Large-scale Service Marketplaces: The Role of the Moderating Firm

Gad Allon
Kellogg School of Management, Northwestern University, Evanston, Illinois, g-allon@kellogg.northwestern.edu

Achal Bassamboo
Kellogg School of Management, Northwestern University, Evanston, Illinois, a-bassamboo@kellogg.northwestern.edu

Eren B. Çil
Lundquist College of Business, University of Oregon, Eugene, Oregon, erencil@uoregon.edu

Recently, large-scale, web-based service marketplaces, where many small service providers compete among themselves in catering to customers with diverse needs, have emerged. Customers who frequent these marketplaces seek quick resolutions and thus are usually willing to trade prices with waiting times. The main goal of the paper is to discuss the role of the moderating firm in facilitating information gathering, operational efficiency, and communication among agents. Surprisingly, we show that operational efficiency may be detrimental to the overall efficiency of the marketplace. Further, we establish that to reap the "expected" gains of operational efficiency, the moderating firm may need to complement the operational efficiency by enabling communication among its agents. The study emphasizes the scale of such marketplaces and the impact it has on the outcomes.

Key words: service operations; fluid models; asymptotic analysis; large games; non-cooperative game theory

1. Introduction

Recently, large-scale, web-based service marketplaces, where many small service providers (agents) compete among themselves in catering to customers with diverse needs, have emerged. Customers who frequent these marketplaces seek quick resolutions for their temporary problems and thus are usually willing to trade prices with waiting times. These marketplaces are typically operated by an independent firm, which we shall refer to as the *moderating firm.* The moderating firm establishes the infrastructure for the interaction between customers and agents. In particular, it provides the customers and the agents with the information required to make their decisions. These moderating firms vary with respect to their involvement in the marketplace. They can introduce operational tools that specify how the customers and the agents are matched together. For instance, while some of the moderating firms allow customers to choose a specific service provider directly, others allow customers to post their needs and let service providers apply, postponing the service provider selection decision of the customers until they obtain enough information about agents' availability. Moreover, moderating firms can introduce strategic tools that allow communication and collaboration among the agents themselves. These different involvements result in different economic and operational systems, and thus vary in their level of efficiency and the outcomes for both customers and service providers.

A typical example of such a marketplace is oDesk.com, where around 1,000,000 programmers compete to provide software solutions. oDesk.com allows for two types of interaction between customers and service providers. Customers can go directly to a programmer and ask him to provide the service. The customers are then queued for this specific agent. In this type of interaction, most of the time is spent waiting for the agent to complete his previous jobs (36% of the waiting time is spent from the moment the customer chooses the agent until the agent begins working.[1]). On the other hand, oDesk.com also allows customers to post jobs and wait while agents apply for the job. In this type of interaction, a negligible amount of time passes until more than 10 agents apply, leaving the decision at the hands of the customer. Another large-scale, online service marketplace is ServiceLive.com, which is a start-up owned by Sears Holding Company. ServiceLive.com (with the slogan of "your price, your time") caters to time and price-conscious customers and service providers in the home repair and improvement arena. ServiceLive.com allows its customers to choose among multiple agents once they describe their projects. This type of interaction between customers and service providers is equivalent to the second one described for oDesk.com. Both oDesk.com and ServiceLive.com receive 10% of the revenue obtained by the providers at service completion. In both marketplaces, the moderating firms allow customers to browse among tens of thousands of agents and communicate with different providers.

Both oDesk.com and ServiceLive.com are part of a growing industry of online service marketplaces. Alok Aggarwal, the chairman of Evalueserve.com, a market research company in Saratoga, CA, said "this market [the market for work outsourcing] is expected to grow 20% to $300 million in sales this year, with transactions between employers and the free-lancers totaling about $1.8 billion" (Flandez, 2008, October 13). In line with this, Gary Swart, CEO of oDesk.com, said that "the number of freelancers registered with the firm in America has risen from 28,000 at the end of 2008 to 247,000 at the end of April" (The Economist, 2010, May 13).

Motivated by these online service marketplaces, we aim to study the moderating firm's role in the service marketplace where the objective of the individual players, customers as well as service providers, is to maximize their own utility. We distinguish between three degrees of moderating

[1] This is based on data obtained from oDesk.com for about 10000 randomly chosen transactions.

firms' involvement in such markets: (1) *No-intervention*: the moderating firm restricts its involvement to providing the facility for agents to advertise their services and set their prices, and for customers to compare the different agents. (2) *Operational efficiency*: the moderating firm provides additional mechanisms that facilitate efficient matching between customers and service providers. These mechanisms aim at reducing the inefficiency associated with having the right agent with the right capability idle while a customer with similar needs is waiting in line for another agent. As we will discuss, a system in which customers post their needs and wait for agents' applications is an example of such a mechanism. (3) *Enabling Communication*: the moderating firm may allow providers to communicate among themselves and exchange information on prices and job requirements.

To study the different configurations possible in such marketplaces we consider a sequence of related games where the set of possible strategies and the solution concepts vary to reflect the different modes of interaction available in the marketplace, either between the customer and the agents or between the agents themselves. Specifically, we study the following three games:

No-intervention Model: In this game, each agent chooses his price and operates as a single-server queue. Customers then choose agents based on prices and waiting times. We characterize the Subgame Perfect Nash equilibrium in this game.

Operational Efficiency Model: In this game, the mechanism introduced by the moderating firm efficiently matches customers interested in purchasing the service at a particular price with the available agents charging that or a lower price. This mechanism achieves the desired level of efficiency by virtually grouping all agents charging the same price. In contrast to the no-intervention model, customers do not need to commit to a specific agent upon their arrival.

Communication Enabled Model: In this game, agents can exchange information in a non-committal, costless manner. As in the model with operational efficiency, all the agents charging the same price are virtually grouped, and customers choose the price/sub-pool. We would be interested in allowing limited pre-play communication among the agents within a noncooperative structure; i.e., the agents are free to discuss their pricing strategies but not allowed to make binding commitments. Ray (1996) claims that the possibility of pre-play communication have motivated the notion of strong Nash equilibrium, see Aumann (1959), which requires stability against deviations by every conceivable coalition. Following this idea, we use a refinement of the Subgame Perfect Nash Equilibrium concept that requires the equilibrium to be (limited size) coalition proof.

We next state our key findings along with the contributions of the paper:

1. We appear to be the first to distinguish between tools aimed at increasing the operational efficiency (which manifest themselves in different routing decisions) and tools aimed at changing the nature of the strategic interaction by enabling communication (which manifest themselves in different solutions concepts). We show these tools have a non-trivial impact on the outcomes for all involved parties, thus creating an opportunity for the moderating firm to exploit these tools to maximize its profit.

2. In analyzing a market with operational efficiency, we first show that only the prices in a small neighborhood of the operating cost of agents are sustained as equilibrium outcomes when supply exceeds demand. Further, when demand exceeds supply, we are able to show that operational efficiency leads to multiple equilibria in markets with a sufficiently large number of agents. In many of these equilibria, the emerging prices are lower than those arising in the market with no-intervention. The fact that operational efficiency may lead to loss in profits for both the agents and the moderating firm (even when supply is scarce) is counter-intuitive. The main intuition behind this result is that the strong pooling benefits associated with operational efficiency serve as a deterrent for deviation, practically from every price when demand exceeds supply.

3. We show that to overcome the deterioration of the profits discussed above, and to reap the benefit that one expects from operational efficiency, the moderating firm can allow for communication among the agents, even if done through a non-binding mechanism. The main contribution of this result is in showing that the operational efficiency needs to be complemented with the ability to communicate in order to obtain desirable outcomes for the involved parties. These desirable outcomes are only achievable in a marketplace where demand exceeds supply. Therefore, the contribution is also in highlighting the fact that it is crucial to understand the specific market conditions in terms of the ratio between demand and supply.

4. On the theoretical front, we introduce a new solution concept to study the communication enabled market. This new solution concept captures the fact that agents are allowed to communicate prior to the game in order to achieve a non-binding agreement regarding their actions.

2. Literature Review

The previous work related with our paper can be divided into two categories. The first category consists of research that studies the applications of queueing theory in service systems. The second one consists of research focused on developing approximations to analyze complex service systems.

Service systems with customers, who are both price and time sensitive, have attracted the attention of researchers for many years. The analysis of such systems dates back to Naor's seminal work

(See Naor, 1969), which analyzes customer behavior in a single-server queueing system. Motivated by his work, many researchers study the service systems facing price- and delay-sensitive customers in various settings. We refer the reader to Hassin and Haviv (2003) for an extensive summary of the early attempts in this line of research. More recently, Cachon and Harker (2002) and Allon and Federgruen (2007) studies the competition between multiple firms offering substitute but differentiated services by modeling the customer behavior implicitly via an exogenously given demand function. An alternative approach is followed in Chen and Wan (2003), where authors examine the customers' choice problem explicitly by embedding it into the firms' pricing problem. Other notable examples focusing on the customers' demand decision in competition models are Ha et al. (2003), and Cachon and Zhang (2007).

The pricing and the capacity planning problem of the service systems can easily become analytically intractable when trying to study more complex models, such as a multi-server queueing systems. Recognizing this difficulty, many researchers seek robust and accurate approximations to analyze multi-server queues. Halfin and Whitt (1981) is the first paper that proposes and analyzes a multi-server framework. This framework is aimed at developing approximations, which are asymptotically correct, for multi-server systems. It has been applied by many researchers to study the pricing and service design problem of a monopoly in more realistic and detailed settings. Armony and Maglaras (2004), and Maglaras and Zeevi (2005) are examples of recent work using the asymptotic analysis to tackle complexity of these problems. Furthermore, Garnett et al. (2002), Ward and Glynn (2003), and Zeltyn and Mandelbaum (2005) extends the asymptotic analysis of markovian queueing system by considering customer abandonments.

The idea of using approximation methods can also be applied to characterize the equilibrium behavior of the firms in a competitive environment. To our knowledge, Allon and Gurvich (2010) is the first paper studying competition among complex queueing systems by using asymptotic analysis to approximate the queueing dynamics. Another recent paper studying the equilibrium characterization of a competitive marketplace using asymptotic analysis is Chen et al. (2008). They consider a marketplace with multiple suppliers competing with each other over their prices and target lead times. There are two main differences between these two papers and our work. First, both of them study a service environment with a fixed number of decision makers (firms) while the number of decision makers in our marketplace (agents) is growing. Second, they only consider a competitive environment where the firms behave individually. In contrast, we study the non-cooperative case as well as the case where the agents have a limited level of collaboration.

In the field of operations management (OM), the majority of the papers employing game-theoretic foundations study non-cooperative settings. For an excellent survey, we refer to Cachon and Netessine (2004). There is also a growing literature that studies the OM problems in the context of cooperative game theory. Nagarajan and Sosic (2008) provide an extensive summary of the applications of cooperative game theory in supply chain management. Notable examples are the formation of coalitions among retailers to share their inventories, suppliers, and marketing powers (See Granot and Sosic (2005), Sosic (2006), and Nagarajan and Sosic (2007)). This body of research is related with our work, where we look for the limited collaboration among agents.

Our work may also be viewed as related to the literature on labor markets that studies the wage dynamics (See Burdett and Mortensen (1998), Manning (2003), Manning (2004), and Michaelides (2010)). In both our model and labor economics literature, people or firms with service needs seek an employee or an agent to perform the job they requested. In our model, service seekers trade-off time they need to wait until their job starts and cost, the phenomenon generally disregarded in labor economics literature. Further, our focus is on a market for temporary help, which means that the engagement between sides ends upon the service completion. This stands in contrast to the labor economics literature in which the engagement is assumed to be permanent. It is also important to note the difference between interventions studied in our model and the ones in the labor economics literature. Unlike the interventions we studied, which focus on improving operational efficiency, the interventions discussed in labor economics are usually aimed at regulating wages directly. Our paper also differs from the literature on market microstructure. This body of literature studies market makers who can set prices and hold inventories of assets in order to stabilize markets (See Garman (1976), Amihud and Mendelson (1980), Ho and Stoll (1983), and a comprehensive survey by Biais et al. (2005)). However, the moderating firm considered in our paper has no direct price-setting power and cannot respond to customers' service requests. Furthermore, papers studying market microstructure disregard the operational details such as waiting and idleness.

3. Model Formulation

Consider a service marketplace where agents and customers make their decisions in order to maximize their individual utilities. Customers' need for the service is generated according to a Poisson process with rate Λ. This forms the "potential demand" for the marketplace. A customer decides whether to join the marketplace or not: If she decides not to join the system, her utility is zero. If she joins the system, she decides who would process her job. The customers who join the marketplace form the "effective demand" for the marketplace. The exact nature of this decision depends on the specific structure of the marketplace, decided upfront by the moderating firm. We shall elaborate on the choices of customers in Sections 4-6. We assume that the service time required to satisfy the requests of a given customer is exponentially distributed with rate μ. Without loss of generality, we let $\mu = 1$. When the service of a customer is successfully completed, she pays the price of the service, earns a reward of R, and incurs a waiting cost of c per unit time until her service commences.[2] As the customers visiting the marketplace seek temporary help, a customer joining the system may become impatient while waiting for her service to start and abandon. In this case, the abandoning customer does not pay any price or earn any reward, but she does incur a waiting cost for the time she spends in the system. We assume that customers' abandonment times are independent of all other stochastic components and are exponentially distributed with mean m_a. Customers decide whether to request service or not and by whom to be served according to their expected utility. The expected utility of a customer is based on the reward, the price and the anticipated waiting time.

The above summarizes the demand arriving to the marketplace. Next, we discuss the service provision in a marketplace with k ex-ante identical agents.[3] The only decision of an agent is to choose a price for his service; each agent makes this decision independently. Let $(p_1, \ldots, p_k)$ denote the resulting price vector with p_n being the price chosen by the n^{th} agent. We normalized the operating cost of the agents to zero for notational convenience. The expected revenue of an agent depends on the price he chooses and his demand volume.

We refer to the ratio $\Lambda/(\mu k)$ as the demand-supply ratio of the system and denote it by ρ. The demand-supply ratio is a first order measure for the mismatch between aggregate demand and the total processing capacity. Marketplaces vary with respect to their demand-supply ratio, ρ, and, as we shall discuss, ρ has a significant impact on the market outcome. We broadly categorize marketplaces into two: Buyer's market where $\rho \leq 1$, and seller's market where $\rho > 1$.

4. No-intervention Model

The essential role of the moderating firm in a large scale marketplace is to set up the infrastructure for the interaction between players. This is crucial because all players have to be equipped with the necessary information, such as prices to make their decisions, yet individual players cannot gather this information on their own. When the moderating firm provides only the required information, it has no impact on the strategic interaction taking place in the marketplace. We thus refer to

[2] Our model can also be used to study a setting where customers incurs waiting cost also during their service. One can incorporate that by modifying the customer reward from R to $R - c/\mu$.

[3] We will discuss a model with heterogenous agents in Section 7.

such a setting as the no-intervention model. We analyze the dynamics of a large-scale marketplace in the no-intervention model not only to derive insights about the behavior of the self-interested and competing players in such a system, but also to build a benchmark for the cases in which the moderating firm introduces additional features which change the nature of the marketplace. Therefore, in this section, we study the behavior of a marketplace where the moderating firm confines itself to aggregating and providing information.

We model the strategic interaction between the agents and the customers as a sequential move game. Given the setup of Section 3, along with the above mentioned role of the moderating firm, the agents first announce their prices. Each arriving customer observes these prices and decides whether to request service or not. Further, if a customer decides to join the system, she also chooses the agent who processes her service request. The service of a customer starts immediately if the agent she chooses is available. Otherwise, she joins the queue in front of the agent and waits for her service to commence. We denote the fraction of customers choosing agent-n by D_n. Then, ΛD_n is the demand volume for agent-n.

More specifically, each agent's operations can be modeled as an $M/M/1+M$ queueing system[4] where the arrival rate of customers depends on the strategies of customers and agents[5]. If the rate of customers who request service from an agent charging price p is λ, the utility of a customer requesting service from this agent is $U(\lambda,p) = (R-p)[1-\beta(\lambda)] - W(\lambda)c$, where $\beta(\lambda)$, which will be referred to as the abandonment function, is the probability of abandonment, and $W(\lambda)$ is the expected waiting time, in an $M/M/1+M$ system with arrival rate λ, service rate 1, and abandonment rate $1/m_a$. Using queueing theory, the utility of customers can be rewritten as $U(\lambda,p) = (R-p+cm_a)[1-\beta(\lambda)] - cm_a$. Similarly, the revenue of that agent is $V(\lambda,p) = p\lambda[1-\beta(\lambda)]$. It is important to note that $V(\lambda,p)$ is the revenue rate of an agent, but throughout the paper we will refer to it as the revenue for ease of exposition.

As we consider a sequential move game, we are interested in the Subgame Perfect Nash Equilibrium (SPNE) of the game. We begin by characterizing the equilibrium in the second stage game where customers make their service requests given the agents' pricing decisions. Then, based on the second stage equilibrium, we derive the equilibrium of the first stage in which only agents make pricing decisions.

[4] $+M$ notation denotes the exponential abandonment times.

[5] Note that an agent can process more than one jobs at the same time in certain settings. In such settings, a processor sharing model will be a more appropriate queueing model, yet these models are known to be significantly more complex than our queueing model. Our model can be viewed as an approximation of such settings.

Fixing the agents' strategies $(p_n)_{n=1}^k$, an arriving customer observes the agents' prices and chooses the agent who maximizes her utility, anticipating the behavior of all other customers. Therefore, in equilibrium a customer chooses an agent only if the utility she obtains from him (weakly) dominates her utility from any other agent. This is also known as "Nash Flow Equilibrium" (See Roughgarden, 2005) in the congestion games literature. We formally define the Customer Equilibrium as follows:

DEFINITION 1 **(Customers Equilibrium)**. *Given $(p_n)_{n=1}^k$, we say that $(D_n)_{n=1}^k$ is a Customers Equilibrium if the following conditions are satisfied:*

1. *For any n with $D_n > 0$, we have that $U(\Lambda D_n, p_n) \geq U(\Lambda D_m, p_m) \geq 0$, for all $m \leq k$.*
2. *If $U(\Lambda D_n, p_n) > 0$ for some $n \leq k$, then $\sum_{n=1}^{k} D_n = 1$.*

The first condition of the Customer Equilibrium requires that customers request service from an agent in equilibrium only if that agent is one of their best alternatives. Moreover, the second condition ensures that all customers join the system if it is possible to earn strictly positive utility by requesting service from an agent. Customer Equilibrium exists by the continuity of the utility functions and Rath (1992). In the following proposition, we show that for any given price vector, the second stage game has a unique equilibrium.

PROPOSITION 1. *Given a price vector $(p_n)_{n=1}^k$, there is a unique Customer Equilibrium.*

Since the Customer Equilibrium is unique for any given price vector, we denote the fraction of customers requesting service from agent-n in equilibrium by $D_n^{CE}(p_1, \ldots, p_k)$ when $(p_1, \ldots, p_k)$ are the prices announced by agents. $D_n^{CE}(p_1, \ldots, p_k)$ is well defined in the light of Proposition 1.

We can now move to the first stage game which is played only among the agents. An equilibrium in this stage requires that none of the agents can improve his revenues by deviating unilaterally while taking the customers' response into account. We formalize this in the following definition:

DEFINITION 2 **(Subgame Perfect Nash Equilibrium)**. *Let $(D_n, p_n)_{n=1}^k$ summarize the strategy of all players in the market for all $n = 1, \ldots, k$. Then, $(D_n, p_n)_{n=1}^k$ is a SPNE if the following conditions are satisfied:*

1. *$D_n = D_n^{CE}(p_1, \ldots, p_k)$ for all $n \leq k$.*
2. *For any $\ell \leq k$, we have $V(\Lambda D_\ell, p_\ell) = \max_{p'} V(\Lambda D_\ell^{CE}(p_1, \ldots, p_{\ell-1}, p', p_{\ell+1}, \ldots, p_k), p')$.*

The first condition requires that $(D_n)_{n=1}^k$ arises in equilibrium in the second stage game. The second condition states that none of the agents has incentive to change his price. Note that agents take into account the impact price changes have on the Customer Equilibrium, and thus on demand.

4.1. Characterization of SPNE

In this section, we our restrict attention to symmetric SPNE where all agents charge the same price p in the first stage. This is a natural choice since all agents are identical. We will discuss non-symmetric equilibria in Section 7.1.

A price p emerges in equilibrium in the first stage if a single agent chooses to charge p to maximize his revenues given that all other agents announce p. When all other $k-1$ agents announce p, a generic agent, say agent-ℓ, solves the following maximization problem to determine his best-response:

$$\max_{p_\ell \geq 0} \; p_\ell \Lambda D_\ell^{CE}(p,\ldots,p,p_\ell,p,\ldots,p)\left[1-\beta(\Lambda D_\ell^{CE}(p,\ldots,p,p_\ell,p,\ldots,p))\right] \tag{1}$$

In this problem, the objective function is the revenue of agent-ℓ when he charges p_ℓ and the remaining agents charge p. Thus, p is a symmetric equilibrium in the first stage game if it is a solution to the above problem. We denote the symmetric SPNE by (D^*, p^*) where all agents charge p^* and each agent has a demand of ΛD^*, i.e. $D_n^{CE}(p,\ldots,p) = D^*$ for any $n \leq k$. We characterize the symmetric SPNE in the following theorem:

THEOREM 1. *If $\beta(\lambda)$ is concave, then there exists a symmetric SPNE. Furthermore, the symmetric SPNE is characterized as follows:*

1. *If $\Lambda \geq k\lambda^0$, then the symmetric SPNE is $(D^*,p^*) = \left(\frac{\min\{\lambda^{mon},\rho\}}{\Lambda}, R+cm_a - \frac{cm_a}{1-\beta(\min\{\lambda^{mon},\rho\})}\right)$.*
2. *If $\Lambda \leq k\lambda^0$, then the symmetric SPNE is $(D^*,p^*) = \left(1/k, (R+cm_a) - \frac{(R+cm_a)(k-1)}{\frac{k}{1-\nu(\rho)}-1}\right)$.*

Here λ^{mon} is the unique solution to $1-\beta(\lambda)-\lambda\beta'(\lambda) = \frac{cm_a}{R+cm_a}$, λ^0 is the unique solution to $(R+cm_a)(k-1) - \frac{cm_a}{1-\beta(\lambda)}\left(\frac{k}{1-\nu(\lambda)}-1\right) = 0$, and $\nu(\lambda) = \frac{\lambda\beta'(\lambda)}{1-\beta(\lambda)}$.

Similar to Theorems 1-3 in Chen and Wan (2003), the above result suggests that agents behave as local monopolists and charge their monopoly prices when the arrival rate is sufficiently high. Moreover, in this case, agents may choose not to cover the market completely. However, once the arrival rate becomes less than λ^0, the equilibrium price will be pushed down as the agents are engaged in a cut-throat competition, where intensity of competition can be quantified by the strictly positive utility left for customers in the equilibrium. It is also worth noting that utility of customers in the equilibrium increases as the arrival rate decreases.

REMARK 1. Concavity of the abandonment function, $\beta(\lambda)$, is a sufficient condition for the existence of symmetric equilibrium. In Lemma 1 in Appendix A, we show that $\beta(\lambda)$ is concave when $m_a \leq 1$, i.e. abandonment rate is higher than service rate. Furthermore, conducting a numerical study, we observe that $\beta(\lambda)$ is concave even for $1 \leq m_a \leq 2$. However, for higher values of m_a, the function $\beta(\lambda)$ is not concave in λ. This is not surprising given the complicated structure of queueing systems with impatient customers. For instance, Armony et al. (2009) shows the difficulty of proving the convexity of the expected head-count in the steady state of a system with customer abandonments. Even though $\beta(\lambda)$ is not concave, there can be a symmetric SPNE, and the above theorem characterizes this symmetric equilibrium. Numerically, we see that the equilibrium candidate characterized above still emerges as the symmetric SPNE when $\beta(\lambda)$ is not concave. In this numerical study, we consider a marketplace where $R = 1$, $c \in \{0.05, 0.06, \ldots, 0.2\}$, and $k = 50$. Then, we study five scenarios that differ in the average abandonment time m_a and lead to non-concave $\beta(\lambda)$. We assume $m_a \in \{5, 6, \ldots, 10\}$. For each of these scenarios, we show that the price proposed as equilibrium price in Theorem 1 is equilibrium by varying the arrival rate Λ on a grid from 10 to 50 with a step size of 1.

5. Operational Efficiency Model

In the previous section, we characterized the market outcome in the absence of any intervention on the part of the moderating firm. We now turn to discuss the impact of different mechanisms used by the moderating firm. As we discussed in the introduction, the moderating firm may provide a mechanism that improves the operational efficiency of the whole system by efficiently matching customers and agents. This mechanism aims at reducing inefficiency due to the possibility of having a customer waiting in line for a busy agent while an agent who can serve her is idle. This efficiency improvement is equivalent to virtually grouping the agents charging the same price. For instance, oDesk.com achieves this goal by allowing customers to post their needs and allowing service providers to apply to these postings. When a customer posts a job at oDesk.com, agents that are willing to serve this customer apply to the posting. Among the applicants charging less than what the customer wants to pay, the customer will favor agents based on their immediate availability. The main driver of the operational efficiency in this setting is the fact that customers no longer need to specify an agent upon their arrival because the job posting mechanism allows customers to postpone their service request decisions until they have enough information about the availability of the providers.

In this section, we modify the service marketplace considered in Section 4 by assuming that the mechanism introduced by the moderating firm ensures that customers do not stay in line when there is an idle agent willing to serve them by charging the price they want to pay or less. This can be modeled as a queuing network where the agents announcing the same price are virtually grouped together. Once each agent announces a price per customer to be served, we can construct a resulting price vector $(p_n)_{n=1}^N$ where $N \leq k$ is the number of different prices announced by the agents. We refer to the agents announcing the price p_n as sub-pool-n and denote the number of agents in the sub-pool-n by y_n. Hence, $(p_n, y_n)_{n=1}^N$ summarizes the strategy of all agents.

Under this mechanism, we model the customer decision making and experience as follows: If there are different prices announced by the agents, i.e., $N > 1$, the customer chooses a sub-pool from which she requests the service. We refer to the price charged by this sub-pool as the "preferred price". Each customer who decides to join the system enters the service immediately if there is an available agent either in the sub-pool she chooses or in any sub-pool announcing a price less than her preferred price. Moreover, the customer is served by the sub-pool offering the lowest price among all available sub-pools. Otherwise, she waits in a queue until an agent, who charges a price less than or equal to her preferred price, becomes available. We denote the fraction of customers requesting service from sub-pool-n by D_n. In this model of customer experience, there are two crucial features: 1) The service of an arriving customer commences immediately when there are available agents charging less than or equal to her preferred price, 2) If they have to wait, customers no longer wait for a specific agent rather for an available agent.

As we model the marketplace as a queuing network, the operations of each sub-pool depend on the operations of the other sub-pools. For instance, each sub-pool may handle customers from the other sub-pools (giving priority to its "own" customers) while some of the other sub-pools are serving its customers. Therefore, given the strategies of agents, $(p_n, y_n)_{n=1}^N$, and the service decisions of customers, $(D_n)_{n=1}^N$, the expected utility of a customer choosing the sub-pool-ℓ depends on all of these decisions, and can be written as:

$$U_\ell(D_1,\dots,D_N;p_1,\dots,p_N;y_1,\dots,y_N) = PServ_{\ell\ell}\big[(R-p_\ell+cm_a)(1-\beta_\ell)-cm_a\big]+\sum_{m\neq\ell} PServ_{\ell m}(R-p_m),$$

where $\beta_\ell(D_1,\dots,D_N;p_1,\dots,p_N;y_1,\dots,y_N)$ denotes the probability of abandonment in the sub-pool-ℓ, and $PServ_{\ell m}(D_1,\dots,D_N;p_1,\dots,p_N;y_1,\dots,y_N)$ denotes the probability that a customer choosing the sub-pool-ℓ is served by the sub-pool-m when ΛD_n is the rate of customer arrival to the sub-pool-n for $n=1,\dots,N$. We want to note that for any sub-pool-ℓ, $PServ_{\ell m}=0$ for any m such that $p_m > p_\ell$ since customer choosing sub-pool-ℓ cannot be served by a sub-pool charging more than p_ℓ. Furthermore, the revenue of an agent in sub-pool-ℓ is: $V_\ell(D_1,\dots,D_N;p_1,\dots,p_N;y_1,\dots,y_N) = p_\ell\sigma_\ell(D_1,\dots,D_N;p_1,\dots,p_N;y_1,\dots,y_N)$, where $\sigma_\ell(\dots;\dots;\dots)$ is utilization of agents in sub-pool-ℓ when ΛD_n is the rate of customer arrival to the sub-pool-n for $n=1,\dots,N$. Here, we assume that a customer choosing the sub-pool-ℓ pays p_m when she is served by sub-pool-m for $m\neq\ell$.

It is also worth noting that a marketplace operates as an $M/M/k+M$ system when all agents charge the same price. This allows us to employ the well-known limiting behavior of the multi-server systems to characterize the market outcome. Furthermore, in the case, where the agents announce different prices, we will show that the interdependency between the sub-pools announcing different prices diminishes as the market grows. In fact, large-scale marketplaces operate "almost like" the combination of independent multi-server systems.

The strategic interaction between the agents and the customers is modeled, as before, as a sequential move game. However, we use a slightly different second stage equilibrium than the one in Definition 1 since the customers decision and utility is changed by the new mechanism. The new customer equilibrium, which we refer to as Market Customer Equilibrium, uses the concept of Nash Flow Equilibrium with the requirement that customers only care for the prices announced by the sub-pools instead of individual prices.

Definition 3 **(Market Customers Equilibrium)**. *Given* $(p_n, y_n)_{n=1}^N$*, we say that* $(D_n)_{n=1}^N$ *is a Market Customers Equilibrium (MCE) if the following conditions are satisfied:*

1. *For any* ℓ *with* $D_\ell > 0$*, we have that* $U_\ell(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N) \geq U_m(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N)$*, for all* $m \leq N$.
2. *If* $U_\ell(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N) > 0$ *for some* $\ell \leq N$*, then* $\sum_{n=1}^N D_n = 1$.

While *MCE* always exists by the continuity of the utility functions and Rath (1992), its uniqueness cannot be guaranteed. For notational convenience, we shall assume that the best outcome from the customer perspective arises when there are multiple *MCE* (In fact, it can be shown that the limit of all *MCE*s is unique as the number of agents in the market grows). As the outcome is assumed to be unique, we denote the fraction of customers requesting service from sub-pool-n in a Market Customer Equilibrium by $D_n^{MCE}(p_1,\ldots,p_N;y_1,\ldots,y_N)$ when $(p_n,y_n)_{n=1}^N$ is a tuple of two vectors whose components are the prices and the number of agents announcing them.

Agents make pricing decisions in the first stage of the game. Unlike the no-intervention model, we need to account for two types of unilateral deviation of agents: an agent can either choose to deviate by joining an existing sub-pool or announce a new price. Therefore, an equilibrium in the first stage should be immune to any of these two deviations. One can show that, as the market grows, there exists a profitable unilateral deviation from any price in a buyer's market. In analyzing such markets, we would like to highlight the following two observations: 1) The arising system dynamic is too complex for exact analysis yet amenable to asymptotic analysis. 2) While a single agent, indeed, may have profitable deviations from every price in a buyer's market, the gains from deviations are small and diminish as the market grows. Thus, following Dixon (1987) and recently Allon and Gurvich (2010), we study a somewhat weaker notion of equilibrium, which allows us to characterize the market outcome (if one exists), as the market grows even when Nash equilibrium does not exist. To this end, we consider a sequence of marketplaces indexed by the number of agents, i.e., there are k agents in the k^{th} marketplace. The arrival rate in the k^{th} marketplace is assumed to be $\Lambda^k = \rho k$. This ensures that the demand-supply ratio is constant along the sequence of marketplaces. Then, in each market, we focus on an equilibrium concept, which requires immunity against only deviations that improve the revenue of an agent by at least $\epsilon \geq 0$ as formally stated in Definition 4 (See below). We refer to ϵ as the level of equilibrium approximation. We denote the level of equilibrium approximation in the k^{th} market by ϵ^k, and we assume that $\epsilon^k \to 0$ and $\epsilon^k \sqrt{k} \to \infty$ as $k \to \infty$. We study the behavior of the equilibrium along the sequence of marketplaces we described above in order to derive the equilibrium in a marketplace with large number of agents.

Definition 4 (**ϵ-Market Equilibrium**). *Let $(D_n^k, p_n^k, y_n^k)_{n=1}^N$ summarize the strategy of all players in the k^{th} market with $y_n^k > 0$ for all $n = 1, \dots, N$. Then, $(D_n^k, p_n^k, y_n^k)_{n=1}^N$ is an ϵ-Market Equilibrium if the following conditions are satisfied:*

1. *$D_n^k = D_n^{MCE}(p_1^k, \dots, p_N^k; y_1^k, \dots, y_N^k)$ for all $n \leq N$.*
2. *For any $\ell \leq N$ and $m \leq N$, we have that $V_\ell(D_1^k, \dots, D_N^k; p_1^k, \dots, p_N^k; y_1^k, \dots, y_N^k) \geq V_\ell(\hat{D}_1^k, \dots, \hat{D}_N^k; p_1^k, \dots, p_N^k; \hat{y}_1^k, \dots, \hat{y}_N^k) - \epsilon^k$, where $\hat{y}_n^k = y_n^k - 1$ if $n = \ell$, $\hat{y}_n^k = y_n^k + 1$ if $n = m$, $\hat{y}_n^k = y_n^k$ otherwise, and $\hat{D}_n^k = D_n^{MCE}(p_1^k, \dots, p_N^k; \hat{y}_1^k, \dots, \hat{y}_N^k)$ for all $n \leq N$.*
3. *For any $\ell \leq N$ and $p' \neq p_n^k$ for all $n = 1, \dots, N$, we have that $V_\ell(D_1^k, \dots, D_N^k; p_1^k, \dots, p_N^k; y_1^k, \dots, y_N^k) \geq V_{N+1}(\hat{D}_1^k, \dots, \hat{D}_{N+1}^k; p_1^k, \dots, p_N^k, p'; \hat{y}_1^k, \dots, \hat{y}_{N+1}^k) - \epsilon^k$ where $\hat{y}_n^k = y_n^k - 1$ if $n = \ell$, $\hat{y}_n^k = 1$ if $n = N+1$, $\hat{y}_n^k = y_n^k$ otherwise, and $\hat{D}_n^k = D_n^{MCE}(p_1^k, \dots, p_N^k, p'; \hat{y}_1^k, \dots, \hat{y}_{N+1}^k)$ for all $n \leq N+1$.*

The first condition in the above definition requires that the vector $(D_n^k)_{n=1}^N$ forms an equilibrium among the customers if the agents choose the strategy $(p_n^k, y_n^k)_{n=1}^N$. The second and third conditions characterize the equilibrium in the first stage game: The second condition states that an agent cannot improve his revenue by more than ϵ^k when he joins an existing sub-pool, while the third condition states that an agent cannot improve his revenue by more than ϵ^k when he introduces a new sub-pool. We next turn to characterize the equilibrium in the k^{th} marketplace. Note that if $\epsilon^k \equiv 0$ for all k, then the above definition reduces to that of the Nash Equilibrium.

5.1. Characterization of the Market Equilibrium

In this subsection, we study the symmetric equilibrium for the sequence of marketplaces we constructed above. As a first step towards characterizing the symmetric equilibrium, we derive the revenues of agents when they announce the same price in the k^{th} marketplace. As we noted before, such a marketplace operates as an $M/M/k+M$ system with arrival rate $\Lambda^k D_1^{MCE}(p^k; k)$, service rate 1, and abandonment rate $1/m_a$, where $D_1^{MCE}(p^k;k)$ is the Market Customer Equilibrium when all k agents charge p^k. Therefore, the revenue of an agent in this case is given by

$$V_1(D_1^{MCE}(p^k;k);p^k;k) = p\rho D_1^{MCE}(p^k;k)[1-\beta^M(\Lambda^k D_1^{MCE}(p^k;k);k)], \tag{2}$$

where $\beta^M(\lambda;k)$ is probability of abandonment in $M/M/k+M$ system with arrival rate λ, service rate 1, and abandonment rate $1/m_a$.

In order to characterize an ϵ^k-symmetric Market Equilibrium, we need to verify that a single agent does not have any incentive to deviate to a price other than p^k in the k^{th} marketplace. Recall that if an agent chooses $p' \neq p^k$, this amounts to creating his own sub-pool, and his revenue is given by $V_2(D_1^{MCE}(p^k,p';k-1,1), D_2^{MCE}(p^k,p';k-1,1);p^k,p';k-1,1)$, where $(D_n^{MCE}(p^k,p';k-1,1))_{n=1}^2$ is the Market Customer Equilibrium given that $k-1$ agents charge p^k and one agent charges p'. We then say that a price p^k emerges as the symmetric ϵ^k-Market Equilibrium if

$$V_1(D_1^{MCE}(p^k;k),p^k,k) \geq \max_{0\leq p'\leq R} V_2(D_1^{MCE}(p^k,p';k-1,1), D_2^{MCE}(p^k,p';k-1,1);p^k,p';k-1,1) - \epsilon^k, \tag{3}$$

where the left-hand side is the revenues of agents when all agents charge p^k, and the right-hand side is the maximum revenue that a single agent can obtain by deviating from p^k.

To understand the behavior of the market outcome in large markets, we shall first study the left-hand side of (3) along the trajectory of marketplaces in which all k agents charge p^k and $p^k \to p$ as $k\to\infty$. In a buyer's market, we show that all customers join the system in equilibrium as long as $p<R$ since they experience negligible waiting times and obtain approximately the utility of $R-p$ by joining in a marketplace with a large number of agents. Therefore, the revenue of each agent is approximated by $p\rho$ in a buyer's market when $p<R$. In a seller's market, some of the customers leave the market immediately due to the high congestion level even if $p<R$, but the rate of customers requesting service should, in equilibrium, be higher than the processing capacity when $p<R$. Therefore, agents are always "over-utilized" in a seller's market and the revenue of each agent is approximately p when $p<R$. When $p=R$, the rate of customers requesting service depends on the convergence rate of p^k both in a buyer's and a seller's market. Thus, $p\min\{\rho,1\}$ constitutes an upper bound for the revenue of each agent if $p=R$. The following proposition presents these results formally.

PROPOSITION 2. *Let $D_1^{MCE}(p^k;k)$ be the Market Customer Equilibrium when all agents charge p^k in the k^{th} marketplace such that $\lim_{k\to\infty} p^k = p$. When $=p<R$, we have that*

$$\lim_{k\to\infty} D_1^{MCE}(p^k;k) = \min\left\{1, \frac{R-p+cm_a}{\rho c m_a}\right\}.$$

Furthermore, $\lim_{k\to\infty} V_1(D_1^{MCE}(p^k;k);p^k;k) = \begin{cases} p\rho & \text{if } \rho \le 1 \\ p & \text{if } \rho > 1 \end{cases}$. *When* $p = R$, *we have that*

$$\limsup_{k\to\infty} D_1^{MCE}(p^k;k) \le \min\left\{1, \frac{1}{\rho}\right\}, \text{ and } \lim_{k\to\infty} V_1(D_1^{MCE}(p^k;k);p^k;k) \le \begin{cases} p\rho & \text{if } \rho \le 1 \\ p & \text{if } \rho > 1 \end{cases}.$$

After approximating the revenue of the agents when they charge the same price, we now focus on the maximum revenue that an agent can obtain by creating his own sub-pool. As we did above, we again distinguish between buyer's and seller's markets.

5.1.1. Buyer's Market: When all agents charge the same price p^k in a buyer's market, we next show that a single agent can improve his revenue when he decreases his price. Such a cut will allow a single agent to serve not only his own customers but also the customers choosing the price p^k. In fact, his revenue can be arbitrarily close to p^k following a small price cut as long as the rate of customers requesting service is bounded away from zero when all agents charge p^k, i.e., $\lim_{k\to\infty} D_1^{MCE}(p^k;k) > 0$. The following proposition proves this observation formally.

PROPOSITION 3. *Let* $V'(p^k;k) = \max_{0\le p' < p^k} V_2(D_1^{MCE}(p^k,p';k-1,1), D_2^{MCE}(p^k,p';k-1,1);p^k,p';k-1,1)$ *for any sequence of* p^k *such that* $\lim_{k\to\infty} p^k = p$. *Then, we have that* $\liminf_{k\to\infty} V_k'(p^k;k) > 0$ *when* $p > 0$. *Furthermore, when* $\lim_{k\to\infty} D_1^{MCE}(p^k;k) > 0$, *we have that* $\lim_{k\to\infty} V'(p^k;k) = p$.

As we established in Proposition 2, the revenue of an agent when all agents charge the same price p^k can be bounded from above by $p^k\rho$ in large marketplaces. Then, Proposition 3 implies that any p^k satisfying $\lim_{k\to\infty} p^k = p > \epsilon^k/(1-\rho)$ cannot emerge as the equilibrium price of a symmetric ϵ^k-Market Equilibrium for large k. Thus, as $\lim_{k\to\infty} \epsilon^k = 0$, we obtain that any sequence of prices except the ones converging to zero cannot be sustained as the equilibrium price of a symmetric ϵ^k-Market Equilibrium along the trajectory of marketplaces. Note that we do not need to analyze the revenue of an agent after a price increase because it is sufficient to demonstrate the existence of one profitable deviation in order to show that a given price cannot be an equilibrium outcome. We formalize these observations in the following theorem.

THEOREM 2. *In a buyer's market with* $\rho < 1$,

1. *Let* p_{EQ}^k *be a price emerging as the equilibrium price of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace. Then, for any* $\xi > 0$, *there exists a* K *such that* $p_{EQ}^k < \xi$ *for all* $k > K$.

2. *There exists a* K *such that zero is an equilibrium price of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for all* $k > K$.

3. *Let Π^k_{OE} and Π^k_{NI} be the total revenue generated in the k^{th} marketplace with and without operational efficiency, respectively. Then, for any $\xi > 0$, there exists a K such that $\frac{\Pi^k_{OE}}{\Pi^k_{NI}} < \xi$ for all $k > K$.*

The above theorem states that if a moderating firm provides efficient matching in a buyer's market, the equilibrium outcome of the marketplace will converge to zero. As the profit of the firm is the share of the revenue generated in the marketplace, providing efficient matching deteriorates the profit of the firm compared to the no-intervention case as well as the revenue of the agents. In fact, we show that the ratio between the total revenue generated in a marketplace under operational efficiency and under the no-intervention converges to zero. We also establish that zero can emerge as the equilibrium price in large marketplaces. In Section 7.2, we discuss the extension of the above theorem, which is based on showing that the revenues of agents converges to zero even in a non-symmetric equilibrium. For a formal treatment, see Proposition 8.

5.1.2. Seller's Market: After discussing the impact of providing efficient matching in a buyer's market, we now focus on a seller's market. Unlike in a buyer's market, a single agent cannot improve his revenue after a price cut since it does not improve his utilization significantly. Note that agents are already "over-utilized," and earning a revenue of p^k while they are charging the same price p^k in a seller's market. Therefore, in a seller's market, the only possible profitable deviation for a single agent is to increase his price in large enough marketplaces. In such a deviation, a single agent loses some of his customers because of his high price, and he also loses the benefits of efficient matching since he becomes an individual provider. Both of these factors will limit his ability to make higher profit. In fact, the following proposition establishes an upper bound on the asymptotic revenue which a single agent can generate by increasing his price.

PROPOSITION 4. *Let $V'(p^k;k) = \max_{p^k \le p' \le R} V_1(D_1^{MCE}(p', p^k; 1, k-1), D_2^{MCE}(p', p^k; 1, k-1); p', p^k; 1, k-1)$ for any given sequence of prices p^k such that $\lim_{k\to\infty} p^k = p$. When $p < R$ in a seller's market ($\rho > 1$), we have that $\limsup_{k\to\infty} V'(p^k;k) \le (R + cm_a)\lambda^{\Delta}(p;R)[1 - \beta(\lambda^{\Delta}(p;R))] - \lambda^{\Delta}(p;R)(\Delta(p;R) + cm_a)$, where $\Delta(p;R) = \max\left\{0, \frac{R-p+cm_a}{\rho} - cm_a\right\}$, and $\lambda^{\Delta}(p;R)$ is the unique solution to $1 - \beta(\lambda) - \lambda\beta'(\lambda) = \frac{\Delta(p;R)+cm_a}{R+cm_a}$.*

When a single provider increases his price, we show that the demand for agents, who do not change their prices, is almost the same as their original demand before deviation. Hence, the utility of customers choosing the sub-pool consisting of $k-1$ agents is $\Delta(p;R)$, which is the utility that the customers obtain in the Market Customer Equilibrium in a large marketplace when all agents charge p^k. These results hold due to the facts that sub-pools operate as if they are independent in a large marketplace when demand exceeds supply and the demand that a single agents can serve is negligible relative to the aggregate demand. Then, to approximate the maximum post-deviation revenue, one can treat the deviating agent as a monopoly whose customers have an outside option with the value of $\Delta(p;R)$. In fact, the above proposition shows that this approximation constitutes an upper bound on the agent's post-deviation revenue. A monopoly always makes sure that the utility of customers is exactly equal to their outside option, by setting the price to $R+cm_a-\frac{\Delta(p;R)+cm_a}{1-\beta(\lambda)}$ for any given target of demand rate λ. He then picks λ, maximizing his revenue and sets his price accordingly. We refer the reader to the proof of Proposition 4 for a more detailed discussion on the revenue maximization problem of a monopoly.

Combining the two observations above, it is clear that in a large marketplace, a price p^k emerges as the symmetric ϵ^k-Market Equilibrium outcome if p^k is greater than the profit of a monopoly serving customers with outside option $\Delta(p;R)$. We state this result in the following theorem.

THEOREM 3. *In a seller's market ($\rho > 1$), let*

$$p^* \in \mathcal{P}(\rho;R) \equiv \left\{p : p > (R+cm_a)\lambda^{\Delta}(p;R)[1-\beta(\lambda^{\Delta}(p;R))] - \lambda^{\Delta}(p;R)(\Delta(p;R)+cm_a),\ 0 \le p < R\right\},$$

where $\Delta(p;R)$, and $\lambda^{\Delta}(p;R)$ are defined as in Proposition 4. Then, for any given sequence of prices p^{*k} that converges to p^* as $k\to\infty$, there exists a K such that p^{*k} emerges as the equilibrium price of a symmetric ϵ^k-Market Equilibrium in the k^{th} marketplace for all $k > K$. Furthermore, for any $\rho_1 > \rho_2$, we have that $\mathcal{P}(\rho_1;R) \subseteq \mathcal{P}(\rho_2;R)$.

The above theorem characterizes the set of symmetric ϵ^k-Market Equilibria for large marketplaces. The theorem does not guarantee the uniqueness of such an equilibrium, i.e. $\mathcal{P}(\rho;R)$ may not be a singleton. In fact, $\mathcal{P}(\rho;R)$ may consist of uncountably many prices. Furthermore, we show that $\mathcal{P}(\rho;R)$ shrinks as ρ increases. As the demand-supply ratio increases, customers experience significant waiting times even if they are served by a price-generated pool. Therefore, the level of customer surplus that a deviating agent has to forego declines as ρ rises. As a result of this, a single agent has more room to deviate and improve his revenue when demand is high. It is also worth highlighting that a single agent has such a profitable deviation opportunity even though the number of agents grows to infinity.

Characterizing the set of symmetric equilibria, $\mathcal{P}(\rho;R)$, is difficult in general. For illustrative purposes, we consider the case where the abandonment rate is equal to the service rate. We show that a similar structure holds for the settings when $\mu \neq m_a$ using a numerical study in Section B.7 in the supporting document. The next corollary characterizes the correspondence $\mathcal{P}(\rho;R)$ as well as the asymptotic behavior of the unique equilibrium price under the no-intervention model.

Corollary 1. *Suppose the abandonment rate is equal to the service rate. Then, we have that*

1. $\lambda^{\Delta}(p;R) = \log\left(\frac{R+c}{\Delta(p;R)+c}\right)$ *where* $\Delta(p;R)$ *is defined as in Proposition 4. Furthermore, the correspondence* $\mathcal{P}(\rho;R)$ *defined in Theorem 3 can be expressed as*

$$\mathcal{P}(\rho;R) = \left\{ p : p > \left[R + c - \left(1 + \log\left(\frac{R+c}{\Delta(p;R)+c}\right)\right)[\Delta(p;R)+c]\right], 0 \leq p < R, \right\}.$$

2. $\lim_{k\to\infty} p^k_{NI} = p_{NI} \equiv (R+c)\min\left\{1 - \frac{\rho}{e^{\rho}-1}, 1 - \frac{c}{R}\log\left(\frac{R+c}{c}\right)\right\}$, *where* p^k_{NI} *is the unique equilibrium price under no-intervention setting in the* k^{th} *marketplace.*

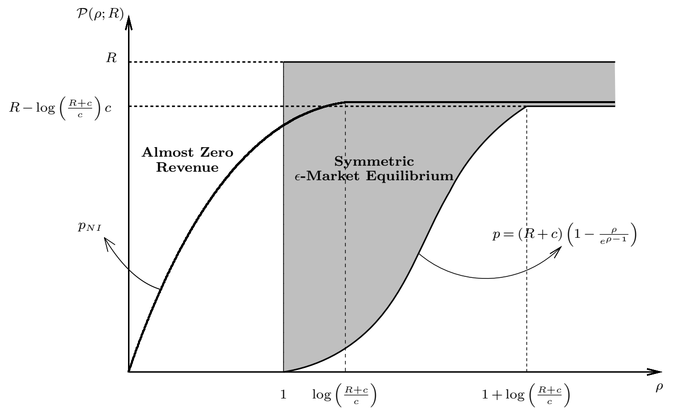


Figure 1 The prices that form a symmetric market equilibrium as a function of the demand-supply mismatch (ρ). The service rates and abandonment rates are assumed to be one.

Figure 1 displays the correspondence $\mathcal{P}(\rho;R)$ and the limit p_{NI}. More specifically, the gray area represents the prices that can emerge as the equilibrium price of a symmetric equilibrium in a large marketplace and the bold curve depicts p_{NI}. We observe that for all $\rho > 1$, the set $\mathcal{P}(\rho;R)$ is not a singleton. In fact, we have a wide range of prices that can form an equilibrium. Furthermore, many of the possible equilibrium prices in $\mathcal{P}(\rho;R)$ are lower than p_{NI}. The intuition behind this result is the following: In a marketplace where the moderating firm efficiently matches customers and agents, a single agent, who deviates by increasing his price, loses benefits of efficient matching, and thus cannot sustain the same quality of service (in terms of waiting times) as his "original" pool. It turns out that the deviating agent cannot improve his "original" revenue by decreasing his price either. Thus, in a seller's market, the price-generated pool serves as a deterrent against single agent deviations even if prices are unappealing from a system point of view. It is also important to note that such lower prices lead to loss in total revenue for the marketplace compared to the no-intervention setting. While one may expect operational efficiency tools to be a leverage for higher revenues in the market, it is surprising to see that reducing the unnecessary waiting and idleness present in a system with no-intervention may deteriorate the revenues.

Myerson (1991) argues that the question of which equilibrium would emerge as the outcome of a game with multiple equilibria can be answered with the focal-point effect phenomenon [6]. Our goal in this paper is not to conclude that only the low prices can be a focal-point. In fact, wphen comparing the equilibrium prices in a market with and without operational efficiency, one should also observe that operational efficiency does not only serve as a deterrent for deviations from low prices but also prevents deviations from high prices for any level of demand-supply ratio. Moreover, when the aggregate demand is sufficiently high, efficient matching always leads to higher profits, although the equilibrium prices under operational efficiency may be slightly lower than the unique equilibrium in a market without operational efficiency.

Our analysis in this section assumes that a customer with a preferred price p_ℓ pays p_m when she is served by sub-pool-m for any $m \neq \ell$. In real marketplaces such as oDesk.com, customers may end up paying a price between their preferred price and the prices asked by the providers when these prices are different. To account for that, one may envision an extension of our model, in which a customer choosing sub-pool-ℓ pays $\phi p_\ell + (1-\phi)p_m$, where $\phi \in (0,1)$, when an agent from sub-pool-m, with $m \neq \ell$ serves her. In such a model, our key findings, which are equilibrium prices are close to zero in a buyer's market, and some of the equilibrium outcomes may lead to profit loss in a seller's market, would continue to hold.

5.2. Supplying Real-Time Congestion Information

In Section 5.1, we have studied the impact of a particular efficient matching mechanism which aims at reducing the mismatch between customers and service providers. Considering the level of technology that online marketplaces have, another way of achieving operational efficiency might be to provide the real-time congestion information of each agent. This way, the firm would again reduce the inefficiency due to unnecessary waits and idleness. Here, we discuss the impact of this strategy on the market outcome.

Buyer's Market: Recall that in Proposition 3, we show that a single agent can improve his revenue significantly by decreasing his price when the firm provides the efficient matching mechanism described in the beginning of Section 5. The key driver of this result is the fact that demand for the agent, who decreases his price, increases drastically. When the firm provides real-time congestion information, a single agent has the same opportunity to improve his revenue after a price cut; customers will always choose him whenever he is idle, and this leads to a significant demand increase for him. Then, similar to Theorem 2, one can show that only the prices very close to zero emerge as the equilibrium outcome in this setting. Hence, the impact of providing real-time congestion data in a buyer's market would be the same as the aforementioned matching mechanism.

[6] Focal-point effects are any psychological or cultural norms that tends to focus players' attention on one equilibrium.

Seller's Market: As the aggregate demand exceeds the total processing capacity in a seller's market, the most profitable deviation, if any, for a single agent may be to increase his price when the firm provides real-time congestion data. Proposition 4 shows that this is true under the efficient matching mechanism described before. Unfortunately, when real-time congestion information is available, the system dynamics of a marketplace arising after one agent increases his price is quite complex, and thus it is analytically intractable. We perform a simulation study to better understand whether a single agent can improve his revenue by changing his price when the moderating firm provides real-time queue information.

In our simulation study, we consider a marketplace where each customer obtains a reward of $R = 1$, incurs a waiting cost of $c \in \{0.01, 0.02, \ldots, 0.05\}$ per unit time, and abandon the system with rate of $1/m_a \in \{1/2, 1, 2\}$. We fix the number of agents k to be 25, the service rate μ to be 1, and assume the arrival rate Λ is either 40, 50, or 60. For each of these different scenarios, we check whether a single agent has an incentive to increase his price to $p' \in \{0.01, 0.02, 0.03, \ldots, 0.99\} \setminus p$ when all the remaining $k-1$ agents still charge a price $p \in \{0.1, 0.2, \ldots, 0.9\}$. We say p emerges as the equilibrium price of a symmetric ϵ-Market Equilibrium, with $\epsilon = 0.01$, if a single agent cannot improve his revenue by more than 0.01 when he raises his price. To this end, for any given instance (Λ, p, p'), we generate 10000 random customer arrivals. Upon each arrival, the customer observes the number of customers waiting for each agent, and chooses the one which provides the highest expected utility. Letting the number of customers waiting for agent-n by Q_n, the expected utility of the customer is $R - p_n - c\mathbf{E}[W(Q_n)]$, where $\mathbf{E}[W(Q)]$ is expected waiting time of an arriving customer in an $M/M/1+M$ system with service rate 1 and abandonment rate $1/m_a$ given that there are Q customers in the queue. Whitt (1999) shows that $\mathbf{E}[W(Q)] = \sum_{j=0}^{Q} \frac{1}{1+j/m_a}$.

We estimate the utilization of agents by simulating four runs (The relative error in all cases was less than 1%). Then, the revenue of the agent is his average utilization multiplied by the price he charges.

Table 1 summarizes the results of our simulation study when $c = 0.03$ and $m_a = 1$. As in our efficient matching model, there is a wide range of prices that can emerge as an equilibrium outcome for a given arrival rate while the range shrinks as the arrival rate increases when the moderating firm provides real-time congestion information. In other words, this simulation study provides strong evidence for the fact that our particular operational efficiency model provides very similar key insights as providing real-time congestion information does. In Section 6.3, we will further analyze the model in which the firm provides real-time congestion information.

	$\Lambda = 40$	$\Lambda = 50$	$\Lambda = 60$
$p = 0.1$	×	×	×
$p = 0.2$	×	×	×
$p = 0.3$	×	×	×
$p = 0.4$	✓	×	×
$p = 0.5$	✓	✓	×
$p = 0.6$	✓	✓	✓
$p = 0.7$	✓	✓	✓
$p = 0.8$	✓	✓	✓
$p = 0.9$	✓	✓	✓

✓: p is an equilibrium price,
×: p is not an equilibrium price.

(a)

	$\Lambda = 40$	$\Lambda = 50$	$\Lambda = 60$
$p = 0.1$	N/A	N/A	N/A
$p = 0.2$	N/A	N/A	N/A
$p = 0.3$	N/A	N/A	N/A
$p = 0.4$	0.396	N/A	N/A
$p = 0.5$	0.495	0.499	N/A
$p = 0.6$	0.593	0.598	0.600
$p = 0.7$	0.694	0.697	0.700
$p = 0.8$	0.790	0.798	0.799
$p = 0.9$	0.892	0.897	0.899

N/A: Not applicable.

(b)

Table 1 **(a) Prices that can emerge as the equilibrium outcome and (b) the equilibrium revenues of agents when the moderating firm provides real-time congestion information.**

6. Communication Enabled Model

In this section, we continue to study the impact of different mechanisms used by the moderating firm. As we mentioned in the introduction, the moderating firm may complement its operational tool discussed in the previous section with a strategic tool, which changes the nature of the interaction among agents. In a marketplace such as oDesk.com, service providers are offered discussion boards in which they are allowed to exchange information. Moreover, the market supports the creation of affiliation groups, which are self-enforcing entities. We will thus focus on the impact of enabling communication among agents on the market outcome.

The economics literature suggests that, when the players have the opportunity to perform non-binding pre-play communication among themselves, the stability of an outcome can be threatened by potential deviations formed by coalitions, even in noncooperative games. Following this idea, the well-know notion of Strong Nash Equilibrium (*SNE*) requires stability against deviations formed by any conceivable coalitions (See Aumann (1959)). The main drawback of *SNE* is that many of the games do not have any *SNE*.

In this section, we modify the marketplace we study in the previous section by assuming that agents have opportunities to make non-binding communication prior to making their decisions, so that they can try to self-coordinate their actions in a mutually beneficial way, despite the fact that each agent selfishly maximizes his own utility.

Echoing the ideas in the economics literature, allowing communication among agents changes the equilibrium concept we use to characterize the outcome in the marketplace. We model this by proposing a new equilibrium concept that allows several agents to deviate together. More specifically, the new concept requires that a strategy of agents should be immune to any coalitions. Since a marketplace tends to be large, e.g., there are hundreds of thousands of agents in oDesk.com, one has to restrict the possible size of a coalition. We denote the largest fraction of agents that is allowed to deviate together by $\delta \in (1/k, 1]$. As in Section 5, we focus on the deviations that improve the revenues of agents at least by $\epsilon \geq 0$. Furthermore, we again study the behavior of the equilibrium along the sequence of marketplaces we described in Section 5. Recall that there are k agents, the arrival rate is $\Lambda^k = \rho k$, and the level of equilibrium approximation is ϵ^k, with the same asymptotic properties as in Section 5, in the k^{th} marketplace. We let δ^k be the largest fraction of agents that is allowed to deviate together in the k^{th} marketplace. We assume that $\delta^k k \to \infty$ as $k \to \infty$. This condition states that the number of agents allowed to deviate increases without bound as the market size increases. We refer to our new equilibrium concept as (δ, ϵ)-Market Equilibrium which is defined as follows:

Definition 5 (**(δ, ϵ)-Market Equilibrium**). *Let $(D_n^k, p_n^k, y_n^k)_{n=1}^N$ summarize the strategy of all players in the k^{th} market with $y_n^k > 0$ for all $n = 1, \ldots, N$. Then, $(D_n^k, p_n^k, y_n^k)_{n=1}^N$ is a (δ^k, ϵ^k)-Market Equilibrium if the following conditions are satisfied:*

1. *$D_n^k = D_n^{MCE}(p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k)$ for all $n \leq N$.*
2. *For any $\ell \leq N$, $m \leq N$, and $0 < d \leq \min\{y_\ell^k, \lfloor \delta^k k \rfloor\}$, we have that $V_\ell(D_1^k, \ldots, D_N^k; p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k) \geq V_\ell(\hat{D}_1^k, \ldots, \hat{D}_N^k; p_1^k, \ldots, p_N^k; \hat{y}_1^k, \ldots, \hat{y}_N^k) - \epsilon^k$, where $\hat{y}_n^k = y_n^k - d$ if $n = \ell$, $\hat{y}_n^k = y_n^k + d$ if $n = m$, $\hat{y}_n^k = y_n^k$ otherwise, and $\hat{D}_n^k = D_n^{MCE}(p_1^k, \ldots, p_N^k; \hat{y}_1^k, \ldots, \hat{y}_N^k)$ for all $n \leq N$.*
3. *For any $\ell \leq N$, $0 < d \leq \min\{y_\ell^k, \lfloor \delta^k k \rfloor\}$, and $p' \neq p_n$ for all $n = 1, \ldots, N$, we have that $V_\ell(D_1^k, \ldots, D_N^k; p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k) \geq V_{N+1}(\hat{D}_1^k, \ldots, \hat{D}_{N+1}^k; p_1^k, \ldots, p_N^k, p'; \hat{y}_1^k, \ldots, \hat{y}_{N+1}^k) - \epsilon^k$, where $\hat{y}_n^k = y_n^k - d$ if $n = \ell$, $\hat{y}_n^k = d$ if $n = N+1$, $\hat{y}_n^k = y_n^k$ otherwise, and $\hat{D}_n^k = D_n^{MCE}(p_1^k, \ldots, p_N^k, p'; \hat{y}_1^k, \ldots, \hat{y}_{N+1}^k)$ for all $n \leq N+1$.*

The above definition is closely related to the definition of ϵ-Market Equilibrium in Section 5. The key difference between these two equilibrium definitions is that (δ, ϵ)-Market Equilibrium allows a group of agents to deviate by either forming a new sub-pool or joining an existing one. In fact, our new equilibrium concept is a refinement of the ϵ-Market Equilibrium. Therefore, any (δ, ϵ)-Market Equilibrium is also a ϵ-Market Equilibrium. Employing the (δ, ϵ)-Market Equilibrium concept, we expect that the set of prices that can be sustained as a ϵ-Market Equilibrium will shrink since (δ, ϵ)-Market Equilibrium is more restrictive. Kalai (2004) and Gradwohl and Reingold (2008) study large games and shows that all Nash Equilibria of certain large games are resilient to deviations by coalitions. Such a phenomena does not exist in our model.[7]

[7] According to the definition in Gradwohl and Reingold (2008), a Nash Equilibrium is resilient to coalitions if players cannot improve their revenues "*too much*" even after a coordinated deviation. In our setting, "*too much*" has to be almost as much as the customer reward, R, in order to apply their results to our game. Clearly, this makes the definition of resilience vacuous because none of the agents can increase his revenue by more than R.

6.1. Characterization of the (δ, ϵ)-Market Equilibrium

Similar to Section 5, we focus on the symmetric (δ, ϵ)-*ME* where all agents charge the same price. The revenue of an agent when all agents charge the same price p^k is the same as in (2), and thus Proposition 2 establishes its asymptotic behavior.

In a buyer's market with $\rho < 1$, we showed that only the prices in a small neighborhood of zero can emerge as a symmetric ϵ-Market Equilibrium in large marketplaces. As a direct implication of the fact that (δ, ϵ)-Market Equilibrium is a refinement of the ϵ-Market Equilibrium, any sequence of prices that emerge as symmetric (δ, ϵ)-Market Equilibrium converges to zero as the market size grows. Furthermore, we show that $p = 0$ can emerge as the equilibrium price in large marketplaces.

Theorem 4. *Let p^k_{EQ} be a price emerging as a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace where $\rho < 1$. Then, for any $\xi > 0$, there exists a K such that $p^k_{EQ} < \xi$ for all $k > K$. Furthermore, when $\lim_{k\to\infty} \delta^k = 0$, there exists a K such that zero is an equilibrium price of a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace for all $k > K$.*

In a seller's market, Proposition 2 shows that the rate of customers requesting service will exceed the processing capacity of agents when all agents charge a price lower than R. Therefore, customers experience significant waiting times, and not only pay the price of the service but also incur a strictly positive waiting cost. Then, we show that a small group of agents can use the fact that customers pay an extra cost to increase their prices while ensuring that they are still "over-utilized" after the price increase. Since this small group of agents increases their prices without hurting their utilization, this deviation clearly improves their revenues (This is in contrast to the setting in Section 5 where the utilization of a single agent does drop after a price decrease). Thus, in a seller's market, only the prices, which are very close to R, can emerge as the equilibrium price of a symmetric (δ, ϵ)-Market Equilibrium in large marketplaces. To contrast this result with the result in Theorem 3, it is worth noting that a single agent has only a limited opportunity to improve his revenue by increasing his price as in most cases, the revenue improvement due to the price increase is overcome by the drop in utilization. Therefore, without the communication opportunity, it was possible to observe low prices as the market outcome even though demand exceeds supply.

Theorem 5. *Let p^k_{EQ} be a price emerging as a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace where $\rho > 1$. Then, for any $\xi > 0$, there exists a K such that $p^k_{EQ} > R - \xi$ and $D_1^{MCE}(p^k_{EQ}; k) > 1/\rho - \xi$ for all $k > K$. Furthermore, there exist a sequence p^{*k} and a K such that p^{*k} forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for all $k > K$.*

The above result shows that agents can sustain a price, which extracts all of the customer surplus, as the equilibrium outcome in a seller's market. Moreover, it also implies that the marketplace cannot be congested in the equilibrium even in a seller's market since any level of congestion can be capitalized by agents through a price increase.

Theorem 5 characterizes the unique limit of symmetric (δ,ϵ)-Market Equilibrium, but this result can be extended by showing that R is indeed the unique limit of all possible (δ,ϵ)-Market Equilibria as discussed in Section 7.3. For a formal treatment, see Proposition 9.

6.2. Numerical Study

We conducted a numerical study to illustrate the above theorem. In this study, we consider a marketplace where customers obtain a reward of $R = 1$, incur a waiting cost of $c = 0.01$ per unit time, and abandon the system with rate of $1/m_a = 1/1.2$. While keeping $\Lambda/k = 1.2$, we assume the number of agents in the marketplace, k, is either 40, 80, or 120. For each of these three marketplaces, we compute the (δ,ϵ)-Market Equilibrium while varying δ on a grid from 0.05 to 0.4 at step size of 0.05 and keeping $\epsilon = 0.001$. Considering an instance, we characterize the set of prices that can form a symmetric (δ,ϵ)-Market Equilibrium. In a system with k agents, we first calculate the revenue of an agent when all agents charge p, i.e. $V_1(D_1^{MCE}(p;k);p^k;k)$, for all $p \in \mathcal{A} = \{0.01, 0.02, \ldots, 0.99\}$. Then, we compare this revenue with the revenue of an agent who deviates with δk number of agents to charge $p' \neq p$, i.e. $V_2(D_1^{MCE}(p',p;\delta k, k-\delta k), D_2^{MCE}(p',p;\delta k, k-\delta k); p',p;\delta k, k-\delta k)$. (To compute the agent utilities and customer equilibria, we need the abandonment function for which we appeal to the exact expressions in Zeltyn and Mandelbaum (2005).) If $V_1(D_1^{MCE}(p;k);p^k;k)+\epsilon \geq V_2(D_1^{MCE}(p',p;\delta k, k-\delta k), D_2^{MCE}(p',p;\delta k, k-\delta k); p',p;\delta k, k-\delta k)$ for any $p' \in \mathcal{A} \setminus p$, we conclude that p can form a symmetric (δ,ϵ)-Market Equilibrium.

As it can be seen in Figure 2, the set of prices that can be sustained as (δ,ϵ)-Market Equilibrium shrinks as the number of agents increases for any given δ. Consistent with Theorem 5, we also observe that the equilibrium prices converge to the reward for customers in a seller's market. Similar results were obtained for other parameter values of the marketplace. (We vary c/R from 0.01 to 0.1, and m_a from 0.5 to 1.2.)

6.3. Supplying Real-Time Congestion Information

In Section 6.1, we show that the moderating firm can ensure that agents charge prices arbitrarily close to R in the equilibrium when it complements its efficient matching mechanism with the ability to communicate in a seller's market. Here we want to discuss whether the ability to communicate leads to high equilibrium prices when the moderating firm provides real-time queue information

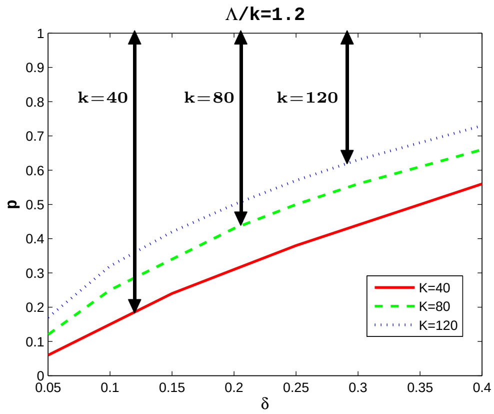


Figure 2 **The set of prices that form a symmetric (δ, ϵ)-Market Equilibrium in the marketplace with k agents is the area above (below) the corresponding curve when $\Lambda/k = 1.2$.**

in order to reduce the mismatch between customers and agents. Note that we already discuss that only the prices very close to zero can be sustained as an equilibrium in a buyer's market when the moderating firm provides real-time congestion information. Since we use a more restrictive equilibrium concept when there is communication opportunity, there will not be any new equilibrium in a buyer's market if the moderating firm provides real-time congestion information and allows agents to make pre-play communication.

In a seller's market, as in Section 5.2, characterizing the equilibrium outcome in general is again analytically intractable. However, considering a special case where $\delta^k = 1$ for all k, we can analytically show that allowing communication leads to higher equilibrium prices even when the moderating firm provides real-time queue information.

Note that the operations of a marketplace, in which the firm provides real-time queue information and all agents charge the same price, behave like a parallel server system where customers are joining the server with the shortest queue length, i.e. an $M/M/k + M/JSQ$ system[8]. There is a huge volume of literature studying such systems without customer abandonments, but none of these papers provides an exact expression for the performance evaluation of the system in a general setting (See Halfin (1985), Grassmann (1980), Blanc (1987), Nelson and Philips (1989)). Fortunately, almost all of these studies highlight the close connection between an $M/M/k/JSQ$ and an $M/M/k$ system, and show that they behave almost the same under certain conditions (For example of the system size k becomes large). Motivated by these studies, we state the following proposition by supposing that the performance of an $M/M/k + M/JSQ$ system and an $M/M/k + M$ system are close to each other when k is large. We show that only the prices above a certain threshold,

[8] In this notation, k denotes the number of parallel and independent servers, and JSQ represents the policy used to route arrivals to the servers.

which depends on R and ζ, which measures the gap in performance between an $M/M/k+M/JSQ$ system and an $M/M/k+M$ system, can emerge as the equilibrium price in a marketplace in which real-time queue information is provided. It is worth noting that if ζ is equal (or close) to zero, as it is argued for the multi-server systems without customer abandonments, the above proposition provides the same conclusion as Theorem 5.

PROPOSITION 5. *Let $\beta^{JSQ}(\lambda;k)$, and $\sigma^{JSQ}(\lambda;k)$ be the probability of abandonment, and agent utilization in a $M/M/k+M/JSQ$ system with arrival rate λ, service rate 1, abandonment rate $1/m_a$. Suppose $\lim_{k\to\infty}\frac{\beta^{JSQ}(\rho k;k)}{\beta^{M}(\rho k;k)}<\infty$, and $\lim_{k\to\infty}\frac{\sigma^{JSQ}(\rho k;k)}{\rho(1-\beta^{M}(\rho k;k))}>1-\zeta$. Then, let p^k_{info} be a price emerging in a symmetric (δ,ϵ^k)-Market Equilibrium in the k^{th} marketplace whit real-time congestion information. If $\rho>1$ and $\delta^k=1$, then for any $\xi>0$, there exists a K such that $p^k_{info}\geq R(1-\zeta)-\xi$ for all $k>K$.*

7. A Marketplace with Non-Identical Agents (Extended Version)

In Section 3, we introduce a model where all of the agents in the marketplace are a priori identical, and thus customers earn a reward of R when their service is completed, regardless of the agent performing the service. However, it is natural to imagine that large service marketplaces attract service providers with different skill sets, which provide their customers different values for the service. In this section, we explore the robustness of the conclusions of the previous sections to the heterogeneity among service providers.

To this end, we extend our original model by considering a marketplace where agents provide the same service but in different quality levels, say low (L) and high (H). We assume α_i fraction of agents provide quality-i service for $i\in\{H,L\}$ while there are still k agents in total. Furthermore, we assume that customers value the service with respect to its quality. Particularly, customers earn a reward of R_i when they are served by a quality-i agent for $i\in\{H,L\}$ where $R_L\leq R_H$. We also distinguish the agents according to their operating costs. We assume the operating cost of quality-i agents is w_i for $i\in\{H,L\}$ where $w_L\leq w_H$. For notational convenience, we let $w_L=0$. We refer to the difference between R_i and w_i as the "quality-cost differential" of quality-i agents for $i\in\{H,L\}$. As in Section 3, the arrival rate is Λ, the abandonment rate is $1/m_a$, the waiting cost is c, and the service rate is 1 for all agents.

In our model with identical agents, our major results are: 1) When the moderating firm does not intervene in the marketplace, the symmetric equilibrium price will be the outcome of a pure competition model and increases as the demand increases. 2) Providing operational efficiency alone may deteriorate the profit of the moderating firm. 3) Complementing operational efficiency with a strategic tool, which allows communication among service providers, helps the moderating firm to achieve the "expected benefit" of the efficient matching in a seller's market. In the next three subsections, we compare these findings with the results for a model with non-identical agents. We will use a similar mode of analysis as in Sections 4-6. We also discuss how the composition of the marketplaces, i.e., the ratio between high-quality and low-quality agents, affects the outcomes.

7.1. No-intervention Model

In the model with non-identical agents, we again start with the behavior of the marketplace when the moderating firm confines itself to setting up the necessary infrastructure. In such a setting, each agent's operations can still be modeled as an $M/M/1+M$ queuing system.

As the agents provide a different quality of service, the expected utility of a customer will not only depend on the price announced by the agent who serves her but also the reward she earns after being served by this particular agent. To account for that, we define the "*net reward*" of a customer from a quality-i agent charging p as $R_i - p$ for $i \in \{H, L\}$. Then, if the rate of customers who request service from an agent offering a net reward of r is λ, the expected utility of a customer requesting service from this agent is $U^N(\lambda, r) = (r + cm_a)[1 - \beta(\lambda)] - cm_a$. Furthermore, the revenue of that agent is $V_i^N(\lambda, p) = (p - w_i)\lambda[1 - \beta(\lambda)]$ if he is a quality-i agent for $i \in \{H, L\}$.

The formal definitions of the Customer Equilibrium and Sub-Game Perfect Nash Equilibrium can be trivially adopted to our new model and are, thus, omitted. We denote the Customer Equilibrium by $D_n^{CE}(r_1, \ldots, r_{k_H}, r_{k_H+1}, \ldots, r_k)$ when $(p_1, \ldots, p_{k_H})$, i.e. the first k_H element of the price vector, are the prices announced by the high-quality agents, $(p_{k_H+1}, \ldots, p_k)$, i.e. the last k_L element of the price vector, are the prices announced by the low-quality agents, and r_n is the net reward offered by agent-n for any $n = \{1, \ldots, k\}$. Note $r_n = R_H - p_n$ if $n \leq k_H$ while $r_n = R_L - p_n$ if $n > k_H$.

7.1.1. Characterization of SPNE: As before, we focus on the symmetric SPNE. Since we have two groups of agents, we define the symmetric equilibrium as one where all the high-quality agents charge p_H and all the low-quality agents charge p_L. Then, a price pair (p_H, p_L) form a symmetric equilibrium price, if they satisfy:

$$p_H \in \underset{p_\ell \geq w_H}{\arg\max} \;\; (p_\ell - w_H)\Lambda D_\ell^{CE}(r_1, \ldots, R_H - p_\ell, \ldots, r_{k_H}, r_{k_H+1}, \ldots, r_k) \times \left[1 - \beta(\Lambda D_\ell^{CE}(r_1, \ldots, R_H - p_\ell, \ldots, r_{k_H}, r_{k_H+1}, \ldots, r_k))\right], \quad (4)$$

$$p_L \in \underset{p_\ell \geq w_L}{\arg\max} \;\; (p_\ell - w_L)\Lambda D_\ell^{CE}(r_1, \ldots, r_{k_H}, r_{k_H+1}, \ldots, R_L - p_\ell, \ldots, r_k) \times \left[1 - \beta(\Lambda D_\ell^{CE}(r_1, \ldots, r_{k_H}, r_{k_H+1}, \ldots, R_L - p_\ell, \ldots, r_k))\right], \quad (5)$$

where for any $n \neq \ell$, $r_n = R_H - p_H$ if $n \leq k_H$ while $r_n = R_L - p_L$ otherwise. Note that the objective function in (4) is the revenue of a high-quality agents when he deviates and charge p_ℓ. Hence, (4) is the best-response problem of a high-quality agent. Similarly, (5) is the best-response problem of a low-quality agent.

We denote the symmetric SPNE by $(D_H^*, D_L^*; p_H^*, p_L^*)$ where all the quality-i (low-quality) agents charge p_i^* and each quality-i agent has a demand of ΛD_i^* for $i = \{H, L\}$. We also denote the equilibrium revenue of quality-i agents by V_i^*. Solving the best-response problems in (4) and (5) for any given (p_H, p_L), we characterize the symmetric SPNE in the following theorem.

THEOREM 6. *Suppose $R_H - w_H \geq R_L$. If $\beta(\lambda)$ is concave and $\frac{\beta'(\lambda)}{1-\beta(\lambda)}$ is decreasing in λ, then there exists a symmetric SPNE. Furthermore, the symmetric SPNE is characterized as follows:*

1. *If $\Lambda > k_H\lambda_H^{mon} + k_L\lambda_H^{mon}$, then the symmetric SPNE is $(D_i^*; p_i^*) = \left(\frac{\lambda_i^{mon}}{\Lambda}; R_i + cm_a - \frac{cm_a}{1-\beta(\lambda_i^{mon})}\right)$ for $i \in \{H, L\}$. Furthermore, $D_H^* \geq D_L^*$, and $V_H^* \geq V_L^*$.*
2. *If $\Lambda(0) \leq \Lambda \leq k_H\lambda_H^{mon} + k_L\lambda_H^{mon}$, then the symmetric SPNE is $(D_i^*; p_i^*) = \left(\tilde{D}_i\Lambda; R_i + cm_a - \frac{cm_a}{1-\beta(\tilde{D}_i)}\right)$ for $i \in \{H, L\}$, where $(\tilde{D}_L, \tilde{D}_H) \in \{(x, y) : \hat{U}_L(x, y) \leq 0, \hat{U}_H(x, y) \leq 0,\ k_L x + k_H y = \Lambda\}$.*
3. *If $\Lambda(R_L) < \Lambda < \Lambda(0)$, then the symmetric SPNE is*

$$(D_i^*; p_i^*) = \left(\hat{D}_i(\Lambda)/\Lambda; R_i + cm_a - \frac{(R_i + cm_a - w_i)\left[k_i - 1 + k_j\vartheta(\hat{D}_i(\Lambda), \hat{D}_j(\Lambda))\right]}{\frac{k_i + k_j\vartheta(\hat{D}_i(\Lambda), \hat{D}_j(\Lambda))}{1-\nu(\hat{D}_i(\Lambda))} - 1}\right),$$

for $i, j \in \{H, L\}$ and $j \neq i$. Furthermore, $D_H^* \geq D_L^*$ and $V_H^* \geq V_L^*$.

4. *If $k_H\lambda_H^{R_L} \leq \Lambda \leq \Lambda(R_L)$, then the symmetric SPNE is*

$$(D_H^*, D_L^*; p_H^*, p_L^*) = \left(1/k_H, 0; R_H + cm_a - \frac{R_L + cm_a}{1-\beta(\Lambda/k_H)}, p\right).$$

where $p \leq R_L$, and $p = 0$ when $\Lambda > k_H\lambda_H^{R_L}$.

5. *If $\Lambda < k_H\lambda_H^{R_L}$, then the symmetric SPNE is*

$$(D_H^*, D_L^*; p_H^*, p_L^*) = \left(1/k_H, 0; R_H + cm_a - (R_H + cm_a - w_H)\left(\frac{k_H - 1}{\frac{k_H}{1-\nu(\Lambda/k_H)} - 1}\right), 0\right).$$

Here λ_i^{mon} is the unique solution to $1 - \beta(\lambda) - \lambda\beta'(\lambda) = \frac{cm_a}{R_i + cm_a - w_i}$ for $i \in \{H, L\}$, $\lambda_H^{R_L}$ is the unique solution to $(R_H + cm_a - w_H)(k_H - 1) - \frac{R_L + cm_a}{1-\beta(\lambda)}\left(\frac{k_H}{1-\nu(\lambda)} - 1\right) = 0$,

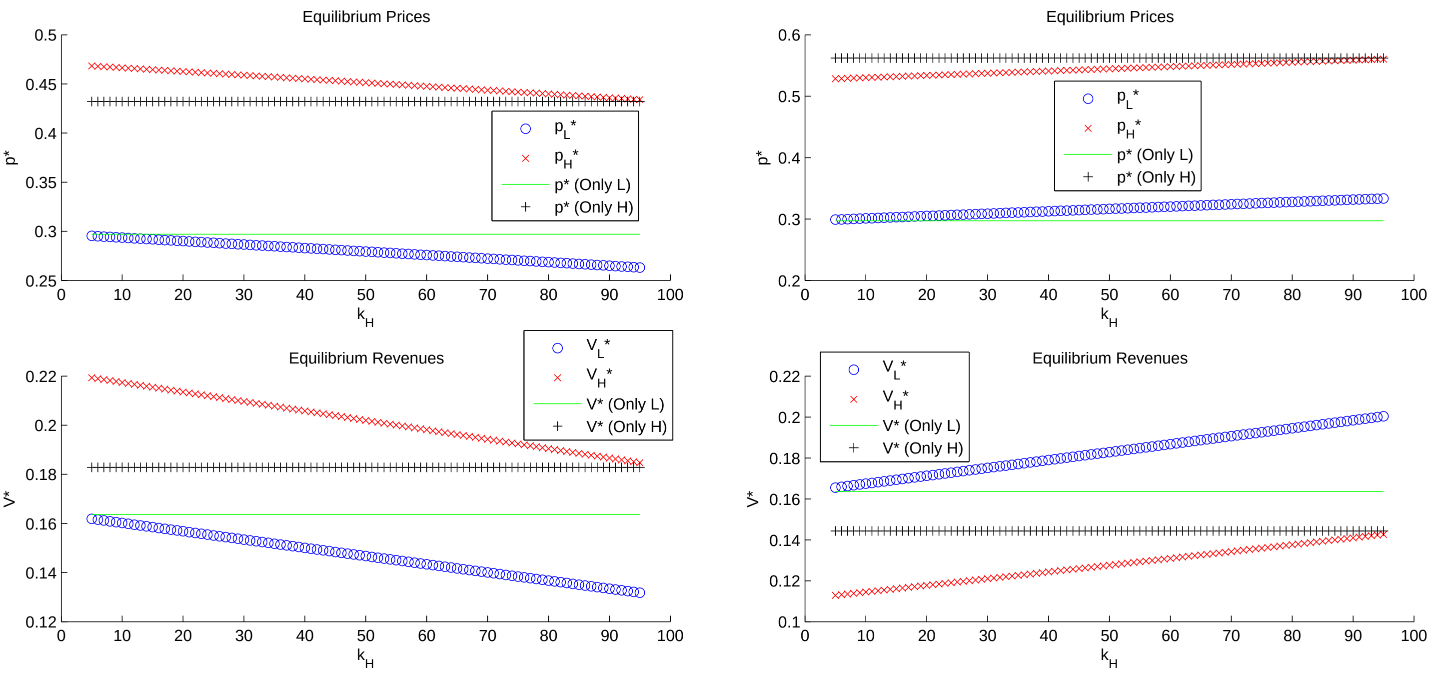
$$\begin{aligned}
(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)) &= \{(x, y) : \hat{U}_L(x, y) = \hat{U}_H(x, y),\ k_L x + k_H y = \Lambda\} \\
\hat{U}_L(x, y) &= (R_L + cm_a)\left[\frac{1 - \nu(x)}{1 + \frac{\nu(x)}{k_L + k_H\vartheta(x,y) - 1}}\right](1 - \beta(x)) - cm_a, \\
\hat{U}_H(x, y) &= (R_H + cm_a - w_H)\left[\frac{1 - \nu(y)}{1 + \frac{\nu(y)}{k_H + k_L\vartheta(y,x) - 1}}\right](1 - \beta(y)) - cm_a,
\end{aligned}$$

$\vartheta(x, y) = \frac{y\nu(x)}{x\nu(y)}$, and $\Lambda(u)$ is the unique solution to $\hat{U}_H(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)) = u$.

The above equilibrium characterization is very similar to the equilibrium in Theorem 1: Agents may behave as local monopolists when the arrival rate is sufficiently high, whereas once the arrival rate is less than $\Lambda(0)$, the competition between agents is intensified. As a result of intensified competition, customers observe lower prices, which allow them to earn strictly positive utility. However, we also encounter new results when we allow for heterogeneous agents. First, unlike the identical agent model, we observe that the main driver of equilibrium outcomes for certain parameters is not only the competition between providers but also the fact that agents offer different quality of service. For instance, when the demand rate is between $k_H\lambda_H^{R_L}$ and $\Lambda(R_L)$, high-quality agents charge a low price and forego a significant customer surplus both because of the low demand and the fact that they want to keep the low-quality agents out of the marketplace. Furthermore, there is a continuum of symmetric equilibria when the demand rate is between $\Lambda(0)$ and $k_H\lambda_H^{mon} + k_L\lambda_H^{mon}$, whereas we always have a unique symmetric equilibrium with the identical agent. Although the prices charged by different groups of agents vary in these equilibria, the expected utility of customers is always zero.

Note that high-quality agents almost always serve more customers and earn more revenue in the equilibrium. Moreover, if the arrival rate is less than $\Lambda(R_L)$, the market is covered solely by the high-quality providers. These findings stem from our assumption that high-quality agents have a higher quality-cost differential, i.e., $R_H - w_H \geq R_L$. The above theorem establishes that only the quality-cost differentials of the agents matter in the equilibrium. In other words, a marketplace where high quality agents have an operating cost of w_H, and generate a reward of R_H is the same as a marketplace where high quality agents have no operating cost, and generate a reward of $R_H - w_H$ in terms of equilibrium outcome. Therefore, if the quality-cost differential of the low-quality agents were higher, the same equilibrium characterization would hold with the only exception that low-quality agents would earn more revenue.

Having two different groups of agents allows us to discuss the impact of the fraction of agents with a certain quality, α_H and α_L, on the equilibrium outcomes. However, the equilibrium characterization in Theorem 6 is not explicit enough to show this impact analytically. Therefore, we explore this question by an extensive numerical study. As Figure 3 illustrates, the equilibrium prices and revenues decrease as we have more high-quality agents in the marketplace when $R_H - w_H \geq R_L$. On the other hand, having more high-quality agents let the prices and revenues go up when $R_H - w_H \leq R_L$. In other words, revenues in a marketplace is deteriorated as a result of having more low-quality agents only when the operating cost of providing high-quality service is significant.



(a) (b)

Figure 3 **Equilibrium prices and revenues as a function of the number of high-quality agents,** k_H**. For all examples** $\Lambda = 80$, $k = 100$, $R_H = 1$, $R_L = 0.8$, $w_L = 0$, $c = 0.05$, $m_a = 1$. **In (a)** $w_H = 0.1$, **in (b)** $w_H = 0.3$.

The implications on the identical agent model: The equilibrium characterization in Theorem 6 also helps us to prove that the non-symmetric equilibrium may exist only for a small range of demand-supply ratio ρ in the no-intervention model with identical agents. Furthermore, we show that this range shrinks to zero as the number of agents grow. The following proposition presents these results formally:

PROPOSITION 6. *When the moderating firm does not intervene in a marketplace with k identical agents, the symmetric equilibrium described in Theorem 1 is the unique equilibrium when* $\rho \notin [\lambda^0, \lambda^{mon}]$*. Furthermore, we have that*

$$\lim_{k\to\infty} \lambda^0 = \lambda^{mon}.$$

7.2. Operational Efficiency Model

After studying a marketplace where there is no intervention by the moderating firm, we now turn our attention to a marketplace where the moderating firm aims at reducing the unnecessary waits and idleness in the system through a matching mechanism. The matching mechanism that the moderating firm provides achieves such an operational efficiency improvement by allowing customers to postpone their agent selection. In particular, the marketplace under this matching mechanism operates as a queuing system where all agents offering the same net reward are virtually grouped together, regardless of the quality of their service. We assume that customers decide which agents to choose based on the net reward and they treat all agents as the same when they offer the same net reward because the nature of the tasks is simple, benefits are tangible, features are clear, and thus rewards are easily quantifiable. Further, one may view these tasks as commodities.[9]

Once each agent announces a price per customer to be served, we can construct a resulting net reward vector $(r_n)_{n=1}^N$ where $N \le k$ is the number of different net rewards announced by the agents. As before, we refer to agents announcing r_n as sub-pool-n, and denote the number of agents of quality-i in the sub-pool-n by y_{i_n} for $i \in \{H, L\}$. Hence, (r_n, y_{H_n}, y_{L_n}) summarizes the strategy of all agents.

In a marketplace with non-identical agents, the customer decision making and experience is the same as in Section 5, and we still denote the fraction of customers requesting service from the sub-pool-n by D_n. Given the decisions of customers and agents, the expected utility of a customer choosing sub-pool-ℓ is

$$U_\ell(D_1,\dots,D_N;r_1,\dots,r_N;y_1,\dots,y_N) = PServ_{\ell\ell}\big[(r_\ell + cm_a)(1-\beta_\ell) - cm_a\big] + \sum_{m\neq\ell} PServ_{\ell m} r_m,$$

where $y_n = y_{H_n} + y_{L_n}$, $\beta_\ell(D_1,\dots,D_N;r_1,\dots,r_N;y_1,\dots,y_N)$ denotes the probability of abandonment in the sub-pool-ℓ, and $PServ_{\ell m}(D_1,\dots,D_N;r_1,\dots,r_N;y_1,\dots,y_N)$ denotes the probability that a customer choosing the sub-pool-ℓ is served by the sub-pool-m when ΛD_n is the rate of customer arrival to the sub-pool-n for $n = 1,\dots,N$. Note that the expected utility of customers depends on the total number of agents in a sub-pool instead of the number of agents with different service quality, because all agents are treated equally by the customers as long as they offer the same net reward. Furthermore, the revenue of a quality-i agent in sub-pool-ℓ for $i \in \{H, L\}$ is $(p_\ell -$

[9] One may envision a model, in which customers strictly prefer high-quality agents even if they provide the same net reward as low-quality agents. Such a model would require additional notation and analysis but our key findings, namely providing operational efficiency may lead to profit loss and enabling communication may help to overcome that loss, continue to hold.

$w_i)\sigma_\ell(D_1,\ldots,D_N;r_1,\ldots,r_N;y_1,\ldots,y_N)$, where $\sigma_\ell(D_1,\ldots,D_N;r_1,\ldots,r_N;y_1,\ldots,y_N)$ is utilization of agents in sub-pool-ℓ when ΛD_n is the rate of customer arrival to the sub-pool-n for $n=1,\ldots,N$.

The Market Customer Equilibrium and ϵ-Market Equilibrium are again the natural extensions of the definitions in Section 5 to a marketplace with non-identical agents and are, thus, omitted. We denote the fraction of customers requesting service from sub-pool-n in a Market Customer Equilibrium by $D_n^{MCE}(r_1,\ldots,r_N;y_1,\ldots,y_N)$ when $(r_n,y_{Hn},y_{Ln})_{n=1}^N$ is a tuple of three vectors whose components are the net rewards and the number of agents announcing them, and $y_n=y_{Hn}+y_{Ln}$. We study the behavior of the equilibrium in large marketplaces by considering the sequence of marketplaces we described in Section 5 along with the following assumption: the number of high-quality and low-quality agents are $\alpha_H k$ and $\alpha_L k$, respectively, in the k^{th} marketplace. This ensures that the ratio of high and low-quality agents is constant along the sequence of marketplaces.

7.2.1. Characterization of the Market Equilibrium: Here we characterize the symmetric equilibrium along the trajectory of marketplaces introduced above. As a first step, we derive the revenue of agents when all the quality-i agents for $i\in\{H,L\}$, charge the same price p_i^k in the k^{th} marketplace where $p_i^k\to p_i$ as $k\to\infty$.

Proposition 7. *Let $V_H^{MCE}(p_H^k,p_L^k;k)$ and $V_L^{MCE}(p_H^k,p_L^k;k)$ be the revenue of an high-quality and a low-quality agent, respectively, when all the high-quality agents charge p_H^k and all the low-quality agents charge p_L^k in the k^{th} marketplace. If $R_i-p_i>R_j-p_j$ for some $i,j\in\{H,L\}$ with $i\neq j$, then we have that*

$$\lim_{k\to\infty} V_i^{MCE}(p_H^k,p_L^k;k)=\begin{cases}\frac{\rho}{\alpha_i}(p_i-w_i) & \text{if } \rho\le\alpha_i,\\ p_i-w_i & \text{if } \rho>\alpha_i.\end{cases}$$

Furthermore, when $p_j<R_j$, we have that

$$\lim_{k\to\infty} V_j^{MCE}(p_H^k,p_L^k;k)=\begin{cases}0 & \rho\le\frac{(R_i-p_i+cm_a)\alpha_i}{R_j-p_j+cm_a},\\ \left(\frac{\rho}{\alpha_j}-\frac{(R_i-p_i+cm_a)\alpha_i}{(R_j-p_j+cm_a)\alpha_j}\right)(p_j-w_j) & \frac{(R_i-p_i+cm_a)\alpha_i}{R_j-p_j+cm_a}<\rho\le\frac{\bar R}{R_j-p_j+cm_a},\\ p_j-w_j & \rho>\frac{\bar R}{R_j-p_j+cm_a},\end{cases}$$

and when $p_j=R_j$, we have that

$$\lim_{k\to\infty} V_j^{MCE}(p_H^k,p_L^k;k)\le\begin{cases}0 & \rho\le\frac{(R_i-p_i+cm_a)\alpha_i}{cm_a},\\ \left(\frac{\rho}{\alpha_j}-\frac{(R_i-p_i+cm_a)\alpha_i}{cm_a\alpha_j}\right)(p_j-w_j) & \frac{(R_i-p_i+cm_a)\alpha_i}{cm_a}<\rho\le\frac{\bar R}{cm_a},\\ p_j-w_j & \rho>\frac{\bar R}{cm_a},\end{cases}$$

where $\bar R=(R_H-p_H)\alpha_H+(R_L-p_L)\alpha_L+cm_a$.

The above proposition establishes that the group of agents offering the higher limiting net reward will always be over-utilized as long as the total demand exceeds their capacity. On the other hand, the group offering the lower net reward will be under-utilized unless ρ is sufficiently high. It is worth highlighting that the net reward does not depend on the operating cost of the agents, and thus the customer equilibrium is independent of the operating cost of an agent.

Using the above proposition, the group offering the lower limiting net reward will always be "under-utilized" in a buyer's market since $\frac{\bar{R}}{R_j - p_j + cm_a} > 1 > \rho$ in a buyer's market for $j \in \{H, L\}$, such that $R_i - p_i > R_j - p_j$ with $i \in \{H, L\}$ and $i \neq j$. In other words, ρ cannot be high enough to let the group offering the lower net reward to be over-utilized in a buyer's market. It turns out this will create an opportunity for the members of that group to improve their revenue by slightly decreasing their price if it is strictly greater than their operating cost. Therefore, we cannot have any symmetric equilibrium where the agents with different quality offer a different level of net reward and both of them earn strictly positive revenue. Furthermore, when they offer the same level of net reward in the limit, at least one group of agents will be "under-utilized" along the trajectory of marketplaces since $\rho < 1$. Then, we show that there will always be room for a single agent to improve his revenue in a buyer's market. Hence, in a buyer's market, we can only have an equilibrium where only one group of agents (high or low) can earn positive revenue in the limit. In fact, Theorem 7 below establishes that group of agents with the higher quality-cost differential can earn positive revenue when demand exceeds their capacity.

The above proposition also states that both of the groups can be over-utilized when they offer different level of limiting net reward in a seller's market. In this case, cutting the price does not help the agents to improve their revenues. However, we show that in this setting, an agent from the group offering the higher net reward will have an opportunity to increase his price and improve his revenue. Thus, even in a seller's market, it will not be possible to see a symmetric equilibrium where the agents with different quality offer a different level of net reward. Finally, when all agents offer the same level of limiting net reward in a seller's market, the pooling benefits associated with operational efficiency will again serve as a deterrent for deviation as in the case of identical agents. Thus, there will be multiple symmetric equilibria in a seller's market as established in Theorem 3. We formally present these observations in the following result:

THEOREM 7. *Suppose* $R_i - w_i > R_j - w_j$ *for some* $i, j \in \{H, L\}$ *with* $i \neq j$. *Let* $(p^k_{EQ_i}, p^k_{EQ_j})$ *be a price pair emerging as the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace.*

1. *If $\rho < \alpha_i$, then for any $\xi > 0$, there exists a K such that $p^k_{EQ_i} < w_i + \xi$ for all $k > K$. Furthermore, there exists a K' such that any price pair (p_i, p_j), with $p_i = w_i$, emerges as the equilibrium price pair of a symmetric ϵ^k-Market Equilibrium in the k^{th} marketplace for all $k > K$.*
2. *If $\alpha_i < \rho < 1$, then*
   (a) *For any $\xi > 0$, we have either one of the following*

   (i) *there exists a K such that $p^k_{EQ_i} < \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\dfrac{\rho}{\alpha_i} - 1\right]\right\} + \xi$ for all $k > K$,*

   (ii) *there exists a K such that $|p^k_{EQ_i} - (R_i - R_j + w_j)| < \xi$, and $p^k_{EQ_j} < w_j + \xi$ for all $k > K$.*

   (b) *For any given sequence of price pairs $(\tilde{p}^k_i, \tilde{p}^k_j)$ where $\tilde{p}^k_i$ converges to $\tilde{p}_i$ as $k \to \infty$, $\tilde{p}_i \in \mathcal{P}(\rho/\alpha_i, R_i - w_i)$ and $w_i \le \tilde{p}_i < (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]$, there exists a K such that $(\tilde{p}^k_i, \tilde{p}^k_j)$ emerges as the equilibrium price pair of a ϵ^k-symmetric Market Equilibrium in the k^{th} marketplace for all $k > K$.*

   (c) *There exists a sequence of prices $\hat{p}^k_i$ and a K such that $\hat{p}^k_i < R_i - R_j + w_j$, $\lim_{k\to\infty} \hat{p}^k_i = R_i - R_j + w_j$, and $(\hat{p}^k_i, w_j)$ is the equilibrium price pair of a symmetric ϵ^k-Market Equilibrium in the k^{th} marketplace for all $k > K$.*
3. *If $\rho > 1$, then*
   (a) *For any given sequence of price pairs $(\tilde{p}^k_i, \tilde{p}^k_j)$ where $\tilde{p}^k_\ell$ converges to $\tilde{p}_\ell$ as $k \to \infty$ for $\ell \in \{H, L\}$, $\tilde{p}_j \in \mathcal{P}(\rho, R_j)$ and $\tilde{p}_i = \tilde{p}_j + R_i - R_j + w_j$, there exists a K such that $(\tilde{p}^k_i, \tilde{p}^k_j)$ emerges as the equilibrium price pair of a symmetric ϵ^k-Market Equilibrium in the k^{th} marketplace for all $k > K$.*

   (b) *For any $\xi > 0$, there exists a K such that $p^k_{EQ_i} < \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\} + \xi$ for all k when $p^k_{EQ_\ell}$ converges to p_{EQ_ℓ} as $k \to \infty$ for $\ell \in \{H, L\}$ $p_{EQ_i} \neq p_{EQ_j} + R_i - R_j + w_j$.*

When we focus on the case where $i = H$ and $j = L$, the above theorem shows that high-quality agents can only charge a price very close to their operating costs in a symmetric equilibrium when demand is low, and in such a setting, only the high-quality agents can serve customers since customers strictly prefer high-quality agents. Thus, the revenue of all agents is in a small neighborhood of zero in large marketplaces when demand is low. This result is similar to the one in the model with identical agents. On the other hand, when demand is sufficiently high, namely $\rho > \alpha_H$, there are multiple symmetric equilibria. The only significant difference here is that there may be multiple equilibria even in a buyer's market as long as demand exceeds the total capacity of high-quality agents. However, similar to Section 5, most of these equilibrium prices may be very low compared to the equilibrium outcome in the no-intervention model. Thus, providing tools to improve the operational efficiency may still deteriorate the moderating firm's profit. It is also worth noting that the best equilibrium outcome from the perspective of agents and the moderating firm is the one where the high-quality agents charge almost $R_H - R_L$ when $\alpha_H < \rho < 1$. Even this outcome may be worse than the outcome in a no-intervention model as long as R_H and R_L are close to each other. In fact, this outcome is equivalent to the almost zero price equilibrium when $R_H - w_H \simeq R_L$ and $w_H = 0$.

When we consider the case where $i = L$ and $j = H$, we would have similar results where the role of high- and low-quality agents are flipped. For instance, the total capacity of the low-quality agent would determine where we have multiple equilibria in a buyer's market, and only the low-quality agents would serve customers in all of these equilibria. Furthermore, the only apparent impact of α_H and α_L is that the region where the agents charge their operating costs in equilibrium expands by any increase in α_H when $R_H - w_H > R_L$.

The implications on the identical agent model: One important result in the above theorem is that any sequence of price pairs (p_H^k, p_L^k) with both $p_H > w_H$ and $p_L > 0$ cannot be an equilibrium in large marketplaces. This result holds even for the setting where $R_H - w_H = R_L$ and $w_H = 0$. Using this observation, we can argue that there cannot be any non-symmetric equilibrium where agents earn positive revenue in the operational efficiency model with identical agents. Thus, even if there are any non-symmetric equilibrium in a buyer's market with identical agents, the revenue of all agents should be almost zero.

Proposition 8. *Let $(p_1^k, \ldots, p_N^k)$ be a price vector emerging as the equilibrium price vector of an ϵ^k-Market Equilibrium in the k^{th} marketplace in the operational efficiency model with identical agents where $p_1^k < p_2^k < \cdots < p_N^k$. Then for any $\xi > 0$, there exists a K such that $p_1^k < \xi$ and $V_n^k < \xi$ for all $k > K$, where V_n^k is the revenue of agents charging p_n^k in the k^{th} marketplace.*

7.3. Communication Enabled Model

As in the model with identical agents, we finally explore the impact of enabling communication among agents in a market with non-identical agents. To this end, we study the behavior of the (δ^k, ϵ^k)-Market Equilibrium in large marketplaces by considering the sequence of marketplaces described in the previous sub-section.

In Theorem 7, we show that the revenue of agents is always in a small neighborhood of zero in large marketplaces when the total capacity of agents with high quality-cost differential exceed demand. Since (δ^k, ϵ^k)-Market Equilibrium is a refinement of ϵ^k-Market Equilibrium, this equilibrium outcome is the only possible (δ^k, ϵ^k)-Market Equilibrium when demand is sufficiently low in a buyer's market.

Theorem 7 also establishes that there are multiple symmetric ϵ^k-Market Equilibria when ρ exceeds the capacity of agents with higher quality-cost differential. In most of these equilibria, agents with higher quality-cost differential will be over-utilized, and thus will have an opportunity to capitalize on this congestion when pre-play communication is allowed as we discussed in Section 6. However there is limit for that capitalization in a buyer's market. Namely, when $R_i - w_i > R_j - w_j$ for some $i, j \in \{H, L\}$ with $i \neq j$, quality-i agents cannot charge a price higher than $R_i - R_j + w_j$ because of the threat of quality-j agents. It turns out such a threat does not exist in a seller's market. Hence, agents can sustain the profit maximizing one among the multiple equilibria arising due to providing operational, when $\rho > 1$. We summarize these results in the following theorem:

THEOREM 8. *Suppose $R_i - w_i > R_j - w_j$ for some $i, j \in \{H, L\}$ with $i \neq j$. Let $(p^k_{EQ_i}, p^k_{EQ_j})$ be a price pair emerging in a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace.*

1. *If $\rho < \alpha_i$, then for any $\xi > 0$, there exists a K such that $p^k_{EQ_i} < w_i + \xi$ for all $k > K$. Furthermore, when $\lim_{k\to\infty} \delta^k = 0$, there exists a sequence $(\hat{p}^k_i, \hat{p}^k_j)$ and a K' such that $(\hat{p}^k_i, \hat{p}^k_j)$ forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for all $k > K'$.*
2. *If $\alpha_i < \rho < 1$, then for any $\xi > 0$, there exists a K such that $|p^k_{EQ_i} - (R_i - R_j + w_j)| < \xi$ and $p^k_{EQ_j} < \xi$ for all $k > K$. Furthermore, when $\lim_{k\to\infty} \delta^k = 0$, there exists a sequence $(\hat{p}^k_i, \hat{p}^k_j)$ and a K' such that $(\hat{p}^k_i, \hat{p}^k_j)$ forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for all $k > K'$ when $\lim_{k\to\infty} \delta^k = 0$.*
3. *If $\rho > 1$, then for any $\xi > 0$, there exists a K such that $p^k_{EQ_\ell} > R_\ell - \xi$ for $\ell \in \{H, L\}$ and all $k > K$. Furthermore, there exists a sequence $(\hat{p}^k_i, \hat{p}^k_j)$ and a K' such that $(\hat{p}^k_i, \hat{p}^k_j)$ forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for all $k > K'$.*

The implications on the identical agent model: Similar to the previous results in this section, the above theorem helps us to show that, even if there are any non-symmetric equilibrium in a seller's market with identical agents, the revenue of agents in equilibrium should converge to R as well as the price they charge.

PROPOSITION 9. *Let $(p^k_1, \ldots, p^k_N)$, where $N > 1$, be an equilibrium price vector of an (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace with identical agents. If $\rho > 1$, then for any $\xi > 0$, there exists a K such that $p^k_n > R - \xi$ for all $n \in \{1, \ldots, N\}$ and all $k > K$.*

8. Conclusion

In this paper, we study a marketplace in which many small service providers compete with each other in providing service to self-interested customers looking for temporary help. The main focus of the paper is on the role of the moderating firm, which sets up the marketplace and creates the infrastructure where agents and customers interact. To this end, we explore the impact of different strategies employed by the moderating firm by considering three market models, where the moderating firm has different degrees of involvement.

We characterize the market outcomes in each of these models. We observe that outcomes critically depend on the moderating firm's involvement and market conditions, i.e., whether it is a buyer's or a seller's market. Since different types of involvement of the moderating firm result in different equilibrium prices and customer demand, the moderating firm aims to intervene in the marketplace in order to make sure that the "right" prices and customer demand emerge in equilibrium. Specifically, the moderating firm tries to maximize the revenues of agents since its profit is a share of the agents' revenues.

We show that when the firm ensures efficient operational matching and enables agent communication in a seller's market, the natural upper-bound on the revenue generated in a marketplace[10] is asymptotically achievable, and thus, using these two tools together dominates any other strategy from the moderating firm's perspective in a seller's market. We also show that efficient operational matching in a buyer's market leads to arbitrarily small total marketplace revenue compared to the total revenue under the no-intervention model. Hence, using the matching mechanism we discuss in this paper is not advisable in a buyer's market despite the fact that it reduces the mismatch between demand and supply. This result is somewhat counter-intuitive, because the efficiency improvement due to better matching is not necessarily translated into additional profits. It seems other tools aimed at improving the operational efficiency, such as providing real-time queue information, will have a similar impact on the moderating firm's profit in a buyer's market.

Both oDesk.com and ServiceLive.com are currently in their growth stage and have not achieved their full potential in terms of demand for their services. However, both firms can and should project the "mature" market conditions and decide on their appropriate measures to adopt. Given the moderate level of congestion in oDesk.com, one may infer that the marketplace can be identified as a seller's market. Following the discussion before, oDesk.com's decision to offer operational tools complemented with strategic tools is well justified.

[10] In a given marketplace, the total revenues of the agents cannot exceed $\min\{\Lambda, k\}R$ since they cannot charge more than R, and their effective demand is the minimum of their processing capacity and the aggregate demand.

There are also other possible ways for a moderating firm to be involved in the marketplace including contracting with agents or providing a suggested price. Particularly, the setting in which the firm provides a suggested price can be viewed as pre-play communication and will indeed shrink the set of equilibria. However, these type of interactions between the moderating firm and agents are outside the scope of this paper as these settings are not a market per-se anymore. In such environments, the firm would decide on prices as well as the allocation of agents to customers.

While modeling operational efficiency, we assume that agents give priority to their own customers. One may consider an extension of our model in which agents are allowed to choose both priority and prices, simultaneously. The equilibria that arise in our model with fixed priority rule would still be sustained in such an extended game. Hence, the main spirit of our findings, namely, the fact that providing operational efficiency may lead to profit loss, would not change. Additional equilibria would be possible in the extended model only when demand exceeds supply.

References

Allon, G., A. Federgruen. 2007. Competition in service industries. *Operations Research* **55** 37–55.

Allon, G., I. Gurvich. 2010. Pricing and dimensioning competing large-scale service providers. *Manufacturing & Service Operations Management* **12** 449–469.

Amihud, Y., H. Mendelson. 1980. Dealership market: Market-making with inventory. *Journal of Financial Economics* **8** 31–53.

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

Armony, M., E. Plambeck, S. Seshadri. 2009. Sensitivity of optimal capacity to customer impatience in an unobservable M/M/S queue (Why you shouldn't shout at the DMV). *Manufacturing & Service Operations Management* **11** 19–32.

Aumann, R. J. 1959. Acceptable points in general cooperative n-person games. *Contributions to the Theory of Games* .

Biais, B., L. Glosten, C. Spattt. 2005. Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets* **8** 217–264.

Blanc, J. P. C. 1987. A note on waiting times in systems with queues in parallel. *Journal of Applied Probability* **24**(2) 540–546.

Burdett, K., D. Mortensen. 1998. Wage differentials, employer size, and unemployment. *International Economic Review* **39** 257273.

Cachon, G., P. T. Harker. 2002. Competition and outsourcing with scale economies. *Management Science* **48** 1314–1333.

Cachon, G., S. Netessine. 2004. *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, chap. Game theory in supply chain analysis. Kluwer.

Cachon, G., F. Zhang. 2007. Obtaining fast service in a queuing system via performance-based allocation of demand. *Management Science* **53** 408–420.

Chen, H., Y. Wan. 2003. Price competition of make-to-order firms. *IIE Transactions* **35** 817–832.

Chen, Y., C. Maglaras, G. Vulcano. 2008. Design of an aggregated marketplace under congestion effects: Asymptotic analysis and equilibrium characterizatio. Working Paper.

Dixon, H. 1987. Approximate bertrand equilibria in a replicated industry. *The Review of Economic Studies* **54**(1) 47–62.

Flandez, R. 2008, October 13. Help wanted and found. *The Wall Street Journal* Http://online.wsj.com/article/SB122347721312915407.html.

Garman, M. 1976. Market microstructure. *Journal of Financial Economics* **3** 257–275.

Garnett, O., A. Mandelbaum, M. Reiman. 2002. Desining a call center with impatient customers. *Manufacturing & Service Operations Management* **4** 208–227.

Gradwohl, R., O. Reingold. 2008. Fault tolerance in large games. Proceedings of EC 2008.

Granot, D., G. Sosic. 2005. Formation of alliance in internet-based supply exchanges. *Management Science* **51** 92–105.

Grassmann, W. K. 1980. Transient and steady state results for two parallel queues. *Omega* **8**(1) 105–112.

Ha, A. Y., L. Li, S. Ng. 2003. Price and delivery logistics competition in a supply chain. *Management Science* **49** 1139–1153.

Halfin, S. 1985. The shortest queue problem. *Journal of Applied Probability* **22**(4) 865–878.

Halfin, S., W. Whitt. 1981. Heavy-traffic limits for queues with many exponential servers. *Operations Research* **29** 567–588.

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

Ho, T., H. Stoll. 1983. The dynamics of dealer markets under competition. *Journal of Finance* **38** 1053–1074.

Kalai, E. 2004. Large robust games. *Econometrica* **72** 1631–1665.

Maglaras, C., A. Zeevi. 2005. Pricing and design of differentiated services: Approximate analysis and structural insights. *Operations Research* **53** 242–262.

Manning, A. 2003. The real thin theory: monopsony in modern labour markets. *Labour Economics* **10** 105131.

Manning, A. 2004. Monopsony and the efficiency of labour market interventions. *Labour Economics* **11** 145–163.

Michaelides, M. 2010. Labour market oligopsonistic competition: The effect of worker immobility on wages. *Labour Economics* **17** 230239.

Myerson, R. B. 1991. *Game Theory: Analysis of Conflict*. Harvard University Press.

Nagarajan, M., G. Sosic. 2007. Stable farsighted coalitions in competetive markets. *Management Science* **53** 29–45.

Nagarajan, M., G. Sosic. 2008. Game-theoretic analysis of cooperation among supply chain agents: Review and extensions. *European Journal of Operational Research* **187** 719–745.

Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica* **37** 15–24.

Nelson, R. D., T. K. Philips. 1989. An approximation to the response time for shortest queue routing. *SIGMETRICS Perform. Eval. Rev.* **17**(1) 181–189.

Rath, K. P. 1992. A direct proof of the existence of pure strategy equilibria in games with a continuum of players. *Economic Theory* **2** 427–433.

Ray, I. 1996. Coalition-proof correlated equilibirum: A definition. *Games and Economic Behavior* **17** 56–79.

Roughgarden, T. 2005. *Selfish Routing and the Price of Anarchy*. The MIT Press.

Rudin, W. 1976. *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill Book Company.

Sosic, G. 2006. Transhipment of inventories among retailers: Myopic vs. farsighted staility. *Management Science* **52** 1491–1508.

The Economist. 2010, May 13. Work in the digital age: A clouded future. *The Economist* Http://www.economist.com/node/16116919.

Ward, A. R., P. W. Glynn. 2003. Diffusion approximation for a markovian queue with reneging. *Queueing Systems* **43** 103–128.

Zeltyn, S., A. Mandelbaum. 2005. Call centers with impatient customers: Many-server asymptotics of the M/M/n+G queues. *Queueing Systems* **51** 361–402.

We provide the formal presentations and the proofs of the supplementary lemmas used in the appendix in the Supporting Document.

Appendix A: Proofs in Section 4

In order to solve the single agent's problem and characterize the equilibrium, the probability of abandonment function has to satisfy some technical properties. Some of these technical requirements are shown as in the following lemma.

LEMMA 1. 1. *$\beta(\lambda)$ is continuous and continuously twice differentiable.*
2. *$\beta(\lambda)$ is strictly increasing in λ for any $\lambda > 0$.*
3. *$\nu(\lambda)$ is strictly increasing in λ for any $\lambda > 0$.*
4. *$\lambda(1-\beta(\lambda))$ is strictly increasing and concave in λ for any $\lambda > 0$.*
5. *If $m_a \le 1$, then $\beta(\lambda)$ is concave in λ.*

A.1. Proof of Proposition 1

Suppose there are two Customer Equilibrium, say $(D_n)_{n=1}^k$ and $(D'_n)_{n=1}^k$, given $(p_n)_{n=1}^k$ with $p_n < R$ for some n (When $p_n = R$ for all $1 \le n \le k$, the unique equilibrium is clearly $D_n = 0$ for all $1 \le n \le k$.). Let $S = \{n \le k : D_n > 0\}$, *and* $S' = \{n \le k : D'_n > 0\}$.

Lemma 2. *We have that* $S = S'$.

As we show in Lemma 2, we have that $S = S'$. Then, let $U(\Lambda D_n, p_n) = u$ for any $n \in S$ and $U(\Lambda D'_n, p_n) = u'$ for any $n \in S'$. Since $S = S'$ and $D_n \ne D'_n$ for some $n \in S$, we have that $u \ne u'$. WLOG, assume $u > u'$. This implies that $\sum_{n=1}^{k} D_n < \sum_{n=1}^{k} D'_n \le 1$. However, since $\sum_{n=1}^{k} D_n < 1$, we have that $u' < u = 0$ which is a contradiction.

A.2. Proof of Theorem 1

Existence and uniqueness of λ^{mon} and λ^0: After a birth-death chain analysis of an $M/M/1+M$ system with arrival rate λ, service rate 1, and abandonment rate $1/m_a$, we have that $\beta(\lambda) = 1 - \frac{g(\lambda)}{\lambda(1+g(\lambda))}$, and $W(\lambda) = m_a \beta(\lambda)$ where $a_0 = 1$, $a_n = \frac{1}{\prod_{i=0}^{n-1}(1+i/m_a)} = \frac{m_a^n}{\prod_{i=0}^{n-1}(m_a+i)}$ for any $n \ge 1$, and $g(\lambda) = \sum_{n=1}^{\infty} a_n \lambda^n$.

Observe that $1 - \beta(\lambda) - \lambda\beta'(\lambda)$ is strictly decreasing in λ since $\lambda[1-\beta(\lambda)]$ is strictly concave by Lemma 1.4. Moreover, $\lim_{\lambda\to 0}\left[1-\beta(\lambda)-\lambda\beta'(\lambda)\right] = \lim_{\lambda\to 0}\frac{g'(\lambda)}{[1+g(\lambda)]^2} = 1$ since $\lim_{\lambda\to 0} g(\lambda) = 0$, and $\lim_{\lambda\to 0} g'(\lambda) = 1$. It is also true that $\lim_{\lambda\to\infty}\left[1-\beta(\lambda)-\lambda\beta'(\lambda)\right] \le 0$ since $\lim_{\lambda\to\infty}\beta(\lambda) = 1$. Therefore, it is clear that λ^{mon} exists and it is unique.

Let $z(\lambda) = (R+cm_a)(k-1) - \frac{cm_a}{1-\beta(\lambda)}\left(\frac{k}{1-\nu(\lambda)} - 1\right)$. Then, $z(\lambda)$ is strictly decreasing in λ since $\nu(\lambda)$ and $\beta(\lambda)$ are strictly increasing in λ by Lemma 1. Moreover, $z(\lambda^{mon}) = \frac{cm_a}{1-\beta(\lambda^{mon})} - (R+cm_a) < 0$. Therefore, it is clear that λ^0 exists, it is unique, and $\lambda^0 < k\lambda^{mon}$.

Necessary conditions for the symmetric equilibrium: The best response problem of agent-ℓ in (1) can be rewritten as follows:

$$\begin{aligned}
\max_{p_\ell \ge 0, D_\ell \ge 0, D_{-\ell} \ge 0} \quad & p_\ell \Lambda D_\ell \left[1-\beta(\Lambda D_\ell)\right] \\
s.to \quad & (R - p_\ell + cm_a)\left[1-\beta(\Lambda D_\ell)\right] - cm_a \ge 0 \\
& (R - p_\ell + cm_a)\left[1-\beta(\Lambda D_\ell)\right] = (R - p + cm_a)\left[1-\beta(\Lambda D_{-\ell})\right] \\
& D_\ell + (k-1)D_{-\ell} \le 1
\end{aligned}$$

In this new problem, we state the conditions of the Customer Equilibrium as the constraints of the problem. In other words, for any $(D_\ell, D_{-\ell})$ satisfying the constraints, we have that $D_\ell = D_\ell^{CE}(p,\ldots,p,p_\ell,p,\ldots,p)$ and $D_{-\ell} = D_n^{CE}(p,\ldots,p,p_\ell,p,\ldots,p)$ for any $n \ne \ell$. We denote the solution to the above problem by $(D_\ell(p), D_{-\ell}(p), p_\ell(p))$ for a given p.

Then, any symmetric SPNE (D,p) should satisfy the following FOC by the definition of the symmetric SPNE:

$$\Lambda D - \eta_1 - \eta_2 = 0, \tag{6}$$

$$\Lambda p[1-\beta(\Lambda D)] - \Lambda^2 D(R+cm_a)\beta'(\Lambda D) - \eta_3 = 0, \tag{7}$$

$$\eta_2 \Lambda (R - p + cm_a)\beta'(\Lambda D) - (k-1)\eta_3 = 0, \tag{8}$$

$$\eta_1\big((R-p+cm_a)[1-\beta(\Lambda D)]-cm_a\big)=0, \tag{9}$$

$$\eta_3(1-D)=0, \tag{10}$$

$$\eta_1,\eta_3\geq 0, \tag{11}$$

where η_1, η_2, and η_3 are the Lagrangian multipliers of the constraints 1, 2, and 3 of the best response problem of agent-ℓ, respectively.

After some algebra the above FOC reduces to the following conditions:

$$D=\frac{\min\{\lambda^{mon},\Lambda/k\}}{\Lambda},\ p=R+cm_a-\frac{cm_a}{1-\beta(\min\{\lambda^{mon},\Lambda/k\})}\Leftrightarrow \Lambda\geq\lambda^0$$
$$D=1/k,\ p=(R+c)-\frac{(R+c)(k-1)}{\frac{k}{1-\frac{\Lambda/k\beta'(\Lambda/k)}{1-\beta(\Lambda/k)}}-1}\Leftrightarrow \Lambda<\lambda^0.$$

Sufficient conditions for the symmetric equilibrium: We establish the existence of the symmetric SPNE when $\beta(\lambda)$ is concave as presented in the following lemma.

LEMMA 3. *Let*

$$p=\begin{cases} R+cm_a-\frac{cm_a}{1-\beta(\min\{\lambda^{mon},\Lambda/k\})} & \text{if } \Lambda\geq\lambda^0\\ (R+cm_a)-\frac{(R+cm_a)(k-1)}{\frac{k}{1-\frac{\Lambda/k\beta'(\Lambda/k)}{1-\beta(\Lambda/k)}}-1} & \text{if } \Lambda<\lambda^0,\end{cases}$$

and suppose $\beta(\lambda)$ is concave. Then, we have that $p_\ell(p)=p$, i.e. the best response of a single agent is p.

Appendix B: Proofs in Section 5

B.1. Proof of Proposition 2

Before proving the proposition, we state the following lemma which we use to prove the result.

LEMMA 4. *Consider two convergent sequences of non-negative numbers a^k and b^k such that $\lim_{k\to\infty} a^k=\tilde{a}$, $a^k k\leq a^{k+1}(k+1)$, and $\lim_{k\to\infty} b^k=\tilde{b}$. Then, we have that*

$$\lim_{k\to\infty}\left(1-\beta^M\left(b^k k;a^k k\right)\right)=\frac{1}{\max\{\tilde{b}\rho/\tilde{a},1\}}.$$

We prove Proposition 2 through a case-by-case analysis focusing two cases: 1) $\mathbf{p}<\mathbf{R}$ and 2) $\mathbf{p}=\mathbf{R}$.

Case-1 ($\mathbf{p}<\mathbf{R}$): To prove our claim in this case, we first argue that $\liminf_{k\to\infty} D_1^{MCE}(p^k;k)\geq \min\left\{1,\frac{R-p+cm_a}{\rho cm_a}\right\}$. Suppose on the contrary that the result does not hold. Then, there exists a convergent subsequence of $D_1^{MCE}(p^k;k)$, say $D_1^{MCE}(p^k;k)$ (we do not use a new notation for the subsequence for notational convenience), such that

$$\lim_{k\to\infty} D_1^{MCE}(p^k;k)=\liminf_{k\to\infty} D_1^{MCE}(p^k;k)<\min\left\{1,\frac{R-p+cm_a}{\rho cm_a}\right\},$$

since $D_1^{MCE}(p^k;k)\in[0,1]$ for any $k=1,2,\ldots$. Let $\tilde{D}=\lim_{k\to\infty} D_1^{MCE}(p^k;k)$. Then using the fact that system behaves as a multi-server queue when all the agents charge the same price, we have that

$$\begin{aligned}\lim_{k\to\infty} U_1(D_1^{MCE}(p^k;k);p^k;k) &= \lim_{k\to\infty}(R-p^k+cm_a)\left(1-\beta^M(\Lambda^k D_1^{MCE}(p^k;k);k)\right)-cm_a\\ &=\frac{R-p+cm_a}{\max\{\tilde{D}\rho,1\}}-cm_a>cm_a-cm_a=0,\end{aligned}$$

where the equality holds by Lemma 4 and the last inequality holds since $\tilde{D}(p) < \min\left\{1, \frac{R-p+cm_a}{\rho cm_a}\right\}$ and $p < R$. Therefore, there exists a K^* such that for any $k > K^*$, we have $U_1(D_1^{MCE}(p^k;k);p^k;k) > 0$ whereas $D_1^{MCE}(p^k;k) < 1$. However, this contradicts with the definition of Market Customer Equilibrium.

We now argue that $\limsup_{k\to\infty} D_1^{MCE}(p^k;k) \le \min\left\{1, \frac{R-p+cm_a}{\rho cm_a}\right\}$. To do this it is sufficient to show $\limsup_{k\to\infty} D_1^{MCE}(p^k;k) \le \frac{R-p+cm_a}{\rho cm_a}$ since $D_1^{MCE}(p^k;k) \le 1$ for any k. Suppose on the contrary that the result does not hold. Then, there exists a convergent subsequence of $D_1^{MCE}(p^k;k)$, say $D_1^{MCE}(p^k;k)$, such that

$$\lim_{k\to\infty} D_1^{MCE}(p^k;k) = \limsup_{k\to\infty} D_1^{MCE}(p^k;k) > \frac{R-p+cm_a}{\rho cm_a},$$

since $D_1^{MCE}(p;k) \in [0,1]$ for any $k = 1, 2, \ldots$.

Let $\tilde{D} = \lim_{k\to\infty} D_1^{MCE}(p^k;k)$. Then, observe that

$$\begin{aligned}\lim_{k\to\infty} U_1(D_1^{MCE}(p^k;k);p^k;k) &= \lim_{k\to\infty} (R - p^k + cm_a)\left(1 - \beta^M(\Lambda^k D_1^{MCE}(p^k;k),k)\right) - cm_a \\ &= \frac{R-p+cm_a}{\max\{\tilde{D}\rho, 1\}} - cm_a < cm_a - cm_a = 0,\end{aligned}$$

where the equality holds by Lemma 4 and the last inequality holds since $\rho\tilde{D}(p) > \frac{R-p+cm_a}{cm_a} \ge 1$. Therefore, there exists a K^* such that for any $k > K^*$, we have $U_1(D_1^{MCE}(p^k;k);p^k;k) < 0$. However, this contradicts with the definition of Market Customer Equilibrium since $D_1^{MCE}(p^k;k) > 0$ for large k.

Once we establish that $\lim_{k\to\infty} D_1^{MCE}(p^k;k) = \min\left\{1, \frac{R-p+cm_a}{\rho cm_a}\right\}$, we have that $\lim_{k\to\infty}[1-\beta^M(\Lambda^k D_1^{MCE}(p^k;k);k)] = \frac{1}{\max\left\{\rho\min\left\{1,\frac{R-p+cm_a}{\rho cm_a}\right\},1\right\}}$ by Lemma 4. Finally, combining these two, we have that $\lim_{k\to\infty} V_1(D_1^{MCE}(p^k;k);p^k;k) = p\min\{\rho, 1\}$.

Case-2 ($\mathbf{p = R}$): Note that we do not use the condition $p < R$ to argue that $\limsup_{k\to\infty} D_1^{MCE}(p^k;k) \le \min\left\{1, \frac{R-p+cm_a}{\rho cm_a}\right\}$. Therefore, $\min\left\{1, \frac{R-p+cm_a}{\rho cm_a}\right\}$ is also the upper-bound for the fraction of customers requesting service in the limit when $p = R$. As a direct implication of that the upper-bound for the revenues of the agents in the limit is $p\min\{\rho, 1\}$.

B.2. Proof of Proposition 3

The following lemma establishes the utility of a single agent when he decreases his price while the remaining agents do not change their price.

LEMMA 5. *For any $\lambda_1 + \lambda_2 \le \Lambda$, and $p' < p$, we have that $\sigma_2(\lambda_1, \lambda_2; p, p'; k-1, 1) \ge \frac{\lambda_1+\lambda_2}{1+\lambda_1+\lambda_2}$.*

We start proving the proposition by considering the case $p < R$. Note that $\lim_{k\to\infty} D_1^{MCE}(p^k;k) > 0$ when $p < R$ by Proposition 2. Thus, for the case where $p < R$, we need to show that $\lim_{k\to\infty} V'(p^k;k) = p$. Note that this statement is trivially true when $p = 0$. In order to prove Proposition 3 for $p > 0$, we consider a deviation by a single agent where he decreases his price by an arbitrary small amount $\varepsilon > 0$.

Let $D_{pool}(k) = D_1^{MCE}(p^k, p-\varepsilon; k-1, 1)$ and $D_{one}(k) = D_2^{MCE}(p^k, p-\varepsilon; k-1, 1)$. We first argue that $\liminf_{k\to\infty} D_{pool}(k) + D_{one}(k) \ge \min\{1, 1/\rho\}$ for any $p < R$. We prove this claim by contradiction, so that we suppose $\liminf_{k\to\infty} D_{pool}(k) + D_{one}(k) < \min\{1, 1/\rho\}$. Then, there should exist convergent subsequences of $D_{pool}(k)$

and $D_{one}(k)$ such that $\lim_{k\to\infty} D_{pool}(k)+D_{one}(k)<\min\{1,1/\rho\}$. Using this observation, and letting $P_{one}(k) = PServ_{12}\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big)$ for notational convenience, we have that

$$\begin{aligned}
&\lim_{k\to\infty} U_1\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big)\\
&\quad = \Big(1-\lim_{k\to\infty} P_{one}(k)\Big)\Big[(R-p+cm_a)\big[1-\lim_{k\to\infty}\beta_1\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big)\big]-cm_a\Big]\\
&\qquad +(R-p+\varepsilon)\lim_{k\to\infty} P_{one}(k)\\
&\quad \geq (R-p+cm_a)\big[1-\lim_{k\to\infty}\beta_1\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big)\big]-cm_a\\
&\quad \geq (R-p+cm_a)\big[1-\lim_{k\to\infty}\beta^M(\Lambda^k D_{pool}(k); k-1)\big]-cm_a = R-p>0,
\end{aligned}$$

where the second inequality holds since some customers choosing sub-pool-1 may be served by sub-pool-2, and the last equality holds since $\lim_{k\to\infty}\frac{\Lambda^k D_{pool}(k)}{k-1}<\min\{1,\rho\}\leq 1$. However, this contradicts with the definition of the customer equilibrium since we suppose $\lim_{k\to\infty} D_{pool}(k)+D_{one}(k)<1$, i.e. some customers choose not to request service for sufficiently large k. Hence, we should have that $\liminf_{k\to\infty} D_{pool}(k)+D_{one}(k)\geq\min\{1,1/\rho\}$.

Then using the fact that $\liminf_{k\to\infty} D_{pool}(k)+D_{one}(k)\geq\min\{1,1/\rho\}$, we have that

$$\begin{aligned}
\liminf_{k\to\infty} V_2\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big) &= (p-\varepsilon)\liminf_{k\to\infty}\sigma_2\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big)\\
&\geq (p-\varepsilon)\lim_{k\to\infty}\frac{\Lambda^k\big(D_{pool}(k)+D_{one}(k)\big)}{1+\Lambda^k\big(D_{pool}(k)+D_{one}(k)\big)} = p-\varepsilon,
\end{aligned}$$

where the inequality holds by Lemma 5. Note that the revenue of a single agent after the deviation we propose is less than the optimal deviation $V'(p^k;k)$, and thus we have that

$$\liminf_{k\to\infty} V'(p^k;k)\geq\liminf_{k\to\infty} V_2\big(D_{pool}(k), D_{one}(k); p^k, p-\varepsilon; k-1, 1\big)\geq p-\varepsilon.$$

Finally, our claim holds since ε can be arbitrarily small and $V'(p^k;k)\leq p^k$ by construction.

Now, we consider the case where $p=R$ and $\liminf_{k\to\infty} D_1^{MCE}(p^k;k)=\tilde{D}>0$ (Note that $\rho\tilde{D}\leq 1$ by Proposition 2). This time, we will show that $\liminf_{k\to\infty} D_{pool}(k)+D_{one}(k)\geq\tilde{D}$. As above, we assume the contrary. Then, there exists a subsequence of $D_{pool}(k)$ such that $\liminf_{k\to\infty} D_{pool}(k)=\tilde{D}_{pool}<\tilde{D}$. Then, we have that

$$\begin{aligned}
U_1\big(D_{pool}(k), D_{one}(k); p^k, p^k-\varepsilon; k-1, 1\big) &\geq (R-p^k+cm_a)\big[1-\beta^M(\Lambda^k D_{pool}(k); k-1)\big]-cm_a,\\
&> (R-p^k+cm_a)\big[1-\beta^M(\Lambda^k D_1^{MCE}(p^k;k); k)\big]-cm_a\geq 0,
\end{aligned}$$

for large k, where the strict inequality holds since $\beta^M(\Lambda^k D_{pool}(k); k-1)\simeq\zeta_1(\rho\tilde{D}_{pool}e^{1-\rho\tilde{D}_{pool}})^k$ and $\beta^M(\Lambda^k D_1^{MCE}(p^k;k);k)\simeq\zeta_2(\rho\tilde{D}e^{1-\rho\tilde{D}})^k$ for some constants ζ_1 and ζ_2 by Theorem 5 in Zeltyn and Mandelbaum (2005) and the fact that $\rho\tilde{D}\leq 1$. Note that $\frac{(\rho\tilde{D}_{pool}e^{1-\rho\tilde{D}_{pool}})^k}{(\rho\tilde{D}e^{1-\rho\tilde{D}})^k}\to 0$ as $k\to 0$ since $\tilde{D}_{pool}<\tilde{D}$ and xe^{1-x} is strictly increasing in x for any $x<1$. However, this contradicts with the definition of the customer equilibrium since we suppose $\lim_{k\to\infty} D_{pool}(k)+D_{one}(k)<1$, i.e. some customers choose not to request service for sufficiently large k. Hence, we should have that $\liminf_{k\to\infty} D_{pool}(k)+D_{one}(k)\geq\tilde{D}$. Using this result, we can again show that the utilization of the deviating agent will converge to one, and thus his revenue will be $R-\varepsilon$.

Finally, we need to show that $\lim_{k\to\infty} V'(p^k;k)>0$ when $p=R$ and $\liminf_{k\to\infty} D_1^{MCE}(p^k;k)=0$. Consider a deviation where a single agent cuts his price and charge $R/2$. Let, $\hat{\lambda}$ solves $(R/2+cm_a)[1-\beta(\lambda)]=cm_a$. There exists such $\hat{\lambda}$ since $\beta(\lambda)$ is increasing in λ and $\lim_{\lambda\to\infty}\beta(\lambda)=1$. Then, by construction $\Lambda^k\big(D_{pool}(k)+D_{one}(k)\big)\geq\hat{\lambda}$ because otherwise the customer choosing the deviating agent would earn a strictly positive utility while $D_{pool}(k)+D_{one}(k)<1$ for large k, and that would be a contradiction. Therefore, using Lemma 5, we have that $\lim_{k\to\infty} V'(p^k;k)\geq R/2\frac{\hat{\lambda}}{1+\hat{\lambda}}>0$.

B.3. Proof of Theorem 2

1. To prove our claim, it is sufficient to show that $\limsup_{k\to\infty} p_{EQ}^k = 0$ because for any $\xi > \limsup_{k\to\infty} p_{EQ}^k$, there is a K such that $p_{EQ}^k < \xi$ for all $k > K$ by Theorem 3.17 in Rudin (1976). We prove that $\limsup_{k\to\infty} p_{EQ}^k = 0$ by contradiction. Thus, we suppose that $\limsup_{k\to\infty} p_{EQ}^k > 0$. Then, there should exist a convergent subsequence of p_{EQ}^k such that $\lim_{k\to\infty} p_{EQ}^k = \tilde{p} > 0$ since the equilibrium prices p_{EQ}^k are bounded from above by R. Let $V'(p_{EQ}^k;k) = \max_{0\le p'\le p_{EQ}^k} V_2(D_1^{MCE}(p_{EQ}^k, p'; k-1, 1), D_2^{MCE}(p_{EQ}^k, p'; k-1, 1); p_{EQ}^k, p'; k-1, 1)$. When $\rho < 1$ and $\lim_{k\to\infty} D_1^{MCE}(p_{EQ}^k;k) > 0$, we have that

$$\liminf_{k\to\infty} V'(p_{EQ}^k;k) = \tilde{p} > \rho\tilde{p} \ge \lim_{k\to\infty} V_1(D_1^{MCE}(p_{EQ}^k;k); p_{EQ}^k;k) + \lim_{k\to\infty} \epsilon^k,$$

by Proposition 3, and by the definition of ϵ^k. Then, for sufficiently large k, we should have that $V'(p_{EQ}^k;k) > V_1(D_1^{MCE}(p_{EQ}^k;k); p_{EQ}^k;k) + \epsilon^k$, which implies that p_{EQ}^k cannot emerge as the equilibrium price of a symmetric ϵ-Market Equilibrium for large k.

Similarly, when $\lim_{k\to\infty} D_1^{MCE}(p_{EQ}^k;k) = 0$ (and thus, $\lim_{k\to\infty} V_1(D_1^{MCE}(p_{EQ}^k;k); p_{EQ}^k;k) = 0$), we have that

$$\liminf_{k\to\infty} V'(p_{EQ}^k;k) > 0 = \lim_{k\to\infty} V_1(D_1^{MCE}(p_{EQ}^k;k); p_{EQ}^k;k) + \lim_{k\to\infty} \epsilon^k,$$

by again Proposition 3 and the fact that $\tilde{p} > 0$. Thus, for sufficiently large k, we should again have that $V'(p_{EQ}^k;k) > V_1(D_1^{MCE}(p_{EQ}^k;k); p_{EQ}^k;k) + \epsilon^k$, which implies that p_{EQ}^k cannot emerge as the equilibrium price of a symmetric ϵ-Market Equilibrium for large k.

2. To prove this claim, we suppose, on the contrary, that for any K, there exists a $k > K$ such that zero cannot be a symmetric equilibrium in the k^{th} market. Thus, there should be a sequence $\hat{p}^k$ such that a single agent can improve his revenue by increasing his price to $\hat{p}^k$ in the k^{th} marketplace. Let $U_{pool}(k)$ and $U_{dev}(k)$ be the utility of customers choosing price zero and $\hat{p}^k$, respectively. As we suppose that the deviating agent improves his revenue, strictly positive fraction of customers should pick him, and thus we should have that $U_{dev}(k) \ge U_{pool}(k)$ for any k. Using this observation we have that

$$(R-\hat{p}^k)[1-P_{12}(k)] + RP_{12}(k) \ge U_{dev}(k) \ge U_{pool}(k) \ge (R - cm_a)(1-\beta^M(\rho k; k-1)) - cm_a,$$

where $P_{12}(k)$ is the probability that a customer picking $\hat{p}^k$ is served by the agents charging zero in the k^{th} marketplace. The first inequality above holds since customers, who pick $\hat{p}^k$ and served by the deviating agents, may abandon, and the last inequality holds since agents charging zero may not serve all customers, and they give priority to their own customers. Since $\rho < 1$ and using Theorem 5.1 Zeltyn and Mandelbaum (2005), we have that $U_{pool}(k)$ converges to R with an exponential speed, i.e. there exists a constant ζ such that $U_{pool}(k) = R - e^{-\zeta k}$. Then, the above inequality implies that $\hat{p}^k[1-P_{12}(k)] \le e^{-\zeta k}$.

Note that the revenue of the agent deviating, say $V_{dev}(k)$, in the k^{th} marketplace is less than $\rho k p_k'[1-P_{12}(k)]$ since the rate of customer he can serve cannot be greater than $\rho k[1-P_{12}(k)]$. As a result of this observation, we have that

$$V_{dev}(k) \le \rho k e^{-\zeta k} \Rightarrow \frac{V_{dev}(k)}{\epsilon^k} \le \frac{\rho k e^{-\zeta k}}{\epsilon^k} \to 0 \ as\ k\to\infty,$$

where the convergence holds since $\epsilon^k\sqrt{k}\to\infty$ and $e^{\zeta k}k^{-3/2}\to\infty$ as $k\to\infty$. Since $\frac{V_{dev}(k)}{\epsilon^k}$ converges to zero, we should have that $V_{dev}(k)<\epsilon^k$ for large k, which contradicts with the fact that $\hat{p}^k$ is a profitable deviation. Hence, there should be a K such that zero emerges as the equilibrium price of a symmetric ϵ^k-Market Equilibrium for all $k>K$.

3 Let λ^0_k be the constant λ^0 defined in the k^{th} marketplace in the no-intervention model. Then, we have that

$$(R+cm_a)\frac{k-1}{k}-\frac{c}{1-\beta(\lambda^0_k/k)}\left(\frac{1-\beta(\lambda^0_k/k)}{1-\beta(\lambda^0_k/k)-\lambda^0_k/k\beta'(\lambda^0_k/k)}-\frac{1}{k}\right)=0.$$

Using that equation we have that $\lim_{k\to\infty}\lambda^0_k/k=\lambda^{mon}$. Thus, we have that

$$\lim_{k\to\infty}p^k_{NI}=\begin{cases}(R+c)\left[1-\frac{1-\beta(\rho)-\rho\beta'(\rho)}{1-\beta(\rho)}\right] & \text{if } \rho\le\lambda^{mon}\\ (R+c)\left[1-\frac{c}{(R+c)(1-\beta(\lambda^{mon}))}\right] & \text{if } \rho>\lambda^{mon},\end{cases}$$

where p^k_{NI} is the unique equilibrium price under no-intervention setting. Furthermore, we have that the utilization of a single agent in the no-intervention model converges to $\rho\big[1-\beta(\rho)\big]$ when $\rho\le\lambda^{mon}$ and $\lambda^{mon}\big[1-\beta(\lambda^{mon})\big]$ otherwise. Thus, it is clear that the revenue of an agent converges to a strictly positive limit, say $\bar{v}$, under the no-intervention model. Thus, there is a K_1 such that $\Pi^k_{NI}>\bar{v}/2$ for all $k>K_1$. Furthermore, by part 1, there is a K_2 such that revenue of an agent under the operational efficiency model is less than $\xi\bar{v}/2$ for all $k>K_2$. Thus, we have that $\frac{\Pi^k_{OE}}{\Pi^k_{NI}}<\frac{k\xi\bar{v}/2}{k\bar{v}/2}<\xi$ for all $k>\max\{K_1,K_2\}$.

B.4. Proof of Proposition 4

Before proving the Proposition 4, we state the following lemma, which we use in the proof.

LEMMA 6. *Let $D^{MCE}_n(k)=D^{MCE}_n(\hat{p}^k,p^k;1,k-1)$ for $n=1,2$ in the k^{th} marketplace where $\hat{p}^k>p^k$, $\lim_{k\to\infty}p^k=p$, $\lim_{k\to\infty}\hat{p}^k=\hat{p}$, $p\le\hat{p}<R$ and $\rho>1$. Then, the following statements are true.*

1. $\liminf_{k\to\infty}\frac{\Lambda^k D^{MCE}_2(k)}{k-1}>1.$
2. $\lim_{k\to\infty}PServ_{12}(D^{MCE}_1(k),D^{MCE}_2(k);\hat{p}^k,p^k;1,k-1)=0.$
3. *Suppose $\tilde{\lambda}=\lim_{k\to\infty}\Lambda^k D^{MCE}_1(k)$, then we have that*

$$\lim_{k\to\infty}\beta_1(D^{MCE}_1(k),D^{MCE}_2(k);\hat{p}^k,p^k;1,k-1)=\beta^M(\tilde{\lambda};1)$$

4. *There exists a K^* such that $\Lambda^k D^{MCE}_1(k)\le\bar{\lambda}$ for any $k>K^*$, where $\bar{\lambda}<\infty$ is the unique solution to $1-\beta^M(\lambda;1)=\frac{cm_a}{R+cm_a}$.*
5. *Suppose $\tilde{D}=\lim_{k\to\infty}D^{MCE}_2(k)$, then we have that*

$$\lim_{k\to\infty}\beta_2(D^{MCE}_1(k),D^{MCE}_2(k);\hat{p}^k,p^k;1,k-1)=1-\frac{1}{\rho\tilde{D}}.$$

6. $\lim_{k\to\infty}D^{MCE}_2(k)=\min\left\{1,\frac{R-p+cm_a}{\rho cm_a}\right\}.$

The proof of this lemma can be seen in the Supporting Document.

Consider the following problem

$$\Pi(p)=\max_{p\le p'\le R,\lambda\ge0}\quad p'\lambda[1-\beta(\lambda)]$$

$$s.to \quad (R - p' + cm_a)[1 - \beta(\lambda)] \geq \Delta(p;R) + cm_a$$

One can easily show that $\Pi(p) = \max_{\lambda} \; (R + cm_a)\lambda[1 - \beta(\lambda)] - \lambda[\Delta(p;R) + cm_a]$. Then, by using the FOC, we have that $\Pi(p) = (R + cm_a)\lambda^{\Delta}(p;R)[1 - \beta(\lambda^{\Delta}(p;R))] - \lambda^{\Delta}(p;R)(\Delta(p;R) + cm_a)$, where $\lambda^{\Delta}(p;R)$ solves $1 - \beta(\lambda) - \lambda\beta'(\lambda) = \frac{\Delta(p;R)+cm_a}{R+cm_a}$.

Now, we are ready to prove Proposition 4. For notational convenience, let $\hat{p}(k) = \arg\max_{p^k \leq p' \leq R} V_1(D_1^{MCE}(p', p^k; 1, k-1), D_2^{MCE}(p', p^k; 1, k-1); p', p^k; 1, k-1)$, $\lambda(k) = \Lambda^k D_1^{MCE}(\hat{p}(k), p^k; 1, k-1)$, and $D_{pool}(k) = D_2^{MCE}(\hat{p}(k), p^k; 1, k-1)$. Observe that for any $k = 1, 2, \ldots$, $\hat{p}(k) \in [0, R]$, $\lambda(k) \in [0, \overline{\lambda}]$, $D_{pool}(k) \in [0,1]$, $V'(p^k;k) \in [0, R]$, where $\overline{\lambda}$ is defined as in Lemma 6. Then, there exists a convergent subsequence of $V'(p^k;k)$, which is denoted by $V'(p^k;k)$ (for notational convenience), such that $\lim_{k\to\infty} V'(p^k;k) = \limsup_{k\to\infty} V'(p^k;k)$. Moreover, $\hat{p}(k)$, $\lambda(k)$, $D_{pool}(k)$ have convergent subsequences, and we let $\tilde{p}^{dev} = \lim_{k\to\infty} \hat{p}(k)$, $\tilde{\lambda} = \lim_{k\to\infty} \lambda(k)$, and $\tilde{D}_{pool} = \lim_{k\to\infty} D_{pool}(k)$.

Then, by the continuity of $\beta(\lambda)$ and definition of Market Customer Equilibrium, we have that

$$\begin{aligned}(R - \tilde{p}^{dev} + cm_a)[1 - \beta(\tilde{\lambda})] &= \lim_{r\to\infty} (R - \hat{p}(k) + cm_a)[1 - \beta(\lambda(k))] = \lim_{k\to\infty} U_1\big(\lambda(k)/\Lambda^k, D_{pool}(k); \hat{p}(k), p^k; 1, k-1\big) + cm_a \\ &= \lim_{k\to\infty} U_2\big(\lambda(k)/\Lambda^k, D_{pool}(k); \hat{p}(k), p^k; 1, k-1\big) + cm_a = \frac{(R - p + cm_a)}{\rho\tilde{D}_{pool}} = \Delta(p;R) + cm_a,\end{aligned}$$

where the second equality follows by Lemmas 6.2 and 6.3, and the last two equalities holds by Lemmas 6.5 and 6.6. Therefore, $(\tilde{p}^{dev}, \tilde{\lambda})$ satisfy the constraint in the limit problem. And, this implies that

$$\limsup_{k\to\infty} V'(p^k;k) = \tilde{p}^{dev}\tilde{\lambda}\big(1 - \beta(\tilde{\lambda})\big) \leq \Pi(p).$$

B.5. Proof of Theorem 3

We first show that when $\rho > 1$, a single provider, who cuts his price, cannot improve his revenue by more than ϵ^k for large enough k. Thus, the only possible profitable deviation for a single agent is to increase his price for large enough, yet finite, k. To argue that let $V^{cut}(p^k;k) = \max_{0 \leq p' \leq p^k} V_2(D_1^{MCE}(p^k, p'; k-1, 1), D_2^{MCE}(p^k, p'; k-1, 1); p^k, p'; k-1, 1)$ for any given sequence p^k with limit $p < R$. Note that $\rho \lim_{k\to\infty} D_1^{MCE}(p;k) > 1$ by Proposition 2. Therefore, using Theorem 6.1 in Zeltyn and Mandelbaum (2005), we have that the revenue of an agent converges to p exponentially as $k \to \infty$ when all agents charge p^k in a seller's market, i.e there exists a constant ζ such that $V_1(D_1^{MCE}(p^k;k); p^k;k) = p^k(1 - e^{-\zeta k})$. Using this observation, for large enough k, we have that $V_1(D_1^{MCE}(p^k;k); p^k;k) = p^k(1 - e^{-\zeta k}) \geq p^k - Re^{-\zeta k} > p^k - \epsilon^k \geq V^{cut}(p;k) - \epsilon^k$. The second inequality holds for large enough k because $\lim_{k\to\infty} \frac{e^{-\zeta k}}{\epsilon^k} = 0$ by our assumption of $\lim_{k\to\infty} \epsilon^k\sqrt{k} = \infty$ and the fact that $\lim_{k\to\infty} e^{\zeta k}/\sqrt{k} = \infty$. This implies that a single agent cannot have a profitable deviation by decreasing his price in large marketplaces. Hence, in order to verify that any sequence of prices p^{*k} with limit $p^* \in \mathcal{P}(\rho;R)$ can emerge as an equilibrium outcome of a symmetric ϵ^k-Market Equilibrium, it is sufficient to check any single agent deviation where the agent increases his price.

Let $V'(p^{*k};k) = \max_{p^{*k} \leq p' \leq R} V_1(D_1^{MCE}(p', p^{*k}; 1, k-1), D_2^{MCE}(p', p^{*k}; 1, k-1); p', p^{*k}; 1, k-1)$. Since $p^* \in \mathcal{P}(\rho;R)$, we have that

$$\lim_{k\to\infty} V_1(D_1^{MCE}(p^{*k};k); p^{*k};k) \;=\; p^* > (R + cm_a)\lambda^{\Delta}(p^*;R)[1 - \beta(\lambda^{\Delta}(p^*;R))] - \lambda^{\Delta}(p^*;R)(\Delta(p^*;R) + cm_a)$$

$$\geq \limsup_{k\to\infty} V'(p^{*k};k),$$

where the last inequality holds by Proposition 4. Then, there should exist a K such that $V'(p^{*k};k) < V_1(D_1^{MCE}(p^{*k};k);p^{*k};k)$ for all $k > K$, which implies that p^{*k} can emerge as the equilibrium price of a symmetric ϵ^k-Market Equilibrium for all $k > K$.

Monotonicity of $\mathcal{P}(\rho;R)$: Let $\Pi(p,\rho) = \max_\lambda (R+cm_a)\lambda[1-\beta(\lambda)] - \lambda[\Delta(p;R)+cm_a]$. We first want to note that $\Pi(p,\rho)$ is increasing in ρ for all p since $\Delta(p;R)$ is decreasing in ρ. Now, suppose $\rho_1 > \rho_2$ and $p \in \mathcal{P}(\rho_1;R)$. Then, we have that $p \in \mathcal{P}(\rho_1;R) \Rightarrow\ p > \Pi(p,\rho_1) \geq \Pi(p,\rho_2) \Rightarrow\ p \in \mathcal{P}(\rho_2;R)$, where the inequality holds since $\Pi(p,\rho)$ is increasing in ρ. Hence, we have that $\mathcal{P}(\rho_1;R) \subseteq \mathcal{P}(\rho_2;R)$.

B.6. Proof of Corollary 1

1. Using the birth-death chain analysis of the corresponding analysis, one can show that $\beta(\lambda) = \frac{\lambda e^\lambda - e^\lambda + 1}{\lambda e^\lambda}$ when $m_a = 1$. Then, $\lambda^\Delta(p;R)$ solves

$$e^{-\lambda} = \frac{\Delta(p;R)+c}{R+c},$$

which implies that $\lambda^\Delta(p;R) = \log\left(\frac{R+c}{\Delta(p)+c}\right)$. Once we have $\lambda^\Delta(p;R)$, $\mathcal{P}(\rho;R)$ can be written as in the corollary immediately.

2. As we showed in the proof of Theorem 2.3 Let λ_k^0 be the constant λ^0 defined in the k^{th} marketplace. Then, we have that

$$(R+cm_a)\frac{k-1}{k} - \frac{c}{1-\beta(\lambda_k^0/k)}\left(\frac{1-\beta(\lambda_k^0/k)}{1-\beta(\lambda_k^0/k)-\lambda_k^0/k\beta'(\lambda_k^0/k)} - \frac{1}{k}\right) = 0.$$

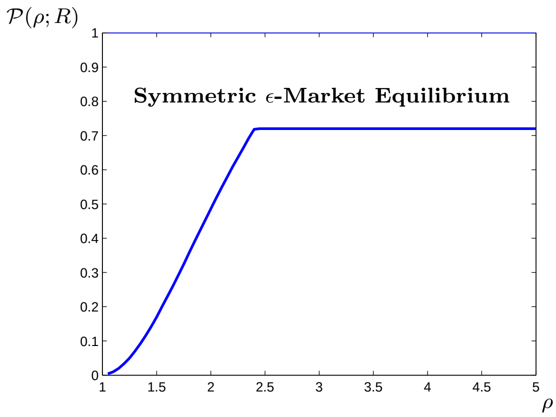
Using that equation we have that $\lim_{k\to\infty} \lambda_k^0/k = \lambda^{mon}$. Thus, we have that

$$\lim_{k\to\infty} p_{NI}^k = \begin{cases} (R+c)\left[1 - \frac{1-\beta(\rho)-\rho\beta'(\rho)}{1-\beta(\rho)}\right] & \text{if } \rho \leq \lambda^{mon} \\ (R+c)\left[1 - \frac{c}{(R+c)(1-\beta(\lambda^{mon}))}\right] & \text{if } \rho > \lambda^{mon} \end{cases}$$

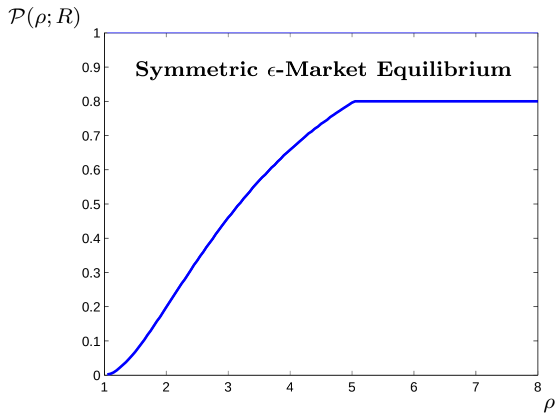
Furthermore, using the fact that $\beta(\lambda) = \frac{\lambda e^\lambda - e^\lambda + 1}{\lambda e^\lambda}$, we can obtain the result in the corollary.

B.7. Numerical Study to Derive $\mathcal{P}(\rho;R)$ for $m_a \neq \mu$

Figure 4 illustrates $\mathcal{P}(\rho;R)$ for the settings where $\mu \neq m_a$. We provide the graphs by fixing $c = 1$ but we carry out the numerical study for other values of c as well.



(a)



(b)

Figure 4 **Any price above the curve is in $\mathcal{P}(\rho;R)$, and thus it forms a symmetric ϵ-Market Equilibrium. For both examples, $R=1$, $c=0.1$, $\mu=1$. For (a), $m_a=2$. For (b), $m_a=0.5$.**

Appendix C: Proofs in Section 6

C.1. Proof of Theorem 4

We showed, in Theorem 2, that $p^k < \xi$ for large k for any $\xi > 0$ even we allowed for only single agent deviations. Thus, it is only necessary to argue that $p=0$ is an equilibrium price. In fact, the proof of such a claim is the same as the proof of Theorem 2.2: Suppose there is a sequence $\hat{p}^k$ such that δ^k fraction of agents can improve his revenue by increasing his price to $\hat{p}^k$ in the k^{th} marketplace while the remaining agents charge zero. Then, as in the proof of Theorem 2.2, we can show that the revenue of the deviating agents should converge to zero in an exponential rate since we assume $\delta^k \to 0$. However, ϵ^k converges to zero in a slower rate. Hence, $p=0$ should emerge as the equilibrium price of a symmetric (δ^k,ϵ^k)-Market Equilibrium for large k. It is worth noting that we could only show the existence of equilibrium for $\rho < 1-\tilde{\delta}$ if $\lim_{k\to\infty}\delta^k = \tilde{\delta}$ for some $\tilde{\delta} > 0$.

C.2. Supplementary Claims for the Proof of Theorem 5

Before proving the theorem, we first state the following proposition. The proof of this proposition can be seen in the technical appendix. This proposition simply proves that a group of agents can improve their revenues by increasing their prices in a seller's market when they are allowed to deviate together.

Lemma 7. *Let $D_n^{MCE}(k) = D_n^{MCE}(\hat{p}^k, p^k; \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor)$ for $n=1,2$ in the k^{th} marketplace, where $\lim_{k\to\infty} p^k = p$, $\lim_{k\to\infty}\hat{p}^k = p'$, $p<p'<R$ and $\rho>1$. Then, the following statements are true.*

1. $\liminf_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k-\lfloor\delta^k k\rfloor} > 1.$
2. $\lim_{k\to\infty} PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor) = 0.$
3. $\lim_{k\to\infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor) = \min\left\{0, 1-\frac{1}{\rho\tilde{D}_1}\right\}$, *when* $\tilde{D}_1 = \lim_{k\to\infty}\frac{D_1^{MCE}(k)}{\delta^k}$.
4. *Suppose* $\tilde{D}_2 = \lim_{k\to\infty} D_2^{MCE}(k)$ *and* $\tilde{\delta} = \lim_{k\to\infty}\delta^k$, *then we have that*

$$\lim_{k\to\infty} \beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor) = 1-\frac{1-\tilde{\delta}}{\rho\tilde{D}_2}.$$

The proof of this lemma can be seen in the Supporting Document.

PROPOSITION 10. *In a seller's market ($\rho > 1$), we have that* $\liminf_{k\to\infty} \frac{D_1^{MCE}(k)}{\delta^k} \geq \frac{1}{\rho}$*, where* $D_n^{MCE}(k) = D_n^{MCE}(\hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor)$*,* $\lim_{k\to\infty} p^k = p < R$*,* $\lim_{k\to\infty} \hat{p}^k = p'$*, and* $p < p' < \min\left\{R, p + \left(1 - \frac{1}{\rho}\right)(R - p + cm_a)\right\}$*. Furthermore, we have that* $\lim_{k\to\infty} V_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) = p'$*.*

Proof:

Similar to the proofs before, we prove our claim by contradiction. Hence, we suppose that $\liminf_{k\to\infty} \frac{D_1^{MCE}(k)}{\delta^k} < \frac{1}{\rho}$. Then, there exists a convergent subsequence of $D_1^{MCE}(k)$ such that $\tilde{D}_1 = \lim_{k\to\infty} \frac{D_1^{MCE}(k)}{\delta^k} < \frac{1}{\rho}$.

For notational convenience, we let

$$\begin{aligned}P_\delta(k) &= PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor)\\ \beta_\delta(k) &= \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor).\end{aligned}$$

Then, using the fact that $\tilde{D}_1 < 1/\rho$, Lemma 7.2, and Lemma 7.3, the limit of the expected utility of customers choosing the providers charging $\hat{p}^k$ can be written as

$$\begin{aligned}&\lim_{k\to\infty} U_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor)\\ &\quad = \left(1 - \lim_{k\to\infty} P_\delta(k)\right)\left((R - p' + cm_a)\big[1 - \lim_{k\to\infty} \beta_\delta(k)\big] - cm_a\right) + \lim_{k\to\infty}(R - p^k)P_\delta(k)\\ &\quad = R - p' > 0. \end{aligned} \tag{12}$$

which implies that the expected utility of customers choosing the price $\hat{p}^k$ is strictly positive for large k, so that we should have $\lim_{k\to\infty} D_1^{MCE}(k) + D_2^{MCE}(k) = 1$ by the definition of customer equilibrium. Moreover, using the fact that $\tilde{D}_1 < 1/\rho$, we have that $\lim_{k\to\infty} D_2^{MCE}(k) \geq 1 - \tilde{\delta}/\rho$ where $\tilde{\delta} = \lim_{k\to\infty} \delta^k$. Then, using Lemma 7.3, the limit of the expected utility of customers choosing the providers charging p^k can be written as

$$\begin{aligned}&\lim_{k\to\infty} U_2(D_1^{MCE}(k),\ D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor)\\ &\quad = (R - p + cm_a)\big[1 - \lim_{k\to\infty} \beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor)\big] - cm_a\\ &\quad \leq (R - p + cm_a)\frac{1 - \tilde{\delta}}{\rho - \tilde{\delta}} - cm_a. \end{aligned} \tag{13}$$

Combining (12) and (13), we have that

$$\begin{aligned}\lim_{k\to\infty} U_2(D_1^{MCE}(k),\ D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) &\leq (R - p + cm_a)\frac{1 - \tilde{\delta}}{\rho - \tilde{\delta}} - cm_a \leq (R - p + cm_a)\frac{1}{\rho} - cm_a\\ &< R - p' = \lim_{k\to\infty} U_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor),\end{aligned}$$

which implies that customers strictly prefer providers charging $\hat{p}^k$ over the ones charging p^k for large k. However, this contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_2^{MCE}(k) > 1 - \tilde{\delta}/\rho > 0$. Therefore, we should have that $\liminf_{k\to\infty} \frac{D_1^{MCE}(k)}{\delta^k} \geq \frac{1}{\rho}$

Furthermore, using the fact that $\liminf_{k\to\infty} \frac{D_1^{MCE}(k)}{\delta^k} \geq \frac{1}{\rho}$, Lemma 7.2, and Lemma 7.3, we have that

$$\liminf_{k\to\infty} V_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) = \liminf_{k\to\infty} p' \rho D_1^{MCE}(k)\left[1 - P_\delta(k)\right]\left[1 - \beta_\delta(k)\right] \geq p'.$$

Finally, the result about the profit of the providers charging $\hat{p}^k$ holds since the revenue of an agent cannot exceed the price he charges , i.e. $V_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) \leq \hat{p}^k$.

■

C.3. Proof of Theorem 5

To prove our claim on the equilibrium prices, it is sufficient to show that $\liminf_{k\to\infty} p^k_{EQ} = R$. We prove this by supposing on the contrary that the result does not hold. Then, we can find a convergent subsequence of equilibrium prices p^k_{EQ} such that $\lim_{k\to\infty} p^k_{EQ} = p < R$. Then, using Proposition 10, agents can improve their utility by increasing their price to p' in a marketplace with sufficiently large number of agents. However, this contradicts with the fact that p^k_{EQ} is a (δ^k, ϵ^k)-Market Equilibrium. Hence, for any given $\xi > 0$, there should exist a large K such that $p^k_{EQ} > R - \xi$ for any $k > K$.

To prove our claim on the customer equilibrium, it is sufficient to show that $\liminf_{k\to\infty} D_1^{MCE}(p^k; k) \geq 1/\rho$. We prove this by supposing on the contrary that the result does not hold. Then, there is a subsequence of equilibrium prices p^k such that $\lim_{k\to\infty} D_1^{MCE}(p^k; k) = \tilde{D} < 1/\rho$. Furthermore, note that $\lim_{k\to\infty} p^k = R$ by our claim on prices. Then, by Proposition 3, the limiting revenue of a single agent cutting his price is R if $\tilde{D} > 0$ and strictly positive otherwise whereas his revenue before the deviation is at most $R\rho\tilde{D} < R$. This implies that the revenue of an agent cutting his price is strictly greater than his revenue before the deviation for large k. This contradicts with the definition of equilibrium, and thus, we should have that $\liminf_{k\to\infty} D_1^{MCE}(p^k; k) \geq 1/\rho$.

To show the existence of the equilibrium sequence, let $p^k = R + cm_a - \frac{cm_a}{1-\beta^M(k;k)}$. By construction, we have that $D_1^{MCE}(p^k; k) = 1/\rho$, and thus the revenue of agents charging p^k, say V^k, is $R - (R + cm_a)\beta^M(k; k)$. By Zeltyn and Mandelbaum (2005) Theorem 5, V^k converges to R with a rate of $1/\sqrt{k}$, i.e. there exists a constant ζ such that $V^k = R - \zeta/\sqrt{k}$. Then, using the fact that $\lim_{k\to\infty} \frac{1}{\epsilon^k\sqrt{k}} = 0$, there exists a K such that $\epsilon^k > \zeta/\sqrt{k}$ for all $k > K$ which implies that $V^k + \epsilon^k > R$ for all $k > K$. Since agents cannot obtain a revenue strictly greater than R after a deviation, the proposed sequence is clearly a (δ^k, ϵ^k)-Market Equilibrium for large k.

C.4. Proof of Proposition 5

Let $D_{info}(p^k; k)$ be the fraction of customers requesting service when all agents charge p^k with limit $p < R$ in a marketplace where the moderating firm provides real-time congestion information. We first argue that $\lim_{k\to\infty} D_{info}(p^k; k) > 1/\rho$ when $\rho > 1$. The proof is very similar to the proof of Proposition 7: We suppose $\lim_{k\to\infty} D_{info}(p^k; k) \leq 1/\rho$. Then, we would have that $\lim_{k\to\infty} \beta^M(\Lambda^k D_{info}(p^k; k); k) = 0$, which implies that $\lim_{k\to\infty} \beta^{info}(\Lambda^k D_{info}(p; k); k) = 0$. As the customers would not abandon, they would not wait as well, and thus the expected utility of each customer would be arbitrarily close to $R - p > 0$ as k grows. However, this would be a contradiction because all customers should request service when the expected utility of customers requesting service is strictly positive.

Once we establish that $\lim_{k\to\infty} D_{info}(p^k, k) > 1/\rho$, we have that the limits of the revenue of the agents charging p^k when the firm provides real-time information, say $V_{info}(p^k; k)$, can be written as follows:

$$\lim_{k\to\infty} V_{info}(p^k; k) = p \lim_{k\to\infty} \sigma_{info}(p^k; k) > p(1-\zeta) \lim_{k\to\infty} \sigma^M(p^k; k) = p(1-\zeta).$$

Our claim in this proposition is equivalent to claiming that $\limsup_{k\to\infty} p^k_{info} \geq R(1-\zeta)$. We show this by contradiction. Thus, we suppose $\limsup_{k\to\infty} p^k_{info} < R(1-\zeta)$. Then, there should exist a sequence of prices

p^k_{info} such that $p^k_{info} < R(1-\zeta) - \xi$ for all k. Then, there must be a convergent subsequence p^k_{info} with $\lim_{k\to\infty} p^k_{info} < R(1-\zeta)$. Let $p = \lim_{k\to\infty} p^k_{info}$. Then, consider a deviation where all agents charge $p' > p/(1-\zeta)$ (Such a deviation is possible since $p/(1-\zeta) < R$). As we show above, the revenue of agents will be arbitrarily close to $p'(1-\zeta)$ which is strictly greater than p. Note that agents revenue cannot be higher than p when everybody charges p. Therefore, for large k, we have that agents improve their profits by raising their prices to p'. However, this contradicts with the definition of equilibrium. Hence, for any $\xi > 0$, we should have that $p^k_{info} \geq R(1-\zeta)$ for large k.

Appendix D: Proofs in Section 7

D.1. Proof of Theorem 6

The proof for the equilibrium characterization is again used the approach we used in the identical agent setting. The rigorous proof can be seen in Section S.4 of the Supporting Document. Here, we only give the proofs for the claims that equilibrium demand for the high-quality agents and their equilibrium revenue is higher, i.e. $D^*_H \geq D^*_L$, and $V^*_H \geq V^*_L$, for regions 1 and 3.

Region 1 ($\Lambda > k_H\lambda^{mon}_H + k_L\lambda^{mon}_L$): In this case both groups solves the following monopoly problem:

$$\begin{aligned} \max_{p\geq w_i,\lambda\geq 0} \quad & (p-w_i)\lambda\,[1-\beta(\lambda)] - \lambda cm_a \\ s.to \quad & \\ & (R_i - p + cm_a)\,[1-\beta(\lambda)] - cm_a \geq 0, \end{aligned}$$

which can be reduced to $\max_{\lambda\geq 0}(R_i + cm_a - w_i)\lambda\,[1-\beta(\lambda)]$. Note that $R_i - w_i$ is a parameter of this optimization problem, and by the envelope theorem the value of the optimum and the optimal solution are both increasing in $R_i - w_i$. Therefore, our claim holds since $R_H - w_H \geq R_L$.

Region 3 ($\Lambda(R_L) < \Lambda < \Lambda(0)$): We first want to note that $y(x) > x$ where $y(x)$ is the unique solution for $\hat{U}_L(x,y) = \hat{U}_H(x,y)$ for any given x by Lemma 15 in the Supporting Document. Thus, we should have that $\Lambda D^*_H > \Lambda D^*_L$ by their definition.

Furthermore, we have that $\Lambda D^*_H < \lambda^{mon}_H$ since

$$\hat{U}_H(x,\lambda^{mon}_H) \;=\; (R_H + cm_a - w_H)\left[\frac{\frac{cm_a}{R_H+cm_a}}{1+\frac{\nu(\lambda^{mon}_H)}{k_L+k_H\vartheta(\lambda^{mon}_H,x)-1}}\right] - cm_a < 0,$$

for any $0 \leq x \leq \lambda^{mon}_L$. Then, using the optimal prices p^*_H and p^*_L, we have that

$$\begin{aligned} V^*_H &= (R_H + cm_a - w_H)\Lambda D^*_H(1-\beta(\Lambda D^*_H)) - \Lambda D^*_H(\hat{U}_H(\Lambda D^*_L, \Lambda D^*_H) + cm_a) \\ &\geq (R_H + cm_a - w_H)\Lambda D^*_L(1-\beta(\Lambda D^*_L)) - \Lambda D^*_L(\hat{U}_H(\Lambda D^*_L, \Lambda D^*_H) + cm_a) \\ &\geq (R_L + cm_a)\Lambda D^*_L(1-\beta(\Lambda D^*_L)) - \Lambda D^*_L(\hat{U}_H(\Lambda D^*_L, \Lambda D^*_H) + cm_a) = V^*_L, \end{aligned}$$

where the first inequality holds since $(R_H + cm_a - w_H)\lambda(1-\beta(\lambda)) - \lambda cm_a$ is increasing for any $\lambda \leq \lambda^{dom}_H$, and the second one holds since $R_H - w_H \geq R_L$.

D.2. Proof of Proposition 6

We prove our claim by contradiction. Therefore, we suppose there exists a non-symmetric equilibrium where $N > 1$ groups of agents charge different prices, i.e., there exists a price vector $(p_n^*)_{n=1}^N$ arising as the equilibrium outcome.

When we let $R_H = R_L = R$ and $w_H = 0$ in Theorem 6, the equilibrium characterization shows that each group should charge the same price in part 1 and 3. Note that $\Lambda(0) = \lambda^0$, and $\lambda_H^{mon} = \lambda_L^{mon} = \lambda^{mon}$ by definition and the fact that $R_H = R_L = R$. Furthermore, we do not have part 4 and 5 since $R_H = R_L = R$. Hence, as a corollary of Theorem 6, we cannot have a non-symmetric equilibrium with $N = 2$.

$N > 2$ requires extra arguments. Similar to the additional functions in Theorem 6, we define N different functions:

$$\hat{U}_\ell(x_1, \ldots, x_N) = (R + cm_a)\left[\frac{1-\nu(x_\ell)}{1+\frac{\nu(x_\ell)}{k_\ell - 1 + \sum_{n\neq\ell} k_n \vartheta(x_\ell, x_n)}}\right]\left[1-\beta(x_\ell)\right] - cm_a,$$

for any $\ell \in \{1, \ldots, N\}$.

We first focus on the case where $\rho < \lambda^0$. Let D_n^* be the fraction of customers picking an agent charging p_n^* in the equilibrium. We now argue that $U(\Lambda D_n^*, p_n^*) > 0$ for all $n \in \{1, \ldots, N\}$, i.e. customer utility is strictly positive in the equilibrium. To see that suppose customer utility is zero on the contrary. Then as in Case-2 of the proof of Theorem 6, we can show that it is necessary to have $\hat{U}_\ell(\Lambda D_1^*, \ldots, \Lambda D_N^*) \le 0$.

Without loss of generality, assume $D_1^* > \cdots > D_N^*$. This implies that $D_N^* < 1/k$ because otherwise we would have that $\sum_{n=1}^{n} k_n D_n^* > 1$. Furthermore, note that $\vartheta(x, y) > 1$ when $x < y$ as $\beta'(x)/(1-\beta(x))$ is assumed to be decreasing. Hence, we have that $\vartheta(\Lambda D_N^*, \Lambda D_n^*) > 1$ for all $n \in \{1, \ldots, N\}$. Moreover, since $D_N^* < 1/k$ and both $\beta(x)$ and $\nu(x)$ are decreasing, we have that

$$\left[1-\nu(\Lambda D_N^*)\right]\left[1-\beta(\Lambda D_N^*)\right] > \left[1-\nu(\Lambda/k)\right]\left[1-\beta(\Lambda/k)\right]$$

Combining these observations, we have that

$$\hat{U}_N(\Lambda D_1^*, \ldots, \Lambda D_N^*) > \hat{U}_N(\rho, \ldots, \rho) > 0,$$

where the last inequality holds since $\rho < \lambda^0$ and $\hat{U}_N(\lambda^0, \ldots, \lambda^0) = 0$ by definition of λ^0.

Once we have that customer utility will be strictly positive in the equilibrium, we can show that utility of customers picking the agent charging p_n^* will be $\hat{U}_n(\Lambda D_1^*, \ldots, \Lambda D_N^*)$ in the equilibrium. Moreover, all customers request service. This means that, $(D_n^*)_{n=1}^N$ should solve that

$$\begin{aligned}&\hat{U}_1(\Lambda D_1^*, \ldots, \Lambda D_N^*) = \cdots = \hat{U}_N(\Lambda D_1^*, \ldots, \Lambda D_N^*), \ \textit{and} \\ &\sum_{n=1}^{n} k_n D_n^* = 1.\end{aligned}$$

We prove our claim for $\rho < \lambda^0$ by showing that the unique solution to the above system is $D_n^* = 1/k$ for all $n \in \{1, \ldots, N\}$. First, it is easy to see that $D_n^* = 1/k$ solves the above system of equations by the definition of

$\hat{U}_n$ functions. Suppose there exists another vector $(D_1,\ldots,D_N)$ which solves these equations. Without loss of generality, assume $D_1 > \cdots > D_N$. As we mentioned above, this implies that

$$\hat{U}_N(\Lambda D_1,\ldots,\Lambda D_N) > \hat{U}_N(\rho,\ldots,\rho) > \hat{U}_1(\Lambda D_1,\ldots,\Lambda D_N).$$

This is a contradiction. Therefore, the only solution to above equation system is $D_n^* = 1/k$ for all $n \in \{1,\ldots,N\}$. So far, we assume that agents charge N different prices, and concluded that they attract the same demand. However, this contradicts with the definition of Customer Equilibrium. Hence, our assumption of non-symmetric equilibrium is wrong for $\rho < \lambda^0$.

Now, we show that there cannot be any non-symmetric equilibrium where agents charge $N > 2$ different prices when $\rho > \lambda^{mon}$. It is very easy to argue that a single agent always attracts a demand less than λ^{mon} (demand of a monopolist) in the equilibrium because otherwise he would have opportunity to deviate and behave like a monopolist. Hence, we should have that $\Lambda D_n^* \le \lambda^{mon}$ for all $n \in \{1,\ldots,N\}$. This implies that $\sum_{n=1}^{n} k_n D_n^* < 1$ since $\Lambda > k\lambda^{mon}$. Then, as in the proof of Case-1 of Theorem 6, we can show that $\sum_{n=1}^{n} k_n D_n^* < 1$ implies that $\Lambda D_n^* = \lambda^{mon}$ for any $n \in \{1,\ldots,N\}$. The intuition is that there are customers not requesting service as $\sum_{n=1}^{n} k_n D_n^* < 1$, so that there is room for any single agent to behave like a monopolist.

D.3. Proof of Proposition 7

Lemma 8. *In the k^{th} marketplace, let*

$$\begin{aligned}
D_H(k) &= D_1^{MCE}(R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k)\\
D_L(k) &= D_2^{MCE}(R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k)\\
\beta_H(k) &= \beta_1(D_H(k), D_L(k); R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k)\\
\beta_L(k) &= \beta_2(D_H(k), D_L(k); R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k)\\
P_{HL}(k) &= PServ_{12}(D_H(k), D_L(k); R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k)\\
P_{LH}(k) &= PServ_{21}(D_H(k), D_L(k); R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k),
\end{aligned}$$

where $\lim_{k\to\infty} p_i^k = p_i \le R_i$ for $i \in \{H,L\}$. the, when $R_i - p_i > R_j - p_j$ for some $i,j \in \{H,L\}$ with $i \ne j$, the following statements are true:

1. *If $\rho \le \alpha_i$, we have that $\lim_{k\to\infty} D_i(k) + P_{ji}(k) D_j(k) = 1$.*
2. *If $\rho > \alpha_i$, we have that*
   (a) $\liminf_{k\to\infty} \frac{\Lambda^k D_i(k)}{\alpha_i k} > 1$.
   (b) $\lim_{k\to\infty} P_{ji}(k) = 0$.
   (c) $\lim_{k\to\infty} \beta_j(k) = \max\left\{0, 1 - \frac{\alpha_j}{\rho \tilde{D}_j}\right\}$, *when* $\tilde{D}_j = \lim_{k\to\infty} D_j(k)$.
   (d) $\lim_{k\to\infty} \beta_i(k) = 1 - \frac{\alpha_i}{\rho \tilde{D}_i}$, *when* $\tilde{D}_i = \lim_{k\to\infty} D_i(k)$.
3. *If $\rho > \alpha_i$, we have that*

$$\lim_{k\to\infty} D_i(k) = \max\left\{\min\left\{1, \left(\frac{R_i - p_i + cm_a}{R_j - p_j + cm_a}\right)\frac{\alpha_i}{\rho}\right\}, \min\left\{\frac{R_i - p_i + cm_a}{\bar{R}}\alpha_i, \left(\frac{R_i - p_i + cm_a}{cm_a}\right)\frac{\alpha_i}{\rho}\right\}\right\}.$$

4. *If* $\rho < \frac{\bar{R}}{cm_a}$ *and* $p_j < R_j$*, we have that*

$$\lim_{k\to\infty} D_j(k) = 1 - \lim_{k\to\infty} D_i(k).$$

Furthermore, $\lim_{k\to\infty} D_j(k) \leq 1 - \lim_{k\to\infty} D_i(k)$ *when* $p_j = R_j$*.*

5. *If* $\rho \geq \frac{\bar{R}}{cm_a}$ *and* $p_j < R_j$*, we have that*

$$\lim_{k\to\infty} D_j(k) = \left(\frac{R_j - p_j + cm_a}{cm_a}\right)\frac{\alpha_j}{\rho}.$$

Furthermore, $\lim_{k\to\infty} D_j(k) \leq \left(\frac{R_j - p_j + cm_a}{cm_a}\right)\frac{\alpha_j}{\rho}$ *when* $p_j = R_j$*.*

The proof this lemma can be seen in the Supporting Document.

Revenue of Group-i: Using the above lemma, we have that

$$\lim_{k\to\infty} V_i^{MCE}(p_H^k, p_L^k; k) = \lim_{k\to\infty}(p_i - w_i)\frac{\rho D_i(k)}{\alpha_i}[1 - \beta_i(k)] + \lim_{k\to\infty}(p_i - w_i)\frac{\rho P_{ji}(k) D_j(k)}{\alpha_i} = (p_i - w_i)\min\{\rho/\alpha_i, 1\}.$$

Revenue of Group-j: Using the above lemma, we have that $\lim_{k\to\infty} D_j(k) = 0$ when $\rho \leq \frac{(R_i - p_i + cm_a)\alpha_i}{R_j - p_j + cm_a}$. Therefore, the revenue goes to zero.

Furthermore, when $\frac{(R_i - p_i + cm_a)\alpha_i}{R_j - p_j + cm_a} < \rho \leq \frac{\bar{R}}{R_j - p_j + cm_a}$, we have that $\lim_{k\to\infty} \rho D_j(k) = \left(\rho - \frac{(R_i - p_i + cm_a)\alpha_i}{R_j - p_j + cm_a}\right) \leq \alpha_j$. Therefore, we have that $\lim_{k\to\infty} \beta_j(k) = 0$, and

$$\lim_{k\to\infty} V_j^{MCE}(p_H^k, p_L^k; k) = \lim_{k\to\infty}(p_j - w_j)\frac{\rho D_j(k)}{\alpha_j}[1 - \beta_j(k)] = \left(\frac{\rho}{\alpha_j} - \frac{(R_i - p_i + cm_a)\alpha_i}{(R_j - p_j + cm_a)\alpha_j}\right)(p_j - w_j),$$

when $p_j < R_j$. When $p_j = R_j$, the above expression is an upper-bound for the agent revenues since $\lim_{k\to\infty} D_j(k) \leq 1 - \lim_{k\to\infty} D_i(k)$.

Finally, when $\rho > \frac{\bar{R}}{R_j - p_j + cm_a}$, we have that $\lim_{k\to\infty} \rho D_j(k) > \alpha_j$, and thus the limit of the revenue goes to $p_j - w_j$ when $p_j < R_j$. When $p_j = R_j$, $p_j - w_j$ is an upper-bound for the agent revenues.

D.4. Supplementary Claims for the Proof of Theorem 7

LEMMA 9. *For any* (p_H^k, p_L^k)*, where* $\lim_{k\to\infty} p_i^k = p_i$ *for* $i \in \{H, L\}$*,* $1 < \frac{R_i - p_i + cm_a}{R_j - p_j + cm_a} < \frac{\rho}{\alpha_i}$*, for some* $i, j \in \{H, L\}$ *with* $i \neq j$*, and* $p_j < R_j$*, we have that*

$$\lim_{k\to\infty} V_{i_{dev}}(k) = p_i - w_i + \varepsilon, \ \lim_{k\to\infty} V_{j_{dev}}(k) = p_j - w_j - \varepsilon,$$

where $\varepsilon < (R_i - p_i) - (R_j - p_j)$*,* $V_{i_{dev}}(k)$ *is the profit of a quality-i agent charging* $p_i + \varepsilon$ *when all other quality-i providers charge* p_i^k*, and all quality-j providers charge* p_j^k*,and* $V_{j_{dev}}(k)$ *is the profit of a quality-j agent charging* $p_j - \varepsilon$ *when all other quality-i providers charge* p_i*, and all quality-j providers charge* p_j*. Furthermore, the same result holds even when* $p_j = R_j$ *as long as the limiting revenue of quality-j agents is strictly positive before deviation.*

The proof of the lemma can be seen in the Supporting Document.

LEMMA 10. *For any given sequence of price pairs* (p_H^k, p_L^k)*, where* $\lim_{k\to\infty} p_H^k = p_H$*,* $\lim_{k\to\infty} p_L^k = p_L$*, and* $R_H - p_H = R_L - p_L > 0$*, we have that*

$$\lim_{k\to\infty} V_{i_{dev}}(k) = p_i - \varepsilon - w_i, \text{ for } i \in \{H, L\}$$

where $0 < \varepsilon < \min\{0, p_i - w_i\}$ *and* $V_{i_{dev}}(k)$ *is the profit of a quality-*i *agent charging* $p_i - \varepsilon$ *when all other low-quality providers charge* p_L^k*, and all other high-quality providers charge* p_H^k *in the* k^{th} *marketplace. Furthermore, the same result holds even when* $R_H - p_H = R_L - p_L = 0$ *as long as the limiting revenue of all agents is strictly positive before deviation.*

The proof of the lemma can be seen in the Supporting Document.

LEMMA 11. *If* $p \in \mathcal{P}(\rho; R)$*, where* $\mathcal{P}(\rho; R)$ *is defined as in Theorem 3, then for any* $\varepsilon > 0$*, we have that*

$$(p + \varepsilon) \in \mathcal{P}(\rho; R + \varepsilon).$$

The proof of the lemma can be seen in the Supporting Document.

D.5. Proof of Theorem 7

1. We first want to note that the equivalent of our claim is that $\limsup_{k\to\infty} p_{EQ_i}^k = w_i$. We show this by contradiction. Thus, we suppose that $\limsup_{k\to\infty} p_{EQ_i}^k > w_i$. Then, there should be convergent subsequence of the equilibrium price pairs with limits p_ℓ for all $\ell \in \{H, L\}$ satisfying $p_{EQ_i} > w_i$ since equilibrium prices are bounded. Following a case-by-case analysis, we show that this assumption leads to a contradiction:

i. $(\mathbf{R_i - p_{EQ_i} = R_j - p_{EQ_j}})$**:** In this case, the utilization of quality-i agents cannot be higher than ρ/α_i even though all customer request service from them. Thus, the upper-bound for their limiting revenue is $\rho/\alpha_i(p_{EQ_i} - w_i)$. However, Lemma 10 establishes that a quality-i agent can secure a revenue of $p_{EQ_i} - w_i - \varepsilon$ when he cuts his price by ε as long as $p_{EQ_j} < R_j$ or the limiting revenue of all agents is positive before the deviation. Clearly, for any $\varepsilon < (1 - \rho/\alpha_i)(p_{EQ_i} - w_i)$, this kind of deviation improves his profit. For the sequence of prices resulting zero limiting revenue for at least one group of agents (which can only happen when $R_i - p_{EQ_i} = R_j - p_{EQ_j} = 0$), any single agent from this group can improve his revenue by trivially deviating to $R_j/2$ as discussed rigorously in the proof of Proposition 3. Hence, any sequence of price pairs with limits satisfying $R_i - p_{EQ_i} = R_j - p_{EQ_j}$ cannot emerge as an equilibrium price pair.

ii. $(\mathbf{R_j - p_{EQ_j} > R_i - p_{EQ_i}})$**:** In this case, we have two sub-cases:

* $\underline{\rho \le \frac{\mathbf{R_j - p_{EQ_j} + cm_a}}{\mathbf{R_i - p_{EQ_i} + cm_a}} \alpha_\mathbf{j}}$**:** As we show in Lemma 8, the profit of a quality-i provider is zero as k goes to infinity. However, when a quality-i provider deviates to charge a price $p' < R_i - R_j + p_{EQ_j}$, he becomes the least expensive provider, and always attracts strictly positive demand (Similar to the discussion at the end of the proof of Proposition 3). This kind of deviation clearly improves his profit. Therefore, in large systems any sequence of price pairs with limits satisfying this sub-case cannot emerge as an equilibrium price pair (Note that there always exists a $p' > w_i$ since $R_i - w_i > R_j - w_j$. When $R_i - w_i = R_j - w_j$, the above proof holds only for $p_{EQ_j} > w_j$).

** $\rho > \frac{\mathbf{R_j - p_{EQ_j} + cm_a}}{\mathbf{R_i - p_{EQ_i} + cm_a}} \alpha_\mathbf{j}$**:** In Lemma 9, we show that a quality-j provider can increase his price while keeping him fully-utilized whenever the limiting revenue of quality-i agents is strictly positive or $p_{EQ_i} < R_i$. This kind of deviation clearly improves his profit. On the other hand, if the limiting revenue of quality-i agents is zero and $p_{EQ_i} = R_i$, a quality-i provider can deviate to charge a price $p' < (R_i - R_j + w_j) + (R_j - w_j + cm_a)\left[1 - \frac{\alpha_j}{\rho}\right]$. As a result of this deviation he always attracts strictly positive demand because the utility of customers would be $(R_j - p_{EQ_j} + cm_a)\alpha_j/\rho - cm_a$ (which is strictly less than $R_i - p'$ even if $p_{EQ_j} = w_j$) if all of them chose quality-j agents. This kind of deviation clearly improves his profit compared to his revenue before deviation, which is zero in the limit. Therefore, in large systems any sequence of price pairs with limits satisfying this sub-case cannot emerge as an equilibrium price pair (Note that $p' > w_i$ is always true, even if $R_i - w_i = R_j - w_j$, since $\rho > \alpha_j$. thus, we do not use our assumption of $R_i - w_i > R_j - w_j$.).

iii. $(\mathbf{R_i - p_{EQ_i} > R_j - p_{EQ_j}})$: In Lemma 8.1, we show that quality-i providers are under-utilized since $\rho < \alpha_i$. Then, as we rigorously show in Section 5, any single-provider can cut his price by a small amount, and he can make sure he will be fully-utilized in a sufficiently large system. Clearly, this kind of deviation improves his profit, so that any sequence of price pairs with limits satisfying this case cannot emerge as an equilibrium price pair.

2.a) (Characterization of the limit): In this case, the equivalent of our claim is that either of the following statements is true:

1. $\limsup_{k\to\infty} p^k_{EQ_i} < \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$.
2. $\limsup_{k\to\infty} p^k_{EQ_i} = \liminf_{k\to\infty} p^k_{EQ_i} = R_i - R_j + w_j$ and $\limsup_{k\to\infty} p^k_{EQ_j} = \liminf_{k\to\infty} p^k_{EQ_j} = w_j$.

As in part 1, we prove this claim by contradiction. Thus, we suppose that both of the following statements should be true:

1. $\limsup_{k\to\infty} p^k_{EQ_i} > \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$
2. Either $\limsup_{k\to\infty} p^k_{EQ_i} \neq R_i - R_j$ or $\liminf_{k\to\infty} p^k_{EQ_i} \neq R_i - R_j$ $\limsup_{k\to\infty} p^k_{EQ_j} \neq 0$ or $\liminf_{k\to\infty} p^k_{EQ_j} \neq 0$.

Then, there should exist a sequence of equilibrium price pairs with limits p_ℓ for all $\ell \in \{H, L\}$ satisfying $p_{EQ_i} > \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$, but not satisfying $R_i - p_{EQ_i} = R_j - p_{EQ_j} = R_j - w_j$. Following a case-by-case analysis, we show that this assumption leads to a contradiction:

i. $(\mathbf{R_i - p_{EQ_i} = R_j - p_{EQ_j} < R_j - w_j})$: In this case, the utilization of at least one group of agents (either low or high) should be less than ρ since the total rate of customers requesting service cannot exceed the total rate of customers. WLOG, suppose the utilization of quality-j agents is less than ρ. Then, the upper-bound for their limiting revenue is ρp_{EQ_j}. However, Lemma 10 establishes that a quality-j agent can secure a revenue of $p_{EQ_j} - \varepsilon$ when he cuts his price by ε as long as $p_{EQ_j} < R_j$ or the limiting revenue of agents is positive before the deviation. Clearly, for any $\varepsilon < (1-\rho)p_{EQ_j}$, this kind of deviation improves his profit Furthermore, the sequence of prices resulting zero limiting revenue for quality-j agents (can happen only when $p_{EQ_j} = R_j$) can be ruled out as in Part 1.i. Thus, any sequence of price pairs with limits satisfying this case cannot emerge as an equilibrium price pair. (Note that we do not use our assumption of $R_i - w_i > R_j - w_j$.)

ii. $(\mathbf{R_j - p_{EQ_j} > R_i - p_{EQ_i}})$: First note that $p_{EQ_i} \geq \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$ since $p_{EQ_i} > R_i - R_j + w_j$ in this case. Then any sequence of price pairs with limits satisfying this case cannot emerge as an equilibrium price as in part 1.ii. (Note that we do not use our assumption of $R_i - w_i > R_j - w_j$ here as well.)

iii. $(\mathbf{R_i - p_{EQ_i} > R_j - p_{EQ_j}})$: In this case, we have two sub-cases:

* $\rho \leq \frac{\mathbf{R_i - p_{EQ_i} + cm_a}}{\mathbf{R_j - p_{EQ_j} + cm_a}} \alpha_\mathbf{i}$**:** As we show in Lemma 8, the profit of a quality-j provider is zero as k goes to infinity. Furthermore, since $p_{EQ_i} > (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]$, the expected utility of customers converges to a limit, which is strictly less than $R_j - w_j$. The lower bound on p_{EQ_i} also implies that $p_{EQ_j} > w_j$ in this case. Then, we can argue that when a quality-j provider deviates to charge a price p' such that $R_j - R_i + p_{EQ_i} < p' < R_j + cm_a - (R_i - p_{EQ_i} + cm_a)\frac{\alpha_i}{\rho} \leq p_{EQ_j}$, he always attracts strictly positive demand (Note that $R_j + cm_a - (R_i - p_{EQ_i} + cm_a)\frac{\alpha_i}{\rho} > w_j$ by the lower bound on p_{EQ_i}.). The reasoning for this argument is similar to previous claims:

1. Demand for quality-i agents exceeds their capacity after the deviation of quality-j agent.
2. Therefore, all customers picking the deviating agent should be served by him.
3. Some customers should pick the deviating agents because otherwise customers' utility would be less than $R_j - p'$ since $(R_i - p_{EQ_i} + cm_a)\frac{\alpha_i}{\rho} - cm_a < R_j - p'$.

This kind of deviation clearly improves his profit. Therefore, in large systems any sequence of price pairs with limits satisfying this sub-case cannot emerge as an equilibrium price pair.

** $\rho > \frac{\mathbf{R_i - p_{EQ_i} + cm_a}}{\mathbf{R_j - p_{EQ_j} + cm_a}} \alpha_\mathbf{i}$**:** In Lemma 9, we show that a quality-i provider can increase his price while ensuring a revenue strictly greater than p_{EQ_i} when $p_{EQ_j} < R_j$ or the limiting revenue of agents is positive before the deviation. This kind of deviation clearly improves his profit. On the other hand, if the limiting revenue of quality-j agents is zero and $p_{EQ_j} = R_j$, a quality-j provider can deviate to charge a price p' such that $R_j - R_i + p_{EQ_i} < p' < p_{EQ_j}$ and always attract strictly positive demand. The reasoning for this argument is similar to previous claims:

1. Demand for quality-i agents exceeds their capacity after the deviation of quality-j agent.
2. Therefore, all customers picking the deviating agent should be served by him.
3. Moreover, since $\rho > \frac{R_i - p_{EQ_i} + cm_a}{cm_a}\alpha_i$ when $p_{EQ_j} = R_j$, not all customers will request service, and thus the utility of customers after deviation will be zero.
4. Hence, some customers should request service from the deviating agent since $R_j - p' > 0$. To be more specific, pick $u > 0$, let $\bar{\lambda} > 0$ solves $(R - p')(1 - \beta(\bar{\lambda})) = u$. Then, the utility of customers picking the deviating agent would be at least u when his demand is less than $\bar{\lambda}$ for large k. Thus, the demand for the deviating agent should be more than $\bar{\lambda}$ for large k.

This kind of deviation clearly improves his profit since all quality-j agents earn zero when $p_{EQ_j} = R_j$. Therefore, in large systems any sequence of price pairs with limits satisfying this sub-case cannot emerge as an equilibrium price pair. (Note that we do not use our assumption of $R_i - w_i > R_j - w_j$ here as well.)

2.b) (Existence of the multiple equilibria): We prove that the proposed price pair emerges as the equilibrium price as in the proof of Theorem 3. First, we want to note that quality-j agents will always be out of the market for large k since $\tilde{p}_i < (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]$ implies that the expected utility of customers converges to a limit, which is strictly greater than $R_j - w_j$. Thus, they cannot improve their revenues.

Furthermore, in order to show that even the quality-i agents cannot improve their revenues, we let $V_i'(k)$ be the maximum profit that a quality-i provider can get by increasing his price. Note that we do not consider a deviation where a quality-i agent cuts his price because quality-i agents are already over-utilized. Let $\Delta^H(p; R) = \max\left\{0, \frac{R - p + cm_a}{\rho}\alpha_i - cm_a\right\}$. Then, we can show that

$$\limsup_{k\to\infty} V_i'(k) \leq \max_{\lambda}(R_i + cm_a - w_i)\lambda[1 - \beta(\lambda)] - \lambda(\Delta^H(\tilde{p}_i; R_i) + cm_a)$$

$$\leq \max_{\lambda}(R_i + cm_a - w_i)\lambda[1-\beta(\lambda)] - \lambda(\Delta^H(\tilde{p}_i; R_i - w_i) + cm_a),$$

where the first inequality holds as in the proof of Proposition 4, and the second one holds since $\Delta^H(p;R)$ is increasing in R. Then, as in the proof of Theorem 3, quality-i agents do not have a profitable deviation since $\tilde{p}_i \in \mathcal{P}(\rho/\alpha_i, R_i - w_i) \Rightarrow \tilde{p}_i > \max_{\lambda}(R_i + cm_a - w_i)\lambda[1-\beta(\lambda)] - \lambda(\Delta^H(\tilde{p}_i; R_i - w_i) + cm_a)$.

2.c) (Existence of the sequence converging to $R_i - R_j + w_j$): Let $\hat{p}_i^k$ be a sequence converging to $R_i - R_j + w_j$ with a rate of $1/\sqrt{k}$ such that $\hat{p}_i^k = R_i + cm_a - \frac{R_j - w_j + cm_a}{1-\beta^M(\alpha_i k; \alpha_i k)}$. By construction, the quality-i agents, who charge $\hat{p}_i^k$, attracts a demand rate of $\alpha_i k$ when all quality-i agents charge $\hat{p}_i^k$ and all quality-j agents charge zero. Furthermore, their utilization will be very close to 1 as k grows. To be more specific, their utilization will converge to 1 with a rate of $1/\sqrt{k}$ by Theorem 5 in Zeltyn and Mandelbaum (2005). Hence, the revenue of a quality-i agent will be in the form of $R_i - R_j - w_i + w_j - \zeta_2/\sqrt{k}$, where ζ_2 is a constant, given that all quality-i agents charge $\hat{p}_i^k$ and all quality-j agents charge zero.

As we assume that $\frac{1}{\epsilon^k\sqrt{k}} \to 0$, we should have that $R_i - R_j - w_i + w_j - \zeta_2/\sqrt{k} + \epsilon^k > R_i - R_j - w_i + w_j$ for large k. This implies that a quality-i agent should charge more than $R_i - R_j + w_j$ in order to ensure a profitable deviation. However, a high quality agent would become the most expensive agent while there is ample capacity to serve customers when he follows such a deviation, and thus his demand would be zero. Therefore, quality-i agents do not have a profitable deviation from this price pair. The same would happen if a quality-j agent would increase his price.

3. The proof to rule out any sequence of price pairs with limits satisfying $R_i - p_{EQ_i} \neq R_j - \tilde{p}_{EQ_j}$ is the same as the proof in part 2.a. Furthermore, we prove that the proposed price pairs emerge as the equilibrium price as in the proof of Theorem 3. Note that $\tilde{p}_j \in \mathcal{P}(\rho, R_j)$ implies that $\tilde{p}_j < R_j$, and thus both groups of agents should be over-utilized given all quality-q agents charge p_q^k for $q \in \{H, L\}$ in the k^{th} marketplace. Hence, it is again enough to just focus on the price increase as a possible profitable deviation.

We let $V_q'(k)$ be the maximum profit that a quality-q provider can get by increasing his price for $q \in \{H, L\}$. Then, we have that

$$\begin{aligned}
\limsup_{k\to\infty} V_q'(k) &\leq \max_{\lambda}(R_q + cm_a - w_q)\lambda[1-\beta(\lambda)] - \lambda(\Delta(\tilde{p}_q; R_q) + cm_a) \\
&\leq \max_{\lambda}(R_q + cm_a)\lambda[1-\beta(\lambda)] - \lambda(\Delta(\tilde{p}_q; R_q) + cm_a) \\
&= (R_q + cm_a)\lambda^{\Delta}(\tilde{p}_q; R_q)[1-\beta(\lambda^{\Delta}(\tilde{p}_q; R_q))] - \lambda^{\Delta}(\tilde{p}_q; R_q)(\Delta(\tilde{p}_q; R_q) + cm_a),
\end{aligned}$$

where the first inequality holds as in the proof of Proposition 4, and the second one holds since $w_q \geq 0$ for $q \in \{H, L\}$. Then, as in the proof of Theorem 3, quality-j providers do not have a profitable deviation since $\tilde{p}_j \in \mathcal{P}(\rho, R_j)$. Moreover, quality-$i$ providers do not have a profitable deviation, either since $\tilde{p}_j \in \mathcal{P}(\rho, R_j)$ implies that $\tilde{p}_i \in \mathcal{P}(\rho, R_i)$ by Lemma 11.

D.6. Proof of Proposition 8

We first want to note that the equivalent of our claim is that $\limsup_{k\to\infty} p_1^k = 0$. We show this by contradiction. Thus, we assume that $\limsup_{k\to\infty} p_1^k > 0$. Then, there should exist a sequence of price vectors $(p_1^k, \ldots, p_N^k)$ with limits $(p_1, \ldots, p_N)$, where $p_n > 0$ for all $n \in \{1, \ldots, N\}$. Suppose α_n fraction of agents charge p_n^k for all

$n \in \{1, \dots, N\}$ in the k^{th} marketplace. We will show that a single agent will have a profitable deviation from this strategy for large k.

When $N = 2$, we can use the results of Theorem 7 by letting $R_H = R_L = R$ and $w_H = w_L = 0$ because this theorem states that any price vector, where different class of agents charge different strictly positive prices, cannot emerge as the equilibrium price for large k (Note that we do not use our assumption of $R_H - w_H > R_L$ to rule out sequences with limit $p_L > 0$. It is only required for the case of $p_L = 0$). However, $N > 2$ needs extra arguments. Suppose there exists a sequence of equilibirum price vectors $(p_1^k, \dots, p_N^k)$ with limits $(p_1, \dots, p_N)$ and $N > 2$ for any k.

We first need to characterize the revenue of agents given the price vector $(p_1^k, \dots, p_N^k)$ and the fraction of agents charging these prices $(\alpha_1, \dots, \alpha_N)$. Let $(V_1^k, \dots, V_N^k)$ be the the revenue of agents in the k^{th} marketplace. Then very similar to Proposition 7, we can show that for any $\ell \in \{1, \dots, N\}$

$$\lim_{k\to\infty} V_\ell^k = \begin{cases} 0 & \text{if } \rho < \rho_\ell^0 \\ p_\ell \left(\frac{\rho - \rho_\ell^0}{\alpha_\ell}\right) & \text{if } \rho \in [\rho_\ell^0, \rho_\ell^0 + \alpha_\ell] \\ p_\ell & \text{if } \rho > \rho_\ell^0 + \alpha_\ell, \end{cases}$$

where $\rho_1^0 = 0$, and $\rho_\ell^0 = \frac{\sum_{n=1}^{\ell-1} (R - p_n + cm_a)\alpha_n}{R - p_\ell + cm_a}$ for all $\ell > 1$. Note that we always have that $\lim_{k\to\infty} V_1^k > 0$ as $\rho > 0$. Furthermore, either of the following two cases also holds always:

1. $\lim_{k\to\infty} V_N^k = 0$: In such a case, we would have that agents charging p_N earn zero revenue while the revenue of agents charging p_1 is strictly positive. This would contradicts with the definition of ϵ^k-Market Equilibrium because a single agent from sub-pool-N could improve his revenue by charging an arbitrarily small price less than $p_1 > 0$.

2. $\lim_{k\to\infty} V_N^k > 0$ and $\lim_{k\to\infty} V_n^k = p_n$ for all $n < N$: In this case, sub-pool-$(N-1)$ earns strictly more than sub-pool-n for all $n < N-1$ since $p_1 < \cdots < p_N$. Then this creates an opportunity for any single agent in sub-pool-1 to improve his revenue by increasing his price as in the proof of Lemma 9. Thus, this case also contradicts with the definition of ϵ^k-Market Equilibrium.

As both of these cases lead to a contradiction, our assumption that $(p_1^k, \dots, p_N^k)$ is equilibrium for any k is wrong, and thus the lowest price should converge to zero. As a direct implication of that we have that $\lim_{k\to\infty} V_1^k = 0$. Moreover, we also need to have that $\lim_{k\to\infty} V_n^k = 0$ for $n > 1$ because otherwise any single agent in sub-pool-1 would have an opportunity to improve his revenue and that would contradict with the definition of ϵ^k-Market Equilibrium.

D.7. Supplementary Claims for the Proof of Theorem 8

LEMMA 12. *Suppose $R_i - w_i > R_j - w_j$ for some $i, j \in \{H, L\}$ with $i \neq j$. Then, if $\rho > \alpha_i$, we have that*

$$\lim_{k\to\infty} V_i'(k) = p_i + \varepsilon - w_i,$$

where $V_i'(k)$ is the profit of the δ fraction of high-quality providers charging $p_i + \varepsilon$ when all other quality-q providers charge p_q^k for $q \in \{H, L\}$ with $0 < \varepsilon < \min\left\{R_i - R_j + w_j - p_i, \left(1 - \frac{\alpha}{\rho}\right)(R_i - p_i + cm_a)\right\}$ and $\lim_{k\to\infty} p_i^k = p_i \leq \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$.

LEMMA 13. *In a seller's market ($\rho > 1$), for any sequence of price pairs (p_H^k, p_L^k), where $\lim_{k\to\infty} p_i^k = p_i$ for $i \in \{H, L\}$ and $R_H - p_H = R_L - p_L > 0$, we have that*

$$\lim_{k\to\infty} V_i'(k) = p_i + \varepsilon - w_i,$$

where $V_i'(k)$ is the profit of the δ fraction of quality-i providers charging $p_i + \varepsilon$ for $i \in \{H, L\}$ when all other high-quality providers charge p_H^k, and all low-quality providers charge p_L^k, and $0 < \varepsilon < \min\left\{R_i - p_i, \left(1 - \frac{1}{\rho}\right)(R_i - p_i + cm_a)\right\}$.

The proofs of these lemmas can be seen in the Supporting Document

D.8. Proof of Theorem 8

1. The fact that $p_i^k < w_i + \xi$ for large k holds directly by Theorem 7.1.

Existence: The existence of the sequence holds as in the proof of Theorem 4. Similar to this proof, we can only show the existence of such a sequence for $\rho < \alpha_i - \tilde{\delta}$ if $\lim_{k\to\infty} \delta^k = \tilde{\delta} > 0$.

2. We have already shown that any sequence of price pairs with limits (p_{EQ_i}, p_{EQ_j}), where $p_{EQ_i} > \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$, except $R_i - p_{EQ_i} = R_j - p_{EQ_j} = R_j - w_j$, can not be an equilibrium in a large marketplace in Theorem 7.2.a. Furthermore, Lemma 12 provides a profitable deviation for a small group of agent when $p_{EQ_i} \le \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$. Hence, any sequence of price pairs with limits (p_{EQ_i}, p_{EQ_j}) cannot emerge as the equilibrium price pair of a symmetric (δ^k, ϵ^k)-Market Equilibrium in large marketplaces even if $p_{EQ_i} \le \max\left\{w_i, (R_i - R_j + w_j) - (R_j - w_j + cm_a)\left[\frac{\rho}{\alpha_i} - 1\right]\right\}$. Hence, the only possible limit for a sequence of equilibrium prices $(p_{EQ_i}^k, p_{EQ_j}^k)$ is $(R_i - R_j + w_j, 0)$.

Existence: The existence of the sequence holds as in the proof of Theorem 7.2.c when we have $\lim_{k\to\infty} \delta^k = 0$. If $\lim_{k\to\infty} \delta^k = \tilde{\delta} > 0$, we can show the existence of such a sequence only for $\rho < 1 - \tilde{\delta}$.

3. Any sequence of price pairs with limit $p_i \ne R_i - R_j + p_j$ can not emerge as a price pair in large marketplaces as shown in part 2, and thus we are only left with the case where $p_i = R_i - R_j + p_j$. Furthermore, we rule out all the sequences with limit $p_j < R_j$ by Lemma 13.

Existence: The existence of the sequence also holds as in Theorem 5.

D.9. Proof of Proposition 9

We first want to note that the equivalent of our claim is that $\lim_{k\to\infty} p_n^k = R$ for all $n \in \{1, \ldots, N\}$. To prove this result, it is sufficient to show that the given sequence of equilibrium prices always converge to one price, i.e. $\lim_{k\to\infty} p_n^k = \tilde{p}$ for all $n \in \{1, \ldots, N\}$. Then, using Theorem 5, we should have that $\tilde{p} = R$.

To prove that $\lim_{k\to\infty} p_n^k = \tilde{p}$ for all $n \in \{1, \ldots, N\}$, suppose $\lim_{k\to\infty} p_n^k = \tilde{p}_n$, where $\tilde{p}_n \ne \tilde{p}_m$ for any $n \ne m$, $n \in \{1, \ldots, N\}$, $m \in \{1, \ldots, N\}$ (Note that prices can also converge to $N' < N$ limits. The proof for such a case is the same). First, we note that the case such that $\tilde{p}_n > 0$ for all $n \in \{1, \ldots, N\}$ is already ruled out in Proposition 8. The only case that is remained to be ruled out is that $\tilde{p}_1 = 0$ by assuming $\tilde{p}_1 < \cdots < \tilde{p}_N$ without loss of generality. Observe that in such a case demand for sub-pool-1 exceeds the capacity of the sub-pool as $\rho > 1$. Therefore, there is always room for a group of agents to increase their prices and improve their revenues as we show in Lemma 12. Hence, this case also cannot be true.

SUPPORTING DOCUMENT

This supporting document presents the proofs of the supplementary results in the paper. It also provides the extended proof of Theorem 6 in Section S.4.

Appendix S.1: No-intervention (Identical Agents)

S.1.1. Proof of Lemma 1

1. After a birth-death chain analysis of an $M/M/1+M$ system with arrival rate λ, service rate 1, and abandonment rate $1/m_a$, we have that

$$\beta(\lambda) = 1 - \frac{g(\lambda)}{\lambda(1+g(\lambda))}, \text{ and } W(\lambda) = m_a \beta(\lambda),$$

where $a_0 = 1$, $a_n = \frac{1}{\prod_{i=0}^{n-1}(1+i/m_a)} = \frac{m_a^n}{\prod_{i=0}^{n-1}(m_a+i)}$ for any $n \geq 1$, and $g(\lambda) = \sum_{n=1}^{\infty} a_n \lambda^n$. As in Ward and Glynn (2003), $g(\lambda)$ can be written as $g(\lambda) = [\lambda m_a]^{1-m_a} e^{\lambda m_a} \int_0^{\lambda m_a} t^{1-m_a} e^{-t} dt$.

The above representation of $g(\lambda)$ is clearly continuous and continuously twice differentiable, so that $\beta(\lambda)$ is also continuous and continuously twice differentiable. Furthermore, using the above representation,

$$\begin{aligned} g'(\lambda) &= m_a[1+g(\lambda)] - (m_a - 1)\frac{g(\lambda)}{\lambda} = \sum_{n=0}^{\infty}(n+1)a_{n+1}\lambda^n \\ g''(\lambda) &= m_a g'(\lambda) - (m_a - 1)\left[\frac{g'(\lambda)}{\lambda} - \frac{g(\lambda)}{\lambda^2}\right] = \sum_{n=0}^{\infty}(n+1)(n+2)a_{n+2}\lambda^n. \end{aligned}$$

2. Observe that

$$\frac{d\beta(\lambda)}{d\lambda} = -\frac{\lambda g'(\lambda) - g(\lambda)(1+g(\lambda))}{\lambda^2(1+g(\lambda))^2}.$$

Since $g(\lambda) > 0$ for any $\lambda > 0$, it is sufficient to show that $g(\lambda)(1+g(\lambda)) - \lambda g'(\lambda) > 0$, and this holds as follows

$$\begin{aligned} \lambda g'(\lambda) - g(\lambda) &= \sum_{n=1}^{\infty}(n-1)a_n\lambda^n. \\ [g(\lambda)]^2 &= \sum_{n=1}^{\infty}\lambda^n\left(\sum_{k=1}^{n-1}a_{n-k}a_k\right) > \sum_{n=1}^{\infty}(n-1)a_n\lambda^n, \end{aligned}$$

where the inequality holds since $a_{n-k}a_k > a_n$.

3 Observe that

$$\frac{\lambda\beta'(\lambda)}{1-\beta(\lambda)} = 1 - \frac{\lambda g'(\lambda)}{g(\lambda)[1+g(\lambda)]}.$$

Therefore, we have that

$$\frac{d}{d\lambda}\left[\frac{\lambda\beta'(\lambda)}{1-\beta(\lambda)}\right] = -\frac{\lambda g(\lambda)\Big[g''(\lambda)[1+g(\lambda)]-2[g'(\lambda)]^2\Big]+g'(\lambda)\Big[g(\lambda)[1+g(\lambda)]-\lambda g'(\lambda)\Big]}{\Big[g(\lambda)[1+g(\lambda)]\Big]^2}.$$

Since $g(\lambda)>0$ for any $\lambda>0$, it is sufficient to show that

$$\lambda g(\lambda)\Big[g''(\lambda)[1+g(\lambda)]-2[g'(\lambda)]^2\Big]+g'(\lambda)\Big[g(\lambda)[1+g(\lambda)]-\lambda g'(\lambda)\Big]<0.$$

As we show above, $g(\lambda)(1+g(\lambda))-\lambda g'(\lambda)>0$. Therefore, we have that

$$\begin{aligned}\lambda g(\lambda)\Big[g''(\lambda)[1+g(\lambda)]-2[g'(\lambda)]^2\Big]+g'(\lambda)\Big[g(\lambda)[1+g(\lambda)]-\lambda g'(\lambda)\Big]<\\ \lambda g(\lambda)\Big[g''(\lambda)[1+g(\lambda)]-2[g'(\lambda)]^2\Big]+2g'(\lambda)\Big[g(\lambda)[1+g(\lambda)]-\lambda g'(\lambda)\Big]\\ =[1+g(\lambda)]\Big[g(\lambda)[\lambda g''(\lambda)+2g'(\lambda)]-2\lambda[g'(\lambda)]^2\Big] \qquad (14)\end{aligned}$$

Using the definition of $g(\lambda)$ and its derivatives, we have that

$$g(\lambda)[\lambda g''(\lambda)+2g'(\lambda)] = \sum_{n=1}^{\infty}\lambda^n\left[\sum_{k=1}^{n}(n+1-k)(n+2-k)a_k a_{n+1-k}\right].$$

Similarly, we also have that

$$2\lambda[g'(\lambda)]^2 = \sum_{n=1}^{\infty}\lambda^n\left[\sum_{k=1}^{n}(n+1-k)k a_k a_{n+1-k}\right].$$

Combining these two equalities, we obtain that

$$\begin{aligned}g(\lambda)[\lambda g''(\lambda)+2g'(\lambda)]-2\lambda[g'(\lambda)]^2 &= \sum_{n=1}^{\infty}\lambda^n\left[\sum_{k=1}^{n}(n+1-k)(n+2-3k)a_k a_{n+1-k}\right]\\ &= 2\lambda^3(a_1a_3-a_2^2)+\sum_{n=4}^{\infty}\lambda^n\left[\sum_{k=1}^{n}(n+1-k)(n+2-3k)a_k a_{n+1-k}\right].\end{aligned}$$

Let $s_k=(n+2)(n+1)-6k(n+1-k)$ for any $1\le k\le\lfloor\frac{n}{2}\rfloor$, and

$$s_{\lfloor\frac{n}{2}\rfloor+1}=\begin{cases}-(n+1)(n-1)/4 & \textit{if } n \textit{ is odd}\\ 0 & \textit{if } n \textit{ is even.}\end{cases}$$

Then, we have that $\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor+1} s_k a_k a_{n+1-k} = \sum_{k=0}^{n+1}(n+1-k)(n+2-3k)a_k a_{n+1-k}$.

Following observations can be proven easily.

1. $\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor+1} s_k = \sum_{k=0}^{n+1}(n+1-k)(n+2-3k)=0$.
2. $s_{k-1}\ge s_k$ for any $2\le k\le\lfloor\frac{n}{2}\rfloor$.

3. $s_1 = (n-2)(n-1) \geq 0$, $s_{\lfloor \frac{n}{2} \rfloor + 1} \leq 0$ for any $n > 3$.
4. $s_{\lfloor \frac{n}{2} \rfloor} = \begin{cases} -(n^2 - 13)/2 & \textit{if } n \textit{ is odd} \\ -(n+2)(n/2-1) & \textit{if } n \textit{ is even} \end{cases}$. Therefore, $s_{\lfloor \frac{n}{2} \rfloor} < 0$ for any $n > 3$.
5. For any $n > 3$, there exists $\bar{k}_n < \lfloor \frac{n}{2} \rfloor$ such that $s_{\bar{k}_n} \geq 0 \geq s_{\bar{k}_n + 1}$.

Using the above observations and $a_k a_{n+1-k} < a_{k+1} a_{n-k}$ for any $k \leq \lfloor \frac{n}{2} \rfloor$,

$$\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor + 1} s_k a_{n+1-k} a_k < \sum_{k=0}^{\bar{k}_n} s_k a_{n+2-\bar{k}_n} a_{\bar{k}_n} + \sum_{k=\bar{k}_n+1}^{\lfloor \frac{n}{2} \rfloor + 1} s_k a_{n+2-\bar{k}_n} a_{\bar{k}_n} = 0,$$

which implies that $g(\lambda)[\lambda g''(\lambda) + 2g'(\lambda)] - 2\lambda[g'(\lambda)]^2 < 0$. Finally by Equation 14, we have that

$$\lambda g(\lambda)\Big[g''(\lambda)[1+g(\lambda)] - 2[g'(\lambda)]^2\Big] + g'(\lambda)\Big[g(\lambda)[1+g(\lambda)] - \lambda g'(\lambda)\Big] < 0.$$

4. Let $f(\lambda) = \lambda(1-\beta(\lambda))$ for notational simplicity. First, observe that

$$\frac{df(\lambda)}{d\lambda} = \frac{g'(\lambda)}{[1+g(\lambda)]^2} > 0,$$

since $g(\lambda)$ is strictly increasing in λ. Moreover, we have that

$$\frac{d^2 f(\lambda)}{d\lambda^2} = \frac{g''(\lambda)[1+g(\lambda)] - 2[g'(\lambda)]^2}{[1+g(\lambda)]^3}.$$

Therefore, $f(\lambda)$ is concave in λ if and only if $g''(\lambda)[1+g(\lambda)] - 2[g'(\lambda)]^2 < 0$. In the previous parts, we show that

$$\begin{aligned} &\lambda g(\lambda)\Big[g''(\lambda)[1+g(\lambda)] - 2[g'(\lambda)]^2\Big] + g'(\lambda)\Big[g(\lambda)[1+g(\lambda)] - \lambda g'(\lambda)\Big] < 0 \\ &g(\lambda)[1+g(\lambda)] - \lambda g'(\lambda) > 0. \end{aligned}$$

Combining these two inequalities, and using the fact that $g(\lambda) > 0$ for any $\lambda > 0$, we have that $g''(\lambda)[1+g(\lambda)] - 2[g'(\lambda)]^2 < 0$.

5. Using $\beta'(\lambda)$ in part 2, we have that

$$\begin{aligned} \beta''(\lambda) &= \frac{\lambda[1+g][2gg' - \lambda g''] - 2\big[1+g+\lambda g'\big]\big[g[1+g] - \lambda g'\big]}{\big[\lambda[1+g]\big]^3} \\ &= \frac{2\lambda g'[\lambda g' + 1 + g] - [1+g]\big[\lambda^2 g'' + 2g[1+g]\big]}{\big[\lambda[1+g]\big]^3}. \end{aligned}$$

Therefore, we need to show $2\lambda g'[\lambda g' + 1 + g] - [1+g]\big[\lambda^2 g'' + 2g[1+g]\big] < 0$ in order to show the concavity of $\beta(\lambda)$. Using the definition of $g(\lambda)$ and its derivative, we have that

$$\begin{aligned} 2\lambda g'[\lambda g' + 1 + g] &= 2\lambda\Big[1 + g + (m_a - 1)[1 + g - g/\lambda]\Big]\Big[\lambda[1+g] + \lambda(m_a-1)[1+g-g/\lambda] + 1 + g\Big] \\ &= 2\lambda^2(\lambda+1)[1+g]^2 + 2\lambda(2\lambda+1)(m_a-1)[1+g][1+g-g/\lambda] \end{aligned}$$

$$
\begin{aligned}
&\quad +2\big[\lambda(m_a-1)[1+g-g/\lambda]\big]^2, \\
[1+g]\big[\lambda^2 g''+2g[1+g]\big] &= [1+g]\left[\lambda^2 m_a\big[1+g+\frac{(m_a-1)(\lambda-1)}{\lambda}[1+g-g/\lambda]\big]+2g[1+g]\right] \\
&= [2g+\lambda^2 m_a][1+g]^2+\lambda(\lambda-1)m_a(m_a-1)[1+g][1+g-g/\lambda].
\end{aligned}
$$

Combining above equations, we have that

$$
\begin{aligned}
&2\lambda g'[\lambda g'+1+g]-[1+g]\big[\lambda^2 g''+2g[1+g]\big] \\
&\quad =[1+g]^2\big[\lambda^2(2-m_a)+2\lambda-2g\big]+\lambda(m_a-1)[1+g][1+g-g/\lambda][4\lambda+2-m_a(\lambda-1)] \\
&\quad +2\big[\lambda(m_a-1)[1+g-g/\lambda]\big]^2 \\
&\quad =[1+g]^2\big[\lambda^2(2-m_a)+2\lambda-2g\big]+\lambda(\lambda+1)(2+m_a)(m_a-1)[1+g][1+g-g/\lambda] \\
&\quad -2\lambda(m_a-1)^2 g[1+g-g/\lambda] \\
&\quad \leq [1+g]\Big[[1+g]\big[\lambda^2(2-m_a)+2\lambda-2g\big]+[1+g-g/\lambda][\lambda(\lambda+1)(2+m_a)(m_a-1)]\Big] \\
&\quad =[1+g]\Big[[1+g]\big[\lambda^2 m_a^2+\lambda m_a(m_a+1)-2g\big]-g\big[(\lambda+1)(2+m_a)(m_a-1)\big]\Big] \\
&\quad =[1+g]\begin{bmatrix}[1+g]\big[\lambda^2 m_a^2+\lambda m_a(m_a+1)-m_a(m_a+1)g\big] \\ -g\big[[1+g][2-m_a(m_a+1)]+(\lambda+1)(2+m_a)(m_a-1)\big]\end{bmatrix} \\
&\quad =[1+g]\Big[[1+g]m_a(m_a+1)\big[a_2\lambda^2+\lambda-g\big]-g\big[g-\lambda\big][2-m_a(m_a+1)]\Big]<0,
\end{aligned}
$$

when $m_a \leq 1$ because $a_2\lambda^2+\lambda-g<0$, $g>\lambda$ for any $\lambda>0$, and $2-m_a(m_a+1)\geq 0$ when $m_a\leq 1$.

S.1.2. Proof of Lemma 2

Let $p_{max} = \max_{n \in S} p_n$, and $p'_{max} = \max_{n \in S'} p_n$. Note that if an agent attracts some demand, then all agents charging a lower price should also attract some demand. Therefore, we can write

$$S = \{n \leq k : p_n \leq p_{max}\}, \ and \ S' = \{n \leq k : p_n \leq p'_{max}\},$$

and it is enough to show $p_{max} = p'_{max}$. Suppose NOT, WLOG assume $p_{max} < p'_{max}$, i.e. $|S| < |S'|$. Then, we have that

$$0 < R - p'_{max} \leq U(\Lambda D_n, p_n) \ for \ any \ n \in S,$$

since $p'_{max} < R$ and by the definition of Customer Equilibrium. The above inequality implies that $\sum_{n=1}^{k} D_n = 1$. Then, since $|S| < |S'|$, there exits an $n^* \in S$ such that $D_{n^*} > D'_{n^*}$ (Otherwise, we would have that $\sum_{n=1}^{k} D'_n > 1$.) However, this leads to the following contradiction

$$R - p'_{max} \leq U(\Lambda D_{n^*}, p_{n^*}) < U(\Lambda D'_{n^*}, p_{n^*}) \leq R - p'_{max}.$$

Hence $p_{max} = p'_{max}$.

S.1.3. Proof of Lemma 3

Recall that the best response problem of agent-ℓ can be rewritten as follows:

$$\begin{aligned} \max_{p_\ell \geq 0, D_\ell \geq 0, D_{-\ell} \geq 0} \quad & p_\ell \Lambda D_\ell \left[1 - \beta(\Lambda D_\ell)\right] \\ s.to \quad & \\ & (R - p_\ell + cm_a)\left[1 - \beta(\Lambda D_\ell)\right] - cm_a \geq 0 \\ & (R - p_\ell + cm_a)\left[1 - \beta(\Lambda D_\ell)\right] = (R - p + cm_a)\left[1 - \beta(\Lambda D_{-\ell})\right] \\ & D_\ell + (k-1)D_{-\ell} \leq 1 \end{aligned}$$

and the FOC of this problem are:

$$\lambda D_\ell - \eta_1 - \eta_2 = 0, \tag{15}$$

$$\lambda p_\ell[1 - \beta(\lambda D_\ell)] - \lambda^2 p_\ell D_\ell \beta'(\lambda D_\ell) - \lambda^2 D_\ell (R - p_\ell + cm_a)\beta'(\lambda D_\ell) - \eta_3 = 0, \tag{16}$$

$$\eta_2 \lambda (R - p + cm_a)\beta'(\lambda D_{-\ell}) - (k-1)\eta_3 = 0, \tag{17}$$

$$\eta_1\big((R - p_\ell + cm_a)[1 - \beta(\lambda D_\ell)] - cm_a\big) = 0, \tag{18}$$

$$\eta_3(1 - D_\ell - (k-1)D_{-\ell}) = 0, \tag{19}$$

$$\eta_1, \eta_3 \geq 0. \tag{20}$$

where η_1, η_2, and η_3 are the Lagrangian multipliers of the constraints 1, 2, and 3 of the best response problem of agent-ℓ, respectively. Moreover, we denote the solution to the above problem by $(D_\ell(p), D_{-\ell}(p), p_\ell(p))$ for a given p.

Case-1 ($\Lambda \geq k\lambda^{mon}$)**:** When $p = R + cm_a - \frac{cm_a}{1-\beta(\lambda^{mon})}$, it is feasible for a single agent to charge its monopoly price which is exactly $R + cm_a - \frac{cm_a}{1-\beta(\lambda^{mon})}$. Then, we have that $p_\ell(p) = p$, and thus p is clearly the symmetric equilibrium in this case.

Case-2 ($\lambda^0 \leq \Lambda < k\lambda^{mon}$)**:**

CLAIM 1. *Let* $(D_\ell(p), D_{-\ell}(p), p_\ell(p))$ *be the solution of single agent's best response problem when all other agents charge the price* p. *If* $p = R + cm_a - \frac{cm_a}{1-\beta(\Lambda/k)}$, *then we have that* $(R - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] = cm_a$.

Proof:

Suppose NOT. Then, we have that $\eta_1 = 0$ and this implies that $\eta_3 > 0$. Moreover, we also have that $D_{-\ell}(p) < 1/k$. Since $D_\ell(p) + (k-1)D_{-\ell}(p) = 1$ when $\eta_3 > 0$, $D_{-\ell}(p) < 1/k$ implies that $D_\ell(p) > 1/k$.

Let $T = \frac{\Lambda/k\beta'(\Lambda/k)}{1-\beta(\Lambda/k)}$ for notational simplicity. Then, using FOC, we have that

$$\begin{aligned}
\eta_2 \frac{cm_a}{1-\beta(\Lambda/k)}\beta'(\Lambda D_{-\ell}(p)) &= (k-1)\eta_3/\Lambda \\
&= (k-1)[p_n^*(p)[1-\beta(\Lambda D_\ell(p))] - (R+cm_a)\Lambda D_\ell(p)\beta'(\Lambda D_\ell(p))] \\
&\leq (k-1)(R+cm_a)[1-\beta(\Lambda D_\ell(p)) - \Lambda D_\ell(p)\beta'(\Lambda D_\ell(p))] - (k-1)cm_a \\
&< (k-1)(R+cm_a)[1-\beta(\Lambda/k) - \Lambda/k\beta'(\Lambda/k)] - (k-1)cm_a \\
&= (k-1)(R+cm_a)(1-T)[1-\beta(\Lambda/k)] - (k-1)cm_a,
\end{aligned}$$

where first inequality holds since $p_\ell(p) \leq R + cm_a - \frac{cm_a}{1-\beta(\Lambda D_\ell(p))}$, second inequality holds since $D_\ell(p) > 1/k$ and $\lambda[1-\beta(\lambda)]$ is strictly concave. Using the facts that $\eta_2 = \Lambda D_\ell(p) > \Lambda/k$, $D_{-\ell}(p) < 1/k$, and $\beta(\lambda)$ is increasing and concave, the above inequality implies that

$$\begin{aligned}
&Tcm_a < \eta_2 \frac{cm_a}{1-\beta(\Lambda/k)}\beta'(\Lambda D_{-\ell}(p)) < (k-1)(R+cm_a)(1-T)[1-\beta(\Lambda/k)] - (k-1)c \\
&\Rightarrow (1-T)[1-\beta(\Lambda/k)]\left[(R+cm_a)(k-1) - \frac{c}{1-\beta(\Lambda/k)}\left(\frac{k}{1-T} - 1\right)\right] > 0 \\
&\Rightarrow z(\Lambda) > 0,
\end{aligned}$$

which is a contradiction since $\Lambda \geq \lambda^0$ and $z(\lambda)$ is decreasing in λ.

■

Using the above claim, we will argue that $\eta_3 > 0$. If $\eta_3 = 0$, we would have that $p_\ell(p) = R + cm_a - \frac{cm_a}{1-\beta(\lambda^{mon})}$ and $\Lambda D_\ell(p) = \lambda^{mon}$ since the customer surplus is zero by the above claim. However, this would imply that $D_\ell(p) + (k-1)D_{-\ell}(p) = \frac{\lambda^{mon}}{\Lambda} + \frac{k-1}{k} > 1$ which is a contradiction. Hence, we should have that $\eta_3 > 0$.

Finally, $\eta_3 > 0$ and the above claim jointly imply that $D_\ell(p) = D_{-\ell}(p) = 1/k$. Therefore, if $p = R + cm_a - \frac{cm_a}{1-\beta(\Lambda/k)}$, then we will have that $D_\ell(p) = 1/k$ and $p_\ell(p) = R + cm_a - \frac{cm_a}{1-\beta(\Lambda/k)}$ under the assumption that $\beta(\lambda)$ is concave.

Case 3 ($\Lambda < \lambda^0$)**:**

CLAIM 2. *Let* $(D_\ell(p), D_{-\ell}(p), p_\ell(p))$ *be the solution of single agent's best response problem when all other agents charge the price* p. *If* $p = R + cm_a - \frac{(R+cm_a)(k-1)}{\frac{k}{1-T}-1}$, *then we have that* $(R - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] = \Delta$, *where*

$$\Delta = \frac{(R-p)(k-1)[1-\nu(\Lambda/k)][1-\beta(\Lambda/k)]}{k-1+\nu(\Lambda/k)}.$$

Proof:
Suppose $(R-p_\ell(p)+cm_a)[1-\beta(\Lambda D_\ell(p))]<\Delta$. Then, we have that $D_{-\ell}(p)>1/k$ since
$$(R-p+cm_a)[1-\beta(\Lambda/k)]=\Delta>=(R-p+cm_a)[1-\beta(\Lambda D_{-\ell}(p))].$$
Moreover $D_{-\ell}(p)>1/k$ implies that $D_\ell(p)<1/k$ since $D_\ell(p)+(k-1)D_{-\ell}(p)\le 1$.

Then, using FOC (16), we have that
$$\begin{aligned}\eta_2 &= \left(\frac{(k-1)}{(R-p+cm_a)\beta'(\Lambda D_{-\ell})}\right)\frac{\eta_3}{\Lambda}\\ &> \left(\frac{(k-1)}{(R-p+cm_a)\beta'(\Lambda/k)}\right)\frac{\eta_3}{\Lambda},\end{aligned}$$
where the inequality holds since $D_{-\ell}>1/k$ and $\beta(\lambda)$ is concave. Moreover, using (17), we have that
$$\begin{aligned}\frac{\eta_3}{\Lambda} &= p_\ell(p)[1-\beta(\Lambda D_\ell(p))]-\Lambda D_\ell(R+cm_a)\beta'(\Lambda D_\ell(p))\\ &> (R+cm_a)[1-\beta(\Lambda D_\ell(p))-\Lambda D_\ell\beta'(\Lambda D_\ell(p))]-\Delta\\ &= (R+cm_a)[1-\nu(\Lambda D_\ell(p))][1-\beta(\Lambda D_\ell(p))]-\Delta\\ &> (R+cm_a)[1-\nu(\Lambda/k)][1-\beta(\Lambda/k)]-\Delta\end{aligned}$$
where the first inequality holds since $p_\ell(p)=R+cm_a-\frac{\Delta}{1-\beta(\Lambda D_\ell(p))}$, the second inequality holds since $D_\ell(p)<1/k$ and $[1-\nu(\Lambda/k)][1-\beta(\Lambda/k)]$ is the derivative of $\lambda[1-\beta(\lambda)]$, which is strictly concave.

Using these two observations, and the facts that $\eta_2\le\Lambda D_\ell(p)<\Lambda/k$ (since $\eta_1\ge 0$) and $R-p+cm_a=\frac{\Delta}{1-\beta(\Lambda/k)}$, we have that
$$\begin{aligned}\Lambda/k &> \left(\frac{(k-1)}{\frac{\Delta}{1-\beta(\Lambda/k)}\beta'(\Lambda/k)}\right)\left[R+cm_a)[1-\nu(\Lambda/k)][1-\beta(\Lambda/k)]-\Delta\right]\\ &\Rightarrow R+cm_a)[1-\nu(\Lambda/k)][1-\beta(\Lambda/k)]-\Delta\left(\frac{k-1+\nu(\Lambda/k)}{k-1}\right)<0\\ &\Rightarrow \Delta>\frac{(R-p)(k-1)[1-\nu(\Lambda/k)][1-\beta(\Lambda/k)]}{k-1+\nu(\Lambda/k)}=\Delta,\end{aligned}$$
which is clearly a contradiction. Hence, we should have that $(R-p_\ell(p)+cm_a)[1-\beta(\Lambda D_\ell(p))]\ge\Delta$.

Now, we suppose $(R-p_\ell(p)+cm_a)[1-\beta(\Lambda D_\ell(p))]>\Delta$. As the same as above (only by reversing the inequality signs and using the fact that $\eta_1=0$ since $\Delta>cm_a$), we can again have a contradiction. Therefore, we should have that $(R-p_\ell(p)+cm_a)[1-\beta(\Lambda D_\ell(p))]=\Delta$.

■

As a direct implication of the above claim, we have that $D_{-\ell}=1/k$ since $(R-p+cm_a)[1-\beta(\Lambda D_{-\ell})]=\Delta$. Furthermore, since $\Delta>0$, we have that $D_\ell=1/k$ since $D_\ell+(k-1)D_{-\ell}=1$. Finally, we have that $p_\ell=p$ since $D_\ell=1/k$ and $(R-p_\ell+cm_a)[1-\beta(\Lambda D_\ell)]=\Delta$.

Appendix S.2: Operational Efficiency (Identical Agents)

S.2.1. Proof of Lemma 4

When $a_k=1$ for any k, the result follows by Theorem 1 in Garnett et al. (2002).

Now, we prove the result for the case when $a_k\neq 1$ for any k. Consider any convergent subsequence of $\{\beta(b_k k,a_k k)\}_{k=1}^{\infty}$, say $\{\beta(b_{k(r)}k(r),a_{k(r)}k(r))\}_{r=1}^{\infty}$. WLOG we can assume that $a_{k(r)}k(r)<a_{k(r+1)}k(r+1)$ for any $r=1,2,\ldots$ (If not consider a subsequence which satisfies that condition).

Let $N(r)=a_{k(r)}k(r)$ and $\hat{b}_{N(r)}=b_{k(r)}/a_{k(r)}$ for any $r=1,2,\ldots$. Then, we have that
$$\beta(b_{k(r)}k(r),a_{k(r)}k(r))=\beta(\hat{b}_{N(r)}N(r),N(r)),$$

for any $r = 1, 2, \ldots$. Observe that $\lim_{r\to\infty} b_{N(r)} = \tilde{b}/\tilde{a}$. Therefore, we have that

$$\lim_{r\to\infty} \beta(\hat{b}_{N(r)} N(r), N(r)) = \max\{0, 1 - \tilde{a}/\tilde{b}\},$$

by Theorem 1 in Garnett et al. (2002). This shows that any convergent subsequence of $\{\beta(b_k k, a_k k)\}_{k=1}^{\infty}$ converges to the same limit, $\max\{0, 1 - \tilde{a}/\tilde{b}\}$. Since $\beta(b_k k, a_k k) \in [0,1]$, it also converges to that limit.

S.2.2. Proof of Lemma 5

Let π_n be the steady state probability of having n customers in sub-pool-2 which consists of a single agent. By the birth-death chain analysis, we have that

$$\begin{aligned}
\pi_n &= \frac{\lambda_2 \pi_{n-1}}{1 + (n-1)/m_a} \text{ for any } n > 1, \\
\pi_1 &= (\lambda_1 + \lambda_2)\pi_0 + \sum_{j=2}^{\infty} \hat{\pi}_{j1}, \\
\pi_0 &= \frac{1 - \left(\sum_{j=2}^{\infty} \hat{\pi}_{j1}\right) \frac{g(\lambda_2)}{\lambda_2}}{1 + (\lambda_1 + \lambda_2)\frac{g(\lambda_2)}{\lambda_2}} \le \frac{1}{1 + (\lambda_1 + \lambda_2)\frac{g(\lambda_2)}{\lambda_2}},
\end{aligned}$$

where $\hat{\pi}_{j1}$ is the steady state probability of having j customers in sub-pool-1 while sub-pool-2 has only one customer, and $g(\lambda) = \sum_{n=1}^{\infty} \frac{\lambda^n}{\prod_{i=0}^{n-1}(1+i/m_a)}$. Note that

$$\pi_0 \le \frac{1}{1 + \lambda_1 + \lambda_2}$$

since $g(\lambda)/\lambda = 1 + \frac{\lambda}{1+1/m_a} + \cdots \ge 1$ for any $\lambda \ge 0$. Using this observation, we have that

$$\sigma_2(\lambda_1, \lambda_2; k-1, 1) \;=\; 1 - \pi_0 \;\ge\; 1 - \frac{1}{1 + \lambda_1 + \lambda_2} \;=\; \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}.$$

S.2.3. Proof of Lemma 6

1. Suppose NOT. Then, there exists a subsequence such that

$$\lim_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k-1} \le 1.$$

We first want to note that

$$\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \le \beta^{MM1}(\Lambda^k D_2^{MCE}(k), k-1).$$

where $\beta^{MM1}(\lambda, k)$ is the probability of abandonment in $M/M/1+M$ system with arrival rate λ, service rate k, and abandonment rate $1/m_a$. Since $\lim_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k-1} \le 1$, we have that

$$\lim_{k\to\infty} \beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \le \lim_{k\to\infty} \beta^{MM1}(\Lambda^k D_2^{MCE}(k), k-1) = 0, \tag{21}$$

where the last equality is due to Ward and Glynn (2003). Using this result, we have that

$$\lim_{k\to\infty} U_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) = R - p > 0, \tag{22}$$

which implies that utility of customers choosing the price p^k is strictly positive for large k, so that we should have $D_1^{MCE}(k) + D_2^{MCE}(k) = 1$ for large k by the definition of Market Customer Equilibrium. Furthermore, using the fact that the rate of arrival to sub-pool-2 is equal to the rate of departure (either by service or abandonment), we have that

$$\begin{aligned}&\Lambda^k D_1^{MCE}(k) PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) + \Lambda^k D_2^{MCE}(k)\\ &= (k-1)\sigma_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) + \Lambda^k D_2^{MCE}(k)\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1).\end{aligned}$$

Dividing both sieds by Λ^k, the above equation implies that

$$\begin{aligned}&PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1)\left[D_1^{MCE}(k) + D_2^{MCE}(k)\right]\\ &\le \frac{k-1}{\Lambda^k}\sigma_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) + D_2^{MCE}(k)\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1).\end{aligned}$$

Letting k go to infinity, we have that

$$\lim_{k\to\infty} PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \le \frac{1}{\rho}. \tag{23}$$

For notational convenience, we let $\hat{P}_{pool}(k) = PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1)$, and $\beta_{one}(k) = \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1)$. Then, we have that

$$\begin{aligned}&\lim_{k\to\infty} U_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1)\\ &= \left[1 - \lim_{k\to\infty}\hat{P}_{pool}(k)\right]\lim_{k\to\infty}\left[(R - \hat{p}^k + cm_a)[1-\beta_{one}(k)] - cm_a\right] + (R-p)\lim_{k\to\infty}\hat{P}_{pool}(k)\\ &\le \left[1 - \lim_{k\to\infty}\hat{P}_{pool}(k)\right]\lim_{k\to\infty}\left[(R - \hat{p}^k + cm_a)[1-\beta^M(\Lambda^k;k)] - cm_a\right] + (R-p)\lim_{k\to\infty}\hat{P}_{pool}(k)\\ &\le \left(1-\frac{1}{\rho}\right)\left[(R-p+cm_a)/\rho - cm_a\right] + \frac{1}{\rho}(R-p)\\ &< \left(1-\frac{1}{\rho}\right)\left[(R-p)/\rho\right] + \frac{1}{\rho}(R-p)\\ &< R-p,\end{aligned} \tag{24}$$

where the first inequality holds since $D_1^{MCE}(k) + D_2^{MCE}(k) = 1$ for large k and customers choosing the deviating agent has to wait customers choosing p^k (thus the system after deviation is less efficient for customers choosing the deviating agent than its multi-server equivalent), and the second inequality holds since $p' \ge p$, $\lim_{k\to\infty}[1-\beta^M(\Lambda^k;k)] = 1/\rho$ (as shown in Lemma 4) and by (23), and the last two strict inequality holds as $\rho > 1$.

Combining (22) and (24), we have that

$$U_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) < U_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1),$$

for large k. However, this contradicts with the definition of Market Customer Equilibrium since

$$\lim_{k\to\infty} D_1^{MCE}(k) = 1 - \lim_{k\to\infty} D_2^{MCE}(k) \geq 1 - \frac{1}{\rho} > 0.$$

Hence, we should have that $\lim\limits_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k-1} > 1$.

2. Let π_n be the steady-state probability of having n customers in sub-pool-2, π_n^H be the steady-state probability of having n customers in a hypothetical sub-pool-2 which serves customers from sub-pool-1 only upon their arrival, and π_n^M be the steady-state probability of having n customers in an $M/M/(k-1)+M$ system with arrival rate $\Lambda^k D_2^{MCE}(k)$, service rate 1, and abandonment rate $1/m_a$. By studying the birth-death chain of all these systems, we have that

$$\begin{aligned}
\sum_{n=0}^{k-1} \pi_n \leq \sum_{n=0}^{k-1} \pi_n^H &= \frac{\sum\limits_{n=0}^{k-1} \frac{\left[\Lambda^k (D_1^{MCE}(k)+D_2^{MCE}(k))\right]^{n-k+1}(k-1)!}{n!}}{\sum\limits_{n=0}^{k-1} \frac{\left[\Lambda^k (D_1^{MCE}(k)+D_2^{MCE}(k))\right]^{n-k+1}(k-1)!}{n!} + \sum\limits_{n=k}^{\infty} \frac{\left[\Lambda^k D_2^{MCE}(k)\right]^{n-k+1}}{\prod_{i=1}^{n-k+1}(k+i/m_a)}} \\
&\leq \frac{\sum\limits_{n=0}^{k-1} \frac{\left[\Lambda^k D_2^{MCE}(k)\right]^{n-k+1}(k-1)!}{n!}}{\sum\limits_{n=0}^{k-1} \frac{\left[\Lambda^k D_2^{MCE}(k)\right]^{n-k+1}(k-1)!}{n!} + \sum\limits_{n=k}^{\infty} \frac{\left[\Lambda^k D_2^{MCE}(k)\right]^{n-k+1}}{\prod_{i=1}^{n-k+1}(k+i/m_a)}} \\
&= \sum_{n=0}^{k-1} \pi_n^M,
\end{aligned}$$

where the first inequality holds since the rate pushing the number of customers from zero towards $k-1$ is lower in the hypothetical sub-pool-2, and the second inequality holds since $\frac{x}{x+A}$, where $A > 0$ is a constant, is increasing in x.

Using the above relation, we have that

$$\begin{aligned}
\lim_{k\to\infty} PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) &= \lim_{k\to\infty} \sum_{n=0}^{k-1} \pi_n \\
&\leq \lim_{k\to\infty} \sum_{n=0}^{k-1} \pi_n^M = 0,
\end{aligned}$$

where the last equality holds since $\liminf\limits_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k-1} > 1$.

3. We first want to note that

$$\beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \leq \beta^M(\Lambda^k D_1^{MCE}(k); 1)$$

because some of the customer choosing $\hat{p}^k$ may be served by sub-pool-2. Therefore, it is sufficient to show that

$$\liminf_{k\to\infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \geq \beta^M(\tilde{\lambda}; 1).$$

To show that we consider a hypothetical situation where any customer choosing the price $\hat{p}^k$ is duplicated when there is an idle agent in sub-pool-2, and one of these copies goes to sub-pool-2 while the other one is colored and goes to sub-pool-1. Furthermore, any non-colored customer in sub-pool-1 has service priority.

This hypothetical sub-pool-1 operates as $M/M/1+M$ system with arrival rate $\Lambda^k D_1^{MCE}(k)$, so that total abandonment rate is $\Lambda^k D_1^{MCE}(k)\beta^M(\Lambda^k D_1^{MCE}(k); 1)$. Moreover, the abandonment rate of non-colored customers is the same as the abandonment rate in the real sub-pool-1. Then, we have that

$$\Lambda^k D_1^{MCE}(k)\beta^M(\Lambda^k D_1^{MCE}(k); 1) = \Lambda^k \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) + \Lambda^k \beta^{color}(k),$$

where $\Lambda^k \beta^{color}(k)$ is the rate that colored customers abandon the hypothetical system. It is clear that $\Lambda^k \beta^{color}(k) \leq \Lambda^k PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1)$. Thus, we have that

$$\begin{aligned}\beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \geq\ & \beta^M(\Lambda^k D_1^{MCE}(k); 1) \\ & -PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1).\end{aligned} \quad (25)$$

Finally, letting $k\to\infty$ and using part 2 provide the result we want.

4. We prove our claim by contradiction. Thus, we suppose that there exists a sub-sequence of $\{\Lambda^k D_1^{MCE}(k)\}_{k=1}^{\infty}$ such that $\Lambda^k D_1^{MCE}(k) > \bar{\lambda}$ for any k. Note that $\overline{\lambda} < \infty$ since we have that $\lim_{\lambda\to\infty} \beta^M(\lambda; 1) = 1$, and $c > 0$.

Let $PServ_{12}(k) = PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1)$ for notational convenience. Then, we have that

$$\begin{aligned}& U_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) \\ &\quad \leq \left[(R-\hat{p}^k + cm_a)\left[1-\beta^M(\Lambda^k D_1^{MCE}(k); 1) + PServ_{12}(k)\right] - cm_a\right]\left[1-PServ_{12}(k)\right] \\ &\quad\quad +(R-p^k)PServ_{12}(k) \\ &\quad < \left[(R-\hat{p}^k + cm_a)\left[1-\beta^M(\bar{\lambda}, 1) + PServ_{12}(k)\right] - cm_a\right]\left[1-PServ_{12}(k)\right] \\ &\quad\quad +(R-p^k)PServ_{12}(k),\end{aligned}$$

where the first inequality holds by (25), and the second one holds since $\Lambda^k D_1^{MCE}(k) > \bar{\lambda}$.

Since $PServ_{12}(k)$ converges to zero by part 2 and by the definition of $\bar{\lambda}$, the above inequality implies that $U_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) < 0$ for sufficiently large k. However, this contradicts with the definition of Market Customer Equilibrium. Hence, there should be a K^* such that $\Lambda^k D_1^{MCE}(k) \le \bar{\lambda}$ for any $k > K^*$.

5. Similar to part 2, we let π_n be the steady-state probability of having n customers in sub-pool-2. By the birth-death chain analysis, we have that $\left[k-1+(n-k+1)/m_a\right]\pi_n = \Lambda^k D_2^{MCE}(k)\pi_{n-1}$, for any $n > k-1$. Furthermore, we have that

$$\begin{aligned}\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; 1, k-1) &= \sum_{n=k}^{\infty}(n-k+1)/m_a \frac{\pi_n}{\Lambda^k D_2^{MCE}(k)} \\ &= \sum_{n=k}^{\infty}\left[\pi_{n-1} - \frac{(k-1)\pi_n}{\Lambda^k D_2^{MCE}(k)}\right] \\ &= \left(\sum_{n=k}^{\infty}\pi_n\right)\left(1 - \frac{k-1}{\Lambda^k D_2^{MCE}(k)}\right) + \pi_{k-1}.\end{aligned}$$

Then, the result follows by letting $k \to \infty$ and using the fact from part 2 that $\lim_{k\to\infty}\sum_{n=0}^{k-1}\pi_n = 0$.

6. Once we establish part 5, the proof of this part is very similar to the proof in Proposition 2. We first show that $\liminf_{k\to\infty} D_2^{MCE}(k) \ge \min\left\{1, \frac{R-p+cm_a}{\rho c m_a}\right\}$. This is true because otherwise we would have that customers choosing sub-pool-2 earn strictly positive utility while $D_1^{MCE}(k) + D_2^{MCE}(k) < 1$ and this would contradicts with the definition of *MCE*. Furthermore, we show that $\limsup_{k\to\infty} D_2^{MCE}(k) \le \min \frac{R-p+cm_a}{\rho c m_a}$. This is also true because otherwise we would have that customers choosing sub-pool-2 earn strictly negative utility while $D_2^{MCE}(k) > 0$, which contradicts with the definition of *MCE*.

Appendix S.3: Communication Enabled Model (Identical Agents)

S.3.1. Proof of Lemma 7

Proof of this lemma is very similar to the proof of Lemma 6.

1. Suppose NOT. Then, there exists a subsequence such that

$$\lim_{k\to\infty}\frac{\Lambda^k D_2^{MCE}(k)}{k - \lfloor \delta^k k \rfloor} \le 1.$$

Since $\lim_{k\to\infty}\frac{\Lambda^k D_2^{MCE}(k)}{k-\lfloor \delta^k k\rfloor} \le 1$, we have that

$$\begin{aligned}\lim_{k\to\infty}\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k\rfloor, k-\lfloor \delta^k k\rfloor) \le \\ \lim_{k\to\infty}\beta^{MM1}(\Lambda^k D_2^{MCE}(k), k-\lfloor \delta^k k\rfloor) = 0,\end{aligned} \tag{26}$$

where $\beta^{MM1}(\lambda, k)$ is the probability of abandonment in $M/M/1+M$ system with arrival rate λ, service rate k, and abandonment rate $1/m_a$. Using this result, we have that

$$\lim_{k\to\infty} U_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) = R - p > 0.$$

Thus, we have that $D_1^{MCE}(k) + D_2^{MCE}(k) = 1$ for large k.

Furthermore, using the fact that the rate of arrival to sub-pool-2 is equal to the rate of departure (either by service or abandonment) and (26), we have that

$$\lim_{k\to\infty} PServ_{12}(D_1^{MCE}(k), \Lambda^k D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) \le \frac{1}{\rho}.$$

Combining the above observations with the fact that $p' > p$, we have that

$$U_1(D_1^{MCE}(k), \Lambda^k D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) < U_2(D_1^{MCE}(k), \Lambda^k D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor),$$

for large k. However, this contradicts with the definition of Market Customer Equilibrium. Hence, we should have that $\lim_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k - \lfloor \delta^k k \rfloor} > 1$.

2. Let π_n be the steady-state probability of having n customers in sub-pool-2, and π_n^M be the steady-state probability of having n customers in an $M/M/(k - \lfloor \delta^k k \rfloor) + M$ system with arrival rate $\Lambda^k D_2^{MCE}(k)$, service rate 1, and abandonment rate $1/m_a$. By studying the birth-death chain of both systems, we have that

$$\begin{aligned} \lim_{k\to\infty} PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) &= \lim_{k\to\infty} \sum_{n=0}^{k-\lfloor \delta^k k \rfloor} \pi_n \\ &\le \lim_{k\to\infty} \sum_{n=0}^{k-\lfloor \delta^k k \rfloor} \pi_n^M = 0, \end{aligned}$$

where the last equality holds since $\liminf_{k\to\infty} \frac{\Lambda^k D_2^{MCE}(k)}{k - \lfloor \delta^k k \rfloor} > 1$.

3. To prove our claim, we first show that

$$\lim_{k\to\infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta k \rfloor) = \lim_{k\to\infty} \beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor).$$

Note that $\beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) \le \beta^M(D_1^{MCE}(k); \lfloor \delta^k k \rfloor)$, since some of the customers choosing sub-pool-1 can be served by sub-pool-2. Therefore, it is sufficient to show that

$$\liminf_{k\to\infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) \ge \lim_{k\to\infty} \beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor).$$

To show that we consider a hypothetical situation where any customer choosing the price $\hat{p}^k$ is duplicated when there is an idle agent in sub-pool-2, and one of these copies goes to sub-pool-2 while the other one is colored and goes to sub-pool-1. Furthermore, any non-colored customer in sub-pool-1 has service priority.

This hypothetical sub-pool-1 operates as $M/M/\lfloor \delta^k k \rfloor + M$ system with arrival rate $\Lambda^k D_1^{MCE}(k)$, so that total abandonment rate is $\Lambda^k D_1^{MCE}(k)\beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor)$. Moreover, the abandonment rate of non-colored customers is the same as the abandonment rate in the real sub-pool-1. Then, we have that

$$\Lambda^k D_1^{MCE}(k)\beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor) = \Lambda^k D_1^{MCE}(k)\beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) + \Lambda^k D_1^{MCE}(k)\beta^{color}(k),$$

where $\beta^{color}(k)$ is the probability that colored customers abandon the hypothetical system. It is clear that $\beta^{color}(k) \le PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor)$. Thus, we have that

$$\beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) \ge \beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor) - PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor).$$

Then, using this inequality and part 2, we have that

$$\liminf_{k \to \infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta k \rfloor) \ge \lim_{k \to \infty} \beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor).$$

Finally the result holds since $\lim_{k \to \infty} \beta^M(\Lambda^k D_1^{MCE}(k); \lfloor \delta^k k \rfloor) = \max\left\{0, 1 - \frac{1}{\rho \tilde{D}_1}\right\}$ by Lemma 4.

4. Similar to part 2, we let π_n be the steady-state probability of having n customers in sub-pool-2. By the birth-death chain analysis, we have that

$$\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); \hat{p}^k, p^k; \lfloor \delta^k k \rfloor, k - \lfloor \delta^k k \rfloor) = \left(\sum_{n = k - \lfloor \delta^k k \rfloor + 1}^{\infty} \pi_n \right) \left(1 - \frac{k - \lfloor \delta^k k \rfloor}{\Lambda^k D_2^{MCE}(k)} \right) + \pi_k.$$

Then, the result follows by letting $k \to \infty$ and using the fact from part 2 that $\lim_{k \to \infty} \sum_{n=0}^{k - \lfloor \delta^k k \rfloor} \pi_n = 0$.

Appendix S.4: No-intervention (Non-Identical Agents)

S.4.1. Supplementary Claims to Characterize SPNE

LEMMA 14. *Let*

$$\hat{U}_L(x,y) = (R_L + cm_a)\left[\frac{1-\nu(x)}{1+\frac{\nu(x)}{k_L+k_H\vartheta(x,y)-1}}\right](1-\beta(x)) - cm_a,$$

$$\hat{U}_H(x,y) = (R_H + cm_a - w_H)\left[\frac{1-\nu(y)}{1+\frac{\nu(y)}{k_H+k_L\vartheta(y,x)-1}}\right](1-\beta(y)) - cm_a,$$

where $\nu(x) = \frac{x\beta'(x)}{1-\beta(x)}$, *and* $\vartheta(x,y) = \frac{y\nu(x)}{x\nu(y)}$. *Then, we have that*

1. $\frac{\partial \hat{U}_L(x,y)}{\partial x} < 0$. *Furthermore,* $\frac{\partial \hat{U}_L(x,y)}{\partial y} > 0$ *when* $\frac{\partial \vartheta(x,y)}{\partial x} < 0$.
2. $\frac{\partial \hat{U}_H(x,y)}{\partial y} < 0$. *Furthermore,* $\frac{\partial \hat{U}_H(x,y)}{\partial x} > 0$ *when* $\frac{\partial \vartheta(x,y)}{\partial x} < 0$.

Proof:

1. We have shown that both $\nu(x)$ and $\beta(x)$ are strictly increasing in x in Lemma 1. Using these observations, it is sufficient to show that $\frac{\nu(x)}{k_L+k_H\vartheta(x,y)-1}$ is increasing in x in order to prove that $\frac{\partial \hat{U}_L(x,y)}{\partial x} < 0$.

Let $h(x,y) = \frac{\nu(x)}{k_L+k_H\vartheta(x,y)-1}$. Then, observe that

$$\begin{aligned}
\frac{\partial h(x,y)}{\partial x} &= \frac{\nu'(x)[k_L + k_H\vartheta(x,y) - 1] - k_H\nu(x)\left[\frac{\partial\vartheta(x,y)}{\partial x}\right]}{[k_L + k_H\vartheta(x,y) - 1]^2} \\
&= \frac{\nu'(x)[k_L + k_H\vartheta(x,y) - 1] - k_H\frac{y\nu(x)}{\nu(y)}\left[\frac{\nu'(x)}{x} - \frac{\nu(x)}{x^2}\right]}{[k_L + k_H\vartheta(x,y) - 1]^2} \\
&= \frac{\nu'(x)[k_L + k_H\vartheta(x,y) - 1] - k_H x\vartheta(x,y)\left[\frac{\nu'(x)}{x} - \frac{\nu(x)}{x^2}\right]}{[k_L + k_H\vartheta(x,y) - 1]^2} \\
&= \frac{\nu'(x)[k_L - 1] + k_H\vartheta(x,y)\frac{\nu(x)}{x}}{[k_L + k_H\vartheta(x,y) - 1]^2} \geq 0.
\end{aligned}$$

Furthermore, after simple algebra we have that

$$\begin{aligned}
\frac{\partial \hat{U}_L(x,y)}{\partial y} &= \left(\frac{\hat{U}_L(x,y) + cm_a}{1+\frac{\nu(y)}{k_H+k_L\vartheta(y,x)-1}}\right)\left(\frac{k_H\nu(x)\frac{\partial\vartheta(x,y)}{\partial y}}{[k_L + k + H\vartheta(x,y)]^2}\right) \\
&= -\left(\frac{\hat{U}_L(x,y) + cm_a}{1+\frac{\nu(y)}{k_H+k_L\vartheta(y,x)-1}}\right)\left(\frac{k_H\nu(x)\frac{\partial\vartheta(x,y)}{\partial x}}{[k_L + k + H\vartheta(x,y)]^2}\right) \geq 0,
\end{aligned}$$

whenever $\frac{\partial\vartheta(x,y)}{\partial x} \leq 0$.

2. The proof is very similar to the proof of Part 1.

■

The demand for High-Quality Agents given the demand for Low-Quality Agents:

LEMMA 15. *Let* $y(x)=\{y:\hat{U}_L(x,y)=\hat{U}_H(x,y)\}$. *Suppose* $\frac{\partial\vartheta(x,y)}{\partial x}<0$. *Then, we have that*

1. $y(x)$ *is a singleton and* $y(x)\geq x$.
2. $y(x)$ *is strictly increasing in* x.
3. $U_L\big(x,y(x)\big)$ *is strictly decreasing in* x.
4. *Let* λ_H^{dom} *be the unique solution to*

$$1-\beta(x)-x\beta'(x)=\frac{R_L+cm_a}{R_H+cm_a-w_H}.$$

Then, there exists a $\Lambda^{R_L}\leq k_H\lambda_H^{dom}$ *such that* $y(0)=\Lambda^{R_L}/k_H$.

Proof:

1. Note that for any given $0\leq x<\infty$, we have that $\hat{U}_H(x,0)=R_H-w_H$ since $\nu(0)=0$, and $\hat{U}_L(x,0)\leq R_L$. Using these observations, we have that

$$\hat{U}_H(x,0)-\hat{U}_L(x,0)\geq R_H-w_H-R_L\geq 0.$$

Moreover, for any given $0\leq x<\infty$, we have that $\lim_{y\to\infty}\hat{U}_H(x,y)=-cm_a$, and $\lim_{y\to\infty}\hat{U}_L(x,y)>-cm_a$. Thus, we have that

$$\lim_{y\to\infty}\big[\hat{U}_H(x,y)-\hat{U}_L(x,y)\big]<0.$$

Then, the claim holds since we have that

$$\frac{\partial\hat{U}_H(x,y)}{\partial y}-\frac{\partial\hat{U}_L(x,y)}{\partial y}<0,$$

by Lemma 14.

Finally, we have that $y(x)\geq x$ since $\frac{\partial\hat{U}_L(x,y)}{\partial y}>0$, $\frac{\partial\hat{U}_H(x,y)}{\partial y}<0$, and

$$\begin{aligned}\hat{U}_L(x,x)&=(R_L+cm_a)\frac{1-\nu(x)}{1+\frac{\nu(x)}{k_L+k_H-1}}(1-\beta(x))-cm_a\\&\leq(R_H+cm_a-w_H)\frac{1-\nu(x)}{1+\frac{\nu(x)}{k_L+k_H-1}}(1-\beta(x))-cm_a=\hat{U}_H(x,x).\end{aligned}$$

2. For any given x, and $\varepsilon>0$, note that

$$\hat{U}_L(\Lambda(x+\varepsilon),y(x))\;<\;\hat{U}_L(x,y(x))\;=\;\hat{U}_H(x,y(x))\;<\;\hat{U}_H(\Lambda(x+\varepsilon),y(x)),$$

since $\frac{\partial\hat{U}_L(x,y)}{\partial x}<0$, and $\frac{\partial\hat{U}_H(x,y)}{\partial x}>0$ by Lemma 14.

Now, suppose $y(x)\geq y(x+\varepsilon)$ for some $x\geq 0$, and $\varepsilon>0$. Then, using the above observation, we would have that

$$\hat{U}_L(\Lambda(x+\varepsilon),y(x+\varepsilon))\leq\hat{U}_L(\Lambda(x+\varepsilon),y(x))\;<\;\hat{U}_H(\Lambda(x+\varepsilon),y(x))\leq\hat{U}_H(\Lambda(x+\varepsilon),y(x+\varepsilon)),$$

since $\frac{\partial\hat{U}_L(x,y)}{\partial y}>0$, and $\frac{\partial\hat{U}_H(x,y)}{\partial y}<0$ again by Lemma 14. However, this contradicts with the definition of $y(x)$. Hence, we should have that $y(x)<y(x+\varepsilon)$.

3. We prove this claim by contradiction. Therefore, we suppose there exists some x_1, and x_2 such that $x_1<x_2$, and

$$\hat{U}_L\big(x_1,y(x_1)\big)\leq\hat{U}_L\big(x_2,y(x_2)\big).\tag{27}$$

Then, we would have that

$$\begin{aligned}\hat{U}_L\big(x_2,y(x_2)\big) \;&\geq\; \hat{U}_L\big(x_1,y(x_1)\big) = (R_L+cm_a)\left[\frac{1-\nu(x_1)}{1+\frac{\nu(x_1)}{k_L+k_H\vartheta(x_1,y(x_1))-1}}\right][1-\beta(x_1)]-cm_a \\ &> (R_L+cm_a)\left[\frac{1-\nu(x_2)}{1+\frac{\nu(x_2)}{k_L+k_H\vartheta(x_1,y(x_1))-1}}\right][1-\beta(x_2)]-cm_a,\end{aligned}$$

where the inequality holds since both $\nu(x)$ and $\beta(x)$ are strictly increasing in x. Then, the above inequality implies that

$$\begin{aligned}&(R_L+cm_a)\left[\frac{1-\nu(x_2)}{1+\frac{\nu(x_2)}{k_L+k_H\vartheta(x_2,y(x_2))-1}}\right][1-\beta(x_2)]-cm_a \\ &\quad= \hat{U}_L\big(x_2,y(x_2)\big) \;>\; (R_L+cm_a)\left[\frac{1-\nu(x_2)}{1+\frac{\nu(x_2)}{k_L+k_H\vartheta(x_1,y(x_1))-1}}\right][1-\beta(x_2)]-cm_a \\ &\quad\Rightarrow \vartheta(x_2,y(x_2))>\vartheta(x_1,y(x_1)). \end{aligned} \tag{28}$$

Note that by Equation 27, we also have that

$$\hat{U}_H\big(x_1,y(x_1)\big)\leq \hat{U}_H\big(x_2,y(x_2)\big).$$

Furthermore, as we show above, we have that

$$\begin{aligned}\hat{U}_H\big(x_2,y(x_2)\big) \;&\geq\; \hat{U}_H\big(x_1,y(x_1)\big) = (R_H+cm_a-w_H)\left[\frac{1-\nu(y(x_1))}{1+\frac{\nu(y(x_1))}{k_L+k_H\vartheta(y(x_1),x_1)-1}}\right][1-\beta(y(x_1))]-cm_a \\ &> \;(R_H+cm_a-w_H)\left[\frac{1-\nu(y(x_2))}{1+\frac{\nu(y(x_2))}{k_L+k_H\vartheta(y(x_1),x_1)-1}}\right][1-\beta(y(x_2))]-cm_a \\ &\Rightarrow\; \vartheta(y(x_2),x_2)>\vartheta(y(x_1),x_1).\end{aligned} \tag{29}$$

where the inequality holds since both $\nu(x)$ and $\beta(x)$ are strictly increasing in x, and $y(x)$ is strictly increasing in x.

Finally, by combining Equations 28 and 29, we have that

$$1=\vartheta(x_2,y(x_2))\vartheta(y(x_2),x_2)>\vartheta(x_1,y(x_1))\vartheta(y(x_1),x_1)=1,$$

which is a contradiction. Hence, $\hat{U}_L(x,y(x))$ is strictly decreasing in x.

4. We first want to note that $\hat{U}_H(0,0)=R_H$,

$$\hat{U}_H(0,\lambda_H^{dom})=(R_H+cm_a-w_H)\left[\frac{\frac{R_L+cm_a}{R_H+cm_a-w_H}}{1+\frac{\nu(\lambda_H^{dom})}{k_H+k_L\vartheta(\lambda_H^{dom},0)-1}}\right]-cm_a\leq R_L,$$

and $\frac{\partial \hat{U}_H(x,y)}{\partial y}<0$. Therefore, there exists a unique $\Lambda^{R_L}\leq k_H\lambda_H^{dom}$ such that

$$\hat{U}_H(0,\Lambda^{R_L}/k_H)=R_L.$$

Furthermore, we have that $\hat{U}_L(0,y)=R_L$ for any $y\geq 0$. Hence, it is clear that $y(0)=\Lambda^{R_L}/k_H$.

■

Candidate for the Customer Equilibrium with Positive Surplus:

PROPOSITION 11. *For any given Λ, let*

$$\big(\hat{D}_L(\Lambda),\hat{D}_H(\Lambda)\big)=\Big\{(x,y):\hat{U}_L(x,y)=\hat{U}_H(x,y),\ k_Lx+k_Hy=\Lambda\Big\}.$$

If $\frac{\partial\vartheta(x,y)}{\partial x}<0$, then the following statements are true:

1. *If $\Lambda\ge\Lambda^{R_L}$, then*
   (a) *$\hat{D}_L(\Lambda)$ is a singleton, and $\hat{D}_L(\Lambda^{R_L})=0$.*
   (b) *$\hat{D}_L(\Lambda)$ is strictly increasing in Λ.*
   (c) *$\hat{U}_L\big(\hat{D}_L(\Lambda),\hat{D}_H(\Lambda)\big)$ is strictly decreasing in Λ.*
2. *Let $\Lambda^{mon}=k_H\lambda_H^{mon}+k_L\lambda_L^{mon}$, where λ_i^{mon}, $i=\{L,R\}$, is the unique solution to*

$$1-\beta(x)-x\beta'(x)=\frac{cm_a}{R_i+cm_a-w_i}.$$

Then, we have that $\hat{U}_L\big(\hat{D}_L(\Lambda^{mon}),\hat{D}_H(\Lambda^{mon})\big)<0$.

3. *For any given $0\le u\le R_L$, there exists a $\Lambda(u)$ such that $\hat{U}_L\big(\hat{D}_L(\Lambda(u)),\hat{D}_H(\Lambda(u))\big)=u$. Furthermore, $\Lambda(u)$ is strictly decreasing in u, and $\Lambda(0)<k_H\lambda_H^{mon}+k_L\lambda_L^{mon}$.*

Proof:

1. a) Note that $y(0)=\Lambda^{R_L}/k_H$, and $k_Lx+k_Hy(x)$ is strictly increasing in x by Lemma 15. Hence, it is clear that there exists a unique $\hat{D}_L(\Lambda)$ such that

$$k_L\hat{D}_L(\Lambda)+k_Hy(\hat{D}_L(\Lambda))=\Lambda,$$

for any given $\Lambda\ge\Lambda^{R_L}$.

Also note that $\hat{D}_L(\Lambda^{R_L})=0$.

b) By definition, for any $\Lambda_1<\Lambda_2$, we have that

$$k_L\hat{D}_L(\Lambda_1)+k_Hy(\hat{D}_L(\Lambda_1))<k_L\hat{D}_L(\Lambda_2)+k_Hy(\hat{D}_L(\Lambda_2)).$$

Then, the claim follows since $k_Lx+k_Hy(x)$ is strictly increasing in x by Lemma 15.

c) Our claim follows since $\hat{D}_L(\Lambda)$ is strictly increasing in Λ as shown in part 2, and $\hat{U}_L\big(x,y(x)\big)$ is strictly decreasing in x by Lemma 15.3.

2. We prove this result by contradiction. Therefore, we suppose

$$\hat{U}_L\big(\hat{D}_L(\Lambda^{mon}),\hat{D}_H(\Lambda^{mon})\big)\ge 0,$$

Observe that

$$\begin{aligned}\hat{U}_L\big(\lambda_L^{mon},y(\lambda_L^{mon})\big)&=(R_L+cm_a)\left[\frac{\frac{cm_a}{R_L+cm_a}}{1+\frac{\nu(\lambda_L^{mon})}{k_L+k_H\vartheta(\lambda_L^{mon},y(\lambda_L^{mon}))-1}}\right]-cm_a\\&=\frac{cm_a}{1+\frac{\nu(\lambda_L^{mon})}{k_L+k_H\vartheta(\lambda_L^{mon},y(\lambda_L^{mon}))-1}}-cm_a<0.\end{aligned}$$

and this inequality implies that $\hat{D}_L(\Lambda^{mon})<\lambda_L^{mon}$ since $\hat{U}_L\big(x,y(x)\big)$ is decreasing in x by Lemma 15.3. We also have that

$$\hat{D}_H(\Lambda^{mon})>\lambda_H^{mon}$$

since $k_L \hat{D}_L(\Lambda^{mon}) + k_H \hat{D}_H(\Lambda^{mon}) = \Lambda^{mon}$.

Furthermore, observe that

$$\hat{U}_H(x, \lambda_H^{mon}) \;=\; (R_H + cm_a - w_H) \left[\frac{\frac{cm_a}{R_H + cm_a}}{1 + \frac{\nu(\lambda_H^{mon})}{k_L + k_H \vartheta(\lambda_H^{mon}, x) - 1}} \right] - cm_a < 0,$$

for any $0 \le x \le \lambda_L^{mon}$. Then, we have that

$$\hat{U}_H\big(\hat{D}_L(\Lambda^{mon}), \hat{D}_H(\Lambda^{mon})\big) \;<\; \hat{U}_H\big(\hat{D}_L(\Lambda^{mon}), \lambda_H^{mon}\big) < 0 \le \hat{U}_L\big(\hat{D}_L(\Lambda^{mon}), \hat{D}_H(\Lambda^{mon})\big),$$

since $\hat{D}_H(\Lambda^{mon}) > \lambda_H^{mon}$ and $\frac{\partial \hat{U}_H(x,y)}{\partial y} < 0$. However, this contradicts with the definition of $\big(\hat{D}_L(\Lambda^{mon}), \hat{D}_H(\Lambda^{mon})\big)$.

3. Note that we have

$$\begin{aligned} \hat{U}_L\big(\hat{D}_L(\Lambda^{R_L}), \hat{D}_H(\Lambda^{R_L})\big) \;&=\; R_L, \\ \hat{U}_L\big(\hat{D}_L(\Lambda^{mon}), \hat{D}_H(\Lambda^{mon})\big) \;&<\; 0, \end{aligned}$$

where the first equality holds since $\hat{D}_L(\Lambda^{R_L}) = 0$ and $\hat{D}_H(\Lambda^{R_L}) = \Lambda^{R_L}/k_H$, and we prove the second one in part 2. Therefore, it is clear that $\Lambda(R_L) = \Lambda^{R_L}$, where Λ^{R_L} is defined as in Lemma 15.

Finally our claim holds since $\hat{U}_L\big(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)\big)$ is strictly decreasing in Λ as we show in part 1.c.

■

Corollary 2. *For any given $\Lambda < \Lambda(0)$, if $\frac{\partial \vartheta(x,y)}{\partial x} < 0$, then we have that*

$$\{(x,y) : \hat{U}_L(x,y) \le 0,\ \hat{U}_H(x,y) \le 0,\ k_L x + k_H y = \Lambda,\ x \ge 0,\ y \ge 0\} = \emptyset.$$

Proof:

Note that since $\Lambda < \Lambda(0)$, we have that

$$\hat{U}_L\big(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)\big) = \hat{U}_H\big(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)\big) > 0,$$

as we shown in Proposition 11. Furthermore, since $k_L x + k_H y = \Lambda$, we have two possible cases:

1. $\mathbf{x} \ge \hat{\mathbf{D}}_{\mathbf{L}}(\boldsymbol{\Lambda})$, and $\mathbf{y} \le \hat{\mathbf{D}}_{\mathbf{H}}(\boldsymbol{\Lambda})$: In this case, we have that

$$\hat{U}_H(x,y) \ge \hat{U}_H\big(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)\big) > 0$$

since $\frac{\partial U_H(x,y)}{\partial y} < 0$, and $\frac{\partial U_H(x,y)}{\partial x} > 0$. Therefore, this case cannot be in the set we defined above.

2. $\mathbf{x} \le \hat{\mathbf{D}}_{\mathbf{L}}(\boldsymbol{\Lambda})$, and $\mathbf{y} \ge \hat{\mathbf{D}}_{\mathbf{H}}(\boldsymbol{\Lambda})$: In this case, we have that

$$\hat{U}_L(x,y) \ge \hat{U}_L\big(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)\big) > 0$$

since $\frac{\partial \hat{U}_L(x,y)}{\partial x} < 0$, and $\frac{\partial \hat{U}_L(x,y)}{\partial y} > 0$. Therefore, this case also cannot be in the set we defined above. Hence, the above set is empty.

■

S.4.2. Single-Agent Best Response

Given that $k_H - 1$ high-quality agents charge p_H, and k_L low-quality agents charge p_L, a single high-quality agents, say agent-ℓ, solves the following problem to find his best response:

$$\max_{p_\ell \ge w_H,\ D_\ell \ge 0,\ D_H \ge 0,\ D_L \ge 0} \quad (p_\ell - w_H)\Lambda D_\ell \left[1 - \beta(\Lambda D_\ell)\right]$$

$$
\begin{aligned}
&s.to \\
&\quad (R_H - p_\ell + cm_a)\,[1-\beta(\Lambda D_\ell)] - cm_a \geq 0 \\
&\quad (R_H - p_\ell + cm_a)\,[1-\beta(\Lambda D_\ell)] = (R_H - p_H + cm_a)\,[1-\beta(\Lambda D_H)] \\
&\quad (R_H - p_H + cm_a)\,[1-\beta(\Lambda D_H)] \geq (R_L - p_L + cm_a)\,[1-\beta(\Lambda D_L)] \\
&\quad D_\ell + (k_H - 1)D_H + k_L D_L \leq 1 \\
&\quad D_L \geq 0
\end{aligned}
$$

Any the symmetric SPNE $(D_L, D_H; p_L, p_H)$ should satisfy the following FOC:

$$\Lambda D_H - \eta_{1_H} - \eta_{2_H} = 0, \tag{30}$$

$$\Lambda(p_H - w_H)[1-\beta(\Lambda D_H)] - \Lambda^2 D_H (R + cm_a - w_H)\beta'(\Lambda D_H) - \eta_{4_H} = 0, \tag{31}$$

$$(\eta_{2_H} - \eta_{3_H})\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H) - (k_H - 1)\eta_{4_H} = 0, \tag{32}$$

$$\eta_{3_H}\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L) - k_L\eta_{4_H} + \eta_{5_H} = 0, \tag{33}$$

$$\eta_{1_H}\big((R_H - p_H + cm_a)[1-\beta(\Lambda D_H)] - cm_a\big) = 0, \tag{34}$$

$$\eta_{3_H}\big((R_H - p_H + cm_a)[1-\beta(\Lambda D_H)] - (R_L - p_L + cm_a)[1-\beta(\Lambda D_L)]\big) = 0, \tag{35}$$

$$\eta_{4_H}(1 - k_L D_L - k_H D_H) = 0, \tag{36}$$

$$\eta_{5_H} D_L = 0, \tag{37}$$

$$\eta_{1_H},\ \eta_{3_H},\ \eta_{4_H},\ \eta_{5_H} \geq 0, \tag{38}$$

where η_{1_H}, η_{2_H}, η_{3_H}, η_{4_H}, and η_{5_H} are the Lagrangian multipliers of the constraints 1, 2, 3, 4, and 5 of the best response problem of agent-ℓ, respectively. Furthermore, given any symmetric SPNE $(D_L, D_H; p_L, p_H)$, we denote the expected utility of a customer choosing the price p_i for $i \in \{H, L\}$ by $U_i^{SPNE}(D_L, D_H; p_L, p_H)$.

Claim 3. *Given any symmetric SPNE $(D_L, D_H; p_L, p_H)$, we have that*

1. $D_H > 0$.
2. $U_H^{SPNE}(D_L, D_H; p_L, p_H) < R_L \Leftrightarrow D_L > 0$.

Proof:

1. Suppose NOT, i.e. $D_H = 0$. Note that $D_H = 0$ implies that $V(\Lambda D_H, p_H) = 0$, and $U_L(\Lambda D_L, p_L) < R_L$ since all customers choose low-quality providers whose service can only give a reward of R_L. Consider the case where a single high-quality agent deviates and charge a price $p < R_H - U_L(\Lambda D_L, p_L) - w_H$. It is clear that some of the customers should choose this agent after deviation since he would be the cheapest agent if all customers would still choose only low-quality providers. Furthermore, it is apparent that the deviating high-quality agent will earn a strictly positive profit after deviation. However, this contradicts with the definition of *SPNE*. Hence, We should have $D_H > 0$.

2. Suppose NOT. First note that $D_L = 0$ implies that $V(\Lambda D_L, p_L) = 0$. Consider the case where a single low-quality agent deviates and charge a price $p < R_L - U_H^{SPNE}(D_L, D_H; p_L, p_H)$. It is clear that some of the customers should choose this agent after deviation since he would be the cheapest agent if all customers would still choose only high-quality providers. Furthermore, it is apparent that the deviating low-quality agent will earn a strictly positive profit after deviation. However, this contradicts with the definition of *SPNE*. Hence, We should have $D_L > 0$.

■

In the remaining of the proof, we perform a case-by-case analysis to show that

- $k_H D_H + k_L D_L < 1 \Leftrightarrow \Lambda > k_H \lambda_H^{mon} + k_L \lambda_L^{mon}$.
- $\begin{array}{c} U_H^{SPNE}(D_L, D_H; p_L, p_H) = 0, \\ and \\ k_H D_H + k_L D_L = 1 \end{array} \Leftrightarrow \Lambda(0) \le \Lambda \le k_H \lambda_H^{mon} + k_L \lambda_L^{mon}$.
- $0 < U_H^{SPNE}(D_L, D_H; p_L, p_H) < R_L \ \Leftrightarrow \Lambda(R_L) < \Lambda < \Lambda(0)$.
- $U_H^{SPNE}(D_L, D_H; p_L, p_H) = R_L \Leftrightarrow k_H \lambda_H^{R_L} \le \Lambda \le \Lambda(R_L)$.
- $U_H^{SPNE}(D_L, D_H; p_L, p_H) > R_L \Leftrightarrow \Lambda < k_H \lambda_H^{R_L}$.

Note that to prove the above $\Leftrightarrow$ statements, it is sufficient to show the $\Rightarrow$ statements because $\Rightarrow$ statements cover all possible values for Λ.

Case-1 ($k_H D_H + k_L D_L < 1$): Note that in this case, we have that $U_H^{SPNE}(D_L, D_H; p_L, p_H) = 0$, so that $D_L > 0$. Then, we have that

$$\begin{aligned} k_H D_H + k_L D_L < 1 &\Rightarrow \eta_{4_H} = 0 \\ &\Rightarrow \eta_{3_H} = 0 \qquad (\text{Since } \eta_{5_H} = 0) \\ &\Rightarrow \eta_{2_H} = 0 \qquad (\text{Since } \eta_{4_H} = 0) \\ &\Rightarrow \eta_{1_H} = \Lambda D_H > 0 \\ &\Rightarrow p_H = R_H + cm_a - \frac{cm_a}{1 - \beta(\Lambda D_H)} \\ &\Rightarrow (R_H + cm_a - w_H)\big[1 - \beta(\Lambda D_H) - \Lambda D_H \beta'(\Lambda D_H)\big] = cm_a \qquad (\text{Since } \eta_{4_H} = 0) \\ &\Rightarrow D_H = \frac{\lambda_H^{mon}}{\Lambda}. \end{aligned}$$

Similarly, we can show that $k_H D_H + k_L D_L < 1 \Rightarrow D_L = \frac{\lambda_L^{mon}}{\Lambda}$. Combining these observations, we have that

$$k_H D_H + k_L D_L < 1 \Rightarrow \Lambda > k_H \lambda_H^{mon} + k_L \lambda_L^{mon}.$$

Case-2 ($U_H^{SPNE}(D_L, D_H; p_L, p_H) = 0$, and $k_H D_H + k_L D_L = 1$): As in the previous case, we have that $D_L > 0$. We first want to note that

$$\begin{aligned} U_H^{SPNE}(D_L, D_H; p_L, p_H) = 0 &\Rightarrow p_H = R_H + cm_a - \frac{cm_a}{1 - \beta(\Lambda D_H)} \\ &\Rightarrow \frac{\eta_{4_H}}{\Lambda} = (R_H + cm_a - w_H)\big[1 - \beta(\Lambda D_H) - \Lambda D_H \beta'(\Lambda D_H)\big] - cm_a \\ &\Rightarrow D_H \le \frac{\lambda_H^{mon}}{\Lambda}, \end{aligned}$$

where the last statement holds since $\eta_{4_H} \geq 0$ and by the definition of λ_H^{mon}.

Similarly, we can show $D_L \leq \frac{\lambda_L^{mon}}{\Lambda}$ in this case. Then, we have that

$$\begin{array}{c} U_H^{SPNE}(D_L, D_H; p_L, p_H) = 0, \\ and \\ k_H D_H + k_L D_L = 1 \end{array} \quad \Rightarrow \Lambda \leq k_H \lambda_H^{mon} + k_L \lambda_L^{mon}.$$

Furthermore, since $p_H = R_H + cm_a - \frac{cm_a}{1-\beta(\Lambda D_H)}$, and $p_L = R_L + cm_a - \frac{cm_a}{1-\beta(\Lambda D_L)}$, we have that

$$\begin{aligned} \eta_{2_H} &= \eta_{3_H} + \frac{(k_H-1)[1-\beta(\Lambda D_H)]}{cm_a \Lambda \beta'(\Lambda D_H)} \eta_{4_H} \\ &= \left[\frac{k_L[1-\beta(\Lambda D_L)]}{cm_a \beta'(\Lambda D_L)} + \frac{(k_H-1)[1-\beta(\Lambda D_H)]}{cm_a \Lambda \beta'(\Lambda D_H)}\right] \eta_{4_H} \\ &= \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L)}{cm_a \frac{\beta'(\Lambda D_H)}{1-\beta(\Lambda D_H)}}\right] \big[(R_H + cm_a - w_H)[1-\nu(\Lambda D_H)][1-\beta(\Lambda D_H)] - cm_a\big], \end{aligned}$$

where the last equality holds by Equation 31.

Then, using Equation 30, we have that

$$\begin{aligned} \eta_{1_H} &= \Lambda D_H - \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L)}{cm_a \frac{\beta'(\Lambda D_H)}{1-\beta(\Lambda D_H)}}\right] \big[(R_H + cm_a - w_H)[1-\nu(\Lambda D_H)][1-\beta(\Lambda D_H)] - cm_a\big] \\ &= \left[\begin{array}{c} -(R_H + cm_a - w_H)[1-\nu(\Lambda D_H)][1-\beta(\Lambda D_H)]\big[k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L)\big] \\ +cm_a\big[k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L) + \nu(\Lambda D_H)\big] \end{array}\right] \frac{1}{cm_a \frac{\beta'(\Lambda D_H)}{1-\beta(\Lambda D_H)}} \\ &= -\hat{U}_H(\Lambda D_L, \Lambda D_H) \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L) + \nu(\Lambda D_H)}{cm_a \frac{\beta'(\Lambda D_H)}{1-\beta(\Lambda D_H)}}\right]. \end{aligned}$$

Similarly, we have that

$$\eta_{1_L} = -\hat{U}_L(\Lambda D_L, \Lambda D_H) \left[\frac{k_L - 1 + k_H \vartheta(\Lambda D_L, \Lambda D_H) + \nu(\Lambda D_L)}{cm_a \frac{\beta'(\Lambda D_L)}{1-\beta(\Lambda D_L)}}\right].$$

As a part of FOC, we should have that $\eta_{1_H} \geq 0$, and $\eta_{1_L} \geq 0$. Then, using above observations, we have that

$$\begin{aligned} \hat{U}_H(\Lambda D_L, \Lambda D_H) &\leq 0 \\ \hat{U}_L(\Lambda D_L, \Lambda D_H) &\leq 0. \end{aligned}$$

Finally, using these inequalities and Corollary 2, we have that

$$\begin{array}{c} U_H^{SPNE}(D_L, D_H; p_L, p_H) = 0, \\ and \\ k_H D_H + k_L D_L = 1 \end{array} \quad \Rightarrow \Lambda \geq \Lambda(0).$$

Case-3 ($0 < U_H^{SPNE}(D_L, D_H; p_L, p_H) < R_L$): First, observe that

$$\begin{aligned}\eta_{2_H} &= \eta_{3_H} + \frac{(k_H-1)}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H)}\eta_{4_H} \\ &= \left[\frac{k_L}{\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L)} + \frac{(k_H-1)}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H)}\right]\eta_{4_H} \\ &= \left[\frac{k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L)}{(R_H - p_H + cm_a)\beta'(\Lambda D_H)}\right]\Big[(p_H - w_H)[1-\beta(\Lambda D_H)] - \Lambda D_H(R + cm_a - w_H)\beta'(\Lambda D_H)\Big],\end{aligned}$$

where the last equality holds since $\frac{R_H - p_H + cm_a}{R_L - p_L + cm_a} = \frac{1-\beta(\Lambda D_L)}{1-\beta(\Lambda D_H)}$, and by Equation 31.

Furthermore, we have that $\eta_{1_H} = 0$, and this implies that $\eta_{2_H} = \Lambda D_H$. Therefore, we have that

$$\begin{aligned}p_H &= (R_H + cm_a - w_H)\frac{\big[k_H + k_L\vartheta(\Lambda D_H, \Lambda D_L)\big]\nu(\Lambda D_H)}{k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L) + \nu(\Lambda D_H)} + w_H \\ &= R_H + cm_a - \frac{(R_H + cm_a - w_H)\big[k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L)\big]}{\frac{k_H + k_L\vartheta(\Lambda D_H, \Lambda D_L)}{1-\nu(\Lambda D_H)} - 1}\end{aligned}$$

Then, using this equation, we have that

$$\begin{aligned}U_H^{SPNE}(D_L, D_H; p_L, p_H) &= \frac{(R_H + cm_a - w_H)\big[k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L)\big]}{k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L) - \nu(\Lambda D_H)}\big[1-\nu(\Lambda D_H\big]\big[1-\beta(\Lambda D_H)\big] - cm_a \\ &= \frac{(R_H + cm_a - w_H)\big[1-\nu(\Lambda D_H\big]\big[1-\beta(\Lambda D_H)\big]}{1 + \frac{\nu(\Lambda D_H)}{k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L)}} - cm_a \\ &= \hat{U}_H(\Lambda D_L, \Lambda D_H),\end{aligned}$$

where $\hat{U}_H(\Lambda D_L, \Lambda D_H)$ is defined in Lemma 14.

Similarly, we can also show that

$$\begin{aligned}&p_L = R_L + cm_a - \frac{(R_L + cm_a)\big[k_L - 1 + k_H\vartheta(\Lambda D_L, \Lambda D_H)\big]}{\frac{k_L + k_H\vartheta(\Lambda D_L, \Lambda D_H)}{1-\nu(\Lambda D_L)} - 1} \\ &U_L^{SPNE}(D_L, D_H; p_L, p_H)) = \hat{U}_L(\Lambda D_L, \Lambda D_H).\end{aligned}$$

Note that we have $U_H^{SPNE}(D_L, D_H; p_L, p_H) = U_L(\Lambda D_L, p_L)$, and $k_L D_L + k_H D_H = 1$ by the definition of Customer Equilibrium. Using the above observations, we have that

$$(\Lambda D_L, \Lambda D_H) \in \{(x,y) : \hat{U}_L(x,y) = \hat{U}_H(x,y),\ k_L x + k_H y = \Lambda\}.$$

Then, by Proposition 11, $(\Lambda D_L, \Lambda D_H)$ is unique, and $\Lambda D_L = \hat{D}_L(\Lambda)$, and $\Lambda D_H = \hat{D}_H(\Lambda)$. Furthermore,

$$0 < U_H^{SPNE}(D_L, D_H; p_L, p_H) = \hat{U}_H(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)) < R_L \ \Rightarrow \Lambda(R_L) < \Lambda < \Lambda(0).$$

Case-4 ($U_H^{SPNE}(D_L,D_H;p_L,p_H)=R_L$)**:** First, we want to note that $\eta_{1_H}=0$ since $U_H^{SPNE}(D_L,D_H;p_L,p_H)>0$, and this implies that $\eta_{2_H}=\Lambda D_H$. Furthermore, in this case, we have $p_H=R_H+cm_a-\frac{R_L+cm_a}{1-\beta(\Lambda D_H)}$, and $D_H=1/k_H$ since $U_H^{SPNE}(D_L,D_H;p_L,p_H)=R_L$. Then, by Equations 31 and 32, we have that

$$\begin{aligned}
\eta_{3_H} &= \eta_{2_H}-\frac{(k_H-1)}{\Lambda(R_H-p_H+cm_a)\beta'(\Lambda/k_H)}\eta_{4_H}\\
&=\Lambda/k_H-\left[\frac{(k_H-1)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]\left[(R_H+cm_a-w_H)\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]-(R_L+cm_a)\right]\\
&=\frac{\left[(R_L+cm_a)\left[k_H-1+\nu(\Lambda/k_H)\right]-(R_H+cm_a-w_H)(k_H-1)\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]\right]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\\
&=-\left[(R_H+cm_a-w_H)(k_H-1)-\frac{R_L+cm_a}{1-\beta(\Lambda/k_H)}\left(\frac{k_H-1+\nu(\Lambda/k_H)}{1-\nu(\Lambda/k_H)}\right)\right]\frac{\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\\
&=-\left[(R_H+cm_a-w_H)(k_H-1)-\frac{R_L+cm_a}{1-\beta(\Lambda/k_H)}\left(\frac{k_H}{1-\nu(\Lambda/k_H)}-1\right)\right]\frac{\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\\
&=-z^{R_L}(\Lambda/k_H)\left[\frac{\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right],
\end{aligned}$$

where $z^{R_L}(\lambda)=(R_H+cm_a-w_H)(k_H-1)-\frac{R_L+cm_a}{1-\beta(\lambda)}\left(\frac{k_H}{1-\nu(\lambda)}-1\right)$.

Note that we have $\eta_{3_H}\geq 0$. Thus,

$$U_H^{SPNE}(D_L,D_H;p_L,p_H)=R_L \Rightarrow z^{R_L}(\Lambda/k_H)\leq 0 \Rightarrow \Lambda\geq k_H\lambda_H^{R_L}, \tag{39}$$

since $z^{R_L}(\lambda)$ is strictly decreasing in λ (as $1-\beta(\lambda)$ and $1-\nu(\lambda)$ are strictly decreasing), and $z^{R_L}(\lambda_H^{R_L})=0$.

Moreover, observe that for any $\Lambda>k_H\lambda_H^{R_L}$, we have that $\eta_{3_H}>0$. This implies that for any $\Lambda>k_H\lambda_H^{R_L}$, 3^{rd} contraint is binding, so that we have $U(0,p_L)=R_L$, i.e. $p_L=0$. Then, again using the Equations 31-33, we have that

$$\begin{aligned}
\eta_{2_H} &= \eta_{3_H}+\frac{(k_H-1)}{\Lambda(R_H-p_H+cm_a)\beta'(\Lambda/k_H)}\eta_{4_H}\\
&=\left[\frac{k_L}{\Lambda(R_L-p_L+cm_a)\beta'(0)}+\frac{(k_H-1)}{\Lambda(R_H-p_H+cm_a)\beta'(\Lambda/k_H)}\right]\eta_{4_H}-\eta_{5_H}\\
&=\left[\frac{k_H-1+k_L\vartheta(0,\Lambda/k_H)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]\left[(R_H+cm_a-w_H)\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]-(R_L+cm_a)\right]-\eta_{5_H}.
\end{aligned}$$

Then, using the fact that $\eta_{2_H}=\Lambda/k_H$, we have that

$$\eta_{5_H}=\left[\frac{k_H-1+k_L\vartheta(0,\Lambda/k_H)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]\left[(R_H+cm_a-w_H)\left[1-\nu(\Lambda/k_H)\right]\left[1-\beta(\Lambda/k_H)\right]-(R_L+cm_a)\right]-\Lambda/k_H$$

$$= \begin{bmatrix} (R_H + cm_a - w_H)\big[k_H - 1 + k_L\vartheta(0,\Lambda/k_H)\big]\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big] \\ -(R_L + cm_a)\big[k_H - 1 + k_L\vartheta(0,\Lambda/k_H) + \nu(\Lambda/k_H)\big] \end{bmatrix} \frac{1}{(R_L + cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}$$

$$= \begin{bmatrix} (R_H + cm_a - w_H)\frac{\left[k_H - 1 + k_L\vartheta(0,\Lambda/k_H)\right]}{\left[k_H - 1 + k_L\vartheta(0,\Lambda/k_H) + \nu(\Lambda/k_H)\right]}\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big] \\ -(R_L + cm_a) \end{bmatrix}$$
$$\times \left[\frac{k_H - 1 + k_L\vartheta(0,\Lambda/k_H) + \nu(\Lambda/k_H)}{(R_L + cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]$$
$$= \left[\hat{U}_H(0,\Lambda/k_H) - R_L\right]\left[\frac{k_H - 1 + k_L\vartheta(0,\Lambda/k_H) + \nu(\Lambda/k_H)}{(R_L + cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right].$$

Note that we have $\eta_{5_H} \geq 0$. Thus,

$$U_H^{SPNE}(D_L, D_H; p_L, p_H) = R_L \Rightarrow \hat{U}_H(0,\Lambda/k_H) \geq R_L \;\Rightarrow \Lambda \leq \Lambda(R_L), \tag{40}$$

since $\frac{\partial \hat{U}_H(x,y)}{\partial y} < 0$, $\hat{U}_H(\hat{D}_L(\Lambda(R_L)), \hat{D}_H(\Lambda(R_L))) = R_L$, $\hat{D}_L(\Lambda(R_L)) = 0$, and $\hat{D}_H(\Lambda(R_L)) = \Lambda(R_L)/k_H$. Finally combining (39) and (40), we have that

$$U_H^{SPNE}(D_L, D_H; p_L, p_H) = R_L \Rightarrow \Lambda_H^{R_L} \leq \Lambda \leq \Lambda^{R_L}.$$

Case-5 ($U_H^{SPNE}(D_L, D_H; p_L, p_H) > R_L$): As in the previous case, we have that $\eta_{1_H} = 0$, $\eta_{2_H} = \Lambda D_H$, and $D_H = 1/k_H$. Moreover, in this case, we have $\eta_{3_H} = 0$ since $U_H^{SPNE}(D_L, D_H; p_L, p_H) > R_L \geq U_L^{SPNE}(D_L, D_H; p_L, p_H)$). Then, using Equations 31 and 32, we have that

$$\begin{aligned} \eta_{2_H} &= \frac{k_H - 1}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda/k_H)}\eta_{4_H} \\ &= \frac{(k_H - 1)\Big[(p_H - w_H)\big[1-\beta(\Lambda/k_H)\big] - \Lambda/k_H(R_H + cm_a - w_H)\beta'(\Lambda/k_H)\Big]}{(R_H - p_H + cm_a)\beta'(\Lambda/k_H)}. \end{aligned}$$

Using $\eta_{2_H} = \Lambda/k_H$, this equation implies that

$$\begin{aligned} p_H &= \frac{(R_H + cm_a - w_H)\Lambda\beta'(\Lambda/k_H)}{(k_H - 1)\big[1-\beta(\Lambda/k_H)\big] + \Lambda/k_H\beta'(\Lambda/k_H)} + w_H \\ &= \frac{(R_H + cm_a - w_H)k_H}{\frac{k_H - 1}{\nu(\Lambda/k_H)} + 1} + w_H \\ &= R_H + cm_a - (R_H + cm_a - w_H)\left(1 - \frac{k_H}{\frac{k_H-1}{\nu(\Lambda/k_H)} + 1}\right) \\ &= R_H + cm_a - (R_H + cm_a - w_H)\left(\frac{(k_H - 1)\big[1-\nu(\Lambda/k_H)\big]}{k_H - 1 + \nu(\Lambda/k_H)}\right) \\ &= R_H + cm_a - (R_H + cm_a - w_H)\left(\frac{k_H - 1}{\frac{k_H}{1-\nu(\Lambda/k_H)} - 1}\right). \end{aligned}$$

Furthermore, since $U_H^{SPNE}(D_L, D_H; p_L, p_H) > R_L$, we have that

$$U_H^{SPNE}(D_L, D_H; p_L, p_H)) = (R_H + cm_a - w_H)\left(\frac{(k_H-1)\left[1-\beta(\Lambda/k_H)\right]}{\frac{k_H}{1-\nu(\Lambda/k_H)}-1}\right) - cm_a > R_L$$
$$\Rightarrow \left[(R_H + cm_a - w_H)(k_H-1) - \frac{R_L + cm_a}{1-\beta(\Lambda/k_H)}\left(\frac{k_H}{1-\nu(\Lambda/k_H)}-1\right)\right]\frac{1-\beta(\Lambda/k_H)}{\frac{k_H}{1-\nu(\Lambda/k_H)}-1} > 0$$
$$\Rightarrow z^{R_L}(\Lambda/k_H) > 0 \;\Rightarrow \Lambda < k_H\lambda_H^{R_L}.$$

S.4.3. Sufficiency for the Equilibrium

LEMMA 16. *Let (p_H^*, p_L^*) be the equilibrium prices defined in Theorem 6. then, the best response of a quality-i agent is p_i^* when all the other quality-i agents charge p_i^* and all quality-j agents charge p_j^* for $i,j \in \{H, L\}$.*

Proof of the Lemma: Given that $k_H - 1$ high-quality agents charge p_H, and k_L low-quality agents charge p_L, a single high-quality agents, say agent-ℓ, solves the following problem to find his best response:

$$\begin{aligned}
\max_{p_\ell \geq 0,\ D_\ell \geq 0,\ D_H \geq 0} \quad & p_\ell \Lambda D_\ell\left[1-\beta(\Lambda D_\ell)\right] \\
s.to \quad & \\
& (R_H - p_\ell + cm_a)\left[1-\beta(\Lambda D_\ell)\right] - cm_a \geq 0 \\
& (R_H - p_\ell + cm_a)\left[1-\beta(\Lambda D_\ell)\right] = (R_H - p_H + cm_a)\left[1-\beta(\Lambda D_L)\right] \\
& (R_H - p_H + cm_a)\left[1-\beta(\Lambda D_H)\right] \geq (R_L - p_L + cm_a)\left[1-\beta(\Lambda D_L)\right] \\
& D_\ell + (k_H-1)D_H + k_L D_L \leq 1 \\
& D_L \geq 0
\end{aligned}$$

and the FOC of this problem are:

$$\Lambda D_H - \eta_{1_H} - \eta_{2_H} = 0, \tag{41}$$
$$\Lambda p_\ell[1-\beta(\Lambda D_\ell)] - \Lambda^2 D_H (R + cm_a)\beta'(\Lambda D_\ell) - \eta_{4_H} = 0, \tag{42}$$
$$(\eta_{2_H} - \eta_{3_H})\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H) - (k_H - 1)\eta_{4_H} = 0, \tag{43}$$
$$\eta_{3_H}\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L) - k_L\eta_{4_H} + \eta_{5_H} = 0, \tag{44}$$
$$\eta_{1_H}\big((R_H - p_\ell + cm_a)[1-\beta(\Lambda D_\ell)] - cm_a\big) = 0, \tag{45}$$
$$\eta_{3_H}\big((R_H - p_H + cm_a)[1-\beta(\Lambda D_H)] - (R_L - p_L + cm_a)[1-\beta(\Lambda D_L)]\big) = 0, \tag{46}$$
$$\eta_{4_H}(1 - k_L D_L - k_H D_H) = 0, \tag{47}$$
$$\eta_{5_H} D_L = 0, \tag{48}$$
$$\eta_{1_H},\ \eta_{3_H},\ \eta_{4_H},\ \eta_{5_H} \geq 0, \tag{49}$$

where η_{1_H}, η_{2_H}, η_{3_H}, η_{4_H}, and η_{5_H} are the Lagrangian multipliers of the constraints 1, 2, 3, 4, and 5 of the best response problem of agent-ℓ, respectively. Moreover, we denote the solution to the above problem by $(D_\ell(p), D_L(p), D_H(p), p_\ell(p))$ for a given $p = (p_H, p_L)$.

Case-2 ($\Lambda^0 \le \Lambda \le \Lambda^{mon}$): Similar to the sufficiency proof in the identical providers case, we first show that given (p_H^*, p_L^*) as in Theorem 6, single high-quality provider will not leave any surplus to the customers when $\Lambda^0 \le \Lambda \le \Lambda^{mon}$.

CLAIM 4. *Let $(D_\ell(p), D_L(p), D_H(p), p_\ell(p))$ be the solution of single agent's best response problem when all other agents charge the price p. If $p = (p_H^*, p_L^*)$ as described in Theorem 6, then we have that $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] = cm_a$.*

Proof:

Suppose NOT. Then, we have that $\eta_1 = 0$ and this implies that $\eta_3 > 0$. Moreover, we have that $\Lambda D_H(p) < \Lambda D_H^*$ since $(R_H - p_H + cm_a)[1 - \beta(\Lambda D_H^*)] = cm_a$ and we suppose that the deviating provider offers a strictly positive utility to customers. Similarly, we have that $\Lambda D_L(p) < \Lambda D_L^*$.

Then, using the fact that $D_\ell(p) + (k_H - 1)D_H(p) + k_L D_L(p) = 1$ when $\eta_3 > 0$, we have that $\Lambda D_\ell(p) > \Lambda D_H^*$ since $\Lambda D_L(p) < \Lambda D_L^*$ and $\Lambda D_H(p) < \Lambda D_H^*$.

Note that using (43) and (44), we have that

$$\begin{aligned}
\eta_{2_H} &= \eta_{3_H} + \frac{(k_H - 1)[1 - \beta(\Lambda D_H^*)]}{cm_a \Lambda \beta'(\Lambda D_H(p))} \eta_{4_H} \\
&= \left[\frac{k_L[1 - \beta(\Lambda D_L^*)]}{cm_a \beta'(\Lambda D_L(p))} + \frac{(k_H - 1)[1 - \beta(\Lambda D_H^*)]}{cm_a \Lambda \beta'(\Lambda D_H(p))} \right] \eta_{4_H} - \eta_{5_H} \frac{[1 - \beta(\Lambda D_L^*)]}{cm_a \beta'(\Lambda D_L(p))} \\
&\le \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H^*, \Lambda D_L^*)}{cm_a \frac{\beta'(\Lambda D_H^*)}{1 - \beta(\Lambda D_H^*)}} \right] \frac{\eta_{4_H}}{\Lambda},
\end{aligned}$$

where the inequality holds since $\Lambda D_L(p) < \Lambda D_L^*$, $\Lambda D_H(p) < \Lambda D_H^*$, and $\beta(\lambda)$ is concave. Moreover, using (43)and the fact that $p_\ell(p) \le R_H + cm_a - \frac{cm_a}{1 - \beta(\Lambda D_\ell(p))}$, we have that

$$\begin{aligned}
\frac{\eta_{4_H}}{\Lambda} &\le \left((R_H + cm_a)[1 - \nu(\Lambda D_H(p))][1 - \beta(\Lambda D_H(p))] - cm_a \right) \\
&< \left((R_H + cm_a)[1 - \nu(\Lambda D_H^*)][1 - \beta(\Lambda D_H^*)] - cm_a \right),
\end{aligned}$$

where the second inequality holds since $\Lambda D_\ell(p) > \Lambda D_H^*$ and $[1 - \nu(\lambda)][1 - \beta(\lambda)]$ is the derivative of $\lambda[1 - \beta(\lambda)]$, which is a strictly concave function.

Combining these two observations with the fact that $\eta_{2_H} = \Lambda D_\ell(p)$ (since $\eta_{1_H} = 0$), we have that

$$\begin{aligned}
&\Lambda D_H^* < \eta_{2_H} < \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H^*, \Lambda D_L^*)}{cm_a \frac{\beta'(\Lambda D_H^*)}{1 - \beta(\Lambda D_H^*)}} \right] \left((R_H + cm_a)[1 - \nu(\Lambda D_H^*)][1 - \beta(\Lambda D_H^*)] - cm_a \right) \\
&\Rightarrow -\hat{U}_H\left(\Lambda D_L^*, \Lambda D_H^*\right) \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L) + \nu(\Lambda D_H)}{cm_a \frac{\beta'(\Lambda D_H)}{1 - \beta(\Lambda D_H)}} \right] < 0 \\
&\Rightarrow \hat{U}_H\left(\Lambda D_L^*, \Lambda D_H^*\right) > 0.
\end{aligned}$$

However, this is a contradiction since $U_H\left(\Lambda D_L^*, \Lambda D_H^*\right) \le 0$ by definition. ■

The above claim proves that $D_H(p) = D_H^*$, and $D_L(p) = D_L^*$ since customer utility is zero in the best-response problem of a single high-quality provider. Moreover, another implication of the above claim is that $\eta_{4_H} > 0$, i.e. all customers request service, as we discussed in the identical providers setting. Hence, we have that $D_\ell = D_H^*$ and $p_\ell = p_H^*$. The proof of the fact that the best-response for a low-quality provider is p_L^* is the same and omitted.

Case-3 **($\Lambda^{R_L} \le \Lambda < \Lambda^0$):** In this case, we first show that customer surplus is positive in the best-response problem of a single high-quality provider.

CLAIM 5. *Let $(D_\ell(p), D_L(p), D_H(p), p_\ell(p))$ be the solution of single agent's best response problem when all other agents charge the price p. If $p = (p_H^*, p_L^*)$ as described in Theorem 6, then we have that $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] = \Delta$, where*

$$\Delta = \frac{(R_H + cm_a)\big[k_H - 1 + k_L \vartheta(\hat{D}_H(\Lambda), \Lambda D_L)\big]}{\frac{k_H + k_L \vartheta(\hat{D}_H(\Lambda), \Lambda D_L)}{1 - \nu(\hat{D}_H(\Lambda))} - 1}[1 - \beta(\hat{D}_H(\Lambda))].$$

Proof:
Suppose $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] < \Delta$. Then, we have that $\Lambda D_i(p) > \hat{D}_i(\Lambda)$ for $i \in \{H, L\}$ since

$$(R_i - p_i + cm_a)[1 - \beta(\hat{D}_i(\Lambda))] = \Delta > (R_i - p_i + cm_a)[1 - \beta(\Lambda D_i(p))],$$

where the inequality holds by the fact that $U_i\big(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)\big) > 0$ for any $\Lambda < \Lambda^0$. Moreover, $\Lambda D_i(p) > \hat{D}_i(\Lambda)$ implies that $\Lambda D_\ell(p) < \hat{D}_H(\Lambda)$ since $D_\ell(p) + (k_H - 1)D_H(p) + k_L D_L(p) = 1$.

Then using (43) and (44), we have that

$$\begin{aligned}
\eta_{2_H} &= \eta_{3_H} + \frac{(k_H - 1)}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H(p))}\eta_{4_H} \\
&= \left[\frac{k_L}{\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L(p))} + \frac{(k_H - 1)}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H(p))}\right]\eta_{4_H} \\
&> \left(\frac{k_H - 1 + k_L \vartheta(\hat{D}_H(\Lambda), \Lambda D_L)}{(R_H - p_H + cm_a)\beta'(\hat{D}_H(\Lambda))}\right)\frac{\eta_{4_H}}{\Lambda},
\end{aligned}$$

where the inequality holds since $\Lambda D_L(p) > \hat{D}_L(\Lambda)$, $\Lambda D_H(p) > \hat{D}_H(\Lambda)$, and $\beta(\lambda)$ is concave. Moreover, using (43)and the fact that $p_\ell(p) > R_H + cm_a - \frac{cm_a}{1 - \beta(\Lambda D_\ell(p))}$, we have that

$$\begin{aligned}
\frac{\eta_{4_H}}{\Lambda} &> \Big((R_H + cm_a)[1 - \nu(\Lambda D_H)][1 - \beta(\Lambda D_H)] - cm_a\Big) \\
&> \Big((R_H + cm_a)[1 - \nu(\hat{D}_H(\Lambda))][1 - \beta(\hat{D}_H(\Lambda))] - cm_a\Big),
\end{aligned}$$

where the second inequality holds since $\Lambda D_\ell(p) < \hat{D}_H(\Lambda)$ and $[1 - \nu(\lambda)][1 - \beta(\lambda)]$ is the derivative of $\lambda[1 - \beta(\lambda)]$, which is a strictly concave function.

Combining these two observations with the fact that $\eta_{2_H} \le \Lambda D_\ell(p) < \hat{D}_H(\Lambda)$ (since $\eta_{1_H} \ge 0$), and $R_H - p_H + cm_a = \frac{\Delta}{1 - \beta(\hat{D}_H(\Lambda))}$, we have that

$$\begin{aligned}
&\hat{D}_H(\Lambda) > \left(\frac{k_H - 1 + k_L \vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))}{\frac{\Delta}{1 - \beta(\hat{D}_H(\Lambda))}\beta'(\hat{D}_H(\Lambda))}\right)\Big((R_H + cm_a)[1 - \nu(\hat{D}_H(\Lambda))][1 - \beta(\hat{D}_H(\Lambda))] - \Delta\Big) \\
&\Rightarrow (R_H + cm_a)[1 - \nu(\hat{D}_H(\Lambda))][1 - \beta(\hat{D}_H(\Lambda))] - \Delta\frac{k_H - 1 + k_L \vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda)) + \nu(\hat{D}_H(\Lambda))}{k_H - 1 + k_L \vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))} < 0 \\
&\Rightarrow \Delta > \frac{(R_H + cm_a)\big[k_H - 1 + k_L \vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))\big]}{\frac{k_H + k_L \vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))}{1 - \nu(\hat{D}_H(\Lambda))} - 1}[1 - \beta(\hat{D}_H(\Lambda))] = \Delta
\end{aligned}$$

Hence, we should have that $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] \ge \Delta$.

Now, we suppose $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] > \Delta$. As the same as above (only by reversing the inequality signs), we can again have a contradiction. Therefore, we should have that $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] = \Delta$.

■

Once we have the above claim, it is clear that $\Lambda D_i(p) = \hat{D}_i(\Lambda)$ for $i \in \{H, L\}$. Moreover, the claim also implies that all customers request service since $\Delta > cm_a$ as $\Lambda < \Lambda^0$. Hence, we also have that $\Lambda D_\ell(p) = \hat{D}_H(\Lambda)$, and $p_\ell(p) = p_H^*$. Similarly, proving a claim as above for the low-quality providers, we can show that the best-response of a low-quality provider is p_L^* given (p_H^*, p_L^*).

Case-4 ($\Lambda_H^{R_L} \leq \Lambda \leq \Lambda^{R_L}$): In this case, we first show that customer surplus is exactly R_L in the best-response problem of a single high-quality provider.

Claim 6. *Let $(D_\ell(p), D_L(p), D_H(p), p_\ell(p))$ be the solution of single agent's best response problem when all other agents charge the price p. If $p = (p_H^*, p_L^*)$ as described in Theorem 6, then we have that $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] = R_L + cm_a$.*

Proof:

Suppose $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] < R_L + cm_a$. Then, we have that $D_\ell(p) > 1/k_H$ since

$$(R_H - p_H^* + cm_a)[1 - \beta(\Lambda/k_H)] = R_L + cm_a > (R_H - p_H^* + cm_a)[1 - \beta(\Lambda D_H(p))].$$

Moreover, $D_H(p) > 1/k_H$ implies that $D_\ell(p) < 1/k_H$ since $D_\ell(p) + (k_H - 1)D_H(p) + k_L D_L(p) \leq 1$.

Then using (43) and (44), we have that

$$\begin{aligned}
\eta_{2_H} &= \left[\frac{k_L}{\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L(p))} + \frac{(k_H - 1)}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H(p))}\right]\eta_{4_H} \\
&> \left[\frac{k_L}{\Lambda(R_L + cm_a)\beta'(0)} + \frac{(k_H - 1)}{\Lambda(R_H - p_H + cm_a)\beta'(\Lambda/k_H)}\right]\eta_{4_H} \\
&= \left(\frac{k_H - 1 + k_L\vartheta(\Lambda/k_H, 0)}{(R_H - p_H + cm_a)\beta'(\Lambda/k_H)}\right)\frac{\eta_{4_H}}{\Lambda},
\end{aligned}$$

where the inequality holds since $D_L(p) \geq 0$, $p_L \geq 0$, $D_H(p) > 1/k_H$, and $\beta(\lambda)$ is concave, and the equality holds since $(R_H - p_H^* + cm_a)[1 - \beta(\Lambda/k_H)] = R_L + cm_a$. Moreover, using (43)and the fact that $p_\ell(p) > R_H + cm_a - \frac{cm_a}{1-\beta(\Lambda D_\ell(p))}$, we have that

$$\frac{\eta_{4_H}}{\Lambda} > \Big((R_H + cm_a - w_H)[1 - \nu(\Lambda/k_H)][1 - \beta(\Lambda/k_H)] - (R_L + cm_a)\Big),$$

where the inequality holds since $D_\ell(p) < 1/k_H$ and $[1 - \nu(\lambda)][1 - \beta(\lambda)]$ is the derivative of $\lambda[1 - \beta(\lambda)]$, which is a strictly concave function.

Combining these two observations with the fact that $\eta_{2_H} \leq \Lambda D_\ell(p) < \Lambda/k_H$ (since $\eta_{1_H} \geq 0$), and $R_H - p_H^* + cm_a = \frac{R_L + cm_a}{1-\beta(\Lambda/k_H)}$, we have that

$$\begin{aligned}
&(R_H + cm_a - w_H)[1 - \nu(\Lambda/k_H)][1 - \beta(\Lambda/k_H)] - (R_L + cm_a)\frac{k_H - 1 + k_L\vartheta(\Lambda/k_H, 0) + \nu(\Lambda/k_H)}{k_H - 1 + k_L\vartheta(\Lambda/k_H, 0)} < 0 \\
&\Rightarrow \hat{U}_H(0, \Lambda/k_H) < R_L,
\end{aligned}$$

However, this is a contradiction because by the definition of $\Lambda(R_L)$ since we have that $\hat{U}_H(0, \Lambda/k_H) \geq R_L$ for any $\Lambda \leq \Lambda(R_L)$.

Now, suppose $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] > R_L + cm_a$. Then, we have that $D_L(p) = 0$, $D_H(p) < 1/k_H$, and $D_\ell(p) > 1/k_H$. Using these and other properties we used in previous proofs, we have that

$$\begin{aligned}
\eta_{2_H} &< \left(\frac{k_H - 1}{\frac{R_L + cm_a}{1-\beta(\Lambda/k_H)}\beta'(\Lambda/k_H)}\right)\frac{\eta_{4_H}}{\Lambda} \\
\frac{\eta_{4_H}}{\Lambda} &< (R_H + cm_a)[1 - \nu(\Lambda/k_H)][1 - \beta(\Lambda/k_H)] - (R_L + cm_a).
\end{aligned}$$

Combining these observations, and the fact that $\eta_{2_H} > \Lambda/k_H$, we have that

$$(R_H + cm_a)[1-\nu(\Lambda/k_H)][1-\beta(\Lambda/k_H)] - (R_L + cm_a)\frac{k_H - 1 + \nu(\Lambda/k_H)}{k_H - 1} > 0$$
$$\Rightarrow z^{R_L}(\Lambda) > 0.$$

However, this is a contradiction because $z^{R_L}(\Lambda)$ is decreasing in Λ, so that by the definition of Λ^{R_L}, we have $z^{R_L}(\Lambda) \le 0$ for any $\Lambda \ge \Lambda_H^{R_L}$.

Hence, we should have that $(R_H - p_\ell(p) + cm_a)[1-\beta(\Lambda D_\ell(p))] = R_L + cm_a$.

■

Similar to other case, the direct implication of the above claim is that $D_\ell(p) = 1/k_H$, and $p_\ell = p_H^*$. Furthermore, it can be shown that the solution of the best-response problem of a low-quality agent is $D_\ell(p) = 1/k_L$, and $p_\ell = p_L^*$ in a very similar way.

Case-5 **($\Lambda < \Lambda_H^{R_L}$):** In this case, only the high-quality providers are in the market, so the proof is very similar to the proof in the identical agents model. Letting $\Delta = \frac{(R_H + cm_a - w_H)(k_H - 1)[1-\nu(\Lambda/k_H)][1-\beta(\Lambda/k_H)]}{k_H - 1 + \nu(\Lambda/k_H)}$, we can show that any high-quality provider leaves $\Delta - cm_a$ surplus in his best-response, and this established the result. The proof of this claim suppose this is not true and come up with contradictions as in the proof of Case-3 of the identical agents model.

Appendix S.5: Operational Efficiency (Non-Identical Agents)

S.5.1. Proof of Lemma 8

Here, we only present the proof for $i = L,\ j = H$. For notational convenience, we let $r_{i_k} = R_i - p_{i_k}$ and $r_i = R_i - p_i$ for $i \in \{H, L\}$.

1. To prove our claim, it is sufficient to show that $\liminf_{k\to\infty} D_L(k) + P_{HL}(k)D_H(k) = 1$. We prove this claim by contradiction. Thus, we suppose $\liminf_{k\to\infty} D_L(k) + P_{HL}(k)D_H(k) < 1$ on the contrary. Then, we have convergent subsequences $D_L(k)$, $D_H(k)$, P_{HL} such that

$$\lim_{k\to\infty} D_L(k) + P_{HL}(k)D_H(k) < 1$$
$$\lim_{k\to\infty} \frac{\Lambda^k D_L(k)}{\alpha_L k} < 1,$$

where the second inequality holds since $\rho \le \alpha_L$.

Similar to the proof of Lemma 6.1, we have that

$$\lim_{k\to\infty} \beta_L(k) \le \lim_{k\to\infty} \beta^{MM1}(\Lambda^k D_L(k), \alpha_L k) = 0,$$

where the last equality is due to Ward and Glynn (2003). Using this result, we have that

$$\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) = r_L > 0, \tag{50}$$

which implies that utility of customers choosing the price p_L is strictly positive for large k, so that we should have $\lim_{k\to\infty} D_H(k) + D_L(k) = 1$ by the definition of Market Customer Equilibrium. Furthermore, we have that

$$\lim_{k\to\infty} P_{HL}(k) < 1$$

since $\lim_{k\to\infty} D_L(k) + P_{HL}(k)D_H(k) < 1$. Then, we have that

$$\begin{aligned}
&\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) \\
&\qquad = \lim_{k\to\infty} \left[1 - P_{HL}(k)\right]\left[(r_H + cm_a)\left[1-\beta_H(k)\right] - cm_a\right] + \lim_{k\to\infty} P_{HL}(k)(r_L) \\
&\qquad \le \lim_{k\to\infty} \left[1 - P_{HL}(k)\right](r_H) + \lim_{k\to\infty} P_{HL}(k)(r_L) < r_L,
\end{aligned} \tag{51}$$

where the first inequality holds since $\beta_H(k) \ge 0$, and the second one holds since $\lim_{k\to\infty} P_{HL}(k) < 1$. Then, combining 50 and 51, we have that

$$\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) < \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k)$$

which implies that customers are strictly better-off by choosing sub-pool-2 over sub-pool-1 for sufficiently large k. This contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_H > 0$, i.e. customers choose sub-pool-1 in sufficiently large systems. Hence, we should have that $\liminf_{k\to\infty} D_L(k) + P_{HL}(k)D_H(k) = 1$.

2.a) (*The proof is very similar to the proof of Lemma 6.1*)
We again prove our claim by contradiction. Thus, we suppose that $\liminf_{k\to\infty} \frac{\Lambda^k D_L(k)}{\alpha_L k} \le 1$. Then, there exists a convergent subsequence such that

$$\liminf_{k\to\infty} \frac{\Lambda^k D_L(k)}{\alpha_L k} \le 1.$$

Similar to the proof of Lemma 6.1, we have that

$$\lim_{k\to\infty} \beta_L(k) \le \lim_{k\to\infty} \beta^{MM1}(\Lambda^k D_L(k), \alpha_L k) = 0,$$

where the last equality is due to Ward and Glynn (2003). Using this result, we have that

$$\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) = r_L > 0,$$

which implies that utility of customers choosing the price p_{L_k} is strictly positive for large k, so that we should have $\lim_{k\to\infty} D_H(k) + D_L(k) = 1$ by the definition of Market Customer Equilibrium. Furthermore, using the fact that the rate of arrival to sub-pool-2 is equal to the rate of departure (either by service or abandonment), we have that

$$\Lambda^k D_H(k)P_{HL}(k) + \Lambda^k D_L(k) = \alpha_L k \sigma_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) + \Lambda^k D_L(k)\beta_L(k).$$

Dividing both sides by Λ^k, the above equation implies that

$$P_{HL}(k)\big[D_H(k)+D_L(k)\big] \le \frac{\alpha_L k}{\Lambda^k}\sigma_2(D_H(k),D_L(k);r_{H_k},r_{L_k};\alpha_H k,\alpha_L k)+D_L(k)\beta_L(k).$$

Letting k go to infinity, we have that

$$\lim_{k\to\infty} P_{HL}(k) \le \frac{\alpha_L}{\rho} < 1.$$

Then, as in the proof of Part a, we have that

$$\lim_{k\to\infty} U_1(D_H(k),D_L(k);r_{H_k},r_{L_k};\alpha_H k,\alpha_L k) < \lim_{k\to\infty} U_2(D_H(k),D_L(k);r_{H_k},r_{L_k};\alpha_H k,\alpha_L k)$$

which implies that customers are strictly better-off by choosing sub-pool-2 over sub-pool-1 for sufficiently large k. This contradicts with the definition of customer equilibrium since

$$\lim_{k\to\infty} D_H(k) = 1 - \lim_{k\to\infty} D_L(k) \ge 1 - \frac{\alpha_L}{\rho} > 0,$$

i.e. customers choose sub-pool-1 in sufficiently large systems. Hence, we should have that $\liminf_{k\to\infty} \frac{\Lambda^k D_L(k)}{\alpha_L k} > 1$.

2.b) The proof is almost the same as the proof of Lemma 6.2.

2.c) To prove our claim, we first show that

$$\lim_{k\to\infty} \beta_H(k) = \lim_{k\to\infty} \beta^M(\Lambda^k D_H(k);\alpha_H k).$$

Note that $\beta_H(k) \le \beta^M(\Lambda^k D_H(k);\alpha_H k)$, since some of the customers choosing sub-pool-1 can be served by sub-pool-2. Therefore, it is sufficient to show that

$$\liminf_{k\to\infty} \beta_H(k) \ge \lim_{k\to\infty} \beta^M(\Lambda^k D_H(k);\alpha_H k).$$

We prove this claim as in the proof of Lemma 7.3. We consider a hypothetical situation where any customer choosing the price p_H is duplicated when there is an idle agent in sub-pool-2, and one of these copies goes to sub-pool-2 while the other one is colored and goes to sub-pool-1. Furthermore, any non-colored customer in sub-pool-1 has service priority.

This hypothetical sub-pool-1 operates as $M/M/\alpha_H k + M$ system with arrival rate $\Lambda^k D_H(k)$, so that total abandonment rate is $\Lambda^k D_H(k)\beta^M(\Lambda^k D_H(k);\alpha_H k)$. In other words, we have that

$$\Lambda^k D_H(k)\beta^M(\Lambda^k D_H(k);\alpha_H k) = \Lambda^k D_H(k)\beta_H(k) + \Lambda^k D_H(k)\beta^{color}(k),$$

where $\beta^{color}(k)$ is the probability that colored customers abandon the hypothetical system.

In the hypothetical sub-pool-1, the abandonment rate of non-colored customers is the same as the abandonment rate in the real sub-pool-1. Moreover, since some of the colored customers can be served before abandoning the system, we have that $\beta^{color}(k) \leq P_{HL}(k)$. Thus, we have that

$$\beta_H(k) \geq \beta^M(\Lambda^k D_H(k); \alpha_H k) - P_{HL}(k).$$

Then, using this result and part ii, we have that

$$\liminf_{k\to\infty} \beta_H(k) \geq \lim_{k\to\infty} \beta^M(\Lambda^k D_H(k); \alpha_H k).$$

Finally the result holds since $\lim_{k\to\infty} \beta^M(\Lambda^k D_H(k); \alpha_H k) = \max\left\{0, 1 - \frac{\alpha_H}{\rho \tilde{D}_H}\right\}$ by by Lemma 4.

2.d) (*The proof is very similar to the proof of Lemma 6.5*)

We let π_n be the steady-state probability of having n customers in sub-pool-2. Then, we have that

$$\begin{aligned}
\beta_L(k) &= \sum_{n=\alpha_L k+1}^{\infty} (n - \alpha_L k + 1)/m_a \frac{\pi_n}{\Lambda^k D_L(k)} \\
&= \left(\sum_{n=\alpha_L k+1}^{\infty} \pi_n\right)\left(1 - \frac{\alpha_L k}{\Lambda^k D_L(k)}\right) + \pi_{\alpha_L k}.
\end{aligned}$$

Then, the result follows by letting $k \to \infty$ and using the fact from part ii that $\lim_{k\to\infty} \sum_{n=0}^{\alpha_L k} \pi_n = 0$.

3. Let $D_L^{eq} = \max\left\{\min\left\{1, \left(\frac{r_L + cm_a}{r_H + cm_a}\right)\frac{\alpha_L}{\rho}\right\}, \min\left\{\frac{r_L + cm_a}{\bar{R}}\alpha_L, \left(\frac{r_L + cm_a}{cm_a}\right)\frac{\alpha_L}{\rho}\right\}\right\}$.

<u>**Liminf:**</u> We first show that $\liminf_{k\to\infty} D_L(k) \geq D_L^{eq}$ by contradiction. Thus, we suppose that $\liminf_{k\to\infty} D_L(k) < D_L^{eq}$. Then, there exists a convergent subsequence such that

$$\tilde{D}_L := \lim_{k\to\infty} D_L(k) < D_L^{eq}.$$

Using this inequality and the fact that $\rho D_L^{eq} > \alpha_L$ (since $\rho > \alpha_L$), we have that

$$\begin{aligned}
\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) &= (r_L + cm_a)\left[1 - \lim_{k\to\infty} \beta_L(k)\right] - cm_a \\
&= (r_L + cm_a)\min\left\{1, \frac{\alpha_L}{\rho \tilde{D}_L}\right\} - cm_a \\
&> (r_L + cm_a)\frac{\alpha_L}{\rho D_L^{eq}} - cm_a \\
&\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \geq 0,
\end{aligned}$$

which implies that utility of customers choosing the price p_L is strictly positive for large k, so that we should have $\lim_{k\to\infty} D_H(k) + D_L(k) = 1$ by the definition of Market Customer Equilibrium. Furthermore, we have that

$$\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) = (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho \tilde{D}_L}\right\} - cm_a$$

$$\leq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho D_L^{eq}}\right\} - cm_a$$
$$\leq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}$$
$$< \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k)$$

which implies that customers are strictly better-off by choosing sub-pool-2 over sub-pool-1 for sufficiently large k. This contradicts with the definition of customer equilibrium since

$$\lim_{k\to\infty} D_H(k) = 1 - \lim_{k\to\infty} D_L(k) > 1 - D_L^{eq} \geq 0,$$

i.e. customers choose sub-pool-1 in sufficiently large systems. Hence, we should have that $\liminf_{k\to\infty} D_L(k) \geq D_L^{eq}$.

Limsup: Now, we show that $\limsup_{k\to\infty} D_L(k) \leq D_L^{eq}$. Note that $D_L^{eq} = 1$ when $\frac{r_L + cm_a}{r_H + cm_a} \geq \frac{\rho}{\alpha_L}$, so that this is claim is obviously true. Thus, it sufficient to show that $\limsup_{k\to\infty} D_L(k) \leq D_L^{eq}$ in the case where $D_L^{eq} < 1$.

As above, we show this claim by contradiction. Therefore, we suppose that $\limsup_{k\to\infty} D_L(k) > D_L^{eq}$. Then, there exists a convergent subsequence such that

$$\tilde{D}_L := \lim_{k\to\infty} D_L(k) > D_L^{eq}.$$

Using this inequality, we have that

$$\begin{aligned}\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) &= (r_L + cm_a)\frac{\alpha_L}{\rho \tilde{D}_L} - cm_a \\ &< \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{52}$$

Furthermore, letting $\tilde{D}_H = \lim_{k\to\infty} D_H(k)$ and using parts 2.b and 2.c, we have that

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho \tilde{D}_H}\right\} - cm_a \\ &\geq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho D_L^{eq}}\right\} - cm_a \\ &\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{53}$$

Combining (52) and (53), we have that

$$\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) > \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k)$$

which implies that customers are strictly better-off by choosing sub-pool-1 over sub-pool-2 for sufficiently large k. This contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_L(k) > 0$, i.e. customers choose sub-pool-2 in sufficiently large systems. Hence, we should have that $\limsup_{k\to\infty} D_L(k) \leq D_L^{eq}$.

4. Note that $\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) \geq \min\left\{r_j, \frac{\bar{R}}{cm_a}\right\} > 0$ when $\rho < \frac{\bar{R}}{cm_a}$ and $r_j > 0$ by part 3. Thus, we should have that the total rate of customers requesting is equal to 1. Furthermore, the statement for $r_j = 0$ is trivially true.

5. <u>Liminf</u>: We first show that $\liminf_{k\to\infty} D_H(k) \geq \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$ by contradiction. Thus, we suppose that $\liminf_{k\to\infty} D_H(k) < \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$. Then, there exists a convergent subsequence such that

$$\tilde{D}_H := \lim_{k\to\infty} D_H(k) < \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}.$$

Using this inequality, we have that

$$\begin{aligned}
\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho\tilde{D}_H}\right\} - cm_a \\
&> (r_H + cm_a)\frac{cm_a}{r_H + cm_a} - cm_a \\
&= 0 = \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k),
\end{aligned}$$

where the last equality holds since $\lim_{k\to\infty} D_L(k) = \left(\frac{r_L + cm_a}{cm_a}\right)\frac{\alpha_L}{\rho}$ when $\rho \geq \frac{\bar{R}}{cm_a}$ by part 3. The above inequality which implies that customers are strictly better-off by choosing sub-pool-1 over sub-pool-2 for sufficiently large k. This contradicts with the definition of customer equilibrium since

$$\lim_{k\to\infty} D_L(k) > 0,$$

i.e. customers choose sub-pool-2 in sufficiently large systems. Hence, we should have that $\liminf_{k\to\infty} D_H(k) \geq \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$.

<u>Limsup</u>: Now, we show that $\limsup_{k\to\infty} D_H(k) \leq \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$. As above, we show this claim by contradiction. Therefore, we suppose that $\limsup_{k\to\infty} D_L(k) > \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$. Then, there exists a convergent subsequence such that

$$\tilde{D}_H := \lim_{k\to\infty} D_H(k) > \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}.$$

Using this inequality, we have that

$$\begin{aligned}
\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_{H_k}, r_{L_k}; \alpha_H k, \alpha_L k) &= (r_H + cm_a)\frac{\alpha_H}{\rho\tilde{D}_H} - cm_a \\
&< (r_H + cm_a)\frac{cm_a}{r_H + cm_a} - cm_a = 0.
\end{aligned}$$

which implies that customers are getting strictly negative utility by choosing sub-pool-1 for sufficiently large k. This contradicts with the definition of customer equilibrium since we suppose $\lim_{k\to\infty} D_H(k) > 0$, i.e. customers choose sub-pool-1 in sufficiently large systems. Hence, we should have that $\limsup_{k\to\infty} D_H(k) \leq \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$.

S.5.2. Proof of Lemma 9

We only the give the proof of $\lim_{k\to\infty} V_{i_{dev}}(k) = p_i + \varepsilon$ for $i = L,\ j = H$. As we need to prove some claims before proving the claim in Lemma 9, we state the lemma in a different way as below. Then we prove the new version of Lemma 9. For notational convenience, we let $r_i^k = R_i - {p_i}^k$ and $r_i = R_i - p_i$ for $i \in \{H, L\}$. The proof for the other three cases are the same.

Restatement of Lemma 9: *Consider a sequence of marketplaces indexed by the number of agents, i.e. there are k agents in the k^{th} marketplace, and assume that arrival rate in the k^{th} marketplace is $\Lambda^k = \rho k$ for some $0 < \rho < 1$.*

$$\begin{aligned}
D_H(k) &= D_1^{MCE}(r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\\
D_{L_{dev}}(k) &= D_2^{MCE}(r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\\
D_L(k) &= D_3^{MCE}(r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\\
P_{HL}(k) &= PServ_{13}(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\\
P_{HL_{dev}}(k) &= PServ_{12}(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\\
P_{L_{dev}L}(k) &= PServ_{23}(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)
\end{aligned}$$

where $\lim_{k\to\infty} r_i^k = r_i \geq 0$ for $i \in \{H, L\}$, and $1 < \frac{r_L + cm_a}{r_H + cm_a} < \frac{\rho}{\alpha_L}$. Then, we have that

1. $\liminf_{k\to\infty} \frac{\Lambda^k D_L(k)}{k_L} > 1$.
2. $\lim_{k\to\infty} P_{HL}(k) + P_{L_{dev}L}(k) = 0$.
3. $\lim_{k\to\infty} D_H(k) P_{HL_{dev}}(k) = 0$.
4. $\lim_{k\to\infty} D_{L_{dev}}(k) = 0$.
5. $\lim_{k\to\infty} \beta_1\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1\big) = \max\left\{0, 1 - \frac{\alpha_H}{\rho \tilde{D}_H}\right\}$,
   where $\tilde{D}_H = \lim_{k\to\infty} D_H(k)$.
6. $\lim_{k\to\infty} \beta_3\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1\big) = 1 - \frac{\alpha_L}{\rho \tilde{D}_L}$,
   where $\tilde{D}_L = \lim_{k\to\infty} D_L(k)$.
7. $\lim_{k\to\infty} D_L(k) = D_L^{eq}$,
   where $D_L^{eq} = \max\left\{\min\left\{1, \left(\frac{r_L + cm_a}{r_H + cm_a}\right)\frac{\alpha_L}{\rho}\right\}, \min\left\{\frac{r_L + cm_a}{\bar{R}}\alpha_L, \left(\frac{r_L + cm_a}{cm_a}\right)\frac{\alpha_L}{\rho}\right\}\right\}$.
8. $\lim_{k\to\infty} V_{L_{dev}}(k) = p_L + \varepsilon$,
   where $\varepsilon < r_L - r_H$, and $V_{L_{dev}}(k)$ is the profit of a low-quality charging $p_L + \varepsilon$ when all other low-quality providers charge p_{L_k}, all high-quality providers charge p_{H_k} in the k^{th} marketplace, and $r_H > 0$.
9. $\lim_{k\to\infty} V_{L_{dev}}(k) = p_L + \varepsilon$,
   where $\varepsilon < r_L - r_H$, and $V_{L_{dev}}(k)$ is the profit of a low-quality charging $p_L + \varepsilon$ when all other low-quality providers charge p_{L_k}, all high-quality providers charge p_{H_k} in the k^{th} marketplace, $r_H = 0$, and the limiting revenue of high-quality agents is strictly positive before deviation.

Proof:

1. We prove our claim by contradiction. Hence, we suppose that $\liminf_{k\to\infty} \frac{\Lambda^k D_L(k)}{k_L} \leq 1$ on the contrary. Then, there exists a convergent subsequence of $D_L(k)$ such that

$$\lim_{k\to\infty} \frac{\Lambda^k D_L(k)}{k_L} \leq 1.$$

Next, similar to the proof of Lemma 6.1, we have that

$$\lim_{k\to\infty} \beta_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) \leq \lim_{k\to\infty} \beta^{MM1}(\Lambda^k D_L(k), \alpha_L k) = 0,$$

where $\beta^{MM1}(\lambda, k)$ is the probability of abandonment in $M/M/1+M$ system with arrival rate λ, service rate k, and abandonment rate $1/m_a$. Since $\lim_{k\to\infty} \frac{\Lambda^k D_L(k)}{k_L} \le 1$ we have the above result due to Ward and Glynn (2003). Using this result, we have that

$$\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) = r_L > 0, \tag{54}$$

which implies that utility of customers choosing the price p_L is strictly positive for large k, so that we should have $\lim_{k\to\infty} D_H(k) + D_{H_{dev}L}(k) + D_L(k) = 1$ by the definition of Market Customer Equilibrium.

Now, we argue that $\lim_{k\to\infty} P_{HL}(k)D_H(k) = \lim_{k\to\infty} D_H(k)$. In order to prove that it is enough to show $\lim_{k\to\infty} P_{HL}(k) = 1$ whenever $\lim_{k\to\infty} D_H(k) > 0$. To show that we suppose $\lim_{k\to\infty} P_{HL}(k) < 1$ when $\lim_{k\to\infty} D_H(k) > 0$. Then, we have that

$$\begin{aligned}
&\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) \\
&\qquad \le \big[1 - \lim_{k\to\infty} P_{HL}(k)\big](r_L - \varepsilon) + \big[\lim_{k\to\infty} P_{HL}(k)\big](r_L) \\
&\qquad < r_L \\
&\qquad = \lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1).
\end{aligned}$$

However, this is a contradiction because customers are strictly better of by choosing sub-pool-3 for large k while $\lim_{k\to\infty} D_H(k) > 0$. Hence, we have that $\lim_{k\to\infty} P_{HL}(k)D_H(k) = \lim_{k\to\infty} D_H(k)$. Similarly, we can show that $\lim_{k\to\infty} P_{L_{dev}L}(k)D_{L_{dev}}(k) = \lim_{k\to\infty} D_{L_{dev}}(k)$.

Furthermore, using the fact that the rate of arrival to sub-pool-3 is equal to the rate of departure (either by service or abandonment), we have that

$$\begin{aligned}
&\Lambda^k D_H(k)P_{HL}(k) + \Lambda^k D_{H_{dev}L}(k)P_{L_{dev}L}(k) + \Lambda^k D_L(k) \\
&\qquad = k_L \sigma_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) \\
&\qquad\quad + \Lambda^k D_L(k)\beta_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1).
\end{aligned}$$

Dividing both sides by Λ^k and letting $k \to \infty$, the above equation implies that

$$\begin{aligned}
1 &= \lim_{k\to\infty} D_H(k) + D_{H_{dev}L}(k) + D_L(k) = \lim_{k\to\infty} D_H(k)P_{HL}(k) + D_{H_{dev}L}(k)P_{L_{dev}L}(k) + D_L(k) \\
&= \frac{\alpha_L}{\rho} \lim_{k\to\infty} \sigma_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) \le \frac{\alpha_L}{\rho} < 1,
\end{aligned}$$

where the second equality holds since $\lim_{k\to\infty} P_{HL}(k)D_H(k) = \lim_{k\to\infty} D_H(k)$, and $\lim_{k\to\infty} P_{L_{dev}L}(k)D_{L_{dev}}(k) = \lim_{k\to\infty} D_{L_{dev}}(k)$. The above inequality is a clear contradiction. Hence, we should have that $\liminf_{k\to\infty} \frac{\Lambda^k D_L(k)}{k_L} > 1$.

2. *The proof is very similar to the proof of Lemma 6.2.*

Let π_n be the steady-state probability of having n customers in sub-pool-2, and π_n^M be the steady-state probability of having n customers in an $M/M/\alpha_L k+M$ system with arrival rate $\Lambda^k D_L(k)$, service rate 1, and abandonment rate $1/m_a$. By studying the birth-death chain of both systems, we have that

$$\sum_{n=0}^{\alpha_L k}\pi_n \le \frac{\sum\limits_{n=0}^{\alpha_L k}\frac{\left[\Lambda^k(D_H(k)+D_{L_{dev}}(k)+D_L(k))\right]^{n-\alpha_L k}(\alpha_L k)!}{n!}}{\sum\limits_{n=0}^{\alpha_L k}\frac{\left[\Lambda^k(D_H(k)+D_{L_{dev}}(k)+D_L(k))\right]^{n-\alpha_L k}(\alpha_L k)!}{n!}+\sum\limits_{n=\alpha_L k+1}^{\infty}\frac{\left[\Lambda^k D_L(k)\right]^{n-\alpha_L k}}{\prod_{i=1}^{n-\alpha_L k}(k+i/m_a)}}$$

$$\le \frac{\sum\limits_{n=0}^{\alpha_L k}\frac{\left[\Lambda^k D_L(k)\right]^{n-\alpha_L k}(\alpha_L k)!}{n!}}{\sum\limits_{n=0}^{\alpha_L k}\frac{\left[\Lambda^k D_L(k)\right]^{n-\alpha_L k}(\alpha_L k)!}{n!}+\sum\limits_{n=\alpha_L k+1}^{\infty}\frac{\left[\Lambda^k D_L(k)\right]^{n-\alpha_L k}}{\prod_{i=1}^{n-\alpha_L k}(k+i/m_a)}}=\sum_{n=0}^{\alpha_L k}\pi_n^M,$$

where the inequality holds since $\frac{x}{x+A}$, where $A>0$ is a constant, is increasing in x.

Using the above relation, we have that

$$\lim_{k\to\infty} P_{HL}(k)+P_{L_{dev}L}(k) = \lim_{k\to\infty}\sum_{n=0}^{\alpha_L k}\pi_n \le \lim_{k\to\infty}\sum_{n=0}^{\alpha_L k}\pi_n^M = 0,$$

where the last equality holds since $\liminf\limits_{k\to\infty}\frac{\Lambda^k D_L(k)}{\alpha_L k}>1$.

3. Using the fact that the rate of arrival to sub-pool-2 is equal to the rate of departure (either by service or abandonment), we have that

$$\begin{aligned}\Lambda^k D_H(k)P_{HL_{dev}}+\Lambda^k D_{L_{dev}} &= \sigma_2\big(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1\big)\\ &\quad +\Lambda^k D_{L_{dev}}\beta_2\big(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1\big),\end{aligned}$$

which implies that

$$D_H(k)P_{HL_{dev}}=\frac{1}{\Lambda^k}+D_{L_{dev}}\left[\beta_2\big(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1\big)-1\right]\le\frac{1}{\Lambda^k}.$$

Then, we obtain the result by letting $k\to\infty$.

4. Note it is sufficient to show that $\lim\limits_{k\to\infty} D_{L_{dev}}(k)=0$. We prove this claim by contradiction. Hence, we suppose that $\lim\limits_{k\to\infty} D_{L_{dev}}(k)>0$ on the contrary. Then, there exists a convergent subsequence of $D_{L_{dev}}(k)$ such that

$$\tilde{D}_{L_{dev}} := \lim_{k\to\infty} D_{L_{dev}}(k)>0.$$

Using this observation and the fact that $\lim_{k\to\infty} P_{L_{dev}L}=0$ from part 2, we have that

$$\begin{aligned}&\lim_{k\to\infty} U_2(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)\\&\quad=(r_L+\varepsilon)\big[1-\lim_{k\to\infty}\beta_2(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)\big]-cm_a\\&\quad\le(r_L+\varepsilon)\big[1-\lim_{k\to\infty}\beta^M(\Lambda^k D_{L_{dev}}(k),1)\big]-cm_a=-cm_a,\end{aligned}$$

where the inequality holds since some of customers choosing sub-pool-1 may be served by sub-pool-2, and the last equality holds since the probability of abandonment goes to 1 in a single server system as the arrival rate goes to infinity. The above inequality implies that the expected utility of customers choosing sub-pool-2 is strictly negative for sufficiently large k. However, this contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_{L_{dev}}(k)>0$. Hence, we should have that $\limsup_{k\to\infty} D_{L_{dev}}(k)=0$.

5. (*The proof is very similar to the proof of Lemma 8 part 2.c*)

As in Lemma 8.b.iii, it is sufficient to show that $\liminf_{k\to\infty}\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)\ge 1-\frac{\alpha_H}{\rho}$. We prove this claim by considering a hypothetical situation where any customer choosing the price p_H is duplicated when there is an idle agent in sub-pool-2 or in sub-pool-3, and one of these copies goes to either sub-pool-2 or sub-pool-3 while the other one is colored and goes to sub-pool-1. Furthermore, any non-colored customer in sub-pool-1 has service priority.

This hypothetical sub-pool-1 operates as $M/M/k_H+M$ system with arrival rate $\Lambda^k D_H(k)$, so that total abandonment rate is $\Lambda^k D_H(k)\beta^M(\Lambda^k D_H(k),k_H)$. In other words, we have that

$$\begin{aligned}\Lambda^k D_H(k)\beta^M(\Lambda^k D_H(k),k_H)&=\Lambda^k D_H(k)\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)\\&\quad+\Lambda^k D_H(k)\beta^{color}(k),\end{aligned}$$

where $\beta^{color}(k)$ is the probability that colored customers abandon the hypothetical system.

In the hypothetical sub-pool-1, the abandonment rate of non-colored customers is the same as the abandonment rate in the real sub-pool-1. Moreover, since some of the colored customers can be served before abandoning the system, we have that $\beta^{color}(k)\le P_{HL_{dev}}(k)+P_{HL}(k)$. Thus, we have that

$$\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)\ge\beta^M(\Lambda^k D_H(k),k_H)-P_{HL_{dev}}(k)-P_{HL}(k).$$

Finally, we have that

$$\liminf_{k\to\infty}\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)\ge\lim_{k\to\infty}\beta^M(\Lambda^k D_H(k),k_H)=1-\frac{\alpha_H}{\rho\tilde{D}_H},$$

by using part 2 and 3, and the fact that $\tilde{D}_H>\frac{\alpha_H}{\rho}$.

6. (*The proof is very similar to the proof of Lemma 8* part 2.4)

We let π_n be the steady-state probability of having n customers in sub-pool-2. Then, we have that

$$\beta_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) = \left(\sum_{n=\alpha_L k+1}^{\infty} \pi_n\right)\left(1 - \frac{\alpha_L k}{\Lambda^k D_L(k)}\right) + \pi_{\alpha_L k}.$$

Then, the result follows by letting $k \to \infty$ and using the fact from part 2 that $\lim_{k\to\infty} \sum_{n=0}^{\alpha_L k} \pi_n = 0$ due to the fact that $\lim_{k\to\infty} \frac{\Lambda^k D_L(k)}{\alpha_L k} > 1$.

7. <u>Liminf</u> : We first show that $\liminf_{k\to\infty} D_L(k) \geq D_L^{eq}$ by contradiction. Thus, we suppose that $\liminf_{k\to\infty} D_L(k) < D_L^{eq}$. Then, there exists a convergent subsequence such that

$$\tilde{D}_L := \lim_{k\to\infty} D_L(k) < D_L^{eq}.$$

Using this inequality, we have that

$$\begin{aligned}\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) &= (r_L + cm_a)\frac{\alpha_L}{\rho \tilde{D}_L} - cm_a \\ &> (r_L + cm_a)\frac{\alpha_L}{\rho D_L^{eq}} - cm_a \\ &\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \geq 0,\end{aligned}$$

which implies that utility of customers choosing the price p_L is strictly positive for large k, so that we should have $\lim_{k\to\infty} D_H(k) + D_L(k) = 1$ by the definition of Market Customer Equilibrium. Using this observation, we have that $\lim_{k\to\infty} D_H(k) > 1 - \tilde{D}_L \geq 1 - D_L^{eq} \geq 0$, which implies that $\lim_{k\to\infty} P_{HL_{dev}}(k) = 0$ by part 3. Furthermore, we show that $\lim_{k\to\infty} P_{HL}(k) = 0$ in part 2. Combining these observations, we have that

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho \tilde{D}_L}\right\} - cm_a \\ &\leq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \\ &< \lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\end{aligned}$$

which implies that customers are strictly better-off by choosing sub-pool-3 over sub-pool-1 for sufficiently large k. This contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_H(k) > 0$, i.e. customers choose sub-pool-1 in sufficiently large systems. Hence, we should have that $\liminf_{k\to\infty} D_L(k) \geq D_L^{eq}$.

<u>Limsup:</u> Now, we show that $\limsup_{k\to\infty} D_L(k) \leq \min\left\{1, \left(\frac{r_L + cm_a}{r_H + cm_a}\right)\frac{\alpha_L}{\rho}\right\}$. Note that $D_L^{eq} < 1$ since $\frac{r_L + cm_a}{r_H + cm_a} \geq \frac{\alpha_L}{\rho}$.

As above, we show this claim by contradiction. Therefore, we suppose that $\limsup_{k\to\infty} D_L(k) > D_L^{eq}$. Then, there exists a convergent subsequence such that

$$\tilde{D}_L := \lim_{k\to\infty} D_L(k) > D_L^{eq}.$$

Using this inequality, we have that

$$\begin{aligned}\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) &= (r_L + cm_a)\frac{\alpha_L}{\rho\tilde{D}_L} - cm_a \\ &< \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{55}$$

Furthermore, letting $\tilde{D}_H = \lim_{k\to\infty} D_H(k)$ and using part 5 and the fact some of the customers choosing sub-pool-1 may be served by other pools, we have that

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) &\geq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho\tilde{D}_H}\right\} - cm_a \\ &\geq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho D_L^{eq}}\right\} - cm_a \\ &\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{56}$$

Combining (55) and (56), we have that

$$\begin{aligned}&\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) \\ &\quad > \lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1)\end{aligned}$$

which implies that customers are strictly better-off by choosing sub-pool-1 over sub-pool-2 for sufficiently large k. This contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_L(k) > 0$, i.e. customers choose sub-pool-2 in sufficiently large systems. Hence, we should have that $\limsup_{k\to\infty} D_L(k) \leq D_L^{eq}$.

8. To prove this claim, we first show that

$$\lim_{k\to\infty} D_H(k) = \min\left\{1 - D_L^{eq}, \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}\right\} > 0.$$

Once we show this result, then the utilization of the low-quality provider charging $p_L + \varepsilon$ will converge to 1 since all customers choosing high-quality providers first visit him.

We first consider the case where $\rho < \frac{\bar{R}}{cm_a}$. In this case, we have that

$$\begin{aligned}\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H^k, r_L - \varepsilon, r_L^k; \alpha_H k, 1, \alpha_L k - 1) &= (r_L + cm_a)\frac{\alpha_L}{\rho D_L^{eq}} - cm_a \\ &\geq \min\{\frac{\bar{R}}{\rho} - cm_a, r_H\} > 0.\end{aligned}$$

Hence, we should have that $\lim_{k\to\infty} D_H(k)+D_{L_{dev}}(k)+D_L(k)=1$, and it implies that $\lim_{k\to\infty} D_H(k)=1-D_L^{eq}$ by part 4.

Now, we focus on the case where $\rho\geq\frac{\bar{R}}{cm_a}$. We first want to note that

$$\lim_{k\to\infty} U_3(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_H k,1,\alpha_L k-1)=0$$

by using part 7. As before, we prove our claim by arguing that

$$\left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}\leq\liminf_{k\to\infty} D_H(k)\leq\limsup_{k\to\infty} D_H(k)\leq\left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}. \tag{57}$$

Suppose $\liminf_{k\to\infty} D_H(k)<\left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$. Then, the expected utility of customers from high-quality pool would converge to a strictly positive limit. This would mean that customers strictly prefer high-quality pool over providers charging p_L for large k. However, this would contradict with the definition of the customer equilibrium since $\lim_{k\to\infty} D_L(k)>0$.

Now, we suppose $\limsup_{k\to\infty} D_H(k)>\left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$. Then, the expected utility of customers from high-quality pool would converge to a strictly negative limit. However, this would contradict with the definition of the customer equilibrium since we suppose $\limsup_{k\to\infty} D_H(k)>0$.

Thus, (57) holds, and we have that $\lim_{k\to\infty} D_H(k)=\left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$ when $\rho\geq\frac{\bar{R}}{cm_a}$.

9. First note that we have that

$$\lim_{k\to\infty} D_H^{pre}(k)=\tilde{D}_H^{pre}>0,$$

where $D_H^{pre}(k)$ is the rate of customers choosing high-quality agents before a single agent cuts his price (Otherwise high-quality agents couldn't have a strictly positive revenue in the limit). Also note that $\tilde{D}_H^{pre}\leq 1-D_L^{eq}$ by Lemma 8.

Under this situation, we will show that the total rate of customers requesting service from the high-quality agents after the deviation is strictly positive. In particular, we show that

$$\lim_{k\to\infty} D_H(k)=\tilde{D}\geq\alpha_H\tilde{D}_H^{pre}>0.$$

We prove this claim by contradiction. Thus, we suppose $\tilde{D}<\alpha_H\tilde{D}_H^{pre}$, on the contrary. Then, there should exist a convergent subsequences of $\{D_H(k)\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} D_H(k)=\tilde{D}<\alpha_H\tilde{D}_H^{pre}.$$

Using this observation, for large k, we have that

$$U_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H^k,r_L-\varepsilon,r_L^k;\alpha_j k-1,\alpha_i k,1)$$

$$
\begin{aligned}
&= \left[1 - P_{HL}(k) - P_{HL_{dev}}(k)\right]\left[(r_H^k + cm_a)\big(1 - \beta_H(k)\big) - cm_a\right] + P_{HL_{dev}}(k)(r_L - \varepsilon) + P_{HL}(k) r_L^k \\
&\geq \left[1 - P_{HL}(k)\right]\left[(r_H^k + cm_a)\big(1 - \beta_H(k)\big) - cm_a\right] + P_{HL}(k) r_L^k \\
&\geq \left[1 - P_{HL}(k)\right]\left[(r_H^k + cm_a)\big(1 - \beta^M(\Lambda^k D_H(k); \alpha_H k)\big) - cm_a\right] + P_{HL}(k) r_L^k \\
&\geq \left[1 - P_{HL}^{pre}(k)\right]\left[(r_H^k + cm_a)\big(1 - \beta^M(\Lambda^k D_H(k); \alpha_H k)\big) - cm_a\right] + P_{HL}^{pre}(k) r_L^k \\
&> \left[1 - P_{HL}^{pre}(k)\right]\left[(r_H^k + cm_a)\big(1 - \beta^M(\Lambda^k D_H^{pre}(k); k)\big) - cm_a\right] + P_{HL}^{pre}(k) r_L^k \\
&\geq \left[1 - P_{HL}^{pre}(k)\right]\left[(r_H^k + cm_a)\big(1 - \beta_H^{pre}(k)\big) - cm_a\right] + P_{HL}^{pre}(k) r_L^k \geq 0,
\end{aligned}
$$

where $P_{HL}^{pre}(k)$ is the probability with which a customer choosing high-quality agents is served by low-quality agents, and $\beta_H^{pre}(k)$ is the probability of abandonment for customers choosing quality-H agents in the equilibrium before deviation. We also denote the probability of abandonment for customers choosing quality-H agents in the equilibrium after deviation by $\beta_H(k)$.

The first inequality holds since $r_L - \varepsilon > r_H^k$. The second inequality holds since not all of the customers choosing high-quality agents are served by high-quality agents. The third inequality holds since the probability with which a customer choosing high-quality agents is served by low-quality agent is higher after the deviation as a result of our assumption that $\tilde{D} < \alpha_H \tilde{D}_H^{pre}$ and by parts 4 and 7.

The forth inequality holds since $\tilde{D} < \alpha_H \tilde{D}_H^{pre}$ and by the fact that $\beta^M(\Lambda^k D_H(k); \alpha_H k) \simeq \left(\rho\tilde{D}/\alpha_H e^{-(1-\rho\tilde{D}/\alpha_H)}\right)^k$, $\beta^M(\Lambda^k D_H(k)^{pre}; k) \simeq \left(\rho\tilde{D}_H^{pre} e^{-(1-\rho\tilde{D}_H^{pre})}\right)^k$, and $\frac{\left(\rho\tilde{D}/\alpha_H e^{-(1-\rho\tilde{D}/\alpha_H)}\right)^k}{\left(\rho\tilde{D}_H^{pre} e^{-(1-\rho\tilde{D}_H^{pre})}\right)^k} \to 0$ as $k \to \infty$ similar to what we discuss in the proof of Proposition 3. Finally, the fifth inequality holds since we $\beta^M(\Lambda^k D_H(k)^{pre}; k)$ is the probability of abandonment in a hypothetical system where the customers choosing high-quality agents are served by k agents instead of $\alpha_H k$. This hypothetical system always serves more of the customers choosing the high-quality agents(or, equivalently, causes less abandonment) because in the real system customers choosing the high-quality agents can be served by low-quality agents only if they don't serve their own customers.

The above inequality implies that customers obtain strictly positive utility for large k. However, this contradicts with the definition of customer equilibrium since

$$
\lim_{k\to\infty} D_H(k) + D_{L_{dev}}(k) + D_L(k) < \alpha_H \tilde{D}_H^{pre} + D_L^{eq} \leq \alpha_H(1 - D_L^{eq}) + D_L^{eq} < 1,
$$

i.e. some customers don't request service for large k (The above inequality holds since $0 < \tilde{D}_H^{pre} \leq 1 - D_L^{eq}$). Hence, we should have that $\lim_{k\to\infty} D_H(k) = \tilde{D} \geq \alpha_H \tilde{D}_H^{pre} > 0$.

As a direct implication of this result, we have that the utilization of the deviating agents converges to 1 as k grows (as in the proof of Proposition 3), so that his revenue converges to the price he charges minus his operating cost, $p_L + \varepsilon - w_L$.

S.5.3. Proof of Lemma 10

In this proof, we only consider the case where $R_i - p_i^k \geq R_j - p_j^k$ for all k for some $i, j \in \{H, L\}$ with $i \neq j$. Furthermore, the below proof focuses on the deviation of a quality-j agent. The proof for the deviation of a quality-i agent is the same.

Similar to the proof of Lemma 9, we let $r_i = R_i - p_i$, $r_i^k = R_i - p_i^k$ for all $i = \{H, L\}$, and

$$\begin{aligned}
D_j(k) &= D_1^{MCE}(r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
D_i(k) &= D_2^{MCE}(r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
D_{j_{dev}}(k) &= D_3^{MCE}(r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
P_{jj_{dev}}(k) &= PServ_{13}(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
P_{ji}(k) &= PServ_{12}(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
P_{ij_{dev}}(k) &= PServ_{23}(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
\beta_j(k) &= \beta_1(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
\beta_i(k) &= \beta_2(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1)
\end{aligned}$$

for notational convenience.

S.5.3.1. $\mathbf{p_i < R_i}$: The proof of this case is very similar to the proof of Proposition 3. We mainly show that the rate of customers requesting service after the deviation of a single agent will be strictly positive in the limit, and this guarantees a 100% utilization for the deviating agents as it is rigorously shown in Lemma 5. Particularly, we show that

$$\liminf_{k\to\infty} D_j(k) + D_i(k) + D_{j_{dev}}(k) = \tilde{D} \geq \min\{\alpha_i/\rho, 1\}.$$

We prove this claim by contradiction. Thus, we suppose $\tilde{D} < \min\{\alpha_i/\rho, 1\}$, on the contrary. Then, there should exist a convergent subsequences of $\{D_j(k)\}_{k=1}^{\infty}$ and $\{D_i(k)\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} D_j(k) + D_i(k) + D_{j_{dev}}(k) = \tilde{D} < \min\{\alpha_i/\rho, 1\} < 1.$$

Using this observation, we have that

$$\begin{aligned}
&\lim_{k\to\infty} U_2(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
&\quad = \big(1 - \lim_{k\to\infty} P_{ij_{dev}}(k)\big)\big(\lim_{k\to\infty}(r_i^k + cm_a))\big(1 - \beta_i(k)\big) - cm_a\big) + \lim_{k\to\infty} P_{ij_{dev}}(k)(r_j + \varepsilon) \\
&\quad \geq \lim_{k\to\infty}(r_i^k + cm_a))\big(1 - \beta_i(k)\big) - cm_a \\
&\quad \geq \lim_{k\to\infty}(r_i^k + cm_a))\big(1 - \beta^M(\Lambda^k(D_j(k) + D_i(k)); \alpha_i k)\big) - cm_a = r_i > 0,
\end{aligned}$$

where the first inequality holds $r_j + \varepsilon > r_i^k$ for large k, the second inequality holds since some of the customers choosing quality-j can be served by quality-j agents and some of the customers choosing quality-i can be served by the deviating agent, and finally the equality holds since $\tilde{D} < \min\{\alpha_i/\rho, 1\} < 1$. This result implies that customers obtain strictly positive utility for large k. However, this contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_j(k) + D_i(k) + D_{j_{dev}}(k) < 1$, i.e. some customers don't request service for large k. Hence, we should have that $\liminf_{k\to\infty} D_j(k) + D_i(k) + D_{j_{dev}}(k) = \tilde{D} \geq \min\{\alpha_i/\rho, 1\}$.

As a direct implication of this result, we have that the utilization of the deviating agents converges to 1 as k grows (as in the proof of Proposition 3), so that his revenue converges to the price he charges minus his operating cost, $p_j - \varepsilon - w_j$.

S.5.3.2. $\mathbf{p_i = R_i}$: First note that we have that

$$\lim_{k\to\infty} D_j^{pre}(k) = \tilde{D}_j^{pre} > 0, \ \textit{and} \ \lim_{k\to\infty} P_{ji}^{pre}(k) = \tilde{P}_{ji}^{pre} > 0,$$

where $D_j^{pre}(k)$ is the rate of customers choosing quality-j agents and $P_{ji}^{pre}(k)$ is the probability with which a customer choosing quality-j before a single agent cuts his price (Otherwise quality-j agents couldn't have a strictly positive revenue in the limit).

Under this situation, we will show that the total rate of customers requesting service after the deviation is strictly positive. In particular, we show that

$$\lim_{k\to\infty} D_j(k) + D_i(k) + D_{j_{dev}}(k) = \tilde{D} \geq \alpha_i \tilde{D}_j^{pre}(1 - \tilde{P}_{ji}^{pre}) > 0.$$

We prove this claim by contradiction. Thus, we suppose $\tilde{D} < \alpha_i \tilde{D}_j^{pre}(1 - \tilde{P}_{ji}^{pre})$, on the contrary. Then, there should exist a convergent subsequences of $\{D_j(k)\}_{k=1}^{\infty}$ and $\{D_i(k)\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} D_j(k) + D_i(k) + D_{j_{dev}}(k) = \tilde{D} < \alpha_i \tilde{D}_j^{pre}(1 - \tilde{P}_{ji}^{pre}).$$

Using this observation, for large k, we have that

$$\begin{aligned}
&U_2(D_j(k), D_i(k), D_{j_{dev}}(k); r_j^k, r_i^k, r_j + \varepsilon; \alpha_j k - 1, \alpha_i k, 1) \\
&\quad = \big(1 - P_{ij_{dev}}(k)\big)\big((r_i^k + cm_a)\big)\big(1 - \beta_i(k)\big) - cm_a\big) + P_{ij_{dev}}(k)(r_j + \varepsilon) \\
&\quad \geq (r_i^k + cm_a)\big(1 - \beta^M(\Lambda^k(D_j(k) + D_i(k)); \alpha_i k)\big) - cm_a \\
&\quad = r_i^k - (r_i^k + cm_a)\beta^M(\Lambda^k(D_j(k) + D_i(k)); \alpha_i k) \\
&\quad > r_i^k - (r_i^k + cm_a)\beta^M(\Lambda^k D_j(k)^{pre}(1 - P_{ji}(k)^{pre}); k)\big(1 - P_{ji}(k)^{pre}\big) \\
&\quad = \big(1 - P_{ji}(k)^{pre}\big)\bigg((r_i^k + cm_a)\big(1 - \beta^M(\Lambda^k D_j(k)^{pre}(1 - P_{ji}(k)^{pre}); k)\big) - cm_a\bigg) + P_{ji}(k)^{pre} r_i^k \\
&\quad \geq \big(1 - P_{ji}(k)^{pre}\big)\bigg((r_j^k + cm_a)\big(1 - \beta_j^{pre}(k)\big) - cm_a\bigg) + P_{ji}(k)^{pre} r_i^k \geq 0,
\end{aligned}$$

where $\beta_j^{pre}(k)$ is the probability of abandonment for customers choosing quality-j agents in the equilibrium before deviation. The first inequality holds since not all of the customers choosing non-deviating agents are served by quality-i agents. The second inequality holds since $\tilde{D} < \alpha_i \tilde{D}_j^{pre}(1-\tilde{P}_{ji}^{pre})$ and by the fact that $\beta^M(\Lambda^k(D_j(k)+D_i(k));\alpha_i k) \simeq \left(\rho\tilde{D}/\alpha_i e^{-(1-\rho\tilde{D}/\alpha_i)}\right)^k$, $\beta^M(\Lambda^k D_j(k)^{pre}(1-P_{ji}(k)^{pre});k) \simeq \left(\rho\tilde{D}_j^{pre}(1-\tilde{P}_{ji}^{pre})e^{-(1-\rho\tilde{D}_j^{pre}(1-\tilde{P}_{ji}^{pre}))}\right)^k$, and $\frac{\left(\rho\tilde{D}/\alpha_i e^{-(1-\rho\tilde{D}/\alpha_i)}\right)^k}{\left(\rho\tilde{D}_j^{pre}(1-\tilde{P}_{ji}^{pre})e^{-(1-\rho\tilde{D}_j^{pre}(1-\tilde{P}_{ji}^{pre}))}\right)^k} \to 0$ as $k \to \infty$ as we discuss in the proof of Proposition 3. Finally, the third inequality holds since we $\beta^M(\Lambda^k D_j(k)^{pre}(1-P_{ji}(k)^{pre});k)$ is the probability of abandonment in a hypothetical system where the arrival rate is equal to the fraction of customers choosing quality-j and served by quality-j agents while the capacity is k instead of $\alpha_j k$. This hypothetical system always serves more customers (or, equivalently, causes less abandonment).

The above inequality implies that customers obtain strictly positive utility for large k. However, this contradicts with the definition of customer equilibrium since $\lim_{k\to\infty} D_j(k)+D_i(k)+D_{j_{dev}}(k) < 1$, i.e. some customers don't request service for large k. Hence, we should have that $\liminf_{k\to\infty} D_j(k)+D_i(k)+D_{j_{dev}}(k) = \tilde{D} \geq \alpha_i \tilde{D}_j^{pre}(1-\tilde{P}_{ji}^{pre})$.

As a direct implication of this result, we have that the utilization of the deviating agents converges to 1 as k grows (as in the proof of Proposition 3), so that his revenue converges to the price he charges minus his operating cost, $p_j - \varepsilon - w_j$.

S.5.4. Proof of Lemma 11

For notational convenience, let $\lambda_\varepsilon^\Delta(p;R) = \lambda^\Delta(p+\varepsilon;R+\varepsilon)$, and $\Delta_\varepsilon(p;R) = \Delta(p+\varepsilon;R+\varepsilon)$. Note that $\Delta_\varepsilon(p;R) = \Delta(p;R)$. Then, since $p \in \mathcal{P}(\rho;R)$, and $\lambda^\Delta(p;R) = \arg\max_{\lambda\geq 0} (R+cm_a)\lambda[1-\beta(\lambda)] - \lambda[\Delta(p;R)+cm_a]$, we have that

$$\begin{aligned}
p &> (R+cm_a)\lambda^\Delta(p;R)[1-\beta(\lambda^\Delta(p;R))] - \lambda^\Delta(p;R)(\Delta(p;R)+cm_a) \\
&\geq (R+cm_a)\lambda_\varepsilon^\Delta(p;R)[1-\beta(\lambda_\varepsilon^\Delta(p;R))] - \lambda_\varepsilon^\Delta(p;R)(\Delta_\varepsilon(p;R)+cm_a) \\
&\geq (R+\varepsilon+cm_a)\lambda_\varepsilon^\Delta(p;R)[1-\beta(\lambda_\varepsilon^\Delta(p;R))] - \lambda_\varepsilon^\Delta(p;R)(\Delta_\varepsilon(p;R)+cm_a) \\
&\quad -\varepsilon\lambda_\varepsilon^\Delta(p;R)[1-\beta(\lambda_\varepsilon^\Delta(p;R))] \\
&\geq (R+\varepsilon+cm_a)\lambda_\varepsilon^\Delta(p;R)[1-\beta(\lambda_\varepsilon^\Delta(p;R))] - \lambda_\varepsilon^\Delta(p;R)(\Delta_\varepsilon(p;R)+cm_a) - \varepsilon,
\end{aligned}$$

where the last inequality holds since $\lambda[1-\beta(\lambda)]$ is the utilization of a provider, and we have that $\lambda[1-\beta(\lambda)] \leq 1$ by definition. Using the above observation, we have that

$$p+\varepsilon > (R+\varepsilon+cm_a)\lambda_\varepsilon^\Delta(p;R)[1-\beta(\lambda_\varepsilon^\Delta(p;R))] - \lambda_\varepsilon^\Delta(p;R)(\Delta_\varepsilon(p;R)+cm_a),$$

which means $(p+\varepsilon) \in \mathcal{P}(\rho;R+\varepsilon)$.

Appendix S.6: Communication Enabled Model (Non-Identical Agents)

S.6.1. Proof of Lemma 12

We prove this claim for the setting $R_H - w_H > R_L - w_L$. For notational convenience we let $\tilde{\delta} = \lim_{k\to\infty} \delta^k$, $r_i^k = R_i - p_i^k$ for $i \in \{H, L\}$, $r' = R_H - p_H + \varepsilon$, and

$$\begin{aligned}
D_L(k) &= D_1^{MCE}(r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor)\\
D'_H(k) &= D_2^{MCE}(r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor)\\
D_H(k) &= D_3^{MCE}(r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor)\\
P_{LH'} &= PServ_{12}(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor)\\
P_{LH} &= PServ_{13}(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor)\\
P_{H'H} &= PServ_{23}(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor).
\end{aligned}$$

Before proving the above claim, we want to note that $\lim_{k\to\infty} D_H(k) > \frac{\alpha - \tilde{\delta}}{\rho}$, and only the customers choosing sub-pool-3 can be served in sub-pool-3, i.e. $\lim_{k\to\infty} P_{LH}(k) + P_{H'H}(k) = 0$. The proof of these claims are the same as the proofs of Lemma 9.1 and Lemma 9.2, respectively.

To prove the above claim, we first show that $\liminf_{k\to\infty} \frac{D'_H(k)}{\delta^k} \geq \frac{1}{\rho}$. In order to prove that by contradiction, we suppose $\liminf_{k\to\infty} \frac{D'_H(k)}{\delta^k} < \frac{1}{\rho}$. Then, there exists a convergent subsequence of $D'_H(k)$ such that $\lim_{k\to\infty} \frac{D'_H(k)}{\delta^k} < \frac{1}{\rho}$. As we argue in the proof of previous claims, the above assumption implies that the probability of abandonment in sub-pool-2 becomes negligible as k becomes large. (See Lemma 6.1) As a direct implication of this fact, we have that

$$\lim_{k\to\infty} U_2(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor) \geq r' > R_L - w_L > 0,$$

which implies that $\lim_{k\to\infty} D_L(k) + D'_H(k) + D_H(k) = 1$, i.e. all customers request service in a large marketplace, by the definition of the customer equilibrium. It also implies that $\lim_{k\to\infty} P_{LH'} D_L(k) = D_L(k)$, i.e. all of the customers choosing sub-pool-1 will be served by sub-pool-2, since the expected utility from sub-pool-2 exceeds $R_L - w_L$. (A similar proof can be seen in the proof of Lemma 9.1)

Furthermore, using the above observations, and considering the balance equation of sub-pool-2, we have that

$$\lim_{k\to\infty} D_L(k) + D'_H(k) = \lim_{k\to\infty} P_{LH'} D_L(k) + \lim_{k\to\infty} D'_H(k) \leq \frac{\delta}{\rho},$$

which implies that $\lim_{k\to\infty} D_H(k) \geq 1 - \tilde{\delta}/\rho$. However, this leads to the following contradiction:

$$\begin{aligned}
&\lim_{k\to\infty} U_3(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor)\\
&\quad = (r_H + cm_a)\frac{\alpha_H - \tilde{\delta}}{\rho - \tilde{\delta}} - cm_a < (r_H + cm_a)\frac{\alpha_H}{\rho} - cm_a
\end{aligned}$$

$$< r' \leq \lim_{k\to\infty} U_2(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor).$$

Hence, we should have that $\liminf_{k\to\infty} \frac{D'_H(k)}{\delta^k} \geq \frac{1}{\rho}$. Once we have that we can further show that $\lim_{k\to\infty} P_{LH'} D_L(k) = 0$ (as in Lemma 9.3), and the probability of abandonment in the sub-pool-2 converges to $1 - \frac{\delta}{\rho \tilde{D}'_H}$ when $\tilde{D}'_H = \liminf_{k\to\infty} D'_H(k)$ (as in Lemma 9.5)

Using these result, we have that

$$\begin{aligned}
\liminf_{k\to\infty} V'_H(k) \geq\ & (p_H + \varepsilon + w_H) \\
& \times \liminf_{k\to\infty} \frac{\Lambda^k D'_H(k)}{\lfloor \delta^k k \rfloor} \left[1 - \liminf_{k\to\infty} \beta_2(D_L(k), D'_H(k), D_H(k); r_L^k, r', r_H^k; \alpha_L k, \lfloor \delta^k k \rfloor, \alpha_H k - \lfloor \delta^k k \rfloor) \right] \\
=\ & p_H + \varepsilon + w_H.
\end{aligned}$$

Finally, the result holds since $V'_H(k) \leq p_H + \varepsilon + w_H$ by definition.

S.6.2. Proof of Lemma 13

In this proof, we only consider a convergent subsequence of (p_H^k, p_L^k) where $R_i - p_i^k \geq R_j - p_j^k$ for some $i, j \in \{H, L\}$ with $i \neq j$. Here, we show that a δ fraction of quality-i agents can improve their revenues by charging $p_i + \varepsilon$. The same proof holds for quality-j agents.

For notational convenience we let $\tilde{\delta} = \lim_{k\to\infty} \delta^k$, $r_i^k = R_i - p_i^k$ for $i \in \{H, L\}$, $r' = R_i - p_i - \varepsilon$, and

$$\begin{aligned}
D'_i(k) &= D_1^{MCE}(r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\
D_j(k) &= D_2^{MCE}(r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\
D_i(k) &= D_3^{MCE}(r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\
P_{i'j} &= PServ_{12}(D'_i(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\
P_{i'i} &= PServ_{13}(D'_i(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\
P_{ji} &= PServ_{23}(D'_i(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor).
\end{aligned}$$

Before proving the above claim, we want to note that $\lim_{k\to\infty} D_i(k) > \frac{\alpha - \tilde{\delta}}{\rho}$, and only the customers choosing sub-pool-3 can be served in sub-pool-3, i.e. $\lim_{k\to\infty} P_{i'i}(k) + P_{ji}(k) = 0$. The proof of these claims are the same as the proofs of Lemma 9.1 and Lemma 9.2, respectively.

To prove our claim, we first show that $\liminf_{k\to\infty} \frac{D'_i(k)}{\delta^k} \geq \frac{1}{\rho}$. In order to prove that by contradiction, we suppose $\liminf_{k\to\infty} \frac{D'_i(k)}{\delta^k} < \frac{1}{\rho}$. Then, there exists a convergent subsequence of $D'_i(k)$ such that $\lim_{k\to\infty} \frac{D'_i(k)}{\delta^k} < \frac{1}{\rho}$. As a direct implication of this, we have that

$$\lim_{k\to\infty} U_1(D'_i(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) = r' > 0,$$

which implies that $\lim_{k\to\infty} D'_i(k) + D_j(k) + D_i(k) = 1$, i.e. all customers request service in a large marketplace, by the definition of the customer equilibrium. Furthermore, this implies that $\lim_{k\to\infty} D_j(k) +$

$D_i(k) = 1 - \tilde{\delta}/\rho$ since $\lim_{k\to\infty} \frac{D_i'(k)}{\delta^k} < \frac{1}{\rho}$. Then, we should have that either $\lim_{k\to\infty} D_j(k) \geq \alpha_j$ or $\lim_{k\to\infty} D_i(k) \geq \alpha_i - \tilde{\delta}$ (or both) holds because otherwise we would have that $\lim_{k\to\infty} D_j(k) + D_i(k) = 1 - \tilde{\delta} < 1 - \tilde{\delta}/\rho$. Combining this with the observation that sub-pool-3 serves only its customers, we have either

$$\begin{aligned}&\lim_{k\to\infty} U_3(D_i'(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\ &\quad = (r_i + cm_a)\frac{\alpha_i - \tilde{\delta}}{\rho \lim_{k\to\infty} D_i(k)} - cm_a \leq (r_i + cm_a)\frac{1}{\rho} - cm_a < r'.\end{aligned}$$

or

$$\begin{aligned}&\lim_{k\to\infty} U_2(D_i'(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor) \\ &\quad \leq (r_j + cm_a)\frac{\alpha_j}{\rho \lim_{k\to\infty} D_j(k)} - cm_a \leq (r_i + cm_a)\frac{1}{\rho} - cm_a < r'.\end{aligned}$$

Both of these results contradicts with the definition of the Customer Equilibrium. Hence, we should have that $\liminf_{k\to\infty} D_i'(k) \geq \frac{\delta}{\rho}$. Finally, as in the above lemmas, we have that

$$\begin{aligned}\liminf_{k\to\infty} V_i'(k) \;&\geq\; (p_i + \varepsilon - w_i) \liminf_{k\to\infty} \frac{\Lambda^k D_i'(k)}{\lfloor \delta^k k \rfloor} \\ &\qquad \times \left[1 - \liminf_{k\to\infty} \beta_2(D_i'(k), D_j(k), D_i(k); r', r_j^k, r_i^k; \lfloor \delta^k k \rfloor, \alpha_j k, \alpha_i k - \lfloor \delta^k k \rfloor)\right] \\ &= p_i + \varepsilon - w_i.\end{aligned}$$

Finally, the result holds since $V_i'(k) \leq p_i + \varepsilon - w_i$ by definition.