt - \lambda, \end{aligned}$$

where we use that fact that $z_{\max}(0) = \lambda$ if $\kappa = \lambda < k$, and that $\Phi(0) = 1$.

Next, for $\kappa = k \le \lambda$, we have for $\lambda - \kappa < x < \xi$,

$$\begin{aligned} s(z_{\max}(x), x) &\ge s(\lambda - x, x) \\ &= \int_x^{\lambda} \Gamma(\lambda - t)dt - (\lambda - x)\Phi(x) \\ &= \int_0^{\lambda - x} \Gamma(t)dt - (\lambda - x)\Phi(x) \\ &\ge \int_0^{\lambda - x} \Gamma(t)dt - \kappa\Phi(\xi). \end{aligned}$$

For $\kappa = k$, by Assumption 1, $\lim_{x \downarrow \lambda - \kappa} s(z_{\max}(x), x) = \infty$. □

Lemma 11. *Suppose $P \ge P_{\max}(\lambda)$. Then, for all $x \in (\lambda - \kappa, \nu)$ and for all $z \in [0, \lambda - x]$, we have $s(z, x) < \lambda P$.*

Proof. The result follows directly from Lemma 9 and Lemma 10. □

For any $0 < P < P_{\max}(\lambda)$, define $\digamma(P)$ as follows:

$$\digamma(P) = \sup\{\lambda - \kappa < x < \nu : s(z_{\max}(x), x) > \lambda P\} \tag{15}$$

From Lemma 9 and Lemma 10, we have $\digamma(P) \in (\lambda - \kappa, \xi)$ for all $P \in (0, P_{\max}(\lambda))$.

Note that for $P \in (0, P_{\max}(\lambda))$, and for each $x \in (\lambda - \kappa, \digamma(P)]$, we have $s(0, x) = 0 < \lambda P$. Also, by continuity of $s(z_{\max}(x), x)$, we have $s(z_{\max}(x), x) > \lambda P$ for all $x \in (\lambda - k, \digamma(P))$. Thus, there exists a unique solution $z = v_1(x, P) \in (0, z_{\max}(x))$ to the equation $s(z, x) = \lambda P$ for each $x \in (\lambda - k, \digamma(P))$ and $0 < P < P_{\max}(\lambda)$. (Here uniqueness follows from the strict concavity of $s(z, x)$ in z). For $x \in [\digamma(P), \nu)$, define $v_1(x, P) = z_{\max}(\digamma(P))$. (It is straightforward to show that $v_1(x, P)$ is continuous at $x = \digamma(P)$.)

Define

$$v_2(x, P) \triangleq \begin{cases} v_1(x, P) + x & \text{if } x \in (\lambda - \kappa, \digamma(P)]; \\ v_1(\digamma(P), P) + \digamma(P) & \text{if } x \in (\digamma(P), \nu). \end{cases}$$

Note that for all $0 < P < P_{\max}(\lambda)$, and $x \in (\lambda - \kappa, \digamma(P)]$, we have $v_2(x, P) = v_1(x, P) + x < z_{\max}(x) + x \le \lambda$. Hence, $v_2(x, P) < \lambda$. Define $\Psi(x, P) \triangleq \Phi^{-1}(\Gamma(\lambda - v_2(x, P)))$ for all $x \in (\lambda - \kappa, \nu)$ and $0 < P < P_{\max}(\lambda)$.

Lemma 12. *For all $P \in (0, P_{\max}(\lambda))$, $\Psi(x, P)$ is strictly decreasing in x over $(\lambda - \kappa, \digamma(P)]$.*

Proof. It suffices to show that $v_1(x, P)$ is strictly increasing in x over $(\lambda - \kappa, \digamma(P)]$. This is because, then so is $v_2(x, P) = v_1(x, P) + x$, and the proof follows from observing that both Γ and Φ are strictly increasing.

Note that for $P \in (0, P_{\max}(\lambda))$, we have by definition, $s(v_1(x, P), x) = \lambda P$ for all $x \in (\lambda - \kappa, \digamma(P)]$. This implies, on differentiating with respect to x,

$$\frac{\partial v_1(x, P)}{\partial x} \left.\frac{\partial s(z, x)}{\partial z}\right|_{z = v_1(x,P)} + \left.\frac{\partial s(z, x)}{\partial x}\right|_{z = v_1(x,P)} = 0.$$

Observe that, for $x \in (\lambda - \kappa, \digamma(P))$,

$$\left.\frac{\partial s(z, x)}{\partial z}\right|_{z = v_1(x,P)} = \Gamma(\lambda - v_1(x, P) - x) - \Phi(x) > 0,$$

and

$$\left.\frac{\partial s(z,x)}{\partial x}\right|_{z=v_1(x,P)} = \Gamma(\lambda - v_1(x,P) - x) - \Gamma(\lambda - x) - v_1(x,P)\Phi'(x) < 0.$$

This implies that $\frac{\partial v_1(x,P)}{\partial x} > 0$, and hence we are done. □

Finally, we have the following characterization of the fixed point of $\Psi(\cdot, P)$ for all $P \in [0, P_{\max}(\lambda))$.

Lemma 13. *For each $P \in (0, P_{\max}(\lambda))$, the equation $\Psi(x,P) = x$ with $x \in (\lambda - \kappa, \nu)$ has a unique solution given by $\digamma(P) \in (\lambda - \kappa, \xi)$.*

Proof. Since $\digamma(P) \in (\lambda - \kappa, \xi)$, from Lemma 8 we obtain that $z_{\max}(\digamma(P)) = \lambda - \digamma(P) - \Gamma^{-1}(\Phi(\digamma(P)))$. By continuity of $s(z_{\max}(x), x)$, we have $s(z_{\max}(\digamma(P)), \digamma(P)) = \lambda P$. This implies that $v_1(\digamma(P), P) = z_{\max}(\digamma(P)) = \lambda - \digamma(P) - \Gamma^{-1}(\Phi(\digamma(P)))$. Substituting the expression for $v_1(\digamma(P), P)$, we obtain

$$\Psi(\digamma(P), P) = \Phi^{-1}(\Gamma(\lambda - v_2(\digamma(P), P))) = \Phi^{-1}(\Gamma(\lambda - v_1(\digamma(P), P) - \digamma(P))) = \digamma(P).$$

Finally, from Lemma 12, we obtain that $\Psi(x,P)$ is strictly decreasing in $(\lambda - \kappa, \digamma(P)]$. Since $\Psi(x,P) = \digamma(P) < x$ for $x \in (\digamma(P), \nu)$, we obtain that $\digamma(P)$ is the only solution to $\Psi(x,P) = x$ in $(\lambda - \kappa, \nu)$ for all $P \in (0, P_{\max}(\lambda))$. □

The following theorem relates the fixed points of $\Psi(\cdot, P)$ to symmetric equilibria.

Theorem 10.

1. *Suppose for some $P > 0$, there exists a symmetric equilibrium $\bar{c} = (c_1, c_2, 1)$ with $0 < c_1 < c_2 < 1$. Then, $P < P_{max}(\lambda)$, and $x \triangleq \lambda(c_2 - c_1) \in (\lambda - \kappa, \nu)$ satisfies $\Psi(x,P) = x$, with $\lambda c_1 = v_1(x,P)$.*
2. *Conversely, suppose for some $x \in (\lambda - \kappa, \nu)$ and $P \in (0, P_{\max}(\lambda))$, we have $\Psi(x,P) = x$. Then, $\bar{c} = (c_1, c_2, 1)$ with $c_1 = \frac{v_1(x,P)}{\lambda} > 0$ and $c_2 = \frac{v_2(x,P)}{\lambda} \in (c_1, 1)$, constitutes a symmetric equilibrium.*

Proof. We provide the proof in two steps corresponding to the two statements.

Step 1. Suppose $\bar{c} = (c_1, c_2, 1)$ is a symmetric equilibrium with $0 < c_1 < c_2 < 1$. Let $x = \lambda(c_2 - c_1)$. By stability of the fixed-price firm in equilibrium, we have $x < \min\{n, \lambda\} = \nu$. By stability of the bid-based firm in equilibrium, we have $\lambda - x < \min\{k, \lambda\}$, implying $x > \lambda - \kappa$. Thus, $x \in (\lambda - \kappa, \nu)$.

Now, in the symmetric equilibrium $\bar{c} = (c_1, c_2, 1)$, we obtain, from (3) and (4), that

$$w_F(\bar{c}) = \Phi(x)$$
$$w_B(t|\bar{c}) = \begin{cases} \Gamma(\lambda - \lambda t) & t \in [c_2, 1]; \\ \Gamma(\lambda - x - \lambda t) & t \in [0, c_1], \end{cases}$$

where we have used the fact that $x = \lambda(c_2 - c_1)$.

Recall, from Theorem 1, that the necessary conditions for $\bar{c}$ to be an equilibrium are

$$\int_0^{c_1} w_B(t|\bar{c})dt = c_1 w_F(\bar{c}) + P,$$
$$w_B(c_1|\bar{c}) = w_B(c_2|\bar{c}) = w_F(\bar{c}).$$

Using the expressions for $w_F(\bar{c})$ and $w_B(t|\bar{c})$, we obtain

$$\int_0^{c_1} \Gamma(\lambda - x - \lambda t)dt = c_1\Phi(x) + P$$
$$\Gamma(\lambda - x - \lambda c_1) = \Phi(x).$$

On substituting $u = x + \lambda t$ in the integral in the first equation and rearranging, we obtain

$$\int_x^{x+\lambda c_1} \Gamma(\lambda - u)du - \lambda c_1\Phi(x) = \lambda P$$

Note that, by definition, the left hand side is equal to $s(\lambda c_1, x)$. Thus, we obtain $s(\lambda c_1, x) = \lambda P$. Now, $x \in (\lambda - \kappa, \nu)$ and $\lambda c_1 = \lambda c_2 - x < \lambda - x$. Hence, the equation $s(\lambda c_1, x) = \lambda P$, together with Lemma 11, implies that $P < P_{\max}(\lambda)$.

Next, from the second necessary condition, $\Gamma(\lambda - \lambda c_1 - x) = \Phi(x)$, we obtain that $\Gamma(\lambda - x) > \Phi(x)$, which yields $x \in (\lambda - \kappa, \xi)$. This, along with Lemma 8, yields $z_{\max}(x) = \lambda c_1$. Hence, $s(z_{\max}(x), x) = \lambda P$. By Lemma 9, we know that $s(z_{\max}(t), t)$ is strictly decreasing in t over $(\lambda - \kappa, \xi)$. Further, by definition of $\digamma(P)$, we obtain $s(z_{\max}(\digamma(P)), \digamma(P)) = \lambda P$. Taken together, we obtain $x = \digamma(P)$, and hence $\Psi(x, P) = x$. Moreover, we have $\lambda c_1 = z_{\max}(x) = z_{\max}(\digamma(P)) = v_1(\digamma(P), P) = v_1(x, P)$.

Step 2. Suppose for some $x \in (\lambda - \kappa, \nu)$ and $P \in (0, P_{\max}(\lambda))$, we have $\Psi(x, P) = P$. By Lemma 13, we obtain that $x = \digamma(P) \in (\lambda - \kappa, \xi)$. Let $c_1 = v_1(x, P)/\lambda > 0$, and $c_2 = v_2(x, P)/\lambda > c_1$. Note that $\lambda c_2 = v_2(x, P) < \lambda$, and hence $c_2 < 1$.

Since $x = \digamma(P)$, we obtain $z_{\max}(x) = v_1(x, P) = \lambda c_1$ and $s(\lambda c_1, x) = s(z_{\max}(x), x) = \lambda P$. Thus, we obtain

$$\int_x^{x+\lambda c_1} \Gamma(\lambda - t)dt - \lambda c_1\Phi(x) = \lambda P.$$

Now, observe that for $\bar{c} = (c_1, c_2, 1)$, we have

$$w_F(\bar{c}) = \Phi(x)$$
$$w_B(t|\bar{c}) = \begin{cases} \Gamma(\lambda - \lambda t) & t \in [c_2, 1]; \\ \Gamma(\lambda - x - \lambda t) & t \in [0, c_1]. \end{cases}$$

Substituting these expressions, and making a change of variables, yields,

$$\begin{aligned} P &= \frac{1}{\lambda}\left(\int_x^{x+\lambda c_1} \Gamma(\lambda - t)dt - \lambda c_1\Phi(x)\right) \\ &= \frac{1}{\lambda}\left(\int_0^{\lambda c_1} \Gamma(\lambda - x - t)dt - \lambda c_1\Phi(x)\right) \\ &= \int_0^{c_1} \Gamma(\lambda - x - \lambda t)dt - c_1\Phi(x) \\ &= \int_0^{c_1} w_B(t|\bar{c})dt - c_1 w_F(\bar{c}). \end{aligned}$$

Thus, we obtain,

$$\int_0^{c_1} w_B(t|\bar{c})dt = c_1 w_F(\bar{c}) + P.$$

Finally, since $\Psi(x, P) = x$, we obtain

$$\Phi(x) = \Gamma(\lambda - v_2(x, P)) = \Gamma(\lambda - x - v_1(x, P)) = \Gamma(\lambda - x - \lambda c_1).$$

This implies that $w_F(\bar{c}) = w_B(c_1|\bar{c}) = w_B(c_2|\bar{c})$. Taken together, this implies that $(c_1, c_2, 1)$ satisfies the sufficient conditions in Theorem 1 for being a symmetric equilibrium. □

Proof of Theorem 2. For $\bar{c} = (1, 1, 1)$ to be a symmetric equilibrium, a necessary condition is that the customer with unit waiting cost $c = 1$ must prefer to obtain service from the bid-based firm over the fixed-price firm. Using the expression for the total expected cost from (2), this yields

$$\int_0^1 w_B(t|\bar{c})dt \leq w_F(\bar{c}) + P.$$

Note that $w_F(\bar{c}) = 1$, whereas, from (4), we have $w_B(t|\bar{c}) = \Gamma(\lambda - \lambda t)$ for all $t \in [0, 1]$. Thus, a necessary condition for $\bar{c} = (1, 1, 1)$ to be a symmetric equilibrium is

$$P \geq \int_0^1 \Gamma(\lambda - \lambda t)dt - 1 = \frac{1}{\lambda}\left(\int_0^\lambda \Gamma(t)dt - \lambda\right) = P_{\max}(\lambda).$$

Next, suppose $P \geq P_{\max}(\lambda)$ with $\kappa = \lambda < k$. From Lemma 11, we obtain $s(z, x) < \lambda P$ for all $x \in (\lambda - \kappa, \nu)$ and $z \in [0, \lambda - x]$. (Note that $\lambda - \kappa = 0$.) This implies

$$z\Phi(x) + \lambda P > \int_x^{z+x} \Gamma(\lambda - t)dt, \quad \text{for all } x \in (0, \nu), \text{ and } z \in [0, \lambda - x].$$

Taking limits as $x \downarrow 0$, and letting $c = z/\lambda$, we obtain

$$c\Phi(0) + P \geq \int_0^c \Gamma(\lambda - \lambda t)dt, \quad \text{for all } c \in [0, 1].$$

For $\bar{c} = (1, 1, 1)$, we have $\Phi(0) = w_F(\bar{c})$, and $\Gamma(\lambda - \lambda t) = w_B(t|\bar{c})$ for all $t \in [0, 1]$. Thus, the preceding equation implies

$$cw_F(\bar{c}) + P \geq \int_0^c w_B(t|\bar{c})dt, \quad \text{for all } c \in [0, 1].$$

This implies that if all other customers follow the strategy $\bar{c}$, then it is preferable for a customer with unit waiting cost $c \in [0, 1]$ to obtain service from the bid-based firm over the fixed-price firm. Thus, $\bar{c} = (1, 1, 1)$ is a symmetric equilibrium. To obtain uniqueness, observe that if there exists another equilibrium $(c_1, c_2, 1)$ with $0 < c_1 < c_2 < 1$, then from the first statement of Theorem 10, we obtain that $P < P_{\max}(\lambda)$, which contradicts our assumption on P. □

Proof of Theorem 3. From Lemma 13, we know that for each $P \in (0, P_{\max}(\lambda))$, there exists a unique solution x^* to the equation $\Psi(x, P) = x$ with $x \in (\lambda - \kappa, \nu)$. From the second statement of Theorem 10, we obtain that there exists a symmetric equilibrium $\bar{c} = (c_1, c_2, 1)$ with $0 < c_1 < c_2 < 1$, and $\lambda c_1 = v_1(x^*, P)$.

To obtain uniqueness, observe that for an equilibrium $\bar{C} = (C_1, C_2, 1)$ with $0 < C_1 < C_2 < 1$, by the first part of Theorem 10, we obtain that $X = \lambda(C_2 - C_1) \in (\lambda - \kappa, \nu)$ satisfies $\Psi(X, P) = X$ with $\lambda C_1 = v_1(X, P)$. But since x^* is the unique solution, we have $X = x^*$, and hence $\lambda C_1 = v_1(X, P) = v_1(x^*, P) = \lambda c_1$. Finally, note that since $P < P_{\max}(\lambda)$, from Theorem 2, there cannot be an equilibrium of the form $\bar{c} = (1, 1, 1)$. This proves the uniqueness of the symmetric equilibrium. □

E Proof of Theorem 4

Let $U_\lambda = (0, 1] \cap (0, \frac{n+k}{\lambda})$. Lemma 4 and Lemma 5, together with Theorem 2 and Theorem 3, imply that there exist functions $\mathcal{C}_i : U_\lambda \to [0, 1]$ for $i = 1, 2$, such that for each $u \in U_\lambda$, we have $\mathcal{C}_1(u) \le \mathcal{C}_2(u) \le u$, and the strategy $\left(\frac{\mathcal{C}_1(u)}{u}, \frac{\mathcal{C}_2(u)}{u}, 1\right)$ is the unique symmetric equilibrium of the system $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$. Define $\mathcal{A}(u) = \mathcal{C}_2(u) - \mathcal{C}_1(u)$. We have the following lemma:

Lemma 14. *The function $\mathcal{A}(u)$ is non-decreasing and continuous in u over U_λ. Further, $u - \mathcal{A}(u)$ is non-decreasing in u over U_λ.*

Proof. First, we consider the set of values of $u \in U_\lambda$ for which $\frac{P}{u} \ge P_{\max}(\lambda u)$. Note that since $P_{\max}(\lambda)$ is non-decreasing in λ, we obtain that this set is an interval $(0, u_0]$ for some $u_0 \in U_\lambda$. Since $\left(\frac{\mathcal{C}_1(u)}{u}, \frac{\mathcal{C}_2(u)}{u}, 1\right)$ is the unique symmetric equilibrium of the system $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$, we obtain from Theorem 2, that $\mathcal{C}_1(u) = \mathcal{C}_2(u) = u$ for all $u \in (0, u_0]$, and hence $\mathcal{A}(u) = 0$ for $u \in (0, u_0]$.

Now consider $u \in U_\lambda$ with $u > u_0$. Again, applying Theorem 3 to the system $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$, we obtain that $0 < \mathcal{C}_1(u) < \mathcal{C}_2(u) < u$, and hence $\mathcal{A}(u) > 0$, for all $u \in U_\lambda$ with $u > u_0$. From the necessary condition (7) for equilibrium for this system, we obtain for all $u \in U_\lambda$ with $u > u_0$,

$$\begin{aligned} P &= \int_0^{\mathcal{C}_1(u)} \Gamma(\lambda u - \lambda \mathcal{C}_2(u) + \lambda \mathcal{C}_1(u) - \lambda t) dt - \mathcal{C}_1(u) \Phi(\lambda (\mathcal{C}_2(u) - \mathcal{C}_1(u))) \\ &= \int_0^{\mathcal{C}_1(u)} \Gamma(\lambda u - \lambda \mathcal{A}(u) - \lambda t) dt - \mathcal{C}_1(u) \Phi(\lambda \mathcal{A}(u)) \\ &= \int_{u - \mathcal{A}(u) - \mathcal{C}_1(u)}^{u - \mathcal{A}(u)} \Gamma(\lambda t) dt - \mathcal{C}_1(u) \Phi(\lambda \mathcal{A}(u)). \end{aligned} \tag{16}$$

Similarly, from the necessary condition (8), we obtain for all $u \in U_\lambda$ with $u > u_0$,

$$\Phi(\lambda \mathcal{A}(u)) = \Gamma(\lambda(u - \mathcal{C}_2(u))) = \Gamma\left(\lambda\left(u - \mathcal{A}(u) - \mathcal{C}_1(u)\right)\right). \tag{17}$$

We begin with the proof of the first statement in lemma. Suppose, for the sake of arriving at a contradiction, we have $\mathcal{A}(u_1) > \mathcal{A}(u_2)$ for $u_1, u_2 \in U_\lambda$ with $u_2 > u_1 > u_0$. Since Γ and Φ are strictly increasing, from (17), we obtain $u_1 - \mathcal{A}(u_1) - \mathcal{C}_1(u_1) > u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_2)$. Since $u_2 - \mathcal{A}(u_2) > u_1 - \mathcal{A}(u_1)$, we obtain $\mathcal{C}_1(u_2) > \mathcal{C}_1(u_1)$. From this, we have

$$\int_{u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_1)}^{u_2 - \mathcal{A}(u_2)} \Gamma(\lambda t) dt - \mathcal{C}_1(u_1) \Phi(\lambda \mathcal{A}(u_2)) > \int_{u_1 - \mathcal{A}(u_1) - \mathcal{C}_1(u_1)}^{u_1 - \mathcal{A}(u_1)} \Gamma(\lambda t) dt - \mathcal{C}_1(u_1) \Phi(\lambda \mathcal{A}(u_1)) = P,$$

which yields,

$$\int_{u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_1)}^{u_2 - \mathcal{A}(u_2)} \Gamma(\lambda t) dt > \mathcal{C}_1(u_1) \Phi(\lambda \mathcal{A}(u_2)) + P. \tag{18}$$

Now, note that since $\mathcal{C}_1(u_2) > \mathcal{C}_1(u_1)$, under the symmetric equilibrium strategy $\left(\frac{\mathcal{C}_1(u_2)}{u_2}, \frac{\mathcal{C}_2(u_2)}{u_2}, 1\right)$ for the system $\mathsf{SYS}_{\mathsf{man}}\left(\lambda u_2, \frac{P}{u_2}\right)$, the customer with unit waiting cost $\frac{\mathcal{C}_1(u_1)}{u_2}$ prefers to obtain service from the bid-based firm as opposed to the fixed-price firm. Using (2) and (4), the expected total cost of this customer in equilibrium is given by

$$\int_0^{\frac{\mathcal{C}_1(u_1)}{u_2}} \Gamma(\lambda u_2 - \lambda \mathcal{A}(u_2) - \lambda u_2 t)dt = \frac{1}{u_2}\int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)} \Gamma(\lambda t)dt.$$

On the other hand, the expected total cost of this customer if she obtains service from the fixed-price firm is given by $\frac{\mathcal{C}_1(u_1)}{u_2}\Phi(\lambda\mathcal{A}(u_2)) + \frac{P}{u_2}$. Thus, in equilibrium for the system $\mathsf{SYS}_{\mathsf{man}}\left(\lambda u_2, \frac{P}{u_2}\right)$, we obtain

$$\frac{1}{u_2}\int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)} \Gamma(\lambda t)dt \leq \frac{\mathcal{C}_1(u_1)}{u_2}\Phi(\lambda\mathcal{A}(u_2)) + \frac{P}{u_2}.$$

This contradicts (18), and hence, we must have $\mathcal{A}(u_2) \geq \mathcal{A}(u_1) > 0$ for all $u_2 > u_1 > u_0$. Since $\mathcal{A}(u) = 0$ for $u \in U_\lambda$ with $u \leq u_0$, this completes the proof of the statement $\mathcal{A}(u)$ is non-decreasing over U_λ.

Next, we show that $u - \mathcal{A}(u)$ is non-decreasing over U_λ. Since $\mathcal{A}(u) = 0$ for $u \in (0, u_0]$, the statement holds trivially over $(0, u_0]$. If $u_0 = 1$, we are done. Hence, suppose $u_0 < 1$. Then, by continuity of $P_{\max}(\cdot)$, we obtain that $u_0 P_{\max}(\lambda u_0) = P$. Using the expression for $P_{\max}(\cdot)$ from (5), we obtain

$$\begin{aligned}
P &= \frac{1}{\lambda}\int_0^{\lambda u_0} \Gamma(t)dt - u_0 \\
&= \int_0^{u_0} \Gamma(\lambda t)dt - u_0 \\
&= \int_{u_0-\mathcal{A}(u_0)-\mathcal{C}_1(u_0)}^{u_0-\mathcal{A}(u_0)} \Gamma(\lambda t)dt - \mathcal{C}_1(u_0)\Phi(\lambda\mathcal{A}(u_0)),
\end{aligned}$$

where, in the last equality we use the fact that $\mathcal{A}(u_0) = 0$, $\mathcal{C}_1(u_0) = u_0$, and $\Phi(0) = 1$. This implies that (16) holds for all $u \in U_\lambda$ with $u \geq u_0$, when $u_0 < 1$. Similarly, it is straightforward to verify that (17) also holds for all $u \in U_\lambda$ with $u \geq u_0$. Thus, for all $u_1, u_2 \in U_\lambda$ with $u_2 > u_1 \geq u_0$, we have

$$\begin{aligned}
0 &= \left(\int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_2)}^{u_2-\mathcal{A}(u_2)} \Gamma(\lambda t)dt - \mathcal{C}_1(u_2)\Phi(\lambda\mathcal{A}(u_2))\right) \\
&\quad - \left(\int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_1-\mathcal{A}(u_1)} \Gamma(\lambda t)dt - \mathcal{C}_1(u_1)\Phi(\lambda\mathcal{A}(u_1))\right) \\
&= \int_{u_1-\mathcal{A}(u_1)}^{u_2-\mathcal{A}(u_2)} \Gamma(\lambda t)dt - \int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_2)} \Gamma(\lambda t)dt \\
&\quad - \mathcal{C}_1(u_2)\Phi(\lambda\mathcal{A}(u_2)) + \mathcal{C}_1(u_1)\Phi(\lambda\mathcal{A}(u_1)) \\
&= \int_{u_1-\mathcal{A}(u_1)}^{u_2-\mathcal{A}(u_2)} \Gamma(\lambda t)dt - \int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)} \Gamma(\lambda t)dt - \int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_2)} \Gamma(\lambda t)dt \\
&\quad - \left(\mathcal{C}_1(u_2) - \mathcal{C}_1(u_1)\right)\Phi(\lambda\mathcal{A}(u_2)) - \mathcal{C}_1(u_1)\left(\Phi(\lambda\mathcal{A}(u_2)) - \Phi(\lambda\mathcal{A}(u_1))\right).
\end{aligned}$$

This implies,

$$\int_{u_1-\mathcal{A}(u_1)}^{u_2-\mathcal{A}(u_2)} \Gamma(\lambda t)dt - \int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)} \Gamma(\lambda t)dt = \int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_2)} \Gamma(\lambda t)dt$$
$$+ \left(\mathcal{C}_1(u_2) - \mathcal{C}_1(u_1)\right)\Phi(\lambda\mathcal{A}(u_2))$$
$$+ \mathcal{C}_1(u_1)\left(\Phi(\lambda\mathcal{A}(u_2)) - \Phi(\lambda\mathcal{A}(u_1))\right).$$

After some algebra and rearranging, we obtain

$$\int_{u_1-\mathcal{A}(u_1)}^{u_2-\mathcal{A}(u_2)} \left(\Gamma(\lambda t) - \Gamma(\lambda t - \mathcal{C}_1(u_1))\right)dt = \int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_2)} \Gamma(\lambda t) - \Phi(\lambda\mathcal{A}(u_2))dt$$
$$+ \mathcal{C}_1(u_2)\left(\Phi(\lambda\mathcal{A}(u_2)) - \Phi(\lambda\mathcal{A}(u_1))\right).$$

Now note that since $\Gamma(\cdot)$ is increasing, from (17), we obtain that $\Gamma(\lambda t) > \Phi(\lambda\mathcal{A}(u_2))$ for $t > u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_2)$, and $\Gamma(\lambda t) < \Phi(\lambda\mathcal{A}(u_2))$ for $t < u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_2)$. Thus the integral on the right hand side is non-negative. Further, since $\mathcal{A}(u)$ is non-decreasing and Φ is strictly increasing, the second term on the right hand side is also non-negative. This implies that the left hand side is non-negative. Since $\Gamma(\lambda t) - \Gamma(\lambda t - \mathcal{C}_1(u_1)) \geq 0$, this implies that $u_2 - \mathcal{A}(u_2) \geq u_1 - \mathcal{A}(u_1)$ for $u_2 > u_1 \geq u_0$. Hence $u - \mathcal{A}(u)$ is non-decreasing over all $u \in U_\lambda$ with $u \geq u_0$. Since $\mathcal{A}(u) = 0$ for $u \in U_\lambda$ with $u \leq u_0$, the statement extends to all $u \in (0, 1]$.

Finally, we show that $\mathcal{A}(u)$ is continuous over $u \in U_\lambda$. For each $u \in U_\lambda$ with $u \geq u_0$, we obtain that $\mathcal{A}(u)$ and $\mathcal{C}_1(u)$ satisfy (16) and (17). By continuity of both sides of these equations in u, we obtain that for a sequence $u_n \to u_\infty \in U_\lambda$ with $u_\infty \geq u_0$, the limits $\lim_{n\to\infty} \mathcal{A}(u_n)$ and $\lim_{n\to\infty} \mathcal{C}_1(u_n)$ (along a subsequence if necessary for existence of the limits) also satisfy the same equations for $u = u_\infty$. However, since (16) and (17) also constitute the sufficient conditions for equilibrium (from Theorem 1), we obtain that $\mathcal{A}(u_\infty) = \lim_{n\to\infty} \mathcal{A}(u_n)$ and $\mathcal{C}_1(u_\infty) = \lim_{n\to\infty} \mathcal{C}_1(u_n)$. This implies that $\mathcal{A}(u)$ is continuous over $u \in U_\lambda$ with $u \geq u_0$. Observe that $\mathcal{A}(u) = 0$ for $u \in (0, u_0]$. Taken together, this implies that $\mathcal{A}(u)$ is continuous over $u \in U_\lambda$. □

Proof of Theorem 4. From Lemma 4, we obtain that to show the existence of a symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$, it suffices to show that there exists a $u \in U_\lambda$ such that the condition (10) holds for the strategy $(\mathcal{C}_1(u), \mathcal{C}_2(u), u)$.

For $u \in U_\lambda$, observe that the total expected cost of a customer with unit waiting cost u under the strategy $\bar{\mathcal{C}}(u) = (\mathcal{C}_1(u), \mathcal{C}_2(u), u)$ is given by

$$TC(u) = \int_0^{\mathcal{C}_1(u)} \Gamma(\lambda u - \lambda\mathcal{A}(u) - \lambda t)dt + \Phi(\lambda\mathcal{A}(u))\left(\mathcal{C}_2(u) - \mathcal{C}_1(u)\right) + \int_{\mathcal{C}_2(u)}^{u} \Gamma(\lambda u - \lambda t)dt$$
$$= \int_0^{u-\mathcal{A}(u)} \Gamma(\lambda t)dt + \mathcal{A}(u)\Phi(\lambda\mathcal{A}(u)).$$

Now, from Lemma 14, we obtain that both $\mathcal{A}(u)$ and $u - \mathcal{A}(u)$ are non-decreasing and continuous over U_λ. Since one of these two functions must strictly increase at any u, we obtain that $TC(u)$ is strictly increasing and continuous over $u \in U_\lambda$. Thus, for all $V \geq 0$, there exists a unique $u = u(\lambda, P, V) \in U_\lambda$ such that either $TC(u) = V$ with $u \leq 1$, or $TC(u) \leq V$ and $u = 1$. Note that this is exactly the condition (10) for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$. Thus, we obtain that the strategy $\bar{\mathcal{C}}(u)$ for $u = u(\lambda, P, V)$ satisfies the (necessary and) sufficient equilibrium conditions for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$, and hence constitutes the unique symmetric equilibrium for the system.

Define $\Delta_\lambda(u) = \int_0^u \Gamma(\lambda t)dt$. Observe that Δ_λ is a strictly increasing function in u over $(0,1] \cap (0, k/\lambda)$. Let $u_\lambda = u_\lambda(V) \triangleq \min\{\Delta_\lambda^{-1}(V), 1\} > 0$ for $V > 0$ and $P(\lambda, V) = \Delta_\lambda(u_\lambda(V)) - u_\lambda(V) = u_\lambda(V)P_{\max}(\lambda u_\lambda(V))$. Now, if $P \geq P(\lambda, V) = u_\lambda P_{\max}(\lambda u_\lambda)$, then from Theorem 2, we obtain that $(1,1,1)$ is the unique symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{man}}(\lambda u_\lambda, \frac{P}{u_\lambda})$. Observe that $u_\lambda \leq \Delta^{-1}(V) \leq k/\lambda$, and hence $\lambda u_\lambda < k < n+k$. Furthermore, using the definition of u_λ, it is straightforward to verify that for this strategy, we have $\Delta_\lambda(u_\lambda) = TC(u_\lambda)$ and that the condition (10) also holds. Hence, from Lemma 4, we obtain that $(u_\lambda, u_\lambda, u_\lambda)$ constitutes the unique symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$ for $P \geq P(\lambda, V)$. This implies that the arrival rate of the customers to the fixed-price firm is zero.

Conversely, if (u,u,u) is a symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$, then we have $\Delta_\lambda(u) = TC(u)$ for this strategy, and condition (10) implies $u = u_\lambda$. Also, from Lemma 5, we obtain that $(1,1,1)$ is a symmetric equilibrium for the system $\mathsf{SYS}_{\mathsf{man}}(\lambda u, \frac{P}{u})$, and hence from Theorem 2, we obtain $P \geq uP_{\max}(\lambda u) = P(\lambda, V)$. $\square$

F Proofs in Section 5

Proof of Theorem 6. Since the strategy space is compact, we only need to show that the revenue of each operator is continuous in both P and r. Then the existence of a mixed strategy equilibrium follows from Glicksberg's theorem (Glicksberg 1952).

To show the continuity of firms' revenue, observe that both firms' revenue is continuous in the equilibrium thresholds $\bar{c} = (c_1, c_2, c_\ell)$. Thus it suffices to show that $\bar{c}$ is continuous in P and r. We show this for $P > r$, and the case with $P \leq r$ can be shown using the same argument.

By the discussion in subsection 4.3, a system with positive r can be mapped to a system with $r = 0$, fixed price $P - r$, and value $V - r$. Thus, it suffices to show that in any system with zero reserve price, $\bar{c}$ is continuous in P and V.

First, consider $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$. Note that by Lemma 9, $s(z_{\max}(x), x)$ is continuous and strictly decreasing, thus $\digamma(P) = \sup\{\lambda - \kappa < x < \nu : s(z_{\max}(x), x) > \lambda P\}$ is continuous in P. Then by Lemma 10, α in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ is given by $\digamma(P)$ for $P < P_{\max}(\lambda)$, and zero otherwise. Thus α is continuous in P in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, for fixed λ. It then follows from condition (8) that c_1 is continuous in P in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$.

Now consider $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$. Recall from the proof of Theorem 4, $TC(u)$, which is the total expected cost of a customer with unit waiting cost u under the strategy $\bar{\mathcal{C}}(u) = (\mathcal{C}_1(u), \mathcal{C}_2(u), u)$, is continuous and strictly increasing in u. Thus, $u(\lambda, P, V)$ that satisfies condition (10) ($TC(u) = V$ with $u \leq 1$, or $TC(u) \leq V$ with $u = 1$) is continuous in V for fixed λ and P. On the other hand, since c_1 and α are continuous in P in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, it follows that $\mathcal{A}(u)$ is continuous in P for fixed u. Thus, $TC(u)$ is continuous in P. It then follows from monotonicity of $TC(u)$ in u that $u(\lambda, P, V)$ that satisfies condition (10) is continuous in both P and V. The result then follows immediately from the observation that $\mathcal{C}_1(u), \mathcal{C}_2(u)$ are continuous in both u and P. $\square$

The rest of the results in this section use the assumption on the expected waiting time expressions, under which we can simplify the conditions (7), (8), and (10). In particular, with the expected waiting time expressions and under the assumption that $r = 0$, the conditions (7) and (8) become

$$\frac{1}{1-\rho_B} - \frac{1}{1-\rho_B+\rho_B\frac{c_1}{1-\alpha}} - \frac{1}{q_B}\frac{c_1}{(1-\rho_B+\rho_B\frac{c_1}{1-\alpha})^2} = \lambda P, \tag{19}$$

$$\frac{1}{q_B(1-\rho_B+\rho_B\frac{c_1}{1-\alpha})^2} - \frac{1}{q_B} = \frac{\rho_F}{1-\rho_F}\frac{1}{q_F}, \tag{20}$$

where $\rho_B = (1-\alpha)/(q_B)$, and $\rho_F = \alpha/(q_F)$. Condition (6) can be obtained similarly. And the expected total cost of the customer with unit waiting cost c_ℓ is given by

$$\int_0^{c_\ell} w(t;\bar{c})dt = \frac{1}{\lambda}\left(\frac{1}{1-\frac{c_\ell-\alpha}{q_B}}-1\right) + c_\ell\left(1-\frac{1}{k}\right) + \frac{\alpha}{k} + \left(\frac{1}{1-\rho_F}-1\right)\frac{\alpha}{n}, \tag{21}$$

which is used in condition (10).

Proof of Lemma 6. We prove the first case with $q_F < 1, q_B < 1$ and $q_F + q_B > 1$ here. The other cases can be shown using the same analysis.

We start by computing the limiting thresholds in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ as λ approaches infinity, when $r = 0$. Note that as λ approaches infinity, the right hand side of condition (19) tends to infinity, thus the left hand side also tends to infinity. As a result, $\frac{1}{1-\rho_B}$ tends to infinity and thus α^λ tends to $1 - q_B$, as λ approaches infinity. Substituting the limiting value of α^λ into (20), we obtain that c_1^λ tends to $\sqrt{q_B/\left(\frac{1}{q_F+q_B-1}-\frac{1}{q_F}+\frac{1}{q_B}\right)}$ as λ approaches infinity. Thus, using equation (21) with $c_\ell = 1$, the expected total cost of the customer with unit waiting cost 1 is given by $P + 1$.

Using the mapping considered in Section 4.2, we can show that in the system $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$ (with $r = 0$), the expected total cost of the customer with unit cost u when all customers follow the strategy given by $\bar{\mathcal{C}}(u) = (\mathcal{C}_1(u), \mathcal{C}_2(u), u)$ is $P + c_\ell$. Thus, we obtain from condition (10) that $c_\ell^\infty(P, r) = \min\{V - P, 1\}$. It immediately follows from the mapping that c_1 tends to $c_1^\infty(P, r) = \sqrt{q_B/\left(\frac{1}{q_F+q_B-c_\ell^\infty(P,r)}-\frac{1}{q_F}+\frac{1}{q_B}\right)}$, and α tends to $\alpha^\infty(P, r) = c_\ell^\infty(P, r) - q_B$, as λ approaches infinity.

The limiting revenue per arrival of the fixed-price firm can be obtained immediately by multiplying P and the proportion of the arrivals obtaining service in the fixed-price firm. Thus $R_F^\infty(P, r) = (c_\ell^\infty(P, r) - q_B)\,P$. To compute the limiting revenue per arrival of the bid-based firm, we consider the expected total revenue of the two firms in the $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, which is

$$\begin{aligned}
&\int_0^1 \left[\int_0^c w(t)dt\right] dc - \int_0^1 cw(c)dc \\
&= -2\alpha^\lambda P + \frac{1}{\lambda}\left[\frac{\alpha^\lambda(1-\alpha^\lambda)}{q_B(1-\frac{1-\alpha^\lambda}{q_B}+\frac{c_1^\lambda}{q_B})^2} + 1 + \frac{1+2\alpha^\lambda}{1-\frac{1-\alpha^\lambda}{q_B}} + 2q_B\log(1-\frac{1-\alpha^\lambda}{q_B})\right]
\end{aligned} \tag{22}$$

Note that by the limit of α and c_1, $\frac{\alpha^\lambda(1-\alpha^\lambda)}{q_B\left(1-\frac{1-\alpha^\lambda}{q_B}\right)} + 1$ is finite in the limit, and thus the limit of $\frac{1}{\lambda}\left[\frac{\alpha^\lambda(1-\alpha^\lambda)}{q_B(1-\frac{1-\alpha^\lambda}{q_B}+\frac{c_1^\lambda}{q_B})^2}+1\right]$ is zero. On the other hand, by condition (19), we have

$$\frac{1}{\lambda}\frac{1}{1-\frac{1-\alpha^\lambda}{q_B}} = P + \frac{1}{\lambda}\left[\frac{1}{1-\rho_B+\rho_B\frac{c_1^\lambda}{1-\alpha^\lambda}} + \frac{1}{q_B}\frac{c_1^\lambda}{(1-\rho_B+\rho_B\frac{c_1^\lambda}{1-\alpha^\lambda})^2}\right],$$

which tends to P in the limit, as the term in the square bracket has a finite limit. And since $\rho_B \to 1$, the limit of $(1-\rho_B)\log(1-\rho_B)$ is zero.

Therefore, the limit of (22) is P, and in $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$, the limiting revenue of the bid-based firm is $q_B P$. By the mapping considered in Section 4.2, its limiting revenue in $\mathsf{SYS}_{\mathsf{op}}(\lambda, P, V)$ (with $r = 0$) is also $q_B P$.

Thus we proved the first case with $q_F < 1, q_B < 1$ and $q_F + q_B > 1$, when $r = 0$. The results when $r > 0$

then immediately follows from the mapping discussed in subsection 4.3. The other cases can be shown using the same techniques. □

Proof of Theorem 7. We only need to show that the firms' revenue is continuous in λ. This proof is analogous to the proof of Lemma 6, so we only briefly outline the argument here.

As shown in the proof of Lemma 6, the thresholds c_1, c_2 that solves $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ is continuous in P. Also observe that conditions (19) and (20) only depend on λ and P through λP, thus the thresholds that solves $\mathsf{SYS}_{\mathsf{man}}(\lambda, P)$ is also continuous in λ. Moreover, from equation (21) we observe that the expected total cost of the customer with unit waiting cost u when all customers follow the strategy $\bar{\mathcal{C}}(u)$ is continuous and strictly decreasing in λ. Recall that it is also continuous and strictly monotone in u. Thus, $u(\lambda, P, V)$ that satisfies condition (10) is continuous in λ. So far we have shown that the equilibrium thresholds are continuous in λ, and thus the revenues are also continuous in λ. □

Proof of Lemma 7. We focus on the first case when $q_B < 1$ and $q_F < 1$. The other cases can be shown using the same technique.

Lemma 6 gives the payoffs in the limiting game R_F^∞ and R_B^∞ as functions of (P, r), from which we can calculate the best response functions $P^{BR}(r)$ and $r^{BR}(P)$.

The best response of the fixed-price operator to the reserve price set by the bid-based firm is given by

$$P^{BR}(r) = \begin{cases} V-1 & \text{if } r \le (V-1)\frac{1-q_B}{q_F} \\ r & \text{if } r > (V-1)\frac{1-q_B}{q_F}. \end{cases}$$

The best response of the bid-based firm to the fixed-price firm's fixed price is given by

$$r^{BR}(P) \in \begin{cases} V-1 & \text{if } P \le (V-1)\frac{1-q_F}{q_B} \\ [0, P) & \text{if } P > (V-1)\frac{1-q_F}{q_B}. \end{cases}$$

Finding the fixed point of the best response functions, we prove the first case of this lemma when $q_B < 1$ and $q_F < 1$. The other cases can be shown using the same approach. □

Proof of Theorem 8. We focus on the case when $q_B < 1, q_F < 1$ and $q_B + q_F \ge 1$.

We first consider the optimization problem under collusion. The firms choose (P, r) to maximize the total payoff, which is given by

$$R_F^\infty(P,r) + R_B^\infty(P,r) = \begin{cases} P & \text{if } P \le V-1, r < P \\ (V-P)P & \text{if } V > P \ge V-1, r < P \\ Pq_F + r(1-q_F) & \text{if } P \le r \le V-1 \\ Pq_F + r(V-r-q_F) & \text{if } r > V-1, r \ge P \\ Pq_F & \text{if } r > V-q_F > P, r \ge P \\ P(V-P) & \text{if } V-q_F < P, r \ge P. \end{cases}$$

Maximizing the quantity above, we get the policy under collusion given by

$$(P, r) \in \begin{cases} \{(P,r) : P = V-1, r \le P\} & \text{if } V \ge 2 \\ \{(P,r) : P = V/2, r \le P\} & \text{if } V < 2. \end{cases}$$

And the total payoff under collusion is $V-1$ if $V \geq 2$, and $V^2/4$ if $V < 2$.

By Lemmas 6 and 7, the total payoff in Nash equilibrium of the limiting game is $V-1$ if $V \geq 2-q_B$, and $\frac{V^2-q_B^2}{4}$ if $V < 2-q_B$. Thus, the price of stability is given by

$$\begin{cases} \frac{V-1}{V-1} = 1 & \text{if } V \geq 2 \\ \frac{V^2/4}{V-1} & \text{if } 2-q_B \leq V < 2 \\ \frac{V^2/4}{(V^2-q_B^2)/4} = \frac{1}{1-q_B^2/V^2} & \text{if } V < 2-q_B. \end{cases}$$

It is easy to verify that the price of stability is non-increasing in V. Moreover, when $q_B \geq 0.5$, the price of stability at $V = 1.5$ is given by $V^2/4(V-1) = 9/8$. When $q_B < 0.5$, the price of stability at $V = 1.5$ is $1/(1-q_B^2/V^2) < 1/(1-0.5^2/1.5^2) = 9/8$. The theorem follows from this when $q_B < 1, q_F < 1$ and $q_B + q_F \geq 1$.

The other cases can be considered using similar argument. In particular, when $q_B + q_F < 1$, we obtain that the price of stability is given by

$$\begin{cases} 1 & \text{if } V \geq 2(q_B+q_F) \\ \frac{V^2/4}{(V-q_B-q_F)(q_B+q_F)} & \text{if } 2q_F+q_B \leq V < 2(q_B+q_F) \\ \frac{1}{1-q_B^2/V^2} & \text{if } V < 2q_F+q_B. \end{cases}$$

When $2q_F + q_B > 1.5$, the price of stability at $V = 1.5$ is $1/(1-q_B^2/V^2) < 1/(1-0.5^2/1.5^2) = 9/8$. When $2q_F + q_B < 1.5 < 2q_F + 2q_B$, the price of stability at $V = 1.5$ is $1.5^2/4(1.5-q_B-q_F)(q_B+q_F) < 9/8$. When $2q_F + 2q_B < 1.5$, the price of stability is 1. Thus, the price of stability at $V = 1.5$ is at most $9/8$. The rest of the results follow by similar argument as the previous case. □

Proof of Theorem 9. We can calculate the best response functions $q_F^{BR}(q_B)$ and $q_B^{BR}(q_F)$. In particular, the fixed-price firm's best response to q_B is given by

$$q_F^{BR}(q_B) = \begin{cases} \frac{V-q_B-\beta_F}{2} & \text{if } 0 \leq q_B \leq V-\beta_F \text{ and } \beta_F \geq V-1; \\ \frac{V-q_B-\beta_F}{2} & \text{if } 0 \leq q_B \leq 2-V+\beta_F \text{ and } \beta_F \leq V-1; \\ 1-q_B & \text{if } \max\{0, 2-V+\beta_F\} \leq q_B \leq 1 \text{ and } \beta_F \leq V-1; \\ 0 & \text{otherwise.} \end{cases}$$

The bid-based firm's best response to q_F is given by

$$q_B^{BR}(q_F) = \begin{cases} \frac{V-q_F-\beta_B}{2} & \text{if } 0 \leq q_F \leq \min\{V-\beta_B, 2-V+\beta_B, 1\} \text{ and } \beta_B > V-1; \\ 1-q_F & \text{if } \max\{2-V+\beta_B, 0\} \leq q_F \leq 1 \text{ and } \beta_B > V-1; \\ \frac{V-q_F-\beta_B}{2} & \text{if } 0 \leq q_F \leq \min\{2-V+\beta_B, \left(\sqrt{V-\beta_B-1}-1\right)^2\} \text{ and } \beta_B \leq V-1; \\ 1 & \text{if } \min\{2-V+\beta_B, \left(\sqrt{V-\beta_B-1}-1\right)^2\} < q_F \text{ and } \beta_B \leq V-1; \\ 0 & \text{otherwise.} \end{cases}$$

By a straightforward (but cumbersome) case analysis, we observe that a duopoly equilibrium arises when $q_B = (V-q_F-\beta_B)/2$ and $q_F = (V-q_B-\beta_F)/2$; imposing the corresponding conditions in the best response function then yields the parameter values for existence of a duopoly equilibrium. The other two cases of monopoly can be similarly shown by considering intersections of the preceding best response functions. We omit the details for brevity. □

Lemma 15. *1. Suppose $q_B = 0$. Then, the fixed-price firm's optimal fixed-price is given by $P^m = \max\{\frac{V}{2}, V - \min\{q_F, 1\}\}$. The revenue per arrival of the fixed-price firm is given by $P^m(V - P^m)$.*

2. Suppose $q_F = 0$. Then, the bid-based firm's optimal reserve is given by $r^m = \max\{\frac{V}{2}, V - \min\{q_B, 1\}\}$. The revenue per arrival of the bid-based firm is given by $r^m(V - r^m)$.

Proof. The proof follows from the revenue expressions in Lemma 6 through straightforward algebra, and is omitted for brevity. □

G Extension to non-preemptive queues for the bid-based firm

To extend our analysis in Sections 3 and 4 to the case where the bid-based firm operates a non-preemptive queue, we need to change our assumptions of $\Gamma(\cdot)$, so that it not only depends on the arrival rate of the customers with higher bids, but also depends on the arrival rate of all customers to the bid-based queue. More specifically, we let $\Gamma_y(x)$ to be the expected waiting time in the bid-based firm when the arrival rate of the customers with higher bids is x, and the arrival rate of all customers to the bid-based firm is y. Similar to our previous assumptions, we assume that $\Gamma_y(x)$ is continuous and strictly increasing in both x and y for $x \leq y < k$, and that $\Gamma_y(0) > 1$. It is straightforward to verify that any $G/M/k$ non-preemptive queues satisfy these assumptions.

In Section 3: Almost all analysis in Section 3 follow in this non-preemptive case. The only change is that in Theorem 1, the strict inequality in $0 < c_1 < c_2 < c_\ell \leq 1$ has to be changed to a less than or equal to as in $0 < c_1 < c_2 \leq c_\ell \leq 1$. In other words, it is possible that in equilibrium, the fixed-price firm is non-empty yet there is no customer bidding higher than P in the bid-based firm. This is because when queues are non-preemptive, $\Gamma_y(0) > 1$, and thus there exists some ϵ_0 such that $\Gamma_y(0) = \Phi(\epsilon_0)$. Thus there may exist an equilibrium in which the fixed-price firm has arrival rate less than ϵ_0 and $c_2 = c_\ell$.

In Section 4: Multiple changes need to be made to Section 4. The idea is to add appropriate subscripts to all occurrence of $\Gamma(\cdot)$ to represent the total arrival rate to the bid-based firm. Some proofs need to be carefully modified, especially when there are inequalities comparing the Γ functions under different total arrival rates. Moreover, the possibility of an equilibrium in which $0 < c_1 < c_2 = c_\ell \leq 1$ as discussed above leads to modification of condition (20), as in such equilibrium the expected waiting time of the highest priority customer in the bid-based firm can be strictly larger than the expected waiting time in fixed-price firm. We discuss the changes in the main text and the appendix separately, as follows.

In the main text of Section 4: In equation (4) which connects $w_B(\cdot)$ with $\Gamma(\cdot)$, the subscript $\lambda - \alpha\lambda$ needs to be added to both occurrences. In equation (5) where we define $P_{\max}(\lambda)$, the subscript λ needs to be added to Γ. In Theorem 3, we need to change $0 < c_1 < c_2 < 1$ to $0 < c_1 < c_2 \leq 1$. In equation (9), the subscript $\lambda - \alpha\lambda$ needs to be added to Γ. And in all occurrences of Γ in subsection 4.2, the subscript $\lambda u - \mathcal{A}(u)$ needs to be added.

Moreover, condition (20) needs to be changed to

$$c_2 < c_\ell, \quad w_B(c_2|\bar{c}) = w_B(c_1|\bar{c}) = w_F(\bar{c}), \qquad \text{OR} \qquad c_2 = c_\ell, \quad w_B(c_1|\bar{c}) \geq w_F(\bar{c}).$$

In Appendix C: Almost all analysis in Appendix C follow without modification. We only need to add appropriate subscripts to the equation in the proof of Lemma 4 involving Γ, and it becomes

$$\int_0^{c_1 u} w_B(t|\bar{c}(u))dt = \int_0^{c_1 u} \Gamma_{\frac{\lambda}{u}u - \frac{\lambda}{u}(c_2 u - c_1 u)} \left(\frac{\lambda}{u}u - \frac{\lambda}{u}(c_2 u - c_1 u) - \frac{\lambda}{u}t\right) dt$$
$$= u \int_0^{c_1} \Gamma_{\lambda - \lambda(c_2 - c_1)}(\lambda - \lambda(c_2 - c_1) - \lambda t)dt = u \int_0^{c_1} w_B(t|\bar{c})dt.$$

In Appendix D: First, define η as

$$\eta \triangleq \sup\{\lambda - \kappa < x < \nu : \Gamma_{\lambda - x}(0) > \Phi(x)\}.$$

Add subscript $\lambda - x$ to Γ in the definition of ξ and $s(z, x)$.

The statement in Lemma 8 becomes *"The function $z_{\max}(x)$ is continuous over $x \in (\lambda - \kappa, \nu)$. Further, $z_{\max}(\cdot) = \lambda - x$ for all $x \in (\lambda - \kappa, \eta)$, $z_{\max}(\cdot)$ is strictly decreasing over (η, ξ) and $z_{\max}(x) = 0$ for all $x \in [\xi, \nu)$. Moreover, $z_{\max}(x) = \lambda - x - \Gamma^{-1}_{\lambda - x}(\Phi(x)) \in (0, \lambda - x)$ for $x \in (\eta, \xi)$, and $z_{\max}(\eta) = \lim_{x \downarrow \eta} z_{\max}(x) = \lambda - \eta$."* The proof can be done using the same argument as the original proof.

In the proof of Lemma 9, the derivative needs to be changed to

$$\frac{ds(z_{\max}(x), x)}{dx} = \Gamma_{\lambda - x}(\lambda - z_{\max}(x) - x) - \Gamma_{\lambda - x}(\lambda - x) - z_{\max}(x)\Phi'(x) + \int_x^{z_{\max}(x) + x} \frac{\partial \Gamma_{\lambda - x}(\lambda - t)}{\partial x} dt.$$

By monotonicity of Γ, $\frac{\partial \Gamma_{\lambda - x}(\lambda - t)}{\partial x} < 0$. Thus the original result still follows.

Lemma 10 and its proof hold after adding subscript λ to Γ in the first two displayed equations, and adding subscript $\lambda - x$ to Γ in the last displayed equation.

Lemmas 11, 12, Theorem 10 and their proofs hold after adding subscript $\lambda - x$ to all occurrences of Γ.

In the proof of Lemma 13, add subscript $\lambda - \digamma(P)$. More importantly, note that $\Gamma^{-1}_{\lambda - \digamma(P)}(\Phi(\digamma(P)))$ may not be well defined when $\Phi(\digamma(P)) < \Gamma_{\lambda - \digamma(P)}(0)$. In those cases, we let $v_1(\digamma(P), P) = \lambda - \digamma(P)$, and it corresponds to an equilibrium in which $c_1 < c_2 = c_\ell$. We can verify that all other statements in the proof hold after we make this change.

The proof of Theorem 2 holds after adding subscript λ.

In Appendix E: The proof of Theorem 4 holds after adding the subscript $\lambda u - \lambda \mathcal{A}(u)$ to all occurrences of Γ, and changing the definition of $\Delta_\lambda(u)$ to $\Delta_\lambda(u) = \int_0^u \Gamma_{u\lambda}(\lambda t)dt$.

The first part of the proof of Lemma 14 holds after adding the subscripts $\lambda u - \lambda \mathcal{A}(u)$, $\lambda u_1 - \lambda \mathcal{A}(u_1)$, $\lambda u_2 - \lambda \mathcal{A}(u_2)$ or λu_0 (which one to add should be clear given context). However, to show that $u - \mathcal{A}(u)$ is non-decreasing in u, we need a new proof as follows. For all $u_1, u_2 \in U_\lambda$ with $u_2 > u_1 \geq u_0$, we have

$$\begin{aligned}
0 = &\left(\int_{u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_2)}^{u_2 - \mathcal{A}(u_2)} \Gamma_{\lambda u_2 - \lambda \mathcal{A}(u_2)}(\lambda t)dt - \mathcal{C}_1(u_2)\Phi(\lambda \mathcal{A}(u_2))\right) \\
&- \left(\int_{u_1 - \mathcal{A}(u_1) - \mathcal{C}_1(u_1)}^{u_1 - \mathcal{A}(u_1)} \Gamma_{\lambda u_1 - \lambda \mathcal{A}(u_1)}(\lambda t)dt - \mathcal{C}_1(u_1)\Phi(\lambda \mathcal{A}(u_1))\right) \\
= &\int_{u_2 - \mathcal{A}(u_2) - \mathcal{C}_1(u_2)}^{u_2 - \mathcal{A}(u_2)} \Gamma_{\lambda u_2 - \lambda \mathcal{A}(u_2)}(\lambda t)dt - \int_{u_1 - \mathcal{A}(u_1) - \mathcal{C}_1(u_1)}^{u_1 - \mathcal{A}(u_1)} \Gamma_{\lambda u_1 - \lambda \mathcal{A}(u_1)}(\lambda t)dt \\
&- \left(\mathcal{C}_1(u_2) - \mathcal{C}_1(u_1)\right)\Phi(\lambda \mathcal{A}(u_2)) - \mathcal{C}_1(u_1)\left(\Phi(\lambda \mathcal{A}(u_2)) - \Phi(\lambda \mathcal{A}(u_1))\right).
\end{aligned}$$

This implies,

$$\int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)} \Gamma_{\lambda u_2-\lambda\mathcal{A}(u_2)}(\lambda t)dt - \int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_1-\mathcal{A}(u_1)} \Gamma_{\lambda u_1-\lambda\mathcal{A}(u_1)}(\lambda t)dt$$
$$= \int_{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_1)}^{u_2-\mathcal{A}(u_2)-\mathcal{C}_1(u_2)} \Gamma_{\lambda u_2-\lambda\mathcal{A}(u_2)}(\lambda t)dt + (\mathcal{C}_1(u_2)-\mathcal{C}_1(u_1))\,\Phi(\lambda\mathcal{A}(u_2)) + \mathcal{C}_1(u_1)\left(\Phi(\lambda\mathcal{A}(u_2)) - \Phi(\lambda\mathcal{A}(u_1))\right).$$

Using the same argument as in the original proof, we can show that the right hand side of the equation above is non-negative. Thus, the left hand side is also non-negative. After rearranging the left hand side, we have

$$\int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_1-\mathcal{A}(u_1)} \Gamma_{\lambda u_2-\lambda\mathcal{A}(u_2)}(\lambda t + (u_2-\mathcal{A}(u_2)) - (u_1-\mathcal{A}(u_1))) - \Gamma_{\lambda u_1-\lambda\mathcal{A}(u_1)}(\lambda t)dt \geq 0.$$

Suppose $u_2 - \mathcal{A}(u_2)) < u_1 - \mathcal{A}(u_1)$, we have by monotonicity of $\Gamma_y(x)$ in both x and y, that

$$\Gamma_{\lambda u_2-\lambda\mathcal{A}(u_2)}(\lambda t + (u_2-\mathcal{A}(u_2)) - (u_1-\mathcal{A}(u_1))) < \Gamma_{\lambda u_2-\lambda\mathcal{A}(u_2)}(\lambda t) < \Gamma_{\lambda u_1-\lambda\mathcal{A}(u_1)}(\lambda t).$$

Thus, $\int_{u_1-\mathcal{A}(u_1)-\mathcal{C}_1(u_1)}^{u_1-\mathcal{A}(u_1)} \Gamma_{\lambda u_2-\lambda\mathcal{A}(u_2)}(\lambda t + (u_2-\mathcal{A}(u_2)) - (u_1-\mathcal{A}(u_1))) - \Gamma_{\lambda u_1-\lambda\mathcal{A}(u_1)}(\lambda t)dt < 0$, which contradicts with the previous inequality.

Therefore $u_2 - \mathcal{A}(u_2)) \geq u_1 - \mathcal{A}(u_1)$, as desired.

References

V. Abhishek, I. A. Kash, and P. Key. Fixed and market pricing for cloud services. *arXiv preprint arXiv:1201.5621*, 2012.

P. Afèche. Incentive-compatible revenue management in queueing systems: optimal strategic delay. *Manufacturing & Service Operations Management*, 15(3):423–443, 2013.

P. Afèche and H. Mendelson. Pricing and priority auctions in queueing systems with a generalized delay cost structure. *Management Science*, 50(7):869–882, 2004.

P. Afèche and J. M. Pavlin. Optimal price/lead-time menus for queues with customer choice: Segmentation, pooling, and strategic delay. *Management Science*, 62(8):2412–2436, 2016.

C. D. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media, 2006.

Amazon. Amazon EC2 instance purchasing options. `http://aws.amazon.com/ec2/purchasing-options`, 2015. [Online; accessed 26-June-2015].

K. Balachandran. Purchasing priorities in queues. *Management Science*, 18(5-part-1):319–326, 1972.

R. Caldentey and G. Vulcano. Online auction and list price revenue management. *Management Science*, 53 (5):795–813, 2007.

U. W. Chohan. Dididache - hailing the future, one cab at a time. `https://www.mcgill.ca/channels/channels/news/dididache-di-di-da-che-hailing-future-one-cab-time-235312`, 2014. [Online; accessed 8-May-2018].

S. Doroudi, M. Akan, M. Harchol-Balter, J. Karp, C. Borgs, and J. T. Chayes. Priority pricing in queues with a continuous distribution of customer valuations (cmu-cs-13-109). 2013.

H. Etzion, E. Pinker, and A. Seidmann. Analyzing the simultaneous use of auctions and posted prices for online selling. *Manufacturing & Service Operations Management*, 8(1):68–91, 2006.

N. Gautam. *Analysis of Queues: Methods and Applications*. CRC Press, 2012.

S. Gavirneni and V. G. Kulkarni. Self-selecting priority queues with burr distributed waiting costs. *Production and Operations Management*, 2016.

A. Glazer and R. Hassin. Stable priority purchasing in queues. *Operations Research Letters*, 4(6):285–288, 1986.

I. L. Glicksberg. A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174, 1952.

R. Hassin. Decentralized regulation of a queue. *Management Science*, 41(1):163–173, 1995.

R. Hassin and M. Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*, volume 59. Springer, 2003.

C. Kilcioglu and C. Maglaras. Revenue maximization for cloud computing services. *SIGMETRICS Perform. Eval. Rev.*, 43(3):76–76, Nov. 2015. ISSN 0163-5999. doi: 10.1145/2847220.2847245. URL `http://doi.acm.org/10.1145/2847220.2847245`.

C. Kilcioglu and J. M. Rao. Competition on price and quality in cloud computing. In *Proceedings of the 25th International Conference on World Wide Web*, pages 1123–1132. International World Wide Web Conferences Steering Committee, 2016.

T. Kittsteiner and B. Moldovanu. Priority auctions and queue disciplines that depend on processing time. *Management Science*, 51(2):236–248, 2005.

L. Kleinrock. Optimum bribing for queue position. *Operations Research*, 15(2):304–318, 1967.

V. Krishna. *Auction theory*. Academic press, 2009.

F. T. Lui. An equilibrium queuing model of bribery. *Journal of Political Economy*, pages 760–781, 1985.

C. Maglaras, J. Yao, and A. Zeevi. Optimal price and delay differentiation in queueing systems. *Available at SSRN 2297042*, 2013.

H. Mendelson and S. Whang. Optimal incentive-compatible priority pricing for the $M/M/1$ queue. *Operations Research*, 38(5):870–883, 1990.

Microsoft. Azure pricing. `https://azure.microsoft.com/en-us/pricing`, 2017. [Online; accessed 4-Apr-2017].

J. A. Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38 (2):175–208, 1971.

M. Mussa and S. Rosen. Monopoly and product quality. *Journal of Economic theory*, 18(2):301–317, 1978.

R. B. Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

H. Nazerzadeh and R. S. Randhawa. Asymptotic optimality of two service grades for customer differentiation in queueing systems. *Available at SSRN 2438300*, 2014.

N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic game theory*, volume 1. Cambridge University Press Cambridge, 2007.

X. Palmer. [How to]: Call a cab with didi chuxing and kuaidi dache. `http://www.smartshanghai.com/articles/smsh/how-to-call-a-taxi-with-didi-chuxing-and-kuaidi-dache`, 2015. [Online; accessed 4-Apr-2017].

R. T. Rockafellar. *Convex analysis*. Princeton university press, 1997.

Springwise. In china, cab passengers bid tips to secure a float home. `https://www.springwise.com/china-taxi-passengers-bid-tips-secure-ride-home/`, 2013. [Online; accessed 4-Apr-2017].

M. Tian. Didi dache-kuaidi dache wants to do more than just drive you home. `http://knowledge.ckgsb.edu.cn/2015/05/12/china-business-strategy/didi-dache-and-kuaidi-dache-to-do-more-than-just-drive-you-home/`, 2015. [Online; accessed 4-Apr-2017].

J. A. Van Mieghem. Price and service discrimination in queuing systems: Incentive compatibility of $gc\mu$ scheduling. *Management Science*, 46(9):1249–1267, 2000.

W. Wang, B. Li, and B. Liang. Towards optimal capacity segmentation with hybrid cloud pricing. In *2012 IEEE 32nd International Conference on Distributed Computing Systems (ICDCS)*, pages 425–434. IEEE, 2012.

T. Yahalom, J. M. Harrison, and S. Kumar. Designing and pricing incentive compatible grades of service in queueing systems. Stanford University Working paper, 2006.

L. Yang, L. Debo, and V. Gupta. Trading time in a congested environment. *Chicago Booth Research Paper*, (15-07), 2015.