A pricing model for cloud computing service

Tayfun Keskin
University of Washington Bothell
keskin@uw.edu

Nazim Taskin
Massey University
n.taskin@massey.ac.nz

Abstract
The major purpose of this study is to examine the cloud services pricing schemes and how they can improve previous pricing models by expanding the consumer set with time inconsistent behavior. The industry of cloud computing services is in its infancy, and firms employ pricing models based on conventional information goods. We offer a new approach to cloud services pricing considering the consumer discounting behavior. First, we propose a baseline model based on a profit maximizing duopolistic market serving to both rational and time inconsistent users. Our results reveal that the firms can profit from impatient users. In addition, we extend the baseline model with the effect of delayed network externalities because, by nature, information goods exhibit this property strongly. For the latter case, we show that the effect of network externalities reduces the impact of low switching costs and the monopolist benefits from time-inconsistent behavior. This study contributes to the theories of pricing information goods, and practitioners who make pricing decisions for cloud computing services.

1. Introduction

Amazon web services (AWS), the incumbent firm in the cloud computing industry, announced that it reduced prices 24 times for their Amazon S3 (cloud storage) service in their year-end newsletter [1]. The competitive environment is very intense in cloud computing industry with high capital investments and aggressive pricing [2]. Cloud providers even test unconventional and predatory pricing strategies including bidding to gain a bigger market share [3].

Retaining the customer base is another goal for cloud computing providers. They lock-in customers by complementary products, and form standards to prevent portability. For example, switching from AWS to Microsoft Azure, or another cloud computing provider, is not a simple copy and paste between two machines since the information is stored in different standards on physical drives. Switching includes additional complexities such as complementary services and training for the new system. To summarize, there are significant costs associated for switching from one cloud computing provider to another even without contractual agreements.

This study includes an analytical model to represent the cloud computing market and to understand the economic rationale behind the intense price competition despite high switching costs for the customers. First, we start with a baseline two-period Hotelling model [4]. Then, we improve the baseline model by incorporating the tiered market structure of the cloud and time-inconsistent consumer discounting behavior.

2. Theoretical Background

2.1. What is cloud computing?

The theory behind cloud computing, computing as utility, dates back to the birth of mainframe computing, where computing time was shared amongst multiple users and terminals. Cloud computing uses internet-based and other contemporary virtualization technologies to combine clusters of computers in different geographic locations into a single computing entity [5]. The basic offerings of cloud computing include software as a service (SaaS), platform as a service (PaaS) or infrastructure as a service (IaaS.) [6, 7] These services are available for consumers and business entities on a range of devices such as smart phones and personal computers. SaaS is used for a variety of social applications. An example is Office 365 for e-mail and conferencing. Since these applications are run in the cloud, there is no need to install and maintain them on local hardware. PaaS provides operating system and/or server capabilities, limiting the need to purchase and maintain software and expensive hardware. A well-known PaaS offering is Windows Azure. Lastly, IaaS is comprised of virtual machines that are available as a utility. If there are spikes in demand, a business can deploy additional
machine capacity automatically or manually as a “spot instance” through the IaaS cloud provider. The major player in IaaS is AWS.

In summary, “the cloud” is a contemporary approach to innovate services for firms, and institutions. It is latest method of service innovation, and it viewed much more than a cost-cutting tool for information technology [8].

2.2. What is different with cloud computing pricing and market?

Cloud computing is a dynamic market at its infancy. The industry leader AWS started in 2006 and almost all others followed suit. The industry was not technically feasible before ubiquitous broadband internet connection. The cost of hardware continuously decreased over time and had the same impact on infrastructure pricing.

There are several players as of today (2013) [5] and a few market leaders such as Amazon, Microsoft, Google, IBM stand out as the main players in the cloud computing industry. The market is inherently asymmetric and dominated by AWS at least in the infrastructure layer. We incorporated this asymmetric market structure in our model development starting with the base line model.

There are different standard and technologies in the cloud and switching is a concern for customers once they start using services of a cloud provider. For example, AWS uses a propriety encryption technology [1] which would delay data transfer to other cloud providers. There is an argument where switching costs hinder competition [9, 10] but it is a reality in the cloud market. Switching costs are not only limited to technical barriers. Some providers employ service contracts, and others use complementary operating systems or software (such as Microsoft) to increase barriers to switch [5].

Consumer behavior also differs greatly in the cloud market. The capability to utilize cloud services whenever and wherever fuels impatience of information goods consumers. To represent this type of customers we employ a time-inconsistent discounting factor: hyperbolic discounting [11].

All aforementioned differences of the cloud market from conventional commodity markets provide an opportunity for interesting extensions to the canonical switching cost models [10, 12] and empirical switching cost research [13]. In our literature review, we could only find a handful of studies [14-16] in the information systems field that investigate information related markets with switching costs and none about the cloud computing industry. Why is there an intense price competition in the cloud services industry? Can we forecast the future of the industry in terms of pricing and market share through analytical methods?

In this study, we answer the questions above and explain the processes in which cloud computing industry players determine their prices using a stylized model.

3. Model

We develop and present a two period Hotelling model [4] with asymmetric market shares for a duopolistic market, considering all the differences mentioned above in the cloud computing. We use this model to investigate how prices and market shares would change over time and to explain why the dominant player (AWS in our case) justifies to engage in price wars despite its dominant market share.

We summarize the notation used in this paper in Table 1.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u )</td>
<td>Utility of the consumer</td>
</tr>
<tr>
<td>( j )</td>
<td>Firm: ( j \in {a, b} )</td>
</tr>
<tr>
<td>( t )</td>
<td>Period of consumption: ( t \in {0, 1, 2} )</td>
</tr>
<tr>
<td>( s )</td>
<td>Switching cost: ( s \sim U[0, 0] )</td>
</tr>
<tr>
<td>( e )</td>
<td>Network effect on consumer’s utility</td>
</tr>
<tr>
<td>( \delta )</td>
<td>Exponential discount factor: ( 0 &lt; \delta &lt; 1 )</td>
</tr>
<tr>
<td>( h )</td>
<td>Hyperbolic discount parameter: ( h &gt; 0 )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Marginal shifting cost</td>
</tr>
<tr>
<td>( p_1, p_2 )</td>
<td>Prices charged in periods 1 and 2, respectively</td>
</tr>
<tr>
<td>( q_1, q_2 )</td>
<td>Quantities sold in periods 1 and 2, respectively</td>
</tr>
</tbody>
</table>

We consider a cloud computing market served by two firms \( \{a, b\} \) with asymmetric market shares \( (q^a_0, q^b_0) \). One firm is dominant with a larger market share: \( 0 \leq q^b_0 < 0.5 < q^a_0 \leq 1 \). This is not only a more realistic representation of today’s cloud computing market, but it also covers a wider range of possible scenarios than a duopoly with equal market shares.

We assume that, in the market setting above, there is a continuum of consumers uniformly distributed between firms \( a \) and \( b \). Consumers have the utility \( u \), which is a constant derived from the consumption of the cloud computing service. Marginal shifting cost reflects the marginal change in the utility derived from using the cloud service according to the distance of the customer from buying firm \( a \) or \( b \).

Firms \( a \) and \( b \) engage in price competition in both periods. Prices announced simultaneously after they observe quantities sold \( (q) \), which also represent the market shares of both firms.

Customers suffer additional costs, \( s \), due to switching from one firm to the other. Customer type does not change over time.
3.1. Baseline model

We consider a one-shot game in which firms commit prices $p_1$, $p_2$. Consumers make purchase decisions based on their utilities. Let $p_j^t$ represent the price of firm $j$ in period $t$. The term $x^j_t$ is the distance of the customer from buying firm $j$. The indifferent consumer for firm $a$ in the second period can be characterized as:

$$u - ax^a - p^a_2 = u - a(1 - x^a) - p^a_2 - s$$

These indifferent consumer boundary determine new market shares for firm $a$ and $b$ at the end of the second period. In the cloud computing industry, utility is derived from using per unit of service but in the baseline model, we do not allow for differentiated services.

We use backward induction to find equilibrium prices and quantities sold to represent market shares. Our primary goal is to determine if there is an optimal solution for firm revenues and pricing.

Firm $b$’s indifferent customer is:

$$u - a(1 - x^b) - p^b_2 = u - ax^b - p^b_2 - s$$

We can determine the new allocation of the market shares for firm $a$ and $b$ at the end of the second period by finding the quantity of switching customers. To find second term market shares in terms of second period prices, we start with switching costs:

$$s^a = \alpha(2x^a - 1) + p^a_2 - p^b_2$$

$$s^b = \alpha(1 - 2x^b) - p^a_2 + p^b_2$$

Let $n^j_k$ be the quantity of customers who bought from $k$ in period $t-1$, and firm $j$ in period $t$. For example, customers who switched to firm $b$ from firm $a$ in period 2 are represented as $n^b_a$.

Customers staying with firm $a$:

$$n^a_a = \int_0^{q_1} \left( \int_0^\theta \frac{1}{\theta} ds \right) dx = q_1(\alpha(q_1 - 1) - p^a_2 + p^b_2 + \theta)$$

Customers switching from firm $a$ to firm $b$:

$$n^b_a = q_1 - n^a_a$$

Customers staying with firm $b$:

$$n^b_b = \int_0^{q_1} \left( \int_0^\theta \frac{1}{\theta} ds \right) dx$$

$$= q_1(\alpha(q_1 - 1) + p^a_2 - p^b_2)$$

Customers switching from firm $b$ to firm $a$:

$$n^{ab} = 1 - q_1 - n^b_b$$

$$= (q_1 - 1)(\alpha q_1 + p^a_2 - p^b_2)$$

Market share for firm $a$ at the end of period 2:

$$q^a_2 = n^a_a + n^{ab}$$

$$= q_2 + \frac{p^b_2 - p^a_2}{\theta}$$

Market share for firm $b$ at the end of period 2:

$$q^b_2 = n^b_b + n^{ba}$$

$$= 1 - q_1 + \frac{p^a_2 - p^b_2}{\theta}$$

Firm $j$ maximizes its second period profit.

$$\max_p n^j_2 = p^j_2 q^j_2$$

First order conditions give us equilibrium prices as:

$$p^*_2 = \frac{(1 + q_1) \theta}{3}$$

$$= \frac{(2 - q_1) \theta}{3}$$

Equilibrium quantities sold are:

$$q^*_2 = \frac{(1 + q_1)}{3}$$
As a result of the profit maximization where customers may switch, we obtain second period profits as a function of quantities sold in the first period:

\[
\pi_2^{a*} = \frac{(1 + q_1)^2 \theta}{9}
\]

\[
\pi_2^{b*} = \frac{(2 - q_1)^2 \theta}{9}
\]

Next, we follow the same procedure for period 1. First, we identify the indifferent customers to find switching costs \( s_i \) in terms of \( x_i^l \) and prices. Then we solve the maximization problem for the first period profits to find equilibrium prices and quantities sold. To minimize repetition, these mathematical derivations are in the appendix.

Total profits can be found as a function of initial quantities sold and switching costs:

\[
\pi_1^{a*} = \frac{(13 + 9q_0)(5\theta + 7q_0\theta)}{529}
\]

\[
\pi_1^{b*} = \frac{(10 - 9q_0)(18\theta - 7q_0\theta)}{529}
\]

**THEOREM 1:** There exists a solution for the maximum revenue, thus there are rational pricing strategies for firm a and b.

This is a unique equilibrium where firms a and b pursue rational strategies and consumers purchase in equilibrium.

One of the main objectives of any firm is the market share. Obviously, the ultimate market shares for firm a and b depends on initially sold units but interestingly they approximate each other regardless of the initial value.

\[
q_2^{a*} = \frac{3}{23}(4 + q_0)
\]

\[
q_2^{b*} = \frac{1}{23}(11 - 3q_0)
\]

The following figure represents the relationship between the units sold in the second period and initial units sold for both firms.

To sum up, market leader loses its share and the smaller firm gains a bigger market over time. This fact may explain the rationale behind predatory pricing in the cloud computing market. If the market leader (For example, Amazon Web Services) follows rational profit maximizing strategy, eventually other firms can catch up with them in terms of the market share.

**PROPOSITION 1a:** The difference between units sold by firm a and b reduces over time.

Pricing decision is another important consideration for firms. Cloud computing industry is relatively new and pricing is a vital concern for players. Therefore, finally, we are interested in determining how prices change along with the value of switching costs.

We know that the second period equilibrium price for firm a is:

\[
p_2^{a*} = \frac{(1 + q_1)\theta}{3}
\]

At this point we can substitute units sold and find a price formula in terms of switching costs and initial units sold.

\[
p_2^{a*} = \frac{3(4 + q_0)\theta}{23}
\]

Since \( q_0 \in (0,1) \) it can be shown that:

\[
\frac{\partial p_2^{a*}}{\partial \theta} > 0
\]

This means that when switching costs decrease, prices will decrease with them. This further explains
our initial question of why the market leader in cloud computing reduced its prices over time. This result holds for firm $b$.

Several factors in the cloud industry affect switching costs. First, standards emerge over time and cloud providers lose their edge with their proprietary technology. Second, availability of services with better technology and faster communication reduces switching costs for consumers. Finally, security and connectivity concerns are addressed both by improvements and training. In short, prices follow downward trend along with the switching costs for consumers.

**PROPOSITION 1b:** There is a positive relationship between the switching costs and prices.

However, this does not mean that over time cloud computing industry will turn into a non-profitable commodity market. Firms can invent new ways to increase switching costs for consumers. For example Amazon’s reputation creates a de-facto barrier. Moreover, their proprietary encryption creates an additional layer.

Competitive behavior with switching costs may look collusive [10]. In fact, our solution revealed that market share decreased for firm $a$ and increased for firm $b$ in the second period. However, tendency for prices to decrease over time alleviates these concerns for consumers if switching costs are not high.

### 3.2. Pricing with time-inconsistent consumers

Consumers lost their patience with the rapid growth of internet technologies and wireless broadband. Ubiquitous nature of information goods further increase impatience of these customers. In this extension, we include consumers with time-inconsistent discounting behavior. These consumers exhibit hyperbolic discounting behavior rather than rational exponential discounting with a constant discount factor $\delta$. In this paper we call impatient customers, time-inconsistent discounters, or hyperbolic discounters to identify a new segment.

Lowenstein and Prelec [17] initially suggested a discounting utility model for time-inconsistent consumers:

$$u(c_t, ..., c_T) = \sum_{k=0}^{T-t} D(k)u(c + k)$$

Where the hyperbolic discount function is:

$$D(k) = \left( \frac{1}{1 + \frac{k}{\delta}} \right)^{\alpha/\alpha}$$

We follow the literature [11] and set $\alpha/\alpha = 1$ and $T=2$ for simplicity. Please note that the hyperbolic discount function $D(k)$ declines faster in the short run than in the long run. This property will introduce an interesting and realistic segment to the cloud computing market.

In this section, consumer market has two different segments. First type “patients”, exhibit exponential discounting behavior. Second type “impatients”, exhibit hyperbolic discounting behavior.

We solve a maximization problem with hyperbolic and exponential discounters similar to the procedure in section 3.1 to observe changes in units sold (market share) and prices. With the discounting included both exponential and hyperbolic, we show that:

$$\left( \frac{\partial q^*_{\alpha}}{\partial \theta} \right)_{\text{exponentials}} > \left( \frac{\partial q^*_{\alpha}}{\partial \theta} \right)_{\text{hyperbolic}} > 0$$

The units sold decrease less for time-inconsistent consumers (hyperbolics) than rational consumers (exponentials). These results hold for firm $b$.

This means that firms that serve impatient customers are less prone to losing market shares when switching costs decrease.

This finding is important because firms can target impatient customers in today’s cloud computing market and protect themselves from future changes in parameters that they cannot entirely control, such as decreasing switching costs.

**PROPOSITION 2a:** The units sold reduce less when switching costs decrease for the hyperbolic discounter segment.

Pricing can also differ among rational and time-inconsistent consumers and firms can benefit from identifying these segments.

$$\left( \frac{\partial p^*_{\alpha}}{\partial \theta} \right)_{\text{exponentials}} > \left( \frac{\partial p^*_{\alpha}}{\partial \theta} \right)_{\text{hyperbolic}} > 0$$

These results also hold for firm $b$.

In our updated model, serving time-inconsistent consumers benefited firms because prices decreased less for hyperbolics than exponentials.
PROPOSITION 2b: The prices reduce less for the hyperbolic discounter segment than exponential discounter segment when switching costs decrease.

Therefore, it is safe to claim that firms can benefit from serving hyperbolic discounters not only in terms of the market share but also in terms of the pricing.

3.3. Pricing with network externalities

Information goods inherently exhibit network externalities [18]. This property can be used to further improve the model to a more realistic setting. Not all players in the cloud computing industry offer all three layers of cloud computing service (SaaS, PaaS, IaaS.) For example, Microsoft delivers all three services but Amazon does not deliver SaaS (as of May 2013.) Firms delivering more than one service layer in the cloud can benefit from this extension of our baseline-pricing model.

In this case, we consider cloud computing services exhibit delayed positive network externalities. This means that the utility cloud user derives from using the service increases with the past sales. Firms better off in the second period if they capture a higher market share in the first period.

Consumer market also has two different segments similar to the section 3.2.

The network effect on consumer utility is represented by the variable $e$. In order to compare and contrast two extremes: $e = 0$ and $e = 1$. We solve the cloud computing pricing model with another extension.

We find that:

$$\left(\frac{\partial q^*_2}{\partial \theta}\right)_{\text{no network externalities}} > \left(\frac{\partial q^*_2}{\partial \theta}\right)_{\text{positive network externalities}} > 0$$

The units sold (market share) decrease less for the market leader when the service or product exhibit network externalities.

PROPOSITION 3a: The units sold reduce less for the market leader when cloud computing service exhibit network externalities.

We also show that:

$$\left(\frac{\partial p^*_2}{\partial \theta}\right)_{\text{no network externalities}} > \left(\frac{\partial p^*_2}{\partial \theta}\right)_{\text{positive network externalities}} > 0$$

PROPOSITION 3b: The prices reduce less for the market leader when cloud computing service exhibit network externalities.

For this extension with network externalities, we only present our findings for the larger firm with the dominant initial market share because inherently network externalities are predisposed to benefit the firm with the higher units sold (market share).

4. Conclusion

Advances in information technologies provided us with products and services that exhibit complex properties such as cloud computing. Network externalities, switching costs and unconventional consumer behavior are the main characteristics of the cloud computing market. In our study, we incorporated all of these factors in our model to investigate pricing strategies, and resulting market shares. To sum, this study contributes to the theories of pricing information goods.

Behavioral economics provides us with tools to investigate deviations from conventional rationality principles. With these tools we can create models for more realistic settings. Hyperbolic discounting is one example for the irrational behavior through which we can model time-inconsistent consumers. In this research we contribute to the behavioral economics literature by developing a model in the intersection of switching costs, time-inconsistent discounting, and network externalities. In the literature there are currently no papers addressing all these properties for information goods. Even cloud computing, as a concept, is relatively young and there are only a handful of papers [2, 5, 7] on this exciting area.

Practitioners in the cloud computing industry are in dire need of new pricing models, because conventional models are limited to assist them with pricing decisions [11, 18]. Another goal of this study is to provide tools to practitioners who make pricing decisions for cloud computing services.

We find that firms can profit from identifying impatient users. Hyperbolic discounters market is not only immune to lower switching costs in terms of market share, but also generate higher revenue for firms when they identify that segment. This finding can help practitioners to survive in the industries similar to the cloud computing, even under a predatory pricing strategy of an incumbent market leader.

We extend the model with the effect of delayed network externalities because, by nature, information goods exhibit strong network externalities. We show that the effect of network externalities reduces the impact of low switching costs and the monopolist benefits from time-inconsistent behavior.

There are several ways this work can be extended. First, we use a relatively simple model of hyperbolic
discounting function. Our model can be improved by including a more sophisticated form of hyperbolic discounting function. Second, the model is based on two periods. Extending the time horizon to multiple finite periods can provide additional insights to the impact of network externalities on consumers’ utility function. Finally, empirical analyses and validation of our findings are possible but that would require more tedious work since cloud computing service providers are large corporations (such as Amazon, Microsoft, Google and IBM [5]), and they might not be willing to disclose their pricing and market share information.
5. References


6. Appendix: Proofs

We utilize this appendix to illustrate our proofs.

3.1.

For the first period maximization problem repeating a similar calculation for the period 1 gives equilibrium prices as:

\[ p^a_1 = \frac{5\theta + 7q_0\theta}{23} \]
\[ p^b_1 = \frac{18\theta - 7q_0\theta}{23} \]

Equilibrium quantities sold are:

\[ q^a_1 = \frac{(13 + 9q_0)}{23} \]
\[ q^b_1 = \frac{(10 - 9q_0)}{23} \]

As a result of the profit maximization where customers may switch, we obtain first period profits as a function of initial quantities sold:

\[ \pi^a_1 = \frac{(13 + 9q_0)(5\theta + 7q_0\theta)}{529} \]
\[ \pi^b_1 = \frac{(10 - 9q_0)(18\theta - 7q_0\theta)}{529} \]

3.2.

In section 3.2 we investigate the case with time inconsistent consumers. Therefore we employ discounting when we shift from the second period to the first.

The maximization problem for the first period can be updated as:

\[ \max_p \pi^a_{\text{exponentials,1}} = p^a_1 q^a_1 + \delta \frac{(1 + q_1)^2\theta}{9} \]
\[ \max_p \pi^b_{\text{exponentials,1}} = p^b_1 q^b_1 + \delta \frac{(2 - q_1)^2\theta}{9} \]

And for hyperbolic discounters:

\[ \max_p \pi^a_{\text{hyperbolic,1}} = p^a_1 q^a_1 + \frac{1}{(1 + ht)} \frac{(1 + q_1)^2\theta}{9} \]

For 0 < \delta < 1 and h > 0:

\[ \left( \frac{\partial q^a_2}{\partial \theta} \right)_{\text{exponentials}} > \left( \frac{\partial q^a_2}{\partial \theta} \right)_{\text{hyperbolic}} > 0 \]

And

\[ \left( \frac{\partial p^a_2}{\partial \theta} \right)_{\text{exponentials}} > \left( \frac{\partial p^a_2}{\partial \theta} \right)_{\text{hyperbolic}} > 0 \]

3.3.

In section 3.3 we include network externalities by considering network effects on consumers’ utility. We update the indifferent consumer equation for firm a in period 2 as:

\[ u - \alpha x^a - p^a_1 + eq^a_1 \]
\[ = u - \alpha(1 - x^a) - p^a_1 - s + eq^b_1 \]

Also, the indifferent consumer for firm b in period 2 turns out to be:

\[ u - \alpha(1 - x^b) - p^b_1 + eq^b_1 \]
\[ = u - \alpha x^b - p^b_1 - s + eq^b_1 \]

Then we solve the profit maximization problem for prices and units sold similar to the previous section.

We find that:

\[ \left( \frac{\partial q^a_2}{\partial \theta} \right)_{\text{no network externalities}} > \left( \frac{\partial q^a_2}{\partial \theta} \right)_{\text{positive network externalities}} > 0 \]

And

\[ \left( \frac{\partial p^a_2}{\partial \theta} \right)_{\text{no network externalities}} > \left( \frac{\partial p^a_2}{\partial \theta} \right)_{\text{positive network externalities}} > 0 \]