Prices and Congestion as Signals of Quality

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

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

Successful firms' demand often exceeds capacity, generating congestion. Given innovative, high-quality products' ramp-up costs from technological constraints or limited labor supply, a question arises: Why do successful firms not raise prices to increase revenues given excess demand? For new products and experience goods, communicating quality to consumers is challenging, and congestion may serve as a quality indicator. We develop an Erlang-loss model: A firm signals privately observed quality via price to consumers uninformed about its quality. In addition, consumers observe congestion (occupancy level) upon arrival. Separating and pooling equilibria exist in prices: When prices are separating, congestion provides no additional information. When prices are pooling, congestion is informative. We demonstrate an *empty-restaurant syndrome* with pooling prices: Low congestion makes arriving consumers suspicious about quality; hence, they join less often. The pooling-price range is lower than the separating price, and both high- and low-quality firms profit more in the price-pooling equilibrium. This offers high-quality firms a rationale to refrain from signaling quality through price, depending on consumers to learn quality through congestion. When high-quality firms also have cost advantages for ramping up, pooling prices remain an equilibrium, but the empty-restaurant syndrome disappears.

Key words: Signaling, prices, congestion, intuitive criterion, consumer behavior and queues.

---

1. Introduction and Motivation

For many new, innovative products or services, ramping up capacity is often prohibitively expensive because of a limited supply of skilled labor or technological constraints due to innovative product specifications. Becker (1991) noted that many successful firms have demand in excess of their capacity, which leads to high levels of congestion. A natural question arises in this context: Why does the firm not increase its price? A judiciously selected price increase might reduce only the congestion without affecting the satisfied demand; thus, the firm could still improve revenues (accrued through the higher margin from sales). In addition, a high price can also be used as a signal of quality because with such new or innovative products or services, it is often difficult for firms to communicate the quality to consumers, due to complex features of the product or the experiential nature of the service. In this paper, we study the joint role that prices and congestion play as signals of quality.

The firm under consideration in our paper has N slots to produce a product or provide a service. A slot could be interpreted as a plant or work bench in the manufacturing context, or as a seat or a table at a restaurant in the service context. Each slot has one server (e.g., a machine or a waiter in the above contexts) that takes some stochastically distributed time to produce the product or serve the customer. For simplicity, we assume that each slot can hold one consumer (order) at a time. If all N slots are full, an arriving consumer that is willing to consume the service (or place an order) is lost. Thus, our model of the firm's service is an Erlang-loss system.

The quality of the firm's good is an unknown variable, chosen by nature to be high or low. Based on the quality realization, the firm sets a price for the good and selects a service rate, which can be *fast* or *slow*. Obviously, consumers observe the price. The service rate of a process often cannot be easily communicated to potential consumers. To estimate the service rate, an arriving consumer would have to spend time inspecting the service or production process.[1] However, she can observe how many slots are occupied with a cursory glance or tour. Hence, we model the fact that the service speed is *not* observed by the consumers; they only observe how many slots are currently 'busy' upon their arrival, which we term the *occupancy level* or the *congestion*. Consumers make their decision as to whether to buy the good or not based on the price *and* congestion observed upon arrival.

Fast service is generally more expensive than slow service (whose cost is normalized to zero without loss of generality). In our base model, the fast service is more expensive for the high-quality firm than for the low-quality firm, as delicate high-quality operations need to be performed in order to deliver superior value. This case resembles a firm whose process improvements lag behind its product innovation: the firm first develops a product with a superior value compared to the existing products.[2] In an extension of our model, we study a case in which the cost of speeding up is lower for the high-quality firm than for the low-quality firm. In such a case, the high-quality firm innovates and improves on process efficiency simultaneously.

Our paper addresses the following questions for our base model and its extension: When will firms signal quality via prices? Which service rates will be chosen? How do consumers infer quality from either prices or congestion level? We now position our paper with respect to the literature.

[1] For example, a consumer can observe the number of occupied tables at a restaurant by looking into the restaurant through the window, but she cannot easily obtain information related to the service rate (such as the number of employees at the restaurant, the size of the back-room operations etc.). Sometimes a service provider announces the expected waiting time. However, such communication need not be credible (see Allon et al. 2007). Instead, we focus on the service rate selection as the main decision variable of the service firm. In Allon et al. (2007), the service rates remain observable.

[2] Typically, firms first innovate their products and produce them with relatively generic machinery/processes, which is expensive to speed up compared to dedicated production lines for the more mature 'regular' or 'low-quality' products.

1.1. Our Position in the Libterature

The goal of our paper is to model a firm's equilibrium pricing and service rate strategy in cohesion. We believe that our modeling framework is innovative and contributes to three different literature streams.

In the operations management literature related to queuing games, there is very little work considering queues with unknown service rates. In almost all papers cited in the extensive overview by Hassin and Haviv (2004), beginning with the seminal papers by Naor (1969) and Edelson and Hildebrand (1975), the service rates are generally assumed to be common knowledge (even when the queues are deemed unobservable). This assumption aids tractability in analysis, but may not be realistic. There are some recent exceptions (such as Guo and Hassin 2009; Veeraraghavan and Debo 2009) that allow service rates to be unknown. However, there is no signaling in the above papers. Additionally, the vast majority of papers do not consider pricing. However, some papers have considered quality and informational uncertainty about the waiting time (e.g., Guo and Zipkin 2007). A paper closely related to ours is Debo et al. (2010) in which the consumer population is heterogeneously informed about the product quality: A fraction of the consumers knows the quality of the firm perfectly and the remaining fraction knows only the quality distribution. However, unlike our paper, they do not consider prices as potential signaling instruments.

To the best of our knowledge, how firms can manipulate service rates *and* prices to influence the consumers' perception about the quality of the firm has not yet been studied in the queuing literature.

Second, our paper introduces a new signaling tool that, to the best of our knowledge, has not yet been studied in the economics literature: the congestion level consumers observe when deciding to buy the firm's good. Signaling literature in economics is extensive, beginning with the seminal paper by Spence (1973). Milgrom and Roberts (1986) show that high prices and intense advertising levels signal product quality. Bagwell and Riordan (1991) show that high and declining prices may signal product quality. Several other instruments have been considered to be signals of quality in the extant literature—for instance, time on the market (Taylor 1999), modest advertising (Orzach et al. 2002), and money-back guarantees (Moorthy and Srinivasan 1995). Nevertheless, congestion has not been analyzed as a signaling tool. This is possibly because congestion levels arise endogenously from stochastic service- or production-rate decisions and consumer arrival rates (that are, in turn, dependent on those decisions), which complicates the signaling-equilibria deduction.

Finally, our explanation about why high-quality firms may refrain from increasing prices to reduce excess demand is new. A well-accepted explanation advanced by Becker (1991) asserts that the excess demand is a consequence of social interactions among consumers. Consumers enjoy a higher utility when the aggregate demand for the product is higher. Our explanation does not assume such direct positive consumption externalities. Instead, in our model, positive externalities are an artifact of consumers' imperfect information about the utility they can obtain from buying the product. In other words, in our model, it is not inherently better to eat at a crowded establishment. Rather, better establishments are more likely to be crowded because of their better food and slower service.

1.2. Preview of Our Approach and Insights

It is well known that signaling games generally have a multitude of equilibria (for example, see the discussion in Cachon and Lariviere 2001, p. 642). To resolve this issue, our paper employs a refinement, based on Cho and Kreps' *intuitive criterion*, that restricts the consumers' beliefs about the quality of the firm off the equilibrium path, where Bayes' rule cannot be applied. The reader is referred to Cho and Kreps (1987) for a detailed development of the intuitive criterion. Although such refinement criteria are *de rigueur* in economics, they have been sparsely employed in operations management literature (for notable exceptions, see Akan et al. 2007; Anand and Goyal 2009). For our specific analysis, we are interested in separating (where firm prices are different) and pooling (when firm prices are identical) equilibria that survive the intuitive criterion. Our main insights follow.

- Under separating equilibria, the prices perfectly signal the quality of the firm's product or service. For our base model, we show that to separate using prices is very expensive for the high-quality firm and this results in a valuable signal being unused: When prices are indicators of quality, higher levels of congestion (in equilibrium) at the high-quality firm provide no additional information to the consumers about product quality.

- We demonstrate the existence of a range of equilibrium pooling prices. These prices are uninformative about the quality and are lower than the high-quality firm's separating price. The volume of consumers served by the high-quality firm is thus higher. As the high-quality firm has a slower service, with pooling prices congestion at a firm is a key indicator of its quality level; consumers associate a busier firm with higher quality. In other words, when the firm is not busy, less consumers join. This finding is reminiscent of the 'empty-restaurant' syndrome that is often described in practice. An implication of the empty-restaurant syndrome is that the greater speed of the low-quality firm becomes less valuable as it leaves the firm empty more often, which results in low joining rates. The difference in profits between high- and low-quality firms hence reduces, which sustains the pooling-price equilibrium.

- We find that within the range of pooling prices (above a certain threshold price), the low-quality firm may *garble* its service rate (as it wants to avoid looking empty too often), even if it has *free* access to an infinitely faster service process. In other words, due to signaling considerations, low-quality firms may not invest in a fast service, even if the service is entirely costless.

- In the pooling-price equilibria, both firms' profits are higher than in the separating-price equilibria. This finding provides a theoretical answer to the question of why firms would signal quality with congestion instead of with their prices.

- We study the case in which the high-quality firm delivers superior value at a higher speed. With this additional advantage, separating with prices is less costly for the high-quality firm. Nevertheless, pooling equilibria continue to exist under several scenarios. When the low-quality firm's service rate is sufficiently low (but not too low), the high-quality firm can almost make as much profit as a situation in which all consumers are fully informed. Thus, congestion is a powerful signal of quality, even when the high-quality firm also possesses a cost advantage in speeding up the service rate.

1.3. Organization of Our Paper

The rest of the paper proceeds as follows. In §2, we provide a formal description of our model, we derive some preliminary results in §3. In §§4 and 5, we consider the equilibrium outcomes that survive the intuitive-criterion refinements. In each section, we first consider the *separating equilibrium*, in which prices signal quality, and then the *pooling equilibria*, in which firm prices are identical. Note that the (unobserved) equilibrium service rates may be different.

In §4, we examine the case in which process efficiency is prohibitively expensive when producing a higher quality product or service. That is, the high-quality firm is primarily involved in product innovation (and has not invested in improving process efficiency). Thus, it serves consumers with a finite service rate. In §5, we examine a case in which the high-quality firm has achieved process efficiency and product innovation. In the main paper, in order to obtain analytical insights, we consider a case in which the higher service rate is infinitely faster than the lower service rate and is either costless or prohibitively expensive. We verify numerically the aforementioned insights in Appendix B when the faster service process is (i) limited in speed and (ii) finitely more expensive. Finally, we summarize by discussing our insights and pointing out future directions in §6.

2. Model

In this section, we describe the players in our model, the game and the equilibrium conditions. We then derive some preliminary results in §3, which are used for specific cases in §§4 and 5.

2.1. The Game

Consider a firm endowed with N slots (tables, production departments, etc.). Each slot has one server (e.g., a waiter, a machine, etc.). The service time is exponentially distributed. When all slots are occupied, any arriving consumer balks. Thus, we model the firm and consumer service as a *loss* system.

There are two types of firms that differ in the quality of the good (service or product). The firm's type is denoted by $\omega \in \{h, \ell\}$. The gross utility of a ω-quality firm, where $\omega \in \{h, \ell\}$, is v_ω, where $v_h > v_\ell > 0$. Without loss of generality, we normalize v_h to 1. We simply call the two types the high-quality firm ($\omega = h$) and the low-quality firm ($\omega = \ell$). The firm's type is a random variable, whose realization is observed (only) by the firm. Thus, product quality is the firm's private information. The probability that the quality is high, $\Pr(\omega = h) = p$, is common knowledge. We denote the prior expected quality by $v \triangleq (1-p)v_\ell + p$.

2.2. Consumer Information

Consumers arrive according to a Poisson arrival stream with rate Λ. Without loss of generality, we write Λ as $\Lambda = \Lambda_0 N$, where Λ_0 is the arrival rate per server. The consumers' (outside) reservation utilities are uniformly distributed over $[0, 1]$ (which is common knowledge) and are privately observed by each consumer. No consumer knows the firm's type. Upon arrival, consumers only observe the price and the congestion or occupancy level (the number of occupied slots). Neither the firm's selected service rate, nor its quality is observed by consumers. Based on the price and congestion upon arrival, if a slot is available, the consumer decides whether to balk (and obtain the outside utility) or join the system. If all N slots are busy, the consumer is blocked.

2.3. Firm strategies

The firm (i) sets a price, P_ω, after observing its quality, ω, and (ii) selects one of the two exponentially distributed service processes: One is a slow process with mean service time $1/\underline{\mu}$. The other process is fast with mean service time $1/\overline{\mu}$. We allow for the cost of speeding up to be dependent on the firm's quality. Without loss of generality, we write the total cost of speeding up as Nk_ω, for $\omega \in \{h, \ell\}$. (That is, k_ω is the cost of speeding up per server or slot for the ω-quality firm.)

In our base model in §§4 and 5, we assume that $\overline{\mu} = +\infty$ and $\underline{\mu} = \mu < +\infty$. One firm is stipulated to serve at μ (i.e., the cost of speeding up is infinite for the firm), while the other firm can select at no cost a service rate of either μ or $+\infty$. That is, the other firm can choose to *mimic* or *differentiate* its service rates.[3] All of the above assumptions are relaxed in the numerical experiments reported in Appendix B, which allows the faster service to be finite. That is, $\overline{\mu} < +\infty$, which is achieved at a strictly positive cost.

We denote the probability that the firm selects the high service rate after observing its type ω by β_ω. (Thus, the low service rate is chosen with probability $(1-\beta_\omega)$.) We write $\boldsymbol{\beta} \triangleq (\beta_\ell, \beta_h)$. Finally, the parameters of our model, $(N, \Lambda_0, v_\ell, p, \underline{\mu}, \overline{\mu}, k_\ell, k_h)$, are common knowledge.

2.4. Consumer Strategies and Beliefs

We denote a consumer's updated belief that the firm is of high type, after observing a firm with n occupied slots and price P, by $\gamma(n, P)$. Let $\boldsymbol{\gamma}(P) \triangleq (\gamma(0,P), \gamma(1,P), \gamma(2,P), \ldots, \gamma(N-1,P))$ be the vector that denotes a consumer's updated belief for every n at price P.

The consumer's decision is to join the firm or to balk from it based on the occupancy levels, updated beliefs, and price. We denote the probability that the consumer joins the system, under price P and state n, by $\alpha(n,P)$. $\alpha(n,P)$ is the probability that the updated utility (net of price) from joining the firm exceeds the reservation price. We write $\boldsymbol{\alpha}(P) \triangleq (\alpha(0,P), \alpha(1,P), \alpha(2,P), \ldots, \alpha(N-1,P))$.

[3] Recall that the service rates remain unobservable. That is, the consumers have to infer the service rate selected by a firm based on the congestion information.

Based on the preceding theoretical model, we are ready to describe the equilibrium strategies of the players (the firms and all consumers)—$\boldsymbol{\alpha}^*(P)$, $\boldsymbol{\beta}^*$, and $\boldsymbol{\gamma}^*(P)$—as well as the high- and low-quality firms' prices, (P_ℓ^*, P_h^*).

2.5. Equilibrium Conditions

The consumer's expected value of joining is $u(n,\boldsymbol{\beta},\boldsymbol{\gamma}) - P$. As the outside reservation utility is uniformly distributed over $[0,1]$, the probability of joining the queue is $[u(n,\boldsymbol{\beta},\boldsymbol{\gamma}) - P]^+$, where $x^+ = \max\{0, x\}$. Without loss of generality, we normalize the firm's profit by the number of slots, N: Let $r(\boldsymbol{\alpha},\mu)$ be throughput per slot; then the profit rate (per slot) is

$$\Pi_\omega(\boldsymbol{\alpha},\beta,P) = \beta(r(\boldsymbol{\alpha},\overline{\mu})P - k_\omega) + (1-\beta)r(\boldsymbol{\alpha},\underline{\mu})P.$$

With this notation, the equilibrium conditions are, $\forall P \in [0,1]$, $n \in \{0,\ldots,N-1\}$, and $\omega \in \{h,\ell\}$,

$$\alpha^*(n,P) = [\boldsymbol{\gamma}^*(n,P) + (1-\boldsymbol{\gamma}^*(n,P))v_\ell - P]^+ \tag{1}$$

$$(\beta_\omega^*, P_\omega^*) \in \underset{(P\geq 0,\beta\in[0,1])}{\arg\max}\, \Pi_\omega(\boldsymbol{\alpha}^*(P),\beta,P) \tag{2}$$

$$\boldsymbol{\gamma}^*(P) \text{ is consistent with } \boldsymbol{\beta}^* \text{ and } \boldsymbol{\alpha}^*(P) \text{ for all } P \text{ and } n \text{ via Bayes' rule.} \tag{3}$$

Conditions (1) and (2) ensure that all players' strategies are rational. Condition (3) describes a consumer's updated beliefs about the quality of the firm after observing its price and the congestion. We let $\Pi_\omega^* = \Pi_\omega(\boldsymbol{\alpha}^*(P_\omega^*), \beta_\omega^*, P_\omega^*)$ denote the ω-quality firm equilibrium profits. In equilibrium, consumers observe one of the prices in $\{P_h^*, P_\ell^*\}$ and congestion level n generated by the consumer joining rate $\boldsymbol{\alpha}^*(P^*)$ and the firm's service-rate strategy, $\boldsymbol{\beta}^*$.

2.6. Off-Equilibrium-Path Considerations

For all prices that are not observed on the equilibrium path, $P \notin \{P_h^*, P_\ell^*\}$, and congestion levels, n, that are not generated by $\boldsymbol{\alpha}^*(P^*)$ and $\boldsymbol{\beta}^*$, Bayes' rule cannot be applied to determine $\boldsymbol{\gamma}^*(P)$. Yet, we need to determine the consumers' belief and joining strategy off the equilibrium path in order to meaningfully characterize the firm's optimization problem expressed in Equation (2).

In accordance with the intuitive criterion propounded by Cho and Kreps (1987) for refinement in signaling games, we impose the following restrictions on every (n,P) pair that is not observed on the equilibrium path. Let $\hat{\pi}(n,\boldsymbol{\alpha},\beta)$ be the probability that a randomly arriving agent observes

n consumers in the system when the joining rate is determined by $\boldsymbol{\alpha}$ and the service rate strategy is β. We define the following set for any off-equilibrium pair, (n,P).

$$\begin{aligned}\Gamma(n,P) &= \{(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) \in [0,1]^{N+1}\times[0,1]^2\times[0,1]^{N+1} : \exists\gamma^o\in[0,1] \text{ such that:}\\ \alpha(k) &= [(1-\gamma(k))v_\ell+\gamma(k)-P]^+,\ \gamma(k)=\frac{\gamma^o\hat{\pi}(k,\boldsymbol{\alpha},\beta_h)}{\gamma^o\hat{\pi}(k,\boldsymbol{\alpha},\beta_h)+(1-\gamma^o)\hat{\pi}(k,\boldsymbol{\alpha},\beta_\ell)},\\ &\forall k\in\{0,\ldots,N-1\},\ \hat{\pi}(n,\boldsymbol{\alpha},\beta_\ell)>0 \text{ and } \hat{\pi}(n,\boldsymbol{\alpha},\beta_h)>0\}.\end{aligned} \tag{4}$$

$\Gamma(n,P)$ is a set of rational consumer strategies at price P and beliefs about the quality that are consistent via Bayes' rule with some prior (γ^o) and service rate strategy, for which the probability of observing (n,P) is strictly positive. We now determine the highest possible profits that a ω-quality firm can obtain with any $(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\in\Gamma(n,P)$:[4]

$$\Pi_\omega^{\text{dev}}(n,P)=\sup_{(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\in\Gamma(n,P)}\Pi_\omega(\boldsymbol{\alpha},\beta_\omega,P). \tag{5}$$

The intuitive criterion imposes the following restrictions on $\boldsymbol{\gamma}^*(n,P)$ for any (n,P) observed off the equilibrium path.

$$\boldsymbol{\gamma}^*(n,P)=\begin{cases}1, & \Pi_\ell^*>\Pi_\ell^{\text{dev}}(n,P) \text{ and } \Pi_h^{\text{dev}}(n,P)>\Pi_h^*\\ 0, & \Pi_h^*>\Pi_h^{\text{dev}}(n,P) \text{ and } \Pi_\ell^{\text{dev}}(n,P)>\Pi_\ell^*\\ \in[0,1], & \text{otherwise.}\end{cases} \tag{6}$$

In other words, if upon observing an off-equilibrium congestion–price pair (n,P), there exists a unique firm type with strictly higher profits than its equilibrium profits, the consumer must assign a belief of probability one to that firm type. If there exists no unique firm type that improves on the equilibrium profits by deviating to a price P (that is, when both firm types improve by deviating, or none them can improve), no restriction is imposed on the off-equilibrium beliefs.

3. Preliminary Results

Before proceeding with the analysis of the game, we first provide some useful results. Let $\pi(n,\boldsymbol{\alpha},\mu)$ be the probability that the number of occupied slots is n when the consumers' joining strategy is $\boldsymbol{\alpha}$ and the firm's service rate is μ.

Lemma 1. *The Birth–Death steady-state probabilities are*

$$\pi(n,\boldsymbol{\alpha},\mu)=((\Lambda_0N/\mu)^n/n!)\prod_{m=0}^{n-1}\alpha(m)\times\left[1+\sum_{k=1}^{N}((\Lambda_0N/\mu)^k/k!)\prod_{m=0}^{k-1}\alpha(m)\right]^{-1}.$$

[4] Notice that, due to the strict inequality in the set $\Gamma(n,P)$ of Equation (4), it is possible that the maximum of $\Pi_\omega(\boldsymbol{\alpha},\beta_\omega,P)$ is not reached in $\Gamma(n,P)$. This is the case when $\overline{\mu}=+\infty$, as then it is possible that $\beta_\omega=1$ maximizes $\Pi_\omega(\boldsymbol{\alpha},\beta_\omega,P)$, leading to $\gamma(n)=0$ for $n>0$. Hence, we must find the supremum of $\Pi_\omega(\ \boldsymbol{\alpha},\beta_\omega,P)$.

With the PASTA property (Wolff, 1982), the probability that a randomly arriving agent observes n slots occupied, $\hat{\pi}(n,\boldsymbol{\alpha},\beta)$, equals the long-run probability that the system is in state n: $\beta\pi(n,\boldsymbol{\alpha},\overline{\mu})+(1-\beta)\pi(n,\boldsymbol{\alpha},\underline{\mu})$. The long-run probability also allows us to define the throughput (per server) for a given service rate μ and a given consumer joining strategy $\boldsymbol{\alpha}$:

$$r(\boldsymbol{\alpha},\mu)=\Lambda_0\sum_{k=0}^{N-1}\alpha(n)\pi(n,\boldsymbol{\alpha},\mu).$$

We define $\Delta(\boldsymbol{\alpha})\triangleq r(\boldsymbol{\alpha},\overline{\mu})-r(\boldsymbol{\alpha},\underline{\mu})$ as the increase in throughput from speeding up when the consumer joining strategy is $\boldsymbol{\alpha}$. We obtain the following properties.

Lemma 2. *(i)* $r((\alpha,\alpha,\alpha,\ldots,\alpha),\mu)$ *increases (weakly) concavely in* α *for* $\mu<+\infty$ *(*$\mu=+\infty$*) and increases in* μ*. (ii)* $\Delta((\alpha,\alpha,\alpha,\ldots,\alpha))>0$.

These properties are intuitive. An increase in arrival rate leads to an increase in throughput, but, as the maximum throughput is bounded by the service rate, μ, the increase in throughput progressively decreases with the arrival rate. When the service rate is infinitely fast, the throughput is equal to $\alpha\Lambda_0$. Furthermore, when the arrival rate does not depend on the state of the system upon arrival, a greater service speed leads to a higher throughput.

Lemma 3. *Let* $(\alpha_0,\alpha_1,\alpha_1,\ldots)$ *be the joining profile such that the joining rates are* α_0 *when* $n=0$ *and* $\alpha_n=\alpha_1$ *for* $n\geq 1$*. Let the service rate be* $\mu<\Lambda_0$ *and let the number of slots,* N*, become arbitrarily large. Then the firm's profit rate (per slot) is*

$$\lim_{N\to+\infty} r((\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)=\min\{\alpha_1,\mu\}\Lambda_0 \text{ and } \lim_{N\to+\infty} r((\alpha_0,\alpha_1,\alpha_1,\ldots),+\infty)=\alpha_0\Lambda_0.$$

When adding servers with rate μ ($<\Lambda_0$) and increasing the arrival rate proportionally, the firm's throughput (per slot) depends only on the joining strategy when the firm is not empty: α_1. This is because the service rate is less than the arrival rate at each server. Hence, the servers are always busy. On the other hand, when adding servers with infinite service rates, the firm's throughput (per slot) is determined by the joining strategy when the firm is empty: α_0. Again, this is because the service rate is larger than the arrival rate at each server.[5] Hence, the firm will be empty all the time. We introduce the following notation that will be used throughout the paper.

$$\alpha_\omega(P,n)=(v_\omega-P)^+, \text{ if } 0\leq P\leq 1 \text{ for all } 0\leq n\leq N-1.$$

[5] When $N\to+\infty$, any $\overline{\mu}>1$ will be sufficient for the throughput (per server) of the fast firm to be equal to α_0 (the joining rate when the firm is empty). Hence, $\overline{\mu}=+\infty$ is sufficient, but not necessary to obtain our results.

We denote the strategy $(\alpha_\omega(P,0), \alpha_\omega(P,1), \ldots)$ by $\boldsymbol{\alpha}_\omega(P)$. $\boldsymbol{\alpha}_\omega(P)$ is the consumer joining strategy *if* the consumer believes that the firm's quality is ω. Because of this belief, this joining strategy is independent of congestion.[6] Now, we analyze the ω-type firm's deviation profits (see Equation (5)).

LEMMA 4. *The deviation profits of Equation (5) are independent of n and solved by $\boldsymbol{\alpha} = \boldsymbol{\alpha}_h(P)$ and $\gamma^o = 1$ for both firm types, $\omega \in \{h, \ell\}$ and*

$$\Pi_\omega^{dev}(n,P) = \hat{\Pi}_\omega^{dev}(P) = \max_{\beta \in [0,1]} \Pi_\omega(\boldsymbol{\alpha}_h(P), \beta, P), \forall n \in \{0, \ldots, N-1\}.$$

For *any* given price and service rate strategy, the rational joining strategy that maximizes any firm's profit is one in which all consumers believe that the quality of the firm is high. This belief maximizes the throughput of the firm. For any observed pair, (n,P), the joining strategy $\boldsymbol{\alpha} = (1-P, 1-P, \ldots)$ is consistent with the belief $\boldsymbol{\gamma} = (1,1,\ldots)$ and any $\boldsymbol{\beta}$ when selecting $\gamma^o = 1$ (see Equation (4)). Notice that when $\overline{\mu} = +\infty$, it is possible that $\Pi_\omega(\boldsymbol{\alpha}, \beta_\omega, P)$ increases in β_ω for any $(\boldsymbol{\alpha}, P)$, which implies that β_ω tends to 1, and implies that $\gamma(n)$ tends to 0 for some $n > 0$. With this Lemma, we only need to place restrictions on the beliefs for off-equilibrium prices, instead of placing restrictions for each off-equilibrium congestion–price pair, (n,P). Henceforth, we denote the ω-firm's deviation profits at price P by $\hat{\Pi}_\omega^{\mathrm{dev}}(P)$.

Before characterizing the equilibria in the following sections, we introduce

$$\underline{\Pi}_\omega \triangleq \max_{0 \le P \le v_\ell, \beta \in [0,1]} \Pi_\omega(\boldsymbol{\alpha}_\ell(P), \beta, P), \tag{7}$$

which is the highest profit that a ω-firm can achieve if all consumers believe that quality is low (and hence the joining strategy is $v_\ell - P$). The aforementioned definition and results allow us to characterize equilibrium prices analytically.

4. Analysis When the Low-Quality Firm is Faster

In this section, we study the case in which the high-quality firm performs only product innovation with no process innovation: The firm delivers superior value but must use a slow service technology with rate $\mu < +\infty$ because the cost of speeding up (while maintaining its high quality) is prohibitively high (i.e., $k_h = +\infty$). To obtain analytical results, we consider the extreme case in

[6] Introducing a waiting room would make the joining strategy dependent on the number of consumers in the system upon arrival, see Debo et al. (2010).

which the low-quality firm can costlessly select a service process with rate μ or $+\infty$ (i.e., $k_\ell = 0$). In Appendix B, we relax these assumptions by considering finite high service rates at a strictly positive cost, and show that our conclusions continue to hold.

However, with these assumptions about the service rates and cost parameters, we achieve analytical tractability on equilibrium refinements, and we can determine the deviation profits: $\hat{\Pi}_\ell^{\text{dev}}(P) = P(1-P)\Lambda_0$ and $\hat{\Pi}_h^{\text{dev}}(P) = P \times r(\boldsymbol{\alpha}_h(P), \mu)$. Furthermore, $\underline{\Pi}_\ell = \frac{v_\ell^2}{4}\Lambda_0$ and $\underline{\Pi}_h = \max_{0 \le P \le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P), \mu)$.

4.1. Separating Prices

Let (P_h^*, P_ℓ^*), where $P_h^* \neq P_\ell^*$, be separating prices, and let μ be the service rate.

PROPOSITION 1. *When $v_\ell < \frac{4}{5}$, there always exists a separating equilibrium with service strategy $\boldsymbol{\beta}^* = (1, 0)$ and prices (P_h^*, P_ℓ^*), where $P_\ell^* = \frac{v_\ell}{2}$ and P_h^* is the highest root of $\hat{\Pi}_\ell^{dev}(P) = \underline{\Pi}_\ell$. That is, $P_h^* = \frac{1}{2} + \frac{1}{2}\sqrt{1 - v_\ell^2}$.*

When $\frac{4}{5} < v_\ell < 1$, a separating equilibrium only exists when $P_h^* \times r(\boldsymbol{\alpha}_\ell(P_h^*), \mu) \geq \underline{\Pi}_h$. There exists an $\epsilon > 0$ such that no separating equilibrium exists for $1 - \epsilon < v_\ell < 1$.

We prove that the separating equilibrium with the above-described prices can be sustained under the intuitive-criterion refinements with the following beliefs: $\boldsymbol{\gamma}^*(n, P_\ell^*) = 0$, $\boldsymbol{\gamma}^*(n, P_h^*) = 1$, and (off-equilibrium) for all $P \notin \{P_\ell^*, P_h^*\}$ and $n \in \{0, \ldots, N-1\}$, $\boldsymbol{\gamma}^*(n, P) = 0$. In Appendix A, we show

A. $\boldsymbol{\beta}^* = (1, 0)$ and (P_ℓ^*, P_h^*) solve the firm's rationality conditions of Equation (2), and

B. The beliefs on the equilibrium path satisfy Bayes' rule and off the equilibrium path satisfy the intuitive-criterion restrictions.

4.1.1. Illustrative example:

Before we discuss the intuition behind the results of Proposition 1, we illustrate the separating equilibrium characterized in Proposition 1 with a numerical example. Consider the parameters $N = 10$, $\mu = 0.5$, $\Lambda_0 = 1$; $v_\ell = 0.8$ and $p = 1/2$ (Figure 1).

With off-equilibrium-path beliefs that the quality of the firm is low, it is obvious that the firm-rationality conditions require that the equilibrium profits are at least equal to $\underline{\Pi}_\omega$ for both firm types, $\omega \in \{\ell, h\}$. With price P_ℓ^*, the low-quality firm's profits are exactly equal to $\underline{\Pi}_\ell$ because in equilibrium it is revealed to be the low-quality firm. The high-quality firm's equilibrium profits may be strictly higher than $\underline{\Pi}_h$ (the profit it accrues when consumers think it is a low-quality firm).

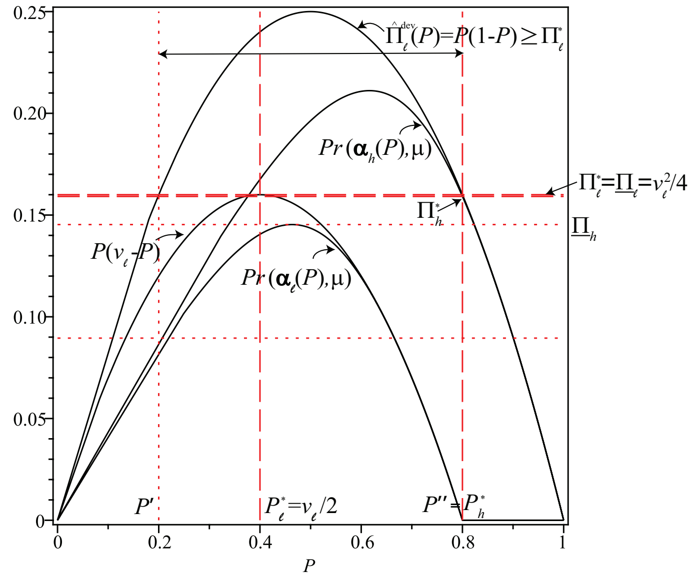

Figure 1 Firm profits and prices under the separating equilibrium when the high-quality firm is slow: $\hat{\Pi}_\ell^{\text{dev}}(P)$, $P\times r(\alpha_h(P),\mu)$, $P\times(v_\ell-P)$ and $P\times r(\alpha_\ell(P),\mu)$. The arrow indicates the price region where the low-quality firm's deviation profits are higher than Π_ℓ^*. The high-quality firm's separating price must be outside that region. It can be seen that the high-quality firm's profits at P_h^* ($\Pi_h^*=P_h^*r(\alpha_h(P),\mu)$) are higher than at P'. Note also that $\Pi_h^*\geq\underline{\Pi}_h=\max_{0\leq P\leq v_\ell}Pr(\alpha_\ell(P),\mu)$ (and $\Pi_\ell^*=\underline{\Pi}_\ell$). Parameters are $N=10$, $\mu=0.5$, $\Lambda_0=1$; $v_\ell=0.8$ and $p=1/2$. Due to the intuitive criterion, we need $P_h^*=P''$.

In Figure 1, examine the range of prices in the region bounded by $P'(=0.2)$ and $P''(=0.8)$. No separating price exists for the high-quality firm in equilibrium in the region where $\hat{\Pi}_\ell^{\text{dev}}(P)>\Pi_\ell^*=\underline{\Pi}_\ell$. In this region, the low-quality firm has an incentive to imitate the high-quality firm's price ($\Pi_\ell^{\text{dev}}(P)$) since its deviation profits are higher than its equilibrium profits (Π_ℓ^*). Thus, the high-quality firm's separating price cannot be inside the region. As a consequence, there are only two regions that the high-quality firm can consider: $[0,P']$ and $[P'',1]$. Since its profits are monotone in these regions, the two candidate separating prices (for the high-quality firm) are P' and P''. It can be easily observed from Figure 1 that the high-quality firm's profits are higher with the highest price, P''. Hence, whenever the low-quality firm has no incentive to deviate from P_ℓ^* (that is in $[0,P']\cup[P'',1]$) the high-quality firm has no incentive to deviate from its equilibrium price either. As a consequence, our off-equilibrium-path beliefs survive the intuitive criterion. Obviously, in (P',P''), the low-quality firm has an incentive to deviate, hence, our off-equilibrium-path belief that the quality of the firm is low survives the intuitive criterion. Due to the "cost" of separation, the high-quality firm's profits, $\Pi_h^*=0.159$, are *slightly* lower than those of the low-quality firm, ($\Pi_\ell^*=(v_\ell^2/4)\Lambda_0=0.16$). Using observations in Figure 1, we now discuss the results of Proposition 1.

Proposition 1 is intuitive: The high-quality firm signals quality with the highest price that prevents the low-quality firm from imitating the high-quality firm's price. Setting such a high price implies low admission rates at the high-quality firm. At low volumes of admission, the high-quality firm's service rate handicap is less pronounced. To see this, note that at some arbitrarily low demand volumes, the blocking probability with finite service is negligible. As a result, there is negligible difference between the throughput with infinitely fast service and finitely fast service. Thus, the high-quality firm sets a high price to prevent the low-quality firm from imitating the high-quality firm's price.

Also note that the high-quality firm's profits are less than the low-quality firm's profits. This may be surprising, but it is also intuitive within the confines of the signaling model: The low-quality firm has a speed advantage that cannot be imitated by the high-quality firm (as the cost of speeding up is high). The low-quality firm, however can imitate any price that the high-quality firm sets.

Furthermore, this separation of prices is possible only when the quality difference between the high- and low-quality firms is large enough. For instance, if v_ℓ is close to v_h, ($v_\ell > 80\%\, v_h$ which is normalized to 1), the highest value of v_ℓ for which a separating equilibrium exists depends on the service rate and the number of slots. This existence condition is determined by the incentive of the high-quality firm to make more equilibrium profit with separation than it would in a case in which consumers would assume its quality were low ($\underline{\Pi}_h$, see Equation (7), $\Pi_h^* > \underline{\Pi}_h$).

When v_ℓ is high enough and close to $v_h (=1)$, it is not *that* damaging for the high-quality firm, if the consumers were to believe it to be a low-quality firm. In fact, the revenue loss for the high-quality firm due to such a belief may be less than the cost it would incur in trying to prevent the low-quality firm from imitating. As a result, the separating equilibrium is not sustained if the quality of the firms are comparable.

4.1.2. When N is large and the arrival rate is high: When the number of slots and the arrival rate are large, the separating equilibrium prices are independent of N. Let $\underline{\mu} = \mu < 1$, $\overline{\mu} = +\infty$ and $\Lambda_0 = 1$. The high-quality firm's equilibrium profits are $\min\left(\frac{1}{4}v_\ell^2, \left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - v_\ell^2}\right)\mu\right)$. In this case, a separating equilibrium exists for any $v_\ell < 1$. That is, in Proposition 1, ϵ tends to 0^+ when N tends to $+\infty$). Thus, the capacity constraints and finite stochastic arrivals limit the occurrence of separating equilibria. In the following subsection, we examine pooling equilibria (in which the firm prices are identical in equilibrium) that survive the intuitive criterion.

4.2. Pooling Price

In this subsection, we study pooling prices: $P_\ell^* = P_h^* = P^*$ for both firm types. Given the prohibitive cost of speeding up, the high-quality firm will select a low service rate in any equilibrium: $\beta_h^* = 0$. Thus, it is sufficient to analyze the case $\boldsymbol{\beta}^* = (\beta_\ell^*, 0)$, with $0 \le \beta_\ell^* \le 1$, where we allow the low-quality firm to randomize between being slow and infinitely fast. Consider a strictly positive price, P^*, that is lower than the ex ante expected value: $0 < P^* < v$. Suppose that consumers who observe $n \in \{0, \ldots, N-1\}$ join with a strictly positive probability. Let $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(P^*)$ denote the consumer joining strategy at the equilibrium price.

4.2.1. Low-quality firm selects high service rate with probability 1: Consider a pooling price, P^* in an equilibrium with a fast, low-quality firm. For $\beta_\ell^* = 1$, as $\overline{\mu} = +\infty$, the low-quality firm is always empty and the high-quality firm's occupancy level is nonempty with a strictly positive probability. As a consequence, any nonempty firm is a high-quality firm; $\gamma^*(n, P^*) = 1$ for $1 \le n \le N$ and the joining probability is $1 - P^*$. It follows that $\boldsymbol{\alpha}^* = (\alpha_0^*, 1-P^*, 1-P^*, \ldots)$ (when $\beta_\ell^* = 1$). Only α_0^* remains to be determined. In equilibrium, the joining probability at the empty queue (α_0^*) should be such that it generates a belief about the firm's quality that is consistent with Bayes' rule $\frac{p\pi(0,\boldsymbol{\alpha}^*,\mu)}{p\pi(0,\boldsymbol{\alpha}^*,\mu)+(1-p)}$. This posterior, then, should rationalize the assumed joining probability at the empty firm. Hence, α_0^* is determined as a fixed point and can be solved analytically as the root of a quadratic equation (of which the second root can be shown to be discarded). The solution of $\alpha_0 = v_\ell + \frac{p\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)}{p\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)+(1-p)}(1-v_\ell) - P$ for α_0 is $H(P, \alpha_1, p/(1-p))$, where

$$H(P,\alpha_1,L) = \frac{1}{2}\left(\frac{1+L}{F(\alpha_1)} + P - v_\ell\right)\left(-1 + \sqrt{1 + 4\frac{(1-P)L + v_\ell - P}{F(\alpha_1)\left(\frac{1+L}{F(\alpha_1)} + P - v_\ell\right)^2}}\right), \text{ and} \tag{8}$$

$$F(\alpha_1) = \sum_{k=1}^{N} \frac{1}{k!}\alpha_1^{k-1}\left(\frac{\Lambda_0 N}{\mu}\right)^k.$$

We denote $\alpha_0^p(P) = H(P, 1-P, p/(1-p))$. Recall that $\Delta((\alpha_0, \alpha_1, \alpha_1, \ldots))$ is the increase in revenue rate for a strategy profile $(\alpha_0, \alpha_1, \alpha_1, \ldots)$. In the proof of the following proposition, we prove that this additional revenue is positive if $\alpha_0 < \alpha_1 - \frac{(\alpha_1)^{N-1}}{N!F(\alpha_1)}\left(\frac{\Lambda_0 N}{\mu}\right)^N$. We denote $\Delta^p(P) = \Delta((\alpha_0^p(P), 1-P, 1-P, \ldots))$, the revenue increase from speeding up at price P when the joining strategy is

$(\alpha_0^p(P), 1-P, 1-P, \ldots)$. The candidate pooling price, P^*, can be part of an equilibrium with a fast, low-quality firm *only* when the low-quality firm has an incentive to speed up. That is, when $\Delta^p(P^*) > 0$. Now we can formulate the pooling equilibrium that survives the intuitive-criterion refinement (for a case in which the low-quality firm is fast), which leads us to Proposition 2.

PROPOSITION 2. *Suppose that a pooling price, $P^* > 0$, exists for which the low-quality firm selects the highest service rate, $\boldsymbol{\beta}^* = (1,0)$, and consumers join according to $\boldsymbol{\alpha}^* = (\alpha_0^p(P^*), 1 - P^*, 1-P^*, \ldots)$. P^* is an intuitive pooling price only*

A. *when $P^* < \hat{P}$, where $\hat{P}$ is the unique root of*

$$\alpha_0^p(P) = 1 - P - \frac{(1-P)^{N-1}}{N!F(1-P)} \left(\frac{\Lambda_0 N}{\mu}\right)^N$$

when $\alpha_0^p(0) < 1 - \frac{1}{N!F(1)} \left(\frac{\Lambda_0 N}{\mu}\right)^N$ and $\hat{P} = 0$ otherwise.

B. *when*

$$P^* \times r(\boldsymbol{\alpha}^p(P^*), \mu) \geq \underline{\Pi}_h,\ P^* \times \alpha^*(0, P^*) \geq \underline{\Pi}_\ell = \frac{1}{4} v_\ell^2 \Lambda_0,\ \underline{P}_\ell(P^*) \leq \underline{P}_h(P^*) \text{ and } \overline{P}_h(P^*) \leq \overline{P}_\ell(P^*),$$

$$\text{where } \underline{P}_\ell(P) = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4P\alpha_0^p(P)} \text{ and } \overline{P}_\ell(P) = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4P\alpha_0^p(P)}$$

$$\text{and } \underline{P}_h(P) \text{ and } \overline{P}_h(P) \text{ are the roots of } P' \times r(\boldsymbol{\alpha}_h(P'), \mu) = P \times r(\boldsymbol{\alpha}^p(P), \mu).$$

Again, the proof (deferred to Appendix A) proceeds as follows. Condition A guarantees that the low-quality firm speeds up. Recall that the consumers' off-equilibrium belief is that the quality of the firm is low. The first and second condition of B ensure that the equilibrium profits, Π_ω^* for both firms ($\omega = h$ or ℓ), are higher than the highest possible profits when all consumers believe the firm's quality is low, ($\underline{\Pi}_\omega$). Hence, these conditions ensure rationality of the firm's pricing strategy. On the equilibrium path (i.e., for $P = P^*$), the beliefs, joining strategy and service rate are consistent with the prior belief (via Bayes' Rule). The third and fourth conditions of B ensure that, off the equilibrium path, the beliefs obey the intuitive criterion. The prices $\overline{P}_\omega(P^*)$ and $\underline{P}_\omega(P^*)$ are the solutions of $\Pi_\omega^{\text{dev}}(P) = \Pi_\omega^*$. The conditions ensure that, when the low-quality firm has no incentive to deviate, the high-quality firm also has no incentive to deviate. As a consequence, the off-equilibrium-path belief (that the quality of the firm is low) survives the conditions of the intuitive criterion.

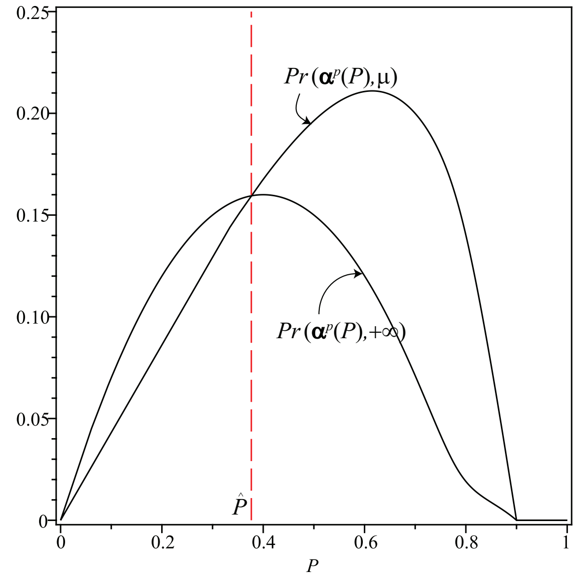

Figure 2 **Section 4.2.1: Illustration of profits of a low-quality firm as a function of its speed, $P \times r(\alpha^p(P), \mu)$, and $P \times r(\alpha^p(P), +\infty)$. The vertical dashed line represents $\hat{P}$. For prices higher than $\hat{P}$, there cannot be an equilibrium with the low-quality firm selecting a high service rate. Parameters are: $N = 10$, $\mu = 0.5$, $\Lambda_0 = 1$; $v_\ell = 0.8$ and $p = 1/2$.**

In Proposition 2, we show that if the low-quality firm has an incentive to speed up at price $P = 0$, i.e., $\Delta^p(0) > 0$, then there always exists a unique price, $\hat{P}$, such that the low-quality firm has an incentive to speed up for all prices in $(0, \hat{P})$, but not for higher prices. This result is interesting. Notice that when the firm is not empty, the joining rate is the highest possible at $1 - P$. (All consumers with a valuation higher than P join.) Clearly, this rate must always be higher than the joining rate at the empty firm, $\alpha_0^p(P)$. The business *stalls* when the firm is *empty*, as a busy firm is associated *with certainty* in the consumer's beliefs with high quality. As a consequence of the business stalling at the empty firm, any speed advantage a low-quality firm possesses is worthless: Fast service can only make the firm empty *more often*. Also, the disadvantage of fast service is pronounced at high prices (above $\hat{P}$). At high prices, the admitted volume is low, so any disadvantage due to the high-quality firm's slow service diminishes further.

Illustrative example (continued): We shall now return to our illustrative example and examine the pooling equilibrium. We now show $\Delta(P)$ and $\hat{P}$ in Figure 2 by plotting the revenue rates for different prices, $P \times r(\boldsymbol{\alpha}^p(P), \mu)$ and $r(\boldsymbol{\alpha}^p(P), +\infty) = P \times \alpha^p(0, P)\Lambda_0$ for all P for the low-quality firm. The curves intersect at $\hat{P}$. For prices above $\hat{P}$, fast service decreases the low-quality firm's profit. In the next subsection, we characterize intuitive pooling prices with the low-quality service firm garbling its service rate.

4.2.2. Low-quality firm garbles its service rate: We now examine a case in which the low-quality firm randomizes its service rate (i.e., $\beta_\ell^* \in (0,1)$). (We can easily show that $\beta_\ell^* = 0$ can never be an equilibrium.[7]) Thus, the low-quality firm adopts a delay tactic (compared to §4.2.1), to hide its private information (low quality)—a result reminiscent of observations in Afèche (2005) from a different modeling context. Suppose a pooling price, $0 \le \hat{P} < P^*$ exists. Then, the low-quality firm selects the highest service rate in equilibrium, with probability $\beta_\ell^* \in (0,1)$. The consumer's belief regarding quality upon observing a congestion level n now also depends on β_ℓ^*:

$$\frac{p\pi(0,\boldsymbol{\alpha}^*,\mu)}{p\pi(0,\boldsymbol{\alpha}^*,\mu)+(1-p)((1-\beta_\ell^*)\pi(n,\boldsymbol{\alpha}^*,\mu)+\beta_\ell^*)}$$

and

$$\frac{p\pi(n,\boldsymbol{\alpha}^*)}{p\pi(n,\boldsymbol{\alpha}^*)+(1-p)(1-\beta_\ell^*)\pi(n,\boldsymbol{\alpha}^*)} \qquad \text{for } 1 \le n \le N.$$

The solution of $\alpha_0 = v_\ell + \frac{p\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)}{p\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)+(1-p)((1-\beta)\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)+\beta)}(1-v_\ell) - P$ for α_0 is $H(P,\alpha_1,p/((1-p)\beta))$, where $H(P,\alpha_1,L)$ was defined in Equation (8). When $1 \le n$, trivially, we obtain

$$\alpha_1(\beta,P) = v_\ell + \frac{p}{p+(1-p)(1-\beta)}(1-v_\ell) - P.$$

Hence, any pooling price P^* for which the low-quality firm randomizes its service rate must have a consumer joining strategy, $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\alpha_0^p(P^*),\beta^p(P^*),P^*)$ and a randomization probability $\beta_\ell^* = \beta^p(P^*)$ where $\boldsymbol{\alpha}(\alpha_0,\beta,P) \triangleq (\alpha_0,\alpha_1(\beta,P),\alpha_1(\beta,P),\ldots,\alpha_1(\beta,P))$ and $(\alpha_0^p(P),\beta^p(P))$ are the solutions of

$$\alpha_0 = H(P,\alpha_1(\beta,P),p/((1-p)\beta)) \text{ and } \alpha_0 = \alpha_1(P,\beta) - \frac{(\alpha_1(\beta,P))^{N-1}}{N!F(\alpha_1(\beta,P))}\left(\frac{\Lambda_0 N}{\mu}\right)^N, \forall P \in [0,v].$$

The first condition above makes the joining probability at the empty firm consistent with the consumer belief, and the second condition guarantees that the low-quality firm has no incentive to speed up from its chosen garbled-service strategy (that is: $\Delta^p(P) = 0$). Now we can formulate the pooling equilibrium (under the intuitive-criterion refinement) when the low-quality firm garbles its service rate.

PROPOSITION 3. *Suppose that a pooling price, $P^* > 0$, exists for which the low-quality firm selects the highest service rate, $\boldsymbol{\beta}^* = (\beta_\ell^*, 0)$ and consumers join according to $\boldsymbol{\alpha}^* =$*

[7] As the joining rate would become independent of the congestion, which provides an incentive for the low-quality firm to speed up (through Lemma 2(ii)).

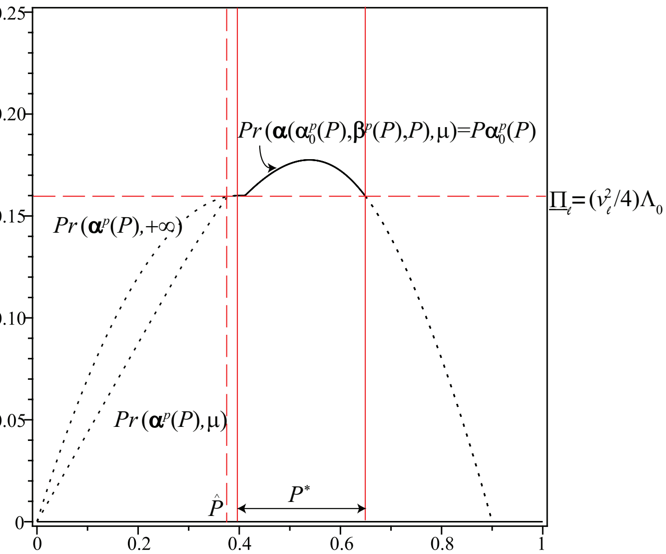

Figure 3 For Section 4.2.2. Range of pooling prices that are supported by a belief that survives the intuitive criterion (see Proposition 3) and corresponding profits (solid line) for: $N = 10$, $\mu = 0.5$, $\Lambda_0 = 1$; $v_\ell = 0.8$ and $p = 1/2$. When the low-quality firm does not garble, the low-quality firm's (high-quality firm's) profits are $Pr(\alpha^p(P), +\infty)$ ($Pr(\alpha^p(P), \mu)$). When the low-quality firm garbles, the high- and low-quality firms' profits are identical ($Pr(\alpha(\alpha_0^p(P), \beta^p(P), P), \mu) = P\alpha_0^p(P)$).

$\boldsymbol{\alpha}(\alpha_0^p(P^*), \beta^p(P^*), P^*)$. *$P^*$ is an intuitive pooling price only when (i) $P^* > \hat{P}$ and (ii) $P^* \times r(\boldsymbol{\alpha}^*, \mu) \geq \underline{\Pi}_\ell = \frac{1}{4}v_\ell^2\Lambda_0$.*

Intuitively, as the low-quality firm garbles its service rate, both firms make the same profits in equilibrium. However, the deviation profits of the low-quality firm are always higher than the deviation profits of the high-quality firm due to the service-rate advantage. As a consequence, the intuitive criterion supports the belief that any deviation from a pooling equilibrium should come from the low-quality firm.

Illustrative example (continued): In Figure 3, we plot a range of pooling prices when the service-rate strategy involves randomization (corresponding to Proposition 3) for our example. Note that there exists a range of prices larger than $\hat{P}$ for which the profits of both firms are higher than $\frac{v_\ell^2}{4}\Lambda_0 = 0.16$. Hence, these are pooling prices. For the prices lower than $\hat{P}$, the intuitive-criterion conditions need to be checked. However, the profits for the low-quality firm at these prices are less than $\frac{v_\ell^2}{4}\Lambda_0 = 0.16$. Hence, pooling prices do not exist with pure-strategy service rates.

The empty-restaurant syndrome: The analysis of pooling equilibrium allows us to discuss some management implications. Recall that for pooling prices below $\hat{P}$, the joining probability at the empty firm was lower than at the nonempty firm. For pooling prices above $\hat{P}$, the joining probabilities satisfy $\alpha_0 = \alpha_0\pi(0) + \alpha_1(1-\pi(0)-\pi(N))$, i.e., $\alpha_0(1-\pi(0)) = \alpha_1(1-\pi(0)-\pi(N))$. From the above condition, it follows that the joining probability at an empty firm (α_0) is *lower* than at a nonempty firm (α_1). The joining rate *increasing* in congestion at the firm is reminiscent of the empty-restaurant syndrome. This phenomenon has been colloquially described in many blogs and discussions.[8]

Similar observations have been made by Becker (1991). In our paper, fast service makes the firm empty more often, and then the business stalls. This is more pronounced at high prices (above $\hat{P}$), when the arrival volumes are lower. Any profit advantage of the low-quality firm, associated with its fast service, vanishes. If high service rates are chosen, the empty-restaurant syndrome persists as an outcome. At the high-price pooling equilibrium, the profits of both firm types are equal.

4.2.3. When N is large and the arrival rate is high: So far, we have characterized the necessary conditions for intuitive pooling prices, but we have not yet established whether such prices exist. Due to the analytical complexity of the conditions, we obtain insights into the pooling prices when the number of slots, N, is large. However, the existence of pooling prices can be verified numerically for any case. For completeness, in Proposition 4, we also characterize the equilibrium joining strategies when the number of available slots and the arrival rate become large, that is, N tends to $+\infty$.

Proposition 4. *Let* $\underline{\mu} = \mu < 1$, $\overline{\mu} = +\infty$ *and* $\Lambda_0 = 1$*: Then* $\lim_{N\to+\infty} \hat{P} = [v_\ell - \mu]^+$*. For* $0 \le P < \hat{P}$*:* $\lim_{N\to+\infty} \alpha_0^p(P) = v_\ell - P$ *and* $\alpha^*(n,P) = 1-P$ *for* $1 \le n$*,* $\beta_\ell^* = 1$ *and* $\beta_h^* = 0$ *and*

$$\lim_{N\to+\infty} r(\boldsymbol{\alpha}^*, \mu) = \mu \text{ and } \lim_{N\to+\infty} r(\boldsymbol{\alpha}^*, +\infty) = v_\ell - P.$$

For $\hat{P} < P < v$*:* $\lim_{N\to+\infty} \alpha_0^p(P) = \min(\mu, v-P)$ *and* $\lim_{N\to+\infty} \alpha^p(n,P) = v - P$ *for* $1 \le n$*,* $\lim_{N\to+\infty} \beta_\ell^* = \beta_h^* = 0$ *and*

$$\lim_{N\to+\infty} r(\boldsymbol{\alpha}^*, \mu) = \min(\mu, v-P).$$

[8] see, for example, `http://www.growcookeat.com/2009/06/ers-empty-restaurant-syndrome.html` and `http://thestar.com.my/metro/story.asp?file=/2009/12/13/sundaymetro/5261952&sec=sundaymetro`, websites accessed on September 03, 2010.

A. *When* $\mu < v_\ell/2$ *and* $\frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right) \le \frac{1}{4}\frac{v_\ell^2}{\mu}$, *there exists no pooling price.*

When $\mu < v_\ell/2$ *and* $\frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right) > \frac{1}{4}\frac{v_\ell^2}{\mu}$, *there exists a range of pooling prices above* $\hat{P}$*:*

$$P^* \in \left[\frac{1}{4}\frac{v_\ell^2}{\mu}, \frac{1}{2}(v + \sqrt{v^2 - v_\ell^2})\right].$$

B. *When* $v_\ell/2 < \mu < v_\ell$, *there exists a range of pooling prices above* $\hat{P}$*:*

$$P^* \in \left[\max\left(\frac{1}{4}\frac{v_\ell^2}{\mu}, \frac{1}{2}\left(v - \sqrt{v^2 - v_\ell^2}\right)\right), \frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right)\right].$$

Assume that $v_\ell > \mu$. The price, $\hat{P}$, below which the low-quality firm selects a high service rate in a pooling equilibrium (see Proposition 2) is now simply $\hat{P} = v_\ell - \mu > 0$.[9] For prices lower than $\hat{P}$, the low-quality firm's profits are higher with slow service. Consumers arriving at the empty firm significantly revise their prior belief of the value (to v_ℓ) and consumers arriving at the nonempty firm know for sure that the quality is high (as the low-quality firm does not randomize its service rate for $0 \le P < \hat{P}$). Here, the empty-restaurant syndrome is at its most extreme.

Hence, the high-quality firm enjoys the additional benefits of empty-restaurant syndrome: Its joining rate is determined by $\alpha^*(n,P) = 1 - P$ for $n \ge 1$. However, the volume of consumers it can serve is also restricted by its finite capacity, μ; the high-quality firm's throughput is $\min(1-P,\mu) = \mu$.

The low-quality firm suffers from the empty-restaurant syndrome: Its throughput is determined by $\alpha_0^p(P) = v_\ell - P$, which can be fully served (as the low-quality firm has no capacity restrictions). Hence, the low-quality firm's throughput is $v_\ell - P$. Thus, there must exist a price, $\hat{P}$, at which the low-quality firm gains no advantage from fast service, the low-quality firm's throughput with fast service ($v_\ell - P$) and slow service (μ) become equal: $\hat{P} = v_\ell - \mu$. For prices above $\hat{P}$, the low-quality firm randomizes its service rate. For all prices above $\hat{P}$ for any finite N, the randomization probability will be strictly greater than 0, but as N becomes larger, the randomization probability, β_ℓ^*, will approach 0^+. The empty-restaurant syndrome gradually vanishes and the consumers simply ignore the congestion level upon arrival; the joining probability is $\min(\mu, v - P)$, irrespective of the congestion. The throughput for *both* firms is the same and is determined by the high-quality firm's capacity (μ), up to price $v - \mu$ (which is higher than $\hat{P} = v_\ell - \mu$). The throughput is determined by the 'average' joining rate $v - P$ for prices in the range $(v - \mu, v)$. Having characterized the

[9] When $v_\ell < \mu$, this case does not exist as $\hat{P} = 0$.

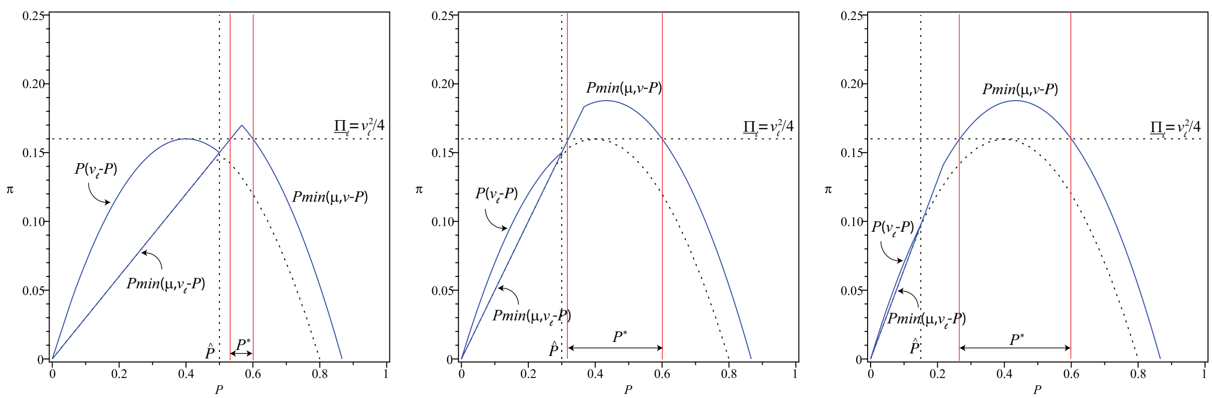

Figure 4 For the limiting case in Section 5.2. Illustration of the low-quality firm's profits ($P(v_\ell - P)$ for $0 \leq P < \hat{P}$ and $P\min(\mu, v-P)$ for $\hat{P} \leq P < v$) and the high-quality firm's profits ($P\min(\mu, 1-P)$ for $0 \leq P < \hat{P}$ and $P\min(\mu, v-P)$ for $\hat{P} \leq P < v$). Parameter values are $v_\ell = 0.8$ and $p_0 = 1/3$, $\mu_0 = 0.3$ (left panel) and $\mu_0 = 0.5$ (middle panel) and $\mu_0 = 0.65$ (right panel). The range of pooling prices is determined by $P\min(\mu, v-P) > v_\ell^2/4 = 0.16$ (horizontal dotted line). The segment between the two solid vertical lines is the region in which the low-quality firm garbles its service rate.

equilibrium queue joining strategy, we now characterize the pooling prices that survive the intuitive criterion.

Illustrative example (continued): The case $\mu < v_\ell/2$ and $\frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right) > \frac{1}{4}\frac{v_\ell^2}{\mu}$ (the second part of *(i)* in Proposition 4) is illustrated in Figure 4, in the left panel: $v_\ell/2$ and $v_\ell - \mu$ maximize $P(v_\ell - P)\Lambda_0$ and $P\min(\mu, v_\ell - P)\Lambda_0$, respectively. Note that there is a small range of intuitive pooling prices. The case $v_\ell/2 < \mu < v_\ell$ (*(ii)* in Proposition 4) is illustrated in Figure 4, in the middle and right panels. For prices higher than $\hat{P}$, the unconstrained profits, $\frac{1}{4}v^2$, exceed $\frac{1}{4}v_\ell^2$. The upper bound of the range of pooling prices is determined by the unconstrained profits. When $(v-\mu)\mu > \frac{1}{4}v_\ell^2$, the lower bound on the range of pooling prices is determined by the constrained (unconstrained) profits as in Figure 4, middle (right) panel. Note that a pooling price always exists as the upper bound is always more than $\frac{1}{2}v_\ell$, which is always more than $\frac{1}{4}\frac{v_\ell^2}{\mu}$ as $v_\ell/2 < \mu$. Observe that the upper bound of the pooling prices, $\frac{1}{2}(v + \sqrt{v^2 - v_\ell^2})$, is always lower than the separating price (which is $\frac{1}{2}(1 + \sqrt{1 - v_\ell^2})$).

We have thus obtained that when N is large, either the high-quality firm is not too slow compared to the low-quality firm, or the prior expected value ($v = (1-p_0)v_\ell + p_0$) is high, leading to a range of pooling prices.

4.3. Comparison of the High-Quality Firm's Separating and Pooling Prices

Comparing the equilibrium results derived, the reader might note that

1. the profits in any pooling price equilibria are higher than $(v_\ell^2/4)\Lambda_0$ in the pooling and lower than $(v_\ell^2/4)\Lambda_0$ in separating equilibria.
2. the highest possible pooling price is less than the high-quality firm's separating price.

These observations illustrate one main tenet of our paper: The profits for both the high- and low-quality firms are *higher* in the pooling-price equilibria than in the separating-price equilibria. In the pooling-price equilibria, prices are not informative, but the firm's congestion level is. The pooling prices are lower than the separating prices and, hence, the demand volume and rejection rate are higher than in the separating equilibrium. Hence, the excess *unmet demand* is higher with a pooling price than with a separating price.

We believe that the pooling-price equilibria has an intuitive appeal as it implies that (i) consumers learn about quality from the congestion they observe upon arrival, and (ii) the low-quality firm hides its fast service process in order not to fall into the empty-restaurant syndrome trap.

5. When the High-Quality Firm is Faster

In this section, we study a case in which the firm simultaneously performs product innovation *and* process improvement. The high-quality firm not only delivers superior value, but also has access to a faster (service/production) process. Again, for analytical tractability, we consider the extreme case in which the low-quality firm has a service process with rate μ (i.e., $k_\ell = +\infty$), while a high-quality firm can select at no cost $(k_h = 0)$ a service process with rate μ or $+\infty$. With these assumptions about the service rates and cost parameters, we follow the same scheme as in §4, and first determine the deviation profits: $\hat{\Pi}_h^{\text{dev}}(P) = P(1-P)\Lambda_0$ and $\hat{\Pi}_\ell^{\text{dev}}(P) = P \times r(\boldsymbol{\alpha}_h(P), \mu)$. Furthermore, $\underline{\Pi}_h = \frac{v_\ell^2}{4}\Lambda_0$ and $\underline{\Pi}_\ell = \max_{0 \le P \le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P), \mu)$.

5.1. Separating Prices

We first characterize the separating prices that survive the intuitive criterion in Proposition 5.

PROPOSITION 5. *When* $v_\ell < \frac{2}{3}$*, there always exists a separating equilibrium with prices* $P_\ell^* = \arg\max_{0 \le P \le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P), \mu)$ *and* P_h^**, which is the lowest root of* $P \times r(\boldsymbol{\alpha}_h(P), \mu) = \underline{\Pi}_\ell$*, and service strategy* $\boldsymbol{\beta}^* = (0, 1)$*. When* $\frac{2}{3} < v_\ell < 1$*, a separating equilibrium only exists when* $P_h^*(1 - P_h^*)\Lambda_0 > \underline{\Pi}_h = \frac{1}{4}v_\ell^2\Lambda_0$*. There exists an* $\epsilon > 0$ *such that no separating equilibrium exists for* $1 - \epsilon < v_\ell < 1$*.*

In the separating-price equilibria, the high-quality firm earns a higher profit than the low-quality firm. The cost of separating is lowest at high volumes (low prices) because of the high-quality

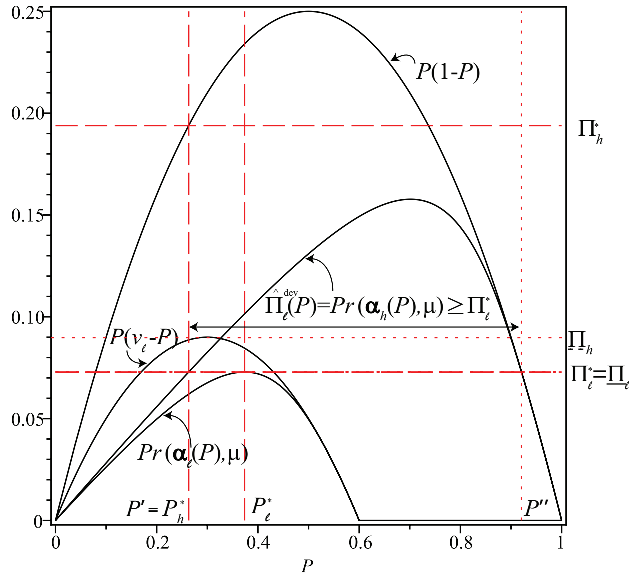

Figure 5 **For Section 5.1. Firm profits and prices under the separating equilibrium when the high-quality firm is fast: $P(v_\ell - P)$ and $P \times r(\alpha_\ell(P), \mu)$. The deviation profits are $\hat{\Pi}_\ell^{\text{dev}}(P)$ and $\hat{\Pi}_h^{\text{dev}}(P)$. The arrow indicates the price region where the low-quality firm's deviation profits are higher than Π_ℓ^*. The high-quality firm's separating price must be outside that region. The high-quality firm's profits at P_h^* (Π_h^*) are higher than at P''. Note also that $\Pi_h^* \geq \underline{\Pi}_h$ (and $\Pi_\ell^* \geq \underline{\Pi}_\ell$).**

firm's service-rate advantage compared to the low-quality firm. As in the former case in Section 4.1 (where the high-quality firm was slow), a separating equilibrium exists only when the high-quality firm adds a high enough premium to the value of the product. The results are similar to the separating-equilibrium results in Proposition 1. If the quality difference between the firms is significant, there always exists a separating equilibrium. When the qualities are comparable, a separating equilibrium may not exist.

5.1.1. Illustrative example: We illustrate the separating prices with the following parameters: $\mu = 0.3$, $N = 7$, $v_\ell = 0.6$, $p = 1/3$. In Figure 5, a separating equilibrium does not exist in this region bounded by P' and P''. In this region, $\hat{\Pi}_\ell^{\text{dev}}(P) \geq \Pi_\ell^*(= \max P(v_\ell - P))$, which indicates that the low-quality firm has an incentive to imitate the high-quality firm, since its deviation profits are higher. As a consequence, there are only two regions that the high-quality firm can consider: $[0, P']$ and $[P'', 1]$, and it can be seen that the profit $P(1-P)$ is higher at P' than at P''. Hence, $P_h^* = P'$.

Note that the high-quality firm still cannot extract the maximum profit (which is $1/4$), because it must keep its price low enough to be able to prevent any imitation from the low-quality firm.

However, the profit difference with the low-quality firm is significant (compared to the separating equilibrium in §4), because the high-quality firm has an advantage in both quality and speed.

5.1.2. When N is large and the arrival rate is high: Let $\underline{\mu} = \mu < 1$, $\overline{\mu} = +\infty$ and $\Lambda_0 = 1$. We easily obtain the limiting prices for large markets (high arrival rates, and a large number of servers): $P_\ell^* = \max(v_\ell - \mu, v_\ell/2)$ and[10]

$$P_h^* = \begin{cases} v_\ell - \mu, & v_\ell - \mu > v_\ell/2 \\ v_\ell^2/(4\mu), & v_\ell - \mu < v_\ell/2. \end{cases}$$

Consider a relatively high value, v_ℓ, such that $v_\ell - \mu < v_\ell/2$ or $\mu < v_\ell/2$. For large, but, finite N, P_ℓ^* and P_h^* will be strictly different, but in the limit P_ℓ^* coincides with P_h^*. Hence, no separating prices exist. For low values of v_ℓ (such that $v_\ell/2 < \mu$), separating prices always exist, even in the limit.[11]

5.2. Pooling Price

Now we analyze pooling equilibria when the high-quality firm is faster. Similar to §4.2, in which the high-quality firm was slow, at any pooling price P^*, we can compute consumers' consistent beliefs and joining strategies when the high-quality firm is fast, $\boldsymbol{\beta}^* = (0, 1)$. Since the high-quality firm is infinitely fast, consumers arriving at a nonempty firm will infer that the quality is low. Thus, for $n \geq 1$, the consumer joining strategy is $[v_\ell - P^*]^+$. Deriving the probability of joining an empty firm is somewhat more involved than in §4.2. Again, we compute a consistent belief at the empty firm and a joining strategy by solving $\alpha_0 = v_\ell + \frac{p}{p+(1-p)\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)}(1 - v_\ell) - P$ for α_0. There may be two roots in $[0, 1]$, characterized by $G_\pm(P, \alpha_1, p/(1-p))$, where $G_\pm(P, \alpha_1, L)$

$$= \begin{cases} \frac{1}{2}\left(1 - \frac{1+L}{F(\alpha_1)L} - P\right)\left(1 \pm \sqrt{1 - 4\frac{L(1-P)+(P-v_\ell)}{F(\alpha_1)L\left(1-\frac{1+L}{F(\alpha_1)L}-P\right)^2}}\right), & \alpha_1 > 0 \text{ and } \frac{L(1-P)+(P-v_\ell)}{F(\alpha_1)L\left(1-\frac{1+L}{F(\alpha_1)L}-P\right)^2} < \frac{1}{4} \\ 1 - P, & \alpha_1 = 0 \end{cases}$$

For some prices, $G_-(P, \alpha_1, p/(1-p))$ may be negative. In that case, G_- does not characterize an equilibrium with a positive joining probability. In this equilibrium, $\boldsymbol{\alpha}^* = (\alpha_0^p(P^*), [v_\ell - P^*]^+, [v_\ell - P^*]^+, \ldots)$, where $\alpha_0^p(P) = [G_\pm(P, [v_\ell - P]^+, p/(1-p))]^+$.[12] With this strategy, we obtain the following results.

[10] P_h^* solves: $P\min(1-P, \mu) = (v_\ell - \mu)\mu$ if $v_\ell - \mu > v_\ell/2$ and $P\min(1-P, \mu) = v_\ell^2/4$ if $v_\ell - \mu < v_\ell/2$. Note also that $\Pi_\ell^* = \underline{\Pi}_\ell$ and $\Pi_h^* \geq \underline{\Pi}_h$.

[11] Notice that, in the limit, the high-quality firm's price can never be $1/2$, which would yield the same profits as in the absence of information asymmetry. When $P_h^* = 1/2$, $v_\ell^2 = 2\mu$ must hold. However, in order to be an equilibrium, $v_\ell < 2\mu$ must be satisfied. As $v_\ell < 1$, this is impossible.

[12] To reduce notational burden, we do not explicitly distinguish between the two roots in $\alpha_0^p(P)$.

Proposition 6. *Suppose that a pooling price, $P^* > 0$, exists such that the high-quality firm selects the highest service rate, $\boldsymbol{\beta}^* = (0,1)$ and consumers join according to $\boldsymbol{\alpha}^* = (\alpha_0^p(P^*), [v_\ell - P^*]^+, [v_\ell - P^*]^+, \ldots)$. P^* is an intuitive pooling price only when*

$$\underline{P}_\ell(P^*) \le \underline{P}_h(P^*) \text{ and } \overline{P}_h(P^*) \le \overline{P}_\ell(P^*) \text{ and } P^* \times r(\boldsymbol{\alpha}(P^*), \mu) > \underline{\Pi}_\ell,$$

$$\text{where } \underline{P}_h(P) = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4P\alpha_0^p(P)} \text{ and } \overline{P}_h(P) = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4P\alpha_0^p(P)}$$

$$\text{and } \underline{P}_\ell(P) \text{ and } \overline{P}_\ell(P) \text{ are the roots of } P' \times r(\boldsymbol{\alpha}_h(P'), \mu) = P \times r(\boldsymbol{\alpha}^p(P), \mu).$$

No intuitive pooling prices with $0 < \beta_h^* < 1$ exist.

Interestingly, the intuition for pooling prices when the high-quality firm is fast is fundamentally different from the intuition for pooling prices when the high-quality firm is slow. In the latter case, the low-quality firm's profit decreases in service rate due to the empty-restaurant syndrome; the busier the restaurant, the higher the rate at which consumers join.

In contrast, when the high-quality firm is fast, the joining rate *always* decreases with congestion level at the firm, as increased congestion is associated with low quality. Therefore, the high-quality firm's revenues always increase with the service rate. In addition, as the high service rate comes at no extra cost, the high-quality firm will *never* randomize service rates. This drives a wedge between the high- and low-quality firms' profits, which makes satisfying the intuitive criterion harder in pooling equilibria than in §4.2.2.

Furthermore, $\alpha_0^p(P)$ may have *two* roots in $(0,1)$. In that case, there are two equilibria. Interestingly, equilibria with positive joining probability may occur at prices that are strictly higher than v, which is the expected valuation. This is because even though there is no information contained in the prices (as both firms charge the same price), there is some positive information about the quality when the firm is empty upon arrival, as the high-quality firm is assumed to be fast (in equilibrium). Therefore, even when the price is *higher* than the expected prior valuation (v), observing an empty firm may make the consumers join it. This is because the prior belief that the quality is high improves on observing an empty firm.

5.2.1. Illustrative example: We illustrate the profits of the high- and low-quality firms under pooling equilibria, $P \times r(\boldsymbol{\alpha}^p(P), \overline{\mu})$ and $P \times r(\boldsymbol{\alpha}^p(P), \underline{\mu})$, in Figure 6. Note that the kink is where $P = v_\ell (= 0.6)$. Above v_ℓ, consumers who observe that one slot (or more) is occupied upon arrival,

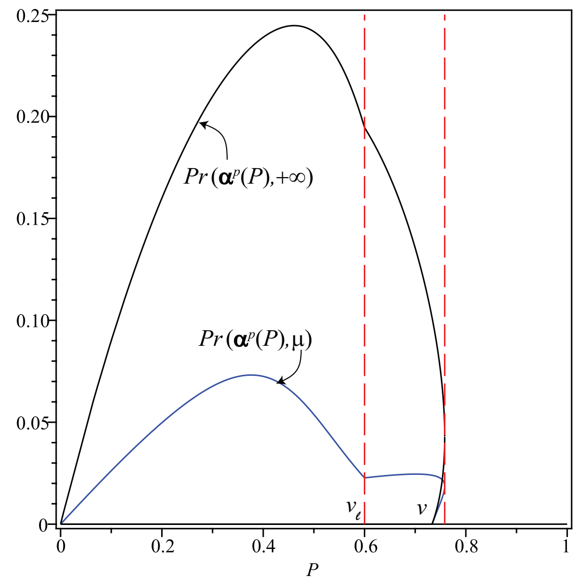

Figure 6 (Corresponds to Section 5.2). Illustration of $P \times r(\alpha^p(P), \overline{\mu})$ and $P \times r(\alpha^p(P), \underline{\mu})$. The parameter values are the same as in Figure 5.

know *for sure* that the quality is low (the consumer at the slot must have arrived when the firm was empty). Therefore, consumers join the firm only when it is empty (irrespective of the firm type). Notice that there exists two equilibria (of which the corresponding profits are plotted) for prices above the prior expected value, v. These correspond to the two possible roots (G_+ and G_-) when solving $\alpha_0 = v_\ell + \frac{p}{p+(1-p)\pi(0,(\alpha_0,\alpha_1,\alpha_1,\ldots),\mu)}(1-v_\ell) - P$ for α_0 in $[0,1]$.

As for any price, P, the extra revenues from speeding up are always positive, $\Delta^p(P) > 0$. No equilibria exist in which the high-quality firm randomizes its service rate. We verified the conditions for a pooling price. The solid lines in Figure 7 indicate the firm profits at the pooling prices.

5.2.2. When N is large and the arrival rate is high: The pooling prices and profits are determined by the following equilibrium joining strategies.

Proposition 7. *Let $\underline{\mu} = \mu < 1$, $\overline{\mu} = +\infty$ and $\Lambda_0 = 1$. For $0 \le P < 1$, $\lim_{N\to+\infty} \alpha_0^p(P) = 1 - P$, $\alpha^*(n,P) = (v_\ell - P)^+$ for $1 \le n$, $\boldsymbol{\beta}^* = (0,1)$, $\lim_{N\to+\infty} r(\boldsymbol{\alpha}^*, +\infty) = 1 - P$, and $\lim_{N\to+\infty} r(\boldsymbol{\alpha}^*, \mu) = \min(\mu, [v_\ell - P]^+)$.*

(i). *When $\mu < v_\ell/2$ and $\mu < v_\ell - 1/2$, no pooling price exists. When $\mu < v_\ell/2$ and $v_\ell - 1/2 < \mu$, $P^* = v_\ell - \mu$ is the unique pooling price.*

(ii). *When $v_\ell/2 < \mu < v_\ell$, $P^* = \frac{v_\ell}{2}$ is the unique pooling price.*

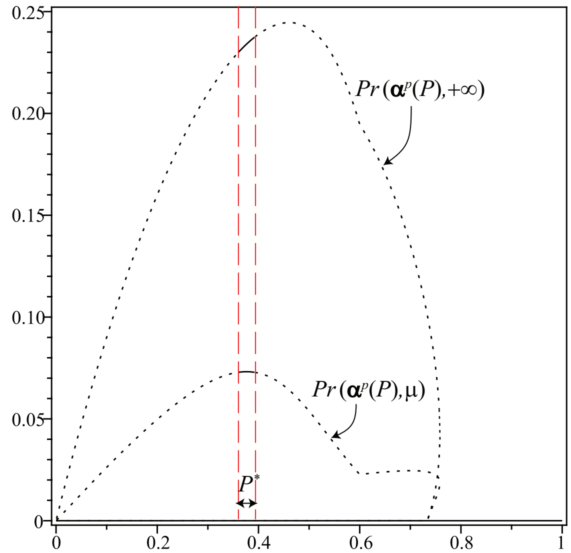

Figure 7 **(Corresponds to Section 5.2) Pooling profits and prices. The solid curves between the vertical dotted lines indicate the pooling profits that survive the intuitive criterion. Note that the existence region for the pooling equilibrium is much reduced when the high-quality firm is faster. Unlike in Section 4.2, where the firm profits were identical (see Figure 3), here the pooling profits of the high-quality firm are higher than those of the low-quality firm due to its service rate advantage.**

As discussed in Proposition 6, there exist pooling prices that are strictly higher than the ex ante prior valuation v. When the number of slots is large, the empty firm will *perfectly reveal* that the quality is high. Hence, at any price below $v_h(=1)$, the empty firm will attract consumers. The fast, high-quality firm enjoys a high joining rate as an empty firm and is not constrained in its throughput. Unfortunately the slow, low-quality firm cannot enjoy this increased joining rate as its profit rate is determined by the joining rate when the firm is not empty (see Lemma 3) and its finite capacity.

If the low-quality firm is substantially slower than the high-quality firm, then the pooling equilibrium does not exist. But as long as $v_\ell - 1/2 < \mu < v_\ell$, (covered in cases (i) and (ii) in Proposition 7), a unique pooling price exists for which the high-quality firm selects a high rate with probability one (and the low-quality firm selects a low rate).

5.2.3. Illustrative example: The case when $\mu < v_\ell/2$ (corresponding to Proposition 7(i)) is illustrated in Figure 8 in the left and middle panels. In the left panel, the unique candidate pooling price, $v_\ell - \mu$, does not satisfy the intuitive-criterion conditions. This is because the *high-quality* firm has an incentive to deviate to, for example, 1/2, where the intuitive criterion forces the belief

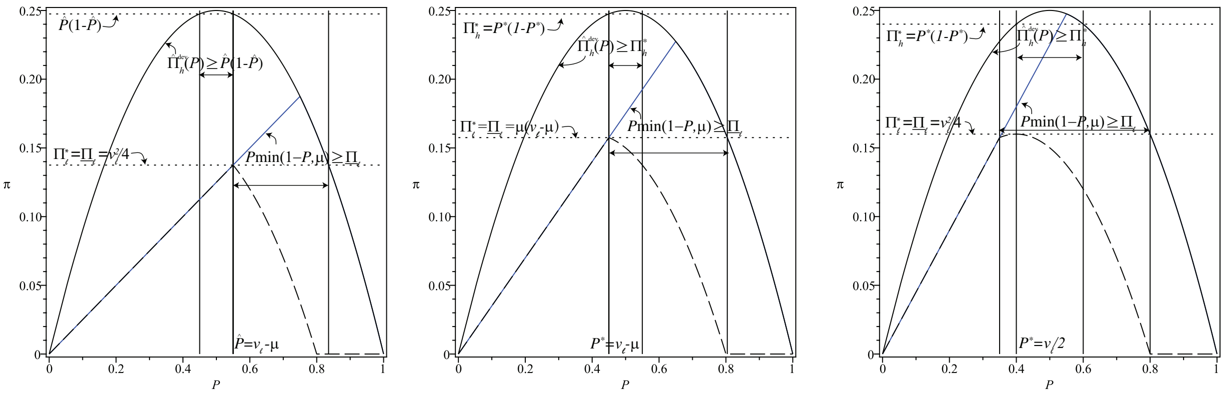

Figure 8 Illustration of the conditions for the intuitive criterion with $P^* = \max(\frac{v_\ell}{2}, v_\ell - \mu)$: $\hat{\Pi}_h^{\text{dev}}(P)$ and $\hat{\Pi}_\ell^{\text{dev}}(P)$ (solid lines), $P\min(v_\ell - P, \mu)$ (dashed line), and Π_ℓ^* and Π_h^* (dotted lines) for $\mu_0 = 0.25$ (left) 0.35 (middle) and 0.45 (right) and $v_\ell = 0.8$.

to be 1 (as only the high-quality firm, not the low-quality firm improves its profits). Hence, with this belief, $v_\ell - \mu$ is not rational for the high-quality firm.

In the middle panel, the pooling price, $v_\ell - \mu$, does satisfy the intuitive criterion, as for any deviation of the high-quality firm, the low-quality firm also has an incentive to deviate, the intuitive-criterion conditions do not restrict the off-equilibrium-path beliefs. Notice that when $v_\ell/2$ is slightly below μ, the high-quality firm's profits are about $\frac{1}{4}$, which are its profits if the firm were able to credibly communicate its service rate to consumers. Hence, the low service rate of the low-quality firm may solve the high-quality firm's communication problem with pooling prices.

The case $v_\ell/2 < \mu < v_\ell$ (corresponding to Proposition 7(ii)) is illustrated in Figure 8 in the right panel. The pooling price, $v_\ell/2$, does satisfy the intuitive-criterion condition as for any deviation of the high-quality firm, the low-quality firm also has an incentive to deviate; hence, again the intuitive-criterion conditions do not restrict the off-equilibrium-path beliefs.

6. Concluding Discussion

Our goal in this paper is to model how congestion level at a firm may serve as a signal of the firm's quality. In particular, our research is motivated by the observation that many firms encourage congestion, implying that they always look busy with excess demand. A key question is why a high-quality firm would not simply raise its price without affecting the fulfilled demand. Furthermore, a high price may signal high quality, and thus helps a firm to be perceived as a good firm.

We have addressed this question by developing a theoretical model in which a firm, knowing the quality of its own product, but unable to communicate that quality to its consumer base, sets a price and selects an (unobserved) service process. Consumers infer quality from the price and from the number of other customers (congestion) in the system upon arrival.

6.1. When the High-Quality Firm Is Slow

In many cases, it is more expensive for the high-quality firm to speed up while maintaining the same quality. For instance, a high-quality firm may produce a truly innovative product at a pre-existing facility. Therefore, the process speed has to be carefully slow. A high-quality restaurant pays attention to minute details of service, and therefore cannot costlessly increase its service rate without sacrificing details.

We find that when the high-quality firm provides a significantly additional value for consumers compared to the low-quality firm, it can prevent a low-quality firm from imitating it by setting a high price. However, this price separation comes at a significant cost. The separating high-quality firm's profits are lower than the low-quality firm's profits reflecting that the high-quality firm is handicapped in two ways: (i) it cannot distinguish itself in any way other than setting a different price from the low-quality firm and, in addition, (ii) it cannot turn potential consumers into sales as efficiently as the low-quality firm due to its lower service rate. With separating prices, the congestion upon arrival does not contain any additional information for the consumers beyond the price.

In equilibria with pooling prices, the congestion upon arrival becomes informative: The high-quality firm is expected to be more congested (controlling for slow-down incentives). Consumers are more eager to join when the restaurant is full than when the restaurant is empty, leading to the empty-restaurant syndrome. Our model confirms the anecdotal observations that empty businesses may not be attractive to consumers.

An important consequence of the empty-restaurant syndrome is that it renders the low-quality firm's speed advantage useless, especially when the price is high enough. Therefore, at high prices, the low-quality firm garbles its service rate, even when it can costlessly improve service speed. This, in turn, reduces the profit gap between the high- and low-quality firms, which supports a pooling-price equilibrium over a range of prices if (i) the high-quality firm's service rate is high enough or (ii) the high-quality firm's service rate is low, but the consumers' prior of the quality is high enough. Thus, pooling is not supported in equilibrium when the high-quality firm's service rate is too low and the prior of the quality is also low.

6.2. When the High-Quality Firm Is Fast

We also study the case in which the high-quality firm is able to produce the innovative product using a faster production process than the low-quality firm. Given that the high-quality firm has a speed advantage, which is more pronounced at high demand volumes, the high-quality firm now prevents imitation by the low-quality firm through low prices (and high volumes). Nevertheless, separating from the low-quality firm is still expensive for the high-quality firm.

Interestingly, the pooling equilibrium is also quite different: The empty-restaurant syndrome does not exist; consumers are more enticed to join an empty firm because emptiness is a sign of high quality (since the high-quality product is produced very quickly). Nevertheless, pooling-price equilibria exist if the service rate of the low-quality firm is high enough. Interestingly, the prior is less important especially when the number of slots becomes large. The reason is that low congestion may become such a strong signal of quality that it almost completely separates the high-quality firm from the low-quality firm, even with pooling prices, irrespective of the prior. There is a smaller range of pooling-equilibrium prices when the high-quality firm is fast, which collapses to a single price when the number of slots tends to infinity.

We can thus conclude that in our model, irrespective of which firm has the service-rate advantage, congestion levels are better signaling instruments than prices alone.

6.3. Assumptions, Extensions and Further Research

To keep the analysis tractable, we have assumed that the fast service is infinitely fast and free. This assumption allowed us to utilize the concavity properties of the profit functions in the price and establish intuitive-criterion refinements. However, our main insights (that pooling prices exist, they are lower than separating prices, and they lead to higher profits in equilibrium) continue to hold without this assumption. In the technical Appendix B, we derive conditions for pooling and separating equilibria when the high service rate is finite and expensive.

We also assumed no reneging in our model. As the service rate is not known, but is informative about the product quality under the pooling equilibrium, consumers could collect more information by 'waiting' in their service, reneging and revisiting the firm. Incorporating reneging decisions would make the consumer decision significantly more difficult and is left for future considerations.

Ours is a single, static, service-rate-determination problem, without any transient effects or even state dependencies. However, a firm could dynamically adjust its service rate based on congestion levels without dynamic pricing decisions (as in Ata and Schneorsen 2006). In such a model, the equilibrium service rate would be a multidimensional vector and the consumers' value updating problem would depend on the entire vector. The best response of each type of firm is the optimal solution to a dynamic program. In equilibrium, two dynamic programs and a complicated consumer inference problem need to be solved jointly. Thus, there may be dimensionality issues.

Finally, we assume that consumers do not observe the selected service rate, only the congestion upon arrival. This is a relaxation of the common-knowledge assumption in many queuing games. In our model, consumers know the costs associated with the different service rates. Considering an imperfect consumer knowledge about the costs would surely add another layer of complexity.

In addition, we assume that the firm can only select from two service processes (one of which is instantaneous service). We conjecture that considering a service process with multiple service-rate alternatives (or a continuum of service rates) would not change the main insights in our paper, but would make the analysis quite intractable under the equilibrium-refinement requirements.

To conclude, we have focused on what we believe to be two fundamental sources of quality information in a queuing context: the price and number of consumers in the system. It would be interesting to enrich the model with other potential sources of learning such as repeat purchases, word-of-mouth effects, and networking with informed consumers (see, e.g., Debo et al., 2010). We believe that modeling such learning mechanisms would surely enhance our understanding of how 'operational' signals influence consumer behavior. We also believe that these mechanisms represent a challenging but interesting future research agenda.

References

Afèche, P. 2005. Revenue Management and Delay Tactics for Competing Service Providers in Time-Sensitive Markets with Private Information. Working Paper, University of Toronto.

Allon, G., A. Bassamboo and I. Gurvich. 2007. We Will be Right with You: Managing Customers with Vague Promises. Working paper, Kellogg School of Management, Northwestern University.

Anand, K. and M. Goyal. 2009. Strategic Information Management under Leakage in a Supply Chain. Management Science, 55(3), 438–452.

Akan, M., B. Ata and M. A. Lariviere. 2007. Asymmetric Information and Economies-of-Scale in Service Contracting. Working Paper, Northwestern University.

Ata, B. and S. Shneorsen. 2006. Dynamic Control of an M/M/1 Service System with Adjustable Arrival and Service Rates. Management Science, 52(11), 1778–1791

Bagwell, K. and M. H. Riordan. 1991. High and Declining Prices Signal Product Quality. The American Economic Review, 81(1), 224–239.

Becker, G. S., 1991. A Note on Restaurant Pricing and Other Examples of Social Influences on Price. The Journal of Political Economy, 99(5), 1109–1116.

Cachon, G. and M. Lariviere. 2001. Contracting to Assure Supply: How to Share Demand Forecasts in a Supply Chain. Management Science, 47(5), 629–646.

Cho, I. and D. M. Kreps. 1987. Signaling Games and Stable Equilibria. Quarterly Journal of Economics, 102(May), 179–221.

Debo, L. G., C. Parlour and U. Rajan, 2010. Inferring Quality from a Queue. Working Paper, Booth School of Business, University of Chicago.

Edelson, N. and D. K. Hildebrand. 1975. Congestion Tolls for Poisson Queuing Processes. Econometrica. 43(1), 81–92.

Guo, P. and R. Hassin. 2009. Strategic Behavior and Social Optimization in Markovian Vacation Queues. Working Paper, Hong Kong University.

Guo, P. and P. Zipkin. 2007. Analysis and Comparison of Queues with Different Levels of Delay Information. Management Science, 53(6), 962–970.

Hassin, R. and M. Haviv. 2003. To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems. Kluwer Academic Publishers.

Milgrom, P. and J. Roberts. 1986. Price and Advertising Signals of Product Quality, Journal of Political Economy, August 1986, 94, 796–821.

Moorthy, S. and K. Srinivasan. 1995. Signaling Quality with a Money-Back Guarantee: The Role of Transaction Costs. Marketing Science, 14(4), 442–466

Naor, P. 1969. The Regulation of Queue Size by Levying Tolls. Econometrica, 37, 15–34.

Orzach, R., P. B. Overgaard and Y. Tauman. 2002. Modest Advertising Signals Strength. The RAND Journal of Economics, 33(2), 340–358.

Spence, M. 1973. Job Market Signaling. Quarterly Journal of Economics, 87(3), 355–374.

Taylor, C. 1999. Time-on-the-Market as a Sign of Quality. Review of Economic Studies. 66(3), 555–578.

Veeraraghavan, S. and L. G. Debo. 2009. Joining Longer Queues: Information Externalities in Queue Choice. M&SOM, 11(4), 543–562.

Veeraraghavan, S. and L. G. Debo. 2010. Herding in Queues with Waiting Costs: Rationality and Regret. Wharton Working Paper.

Wolff, R. W. 1982. Poisson arrivals see time averages. Operations Research, 30, 223–231.

Appendix A. Proofs of Lemmas and Propositions

Proof of Lemma 1: Follows from the steady state probabilities of birth and death processes. See, for example, Ross (1996). ■

Proof of Lemma 2: (i) For notational convenience, we rewrite $Nr((\alpha,\alpha,\alpha,\ldots),\mu) = \sum_{n=0}^{N-1} a\pi(n,(a,a,\ldots),\mu)$ as $R(a)$ using the expression of $\pi(n,(a,a,\ldots),\mu)$ from Lemma 1. As $\prod_{m=0}^{k-1}\alpha(m) = a^k$, with $z_k = \frac{(1/\mu)^k}{k!}$, we can write

$$R(a) = a\frac{\sum_{k=0}^{N-1} z_k a^k}{\sum_{k=0}^{N} z_k a^k} = \frac{\overbrace{\sum_{k=1}^{N} z_{k-1} a^k}^{=A(a)}}{\underbrace{\sum_{k=0}^{N} z_k a^k}_{=B(a)}}.$$

Taking the first and second derivatives, we obtain

$$\begin{aligned}
\frac{\mathrm{d}R(a)}{\mathrm{d}a} &= \frac{\mathrm{d}}{\mathrm{d}a}\frac{A(a)}{B(a)} = \frac{A'(a)B(a) - A(a)B'(a)}{(B(a))^2} \\
\frac{\mathrm{d}^2R(a)}{\mathrm{d}a^2} &= \frac{\mathrm{d}}{\mathrm{d}a}\frac{A'(a)B(a) - A(a)B'(a)}{(B(a))^2} \\
&= \frac{\left(A''(a)B(a) - A(a)B''(a)\right)(B(a))^2 - 2\left(A'(a)B(a) - A(a)B'(a)\right)B(a)B'(a)}{(B(a))^4}.
\end{aligned}$$

For notational convenience, we drop a from the arguments and add $_N$ to R, A and B. Thus, as $B'_N > 0$, we obtain that

$$\frac{A'_N}{A_N} - \frac{B'_N}{B_N} > 0 \Rightarrow \frac{\mathrm{d}R}{\mathrm{d}a} > 0 \tag{9}$$

$$\frac{A''_N B_N - A_N B''_N}{A'_N B_N - A_N B'_N} < \frac{2B'_N}{B_N} \Leftrightarrow \frac{\frac{A''_N}{A_N} - \frac{B''_N}{B_N}}{\frac{A'_N}{A_N} - \frac{B'_N}{B_N}} < 2\frac{B'_N}{B_N} \Rightarrow \frac{\mathrm{d}^2R}{\mathrm{d}a^2} < 0. \tag{10}$$

Via Equations (9) and (10), we obtain that R is concavely increasing in a when $\frac{A'_N}{A_N} - \frac{B'_N}{B_N} > 0$ and

$$\frac{\frac{A''_N}{A_N} - \frac{B''_N}{B_N}}{\frac{A'_N}{A_N} - \frac{B'_N}{B_N}} < 2\frac{B'_N}{B_N}.$$

Now, we prove the latter two inequalities. We obtain the following expressions for A'_N, B'_N, A''_N and B''_N from simple derivation with respect to a:

$$\begin{aligned} A'_N &= \textstyle\sum_{k=1}^{N} k z_{k-1} a^{k-1}, B'_N = \sum_{k=1}^{N} k z_k a^{k-1} \text{ and} \\ A''_N &= \textstyle\sum_{k=2}^{N} k(k-1) z_{k-1} a^{k-2}, B''_N = \sum_{k=2}^{N} k(k-1) z_k a^{k-2}. \end{aligned}$$

We can describe the following relationships.

$$\begin{aligned} k z_k &= k\frac{(1/\mu)^k}{k!} = (1/\mu)\frac{(1/\mu)^{k-1}}{(k-1)!} = (1/\mu) z_{k-1} \\ k(k-1) z_k &= (1/\mu)^2 \frac{(1/\mu)^{k-2}}{(k-2)!} = (1/\mu)^2 z_{k-2} \\ k(k-1) z_{k-1} &= k(k-1) z_k \frac{z_{k-1}}{z_k} = (1/\mu)^2 z_{k-2}\frac{z_{k-1}}{z_k} = (1/\mu)^2 z_{k-2} k\mu. \end{aligned}$$

$$\begin{aligned} \text{Therefore: } A'_N &= \textstyle\sum_{k=0}^{N-1} (k+1) z_k a^k = \sum_{k=0}^{N-1} z_k a^k + \sum_{k=1}^{N-1} k z_k a^k \\ &= \textstyle\sum_{k=0}^{N-1} z_k a^k + (1/\mu)\sum_{k=1}^{N-1} z_{k-1} a^k = \sum_{k=0}^{N-1} z_k a^k + (1/\mu) a \sum_{k=0}^{N-2} z_k a^k \\ B'_N &= \textstyle (1/\mu)\sum_{k=1}^{N} z_{k-1} a^{k-1} = (1/\mu)\sum_{k=0}^{N-1} z_k a^k \\ A''_N &= \textstyle (1/\mu)^2 \sum_{k=2}^{N} z_{k-2} k\mu a^{k-2} = (1/\mu)^2 \sum_{k=0}^{N-2} z_k (k+2)\mu a^k \\ &= \textstyle (1/\mu)^2 \sum_{k=1}^{N-2} z_{k-1} a^k + 2(1/\mu)\sum_{k=0}^{N-2} z_k a^k \\ &= \textstyle (1/\mu)^2 \sum_{k=0}^{N-3} z_k a^{k+1} + 2(1/\mu)\sum_{k=0}^{N-2} z_k a^k \\ B''_N &= \textstyle (1/\mu)^2 \sum_{k=0}^{N-2} z_k a^k \end{aligned}$$

and we can write sufficient conditions for $\frac{\mathrm{d}R}{\mathrm{d}a} > 0$ and $\frac{\mathrm{d}^2 R}{\mathrm{d}a^2} < 0$ as:

$$\frac{A'_N}{A_N} - \frac{B'_N}{B_N} > 0 \Leftrightarrow \frac{\mu}{a} + \frac{\sum_{k=0}^{N-2} z_k a^k}{\sum_{k=0}^{N-1} z_k a^k} - \frac{\sum_{k=0}^{N-1} z_k a^k}{\sum_{k=0}^{N} z_k a^k} > 0 \Leftrightarrow \mu + R_{N-1} - R_N > 0 \tag{11}$$

and

$$\frac{\frac{A''_N}{A_N} - \frac{B''_N}{B_N}}{\frac{A'_N}{A_N} - \frac{B'_N}{B_N}} < 2\frac{B'_N}{B_N} \Leftrightarrow \frac{\frac{\sum_{k=0}^{N-3} z_k a^k}{\sum_{k=0}^{N-1} z_k a^k} - \frac{\sum_{k=0}^{N-2} z_k a^k}{\sum_{k=0}^{N} z_k a^k} + 2\frac{\mu}{a}\frac{\sum_{k=0}^{N-2} z_k a^k}{\sum_{k=0}^{N-1} z_k a^k}}{\frac{\mu}{a} + \frac{\sum_{k=0}^{N-2} z_k a^k}{\sum_{k=0}^{N-1} z_k a^k} - \frac{\sum_{k=0}^{N-1} z_k a^k}{\sum_{k=0}^{N} z_k a^k}} < 2\frac{\sum_{k=0}^{N-1} z_k a^k}{\sum_{k=0}^{N} z_k a^k}$$

Now, we rewrite the above inequality as

$$\frac{\frac{1}{a^2}\frac{a\sum_{k=0}^{N-3} z_k a^k}{\sum_{k=0}^{N-2} z_k a^k}\frac{a\sum_{k=0}^{N-2} z_k a^k}{\sum_{k=0}^{N-1} z_k a^k} - \frac{1}{a^2}\frac{a\sum_{k=0}^{N-2} z_k a^k}{\sum_{k=0}^{N-1} z_k a^k}\frac{a\sum_{k=0}^{N-1} z_k a^k}{\sum_{k=0}^{N} z_k a^k} + 2\frac{\mu}{a^2}R_{N-1}}{\frac{\mu}{a}+\frac{1}{a}R_{N-1}-\frac{1}{a}R_N} < \frac{2}{a}R_N \Leftrightarrow$$
$$\frac{\frac{1}{a^2}R_{N-2}R_{N-1} - \frac{1}{a^2}R_{N-1}R_N + 2\frac{\mu}{a^2}R_{N-1}}{\frac{\mu}{a}+\frac{1}{a}R_{N-1}-\frac{1}{a}R_N} < \frac{2}{a}R_N.$$

Simplifying further, we obtain

$$\frac{R_{N-2}R_{N-1} - R_{N-1}R_N + 2\mu R_{N-1}}{\mu + R_{N-1} - R_N} < 2R_N \Leftrightarrow$$
$$\frac{1}{2}\left(\left(R_{N-1}-R_{N-2}\right)+\left(R_N - R_{N-1}\right)\right)R_{N-1} + \mu\left(R_N - R_{N-1}\right) > R_N\left(R_N - R_{N-1}\right).$$

We thus obtain a sufficient condition for the above inequality: If

$$R_N\left(R_N - R_{N-1}\right) - \left(R_N - R_{N-1}\right)R_{N-1} < \mu\left(R_N - R_{N-1}\right) \Leftrightarrow \left(R_N - R_{N-1}\right)^2 < \mu\left(R_N - R_{N-1}\right)$$
$$\Leftrightarrow \left(R_N - R_{N-1}\right) < \mu, \tag{12}$$

then the above inequality is satisfied. Notice that Equations (11) and (12) are identical ($R_N - R_{N-1} < \mu$) and sufficient conditions for Equations (9) and (10). This condition is intuitive: It states that the throughput never increases by more than the capacity of a server when a server is added to an Erlang-loss system. We now prove this statement. Messerli (1972) proved that the blocking probability, L_N, is convex in N, in other words,

$$R_N = \lambda\left(1 - \overbrace{\frac{z_N\lambda^N}{\sum_{k=0}^{N} z_k\lambda^k}}^{=L_N=\text{blocking probability}}\right).$$

The result from Messerli (1972) implies that R_N is concave in N. As

$$R_2 - R_1 = \lambda\left(\frac{z_1\lambda}{\sum_{k=0}^{1} z_k\lambda^k} - \frac{z_2\lambda^2}{\sum_{k=0}^{2} z_k\lambda^k}\right) = \lambda\left(\frac{(1/\mu)\lambda}{1+(1/\mu)\lambda} - \frac{\frac{(1/\mu)^2}{2}\lambda^2}{1+(1/\mu)\lambda+(1/\mu)^2\lambda^2/2}\right)$$
$$= \mu\frac{1+2\frac{\mu}{\lambda}}{\left(1+\frac{2\mu}{\lambda}+2\frac{\mu^2}{\lambda^2}\right)\left(1+\frac{\mu}{\lambda}\right)} < \mu$$

and $R_N - R_{N-1}$ is decreasing in N, it follows that for all N, $R_N - R_{N-1} < \mu$ As a result, R_N is concave, increasing in λ.

(ii) Now, we prove that $\Delta((a,a,\ldots)) > 0$. We take the derivative of R with respect to μ and show that the derivative is positive:

$$\frac{\mathrm{d}}{\mathrm{d}\mu}R = \frac{\mathrm{d}}{\mathrm{d}\mu}\left(1 - \left(\frac{1}{N!}\left(\frac{a}{\mu}\right)^N\right) / \left(\sum_{k=0}^N \frac{1}{k!}\left(\frac{a}{\mu}\right)^k\right)\right)$$

$$= a\frac{1}{(N-1)!}\left(\frac{a}{\mu}\right)^{N-1} \frac{1 - \frac{1}{N}\frac{a}{\mu}\frac{\sum_{k=0}^{N-1}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k}{\sum_{k=0}^{N}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k}}{\sum_{k=0}^{N}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k}\frac{1}{\mu^2}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}R > 0 \Leftrightarrow \frac{N\mu}{a} > \frac{\sum_{k=0}^{N-1}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k}{\sum_{k=0}^{N}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k} = \frac{\mu}{a}\frac{\sum_{k=0}^{N}\frac{k}{k!}\left(\frac{a}{\mu}\right)^k}{\sum_{k=0}^{N}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k}.$$

It is easily observed that

$$\frac{\sum_{k=0}^{N}\frac{k}{k!}\left(\frac{a}{\mu}\right)^k}{\sum_{k=0}^{N}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k} \leq \frac{\sum_{k=0}^{N}\frac{N}{k!}\left(\frac{a}{\mu}\right)^k}{\sum_{k=0}^{N}\frac{1}{k!}\left(\frac{a}{\mu}\right)^k} = N$$

because $k \leq N$ in the summation. The inequality follows. ■

We now introduce an auxiliary lemma that will be useful in the following lemmas and propositions.

LEMMA 5. *For any $a > 0$: $H \doteq \lim_{N\to+\infty} H_N = \begin{cases} 1-\frac{1}{a}, & a > 1 \\ 0, & 0 < a < 1 \end{cases}$, where $H_N \doteq \frac{(aN)^N \frac{1}{N!}}{\sum_{k=1}^N \frac{1}{k!}(aN)^k}$.*

Proof of Lemma 5: We rewrite the limit

$$\frac{1}{1-H_N} = \frac{\sum_{k=1}^N \frac{1}{k!}(aN)^k}{\sum_{k=1}^{N-1}\frac{1}{k!}(aN)^k} = \frac{\sum_{k=1}^N a^k\frac{N^k}{k!}}{\sum_{k=1}^{N-1} a^k\frac{N^k}{k!}}$$

Notice that $\frac{N^N}{N!} \to +\infty$ when N is large. This is because, with Stirling's formula for $N!$,

$$\frac{N^N}{N!} \approx \frac{N^N}{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N} = \frac{e^N}{\sqrt{2\pi N}}.$$

If $a > 1$, then the highest power dominates: $\lim_{N\to+\infty}\frac{\sum_{k=1}^N a^k\frac{N^k}{k!}}{\sum_{k=1}^{N-1} a^k\frac{N^k}{k!}} = \lim_{N\to+\infty}\frac{a^N\frac{N^N}{N!}}{a^{N-1}\frac{N^{N-1}}{(N-1)!}} = \lim_{N\to+\infty} a\frac{N^N}{N^{N-1}}\frac{(N-1)!}{N!} = a$. If $a < 1$, then the lowest power dominates:

$$\lim_{N\to+\infty}\frac{\sum_{k=1}^N a^k\frac{N^k}{k!}}{\sum_{k=1}^{N-1} a^k\frac{N^k}{k!}} = \lim_{N\to+\infty}\frac{a\frac{1}{1!}}{a\frac{1}{1!}} = 1.$$

It follows that $\lim_{N\to+\infty}\frac{1}{1-H_N} = \max(1,a)$, or $H = \max(0, 1-\frac{1}{a})$. ■

Proof of Lemma 3: The throughput (per slot) is

$$
\begin{aligned}
r(\boldsymbol{\alpha},\mu) &= \Lambda_0\alpha_0\pi_0(\boldsymbol{\alpha},\mu) + \Lambda_0\sum_{k=1}^{N-1}\alpha_1\frac{(N\alpha_0)(N\alpha_1)^{k-1}}{k!}\left(\frac{1}{\mu}\right)^k\pi_0(\boldsymbol{\alpha},\mu)\\
&= \Lambda_0\alpha_0\frac{1+(\alpha_1 N)\sum_{k=1}^{N-1}\frac{(N\alpha_1)^{k-1}}{k!}\left(\frac{1}{\mu}\right)^k}{1+\sum_{k=1}^{N}\frac{(N\alpha_0)(N\alpha_1)^{k-1}}{k!}\left(\frac{1}{\mu}\right)^k} = \Lambda_0\alpha_0\frac{1+\alpha_1\sum_{k=1}^{N-1}\frac{(\alpha_1)^{k-1}}{k!}\left(\frac{1}{\mu/N}\right)^k}{1+\alpha_0\sum_{k=1}^{N}\frac{(\alpha_1)^{k-1}}{k!}\left(\frac{1}{\mu/N}\right)^k}\\
&= \Lambda_0\alpha_0\frac{1+\alpha_1\overbrace{\sum_{k=1}^{N}\frac{\alpha_1^{k-1}}{k!}\left(\frac{1}{\mu/N}\right)^k}^{=F(\alpha_1,\mu/N)} - \alpha_1\frac{(\alpha_1)^{N-1}}{N!}\left(\frac{1}{\mu/N}\right)^N}{1+\alpha_0\underbrace{\sum_{k=1}^{N}\frac{\alpha_1^{k-1}}{k!}\left(\frac{1}{\mu/N}\right)^k}_{=F(\alpha_1,\mu/N)}},
\end{aligned}
$$

with $F(\alpha,\mu) = \sum_{k=1}^{N}\frac{\alpha^{k-1}}{k!}\left(\frac{1}{\mu}\right)^k$. Hence, we rewrite

$$
\begin{aligned}
\lim_{N\to+\infty} r(\boldsymbol{\alpha},\mu) &= \alpha_0\Lambda_0\lim_{N\to+\infty}\frac{1+\alpha_1 F(\alpha_1,\mu/N) - \alpha_1^N\frac{(N/\mu)^N}{N!}}{1+\alpha_0 F(\alpha_1,\mu/N)}\\
&= \alpha_0\Lambda_0\lim_{N\to+\infty}\left(1-\frac{(\alpha_0-\alpha_1)F(\alpha_1,\mu/N)+\alpha_1^N\frac{(N/\mu)^N}{N!}}{1+\alpha_0 F(\alpha_1,\mu/N)}\right).
\end{aligned}
$$

With Lemma 5, we obtain that $\lim_{N\to+\infty}\frac{\left(\frac{\alpha_1}{\mu/N}\right)^N\frac{1}{N!}}{F(\alpha_1,\mu/N)} = (\alpha_1-\mu)^+$. Hence,

$$
\begin{aligned}
\lim_{N\to+\infty} r(\boldsymbol{\alpha},\mu) &= \alpha_0\Lambda_0\lim_{N\to+\infty}\left(1-\frac{\alpha_0-\alpha_1}{\alpha_0}-\frac{\max(0,\alpha_1-\mu)}{\alpha_0}\right) = \alpha_1\Lambda_0 - \Lambda_0\max(0,\alpha_1-\mu)\\
&= \Lambda_0\min\{\alpha_1,\mu\}.
\end{aligned}
$$

When $\mu = +\infty$, $\lim_{N\to+\infty}\pi_0(\boldsymbol{\alpha},\mu) = 1$, so $\lim_{N\to+\infty} r(\boldsymbol{\alpha},\mu) = \Lambda_0\alpha_0$. ■

Proof of Lemma 4: We prove that for any $\boldsymbol{\alpha}$, the throughput, and hence, the firm's profits increase in any $\alpha(n)$. To maximize the firm's profit, $\alpha(n)$ must be as high as possible. This is achieved by selecting $\gamma^o = 1$ in the set $\Gamma(n,P)$ as $\alpha(n)$ increases in $\gamma(n)$, and $\gamma(n)$ increases in γ^o, while γ^o is bounded from above by 1.

Now, we prove that $\frac{\partial r(\boldsymbol{\alpha},\mu)}{\partial\alpha(n)} > 0$. Let $\rho = \Lambda_0 N/\mu$.

$$
\begin{aligned}
r(\boldsymbol{\alpha},\mu) &= \frac{\sum_{k=0}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)}{1+\sum_{k=1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)}\Lambda_0, \text{ where } z_k = \rho^k/k! \text{ and}\\
\frac{\partial}{\partial\alpha(n)}r(\boldsymbol{\alpha},\mu) > 0 &\Leftrightarrow \frac{\frac{\partial}{\partial\alpha(n)}\left(\sum_{k=0}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)\right)}{\sum_{k=0}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)} > \frac{\frac{\partial}{\partial\alpha(n)}\left(1+\sum_{k=1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)\right)}{1+\sum_{k=1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)}.
\end{aligned}
$$

$$\text{Notice that } \frac{\partial}{\partial\alpha(n)}\left(\sum_{k=0}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)\right) = \frac{\sum_{k=n}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)}{\alpha(n)} \text{ and}$$

$$\frac{\partial}{\partial\alpha(n)}\left(1+\sum_{k=1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)\right) = \frac{\sum_{k=n+1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)}{\alpha(n)},$$

$$\text{from which } \frac{\partial}{\partial\alpha(n)}r(\boldsymbol{\alpha},\mu)>0 \Leftrightarrow \frac{\sum_{k=n}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)}{\sum_{k=0}^{N-1}\alpha(k)z_k\prod_{m=0}^{k-1}\alpha(m)} > \frac{\sum_{k=n+1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)}{1+\sum_{k=1}^{N}z_k\prod_{m=0}^{k-1}\alpha(m)}.$$

With $\prod_{m=0}^{k}\alpha(m)=A_k$, we obtain

$$\frac{\partial}{\partial\alpha(n)}r(\boldsymbol{\alpha},\mu)>0 \Leftrightarrow \frac{\sum_{k=n}^{N-1}z_kA_k}{\sum_{k=n+1}^{N}z_kA_{k-1}} > \frac{\sum_{k=0}^{N-1}z_kA_k}{1+\sum_{k=1}^{N}z_kA_{k-1}} = r(\boldsymbol{\alpha},\mu)/\Lambda_0.$$

Notice that the throughput per server, $r(\boldsymbol{\alpha},\mu)$, can never exceed the server capacity, μ, or $r(\boldsymbol{\alpha},\mu)<\mu$. Consider

$$\frac{\partial}{\partial\alpha_{N-1}}r(\boldsymbol{\alpha},\mu)>0 \Leftrightarrow \frac{z_{N-1}}{z_N} = \frac{1}{\rho/N} = \frac{\mu}{\Lambda_0} > r(\boldsymbol{\alpha},\mu)/\Lambda_0,$$

which is always true. Now, assume that $\frac{\partial}{\partial\alpha(n)}r(\boldsymbol{\alpha},\mu)>0$, or

$$\frac{\sum_{k=n+1}^{N-1}z_kA_k}{\sum_{k=n+2}^{N}z_kA_{k-1}} > \frac{\sum_{k=0}^{N-1}z_kA_k}{1+\sum_{k=1}^{N}z_kA_{k-1}},$$

$$\text{then } \frac{z_nA_n}{z_{n+1}A_n} > \frac{\sum_{k=0}^{N-1}z_kA_k}{1+\sum_{k=1}^{N}z_kA_{k-1}} \Rightarrow \frac{z_nA_n+\sum_{k=n+1}^{N-1}z_kA_k}{z_{n+1}A_n+\sum_{k=n+2}^{N}z_kA_{k-1}} > \frac{\sum_{k=0}^{N-1}z_kA_k}{1+\sum_{k=1}^{N}z_kA_{k-1}}.$$

This is because $\frac{r_n}{s_n}>\frac{R_n}{S_n}\Leftrightarrow S_nr_n>s_nR_n$ and

$$\frac{z_nA_n+r_n}{z_{n+1}A_n+s_n} > \frac{R_n}{S_n} \Leftrightarrow z_nA_nS_n+S_nr_n > z_{n+1}A_nR_n+s_nR_n.$$

Hence, $\frac{z_nA_n}{z_{n+1}A_n}>\frac{R_n}{S_n}\Leftrightarrow z_nA_nS_n>z_{n+1}A_nR_n$ and $S_nr_n>s_nR_n$ (by induction) from which follows that $\frac{z_nA_n+r_n}{z_{n+1}A_n+s_n}>\frac{R_n}{S_n}$. So, we only need to argue that $\frac{z_nA_n}{z_{n+1}A_n}=\frac{1}{\rho/N}>\frac{\sum_{k=0}^{N-1}z_kA_k}{1+\sum_{k=1}^{N}z_kA_{k-1}}=r(\boldsymbol{\alpha},\mu)//\Lambda_0$, which is always true (as the throughput can never exceed the service rate). It follows immediately that for any (β,P): $\frac{\partial}{\partial\alpha(n)}\Pi_\omega(\boldsymbol{\alpha},\beta,P)>0$. In other words, increasing the arrival rate at any state makes the profits increase. It is clear that $\gamma(k)=1$ for all $k\in\{0,\dots,N-1\}$ maximizes every component of $\boldsymbol{\alpha}$. It is easy to see that $\Pi_\omega(\boldsymbol{\alpha},\beta,P)$ increases in every $\alpha(k)$, $k\in\{0,\dots,N-1\}$. Within the set $\Gamma(n,P)$, $\gamma(k)=1$ is achieved by selecting $\gamma^o=1$. When $\gamma(k)=1$ for all $k\in\{0,\dots,N-1\}$ and $\gamma^o=1$, $\boldsymbol{\alpha}=\boldsymbol{\alpha}_h(P)$. For all finite $\overline{\mu}$, $\frac{\gamma^o\hat{\pi}(n,\boldsymbol{\alpha},\beta_h)}{\gamma^o\hat{\pi}(n,\boldsymbol{\alpha},\beta_h)+(1-\gamma^o)\hat{\pi}(n,\boldsymbol{\alpha},\beta_\ell)}$ is well defined and strictly positive and thus $\gamma(n)>0$. Hence, $(\boldsymbol{\alpha}_h(P),(\beta_\ell,\beta_h),(1,\dots))\in\Gamma(n,P)$ and $\max_{\beta\in[0,1]}\Pi_\omega(\boldsymbol{\alpha}_h(P),\beta,P)$ gives Π_ω^{dev}, which is independent of n. Only when $\overline{\mu}=+\infty$ and $n>0$,

$\lim_{\beta_h \to 1-} \frac{\gamma^o \hat{\pi}(n,\boldsymbol{\alpha},\beta_h)}{\gamma^o \hat{\pi}(n,\boldsymbol{\alpha},\beta_h)+(1-\gamma^o)\hat{\pi}(n,\boldsymbol{\alpha},\beta_\ell)} = 0$ and $(\boldsymbol{\alpha}_h(P),(\beta_\ell,1),(1,\ldots)) \notin \Gamma(n,P)$ and $\beta_h \in [0,1)$. However, for any $\epsilon > 0$, $(\boldsymbol{\alpha}_h(P),(\beta_\ell,1-\epsilon),(1,\ldots)) \in \Gamma(n,P)$ and, by continuity of $\Pi_\omega(\boldsymbol{\alpha}_h(P),\beta_\omega,P)$ in β, the supremum of $\Pi_\omega(\boldsymbol{\alpha}_h(P),\beta_\omega,P)$ is achieved for $\beta_h = 1$. ■

Proof of Proposition 1: We prove that the equilibrium prices can be supported by the following beliefs: $\boldsymbol{\gamma}^*(n,P_\ell^*) = 0$, $\boldsymbol{\gamma}^*(n,P_h^*) = 1$ and $P \notin \{P_\ell^*, P_h^*\}$ and, for all $n \in \{0,\ldots,N-1\}$, $\boldsymbol{\gamma}^*(n,P) = 0$. That is, consumers attribute any off-equilibrium price or congestion level to the low-quality firm. We show that (A) $\boldsymbol{\beta}^* = (1,0)$ and (P_ℓ^*, P_h^*) solve the firm's rationality conditions of Equation (2) and (B) the equilibrium belief satisfies Bayes' rule on the equilibrium path and the intuitive-criterion restrictions off the equilibrium path. We first provide the main steps of the proof and prove some minor supporting details under 'Remaining Details'. The main steps are:

A. Rationality of $\boldsymbol{\beta}^*$ and (P_ℓ^*, P_h^*). With Lemma 2, it is immediately clear that for any separating equilibrium in which the queue-joining strategy does not depend on congestion, the high-quality firm's equilibrium service rate is $\overline{\mu}$, as the costs of speeding up are zero. Given the infinite costs of speeding up for the low-quality firm, its equilibrium service rate is $\underline{\mu}$. Hence, any separating equilibrium must have $\boldsymbol{\beta}^* = (1,0)$. For any price that is different from P_h^*, the consumer's joining strategy is determined by the belief that the quality of the firm is low; $\boldsymbol{\alpha}_\ell(P)$. Now, we show that P_h^* and P_ℓ^* maximize

$$P_h^* \in \arg\max_{0 \le P \le 1} \begin{cases} P \times r(\boldsymbol{\alpha}_\ell(P),\mu), & P \neq P_h^* \\ P_h^* \times r(\boldsymbol{\alpha}_h(P_h^*),\mu), & P = P_h^* \end{cases} \text{ and}$$
$$P_\ell^* \in \arg\max_{0 \le P \le 1} \begin{cases} P \times \boldsymbol{\alpha}_\ell^*(P)\Lambda_0, & P \neq P_h^* \\ P_h^* \times \boldsymbol{\alpha}_h(P_h^*)\Lambda_0, & P = P_h^*. \end{cases}$$

These conditions are rewritten as

$$P_h^* \times r(\boldsymbol{\alpha}_h(P_h^*),\mu) \ge \max_{0 \le P \le 1} P \times r(\boldsymbol{\alpha}_\ell(P),\mu) = \underline{\Pi}_h \text{ and}$$
$$P_\ell^*(v_\ell - P_\ell^*)\Lambda_0 \ge \max_{0 \le P \le 1} \begin{cases} P[v_\ell - P]^+\Lambda_0, & P \neq P_h^* \\ P_h^* \times \boldsymbol{\alpha}_h(P_h^*)\Lambda_0, & P = P_h^* \end{cases} = \max\left\{\frac{v_\ell^2}{4}\Lambda_0, P_h^*(1-P_h^*)\Lambda_0\right\}.$$

Notice that P_h^* is by definition the highest root of $\frac{v_\ell^2}{4} = P(1-P)$. Therefore, $\max\left\{\frac{v_\ell^2}{4}, P_h^*(1-P_h^*)\right\} = \frac{v_\ell^2}{4}$ and, by definition, $P_\ell^*(v_\ell - P_\ell^*) = \frac{v_\ell^2}{4}$. Hence, the second inequality is always satisfied. Later in this proof, we prove that for $v_\ell < \frac{4}{5}$, the first inequality is always satisfied and that there exists an $\epsilon > 0$ such that the first inequality can never be satisfied for $1 - \epsilon < v_\ell < 1$. Otherwise, we impose the first inequality.

B. Consistency of $\boldsymbol{\gamma}^*(P)$. On the equilibrium path, it is immediately clear that the beliefs are consistent with Bayes' rule. Off the equilibrium path, $\boldsymbol{\gamma}^*(n,P)=0$, satisfying the intuitive criterion (see Equation (4)) if there exists no price $P \notin \{P_\ell^*, P_h^*\}$ such that

$$\Pi_\ell^* > \hat{\Pi}_\ell^{\text{dev}}(P) \text{ and } \hat{\Pi}_h^{\text{dev}}(P) > \Pi_h^*. \tag{13}$$

If such a price did exist, then the intuitive criterion would restrict the belief to 1 at such a price, which would violate the rationality conditions of $\boldsymbol{\beta}^*$ and (P_ℓ^*, P_h^*), as discussed under (A). If no such price exists, either the intuitive criterion restricts the belief to 0, or the intuitive criterion does not impose any restriction. This allows us to select 0, which would support the rationality conditions of $\boldsymbol{\beta}^*$ and (P_ℓ^*, P_h^*) discussed under (A). To show that for all prices, the conditions of Equation (13) are satisfied, consider all prices for which $\Pi_\ell^* = \frac{v_\ell^2}{4}\Lambda_0 > \hat{\Pi}_\ell^{\text{dev}}(P) = P(1-P)\Lambda_0$. Due to the concavity of $P(1-P)$, it is obvious that this set is $[0,P'] \cup [P'',1]$ where $P' = \frac{1}{2} - \frac{1}{2}\sqrt{1-v_\ell^2}$ and $P'' = P_h^*$. We must now show that in this set, $\hat{\Pi}_h^{\text{dev}}(P) = P \times r(\boldsymbol{\alpha}_h(P),\mu) \leq \Pi_h^*$. From Lemma 2, it follows that $\hat{\Pi}_h^{\text{dev}}(P) = P \times r(\boldsymbol{\alpha}_h(P),\mu)$ is concave in P. Due to the concavity of $\hat{\Pi}_h^{\text{dev}}(P)$, $\hat{\Pi}_h^{\text{dev}}(P)$ is increasing over $[0,P']$ and decreasing over $[P'',1]$. It follows that over $[0,P']$, the maximum value is $\hat{\Pi}_h^{\text{dev}}(P')$ and over $[P'',1]$, the maximum value is $\hat{\Pi}_h^{\text{dev}}(P'')$. Below, we show that $\hat{\Pi}_h^{\text{dev}}(P'') > \hat{\Pi}_h^{\text{dev}}(P')$. As $\Pi_h^* = P_h^* \times r(\boldsymbol{\alpha}_h(P_h^*),\mu) = \hat{\Pi}_h^{\text{dev}}(P'')$, we have obtained that $\Pi_h^* \geq \hat{\Pi}_h^{\text{dev}}(P)$ for all prices in $[0,P'] \cup [P'',1]$, from which it follows that there exists no price for which the condition of Equation (13) is satisfied. Hence, the off-equilibrium-path belief satisfies the intuitive criterion.

Remaining Details. We have argued above that the low-quality firm's separating price is $\frac{v_\ell}{2}$ and the high-quality firm's separating price that satisfies the intuitive criterion is a root of $P(1-P) = \frac{v_\ell^2}{4}$. Under point (A), it remains to be shown that there exists an $\epsilon > 0$ such that

$$P'' \times r(\boldsymbol{\alpha}_h(P''),\mu) \geq \max_{0\leq P\leq 1} P \times r(\boldsymbol{\alpha}_\ell(P),\mu) = \underline{\Pi}_h, \tag{14}$$

where $P_h^* = P'' = \frac{1}{2} + \frac{1}{2}\sqrt{1-v_\ell^2}$, and, for $v_\ell < \frac{4}{5}$, the condition of Equation (14) is always satisfied. Under point (B), it remains to be shown that

$$P' r(\boldsymbol{\alpha}_h(P'),\mu) < P'' r(\boldsymbol{\alpha}_h(P''),\mu). \tag{15}$$

Now, we prove the remaining Equations (14) and (15).

A. We verify whether the high-quality firm has an incentive to deviate at price P'' to the price that maximizes the its profits when consumers believe that the quality is low. By definition of P'', $P''r(\boldsymbol{\alpha}_h(P''),+\infty) = \max_{0\le P\le v_\ell} Pr(\boldsymbol{\alpha}_\ell(P),+\infty)$. Notice that if $v_\ell = 1-\epsilon$, where $\epsilon > 0$, then $r(\boldsymbol{\alpha}_\ell(P),\mu)$ is arbitrarily close to $r(\boldsymbol{\alpha}_h(P),\mu)$, and $P'' = \frac{1}{2}+\frac{1}{2}\sqrt{1-v_\ell^2}$ is arbitrarily close to $\frac{1}{2}$, but due to the finite service rate, $P_h = \arg\max_{0\le P\le 1} Pr(\boldsymbol{\alpha}_h(P),\mu)$ is different from $\frac{1}{2}$. As a consequence, $P''r(\boldsymbol{\alpha}_h(P''),+\infty) \approx \frac{1}{2}r(\boldsymbol{\alpha}_h(\frac{1}{2}),\mu) < \max_{0\le P\le 1} Pr(\boldsymbol{\alpha}_h(P),\mu) \approx \max_{0\le P\le v_\ell} Pr(\boldsymbol{\alpha}_\ell(P),\mu)$ and hence Equation (14) is not satisfied. Therefore, in order to have $P''r(\boldsymbol{\alpha}_h(P''),\mu) > \max_{0\le P\le v_\ell} Pr(\boldsymbol{\alpha}_\ell(P),\mu)$, v_ℓ must be low enough. From Equation (14) and $P''r(\boldsymbol{\alpha}_h(P''),+\infty) = \max_{0\le P\le v_\ell} P\times r(\boldsymbol{\alpha}_\ell(P),+\infty)$, we obtain

$$\text{Equation (14)} \Leftrightarrow \frac{r(\boldsymbol{\alpha}_h(P''),\mu)}{r(\boldsymbol{\alpha}_h(P''),+\infty)} > \frac{\max_{0\le P\le v_\ell} P\times r(\boldsymbol{\alpha}_\ell(P),\mu)}{\max_{0\le P\le v_\ell} P\times r(\boldsymbol{\alpha}_\ell(P),+\infty)}. \tag{16}$$

Letting $P_\ell = \arg\max_{0\le P\le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P),\mu)$, $\max_{0\le P\le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P),+\infty) > P_\ell \times r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)$. We have that

$$\frac{r(\boldsymbol{\alpha}_\ell(P_\ell),\mu)}{r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)} > \frac{\max_{0\le P\le v_\ell} Pr(\boldsymbol{\alpha}_\ell(P),\mu)}{\max_{0\le P\le v_\ell} Pr(\boldsymbol{\alpha}_\ell(P),+\infty)}.$$

Therefore, if

$$\frac{r(\boldsymbol{\alpha}_h(P''),\mu)}{r(\boldsymbol{\alpha}_h(P''),+\infty)} > \frac{r(\boldsymbol{\alpha}_\ell(P_\ell),\mu)}{r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)}, \tag{17}$$

Equation (16) is satisfied. We rewrite the inequality of Equation (17):

$$\frac{r(\boldsymbol{\alpha}_h(P''),\mu)}{r(\boldsymbol{\alpha}_h(P''),+\infty)} > \frac{r(\boldsymbol{\alpha}_\ell(P_\ell),\mu)}{r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)} \Leftrightarrow \frac{r(\boldsymbol{\alpha}_h(P''),+\infty)}{r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)} < \frac{r(\boldsymbol{\alpha}_h(P''),\mu)}{r(\boldsymbol{\alpha}_\ell(P_\ell),\mu)} \Leftrightarrow$$

$$\frac{1-P''}{1-P_\ell} < \frac{1-P''}{1-P_\ell}\frac{1-\frac{\frac{1}{N!}\left(\frac{1-P''}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P''}{\mu}\right)^k}}{1-\frac{\frac{1}{N!}\left(\frac{1-P_\ell}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P_\ell}{\mu}\right)^k}} \Leftrightarrow \frac{\frac{1}{N!}\left(\frac{1-P''}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P''}{\mu}\right)^k} < \frac{\frac{1}{N!}\left(\frac{1-P_\ell}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P_\ell}{\mu}\right)^k}.$$

As

$$\frac{\frac{1}{N!}\left(\frac{\lambda}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}$$

increases in λ, we obtain that $\frac{r(\boldsymbol{\alpha}_h(P''),\mu)}{r(\boldsymbol{\alpha}_h(P''),+\infty)} > \frac{r(\boldsymbol{\alpha}_\ell(P_\ell),\mu)}{r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)}$ if $1-P'' < 1-P_\ell$, or $P'' > P_\ell$. From $\frac{r(\boldsymbol{\alpha}_h(P''),\mu)}{r(\boldsymbol{\alpha}_h(P''),+\infty)} > \frac{r(\boldsymbol{\alpha}_\ell(P_\ell),\mu)}{r(\boldsymbol{\alpha}_\ell(P_\ell),+\infty)}$ follows the inequality of Equation (16), and hence the inequality of Equation (14). Notice that it must be true that $P_\ell < v_\ell$. Hence, a sufficient condition for

$P'' > P_\ell$ (and therefore for the inequality of Equation (14)) is $v_\ell < P''$, or, with the definition of P'',

$$v_\ell < \frac{1}{2} + \frac{1}{2}\sqrt{1 - v_\ell^2} \Leftrightarrow v_\ell < \frac{1}{2} + \frac{1}{2}\sqrt{1 - v_\ell^2} \Leftrightarrow v_\ell < \frac{4}{5}.$$

In conclusion, we obtained that $v_\ell < \frac{4}{5}$ is a sufficient condition for the inequality of Equation (14) and that for v_ℓ close enough to $v_h(=1)$, the inequality of Equation (14) cannot be satisfied. For other values of v_ℓ, the inequality of Equation (14) needs to be imposed (and will be a function of all model parameters).

B. In order to determine which of the two roots, P' and P'', is a separating price, we need to find which root maximizes the high-quality firm's profits. Now, we can write

$$P' r(\boldsymbol{\alpha}_h(P'), \mu) = P' r(\boldsymbol{\alpha}_h(P'), +\infty)\frac{r(\boldsymbol{\alpha}_h(P'), \mu)}{r(\boldsymbol{\alpha}_h(P'), +\infty)} = P_\ell^* r(\boldsymbol{\alpha}_\ell(P_\ell^*), +\infty)\frac{r(\boldsymbol{\alpha}_h(P'), \mu)}{r(\boldsymbol{\alpha}_h(P'), +\infty)}$$

$$\text{and } P'' r(\boldsymbol{\alpha}_h(P''), \mu) = P_\ell^* r(\boldsymbol{\alpha}_\ell(P_\ell^*), +\infty)\frac{r(\boldsymbol{\alpha}_h(P''), \mu)}{r(\boldsymbol{\alpha}_h(P''), +\infty)}.$$

Hence, the high price (P'') is more profitable for the high-quality firm; that is, the inequality of Equation (15) is satisfied when

$$1 > \frac{P' r(\boldsymbol{\alpha}_h(P'), \mu)}{P'' r(\boldsymbol{\alpha}_h(P''), \mu)} = \frac{\frac{r(\boldsymbol{\alpha}_h(P'), \mu)}{r(\boldsymbol{\alpha}_h(P'), +\infty)}}{\frac{r(\boldsymbol{\alpha}_h(P''), \mu)}{r(\boldsymbol{\alpha}_h(P''), +\infty)}}.$$

The latter is true when

$$\frac{r(\boldsymbol{\alpha}_h(P), \mu)}{r(\boldsymbol{\alpha}_h(P), +\infty)}$$

is increasing in P. We now prove that $\frac{r(\boldsymbol{\alpha}_h(P), \mu)}{r(\boldsymbol{\alpha}_h(P), +\infty)}$ is increasing in P. Notice that

$$r(\boldsymbol{\alpha}_h(P), \mu) = (1-P)\left(1 - \frac{\frac{1}{N!}\left(\frac{1-P}{\mu}\right)^N}{1 + \sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P}{\mu}\right)^k}\right)\Lambda_0 \text{ and } r(\boldsymbol{\alpha}_h(P), +\infty) = (1-P)\Lambda_0.$$

Hence,

$$\frac{r(\boldsymbol{\alpha}_h(P), \mu)}{r(\boldsymbol{\alpha}_h(P), +\infty)} = 1 - \frac{\frac{1}{N!}\left(\frac{1-P}{\mu}\right)^N}{1 + \sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P}{\mu}\right)^k}.$$

The right-hand side of the above expression is $L(1-P)$, where $L(\lambda) = \frac{\frac{1}{N!}\left(\frac{\lambda}{\mu}\right)^N}{\sum_{k=0}^N \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}$ is the loss rate, which increases in the arrival rate as:

$$\frac{\mathrm{d}}{\mathrm{d}\rho}\frac{\frac{1}{N!}\rho^N}{\sum_{k=0}^N \frac{1}{k!}\rho^k} = \frac{N\rho^{N-1}\sum_{k=0}^N \frac{\rho^k}{k!} - \rho^N \sum_{k=1}^N k\frac{\rho^{k-1}}{k!}}{N!\left(\sum_{k=1}^N \frac{\rho^k}{k!} + 1\right)^2} > 0 \Leftrightarrow N\sum_{k=0}^N \frac{\rho^k}{k!} > \sum_{k=0}^{N-1} \frac{\rho^k}{k!}.$$

The latter inequality is always true (because $N \geq k$ in the summation) and hence $\frac{r(\boldsymbol{\alpha}_h(P),\mu)}{r(\boldsymbol{\alpha}_h(P),+\infty)}$ is decreasing in α and increasing in P. It follows that $P''r(\boldsymbol{\alpha}_h(P''),\mu) > P'r(\boldsymbol{\alpha}_h(P'),\mu)$, or the separating price of the high-quality firm is P''. ■

Proof of Proposition 2: On the equilibrium path (i.e., for $P = P^*$), by construction, the beliefs, joining strategy and service rate are consistent with the prior belief that the quality is high. It is also easily observable that P^* maximizes each firm type's profits with an off-equilibrium-path belief that the firm's quality is low (i.e., $P \neq P^*$ and, for all $n \in \{0,\dots,N-1\}$, $\boldsymbol{\gamma}^*(n,P) = 0$) if $P^* \times r(\boldsymbol{\alpha}^p(P^*),\mu) \geq \underline{\Pi}_h$ and $P^* \times r(\boldsymbol{\alpha}^p(P^*),+\infty) = P^* \times \alpha^*(0,P^*) \geq \underline{\Pi}_\ell$. We only need to check whether the off-equilibrium-path belief satisfies the intuitive criterion. Similar to Proposition 1, there should be no price such that $\Pi_\ell^* > \hat{\Pi}_\ell^{\text{dev}}(P)$ and $\hat{\Pi}_h^{\text{dev}}(P) > \Pi_h^*$. Due to the concavity of $\hat{\Pi}_\omega^{\text{dev}}(P)$ for $\omega \in \{\ell,h\}$, these conditions reduce to $\underline{P}_\ell(P^*) \leq \underline{P}_h(P^*)$ and $\overline{P}_h(P^*) \leq \overline{P}_\ell(P^*)$. If the consumers believed that the firm's quality were high, then the high-quality firm would have an incentive to deviate from the pooling price, P^*.

We argued above that $\pi_\ell(0,P) = 1$ and $\pi_\ell(n,P) = 0$ for $1 \leq n \leq N$ and, as a consequence, any nonempty restaurant is a high-quality restaurant. Thus, the joining probability is $\alpha(n,P) = \alpha_1(P) = 1 - P$ for $1 \leq n \leq N$. Let $\alpha(0,P) = \alpha_0(P)$ be the joining probability at the empty restaurant. When the service rate is μ (for the high-quality firm), the steady-state probabilities are

$$\begin{aligned}
\pi(0,\boldsymbol{\alpha}) &= \frac{1}{1 + \sum_{k=1}^N \frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{\Lambda_0 N}{\mu}\right)^k} \\
\pi(n,\boldsymbol{\alpha}) &= \frac{\frac{1}{n!}\alpha_0\alpha_1^{n-1}\left(\frac{\Lambda_0 N}{\mu}\right)^n}{1 + \sum_{k=1}^N \frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{\Lambda_0 N}{\mu}\right)^k}, 1 \leq n \leq N.
\end{aligned}$$

For the low-quality firm, as the service rate is $+\infty$, with probability 1, the firm is empty (that is, the state is $n = 0$). At the empty firm, the updated value is thus

$$v(0,\boldsymbol{\alpha}(P)) = v_\ell + \frac{p\pi_h(0,\boldsymbol{\alpha}(P))/\pi_\ell(0,\boldsymbol{\alpha}(P))}{p\pi_h(0,\boldsymbol{\alpha}(P))/\pi_\ell(0,\boldsymbol{\alpha}(P)) + (1-p)}(1 - v_\ell) \tag{18}$$

and, hence, $\alpha_0(P)$ is rational when

$$\alpha_0(P) = v(0,\boldsymbol{\alpha}(P)) - P. \tag{19}$$

Now, let $\sum_{k=1}^{N}\frac{1}{k!}\alpha_1^{k-1}\left(\frac{\Lambda_0 N}{\mu}\right)^k = F(\alpha_1)$. Then $\pi(0,\boldsymbol{\alpha}) = \frac{1}{1+\alpha_0 F(\alpha_1)}$. We thus obtain the following condition.

$$\begin{aligned}
\alpha_0(P) \;:\; \alpha_0 &= v_\ell + \frac{p\frac{1}{1+\alpha_0 F(\alpha_1(P))}}{p\frac{1}{1+\alpha_0 F(\alpha_1(P))}+(1-p)}(1-v_\ell) - P, \text{ from which we solve}\\
\alpha_0(P) &= -\frac{1}{2}\left(\frac{\frac{p}{1-p}+1}{F(\alpha_1(P))} - v_\ell + P\right)\\
&\quad + \frac{1}{2}\sqrt{\left(\frac{\frac{p}{1-p}+1}{F(\alpha_1(P))} - v_\ell + P\right)^2 + 4\frac{(v_\ell - P)\left(\frac{p}{1-p}+1\right)+\frac{p}{1-p}(1-v_\ell)}{F(\alpha_1(P))}}.
\end{aligned}$$

It is easy to observe that the negative root can be discarded and $\alpha_0(P) \le v - P < 1$ because, in Equation (19), $v(0,\boldsymbol{\alpha}(P)) < v$ (see Equation (18): $\frac{p\pi_h(0,\boldsymbol{\alpha}(P))/\pi_\ell(0,\boldsymbol{\alpha}(P))}{p\pi_h(0,\boldsymbol{\alpha}(P))/\pi_\ell(0,\boldsymbol{\alpha}(P))+(1-p)} \in [0,p]$ because $\frac{1}{1+\alpha_0 F(\alpha_1(P))} < 1$).

Now, we prove that with the queue-joining strategy $(\alpha_0(P), 1-P, \ldots)$, there exists at most one price, $\hat{P}$, in $(0,v]$ such that, above (below) $\hat{P}$, the increase in profits from speeding up is positive (negative). We write this increase first as a function of the joining strategy, (α_0,α_1):

$$\begin{aligned}
\Delta(\alpha_0,\alpha_1) &= \alpha_0\Lambda_0 - \frac{\alpha_0 + \sum_{k=1}^{N-1}\frac{1}{k!}\alpha_0\alpha_1^k\left(\frac{1}{\mu}\right)^k}{1+\sum_{k=1}^{N}\frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k}\Lambda_0 = \alpha_0\frac{(\alpha_0-\alpha_1)\sum_{k=1}^{N-1}\frac{1}{k!}\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k + \frac{1}{k!}\alpha_0\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N}{1+\sum_{k=1}^{N}\frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k}\Lambda_0\\
&= \alpha_0\frac{(\alpha_0-\alpha_1)F(\alpha_1) - (\alpha_0-\alpha_1)\frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N + \frac{1}{k!}\alpha_0\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N}{1+\sum_{k=1}^{N}\frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k}\Lambda_0\\
&= \alpha_0\frac{(\alpha_0-\alpha_1)F(\alpha_1) + \alpha_1\frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N}{1+\sum_{k=1}^{N}\frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k}\Lambda_0 \text{ and}
\end{aligned}$$

$$\begin{aligned}
\Delta(\alpha_0,\alpha_1) > 0 &\Leftrightarrow (\alpha_0-\alpha_1)F(\alpha_1) + \frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N > 0 \Leftrightarrow \alpha_0 > \alpha_1 - \frac{\alpha_1^{N-1}}{F(\alpha_1)}\frac{\left(\frac{1}{\mu}\right)^N}{N!}\\
&\Leftrightarrow \frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N > (\alpha_1-\alpha_0)F(\alpha_1).
\end{aligned}$$

Now we prove the existence of a unique root of $\Delta^p(P) = \Delta(\alpha_0(P), 1-P)$ in $(0,v)$ when $\Delta^p(0) > 0$. To that end, we prove three properties: (i) when $\Delta^p(P) \ge 0$, $\Delta^p(P)$ strictly decreases in P, (ii) $\lim_{P\to v-}\Delta^p(P) = 0$ and (iii) $\lim_{P\to v-}\frac{\mathrm{d}\Delta^p(P)}{\mathrm{d}P} > 0$. Taken together, $(i$–$iii)$ imply that when $\Delta^p(0) > 0$, it must decrease. In addition, there must exist an $\epsilon > 0$ such that at price $v-\epsilon$ $\Delta^p(v-\epsilon) < 0$. By continuity, there must exist at least one $\hat{P}$ where $\Delta^p(\hat{P}) = 0$. There cannot exist two prices $\hat{P}'$ and $\hat{P}''$ such that $\Delta^p(\hat{P}') = \Delta^p(\hat{P}'') = 0$. Assuming this were true at $\hat{P}''$, $\Delta^p(\hat{P}'')$ increases (because it comes from below). However, this contradicts that when $\Delta^p(\hat{P}'')=0$, $\Delta^p(\hat{P}'')$ strictly decreases.

Now we prove (i). We take the derivative of $\alpha_0(P)$ with respect to P:

$$\begin{aligned}
\frac{\mathrm{d}\alpha_0}{\mathrm{d}P} &= \frac{\mathrm{d}}{\mathrm{d}P}\frac{p}{p+(1-p)\left(1+\alpha_0 F(\alpha_1(P))\right)}(1-v_\ell)-1 \Rightarrow \\
\frac{\mathrm{d}\alpha_0}{\mathrm{d}P} &= -p(1-p)\frac{\frac{\mathrm{d}\alpha_0}{\mathrm{d}P}F(\alpha_1)+\alpha_0 F'(\alpha_1)\frac{\mathrm{d}\alpha_1}{\mathrm{d}P}}{\left(p+(1-p)\left(1+\alpha_0 F(\alpha_1)\right)\right)^2}(1-v_\ell)-1 \Rightarrow \\
\frac{\mathrm{d}\alpha_0}{\mathrm{d}P} &= -\frac{1-\frac{p(1-p)(1-v_\ell)\alpha_0 F'(\alpha_1)}{(p+(1-p)(1+\alpha_0 F(\alpha_1)))^2}}{1+\frac{p(1-p)(1-v_\ell)F(\alpha_1)}{(p+(1-p)(1+\alpha_0 F(\alpha_1)))^2}}
\end{aligned}$$

Note that as $F'(\alpha_1)>0$,

$$\frac{\mathrm{d}\alpha_0}{\mathrm{d}P} < \frac{\mathrm{d}\alpha_1}{\mathrm{d}P} = -1. \tag{20}$$

Also, note that

$$\begin{aligned}
(N-1)F(\alpha_1) &= \textstyle\sum_{k=1}^N \frac{N-1}{k!}\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k \text{ and } F'(\alpha_1)=\frac{1}{\alpha_1}\sum_{k=1}^N \frac{k-1}{k!}\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k \\
F'(\alpha_1)\alpha_1 &= \textstyle\sum_{k=1}^N \frac{k-1}{k!}\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k \text{ and } (N-1)F(\alpha_1)=\sum_{k=1}^N \frac{N-1}{k!}\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k \\
&\Rightarrow (N-1)F(\alpha_1) \geq F'(\alpha_1)\alpha_1.
\end{aligned}$$

Now we differentiate $\left(\alpha_0-\alpha_1\right)F(\alpha_1)+\frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N$ with respect to P. Assume now that $\Delta^p(P)\geq 0$, that is $\Delta(\alpha_0,\alpha_1)\geq 0$, then

$$\begin{aligned}
&\left(\frac{\mathrm{d}\alpha_0}{\mathrm{d}P}-\frac{\mathrm{d}\alpha_1}{\mathrm{d}P}\right)\alpha_1 F(\alpha_1)+\left\{\left(\alpha_0-\alpha_1\right)\alpha_1 F'(\alpha_1)+\frac{N-1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N\right\}\frac{\mathrm{d}\alpha_1}{\mathrm{d}P} \\
&< \left(\frac{\mathrm{d}\alpha_0}{\mathrm{d}P}-\frac{\mathrm{d}\alpha_1}{\mathrm{d}P}\right)\alpha_1 F(\alpha_1)+\left\{\left(\alpha_0-\alpha_1\right)(N-1)F(\alpha_1)+\frac{N-1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N\right\}\frac{\mathrm{d}\alpha_1}{\mathrm{d}P} \\
&= \left(\frac{\mathrm{d}\alpha_0}{\mathrm{d}P}-\frac{\mathrm{d}\alpha_1}{\mathrm{d}P}\right)\alpha_1 F(\alpha_1)+(N-1)\left\{\left(\alpha_0-\alpha_1\right)F(\alpha_1)+\frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N\right\}\frac{\mathrm{d}\alpha_1}{\mathrm{d}P} \\
&= \underbrace{\left(\frac{\mathrm{d}\alpha_0}{\mathrm{d}P}-\frac{\mathrm{d}\alpha_1}{\mathrm{d}P}\right)}_{-ve}\alpha_1 F(\alpha_1)+(N-1)\underbrace{\Delta(\alpha_0,\alpha_1)}_{+ve}\underbrace{\frac{\mathrm{d}\alpha_1}{\mathrm{d}P}}_{-ve} \\
&< 0.
\end{aligned}$$

Therefore, when $\Delta^p(0)\geq 0$, it decreases in P. This proves (i).

Now we prove (ii). If $P=0$, then $\alpha_0(0)<\alpha_1(0)=1$ and

$$\begin{aligned}
\Delta^p(0)>0 &\Leftrightarrow \alpha_0(0)>1-\frac{1}{F(1)}\frac{\left(\frac{1}{\mu}\right)^N}{N!} \Leftrightarrow \alpha_0(0)=1-\frac{(1-p)}{p\frac{1}{1+\alpha_0 F(1)}+(1-p)}(1-v_\ell)>1-\frac{1}{F(1)}\frac{\left(\frac{1}{\mu}\right)^N}{N!} \\
&\Leftrightarrow \frac{(1-p)}{p\frac{1}{1+\alpha_0(0)F(1)}+(1-p)}(1-v_\ell)<\frac{1}{F(1)}\frac{\left(\frac{1}{\mu}\right)^N}{N!}.
\end{aligned}$$

When $P \to v_\ell + p(1-v_\ell)$, α_0 goes to zero because for $\alpha_0 = 0$ and $P = v_\ell + p(1-v_\ell)$. We obtain

$$0 = v_\ell + \frac{p}{p+(1-p)}(1-v_\ell) - (v_\ell + p(1-v_\ell))$$

and $\alpha_1 \to 1-(v_\ell + p(1-v_\ell))$ and, hence, $\Delta^p(P) \to 0$.

Finally, we prove (iii):

$$\begin{aligned}\lim_{P\to v_\ell+p(1-v_\ell)} \frac{\mathrm{d}\Delta^p(P)}{\mathrm{d}P} &= \frac{(\alpha_0-\alpha_1)\,F(\alpha_1) + \frac{1}{N!}\alpha_1^N\left(\frac{1}{\mu}\right)^N}{1+\alpha_0 F(\alpha_1)} \lim_{P\to v_\ell+p(1-v_\ell)} \frac{\mathrm{d}\alpha_0}{\mathrm{d}P} \\ &\quad + \frac{\mathrm{d}}{\mathrm{d}P}\frac{(\alpha_0-\alpha_1)\,F(\alpha_1) + \frac{1}{N!}\alpha_1^N\left(\frac{1}{\mu}\right)^N}{1+\alpha_0 F(\alpha_1)} \lim_{P\to v_\ell+p(1-v_\ell)} \alpha_0 \\ &= \left(-\alpha_1 F(\alpha_1) + \frac{1}{N!}\alpha_1^N\left(\frac{1}{\mu}\right)^N\right) \lim_{P\to v_\ell+p(1-v_\ell)} \frac{\mathrm{d}\alpha_0}{\mathrm{d}P},\end{aligned}$$

from which follows that

$$\lim_{P\to v_\ell+p(1-v_\ell)} \frac{\mathrm{d}\alpha_0}{\mathrm{d}P} = -\frac{1}{1+p(1-p)(1-v_\ell)F(\alpha_1)} < 0,$$

$$\begin{aligned}\text{As } -\alpha_1 F(\alpha_1) + \frac{1}{N!}\alpha_1^N\left(\frac{1}{\mu}\right)^N < 0 \Leftrightarrow \frac{1}{N!}\alpha_1^N\left(\frac{1}{\mu}\right)^N &< \alpha_1 F(\alpha_1) = \textstyle\sum_{k=1}^{N} \frac{1}{k!}\alpha_1^k\left(\frac{1}{\mu}\right)^k \\ &= \textstyle\sum_{k=1}^{N-1} \frac{1}{k!}\alpha_1^k\left(\frac{1}{\mu}\right)^k + \frac{1}{N!}\alpha_1^N\left(\frac{1}{\mu}\right)^N,\end{aligned}$$

and that $\lim_{P\to v_\ell+p(1-v_\ell)} \frac{\mathrm{d}\Delta^p(P)}{\mathrm{d}P} > 0$. With (i), (ii), (iii), we have obtained that when $\Delta^p(0) > 0$, there exists exactly one price, $\hat{P}$, such that $\Delta^p(\hat{P}) = 0$. ■

Proof of Proposition 3: For candidate pooling prices, P^*, higher than $\hat{P}$, the price where the low-quality firm does not speed up, the equilibrium service-rate selection by the low-quality firm must be mixed. As for $P^* > \hat{P}$, $\Delta^p(P^*) < 0$. That is, the low-quality firm has an incentive to reduce its service rate. Assume that the low-quality firm selects μ. Then both firms select μ. The consumer joining strategy becomes independent of the congestion level upon arrival: $\alpha_0 = \alpha_1 = \alpha$, from which follows that $\Delta(\alpha, \alpha, \alpha, \ldots) > 0$ (see Lemma 2). Therefore, the low-quality firm has an incentive to increase its service rate. As a consequence, the equilibrium service rate of the low-quality firm cannot be a pure strategy. It will be mixed in service rates and the probability that the queue length is in $1 \le n \le N$ for the low-quality firm is strictly positive. Assume that with probability β, the low-quality firm's service rate is $+\infty$ (and with probability $1-\beta$, the service rate is μ). Then, the updated service value is

$$v(0, \boldsymbol{\alpha}, \beta) = v_\ell + \frac{p\pi(0, \boldsymbol{\alpha})}{p\pi(0, \boldsymbol{\alpha}) + (1-p)((1-\beta)\pi(n, \boldsymbol{\alpha}) + \beta)}(1-v_\ell) \text{ and}$$

$$v(n,\boldsymbol{\alpha},\beta) = v_\ell + \frac{p\pi(n,\boldsymbol{\alpha})}{p\pi(n,\boldsymbol{\alpha}) + (1-p)(1-\beta)\pi(n,\boldsymbol{\alpha})}(1-v_\ell),\ 1 \le n \le N.$$

The long-run queue-length distribution (Lemma 1) implies that

$$\alpha_0(P,\beta)\ :\ \alpha_0 = v_\ell + \frac{L(\beta)\frac{1}{1+\alpha_0 F(\alpha_1(P))}}{L(\beta)\frac{1}{1+\alpha_0 F(\alpha_1(P))} + \frac{\beta}{1-\beta}}(1-v_\ell) - P \text{ and}$$
$$\alpha_n(P,\beta) = v_\ell + \frac{L(\beta)}{L(\beta)+1}(1-v_\ell) - P,\ 1 \le n \le N, \text{ where } L(\beta) = \frac{p}{(1-p)(1-\beta)}.$$

We can solve a quadratic equation for $\alpha_0(P,\beta)$:

$$\alpha_0(P,\beta) = -\frac{1}{2}\left(\frac{\frac{p}{(1-p)\beta}+1}{F(\alpha_1(P,\beta))} + P - v_\ell\right) \pm \frac{1}{2}\sqrt{\left(\frac{\frac{p}{(1-p)\beta}+1}{F(\alpha_1(P,\beta))} + P - v_\ell\right)^2 + \frac{(1-P)\frac{p}{(1-p)\beta} + (v_\ell - P)}{F(\alpha_1(P,\beta))}}.$$

With the definition of $H(P,\alpha_1,L)$, we obtain that $\alpha_0(P,\beta) = H(P,\alpha_1(P,\beta),p/((1-p)\beta))$. Notice that the joining probability for all strictly positive congestion levels is equal to $\alpha_1(P,\beta)$. Therefore, from Proposition 2, the low-quality firm has no incentive to speed up when $\alpha_0(P,\beta) = \alpha_1(P,\beta) - \frac{(\alpha_1(P,\beta))^N}{F(\alpha_1(P,\beta))}\frac{\left(\frac{1}{\mu}\right)^N}{N!}$. The equilibrium conditions are determined by $\alpha_0^p(P)$ and $\beta^p(P)$, which solve

$$\alpha_0 = H(P,\alpha_1(P,\beta),p/((1-p)\beta)) \text{ and } \alpha_0 = \alpha_1(P,\beta) - \frac{(\alpha_1(P,\beta))^N}{F(\alpha_1(P,\beta))}\frac{\left(\frac{1}{\mu}\right)^N}{N!}.$$

Because of the randomization, the high- and the low-quality firms' equilibrium profits are equal; $P^* \times \alpha_0^p(P^*) = P^* \times r(\boldsymbol{\alpha}^*,\mu)$, where $\boldsymbol{\alpha}^* = (\alpha_0^p(P^*), \alpha_1(P^*,\beta^p(P^*)),\ldots)$. As $\underline{\Pi}_\ell = \frac{v_\ell^2}{4} \ge \underline{\Pi}_h = \max_{0\le P\le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P),\mu)$, when $P^* \times r(\boldsymbol{\alpha}^*,\mu) \ge \frac{1}{4}v_\ell^2$, the price P^* maximizes both firms' profits as off equilibrium, consumers believe that the quality of the firm is low and hence, $P^* \times r(\boldsymbol{\alpha}^*,\mu) \ge \frac{1}{4}v_\ell^2$ is sufficient. The condition ensures that the low-quality firm's (pooling) equilibrium profits are at least the maximum profits with the belief that the firm's quality is low, $\underline{\Pi}_\ell$, which are equal to $\frac{1}{4}v_\ell^2$. The high-quality firm's equilibrium profits are the same as the low-quality firm's equilibrium profits, but due to the lower service rate, the high-quality firm's profits with the belief that its quality is low are always less than $\frac{1}{4}v_\ell^2$. Hence, the condition $P^* \times r(\boldsymbol{\alpha}^*,\mu) \ge \frac{1}{4}v_\ell^2$ is sufficient to motivate both firms to select P^*.

The restrictions imposed by the intuitive criterion are satisfied. The reason is that, due to the low-quality firm's garbling of the service rate, the low- and high-quality firms' profits in equilibrium are equal: $\Pi_h^* = \Pi_\ell^* = \Pi^*$. Furthermore, the low-quality firm's deviation profits are always higher than the high-quality deviation profits—$\hat{\Pi}_\ell^{\text{dev}}(P) \ge \hat{\Pi}_h^{\text{dev}}(P)$—due to the low-quality firm's higher service rate. As a consequence, it is always the case that the intuitive criterion forces the off-equilibrium belief to be the low-quality firm: $\Pi^* \geq \hat{\Pi}_h^{\text{dev}}(P)$ and $\hat{\Pi}_\ell^{\text{dev}}(P) \geq \Pi^*$. ■

Proof of Proposition 4: For notational simplicity, we set $\Lambda_0 = 1$. In (A), we write the equilibrium conditions for α_0^*, α_1^* and β_ℓ^* and take the limit for $N \to +\infty$. In (B), we compute Π_h^*, Π_ℓ^*, $\Pi_h^{\text{dev}}(P)$ and $\Pi_\ell^{\text{dev}}(P)$, and impose the intuitive-criterion conditions in order to characterize the intuitive pooling prices.

A. Characterization of $\lim_{N\to+\infty} \alpha_0^*$, $\lim_{N\to+\infty} \alpha_1^*$, $\lim_{N\to+\infty} \beta_\ell^*$ and $\lim_{N\to+\infty} \hat{P}$: The derivations of the limit when $P < \lim_{N\to+\infty} \hat{P}$, for which $\lim_{N\to+\infty} \beta_\ell^* = 1$ are not reported for brevity as they are less involved than as no randomization needs to be considered. Therefore, we introduce $(\beta_\ell^* =)\beta \in [0,1)$. With $L = p/(1-p)$, and $F(\alpha_1, \mu/N) = \sum_{k=1}^N \frac{1}{k!}\alpha_1^{k-1}\left(\frac{1}{\mu/N}\right)^k$, to explicitly show the dependency on N (see Lemma 1), the equilibrium conditions when the low-quality firm does not speed up for sure are

$$\begin{cases} \alpha_0 = (v_\ell - P) + \frac{\frac{L}{1-\beta}\frac{1}{1+\alpha_0 F(\alpha_1,\mu/N)}}{\left(\frac{L}{1-\beta}+1\right)\frac{1}{1+\alpha_0 F(\alpha_1,\mu/N)}+\frac{\beta}{1-\beta}}(1-v_\ell) \\ \alpha_1 = (v_\ell - P) + \frac{\frac{L}{1-\beta}}{\frac{L}{1-\beta}+1}(1-v_\ell) \\ \alpha_0 = \alpha_1 - \frac{\alpha_1^N}{F(\alpha_1,\mu/N)}\frac{\left(\frac{1}{\mu/N}\right)^N}{N!}. \end{cases}$$

(Notice that the derivations for which $\lim_{N\to+\infty} \beta_\ell^* = 1$, the third case relaxes to: $\alpha_0 \leq \alpha_1 - \alpha_1^N (1/(\mu/N))^N / (F(\alpha_1, \mu/N)N!)$, which greatly simplifies the derivations, hence, we omit this case.) We rewrite the above conditions as

$$\begin{cases} \beta & = 1 + L - L\frac{1-v_\ell}{\alpha_1-(v_\ell-P)} \\ \alpha_0 - \alpha_1 & = -\frac{\alpha_1^N}{F(\alpha_1,\mu/N)}\frac{\left(\frac{1}{\mu/N}\right)^N}{N!} \\ \alpha_0 - \alpha_1 & = -\left(\frac{L}{L+1-\beta} - \frac{L}{1+L-\beta+\beta(1+\alpha_0 F(\alpha_1,\mu/N))}\right)(1-v_\ell) \end{cases}$$
$$\Leftrightarrow \begin{cases} \beta & = 1 + L - L\frac{1-v_\ell}{\alpha_1-(v_\ell-P)} \\ \alpha_0 - \alpha_1 & = -\frac{\alpha_1^N}{F(\alpha_1,\mu/N)}\frac{\left(\frac{1}{\mu/N}\right)^N}{N!} \\ \frac{\alpha_1^N\left(\frac{1}{\mu/N}\right)^N/N!}{F(\alpha_1,\mu/N)} & = \left(\frac{L}{L+1-\beta} - \frac{L}{L+1+\beta\alpha_1 F(\alpha_1,\mu/N)-\beta\alpha_1^N\left(\frac{1}{\mu/N}\right)^N/N!}\right)(1-v_\ell). \end{cases} \tag{21}$$

In Propositions 3, we have proven that for a finite N, there must exist at least one strict randomization strategy $\beta \in (0,1)$ when P is above some (unique) threshold $\hat{P}$. To reduce the notational burden, we let $F_N = F(\alpha_1, \mu/N)$ and $G_N = \frac{1}{N!}\left(\frac{\alpha_1}{\mu/N}\right)^N / F_N$ and indicate explicitly the dependency on N. With this notation (i.e., for a given α_1), the last equation in the equation set (21) only depends on α_1 and β:

$$G_N = \left(\frac{L}{L+1-\beta} - \frac{L}{L+1+\beta F_N\left(\alpha_1 - G_N\right)}\right)(1-v_\ell). \tag{22}$$

Equation (22) is quadratic in β. Hence, we can solve for β as a function of α_1 and obtain

$$\beta_N^{\pm}(\alpha_1) = -\frac{1}{2}\left(\begin{array}{c} \frac{L(1-v_\ell)+(1+L)G_N}{G_N F_N(\alpha_1-G_N)} + \frac{L(1-v_\ell)}{G_N} - (1+L) \\ \pm\sqrt{\left(\frac{L(1-v_\ell)+(1+L)G_N}{G_N F_N(\alpha_1-G_N)} + \frac{L(1-v_\ell)}{G_N} - (1+L)\right)^2 + 4\frac{(1+L)^2}{F_N(\alpha_1-G_N)}} \end{array}\right). \tag{23}$$

Notice that

$$G_N = \frac{\alpha_1^N \left(\frac{1}{\mu/N}\right)^N / N!}{\sum_{k=1}^N \frac{\alpha_1^{k-1}}{k!}\left(\frac{1}{\mu/N}\right)^k} = \alpha_1 \frac{\left(\frac{\alpha_1}{\mu}N\right)^N / N!}{\sum_{k=1}^N \frac{1}{k!}\left(\frac{\alpha_1}{\mu}N\right)^k}$$

and, with Lemma 5, $\lim_{N\to+\infty} G_N = \alpha_1\left(1-\frac{\mu}{\alpha_1}\right)^+ = (\alpha_1-\mu)^+$.

Now, we consider two cases: $1 < \frac{\alpha_1}{\mu}$ and $1 > \frac{\alpha_1}{\mu}$. Assume first that $1 < \frac{\alpha_1}{\mu}$, such that $\lim_{N\to+\infty} G_N = \alpha_1 - \mu$. As $\lim_{N\to+\infty} F_N = +\infty$, the roots corresponding with the + and − signs in Equation (23) tend to

$$\begin{aligned} \lim_{N\to+\infty} \beta_N^+ &= -\frac{1}{2}\lim_{N\to+\infty}\left(\frac{L(1-v_\ell)}{G_N} - (1+L) + \left(\frac{L(1-v_\ell)}{G_N} - (1+L)\right)\right) \\ \lim_{N\to+\infty} \beta_N^- &= -\frac{1}{2}\lim_{N\to+\infty}\left(\frac{L(1-v_\ell)}{G_N} - (1+L) - \left(\frac{L(1-v_\ell)}{G_N} - (1+L)\right)\right). \end{aligned}$$

Hence,

$$\lim_{N\to+\infty} \beta_N^+ = 0 \text{ and } \lim_{N\to+\infty} \beta_N^- = 1 + L - L\frac{1-v_\ell}{\alpha_1-\mu}$$

are two solutions when $\frac{\alpha_1}{\mu} > 1$. Next, assume that $\frac{\alpha_1}{\mu} < 1$, such that $\lim_{N\to+\infty} G_N = 0$ and $\lim_{N\to+\infty} F_N = +\infty$:

$$\beta_N^{\pm} = -\frac{1}{2}\left(\begin{array}{c} \frac{L(1-v_\ell)+(1+L)G_N}{G_N F_N(\alpha_1-G_N)} + \frac{L(1-v_\ell)}{G_N} - (1+L) \\ \pm\sqrt{\left(\frac{L(1-v_\ell)+(1+L)G_N}{G_N F_N(\alpha_1-G_N)} + \frac{L(1-v_\ell)}{G_N} - (1+L)\right)^2 + 4\frac{(1+L)^2}{F_N(\alpha_1-G_N)}} \end{array}\right)$$

$$\begin{aligned} \lim_{N\to+\infty} \beta_N^+ &= -\frac{1}{2}\left(\frac{L(1-v_\ell)}{G_N} - (1+L) + \left(\frac{L(1-v_\ell)}{G_N} - (1+L)\right)\right) = -\infty \\ \lim_{N\to+\infty} \beta_N^- &= -\frac{1}{2}\left(\frac{L(1-v_\ell)+(1+L)G_N}{G_N F_N\left(\alpha_1-G_N\right)} + \frac{L(1-v_\ell)}{G_N} - (1+L)\right) \times \\ &\quad \left(1 - \sqrt{1 + 4\frac{\frac{(1+L)^2}{F_N(\alpha_1-G_N)}}{\left(\frac{L(1-v_\ell)+(1+L)G_N}{G_N F_N(\alpha_1-G_N)} + \frac{L(1-v_\ell)}{G_N} - (1+L)\right)^2}}\right) \\ &= \lim_{N\to+\infty} \frac{(1+L)^2}{\frac{L(1-v_\ell)+(1+L)G_N}{G_N} + F_N\left(\alpha_1-G_N\right)\left(\frac{L(1-v_\ell)}{G_N} - (1+L)\right)} = 0 \end{aligned}$$

We can discard β_N^+.

Now we can find a $\hat{P}$ such that $\lim_{N\to+\infty}\beta_N = 1$ and $\alpha_1 = 1-\hat{P}$ solves Equation set (21). That is only possible when $1 < \frac{\alpha_1}{\mu}$, with the negative root: β_N^-, 1 or, equivalently, $1 = 1+L-L\frac{1-v_\ell}{1-\hat{P}-\mu}$ is solved by $\hat{P} = v_\ell - \mu$ (as $1+L-L\frac{1-v_\ell}{1-v_\ell} = 1+L-L = 1$). Combining the two cases, $\frac{\alpha_1}{\mu} < 1$ and $\frac{\alpha_1}{\mu} > 1$, we substitute the feasible limiting values of $\beta_N^\pm$, satisfying the third case of Equation set (21) into the first case and obtain the following set of equations in (α_0, α_1) only:

$$\frac{\alpha_1}{\mu} > 1: \begin{cases} 1+L-L\frac{1-v_\ell}{\alpha_1-\mu} &= 1+L-L\frac{1-v_\ell}{\alpha_1-(v_\ell-P)} \\ \alpha_0-\alpha_1 &= -(\alpha_1-\mu)^+ \end{cases} (\beta_N^+) \text{ or } \begin{cases} 0 &= 1+L-L\frac{1-v_\ell}{\alpha_1-(v_\ell-P)} \\ \alpha_0-\alpha_1 &= -(\alpha_1-\mu)^+ \end{cases} (\beta_N^-)$$

$$\frac{\alpha_1}{\mu} < 1: \begin{cases} 0 &= 1+L-L\frac{1-v_\ell}{\alpha_1-(v_\ell-P)} \\ \alpha_0-\alpha_1 &= -(\alpha_1-\mu)^+. \end{cases} (\beta_N^-)$$

Notice that when $\frac{\alpha_1}{\mu} > 1$, for the left equation set (corresponding with β_N^+) a solution for α_1 exists only when $v_\ell - P = \mu \Leftrightarrow P = \hat{P}$. In that case, any α_1 satisfies the first case. Hence, when $P = \hat{P}$, solution is degenerate; any $\alpha_1 \in (\mu, 1]$ satisfies the first case. When $P > \hat{P}$, the left equation set has no solution for α_1. Hence, it a solution exists, it must satisfy the right equation set (corresponding with β_N^-). Also, when $\frac{\alpha_1}{\mu} < 1$, if a solution exists, it must satisfy the equation set that corresponds with β_N^-. The only solution that satisfies Equation set (21) for large values of N is thus $\lim_{N\to+\infty}\beta_N^- = 0$. Note that this is a limiting result as, for all finite N, $\beta_N > 0$. Below, we solve the first case in these sets for α_1. α_0 is determined by substituting the solution of α_1 in the second case in these sets.

(i) For $0 \le P < \hat{P}$: (recall that the derivations are not reported for brevity), As $\beta = 1$, we obtain

$$\alpha_1 = 1-P \text{ and } \alpha_0 = v_\ell - P$$

from which follow via Lemma 3: $\lim_{N\to+\infty} r(\boldsymbol{\alpha},\mu) = \min\{1-P,\mu\}$ and $\lim_{N\to+\infty} r(\boldsymbol{\alpha},\mu) = v_\ell - P$.

(ii) For $\hat{P} < P \le v = v_\ell + \frac{L}{L+1}(1-v_\ell)$. The only possible solution is $\lim_{N\to+\infty}\beta_N^+ = 0$ and

$$\begin{cases} 0 &= 1+L-L\frac{1-v_\ell}{\alpha_1-(v_\ell-P)} \\ \alpha_0-\alpha_1 &= -(\alpha_1-\mu)^+ \end{cases}$$
$$\Leftrightarrow \begin{cases} \alpha_1 &= \underbrace{v_\ell + L\frac{1-v_\ell}{L+1}}_{=v} - P \\ \alpha_0 &= \alpha_1 - (\alpha_1-\mu)^+ \end{cases}$$

or $\alpha_1 = v - P$ and $\alpha_0 = v - P - (v - P - \mu)^+ = \min(\mu, v - P)$, which, following Lemma 3, yields $\lim_{N\to+\infty} r(\boldsymbol{\alpha}, \mu) = \min\{v - P, \mu\}$ and $\lim_{N\to+\infty} r(\boldsymbol{\alpha}, \mu) = v_\ell - P$.

B. Intuitive Equilibrium Prices: Thus, profits for the high-quality firm in any pooling-price equilibria must be higher than $\Pi_h^{\text{dev}} = \max_{0\le P\le v_\ell} P\min(\mu, v_\ell - P)$. For the low-quality firm, the equilibrium profits must be higher than $\Pi_\ell^{\text{dev}} = \max_{0\le P\le v_\ell} P(v_\ell - P) = \frac{1}{4}v_\ell^2$. We first discuss the case when the high-quality firm's deviation profits, Π_h^{dev}, are determined by its capacity, $(v_\ell - \mu)\mu$. This is when its *service rate is low*, $\mu < v_\ell/2$. Then, we discuss a case in which the high-quality firm's deviation profits are $\frac{1}{4}v_\ell^2$, for *high service* $v_\ell/2 < \mu (< v_\ell)$.

Low service rate:

We assess now whether a price P can be a pooling price for $P \le \hat{P} = v_\ell - \mu$. For these low prices, even when business is slow (the firm is empty), the arrival rate, $v_\ell - P$, exceeds the service rate, μ. Two necessary conditions for a pooling price are $P\mu \ge \Pi_h^{\text{dev}} = (v_\ell - \mu)\mu$ and $P(v_\ell - P) \ge \Pi_\ell^{\text{dev}} = \frac{1}{4}v_\ell^2$. Equivalently, $P \ge v_\ell - \mu$ and $P = v_\ell/2$. As $P \le v_\ell - \mu$, the first inequality is only satisfied for $P = v_\ell - \mu$, which contradicts the second condition. Hence, no low pooling price is possible below $\hat{P}$.

We now assess whether there can be a pooling price for $v_\ell - \mu < P < v$. Recall that the intuitive-criterion conditions are trivially satisfied as the high- and low-quality firms' profits are equal. We only need to confirm that

$$v_\ell - \mu < P \le v : P\min(\mu, v - P) \ge \frac{1}{4}v_\ell^2.$$

Notice that $v - \mu < v/2 \Leftrightarrow v/2 < \mu$. Because $\mu < v_\ell/2$ the maximum of $P(v - P)$ can never be unconstrained. As a result, $v - \mu$ is the maximum of $P\min(\mu, v - P)$ and we can determine the range of pooling prices as

$$\frac{1}{4}\frac{v_\ell^2}{\mu} \le P^* \le \frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right). \tag{24}$$

Notice that it is possible that no pooling price (higher than $v_\ell - \mu$) exists when

$$\frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right) \le \frac{1}{4}\frac{v_\ell^2}{\mu}$$

because then the region is empty. This will happen when v is small (close to v_ℓ)—$\frac{1}{2}v_\ell \le \frac{1}{4}\frac{v_\ell^2}{\mu}$—which is possible because $\mu < v_\ell/2$.

High service rate:

Now we consider the case $v_\ell/2 < \mu (< v_\ell)$. We assess whether pooling prices can exist for $P < \hat{P} = v_\ell - \mu$. First, we check

$$P\mu \geq \Pi_h^{\text{dev}} = \frac{1}{4}v_\ell^2 \text{ and } P(v_\ell - P) \geq \Pi_\ell^{\text{dev}} = \frac{1}{4}v_\ell^2.$$

The only possibility is $P = v_\ell/2$, which cannot be an equilibrium as $v_\ell/2 > v_\ell - \mu$. Hence, no pooling price can exist for low prices (below $\hat{P}$). Next, we assess whether a pooling price exists for $v_\ell - \mu < P < v$. Again, the intuitive-criterion conditions are trivially satisfied. We only need to check whether

$$P \min(\mu, v - P) \geq \frac{1}{4}v_\ell^2.$$

Now we can characterize the pooling prices for the low service rates

$$(v-\mu)\,\mu < \frac{1}{4}v_\ell^2 \Rightarrow \frac{1}{2}\left(v - \sqrt{v^2 - v_\ell^2}\right) \leq P^* \leq \frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right)$$
$$(v-\mu)\,\mu > \frac{1}{4}v_\ell^2 \Rightarrow \frac{1}{4}\frac{v_\ell^2}{\mu} \leq P^* \leq \frac{1}{2}\left(v + \sqrt{v^2 - v_\ell^2}\right) \qquad \blacksquare$$

Proof of Proposition 5: This proof follows exactly the same logic as for Proposition 1. For brevity, we only discuss the key differences. Concavity of the profit function in the price follows from Lemma 2. For price P' satisfying $P' r(\boldsymbol{\alpha}_h(P'), \mu) = \max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), \mu)$, in order to have $P' r(\boldsymbol{\alpha}_h(P'), +\infty) > \max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), +\infty)$, v_ℓ must be low enough. We rewrite

$$\frac{r(\boldsymbol{\alpha}_h(P'), +\infty)}{r(\boldsymbol{\alpha}_h(P'), \mu)} > \frac{\max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), +\infty)}{\max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), \mu)}.$$

Letting $P_\ell = \arg\max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), +\infty) = \frac{v_\ell}{2}$, then $\max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P) > \frac{v_\ell}{2} r(\boldsymbol{\alpha}_\ell(\frac{v_\ell}{2}), \mu)$ we have that

$$\frac{\frac{1}{2}v_\ell}{r(\boldsymbol{\alpha}_\ell(\frac{v_\ell}{2}), \mu)} > \frac{\max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), +\infty)}{\max_{0 \leq P \leq v_\ell} Pr(\boldsymbol{\alpha}_\ell(P), \mu)}.$$

Then we should check whether

$$\frac{1 - P'}{r(\boldsymbol{\alpha}_h(P'), \mu)} > \frac{\frac{1}{2}v_\ell}{r(\boldsymbol{\alpha}_\ell(\frac{v_\ell}{2}), \mu)} \Leftrightarrow \frac{1 - P'}{\frac{1}{2}v_\ell} > \frac{r(\boldsymbol{\alpha}_h(P'), +\infty)}{r(\boldsymbol{\alpha}_\ell(\frac{v_\ell}{2}), \mu)} \Leftrightarrow$$

$$\frac{1 - P'}{\frac{1}{2}v_\ell} > \frac{(1 - P')\left(1 - \frac{\frac{1}{N!}\left(\frac{1-P'}{\mu}\right)^N}{1 + \sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P'}{\mu}\right)^k}\right)}{\frac{v_\ell}{2}\left(1 - \frac{\frac{1}{N!}\left(\frac{\frac{v_\ell}{2}}{\mu}\right)^N}{1 + \sum_{k=1}^N \frac{1}{k!}\left(\frac{\frac{v_\ell}{2}}{\mu}\right)^k}\right)} \Leftrightarrow$$

$$\frac{\frac{1}{N!}\left(\frac{\frac{v_\ell}{2}}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{\frac{v_\ell}{2}}{\mu}\right)^k} < \frac{\frac{1}{N!}\left(\frac{1-P'}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{1-P'}{\mu}\right)^k}$$

As $\frac{\frac{1}{N!}\left(\frac{\lambda}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}$ is increasing in λ, the inequality is satisfied if $\frac{v_\ell}{2} < 1-P' \Leftrightarrow P' < 1-\frac{v_\ell}{2}$. It is obvious that $P' < v_\ell$. Hence, if $v_\ell < 1-\frac{v_\ell}{2}$ or $v_\ell < \frac{2}{3}$, the inequality is always satisfied. ■

Proof of Proposition 6: It is obvious that when the high-quality firm's service rate is $+\infty$ and the low-quality firm's service rate is μ, consumers know for sure that the quality is low when arriving at a nonempty firm: $\alpha_1 = [v_\ell - P]^+$. From Proposition 2, the marginal return to speeding up from μ to $+\infty$ is

$$\Delta = \alpha_0 \frac{(\alpha_0-\alpha_1)\, F(\alpha_1) + \alpha_1 \frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N}{1+\sum_{k=1}^N \frac{1}{k!}\alpha_0\alpha_1^{k-1}\left(\frac{1}{\mu}\right)^k}$$

and

$$\Delta > 0 \Leftrightarrow \frac{1}{N!}\alpha_1^{N-1}\left(\frac{1}{\mu}\right)^N > (\alpha_1-\alpha_0)\, F(\alpha_1).$$

The joining probability at $n=0$ is determined by

$$\alpha_0 = (v_\ell - P) + \frac{\frac{p}{1-p}\,(1+\alpha_0 F(\alpha_1,\mu/N))}{\frac{p}{1-p}\,(1+\alpha_0 F(\alpha_1,\mu/N))+1}(1-v_\ell),$$

from which it follows that $\alpha_0 > v_\ell - P$ and

$$\alpha_0 = -\frac{1}{2}\left(\frac{1}{F(\alpha_1,\mu/N)p} + P - 1 \pm \sqrt{\left(\frac{1}{F(\alpha_1,\mu/N)p}+P-1\right)^2 + 4\frac{(v_\ell-P)+\frac{p}{1-p}\,(1-P)}{F(\alpha_1,\mu/N)\frac{p}{1-p}}}\right).$$

When $\alpha_1 = 0$, $F(\alpha_1,\mu/N) = +\infty$ and, hence,

$$\alpha_0 = (v_\ell - P) + (1-v_\ell) = 1-P.$$

Notice that $\alpha_0 > \alpha_1 = [v_\ell - P]^+$. Hence, $\Delta > 0$. That is, the high-quality firm is in equilibrium and never randomizes. Firm-rationality and the intuitive-criterion conditions follow exactly the same logic as they did in the proof of Proposition 2 and are not reported for brevity. ■

Proof of Proposition 7: For notational simplicity, we set $\Lambda_0 = 1$. When the high-quality firm is fast, the condition for α_0 is

$$\alpha_0 = (v_\ell - P) + \frac{L\,(1+\alpha_0 F(\alpha_1,\mu/N))}{L\,(1+\alpha_0 F(\alpha_1,\mu/N))+1}(1-v_\ell).$$

For prices above v_ℓ, $\alpha_1 = 0$. As $F(\alpha_1, \mu/N) \to +\infty$ for $\alpha_1 > 0$ and for $\alpha_1 \to 0^+$, we obtain

$$\lim_{\alpha\to 0+} F(\alpha,\mu) = \sum_{k=1}^{N} \frac{\alpha^{k-1}}{k!}\left(\frac{1}{\mu}\right)^k = \frac{1}{\mu} \text{ and } \lim_{\alpha_1\to 0^+}\lim_{N\to+\infty} F(\alpha_1, \mu/N) = +\infty.$$

The condition of α_0 becomes $\alpha_0 = (v_\ell - P) + (1 - v_\ell) = 1 - P$, from which

$$\alpha_0^p(P) = 1 - P.$$

Recall that $\alpha_1 = [v_\ell - P]^+$. The low-quality firm's profits are determined by $P\min\{\alpha_1, \mu\} = P\min(\mu, [v_\ell - P]^+)$ as with Lemma 3. The high-quality firm's profits are determined by $P\alpha_0 = P(1-P)$.

Intuitive pooling prices:

Two necessary conditions for a pooling price are

$$P(1-P) \geq \Pi_h^{\text{dev}} = \frac{1}{4}v_\ell^2 \text{ and}$$
$$P\min(\mu, (v_\ell - P)^+) \geq \Pi_\ell^{\text{dev}} = \max_{0\leq P\leq v_\ell} P\min(\mu, (v_\ell - P)^+).$$

From the second condition, it follows that P must be equal to the maximum of $P\min(\mu, (v_\ell - P)^+)$, which is $\max(\frac{v_\ell}{2}, v_\ell - \mu)$. Hence,

$$\frac{1}{2}\left(1 - \sqrt{1 - v_\ell^2}\right) \leq \max(\frac{v_\ell}{2}, v_\ell - \mu) \leq \frac{1}{2}\left(1 + \sqrt{1 - v_\ell^2}\right). \tag{25}$$

If the above condition is satisfied, we need to check the intuitive-criterion conditions. It is easy to see[13] that $\frac{1}{2}\left(1 - \sqrt{1 - v_\ell^2}\right) \leq \frac{v_\ell}{2} \leq \frac{1}{2}\left(1 + \sqrt{1 - v_\ell^2}\right)$ for any $0 < v_\ell < 1$. Hence, when $\frac{v_\ell}{2} > v_\ell - \mu$, the condition of Equation (25) is always satisfied. When $\frac{v_\ell}{2} < v_\ell - \mu \Leftrightarrow \mu < \frac{v_\ell}{2}$, it is easy to observe that $\frac{1}{2}\left(1 + \sqrt{1 - v_\ell^2}\right) < v_\ell - \mu$ exists for some (μ, v_ℓ) values such that $0 \leq \mu \leq \frac{1}{2}$ and $v_\ell \geq \frac{4}{5}$. In these cases, the condition of Equation (25) is not satisfied. In other words, when $v_\ell < \frac{4}{5}$, $P = v_\ell - \mu$ always satisfies the condition of Equation (25) and is thus a candidate pooling price if the intuitive-criterion conditions are also satisfied. Only if $v_\ell \geq \frac{4}{5}$, and then, only if μ is high enough—$\mu > v_\ell - \frac{1}{2}\left(1 + \sqrt{1 - v_\ell^2}\right)$[14]—$P = v_\ell - \mu$ becomes a candidate pooling price if the intuitive-criterion conditions are also satisfied.

[13] Because $1 - v_\ell \leq \pm\sqrt{1 - v_\ell^2} \Leftrightarrow (1 - v_\ell)^2 \leq 1 - v_\ell^2 \Leftrightarrow 1 - v_\ell \leq 1 + v_\ell$.

[14] Note that $\frac{4}{5} - \frac{1}{2}\left(1 + \sqrt{1 - (\frac{4}{5})^2}\right) = 0$.

Notice from Proposition 7 that the equilibrium profits, $P\alpha_0^p(P) = P(1-P)$, are the same as $\hat{\Pi}_h^{\text{dev}}(P) = P(1-P)$ because customers only join the empty firm at which they infer that the quality is high. Therefore,

$$\underline{P}_h(P) = \min(1-P, P) \text{ and } \overline{P}_h(P) = \max(1-P, P).$$

For the low-quality firm, $\hat{\Pi}_\ell^{\text{dev}}(P) = P\min(\mu, 1-P)$. We find the roots of $\hat{\Pi}_\ell^{\text{dev}}(P') = P\min(\mu, [v_\ell - P]^+)$ as follows.

$$\begin{aligned}\underline{P}_\ell(P) &= \max\left(\frac{1}{\mu}P\min(\mu,[v_\ell-P]^+), \frac{1}{2}\left(1-\sqrt{1-4P\min(\mu,[v_\ell-P]^+)}\right)\right)\\ \text{and } \overline{P}_\ell(P) &= \frac{1}{2}\left(1+\sqrt{1-4P\min(\mu,[v_\ell-P]^+)}\right).\end{aligned}$$

The necessary conditions for $P^* = \max(\frac{v_\ell}{2}, v_\ell - \mu)$ to be a pooling price are

$$\underline{P}_\ell(P^*) \le \underline{P}_h(P^*) \text{ and } \overline{P}_h(P^*) \le \overline{P}_\ell(P^*) \text{ and } \frac{1}{2}\left(1-\sqrt{1-v_\ell^2}\right) \le P^* \le \frac{1}{2}\left(1+\sqrt{1-v_\ell^2}\right).$$

We can now analyze the above conditions by considering a low service rate, $\mu < \frac{v_\ell}{2}$, for which $P^* = v_\ell - \mu$ and a fast service rate, $\mu > \frac{v_\ell}{2}$, for which $P^* = \frac{v_\ell}{2}$:

Low service rate:

Assume that $\mu < \frac{v_\ell}{2}$, then $P^* = v_\ell - \mu$,

$$\underline{P}_h(P^*) = \min(v_\ell - \mu, 1 - v_\ell + \mu),\ \overline{P}_h(P^*) = \max(v_\ell - \mu, 1 - v_\ell + \mu)$$

$$\text{and } \underline{P}_\ell(P^*) = v_\ell - \mu \text{ and } \overline{P}_\ell(P^*) = \frac{1}{2}\left(1+\sqrt{1-4(v_\ell-\mu)\mu}\right).$$

$P^* = v_\ell - \mu$ is a pooling price when

$$\begin{aligned}&\min(v_\ell-\mu, 1-v_\ell+\mu) \le v_\ell - \mu \text{ and} \max(v_\ell-\mu, 1-v_\ell+\mu) \le \frac{1}{2}\left(1+\sqrt{1-4(v_\ell-\mu)\mu}\right)\\ &\text{and } \frac{1}{2}\left(1-\sqrt{1-v_\ell^2}\right) \le v_\ell - \mu \le \frac{1}{2}\left(1+\sqrt{1-v_\ell^2}\right).\end{aligned}$$

If $v_\ell - \mu < 1/2$, then the conditions are always satisfied; the prices for which the high-quality firm can obtain higher profits are then to the right of P^*, similarly to the low-quality firm. The region for which the high-quality firm can obtain higher profits is smaller than for the low-quality firm. Therefore, no prices at which only the high-quality firm can deviate exist and the intuitive-criterion conditions are satisfied. It can be shown that for all $v_\ell - \mu < 1/2$, the condition of Equation (25) is also satisfied. As a consequence, $P^* = v_\ell - \mu$ is a (unique) pooling price. If $v_\ell - \mu > 1/2$, the prices for which the high-quality firm has an incentive to deviate are to the left of P^*, while the prices for which the low-quality firm can obtain higher profits are to the right of P^*. It follows that the intuitive-criterion condition cannot be satisfied.

High service rate:

Assume that $\mu > \frac{v_\ell}{2}$. Then, $P^* = \frac{v_\ell}{2}$ and

$$\underline{P}_h(P^*) = \frac{v_\ell}{2} \text{ and } \overline{P}_h(P^*) = 1 - \frac{v_\ell}{2}$$

and $\min(\mu, (v_\ell - P^*)^+) = \min(\mu, \frac{v_\ell}{2}) = \frac{v_\ell}{2}$. Therefore,

$$\begin{aligned} \underline{P}_\ell(P^*) &= \max\left(\frac{1}{\mu}\left(\frac{v_\ell}{2}\right)^2, \frac{1}{2}\left(1 - \sqrt{1 - v_\ell^2}\right)\right) \\ \text{and } \overline{P}_\ell(P^*) &= \frac{1}{2}\left(1 + \sqrt{1 - v_\ell^2}\right). \end{aligned}$$

Again, the prices for which the high-quality firm can obtain higher profits are then to the right of P^*. It is easy to see that $\overline{P}_\ell(P^*) > 1 - \frac{v_\ell}{2}$ (because the low-quality firm's profits are lower). It is also easy to see that $\underline{P}_\ell(P^*) < \frac{v_\ell}{2}$. If the capacity constraint is binding (i.e., $\underline{P}_\ell(P^*) = \frac{1}{\mu}\left(\frac{v_\ell}{2}\right)^2$), then from $\mu > \frac{v_\ell}{2}$ follows $\frac{1}{\mu}\left(\frac{v_\ell}{2}\right)^2 < \frac{v_\ell}{2}$. Otherwise, $\frac{v_\ell}{2} > \underline{P}_\ell(P^*)$ (because the low-quality firm's profits are lower). It can be shown that for all $\mu > \frac{v_\ell}{2}$, the condition of Equation (25) is also satisfied. As a consequence, $P^* = \frac{v_\ell}{2}$ is a (unique) pooling price. ∎

References

Messerli, E. J. 1972. Proof of a Convexity Property of the Erlang B Formula, Bell System Technical Journal, 51, 951–953.

Ross, S. M. 1996. Stochastic Processes. Wiley and Sons.

Appendix B. Finite Fast Service at Positive Cost

In this section, we relax the restriction that fast service is infinitely fast at no cost. We consider again two cases: In the first case, the high-quality firm's service rate is $\underline{\mu}$ and cannot be changed. The low-quality firm's service rate is either $\underline{\mu}$ or $\overline{\mu} > \underline{\mu}$ at some cost $k > 0$. In the second case, the low-quality firm's service rate is $\underline{\mu}$ and cannot be changed, while the high-quality firm's service rate is either $\underline{\mu}$ or $\overline{\mu} > \underline{\mu}$ at some cost $k > 0$. For conciseness, we present both cases together. The insights obtained are similar to those of the the base model. Due to the analytical challenge to obtaining structural expressions for the pooling-price equilibria, we present numerical examples.

For separating prices:

PROPOSITION 8. *When $\underline{\mu} < \overline{\mu} < +\infty$, $k_\ell = k < k_h = +\infty$, intuitive separating prices exist where P_h^* is the highest root of*

$$\max\{P \times r(\boldsymbol{\alpha}_h(P), \underline{\mu}), P \times r(\boldsymbol{\alpha}_h(P), \overline{\mu}) - k\} = \underline{\Pi}_\ell$$

and P_ℓ^* is the maximizer of $\underline{\Pi}_\ell = \max_{0 \le P \le v_\ell} \max\{P \times r(\boldsymbol{\alpha}_\ell(P), \underline{\mu}), P \times r(\boldsymbol{\alpha}_\ell(P), \overline{\mu}) - k\}$ if

$$P \times r(\boldsymbol{\alpha}_h(P_h^*), \underline{\mu}) \ge \underline{\Pi}_h = \max_{0 \le P \le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P), \underline{\mu}).$$

When $\underline{\mu} < \overline{\mu} < +\infty$, $k_h = k < k_\ell = +\infty$, intuitive separating prices exist where P_h^* is the lowest root of

$$P \times r(\boldsymbol{\alpha}_h(P), \underline{\mu}) = \underline{\Pi}_\ell$$

and P_ℓ^* is the solution of $\underline{\Pi}_\ell = \max_{0 \le P \le v_\ell} P \times r(\boldsymbol{\alpha}_\ell(P), \underline{\mu})$ if

$$\max\{P \times r(\boldsymbol{\alpha}_h(P_h^*), \underline{\mu}), P \times r(\boldsymbol{\alpha}_h(P_h^*), \overline{\mu}) - k\} \ge \underline{\Pi}_h$$
$$= \max_{0 \le P \le v_\ell} \max\{P \times r(\boldsymbol{\alpha}_\ell(P), \underline{\mu}), P \times r(\boldsymbol{\alpha}_\ell(P), \overline{\mu}) - k\}.$$

Proof of Proposition 8: Follows exactly the same arguments as Proposition 1. ■

Discussion: This proposition is a straightforward extension of Propositions 1 and 5.

For pooling prices: We analyze the case when the high-quality firm's speed is restricted to $\underline{\mu}$. Let the service rate strategy be $\boldsymbol{\beta}^* = (\beta_\ell^*, 0)$, with $0 \le \beta_\ell^* \le 1$. That is, we allow the low-quality firm to randomize between slow, $\underline{\mu}$, and fast, $\overline{\mu}$. Let $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}^*(P^*)$ denote the consumers' joining strategy and $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}^*(P^*)$ be the updated belief that the quality is high at the equilibrium (pooling) price. Now, for a given consumer's joining strategy, $\boldsymbol{\alpha}$, we introduce Φ, which is the ratio of the probability that the firm is empty upon arrival when the service rate is $\underline{\mu}$ and the probability that the firm is empty upon arrival when the service rate is $\overline{\mu}$; $\frac{\pi(0,\boldsymbol{\alpha},\underline{\mu})}{\pi(0,\boldsymbol{\alpha},\overline{\mu})} = \Phi$. From Lemma 1, we can write the ratio of the probabilities at any queue length, $\frac{\pi(n,\boldsymbol{\alpha},\underline{\mu})}{\pi(n,\boldsymbol{\alpha},\overline{\mu})}$, as a function of Φ: $\frac{\pi(n,\boldsymbol{\alpha},\underline{\mu})}{\pi(n,\boldsymbol{\alpha},\overline{\mu})} = \frac{\pi(0,\boldsymbol{\alpha},\underline{\mu})}{\pi(0,\boldsymbol{\alpha},\overline{\mu})}(\frac{\overline{\mu}}{\underline{\mu}})^n = \Phi(\frac{\overline{\mu}}{\underline{\mu}})^n$. Let $(\boldsymbol{\alpha}^p(P), \beta^p(P), \Phi^p(P))$ satisfy

$$\alpha(n) = v_\ell + \frac{p\Phi(\frac{\overline{\mu}}{\underline{\mu}})^n}{p\Phi(\frac{\overline{\mu}}{\underline{\mu}})^n + (1-p)(\beta + (1-\beta)\Phi(\frac{\overline{\mu}}{\underline{\mu}})^n)}(1 - v_\ell) - P, \forall n \in [0, N-1]$$
$$\Phi = \frac{1 + \sum_{k=1}^N ((\Lambda_0 N/\overline{\mu})^k / k!) \prod_{m=0}^{k-1} \alpha(m)}{1 + \sum_{k=1}^N ((\Lambda_0 N/\underline{\mu})^k / k!) \prod_{m=0}^{k-1} \alpha(m)}$$
$$P\Delta(\boldsymbol{\alpha}^p(P)) > (<) k \Rightarrow \beta^p(P) = 1(0) \text{ and if } P\Delta(\boldsymbol{\alpha}^p(P)) = k \Rightarrow \beta^p(P) \in [0,1].$$

In equilibrium, at the pooling price P^*, $\boldsymbol{\alpha}^*$ is consistent with $\boldsymbol{\beta}^*$ only if $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}^p(P^*)$ and $\boldsymbol{\beta}^* = \beta^p(P^*)$ with the belief

$$\gamma^*(n, P^*) = \frac{p\Phi^*(\frac{\overline{\mu}}{\underline{\mu}})^n}{p\Phi^*(\frac{\overline{\mu}}{\underline{\mu}})^n + (1-p)(\beta_\ell^* + (1-\beta_\ell^*)\Phi^*(\frac{\overline{\mu}}{\underline{\mu}})^n)}$$

that the quality of the firm is high after observing a congestion level of n, where $\Phi^* = \Phi^p(P^*)$. Obviously these are necessary, but not sufficient conditions for a pooling equilibrium. It remains to be verified whether the pooling prices can be sustained by a belief that satisfies the intuitive criterion. The conditions are similar to the case of $\overline{\mu} = +\infty$. For brevity, we do not reproduce the conditions, but provide an illustrative numerical example.

Example: In Figure 9, we illustrate for $P = 0.62$ the updated belief that the quality is high due to observed congestion upon arrival. For this price, the low-quality firm randomizes the service rate. Assume that $\beta^p(P) = 1$. That is, the low-quality firm speeds up. It is analytically challenging to determine the sign of $P\Delta(\boldsymbol{\alpha}^p(P)) - k$, i.e., where it is rational for the low-quality firm to speed up. The price has to be high enough to justify the fixed costs of increased speed. However, from the base model, if the price is too high, then the marginal benefit of speeding up will be low (and can even become negative). Hence, it is intuitive that speeding up for the low-quality firm may be rational only when the prices are neither too high nor too low. In the case that speeding up

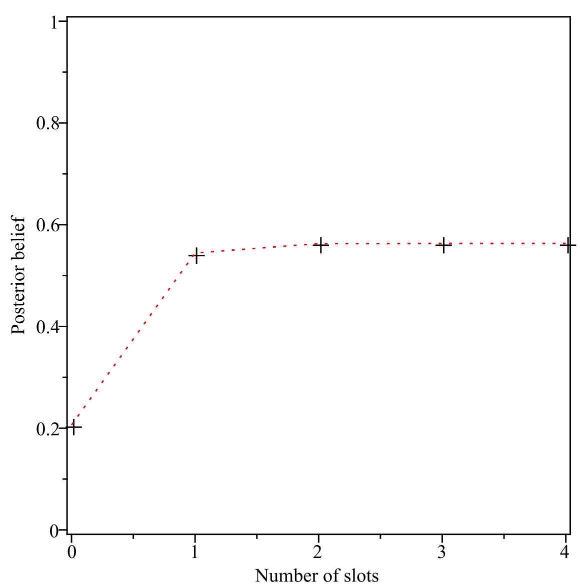

Figure 9 Posterior that the quality is high as a function of the number of slots occupied for $P = 0.62$ and the high-quality firm is slower than the low-quality firm (which randomizes; $\beta_\ell(P) = 0.224$). The other parameters are $N = 5$, $\overline{\mu} = 0.5$, $\underline{\mu} = 0.1$, $v_\ell = 0.8$, $p = 1/2$ and $K = 0.0125$.

is not rational, it is easy to observe that $\beta^p(P) = 0$ is an equilibrium when $k > \Delta(\hat{\boldsymbol{\alpha}}(P))P$, where $\hat{\alpha}(n,P) = v - P, \forall n \in \{0,\ldots,N\}$. Hence, when $P\Delta(\boldsymbol{\alpha}^p(P)) > k > P\Delta(\hat{\boldsymbol{\alpha}}(P))$, the low-quality firm must randomize its service rate strategy; $\beta^p(P) \in (0,1)$.

We plot $P\Delta(\hat{\boldsymbol{\alpha}}(P))$, $P\Delta(\boldsymbol{\alpha}^p(P))$ and k. In Figure 10, left panel, notice that $P\Delta(\hat{\boldsymbol{\alpha}}(P)) \geq P\Delta(\boldsymbol{\alpha}^p(P))$. Hence, there are five price regions with different service rate strategies. When $k > P\Delta(\hat{\boldsymbol{\alpha}}(P))$ (which holds for very low or very high prices), $\beta^*(P) = 0$. The returns to speeding up are too low because of the low price or low volume. Consumers do not learn from queues. For intermediate prices, $P\Delta(\boldsymbol{\alpha}^p(P)) > k$. It follows that the low-quality firm has incentives to speed up, making congestion upon arrival informative. However, the profit difference is large, which eliminates many pooling prices via the intuitive criterion. Finally, there exist two transition regions where $P\Delta(\hat{\boldsymbol{\alpha}}(P)) > k > P\Delta(\boldsymbol{\alpha}^p(P))$. The transition region at lower prices is very small. In these regions, if the low-quality firm speeds up and the consumers learn from the congestion information, the returns from speeding-up are less than the cost of speeding up. However, when the low-quality firm does not speed up and consumers do not learn from congestion, the low-quality firm does have an incentive to speed up. Hence, the low-quality firm randomizes its service rate and makes the

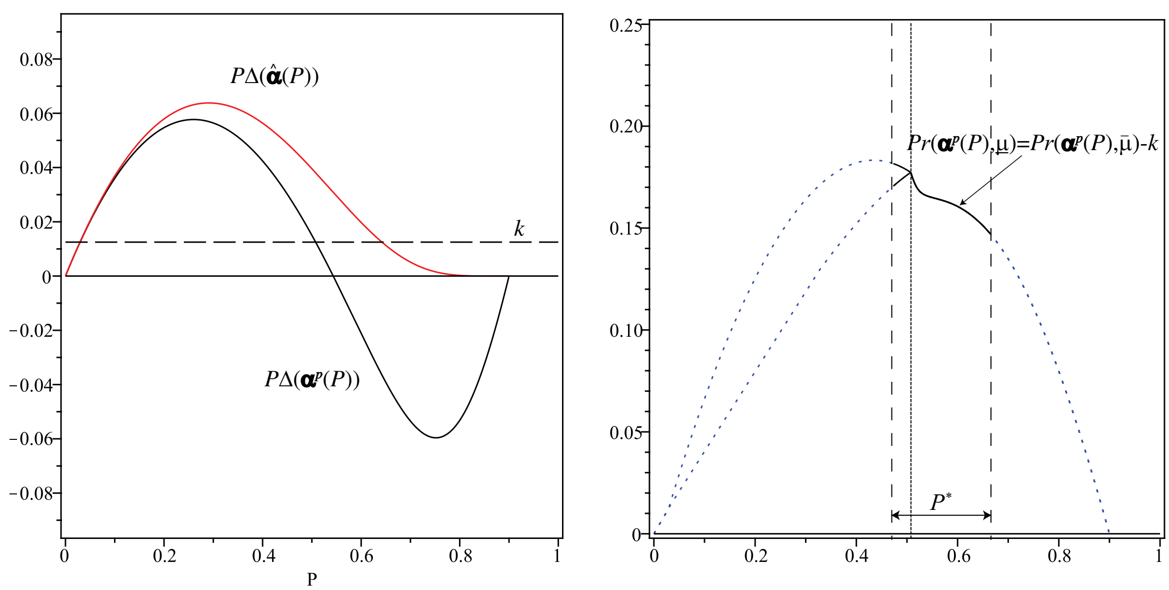

Figure 10 **Left panel: Illustration of $P\Delta(\hat{\alpha}(P))$ (top curve), $P\Delta(\alpha^p(P))$ (bottom curve) and k (dashed line). When for low or for high prices, $k > P\Delta(\hat{\alpha}(P))$, the low-quality firm is slow for sure; $\beta^*(P) = 0$. When for medium-low prices, $P\Delta(\alpha^p(P)) > k$, the low-quality firm is fast for sure; $\beta^*(P) = 1$. Finally, when for medium-high prices, $P\Delta(\hat{\alpha}(P)) > k > P\Delta(\alpha^p(P))$, the low-quality firm garbles its service rate; $\beta^*(P) \in (0,1)$. Right panel: Illustration of the profits for pooling prices that survive the intuitive criterion. For brevity, we do not indicate the profit expressions at very price segment, we only indicate the profits of the high- and low-quality firm when the low-quality firm randomizes (that is $\beta^p(P) < 1$).**

same profits as the high-quality firm. These pooling prices survive the intuitive criterion. We also illustrate the pooling profits in Figure 10, right panel.