7-1-2012

Service and price competition when customers are naive

Li LI
Hong Kong University of Science and Technology

Li JIANG
Hong Kong Polytechnic University

Liming LIU
Lingnan University, Hong Kong

Recommended Citation

Li, L., Li, J., & Liu, L. (2012). Service and price competition when customers are naive. Production and Operations Management, 21(4), 747-760. doi: 10.1111/j.1937-5956.2011.01304.x

Service and Price Competition When Customers Are Naive

Li LI
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Bay, Hong Kong SAR, China, ielli@ust.hk

Li JIANG
Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom Hong Kong SAR, China, lgtjiang@polyu.edu.hk

Liming LIU
Faculty of Business, Lingnan University, Tuen Mun, Hong Kong SAR, China, limingliu@ln.edu.hk

Abstract
We consider a system of two service providers each with a separate queue. Customers choose one queue to join upon arrival and can switch between queues in real time before entering service to maximize their spot utility, which is a function of price and queue length. We characterize the steady-state distribution for queue lengths, and then investigate a two-stage game in which the two service providers first simultaneously select service rates and then simultaneously charge prices. Our results indicate that neither service provider will have both a faster service and a lower price than its competitor. When price plays a less significant role in customers' service selection relative to queue length or when the two service providers incur comparable costs for building capacities, they will not engage in price competition. When price plays a significant role and the capacity costs at the service providers sufficiently differ, they will adopt substitutable competition instruments: the lower-cost service provider will build a faster service and the higher-cost service provider will charge a lower price. Comparing our results to those in the existing literature, we find that the service providers invest in lower service rates, engage in less intense price competition, and earn higher profits, while customers wait in line longer when they are unable to infer service rates and naive in service selection than when they can infer service rates to make sophisticated choices. The customers' jockeying behavior further lowers the service providers' capacity investment and lengthens the customers' duration of stay.

Key words: Queues, customer behavior, price and service rate competition, two-stage game
History: Originally submitted: January 2010; First revision submitted: November 2010; Second revision submitted: April 2011; Accepted: June 2011.

1. Introduction

The service sector encompasses a vast spectrum of business activities including hospitality, trade, health care, communication, and more. A recent report by EconomyWatch estimates that it will account for 63.4% of global GDP by the end of 2011. The continuously growing market and expanding service networks expose service providers to fierce competition, for which service speed and price, in their different forms, are two of the frequently used instruments. It is imperative for the managers to understand their interplay to align operations and marketing initiatives. In academic literature, service systems have long attracted the interest of researchers. Service rate has been a main decision variable, with the expected waiting time as the primary measure for allocating customers into service. Gilbert and Weng (1998), for instance, consider a setting where the customers are allocated to servers by a system manager to equalize expected sojourn times at the two servers. Price has often been another factor in customer allocation. Li and Lee (1994) analyze price competition in a two-server system where the customer utility is modeled as a function of price and waiting time. Notably in the existing literature, service rate directly affects the allocation of customers.

In practice, a customer often chooses a queue to join upon arrival according to the prices that service providers charge and actual queue lengths. Without a prior knowledge or observing over a period of time carefully, the customer is usually unable to estimate service speeds upon arrival to make choices. Also common is the phenomenon of customer jockeying. The changing of relative queue length often acts as a trigger for the customers to jockey, particularly when it is difficult for them to infer service speeds with sufficient confidence from service completions in the past. To the customer at the end of the queue that becomes relatively longer, the shortening of the other queue can be an indication of the difference in service speeds at the service providers to drive their jockeying behavior. Thus, service speed is not frequently used in practice for queue selection, although some imprecise estimate of it may drive jockeying.

The competition of gas stations located at close proximity to each other provides a compelling example. The price at every gas station is easily observable. Drivers can check the posted prices and take queue lengths into consideration in selecting the station. When stations are comparably crowded, it is hard to tell which one has a faster service. While waiting, the drivers at the end of the line for one gas station can readily move the vehicle to another gas station, when the line of the latter seems to shorten faster. The service at the gas stations is fast relative to the arrival rate of drivers. This makes it impractical for the drivers to observe enough service completions to infer service speeds before being served by a particular station. On the other hand, the waiting lines at gas stations are usually not long, and it is fair to argue that the drivers who intend to fill the gas will complete service at one station in the end.

In this paper, we consider a system of two service providers each with a separate queue. Customers join one queue upon arrival and can switch in real time between queues to maximize their spot utility that is a function of queue length and price. We assume the existence of a *maximum price*, such that no customer will patronize a service provider if it charges a price higher than the maximum price, even when no customer is waiting for service. Customers do not incur costs to switch between queues. We apply balance equations to capture the state transitions and derive the steady-state performance for given service rates and prices, and then explore service and price competition between the service providers in a two-stage game. As service capacity takes time to build and its investment is made early, the two service providers simultaneously invest in service in the first stage. Their service rates, once built, are known to each other. Then, they simultaneously announce their respective prices in the second stage.

We characterize the equilibrium for the above game setting. This requires us to tackle a non-trivial system with both continuous and discrete elements. Our results indicate that no service provider will simultaneously have a lower price and a faster service than its competitor. When price plays a less significant role in customers' service selection relative to queue length or when the two service providers incur comparable costs for building capacity, they will not engage in price competition and will each charge the maximum price. Otherwise, the two service providers will adopt substitutable competition instruments: the lower-cost service provider will build a faster service while its higher-cost competitor will charge a lower price. As the capacity cost of one service provider increases, this particular provider will lower service speed and its competitor may reduce service speed as well.

Comparing the results in our model to those in the existing literature, we find that the service providers build slower service, engage in less intense price competition, but earn higher profits when the customers are unable to infer service rates and hence naive in choosing service based on queue length and price than when they incorporate service rates to make sophisticated service selections; the customers on the other hand are expected to spend longer time waiting in queue. Such effects are more significant when capacity costs differ across the service providers.

Moreover, customers' jockeying behavior further lowers service providers' capacity investment and lengthens the customers' stay in the system. The service system is more efficient when the customers can switch queues in real time, since that will reduce the chance for idle server and positive queue length to coexist.

The remainder of this paper is organized as follows. We review literature in section 2. In section 3, we introduce the model, discuss customers' service selection, and derive the steady-state distribution for the queue lengths. We analyze price competition for given service rates in section 4; and explore service competition in section 5. In section 6, we conduct a performance comparison between our model and those in the existing literature. Finally, we conclude the paper in section 7.

2. Literature Review

The Join-the-Shortest queue problem has been studied extensively in the queueing literature. Haight (1958) derives the steady-state distribution for a system of two servers when arriving customers join the shorter queue. Koenigsberg (1966) allows customers to jockey when queue lengths differ by more than one. His model is later generalized by Zhao and Grassmann (1990). Nakamura (1989) adds service speed as a second attribute for customers' choice, i.e., an arriving customer joins the queue of the faster server when the difference between the two queue lengths is below certain threshold.

In the existing literature on competitive two-server systems, there are in general two ways to model customers' behaviors in service selection. Some works consider customers' *collective* behaviors, with the presence of a system manager to allocate customers to servers to achieve certain long term goal, and, each individual customer, once assigned to a server, stays with it until service completion. Gilbert and Weng (1998) characterize equilibrium service rates when the customers are allocated to equalize the expected sojourn times at different servers. Bell and Stidham (1983) assume that customers are allocated to minimize their expected waiting time. Cachon and Zhang (2007) show that performance - based allocation can induce competition among the servers to improve delivery times. Other works assume that customers *individually* select queues. Li and Lee (1994) have customers choose service to maximize their utility as a function of price and expected delivery time, and allow them to jockey. So (2000) builds a multiplicative competitive interaction model for the demand at each server. In Shang and Liu (2011), customers are allocated by a multinomial logit rule in which customers' utility depends on announced waiting times and the rates the service providers can meet the announced waiting times in an oligopoly market with service rate and quality competition. Ha et al. (2003) consider two suppliers competing in price and delivery frequency, when the customers select servers according to a state-independent rule. Li (1992) investigates the role of inventory in time-based competition. Cachon and Harker (2002) develop a model where the customers are sensitive to price and time, and service providers face economies of scale.

Similar to that in LL (1994), we assume the customers select queues individually upon arrival and can switch in real time between queues before entering service. The fundamental difference between our model and LL (1994) is that the customers are unaware of the service rates and their utility is determined by price and queue length in our model, while the customers are aware of the service rates to derive expected sojourn time for service selection in LL (1994). While queue length is generally assumed observable and used in service selection in the existing literature and in this paper as well, Hassin (1986) analyzes one server that can suppress the information on queue length, leaving customers to decide whether to join the queue based on a known distribution of waiting times. Hassin and Haviv (1994) consider a two-server system in which information on queue length is not immediately available to the customers who, upon arrival, have the option to either purchase information or randomly select one line to join. They analyze the equilibrium to evaluate the value of information.

3. The Model

We consider a system of two service providers each with a separate queue. Customers arrive to the system according to a Poisson process with a constant rate of λ. The service time at service provider i follows an exponential distribution with rate μ_i, $i=1,2$. We call $\mu_1+\mu_2$ aggregate service rate and $\rho \equiv \frac{\lambda}{\mu_1+\mu_2}$ system load factor. Service provider i incurs a non-negative increasing cost $C_i(\mu_i)$ to attain service rate μ_i and charges p_i for its service. The marginal cost for each provider to serve its customers is normalized to zero. We let $\overline{P}$ be the *maximum price* that customers can accept, or the value of service perceived by the customers. No customer will select a service provider if it charges a price higher than $\overline{P}$, even when no customers are waiting in line. We say that the two service providers engage in price competition when at least one of them charges a price below $\overline{P}$.

Customers are *spot*-utility maximizers. We let U_i be the spot utility a customer obtains by choosing service provider i with price p_i and queue length n_i. When service speed or waiting time is unknown to the customers and hard to estimate, queue length is the only available proxy for waiting time. Assuming that the utilities a customer receives from the service of the two providers are identical and normalizing them to zero, we assume:

$$U_i = -\beta n_i - p_i \quad \text{for } i = 1, 2, \tag{1}$$

and call β the *marginal queue length disutility*. In case that the services are not completely substitutable (with different utilities), we can reflect it in the choice model (1) without changing the way in which price and queue length affect the choice. We conjecture that the main insights will not be significantly altered. In the existing literature, service rates have been prevalently adopted to allocate customers into service. See, for instance, LL (1994), So (2000), and etc. The reality, however, is more likely to lie somewhere between the customers knowing the service rates and they relying only on observations of the instantaneous queue lengths to make choices.

Informed of the prices and observing the queue lengths, an arriving customer selects one queue to join according to the following rule: He will join:

i) queue i if $U_i > U_j$, $i, j = 1,2$ and $i \neq j$;

ii) the queue for the service provider that charges a lower price if $U_1 = U_2$ but $p_1 \neq p_2$;

iii) the queue for either service provider with equal probability if $U_1 = U_2$ and $p_1 = p_2$.

Waiting customers can switch between queues to obtain a strictly higher spot utility. A jockeying customer, regardless of his position in the original queue, has to move to the end of the destination queue. Note that, when a chance arises for customers to jockey, it must happen after one service provider has completed a service and it must be the customer at the end of the queue to make the first move. So, while they are waiting in line, it is the change in the relative queue length, rooted in the different service rates by the two providers, that causes customers to jockey. We assume that the customers do not incur costs to switch from one queue to the other.

3.1. Steady-State Distribution

When $\mu_1 + \mu_2 \leq \lambda$ so that $\rho \geq 1$, the two service providers are unable to accommodate all the customers in the long run, so that infinitely many customers are expected to be waiting in each queue. We next analyze the situation when $\mu_1 + \mu_2 > \lambda$ so that $\rho < 1$, to derive the steady-state distribution for queue lengths by analyzing state transitions, taking into account of customers' queue selection upon arrival and jockeying behaviors.

We denote $\pi(n_1, n_2)$ as the joint probability for the state where n_i customers are waiting or being served by service provider i; and $\pi_i(n)$, $i = 1,2$, the marginal probability. Then,

$$\pi_1(n) = \sum_{n_2=0}^{\infty} \pi(n \;, n_2), \text{ and } \pi_2(n) = \sum_{n_1=0}^{\infty} \pi(n_1, n \;).$$

3.1.1. $p_1 = p_2$

When the two service providers charge identical prices, customers select service by the relative queue lengths. Note that the difference in queue lengths must be no larger than one. Suppose that, at some time, $n > 1$ customers are waiting in queue i, while $n+2$ customers in queue $j \neq i$. The customer at the end of queue j will have the incentive to switch to queue i, since such a move can earn him a strictly higher spot utility. Associated with his move, both queues will have the same length of $n+1$, and no customers will have further tendencies to jockeying. Consequently, the possible states are (n, n), (n, $n+1$), and ($n+1$, n) for $n \geq 0$. By use of difference equations to capture the state transitions[1], we can derive the steady-state distributions as follows:

$$\pi(0,0) = \frac{2\mu_1\mu_2(\mu_1 + \mu_2 - \lambda)}{\mu_1^2\lambda + \mu_2^2\lambda + 2\mu_1^2\mu_2 + 2\mu_2^2\mu_1}, \; \pi(1,0) = \frac{\lambda\mu_2(\mu_1 + \mu_2 - \lambda)}{\mu_1^2\lambda + \mu_2^2\lambda + 2\mu_1^2\mu_2 + 2\mu_2^2\mu_1},$$

$$\pi(0,1) = \frac{\lambda\mu_1(\mu_1 + \mu_2 - \lambda)}{\mu_1^2\lambda + \mu_2^2\lambda + 2\mu_1^2\mu_2 + 2\mu_2^2\mu_1};$$

$$\pi(n+1, n+1) = \pi(n,n)\rho^2, \; \pi(n+1,n) = \pi(n,n-1)\rho^2,$$

$$\pi(n,n+1) = \pi(n-1,n)\rho^2, \text{ for } n \geq 1.$$

The probability that service provider i is idle is:

$$\pi_i(0) = \frac{\mu_i(\lambda + 2\mu_j)(\mu_1 + \mu_2 - \lambda)}{\mu_1^2\lambda + \mu_2^2\lambda + 2\mu_1^2\mu_2 + 2\mu_2^2\mu_1}, \text{ for } i, j = 1,2 \text{ and } i \neq j \qquad (2)$$

3.1.2. $p_1 \neq p_2$

When service provider 1 charges a higher price than service provider 2, i.e., $p_1 > p_2$, an arriving customer will choose service provider 1 only when the queue at service provider 2 is sufficiently longer than that at service provider 1, i.e., $n_2 > n_1 + \beta^{-1}(p_1 - p_2)$. We let

[1] The detailed state transitions and balance equations can be obtained upon request.

$$B \equiv \left\lceil \beta^{-1} \cdot (p_1 - p_2) \right\rceil, \tag{3}$$

where $\lceil x \rceil$ is the smallest integer greater than x, be the *normalized premium* for service provider 1. B is the gap in queue length needed to equalize the price difference at the service providers. The two service providers charge identical prices when $B = 0$, but service provider 2 (1, resp.) charges a lower price and its competitor a higher price when $B > 0$ ($B < 0$, resp.).

By the same logic as that for the case when $p_1 = p_2$**,** we can argue that queue lengths at the two service providers will not differ by more than $B+1$ when $p_1 \neq p_2$. The steady-state distributions can be derived as follows:

$$\pi(0,n_2) = \frac{(\mu_2 - \lambda)\mu_2^{B+1}(\mu_1 + \mu_2 - \lambda)}{-\mu_1\lambda^{B+2} + \mu_2^{B+2}(\mu_1 + \mu_2 - \lambda)} \cdot \frac{\lambda^{n_2}}{\mu_2^{n_2}}, \text{ for } 0 \le n_2 \le B,$$
$$\pi(0,B+1)\mu_2 = \pi(0,B)(\lambda + \mu_2) - \pi(0,B-1)\lambda, \ \pi(1,B+1) = \pi(0,B+1)\rho;$$
$$\pi(n_1+1,n_1+B+1) = \pi(n_1,n_1+B)\rho^2, \ \pi(n_1,n_1+B+1) = \pi(n_1-1,n_1+B)\rho^2,$$
$$\pi(n_1,n_1+B) = \pi(n_1-1,n_1+B)\rho, \text{ for } n \ge 1.$$

The probabilities that the two service providers are idle are:

$$\pi_1(0) = \frac{(\mu_2^{B+2} - \lambda^{B+2})(\mu_1 + \mu_2 - \lambda)}{-\mu_1\lambda^{B+2} + \mu_2^{B+2}(\mu_1 + \mu_2 - \lambda)} \text{ and } \pi_2(0) = \frac{\mu_2^{B+1}(\mu_2 - \lambda)(\mu_1 + \mu_2 - \lambda)}{-\mu_1\lambda^{B+2} + \mu_2^{B+2}(\mu_1 + \mu_2 - \lambda)}. \tag{4}$$

We can apply a similar procedure as above to analyze the steady-state distribution when $p_1 < p_2$**.** The corresponding probabilities that the two service providers are idle are shown below:

$$\pi_1(0) = \frac{\mu_1^{-B+1}(\mu_1 - \lambda)(\mu_1 + \mu_2 - \lambda)}{-\mu_2\lambda^{-B+2} + \mu_1^{-B+2}(\mu_1 + \mu_2 - \lambda)} \text{ and } \pi_2(0) = \frac{(\mu_1^{-B+2} - \lambda^{-B+2})(\mu_1 + \mu_2 - \lambda)}{-\mu_2\lambda^{-B+2} + \mu_1^{-B+2}(\mu_1 + \mu_2 - \lambda)}. \tag{5}$$

With these steady-state distributions for given service rates and prices, we investigate the service and price competition between the two service providers.

4. Price Competition

We will analyze the equilibrium of the price competition between the two service providers for given service rates in this section; and explore service competition to complete the equilibrium characterization in the next section.

4.1. Profit Functions and Response Decisions

For given service rates at the two service providers, the price at a service provider i determines the revenue for each service it provides and competes against the price by its competitor to affect the initial service selection by the arriving customers.

If $\mu_1 + \mu_2 \le \lambda$, or $\rho \ge 1$, both service providers will be fully utilized, and the expected profit of service provider i is:

$$\Pi_i = p_i \cdot \mu_i - C_i(\mu_i) \text{ for } i = 1, 2. \quad (6)$$

It is obvious that each service provider will charge the highest price $\overline{P}$ the customers can accept. Then, price competition is not in effect when the market demand is so heavy that the system is fully congested.

We next consider the situation when $\rho < 1$, for which the distributions of the steady-state queue lengths were derived in section 3.1. The expected number of customers service provider i serves is $[1-\pi_i(0)]\mu_i$, $i = 1,2$, and we can write its expected profit as:

$$\Pi_i(p_i \mid p_j) = p_i \cdot [1-\pi_i(0)] \cdot \mu_i - C_i(\mu_i) = p_i \cdot f_i(B) \cdot \mu_i - C_i(\mu_i), \quad (7)$$

where, $f_i(B) \equiv 1-\pi_i(0)$ is the rate at which service provider i is busy. We refer to $f_i(B)$ as the *occupancy rate* at service provider i, and obtain its expression by (2)-(5) for different values of B. Note that $f_i(B)$ is a discrete function. Given service rates, it captures the effects of price competition on the allocation of customers and the market shares of the service providers. In Lemma A1 in the appendix, we show that $f_1(B)$ decreases but $f_2(B)$ increases in B. That is, one service provider will be less patronized when it raises price. Moreover, $f_i(B)$ for $i = 1, 2$ decreases in both μ_1 and μ_2. Then, the occupancy rate at one service provider will drop as either service provider speeds up. A faster service at one provider will shorten the service time at this particular provider, and its resultant shorter waiting line will attract those customers that may have otherwise chosen the other service provider, thus downsizing the market share and lowering the occupancy rate of its competitor.

Given $p_2 \in [0, \overline{P})$, service provider 1 can charge any price in $(p_2 + \beta(B-1), p_2 + \beta B]$ to achieve a normalized premium of B, with a resultant occupancy rate of $f_1(B)$. To attain the best unit revenue, it will obviously fix price at $p_1 = p_2 + \beta B$. Its decision of finding the best price is then equivalent to that of finding the best normalized premium. On the other hand, given service provider 1's price p_1, service provider 2 can apply a similar logic to make price decision, except that it will *mark down* its price relative to p_1 and charge $p_2 = p_1 - \beta B$ to achieve a normalized premium of $B \in I$, where I is the set of integers. Later in this paper, we let I^+ and I^- denote, respectively, the sets of non-negative and non-positive integers.

With the occupancy rate defined as above, we next define, for given normalized premium $B \in I$,

$$H_i(B) \equiv \beta(B-1)\cdot \varepsilon_i^f(B-1), \text{ for } i = 1,2,$$

where,
$$\varepsilon_i^f(B-1) = \frac{Bf_i(B)-(B-1)f_i(B-1)}{(B-1)f_i(B-1)}\cdot\frac{f_i(B-1)}{\left|f_i(B-1)-f_i(B)\right|} \tag{8}$$

is the percentage change in $Bf_i(B)$ relative to that in $f_i(B)$ as the normalized premium increases. $H_i(B)$ is the marginal gain in revenue to service provider i with respect to a marginal change in its occupancy rate, corresponding to an increase in the normalized premium from $B-1$ to B. The effect of a larger value of B on $H_i(B)$ is affected by the service rate at provider $j \neq i$ as well as the value of the prevailing normalized premium B.

By (7) and the steady-state distributions derived in section 3.1, we can write the expected profit of service provider 1 for given $p_2 \in [0, \bar{P})$, $\Pi_1(p_1 \mid p_2)$, as:

$$\Pi_1(p_1 \mid p_2) = \begin{cases} p_1\cdot\left[\dfrac{\mu_2\lambda^{-B+2}+\lambda\mu_1^{-B+1}(\lambda-\mu_1-\mu_2)}{\mu_2\lambda^{-B+2}+\mu_1^{-B+2}(\lambda-\mu_1-\mu_2)}\right]\cdot\mu_1 - C_1(\mu_1) & p_1\in[0,p_2) \\ p_1\cdot\left[\dfrac{(\lambda-\mu_2)\lambda^{B+2}}{\mu_1\lambda^{B+2}+\mu_2^{B+2}(\lambda-\mu_1-\mu_2)}\right]\cdot\mu_1 - C_1(\mu_1) & p_1\in[p_2,\bar{P}] \end{cases}, \tag{9}$$

where B, as defined in (3), is the normalized premium determined by the prices at both providers. The shape of $\Pi_1(p_1 \mid p_2)$ is affected by the service rate at service provider 2, as shown in Lemma 1.

Lemma 1: *For given* $p_2 \in [0, \bar{P})$,

1. *when* $\mu_2 > \lambda$, *there exists a* B_1 *with* $H_1(B_1+1) < p_2 \le H_1(B_1)$ *such that* $\Pi_1(p_1 \mid p_2)$ *increases in* p_1 *for* $p_1 \in [0, p_2+\beta B_1]$; *but decreases for* $p_1 \in (p_2+\beta B_1, \bar{P}]$;
2. *when* $\mu_2 < \lambda$, *there exist* $B_1^L \le 0 < B_1^U$, $H_1(B_1^U) < p_2 \le H_1(B_1^U+1)$, $H_1(B_1^L+1) < p_2 \le H_1(B_1^L)$, *such that* $\Pi_1(p_1 \mid p_2)$ *increases in* p_1 *for* $p_1 \in [0, p_2+\beta B_1^L]$, *decreases for* $p_1 \in [p_2+\beta B_1^L, p_2+\beta B_1^U]$; *and increases for* $p_1 \in (p_2+\beta B_1^U, \bar{P}]$.

When $\mu_2 > \lambda$, service provider 2 is fast enough to accommodate all the customers, price is the main weapon for service provider 1 to grab market share by influencing customers' service selection at arrival. Given service provider 2'price, p_2, service provider 1 chooses normalized premium B. A higher value of B will cause it to lose market share but boost its price. Initially, the price increase is enough to offset the loss in market share and earn service provider 1 a better revenue. As its price is sufficiently high, however, a further increase in the normalized premium will make the gain in price be eroded by the loss in market share. B_1, as defined in Lemma 1, is a threshold normalized premium above which service provider 1 ceases to reap in more profit by further raising price. Notice that B_1 is affected by service provider 2's price p_2 as the reference for service provider 1 to charge premium.

Suppose that service provider 2 is slow in service, i.e., $\mu_2 < \lambda$. When service provider 1 charges a lower price than service provider 2, the effect of a higher normalized premium on its profit is similar to that when $\mu_2 > \lambda$, with $B_1^L \leq 0$ replacing B_1 as the threshold value for the normalized premium. When service provider 1 charges a higher price than its competitor, as it raises price, its price increase will initially render a loss in market share that is unrecoverable by the gain in unit revenue. When p_1 is high enough, $p_1 > p_2 + B_1^U$, the loss in market share due to a higher price will be negligible as relative to the price increase itself to render a higher profit to service provider 1.

Best-Response

We next derive the best-response prices for the two service providers, for given service rates, and investigate the equilibrium of price game. To simplify expressions, we introduce four thresholds values for the normalized premium. Let $\overline{B}_0$ and $\underline{B}_P$ be the best normalized premiums for service provider 1 when service provider 2 charges a zero price and the maximum price $\overline{P}$, respectively. They establish the upper and lower bounds for the best-response normalized premium of service provider 1. Similarly, let $\underline{B}_0$ and $\overline{B}_P$ be the best normalized premiums for service provider 2 when service provider 1 charges a zero price and $\overline{P}$, respectively. They bound service provider 2's best-response normalized premium. Their formal definitions are given in the Appendix after the proof for Lemma1. Proposition 1 characterizes the best-response price by service provider 1 for given p_2, where, by convention, $x \wedge y \equiv \min\{x, y\}$ and $x \vee y \equiv \max\{x, y\}$.

Proposition 1: *Suppose that the two service providers have service rates μ_1 and μ_2, for given price p_2 by service provider 2, the best-response price of service provider 1 is:*

1) When $\mu_2 > \lambda$,

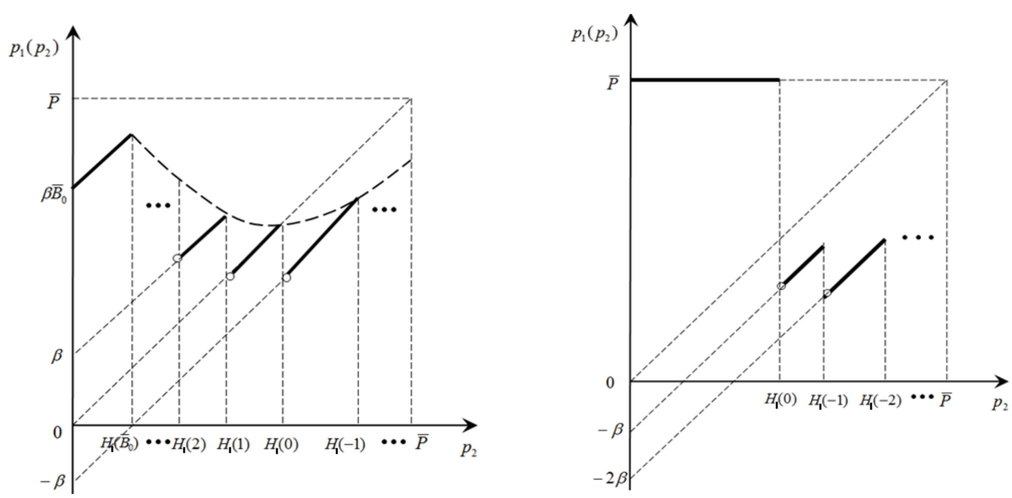
$$p_1(p_2) = \left(\left(p_2 + \beta \mathrm{B} \right) \wedge \bar{P} \right) \vee 0, \text{ where } H_1(B+1) < p_2 \leq H_1(B) \text{ and } B \in [\underline{B}_P, \bar{B}_0].$$

2) When $\mu_2 < \lambda$,

$$p_1(p_2) = \begin{cases} \bar{P} & 0 < p_2 \leq H_1(0) \\ \vdots & \\ (p_2 + \beta B) \vee 0 & H_1(B+1) < p_2 \leq H_1(B) \quad \underline{B}_P \leq B < 0 \end{cases}.$$

3) $p_1(p_2) - p_2$ *decreases in* p_2.

Figure 1 illustrates service provider 1's best-response curves in highlighted solid lines. Observe that the curves are piece-wise linear in p_2 and discontinuous with breakpoints at $H_1(B)$ for $B \in I$. Recall that, when service provider 2 is fast enough, i.e., $\mu_2 > \lambda$, price is the main competition instrument for service provider 1. As shown in Figure 1.a), when service provider 2 charges a low price, i.e., $0 < p_2 \leq H_1(0)$, service provider 1 is safe to charge a higher price, with the negative effect from the resultant smaller market share outweighed by the benefit from a higher unit revenue. As p_2 increases, the room for service provider 1 to charge a higher price will be smaller and its best normalized premium will drop. When p_2 is high, i.e., $p_2 > H_1(0)$, service provider 1 will charge a lower price than its competitor to grasp a larger market share without sacrificing the unit revenue too much.



a) $\mu_2 > \lambda$ *b)* $\mu_2 < \lambda$

Figure 1. Best Response Curve of Service Provider 1

If service provider 2 does not have enough service capacity, i.e., $\mu_2 < \lambda$, service provider 1 has the option to balance between the two competition instruments in service and price. When service provider 2 charges a low price, i.e., $0 < p_2 \leq H_1(0)$, charging an even lower price than

p_2 will hurt the unit revenue for service provider 1, with a negligible effect on its market share. Instead, as shown in Figure 1.b), service provider 1 will forfeit price as a competition instrument by charging the maximum price and rely solely on service speed. Otherwise, i.e., $p_2 > H_1(0)$, it will have the incentive to charge a lower price to affect customers' initial service selection, but the value of normalized premium is affected by $H_1(\cdot)$. The jumps and hence discontinuity in the best-response curves for service provider 1 are attributed to the discrete nature of the normalized premium. A specific best normalized premium for service provider 1 applies to a range of prices charged by service provider 2. p_1 will linearly increase as p_2 increases in this range. As p_2 crosses the boundary of the range, however, the best-response normalized premium will decrease by one to cause a drop in p_1.

An important property of the best response price is that service provider 1's normalized premium, $\lceil (p_1(p_2) - p_2)/\beta \rceil$, decreases with p_2, reaching maximum value $\overline{B}_0$ at $p_2 = 0$ and the minimum value $\underline{B}_P$ at $p_2 = \overline{P}$. When service provider 2 charges a low price, the negative effect of a smaller market share is negligible on the profit of service provider 1, who will charge a high price to earn a high unit revenue. When service provider 2 bids higher, lowering the normalized premium can attract more customers for service provider 1 and its absolute price may not significantly drop. Charging a lower price will become a more attractive option to service provider 1 as its competitor raises price. The best-response price of service provider 2 for given service provider 1's price can be obtained by symmetry.

4.2 Price Equilibrium For Given Service Rates

With the best-response by the two service providers, we can establish the existence and give the specific forms of the equilibrium of the price game, which are characterized in Theorem 1.

Theorem 1: *Suppose the two service providers have service rates μ_1 and μ_2 respectively. The equilibrium of the price competition game, (p_1^*, p_2^*), is unique:*

1) *When $\mu_1 > \lambda$ and $\mu_2 > \lambda$,*
   $p_1^* = 0 \vee ((H_1(B) + \beta B) \wedge H_2(B+1) \wedge \overline{P})$ *and* $p_2^* = p_1^* - \beta B$ *, if* $Y(B) > 0$ *and* $Y(B+1) < 0$ *for* $\underline{B} \le B \le \overline{B}$*, where* $Y(B) = H_1(B) + \beta B - H_2(B)$, $\underline{B} \equiv (\underline{B}_0) \vee (\underline{B}_P)$*, and* $\overline{B} \equiv \overline{B}_0 \wedge \overline{B}_P$.
2) *When $\mu_1 > \lambda$ and $\mu_2 < \lambda$,*
   $p_1^* = \overline{P}$ *and* $p_2^* = (\overline{P} - \beta \overline{B}_P) \vee 0$*, if* $\overline{P} - \overline{B}_P \le H_1(0)$ *and* $H_2(\overline{B}_P) < \overline{P} \le H_2(\overline{B}_P + 1)$.
3) *When $\mu_1 < \lambda$ and $\mu_2 > \lambda$,*

$p_1^* = (\overline{P} + \beta \underline{B}_P) \vee 0$ *and* $p_2^* = \overline{P}$, *if* $\overline{P} + \beta \underline{B}_P \le H_2(1)$ *and* $H_1(\underline{B}_P + 1) \le \overline{P} \le H_1(\underline{B}_P)$.

4) *When* $\mu_1 < \lambda$ *and* $\mu_2 < \lambda$,

a. $p_1^* = \overline{P}$ *and* $p_2^* = (\overline{P} - \beta \overline{B}_P) \vee 0$, *if* $\overline{P} - \overline{B}_P \le H_1(0)$ *and* $H_2(\overline{B}_P) < \overline{P} \le H_2(\overline{B}_P + 1)$.

b. $p_1^* = p_2^* = \overline{P}$, *if* $\overline{P} \le H_2(1)$ *and* $\overline{P} \le H_1(0)$.

c. $p_1^* = (\overline{P} + \beta \underline{B}_P) \vee 0$ *and* $p_2^* = \overline{P}$, *if* $\overline{P} + \beta \underline{B}_P \le H_2(1)$ *and* $H_1(\underline{B}_P + 1) < \overline{P} \le H_1(\underline{B}_P)$.

Theorem 1 reveals that the equilibrium for the price game is crucially affected by service rates; and has embedded in it a critical relationship between the two competition instruments, as shown in Proposition 2.

Proposition 2: *In the equilibrium of price competition,* $p_i^* > p_j^*$ *only if* $\mu_i > \mu_j$ *, for* $i, j = 1,2$ *and* $i \ne j$ *; and* $p_1^* = p_2^* = \overline{P}$ *if* $\mu_1 = \mu_2$ *.*

Hence, it is not to the best interest of an individual service provider to simultaneously charge a lower price and build a faster service than its competitor. Price is substitutable to service, and is adopted as one competition instrument only when service providers build different service speeds. In particular, an individual service provider will charge a higher price when it is faster than its competitor. When both service providers have identical service rates, they will not engage in price competition and will each charge the maximum price of $\overline{P}$.

We next explore two scenarios with respect to the service rates at the service providers: one in which each service provider has enough capacity to serve the entire market, and the other in which at least one service provider does not have enough capacity. As we will show later, the two service providers engage in price competition only in the first case.

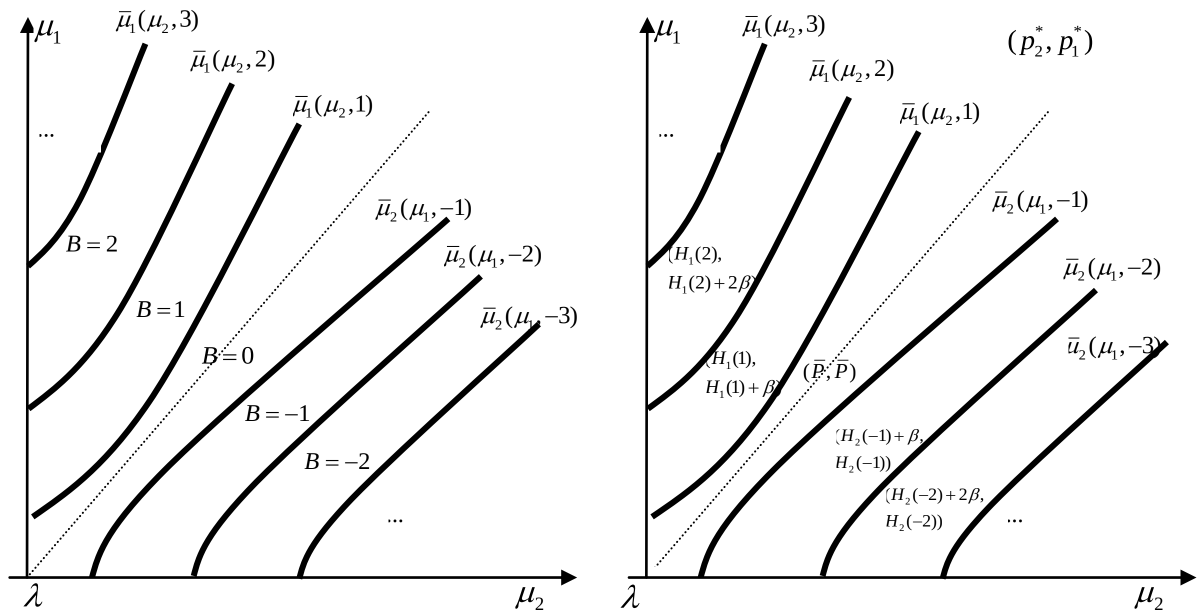
*Both Service Providers Have Enough Capacity***:** $\mu_1 > \lambda$ and $\mu_2 > \lambda$

As shown in part 1) of Theorem 1, the function $Y(B)$ plays a key role to determine equilibrium outcomes. Let us treat the normalized premium B as a continuous variable for the moment, and assume that the first-order derivative of $f_i(B)$ for $i = 1,2$ exists and is continuous. For a given value of B , $H_1(B)$ is *the* price by service provider 2 to which service provider 1's best price is $p_1 = H_1(B) + \beta B$; and $H_2(B)$ *the* price by service provider 1 to which service provider 2's best decision is to accept B as the normalized premium. For a specific normalized premium B to arise in equilibrium, $H_1(B) + \beta B$ and $H_2(B)$ should equal, or $Y(B) = 0$, to make the incentives of the two providers be consistent with each other. The conditions on $Y(B)$ in Theorem 1 attend to the discrete nature of B .

Lemma 2: *Let* $Y(B)$ *be as defined in Theorem 1. When* $\mu_1 > \lambda$ *and* $\mu_2 > \lambda$*:*

1. $Y(B)$ *decreases in* B*, increases in* μ_1*, and decreases in* μ_2*.*
2. *When* $\mu_1 > \mu_2 > \lambda$*, for given* μ_2 *and* $B \in I^+$*, there exists* $\bar{\mu}_1(\mu_2, B)$ *at which* $Y(B) = 0$*; and* $\bar{\mu}_1(\mu_2, B)$ *increases in both* μ_2 *and* B*; when* $\mu_2 > \mu_1 > \lambda$*, for given* μ_1 *and* $B \in I^-$*, there exists* $\bar{\mu}_2(\mu_1, B)$ *at which* $Y(B) = 0$*; and* $\bar{\mu}_2(\mu_1, B)$ *increases in* μ_2*, but decreases in* B*.*

In Figure 2, we illustrate the equilibrium normalized premium and prices. In the subspace of $\mu_i > \mu_{3-i} > \lambda$ for $i = 1, 2$, the $\bar{\mu}_i(\mu_{3-i}, B)$'s are the threshold curves, as defined in part 2) of Lemma 2, for the service rate at supplier i, given service rate at supplier $(3-i)$ and normalized premium B.



***a)* Normalized Premiums** ***b)* Prices**

Figure 2. Equilibrium Normalized Premium and Prices, $\mu_1 > \lambda$ **and** $\mu_2 > \lambda$

The discrete nature of the normalized premium causes the price equilibria to reside in strips. Consider the case when service provider 1 is faster than service provider 2, $\mu_1 > \mu_2 > \lambda$. By the definition of $\bar{\mu}_1(\mu_2, B)$ in Lemma 2, we can write the conditions on $Y(B)$ in part 1) of Theorem 1 as $\bar{\mu}_1(\mu_2, B+1) \geq \mu_1 > \bar{\mu}_1(\mu_2, B)$ for $B \in I^+$. Given the service rate at service provider 2 μ_2, the equilibrium normalized premium will be $B \in I^+$ and service provider 2 will charge a price of $H_1(B)$ to induce service provider 1 to accept that normalized premium, when the service rate at service provider 1 satisfies $\mu_1 \in (\bar{\mu}_1(\mu_2, B), \bar{\mu}_1(\mu_2, B+1)]$, i.e., when the two providers' service rates fall in the strip delimited by the two curves of $\bar{\mu}_1(\mu_2, B)$ and $\bar{\mu}_1(\mu_2, B+1)$ in Figure 2. As service provider 1 speeds up, i.e., μ_1 increases, service provider 2 will have a stronger incentive to lower price to attract more customers upon arrival; and the normalized premium will increase. The analysis for the case $\mu_2 > \mu_1 > \lambda$ follows by symmetry. In this latter case, service provider 1 will charge a lower price, and $\bar{\mu}_2(\mu_1, B)$ will play the same role as that by $\bar{\mu}_1(\mu_2, B)$ when $\mu_1 > \mu_2 > \lambda$ to delimit the space for the equilibrium outcomes. $B \in I^-$ will arise as the equilibrium normalized premium when the two providers' service rates fall in the strip delimited by the two curves of $\bar{\mu}_2(\mu_1, B)$ and $\bar{\mu}_2(\mu_1, B-1)$ in Figure 2.

At least One Service Provider Lacks Enough Capacity

There are three possible scenarios: 1) $\mu_1 > \lambda$ and $\mu_2 < \lambda$; 2) $\mu_1 < \lambda$ and $\mu_2 > \lambda$; 3) $\mu_1 < \lambda$ and $\mu_2 < \lambda$. In the first case, service provider 2, due to its lower service rate, will charge a lower price, while service provider 1 will weigh the options of taking on one more layer of competition in price against purely relying a faster service. Part 2) of Theorem 1 shows that, in equilibrium, service provider 1 will not resort to price as a competition instrument and will charge maximum price, and the value of the normalized premium depends on the extent of service competition influenced by the service rates at both providers. To identify the specific equilibrium, we let

$$Z(\mu_1, \mu_2, \bar{B}_P) = (H_1(0) + \beta \bar{B}_P) \wedge H_2(\bar{B}_P + 1)\,. \tag{10}$$

The existence condition, as given in part 2) of Theorem 1, can be rewritten as

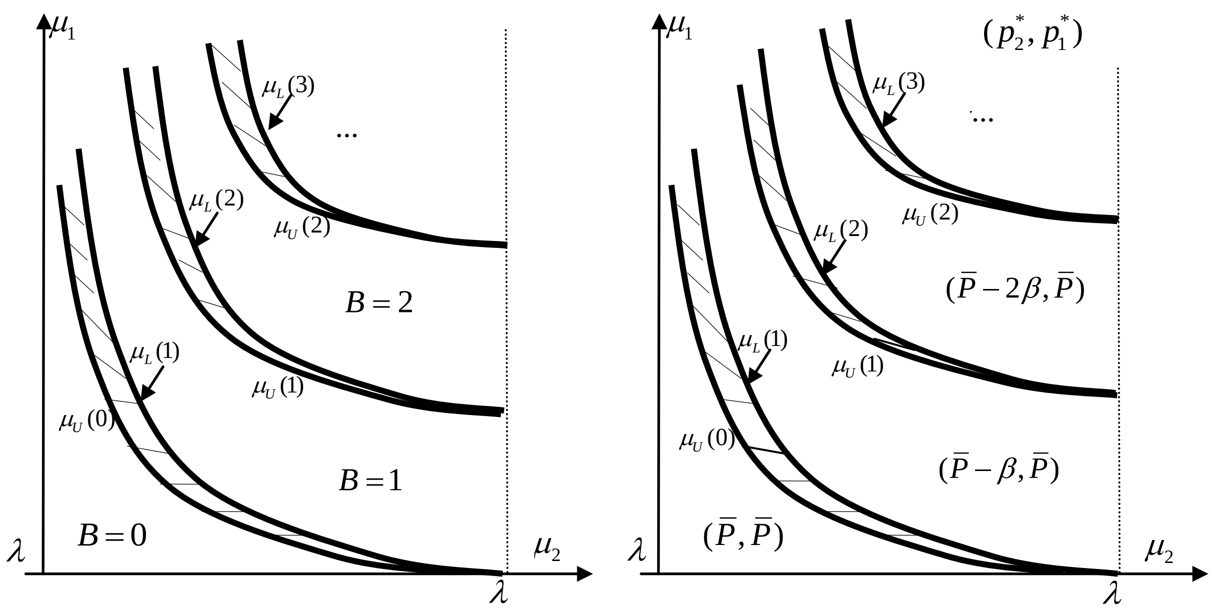
$$H_2(\bar{B}_P) < \bar{P} < Z(\mu_1, \mu_2, \bar{B}_P)\,. \tag{11}$$

Recall that $\bar{B}_P$ is service provider 2's best normalized premium when service provider 1 charges $\bar{P}$. $Z(\mu_1, \mu_2, \bar{B}_P)$, as defined in (10), essentially locates the value for $\bar{B}_P$.

Lemma 3: *Let* $Z(\mu_1, \mu_2, \bar{B}_P)$ *be as defined in* (10).

1. *It decreases in* μ_1 *and* μ_2.
2. *For given* $\bar{P}$ *and* μ_2, *there exist* $\mu_U(\bar{B}_P)$ *such that* $Z(\mu_1, \mu_2, \bar{B}_P) \geq \bar{P}$ *for* $\mu_1 \leq \mu_U(\bar{B}_P)$ *and* $Z(\mu_1, \mu_2, \bar{B}_P) < \bar{P}$ *otherwise; and* $\mu_L(\bar{B}_P)$ *such that* $H_2(\bar{B}_P) \geq \bar{P}$ *for* $\mu_1 \leq \mu_L(\bar{B}_P)$ *and* $H_2(\bar{B}_P) < \bar{P}$ *otherwise.* $\mu_L(\bar{B}_P)$ *and* $\mu_U(\bar{B}_P)$ *decrease in* μ_2 *and increase in* $\bar{B}_P$.

In Figure 3, on the space of (μ_1, μ_2) on $[0, \lambda] \times [\lambda, \infty)$, each heavy curve represents $\mu_L(B)$ or $\mu_U(B)$ for some $B \in I^+$, defined in part 2) of Lemma 3. When the service rates at the two service providers fall into the area delimited by $\mu_L(B)$ and $\mu_U(B)$, service provider 1 will charge the maximum price and its competitor will achieve a normalized premium of B. When the service rates fall into a shaded strip bounded by $\mu_U(B)$ and $\mu_L(B+1)$, while service provider 1 will still charge the maximum price, service provider 2 will be indifferent between the two normalized premiums of B and $B+1$, sustained as the equilibria in the two adjacent strips.



a) **Normalized Premiums** ***b)*** **Prices**

Figure 3. Equilibrium Normalized Premium and Prices, $\mu_1 > \lambda$ and $\mu_2 < \lambda$

We still observe that, as service provider 1 speeds up, service provider 2 will further lower price to attain a larger value of the normalized premium. The analysis for the equilibrium of case 2) follows by symmetry, when service provider 2 will charge the maximum price and the value of the normalized premium will depend on service rates. In case 3) when neither service provider has enough capacity to serve all the customers, they each will charge the maximum price. The price equilibrium is continuous with respect to the service rates. The two service providers engage in price competition only when they each can accommodate all the customers *and* their service rates sufficiently differ. As a result of price competition, the faster (slower) service provider will always charge a higher (lower) price. An individual service provider will forfeit price as a competitive instrument and charge the maximum price when its service rate alone is unable to accommodate all the market demand.

5. Service Competition

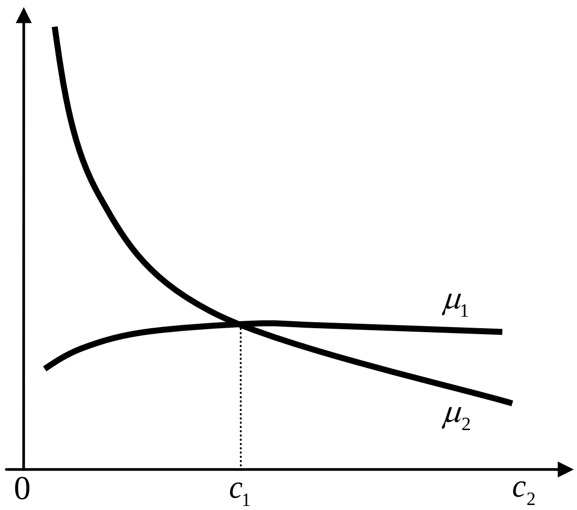
We next explore the competition by the service providers in service rates, by the equilibrium of the price game derived in section 4. We can establish the existence and uniqueness of the equilibrium service rates for general capacity cost functions. To obtain explicit solutions and facilitate comparison, we let the capacity cost function take a quadratic form, $C_i(\mu_i) = c_i \mu_i^2$, $i = 1, 2$. In Proposition 3, we characterize the equilibrium service rates and system performance for the symmetric systems, where the service providers incur identical capacity costs.

Proposition 3: *In a symmetric system where the two service providers incur the same capacity costs, i.e.,* $c_1 = c_2 = c$, *if* $2\lambda c < \overline{P}$, *then, in the unique symmetric equilibrium,*

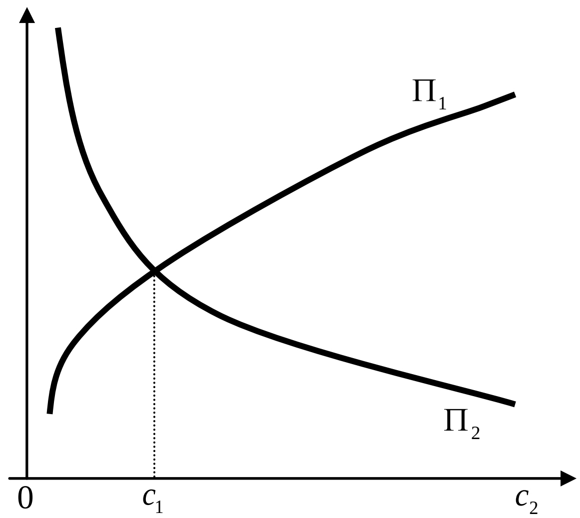
1. *The service rates are* $\mu_1^* = \mu_2^* \equiv \mu^* = \frac{\sqrt[3]{4K}}{12c} + \frac{\sqrt[3]{16}\lambda^2 c}{12\sqrt[3]{K}} - \frac{1}{6}\lambda$, *where* $a \equiv \sqrt{\overline{P}(27\overline{P} - 4c\lambda)}$ *and* $K \equiv \lambda^2 c^2 [27\overline{P} - 2c\lambda + 3\sqrt{3}a]$. μ^* *decreases in capacity cost* c, *increases in the maximum price* $\overline{P}$ *and arrival rate* λ.
2. *The prices are* $p_1^* = p_2^* = \overline{P}$.
3. *The profits of service providers are* $\Pi_1^* = \Pi_2^* = \frac{\overline{P}\lambda}{2} - c\left[\frac{\sqrt[3]{4K}}{12c} + \frac{\sqrt[3]{16}\lambda^2 c}{12\sqrt[3]{K}} - \frac{1}{6}\lambda\right]^2$.

The two service providers build identical service rates in the symmetric systems. They will not compete in price and will each charge the maximum price. As maximum price increases, a service provider can reap in a higher revenue from each completed service and can afford a higher capacity cost to build a faster service. It is obvious that the service rate decreases with capacity cost, but increases with arrival rate.

When the capacity costs differ across the providers, the equilibrium service rates are very difficult, if possible, to derive analytically, due to the discrete nature of the normalized premium. We turn to a numerical study to gain further insights. In our study, for given capacity costs (c_1, c_2), arrival rate (λ), maximum price ($\overline{P}$), and marginal disutility (β), we apply Theorem 1 to obtain the equilibrium prices and the revenues of the two service providers for any pair of service rates, and then use the outcomes to search for the equilibrium service rates and obtain the resultant prices. We observe that the impact of price relative to queue length in the customers' utility and hence their service selection affect the competition behaviors of the service providers in the *asymmetric* systems. When price plays a less significant role in service selection relative to queue length, i.e., the marginal queue length disutility β is high, price competition will be of little value and service providers will each charge the maximum price. Figure 4 displays the typical results for equilibrium service rates and profits with respect to c_2, for given c_1.



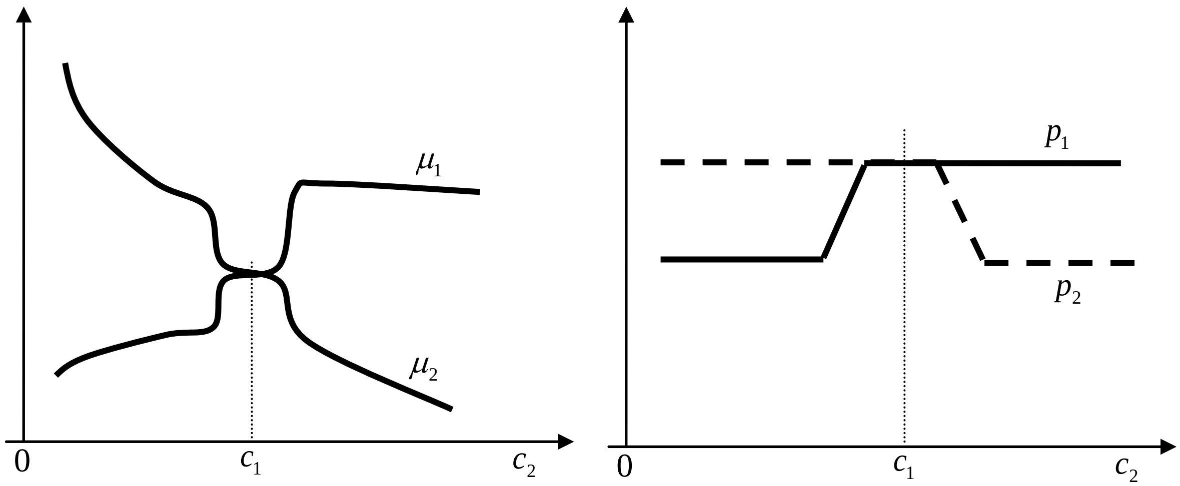
a) μ_1^* **vs** c_2



b) Π_i^* **vs** c_2

Figure 4. Equilibrium in Asymmetric System When Customers are Insensitive to Price

Observe that the service provider with a higher capacity cost is slower in service, i.e., $\mu_i^* \geq \mu_j^*$ when $c_j \geq c_i$, $i, j = 1,2$ and $i \neq j$. As the capacity cost of service provider 2 increases, this particular provider will lower service rate. The effect of a larger value of c_2 on the service rate of service provider 1 is, however, not monotone. As c_2 increases but below c_1, service provider 1 will raise service rate as its cost disadvantage fades until $c_2 = c_1$, when both service providers build the same service rates. When $c_2 > c_1$, service provider 1 will outperform its competitor by a faster service. But, as c_2 continues to increase and service provider 2 slows down, increasing service rate is unlikely to bring service provider 1 a gain in revenue from a larger market share that is enough to offset the mounted capacity cost. Instead, as Figure 4.a) shows, service provider 1 will lower service rate, but remain faster than its competitor. Figure 4.b) shows that, as c_2 increases, service provider 1 will earn a higher profit but service provider 2 suffers a profit loss; they earn the same profits when their capacity costs are identical.



a) μ_i^* **vs** c_2 ***b)*** p_i^* **vs** c_2

Figure 5. Equilibrium in Asymmetric System When Customers are Price Sensitive

We next examine the situation when price plays a significant role in customers' service selection. Figure 5 displays the typical results of equilibrium service rates and prices at the two service providers with respect to c_2, for given c_1. Observe that, when their capacity costs differ sufficiently, the service providers will adopt substitutable competition instruments: the one with a cost advantage will build a faster service and the other will charge a lower price. This echoes our earlier finding that an individual service provider will not simultaneously have a lower price and a faster service than its competitor. Between the two competition instruments, prices affect the customers' queue selections upon arrival, while service rates, through queue lengths, affect their behaviors at arrival and during their stay. The competition in service rates delivers a more influential impact than that in price on customers' behaviors and resultant system performance. The service providers will not compete in price when their service speeds are identical. Price, when adopted as a competition instrument, is used by the slower provider only but exerts a minor effect on the market shares of both service providers. The other observations are similar to those when customers are insensitive to price relative to queue length.

6. Performance Comparison

In the majority of the existing literature, service rates or expected waiting times directly affect customer allocation or service selection. In our model, however, the customers do not infer service rates and are naïve in selecting service based on queue length and price. Moreover, we permit customers to jockey, a common behavior in service systems, but many existing models preclude such a possibility. In this section, we conduct a comparative study on the effects of the availability of service rate in service selection and customers' jockeying behaviors on system performance, by comparing the findings in our model to those in Li and Lee (1994), hereinafter referred to as LL (1994), and So (2000). LL (1994) and this paper define utility functions for the customers to choose service, while So (2000) adopts a choice model for service selection.

LL (1994) assumes that each customer selects a service provider to maximize his utility, which is similar to the one in our model and is shown below:

$$U_i = \begin{cases} -\dfrac{r(n_i+1)}{\mu_i} - \beta p_i & \textit{if } p_i \le \bar{P} \\ -\infty & \textit{Otherwise} \end{cases} \quad \text{for } i = 1, 2, \tag{12}$$

where p_i is the price by service provider i, $\bar{P}$ the customers' reservation price (equivalent to the maximum price in our model), r and β are the respective marginal disutilities of waiting time and price. While LL (1994) also has queue length as a consideration factor in customers' service selection and permits customers to jockey, it differs fundamentally from our model in that the customers use service rates in service selection, but the customers in our model do not have such a privilege. The result in LL (1994) for given service rates provides a benchmark to examine the effects of the service-rate based service selection.

We have defined the normalized premium for service provider 1, B, to capture the result of price competition. The corresponding normalized premium in LL (1994), for given μ_1 and μ_2, can be shown to be:

$$B_L = \left\lceil \beta^{-1}(p_1 - p_2)\mu_2 - \mu_1^{-1}(\mu_1 - \mu_2) \right\rceil, \tag{13}$$

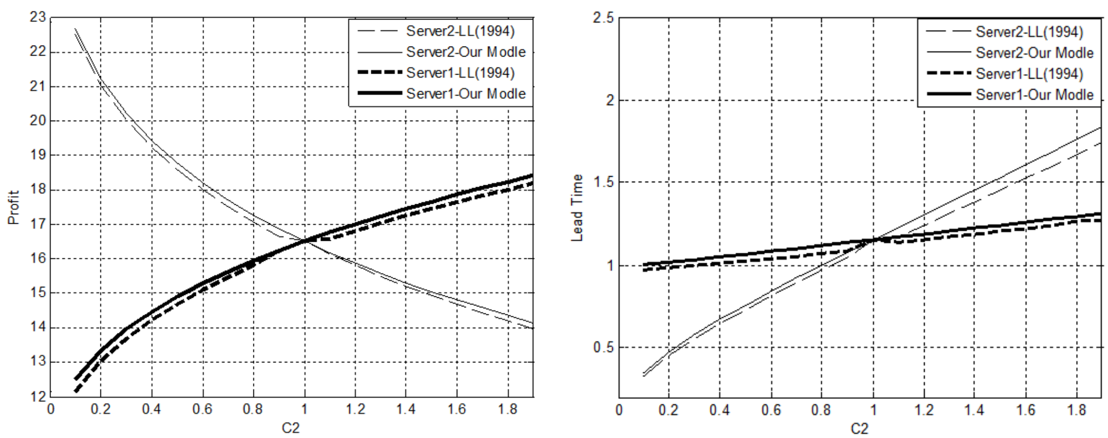
for which we let $\beta = 1$ and $r = \beta$ in the utility function (12).

Proposition 4 compares the normalized premiums and the profits of the service providers in our model and those in LL (1994).

Proposition 4: *For given μ_1, μ_2, and λ ($\mu_1 + \mu_2 > \lambda$), in the equilibrium, the absolute value of the normalized premium for service provider 1 in our model is no larger than that in LL (1994), and the profits of the two service providers in our model are no less than those in LL (1994).*

For given service rates, the two service providers earn higher profits and engage in less intense price competition in our model than in LL (1994). It is because that, the slower service provider will less likely be patronized by the better informed customers in LL (1994) than by those in our model, and will hence be more aggressive in pricing to grab market share, as reflected in a larger normalized premium. A fiercer price competition, however, hurts the profit of each individual service provider.

We apply the results in LL (1994) to derive the equilibrium service rates and compare them to those in our model. We find that, in the symmetric systems where the service providers incur identical capacity costs, the same equilibrium outcomes are sustained in LL (1994) and our model. That is, the availability of service rates in service selection does not deliver a substantial impact on the performance of a symmetric system, in which price is not a competition instrument and the two service providers build the same service rates. Since service selection is based on the relative utilities, the specific values of service rates essentially have no impacts on customers' preferences, and the system dynamics in LL (1994) are exactly identical to those in our model.



a) Π_i^* **vs** c_2 ***b)*** t_i^* **vs** c_2

Figure 6. Effects of Availability of Service Rate Information

For the asymmetric systems, we turn to numerical studies to examine the relevant effects. Figure 6 displays the typical outcomes with base parameters $c_1 = 1.0$, $\lambda = 2.0$, $\overline{P} = 20$, $r = 1$, and $\beta = 1.0$. Service rates do deliver an impact in this case. Observe that service providers build faster services and customers spend less time waiting for service in LL (1994) than in our model. This is quite intuitive since service rate directly enters into customers' utility to affect their service selection in LL (1994), which provides a strong incentive for the providers to speed up. The faster service shortens customers' expected waiting time. But, the added capacity cost to build a faster service in LL (1994) is not paid off by the gain due to the fiercer price competition the two service providers engage in.

We next compare our model to the model of So (2000) to shed light in the effects of the customers' jockeying behavior. Like LL (1994), this paper assumes that service rates are directly used in service selection. But, state-independent policies are applied therein to allocate customers into service upon arrival, and once assigned to a queue, a customer stays with the queue until service completion. In particular, So (2000) assumes that customers select servers based on the expected lead time and price, and applies a multiplicative competitive interaction (MCI) model for demand allocation, which essentially translates into the same utility function as that in LL (1994):

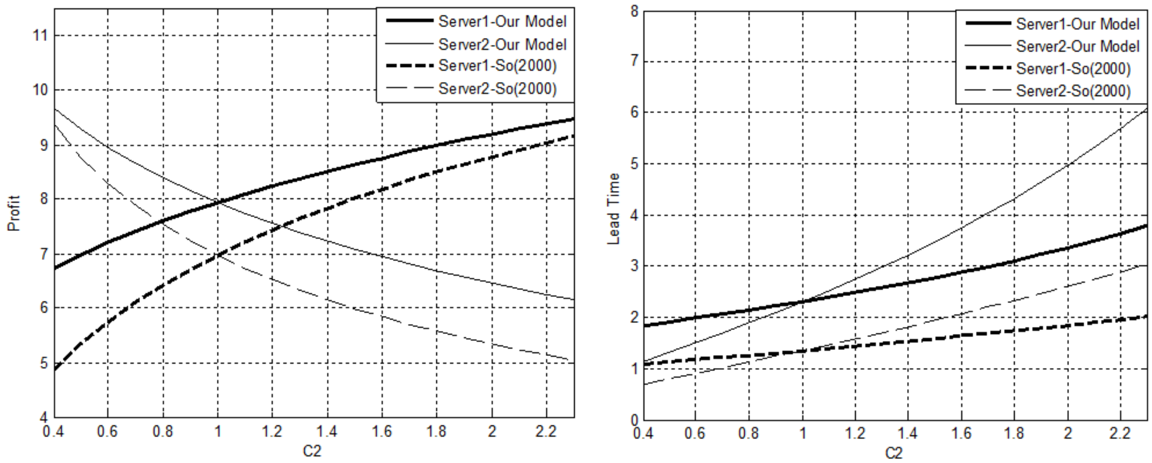
$$\lambda_i = \lambda\left(\frac{p_i^{-\beta} t_i^{-1}}{p_i^{-\beta} t_i^{-1} + p_j^{-\beta} t_j^{-1}}\right), \quad \text{for } i = 1, 2, \tag{14}$$

where p_i is service provider i's price, t_i the expected lead time, and β price elasticity.

Proposition 5: *In a symmetric system where the two service providers incur the same capacity costs, the two service providers build slower service but earn higher profits in our model than in So (2000). The customers, however, endure longer waiting time in our model.*

The service providers will not compete in price in the symmetric systems where, by our earlier discussions, the availability of service rate in customers' service selection delivers an insubstantial effect on system performance. Proposition 5 reveals that, however, the customers' jockeying behavior leads to lower service rates and higher profits, but detains customers for a longer duration. This is because that the chance for waiting customers and idle providers to co-exist is lowered when the customers can switch in real time. The service providers achieve higher capacity utilizations when the customers can jockey than when they commit to one of them. As a consequence, they invest less in capacity when the customers can switch in real time. Their slower service, however, detains the customers longer in waiting for service.

For asymmetric systems, we use numerical studies to examine the impact of customers' jockeying behavior. Figure 7 displays the outcomes with base parameters of $c_1 = 1.0$, $\lambda = 1.0$, $\overline{P} = 20$, and $\beta = 1.0$. Observe that the curves have the same trends as those in Figure 6 for the symmetric systems. The performance gaps in Figure 7 are attributed to both the availability of the service rate in service selection and the customers' jockeying behaviors. With the same parameters, the gaps in Figure 7 are notably wider than those in Figure 6, which strengthens our earlier observation for symmetric systems that, on top of the effects of the availability of service rate, the customers' jockeying behavior further lowers service providers' capacity investment and lengths the customers' duration of stay.



***a*)** Π_i^* **vs** c_2 ***b*)** t_i^* **vs** c_2

Figure 7. Effects of Customers' Service Selection Behavior in Asymmetric Systems

7. Concluding Remarks

We have considered a service system consisting of two service providers each with a separate queue. Customers choose queues to join upon arrival and can switch between queues in real time while waiting for service. The customers are unable to infer service rates and naïve in choosing service based on queue length and price. We have analyzed the service and price competition between the two service providers under such assumptions on customer behaviors, and found that neither service provider will simultaneously charge a lower price and build a faster service than its competitor. When price plays a less significant role in the customers' service selection relative to queue length or their capacity costs are comparable, the service providers will not compete in price and will each charge the maximum price. Otherwise, the lower-cost provider will build a faster service, and its higher-cost competitor will charge a lower price. Comparing the results in our model to those in the existing literature, we find that service providers build lower service rates, engage in less intense price competition, earn higher profits, but customers are expected to wait in line longer when they are naive in selecting service by queue length and price than when they can use service rates to make sophisticated choices. Customers' jockeying behavior further lowers the service providers' capacity investment and lengthens the customers' duration of stay.

This research paves the ground work for further exploration. We assumed the customers choose service by comparing the spot utilities they can obtain from the two providers. The spot utility defined in the current model is a simple function of queue length and price. It is probable that the customers' service selection is affected by queue length and price in a more involved manner. For instance, an arriving customer can choose to join the shorter queue or the lower-priced provider if the relative queue length or price differs too much. Further, customers may not choose a queue if it is too long. Incorporating such features into the customers' service selection criterion can make our model more realistic.

Acknowledgement

This research is supported in part by Niche Area Project J-BB7B and Project A-SA22 by The Hong Kong Polytechnic University.

References

1. Allon, G., A. Federgruen. 2007. Competition in Service Industries. *Operations Research,* **55**(1), 37-55.
2. Bell, C. E., S. Stidham. 1983. Individual versus Social Optimization in the Allocation of Customers to Alternative Servers. *Management Science,* **29** (7), 831-839
3. Cachon, G. P., P. T. Harker. 2002. Competition and Outsourcing with Scale Economies. *Management Science,* **48**(10), 1314-1333.
4. Cachon, G. P., F. Zhang. 2007. Obtaining fast service in a queueing system via performance-based allocation of demand. *Management Science*, **53**(3) 408-420.
5. Christ, D., B. Avi-Itzhak. 2002. Strategic equilibrium for a pair of competing servers with convex cost and balking. *Management Science*, **48**(6), 813-820.
6. Gilbert, S. M., Z. K. Weng. 1998. Incentive effects favor nonconsolidating queues in a service system: the principle-agent perspective. *Management Science*, **44**(12) 1662-1669.
7. Grassmann, W. K. 1980. Transient and steady state results for two parallel queues. *The International Journal of Management Science*, **8**(1), 105-112.
8. Ha, A. Y., Li, L., S. M. Ng. 2003. Price and delivery logistics competition in a supply chain. *Management Science,* **49**(9), 1139-1153.
9. Haight, F. A. 1958. Two queues in parallel. *Biometrica,* **45**(34) 401-410.
10. Kalai, E., M. I. Kamien, M. Rubinovitch. 1992. Optimal service speeds in a competitive environment. *Management Science*. **38**(8) 1154-1163.
11. Knessl, C., J. M. Bernard, Z. Svhuss, C. Tier. 1986. Two parallel queues with dynamic routing. *IEEE Transactions on Communications*, **34**(12), 1170-1175.
12. Koenigsberg, E. 1966. On Jocketing in Queues. *Management Science*. **12** (5) 412-436
13. Hassin, R. 1986. Consumer information in markets with random products quality: the case of queues and balking. *Econometrica*, **54**(5), 1185-1195.
14. Hassin, R., M. Haviv. 1994. Equilibrium strategies and the value of information in a two

line queueing system with threshold jockeying. *Communications in Statistics - Stochastic Models*, **10**(2), 415-436.

15. Lederer, P. J., L. Li. 1997. Pricing, production, scheduling, and delivery time competition. *Operations Research*, **45**(3), 407-420.
16. Lee, H., M. Cohen. 1985. Multi-agent customer allocation in a stochastic service system. *Management Science*, **31**(6), 752-763.
17. Li, L. 1992. The role of inventory in delivery-time competition. *Management Science*, **38**(2), 182-197.
18. Li, L., Y. S. Lee. 1994. Pricing and delivery-time performance in a competitive environment. *Management Science*, **40**(5) 633-646.
19. Nakamura, M., I. Sasase, S. Mori. 1989. Two parallel queues with dynamic routing under a threshold-type scheduling. *IEEE GLOBECON*, **3**,1445-1449.
20. Reitman, D. 1991. Endogenous quality differentiation in congested markets. *Journal of Industrial Economics*, **39**(6), 621-647.
21. Rubinovitch, M. 1985. The slow server problem. *Journal of Applied Probability*, **22**(1), 205-213.
22. Shang, W. and L. Liu. 2011. Promised delivery time and capacity games in time-based competition. *Management Science.* Forthcoming.
23. So, K. C. 2000. Price and time competition for service delivery. *Manufacturing & Service Operations Management*, **2**(4), 392-409.
24. Zhao, Y., W. K. Grassmann. 1990. The shortest queue model with jockeying. *Naval Research logistics*, **37**(5), 773-787.