**When To Fire Customers?**

**Customer Cost Based Pricing**

Jiwoong Shin, K. Sudhir, and Dae-Hee Yoon*

January 2011

*Associate Professor of Marketing, Yale School of Management, Yale University, 135 Prospect St. New Haven, CT 06520 (E-mail: jiwoong.shin@yale.edu, Tel: (203) 432-6665); James L. Frank Professor of Private Enterprise and Management and Professor of Marketing, Yale School of Management, Yale University, 135 Prospect St. New Haven, CT 06520 (E-mail: k.sudhir@yale.edu, Tel: (203) 432-3289); Assistant Professor of Accounting, Yonsei School of Business, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul, Korea 120-749 (E-mail: daehee.yoon@yonsei.ac.kr, Tel: (822) 2123-6579). We thank Raphael Thomadsen, Robert Zeithammer, and seminar participants in the 8th Choice Symposium for useful comments.
**When To Fire Customers?**

**Customer Cost Based Pricing**

**Abstract**

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 the use of such customer cost information affects a firm’s customer acquisition and retention strategies, and ultimately its profit using a two period monopoly model with high and low cost customer segments. While past purchase and cost information help 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 CCP is profitable. Interestingly, 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. Our results also shed insight on retention-acquisition dynamics, on when firms can improve their profitability by selectively firing known old “bad” customers, and replacing them 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
1 Introduction

Customers differ in their costs to serve. Some customers tend to be considerably more costly to serve than others: a bank customer who insists on using a teller, as opposed to using the ATM machine for balance verification, cash withdrawal and deposits; a retail customer who returns products excessively often taking advantage of a liberal return policy; a Yahoo e-mail customer who uses much larger disk space than average taking advantage of “unlimited” hard drive space; a DVD rental customer for Netflix, who rents an abnormally large number of DVDs, taking advantage of the “unlimited rentals”. As 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. Recently, the excessive cost of customer support imposed by some customers was highlighted when Sprint "fired" these customers. Sprint wrote,

"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..." (Wall Street Journal 2007).

In general, marketing scholars have not been very sensitive to the issue of differential customer costs, because of the view that profit margin from products far exceed any differences in cost to serve. But as augmented services get increasingly bundled with the price of the product, the impact of cost to serve customers on customer profits has gained in relative importance. Further, until recently, even if a marketer was sensitive to differences in cost to serve, accounting systems were not capable of tracking the cost to serve individual customers. Accountants simply allocated the cost of augmented services evenly across to all customers. In recent years, the widespread acceptance of activity based customer costing has allowed firms to precisely allocate overhead costs to specific customers (Kaplan and Anderson 2004).

The precise measurement of customer cost has served as an eye-opener to many firms about the importance of cost-to-serve in guiding customer management strategies. For many firms,

---

2Activity based costing is difficult to implement, but firms have made considerable effort to implement it because of its importance for customer management. As a Charles Schwab executive states "One can never underestimate the difficulty in allocating the back-office and shared services costs, when they represent about 40% of our expenses and we want to link them back to product, channel or customer strategies," (“Understanding Customer Profitability at Charles Schwab,” HBS Case No 9-106-002.)
a significant number of customers cost more to serve than the revenues they bring in; these customers, thus destroy rather than augment firm profits. 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. Accounting scholars refer to this inverted-U curve as the “whale” curve in recognition of the “hump” in the curve. The figure shows that 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 value. The phenomenon of profit-sapping customers is not unique to this firm; according to Kaplan and Narayanan (2001), in many B2B firms, generally the top 20% of customers generate 150-300% of total profits, while the middle 70% of customers break even, and the bottom 10% of customers reduce 50-200% of profits. 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).

Figure 1: Cumulative Profit Distribution at an Industrial Firm

Despite the advances in the activity costing system which provides us with insight into customer profitability, there is little research on how the customer cost information affects firm strategies. Our goal in this paper is to address this gap in the literature by seeking theoretical

---

\(^3\text{Kanthal (A), HBS Case No 9-106-002. 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), but marketers have used it to show relative concentration of sales and revenues from customers (Schmittlein et al. 1993). Unlike the inverse Lorenz curves for cumulative income and sales or revenues, which are monotonically increasing, the inverse Lorenz curve for cumulative \textit{profitability} is non-monotonic and often shows a hump beyond which additional customers reduce cumulative profit (Kaplan 2001).}
insight into two sets of related questions. The first set of questions focuses on what actions a firm should take given the availability of customer information. For example, how should a firm balance retention of its existing customers against acquisition of new customers when setting prices, given the availability of customer cost information? 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? Given the additional information obtained from first period purchases on customer revenues and cost, how should a firm set prices across time in order to dynamically balance the mix of high and low cost customers? Firms often use a static picture of the cumulative profit (whale) curve as in Figure 1, but how does that curve evolve dynamically over time as firms selectively retain older customers and acquire new ones?

The second set of questions focuses more broadly on 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 customer cost-based pricing (CCP) and therefore investments in activity based customer costing be profitable?

Our modeling framework focuses on whether and how a firm can profitably discriminate among customers by offering targeted prices that is 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 customer cost-based pricing (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.

On the surface, this suggests that 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 if customers

\[\text{We focus on past cost-based pricing in this paper. There are other possible options to improve customer profitability. In the extreme, one can refuse service to unprofitable customers like Sprint, but this has the potential for bad public relations. Alternatively, one could seek to convert unprofitable customers to profitable ones by either raising revenues through cross-selling or lowering costs through less costly interactions (Mittal, Sarkees and Murshed 2008). In B2B markets, some firms tailor prices to services used and costs imposed on firms (Narayanan and Brem 2002). However, there is a negative perception of "nickel and diming" customers especially when charging for such traditional services as customer support and order processing. Also, researchers have shown that customers have a "flat-fee" bias (Train 1991); that is even in markets where metering is routine (e.g., telephony), customers tend to disproportionately prefer flat rate plans over metered plans (Lambrecht and Skiera 2006, Nunes 2000, and Train et al. 1987).} \]
are strategic and forward-looking. If firms increase prices for those who purchased in the previous period using the fact that these customers tend to have higher valuation for the good, some of the high valuation customers anticipate the price increase (the ratchet effect) for “old” customers and may defer their purchase. To prevent 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 in Hart and Tirole 1988). Therefore, it remains an open question as to whether profits will increase if firms use customer cost information for price discrimination, which is the focus of this paper.

To capture the tradeoff 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 valuations, but also customer cost.

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. Moreover, we find that it is optimal to fire even some profitable customers through higher targeted prices, which suggests that firms should not retain all of its profitable customers. (2) At low levels of heterogeneity, a firm should not fire any of its customers but CCP will lead to lower profits, which is consistent with Villas-Boas (2004). (3) As firms incorporate customer cost information into their pricing, the customer mix evolves to being more profitable, with few unprofitable customers.

Interestingly, unlike pure demand behavior based discrimination, the use of customer cost information can lead to higher profits when customer cost heterogeneity is sufficiently large. At high levels of cost heterogeneity, the benefit from being able to discriminate customers overwhelms the negative ratcheting effects due to consumer’s strategic actions that prevent the firm from learning the information. Importantly, CCP remains profitable, even if we allow customers to 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 in order to avoid facing higher prices in future.
2 Literature Review

This paper intersects with three key research streams: behavior based pricing, adverse selection in marketing and economics and activity based costing in accounting. Behavior based pricing is the practice of offering different prices based on a customer’s past purchasing behavior (e.g., Acquisti and Varian 2005, Fudenberg and Tirole 1998, 2000, Shaffer and Zhang 2000, Villas-Boas 1999, 2004). In contrast to our emphasis on cost information from purchase behavior, behavior based price discrimination has typically focused only on the demand side dimension of information that firms learn from the customers’ past purchase behavior – willingness to pay (or relative preference in a competitive market) of the product. Much of this focus translates to understanding the effects of price discrimination between a firm’s own past customers and new customers. As discussed in the introduction, if consumers are forward-looking, they recognize the possibility that the firm may take advantage of this preference information to discriminate against them (ratchet effect) and may therefore modify their purchase behavior to prevent firms from inferring their true preferences. In one set of models (Villas-Boas 2004, Acquisti and Varian 2005), when a monopolist faces strategic, 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 future. It turns out that such strategic behavior by customers reduces firm profits in equilibrium, relative to when firms commit not to use purchase information in pricing (Hart and Tirole 1988). These issues become magnified in the presence of competition. Competing firms may suffer compared with a scenario without past purchase information and, therefore, in a competitive market, firms confront a prisoner’s dilemma regarding the use of information about customer purchase history (Fudenberg and Tirole 2000; Villas-Boas 1999). According to Fudenberg and Villas-Boas’ (2006, p. 378) succinct summary of existing literature on behavior-based pricing 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 research has focused on adverse selection issues when setting prices. In financial models of credit markets with adverse selection (Pagano and Jappelli 1993, Villas-Boas and Schmidt-Mohr 1999), the firms learn about their own customers’ type (i.e., their ability to repay loans during the lending process) from their relationships with customers, but not about

\[ ^{5}\text{An exception is Shin and Sudhir (2009) who incorporate a quantity (rather than costs) dimension to the behavior-based pricing literature.}\]
customers who do businesses with competitors. The firm then uses this information asymmetry to their advantage when determining interest rates for future loans to their own customers. Given the focus on credit markets, there is no difference in customer’s willingness to pay for different firms, hence the discrimination is on types alone (as in adverse selection models). Our model combines both research streams in that we consider both past purchase (as in behavior based pricing) and cost type information (as in adverse selection models). Further, unlike the credit markets literature, where types (ability to repay) are reasonably treated as exogenous, we also consider the case of endogenous cost types, by allowing customers to choose the level of service (and thus cost-to-serve).

The third stream of literature related to this research is customer activity based costing in accounting. Not surprisingly, accounting researchers have paid greater attention to this issue than marketing scholars. Their focus has primarily been on customer activity-based pricing models (for example, Banker and Hughes 1994, Narayanan 2003). Narayanan (2003) investigates the benefits of activity based pricing compared to traditional pricing models using a static model, when the monopolist is able to price based on the metered use of services in a B2B environment. He concludes that activity based pricing is beneficial when there is more variability in the cost-to-serve among customers in a monopoly setting. Niraj, Gupta and Narasimhan (2001) empirically study the profitability of a distributor supplying to several grocery and retail businesses using activity based costing methods. Related research on the effect of customer costs on firm strategy is Shin (2005) who investigates the strategic implications of selling costs on a firm’s advertising strategy. Haenlein and Kaplan (2008) address the consequence of cost based pricing strategy on a firm’s long-term profitability, when firms are vulnerable to the negative word of mouth which cost based pricing may generate. However, none of these papers addresses the issues of behavior (specifically cost) based pricing in a dynamic setting.

Finally, this research is related to customer relationship management research which emphasizes the importance of identifying the right customers for a successful CRM program (Boulding et al. 2005, Winer 2001). Researchers have, therefore, focused on the identification of good customers (Rigby et al. 2002) by estimating customer lifetime value (Gupta et al. 2004, Jain and Singh 2002) and providing them with differentiated value propositions through different price levels (Shin and Sudhir 2009). These researchers show 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 2003). Our analysis suggests that this may indeed be a
dynamically optimal strategy. The paper formalizes this adverse selection idea, and analyzes the
dynamics of the mix of good and bad customers when 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
that costs $s^L$ to serve ($s^L < s^H$). This exogenous fixed cost type assumption is relaxed in a later
section.

We allow for heterogeneity in consumers willingness to pay $w$ for the product. Consumer
willingness to pay $w$ follows a uniform distribution, $w \sim U[0, v]$, where $v > s^H$ 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 of $u(p|w) = w - p$ from purchasing the
product. The market demand at price $p$ is $D(p) = v - p$.

In the first period, the firm has no specific information about individual consumers. Therefore,
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 identify three groups of customers
and can offer three different prices: (1) a price for low cost type customers, who purchased in
the first period ($p^L_2$), (2) a price for high cost type customers, who purchased in the first period
($p^H_2$), and (3) a price for “others,” who did not purchase in the first period ($p^O_2$).

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
their purchase. Since high and low cost customers receive different prices in the second period,

\begin{itemize}
\item[6] With non-zero marginal cost $c$, we can interpret $w = w^* - c$, as the willingness to pay net the cost of product
and $w \sim U[c, v + c]$.
\end{itemize}
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 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.

Second Period

Let \( \hat{w}_j^H \) be the first period marginal consumer of type \( j \in \{L, H\} \) who is indifferent between purchasing the product and not purchasing. Let \( D_j^2 \) 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:

\[
\Pi_2 = (p^H_2 - s^H)D^H_2(p^H_2) + (p^L_2 - s^L)D^L_2(p^L_2) + (p^O_2 - s^H)(\hat{w}_1^H - p^O_2) + (p^O_2 - s^L)(\hat{w}_1^L - p^O_2).
\]

The first and second term captures profits from high and low cost consumers who purchased in the first period, while 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. Since each segment’s purchase behavior is independent of the other, the firm can set prices \( p^H_2, p^L_2, p^O_2 \) independently by maximizing profits from each segment.

Let us first consider the profits from the high cost type segment who purchased in the first period. The firm’s price to the high cost type \( p^H_2 \) is determined by \( \arg\max_p (p - s^H)D^H_2(p) \), where \( D^H_2(p) = \min(v - \hat{w}_1^H, v - p) \) since 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 \( p^H_2 > \hat{w}_1^H \). The first period marginal consumer \( \hat{w}_1^H \) decides not to purchase in the second period and the demand is given by \( D^H_2(p^H_2) = \ldots \)
This is the “partial coverage” case since only a fraction of customers who bought in the first period purchase in the second period. In this case, the optimal price would be
\[ p_2 = \arg \max p (p - s^H) (v - p) = \frac{v + s^H}{2} \]. Hence, when \( \tilde{w}_1^H < \frac{v + s^H}{2} \), the firm will charge \( p_2^H = \frac{v + s^H}{2} \) in the second period.

On the other hand, if \( \tilde{w}_1^H \geq \frac{v + s^H}{2} \), all first period consumers decide to purchase in the second period when the firm charges \( p = \frac{v + s^H}{2} \); the first period customers will be fully covered. Then, the firm will increase its price and charges \( p_2^H = \tilde{w}_1^H \) and the demand is given by \( D^H_2(p) = v - \tilde{w}_1^H \).

Similarly, the demand function for the low cost type is \( D^L_2(p) = \min(v - \tilde{w}_1^L, v - p) \). The corresponding prices for the full and partial coverage cases are \( p_2^L = \tilde{w}_1^L \) and \( p_2^L = \frac{v + s^L}{2} \), 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 (Frexias et al. 1985, Fudenberg and Villas-Boas 2007).

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

\[ p_2^O = \arg \max p (p - s^H) (\tilde{w}_1^H - p) + (p - s^L) (\tilde{w}_1^L - p) = \frac{\tilde{w}_1^H + \tilde{w}_1^L + s^H + s^L}{4}. \]

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

To summarize, the optimal prices in the second period are,

\[ p_2^j = \max \left\{ \frac{v + s^j}{2}, \tilde{w}_1^j \right\}, \quad p_2^O = \frac{\tilde{w}_1^H + \tilde{w}_1^L + s^H + s^L}{4}. \]  

**First Period**

In the first period, a consumer with willingness to pay \( w \) decides to purchase a product if

\[ w - p_1 + \delta \cdot \max \left\{ w - p_2^j, 0 \right\} \geq \delta \cdot \max \left\{ w - p_2^O, 0 \right\}. \]

\(^7\)In equilibrium, we can easily see that \( p_2^O \leq \tilde{w}_1^j \). This is so since \( \tilde{w}_1 = \tilde{w}_1^L = \tilde{w}_1^H \) in equilibrium (we will show this subsequently) and thus, \( p_2^O \leq \tilde{w}_1^j \iff s^H + s^L \leq 2\tilde{w}_1 \). The last inequality always satisfies in equilibrium.
From the 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 $w \geq \hat{w}_1$ 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, $\hat{w}_j^1 = \hat{w}_j^1(p_1)$, 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:

$$\hat{w}_j^1 - p_1 = \delta \cdot \max \left\{ \hat{w}_j^1 - p_2^O, 0 \right\}$$

$$\Leftrightarrow \hat{w}_j^1 = \begin{cases} p_1 & \text{if } p_1 < p_2^O \\ \frac{p_1 - \delta p_2^O}{1-\delta} & \text{if } p_2^O \leq p_1 \end{cases} \quad (5)$$

It can be easily shown that the monopolist always lowers its price to non-buyers ($p_2^O$) in the second period relative to the first period price ($p_1$) as in Stokcey (1979) and Hart and Tirole (1988). Suppose $p_1 < p_2^O$. From Equation (5), $\hat{w}_j^1 = p_1$ and thus $\hat{w}_1^1 - p_2^O < 0$ since $p_1 < p_2^O$..
Hence, $u(p_2^O | w) = w - p_2^O < 0$ for all $w \leq \hat{w}_1^1$, 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 $p_2^O$ to be below $p_1$. This clearly increases the demand and profit. Hence, $p_2^O \leq p_1$ in equilibrium.

Let us define $\bar{s} = \frac{s^H + s^L}{2}$ 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 $p_2^O = \hat{w}_1^H + \hat{w}_1^L + s^H + s^L$ from Equation (3).

$$\hat{w}_1 = \hat{w}_1^H = \hat{w}_1^L = \frac{p_1 - \delta p_2^O}{1-\delta}$$

$$\Leftrightarrow \hat{w}_1 = \frac{2p_1 - \delta \bar{s}}{2 - \delta}. \quad (6)$$

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

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

$$\Pi(p_1) = (p_1 - s^H)(v - \hat{w}_1) + (p_1 - s^L)(v - \hat{w}_1)$$

$$+ \delta \left\{ (p_2^H - s^H)D_2^H(p_2^H) + (p_2^L - s^L)D_2^L(p_2^L) + (2p_2^O - s^H - s^L)(\hat{w}_1 - p_2^O) \right\},$$

where $\hat{w}_1 = \frac{2p_1 - \delta s}{2 - \delta}$, $D_2^j(p_2^j) = v - p_2^j$, $p_2^j = \max \left\{ \frac{v + s^j}{2}, \hat{w}_1 \right\}$, and $p_2^O = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L}{4}$.

The first term in Equation (7) represents the first period profit and the second term within braces represents the second period profit from its previous high cost type customers, from its previous low cost 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 $v_H^{\text{max}} = v - s^H > 0$ as the maximum extractable value from the high cost type customer given the cost to serve $s^H$, and $\Delta s = s^H - s^L$ as the difference in the cost to serve between the high and low cost types. Thus, $\Delta s$ captures the extent of heterogeneity in cost to serve across the high and low cost type customers. The analysis consists of two parts based on the extent of service cost heterogeneity: $\Delta s \leq \frac{\delta v_H^{\text{max}}}{2}$ and $\Delta s > \frac{\delta v_H^{\text{max}}}{2}$.

**When service cost heterogeneity is sufficiently small: $\Delta s \leq \frac{\delta v_H^{\text{max}}}{2}$.**

In this condition, $\frac{v + s^j}{2} \leq \hat{w}_1^j$ 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,

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

(8)

The first period marginal consumer’s valuations are:

$$\hat{w}_1^1 = \hat{w}_1^H = \hat{w}_1^L = \frac{(2 + \delta)v + 2s}{4 + \delta}.$$  

(9)
It can be easily seen that \( \frac{v + s_j}{2} \leq \hat{w}_1 \) in equilibrium if \( \Delta s \leq \frac{\delta (v - s^H)}{2} = \frac{\delta v^\text{max}}{2} \).

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

**Lemma 1.** When the service cost heterogeneity is sufficiently small such that \( \Delta s \leq \frac{\delta v^\text{max}}{2} \), the equilibrium outcomes are as follows:

\[
\begin{align*}
p_1 &= \frac{v (4 - \delta^2) + (4 + 2\delta + \delta^2) \bar{\pi}}{2(4 + \delta)}; \\
p_2^j &= \frac{(2 + \delta) v + 2 \bar{\pi}}{4 + \delta}; \\
p_2^O &= \frac{v(2 + \delta) + (6 + \delta) \bar{\pi}}{2(4 + \delta)}; \\
\Pi^{\text{CCP}} &= \frac{(v - \bar{\pi})^2(2 + \delta)^2}{2(4 + \delta)}.
\end{align*}
\]

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

**Proposition 1.** When the service cost heterogeneity is sufficiently small such that \( \Delta s \leq \frac{\delta v^\text{max}}{2} \),

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 \( w \geq \hat{w}_1 = \frac{(2 + \delta) v + 2 \pi}{4 + \delta} \) 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 (\( \Pi^{\text{CCP}} \)) is lower than the profit without price discrimination (\( \Pi^{\text{NoPD}} \)).

**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, firms find it not optimal to price discriminate among the high and low cost types. This is particularly surprising given the fact that the firm would have charged different prices by cost types in the absence of purchase history information (in that case, it is easy to see that the firm would have charged \( p_2^L = \frac{v + s^L}{2} \) and \( p_2^H = \frac{v + s^H}{2} \)).

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 the customer’s service cost heterogeneity is low, there is little gain in discriminating between the two types of customers relative to the gain from using the customer purchase history information. The entire surplus of both the high and low cost marginal "old" customer (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. If a firm sought to discriminate the high cost-to-serve customer by charging a higher price, that price would be 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 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. The Figure 2 below illustrates this graphically.

Figure 2: Customer’s Willingness to Pay and Prices when $\Delta s \leq \frac{\delta v^\max_H}{2}$

Second, the firm is worse off using customer’s past purchase and cost information. In this case, since only purchase information is used in setting its prices, the result is consistent with the previous literature using only purchase information (Acquisti and Varian 2005, Hart and Tirole 1985, Villas-Boas 2004). The main intuition in these papers is as follows: Consumers are forward-looking and marginal consumers defer purchasing in the first period because (1) they know that they will face lower price in the future ($p_2^O \leq p_1$) if they opt to not purchase in the first period, and (2) they will be ripped off with a high price in the future ($p_2^i > p_1$) by the firm if they purchase a product (ratchet effect). Now, when customer heterogeneity in cost is small, the effect of valuation information revealed through a customer’s first period purchase dominates any effect of the differential cost of serving them and so firms do not discriminate on the basis of cost to serve. In essence, when service cost heterogeneity 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.

Furthermore, a comparison of the profits between with and without price discrimination gives us the following corollary.

**Corollary 1.** The difference in profits \( \Delta \Pi = \Pi^{CCP} - \Pi^{NoPD} \) increases in \( \delta \), i.e., \( \frac{\partial (\Delta \Pi)}{\partial \delta} > 0 \) for \( \delta \in [0,1] \).

As consumers become more forward looking (greater \( \delta \)), more consumers tend to wait for the lower future price. The ratchet effect becomes more pronounced and the profit decreases even further because the number of first period customers who are willing to buy at a higher first period price falls. Overall, the number of customers who purchase only in the second period increases.

**When service cost heterogeneity is sufficiently large:** \( \Delta s > \frac{\delta v_H^{max}}{2} \).

We next consider the case when the service cost heterogeneity is large such that \( \Delta s > \frac{\delta v_H^{max}}{2} \), which ensures that \( \tilde{w}_1^L > \frac{v + s^L}{2} \) and \( \tilde{w}_1^H < \frac{v + s^H}{2} \) are satisfied in equilibrium (we will confirm this subsequently). Therefore, the firm charges \( p_2^L = \tilde{w}_1^L \), \( p_2^H = \frac{v + s^H}{2} \) and \( D_2^L(p) = v - \tilde{w}_1^L \), \( D_2^H(p) = v - \frac{v + s^H}{2} = \frac{v - s^H}{2} \). Taking the first order condition from Equation (7) gives us the first period optimal price \( p_1 \):

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

And the first period cutoff line is now obtained as

\[
\tilde{w}_1 = \tilde{w}_1^H = \tilde{w}_1^L = \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta}.
\]

By plugging \( \tilde{w}_1 = \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta} \) in equations, we can check that \( \tilde{w}_1^L > \frac{v + s^L}{2} \) and \( \tilde{w}_1^H < \frac{v + s^H}{2} \).

---

*The term, *customer cost based pricing*, can be misleading since the firm *chooses* not to use customer cost information for setting the prices. More precisely, when the customer’s cost heterogeneity is low, the firm uses only the past purchase information and this is equivalent to the case of behavior-based price discrimination (Fudenberg and Tirole 2000) or pricing with customer recognition (Villas-Boas 1999).*
are satisfied in equilibrium when \( \Delta s > \frac{\delta v_H^{\max}}{2} \).

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

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

We summarize the equilibrium outcomes in the following lemma.

**Lemma 2.** When the service cost heterogeneity is sufficiently large such that \( \Delta s > \frac{\delta v_H^{\max}}{2} \), the equilibrium outcomes are as follows:

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

\[
p_2^H = \frac{v + s^H}{2}; \quad p_2^L = \hat{w}_1 = \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta}; \quad p_2^O = \frac{4v + 3(2 - \delta)s^H + (6 - \delta)s^L}{4(4 - \delta)};
\]

\[
\Pi^{CCP} = \frac{16(v - \bar{s})^2 + 4\delta((v - 2\bar{s})^2 + v^2 - 2s^H s^L) - (2v(v - 2s^L) + 2(s^L)^2 - (\Delta s)^2)\delta^2}{8(4 - \delta)}.
\]

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

**Proposition 2.** When the service cost heterogeneity is sufficiently large such that \( \Delta s > \frac{\delta v_H^{\max}}{2} \),

1. The second period price to high cost customers are higher than the price to low cost customers; i.e., 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 \( w \geq p_2^H = \frac{v + s^H}{2} \) are retained in the second period while high type consumers with \( w \in \left[ \hat{w}_1, p_2^H \right] \) will be fired in the second period.

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

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

**Proof.** See the Appendix.  \( \square \)
In contrast to the case when service cost heterogeneity is small, the firm uses the customer cost information. In this case, the firm discriminates between its first period high and low cost customers and offers each group different equilibrium prices. In particular, the firm charges higher price to its high cost customers ($p^L_2 < p^H_2$). At this higher price for the high cost customers, some of the first period high type customers with relatively low willingness to pay ($w \in \left[ \hat{w}_1, p^H_2 \right]$) do not have a high enough willingness to pay to purchase in the second period. The firm, therefore, “fires” some of high cost customers. In contrast, all of the low cost customers are retained in the second period. Overall, 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.

Interestingly, even though the firm fires customers with moderate valuations in the range $w \in \left[ \hat{w}_1, p^H_2 \right]$, they acquire new customers with even lower valuations by offering a lower price to new customers. New customers with lower willingness to pay $w \in \left[ p^O_2, \hat{w}_1 \right]$, in both segments purchase in the second period; i.e., the firm fires moderate valuation customer with known high costs, but acquires a mix of high and low cost customers at a lower price, for whom the cost is unknown.

The impact of CCP on profits is more subtle. For moderate levels of service cost heterogeneity, i.e., $\frac{\delta v^\text{max}_H}{2} < \Delta s < v^\text{max}_H \left(2 + \sqrt{2(4 - \delta)}\right)$, CCP reduces profits, even though the firm price discriminates 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., $\Delta s > v^\text{max}_H \left(2 + \sqrt{2(4 - \delta)}\right)$, 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 service cost heterogeneity, we plot $\Pi^{CCP}$ against $\Delta s$ for low and high levels of $\Delta s$ 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 ($\Pi^{NoPD}$). The plot is shown below in Figure 4.

![Figure 4: Profit Comparison between Customer Cost-based Price Discrimination ($\Pi^{CCP}$), Purchase Behavior-based Price Discrimination ($\Pi^{Purchase}$), and No Price Discrimination ($\Pi^{NoPD}$).](image)

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 (Proposition 1). Hence, $\Pi^{CCP}$ is identical to $\Pi^{Purchase}$. Moreover, this profit is lower than the case of profit without price discrimination ($\Pi^{NoPD}$).

In contrast, from Proposition 2 we know, that when the service cost heterogeneity 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, $\Pi^{CCP}$ is greater than $\Pi^{Purchase}$, the firm is better off using cost information. 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.

We summarize the qualitative conclusions from the analysis along three dimensions in Table 1. When the cost heterogeneity is low \((\Delta s \leq \frac{\delta v_{H}^{\text{max}}}{2})\), 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 cost heterogeneity is relatively large \((\Delta s > \frac{\delta v_{H}^{\text{max}}}{2})\), the effect of cost for serving customers 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 the 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, at moderate levels of \(\Delta s\), CCP negatively impacts profits. But beyond a critical value of \(\Delta s\), CCP becomes profitable.

<table>
<thead>
<tr>
<th>Service-Cost Heterogeneity ((\Delta s = s^H - s^L))</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrimination by Cost Type</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Customer Firing</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Effect of CCP on Profit</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
</tr>
</tbody>
</table>

3.3 Numerical Example

To delve deeper into the dynamics of the acquisition and retention strategies of firms, and their impact on profits, we consider a specific numerical example. For the example, we set \(\delta = 0.9\), \(v = 1.2\), \(s^L = 0\). For the low and high service cost heterogeneity cases, we set \(s^H = 0.3\) and \(s^H = 1\), respectively.

Diagnostic Metrics

Several diagnostic metrics for the first and second periods are provided in Table 2. Customers who buy in the first period are referred to as “Old” customers and those who buy only in the
second period are referred to as “New” customers. There are two key takeaways from Table 2. First, consistent with propositions 1 and 2 (summarized in Table 1), CCP leads to lower aggregate profits ($\Pi^{CCP} = 0.946$) than the benchmark profits without CCP ($\Pi^{NoCCP} = 1.047$) when the service cost heterogeneity is low ($\Delta s = 0.3$) the cost difference is large, the price discrimination effect overwhels; but CCP has higher profit ($\Pi^{CCP} = 0.470 > \Pi^{NoCCP} = 0.466$) when the service cost heterogeneity is high ($\Delta s = 1$). One can also see the intuition about the ratchet effect being stronger when service cost heterogeneity is low. When service cost heterogeneity is low ($\Delta s = 0.3$), second period prices rise by 57% for old customers. No distinction is made between the high and low cost customers, even though customer cost and purchase information are available in the first period; only the purchase information is used in setting prices (as in Villas-Boas 2004). In contrast, when service cost heterogeneity is high ($\Delta s = 1$), both cost and purchase information are used. Second period prices rise by 65% for the high cost old customers, but only by 21% for the low cost customers. The weaker ratcheting effect for the old low cost customers coupled with the price discrimination effect between high and low cost customers makes CCP more profitable.

Second, the optimal retention strategy depends on the service cost heterogeneity. When the service cost heterogeneity is low, the average retention rate is 100% because there is no reason to “fire” customers. The average retention rate declines to 62.5% when the service cost heterogeneity is high, but this lower retention is differentiated by high and low cost customers. 100% of the low cost customers are retained, but only 25% of the high cost customers are retained. In contrast to general exhortations to raise retention rates across customers, our results demonstrate that optimal retention rates should be managed based on customer characteristics. Firms should lower retention rates among its higher cost customers in order to obtain a more favorable mix of low to high cost customers.9

9Incidentally, the result here also sheds insight on a concern among accounting scholars and practitioners about how firing existing customers may impact the profitability of remaining customers, when the service costs (which may be fixed in the short run) needs to be re-allocated across fewer customers. Our result shows why this need not be a serious cause of concern. Even though the firm “fires” 75% of its sure “bad” customers from the first period, it replaces those bad unprofitable demand with new, more profitable customers (new customers are a mix of good and bad customers). Thus overall demand need not fall despite firing existing customers because firms will balance the firing of known unprofitable current customers with new more profitable acquisitions.

*** Table 2a: Low Cost Heterogeneity ***

*** Table 2b: High Cost Heterogeneity ***
Whale Curve Dynamics

To understand how customer profitability changes as firms use customer cost based pricing, we next consider how the inverse Lorenz (whale) curve (shown in Figure 5) evolves from the first period to the second period for the numerical example above.10

*** Figure 5. Inverse Lorenz Curves of Cumulative Profit ***

Consider the case when service cost heterogeneity is low ($\Delta s = 0.3$). In the first period, the curve is monotonically increasing, i.e., there is no “hump” in the “whale” curve. It follows easily 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 ($p_1 > s^H$). Thus we do not see a “hump” in the whale curve in Figure 5-(a). Even though the high and low cost customers have equal share in the customer mix, the low cost customers contribute 72% of profits, while 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 \textit{a priori} identify. The inverse Lorenz curve continues to be monotonically increasing in the second period, but overall the curve is flatter, indicating that the customer contribution to profits are more equitable. Consistent with the valuation information that can be extracted merely from past purchases, the most profitable customers are the “old” customers (low cost and high cost in that order). Together, they contribute 73% of profits, while 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, while the new high cost customers contribute only 6% of profits.

When the service cost heterogeneity 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

---

10The “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. To the best of our knowledge, there has been no characterization of the dynamics of the whale curve, in response to strategic marketing actions.
contribute 198% of the firm profits, while the remaining 50% high cost customers destroy 98% of the overall profits.

In the second period, as we know from Proposition 2, the firm will not retain all of the high cost customers. Therefore, even though there continues to remain a hump in the second period curve, it is much flatter. With selective retention and firing, as many as 81% of customers are profitable compared to 50% of customers in the previous period. Further, 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 case of low heterogeneity, 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. Such a customer mix is consistent with the relative importance of valuation/customer cost information revealed from past purchases.

Our analysis highlights why marketers often seek to raise average retention rates across all customers. In many 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 tolerate some attrition from the high cost customer base.

4 Extension: Endogenous service demand

Thus far, we assume that the cost to serve a given customer is exogenous (i.e., the customer type 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. How would a firm’s acquisition and retention strategies change if customer cost itself were endogenous? Would CCP still be profitable if customers can choose their service demand to prevent the firm from unambiguously learning their cost type?

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. Further, we focus on the case when the service cost heterogeneity is sufficiently large ($\Delta s > \frac{\delta v_{H}}{2}$) since this is the region in which we found CCP to be profitable when the types are exogenous.

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

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

where $\tau$ is the extra utility that a customer obtains from consuming 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 $H$-type customer to demand extra service. In this modified setting, the high cost type customers may strategically conceal their type and mimic the low cost type customers by demanding a low level of service in the first period in order to get a better price in the future.

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

$$\Pi_2 = (p^H_2 - s^H) \cdot \min\{v + \tau - p^H_1, v - \hat{w}^H_1\} + (p^L_2 - s^L) \cdot \min\{v - p^L_1, v - \hat{w}^L_1\} + (p^O_2 - s^H) (\hat{w}^H_1 + \tau - p^O_2) + (p^O_2 - s^L) (\hat{w}^L_1 - p^O_2),$$

Subject to

$$\begin{align*}
(I\text{C-H}) & \quad w + \tau - p_1 + \delta(w + \tau - p^H_2) \geq w - p_1 + \delta(w + \tau - p^L_2), \\
(I\text{C-L}) & \quad w - p_1 + \delta(w - p^L_2) \geq w + \tau - p_1 + \delta(v - p^H_2). 
\end{align*}$$

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. This potential strategic behavior imposes additional constraints on the monopolist’s objective function. The first constraint (IC-H) represents the self-selection constraint for $H$-type that the high cost type customers do not gain from pretending to be a low cost type customer in the first period while the second constraint (IC-L) is for $L$-type customers to not pretend to be the high cost type customer.

The low cost type customer has no incentive to mimic a high type because he does not
value service, i.e., \( \tau = 0 \) for \( L \)-type consumers. Hence, (IC-L) is trivially satisfied in equilibrium where \( p^L_2 < p^H_2 \). On the other hand, the \( H \)-type customers may mimic \( L \)-type by altering service demand, if the utility loss from foregoing service in the first period (\( \tau \)) is lower than the discounted gain from the price differential in the second period \( \delta(p^H_2 - p^L_2) \). The left hand side of (IC-H) represents a \( H \)-type customer’s total utility over two periods when he truthfully reveals his type, \( w + \tau - p_1 + \delta(w + \tau - p^H_2) \) and the right hand side is the total utility he can get when he mimics the \( L \)-type in the first period, \( w - p_1 + \delta(w + \tau - p^L_2) \). Then, (IC-H) can be rewritten as

\[
\begin{align*}
& w + \tau - p_1 + \delta(w + \tau - p^H_2) \geq w - p_1 + \delta(w + \tau - p^L_2) \\
\iff & p^L_2 \geq p^H_2 - \frac{\tau}{\delta}.
\end{align*}
\]

When \( \tau \) is large enough (i.e., the IC for \( H \)-type is not binding), the monopolist’s problem reverts to the basic model where the cost types are exogenously fixed. We first look at this case.

**When \( \tau \) is sufficiently large**

Recall we only consider the case of \( \Delta s > \frac{\delta v^{max}}{2} \), where cost information was used by the monopolist in the *exogenous case*. This condition ensures that \( \hat{w}_1^L > \frac{v + s^L}{2} \) and \( \hat{w}_1^H < \frac{v + \tau + s^H - \tau}{2} \) are satisfied in equilibrium.\(^ {12} \) Therefore, if (IC) constraint is not binding, the firm charges

\[
\begin{align*}
p^L_2 &= \hat{w}_1^L \quad \text{and} \quad p^H_2 = \arg\max_p (p - s^H)(v + \tau - p) = \frac{v + \tau + s^H}{2},
\end{align*}
\]

and

\[
\begin{align*}
D^L_2(p) &= \min\{v - p^L_2, v - \hat{w}_1^L\} = v - \hat{w}_1^L, \quad D^H_2(p) &= \min\{v + \tau - p^H_2, v - \hat{w}_1^H\} = v + \tau - \frac{v + \tau + s^H}{2} = \frac{v + \tau - s^H}{2}.
\end{align*}
\]

Hence, the monopolist does not serve all of the previous customers.

Like the exogenous model, the firm sets price for customers who have not purchased in the first period using the first-order condition from Equation (14) as follows:

\[
\begin{align*}
p^O_2 &= \arg\max_p (p - s^H)(\hat{w}_1^H - p + \tau) + (p - s^L)(\hat{w}_1^L - p) = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L + \tau}{4},
\end{align*}
\]

We summarize the equilibrium outcomes in the following lemma.

**Lemma 3.** When service demanded by consumers (and therefore cost-to-serve) is endogenous and the utility from service \( \tau \) is sufficiently large that (IC-H) constraint will be always satisfied,

\( ^{12} \)We confirm that \( \frac{v + s^L}{2} < \hat{w}_1^L \) and \( \frac{v + \tau + s^H - \tau}{2} > \hat{w}_1^H \) are indeed satisfied when \( \Delta s > \frac{\delta v^{max}}{2} \). Please see the proof of Lemma 3 in the Appendix.
The marginal customers in the first period differ by customer type:

\[ \hat{w}_1^L = \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta} + \frac{(2 - \delta)\tau}{2(4 - \delta)}, \quad \hat{w}_1^H = \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta} - \frac{(6 - \delta)\tau}{2(4 - \delta)}. \]

The equilibrium outcomes are as follows:

\[ p_1 = \frac{(4v - (2 - \delta)\tau)(2 - \delta) + 8\bar{s} - \Delta s(2\delta - \delta^2)}{4(4 - \delta)}, \]

\[ p_2^H = \frac{v + s^H + \tau}{2}, \quad p_2^L = \hat{w}_1^L = \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta} + \frac{(2 - \delta)\tau}{2(4 - \delta)}, \quad \text{and} \quad p_2^O = \frac{\hat{w}_1^H + \hat{w}_1^L + s^H + s^L + \tau}{4}. \]

Proof. See the Appendix. \(\square\)

Unlike the exogenous case, where customer type is fixed, now the marginal customer in the first period from the high and low type differ in their willingness to pay. This is because the high type customer gets an extra utility \(\tau\) from consuming the firm’s augmented services.

From Lemma 3, we identify the \(\tau\)-condition for (IC-H) constraint to be satisfied:

\[ p_2^L \geq p_2^H - \frac{\tau}{\delta} \iff \frac{2v + 2\bar{s} - \delta s^H}{4 - \delta} + \frac{(2 - \delta)\tau}{2(4 - \delta)} \geq \frac{v + s^H + \tau}{2} - \frac{\tau}{\delta} \]

\[ \iff \tau \geq \tau^{IC} = \frac{(2\Delta s - (v - s^H)\delta)\delta}{4(2 - \delta)}. \]

That is, when \(\tau\) is sufficiently large (\(\tau \geq \tau^{IC}\)), the high type consumer will reveal his type in the first period by choosing a high level of service.

**Proposition 3.** Suppose that the service cost heterogeneity is sufficiently large that \(\Delta s > \frac{\delta v_{H}^{\max}}{2}\), and the utility from service is large enough that \(\tau \geq \tau^{IC} = \frac{(2\Delta s - (v - s^H)\delta)\delta}{4(2 - \delta)}\).

1. The monopolist uses CCP and charges different prices to the \(H\) and \(L\)-type customers:

\[ p_2^H = \frac{v + s^H + \tau}{2} > p_2^L = \hat{w}_1^L. \]

2. Further, the total profit with CCP is greater than the total profit without price discrimination \((\Pi^{CCP} > \Pi^{NoPD})\) when \(\Delta s\) becomes large.

Proof. See the Appendix. \(\square\)

When \(\tau \geq \tau^{IC}\), the (IC) constraint for \(H\)-type is not binding. Therefore, the customers will reveal their types in the first period even under optimal prices that the monopolist would have
charged when the customer type is fixed. Hence, the equilibrium outcome is consistent with the
our main model where the customer types are exogenous and CCP can increase a firm’s profit.

When $\tau$ is small

The more challenging case occurs when the $H$-type customer may mimic the $L$-type customers.
When $\tau < \tau^{IC}$, the high cost customer is less sensitive to the utility change due to downgraded
service. In this case, (IC-H) is binding and $p^H_p = p^L_p + \frac{\tau}{\delta}$ from Equation (15).

Plugging this into the profit function in Equation (14), we can restate the second period
profit function as follows:

\[
\max_{p^L_p, p^O_p} \Pi_2 = (p^L_p + \frac{\tau}{\delta} - s^H) \cdot \min \{ v + \tau - p^L_p - \frac{\tau}{\delta}, v - \hat{\omega}^H \}
+ (p^L_p - s^L) \cdot \min \{ v - p^L_p, v - \hat{\omega}^L \} + (p^O_p - s^H)(\hat{\omega}^H + \tau - p^O_p) + (p^O_p - s^L)(\hat{\omega}^L - p^O_p).
\]

As we will confirm below, $\min \{ v - p^L_p, v - \hat{\omega}^L \} = v - \hat{\omega}^L$ in equilibrium and, therefore, $p^L_p = \hat{\omega}^L$
and $p^H_p = \hat{\omega}^L + \frac{\tau}{\delta}$ since $\hat{\omega}^L > p^L_p$, where $p^L_p$ is the optimal price the monopolist would charge if
$\min \{ v - p^L_p, v - \hat{\omega}^L \} = v - p^L_p$. In other words, $p^L_p = \arg\max_{p^L_p} (p^L_p + \frac{\tau}{\delta} - s^H)(v + \tau - p^L_p - \frac{\tau}{\delta}) + (p^L_p - s^L)(v - p^L_p) = \frac{(s^H + s^L + 2v)\delta - (2 + \delta)\tau}{4\delta}$. Further, we can see that $\hat{\omega}^L + \frac{\tau}{\delta} - \tau > \hat{\omega}^H$ is
satisfied in equilibrium and, therefore, $\min \{ v + \tau - \hat{\omega}^L - \frac{\tau}{\delta}, v - \hat{\omega}^H \} = v + \tau - \hat{\omega}^L - \frac{\tau}{\delta}$.
Also, $p^O_p = \arg\max_{p} (p - s^H)(\hat{\omega}^H - p + \tau) + (p - s^L)(\hat{\omega}^L - p) = \hat{\omega}^H + \hat{\omega}^L + s^H + s^L + \tau$.

Therefore, the monopolist’s optimal second period price is

\[
p^L_p = \hat{\omega}^L; \quad p^H_p = \hat{\omega}^L + \frac{\tau}{\delta}; \quad p^O_p = \frac{\hat{\omega}^H + \hat{\omega}^L + s^H + s^L + \tau}{4}.
\]

Similar to our main model, the marginal consumer in the first period, $\hat{\omega}^L = \hat{\omega}^L(p_1)$, can be
calculated by using the fact that $p^O_p \leq p_1$ and the marginal consumer does not get any surplus in
the second period if she already purchased it in the first period: $\hat{\omega}^H + \tau - p_1 = \delta(\hat{\omega}^H + \tau - p^O_p) \iff
\hat{\omega}^H = \frac{p_1 - \tau - \delta(p^O_p - \tau)}{1 - \delta}$, and $\hat{\omega}^L - p_1 = \delta(\hat{\omega}^L - p^O_p) \iff \hat{\omega}^L = \frac{p_1 - \delta p^O_p}{1 - \delta}$. Therefore, the willingness
to pay for the marginal customer in the first period can be obtained by using the fact that

\[\text{Note that the firm would charge } p^H_p = p^H_p = \arg\max_{p^H_p} (p^H_p - s^H)(v + \tau - p^H_p) = \frac{v + s^H + \tau}{2}, \text{ without}
\text{considering (IC-H) condition. However, when } \tau \geq \tau^{IC}, \text{ we can also easily check that in equilibrium } p^H_p = \hat{\omega}^L + \frac{\tau}{\delta} < p^H_p, \text{ and, therefore, (IC-H) condition in Equation (15) does not hold.}\]
\[ p_2^O = \frac{\tilde{w}_1^H + \tilde{w}_1^L + s^H + s^L + \tau}{4}, \]  
where \( \tilde{w}_1^H = \frac{2p_1 - \delta \tau}{2 - \delta} - \tau, \tilde{w}_1^L = \frac{2p_1 - \delta \tau}{2 - \delta}. \) Again, unlike the case where the customer type is exogenously fixed, the first period cutoff line for the high cost type and low cost type customers are different.

In the first period, the monopolist maximizes the following total expected profit:

\[ \Pi(p_1) = (p_1 - s^H)(v - \tilde{w}_1^H) + (p_1 - s^L)(v - \tilde{w}_1^L) + \delta \Pi_2, \]

(20)

where \( \Pi_2 = (\tilde{w}_1^L + \frac{\tau}{\delta} - s^H)(v + \tau - \tilde{w}_1^L - \frac{\tau}{\delta}) + (\tilde{w}_1^L - s^L)(v - \tilde{w}_1^L) + (p_2^O - s^H)(\tilde{w}_1^H + \tau - p_2^O) + (p_2^O - s^L)(\tilde{w}_1^L - p_2^O), \) \( \tilde{w}_1^H = \frac{2p_1 - \delta \tau}{2 - \delta} - \tau, \) and \( \tilde{w}_1^L = \frac{2p_1 - \delta \tau}{2 - \delta}. \)

Taking the first order condition, we obtain the first period price \( p_1: \)

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

(21)

Using this first period price, the first period marginal consumer has willingness to pay of,

\[ \tilde{w}_1^L = \frac{2 ((2 + \delta)v + s^H + s^L) - (2 - \delta)\tau}{2(4 + \delta)}, \quad \tilde{w}_1^H = \frac{2 ((2 + \delta)v + s^H + s^L) - (10 + \delta)\tau}{2(4 + \delta)}. \]

(22)

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

\[ \tilde{w}_1^L - p_2^{L_s} = \frac{\delta^2 \{2v - (s^H + s^L)\} + (8 - 6\delta + \delta^2)\tau}{4\delta(4 + \delta)} > 0, \]

\[ \tilde{w}_1^H - \left( \tilde{w}_1^L + \frac{\tau}{\delta} - \tau \right) = -\frac{\tau}{\delta} < 0. \]

With the following Lemma 4, we summarize the equilibrium results under endogenous service demand.

**Lemma 4.** When service demanded is endogenous and the utility from service \( \tau \) is sufficiently small that \( (IC-H) \) constraint will be binding, the equilibrium outcomes are as follows:

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

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

and \( p_2^O = \frac{(2(2 + \delta)v + (6 + \delta)(s^H + s^L)) - (2 - \delta)\tau}{4(4 + \delta)}. \)
Proposition 4. Suppose that the service cost heterogeneity is sufficiently large that \( \Delta s > \frac{\delta v_{H}^{\text{max}}}{2} \), and the utility from service is small such that \( \tau < \tau^{IC} = \frac{(2\Delta s - (\nu_s - \delta))\delta}{4(2-\delta)} \).

1. The monopolist uses CCP and charges different prices to the H and L-type customers: \( p_{H2} > p_{L2} \).

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

Proof. See the Appendix.

The proposition states that as long as the incremental value of the service, \( \tau \), is not too small, CCP remains profitable as in the exogenous case. Not surprisingly, when \( \tau \) is sufficiently 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. In order to prevent such strategic behavior of the customers, the firm needs to distort price to induce the high type customer to reveal his type. However, when \( \tau \) is very low (\( \tau < \tau^* \)), it is too costly for firm to make customers to reveal their type truthfully; in other words, the efficiency loss from (IC-H) is too high to be profitable. As the firm cannot charge a sufficiently high price to its own high cost type customers (for fear of their mimicking the low cost type), the benefit from the CCP is limited and cannot compensate for the negative indirect ratchet effect on profit.

Overall, we conclude from Proposition 3 and 4 that our main result of increased profit under CCP is robust even if customers can endogenously choose their level of service.

Finally, we compare the prices and profits under the exogenous and endogenous cases to gain insight into how the incentive compatibility constraint affects a firms pricing, retention and acquisition strategies. Under the exogenous case, the customer types is exogenously fixed, so that the firm does not need to consider the incentive constraints (IC-H). Hence, the equilibrium results under exogenous case are the same as those identified in Lemma 3 in which the utility from service \( \tau \) is sufficiently large that (IC-H) constraint will always be satisfied.

Proposition 5. Let \( p_{1,ex}, p_{L2,ex}, p_{H2,ex} \) be the equilibrium price charged under exogenous case, and \( p_{1,en}, p_{L2,en}, p_{H2,en} \) be the equilibrium price charged under endogenous case.

1. The price gap (i.e., the difference between the high and low type price) is higher under the exogenous case than in the endogenous case: \( \Delta_{ex} = p_{H2,ex} - p_{L2,ex} > \Delta_{en} = p_{H2,en} - p_{L2,en} \).

Moreover, \( p_{L2,ex} < p_{L2,en} \) and \( p_{H2,ex} > p_{H2,en} \).
2. The marginal customer in the first period for both high and low type has a lower willingness to pay under the exogenous case: $\hat{w}_{1,ex}^L < \hat{w}_{1,en}^L$ and $\hat{w}_{1,ex}^H < \hat{w}_{1,en}^H$. This implies that the first period demand is higher under the exogenous case than in the endogenous case. Moreover, prices are lower: $p_{1,ex} < p_{1,en}$.

3. The profit is higher under exogenous case: $\Pi_{en}^{CCP} < \Pi_{ex}^{CCP}$.

Proof: See the Appendix.

The proposition provides several key insights. To make the exposition clearer, Figure 5 below shows the second period prices for the old 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 of the price gap is intuitive. The price gap in the second period between the high and low type customers is larger in the exogenous case ($\Delta_{ex} = p_{2,ex}^H - p_{2,ex}^L > \Delta_{en} = p_{2,en}^H - p_{2,en}^L$), reflecting that we can price discriminate more effectively under the exogenous case where the firm is not constrained by incentive compatibility constraint in setting the prices. However, unlike in a standard static model of price discrimination where prices are distorted only for the “low” types, the price for both the high and the low cost type changes and moves closer to each other for the endogenous case ($p_{2,ex}^L < p_{2,en}^L$ and $p_{2,ex}^H > p_{2,en}^H$). This is due to the interaction of incentives of purchase history based with cost based discrimination. In contrast to the static second degree price discrimination, in which the focus of the (IC) constraint is to induce truth-telling by customers in the same period (i.e., second period), here the incentive compatibility constraint is across time. The goal of the firm in the our intertemporal CCP model (which is based on both purchase history and customer cost type) is to induce “truth-telling” of customers in the previous period (i.e., the first period).

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

In this paper, we study the impact of customer cost based pricing (CCP) on a firm’s dynamic customer acquisition and retention strategies and overall firm profits, when both consumers and firms are strategic and forward looking. We develop a two period monopoly model where firms 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 customer cost-to-serve. When customer service cost heterogeneity 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, B2B markets etc., 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 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, when the heterogeneity in cost-to-serve across customers is high, CCP can be profitable, but when heterogeneity in cost-to-serve is not large enough, CCP is unprofitable. Our results are consistent with the behavior-based pricing literature that behavior based pricing leads to lower profits for the monopolist when facing strategic consumers, and when the consumers are relatively homogeneous in their service costs. However, when the heterogeneity in cost to serve reaches a certain threshold, CCP can be profitable. We demonstrate these results initially in a setting where customer cost-to-serve is exogenous. We then show that the results remain robust even if the level of service demanded (and therefore cost-to-serve) is endogenous.

Finally, the 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 contribution to profits more equal across the customer base. When customer service cost heterogeneity is small, all consumers are profitable and there is no lump in the so-called “whale” curve. With all customers being profitable, the most valuable information revealed from the first period purchase is information about the customer’s willingness to pay. Using this information for CCP, the most profitable customers in the second period are the old customers (low and high cost in that order). However, when customer service cost heterogeneity is large, customer cost information revealed is relatively more valuable; using this information for CCP, the low cost type customers (old and new in that order) becomes the most profitable customers in the second period. Managers should judge the efficacy of their customer acquisition and retention strategies 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. We believe the modeling can be extended along a number of dimensions. Currently, we use a two-period monopoly framework. Two natural areas of extension in future research would be: (1) to consider an overlapping generations model where new consumers arrive at a steady rate and the monopolist has to model the tradeoff between acquiring a new generation of customers versus retaining an old generation of customers in steady state (e.g., Villas-Boas 2004), (2) to extend it to a competitive market (e.g., Fudenberg and Tirole 2000, Villas-Boas 1999). 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 worthy of future work.

In this paper, we focused on a setting, where customers are discriminated based on past service cost. This is typical in many markets, because consumers prefer to have fixed prices that are not linked to service usage (for example, Train 1991). The alternative pricing strategy of activity based pricing sets prices based on current activities or services demanded (Narayanan 2003). Ultimately, the choice between these two types of pricing strategies would depend on the level of aversion that consumers may have for being nickel and dimed or being metered constantly on their service usage. A systematic investigation of the tradeoffs between setting fixed prices based on past service costs, or metered usage would be an interesting area of future research. Also, we only focus on short-term pricing. We abstract away from the possibility of long-term commitment. For example, firms may try to escape the Coasian dynamics considered in the current paper by committing to a higher second-period price before making the first period sales. Finally, this paper focuses only on pricing strategies based on customer cost information. Customer costs should also affect other marketing tactics. For example, if churn were to occur naturally among high 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.
Appendix

Proof of Proposition 1

(3) 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 \( p^S = \frac{v + \pi}{2} \), and the total profit will be \( \Pi^S = \frac{(1 + \delta)(v - \pi)^2}{2} \). It immediately follows that \( \Pi^S = \frac{(1 + \delta)(v - \pi)^2}{2} > \frac{(v - \pi)(2 + \delta)^2}{2(4 + \delta)} = \Pi^* \) for all \( \delta \leq 1 \). Q.E.D. ■

Proof of Proposition 2

(3) From the proof of Proposition 1, we know that \( \Pi^S = \frac{(1 + \delta)(v - \pi)^2}{2} \). It immediately follows that \( \Delta \Pi = \Pi^* - \Pi^S = \frac{\delta(2\delta(v - s^H)^2 + (\Delta s)^2 - 4v(v - 2\pi) - 4sL^s)}{8(4 - \delta)} \). Therefore, \( \Delta \Pi \geq 0 \) if and only if

\[
\begin{align*}
&\Rightarrow s^H - s^L \leq \frac{-v - s^L}{2(4 - \delta)} + (1 - \delta) & \text{or } s^H - s^L \geq \frac{(v - s^L)}{2(1 + 2\delta)} - (1 - \delta)\\
&\Leftrightarrow s^H - s^L \leq \frac{-v - s^L}{2(4 - \delta)} + (1 - \delta) & \text{or } s^H - s^L \geq \frac{(v - s^L)}{2(1 + 2\delta)} - (1 - \delta)
\end{align*}
\]

Also, \( \frac{-v - s^L}{2(4 - \delta)} + (1 - \delta) < 0 \) and \( \frac{\delta v^\max_H}{2} < \frac{(v - s^L)}{2(1 + 2\delta)} - (1 - \delta) \) because \( v^\max_H < v^\max_L = v - s^L \) and \( \delta(1 + 2\delta) < 2\delta(1 + 2\delta) \) for all \( \delta \in [0, 1] \). Also, note that \( v^\max_L = v - s^H + s^H - s^L = v^\max_H + \Delta s \). Using this, we can see that \( \Delta \Pi = \Pi^* - \Pi^S > 0 \) if \( \Delta s \geq \frac{v^\max_L (2 + \sqrt{2(4 - \delta)})}{2(1 + 2\delta)} \Leftrightarrow v^\max_H (2 + \sqrt{2(4 - \delta)}) \leq \Delta s \).

Q.E.D. ■

Proof of Lemma 3

The marginal consumer in the first period, \( \tilde{w}_1^j = \tilde{w}_1^j(p_1) \), can be calculated by using the fact that \( p_2^O < p_1 \) and the marginal consumer does not get any surplus in the second period if she already purchased it in the first period: \( \tilde{w}_1^H + \tau - p_1 = \delta (\tilde{w}_1^H + \tau - p_2^O) \Leftrightarrow \tilde{w}_1^H = \frac{p_1 - \tau - (p_2^O - \tau)}{1 - \delta} \), and \( \tilde{w}_1^L - p_1 = \delta (\tilde{w}_1^L - p_2^O) \Leftrightarrow \tilde{w}_1^L = \frac{p_1 - \delta p_2^O}{1 - \delta} \).

Therefore, we can obtain the first period cutoff line for purchasing a product by using the fact that \( p_2^O = \tilde{w}_1^H + \tilde{w}_1^L + s^H + s^L + \tau : \tilde{w}_1^H = \frac{2p_1 - \delta \pi}{2 - \delta} - \tau, \tilde{w}_1^L = \frac{2p_1 - \delta \pi}{2 - \delta} \).

In the first period, the monopolist maximizes the following total expected profit with a common
discount factor $\delta < 1$:

$$
\Pi(p_1) = (p_1 - s^H)(v - \hat{w}^H) + (p_1 - s^L)(v - \hat{w}^L) + \delta \{(p_2^H - s^H)(v + \tau - p_2^H) + (p_2^L - s^L)(v - p_2^L)
+ (p_2^O - s^H)(\hat{w}^H_1 + \tau - p_2^O) + (p_2^O - s^L)(\hat{w}^L_1 - p_2^O)\},
$$

where $\hat{w}^H_1 = \frac{2p_2^H - \delta s^H}{2 - \delta} - \tau, \hat{w}^L_1 = \frac{2p_2^L - \delta s^L}{2 - \delta}, p_2^L = \hat{w}^L_1, p_2^H = \frac{v + \tau + s^H}{2}$.

Taking the first order condition gives us the first period price $p_1$:

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

And the first period marginal consumer is now,

$$
\hat{w}^L_1 = \frac{2v + 2\bar{\pi} - \delta s^H}{4 - \delta} + \frac{(2 - \delta)\tau}{2(4 - \delta)}, \quad \hat{w}^H_1 = \frac{2v + 2\bar{\pi} - \delta s^H}{4 - \delta} - \frac{(6 - \delta)\tau}{2(4 - \delta)}.
$$

Hence,

$$
p_2^H = \frac{v + s^H + \tau}{2}, \quad p_2^L = \hat{w}^L_1 = \frac{2v + 2\bar{\pi} - \delta s^H}{4 - \delta} + \frac{(2 - \delta)\tau}{2(4 - \delta)}.
$$

Further, we can check that $\frac{v + s^L}{2} < \hat{w}^L_1$ and $\frac{v + \tau + s^H}{2} - \tau > \hat{w}^H_1$ are satisfied in equilibrium when $\frac{\delta(v - s^H)}{2} < \Delta s$:

$$
\hat{w}^L_1 - \frac{v + s^L}{2} = \frac{\delta(v - s^L) + 2(1 - \delta)(s^H - s^L) + (2 - \delta)\tau}{2(4 - \delta)} > 0,
$$

$$
\hat{w}^H_1 - \left(\frac{v + \tau + s^H}{2} - \tau\right) = \frac{\delta(v - s^H) - 2(s^H - s^L) - 2\tau}{2(4 - \delta)} < 0.
$$

Q.E.D. ■

**Proof of Proposition 5**

1. $\Delta = p^H_{2, ex} - p^L_{2, ex} = \frac{2(s^H - s^L) - \delta(v - s^H)}{s^H - s^L}$ and $\Delta_{en} = p^H_{2, en} - p^L_{2, en} = \frac{\tau}{s^H - s^L}$.

   If we compare $\Delta_{ex}$ and $\Delta_{en}$, $\Delta_{ex} > \Delta_{en}$ when $\tau < \tau_{IC}$.

   Also, $p^L_{2, en} - p^L_{2, ex} = \frac{\delta(2(s^H - s^L) - \delta(v - s^H) + 4\tau) - 8\tau}{16 - 8\delta} > 0$ and $p^H_{2, en} - p^H_{2, ex} = -\frac{\delta(2(s^H - s^L) - \delta(v - s^H) + 4\tau) - 8\tau}{2\delta(4 + \delta)} < 0$ when $\tau < \tau_{IC}$.

2. $\hat{w}^L_{1, ex} - \hat{w}^L_{1, en} = -\frac{2\delta(s^H - s^L) - \delta^2(v - s^H)}{16 - 8\delta} > 0$ and $\hat{w}^H_{1, ex} - \hat{w}^H_{1, en} = -\frac{2\delta(s^H - s^L) - \delta^2(v - s^H)}{16 - 8\delta} < 0$. 

33
0 when \( \tau < \tau^{IC} \).

Also, \( p_{1,ex} - p_{1,en} = -\frac{(2-\delta)\tau_4}{4(4-\delta)} < 0 \).

3. \( \Pi_{ex}^{CP} - \Pi_{en}^{CP} = \frac{(28(s^H - s^L) - \delta^2(v-s^H) - 4(2-\delta)\tau)^2}{88(16-\delta^2)} > 0 \) when \( \tau < \tau^{IC} \).

Q.E.D. ■

Proof of Proposition

First, we calculate the profit under no discrimination. Without price discrimination, the monopolist simply maximizes the following per-period profit function \( \Pi^S = (p^S - s^H)(v - p^S + \tau) + (p^S - s^L)(v - p^S) \). The optimal price is \( p^S = \frac{2v + s^H + s^L + \tau}{4} \), and the total profit is \( \Pi^S = (1 + \delta) \left[ \Pi^T \right] = (1 + \delta) \left[ \frac{(2v - s^H - s^L)^2 + 2(2v - 3s^H + s^L)\tau + \tau^2}{8} \right] \). Using the results of \( p_1, p_2^H \), and \( p_2^L \) in the profit function of Equation [20], we get

\[
\Pi^* = \frac{2v^2(4+\Omega) + (s^H)^2(2+\delta)^2 + (s^L)^2 + 2s^L(2v - \tau) - 2s^H(2+\delta)(4v - s^L(2-\delta)) + \Omega(4v - 6s^H + \tau^2)}{8(4-\delta)}
\]

\( \Delta \Pi = \Pi^* - \Pi^S = \frac{\delta(2v^2(2-\delta) + 4v\tau - (s^H)^2(1+2\delta) - (s^L)^2 + 2(s^H - s^L)\tau + \tau^2)}{8(4-\delta)} \).

Therefore, \( \Delta \Pi \geq 0 \) if and only if \( \Delta s \geq \frac{(2v - s^L + \tau) \left( \sqrt{2(4-\delta)} - 2(1-\delta) \right) - (1 + 2\delta)\tau}{1 + 2\delta} \). Hence, as \( \Delta s \) becomes larger, \( \Pi^* > \Pi^S \).

Using the results of \( p_1, p_2^H \), and \( p_2^L \) in the profit function of Equation [20], we get

\[
\Pi^* = \frac{1}{88(4+\delta)} \left\{ (2v - s^H - s^L)^2(2+\delta) + 2s^L(2v(4 + \delta)A - s^L(4 - \delta(4 + A)) - s^H(4 + \delta(10 + 3\delta)))\tau - (32 - \delta(28 + \delta(2 + A)))\tau^2 \right\},
\]

where \( A = 2 + \delta \). Also, we know that \( \Pi^S = (1 + \delta) \left[ \Pi_T^S \right] = (1 + \delta) \left[ \frac{(2v - s^H - s^L)^2 + 2(2v - 3s^H + s^L)\tau + \tau^2}{8(4+\delta)} \right] \).

Hence, \( \Delta \Pi = \Pi^* - \Pi^S = \frac{-(2v - s^H - s^L)^2 + 2s^L(8 + 5\delta - 6\delta - s^L(8 - \delta))\tau - (32 - (24 + \delta)\delta)\tau^2}{88(4+\delta)} \), where \( \Delta \Pi \) is a concave function of \( \tau \). So if we show that \( \Delta \Pi \) is positive when \( \tau = \tau^{IC} \), there exists \( \tau^* \) which makes \( \Pi^* > \Pi^S \). First, we note that when \( \tau = 0 \), \( \Delta \Pi = -\frac{(2v - s^H - s^L)^2}{8(4+\delta)} < 0 \). Also, when \( \tau = \tau^{IC} \), \( \Delta \Pi \) is positive when \( \Delta s > \Delta s^* \).

\[
\Delta s^* = \frac{(v - s^H)(32 - 16\sqrt{2(4-\delta)} - \delta(8 - \sqrt{2(4-\delta)} - \delta(18 + \delta)))}{16 - 6(4 - \delta)(16 + \delta)} > \frac{\delta(v - s^H)}{2}.
\]

Hence, as \( \Delta s \) becomes larger, \( \Pi^* > \Pi^S \). Therefore, when \( \Delta s > \Delta s^* \), there exists \( \tau^* \) so that \( \Pi^* > \Pi^S \) when \( \tau^* < \tau < \tau^{IC} \). ■

Q.E.D. ■

\footnote{For example, when \( v = 3, s^H = 2.9, s^L = 2, \delta = 1 \), we can see that \( \Pi^* > \Pi^S \) if \( 0.167 < \tau \leq \tau^{IC} = 0.425 \).}
References


Table 2a: Low Service Cost Heterogeneity ($\Delta s = 0.3$)

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old</td>
<td>New</td>
<td>Benchmark 1</td>
<td>Benchmark 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low Cost</td>
<td>High Cost</td>
<td>Old</td>
<td>New</td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>Margin</td>
<td>0.49</td>
<td>0.19</td>
<td>0.34</td>
<td>0.525</td>
<td></td>
</tr>
<tr>
<td>Quantity</td>
<td>0.43</td>
<td>0.43</td>
<td>0.86</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>Customer Share</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Period 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old</td>
<td>New</td>
<td>Benchmark 1</td>
<td>Benchmark 2</td>
</tr>
<tr>
<td></td>
<td>Low Cost</td>
<td>High Cost</td>
<td>Old</td>
<td>New</td>
</tr>
<tr>
<td>Price</td>
<td>0.77</td>
<td>0.77</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>Margin</td>
<td>0.77</td>
<td>0.47</td>
<td>0.46</td>
<td>0.16</td>
</tr>
<tr>
<td>Quantity</td>
<td>0.43</td>
<td>0.43</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>Customer Share</td>
<td>29%</td>
<td>29%</td>
<td>21%</td>
<td>21%</td>
</tr>
<tr>
<td>Retention</td>
<td>100%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Benchmark 1: traditional behavior-based price discrimination based only on the past purchase history.
†Benchmark 2: the case without price discrimination in which the firms uses neither the past purchase information nor the customer cost type information.
### Table 2b: High Service Cost Heterogeneity ($\Delta s = 1$)

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th></th>
<th>Period 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old</td>
<td>New</td>
<td>Benchmark 1</td>
<td>Benchmark 2</td>
</tr>
<tr>
<td></td>
<td>Low Cost</td>
<td>High Cost</td>
<td>Low Cost</td>
<td>High Cost</td>
</tr>
<tr>
<td>Price</td>
<td>0.67</td>
<td>0.67</td>
<td>0.73</td>
<td>0.85</td>
</tr>
<tr>
<td>Margin</td>
<td>0.67</td>
<td>-0.33</td>
<td>0.23</td>
<td>0.35</td>
</tr>
<tr>
<td>Quantity</td>
<td>0.39</td>
<td>0.39</td>
<td>0.571</td>
<td>0.70</td>
</tr>
<tr>
<td>Customer Share</td>
<td>50%</td>
<td>50%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                  | Low Cost | High Cost        | Low Cost | High Cost        | Old    | New    |
| Price            | 0.81     | 1.1              | 0.65     | 0.65             | 0.91   | 0.71   | 0.85   |
| Margin           | 0.81     | 0.1              | 0.65     | -0.35            | 0.41   | 0.21   | 0.35   |
| Quantity         | 0.39     | 0.1              | 0.15     | 0.15             | 0.571  | 0.42   | 0.70   |
| Customer Share   | 49%      | 13%              | 19%      | 19%              |
| Retention        | 100%     | 25%              |          |                  |

*Benchmark 1: traditional behavior-based price discrimination based only on the past purchase history.

†Benchmark 2: the case without price discrimination in which the firm uses neither the past purchase information nor the customer cost type information.
Figure 5: Inverse Lorenz Curves of Cumulative Profit

(a) Low Service Cost Heterogeneity: $\Delta s \leq \frac{\delta v_{H}}{2}$

First Period

(b) High Service Cost Heterogeneity: $\Delta s > \frac{\delta v_{H}}{2}$

First Period