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

Eren B. Çil
Kellogg School of Management, Northwestern University, Evanston, Illinois, e-cil@kellogg.northwestern.edu

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

Recent times have witnessed the emergence of large-scale, web-based service marketplaces where many small service providers compete among themselves in catering to customers with diverse needs. 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, operational efficiency may be detrimental to the overall efficiency of the marketplace. Further, we show 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

Recent times have witnessed the emergence of large-scale, web-based service marketplaces where many small service providers (agents) compete among themselves in catering to customers with diverse needs. 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 involvements 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 the 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. 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 250,000 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 customers to choose among multiple agents after naming their price and describing their project. 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 price of the project at service completion. In both marketplaces, the moderating firms allow the customers to browse among tens of thousands of agents and communicate with different providers to make the service transaction "one-click-away."

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 and analytics firm in Saratoga, California, recently 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 the above, 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).

---

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

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 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 (with the right price in mind) 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 name their price is an example of such a mechanism, as well as a system in which the moderating firm provides real-time congestion information. (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 in such a case, Nash best-response is certainly a requirement for self-enforceability but is not, in general, sufficient. Considerations of this sort have motivated the notion of strong Nash equilibrium, see Aumann (1959), which requires stability against deviations by every conceivable coalition. Following these ideas, 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 no-intervention, we show that there exists a unique symmetric equilibrium. This result extends the existing results on markets with a small number of players. However, in a marketplace with operational efficiency, we first show that any price but the operating cost of agents fails to be sustained as a symmetric equilibrium when supply exceeds demand. Furthermore, 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 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. We also show that the insights about the equilibrium outcome would be similar if we assumed that the moderating firm achieve the desired operational efficiency by providing customers with real-time congestion information.

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. For the specific case where the operational efficiency is achieved via providing real-time congestion information, we show that the revenue loss (from the best achievable) is bounded. Further, this gap depends on the level of efficiency the mechanism delivers.

4. We also extend our model to study a setting in which agents are heterogeneous in terms of the quality of their service and their operating costs. In particular, we consider a model with two types of agents distinguished by the value they generate for the customers. As in our original model, we again study the impact of the moderating firm on the market outcome. Our results show that an operational tool, which reduces the mismatch between demand and supply, may lower the profit of a moderating firm unless it is complemented with the ability to communicate, and demand exceeds capacity. Moreover, a model with heterogeneous agents allows us to discuss the impact of the market composition on the equilibrium outcomes as well as the role of the moderating firm.

5. 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. The new solution concept, while inspired by the existing game theoretic concept of Strong Nash, allows agents to communicate only with a limited number of peers. Moreover, the concept is more suited for large games where single agent deviations are practically negligible. We refer to this new equilibrium concept as (δ, ϵ)-Market Equilibrium.

The rest of the paper is organized as follows: In §2, we present a literature review on the existing work related to our paper. §3 introduces a basic model of the marketplace under consideration. We introduce and analyze the no-intervention, the operational efficiency, and the communication enabled models in §4, §5, and §6, respectively. In §7, we study the impact of heterogeneity among agents on the market. Finally, §8 concludes the paper. All proofs are relegated to the appendix.

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 pricing problem of a monopoly facing price- and delay-sensitive customers in various settings (See De Vany (1976), Mendelson and Whang (1990), Afeche and Mendelson (2004)). Another body of research that is motivated by Naor (1969) considers the competition among service providers who make pricing and/or service capacity decisions. Luski (1976) and Levhari and Luski (1978) focus on the competition between two firms under markovian assumptions. Natural extensions of the competition models assume general service time distributions, observable queue lengths, many firms, and multiple customer classes (See Loch (1991), Li and Lee (1994), and Lederer and Li (1997)). We refer the reader to Hassin and Haviv (2003) for an extensive summary of the early attempts to model price and service competition. More recently, Cachon and Harker (2002) studies the competition between two firms offering substitute but differentiated services. The paper establishes sufficient conditions for the existence of the equilibrium, and characterizes the equilibrium behavior of the competing firms. Furthermore, it shows that firms can reattain their foregone profits due to the competition by outsourcing their production, even if outsourcing does not bring any cost benefits. In another differentiated services setting, Allon and Federgruen (2007) considers the price and waiting time as completely independent firm attributes by employing a general demand model rather than a full-price model as in the previous papers. In this paper, authors study the competition between single-server queueing systems (under markovian assumptions) in which the price and service level are determined sequentially or simultaneously. In Allon and Federgruen (2008), they relax the markovian assumptions. Most of the above papers model 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. Maglaras and Zeevi (2003), Armony and Maglaras (2004), and Maglaras and Zeevi (2005) are examples of recent work using the asymptotic analysis to tackle complexity of these problems.

An important phenomena in service systems is that customers abandon while waiting for their service to commence. This behavior of customers may have a non-trivial impact on the system under consideration. Furthermore, ignoring the possibility of abandonment may cause misleading operational conclusions such as overstaffing in call-centers. To address these issues, Garnett et al. (2002) extends the asymptotic analysis of multi-server queueing system by considering a markovian system (exponentially distributed arrival, service, and abandonment times). Zeltyn and Mandelbaum (2005) provides approximations for the systems with generally distributed abandonment times. Furthermore, Whitt (2006) introduces fluid approximations for these systems relaxing by markovian assumptions.

Most of the papers that use approximation methods aim to provide prescriptions for a single firm maximizing its profits. However, 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 (2008) is the first paper studying competition among complex queueing systems by using asymptotic analysis to approximate the queueing dynamics. They not only characterize the equilibrium in a market consisting of large-scale service providers but also introduce a framework to analyze such games by combining game theoretic foundations and asymptotic analysis of queueing systems. 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. They show that the first best solution can be induced in the market via a compensation scheme which rewards idle servers. 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 phenomena 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.

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

---

[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.

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 and, as we shall discuss, the level of demand-supply ratio 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 $\rho > 1$.

The reader will notice that the above description of the model, though specifying demand (customer arrivals) and capacity (agents), lacks the characterization of their interaction. As mentioned before, their interaction would depend on the "rules" the moderating firm puts in place when creating the marketplace. We shall provide a detailed description of these rules in the following sections. In Section 4, we study the market in which the moderating firm does not intervene in the marketplace at all and only plays the role of information provider. In Section 5, the moderating firm modifies the interaction between customers and agents to allow for efficient operational matching. Apart from this efficient matching, the moderating firm provides infrastructure for the agents to communicate with each other in Section 6.

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 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 commences. 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.

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 $U(D_n, p_n) > 0$ for some $n \leq k$, then $\sum_{n=1}^{k} D_n = 1$.*

---

[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.

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* $n \le k$.
2. *For any* $\ell \le k$, *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 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. (See the proposition at the end of 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:

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

In this problem, the objective function is the utility 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 $n \leq k$. Solving the above problem for any given p, we characterize the symmetric SPNE as in the following theorem:

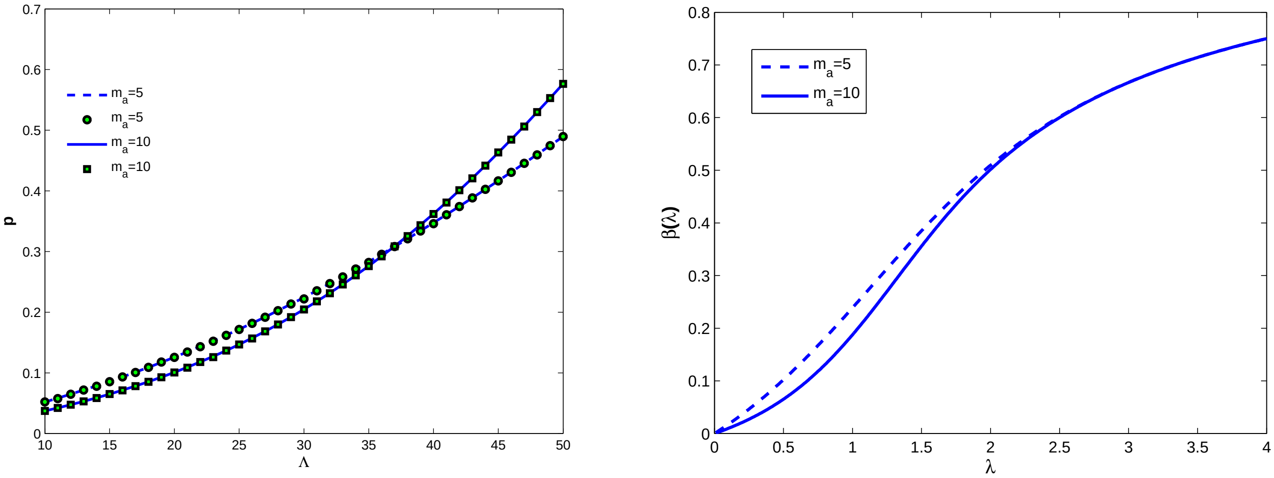
Theorem 1. *If $\beta(\lambda)$ is concave, then there exists a symmetric SPNE. Furthermore, the symmetric SPNE is characterized as follows:*

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 monopolies 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, allowing customers to earn strictly positive utility in the equilibrium.

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. 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 each customer obtains a reward of $R = 1$, incurs a cost of $c = 0.05$, and the number of agents k is 50. Then, we study two scenarios that differ in the average abandonment time m_a. We assume m_a is either 5 and 10. For each of these scenarios, we check whether p^*, 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. Considering an instance, we show that the best-response of a single agent is p^* when the remaining 49 agents charge p^*. We present the results of our study in the following figure. We also illustrate the functional form of the abandonment function, $\beta(\lambda)$, for $m_a = \{5, 10\}$.



(a) (b)

Figure 1 **(a) Best-response of a single agent (the markers) as a function of the arrival rate Λ when all the remaining agents charge p^* characterized in Theorem 1 (the curves). For all examples $k = 50$, $R = 1$, $c = 0.05$. In (a) $w_H = 0.1$, in (b) $w_H = 0.3$. (b) The abandonment function, $\beta(\lambda)$.**

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 price on oDesk.com, agents that are willing to serve a customer for that price apply to the customer's posting, and among these 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. There are other mechanisms, such as providing real-time congestion information, that may be used to achieve operational efficiency. We shall discuss the implications of providing real-time congestion information in Section 5.2.

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 in front of the sub-pool she chooses 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:

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

where $\beta_\ell(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N)$ denotes the probability of abandonment in the sub-pool-ℓ, and $PServ_{\ell m}(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N)$ 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 $n = 1,\ldots,N$. 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 serve by a sub-pool charging more than p_ℓ. Similarly, the revenue of an agent in the sub-pool-ℓ is:

$$V_\ell(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N) = p_\ell\sigma_\ell(D_1,\ldots,D_N;p_1,\ldots,p_N;y_1,\ldots,y_N),$$

where $\sigma_\ell(\ldots;\ldots)$ is utilization of agents in sub-pool-ℓ when ΛD_n is the rate of customer arrival to the sub-pool-n for $n = 1,\ldots,N$. 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-know 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 (2008), 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 a weaker equilibrium concept, which is an approximation of Nash Equilibrium. Particularly, this equilibrium concept 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* $n = 1, \ldots, N$. *Then,* $(D_n^k, p_n^k, y_n^k)_{n=1}^N$ *is a 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* $\ell \leq N$ *and* $m \leq N$, *we have that* $V_\ell(D_1^k, \ldots, D_N^k; p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k) \geq V_\ell(D_1^{k\prime}, \ldots, D_N^{k\prime}; p_1^k, \ldots, p_N^k; y_1^{k\prime}, \ldots, y_N^{K\prime}) - \epsilon^k$, *where*

$$y_n^{k\prime} = \begin{cases} y_n^k & \textit{if } n \leq N,\ n \neq \ell,\ n \neq m \\ y_n^k - 1 & \textit{if } n = \ell \\ y_n^k + 1 & \textit{if } n = m, \end{cases}$$

and $D_n^{k\prime} = D_n^{MCE}(p_1^k, \ldots, p_N^k; y_1^{k\prime}, \ldots, y_N^{k\prime})$ *for all* $n \leq N$.

3. *For any* $\ell \leq N$ *and* $p' \neq p_n^k$ *for all* $n = 1, \ldots, N$, *we have that* $V_\ell(D_1^k, \ldots, D_N^k; p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k) \geq V_{N+1}(D_1^{k\prime}, \ldots, D_{N+1}^{k}{}'; p_1^k, \ldots, p_N^k, p'; y_1^{k\prime}, \ldots, y_{N+1}^{k}{}') - \epsilon^k$ *where*

$$y_n^{k\prime} = \begin{cases} y_n^k & \textit{if } n \leq N,\ n \neq \ell \\ y_n^k - 1 & \textit{if } n = \ell \\ 1 & \textit{if } n = N+1, \end{cases}$$

and $D_n^{k\prime} = D_n^{MCE}(p_1^k, \dots, p_N^k, p'; y_1^{k\prime}, \dots, y_{N+1}^{k\,\prime})$ *for all* $n \le N+1$.

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 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 construct 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 $\Lambda D_1^{MCE}(p;k)$, service rate 1, and abandonment rate $1/m_a$, where $D_1^{MCE}(p;k)$ is the Market Customer Equilibrium when all k agents charge p. Therefore, the revenue of an agent in this case is given by

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

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 a 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 in the k^{th} marketplace. Recall that if an agent chooses $p' \neq p$, this amounts to creating his own sub-pool, and his revenue is given by

$$V_2(D_1^{MCE}(p', p; 1, k-1), D_2^{MCE}(p', p; 1, k-1); p', p; 1, k-1),$$

where $(D_n^{MCE}(p', p; 1, k-1))_{n=1}^2$ is the Market Customer Equilibrium given that $k-1$ agents charge p and one agent charges p'. We then say that a price p emerges as the symmetric ϵ^k-Market Equilibrium if

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

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

To understand the behavior of the market outcome in large markets, we shall first study the left-hand side and then the right-hand side of (3) along the trajectory of marketplaces. For the former, we only need to consider the case $p < R$ because none of the customers joins the system when all agents charge R, i.e. $D_1^{MCE}(R;k) = 0$. This cannot constitute a market equilibrium since any agent can earn positive revenue, which is bounded away from zero, by decreasing his price. While studying the left hand side, we distinguish between a buyer's and a seller's market. In a buyer's market where $\rho \leq 1$, customers experience negligible waiting times in a marketplace with a large number of agents, even if all customers request service. Thus, in a buyer's market, all customers join the system in equilibrium since they obtain approximately the utility of $R - p$ by joining. Furthermore, the revenue of each agent is approximated by $p\rho$. In a seller's market where $\rho > 1$, it may still be true that all of the customers request service. In such a case, customers obtain a strictly positive utility despite incurring a waiting cost. However, once the aggregate demand is sufficiently high, some of the customers leave the market immediately due to the high congestion level. Regardless of how customers behave in the equilibrium, the rate of customers requesting service in a seller's market should, in equilibrium, be higher than the processing capacity. Otherwise, a customer joining the system would earn strictly positive utility while the customers who do not request service would obtain zero utility. Therefore, agents are always "over-utilized" in a seller's market and the revenue of each agent is approximately p. The following proposition presents these results formally.

PROPOSITION 2. *Let $D_1^{MCE}(p;k)$ be the Market Customer Equilibrium when all agents charge p in the k^{th} marketplace. Then, for any $p < R$, we have that*

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

Furthermore, $\lim_{k\to\infty} V_1(D_1^{MCE}(p;k);p;k) = \begin{cases} p\rho & \text{if } \rho \leq 1 \\ p & \text{if } \rho > 1 \end{cases}.$

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 in a buyer's market, we next show that a single agent can improve his utilization significantly when he decreases his price. Hence, he can approximately secure a revenue of p following a small price cut. Such a cut will allow a single agent to serve not only his own customers but also the customers choosing the price p. The following proposition proves this observation formally.

PROPOSITION 3. *Let* $V'(p;k) = \max_{0\leq p'\leq R} V_2(D_1^{MCE}(p',p;1,k-1), D_2^{MCE}(p',p;1,k-1);p',p;1,k-1)$ *for any given $p > 0$. Then, in a buyer's market ($\rho \geq 1$), we have that*

$$\liminf_{k\to\infty} V_k'(p) \geq p.$$

As we established in Proposition 2, the revenue of an agent when all agents charge the same price p can be bounded from above by $p\rho$ in large marketplaces. Then, Proposition 3 implies that any p satisfying $p(1-\rho) > \epsilon^k$ cannot emerge as the equilibrium price of a symmetric ϵ^k-Market Equilibrium. Thus, as $\lim_{k\to\infty} \epsilon^k = 0$, we obtain that all prices except $p = 0$ cannot be sustained as the equilibrium price of a symmetric ϵ^k-Market Equilibrium in large marketplaces. Furthermore, we can show $p = 0$ can emerge as the approximate equilibrium price in large marketplaces. We formalize these observations in the following theorem.

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

1. $p > 0$ *cannot emerge as the equilibrium price of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace, for large* k.
2. $p = 0$ *is an equilibrium price of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace, for large* 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 as well as the revenue of the agents.

In the above theorem, we only rule out the possibility that all agents charge a strictly positive price in a buyer's market. Section 7.2 will show that even an asymmetric equilibrium where all agents charge strictly positive prices cannot arise. (See the proposition at the end of Section 7.2)

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 while they are charging the same price p 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 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* $V'(p;k) = \max_{p \le p' \le R} V_2(D_1^{MCE}(p,p';k-1,1), D_2^{MCE}(p,p';k-1,1); p', p; k-1, 1)$ *for a given* $p > R$. *Then, in a seller's market* ($\rho > 1$), *we have that*

$$\limsup_{k\to\infty} V'(p;k) \le (R + cm_a)\lambda^{\Delta}(p;R)[1 - \beta(\lambda^{\Delta}(p;R))] - \lambda^{\Delta}(p;R)(\Delta(p;R) + cm_a),$$

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* $1-\beta(\lambda) - \lambda\beta'(\lambda) = \frac{\Delta(p;R)+cm_a}{R+cm_a}$.

The agents, who do not change their prices, remain “over-utilized” when a single provider increases his price as long as the number of agents is large in a seller’s market. Therefore, the agents in the sub-pool charging the lower price will always be busy, and none of the customers requesting service from the deviating agent will have a chance to get their service for a lower price. In other words, the deviating agent has to serve all of the customers requesting service from him. Since a single agent can only serve a negligible amount of customers relative to the aggregate demand, deviation of a single agent does not have a significant impact on the demand for the remaining agents charging p. Hence, the demand for the sub-pool consisting of $k-1$ agents is almost the same as their “original” demand before deviation. Note that $\Delta(p;R)$, as defined in Proposition 4, denotes the utility that the customers obtains in the Market Customer Equilibrium when all agents charge p. 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 emerges as the symmetric ϵ^k-Market Equilibrium outcome if p 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,* p *emerges as the equilibrium price of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace, for large* 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 $\mu \neq m_a$ in Appendix B using a numerical study. 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.*

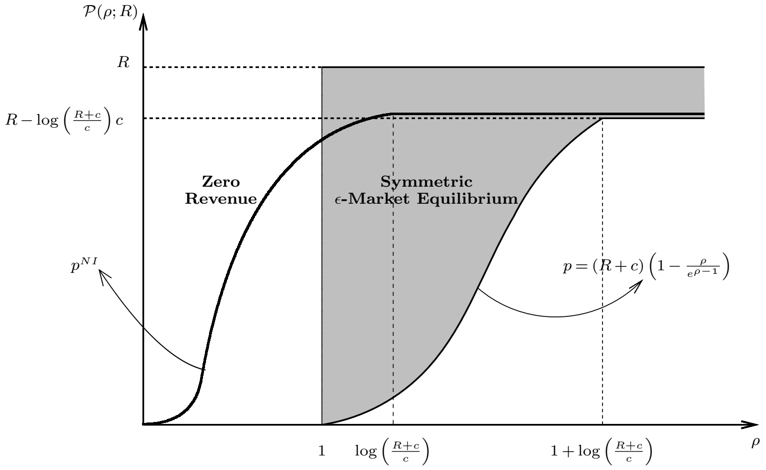


Figure 2 **The prices that form a symmetric market equilibrium as a function of the demand-supply mismatch (ρ). The service rates and abandonment rates are assumed to be one.**

Figure 2 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. 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.

When comparing the equilibrium prices in a market with operational efficiency with the prices in a market with no intervention, 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.

5.2. Other Means of Achieving Operational Efficiency

Up to this point we have studied the impact of a particular efficient matching mechanism which aims at reducing the mismatch between customers and service providers. This mechanism achieves the desired level of operational efficiency by ensuring that customers will not wait for their service as long as there is an available agent willing to serve them for a price that they want to pay. We model this mechanism by assuming that all agents charging the same price are virtually grouped together, and an arriving customer is served by the available agent charging the lowest price. Therefore, compared to the No-Intervention setting, customers may pay less than their preferred price, and they experience lower waiting times as a result of the resource pooling.

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. Similar to Theorem 2, one can show that any price but zero fails to emerge as the equilibrium outcome in this setting. In other words, the impact of providing real-time congestion data in a buyer's market would be the same as the aforementioned matching mechanism.

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 $c = 0.03$ per unit time, and abandon the system with rate of $1/m_a = 1$. We fix the number of agents k to be 25, then we study three scenarios that differ in the arrival rate Λ. We assume the arrival rate Λ is either 40, 50, or 60. For each of these three scenarios, we check whether a single agent has an incentive to increase his price to $p' \in \{0.01, 0.02, 0.03, \dots, 0.99\} \setminus p$ when all the remaining $k-1$ agents still charge a price $p \in \{0.1, 0.2, \dots, 0.9\}$. 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. 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.

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

✓: p is an equilibrium price,

×: p is not an equilibrium price.

Table 1 **Prices that can emerge as the equilibrium outcome when the moderating firm provides real-time congestion information.**

In Section 6.3, we will further analyze the model in which the 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 in the k^{th} marketplace. 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* $V_\ell(D_1^k, \ldots, D_N^k; p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k) \geq V_\ell(D_1^{k'}, \ldots, D_N^{k\,'}; p_1^k, \ldots, p_N^k; y_1^{k'}, \ldots, y_N^{K'}) - \epsilon^k$ *where*

$$y_n^{k'} = \begin{cases} y_n^k & \text{if } n \leq N,\ n \neq \ell,\ n \neq m \\ y_n^k - d & \text{if } n = \ell \\ y_n^k + d & \text{if } n = m, \end{cases}$$

and $D_n^{k'} = D_n^{MCE}(p_1^k, \ldots, p_N^k; y_1^{k'}, \ldots, y_N^{k\,'})$ *for all* $n \leq N$*.*

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* $V_\ell(D_1^k, \ldots, D_N^k; p_1^k, \ldots, p_N^k; y_1^k, \ldots, y_N^k) \geq V_{N+1}(D_1^{k\prime}, \ldots, D_{N+1}^{k}{}'; p_1^k, \ldots, p_N^k, p'; y_1^{k\prime}, \ldots, y_{N+1}^{k}{}') - \epsilon^k$ *where*

$$y_n^{k\prime} = \begin{cases} y_n^k & \text{if } n \leq N,\ n \neq \ell \\ y_n^k - d & \text{if } n = \ell \\ d & \text{if } n = N+1, \end{cases}$$

and $D_n^{k\prime} = D_n^{MCE}(p_1^k, \ldots, p_N^k, p'; y_1^{k\prime}, \ldots, y_{N+1}^{k}{}')$ *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. While their result may seem against our expectations, we will show that such a phenomena does not exist in our model.[6]

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 is the same as in (2), and thus Proposition 2 establishes its asymptotic behavior.

In a buyer's market with $\rho < 1$, we already showed that any price but zero fails to emerge as a symmetric ϵ-Market Equilibrium in large marketplaces. Since (δ, ϵ)-Market Equilibrium is a refinement of the Market Equilibrium, any strictly positive price cannot be a symmetric (δ, ϵ)-Market Equilibrium in large marketplaces. As a direct implication of that, any sequence of prices that emerge as symmetric (δ, ϵ)-Market Equilibrium converge 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 *be a price emerging as a symmetric* (δ^k, ϵ^k)-*Market Equilibrium in the* k^{th} *marketplace. If* $\rho < 1$, *then we have that*

$$\lim_{k \to \infty} p_k = 0, \text{ and } \lim_{k \to \infty} D_1^{MCE}(p_k, 1) = 1.$$

---

[6] 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 three times 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.

Furthermore, when $\lim_{k\to\infty} \delta^k = 0$*,* $p = 0$ *is an equilibrium price of a symmetric* (δ^k, ϵ^k)*-Market Equilibrium in the* k^{th} *marketplace, for large 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 do 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 increase 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 overcomed 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 *be a price emerging as a symmetric* (δ^k, ϵ^k)*-Market Equilibrium in the* k^{th} *marketplace. If* $\rho > 1$*, then we have that*

$$\lim_{k\to\infty} p_k = R, \text{ and } \lim_{k\to\infty} D_1^{MCE}(p_k, 1) = 1/\rho.$$

Furthermore, there exists a sequence p_k *that forms a symmetric* (δ^k, ϵ^k)*-Market Equilibrium in the* k^{th} *marketplace, for large 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 (See the proposition at the end of Section 7.3 for a more detailed discussion.).

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. $V_1(D_1^{MCE}(p;k);p;k)$, 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 $V_1(D_1^{MCE}(p;k);p;k)+\epsilon\geq V_2(D_1^{MCE}(p',p;\delta k,k-\delta k),D_2^{MCE}(p',p;\delta k,k-\delta k);p',p;\delta k,k-\delta k)$ 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 3, 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.)

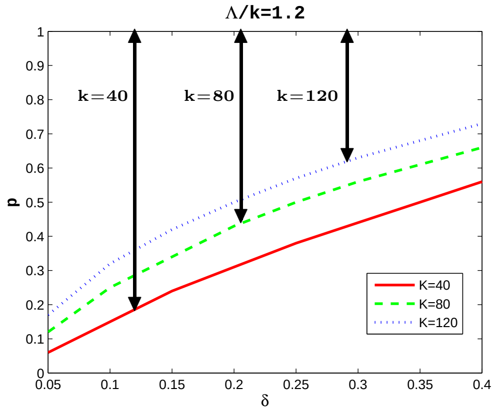


Figure 3 **The set of prices that form a symmetric (δ,ϵ)-Market Equilibrium in the marketplace with k agents is the area above (below) the corresponding curve when $\Lambda/k=1.2$.**

6.3. Other Means of Achieving Operational Efficiency

In the previous subsection, 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 uses other means of operational efficiency. As in Section 5.2, we consider the setting where the moderating firm provides real-time queue information in order to reduce the mismatch between customers and agents. Note that we already discuss that no positive price 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 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[7]. 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, which depends on R and ζ that 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.

---

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

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 arrival rate λ, with service rate 1, abandonment rate $1/m_a$. Suppose*

$$\lim_{k\to\infty}\frac{\beta^{JSQ}(\rho k;k)}{\beta^{M}(\rho k;k)}<\infty,\ \ \textit{and}\ \lim_{k\to\infty}\frac{\sigma^{JSQ}(\rho k;k)}{\rho(1-\beta^{M}(\rho k;k))}>1-\zeta.$$

Then, let p_k^{info} be a price emerging in a symmetric (δ,ϵ^k)-Market Equilibrium in the k^{th} marketplace when the firm provides real-time congestion information. If $\rho>1$ and $\delta=1$, then we have that

$$\liminf_{k\to\infty} p_k^{info}\geq R(1-\zeta).$$

7. A Marketplace with Non-Identical Agents

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. We next explore the robustness of the conclusions of the previous sections to the introduction of 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 there are k_H agents providing a high-quality service and k_L agents providing a low-quality service while there are still k agents in total. We denote the fraction of high-quality and low-quality agents by α_H, α_L, respectively. Furthermore, we assume that customers value the service with respect to its quality. Particularly, customers earn a reward of R_H when they are served by a high-quality agent whereas they earn a reward of $R_L\leq R_H$ when a low-quality agent completes their service. We also distinguish the agents according to their operating costs. We assume the operating cost of high-quality agents is w_H, and the operating cost of low-quality agents is $w_L\leq w_H$. 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 $i=\{H,L\}$. As in Section 3, the service rate is 1, arrival rate is Λ, abandonment rate is $1/m_a$, and waiting cost is c.

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 as the difference between these two. 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 the same as before, $V^N(\lambda, p) = p\lambda[1 - \beta(\lambda)]$.

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 $(p_1, \ldots, p_{k_H})$ are the prices announced by the high-quality agents, $(p_{k_H+1}, \ldots, p_k)$ 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 $n \le k_H$ 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:

$$\begin{aligned} p_H \in \arg\max_{p_\ell \ge w_H} \ & (p_\ell - w_H)\Lambda D_\ell^{CE}(r_1, \ldots, R_H - p_\ell, \ldots, r_{k_H}, r_{k_H+1}, \ldots, r_k) \\ & \times \left[1 - \beta(\Lambda D_\ell^{CE}(r_1, \ldots, R_H - p_\ell, \ldots, r_{k_H}, r_{k_H+1}, \ldots, r_k))\right], \qquad (4) \\ p_L \in \arg\max_{p_\ell \ge w_H} \ & (p_\ell - w_H)\Lambda D_\ell^{CE}(r_1, \ldots, r_{k_H}, r_{k_H+1}, \ldots, R_L - p_\ell, \ldots, r_k) \\ & \times \left[1 - \beta(\Lambda D_\ell^{CE}(r_1, \ldots, r_{k_H}, r_{k_H+1}, \ldots, R_L - p_\ell, \ldots, r_k))\right], \qquad (5) \end{aligned}$$

where for any $n \ne \ell$, $r_n = R_H - p_H$ if $n \le k_H$ 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* $(\tilde{D}_L, \tilde{D}_H) \in \{(x,y) : \hat{U}_L(x,y) \leq 0, \hat{U}_H(x,y) \leq 0,\ k_Lx + k_Hy = \Lambda\}$.
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* $(R_H + cm_a - w_H)(k_H - 1) - \frac{R_L + cm_a}{1-\beta(\lambda)}\left(\frac{k_H}{1-\nu(\lambda)} - 1\right) = 0$,

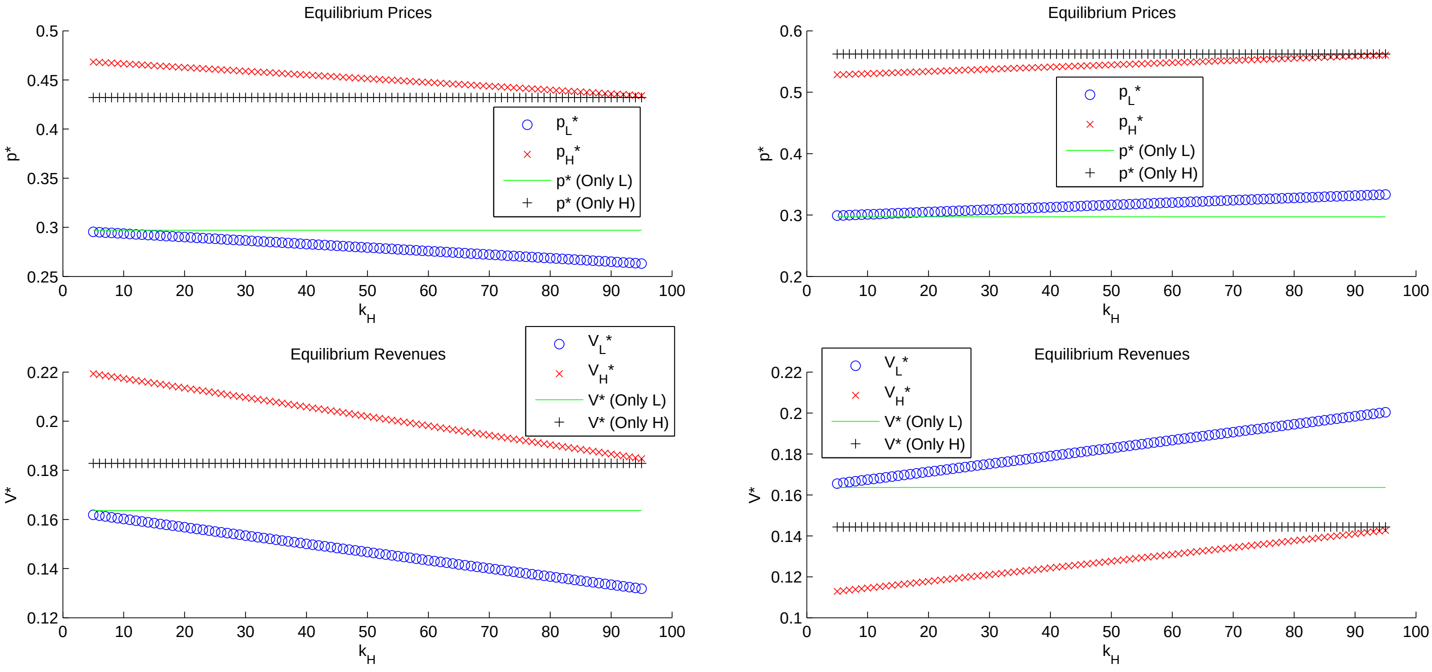
$$\begin{aligned}
(\hat{D}_L(\Lambda), \hat{D}_H(\Lambda)) &= \{(x,y) : \hat{U}_L(x,y) = \hat{U}_H(x,y),\ k_Lx + k_Hy = \Lambda\} \\
\hat{U}_L(x,y) &= (R_L + cm_a)\left[\frac{1 - \nu(x)}{1 + \frac{\nu(x)}{k_L + k_H\vartheta(x,y) - 1}}\right](1 - \beta(x)) - cm_a, \\
\hat{U}_H(x,y) &= (R_H + cm_a - w_H)\left[\frac{1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

$\vartheta(x,y) = \frac{y\nu(x)}{x\nu(y)}$, *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 a local monopolist 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 4 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 4 **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. Therefore, as in Section 5, 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. 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 $N \leq k$ 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_{in} for $i \in \{H, L\}$. Hence, (r_n, y_{Hn}, y_{Ln}) 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

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

where $y_n = y_{Hn} + y_{Ln}$, $\beta_\ell(D_1, \ldots, D_N; r_1, \ldots, r_N; y_1, \ldots, y_N)$ denotes the probability of abandonment in the sub-pool-ℓ, and $PServ_{\ell m}(D_1, \ldots, D_N; r_1, \ldots, r_N; y_1, \ldots, y_N)$ 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 $n = 1, \ldots, N$. 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 an agent in sub-pool-ℓ is the same as the revenue function described in Section 5.

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. The latter ensures that the ratio of high and low-quality agents constant along the sequence of marketplaces.

7.2.1. Characterization of the Market Equilibrium: Here we characterize the symmetric equilibrium in which all the high-quality agents charge the same price p_H and all the low-quality agents charge p_L. As a first step, we derive the revenue of agents when the agents with the same quality charge the same price.

PROPOSITION 7. *Let* $V_H^{MCE}(p_H,p_L;k)$ *and* $V_L^{MCE}(p_H,p_L;k)$ *be the revenue of an high-quality and a low-quality agent, respectively, when all the high-quality agents charge* p_H *and all the low-quality agents charge* p_L *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*

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

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

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 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 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 $R_j - p_j > R_i - p_i$ 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 positive. 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, all the agents are virtually grouped together, and we go back to the model with identical agents. As we show in Section 5, there is always room for a single agent to improve his revenue in a buyer’s market. Hence, we can only have an equilibrium in a buyer’s market where only one group of agents (high or low) can earn positive revenue. In fact, Theorem 7 below establishes that high-quality agents can earn positive revenue when demand exceeds the capacity of high-quality agents.

The above proposition also states that both of the groups can be “over-utilized” when they offer different level of 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, this model reduces to one with identical agents when all agents offer the same level of net reward in a seller's market, and thus there will be multiple symmetric equilibria in a seller's market as established in Theorem 3. We summarize these observations in the following result:

THEOREM 7. *Suppose* $R_H - w_H > R_L$.

1. *If* $\rho < \alpha_H$, *then any* (p_H, p_L), *with* $p_H > w_H$, *cannot emerge as the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k. *Furthermore, any* (p_H, p_L), *with* $p_H = w_H$, *emerges as the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k.

2. *If* $\alpha_H < \rho < 1$, *then*

   (a) *Any* (p_H, p_L), *where* $p_H \in \mathcal{P}(\rho/\alpha_H, R_H - w_H)$ *and* $w_H \leq p_H < (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$, *emerges as the equilibrium price pair of a* ϵ^k*-symmetric Market Equilibrium in the* k^{th} *marketplace for large* k. *Furthermore, any* (p_H, p_L), *where* $p_H > (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$, *cannot emerge as the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k.

   (b) *There exists sequence of prices* p_{H_k} *such that* $p_{H_k} < R_H - R_L$, $\lim_{k\to\infty} p_{H_k} = R_H - R_L$, *and* $(p_{H_k}, 0)$ *is the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k.

3. *If* $\rho > 1$, *then any* (p_H, p_L), *where* $p_L \in \mathcal{P}(\rho, R_L)$ *and* $p_H = p_L + R_H - R_L$, *emerges as the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k. *Furthermore, any* (p_H, p_L), *where* $p_H \neq p_L + R_H - R_L$, *and* $p_H > (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$, *cannot emerge as the equilibrium price pair of a symmetric* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k.

Similar to the results in the model with identical agents, the above theorem shows that high-quality agents cannot charge more than 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. On the other hand, when demand is sufficiently high, 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 compare 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 zero price equilibrium when $R_H - w_H \simeq R_L$ and $w_H = 0$.

Our assumption that $R_H - w_H > R_L$ is again almost immaterial because, if we reversed this assumption, 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 (p_H, p_L) 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. *In the Operational Efficiency model with identical agents, any price vector* $(p_1, \ldots, p_N)$*, where* $N > 1$ *and* $p_n > 0$ *for all* $n \in \{1, \ldots, N\}$*, cannot emerge as the equilibrium price vector of an* ϵ^k*-Market Equilibrium in the* k^{th} *marketplace for large* k*.*

7.3. Communication Enabled Model

As in the model with identical agents, we now 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 Section 6 with the assumption that the fraction of agents of quality-i is α_i for $i \in \{H, L\}$ for all markets.

In Theorem 7, we show that only the high-quality agents serve customers by charging their operating costs in equilibrium when $\rho < \alpha$. Since (δ^k, ϵ^k)-Market Equilibrium is a refinement of ϵ^k-Market Equilibrium, this equilibrium is the only possible (δ^k, ϵ^k)-Market Equilibrium when $\rho < \alpha_H$.

Theorem 7 also establishes that there are multiple symmetric ϵ^k-Market Equilibria when $\rho > \alpha_H$. In most of these equilibria, the high-quality agents will be "over-utilized" since $\rho > \alpha_H$, and thus will have an opportunity to capitalize on this congestion when pre-play communication is allowed as we discussed in Section 6. When $\alpha_H < \rho < 1$, high-quality agents cannot charge a price higher than $R_H - R_L$ because of the threat of low-quality agents. On the other hand, agents can sustain a price, which extracts all of the customer surplus, when $\rho > 1$. We summarize these results in the following theorem:

THEOREM 8. *Let (p_H^k, p_L^k) be a price pair emerging in a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace.*

1. *If $\rho < \alpha_H$, then we have that*

$$\lim_{k\to\infty} p_H^k = w_H, \ \textit{and} \ \lim_{k\to\infty} D_1^{MCE}(R_H - p_H^l, R_L - p_L^k; \alpha_H k, \alpha_L k) = 1.$$

Furthermore, when $\lim_{k\to\infty} \delta^k = 0$, there exists a sequence (p_H^k, p_L^k) that forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for large k.

2. *If $\alpha_H < \rho < 1$, then we have that*

$$\lim_{k\to\infty} p_H^k = R_H - R_L, \ \lim_{k\to\infty} p_L^k = 0, \ \textit{and}$$
$$\lim_{k\to\infty} D_1^{MCE}(R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k) = \alpha_H/\rho, \ \lim_{k\to\infty} D_2^{MCE}(R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k) = (\rho - \alpha_H)/\rho.$$

Furthermore, when $\lim_{k\to\infty} \delta^k = 0$, there exists a sequence (p_H^k, p_L^k) that forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for large k when $\lim_{k\to\infty} \delta^k = 0$.

3. *If $\rho > 1$, then we have that*

$$\lim_{k\to\infty} p_H^k = R_H, \ \lim_{k\to\infty} p_L^k = R_L, \ \textit{and}$$
$$\lim_{k\to\infty} D_1^{MCE}(R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k) = 1/\rho, \ \lim_{k\to\infty} D_2^{MCE}(R_H - p_H^k, R_L - p_L^k; \alpha_H k, \alpha_L k) = 1/\rho.$$

Furthermore, there exists a sequence (p_H^k, p_L^k) that forms a symmetric (δ^k, ϵ^k)-Market Equilibrium in the k^{th} marketplace, for large k.

The implications on the identical agent model: Similar to Proposition 8, 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 $(p_1^k, \ldots, p_N^k)$, 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 $\rho > 1$, we have that*

$$\lim_{k\to\infty} p_n^k = R \ \textit{for all} \ n \in \{1, \ldots, N\}.$$

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 the different strategies employed by the moderating firm. In particular, we distinguish between operational tools and strategic tools and study the interplay between them in the way they alter the market outcome. In order to study the different strategies of the moderating firm, we consider three market models where the moderating firm has different degrees of involvement: (1) *No-Intervention Model* where the moderating firm does not intervene in the marketplace, (2) *Operational Efficiency Model* where the moderating firm efficiently matches customers and agents reducing the wasteful idleness and waiting times, (3) *Communication Enabled Model* where the moderating firm combines the efficient matching mechanism and communication among agents.

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, broadly divided into a buyer's market in which aggregate demand is less than the capacity and a seller's market in which aggregate demand exceeds the capacity. 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 revenue of agents since its profit is a share of the agents' revenue. In a given marketplace, the total revenue 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.

We show that when the firm ensures efficient operational matching and enables agent communication in a seller's market, the equilibrium prices as well as the revenues of agents approach R as the market size grows. Thus, the natural upper-bound on the revenue generated in a marketplace, $\min\{\Lambda, k\}R$, is asymptotically achievable when the moderating firm uses operational and strategic tools together and $\rho > 1$. In other words, the profit of the moderating firm from each agent is approximately R in large marketplaces when $\rho > 1$. Since the moderating firm can achieve the highest possible profit by providing efficient matching and communication, using these two tools together dominates any other strategy from the moderating firm's perspective in a seller's market.

We also show that any possible price but zero fails to emerge as the equilibrium price in a buyer's market. 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 inefficiency, such as providing real-time queueing information, will have a similar impact on the moderating firm's profit in a buyer's market.

Both oDesk.com and ServiceLive.com are 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.

There are also other possible ways for a moderating firm to be involved in the marketplace including contracting with the service providers 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 the 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.

References

Afeche, F., H. Mendelson. 2004. Pricing and priority auctions in queueing systems with generalized delay cost structure. *Management Science* **50** 869–882.

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

Allon, G., A. Federgruen. 2008. Service competition with general queueing facilities. *Operations Research* **56** 827–849.

Allon, G., I. Gurvich. 2008. Pricing and dimensioning competing large-scale service providers. Working Paper, Kellogg School of Management, Northwestern University.

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.

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

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.

De Vany, A. 1976. Uncertainty, waiting time, and capacity utilization: A stochastic theory of product quality. *Journal of Political Economy* **84** 523–540.

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.

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.

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

Lederer, P. J., L. Li. 1997. Pricing, production, scheduling, and delivery-time competition. *Operations Research* **43** 407–420.

Levhari, D., I. Luski. 1978. Duopoly pricing and waiting lines. *European Economic Review* **11** 17–35.

Li, L., Y. S. Lee. 1994. Pricing and delivery-time performance in a competitive environment. *Management Science* **40** 633–646.

Loch, C. H. 1991. Pricing markets sensitive to delay. Ph.D. thesis, Stanford University, Stanford, CA 94305.

Luski, I. 1976. On partial equilibrium in a queueing system with two servers. *Review of Economic Studies* **43** 519–525.

Maglaras, C., A. Zeevi. 2003. Pricing and capacity sizing for systems with shared resources: Scaling relations and approximate solutions. *Management Science* **49** 1018–1038.

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.

Mendelson, H., S. Whang. 1990. Optimal incentive-compatitable priority pricing for the M/M/1 queue. *Operations Research* **38** 870–883.

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

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.

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.

Ward, A. R., P. W. Glynn. 2005. A diffusion approximation for a GI/GI/1 queue with balking or reneging. *Queueing Systems* **50** 371–400.

Whitt, W. 2006. Fluid models for multiserver queues with abandonments. *Operations Research* **54** 37–54.

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 relegate the proofs of the lemmas used in the appendix to our Technical Appendix.

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* $m_a \leq 1$, *then* $\beta(\lambda)$ *is concave in* λ.

A.1. Proof of Proposition 1

Suppose there 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 $1 \leq n \leq k$, the unique equilibrium is clearly $D_n = 0$ for all $1 \leq n \leq k$.). Let

$$S = \{n \leq k : D_n > 0\}, \text{ and } S' = \{n \leq k : D'_n > 0\}.$$

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

Now, 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 $D_n \neq D'_n$ for some $n \in S$, we have that $u \neq u'$. WLOG, assume $u > u'$. This implies that $\sum_{n=1}^{k} D_n < \sum_{n=1}^{k} D'_n \leq 1$. 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

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

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 $g(\lambda) = \sum_{n=1}^{\infty} a_n \lambda^n$.

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:

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

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 $n\neq\ell$. 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:

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

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

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

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

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

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

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 using the above FOCs, any symmetric SPNE (D,p) should satisfy the following conditions:

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

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

Sufficient conditions for the symmetric equilibrium: The below lemma establish the existence of the symmetric SPNE when $\beta(\lambda)$ is concave.

LEMMA 3. *Let*

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

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*

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

To prove Proposition 2, we first argue that $\liminf_{k\to\infty} D_1^{MCE}(p;k) \geq \min\left\{1, \frac{R-p+cm_a}{\rho cm_a}\right\}$. Suppose on the contrary that the result does not hold. Then, there exists a convergent subsequence of $\{D_1^{MCE}(p;k)\}_{k=1}^{\infty}$, say $\{D_1^{MCE}(p;k)\}_{k=1}^{\infty}$ (we do not use a new notation for the subsequence for notational convenience), such that

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

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

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

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 $U_1(D_1^{MCE}(p;k);p;k) > 0$ whereas $D_1^{MCE}(p;k) < 1$. However, this contradicts with the definition of Market Customer Equilibrium.

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

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

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

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

$$\lim_{k\to\infty} U_1(D_1^{MCE}(p;k);p;k) = \lim_{k\to\infty} (R - p + cm_a)\left(1 - P\left(D_1^{MCE}(p;k),k\right)\right) - cm_a$$

$$= \frac{R-p+cm_a}{\max\{\tilde{D}(p)\rho, 1\}} - cm_a < cm_a - cm_a = 0,$$

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

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

B.2. Proof of Proposition 3

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

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

In order to prove Proposition 3, we consider a deviation by a single agent where he decreases his price by an arbitrary small amount $\varepsilon > 0$. Let

$$\begin{aligned} D_{pool}(k) &= D_1^{MCE}(p, p-\varepsilon; k-1, 1) \\ D_{one}(k) &= D_2^{MCE}(p, p-\varepsilon; k-1, 1), \end{aligned}$$

We first argue that $\lim_{k\to\infty} D_{pool}(k) + D_{one}(k) = 1$. Note that it is sufficient to show that $\liminf_{k\to\infty} D_{pool}(k) + D_{one}(k) = 1$ as $D_{pool}(k) + D_{one}(k) \le 1$ by definition. We prove this claim by contradiction, so that we suppose $\liminf_{k\to\infty} D_{pool}(k) + D_{one}(k) < 1$. Then, there should exist convergent subsequences of $D_{pool}(k)$ and $D_{one}(k)$ such that

$$\lim_{k\to\infty} D_{pool}(k) + D_{one}(k) < 1.$$

Using this observation, and letting $P_{one}(k) = PServ_{12}\big(D_{pool}(k), D_{one}(k); p, p-\varepsilon; k-1, 1\big)$ for notational convenience, we have that

$$\begin{aligned} &\lim_{k\to\infty} U_1\big(D_{pool}(k), D_{one}(k); p, p-\varepsilon; k-1, 1\big) \\ &\quad = \Big(1 - \lim_{k\to\infty} P_{one}(k)\Big)\Big[(R-p+cm_a)\big[1-\lim_{k\to\infty}\beta_2\big(D_{pool}(k), D_{one}(k); p, p-\varepsilon; k-1, 1\big)\big] - cm_a\Big] \\ &\qquad +(R-p+\varepsilon)\lim_{k\to\infty} P_{one}(k) \\ &\quad \ge (R-p+cm_a)\big[1-\lim_{k\to\infty}\beta_2\big(D_{pool}(k), D_{one}(k); p, p-\varepsilon; k-1, 1\big)\big] - cm_a \\ &\quad \ge (R-p+cm_a)\big[1-\lim_{k\to\infty}\beta^M(\Lambda_k D_{pool}(k); k-1)\big] - cm_a = R-p > 0, \end{aligned}$$

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 $\lim_{k\to\infty}\frac{\Lambda_k D_{pool}(k)}{k-1} < \rho < 1$. 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 $\liminf_{k\to\infty} D_{pool}(k) + D_{one}(k) = 1$.

Then using the fact that $\lim_{k\to\infty} D_{pool}(k)+D_{one}(k)=1$, we have that

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

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 $V'(p;k)$, and thus we have that

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

Finally, our claim holds since ε can be arbitrarily small.

B.3. Proof of Theorem 2

1. Let $V'(k;p)=\max_{0\leq p'\leq R} V_1(D_1^{MCE}(p',p;1,k-1),D_2^{MCE}(p',p;1,k-1);p',p;1,k-1)$. When $\rho<1$, we have that

$$\liminf_{k\to\infty} V'(p;k)=p>\rho p\geq \lim_{k\to\infty} V_1(D_1^{MCE}(p;k);p;k)+\lim_{k\to}\epsilon_k,$$

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

2. To prove this claim, we suppose, on the contrary, that $p=0$ cannot be a symmetric equilibrium, so that there is a sequence p'_k such that a single agent can improve his revenue by increasing his price to p'_k in the k^{th} marketplace. Let $U_{pool}(k)$ and $U_{dev}(k)$ be the utility of customers choosing price zero and 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 $U_{dev}(k)\geq U_{pool}(k)$ for any k. Using this observation we have that

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

where $P_{12}(k)$ is the probability that a customer picking p'_k is served by the agents charging zero in the k^{th} marketplace. The first inequality above holds since customers, who pick p'_k and served by the deviating agents, may abandon, and the last inequality holds as we show in the proof of Proposition 3. 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

$$p'_k[1-P_{12}(k)]\leq e^{-\zeta k}.$$

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

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

where the convergence holds as we assume that ϵ^k converges to zero at a speed of $1/k$. 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 p'_k is a profitable deviation. Hence, $p=0$ should emerge as the equilibrium price of a symmetric ϵ^k-Market Equilibrium for large k.

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* $D_n^{MCE}(k) = D_n^{MCE}(p', p; 1, k-1)$ *for* $n = 1, 2$, *and* $D^p(k) = D_1^{MCE}(p, ; k)$ *in a marketplace with* k *agents and demand rate* $\Lambda_k = \rho k$ *where* $p < p' < R$ *and* $\rho > 1$. *Then, the following statements are true.*

1. $\liminf_{k\to\infty} \frac{\Lambda_k D_2^{MCE}(k)}{k-1} > 1$.
2. $\lim_{k\to\infty} PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1) = 0$.
3. *Suppose* $\tilde{\lambda} = \lim_{k\to\infty} \Lambda_k D_1^{MCE}(k)$, *then we have that*
$$\lim_{k\to\infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1) = \beta^M(\tilde{\lambda}; 1)$$
4. *There exists a* K^* *such that* $\Lambda_k D_1^{MCE}(k) \le \bar{\lambda}$ *for any* $k > K^*$, *where* $\bar{\lambda}$ *is the unique solution to* $1 - \beta^M(\lambda; 1) = \frac{cm_a}{R + cm_a}$.
5. *Suppose* $\tilde{D} = \lim_{k\to\infty} D_2^{MCE}(k)$, *then we have that*
$$\lim_{k\to\infty} \beta_2(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1) = 1 - \frac{1}{\rho \tilde{D}}.$$

The proof of this lemma can be seen in the Online Appendix. Consider the following problem

$$\begin{aligned}
\Pi(p) = \max_{p \le p' \le R, \lambda \ge 0, D \ge 0} \quad & p'\lambda[1 - \beta(\lambda)] \\
s.to \quad & \\
& (R - p' + cm_a)[1 - \beta(\lambda)] \ge cm_a \\
& (R - p' + cm_a)[1 - \beta(\lambda)] = (R - p + cm_a)\frac{1}{\max\{D\rho, 1\}} \\
& D \le 1
\end{aligned}$$

We refer to this problem as the limit problem. After a case by case analysis, one can easily show that $\Pi(p) = \max_{\lambda} (R + cm_a)\lambda[1 - \beta(\lambda)] - \lambda[\Delta(p; R) + cm_a]$, where $\Delta(p; R) = \max\left\{0, \frac{R - p + cm_a}{\rho} - cm_a\right\}$. 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 $p'(k) = \arg\max_{p \le p' \le R} V_1(D_1^{MCE}(p', p; 1, k-1), D_2^{MCE}(p', p; 1, k-1); p', p; 1, k-1)$, $\lambda(k) = \Lambda_k D_1^{MCE}(p'(k), p; 1, k-1)$, and $D_{pool}(k) = D_2^{MCE}(p'(k), p; 1, k-1)$. Observe that for any $k = 1, 2, \ldots$,

$$p'(k) \in [p, R], \ \lambda(k) \in [0, \overline{\lambda}], \ D_{pool}(k) \in [0, 1], \ V'(k; p) \in [0, R],$$

where $\overline{\lambda}$ is defined as in Lemma 6. Observe that $\overline{\lambda} < \infty$ since we have that $\lim_{\lambda\to\infty} \beta^M(\lambda; 1) = 1$, and $c > 0$. In other words, the rate of customers a single agent serves is finite.

Then, there exists a convergent subsequence of $\{V'(p;k)\}_{k=1}^{\infty}$, which is denoted by again $V'(p;k)$ for notational convenience, such that

$$\lim_{k\to\infty} V'(p;k) = \limsup_{k\to\infty} V'(p;k).$$

Moreover, $\{p'(k)\}_{k=1}^{\infty}$, $\{\lambda(k)\}_{k=1}^{\infty}$, $\{D_{pool}(k)\}_{k=1}^{\infty}$ have convergent subsequences, and we let

$$\begin{aligned} \tilde{p}' &= \lim_{r\to\infty} p'(k) \\ \tilde{\lambda} &= \lim_{r\to\infty} \lambda(k) \\ \tilde{D}_{pool} &= \lim_{r\to\infty} D_{pool}(k). \end{aligned}$$

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

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

where the second equality follows by Lemma 6.2 and 6.3, and the last equality holds by 6.5.

Therefore, $(\tilde{p}', \tilde{\lambda}, \tilde{D}_{pool})$ satisfy all of the constraints in the limit problem. And, this implies that

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

B.5. Proof of Theorem 3

We first want to note that when $\rho > 1$, a single provider can not improve his revenue by more than ϵ^k when he cuts his price in a large marketplace since he is already fully-utilized by charging the same as everyone else. To argue that let $V^{cut}(p;k) = \max_{p'<p} V_1(D_1^{MCE}(p',p;1,k-1), D_2^{MCE}(p',p;1,k-1); p',p;1,k-1)$ in a marketplace with k^{th} for any $p < R$. 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 goes to infinity when all agents charge p in a seller's market, i.e there exists a constant ζ such that $V_1(D_1^{MCE}(p;k);p;k) = p(1-e^{-\zeta k})$. Using this observation, for large k, we have that

$$V_1(D_1^{MCE}(p;k);p;k) = p(1-e^{-\zeta k}) > p - \epsilon^k > V^{cut}(p;k) - \epsilon^k,$$

where the inequality holds since $\lim_{k\to\infty} \frac{e^{-\zeta k}}{\epsilon^k} = 0$ by our assumption that $\lim_{k\to\infty} -\frac{\ln(\epsilon^k)}{k} = 0$. This implies that a single agent cannot have a profitable deviation by decreasing his price. Hence, in order to verify that a price can emerge as an equilibrium outcome of a symmetric Market Equilibrium, it is sufficient to check any single agent deviation where the agent increases his price.

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

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

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

where the last inequality holds by Proposition 4. Then, for sufficiently large k, we have that $V'(p;k) < V_1(D_1^{MCE}(p;k);p;k)$, which implies that p can emerge as the equilibrium price of a symmetric Market Equilibrium.

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

$$p > \Pi(p,\rho_1) \geq \Pi(p,\rho_2) \Rightarrow \; p \in \mathcal{P}(\rho_2;R),$$

where the inequality holds since $\Pi(p,\rho)$ is increasing in rho. 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, $\lambda^\Delta(p)$ solves

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

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

2. Let λ_k^0 be the constant λ^0 defined in the k^{th} marketplace. Then, we have that

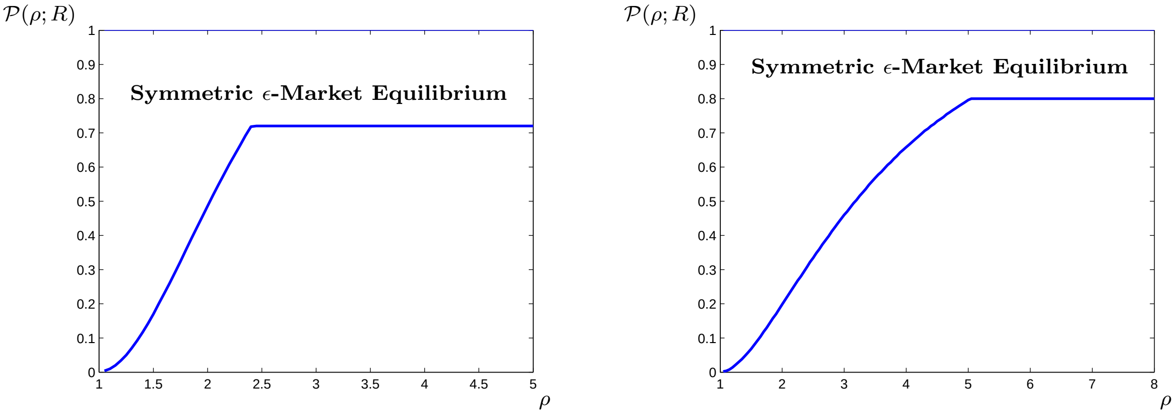
$$(R+cm_a)\frac{k-1}{k} - \frac{c}{1-\beta(\lambda_k^0/k)}\left(\frac{1-\beta(\lambda_k^0/k)}{1-\beta(\lambda_k^0/k) - \lambda_k^0/k\beta'(\lambda_k^0/k)} - \frac{1}{k}\right) = 0.$$

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

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

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 $m_a \neq \mu$



(a) **(b)**

Figure 5 **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 first show that $\lim_{k\to\infty} p_k = 0$. It is enough to show that $\limsup_{k\to\infty} p_k = 0$. To prove that we suppose $\limsup_{k\to\infty} p_k > 0$ on the contrary. Then, we would have a convergent sub-sequence of p_k which has a positive limit, say $\tilde{p}$. This also implies the revenue of the agents will converge to $\rho\tilde{p} < \tilde{p}$ as $k \to \infty$. However, as we showed in Theorem 2, even a single agent can ensure a revenue greater than $\rho\tilde{p}$ by cutting his price slightly, and this contradicts with the definition of market equilibrium. Hence, we should have that $\limsup_{k\to\infty} p_k = 0$

Next, we 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 p'_k such that δ^k fraction of agents can improve his revenue by increasing his price to 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 lemma and proposition:

Lemma 7. *Let $D_n^{MCE}(k) = D_n^{MCE}(p',p;\lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor)$ for $n=1,2$, and $D^p(k) = D_1^{MCE}(p,;k)$ in a marketplace with k agents and demand rate $\Lambda_k = \rho k$ where $p < p' < R$ 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. $\lim_{k\to\infty} \beta_1(D_1^{MCE}(k), D_2^{MCE}(k); \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor) = \min\left\{0, 1-\frac{1}{\rho\tilde{D}_1}\right\}$, *where* $\tilde{D}_1 = \lim_{k\to\infty} \frac{D_1^{MCE}(k)}{\delta^k}$.
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 Online Appendix.

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 $D_n^{MCE}(k) = D_n^{MCE}(p', p(k); \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor)$ $\lim_{k\to\infty} p(k) = p < R$, *and* $p < p' < \min\left\{R, p + \left(1-\frac{1}{\rho}\right)(R-p+cm_a)\right\}$. *Furthermore, we have that*

$$\lim_{k\to\infty} V_1(D_1^{MCE}(k), D_2^{MCE}(k); p', p(k); \lfloor\delta^k k\rfloor, k-\lfloor\delta^k k\rfloor) = p'.$$

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

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

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 p' can be written as

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

which implies that the expected utility of customers choosing the price p' 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 can be written as

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

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

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

which implies that customers strictly prefer providers charging p' over the ones charging p 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

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

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

■

C.3. Proof of Theorem 5

Suppose on the contrary that the result does not hold. Then, we can find a convergent subsequence, say p_k, such that $\lim_{k\to\infty} p_k < R$. Let $p = \lim_{k\to\infty} p_k$. 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 is a δ-Market Equilibrium. Hence, we should have that $\liminf_{k\to\infty} p_k = R$ which implies that $\lim_{k\to\infty} p_k = R$.

To show the existence of the equilibrium sequence, let $p_k = R - \beta^M(k;k)$. 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 $(R-\beta^M(k;k)(1-\beta^M(k;k)$. 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 $\lim_{k\to\infty} \frac{1}{\epsilon^k\sqrt{k}} = 0$, we have that $\epsilon^k > \zeta/\sqrt{k}$ for large k which implies that $V^k + \epsilon^k > R$ for large 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 $D^{info}(p;k)$ be the fraction of customers requesting service when all agents charge p in a marketplace where the moderating firm provides real-time congestion information. We first argue that $\lim_{k\to\infty} D^{info}(p;k) > 1/\rho$ when $\rho > 1$. The proof is very similar to the proof of Proposition 7: We suppose $\lim_{k\to\infty} D^{info}(p;k) \leq 1/\rho$. Then, we would have that $\lim_{k\to\infty} \beta^M(\Lambda_k D^{info}(p;k);k) = 0$, 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 $\lim_{k\to\infty} D^{info}(p,k) > 1/\rho$, we have that the limits of the revenue of the agents charging p when the firm provides real-time information, say $V^{info}(p;k)$, can be written as follows:

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

Now we argue that $\liminf_{k\to\infty} p_k^{info} \geq R(1-\zeta)$. Suppose not, i.e. $\lim_{k\to\infty} p_k^{info} < R(1-\zeta)$. Let $\tilde{p} = \lim_{k\to\infty} p_k^{info}$. Then, consider a deviation where all agents charge $p' > p/(1-\zeta)$. 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, we should have that $\liminf_{k\to\infty} p_k^{info} \geq R(1-\zeta)$.

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 the Technical Appendix. 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 ($\Lambda > k_H\lambda_H^{mon} + k_L\lambda_L^{mon}$): In this case both groups solves the following monopoly problem:

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

which can be reduced to $\max_{\lambda\geq 0}(R_i + cm_a - w_i)\lambda\left[1-\beta(\lambda)\right]$. 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 13 in the Technical Appendix. Thus, we should have that $\Lambda D_H^* > \Lambda D_L^*$ by their definition.

Furthermore, we have that $\Lambda D_H^* < \lambda_H^{mon}$ since

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

for any $0 \leq x \leq \lambda_L^{mon}$. 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 $\lambda \leq \lambda_H^{dom}$, 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:

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

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

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 $n \in \{1,\dots,N\}$, 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 $\hat{U}_\ell(\Lambda D_1^*,\dots,\Lambda D_N^*) \le 0$.

Without loss of generality, assume $D_1^* > \cdots > D_N^*$. This implies that $D_N^* < 1/k$ because otherwise we would have that $\sum\limits_{n=1}^{n} k_n D_n^* > 1$. 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 $n \in \{1,\dots,N\}$. Moreover, since $D_N^* < 1/k$ and both $\beta(x)$ and $\nu(x)$ are decreasing, we have that

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

Combining these observations, we have that

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

where the last inequality holds since $\rho < \lambda^0$ and $\hat{U}_N(\lambda^0,\dots,\lambda^0) = 0$ 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 $\hat{U}_n(\Lambda D_1^*,\dots,\Lambda D_N^*)$ in the equilibrium. Moreover, all customers request service. This means that, $(D_n^*)_{n=1}^N$ should solve that

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

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 $n \in \{1,\dots,N\}$. 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 $(D_1,\dots,D_N)$ which solves these equations. Without loss of generality, assume $D_1 > \cdots > D_N$. As we mentioned above, this implies that

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

This is a contradiction. Therefore, the only solution to above equation system is $D_n^* = 1/k$ for all $n \in \{1,\dots,N\}$. 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), i.e. $\Lambda D_n^* \leq \lambda^{mon}$ 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. *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 $\rho > 0$.*

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

where $p_H \leq R_H$, $p_L \leq R_L$.For any given (p_H, p_L) such that $R_i - p_i > R_j - p_j$, the following statements are true:

1. *If $\rho \leq \alpha_i$, 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) $\lim_{k\to\infty} \beta_j(k) = \max\left\{0, 1 - \frac{\alpha_j}{\rho \tilde{D}_j}\right\}$, *where* $\tilde{D}_j = \lim_{k\to\infty} D_j(k)$.
   (d) $\lim_{k\to\infty} \beta_i(k) = 1 - \frac{\alpha_i}{\rho \tilde{D}_i}$, *where* $\tilde{D}_i = \lim_{k\to\infty} D_i(k)$.
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*

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

The proof this lemma can be seen in the Technical Appendix.

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

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

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

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

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

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 $p_j - w_j$.

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

Lemma 9. *For any (p_H,p_L), where $1 < \frac{R_i-p_i+cm_a}{R_j-p_j+cm_a} < \frac{\rho}{\alpha_i}$, for some $i,j \in \{H,L\}$, $i \ne j$ and $p_j < R_j$, we have that*

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

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, and all quality-j providers charge p_j,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.

The proof of the lemma can be seen in the Technical Appendix.

Lemma 10. *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 Technical Appendix.

D.5. Proof of Theorem 7

1. We prove our claim by a case-by-case analysis:

i. $(\mathbf{R_H - p_H = R_L - p_L})$: In this case, all providers are pooled together regardless of the quality of their service, and all of them are under-utilized since $\rho < 1$. 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 (p_H,p_L) in this case can not emerge as an equilibrium price pair.

ii. $(\mathbf{R_L - p_L > R_H - p_H})$: In this case, we have two sub-cases:

a) $(\rho \le \frac{\mathbf{R_L-p_L+cm_a}}{\mathbf{R_H-p_H+cm_a}}\alpha_\mathbf{L})$: As we show in Lemma 8, the profit of a high-quality provider is zero as k goes to infinity. However, when a high-quality provider deviates to charge a price $p < R_H - R_L$, he becomes the least expensive provider, and always attracts strictly positive demand. This kind of deviation clearly improves his profit. Therefore, in large systems any (p_H,p_L) in this sub-case can not emerge as an equilibrium price pair.

b) $(\rho > \frac{\mathbf{R_L-p_L+cm_a}}{\mathbf{R_H-p_H+cm_a}}\alpha_\mathbf{L})$: In Lemma 9, we show that a low-quality provider can increase his price while keeping him fully-utilized when $p_H < R_H$. This kind of deviation clearly improves his profit. On the other hand, if $p_H = R_H$, a high-quality provider can deviate to charge a price $p < R_H - R_L$ and become the least expensive provider. As a result of this deviation he always attracts strictly positive demand. This kind of deviation clearly improves his profit since all high-quality agents earn zero when $p_H = R_H$. Therefore, in large systems any (p_H, p_L) in this sub-case can not emerge as an equilibrium price pair. Alternatively: In Lemma 9, we show that a high-quality provider earns a profit of $p_H - \varepsilon$ if he cuts his price when $p_H < R_H$. Moreover, by Lemma 8, we know that the profit of a high-quality provider when all high-quality providers charge p_H, and all low-quality providers charge p_L is less than $\frac{\rho-\alpha_L}{1-\alpha_L}p_H < \rho p_H$ since the demand rate of low-quality provider is greater than α_L. Then, cutting the price is a profitable deviation for a high-quality provider as long as $\varepsilon < p_H(1-\rho)$.

iii. $(\mathbf{R_H - p_H > R_L - p_L})$**:** In Lemma 8.1, we show that high-quality providers are under-utilized since $\rho < \alpha_H$. 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 (p_H, p_L) in this case can not emerge as an equilibrium price pair.

2. (Non-existence): As in part 1, we prove this claim by a case-by-case analysis:

i. $(\mathbf{R_H - p_H = R_L - p_L})$**:** The same as in the proof of the part 1, so that we omit the proof here.

ii. $(\mathbf{R_L - p_L > R_H - p_H})$**:** First note that $p_H > R_H - R_L \geq (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$ in this case. Then any (p_H, p_L) can not emerge as an equilibrium price as in part 1.

iii. $(\mathbf{R_H - p_H > R_L - p_L})$**:** In this case, we have two sub-cases:

a) $(\rho \leq \frac{\mathbf{R_H - p_H + cm_a}}{\mathbf{R_L - p_L + cm_a}}\alpha_\mathbf{H})$**:** As we show in Lemma 8, the profit of a low-quality provider is zero as k goes to infinity. Furthermore, since $p_H > (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$, the expected utility of customers converges to a limit, which is strictly less than R_L. This also implies that $p_L > 0$ in this case. Then, we can argue that when a low-quality provider deviates to charge a arbitrary small price $p' < R_L + cm_a - (R_H - p_H + cm_a)\frac{\alpha_H}{\rho} \leq p_L$, he always attracts strictly positive demand (Note that $R_L + cm_a - (R_H - p_H + cm_a)\frac{\alpha_H}{\rho} > 0$ by the lower bound on p_H.). The reasoning for this argument is similar to previous claims:

1. Demand for high-quality agents exceeds their capacity after the deviation of low-quality 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 $R_L - p'$ since $(R_H - p_H + cm_a)\frac{\rho}{\alpha_H} - cm_a < R_L$.

This kind of deviation clearly improves his profit. Therefore, in large systems any (p_H, p_L) in this sub-case can not emerge as an equilibrium price pair.

b) $(\rho > \frac{\mathbf{R_H - p_H + cm_a}}{\mathbf{R_L - p_L + cm_a}}\alpha_\mathbf{H})$**:** In Lemma 9, we show that a high-quality provider can increase his price while ensuring a revenue strictly greater than p_H when $p_L < R_L$. This kind of deviation clearly improves his profit. On the other hand, if $p_L = R_L$, a low-quality provider can deviate to charge a arbitrary small price $p' < \min\{p_L, R_L + cm_a - (R_H - p_H + cm_a)\frac{\alpha_H}{\rho}\}$ and always attract strictly positive demand. The reasoning for this argument is similar to previous claims:

1. Demand for high-quality agents exceeds their capacity after the deviation of low-quality agent.
2. Therefore, all customers picking the deviating agent should be served by him.
3. Moreover, since $\rho > \frac{R_H - p_H + cm_a}{cm_a}\alpha_H$ when $p_L = R_L$, the utility of customers after deviation will be zero.
4. Hence, some customers should request service from the deviating agent since $R_L - p' > 0$. 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 low-quality agents earn zero when $p_L = R_L$. Therefore, in large systems any (p_H, p_L) in this sub-case can not emerge as an equilibrium price pair.

2. (Existence): We prove that the proposed price pair emerges as the equilibrium price like the proof of Theorem 3. First, we want to note that low-quality agents will always be out of the market for large k since $p_H < (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$ implies that the expected utility of customers converges to a limit, which is strictly greater than R_L.

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

$$\begin{aligned}\limsup_{k\to\infty} V'_H(k) &\leq \max_{\lambda}(R_H + cm_a - w_H)\lambda[1-\beta(\lambda)] - \lambda(\Delta^H(p_H; R_H) + cm_a)\\ &\leq \max_{\lambda}(R_H + cm_a - w_H)\lambda[1-\beta(\lambda)] - \lambda(\Delta^H(p_H; R_H - w_H) + cm_a),\end{aligned}$$

where the first inequality holds as in the proof of Proposition 4, and the second one holds by the envelope theorem and the fact that $\Delta^H(p_H; R_H) \geq \Delta^H(p_H; R_H - w_H)$. Then, as in the proof of Theorem 3, high-quality agents do not have a profitable deviation since $p_H \in \mathcal{P}(\rho/\alpha_H, R_H - w_H) \Rightarrow p_H > \max_{\lambda}(R_H + cm_a - w_H)\lambda[1 - \beta(\lambda)] - \lambda(\Delta^H(p_H; R_H - w_H) + cm_a)$.

2. (Existence of the sequence): Let p_{H_k} be a sequence converging to $R_H - R_L - w_H$ with a rate of $1/\sqrt{k}$, i.e. $p_{H_k} = R_H - R_L - \beta^M(\alpha_H k; \alpha_H k)$. By construction, the high-quality agents, who charges p_{H_K}, attracts a demand rate of $\alpha_H k$ when all high-quality agents charge p_{H_K} and all low-quality 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 high quality agents will be in the form of $R_H - R_L - w_H - \zeta_2/\sqrt{k}$, where ζ_2 is a constant, given that all high-quality agents charge p_{H_K} and all low-quality agents charge zero.

As we assume that $\frac{1}{\epsilon_k\sqrt{k}} \to 0$, we should have that $R_H - R_L - w_H - \zeta_2/\sqrt{k} + \epsilon_k > R_H - R_L - w_H$. This implies that a high-quality agent should charge more than $R_H - R_L$ 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, high-quality agents do not have a profitable deviation from this price pair. The same would happen if a low-quality agent would increase his price.

Note that such a sequence of price pair, say (p_{H_k}, p_{L_k}), cannot be created with another limit: Any limit $(\tilde{p}_H, \tilde{p}_L)$ such that $R_H - \tilde{p}_H \neq R_L - \tilde{p}_L$ can be ruled out by part a (such a proof is similar to the proof of Theorem 2.2). Moreover, Any limit $(\tilde{p}_H, \tilde{p}_L)$ such that $R_H - \tilde{p}_H = R_L - \tilde{p}_L$ and $\tilde{p}_L > 0$ can be ruled out by the fact that low quality-agents could have a profitable deviation in such a case since they charge a strictly positive price and they are under-utilized.

3. The proof of the non-existence part is the same as the proof in part 2. Furthermore, we prove that the proposed price pair emerges as the equilibrium price like the proof of Theorem 3.

We let $V'_L(k)$ ($V'_H(k)$ be the maximum profit that a low-quality (high-quality) provider can get by increasing his price. Then, we can show that

$$\begin{aligned}
\limsup_{k\to\infty} V'_L(k) &\le (R_L + cm_a)\lambda^\Delta(p_L;R_L)[1-\beta(\lambda^\Delta(p_L;R_L))] - \lambda^\Delta(p_L;R_L)(\Delta(p_L;R_L)+cm_a)\\
\limsup_{k\to\infty} V'_H(k) &\le \max_\lambda (R_H + cm_a - w_H)\lambda[1-\beta(\lambda)] - \lambda(\Delta(p_H;R_H)+cm_a)\\
&\le \max_\lambda (R_H + cm_a)\lambda[1-\beta(\lambda)] - \lambda(\Delta(p_H;R_H)+cm_a)\\
&= (R_H + cm_a)\lambda^\Delta(p_H;R_H)[1-\beta(\lambda^\Delta(p_H;R_H))] - \lambda^\Delta(p_H;R_H)(\Delta(p_H;R_H)+cm_a),
\end{aligned}$$

where the first two inequality holds as in the proof of Proposition 4, and the last inequality holds by the envelope theorem and the fact that $w_H \ge 0$. Then, as in the proof of Theorem 3, low-quality providers do not have a profitable deviation since $p_L \in \mathcal{P}(\rho, R_L)$. Moreover, high-quality provides also do not have a profitable deviation since $p_L \in \mathcal{P}(\rho, R_L)$ implies that $p_H \in \mathcal{P}(\rho, R_H)$ by Lemma 10.

D.6. Proof of Proposition 8

Consider a price vector $(p_1,\dots,p_N)$, where $N > 1$ and $p_n > 0$ for all $n \in \{1,\dots,N\}$. Suppose α_n fraction of agents charge p_n for all $n \in \{1,\dots,N\}$ in the k^{th} marketplace. We will show that a single agents will have a profitable deviation from this strategy for large k.

When $N = 2$, we can use the result of Theorem 7 by letting $R_H = R_L = R$ and $w_H = 0$ because this theorem states that any price vector where different class of agents charge different strictly positive prices cannot emerge as the equilibrium price as k grows. However, $N > 2$ needs extra arguments. Suppose there exists a $(p_1,\dots,p_N)$ which emerges as equilibrium in the k^{th} marketplace as $k \to \infty$.

We first need to characterize the revenue of agents given the price vector $(p_1,\dots,p_N)$ 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. Without loss of generality, we assume that $p_1 < \dots < p_N$. 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 $\lim_{k\to\infty} V_\ell^k > 0$ 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 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 $n < N$: This case also contradicts with the definition of Market Equilibrium because sub-pool-$(N-1)$ earns strictly more than sub-pool-n for all $n < N-1$ since $p_1 < \dots < p_N$.

As both of these cases lead to a contradiction, our assumption that $(p_1,\dots,p_N)$ is equilibrium as $k \to \infty$ is wrong.

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

Lemma 11. *When $\rho > \alpha_H$, we have that*

$$\lim_{k\to\infty} V_H'(k) = p' - w_H,$$

where $V_H'(k)$ is the profit of the δ fraction of high-quality providers charging p' when all other high-quality providers charge p_H, and all low-quality providers charge p_L, with $p_H \leq (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$, and $p_H < p' < \min\left\{R_H - R_L, p_H + \left(1 - \frac{\alpha}{\rho}(R_H - p_H + cm_a)\right)\right\}$.

D.8. Proof of Theorem 8

1. The non-existence of any other limit holds as in Theorem 7. Furthermore, 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 $\rho < \alpha_H - \tilde{\delta}$ if $\lim\limits_{k\to\infty} \delta^k = \tilde{\delta} > 0$.

2. We have already shown that any (p_H, p_L), where $R_H - p_H \neq R_L - p_L$ and $p_H > (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$, can not be an equilibrium in a large marketplace in Theorem 7.2. Furthermore, Lemma 11 provides a profitable deviation for a small group of agent when $p_H \leq (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$. Hence, any (p_H, p_L) with $R_H - p_H \neq R_L - p_L$ cannot emerge as the equilibrium price pair of a symmetric (δ^k, ϵ^k)-Market Equilibrium in large marketplaces even if $p_H \leq (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$. As a direct implication of that any sequence (p_H^k, p_L^K) which converges to a limit (p_H, p_L) with $R_H - p_H \neq R_L - p_L$ cannot be an equilibrium because agents could improve their revenues by following a deviation similar to the deviation improving their revenues from (p_H, p_L). Finally, we can rule out any sequence (p_H^k, p_L^K) which converges to a limit (p_H, p_L) with $R_H - p_H = R_L - p_L$ and $p_L > 0$ as we discussed in the proof of Theorem 7.2.b. Hence, the only possible limit for a sequence of equilibrium prices (p_H^k, p_L^K) is $(R_H - R_L, 0)$.

The limits of customers holds as follows: Let the limit of the fraction of customers choosing high-quality agents be $\tilde{D}_H$ and suppose $\tilde{D}_H < \alpha_H/\rho$ on the contrary. Then, we can argue that high quality agents have an incentive to cut their prices since their utilization is less than 100%. Hence, we should have that $\tilde{D}_H \geq \alpha_H/\rho$. Now, suppose $\tilde{D}_H > \alpha_H/\rho$. In such a case, the utility of customers choosing high-quality agents would be strictly less than R_L while the utility of customers choosing low-quality agents was R_L. However, this contradicts with the definition of Customer Equilibrium. Thus, we should have that $\tilde{D}_H = \alpha_H/\rho$.

Finally, the existence of the sequence holds as in the proof of Theorem 7.2.b with the help of our assumption that $\lim\limits_{k\to\infty} \delta^k = 0$. If $\lim\limits_{k\to\infty} \delta^k = \tilde{\delta} > 0$, we can show the existence of such a sequence only for $\rho < 1 - \tilde{\delta}$.

3. Again by Lemma 11, (p_H, p_L) can not emerge as the equilibrium price pair of a symmetric (δ^k, ϵ^k)-Market Equilibrium in large marketplaces even if $p_H \leq (R_H - R_L) - (R_L + cm_a)\left[\frac{\rho}{\alpha_H} - 1\right]$. Therefore, any price pairs with $p_H \neq R_H - R_L + p_L$ can not emerge as a price pair in large marketplaces, and we are only left with the case where $p_H = R_H - R_L + p_L$. In this case, all providers are pooled together regardless of their quality, so that our claim holds as in Theorem 5. The existence of the sequence also holds as in Theorem 5.

D.9. Proof of Proposition 9

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 $\tilde{p}_n \neq \tilde{p}_m$ for any $n \neq m$, $n \in \{1,\ldots,N\}$, $m \in \{1,\ldots,N\}$ (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 rules 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 11. Hence, this case also cannot be true.

TECHNICAL APPENDIX

This Technical Appendix presents the proofs of all the supporting results in the paper. The reader can find the proof of main results in the paper. This Technical Appendix will be published in the doctoral dissertation of the first author.

Appendix T.A. 1: No-Intervention (Identical Agents)

T.A. 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

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

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 $1 \leq k \leq \lfloor \frac{n}{2} \rfloor$, 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. $s_{k-1} \geq s_k$ for any $2 \leq k \leq \lfloor \frac{n}{2} \rfloor$.
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

$$2\lambda g'[\lambda g' + 1 + g] = 2\lambda \Big[1 + g + (m_a - 1)[1 + g - g/\lambda] \Big] \Big[\lambda[1+g] + \lambda(m_a - 1)[1 + g - g/\lambda] + 1 + g \Big]$$

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

Combining above equations, we have that

$$
\begin{aligned}
&2\lambda g'[\lambda g'+1+g] - [1+g]\bigl[\lambda^2 g''+2g[1+g]\bigr] \\
&\quad = [1+g]^2\bigl[\lambda^2(2-m_a)+2\lambda-2g\bigr] + \lambda(m_a-1)[1+g][1+g-g/\lambda][4\lambda+2-m_a(\lambda-1)] \\
&\quad +2\bigl[\lambda(m_a-1)[1+g-g/\lambda]\bigr]^2 \\
&\quad = [1+g]^2\bigl[\lambda^2(2-m_a)+2\lambda-2g\bigr] + \lambda(\lambda+1)(2+m_a)(m_a-1)[1+g][1+g-g/\lambda] \\
&\quad -2\lambda(m_a-1)^2 g[1+g-g/\lambda] \\
&\quad \le [1+g]\left[[1+g]\bigl[\lambda^2(2-m_a)+2\lambda-2g\bigr] + [1+g-g/\lambda][\lambda(\lambda+1)(2+m_a)(m_a-1)]\right] \\
&\quad = [1+g]\left[[1+g]\bigl[\lambda^2 m_a^2+\lambda m_a(m_a+1)-2g\bigr] - g\bigl[(\lambda+1)(2+m_a)(m_a-1)\bigr]\right] \\
&\quad = [1+g]\left[\begin{array}{c} [1+g]\bigl[\lambda^2 m_a^2+\lambda m_a(m_a+1)-m_a(m_a+1)g\bigr] \\ -g\bigl[[1+g][2-m_a(m_a+1)]+(\lambda+1)(2+m_a)(m_a-1)\bigr] \end{array}\right] \\
&\quad = [1+g]\left[[1+g]m_a(m_a+1)\bigl[a_2\lambda^2+\lambda-g\bigr] - g\bigl[g-\lambda\bigr][2-m_a(m_a+1)]\right] < 0,
\end{aligned}
$$

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

T.A. 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

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

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

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

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

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

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

T.A. 1.3. Proof of Lemma 3

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

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

and the FOC of this problem are:

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

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

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

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

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

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

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 $D_\ell(p)+(k-1)D_{-\ell}(p)\leq 1$.

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 $\eta_2\leq\Lambda D_\ell(p)<\Lambda/k$ (since $\eta_1\geq 0$) 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 $(R-p_\ell(p)+cm_a)[1-\beta(\Lambda D_\ell(p))]\geq\Delta$.

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 T.A. 2: Operational Efficiency (Identical Agents)

T.A. 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.

T.A. 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

$$\begin{aligned}
\pi_n &= \frac{\lambda_2}{1 + (n-1)/m_a} \text{ for any } n > 1,\\
\pi_1 &= (\lambda_1 + \lambda_2)\pi_0,\\
\pi_0 &= \frac{1}{1 + (\lambda_1 + \lambda_2)\frac{g(\lambda_2)}{\lambda_2}},
\end{aligned}$$

where $g(\lambda) = \sum_{n=1}^{\infty} \frac{\lambda^n}{\prod_{i=0}^{n-1}(1+i/m_a)}$. Note that

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

since $g(\lambda)/\lambda \geq 1$ for any $\lambda \geq 0$. Using this observation, we have that

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

T.A. 2.3. Proof of Lemma 6

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

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

We first want to note that

$$\beta_2(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1) \leq \beta^{MM1}(\Lambda_k D_2^{MCE}(k), k-1).$$

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 $\lim_{k\to\infty} \frac{\Lambda_k D_2^{MCE}(k)}{k-1} \leq 1$, we have that

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

where the last equality is due to Ward and Glynn (2005). Using this result, we have that

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

which implies that utility of customers choosing the price p 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 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

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

Dividing both sieds by Λ_k, the above equation implies that

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

Letting k go to infinity, we have that

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

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

$$\begin{aligned}
&\lim_{k\to\infty} U_1(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1) \\
&\quad = \left[1 - \lim_{k\to\infty} \hat{P}_{pool}(k)\right]\left[(R - p' + cm_a)(1 - \beta_{one}(k)) - cm_a\right] + (R-p)\lim_{k\to\infty} \hat{P}_{pool}(k)
\end{aligned}$$

$$\begin{aligned}
&\leq \left(1-\frac{1}{\rho}\right)(R-p')+\frac{1}{\rho}(R-p) \\
&< R-p, \qquad (24)
\end{aligned}$$

where the first inequality holds by (23).

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

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

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 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

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

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

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

because some of the customer choosing p' may be served by sub-pool-2. Therefore, it is sufficient to show that

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

To show that we consider a hypothetical situation where any customer choosing the price p' 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

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

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

$$\beta_1(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1) \geq \beta^M(\Lambda_k D_1^{MCE}(k), 1) - PServ_{12}(D_1^{MCE}(k), D_2^{MCE}(k); p', p; 1, k-1).$$

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

4. Suppose NOT. Then 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.

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

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

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

$$\big[k-1+(n-k+1)/m_a\big]\pi_n = \Lambda_k D_2^{MCE}(k)\pi_{n-1},$$

for any $n > k-1$. Furthermore, we have that

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

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

Appendix T.A. 3: Communication Enabled Model (Identical Agents)

T.A. 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

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

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

$$\lim_{k\to\infty} \beta_2(D_1^{MCE}(k), D_2^{MCE}(k); p', p; \lfloor \delta^k k\rfloor, k-\lfloor \delta^k k\rfloor) \leq \lim_{k\to\infty} \beta^{MM1}(\Lambda_k D_2^{MCE}(k), k-\lfloor \delta^k k\rfloor) = 0, \quad (25)$$

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

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

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 (25), we have that

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

Using the above observations, we have that

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

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

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

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

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

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

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

To show that we consider a hypothetical situation where any customer choosing the price p' 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

$$\begin{aligned}\Lambda_k D_1^{MCE}(k)\beta^M(\Lambda_k D_1^{MCE}(k);\lfloor\delta^k k\rfloor) = {}&\Lambda_k D_1^{MCE}(k)\beta_1(D_1^{MCE}(k),D_2^{MCE}(k);p',p;\lfloor\delta^k k\rfloor,k-\lfloor\delta^k k\rfloor)\\ &+\Lambda_k D_1^{MCE}(k)\beta^{color}(k),\end{aligned}$$

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

$$\begin{aligned}\beta_1(D_1^{MCE}(k),D_2^{MCE}(k);p',p;\lfloor\delta^k k\rfloor,k-\lfloor\delta^k k\rfloor) \ge {}&\beta^M(\Lambda_k D_1^{MCE}(k);\lfloor\delta^k k\rfloor)\\ &-PServ_{12}(D_1^{MCE}(k),D_2^{MCE}(k);p',p;\lfloor\delta^k k\rfloor,k-\lfloor\delta^k k\rfloor).\end{aligned}$$

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

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

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

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

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

Appendix T.A. 4: No-Interventions Efficiency (Non-Identical Agents)

T.A. 4.1. Supplementary Claims to Characterize SPNE

LEMMA 12. *Let*

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

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 13. *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 12.

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 12.

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 12. 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

$$\hat{U}_L\big(x_1,y(x_1)\big)\leq\hat{U}_L\big(x_2,y(x_2)\big). \tag{26}$$

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

$$\begin{aligned}& (R_L + cm_a)\left[\frac{1-\nu(x_2)}{1+\frac{\nu(x_2)}{k_L+k_H\vartheta(x_2,y(x_2))-1}}\right][1-\beta(x_2)] - cm_a \\ & = \; \hat{U}_L\big(x_2, y(x_2)\big) \; > \; (R_L + cm_a)\left[\frac{1-\nu(x_2)}{1+\frac{\nu(x_2)}{k_L+k_H\vartheta(x_1,y(x_1))-1}}\right][1-\beta(x_2)] - cm_a \\ & \Rightarrow \vartheta(x_2, y(x_2)) > \vartheta(x_1, y(x_1)). \end{aligned} \tag{27}$$

Note that by Equation 26, 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

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

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 27 and 28, 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*

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

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

1. *If $\Lambda \geq \Lambda^{R_L}$, 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 $0 \leq u \leq R_L$, 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 $k_L x + k_H y(x)$ is strictly increasing in x by Lemma 13. Hence, it is clear that there exists a unique $\hat{D}_L(\Lambda)$ such that

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

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

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

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

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

Then, the claim follows since $k_L x + k_H y(x)$ is strictly increasing in x by Lemma 13.

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 13.3.

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

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

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 13.3. 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 13.

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*

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

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 $k_Lx + k_Hy = \Lambda$, 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.

■

T.A. 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:

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

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

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

$$\Lambda D_H - \eta_{1_H} - \eta_{2_H} = 0, \tag{29}$$

$$\Lambda(p_H - w_H)[1-\beta(\Lambda D_H)] - \Lambda^2 D_H(R + cm_a - w_H)\beta'(\Lambda D_H) - \eta_{4_H} = 0, \tag{30}$$

$$(\eta_{2_H} - \eta_{3_H})\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H) - (k_H - 1)\eta_{4_H} = 0, \tag{31}$$

$$\eta_{3_H}\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L) - k_L\eta_{4_H} + \eta_{5_H} = 0, \tag{32}$$

$$\eta_{1_H}\big((R_H - p_H + cm_a)[1-\beta(\Lambda D_H)] - cm_a\big) = 0, \tag{33}$$

$$\eta_{3_H}\big((R_H - p_H + cm_a)[1-\beta(\Lambda D_H)] - (R_L - p_L + cm_a)[1-\beta(\Lambda D_L)]\big) = 0, \tag{34}$$

$$\eta_{4_H}(1 - k_L D_L - k_H D_H) = 0, \tag{35}$$

$$\eta_{5_H} D_L = 0, \tag{36}$$

$$\eta_{1_H},\ \eta_{3_H},\ \eta_{4_H},\ \eta_{5_H} \geq 0, \tag{37}$$

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

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

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 30.

Then, using Equation 29, 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

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

Case-3 ($0 < U_H^{SPNE}(D_L, D_H; p_L, p_H) < R_L$): First, observe that

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

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 30.

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

$$\begin{aligned}
U_H^{SPNE}(D_L, D_H; p_L, p_H) &= \frac{(R_H + cm_a - w_H)\big[k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L)\big]}{k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L) - \nu(\Lambda D_H)}\big[1 - \nu(\Lambda D_H]\big[1 - \beta(\Lambda D_H)\big] - cm_a \\
&= \frac{(R_H + cm_a - w_H)\big[1 - \nu(\Lambda D_H]\big[1 - \beta(\Lambda D_H)\big]}{1 + \frac{\nu(\Lambda D_H)}{k_H - 1 + k_L\vartheta(\Lambda D_H, \Lambda D_L)}} - cm_a \\
&= \hat{U}_H(\Lambda D_L, \Lambda D_H),
\end{aligned}$$

where $\hat{U}_H(\Lambda D_L, \Lambda D_H)$ is defined in Lemma 12.

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 30 and 31, we have that

$$\begin{aligned}
\eta_{3_H} &= \eta_{2_H} - \frac{(k_H-1)}{\Lambda(R_H-p_H+cm_a)\beta'(\Lambda/k_H)}\eta_{4_H}\\
&= \Lambda/k_H - \left[\frac{(k_H-1)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]\left[(R_H+cm_a-w_H)\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]-(R_L+cm_a)\right]\\
&= \frac{\left[(R_L+cm_a)\big[k_H-1+\nu(\Lambda/k_H)\big]-(R_H+cm_a-w_H)(k_H-1)\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]\right]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\\
&= -\left[(R_H+cm_a-w_H)(k_H-1)-\frac{R_L+cm_a}{1-\beta(\Lambda/k_H)}\left(\frac{k_H-1+\nu(\Lambda/k_H)}{1-\nu(\Lambda/k_H)}\right)\right]\frac{\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\\
&= -\left[(R_H+cm_a-w_H)(k_H-1)-\frac{R_L+cm_a}{1-\beta(\Lambda/k_H)}\left(\frac{k_H}{1-\nu(\Lambda/k_H)}-1\right)\right]\frac{\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\\
&= -z^{R_L}(\Lambda/k_H)\left[\frac{\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right],
\end{aligned}$$

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,

$$U_H^{SPNE}(D_L,D_H;p_L,p_H)=R_L \Rightarrow z^{R_L}(\Lambda/k_H)\leq 0 \ \Rightarrow \Lambda \geq k_H\lambda_H^{R_L}, \tag{38}$$

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 30-32, we have that

$$\begin{aligned}
\eta_{2_H} &= \eta_{3_H} + \frac{(k_H-1)}{\Lambda(R_H-p_H+cm_a)\beta'(\Lambda/k_H)}\eta_{4_H}\\
&= \left[\frac{k_L}{\Lambda(R_L-p_L+cm_a)\beta'(0)}+\frac{(k_H-1)}{\Lambda(R_H-p_H+cm_a)\beta'(\Lambda/k_H)}\right]\eta_{4_H}-\eta_{5_H}\\
&= \left[\frac{k_H-1+k_L\vartheta(0,\Lambda/k_H)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]\left[(R_H+cm_a-w_H)\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]-(R_L+cm_a)\right]-\eta_{5_H}.
\end{aligned}$$

Then, using the fact that $\eta_{2_H}=\Lambda/k_H$, we have that

$$\eta_{5_H} = \left[\frac{k_H-1+k_L\vartheta(0,\Lambda/k_H)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right]\left[(R_H+cm_a-w_H)\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big]-(R_L+cm_a)\right]-\Lambda/k_H$$

$$
\begin{aligned}
&= \begin{bmatrix} (R_H+cm_a-w_H)\big[k_H-1+k_L\vartheta(0,\Lambda/k_H)\big]\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big] \\ -(R_L+cm_a)\big[k_H-1+k_L\vartheta(0,\Lambda/k_H)+\nu(\Lambda/k_H)\big] \end{bmatrix} \frac{1}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}} \\
&= \begin{bmatrix} (R_H+cm_a-w_H)\frac{\big[k_H-1+k_L\vartheta(0,\Lambda/k_H)\big]}{\big[k_H-1+k_L\vartheta(0,\Lambda/k_H)+\nu(\Lambda/k_H)\big]}\big[1-\nu(\Lambda/k_H)\big]\big[1-\beta(\Lambda/k_H)\big] \\ -(R_L+cm_a) \end{bmatrix} \\
&\quad \times \left[\frac{k_H-1+k_L\vartheta(0,\Lambda/k_H)+\nu(\Lambda/k_H)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right] \\
&= \Big[\hat{U}_H(0,\Lambda/k_H)-R_L\Big]\left[\frac{k_H-1+k_L\vartheta(0,\Lambda/k_H)+\nu(\Lambda/k_H)}{(R_L+cm_a)\frac{\beta'(\Lambda/k_H)}{1-\beta(\Lambda/k_H)}}\right].
\end{aligned}
$$

Note that we have $\eta_{5_H} \geq 0$. Thus,

$$U_H^{SPNE}(D_L,D_H;p_L,p_H)=R_L \Rightarrow \hat{U}_H(0,\Lambda/k_H)\geq R_L \ \Rightarrow \Lambda \leq \Lambda(R_L), \tag{39}$$

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 (38) and (39), 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 30 and 31, 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

$$
\begin{aligned}
U_H^{SPNE}(D_L, D_H; p_L, p_H)) &= (R_H + cm_a - w_H)\left(\frac{(k_H-1)\big[1-\beta(\Lambda/k_H)\big]}{\frac{k_H}{1-\nu(\Lambda/k_H)} - 1}\right) - cm_a > R_L \\
&\Rightarrow \left[(R_H + cm_a - w_H)(k_H - 1) - \frac{R_L + cm_a}{1-\beta(\Lambda/k_H)}\left(\frac{k_H}{1-\nu(\Lambda/k_H)} - 1\right)\right]\frac{1-\beta(\Lambda/k_H)}{\frac{k_H}{1-\nu(\Lambda/k_H)} - 1} > 0 \\
&\Rightarrow z^{R_L}(\Lambda/k_H) > 0 \ \Rightarrow \Lambda < k_H \lambda_H^{R_L}.
\end{aligned}
$$

T.A. 4.3. Sufficiency for the Equilibrium

LEMMA 14. *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:

$$\Lambda D_H - \eta_{1_H} - \eta_{2_H} = 0, \tag{40}$$

$$\Lambda p_\ell[1 - \beta(\Lambda D_\ell)] - \Lambda^2 D_H (R + cm_a)\beta'(\Lambda D_\ell) - \eta_{4_H} = 0, \tag{41}$$

$$(\eta_{2_H} - \eta_{3_H})\Lambda(R_H - p_H + cm_a)\beta'(\Lambda D_H) - (k_H - 1)\eta_{4_H} = 0, \tag{42}$$

$$\eta_{3_H}\Lambda(R_L - p_L + cm_a)\beta'(\Lambda D_L) - k_L \eta_{4_H} + \eta_{5_H} = 0, \tag{43}$$

$$\eta_{1_H}\big((R_H - p_\ell + cm_a)[1 - \beta(\Lambda D_\ell)] - cm_a\big) = 0, \tag{44}$$

$$\eta_{3_H}\big((R_H - p_H + cm_a)[1 - \beta(\Lambda D_H)] - (R_L - p_L + cm_a)[1 - \beta(\Lambda D_L)]\big) = 0, \tag{45}$$

$$\eta_{4_H}(1 - k_L D_L - k_H D_H) = 0, \tag{46}$$

$$\eta_{5_H} D_L = 0, \tag{47}$$

$$\eta_{1_H},\ \eta_{3_H},\ \eta_{4_H},\ \eta_{5_H} \geq 0, \tag{48}$$

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 ($\Lambda^0 \leq \Lambda \leq \Lambda^{mon}$): 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 $\Lambda^0 \leq \Lambda \leq \Lambda^{mon}$.

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 (42) and (43), we have that

$$\begin{aligned}
\eta_{2_H} &= \eta_{3_H} + \frac{(k_H - 1)[1 - \beta(\Lambda D_H^*)]}{cm_a \Lambda \beta'(\Lambda D_H(p))} \eta_{4_H} \\
&= \left[\frac{k_L[1 - \beta(\Lambda D_L^*)]}{cm_a \beta'(\Lambda D_L(p))} + \frac{(k_H - 1)[1 - \beta(\Lambda D_H^*)]}{cm_a \Lambda \beta'(\Lambda D_H(p))} \right] \eta_{4_H} - \eta_{5_H} \frac{[1 - \beta(\Lambda D_L^*)]}{cm_a \beta'(\Lambda D_L(p))} \\
&\leq \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H^*, \Lambda D_L^*)}{cm_a \frac{\beta'(\Lambda D_H^*)}{1 - \beta(\Lambda D_H^*)}} \right] \frac{\eta_{4_H}}{\Lambda},
\end{aligned}$$

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 (42)and the fact that $p_\ell(p) \leq R_H + cm_a - \frac{cm_a}{1 - \beta(\Lambda D_\ell(p))}$, we have that

$$\begin{aligned}
\frac{\eta_{4_H}}{\Lambda} &\leq \Big((R_H + cm_a)[1 - \nu(\Lambda D_H(p))][1 - \beta(\Lambda D_H(p))] - cm_a \Big) \\
&< \Big((R_H + cm_a)[1 - \nu(\Lambda D_H^*)][1 - \beta(\Lambda D_H^*)] - cm_a \Big),
\end{aligned}$$

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

$$\begin{aligned}
&\Lambda D_H^* < \eta_{2_H} < \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H^*, \Lambda D_L^*)}{cm_a \frac{\beta'(\Lambda D_H^*)}{1 - \beta(\Lambda D_H^*)}} \right] \Big((R_H + cm_a)[1 - \nu(\Lambda D_H^*)][1 - \beta(\Lambda D_H^*)] - cm_a \Big) \\
&\Rightarrow -\hat{U}_H\big(\Lambda D_L^*, \Lambda D_H^*\big) \left[\frac{k_H - 1 + k_L \vartheta(\Lambda D_H, \Lambda D_L) + \nu(\Lambda D_H)}{cm_a \frac{\beta'(\Lambda D_H)}{1 - \beta(\Lambda D_H)}} \right] < 0 \\
&\Rightarrow \hat{U}_H\big(\Lambda D_L^*, \Lambda D_H^*\big) > 0.
\end{aligned}$$

However, this is a contradiction since $U_H\big(\Lambda D_L^*, \Lambda D_H^*\big) \leq 0$ 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 ($\Lambda^{R_L} \leq \Lambda < \Lambda^0$)**:** 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 (42) and (43), 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 (42)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 $\eta_{2_H} \leq \Lambda D_\ell(p) < \hat{D}_H(\Lambda)$ (since $\eta_{1_H} \geq 0$), and $R_H - p_H + cm_a = \frac{\Delta}{1-\beta(\hat{D}_H(\Lambda))}$, we have that

$$\begin{aligned}
&\hat{D}_H(\Lambda) > \left(\frac{k_H - 1 + k_L\vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))}{\frac{\Delta}{1-\beta(\hat{D}_H(\Lambda))}\beta'(\hat{D}_H(\Lambda))}\right)\Big((R_H + cm_a)[1 - \nu(\hat{D}_H(\Lambda))][1 - \beta(\hat{D}_H(\Lambda))] - \Delta\Big) \\
&\quad \Rightarrow (R_H + cm_a)[1 - \nu(\hat{D}_H(\Lambda))][1 - \beta(\hat{D}_H(\Lambda))] - \Delta\frac{k_H - 1 + k_L\vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda)) + \nu(\hat{D}_H(\Lambda))}{k_H - 1 + k_L\vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))} < 0 \\
&\quad \Rightarrow \Delta > \frac{(R_H + cm_a)\big[k_H - 1 + k_L\vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))\big]}{\frac{k_H + k_L\vartheta(\hat{D}_H(\Lambda), \hat{D}_L(\Lambda))}{1-\nu(\hat{D}_H(\Lambda))} - 1}[1 - \beta(\hat{D}_H(\Lambda))] = \Delta
\end{aligned}$$

Hence, we should have that $(R_H - p_\ell(p) + cm_a)[1 - \beta(\Lambda D_\ell(p))] \geq \Delta$.

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 ($\Lambda^{R_L}_H \leq \Lambda \leq \Lambda^{R_L}$): 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 (42) and (43), 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 (42)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

$$\begin{aligned}
&(R_H + cm_a - w_H)[1 - \nu(\Lambda/k_H)][1 - \beta(\Lambda/k_H)] - (R_L + cm_a)\frac{k_H - 1 + k_L\vartheta(\Lambda/k_H, 0) + \nu(\Lambda/k_H)}{k_H - 1 + k_L\vartheta(\Lambda/k_H, 0)} < 0 \\
&\qquad \Rightarrow \hat{U}_H(0, \Lambda/k_H) < R_L,
\end{aligned}$$

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

$$\begin{aligned}&(R_H + cm_a)[1 - \nu(\Lambda/k_H)][1 - \beta(\Lambda/k_H)] - (R_L + cm_a)\frac{k_H - 1 + \nu(\Lambda/k_H)}{k_H - 1} > 0\\ &\quad\Rightarrow z^{R_L}(\Lambda) > 0.\end{aligned}$$

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 $z^{R_L}(\Lambda) \leq 0$ for any $\Lambda \geq \Lambda_H^{R_L}$.

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 T.A. 5: Operational Efficiency (Non-Identical Agents)

T.A. 5.1. Proof of Lemma 8

Here, we only present the proof for $i = L,\ j = H$. For notational convenience, we let $r_i = R_i - p_i$ 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 (2005). Using this result, we have that

$$\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) = r_L > 0, \tag{49}$$

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

$$\begin{aligned}
&\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) \\
&\quad = \lim_{k\to\infty} \big[1 - P_{HL}(k)\big]\Big[(r_H + cm_a)\big[1 - \beta_H(k)\big] - cm_a\Big] + \lim_{k\to\infty} P_{HL}(k)(r_L) \\
&\quad \le \lim_{k\to\infty} \big[1 - P_{HL}(k)\big](r_H) + \lim_{k\to\infty} P_{HL}(k)(r_L) < r_L,
\end{aligned} \tag{50}$$

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 49 and 50, we have that

$$\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) < \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k)$$

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 $\liminf_{k\to\infty} \frac{\Lambda_k D_L(k)}{\alpha_L k} \leq 1$. Then, there exists a convergent subsequence such that

$$\liminf_{k\to\infty} \frac{\Lambda_k D_L(k)}{\alpha_L k} \leq 1.$$

Similar to the proof of Lemma 6.1, we have that

$$\lim_{k\to\infty} \beta_L(k) \leq \lim_{k\to\infty} \beta^{MM1}(\Lambda_k D_L(k), \alpha_L k) = 0,$$

where the last equality is due to Ward and Glynn (2005). Using this result, we have that

$$\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) = r_L > 0,$$

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, 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

$$\Lambda_k D_H(k) P_{HL}(k) + \Lambda_k D_L(k) = \alpha_L k \sigma_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) + \Lambda_k D_L(k)\beta_L(k).$$

Dividing both sides by Λ_k, the above equation implies that

$$P_{HL}(k)\big[D_H(k) + D_L(k)\big] \leq \frac{\alpha_L k}{\Lambda_k}\sigma_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) + D_L(k)\beta_L(k).$$

Letting k go to infinity, we have that

$$\lim_{k\to\infty} P_{HL}(k) \leq \frac{\alpha_L}{\rho} < 1.$$

Then, as in the proof of Part a, we have that

$$\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) < \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k)$$

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

$$\lim_{k\to\infty} D_H(k) = 1 - \lim_{k\to\infty} D_L(k) \geq 1 - \frac{\alpha_L}{\rho} > 0,$$

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 $\beta^{color}(k) \le P_{HL}(k)$. Thus, we have that

$$\beta_H(k) \ge \beta^M(\Lambda_k D_H(k);\alpha_H k) - P_{HL}(k).$$

Then, using this result and part ii, we have that

$$\liminf_{k\to\infty}\beta_H(k) \ge \lim_{k\to\infty}\beta^M(\Lambda_k D_H(k);\alpha_H k).$$

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 $D_L^{eq} = \max\left\{\min\left\{1, \left(\frac{r_L+cm_a}{r_H+cm_a}\right)\frac{\alpha_L}{\rho}\right\}, \min\left\{\frac{r_L+cm_a}{\bar{R}}\alpha_L, \left(\frac{r_L+cm_a}{cm_a}\right)\frac{\alpha_L}{\rho}\right\}\right\}$.

Liminf: 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

$$\begin{aligned}
\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) &= (r_L + cm_a)\left[1 - \lim_{k\to\infty}\beta_L(k)\right] - cm_a \\
&= (r_L + cm_a)\frac{\alpha_L}{\rho\tilde{D}_L} - cm_a \\
&> (r_L + cm_a)\frac{\alpha_L}{\rho D_L^{eq}} - cm_a \\
&\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \geq 0,
\end{aligned}$$

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. Furthermore, we have that

$$\begin{aligned}
\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho\tilde{D}_L}\right\} - cm_a \\
&\leq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho D_L^{eq}}\right\} - cm_a \\
&\leq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \\
&< \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k)
\end{aligned}$$

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 $\limsup_{k\to\infty} D_L(k) \leq \min\left\{1, \left(\frac{r_L+cm_a}{r_H+cm_a}\right)\frac{\alpha_L}{\rho}\right\}$. Note that $D_L^{eq} = 1$ when $\frac{r_L+cm_a}{r_H+cm_a} \geq \frac{\alpha_L}{\rho}$, 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

$$\begin{aligned}\lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) &= (r_L + cm_a)\frac{\alpha_L}{\rho \tilde{D}_L} - cm_a \\ &< \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{51}$$

Furthermore, letting $\tilde{D}_H = \lim_{k\to\infty} D_H(k)$ and using parts b.ii and b.iii, we have that

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho \tilde{D}_H}\right\} - cm_a \\ &\geq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho D_L^{eq}}\right\} - cm_a \\ &\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{52}$$

Combining (51) and (52), we have that

$$\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) > \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k)$$

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. <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

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho \tilde{D}_L}\right\} - cm_a \\ &> (r_H + cm_a)\frac{cm_a}{r_H + cm_a} - cm_a \\ &= 0 = \lim_{k\to\infty} U_2(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k), \end{aligned}$$

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 $\rho > \frac{\bar{R}}{cm_a}$ by part c. 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 $\liminf_{k\to\infty} D_H(k) \geq \left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$.

Limsup: 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

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_L(k); r_H, r_L; \alpha_H k, \alpha_L k) &= (r_H+cm_a)\frac{\alpha_H}{\rho\tilde{D}_H} - cm_a \\ &< (r_H+cm_a)\frac{cm_a}{r_H+cm_a} - cm_a = 0.\end{aligned}$$

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 $\limsup_{k\to\infty} D_H(k) \leq \left(\frac{r_H+cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}$.

T.A. 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 above. Then we prove the new version of Lemma 9. For notational convenience, we let $r_i = R_i - p_i$ 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 $0 < \rho < 1$.*

$$\begin{aligned}
D_H(k) &= D_1^{MCE}(rp_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1)\\
D_{L_{dev}}(k) &= D_2^{MCE}(rp_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1)\\
D_L(k) &= D_3^{MCE}(rp_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1)\\
P_{HL}(k) &= PServ_{13}(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1)\\
P_{HL_{dev}}(k) &= PServ_{12}(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1)\\
P_{L_{dev}L}(k) &= PServ_{23}(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1)
\end{aligned}$$

where $p_H < R_H$, $p_L \leq R_L$, 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. $\lim_{k\to\infty} \beta_1\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1\big) = \max\Big\{0, 1 - \frac{\alpha_H}{\rho \tilde{D}_H}\Big\}$,
   where $\tilde{D}_H = \lim_{k\to\infty} D_H(k)$.
6. $\lim_{k\to\infty} \beta_3\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1\big) = 1 - \frac{\alpha_L}{\rho \tilde{D}_L}$,
   where $\tilde{D}_L = \lim_{k\to\infty} D_L(k)$.
7. $\lim_{k\to\infty} D_L(k) = D_L^{eq}$,
   where $D_L^{eq} = \max\Big\{\min\Big\{1, \Big(\frac{r_L + cm_a}{r_H + cm_a}\Big)\frac{\alpha_L}{\rho}\Big\}, \min\Big\{\frac{r_L + cm_a}{\bar{R}}\alpha_L, \Big(\frac{r_L + cm_a}{cm_a}\Big)\frac{\alpha_L}{\rho}\Big\}\Big\}$.
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, and all high-quality providers charge p_H.

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

$$\lim_{k\to\infty} \beta_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1) \leq \lim_{k\to\infty} \beta^{MM1}(\Lambda_k D_L(k), \alpha_L k) = 0,$$

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 $\lim\limits_{k\to\infty} \frac{\Lambda_k D_L(k)}{k_L} \leq 1$ we have the above result due to Ward and Glynn (2005). Using this result, we have that

$$\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1) = r_L > 0, \tag{53}$$

which implies that utility of customers choosing the price p_L is strictly positive for large k, so that we should have $\lim\limits_{k\to\infty} D_H(k) + D_{H_{dev}L}(k) + D_L(k) = 1$ by the definition of Market Customer Equilibrium.

Now, we argue that $\lim\limits_{k\to\infty} P_{HL}(k)D_H(k) = \lim\limits_{k\to\infty} D_H(k)$. In order to prove that it is enough to show $\lim\limits_{k\to\infty} P_{HL}(k) = 1$ whenever $\lim\limits_{k\to\infty} D_{HL}(k) > 0$. To show that we suppose $\lim\limits_{k\to\infty} P_{HL}(k) < 1$ when $\lim\limits_{k\to\infty} D_{HL}(k) > 0$. Then, we have that

$$\begin{aligned}
&\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1) \\
&\qquad \leq \big[1 - \lim_{k\to\infty} P_{HL}(k)\big](r_L - \varepsilon) + \big[\lim_{k\to\infty} P_{HL}(k)\big](r_L) \\
&\qquad < r_L \\
&\qquad = \lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1).
\end{aligned}$$

However, this is a contradiction because customers are strictly better of by choosing sub-pool-3 for large k while $\lim\limits_{k\to\infty} D_{HL}(k) > 0$. Hence, we have that $\lim\limits_{k\to\infty} P_{HL}(k)D_H(k) = \lim\limits_{k\to\infty} D_H(k)$. Similarly, we can show that $\lim\limits_{k\to\infty} P_{L_{dev}L}(k)D_{L_{dev}}(k) = \lim\limits_{k\to\infty} D_{L_{dev}}(k)$.

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

$$\begin{aligned}
&\Lambda_k D_H(k) P_{HL}(k) + \Lambda_k D_{H_{dev}L}(k) P_{L_{dev}L}(k) + \Lambda_k D_L(k) \\
&\qquad = k_L \sigma_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1) \\
&\qquad\quad + \Lambda_k D_L(k) \beta_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1).
\end{aligned}$$

Dividing both sides by Λ_k and letting $k \to \infty$, the above equation implies that

$$\begin{aligned}
1 &= \lim_{k\to\infty} D_H(k) + D_{H_{dev}L}(k) + D_L(k) = \lim_{k\to\infty} D_H(k)P_{HL}(k) + D_{H_{dev}L}(k)P_{L_{dev}L}(k) + D_L(k) \\
&= \frac{\alpha_L}{\rho} \lim_{k\to\infty} \sigma_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1) \leq \frac{\alpha_L}{\rho} < 1,
\end{aligned}$$

where the second equality holds since $\lim\limits_{k\to\infty} P_{HL}(k)D_H(k) = \lim\limits_{k\to\infty} D_H(k)$, and $\lim\limits_{k\to\infty} P_{L_{dev}L}(k)D_{L_{dev}}(k) = \lim\limits_{k\to\infty} D_{L_{dev}}(k)$. The above inequality is a clear contradiction. Hence, we should have that $\liminf\limits_{k\to\infty} \frac{\Lambda_k D_L(k)}{k_L} > 1$.

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

$$\begin{aligned}\sum_{n=0}^{\alpha_L k} \pi_n &\leq \frac{\sum\limits_{n=0}^{\alpha_L k} \frac{\left[\Lambda_k (D_H(k)+D_{L_{dev}}(k)+D_L(k))\right]^{n-\alpha_L k} (\alpha_L k)!}{n!}}{\sum\limits_{n=0}^{\alpha_L k} \frac{\left[\Lambda_k (D_H(k)+D_{L_{dev}}(k)+D_L(k))\right]^{n-\alpha_L k} (\alpha_L k)!}{n!} + \sum\limits_{n=\alpha_L k+1}^{\infty} \frac{\left[\Lambda_k D_L(k)\right]^{n-\alpha_L k}}{\prod_{i=1}^{n-\alpha_L k} (k+i/m_a)}} \\ &\leq \frac{\sum\limits_{n=0}^{\alpha_L k} \frac{\left[\Lambda_k D_L(k)\right]^{n-\alpha_L k} (\alpha_L k)!}{n!}}{\sum\limits_{n=0}^{\alpha_L k} \frac{\left[\Lambda_k D_L(k)\right]^{n-\alpha_L k} (\alpha_L k)!}{n!} + \sum\limits_{n=\alpha_L k+1}^{\infty} \frac{\left[\Lambda_k D_L(k)\right]^{n-\alpha_L k}}{\prod_{i=1}^{n-\alpha_L k} (k+i/m_a)}} = \sum_{n=0}^{\alpha_L k} \pi_n^M,\end{aligned}$$

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

$$\lim_{k\to\infty} P_{HL}(k) + P_{L_{dev}L}(k) = \lim_{k\to\infty} \sum_{n=0}^{\alpha_L k} \pi_n \leq \lim_{k\to\infty} \sum_{n=0}^{\alpha_L k} \pi_n^M = 0,$$

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

$$\begin{aligned}\Lambda_k D_H(k) P_{HL_{dev}} + \Lambda_k D_{L_{dev}} &= \sigma_2\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1\big) \\ &\quad + \Lambda_k D_{L_{dev}} \beta_2\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1\big),\end{aligned}$$

which implies that

$$D_H(k) P_{HL_{dev}} = \frac{1}{\Lambda_k} + D_{L_{dev}} \left[\beta_2\big(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1\big) - 1\right] \leq \frac{1}{\Lambda_k}.$$

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

$$\begin{aligned}
&\lim_{k\to\infty} U_2(D_H(k),D_{L_{dev}}(k),D_L(k);r_H,r_L-\varepsilon,r_L;\alpha_H k,1,\alpha_L k-1)\\
&\qquad=(r_L+\varepsilon)\big[1-\lim_{k\to\infty}\beta_2(D_H(k),D_{L_{dev}}(k),D_L(k);r_H,r_L-\varepsilon,r_L;\alpha_H k,1,\alpha_L k-1)\big]-cm_a\\
&\qquad\le(r_L+\varepsilon)\big[1-\lim_{k\to\infty}\beta^M(\Lambda_k D_{L_{dev}}(k),1)\big]-cm_a=-cm_a,
\end{aligned}$$

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 0 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 $\liminf_{k\to\infty}\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H,r_L-\varepsilon,r_L;\alpha_H k,1,\alpha_L k-1)\ge 1-\frac{\alpha_H}{\rho}$. 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

$$\begin{aligned}
\Lambda_k D_H(k)\beta^M(\Lambda_k D_H(k),k_H) &= \Lambda_k D_H(k)\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H,r_L-\varepsilon,r_L;\alpha_H k,1,\alpha_L k-1)\\
&\quad+\Lambda_k D_H(k)\beta^{color}(k),
\end{aligned}$$

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

$$\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H,r_L-\varepsilon,r_L;\alpha_H k,1,\alpha_L k-1)\ge\beta^M(\Lambda_k D_H(k),k_H)-P_{HL_{dev}}(k)-P_{HL}(k).$$

Finally, we have that

$$\liminf_{k\to\infty}\beta_1(D_H(k),D_{L_{dev}}(k),D_L(k);r_H,r_L-\varepsilon,r_L;\alpha_H k,1,\alpha_L k-1)\ge\lim_{k\to\infty}\beta^M(\Lambda_k D_H(k),k_H)=1-\frac{\alpha_H}{\rho\tilde{D}_H},$$

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

$$\beta_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) = \left(\sum_{n=\alpha_L k+1}^{\infty} \pi_n\right)\left(1-\frac{\alpha_L k}{\Lambda_k D_L(k)}\right) + \pi_{\alpha_L k}.$$

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

$$\begin{aligned}
\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) &= (r_L + cm_a)\frac{\alpha_L}{\rho \tilde{D}_L} - cm_a \\
&> (r_L + cm_a)\frac{\alpha_L}{\rho D_L^{eq}} - cm_a \\
&\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \geq 0,
\end{aligned}$$

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

$$\begin{aligned}
\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) &= (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho\tilde{D}_L}\right\} - cm_a \\
&\leq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\} \\
&< \lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1)
\end{aligned}$$

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

$$\begin{aligned}\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) &= (r_L + cm_a)\frac{\alpha_L}{\rho\tilde{D}_L} - cm_a \\ &< \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{54}$$

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

$$\begin{aligned}\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) &\geq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho\tilde{D}_H}\right\} - cm_a \\ &\geq (r_H + cm_a)\min\left\{1, \frac{\alpha_H}{\rho - \rho D_L^{eq}}\right\} - cm_a \\ &\geq \min\left\{r_H, \max\left\{\frac{\bar{R}}{\rho} - cm_a, 0\right\}\right\}. \end{aligned} \tag{55}$$

Combining (54) and (55), we have that

$$\begin{aligned}&\lim_{k\to\infty} U_1(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) \\ &\quad > \lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1)\end{aligned}$$

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

$$\begin{aligned}\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L-\varepsilon, r_L; \alpha_H k, 1, \alpha_L k-1) &= (r_L + cm_a)\frac{\alpha_L}{\rho D_L^{eq}} - cm_a \\ &\geq \min\{\frac{\bar{R}}{\rho} - cm_a, r_H\} > 0.\end{aligned}$$

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

$$\lim_{k\to\infty} U_3(D_H(k), D_{L_{dev}}(k), D_L(k); r_H, r_L - \varepsilon, r_L; \alpha_H k, 1, \alpha_L k - 1) = 0$$

by using part 7. As before, we prove our claim by arguing that

$$\left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho} \leq \liminf_{k\to\infty} D_H(k) \leq \limsup_{k\to\infty} D_H(k) \leq \left(\frac{r_H + cm_a}{cm_a}\right)\frac{\alpha_H}{\rho}. \tag{56}$$

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, (56) 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 $\rho \geq \frac{\bar{R}}{cm_a}$.

T.A. 5.3. Proof of Lemma 10

For notational convenience, let $\lambda^{\Delta}_{\varepsilon}(p;R) = \lambda^{\Delta}(p+\varepsilon;R+\varepsilon)$, and $\Delta_{\varepsilon}(p;R) = \Delta(p+\varepsilon;R+\varepsilon)$. Note that $\Delta_{\varepsilon}(p;R) = \Delta(p;R)$

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

$$\begin{aligned}
p &> (R+cm_a)\lambda^{\Delta}(p;R)[1-\beta(\lambda^{\Delta}(p;R))] - \lambda^{\Delta}(p;R)(\Delta(p;R)+cm_a) \\
&\geq (R+cm_a)\lambda^{\Delta}_{\varepsilon}(p;R)[1-\beta(\lambda^{\Delta}_{\varepsilon}(p;R))] - \lambda^{\Delta}_{\varepsilon}(p;R)(\Delta_{\varepsilon}(p;R)+cm_a) \\
&\geq (R+\varepsilon+cm_a)\lambda^{\Delta}_{\varepsilon}(p;R)[1-\beta(\lambda^{\Delta}_{\varepsilon}(p;R))] - \lambda^{\Delta}_{\varepsilon}(p;R)(\Delta_{\varepsilon}(p;R)+cm_a) \\
&\quad -\varepsilon\lambda^{\Delta}_{\varepsilon}(p;R)[1-\beta(\lambda^{\Delta}_{\varepsilon}(p;R))] \\
&\geq (R+\varepsilon+cm_a)\lambda^{\Delta}_{\varepsilon}(p;R)[1-\beta(\lambda^{\Delta}_{\varepsilon}(p;R))] - \lambda^{\Delta}_{\varepsilon}(p;R)(\Delta_{\varepsilon}(p;R)+cm_a) - \varepsilon,
\end{aligned}$$

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

$$p+\varepsilon > (R+\varepsilon+cm_a)\lambda^{\Delta}_{\varepsilon}(p;R)[1-\beta(\lambda^{\Delta}_{\varepsilon}(p;R))] - \lambda^{\Delta}_{\varepsilon}(p;R)(\Delta_{\varepsilon}(p;R)+cm_a),$$

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

Appendix T.A. 6: Communication Enabled Model (Non-Identical Agents)

T.A. 6.1. Proof of Lemma 11

For notational convenience we let $r_i = R_i - p_i$ for $i \in \{H, L\}$, $r' = R_H - p'$, and

$$\begin{aligned}
D_L(k) &= D_1^{MCE}(r_L, r', r_H; k_L, \lfloor \delta k \rfloor, k_H - \lfloor \delta k \rfloor)\\
D'_H(k) &= D_2^{MCE}(r_L, r', r_H; k_L, \lfloor \delta k \rfloor, k_H - \lfloor \delta k \rfloor)\\
D_H(k) &= D_3^{MCE}(r_L, r', r_H; k_L, \lfloor \delta k \rfloor, k_H - \lfloor \delta k \rfloor)\\
P_{LH'} &= PServ_{12}(D_L(k), D'_H(k), D_H(k); r_H, r', r_L; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor)\\
P_{LH} &= PServ_{13}(D_L(k), D'_H(k), D_H(k); r_H, r', r_L; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor)\\
P_{H'H} &= PServ_{23}(D_L(k), D'_H(k), D_H(k); r_H, r', r_L; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor).
\end{aligned}$$

Before proving the above claim, we want to note that $\lim_{k\to\infty} D_H(k) > \frac{\alpha-\delta}{\rho}$, 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 $\liminf_{k\to\infty} D'_H(k) \geq \frac{\delta}{\rho}$. In order to prove that by contradiction, we suppose $\liminf_{k\to\infty} D'_H(k) < \frac{\delta}{\rho}$. Then, there exists a convergent subsequence of $D'_H(k)$ such that

$$\tilde{D}'_H := \lim_{k\to\infty} D'_H(k) < \frac{\delta}{\rho}.$$

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

$$\begin{aligned}
&\lim_{k\to\infty} U_2(D_L(k), D'_H(k), D_H(k); r_L, r', r_H; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor)\\
&\qquad \geq r' > R_L > 0,
\end{aligned}$$

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. (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 $\lim_{k\to\infty} D_H(k) \geq 1-\delta/\rho$. However, this leads to the following contradiction:

$$\begin{aligned}
&\lim_{k\to\infty} U_3(D_L(k), D'_H(k), D_H(k); r_L, r', r_H; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor) \\
&\qquad = (r_H + cm_a)\frac{\alpha_H - \delta}{\rho - \delta} - cm_a < (r_H + cm_a)\frac{\alpha_H}{\rho} - cm_a \\
&\qquad < r' \leq \lim_{k\to\infty} U_2(D_L(k), D'_H(k), D_H(k); r_L, r', r_H; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor).
\end{aligned}$$

Hence, we should have that $\liminf_{k\to\infty} D'_H(k) \geq \frac{\delta}{\rho}$. 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

$$\begin{aligned}
\liminf_{k\to\infty} V'_H(k) &\geq p' \liminf_{k\to\infty} \frac{\Lambda_k D'_H(k)}{\lfloor \delta k \rfloor}\left[1 - \liminf_{k\to\infty} \beta_2(D_L(k), D'_H(k), D_H(k); r_L, r', r_H; \alpha_L k, \lfloor \delta k \rfloor, \alpha_H k - \lfloor \delta k \rfloor)\right] \\
&= p'.
\end{aligned}$$

Finally, the result holds since $V'_H(k) \leq p'$ by definition.