Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Time-Based Competition with Benchmark Effects

Liu Yang, Francis de Véricourt, Peng Sun

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 16, No. 1, Winter 2014, pp. 119–132
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)

http://dx.doi.org/10.1287/msom.2013.0462

Time-Based Competition with Benchmark Effects

Liu Yang
School of Economics and Management, Tsinghua University, Beijing 100084, China, yangliu@sem.tsinghua.edu.cn

Francis de Véricourt
INSEAD, 77305 Fontainebleau, France, francis.devericourt@insead.edu

Peng Sun
Fuqua School of Business, Duke University, Durham, North Carolina 27708, peng.sun@duke.edu

We consider a duopoly where firms compete on waiting times in the presence of an industry benchmark. The demand captured by a firm depends on the gap between the firm's offer and the benchmark. We refer to the benchmark effect as the impact of this gap on demand. The formation of the benchmark is endogenous and depends on both firms' choices. When the benchmark is equal to the shorter of the two offered delays, we characterize the unique Pareto optimal Nash equilibrium. Our analysis reveals a *stickiness effect* in which firms equate their delays at the equilibrium when the benchmark effect is sufficiently strong. When the benchmark corresponds to a weighted average of the two offered delays, we show the existence of a pure Nash equilibrium. In this case, we reveal a *reversal effect*, in which the market leader, i.e., the firm that offers a shorter delay, becomes the follower when the benchmark effect is sufficiently strong. In both cases, we show that customers' equilibrium waiting times are shorter with the benchmark effect than without it. Our models also capture customers' loss aversion, which, in our setting, states that demand is more sensitive to the gap between the delay and the benchmark when the delay is longer than the benchmark (loss) than when it is shorter (gain). We characterize the impact of this loss aversion on the equilibrium in both settings. Finally, we show numerically that the stickiness and reversal effects still exist when firms also compete on price.

Key words: waiting time competition; benchmark effect; loss aversion; queues; game theory
History: Received: May 9, 2012; accepted: June 15, 2013. Published online in *Articles in Advance* November 4, 2013.

1. Introduction

Service firms often compete on waiting times (see Allon and Federgruen 2007 and references therein). In the classical context, a firm adjusts the expected delay it offers to the market to attract additional demand so as to maximize profit. Customers, in turn, derive their utility from the waiting time they experience at the firm they choose to join. However, customers' utility may also directly depend on other offers in the competition. For instance, a given level of delay can be more or less displeasing depending on whether it is longer or shorter than some industry benchmark. This is because customers are subject to reference effects (Kahneman and Tversky 1979) when evaluating waiting times (Leclerc et al. 1995) or, more generally, service quality (Cadotte et al. 1987). When setting an expected waiting time, a firm also influences the benchmark against which customers evaluate the delays they experience at other firms. This paper's main contribution is to offer a theoretical analysis of how benchmark effects such as these influence waiting time competition between firms.

To that end, we consider a simple duopoly where firms offer expected waiting times for their service by adjusting their service rates at a capacity cost. Demand is then split between firms according to the different customers' utilities. More precisely, a customer's satisfaction level with a firm is a function of both firms' offered expected waiting times as in the classical duopoly, and also of the gap between the firm's offered delay and the benchmark. We refer to the *benchmark effect* as the impact the gap has on demand. Our model also allows for customers to be loss averse, in the sense that a positive gap, i.e., when the offered delay exceeds the benchmark, has a greater impact on utility than the corresponding negative gap, i.e., when the benchmark exceeds the delay by the same amount. Thus, a firm's decision has a direct effect on demand through its choice of waiting time, and an indirect one through the benchmark effect. This situation gives rise to a game in which each firm strategically chooses a waiting time to maximize profit.

Following Cadotte et al. (1987), we consider two different settings depending on how the benchmark is formed. In the first setting, the benchmark equals the smaller of the two offered waiting times, whereas in the second setting, the benchmark corresponds to a weighted average. Without the benchmark effect, our model corresponds to a classical setup of time-based competition.

We show the existence of and characterize the Nash equilibria for this problem. We find that the presence of a benchmark benefits customers, in that customers experience shorter delays in equilibrium with the benchmark effect compared to the classical duopoly case without the benchmark. In fact, the expected waiting times in equilibrium decrease as the benchmark effect increases.

In our first setting, where the benchmark corresponds to the smaller expected waiting time, our analysis reveals a *stickiness effect* in which firms equate their offers in equilibrium as long as the ratio of their profit margins belongs to a given interval. In contrast, in the absence of the benchmark, the offered waiting times are equal only when the ratio takes a specific value. We further show that whereas the interval expands as the benchmark effect intensifies, it may either shrink or expand with the level of loss aversion depending on the strength of the benchmark effect. In other words, the intensity of the benchmark effect influences the direction of the impact of loss aversion on the stickiness of the equilibrium waiting time.

In our second setting, where the benchmark corresponds to the weighted average waiting time between the two firms, we identify a *reversal effect* in that the leader in a duopoly without a benchmark effect, i.e., the firm that offers a shorter waiting time, can become the follower (i.e., can offer a longer waiting time) if the benchmark effect is sufficiently strong. The market leader is determined by a profit-margin ratio threshold, which is monotone in both the benchmark effect and loss aversion effects. Similar to the minimum benchmark case, the threshold may either decrease or increase with the level of loss aversion depending on the intensity of the benchmark effect.

The presence of a benchmark effect is supported by results of prospect theory (Kahneman and Tversky 1979), which demonstrates that people generally view a final outcome as a gain or a loss with respect to a certain context-dependent reference point. Although the original applications of prospect theory mainly studied people's preferences toward monetary payoffs (Niedrich et al. 2001, Erdem et al. 2001, Popescu and Wu 2007), empirical studies have demonstrated that a similar reference effect also exists for waiting times (Leclerc et al. 1995). In our context, the service benchmark determines the waiting time reference point. The benchmark effect corresponds, here, to the gap between the benchmark and the actual waiting time. In other words, customers might experience "waiting time gain" if the actual waiting time is shorter than the benchmark, and "waiting time loss" otherwise.

The benchmark effect is also highly consistent with the stream of research on service quality and customer satisfaction (Parasuraman et al. 1985). For example, according to Anderson and Sullivan (1993), customer satisfaction from a service is a function of perceived quality as well as of the gap between customers' expectations (the reference point) and perceived quality. Moreover, when perceived quality falls short of expectations, the gap has a greater impact on satisfaction than does the corresponding gap when quality exceeds expectation, a result consistent with prospect theory's loss aversion. Lin et al. (2008) empirically studied the impact of the service quality gap on customers' behavior intentions. They showed that a customer's behavior depends on the service quality gap according to a value function that is "kinked" at the reference point, with the loss of service quality influencing the customer's behavior more than does a corresponding service quality gain. This is also consistent with prospect theory.

Empirical studies further suggest how firms' decisions might together determine the benchmark, i.e., the reference point (see Zeithaml et al. 1993). For instance, Cadotte et al. (1987) showed that two types of comparisons better explain customer satisfaction with service quality: "best brand norm" and "product type norm." According to the best brand norm, customers select the best brand in the category as their reference. The product type norm, on the other hand, refers to a situation in which the average performance is perceived as typical of a group of similar brands. This latter norm is also consistent with adaptation-level theory in which a reference standard is perceived as the mean of a set of presented stimuli (Helson 1964). In our setting, service quality corresponds to offered waiting time, and therefore, the minimum and average delays between the two firms correspond to the best brand norm and product type norms, respectively.

The conditions under which customers use the best or average quality as a benchmark remains an open research question. Nonetheless, results from Cadotte et al. (1987) indicate that benchmarks are context specific. In their study, the average case (a product type brand in their setting) appeared to be more prevalent when the service was a commodity (e.g., fast food) as opposed to upscale services (e.g., an elegant dining experience). This suggests, in our setting, that the minimum benchmark is more likely to be used for premium services, whereas the average case should be more prominent for regular services. However, other aspects of a service might also influence the benchmark. In particular, marketing research on pricing suggests that the minimum benchmark might be used when the difference between waiting times is more salient. For instance, Rajendran and Tellis (1994) empirically studied the reference price of one grocery product category. They showed that the lowest prices influence customers more than does the mean, because of the prices' salience and availability. In our setting, the minimum waiting time is likely to be more salient when the service is offered on the Internet, where information about the firm's performance is more readily accessible, as opposed to services offered through the brick-and-mortar channel. This also suggests that the minimum benchmark might be more prevalent in our setting precisely because we are interested in time-based competition, for which differences in delays are more salient.

This work is closely related to the literature on competition between service firms when customers are sensitive to waiting time or service quality (Gans 2002, Png and Reitman 1994). Service competition has been the subject of many studies (see Hassin and Haviv 2003 for a comprehensive review in queueing settings). In some of these models (De Vany and Saving 1983, Kalai et al. 1992, So 2000, Cachon and Harker 2002, Ho and Zheng 2004, Allon and Federgruen 2007), firms compete on both waiting times and prices either in an aggregated form (full price) or as separate attributes. Gaur and Park (2007), Liu et al. (2007), and Hall and Porteus (2000), instead, considered the situations in which customers' demand solely depends on waiting time, or service quality. In fact, certain industries experience a higher level of price rigidity compared to their ability to vary service rates. Blinder et al. (1998) provides extensive empirical evidence of this phenomenon. For example, half of the businesses in their study change prices no more than once per year. Among all the industries they studied, service companies adjust prices most slowly. Therefore, this paper mainly focuses on time-based competition. Nonetheless, we allow firms to compete on both time and price in §5 and show numerically that the stickiness and reversal effects still exist in this setting.

Our model is most closely related to that of Allon and Federgruen (2007), which investigated service competition in both service levels and prices. In the absence of a benchmark effect, our model corresponds to a special case of their "price first" scenario, in which our main findings disappear. Ho and Zheng (2004) also studied a similar duopoly competition in waiting time announcements when demand is affected by service quality. Their paper studied the case where firms do not need to comply with the waiting times they offer as Allon and Federgruen (2007) and we do. On the other hand, Ho and Zheng (2004) do consider benchmark effects and loss aversion, the main focus of our paper.

We model each firm's demand function from individual customer's choice level using a *spokes model*, which is a generalization of the Hotelling model (see, for example, Chen and Riordan 2007; Amaldoss and He 2010, 2013). The spokes model was recently introduced by the marketing literature to allow for arbitrary numbers of possible product and service varieties. As a result, compared with the Hotelling model, the spokes model captures the more realistic situation where two competing firms do not always cover the entire potential market (the size of which is determined by the number of spokes in the model). The market shared by the firms is in fact endogenous to the model and depends on the firms' strategies. In our context, this implies that the resulting aggregate demand is an asymmetric function, in the sense that a firm's demand is affected more by its own waiting time than by the other firm's waiting time. The Hotelling model, on the other hand, generates a symmetric demand function, which stands in contrast with the asymmetric structure found in relevant studies, such as Allon and Federgruen (2007).

The aggregate demand resulting from the spokes model is, in general, piecewise linear. Nonetheless, for the minimum benchmark case, we deduce conditions under which it is sufficient to consider linear demands for the equilibrium analysis. We show numerically that similar conditions exist for the average case. In the context of the spokes model, linear demand corresponds to the situation where both firms are able to attract and compete for customers but do not together cover the full potential market.

Our paper contributes to the emerging literature on competition between firms with customer reference effects. Heidhues and Koszegi (2008) studied price competition when customers base their reference on their recent expectations about a product. They showed the existence of a focal price equilibrium in the presence of a reference effect. Zhou (2011) also examined firms' price competition, but with customers' reference point based more on the "prominent" firm. In this setting, the equilibrium price randomizes between high and low levels. These two papers consider price competition rather than service operations competition. They also consider exogenous reference points, whereas in our setup the benchmark is endogenously determined by the firms' strategies.

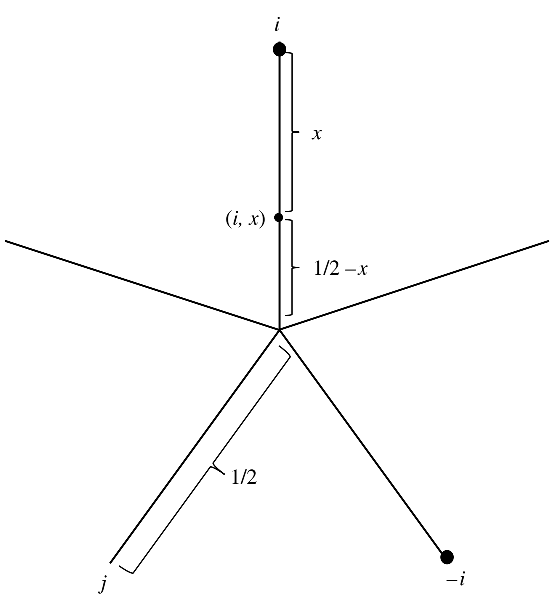
The remainder of this paper is organized as follows. Section 2 presents the spokes model. Section 3 introduces the minimum benchmark model and the corresponding waiting time competition game. After characterizing the unique (Pareto) Nash equilibrium structure, we analyze the "stickiness" effect of equilibrium waiting times, and how they are affected by the benchmark and loss aversion effects. Section 4 analyzes the average benchmark model, its Nash equilibrium characterization, and the impacts of the benchmark and loss aversion effects on the "reversal" phenomenon. We show numerically in §5 that the stickiness and reversal effects still exist in the presence of price competition. We conclude this paper and discuss future research directions in §6. Unless specified, all proofs are presented in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/msom.2013.0462).

2. The Model Setup

To model customer demand, we follow the spokes model (Chen and Riordan 2007; Amaldoss and He 2010, 2013). According to the spokes model, customers are distributed uniformly on a star-shaped network consisting of $N(\geq 2)$ spokes that connect at a hub, as shown in Figure 1. The number of spokes represents the number of different service varieties sought by potential customers. Using an example from the restaurant industry, different spokes may stand for different cuisine types (seafood or steakhouse for example). Each line has length equal to 1/2. Two service firms occupy the terminals of two different spokes, reflecting that only two preference varieties are offered by the market. We denote these firms and their respective spokes with the subscripts i and $-i$, respectively. Any other spoke with $k \neq -i, i$ is referred to as a no-service spoke.

A customer in the market resides on spoke j at distance x from the terminal of the spoke and is represented by pair (j, x). Customers who join a firm incur a distance cost equal to mx if the firm is on their local spoke (which requires $j \in \{i, -i\}$). If the firm is not local, they incur a cost equal to $m(1-x)$. (Note here that distance does not necessarily denote geographical distances, but generally describes customers' preference intensity; in the restaurant example, a customer located close to the center feels more or less indifferent between, say, a seafood restaurant and a steakhouse, whereas a customer much closer to a terminal strongly prefers one of the two.)

Figure 1 An Illustration of the Spokes Model with Eight Possible Varieties ($N = 5$) and Two Competing Firms (Firm i and Firm $-i$)



Furthermore, a customer will typically only consider a limited number of varieties because of potential search costs (Wolinsky 1986). In particular, Chen and Riordan (2007) assume that each customer only considers two possible varieties. We follow the exact same approach in assuming that a customer's choice set consists of her local spoke and only one of the remaining $N-1$ spokes. Customers and their choice sets are uniformly distributed on the spoke network. This means, for instance, that $1/(N-1)$ of the customers on spoke i have both firms i and $-i$ as their choice set. All other spoke i customers choose between spoke i and spoke $k \neq -i$, that is, either receiving service from firm i or not receiving any service. (See Online Appendix A for more details.)

A customer chooses between her two alternatives by maximizing her overall satisfaction from receiving the service. Following Koszegi and Rabin (2006), we assume a customer's overall satisfaction has two components: "consumption utility" and the "gain-loss utility." When a customer at (j, x) receives service from firm i and waits t_i units of time, her consumption utility equals $u_1(i, x, t_i) = v - p - \alpha t_i - mx$ if $j = i$, or $u_1(j, x, t_i) = v - p - \alpha t_i - m(1-x)$ if $j \neq i$, where v and p are the service values and price, respectively, and α is the customer's time sensitivity.

The "gain-loss utility" is based on prospect theory (Tversky and Kahneman 1991); that is, a customer's satisfaction level is affected by the comparison of actual waiting time t_i to benchmark (reference point) r. Specifically, the gain-loss utility of customer (j, x) who chooses firm i is

$$u_2(j, x, t_i \mid r) = \begin{cases} \beta_r(r - t_i) & \text{if } t_i < r, \\ \beta_r\beta_l(r - t_i) & \text{if } t_i \geq r. \end{cases}$$

Here we use a piecewise linear function to simplify the model while keeping the essential idea of loss aversion. Parameter $\beta_r > 0$ represents the benchmark effect, whereas $\beta_l > 1$ captures the level of loss aversion (Kahneman and Tversky 1979), which, in our context, states that customers are more sensitive to waiting times that exceed the benchmark.

We assume that demand for the service arrives according to a Poisson process and that both firms are $M/M/1$ facilities. In particular, firms choose and commit to average waiting times w_i and w_{-i}, respectively. Therefore, waiting time t_i follows an exponential distribution with rate parameter $1/w_i$. In line with the literature, we assume that customers know the distributions of waiting times and are able to take expectation of satisfaction levels. This means that when a customer (i, x) chooses the local firm, her overall satisfaction level is

$$EU_{i,x,i} = E[u_1(i,x,t_i) + u_2(i,x,t_i)] = v - p + S_i - mx,$$

where the term

$$S_i = S(w_i, r) = -\alpha w_i + \beta_r[(r - w_i) - e^{-r/w_i}(\beta_\ell - 1)w_i] \quad (1)$$

captures the contribution of waiting time to customers' satisfaction from firm i's service. Similarly, the overall satisfaction for customer (j, x) to join a nonlocal firm i is $EU_{j,x,i} = v - p + S_i - m(1-x)$.

This setup enables us to construct the aggregate demand for each firm in the form of customer arrival rates λ_i and λ_{-i}. Normalizing the total market size to be 1, firm i's demand function is characterized by the following proposition.

Proposition 1. *Firm i's demand λ_i is a piecewise linear function of S_i and S_{-i}. In particular, when both firms i and $-i$ satisfy*

$$\begin{aligned} 1/2 < (v - p + S_i)/m < 1 \quad &\textit{and} \\ 1/2 < (v - p + S_{-i})/m < 1, \end{aligned} \quad (2)$$

demand is a linear function

$$\lambda_i = aS_i - bS_{-i} + (a-b)(v-p) + bm, \quad (3)$$

where

$$a = \frac{2N-3}{N(N-1)m} > b = \frac{1}{N(N-1)m}.$$

Condition (2) corresponds to the case where customers on both spokes i and $-i$ prefer either firm over no service, but some customers on other spokes prefer not to purchase the service (see Online Appendix A). This corresponds to the most general situation where firms compete but do not fully cover the potential market. The linear demand function also dramatically simplifies the analysis and will be our focus in the rest of this paper.

For a given demand arrival rate λ_i, firm i's committed average waiting time w_i translates to providing a service rate μ_i such that $w_i = 1/(\mu_i - \lambda_i)$. Considering competition in terms of average service time rather than in terms of service rate reduces customers' computational burden, as argued in Cachon and Harker (2002). That is, a customer does not need to consider other customers' decisions and only observes both firms' waiting time commitments for the joining decision. Furthermore, firm i incurs capacity cost c_i per unit of service rate. Let us denote quantity $\rho_i = (p - c_i)/c_i$ to represent firm i's profit margin. Finally, define firm i's profit as

$$P_i(w_i, w_{-i}) \equiv c_i\left(\rho_i \lambda_i(w_i, w_{-i}) - \frac{1}{w_i}\right), \quad (4)$$

where demand rate λ_i is a function of w_i and w_{-i} through S_i and S_{-i}. Firm i's objective is to maximize $P_i(\cdot, w_{-i})$ by choosing w_i.

3. The Minimum Benchmark Case

As mentioned in the introduction, customers sometimes use a "best brand norm" to form a benchmark of service quality, i.e., customers select the best brand in the category as their benchmark (Cadotte et al. 1987). In our setting, this corresponds to the shorter waiting time, or, $r \equiv \min(w_i, w_{-i})$, where r denotes the waiting time benchmark.

We first provide a condition under which it is sufficient to consider linear demands for the equilibrium analysis. In particular, condition (2) depends on firms' waiting time decisions through S_i and S_{-i}. The following lemma provides a sufficient condition under which demand is linear, independently of firms' choices.

Lemma 1. *When $r \equiv \min(w_i, w_{-i})$, model parameters that satisfy*

$$\frac{m}{2} + \sqrt{\frac{\alpha + \beta_r \beta_l}{a \min(\rho_i, \rho_{-i})}} < v - p < m \quad (5)$$

also satisfy condition (2).

In the remainder of this section we assume condition (5), under which the demand functions for both firms are linear regardless of the firms' decisions.

The following proposition reveals important properties of profit function P_i, which will prove useful for analyzing and providing insight into the competition between the two firms.

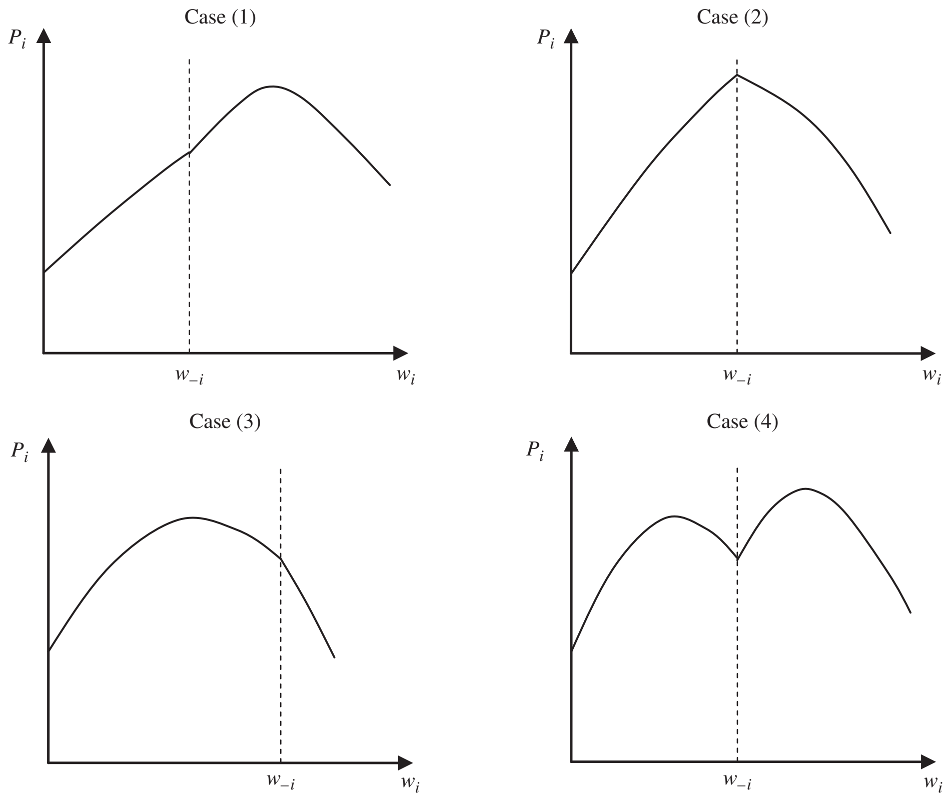
Proposition 2. *Firm i's profit P_i, defined in Equation* (4), *has the following two structural properties*:

1. *P_i is a quasi-concave function of waiting time w_i;*
2. *P_i is supermodular when $w_i > w_{-i}$, and submodular when $w_i < w_{-i}$.*

To prove quasi concavity, we first show that firm i's profit P_i is concave in w_i when $w_i \geq w_{-i}$, and quasi concave in w_i when $w_i < w_{-i}$. Hence, P_i can be reduced to the four cases shown in Figure 2. In all cases except case (4), P_i is quasi concave. Furthermore, we show that if the left derivative of P_i with respect to w_i at $w_i = w_{-i}$ is negative, then the right derivative cannot be positive; thus, case (4) is not possible. (See the online appendix for more details.)

As for the supermodularity and submodularity properties, when $w_i > w_{-i}$, increasing w_{-i} lowers customers' expectations, making them easier to satisfy. This provides an incentive for firm i to extend its waiting time w_i, which leads to the supermodularity of P_i. On the other hand, when $w_i < w_{-i}$, the increase

Figure 2 **Four Possibilities of a Function of w_i That Is Concave When $w_i \le w_{-i}$ and Quasi Concave When $w_i > w_{-i}$**



Note. In the proof of Proposition 2, we show that case (4) is not a possible scenario for the profit function P_i, which implies that P_i is a quasi-concave function.

of w_{-i} makes firm $-i$ even less attractive, and thus increases firm i's demand. This makes it more affordable for firm i to lower its waiting time w_i, which leads to the submodularity of P_i.

3.1. Nash Equilibrium

Given the choice made by the other firm, each firm sets its waiting time so as to maximize its profit. Because demand and, hence, profits depend on both firms' decisions, the situation gives rise to a game. In the following, we show the existence of and fully characterize the Nash equilibrium for this game.

Using Proposition 2, we first study firm i's best response curve as a function of w_{-i}. The quasi concavity of P_i implies that the first-order condition is a sufficient condition for firm i's best response. However, since $r = \min(w_i, w_{-i})$, firm i's demand function λ_i is not differentiable at $w_i = w_{-i}$. Nonetheless, we can define the left and right derivatives of λ_i with respect to w_i when $w_i = w_{-i}$, respectively, as

$$\underline{\delta} \equiv -\frac{\partial \lambda_i}{\partial w_i}\bigg|_{w_i \uparrow w_{-i}} = a\alpha + b\beta_r + (a+b)e^{-1}\beta_r(\beta_l - 1),$$

$$\bar{\delta} \equiv -\frac{\partial \lambda_i}{\partial w_i}\bigg|_{w_i \downarrow w_{-i}} = a[\alpha + \beta_r + 2e^{-1}\beta_r(\beta_l - 1)].$$

Partial derivatives $\underline{\delta}$ and $\bar{\delta}$ represent the impact of waiting time decision w_i on arrival rate λ_i, when w_i approaches the competitor's decision w_{-i} from below and above, respectively. Clearly, $\underline{\delta} < \bar{\delta}$ when $\beta_r > 0$, which means that the marginal increase in demand from setting a shorter waiting time is smaller than the marginal decrease in demand from setting a longer waiting time. Given these two quantities, firm i's best response curve is determined by the following proposition.

Proposition 3. *Firm i's best response curve, $w_i^\star(w_{-i})$, is a piecewise function on the following three intervals*:

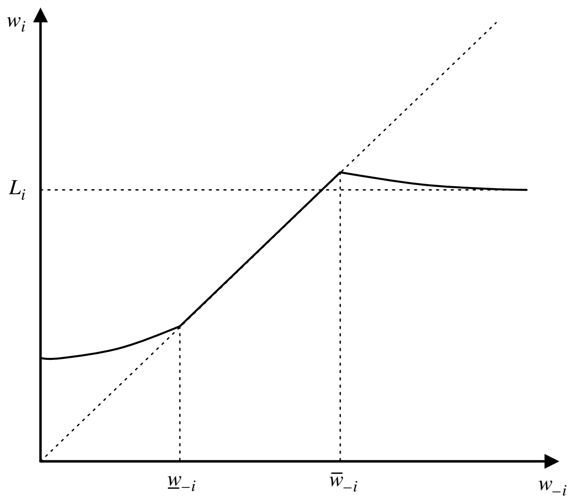
1. *when $w_{-i} \in [0, \underline{w}_{-i})$, $w_i^\star(w_{-i}) > w_{-i}$ and is an increasing function of w_{-i};*
2. *when $w_{-i} \in [\underline{w}_{-i}, \bar{w}_{-i}]$, $w_i^\star = w_{-i}$;*
3. *when $w_{-i} \in (\bar{w}_{-i}, \infty)$, $w_i^\star(w_{-i}) < w_{-i}$ and is a decreasing function of w_{-i}; furthermore, $w_i^\star$ converges to a limit $L_i > 0$;*

where $\underline{w}_{-i} = (\rho_i \bar{\delta})^{-(1/2)}$ and $\bar{w}_{-i} = (\rho_i \underline{\delta})^{-(1/2)}$.

The three cases of Proposition 3 correspond to cases (1)–(3) in Figure 2, respectively. Firm i's best response is calculated from the first-order condition and is illustrated in Figure 3.

Case (1) in Figure 2 illustrates the scenario in which w_{-i} is below $\underline{w}_{-i}$. In this case, the competitor's waiting time is so short that decreasing w_i to below w_{-i} costs

Figure 3 Properties of Firm *i*'s Best Response Curve Following Proposition 3



firm i too much to be offset by the gain in revenue. Hence, firm i chooses a waiting time $w_i^\star$ longer than w_{-i}. Furthermore, $w_i^\star$ increases in w_{-i} due to the supermodularity of profit function P_i.

Case (2) in Figure 2 illustrates the scenario in which w_{-i} is between $\underline{w}_{-i}$ and $\bar{w}_{-i}$. In this case, firm i maximizes profit when matching firm $-i$'s waiting time.

Case (3) in Figure 2, on the other hand, illustrates the scenario in which w_{-i} is longer than $\bar{w}_{-i}$. In this case, the competitor's waiting time w_{-i} is so long that the marginal cost increase for firm i to set w_i to be below w_{-i} can be offset by the revenue increase. It follows that $w_i^\star < w_{-i}$. Furthermore, $w_i^\star$ is decreasing in w_{-i} since P_i is submodular.

In other words, bound $\underline{w}_{-i}$ can be interpreted as the shortest waiting time that firm i is able to match, whereas $\bar{w}_{-i}$ corresponds to the longest waiting time that firm i is willing to match.

We are now ready to state one of our main results, which characterizes the equilibrium of the game.

Theorem 1. *The game has the following three equilibrium scenarios*:

1. $\underline{w}_i > \bar{w}_{-i}$. *There exists a unique Nash equilibrium* (w_i^*, w_{-i}^*) *with* $w_i^* < w_{-i}^*$.
2. $\underline{w}_i \leq \bar{w}_{-i}$ *and* $\bar{w}_i \geq \underline{w}_{-i}$. *Any strategy profile* (w_i^*, w_{-i}^*) *such that*

$$w_i^* = w_{-i}^* \in [\max(\underline{w}_i, \underline{w}_{-i}), \min(\bar{w}_i, \bar{w}_{-i})]$$

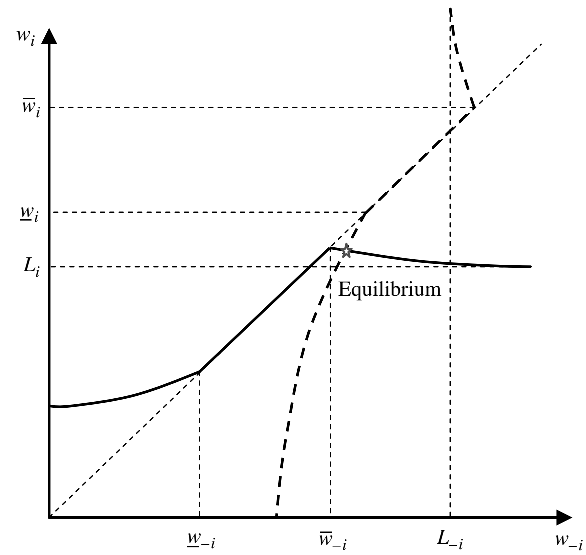
is a Nash equilibrium. In particular,

$$w_i^* = w_{-i}^* = \min(\bar{w}_i, \bar{w}_{-i})$$

is the Pareto Nash equilibrium.

3. $\bar{w}_i < \underline{w}_{-i}$. *There exists a unique Nash equilibrium* (w_i^*, w_{-i}^*) *with* $w_i^* > w_{-i}^*$.

Figure 4 Scenario 1 in Theorem 1

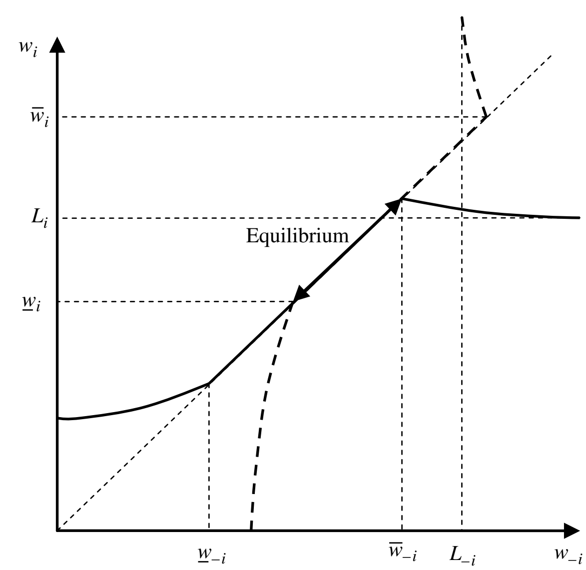


Notes. The solid curve is firm i's response curve, whereas the dashed curve is firm $-i$'s response curve. They have one intersection, which is the pure strategy Nash equilibrium represented by the star, below the 45° line.

Proof. The quasi concavity of P_i implies that there exists a pure strategy Nash equilibrium (Fudenberg and Tirole 1991). The first two scenarios presented in Theorem 1 are illustrated in Figures 4 and 5.

The scenario with $\underline{w}_i > \bar{w}_{-i}$ corresponds to Figure 4. Given that $w_i^*(w_{-i})$ decreases in w_{-i} when $w_{-i} > \bar{w}_{-i}$ and is bounded below by a positive lower bound, whereas $w_{-i}^*(w_i)$ increases in w_i when $w_i < \underline{w}_i$, there exists a unique Nash equilibrium (w_i^*, w_{-i}^*) with

Figure 5 Scenario 2 in Theorem 1



Notes. The solid curve is firm i's response curve, whereas the dashed curve is firm $-i$'s response curve. The double-sided arrow segment represents the intersection of the two response curves.

strategy profile $w_i^* < w_{-i}^*$, which can be characterized by the following first-order conditions:

$$\begin{cases} (p-c_i)[a\alpha+ae^{-1}\beta_r(\beta_\ell-1)+b\beta_r(1+e^{-w_i^*/w_{-i}^*}(\beta_\ell-1))] \\ \qquad =c_i/w_i^{*2}, \\ (p-c_{-i})a[\alpha+\beta_r+e^{-w_i^*/w_{-i}^*}\beta_r(\beta_\ell-1)(1+w_i^*/w_{-i}^*)] \\ \qquad =c_{-i}/w_{-i}^{*2}. \end{cases}$$

The scenario with $\underline{w}_i \le \bar{w}_{-i}$ and $\bar{w}_i \ge \underline{w}_{-i}$ corresponds to Figure 5. Any point in the interval $[\max(\underline{w}_i, \underline{w}_{-i}), \min(\bar{w}_{-i}, \bar{w}_i)]$ is a Nash equilibrium. We next show that a unique Pareto optimal Nash equilibrium exists. Let $P_i(w_i, w_{-i})$ be firm i's profit function and $w^* = \min(\bar{w}_{-i}, \bar{w}_i)$; then

$$P_i(w^*, w^*) > P_i(w, w^*) > P_i(w, w)$$

for any $w < w^*$. The first inequality is based on the fact that (w^*, w^*) is an equilibrium, whereas the second inequality occurs because it is always better for firm i to have firm $-i$ choose longer waiting times since

$$\frac{\partial P_i(w_i, w_{-i})}{\partial w_{-i}} = (p-c_i)b\left[\alpha+\beta_r+e^{-w_i/w_{-i}}(\beta_l-1)\left(1+\frac{w_i}{w_{-i}}\right)\right] > 0$$

for any $w_i \le w_{-i}$. The same inequalities apply to firm $-i$'s profit function. Therefore, $w_i^* = w_{-i}^* = \min(\bar{w}_i, \bar{w}_{-i})$ is the unique Pareto optimal Nash equilibrium.

The third scenario is a mirror reflection of Figure 4. □

Note that in scenario 2 of Theorem 1, when there are multiple equilibria, the equilibrium with the longest waiting time is Pareto optimal. In this scenario, the two firms match each other's waiting times, and they both set the benchmark. But with longer waiting times, both firms can reduce their operating costs. Therefore, both firms prefer the longest possible benchmark.

3.2. The Stickiness Effect

One important insight from Theorem 1 is that for a range of parameters, both firms equate their waiting times. In this section, we explore further this stickiness effect. More generally, we study the impact of benchmark effects and loss aversion on the equilibrium. To that end, we first consider the situation without a benchmark effect, that is, when $\beta_r = 0$. In this setting, the demand function becomes linear, and the model reduces to a special case of the model described in Allon and Federgruen (2007). Each firm's best response is, therefore, constant over the other firm's decision, and the game is effectively degenerate, as stated in the following proposition.

Proposition 4. *When $\beta_r = 0$, the game is degenerate, and the firms' optimal strategies are*

$$w_i^0 = (\rho_i \alpha a)^{-(1/2)}, \quad w_{-i}^0 = (\rho_{-i}\alpha a)^{-(1/2)}.$$

According to Proposition 4, in the absence of a benchmark, firms i and $-i$ generally choose different waiting times unless their profit margins are the same ($\rho_i = \rho_{-i}$). To highlight the impact of model parameters on equilibrium waiting times, we restate Theorem 1 as the following corollary:

Corollary 1. *There exist two thresholds, $\underline{R}(\beta_r, \beta_l)$ and $\bar{R}(\beta_r, \beta_l)$, such that the structure of the equilibrium strategies can be characterized by the following three intervals*:

1. *if $\rho_{-i}/\rho_i < \underline{R}(\beta_r, \beta_l)$, then $w_i^* < w_{-i}^*$;*
2. *if $\underline{R}(\beta_r, \beta_l) \le \rho_{-i}/\rho_i \le \bar{R}(\beta_r, \beta_l)$, then $w_i^* = w_{-i}^*$;*
3. *if $\rho_{-i}/\rho_i > \bar{R}(\beta_r, \beta_l)$, then $w_i^* > w_{-i}^*$;*

where

$$\underline{R} = \underline{\delta}/\bar{\delta} < 1 < \bar{R} = \bar{\delta}/\underline{\delta}.$$

Corollary 1 is a restatement of Theorem 1. (The complete derivation is presented in the online appendix.) This representation emphasizes that with the benchmark effect, the two firms will match each other's waiting time in equilibrium as long as their profit margin ratio is within a certain range. We call this phenomenon the stickiness effect. It is similar to the "focal price equilibrium" in a different setting involving price competition with a reference effect, as shown in Heidhues and Koszegi (2008). Furthermore, we denote the interval $[\underline{R}, \bar{R}]$ as the *stickiness interval*. When the profit margin ratio falls within this interval, the equilibrium waiting times of the two firms are the same.

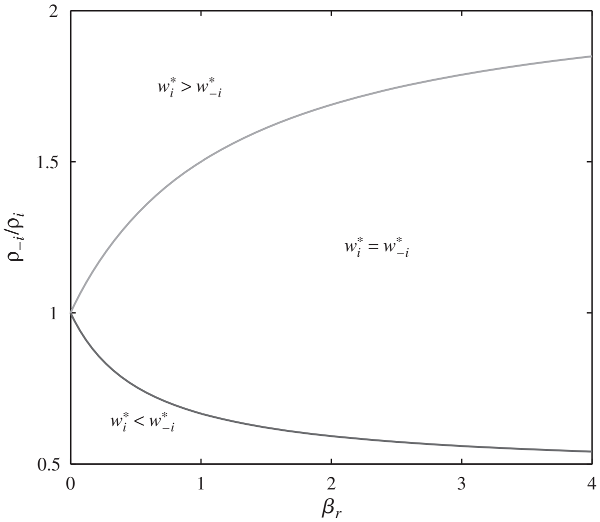
The following result characterizes the impact of the benchmark effect β_r and the loss aversion effect β_l on the stickiness interval.

Proposition 5. *$\underline{R}(\beta_r, \beta_l)$ and $\bar{R}(\beta_r, \beta_l)$ have the following structural properties*:

1. *When $\beta_r = 0$, $\underline{R} = 1 = \bar{R}$; when $\beta_r > 0$, $\underline{R} < 1 < \bar{R}$.*
2. *For fixed β_l, $\underline{R}$ is decreasing in β_r; $\bar{R}$ is increasing in β_r.*
3. *For fixed β_r,*
   (i) *if $\beta_r < \alpha$, $\underline{R}$ is decreasing in β_l; $\bar{R}$ is increasing in β_l;*
   (ii) *if $\beta_r = \alpha$, $\underline{R} = (a+b)/(2a)$, which does not change with β_l;*
   (iii) *if $\beta_r > \alpha$, $\underline{R}$ is increasing in β_l; $\bar{R}$ is decreasing in β_l.*

Proposition 5 essentially states that the stickiness interval expands with β_r, but may either expand or shrink with β_l depending on the relative strength of the benchmark effect (β_r) compared to the direct impact of waiting time (α).

Figure 6 **Combinations of ρ_{-i}/ρ_i and β_r Such That $w_i^* > w_{-i}^*$, $w_i^* = w_{-i}^*$, and $w_i^* < w_{-i}^*$**



Note. The upper curve illustrates how $\bar{R}$ changes with β_r, whereas the lower curve illustrates how $\underline{R}$ changes with β_r ($a = 1/160$, $b = 1/480$, $\rho_i = 2$, $\alpha = 1$, $\beta_l = 2$).

Figure 6 depicts the effect of β_r on the equilibrium for a particular example. Without the benchmark effect, i.e., when $\beta_r = 0$, the firms equate their waiting times if and only if $\rho_{-i}/\rho_i = 1$. As the benchmark effect gets stronger, i.e., as β_r increases, both firms are increasingly incentivized to influence the benchmark. Since a firm cannot influence the benchmark as long as its waiting time is longer than the other firm's waiting time, both firms match each other's offers and the stickiness interval expands.

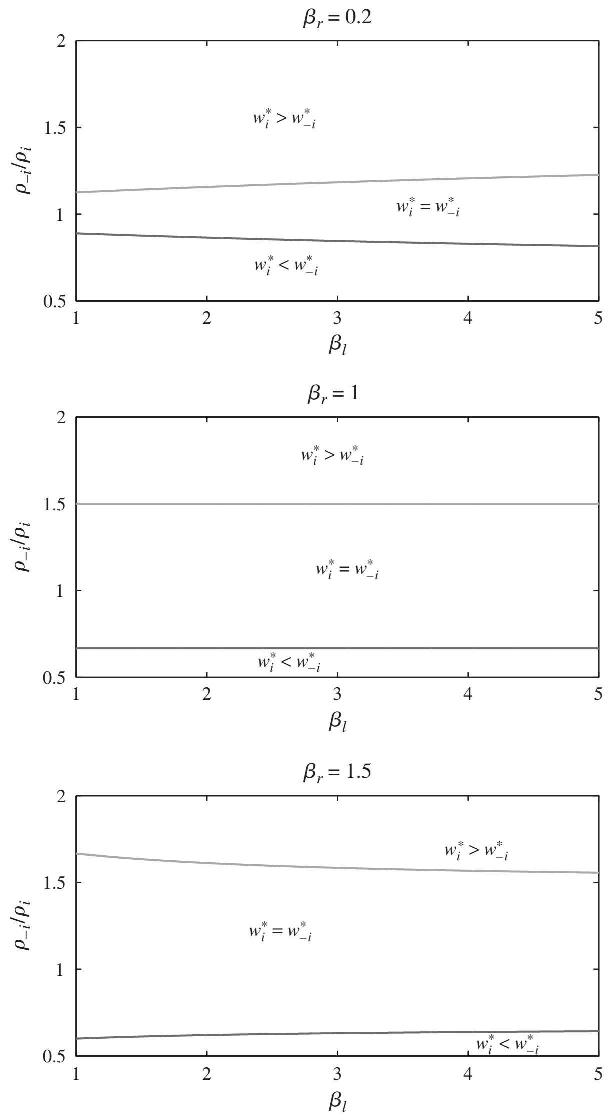
Similarly, Figure 7 illustrates the impact of β_l on the equilibrium. In general, the impact of β_l on the stickiness interval is also monotone, but the impact depends on whether or not $\beta_r > \alpha$. In particular, in the top subfigure, $\beta_r = 0.2 < 1 = \alpha$, and therefore the stickiness interval increases, illustrating case 3(i) of Proposition 5. The bottom subfigure depicts the case $\beta_r = 1.5 > \alpha$, in which the stickiness decreases (case 3(iii) of Proposition 5). When $\beta_r = \alpha = 1$, β_l has no effect on the interval, as illustrated in the middle subfigure.

Thus far, we have studied how the benchmark influences how equilibrium waiting times compare to each other. We conclude this section by analyzing the impact of β_r and β_l on $w_i^*(\beta_r, \beta_l)$ and $w_{-i}^*(\beta_r, \beta_l)$, the unique Pareto equilibrium of the game.

Proposition 6. *Unique (Pareto) equilibria $w_i^*(\beta_r, \beta_l)$, $w_{-i}^*(\beta_r, \beta_l)$ are nonincreasing in both β_r and β_l.*

Proposition 6 implies that both equilibrium waiting times are shorter with the benchmark effect than are the corresponding waiting times without the benchmark effect ($\beta_r = 0$). The benchmark effect, and also loss aversion, make customers more sensitive to delays, and thus intensify competition. This leads to shorter waiting times for customers.

Figure 7 **Combinations of ρ_{-i}/ρ_i and β_l Such That $w_i^* > w_{-i}^*$, $w_i^* = w_{-i}^*$, and $w_i^* < w_{-i}^*$ When $\beta_r = 0.2$, 1, and 1.5**



Note. In each figure, the upper curve denotes the $\bar{R}$ function, whereas the lower curve denotes the $\underline{R}$ ($a = 1/160$, $b = 1/480$, $\rho_i = 2$, $\alpha = 1$).

4. Average Benchmark and Reversal Effect

In this section, we study the case where the reference point is a weighted average of the two firms' committed waiting times, which corresponds to the so-called "product type norm" (Cadotte et al. 1987); that is, $r = \gamma w_i + (1-\gamma) w_{-i}$ for some $\gamma \in (0, 1)$. In the following, we provide conditions for the existence of a Nash equilibrium when demands are linear functions as defined by (3). Recall, however, that demand is in general piecewise linear according to Proposition 2. Unlike the minimum benchmark case (see Lemma 1), we cannot guarantee a priori that demand is linear for all firm's decisions. This means, in particular, that other equilibria may exist when demand is not linear. Nonetheless, we check numerically that the Nash equilibrium we characterize is unique when considering the full original spokes model (see Online Appendix B).

Similar to the minimum benchmark case, we define the derivative of the linear demand function $\lambda_i(\lambda_{-i})$ with respect to $w_i(w_{-i})$ at $w_i = w_{-i}$ as

$$\hat{\delta}_i \equiv -\frac{\partial \lambda_i}{\partial w_i}\bigg|_{w_i=w_{-i}} = a(\alpha-\beta_r)+[a(2-\gamma)+b\gamma]\cdot\beta_r[1+e^{-1}(\beta_l-1)],$$

$$\hat{\delta}_{-i} \equiv -\frac{\partial \lambda_{-i}}{\partial w_{-i}}\bigg|_{w_{-i}=w_i} = a(\alpha-\beta_r)+[a(1+\gamma)+b(1-\gamma)]\cdot\beta_r[1+e^{-1}(\beta_l-1)],$$

which are useful in characterizing the existence of pure strategy Nash equilibrium:

Theorem 2. *Assume that $\sqrt{b/a} < \gamma/(1-\gamma) < \sqrt{a/b}$. Following the spokes model, if $m/2 - S_L < v - p < m - S_H$, in which S_L and S_H only depend on model parameters, then there exists a pure strategy Nash equilibrium (w_i^*, w_{-i}^*) in the game. Furthermore,*
1. *if $\hat{w}_i > \hat{w}_{-i}$, then $w_i^* < w_{-i}^*$;*
2. *if $\hat{w}_i < \hat{w}_{-i}$, then $w_i^* > w_{-i}^*$;*
3. *if $\hat{w}_i = \hat{w}_{-i}$, then $w_i^* = w_{-i}^*$;*

where $\hat{w}_{-i} = (\rho_i \hat{\delta}_i)^{-1/2}$.

The definitions of S_L and S_H are provided in the proof presented in the online appendix. The proof involves showing the quasi concavity of both firms' profit functions, and then a complex combined quasi-convex and quasi-concave structure of both firms' best response functions. Condition $(m/2 - S_L, m - S_H)$ is related to condition (2), which guarantees that the Nash equilibrium described in the theorem indeed corresponds to the linear demand function on which the proof is based. Details of the proof can be found in the online appendix.

Note that the structure described in Theorem 2 differs from Theorem 1, which states that for the minimum benchmark case the two firms match each other's waiting times when the model parameters are within a certain interval. Theorem 2, on the other hand, states that equilibrium waiting times rarely are equal to each other. This is because in the average form case, both firms' decisions always directly influence the benchmark. In contrast, in the minimum case, the benchmark is not affected by the firm with the longer waiting time.

Although we are not able to prove the uniqueness of the pure strategy Nash equilibrium analytically, in Online Appendix B we tested 625 cases of model parameters that satisfy the conditions described in Theorem 2. Numerical tests show that in each of these cases, an equilibrium does exist as characterized in Theorem 2, and it is also unique. Our numerical study shows further that both waiting times at the equilibrium decrease with parameters β_r and β_l, which is also consistent with our findings for the minimum benchmark case.

We next study the impact of the benchmark effect on the equilibrium. To that end, we present the following Corollary from Theorem 2, which shows that the equilibrium waiting time commitments can be characterized by a threshold on the profit margin ratio ρ_{-i}/ρ_i.

Corollary 2. *There exists a threshold $\hat{R}(\beta_r, \beta_l)$ such that the structure of the equilibrium strategies can be characterized by the following three cases:*
1. *if $\rho_{-i}/\rho_i < \hat{R}(\beta_r, \beta_l)$, then $w_i^* < w_{-i}^*$;*
2. *if $\rho_{-i}/\rho_i = \hat{R}(\beta_r, \beta_l)$, then $w_i^* = w_{-i}^*$;*
3. *if $\rho_{-i}/\rho_i > \hat{R}(\beta_r, \beta_l)$, then $w_i^* > w_{-i}^*$;*

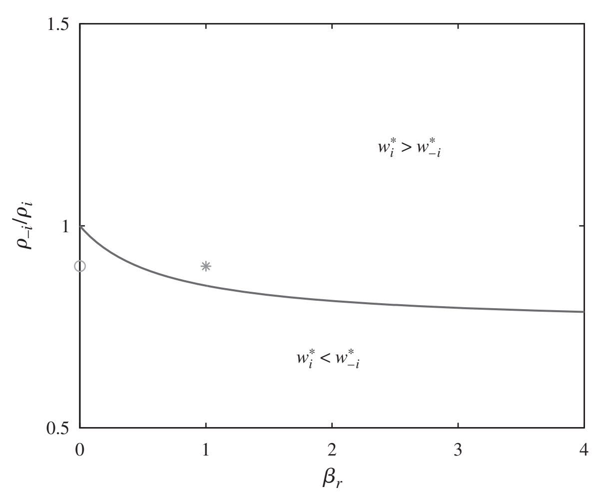
where $\hat{R} = \hat{\delta}_i/\hat{\delta}_{-i}$.

Similar to Proposition 5, we can further study how the threshold $\hat{R}$ varies with the reference effect β_r and the loss aversion effect β_l.

Proposition 7. *Without loss of generality, assume that $\gamma > 1/2$. We have the following:*
1. *When $\beta_r = 0$, $\hat{R} = 1$; when $\beta_r > 0$, $\hat{R} < 1$.*
2. *For fixed β_l, $\hat{R}$ is decreasing in β_r.*
3. *For fixed β_r,*
   (i) *if $\beta_r < \alpha$, $\hat{R}$ is decreasing in β_l;*
   (ii) *if $\beta_r = \alpha$, $\hat{R} = [a(2-\gamma)+b\gamma]/[a(1+\gamma)+b(1-\gamma)]$, which does not change with β_l;*
   (iii) *if $\beta_r > \alpha$, $\hat{R}$ is increasing in β_l.*

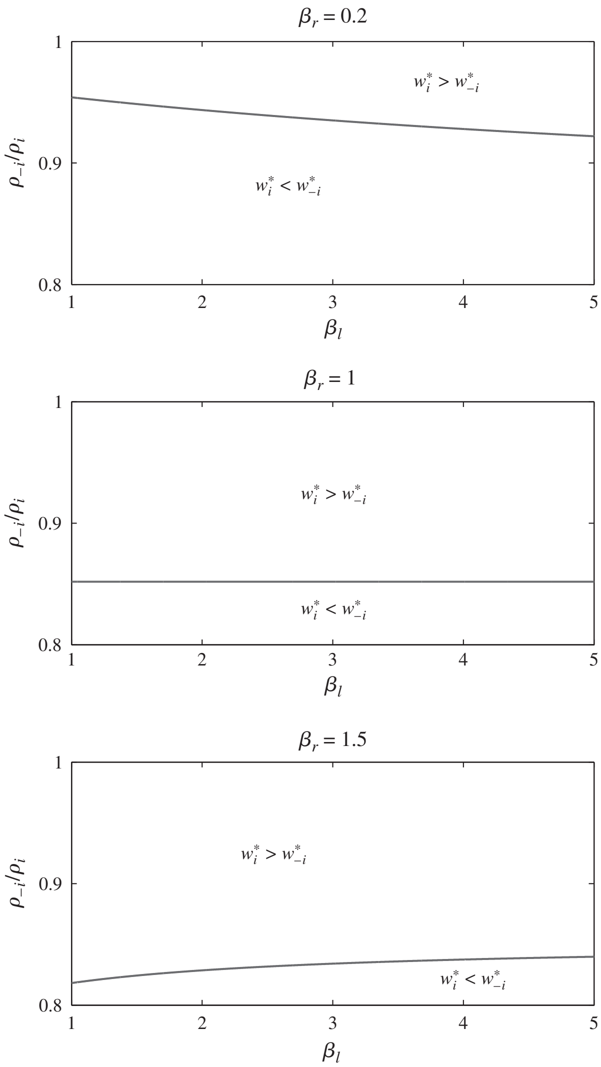
Figures 8 and 9 illustrate the proposition. The overall structure is akin to the minimum benchmark case

Figure 8 **Combinations of ρ_{-i}/ρ_i and β_r Such That $w_i^* > w_{-i}^*$ and $w_i^* < w_{-i}^*$**



Note. The decreasing curve denotes the function $\hat{R}$ ($a = 1/160$, $b = 1/480$, $\rho_i = 2$, $\gamma = 0.7$, $\alpha = 1$, $\beta_l = 2$).

Figure 9 **Combinations of ρ_{-i}/ρ_i and β_l Such That $w_i^* > w_{-i}^*$ and $w_i^* < w_{-i}^*$ When $\beta_r = 0.2$, 1, and 1.5**



Note. The curve in each figure denotes the function $\hat{R}$ ($a = 1/160$, $b = 1/480$, $\rho_i = 2$, $\gamma = 0.7$, $\alpha = 1$).

of the previous section. In particular, the threshold is increasing when the benchmark effect intensifies and either increases or decreases with loss aversion β_l, depending on whether β_r is larger or smaller than α. And there are some similarities between the threshold $\hat{R}$ and the two thresholds obtained in the minimum benchmark setting. Specifically, compare Propositions 5 and 7: if $\gamma = 1$, then the threshold $\hat{R} = \underline{R}$; on the other hand, if $\gamma = 0$, then $\hat{R} = \bar{R}$.

The previous analysis also reveals that a firm that in the absence of a benchmark effect is the market leader, i.e., it offers the shorter waiting time, can sometimes become the follower (i.e., it offers a longer waiting time) when the benchmark effect is strong enough. More specifically, without the reference effect ($\beta_r = 0$), the firms' equilibrium waiting times are at w_i^0 and w_{-i}^0 according to Proposition 4. Assume, for instance, that firm i is the leader, or, $w_i^0 < w_{-i}^0$. Following Corollary 2, firm i becomes the follower (i.e., $w_i^* > w_{-i}^*$) if the profit margin ratio ρ_{-i}/ρ_i is above threshold $\hat{R}(\beta_r, \beta_l)$. Figure 8 illustrates such a reversal effect: When profit margin ratio ρ_{-i}/ρ_i is equal to 0.9, firm i's equilibrium waiting time is smaller than firm $-i$'s for $\beta_r = 0$ (circle), but when the benchmark effect is strong enough ($\beta_r > 1$, for instance), the order is reversed with firm $-i$'s equilibrium waiting time w_{-i}^* being smaller than w_i^* (star).

We conclude this section by providing the necessary and sufficient conditions for this reversal effect to occur, as stated by the following result, which is a direct consequence of Corollary 2 and Proposition 7:

Proposition 8. *Without loss of generality, assume that $\gamma > 1/2$. There exist values of β_r and β_l such that the leader of the duopoly in the absence of a benchmark effect becomes the follower in the presence of a benchmark effect if and only if*

$$\frac{a(1-\gamma)+b\gamma}{a\gamma+b(1-\gamma)} < \frac{\rho_{-i}}{\rho_i} < 1.$$

According to the proposition, the reversal effect only occurs when the ratio of profit margins is within a certain range.

5. Price and Waiting Time Competition

We have focused thus far on situations where service firms experience a relatively high level of price rigidity compared to their ability to vary service rates (Blinder et al. 1998). In this section, we numerically show that the stickiness and reversal effects continue to exist in the presence of price competition. To account for a firm's price decision, we follow the service-level first model of Allon and Federgruen (2007) and assume that firms first select their waiting time standards (or equivalently their capacity), and then select prices.

When the two firms compete on both price and waiting time, firm i's overall demand from the spokes model becomes

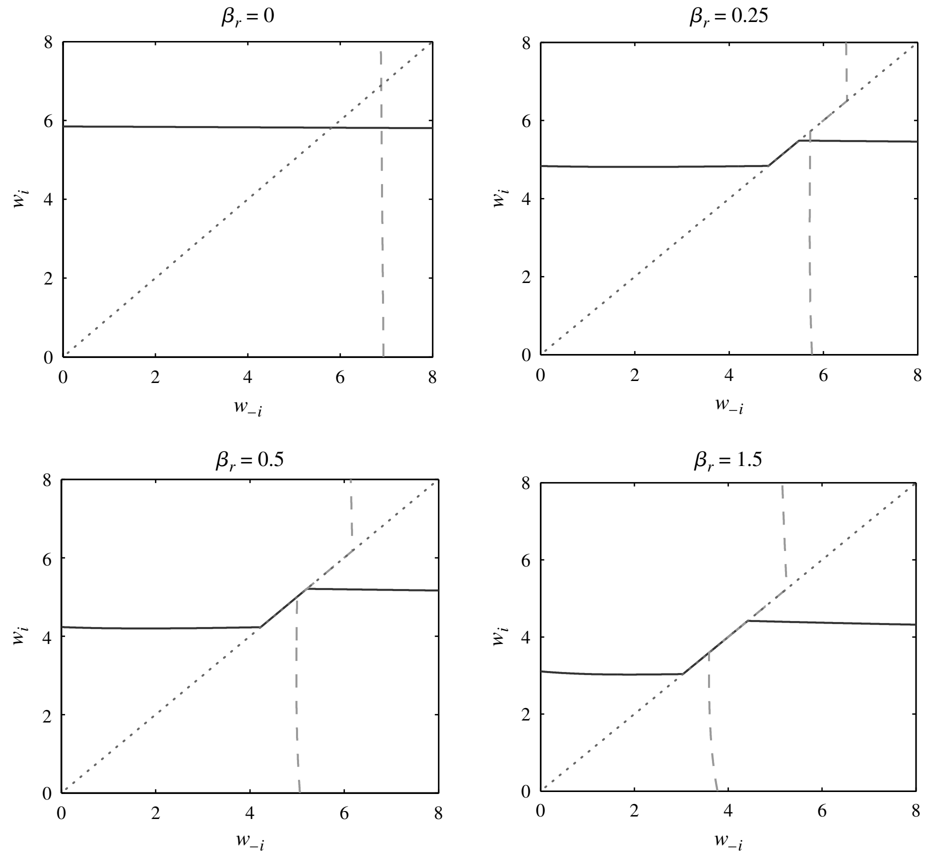
$$\lambda_i = a(S_i - p_i) - b(S_{-i} - p_{-i}) + (a-b)v + bm,$$

under condition (2). We first analyze the price competition subgame, where waiting times w_i and w_{-i}, and therefore S_i and S_{-i}, are given.

Lemma 2. *The price competition subgame has a unique Nash equilibrium (p_i^*, p_{-i}^*) with*

$$p_i^* = \big((2a^2-b^2)S_i - abS_{-i} + 2a^2c_i + abc_{-i} + (2a+b)[(a-b)v+bm]\big)\cdot(4a^2-b^2)^{-1}.$$

Figure 10 Illustration of How the Best Response Curves and the Nash Equilibria Change with β_r in the Minimum Benchmark Case



Notes. The solid and dashed curves correspond to the response curves of firm i and $-i$, respectively. The curves intersect at the Nash equilibria ($v = 70$, $a = 1/96$, $b = 1/480$, $c_i = 10$, $c_{-i} = 13$, $\alpha = 1$, $\beta_l = 2$).

Equipped with the Nash equilibrium prices of the second stage, we now turn to the first stage waiting time competition game, where firm i's profit function is

$$P_i = a(p_i^* - c_i)^2 - \frac{c_i}{w_i}.$$

The profit function is no longer quasi concave, which makes the equilibrium analysis very challenging. Thus, we resort to a numerical approach and provide examples of the stickiness and reversal effects for the minimum and average benchmark cases, respectively.

For the minimum benchmark case, Figure 10 shows cases where both firms set the same equilibrium waiting time when the benchmark effect is strong enough (e.g., β_r takes value 0.5 or 1.5). In these cases, the firm with the lower capacity cost also sets a lower price to attract more customers.

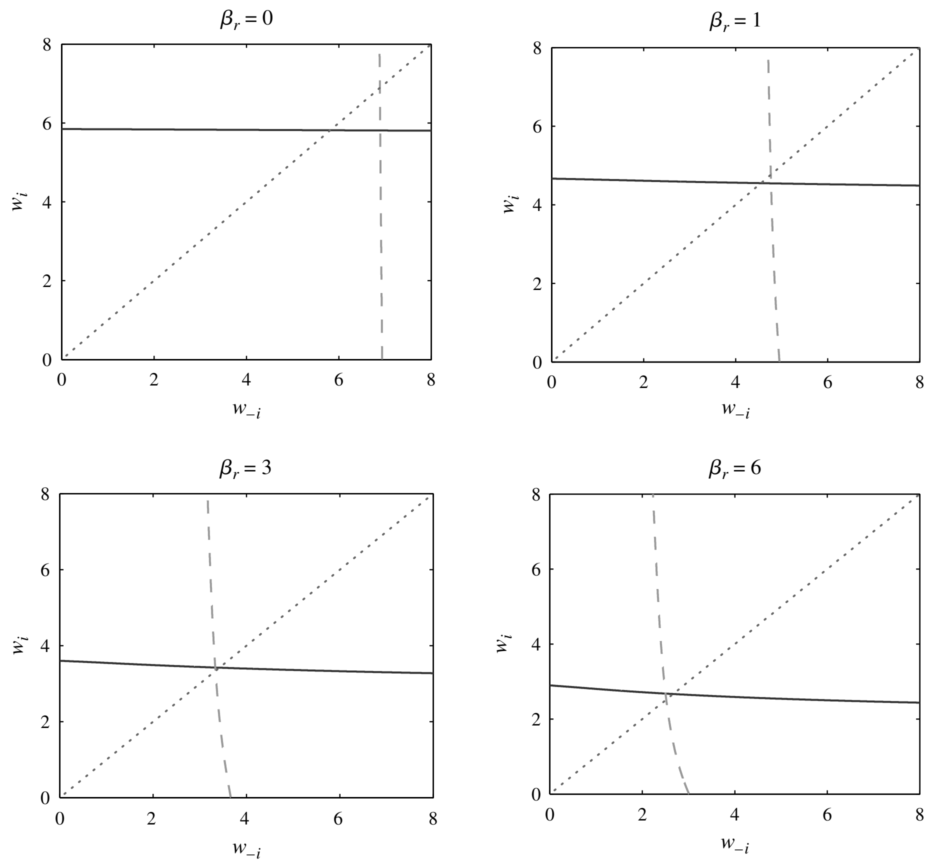
For the weighted average benchmark case, Figure 11 illustrates that when the reference effect β_r is low (say $\beta_r = 0$), the equilibrium waiting times (the intersection of the two best response curves) are such that firm i offers a shorter waiting time. When the benchmark effect increases to, for example, $\beta_r = 3$, firm i offers a longer waiting time in equilibrium instead. This demonstrates a reversal effect similar to the time-based competition case of §4. This happens when firm i has a lower capacity cost and higher influence on the benchmark ($\gamma = 0.7$). Firm i then sets a lower price to attract more customers by exploiting her cost advantage.

6. Conclusions

Empirical studies from both the marketing and decision making literatures suggest that customers can be influenced by the gap between their expected delays and a market benchmark. We posit that this aspect of customer behavior, in turn, should influence the way firms compete. The key feature of our approach is that the benchmark is endogenous. This means that companies can affect demand directly through their choice of waiting times, and indirectly by manipulating the benchmark and making the competitor's offer look worse.

Using an analytical approach, our study reveals several new findings. First, the presence of a benchmark decreases equilibrium waiting times. Second, depending on how the benchmark is formed, either a stickiness effect or a reversal effect can occur. We have

Figure 11 Illustration of How the Best Response Curves and the Nash Equilibria Change with β_r in the Weighted Average Benchmark Case ($r = \gamma w_i + (1-\gamma)w_{-i}$)



Notes. The solid and dashed curves correspond to the response curves of firm i and $-i$, respectively. The curves intersect at the Nash equilibria ($v = 70$, $a = 1/96$, $b = 1/480$, $c_i = 10$, $c_{-i} = 13$, $\gamma = 0.7$, $\alpha = 1$, $\beta_l = 1.5$).

also disentangled the impacts that loss aversion and the benchmark effect have on the equilibrium. In particular, we show that in both cases, the direction of the impact of loss aversion (β_l) changes with the strength of the benchmark effect (β_r).

Our paper appears to be the first to study benchmark effects between service firms competing on waiting times. One limitation of our model is that, following Chen and Riordan (2007), we assume that each customer's consideration set has at most two service firms. This assumption helps us to model the general situation where the competing firms do not cover the entire market, while still keeping the model tractable. Interesting extensions of our paper include generalizing our model to other queueing systems. Taking the $M/G/1$ queue, for example, the customers' aggregate satisfaction with one firm's service will depend not only on the waiting time expectation, but also on the higher moments (e.g., variance), which will greatly complicate the model and the analysis. We could also consider the segmentation of customers with respect to their degree of benchmark dependence and loss aversion or the presence of more than two firms. The equilibrium may no longer be unique for these extensions. However, we suspect that effects akin to the stickiness or reversal effects we identify in this paper should continue to exist. Finally, we have focused thus far on the average waiting time case. Possible extensions of our work also include considering other information regarding delays such as bounds on actual waiting times as in Ho and Zheng (2004).

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/msom.2013.0462.

Acknowledgments

The authors appreciate useful feedback from Wilfred Amaldoss at Duke University. The first author was supported in part by the National Natural Science Foundation of China [Grant 71201090].

References

Allon G, Federgruen A (2007) Competition in service industries. *Oper. Res.* 55(1):37–55.

Amaldoss W, He C (2010) Product variety, information advertising and price competition. *J. Marketing Res.* 47(1):146–156.

Amaldoss W, He C (2013) Pricing prototypical products. *Marketing Sci.* 32(5):733–752.

Anderson EW, Sullivan MW (1993) The antecedents and consequences of customer satisfaction for firms. *Marketing Sci.* 12(2):125–143.

Blinder AS, Canetti ERD, Lebow DE (1998) *Asking About Prices: A New Approach to Understanding Pricing Stickiness* (Russell Sage Foundation, New York).

Cachon GP, Harker PT (2002) Competition and outsourcing with scale economies. *Management Sci.* 48(10):1314–1333.

Cadotte ER, Woodruff RB, Jenkins RL (1987) Expectations and norms in models of consumer satisfaction. *J. Marketing Res.* 24(3):305–314.

Chen Y, Riordan MH (2007) Price and variety in the spokes model. *Econom. J.* 117(522):897–921.

De Vany AS, Saving TR (1983) The economics of quality. *J. Political Econom.* 91(6):979–1000.

Erdem T, Mayhew G, Sun B (2001) Understanding reference-price shoppers: A within- and cross-category analysis. *J. Marketing Res.* 38(4):445–457.

Fudenberg D, Tirole J (1991) *Game Theory* (MIT Press, Cambridge, MA).

Gans N (2002) Customer loyalty and supplier quality competition. *Management Sci.* 48(2):207–221.

Gaur V, Park YH (2007) Asymmetric consumer learning and inventory competition. *Management Sci.* 53(2):227–240.

Hall J, Porteus E (2000) Customer service competition in capacitated systems. *Manufacturing Service Oper. Management* 2(2):144–165.

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

Heidhues P, Koszegi B (2008) Competition and price variation when consumers are loss averse. *Amer. Econom. Rev.* 98(4):1245–1268.

Helson H (1964) *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior* (Harper and Row, New York).

Ho TH, Zheng YS (2004) Setting customer expectation in service delivery: An integrated marketing-operations perspective. *Management Sci.* 50(4):479–488.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Kalai E, Kamien MI, Rubinovitch M (1992) Optimal service speeds in a competitive environment. *Management Sci.* 38(8):1154–1163.

Koszegi B, Rabin M (2006) A model of reference-dependent preferences. *Quart. J. Econom.* 121(4):1133–1165.

Leclerc F, Schmitt BH, Dube L (1995) Waiting time and decision making: Is time like money? *J. Consumer Res.* 22(1):110–119.

Lin J-H, Lee T-R, Jen W (2008) Assessing asymmetric response effect of behavioral intention to service quality in an integrated psychological decision-making process model of intercity bus passengers: A case of Taiwan. *Transportation* 35:129–144.

Liu BS, Petruzzi NC, Sudharshan D (2007) A service effort allocation model for assessing customer lifetime value in service marketing. *J. Services Marketing* 21(1):24–35.

Niedrich RW, Sharma S, Wedell D (2001) Reference price and price perceptions: A comparison of alternative models. *J. Consumer Res.* 28(3):339–354.

Parasuraman A, Zeithaml VA, Berry L (1985) A conceptual model of service quality and implications for future research. *J. Marketing* 49(4):41–50.

Png IP, Reitman D (1994) Service time competition. *RAND J. Econom.* 25(4):619–634.

Popescu I, Wu Y (2007) Dynamic pricing strategies with reference effects. *Oper. Res.* 55(3):413–429.

Rajendran KN, Tellis GJ (1994) Contextual and temporal components of reference price. *J. Marketing* 58(1):22–34.

So KC (2000) Price and time competition for service delivery. *Manufacturing Service Oper. Management* 2(4):392–409.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent modal. *Quart. J. Econom.* 106(4): 1039–1061.

Wolinsky A (1986) True monopolistic competition as a result of imperfect information. *Quart. J. Econom.* 101(3):493–511.

Zeithaml VA, Berry LL, Parasuraman A (1993) The nature and determinants of customer expectations of service. *J. Acad. Marketing Sci.* 21(1):1–12.

Zhou J (2011) Reference dependence and market competition. *J. Econom. Management Strategy* 20(4):1073–1097.