MANAGEMENT SCIENCE
Vol. 58, No. 5, May 2012, pp. 932–947
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1110.1453

When to "Fire" Customers: Customer Cost-Based Pricing

Jiwoong Shin, K. Sudhir
Yale School of Management, Yale University, New Haven, Connecticut 06520
{jiwoong.shin@yale.edu, k.sudhir@yale.edu}

Dae-Hee Yoon
Yonsei School of Business, Yonsei University, Seodaemun-gu, Seoul 120-749, Korea,
dae-hee.yoon@yonsei.ac.kr

The widespread adoption of activity-based costing enables firms to allocate common service costs to each customer, allowing for precise measurement of both the cost to serve a particular customer and the customer's profitability. In this paper, we investigate how pricing strategies based on customer cost information affects a firm's customer acquisition and retention dynamics, and ultimately its profit, using a two-period monopoly model with high- and low-cost customer segments. Although past purchase and cost information helps firms to increase profits through differential prices for good and bad customers in the second period ("price discrimination effect"), it can hurt firms because strategic forward-looking consumers may delay purchases to avoid higher future prices ("ratchet effect"). We find that when the customer cost heterogeneity is sufficiently large, it is optimal for firms to "fire" some of its high-cost customers, and customer cost-based pricing is profitable. Surprisingly, it is optimal to fire even some profitable customers. This result is robust even when the cost to serve is endogenous and determined by the consumer's choice of service level. We also shed insight on acquisition–retention dynamics, on when firms can improve their profitability by selectively firing known old "bad" customers, and on replacing the old "bad" customers with a mix of new "good" and "bad" customers.

Key words: customer cost information; activity-based costing; behavior-based price discrimination; forward-looking customers; customer relationship management
History: Received January 10, 2011; accepted August 3, 2011, by J. Miguel Villas-Boas, marketing. Published online in *Articles in Advance* December 2, 2011.

> I think we will have to [discount] more . . . .
> (Sales Manager)
> Saver Superstore is a great customer. They buy lots . . . place orders on a regular basis . . . . I don't remember . . . pulling a rush order. . . . Who wouldn't want that business? (Operations Manager) (Lynch 2009)

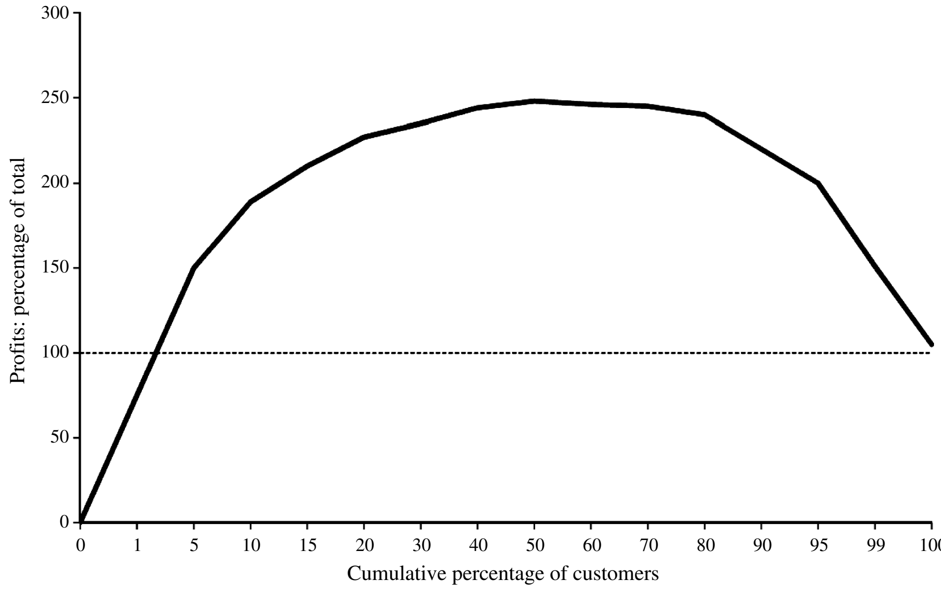
1. Introduction

Customers differ in their cost to serve. Some customers tend to be considerably more costly to serve than others: a bank customer who insists on using tellers rather than ATMs for balance verification and cash withdrawal and deposits; a retail customer who returns purchased items frequently; a Yahoo! e-mail customer who uses high levels of storage and bandwidth, taking advantage of "unlimited" disk space and bandwidth; a Netflix customer who rents an abnormally large number of DVDs, taking advantage of the "unlimited" rentals. Because firms routinely augment their core product or service with additional services as part of the purchase package at a given price, the cost of serving the customer is related to how often or how extensively they use augmented services. As in the excerpt above, a customer placing infrequent, large, or stable orders is far less costly and thus more profitable than one making frequent, small, or rush orders. For any given level of revenues, a firm would prefer to retain more of its low-cost customers and remove high-cost customers. In rare cases, firms explicitly fire high-cost customers. For example, Sprint generated much adverse publicity when it recently wrote some of its high-cost customers, who contacted customer service very often:

> The number of inquiries you have made has led us to determine that we are unable to meet your current wireless needs. Therefore after careful consideration, the decision has been made to terminate your wireless service agreement. (Srivastava 2007, p. A-3)

This paper addresses the more common situation, where firms manage customers through a customized marketing mix (Forsyth et al. 2000; Rust et al. 2000, p. 188). Specifically, we consider the case where firms implicitly "fire" customers by inducing their high-cost-to-serve, unprofitable customers to leave through customized higher prices. For example, Johnson

Figure 1 **Cumulative Profit Distribution at an Industrial Firm**



Beverage (Lynch 2009) and Fedex (Selden and Colvin 2003) selectively fire their high-cost, unprofitable customers through higher prices while offering discounts to their low-cost, profitable customers.[1]

In general, marketing scholars have not been very sensitive to the issue of differential customer cost on customer profitability because of the view that profit margins far exceed any differences in service cost. But as augmented services have become increasingly bundled with the price of the product, the relative impact of cost to serve on customer profit has gained in importance. Yet, even if a marketer was sensitive to service cost differences, accounting systems typically were not set up to track the cost to serve individual customers. Only recently with the widespread adoption of *activity-based customer costing* (ABC) are firms able to meaningfully allocate overhead costs to specific customers.

ABC has served as an eye opener about how many seemingly profitable customers destroy firm profits because of their high cost to serve (Kaplan and Anderson 2004). Figure 1 shows the inverse Lorenz curve of cumulative customer profit for an industrial firm over the cumulative percentage of customers, ordered in descending order of profitability (Kaplan 1989).[2] Accounting scholars refer to this inverted-U curve as the "whale" curve in recognition of the "hump" in the curve. The top 20% of customers contribute about 225% of customer profits, and the top 50% of customers contribute 250% of the firm's profits. The remaining 50% of customers actually destroy 150% of the firm's value. The phenomenon of profit-sapping customers is not unique to this firm; according to Kaplan and Narayanan (2001), in many business-to-business (B2B) firms, generally the top 20% of customers generate 150%–300% of the total profits, whereas the middle 70% of customers break even, and the bottom 10% of customers reduce firm profits by 50%–200%. A multi-industry study by McKinsey found that bad customers may account for 30%–40% of a typical company's revenue (Leszinski et al. 1995).

Despite the rising prevalence of customer cost-based pricing (CCP), there is little research on this topic. Our goal in this paper is to gain theoretical insight on how the availability of customer cost information will impact customer acquisition, retention, and firm profits. First, we investigate how, given cost-to-serve information, a firm should set prices over time to dynamically balance customer retention relative to acquisition. Firms often use a static picture of the cumulative profit (whale) curve as in Figure 1, but how should that curve evolve dynamically over time as firms selectively retain older customers and acquire

[1] There are other ways that firms can manage their customer profitability. For example, Royal Bank of Canada reduced service to unprofitable customers; Fidelity educates unprofitable customers to use less costly service channels (Selden and Colvin 2003, pp. 157–159). Though our focus here is pricing, we also allow customers to endogenously choose their level of service.

[2] If the *x*-axis started with the least valuable profitable customer, then it would be the Lorenz curve; the curve would be U shaped. Lorenz curves were initially used to demonstrate income inequality within countries (e.g., Atkinson 1970). Marketers have used them to show relative concentrations of sales and revenues from customers (Schmittlein et al. 1993). The key difference is that cumulative income, sales, and revenue curves are always monotonic, unlike the potentially nonmonotonic cumulative profit curve.

new ones? Should a firm "fire" its old customers at all, and if so, when? It is often suggested that firms fire unprofitable customers, but does that imply that a firm should retain all of its profitable customers?

Second, we assess the profitability of using targeted pricing strategies based on customer cost information. Should a firm use customer cost information to price discriminate among its consumers, even when consumers anticipate this and behave strategically in response to the price discrimination? More broadly, under what conditions will CCP and therefore investments in activity-based customer costing be profitable?

Finally, we examine whether the results on pricing and profits remain robust even when customers endogenously choose the level of service (and thus cost to serve), i.e., even if high-cost customers can pretend that their costs are low to avoid facing higher prices in future. How will that affect customer acquisition in the first period and the price differentials between high- and low-cost customers in the second period? How does it affect a firm's profit?

Our modeling framework focuses on whether and how a firm can profitably discriminate among customers by offering targeted prices that are based on the past costs of serving the customer. A "bad" customer with high cost will be charged a higher price than a "good" customer with low cost. Such CCP can have two effects: some bad customers will leave the firm (voluntarily "fired"), but those bad customers who choose to stay become more profitable.[3]

On the surface, the ability to price discriminate based on customer cost information should improve profits. Villas-Boas (2004), however, demonstrates that charging customers higher future prices based on their past behavior may not be profitable because customers anticipate such price increase and may defer their purchase (*ratcheting effect*). To reduce such a purchase deferral, firms are forced to lower first-period prices. This negative ratchet effect dominates the price discrimination benefit, leaving the monopolist worse relative to not using past purchase information (a similar result is identified by Hart and Tirole 1988). Which of these two effects will dominate in the presence of customer cost information remains an open question.

[3] Our focus is price discrimination based on past cost to serve. In B2B markets, some firms tailor prices to services used and costs imposed on firms (Narayanan and Brem 2002a). However, customers have negative perceptions around "nickel-and-diming" customers for traditional services such as customer support, order processing, teller services, etc. (see Narayanan and Brem 2002a, b). Even in markets where metering is routine (e.g., telephony), customers disproportionately prefer flat prices ("flat-fee bias") over metered plans based on used services (Lambrecht and Skiera 2006, Nunes 2000, Train et al. 1987).

To capture this trade-off between the benefits of price discrimination and the costs due to ratcheting, we model a two-period market with a monopoly firm facing strategic customers that can anticipate the effect of their current behavior on future prices. Essentially, our model is a two-period version of the infinite period, overlapping generations model of Villas-Boas (2004), which we augment with differential customer cost information. Thus, in the present paper, customers' past purchases not only reveal their product valuations, but also their cost to serve.

Our key results are as follows: (1) When the customer cost heterogeneity is sufficiently large, it is optimal for firms to fire some of its high-cost customers, and CCP is profitable. In contrast to conventional wisdom, it is not optimal to retain all profitable customers; some profitable customers must be fired. (2) At low levels of cost heterogeneity, a firm will not discriminate between high- and low-cost customers; all customers are retained. Yet, discrimination between existing and new customers leads to lower profits as in Villas-Boas (2004). Thus, activity-based customer costing is valuable to a firm only when customer cost heterogeneity is sufficiently large. (3) Even in the endogenous case where customers can choose the level of service and cost to serve, CCP can remain profitable. Customers are offered a higher initial price, but the future price differential between high- and low-cost customers becomes narrow relative to the exogenous case. An important distinction is that in contrast to second-degree price discrimination, price distortion relative to the exogenous case occurs for *both* high- and low-cost customers.

Overall, unlike discrimination based on only information from demand behavior, the use of customer cost information can lead to higher profits. Moreover, CCP remains profitable, even when cost to serve is endogenous. By selectively sifting out high-cost customers and retaining low-cost customers, CCP causes the whale curve of cumulative profits to dynamically evolve into a flatter curve, with few unprofitable customers. This implies that in assessing the efficacy of acquisition and retention strategies, managers should focus on the rate at which the whale curve flattens, not merely focus on the "static" current view.

2. Literature Review

This paper intersects with three key research streams: behavior-based pricing (BBP), adverse selection in marketing and economics, and activity-based costing in accounting. BBP is the practice of offering different prices based on a customer's past purchase behavior (e.g., Fudenberg and Tirole 2000; Villas-Boas 1999, 2004; Pazgal and Soberman 2008; Shin and Sudhir 2010). In contrast to our emphasis on customer cost information from purchase behavior, the BBP literature has typically focused only on demand-side information revealed through the customer's past purchases (e.g., willingness to pay or relative preference in a competitive market). In one set of models (Villas-Boas 2004, Acquisti and Varian 2005), when a monopolist faces strategic and forward-looking customers, customers choose not to purchase initially to prevent the firm from inferring their true preferences, which could be used to hurt the customer in the form of future higher prices through BBP. In equilibrium, such strategic deferral leads to lower profits relative to when firms commit to not using purchase information in pricing (Hart and Tirole 1988). Under competition, firms confront a prisoner's dilemma regarding the use of information about customer purchase history (Fudenberg and Tirole 2000, Villas-Boas 1999), making BBP unprofitable in equilibrium. Fudenberg and Villas-Boas (2006, p. 378) succinctly summarize the literature in their comprehensive review: "the seller may be better off if it can commit to ignore information about buyer's past decisions . . . more information will lead to more intense competition between firms."

A second stream of related research is the area of adverse selection in banking and credit markets (Dell'Ariccia et al. 1999, Padilla and Pagano 1997, Pagano and Jappelli 1993, Sharpe 1990, Villas-Boas and Schmidt-Mohr 1999; see Fudenberg and Villas-Boas 2006, pp. 426-429, for an excellent review). In general, this literature investigates how information about their own customers' types (i.e., ability to repay loans) learned from their relationships with customers leads to information asymmetries that impact future prices (interest rates on loans) to their own customers.[4] While the literature on adverse selection dynamics focuses on revealed customer types,[5] that on behavior-based pricing dynamics focuses on revealed customer preferences from past purchase behavior. Our paper combines the two literatures, by modeling dynamic pricing due to revealed preferences and customer types. In addition, unlike the adverse selection literature, which assumes that customer types are exogenous, we allow for customer types (cost to serve) to be endogenous.

[4] For example, Pagano and Jappelli (1993) and Padilla and Pagano (1997) focus on how information sharing about existing customers moderates adverse selection in credit markets, whereas Dell'Ariccia et al. (1999) show that incumbent's information on existing customers serves as an entry barrier for new entrants.

[5] A notable exception in accounting for customer preferences in the literature on adverse selection is Villas-Boas and Schmidt-Mohr (1999). Using a static and competitive model, they demonstrate that when banks are less horizontally differentiated, firms will invest more in customer screening because choices reveal little about customer preferences.

This work is also related to the literature on customer activity-based costing in accounting, which focuses on static models (for example, Banker and Hughes 1994, Narayanan 2003). Narayanan (2003) compares activity-based pricing with traditional pricing models in a *static* monopoly setting and concludes that activity-based pricing is beneficial when there is high variability in the cost of serving customers. Niraj et al. (2001) empirically study the profitability of a distributor supplying to several grocery and retail businesses using activity-based costing methods.[6]

Finally, the literature on CRM (customer relationship management) has focused on identifying the right customers and offering targeted value propositions (Boulding et al. 2005, Musalem and Joshi 2009) based on customer lifetime value (Gupta et al. 2004) and providing them with targeted value propositions through price (Shin and Sudhir 2010) or product (Zhang 2011). This literature demonstrates that in firms' attempts to acquire new customers, they often acquire the type of customer they wish to avoid—*bad* customers (Cao and Gruca 2005, Venkatesan and Kumar 2004). However, such acquisition of (unknown) bad customers initially should be an integral part of a dynamically optimal strategy; our analysis suggests that the key is to ensure that the whale curve progressively flattens over time. We formalize this adverse selection idea, by analyzing the dynamics of acquiring an initial mix of good and bad customers, followed by selective firing of bad customers over time, in a model where *both* the firm and customers are *strategic* and *forward looking*.

3. Model

Consider a market served by a monopolist who sells one product. The product has a constant marginal cost, which we normalize to 0 without loss of generality. The market exists for two periods, and consumers decide whether to purchase the product or not in each period. Given our focus on CCP, we assume that some customers are more costly to serve than others. Specifically, there are two customer segments: a high-type segment that costs s^H to serve and a low-type segment that costs s^L to serve ($s^L < s^H$). This exogenous fixed-cost-type assumption is relaxed later.

We account for heterogeneity in consumers' willingness to pay w by allowing it to follow a uniform distribution, $w \sim U[0, v]$, where $v > s^H$,[7] that is identical across both segments. The size of both segments is normalized to v. A consumer with willingness to pay w paying a price p for the product obtains utility

[6] Shin (2005) addresses the impact of cost to sell (rather than cost to serve) on a firm's advertising strategy.

[7] With nonzero marginal cost c, we can interpret $w = w^* - c$, as the willingness to pay net the cost of product and $w \sim U[c, v + c]$.

of $u(p \mid w) = w - p$, and thus the market demand at price p is $D(p) = v - p$.

In the first period, the firm has no specific information about individual consumers. So, it offers a single price p_1 to all consumers. At the beginning of the second period, the firm has two types of information that differentiates consumers: (1) whether they purchased from the firm in the first period and (2) how costly it is to serve customers who purchased in the first period. Given these two pieces of information, the firm can offer three different prices to three identifiable groups: (1) a price for low-cost-type customers, who purchased in the first period (p_2^L); (2) a price for high-cost-type customers, who purchased in the first period (p_2^H); and (3) a price for "others," who did not purchase in the first period (p_2^O).

Both consumers and firms are strategic and forward looking in their purchase and pricing decisions, respectively. Consumers realize that their decision to purchase in the first period can affect the price they receive in the second period. Specifically, if they expect that their price might rise in the second period because of their purchase in the first period, they may defer purchase. Because high- and low-cost customers receive different prices in the second period, their decision to purchase in the first period may differ. Similarly, when setting price in the first period, the monopolist will anticipate the effect of the second-period prices and their differential impact on the first-period purchase decisions of the high- and low-cost-type consumers.

3.1. Analysis

We solve for the firm's prices by backward induction, first by solving for the second-period prices, conditional on the consumer purchase decisions in the first period, and then for the first-period prices given the second-period solution.

3.1.1. Second Period. Let $\hat{w}_1^j$ be the first-period marginal consumer of type $j \in \{L, H\}$ who is indifferent between purchasing the product and not purchasing. Let D_2^j be the second-period demand from the type j customers who purchased in the first period. Then, the firm maximizes the following profit function in the second period:

$$\Pi_2 = (p_2^H - s^H)D_2^H(p_2^H) + (p_2^L - s^L)D_2^L(p_2^L) + (p_2^O - s^H)(\hat{w}_1^H - p_2^O) + (p_2^O - s^L)(\hat{w}_1^L - p_2^O). \quad (1)$$

The first and second terms capture profits from high- and low-cost consumers who purchased in the first period, whereas the third and fourth terms capture profits from the high- and low-cost consumers who did not purchase previously. As discussed before, the firm cannot identify the high- and low-cost consumers among those who have not purchased; the firm sets one price to them. Because each segment's purchase behavior is independent of the other, the firm can set prices p_2^H, p_2^L, p_2^O independently by maximizing profits from each segment.

First, consider the profits from the high-cost-type segment who purchased in the first period. The firm's optimal price to the high-cost type p_2^H is determined by $\arg\max_p (p - s^H)D_2^H(p)$, where $D_2^H(p) = \min(v - \hat{w}_1^H, v - p)$. The demand increases as price decreases up to $v - p$, but is truncated by an upper bound of $v - \hat{w}_1$, which is the first-period demand. It does not increase beyond the initial demand from the first period ($\hat{w}_1$), even by lowering the price below $\hat{w}_1^H$.

Suppose that the second-period price is such that $p_2^H > \hat{w}_1^H$. The first-period marginal consumer $\hat{w}_1^H$ decides not to purchase in the second period, and the demand is given by $D_2^H(p_2^H) = v - p_2^H$. We call this the "partial coverage" case because only a fraction of customers who bought in the first period purchase in the second period. In this case, the optimal price would be $p_2 = \arg\max_p (p - s^H)(v - p) = (v + s^H)/2$. Hence, when $\hat{w}_1^H < (v + s^H)/2$, the firm will charge $p_2^H = (v + s^H)/2$ in the second period.

On the other hand, if $\hat{w}_1^H \geq (v + s^H)/2$, all first-period consumers decide to purchase in the second period when the firm charges $p = (v + s^H)/2$; the first-period customers will be fully covered. Then, the firm will increase its price and charge $p_2^H = \hat{w}_1^H$, and the demand is given by $D_2^H(p) = v - \hat{w}_1^H$.

Similarly, the demand function for the low-cost type is $D_2^L(p) = \min(v - \hat{w}_1^L, v - p)$. The corresponding prices for the full and partial coverage cases are $p_2^L = \hat{w}_1^L$ and $p_2^L = (v + s^L)/2$, respectively. It is important to note that in either case (full or partial coverage), the marginal first-period customer gets zero utility in the second period. The monopolist takes advantage of the preference information revealed from customer's purchase in the first period, and the customer ends up being charged a higher price in the second period. This is the ratchet effect identified in the previous literature (Freixas et al. 1985, Fudenberg and Villas-Boas 2006).

For customers who did not purchase in the first period, the monopolist sets the price as follows:

$$p_2^O = \arg\max_p (p - s^H)(\hat{w}_1^H - p) + (p - s^L)(\hat{w}_1^L - p) = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L}{4}. \quad (2)$$

Here, the firm utilizes the fact that these customers have a lower willingness to pay than the first-period customers ($w < \hat{w}_1^j$).[8]

[8] In equilibrium, we can easily see that $p_2^O \leq \hat{w}_1^j$. This is so because $\hat{w}_1 = \hat{w}_1^L = \hat{w}_1^H$ in equilibrium (we will show this subsequently), and thus $p_2^O \leq \hat{w}_1^j \Leftrightarrow s^H + s^L \leq 2\hat{w}_1$. The last inequality always satisfies in equilibrium.

To summarize, the optimal prices in the second period are

$$p_2^j = \max\left\{\frac{v+s^j}{2}, \hat{w}_1^j\right\} \quad \text{and} \quad p_2^O = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L}{4}. \tag{3}$$

3.1.2. First Period. In the first period, a consumer with willingness to pay w decides to purchase a product if

$$w - p_1 + \delta \cdot \max\{w - p_2^j, 0\} \ge \delta \cdot \max\{w - p_2^O, 0\}. \tag{4}$$

From Equation (4), we can see that if a consumer with $\hat{w}_1$ decides to purchase a product in the first period, all consumers with $w \ge \hat{w}_1$ will also purchase a product in the first period. In other words, consumers who purchase a product for the first time in the second period must value the product less than consumers who purchase in the first period.

The marginal consumer in the first period, $\hat{w}_1^j = \hat{w}_1^j(p_1)$, can be calculated from the Equation (4) using the fact that the marginal consumer does not get any surplus in the second period if she already purchased it in the first period:

$$\hat{w}_1^j - p_1 = \delta \cdot \max\{\hat{w}_1^j - p_2^O, 0\} \Leftrightarrow \hat{w}_1^j = \begin{cases} p_1 & \text{if } p_1 < p_2^O, \\ \dfrac{p_1 - \delta p_2^O}{1-\delta} & \text{if } p_2^O \le p_1. \end{cases} \tag{5}$$

It can be easily shown that the monopolist always lowers its price to nonbuyers (p_2^O) in the second period relative to the first-period price (p_1), as in Stokey (1979) and Hart and Tirole (1988). Suppose $p_1 < p_2^O$. From Equation (5), $\hat{w}_1^j = p_1$ and $\hat{w}_1^j - p_2^O < 0$ because $p_1 < p_2^O$. Hence, $u(p_2^O \mid w) = w - p_2^O < 0$ for all $w \le \hat{w}_1^j$, which implies that there will be no new customers in the second period. But the monopolist can deviate and increase profit by lowering its second-period price p_2^O below p_1. This clearly increases the demand and profit. Hence, $p_2^O \le p_1$ in equilibrium.

Let us define $\bar{s} = (s^H + s^L)/2$ as the average cost across the high- and low-cost-type customers. The cutoff in willingness to pay for purchasing in the first period can be obtained by using the fact that $p_2^O = (\hat{w}_1^H + \hat{w}_1^L + s^H + s^L)/4$ from Equation (3):

$$\hat{w}_1 = \hat{w}_1^H = \hat{w}_1^L = \frac{p_1 - \delta p_2^O}{1-\delta} \Leftrightarrow \hat{w}_1 = \frac{2p_1 - \delta\bar{s}}{(2-\delta)}. \tag{6}$$

Note that the first-period cutoff is the same in equilibrium for both the high and low types.

The monopolist maximizes the following total discounted expected profit over the two periods (discount factor $\delta \le 1$):

$$\begin{aligned}\Pi(p_1) = {} & (p_1 - s^H)(v - \hat{w}_1) + (p_1 - s^L)(v - \hat{w}_1) \\ & + \delta\{(p_2^H - s^H)D_2^H(p_2^H) + (p_2^L - s^L)D_2^L(p_2^L) \\ & + (2p_2^O - s^H - s^L)(\hat{w}_1 - p_2^O)\},\end{aligned} \tag{7}$$

where

$$\hat{w}_1 = \frac{2p_1 - \delta\bar{s}}{2-\delta}, \quad D_2^j(p_2^j) = v - p_2^j,$$
$$p_2^j = \max\left\{\frac{v+s^j}{2}, \hat{w}_1^j\right\},$$

and

$$p_2^O = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L}{4}.$$

The first two terms in Equation (7) represent the first-period profit, and the terms within braces represent the second-period profit from its previous high-type customers, from its previous low-type customers, and from new customers who did not purchase in the first period, respectively. Note that these second-period prices are expressed as functions of the first-period marginal customer's willingness to pay ($\hat{w}_1^j$), which is itself a function of the first-period price p_1. We can now use these relationships to solve for the equilibrium prices in terms of market primitives.

3.2. Equilibrium Results

Define $v_H^{\max} = v - s^H > 0$ as the maximum extractable value from the high-cost type customer given the cost to serve s^H, and define $\Delta s = s^H - s^L$ as the difference in the cost to serve between the high- and low-cost types. Thus, Δs captures the extent of heterogeneity in cost to serve across two types. The analysis consists of two parts based on the extent of heterogeneity in cost to serve: $\Delta s \le \delta v_H^{\max}/2$ and $\Delta s > \delta v_H^{\max}/2$.

3.2.1. When Heterogeneity in Cost to Serve Is Sufficiently Small: $\Delta s \le \delta v_H^{\max}/2$. In this condition $(v+s^j)/2 \le \hat{w}_1^j$ is satisfied for both cost types $j \in \{L, H\}$ in equilibrium (we will confirm this subsequently). Therefore, the monopolist charges $p_2^j = \hat{w}_1^j$ and $D_2^j(p) = v - \hat{w}_1^j$. Using the first-order condition from Equation (7), the firm's optimal first-period price is

$$p_1 = \frac{v(4-\delta^2) + (4+2\delta+\delta^2)\bar{s}}{2(4+\delta)}. \tag{8}$$

The first-period marginal consumer's valuations are

$$\hat{w}_1 = \hat{w}_1^H = \hat{w}_1^L = \frac{(2+\delta)v + 2\bar{s}}{4+\delta}. \tag{9}$$

It can be easily seen that $(v+s^j)/2 \le \hat{w}_1^j$ in equilibrium if $\Delta s \le \delta(v - s^H)/2 = \delta v_H^{\max}/2$.

The equilibrium outcomes of the above analysis are presented in the following Lemma 1.

Lemma 1. *When the heterogeneity in cost to serve is sufficiently small such that $\Delta s \le \delta v_H^{\max}/2$, the equilibrium outcomes are as follows: using customer's past purchase and their cost information,*

$$p_1 = \frac{v(4-\delta^2) + (4+2\delta+\delta^2)\bar{s}}{2(4+\delta)};$$

$$p_2^j = \hat{w}_1 = \frac{(2+\delta)v + 2\bar{s}}{4+\delta}; \quad p_2^O = \frac{v(2+\delta) + (6+\delta)\bar{s}}{2(4+\delta)};$$

$$\Pi^{\mathrm{CCP}} = \frac{(v-\bar{s})^2(2+\delta)^2}{2(4+\delta)}.$$

Given Lemma 1, we now summarize the main findings in the following proposition.

Proposition 1. *When the heterogeneity in cost to serve is sufficiently small such that $\Delta s \le \delta v_H^{\max}/2$:*

(1) *The second-period prices for both the high- and low-cost customers who purchased in the first period are the same: $p_2^H = p_2^L = \hat{w}_1$.*

(2) *Consumers with willingness to pay $w \ge \hat{w}_1 = ((2+\delta)v + 2\bar{s})/(4+\delta)$ in both segments purchase in the first period. All of these customers will be retained in the second period.*

(3) *Consumers with willingness to pay $w \in [p_2^O, \hat{w}_1]$ in both segments purchase only in the second period.*

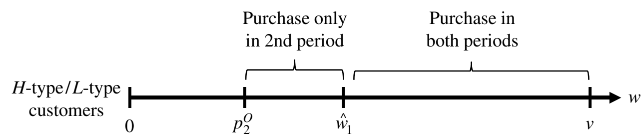
(4) *The total profit with CCP (Π^{CCP}) is lower than the profit without price discrimination (Π^{NoPD}) where the firm uses neither past purchase nor customer cost type information.*

Proof. See the appendix. □

The proposition highlights two key aspects of a firm's pricing and acquisition/retention strategies and its impact on profit. First, even with cost information, it is not optimal for the firm to price discriminate between the high- and low-cost types. This is particularly surprising given the fact that the firm would have charged different prices by cost type in the absence of purchase history information (in that case, it is easy to see that the firm would have charged $p_2^L = (v+s^L)/2$ and $p_2^H = (v+s^H)/2$).

Why does a firm charge the same price to both customer types, i.e., ignore cost information, when the cost heterogeneity is limited relative to the product valuation? The intuition is that when cost-to-serve heterogeneity is low, the effect of valuation information revealed through a customer's first-period purchase dominates any effect of the information about differential cost to serve. The entire surplus of both the high- and low-cost marginal "old" customers (who have revealed their higher willingness to pay through first-period purchase) can be extracted through a common high second-period price to both customers. This is the effect of purchase history information. Even if a firm sought to discriminate the high cost-to-serve customer by charging a higher price, that price would be still lower than the willingness to pay of the marginal high-cost customer who bought in the first period. Hence, the firm charges a common second-period price where all first-period customers (high and low cost) continue to purchase in the second period. Thus, the second-period price depends only on the first-period marginal customer's willingness to pay, which is identical across both customer type segments. Thus, the impact of information on preference revealed through purchase completely dominates the impact of information on cost to serve when the consumer valuation of the product is sufficiently high. Figure 2 illustrates this graphically.

Figure 2 **Customer's Willingness to Pay and Prices When $\Delta s \le \delta v_H^{\max}/2$**



Second, the firm is worse off using customer's past purchase and cost information. When customer heterogeneity in cost is small, the model reverts to the case in Hart and Tirole (1988) and Villas-Boas (2004), where the firm only considers the customers' past purchase information.[9] In this case, consumers are forward looking and they know that (1) they will face a lower price in the future ($p_2^O \le p_1$) if they defer purchase in the first period, and (2) they will be ripped off with a high price in the future ($p_2^j > p_1$) if they purchase (*ratchet effect*). Hence, the marginal consumers defer purchasing in the first period, and therefore the firm is worse off using customer's past purchase information, consistent with the previous literature (Acquisti and Varian 2005, Hart and Tirole 1985, Villas-Boas 2004).

3.2.2. When Heterogeneity in Cost to Serve Is Sufficiently Large: $\Delta s > \delta v_H^{\max}/2$. We next consider the case when the service cost heterogeneity is large such that $\Delta s > \delta v_H^{\max}/2$, which ensures that $\hat{w}_1^L > (v+s^L)/2$ and $\hat{w}_1^H < (v+s^H)/2$ are satisfied in equilibrium (we will confirm this subsequently). Therefore,

[9] In this case of low customer cost heterogeneity, the term *customer cost-based pricing* may be misleading because the firm *chooses* not to use customer cost information for setting the prices. The firm finds it optimal to use only the past purchase information and therefore, is equivalent to behavior-based price discrimination (Fudenberg and Tirole 2000) or pricing with customer recognition (Villas-Boas 1999).

the firm charges $p_2^L = \hat{w}_1^L$, $p_2^H = (v+s^H)/2$ and $D_2^L(p) = v - \hat{w}_1^L$, $D_2^H(p) = v-(v+s^H)/2 = (v-s^H)/2$. Taking the first-order condition from Equation (7) gives us the first-period optimal price p_1:

$$p_1 = \frac{4v(2-\delta)+8\bar{s}-\Delta s(2\delta-\delta^2)}{4(4-\delta)}. \tag{10}$$

And the first-period cutoff line is now obtained as

$$\hat{w}_1 = \hat{w}_1^H = \hat{w}_1^L = \frac{2v+2\bar{s}-\delta s^H}{4-\delta}. \tag{11}$$

By plugging $\hat{w}_1 = (2v+2\bar{s}-\delta s^H)/(4-\delta)$ in equations, we can check that $\hat{w}_1^L > (v+s^L)/2$ and $\hat{w}_1^H < (v+s^H)/2$ are satisfied in equilibrium when $\Delta s > \delta v_H^{\max}2$:

$$\hat{w}_1^L - \frac{v+s^L}{2} = \frac{\delta v_H^{\max}+(2-\delta)\Delta s}{2(4-\delta)} > 0, \tag{12}$$

$$\hat{w}_1^H - \frac{v+s^H}{2} = \frac{\delta v_H^{\max}-2\Delta s}{2(4-\delta)} < 0. \tag{13}$$

We summarize the equilibrium outcomes in the following lemma.

Lemma 2. *When the heterogeneity in cost to serve is sufficiently large such that $\Delta s > \delta v_H^{\max}/2$, the equilibrium outcomes are as follows*:

$$p_1 = \frac{4v(2-\delta)+8\bar{s}-\Delta s(2\delta-\delta^2)}{4(4-\delta)};$$

$$p_2^H = \frac{v+s^H}{2}; \quad p_2^L = \hat{w}_1 = \frac{2v+2\bar{s}-\delta s^H}{4-\delta};$$

$$p_2^O = \frac{4v+3(2-\delta)s^H+(6-\delta)s^L}{4(4-\delta)};$$

$$\Pi^{\text{CCP}} = [16(v-\bar{s})^2 + 4\delta((v-2\bar{s})^2+v^2-2s^Hs^L) - (2v(v-2s^L)+2(s^L)^2-(\Delta s)^2)\delta^2]\cdot(8(4-\delta))^{-1}.$$

Using the equilibrium outcome results in Lemma 2, we now summarize the main findings in the following propositions.

Proposition 2. *When the heterogeneity in cost to serve is sufficiently large such that $\Delta s > \delta v_H^{\max}/2$:*

(1) *The second-period price to high-cost customers is higher than the price to low cost customers. In other words, the firm will price discriminate on the basis of cost*: $p_2^H > p_2^L = \hat{w}_1$.

(2) *Consumers with willingness to pay $w \geq \hat{w}_1 = p_2^L$ in both segments purchase in the first period. All low-cost customers will be retained. Only high-type consumers with $w \geq p_2^H = (v+s^H)/2$ are retained in the second period, whereas high-type consumers with $w \in [\hat{w}_1, p_2^H]$ will be fired in the second period.*

(3) *Consumers with willingness to pay $w \in [p_2^O, \hat{w}_1]$ in both segments purchase only in the second period.*

(4) *The profit with* CCP (Π^{CCP}) *is greater than the profit without price discrimination* (Π^{NoPD}) *if the service cost heterogeneity is sufficiently large, i.e., $\Delta s > v_H^{\max}(2+\sqrt{2(4-\delta)})$; otherwise, profit is lower with* CCP.

Proof. See the appendix. □

In contrast to the case when heterogeneity in cost to serve is small, now the firm uses customer cost information in setting prices. The firm discriminates and offers different equilibrium prices to its first-period high- and low-cost customers. In particular, the firm charges a higher price to its high-cost customers ($p_2^L < p_2^H$). At this higher price for the high type, some of the first-period high-type customers with relatively low willingness to pay ($w \in [\hat{w}_1, p_2^H]$) do not purchase in the second period. The firm, thus, lets go some high-cost customers. Interestingly, there are always some profitable customers among those fired. This is because there is always some profitable customer among the high type with willingness to pay w such that $s^H < w < p_2^H$. However, all low-cost customers are retained in the second period. An alternative way of viewing the result is that the customer cost information outweighs the higher valuation information revealed from first-period purchases for the high-cost customers, but the valuation information dominates for the low-cost customers. Figure 3 illustrates this graphically.

In terms of acquisition–retention dynamics, even though the firm fires customers with moderate valuations in the range $w \in [\hat{w}_1, p_2^H]$, they acquire new customers with even lower valuations by offering a lower price to new customers. New customers with lower willingness to pay $w \in [p_2^O, \hat{w}_1]$ in both segments purchase in the second period; i.e., the firm fires known high-cost moderate valuation customers, but acquires a mix of high- and low-cost customers at a lower price, whose cost is unknown.

The impact of CCP on profits is more subtle. For moderate levels of heterogeneity in cost to serve, i.e., $\delta v_H^{\max}/2 < \Delta s < v_H^{\max}(2+\sqrt{2(4-\delta)})$, CCP reduces profits, even though the firm price discriminates

Figure 3 **Customer's Willingness to Pay and Prices When $\Delta s > \delta v_H^{\max}/2$**

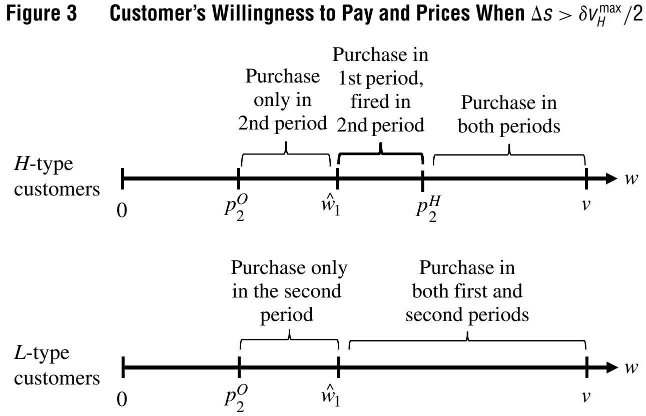
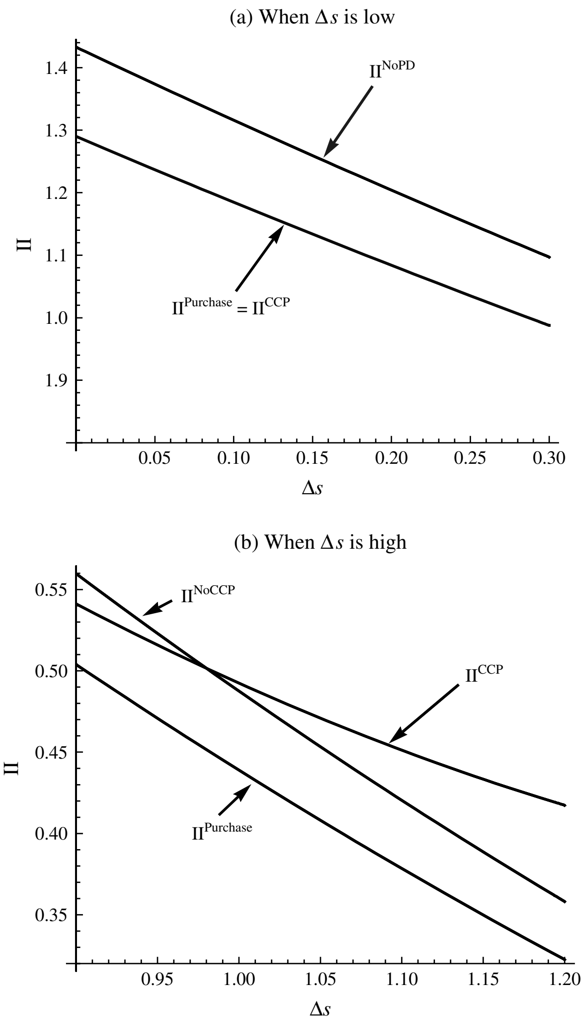


Figure 4 **Profit Comparison Between Customer Cost-Based Price Discrimination (Π^{CCP}), Purchase Behavior-Based Price Discrimination ($\Pi^{Purchase}$), and No Price Discrimination (Π^{NoPD})**



between the high- and low-cost-type customers. In this case, the ratchet effect continues to be greater than the price discrimination effect. But when the service cost heterogeneity becomes large enough, i.e., $\Delta s > v_H^{\max}(2+\sqrt{2(4-\delta)})$, the price discrimination effect outweighs the ratchet effect, and CCP becomes profitable.

To get greater clarity on how profits are affected as a function of heterogeneity in cost to serve, we plot Π^{CCP} against Δs for low and high levels of Δs for a specific set of parameters $v = 1.2$, $\delta = 0.9$. We compare these profits against two benchmarks: traditional behavior-based price discrimination based only on the past purchase history ($\Pi^{Purchase}$) and profits without price discrimination where the firm uses neither past purchase nor customer cost type information (Π^{NoPD}). The plot is shown in Figure 4.

As we know from Proposition 1, when the service cost heterogeneity is low, which is the case of Figure 4(a), the firm does not use cost information but only uses customers' purchase information. Hence, Π^{CCP} is identical to $\Pi^{Purchase}$. Moreover, this profit is lower than the profit without price discrimination (Π^{NoPD}).

In contrast, from Proposition 2, we know that when the heterogeneity in cost to serve is high, which is the case of Figure 4(b), the firm uses the customer cost information in setting price. Hence, profits under customer cost-based price discrimination and the traditional purchase behavior-based price discrimination become different. Clearly, Π^{CCP} is greater than $\Pi^{Purchase}$; the firm is better off using cost information (i.e., more information is better if the firm uses information at all). But, using the customer cost information does not necessarily increase the firm profit relative to not using the information at all because of the negative purchase deferral effects of ratcheting. Only beyond a certain threshold level of service cost heterogeneity does the gain from customer cost price discrimination overcome the ratcheting effect. Still, the price discrimination using only the customers' past purchase information makes the firm worse off.

Table 1 **Summary of Results**

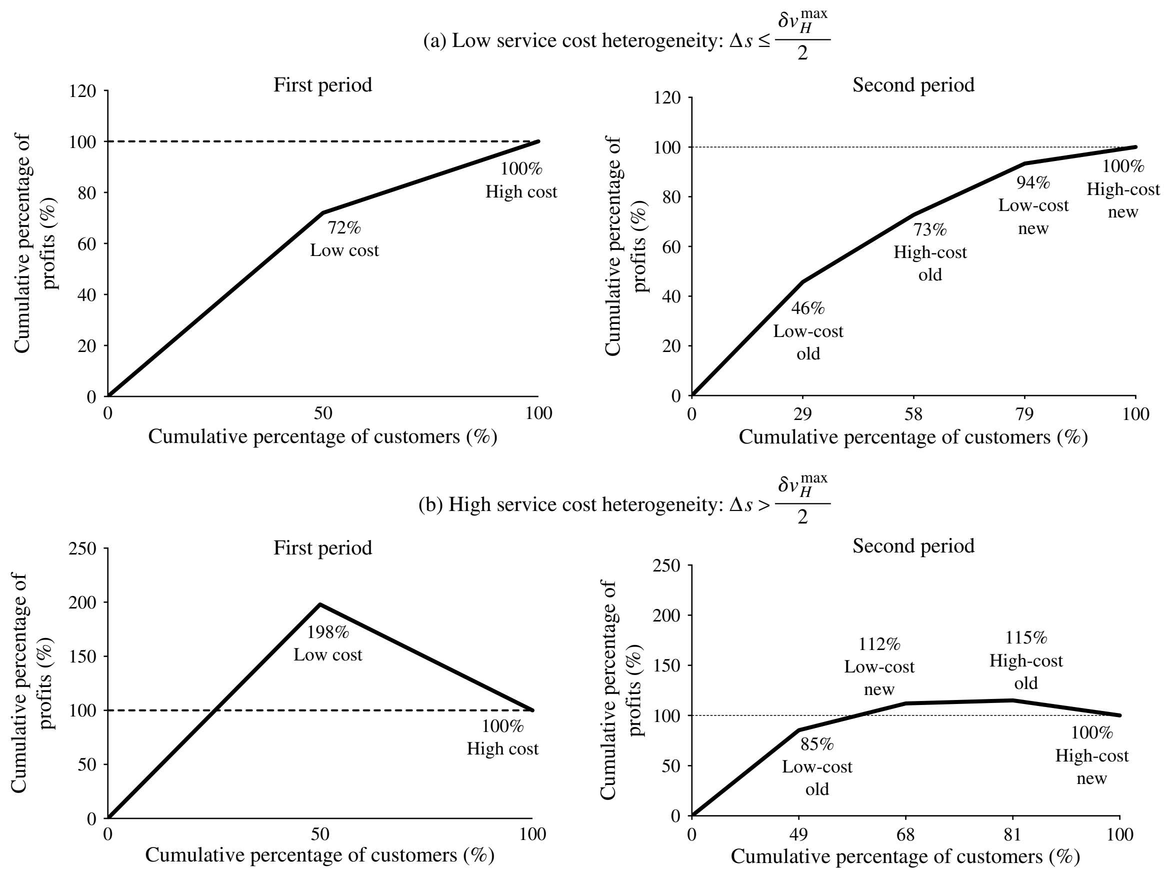
	Service cost heterogeneity ($\Delta s = s^H - s^L$)		
	Low	Moderate	High
Discrimination by cost type	No	Yes	Yes
Customer firing	No	Yes	Yes
Effect of CCP on profit	Negative	Negative	Positive

We summarize the qualitative conclusions along three dimensions in Table 1. When the heterogeneity in cost to serve is low ($\Delta s \le \delta v_H^{\max}/2$), the monopolist will *not* discriminate between high- and low-cost-type "old" customers in the second period but only discriminate between new and old customers. All customers who purchased in the first period repeat purchase in the second period, i.e., no customer is fired. However, the firm's profit is reduced by using the customer's past purchase information. On the other hand, when the heterogeneity in cost to serve is relatively large ($\Delta s > \delta v_H^{\max}/2$), the effect of cost to serve becomes more pronounced. Hence, the monopolist will discriminate customers in the second period based on customer's cost type (i.e., charging the different price for high- and low-cost old customers). Further, some of high-cost old customers, will not buy in the second period; that is, the firm fires some of its customers by raising price. In terms of profit, CCP negatively impacts profit at moderate levels of Δs, but improves profit beyond a critical value of Δs.

3.3. Numerical Example: Whale Curve Dynamics

To delve deeper into the dynamics of the acquisition and retention strategies of firms and their impact on

Figure 5 Inverse Lorenz Curves of Cumulative Profit



profits, we consider a numerical example and see how the inverse Lorenz (whale) curve evolves from the first period to the second period (Figure 5). For the example, we set $\delta = 0.9$, $v = 1.2$, and $s^L = 0$. For the low and high service cost heterogeneity cases, we set $s^H = 0.3$ and $s^H = 1$, respectively.[10]

Consider the case when heterogeneity in cost to serve is low ($\Delta s = 0.3$). In the first period, the curve is monotonically increasing; there is no "hump" in the whale curve. It follows from Lemma 1 that when customer cost heterogeneity is low, all first-period customers are profitable because the first-period price is higher than the cost to serve, even for the high type ($p_1 > s^H$). Even though the high- and low-cost customers have equal share in the customer mix, the low-cost customers contribute 72% of profits, whereas the high-cost customers contribute much less, 28% of profits.

In the second period, from Proposition 1, we know that both high- and low-cost "old" customers will be retained. The firm also acquires new customers: a mix of high- and low-cost types that it cannot identify a priori. The inverse Lorenz curve remains monotonically increasing in the second period, but the curve is flatter, indicating more equitable contribution across customers to profits. The most profitable customers are the "old" customers (low and high cost, in that order) because the value these old customers place on the product exceeds the cost differential. Together, they contribute 73% of profits, whereas the new customers contribute only 27% of the profits. Though equal in quantity sold, the 73% of profits from old customers come disproportionately from the low-cost type (46%). Also, among the new customers, we have an equal number of high- and low-cost customers, but the new low-cost customers contribute 21% of profits, whereas the new high-cost customers contribute only 6% of profits.

When the heterogeneity in cost to serve is high ($\Delta s = 1$), the first-period price is such that high-cost customers are unprofitable. We now see the hump in the "whale" curve in the first period. Even though both customers are equal in the customer mix, the 50% of low-cost customers contribute 198% of the firm

[10] The "whale curve" is typically shown as a static snapshot of the customer mix, but clearly the strategic actions of firms in terms of differential retention and acquisition of high- and low-cost customers should cause the whale curve to evolve over time. Also, please see the online technical appendix (available at http://faculty.som.yale.edu/JiwoongShin/Downloads/articles/TechnicalAppendix_WhenToFireCustomers.pdf) for exact prices, demands, and customer share for each case.

profits, whereas the remaining 50% of high-cost customers destroy 98% of the overall profits. This is very similar to the pattern we observed in Figure 1, where the top 50% of customers generate 250% of total profits, and the rest of customers destroy 150% of profits.

In the second period, as we know from Proposition 2, the firm will *not* retain all high-cost customers. The second-period whale curve becomes much flatter. With selective retention and firing, as many as 81% of customers are profitable compared to 50% of customers in the previous period. Furthermore, with customer cost information and the ability to differentially raise prices for the high-type customers, even the old high-type customers have now become profitable. In contrast to the low-heterogeneity case, where the two most profitable segments are the "old" low- and high-cost customers, here the most profitable customer segments are the "low-cost" customers: both old and new. This is because customer cost becomes relatively more important than the valuation information revealed from past purchases.

The analysis highlights the importance of selecting the right mix of customers in customer management strategy of retention and acquisition.[11] When the customer heterogeneity is low, a firm should seek to raise average retention rates across all customers. In many business-to-consumer (B2C) market situations (e.g., direct marketing, online businesses, casino gambling such as Harrah's), the cost to serve is relatively low and homogeneous, and all customers tend to be profitable; in this case, seeking high levels of retention is indeed a very profitable strategy. However, when cost to serve is relatively high and heterogeneous, firms need to do selective retention and firing by taking advantage of customer cost information and should actively induce attrition from the high-cost customer base. Overall, we find that the cumulative profit curve becomes progressively "flatter" over time under both conditions when firms optimally manage customer acquisition and retention. We suggest that managers as well as empirical scholars look for the "progressive flattening of the whale curve" as a useful diagnostic to assess the efficacy of a firm's customer management strategies over time.

4. Extension: Endogenous Service Demand

Thus far, we have assumed that the customer type (or the cost to serve a given customer) is exogenous and fixed. In reality, strategic and forward-looking customers may be able to reduce their service demand if they believe that higher demand for service may lead to higher future prices, which may prevent the firm from learning about the customer's true cost type. How would a firm's acquisition and retention strategies change if customer cost itself were endogenous? Would CCP still be profitable?

To investigate this, we relax the assumption of exogenous cost type and allow consumers to choose the level of service. The customer's strategy space is now extended to two variables: (1) whether to purchase or not, and (2) how much service to consume. Furthermore, we focus on the case when the heterogeneity in cost to serve is sufficiently large ($\Delta s > \delta v_H^{\max}/2$) because this is the region in which we found CCP to be profitable when the cost types are exogenous.

To allow endogenous service demand, we modify the consumer's utility function as follows:

$$u(p \mid w) = \begin{cases} w + \tau - p & \text{if } s = s^H, \\ w - p & \text{if } s = s^L, \end{cases} \tag{14}$$

where τ is the extra utility that a customer obtains from the firm's augmented services. We assume that this extra utility is $\tau > 0$ for H-type consumers and $\tau = 0$ for L-type consumers, which explicitly captures the incentive of the H-type customer to demand extra service. In this modified setting, high-cost-type customers may strategically conceal their type and mimic low-type customers by demanding a low level of service in the first period to get a better price in the future.

Similar to our main model, the firm maximizes the following profit function in the second period:

$$\begin{aligned} \Pi_2 = {} & (p_2^H - s^H) \cdot \min\{v + \tau - p_2^H, v - \hat{w}_1^H\} \\ & + (p_2^L - s^L) \cdot \min\{v - p_2^L, v - \hat{w}_1^L\} \\ & + (p_2^O - s^H)(\hat{w}_1^H + \tau - p_2^O) \\ & + (p_2^O - s^L)(\hat{w}_1^L - p_2^O), \end{aligned} \tag{15}$$

subject to

$$\begin{aligned} &\text{(IC-H)} \quad w + \tau - p_1 + \delta(w + \tau - p_2^H) \\ &\qquad\qquad \geq w - p_1 + \delta(w + \tau - p_2^L), \\ &\text{(IC-L)} \quad w - p_1 + \delta(w - p_2^L) \geq w + \tau - p_1 + \delta(w - p_2^H). \end{aligned}$$

The monopolist anticipates that the customer can strategically alter service demand in the first period to gain in the second period through a better price tailored to the other type. To facilitate price discrimination, the offered prices need to satisfy two incentive compatibility (IC) constraints, (IC-H) and (IC-L). (IC-H) induces the high-cost-type customers to reveal their true type, because it ensures that a

[11] Musalem and Joshi's (2009) finding that most responsive customers are not necessarily the most profitable customers and thus, a firm should focus on the right mix of customers has a similar spirit.

high-type customer does not gain by choosing the offer for the low-type customers. The (IC-L) constraint is the equivalent truth-telling constraint for the low-cost type.

The low-cost-type customer has no incentive to mimic a high-type because the low-type has no value for service ($\tau = 0$). Hence, (IC-L) is trivially satisfied in equilibrium, where $p_2^L < p_2^H$. On the other hand, the H-type customers may mimic L-type by altering service demand, if the utility loss from forgoing service in the first period (τ) is lower than the discounted gain from the price differential in the second period $\delta(p_2^H - p_2^L)$. The left-hand side of (IC-H) represents a H-type customer's total utility over two periods when he truthfully reveals his type, $w + \tau - p_1 + \delta(w + \tau - p_2^H)$, and the right-hand side is the total utility he can get when he mimics the L-type in the first period, $w - p_1 + \delta(w + \tau - p_2^L)$. Then, (IC-H) can be rewritten as

$$\begin{aligned} &w + \tau - p_1 + \delta(w + \tau - p_2^H) \\ &\quad \geq w - p_1 + \delta(w + \tau - p_2^L) \;\Leftrightarrow\; p_2^L \geq p_2^H - \frac{\tau}{\delta}. \end{aligned} \tag{16}$$

When τ is large enough (i.e., $\tau \geq \tau^{\mathrm{IC}} = ((2\Delta s - (v - s^H)\delta)\delta)/4(2-\delta)$; see the online technical appendix for the derivation), the IC constraint for H-type is not binding. Therefore, the monopolist's problem reverts to the exogenous case; the customers reveal their types even under optimal prices that the monopolist would have charged when the customer type is fixed. Hence, the equilibrium outcome is consistent with the earlier analysis, and CCP can increase a firm's profit.[12]

The more interesting and challenging case occurs when τ is small (i.e., $\tau < \tau^{\mathrm{IC}} = ((2\Delta s - (v - s^H)\delta)\delta)/4(2-\delta)$) such that the H-type customers may mimic the L-type customers. When $\tau < \tau^{\mathrm{IC}}$, the high-cost customer is less sensitive to the utility change due to downgraded service. In this case, (IC-H) is binding, and $p_2^H = p_2^L + \tau/\delta$ from Equation (16).

Plugging this into the second-period profit function in Equation (15), we obtain

$$\begin{aligned} \Pi_2 = {} & \left(p_2^L + \frac{\tau}{\delta} - s^H\right) \cdot \min\left\{v + \tau - p_2^L - \frac{\tau}{\delta},\, v - \hat{w}_1^H\right\} \\ & + (p_2^L - s^L) \cdot \min\{v - p_2^L,\, v - \hat{w}_1^L\} \\ & + (p_2^O - s^H)(\hat{w}_1^H + \tau - p_2^O) \\ & + (p_2^O - s^L)(\hat{w}_1^L - p_2^O). \end{aligned} \tag{17}$$

Also, the monopolist maximizes the following total expected profit in the first period:

$$\Pi(p_1) = (p_1 - s^H)(v - \hat{w}_1^H) + (p_1 - s^L)(v - \hat{w}_1^L) + \delta\Pi_2. \tag{18}$$

[12] See the online technical appendix for the analysis when τ is large.

We summarize the equilibrium results under endogenous service demand in the following lemma.

Lemma 3. *When service demand is endogenous and the utility from service, τ, is sufficiently small such that the (IC-H) constraint is binding (i.e., $\tau < \tau^{\mathrm{IC}} = ((2\Delta s - (v - s^H)\delta)\delta)/4(2-\delta)$),*

$$p_1 = \frac{(2-\delta)(2(2+\delta)v - (2-\delta)\tau) + (4 + 2\delta + \delta^2)(s^H + s^L)}{4(4+\delta)},$$

$$p_2^L = \frac{2((2+\delta)v + s^H + s^L) - (2-\delta)\tau}{2(4+\delta)},$$

$$p_2^H = \frac{2((2+\delta)v + s^H + s^L) - (2-\delta)\tau}{2(4+\delta)} + \frac{\tau}{\delta}, \quad \textit{and}$$

$$p_2^O = \frac{(2(2+\delta)v + (6+\delta)(s^H + s^L)) - (2-\delta)\tau}{4(4+\delta)}.$$

The marginal customers in the first period differ by customer type:

$$\hat{w}_1^L = \frac{2\left((2+\delta)v + s^H + s^L\right) - (2-\delta)\tau}{2(4+\delta)},$$

$$\hat{w}_1^H = \frac{2\left((2+\delta)v + s^H + s^L\right) - (10+\delta)\tau}{2(4+\delta)}.$$

Proof. See the appendix. □

Note that unlike the exogenous case, where customer type is fixed, now the willingness to pay of the marginal customers in the first period from the high and low types differs. This is because the high-type customer gets an extra utility τ from consuming the firm's augmented services.

Proposition 3. *Suppose that the heterogeneity in cost to serve is sufficiently large ($\Delta s > \delta v_H^{\max}/2$) and the utility from service τ is small ($\tau < \tau^{\mathrm{IC}} = ((2\Delta s - (v - s^H)\delta)\delta)/4(2-\delta)$).*

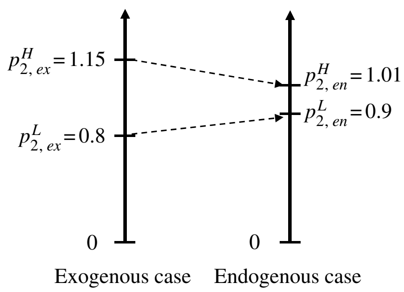
(1) *The monopolist uses* CCP *and charges different prices to the H- and L-type customers*: $p_2^H > p_2^L$.

(2) *There exists a cutoff $\tau^*(<\tau^{\mathrm{IC}})$ such that when the service utility τ is $\tau^* < \tau < \tau^{\mathrm{IC}}$, the firm's profit is greater under* CCP *than without* CCP ($\Pi^{\mathrm{CCP}} > \Pi^{\mathrm{NoPD}}$).

Proof. See the appendix. □

Our main result that CCP can increase profit is robust even if customers can endogenously choose their level of service as long as the incremental value of the service, τ, is not too small. Not surprisingly, when τ is very low such that $\tau < \tau^*$, the high-type customers can easily pretend to be low-type customers to get a better price in the future. To prevent such strategic behavior of the customers, the firm needs to distort prices to induce the high-type customer to reveal his type. But, when τ is very low ($\tau < \tau^*$), the price distortion to make high-type customers reveal their type truthfully makes the

Figure 6 Price Comparison Between Exogenous and Endogenous Cases



incremental gain from price discrimination not high enough to compensate for the negative effect of ratcheting on firm profit.

Finally, we compare prices and profits under the exogenous and endogenous cases to gain insight into how the endogenous choice of service affects a firm's pricing, retention, and acquisition strategies.[13]

Proposition 4. *Let $p_{1,\text{ex}}$, $p^L_{2,\text{ex}}$, $p^H_{2,\text{ex}}$ be the equilibrium price charged under exogenous case, and let $p_{1,\text{en}}$, $p^L_{2,\text{en}}$, $p^H_{2,\text{en}}$ be the equilibrium price charged under endogenous case.*

(1) *The price gap between the high and low types is larger under the exogenous case than that under the endogenous case: $\Delta_{\text{ex}} = p^H_{2,\text{ex}} - p^L_{2,\text{ex}} > \Delta_{\text{en}} = p^H_{2,\text{en}} - p^L_{2,\text{en}}$. Moreover, $p^L_{2,\text{ex}} < p^L_{2,\text{en}}$ and $p^H_{2,\text{ex}} > p^H_{2,\text{en}}$.*

(2) *The first-period price is higher under the endogenous case: $p_{1,\text{ex}} < p_{1,\text{en}}$. Thus, the first-period demand is lower under the endogenous case: $\hat{w}^L_{1,\text{ex}} < \hat{w}^L_{1,\text{en}}$ and $\hat{w}^H_{1,\text{ex}} < \hat{w}^H_{1,\text{en}}$.*

(3) *The profit is higher under the exogenous case: $\Pi^{\text{CCP}}_{\text{en}} < \Pi^{\text{CCP}}_{\text{ex}}$.*

Proof. See the appendix. □

The proposition provides several insights. To make the exposition clearer, Figure 6 compares the second-period prices for existing customers in the exogenous and endogenous cases for a particular numerical example when $v = 1.2$, $\tau = 0.1$, $\delta = 0.95$, $s^H = 1$, and $s^L = 0$.

The first result on the larger price gap in the exogenous case ($\Delta_{\text{ex}} > \Delta_{\text{en}}$) result is intuitive, reflecting that the firm can price discriminate more effectively in the exogenous case when it is not constrained by the incentive compatibility constraint.[14] What is particularly interesting is that unlike the standard *static* second-degree price discrimination model, where prices are distorted only for the "low" types under self-selection, here the prices for both the high- and the low-cost types change. Prices for both types move closer to each other in the endogenous case ($p^L_{2,\text{ex}} < p^L_{2,\text{en}}$ and $p^H_{2,\text{ex}} > p^H_{2,\text{en}}$). This is because, our IC constraint operates across time, in contrast to static second-degree price discrimination, where the IC constraint seeks to induce truth telling by customers in the same period (i.e., second period). Specifically, the goal of the price distortion in the second period in our *intertemporal* CCP model (based on both purchase history and customer cost type) is to induce "truth-telling" behavior by customers in the first period.

Given that the positive benefits of price discrimination in the second period is lower under the endogenous case, the firm does not invest in customer acquisition in the first period as intensely as in the exogenous case. Therefore, the first-period prices are higher (i.e., $p_{1,\text{ex}} < p_{1,\text{en}}$), and fewer customers are acquired in the first period ($\hat{w}^L_{1,\text{ex}} < \hat{w}^L_{1,\text{en}}$) under the endogenous case. Finally, the profit is higher under the exogenous case.

5. Conclusion

This paper introduces the importance of accounting for customer cost to serve in customer management. In a setting where both the firm and consumers are strategic and forward looking, we investigate how CCP impacts a firm's dynamic customer acquisition/retention strategies and profits. To this end, we develop a two-period monopoly model where a firm can set prices in the second period based on customer actions in the first period (purchase and cost to serve).

We find that the emphasis on acquisition versus retention should differ as a function of heterogeneity in cost to serve. When heterogeneity in cost to serve is small, all customers tend to be profitable, and the firms should retain all their current customers. Such a scenario is potentially true in many B2C markets (e.g., direct marketing, online businesses, casino gambling such as Harrah's) where the impact of service cost heterogeneity is small relative to the product's profit margin. In contrast, when this heterogeneity is sufficiently large as with financial institutions (e.g., Royal Bank of Canada) and B2B markets (e.g., Fedex, Johnson Beverages), it pays to selectively "fire" high-cost customers by raising their prices while offering a lower price to lower-cost customers. Interestingly, we find that firms may fire even profitable customers, when the cost-to-serve

[13] Because we modify the consumer's utility function to allow endogenous service demand (Equation (14)), the exogenous results should be reanalyzed accordingly. Under the exogenous case, the firm does not need to consider the incentive constraints (IC-H). Hence, the equilibrium results under exogenous case are the same as the case in which the utility from service τ is sufficiently large that the (IC-H) constraint will always be satisfied. We derive the exogenous case results in the online technical appendix.

[14] Note that the price difference is smaller than the cost difference for both the exogenous and endogenous cases ($s_H - s_L > \Delta_{\text{ex}} > \Delta_{\text{en}}$) when $\tau < \tau^{\text{IC}}$. See the online technical appendix for the proof.

differential is large. Our analysis also suggests that the common apprehension among practitioners that firing customers may lead to allocation of short-term fixed service costs among fewer customers (making them also unprofitable) is misplaced, because the firing is accompanied by new customer acquisition of a mix of good and bad customers, who are on average more profitable than the sure "bad" customers that are fired.

Second, CCP can be profitable when the service cost heterogeneity across customers is high. This result complements the finding in the existing literature that suggests that a monopolist using purchase history when setting prices will obtain lower profits if consumers are strategic and forward looking. This paper shows that when the purchase history is augmented with customer cost, CCP can become profitable. We also demonstrate that the results remain robust even if the level of service demand (and, therefore, cost to serve) is endogenous.

Finally, this paper provides insight into the dynamics of the customer mix by showing the evolution of the inverse Lorenz curve of cumulative customer profit (the "whale" curve). Overall, we see that CCP flattens the whale curve in all scenarios, making customer contributions to profits more equal. When heterogeneity in cost to serve is small, all consumers are profitable, and there is no hump in the "whale" curve; the most valuable information about existing customers is that they have higher willingness to pay than potential new customers. Given their higher willingness to pay, these existing customers remain the most profitable (low and high cost, in that order), and therefore, all of them should be retained. However, when heterogeneity in customer cost to serve is large, customer cost information revealed is relatively more valuable. The existing high-cost customers (with higher willingness to pay) are not as valuable as the new potential low-cost customers (with lower willingness to pay). Hence the *low-cost* customers (old and new, in that order) become the most profitable customers in the second period. Therefore, the firm should selectively fire more of the high-cost customers. Our analysis highlights why managers should judge the efficacy of their customer acquisition and retention strategies, not through a static view of the whale curve, but by checking whether the whale curve becomes flatter over time.

This paper is an initial attempt at studying the impact of customer cost information revealed through activity-based customer costing. The modeling can be extended along a number of dimensions. Currently, we use a two-period monopoly framework. Two natural areas of extension would be (1) to consider an overlapping generations model, where new consumers arrive at a steady rate and the monopolist has to model retaining an old generation of customers in steady state (e.g., Villas-Boas 2004), and (2) to extend it to a competitive market (e.g., Fudenberg and Tirole 2000, Villas-Boas 1999, Shin and Sudhir 2010). Also, to focus on customer cost heterogeneity, we assumed identical willingness to pay distributions across the high- and low-cost segments. Future work can allow for the possibility of higher willingness to pay for the high-cost segment. Also, we assumed that firms can perfectly "identify" old customers in the second period. In many B2C markets, customer identification is unlikely to be perfect. Understanding how this affects firm strategies would be a fruitful research venue.

In this paper, customers are price discriminated based on past cost to serve. Alternately, one could consider pricing based on current activities or service usage. Although consumers do have an aversion for being nickel-and-dimed, or metered constantly for service use (for example, Train et al. 1987), a systematic investigation of the trade-offs between the two pricing approaches would be an interesting area of future research. Finally, customer costs can also affect other marketing tactics. For example, if churn were to occur naturally among highs- and low-cost customers, and firms can invest in advertising to manage churn rate, how should the firm trade off between retention and new customer acquisition advertising? We hope that the current paper serves as an impetus to broader investigation of the impact of customer activity-based costing.

Acknowledgments
The authors thank the department editor, the associate editor, and two anonymous reviewers for very constructive comments during the review process. They also thank Dina Mayzlin, Raphael Thomadsen, Jake Thomas, Robert Zeithammer, and seminar participants in the Eighth Choice Symposium and marketing seminar at Yale University for their comments.

Appendix

Proof of Proposition 1. (4) Without price discrimination, the monopolist simply maximizes the following per-period profit function $\Pi_t = (p - s^H)(v - p) + (p - s^L)(v - p)$. The optimal price is $p^S = (v + \bar{s})/2$, and the total profit will be $\Pi^S = ((1+\delta)(v-\bar{s})^2)/2$. It immediately follows that $\Pi^S = ((1+\delta)(v-\bar{s})^2)/2 > ((v-\bar{s})^2(2+\delta)^2)/2(4+\delta) = \Pi^*$ for all $\delta \leq 1$. □

Proof of Proposition 2. (4) From the proof of Proposition 1, we know that $\Pi^S = ((1+\delta)(v-\bar{s})^2)/2$. It immediately follows that $\Delta\Pi = \Pi^{\text{CCP}} - \Pi^S = \delta(2\delta(v-s^H)^2 + (\Delta s)^2 - 4v(v-2\bar{s}) - 4s^H s^L)/(8(4-\delta))$. Therefore, $\Delta\Pi \geq 0$ if and only if

$$s^H \leq \frac{3s^L - 2v(1-\delta) - (v-s^L)\sqrt{2(4-\delta)}}{1+2\delta} \quad \text{or}$$

$$s^H \geq \frac{3s^L - 2v(1-\delta) + (v-s^L)\sqrt{2(4-\delta)}}{1+2\delta}$$

$$\Leftrightarrow \quad s^H - s^L \le \frac{-(v-s^L)\{\sqrt{2(4-\delta)}+2(1-\delta)\}}{1+2\delta} \quad \text{or}$$

$$s^H - s^L \ge \frac{(v-s^L)\{\sqrt{2(4-\delta)}-2(1-\delta)\}}{1+2\delta}.$$

Also, $(-(v-s^L)\{\sqrt{2(4-\delta)}+2(1-\delta)\})/(1+2\delta) < 0$ and $\delta v_H^{\max}/2 < ((v-s^L)\{\sqrt{2(4-\delta)}-2(1-\delta)\})/(1+2\delta)$ because $v_H^{\max} < v_L^{\max} = v - s^L$ and $\delta(1+2\delta) < 2\delta(1+2\delta)$ for all $\delta \in [0,1]$. Also, note that $v_L^{\max} = v - s^H + s^H - s^L = v_H^{\max} + \Delta s$. Using this, we can see that $\Delta\Pi = \Pi^{\text{CCP}} - \Pi^S > 0$ if

$$\Delta s \ \ge\ v_L^{\max}\frac{\sqrt{2(4-\delta)}-2(1-\delta)}{1+2\delta}$$

$$\Leftrightarrow v_H^{\max}(2+\sqrt{2(4-\delta)}) \le \Delta s. \quad \square$$

Proof of Lemma 3. First, note that in Equation (17), as we will confirm below, $\min\{v-p_2^L, v-\hat{w}_1^L\} = v - \hat{w}_1^L$ and, therefore, $p_2^L = \hat{w}_1^L$ and $p_2^H = \hat{w}_1^L + \tau/\delta$ in equilibrium. This is so because $\hat{w}_1^L > p_2^{L*}$, where $p_2^{L*} = \arg\max_{p^L}(p^L + \tau/\delta - s^H)(v+\tau-p^L-\tau/\delta) + (p_2^L - s^L)(v-p_2^L) = ((s^H+s^L+2v)\delta - (2+\delta)\tau)/(4\delta)$; that is, p_2^{L*} is the optimal price the monopolist would charge if $\min\{v-p_2^L, v-\hat{w}_1^L\} = v - p_2^L$ using customer's past purchase and their cost information. Furthermore, we can see that $\hat{w}_1^L + \tau/\delta - \tau > \hat{w}_1^H$ is satisfied in equilibrium and, therefore, $\min\{v+\tau-\hat{w}_1^L-\tau/\delta, v-\hat{w}_1^H\} = v+\tau-\hat{w}_1^L-\tau/\delta$.[15] Also, $p_2^O = \arg\max_p(p-s^H)(\hat{w}_1^H - p + \tau) + (p-s^L)(\hat{w}_1^L - p) = (\hat{w}_1^H + \hat{w}_1^L + s^H + s^L + \tau)/4$. Taking the first-order condition, we obtain the optimal second-period prices:

$$p_2^L = \hat{w}_1^L, \quad p_2^H = \hat{w}_1^L + \frac{\tau}{\delta},$$
$$p_2^O = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L + \tau}{4}.$$

Similar to our main model, the marginal consumer in the first period, $\hat{w}_1^j = \hat{w}_1^j(p_1)$, can be calculated by using the fact that $p_2^O \le p_1$ and that the marginal consumer does not get any surplus in the second period if she already purchased it in the first period: $\hat{w}_1^H + \tau - p_1 = \delta(\hat{w}_1^H + \tau - p_2^O) \Leftrightarrow \hat{w}_1^H = (p_1 - \tau - \delta(p_2^O - \tau))/(1-\delta)$, and $\hat{w}_1^L - p_1 = \delta(\hat{w}_1^L - p_2^O) \Leftrightarrow \hat{w}_1^L = (p_1 - \delta p_2^O)/(1-\delta)$. Therefore, the willingness to pay for the marginal customer in the first period can be obtained by using the fact that $p_2^O = (\hat{w}_1^H + \hat{w}_1^L + s^H + s^L + \tau)/4$, where $\hat{w}_1^H = (2p_1 - \delta\bar{s})/(2-\delta) - \tau$, $\hat{w}_1^L = (2p_1 - \delta\bar{s})/(2-\delta)$.

Also, taking the first-order condition from Equation (18), we obtain the first-period price p_1:

$$p_1 = \frac{(2-\delta)(2(2+\delta)v - (2-\delta)\tau) + (4+2\delta+\delta^2)(s^H+s^L)}{4(4+\delta)}. \tag{19}$$

Using this first-period price, the first-period marginal consumer has a willingness to pay as follows:

$$\hat{w}_1^L = \frac{2((2+\delta)v + s^H + s^L) - (2-\delta)\tau}{2(4+\delta)},$$
$$\hat{w}_1^H = \frac{2((2+\delta)v + s^H + s^L) - (10+\delta)\tau}{2(4+\delta)}. \tag{20}$$

Here, we can check that the conditions $p_2^{L*} = ((s^H + s^L + 2v)\delta - (2-\delta)\tau)/(4\delta) < \hat{w}_1^L$ and $p_2^H - \tau = \hat{w}_1^L + \tau/\delta - \tau > \hat{w}_1^H$ are satisfied in equilibrium:

$$\hat{w}_1^L - p_2^{L*} = \frac{\delta^2\{2v - (s^H + s^L)\} + (8 - 6\delta + \delta^2)\tau}{4\delta(4+\delta)} > 0,$$
$$\hat{w}_1^H - \left(\hat{w}_1^L + \frac{\tau}{\delta} - \tau\right) = -\frac{\tau}{\delta} < 0. \quad \square$$

Proof of Proposition 3. First, we calculate the profit under no discrimination. Without price discrimination, the monopolist simply maximizes the following per-period profit function $\Pi_t^S = (p^S - s^H)(v - p^S + \tau) + (p^S - s^L)(v - p^S)$. The optimal price is $p^S = (2v + s^H + s^L + \tau)/4$, and the total profit is $\Pi^S = (1+\delta)[\Pi_t^S] = (1+\delta)([(2v - s^H - s^L)^2 + 2(2v - 3s^H + s^L)\tau + \tau^2])/8$. Using the results of p_1, p_2^H, and p_2^L in Lemma 3, we get

$$\begin{aligned}\Pi^{\text{CCP}} = \frac{1}{8\delta(4+\delta)}\big\{&((2v - s^H - s^L)^2\delta A)^2 \\ &+ 2\delta\big(2v(4+\delta A) - s^L\big(4 - \delta(4+A)\big) \\ &\qquad - s^H\big(4 + \delta(10+3\delta)\big)\big)\tau \\ &- \big(32 - \delta(28 + \delta(2+A))\big)\tau^2\big\},\end{aligned}$$

where $A = 2 + \delta$. It immediately follows that

$$\begin{aligned}\Delta\Pi &= \Pi^{\text{CCP}} - \Pi^S \\ &= \big[\delta\big(2v^2(2-\delta) + 4v\tau - (s^H)^2(1+2\delta) - (s^L)^2 \\ &\quad - s^L(4v - 2\tau) - \tau^2 + s^H(6s^L - 4v(1-\tau) + 6\tau)\big)\big] \\ &\quad \cdot (8(4-\delta))^{-1},\end{aligned}$$

where $\Delta\Pi$ is a concave function of τ. So if we show that $\Delta\Pi$ is positive when $\tau = \tau^{\text{IC}}$, there exists τ^* that makes $\Pi^* > \Pi^S$. First, we note that when $\tau = 0$, $\Delta\Pi = -(2v - s^H - s^L)^2/8(4+\delta) < 0$. Also, when $\tau = \tau^{\text{IC}}$, $\Delta\Pi$ is positive when $\Delta s > \Delta s^* = ((v - s^L)(32 - 16\sqrt{2(4-\delta)} - \delta(8(8 - \sqrt{2(4-\delta)}) - \delta(18+\delta))))/(-16 - \delta(44 - \delta(16+\delta)))\ (>(\delta(v - s^H))/2)$. Hence, as Δs becomes larger, $\Pi^{\text{CCP}} > \Pi^S$. Therefore, when $\Delta s > \Delta s^*$, there exists τ^* so that $\Pi^{\text{CCP}} > \Pi^S$ when $\tau^* < \tau < \tau^{\text{IC}}$.[16] $\square$

Proof of Proposition 4. (1) $\Delta = p_{2,\text{ex}}^H - p_{2,\text{ex}}^L = (2(s^H - s^L) - \delta(v - s^H) + 2\tau)/(8 - 2\delta)$ and $\Delta_{\text{en}} = p_{2,\text{en}}^H - p_{2,\text{en}}^L = \tau/\delta$. If we compare Δ_{ex} and Δ_{en}, $\Delta_{\text{ex}} > \Delta_{\text{en}}$ when $\tau < \tau^{\text{IC}}$. Also, $p_{2,\text{en}}^L - p_{2,\text{ex}}^L = (\delta(2(s^H - s^L) - \delta(v - s^H) + 4\tau) - 8\tau)/(16 - \delta^2) > 0$ and $p_{2,\text{en}}^H - p_{2,\text{ex}}^H = -(\delta(2(s^H - s^L) - \delta(v - s^H) + 4\tau) - 8\tau)/(2\delta(4+\delta)) < 0$ when $\tau < \tau^{\text{IC}}$.

(2) $\hat{w}_{1,\text{ex}}^L - \hat{w}_{1,\text{en}}^L = -(2\delta(s_H - s_L) - \delta^2(v - s^H) - 4(2-\delta)\tau)/(16 - \delta^2) < 0$ and $\hat{w}_{1,\text{ex}}^H - \hat{w}_{1,\text{en}}^H = -(2\delta(s_H - s_L) - \delta^2(v - s^H) - 4(2-\delta)\tau)/(16 - \delta^2) < 0$ when $\tau < \tau^{\text{IC}}$. Also, $p_{1,\text{ex}} - p_{1,\text{en}} = -((2-\delta)^2\tau)/4(4-\delta) < 0$.

(3) $\Pi_{\text{ex}}^{\text{CCP}} - \prod_{\text{en}}^{\text{CCP}} = (2\delta(s^H - s^L) - \delta^2(v - s^H) - 4(2-\delta)\tau)^2/(4\delta(16 - \delta^2)) > 0$ when $\tau < \tau^{\text{IC}}$. $\square$

[15] Note that the firm would charge $p_2^H = p_2^{H*} = \arg\max_{p^H}(p^H - s^H)\cdot(v + \tau - p^H) = (v + s^H + \tau)/2$ without considering the (IC-H) condition. However, when $\tau \ge \tau^{\text{IC}}$, we can also easily check that in equilibrium $p_2^H = \hat{w}_1^L + \tau/\delta < p_2^{H*}$, and therefore, the (IC-H) condition in Equation (16) does not hold.

[16] For example, when $v = 3$, $s^H = 2.9$, $s^L = 2$, and $\delta \simeq 1$, we can see that $\Pi^{\text{CCP}} > \Pi^S$ if $0.167 < \tau \le \tau^{\text{IC}} = 0.425$.

References

Acquisti, A., H. R. Varian. 2005. Conditioning prices on purchase history. *Marketing Sci.* **24**(3) 367–381.

Atkinson, A. B. 1970. On the measurement of inequality. *J. Econom. Theory* **2**(3) 244–263.

Banker, R. D., J. S. Hughes. 1994. Product costing and pricing. *Accounting Rev.* **69**(3) 479–494.

Boulding, W., R. Staelin, M. Ehret, W. J. Johnston. 2005. A customer relationship management roadmap: What is known, potential pitfalls, and where to go. *J. Marketing* **69**(4) 155–166.

Cao, Y., T. S. Gruca. 2005. Reducing adverse selection through customer relationship management. *J. Marketing* **69**(October) 219–229.

Dell'Ariccia, G., E. Friedman, R. Marquez. 1999. Adverse selection as a barrier to entry in the banking industry. *RAND J. Econom.* **30**(3) 515–534.

Forsyth, J., A. Gupta, S. Haldar, M. V. Marn. 2000. Shedding the commodity mind-set. *McKinsey Quart.* (November) 78–85.

Freixas, X., R. Guesnerie, J. Tirole. 1985. Planning under incomplete information and the ratchet effect. *Rev. Econom. Stud.* **52**(2) 173–191.

Fudenberg, D., J. Tirole. 2000. Customer poaching and brand switching. *RAND J. Econom.* **31**(4) 634–657.

Fudenberg, D., J. M. Villas-Boas. 2006. Behavior-based price discrimination and customer recognition. T. J. Hendershott, ed. *Handbook on Economics and Information Systems*, Vol. 1. Elsevier, Amsterdam, 377–436.

Gupta, S., D. R. Lehmann, J. A. Stuart. 2004. Valuing customers. *J. Marketing* **41**(1) 7–18.

Hart, O. D., J. Tirole. 1988. Contract renegotiation and coasian dynamics. *Rev. Econom. Stud.* **55**(4) 509–540.

Kaplan, R. S. 1989. Kanthal (A). HBS Case 9-190-002, July 27. Harvard Business School, Boston.

Kaplan, R. S., S. R. Anderson 2004. Time-driven activity-based costing. *Harvard Bus. Rev.* **82**(November) 131–138.

Kaplan, R. S., V. G. Narayanan. 2001. Measuring and managing customer profitability. *J. Cost Management* **15** 5–15.

Lambrecht, A., B. Skiera. 2006. Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *J. Marketing Res.* **43**(2) 212–223.

Leszinski, R., F. A. Weber, R. Paganoni, T. Baumgartner. 1995. Profits in your backyard. *MacKinsey Quart.* (November) 118–127.

Lynch, L. J. 2009. Johnson Beverage, Inc. Darden Business Publishing Case UVA-C-2292, University of Virginia, Charlottesville.

Musalem, A., Y. V. Joshi. 2009. How much should you invest in each customer relationship? A competitive strategic approach. *Marketing Sci.* **28**(3) 555–565.

Narayanan, V. G. 2003. Activity-based pricing in a monopoly. *J. Accounting Res.* **41**(3) 473–502.

Narayanan, V. G., L. Brem. 2002a. Case study: Customer profitability and customer relationship management at RBC financial group. *J. Interactive Marketing* **16**(3) 76–98.

Narayanan, V. G., L. Brem. 2002b. Owens & Minor, Inc. (A). HBS Case 9-100-055, February 14. Harvard Business School, Boston.

Niraj, R., M. Gupta, C. Narasimhan. 2001. Customer profitability in a supply chain. *J. Marketing* **65**(3) 1–16.

Nunes, J. C. 2000. A cognitive model of people's usage estimations. *J. Marketing Res.* **37**(4) 397–409.

Padilla, A. J. M. Pagano. 1997. Endogenous communication among lenders and entrepreneurial incentives. *Rev. Financial Stud.* **10**(1) 205–236.

Pagano, M., T. Jappelli. 1993. Information sharing in credit markets. *J. Finance* **48**(December) 1693–1718.

Pazgal, A., D. Soberman. 2008. Behavior-based discrimination: Is it a winning play, and if so, when?" *Marketing Sci.* **27**(6) 977–994.

Rust, R. T., V. A. Zeithaml, K. N. Lemon. 2000. *Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy*. Free Press, New York.

Schmittlein, D. C., L. G. Cooper, D. G. Morrison. 1993. Truth in concentration in the land of (80/20) laws. *Marketing Sci.* **12**(2) 167–183.

Selden, L., G. Colvin. 2003. *Angel Customers & Demon Customers*. Penguin Books, London.

Sharpe, S. A. 1990. Asymmetric information, Bank lending and implicit contracts: A stylized model of customer relationships. *J. Finance* **45**(4) 1069–1087.

Shin, J. 2005. The role of selling costs in signaling price image. *J. Marketing Res.* **32**(3) 302–312.

Shin, J., K. Sudhir. 2010. A customer management dilemma: When is it profitable to reward your own customers? *Marketing Sci.* **29**(4) 671–689.

Srivastava, S. 2007. Sprint drops clients over excessive inquiries. *Wall Street Journal* (July 7) A-3.

Stokey, N. L. 1979. Intertemporal price discrimination. *Quart. J. Econom.* **93**(3) 355–371.

Train, K. E., D. L. McFadden, M. Ben-Akiva. 1987. The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *RAND J. Econom.* **18**(1) 109–123.

Venkatesan, R., V. Kumar. 2004. A customer lifetime value framework for customer selection and resource allocation strategy. *J. Marketing* **68**(4) 106–125.

Villas-Boas, J. M. 1999. Dynamic competition with customer recognition. *RAND J. Econom.* **30**(4) 604–631.

Villas-Boas, J. M. 2004. Price cycles in markets with customer recognition. *RAND J. Econom.* **35**(3) 486–501.

Villas-Boas, J. M., U. Schmidt-Mohr. 1999. Oligopoly with asymmetric information: Differentiation in credit markets. *RAND J. Econom.* **30**(2) 375–396.

Zhang, J. 2011. The perils of behavior-based personalization. *Marketing Sci.* **30**(1) 170–186.