INSEAD
The Business School for the World®

Working Paper Series
2018/12/DSC

Optimal Pricing and Introduction Timing of New Virtual Machines

Ian A. Kash
Microsoft Corporation - Microsoft Research Cambridge

Peter B. Key
Microsoft Corporation - Microsoft Research Cambridge

Spyros Zoumpoulis
INSEAD, spyros.zoumpoulis@insead.edu

March 8, 2018

Cloud service providers can provide increasingly powerful upgrades of their virtual machines to their customers, but at a launching cost, and at the expense of the sales of existing products. We propose a model of product introduction and characterize the optimal pricing and timing of introductions of new virtual machines for a cloud service provider in the face of customers who are averse to upgrading to improved offerings. Overall, we show that a simple policy of Myerson (i.e., myopic) pricing and periodic introductions is near-optimal. We first show that under a Myerson pricing rule, there is no loss of optimality with a periodic schedule of introductions, and that under periodic introductions, the potential additional revenue of any pricing policy over Myerson pricing decays to zero after sufficiently many introductions. We then show that, given arbitrary fixed introduction times, Myerson pricing is approximately optimal. To do so, we characterize the prices that achieve optimal revenue in a single introduction period and provide a bound for the competitive ratio of Myerson pricing over the optimal single-period pricing. This bound shows that Myerson pricing is approximately optimal when switching costs for the customers who upgrade are small or large. Following our analysis, we examine our analytical bounds for Myerson pricing with simulations and show that they can provide strong guarantees for all values of the switching cost, for several natural distributions for the customer type. Furthermore, when we numerically compute optimal prices, rather than using our bounds, we find that Myerson pricing is often several orders of magnitude closer to optimal than our analytical bounds suggest.

Keywords: New Product Introduction; Pricing; Cloud Computing; Competitive Analysis

Electronic copy available at: https://ssrn.com/abstract=3137038

A Working Paper is the author's intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu
Find more INSEAD papers at https://www.insead.edu/faculty-research/research

1 Introduction

As the quality of computer hardware increases over time, cloud service providers have the ability to offer more powerful virtual machines (VMs) and other resources to their customers. But providers face several trade-offs as they seek to make the best use of improved technology. On one hand, more powerful machines are more valuable to customers and command a higher price. On the other hand, there is a cost to develop and launch a new product. Further, the new product competes with existing products. Thus, the provider faces two questions. First, when should new classes of VMs be introduced? Second, how should they be priced, taking into account both the VM classes that currently exist and the ones that will be introduced in the future?

This decision problem, combining scheduling and pricing new product introductions, is common in a variety of settings. One aspect that is more specific to the cloud setting is that VMs are rented rather than sold. Thus existing customers can switch to a new offering, albeit with some inconvenience. There is indeed evidence of customers' aversion to upgrades in the cloud computing services market. Based on a study of Microsoft Azure, we estimate that customers who arrive after a new VM class is launched are 50% more likely to use it than existing customers, indicating that these switching costs may be substantial (see Appendix A for the analysis).

This opens up a wide range of possible policies for the cloud service provider. Our main result is that a surprisingly simple policy is close to optimal in many situations: new VM classes are introduced on a periodic[1] schedule and each is priced as if it were the only product being offered. We refer to this pricing policy as Myerson pricing, as these prices can be computed as in Myerson's classic paper (Myerson, 1981). This policy produces a marketplace where new customers always select the newest and best offering, while existing customers may stick with older VMs due to switching costs.

In more detail, we model product introduction as a discrete-time process over an infinite time horizon, with future rewards discounted. The provider seeks to maximize expected discounted return. At each time period, the provider decides whether or not to introduce a new VM class (at a fixed cost) and at what price. The quality of a new VM class is assumed to grow linearly with the

[1] Periodic introductions have been noticeable in the practice of cloud computing. For example, Amazon Elastic Compute Cloud (EC2) launched new classes of the m.xlarge series in October 2007, February 2010, October 2012, and June 2015, i.e., in intervals of 28 months, 32 months, and 32 months (Barr, 2015).

Electronic copy available at: https://ssrn.com/abstract=3137038

introduction time. Customers act as simple utility maximizers, and choose the VM offering that is most valuable to them; if they are existing customers, they incur a switching cost for switching technologies. New customers arrive at each time period, and stay for a fixed number of time periods.

For our first set of results, we analyze policies which introduce VM classes on a periodic schedule and establish two key results. First, if the Myerson pricing rule is always used, then the optimal introduction schedule is in fact periodic. Second, if a periodic schedule is used, then Myerson pricing is optimal in the limit, in the sense that for introduction times sufficiently far in the future, the (undiscounted) benefit of shading prices downward from the Myerson prices goes to zero.

With our second set of results, we argue that, given arbitrary fixed introduction times, Myerson pricing is approximately optimal. Combined with our first result, this means that Myerson pricing with periodic introductions is approximately optimal. Our analysis, which assumes the distribution of customer types satisfies the monotone hazard rate condition, establishes an upper bound on the possible revenue in a single time period and characterizes prices which achieve this upper bound (Theorem 1). We extend this bound to discriminatory policies, which are permitted to charge different prices to new and existing customers (intuitively to provide existing customers a discount to entice them to switch), and provide a bound on the competitive ratio between such policies and Myerson pricing (Proposition 3). This bound shows that Myerson pricing is approximately optimal when switching costs are small or large.

Following our analysis, we conduct numerical simulations of our model for both regular and discriminatory pricing. Our bounds can be used to show that, for several natural distributions, gains over the Myerson policy are less than 10% even at intermediate values of the switching cost. Furthermore, when we numerically compute optimal prices, rather than using our bounds, we find that Myerson pricing is often several orders of magnitude closer to optimal than even this already good bound suggests.

2 Literature Review

We first review works that study the optimal decision making of a firm which launches a new product, or successive generations of a new product. Product launch policies have been studied, among others, in the operations management, marketing, and economics literatures. The most commonly examined aspects of the firm's decision about a new product launch are timing, level of technology, and pricing, all directly relevant for this current work.

We start with the operations management literature. Perhaps the closest papers to ours are the works by Krankel et al. (2006) and Lobel et al. (2016), which both consider a firm that introduces successive generations of a product over an infinite time horizon. Both papers study a trade-off between waiting for further technology improvements, or capturing the gains of technology improvements sooner, possibly at the cost of slowing sales for the existing product. Krankel et al. (2006) specify a state-based model of demand diffusion and construct a decision model to solve the firm's introduction timing problem. They prove the optimality of a state-dependent threshold policy governing the firm's product introduction decisions. Our setup is different in two important ways. First, Krankel et al. (2006) look at durable goods and do not allow for upgrades or switches: each purchase uses a unit from the market potential. Second, they assume a specific pricing strategy (with constant unit profit margins) and don't endogenize the pricing decision. In terms of results, although Krankel et al. (2006) characterize the optimal introduction decisions, their setup does not allow them to conclude optimality of a simple pattern of introduction times (such as periodic). Lobel et al. (2016) consider forward-looking consumers and show that when the firm makes product launch decisions "on the go", it is optimal to release products cyclically, i.e., whenever the developed technology is better than the one available in the market by a constant margin. When the firm is able to precommit to a schedule of releases, the optimal policy generally consists of alternating minor and major technology launch cycles. Our model is different in that we are studying a subscription-based service, so revenue can be picked up all the time; and buyers want to maximize their utility at each period, so they are not forward-looking.

Cohen et al. (1996) consider a firm deciding when to introduce a new product generation to the market. They assume that the product can only be sold during a fixed time window and study the following trade-off: delaying the introduction can lead to more development, and therefore a better product and higher revenues, but over a shorter time. They show that it is better to delay the introduction of the new generation (and develop a better product) if the existing product has a high margin, and when the firm is faced with an intermediate level of competition.

Paulson Gjerde et al. (2002) model a firm's decision regarding the level of innovation to incorporate into successive product generations and show that the structure of the internal and external environment in which the firm operates suggests when to innovate to the technology frontier (as opposed to pursuing incremental improvements). Kumar and Swaminathan (2003) consider a firm that sells an innovative product with a given market potential and that may not be able to meet demand due to capacity constraints. Their demand model, modified from the original model of Bass (1969), captures the effect of unmet past demand on future demand. They show that a heuristic "build-up" policy, in which the firm does not sell at all for a period of time and builds up enough inventory to never lose sales once it begins selling, is a robust approximation to the optimal policy.

Klastorin and Tsai (2004) propose a game-theoretic model with two profit maximizing firms that enter a new market and decide on the timing, design and pricing of their product introduction. Their model shows that product differentiation always arises at equilibrium due to the joint effects of resource utilization, price competition, and product life cycle. They conclude that it is not wise for profit-maximizing firms to arbitrarily shorten product life cycle for the sake of competition, because all firms are worse off.

Casadesus-Masanell and Yoffie (2007) study competitive interactions between Intel and Microsoft through a duopoly model between producers of complementary products. Contrary to the popular view that two tight complements will generally have well aligned incentives, they demonstrate that natural conflicts emerge over pricing, the timing of new product releases, and who captures the greatest value at different phases of product generations.

Plambeck and Wang (2009) study the impact of e-waste regulation on new product introduction in a stylized model of the electronics industry. The manufacturer decides on the development time for the next product generation and on the expenditure level, making different decisions for different types of environmental regulation. Consumers purchase the new product and dispose of the previous generation product, which becomes e-waste. According to their model, "fee-upon-sale" types of e-waste regulation cause manufacturers to increase their equilibrium development time and expenditure, and thus the incremental quality for each new product. As new products are introduced (and disposed of) less frequently, the quantity of e-waste decreases and, even excluding the environmental benefits, social welfare may increase.

Araman and Caldentey (2016) consider a model for new product introduction where the seller has the ability to first test the market and gather demand information through crowdvoting before deciding whether or not to launch a new product. Eventually, the seller stops the voting phase and either discards the product or launches it and starts a regular selling phase. Araman and Caldentey (2016) propose a stopping time formulation to determine the optimal duration of the voting process (and therefore also whether and when to introduce a new product). Their model allows to quantify the trade-off between the value of demand information and the financial impact of delaying the product introduction to accumulate pre-orders, and also sheds some light on how to price the voting phase to stimulate effectively this voting process.

We continue with the relevant marketing literature. The seminal work of Bass (1969) proposes a growth model for the timing of initial purchase of a single innovative product based on diffusion from innovators to imitators. A stream of papers build upon the work by Bass (1969) on product diffusion, by incorporating multiple product generations in their models. Mahajan et al. (1990) review and evaluate the various new product diffusion models proposed in the first two decades after the work by Bass (1969). Bayus (1992) investigates the pricing problem for durables with two successive generations. Norton and Bass (1987) propose a product growth model that encompasses both diffusion and substitution between successive generations of a technology. Pae and Lehmann (2003) focus on the impact of intergeneration time (i.e., time in between two generations) on product diffusion, and show that predictions based on intergeneration time achieve improved accuracy. Stremersch et al. (2010) empirically investigate whether introducing new product generations accelerates demand growth, and find that passage of time, as opposed to generational shifts, is what accelerates growth. Wilson and Norton (1989) consider the one-time introduction timing decision for a new product generation and, under the assumption that the line extension has a lower profit margin, show that it is best either to introduce the line extension early in the life cycle, or not to introduce it at all. Mahajan and Muller (1996) extend the work of Wilson and Norton (1989) to allow for general profit margins and conclude that it will be optimal to either introduce the improved product early, or wait until the previous generation becomes mature.

Technology adoption and launch policies have also been studied in the economics literature. Balcer and Lippman (1984) consider the problem of the adoption of new technology, which improves over time. They show that the firm will adopt the current best practice if its technological lag exceeds a certain threshold; moreover, as time passes without new technological advances, it may become profitable to purchase a technology that has been available even though it was not profitable to do so in the past. Farzin et al. (1998) investigate the optimal timing of technology adoption by a competitive firm when technology choice is irreversible and the firm faces a stochastic innovation process with uncertainties about both the speed of the arrival and the degree of improvement of new technologies. They explicitly address the option value of delaying adoption, compare the optimal decision rule to traditional net present value methods, and observe that the optimal timing decision is greatly affected by technological parameters. Goettler and Gordon (2011) study the effect of competition on innovation in the personal computer microprocessor industry. They propose a dynamic model where firms make dynamic pricing and investment decisions while consumers make dynamic upgrade decisions, anticipating product improvements and price declines. They find that the rate of innovation in product quality would be higher if Intel were a monopolist, though higher prices would reduce consumer surplus. Gowrisankaran and Rysman (2012) propose a dynamic model of consumer preferences for new durable goods that allows for consumers to upgrade to new durable goods as features improve. They estimate their model on digital camcorder purchase data and find that the 1-year elasticity in response to a transitory industrywide price shock is about 25 percent less than the 1-month elasticity.

A recent strand of the literature focuses specifically on the pricing problem for cloud services. Borgs et al. (2014), motivated by the cloud computing market, study a multiperiod pricing problem of a service firm with capacity levels that vary over time, where customers strategically choose the timing of their purchases, and where the firm wants to maximize its revenue while guaranteeing service to all paying customers. They provide a dynamic programming based algorithm that computes the optimal sequence of prices in polynomial time, and their optimal policies only use a limited number of different price levels. Kilcioglu and Maglaras (2015) study a problem of market segmentation for a revenue maximizing cloud service provider that offers two classes of service: guaranteed service (on-demand instances) and best effort (spot instances), in a market with heterogeneous customers with respect to their valuation and congestion sensitivity. They show that in settings where the user congestion cost rate grows faster than the valuation rate, it is optimal for the service provider to make the spot service option stochastically unavailable. Abhishek et al. (2012) model a cloud computing service as a hybrid system where customers can choose to obtain service from a fixed-price queue with infinite capacity, or enter a bid-based priority queue. They characterize user equilibrium behavior and show its insensitivity to the precise market design mechanism used. They provide evidence suggesting that a fixed price typically generates a higher expected revenue than the hybrid system for the provider.

Along the same strand, recent works have modelled the strategic interactions between competing cloud computing service providers. In the same vein as the work by Abhishek et al. (2012), Gao et al. (2017), motivated by the cloud computing services of Microsoft Azure and Amazon EC2, consider a service system with two competing firms: a fixed-price firm and a bid-based firm. They characterize the structure of the resulting equilibrium strategy showing that customer equilibrium behavior has a simple threshold structure, and use this characterization to study the price competition between the two firms. Anselmi et al. (2014) propose a model to study the interaction of price competition and congestion. They characterize competitive equilibria within each level of a three-tier market model that captures a marketplace with users purchasing services from Software-as-Service (SaaS) providers, which in turn purchase computing resources from either Provider-as-a-Service (PaaS) providers or Infrastructure-as-a-Service (IaaS) providers. Nair et al. (2014) consider the interplay between network effects, congestion, and competition in cloud services with ad-supported revenues, and they find that users are generally no better off due to competition, as the congestion levels are of the same order as if there were only one firm. Further, their analysis highlights an important contrast between firm-specific and industry-wide network effects: multiple firms can coexist in a marketplace with industry-wide network effects, but near-monopolies tend to emerge in marketplaces with firm-specific network effects. Kilcioglu and Rao (2016) document empirically the performance differentiation in the cloud Infrastructure-as-a-Service (IaaS) market and use it to study optimal pricing, employing a theoretical model that focuses on capturing competition. Finally, Kash and Key (2016) provide a survey of some of the issues inherent in pricing the cloud, and related research work.

Our model assumes myopic customers, which we argue in Section 7 is a mild assumption, given the structure of the optimal policies. In the canonical formulation of the revenue management problem where a monopolist seller seeks to maximize revenues from selling a fixed inventory of a product to *myopic* customers who arrive over time, maintaining prices fixed at an appropriate level over the selling horizon is asymptotically optimal (Gallego and van Ryzin, 1994). Recent works have allowed for *forward-looking* customers (i.e, customers that strategize about their time of purchase) and have characterized optimal policies that are simple, or admit simple interpretations. Besbes and Lobel (2015) provide a general formulation that allows for arbitrary correlation in customers' patience and valuation, prove that the firm can restrict attention to cyclic pricing policies which have length, at most, twice the maximum willingness to wait of the customer population, and develop a dynamic programming approach that efficiently computes optimal policies. Chen and Farias (2015) propose a "robust" pricing mechanism that guarantees to achieve at least 29% of the expected revenues of an optimal dynamic mechanism. Their robust pricing mechanism enjoys the simple interpretation of solving a dynamic pricing problem for myopic customers, with the additional requirement of a price constraint that discourages rapid discounting. Chen et al. (2017) demonstrate that for a broad class of customer utility models, static prices surprisingly continue to remain asymptotically optimal, and that, irrespective of regime, an optimally set static price guarantees the seller revenues that are within at least 63.2% of the revenues under an optimal dynamic mechanism. Chen and Hu (2017), motivated by the sharing economy, study a model with forward-looking buyers and sellers and a single market-making intermediary, and find a simple heuristic policy to be asymptotically optimal. Under their heuristic policy, forward-looking buyers and sellers behave myopically. Caldentey et al. (2017) consider the dynamic pricing problem in a robust formulation that is based on the minimization of the seller's worst-case regret, without distributional assumptions about customers' willingness-to-pay or arrival times. They characterize optimal price paths for both myopic and strategic customer purchasing behavior. Finally, Liu and Cooper (2015) and Lobel (2017) deviate from strategic customers to study dynamic pricing in the face of patient customers: a patient customer is willing to wait up to a certain number of periods for a lower price and will make a purchase as soon as the price falls below her valuation. Liu and Cooper (2015) prove that there is an optimal dynamic pricing policy comprised of repeating cycles of decreasing prices, yet such cycles may no longer be optimal when customers have variable levels of patience. Lobel (2017) proposes an efficient dynamic programming algorithm for finding optimal pricing policies for arbitrary joint distributions of patience levels and valuations.

3 Model

Time is discrete with an infinite horizon. At each time period t, the cloud service provider can introduce up to one new class of virtual machine (VM) at provisioning cost $C > 0$, and price it at price x_t. The provider thus decides when to introduce new VM classes and how to price them.

We generally assume that VM classes, once introduced, remain available for customers to choose thereafter.[2] We assume an unlimited capacity of VMs for all introduced VM classes.[3]

At each period, a unit mass of new customers arrives. Customers stay in the system for two periods before departing. Each customer has a type $\theta \geq 0$ that is drawn i.i.d. from a distribution with differentiable density f and c.d.f. F.

We assume that the quality of the offered services grows linearly with time, so that a customer of type θ enjoys benefit $\theta \cdot t$ when using a VM of a class that was introduced in period t. A customer incurs cost x_t for using a VM class introduced at time t, and a switching cost $c > 0$ for switching to a different VM class. We assume that customers make decisions myopically, maximizing their utility in the current time period. The assumption of myopic customers is a mild one in our setting, as we discuss in Section 7.

We denote a policy for the provider $\pi = ((s_0 = 0, x_0 = 0), (s_1, x_1), (s_2, x_2), \ldots)$, which specifies a (possibly infinite) sequence of pairs of introduction time and price, where we use (s_i, x_i) to denote the time and price of the ith introduction, respectively. A *newly arriving* customer of type θ at time t simply chooses her preferred quality, $q_1 \in \{0, 1, ...\}$, among the introduced VM classes, so her choice is

$$q_1(\pi, t, \theta) = \underset{i \text{ s.t. } s_i \leq t}{\arg\max}\ \theta \cdot s_i - x_i.$$

Note that $q_1 = 0$ encodes the customer opting out. We assume that customers who decline service in the current period are not available as customers in future periods. An *existing* customer will either stay with her previous choice or pay the switching cost to adopt a new technology introduced this period, so her choice is

$$q_2(\pi, t, \theta) = \begin{cases} \arg\max_{i \text{ s.t. } s_i \leq t} \theta \cdot s_i - x_i - \mathbb{1}_{s_i = t} c \\ \qquad\qquad \text{if } \arg\max_{i \text{ s.t. } s_i < t} \theta \cdot s_i - x_i \geq 0 \\ 0 \qquad\qquad \text{otherwise.} \end{cases}$$

There is an inherent asymmetry between the first introduction period and subsequent periods: when

[2]In Section 7, we discuss how this and other model assumptions affect our results.

[3]This is consistent with the wide belief that cloud service providers are not capacity constrained in this stage, but rather going through a phase of infrastructure investment aiming to increase their market share (Kilcioglu and Maglaras, 2015).

$t = s_1$, only new customers can choose the new technology and necessarily $q_2(\pi, s_1, \theta) = 0$. We assume without loss of generality that in case of ties in the definitions of q_1 and q_2, the customer chooses the latest VM class. In particular, we assume that a new customer who is indifferent between opting out and buying will buy.

The (expected) revenue of policy π at time t is

$$\text{Revenue}(\pi, t) = \int (x_{q_1(\pi,t,\theta)} + x_{q_2(\pi,t,\theta)}) f(\theta) d\theta.$$

The provider discounts future utility at a rate of δ per period, so the total revenue of a policy π is

$$\text{Revenue}(\pi) = \sum_{t=1}^{\infty} \delta^t \text{Revenue}(\pi, t).$$

The cost of policy π at time t is

$$\text{Cost}(\pi, t) = C\mathbb{1}_{t\in\pi},$$

where we write $t \in \pi$ as shorthand for t being an introduction time in policy π, i.e., for the existence of some $s_j \in \pi$ such that $s_j = t$. The total cost of policy π is

$$\text{Cost}(\pi) = \sum_{t=1}^{\infty} \delta^t Cost(\pi, t) = C \sum_{j\geq 1} \delta^{s_j}.$$

Last, we define the utility of a policy π at time t to be the net gain,

$$U(\pi, t) = \text{Revenue}(\pi, t) - \text{Cost}(\pi, t)$$

with total net gain $U(\pi) = \text{Revenue}(\pi) - \text{Cost}(\pi)$.

4 Myerson Pricing

In this section we analyze a simple, natural pricing policy: simply compute the optimal price for each class of VM as if it were the only item offered for sale, as per Myerson's approach. We show this has several nice properties. First, with this pricing all newly arriving customers will select the latest quality. Second, with Myerson pricing the optimal policy has a periodic pattern of introductions.

Third, if a periodic schedule is used, then Myerson pricing is optimal in the limit, in the sense that Myerson pricing gets arbitrarily close to the optimal policy after sufficient introductions.

Consider the introduction s_j, with Myerson pricing, so that $x_j = s_j p^*$, where p^* maximizes expected revenue $\mathbb{P}(\theta \geq p) \cdot p$, and therefore satisfies $p^* = \frac{1-F(p^*)}{f(p^*)}$. *New* customers prefer buying technology i to buying nothing if $s_i(\theta - p^*) \geq 0$, or $\theta \geq p^*$, so the set of customers willing to buy each technology is the same, and it is exactly the set the optimal mechanism wants to sell to. Customers prefer technology j to $j-1$ if $s_j(\theta - p^*) \geq s_{j-1}(\theta - p^*)$, or $\theta \geq p^*$. Thus all new customers choose the latest technology and Myerson pricing is optimal for new customers. *Existing* customers prefer to switch to the new technology at a time $t = s_j$ if $s_j(\theta - p^*) - c \geq s_{j-1}(\theta - p^*)$, or $\theta \geq p^* + c/(s_j - s_{j-1})$.

Assuming a policy π_M which follows Myerson pricing, we write down the revenue for the provider at time $t = s_j$:

$$\begin{aligned} \text{Revenue}(\pi_M, s_j) = (1 - F(p^*))\, s_j p^* + \left(1 - F\left(p^* + \frac{c}{s_j - s_{j-1}}\right)\right) s_j p^* \\ + \left(F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F(p^*)\right) s_{j-1} p^* \end{aligned} \tag{1}$$

The first summand is the revenue from the customers who arrive at period s_j; the second summand is the revenue from the customers who arrive at period $s_j - 1$ and switch to the new VM class introduced at time s_j; and the third summand is the revenue from the customers who arrive at period $s_j - 1$ and do not switch to the new VM class at time s_j. (In the special case of $j = 1$, we have only the first summand.)

We also write down the revenue for the provider at time $t > 0, s_j < t < s_{j+1}$, at which no VM class is introduced:

$$\text{Revenue}(\pi_M, t) = 2\,(1 - F(p^*))\, s_j p^* \tag{2}$$

4.1 Under Myerson pricing, periodic introductions are optimal

Our first result shows that there exists a policy that is optimal within the class of policies which use Myerson pricing, which uses periodic introductions. As previously discussed, there is an asymmetry with the first introduction because there are no existing customers, but after that the optimization problem is invariant to being shifted by one introduction. Thus the proof inductively constructs a periodic optimal policy from an arbitrary optimal policy.

Proposition 1. *Assuming Myerson pricing, periodic introductions are optimal after the first introduction.*

All proofs are deferred to the Appendix.

We remark that the optimal policy could be to not offer a service ($\pi_M = ((0,0))$. However, a sufficient condition to prefer to first introduce the service at a time $t = s$ is $\delta^s C <$ Revenue$((0,0),(s, sp^*))$, which simplifies as

$$\begin{aligned} C &< \frac{1}{\delta^s}\left(\delta^s(1-F(p^*))sp^* + \sum_{t=s+1}^{\infty}\delta^t 2(1-F(p^*))sp^*\right) \\ &= \left(1 + 2\delta\sum_{t=0}^{\infty}\delta^t\right)(1-F(p^*))sp^* \\ &= \frac{1+\delta}{1-\delta}(1-F(p^*))sp^*. \end{aligned}$$

Hence for finite C and $\delta < 1$ it is always optimal to introduce at some time s. The optimal policy depends upon C and δ (as well as c). However, the dependence upon C, ceteris paribus, essentially constrains the periodicity and the time of the first introduction. Henceforth without loss of generality we assume that the provider is only interested in maximizing revenue.

4.2 Under periodic introductions, Myerson pricing is optimal in the limit of many introductions

We now show that, under periodic introductions, the potential additional revenue of any pricing policy over Myerson pricing decays to zero. In particular, shading down prices doesn't gain much additional revenue. Informally, we have an incentive to increase the first introduction price, sacrificing short-term revenue; but shade down subsequent prices, giving extra incentive for existing customers to switch. However, the latter effect diminishes with time, as we now prove formally.

Proposition 2. *Let π be a policy with periodic introductions and π_M be a policy that uses the same* *introductions as π, but Myerson pricing. Then we have*

$$\lim_{j\to\infty} \frac{Revenue(\pi, s_j)}{Revenue(\pi_M, s_j)} \leq 1.$$

The key insight is that the potential gains from alternate prices can be bounded in terms of the length of the periodicity used by π, independent of introduction index j. As a result, assuming periodic introductions, the potential additional revenue of any policy over the Myerson policy decays to zero as the introduction time increases. In particular, the potential additional revenue earned by shading prices down from the Myerson prices decays to zero as the introduction time increases. Among the class of policies, in other words, which use multiples of a fixed base rate (i.e., they charge $x_j = (1-h)s_j p^*$, with $0 \leq h < 1$) and introduce periodically, Myerson pricing gets arbitrarily close to the optimal policy, after sufficient introductions.

5 General Pricing

In this section we argue that in many cases Myerson pricing is near optimal even with arbitrary patterns of introduction times. In view of Proposition 1, this shows that Myerson pricing combined with periodic introductions is near optimal.

To begin, we consider the set of all policies with a particular pattern of introductions.

Definition 1. For a set of introduction times $\mathbf{s} = (s_0, s_1, s_2, \ldots)$, the set of all policies with these introduction times is denoted $\Pi(\mathbf{s}) = \{\pi' = ((s'_0, x'_0), (s'_1, x'_1), (s'_2, x'_2), \ldots) \mid s_i = s'_i \ \forall i\}$.

Recall that the revenue of a policy π at time t is

$$\text{Revenue}(\pi, t) = \int (x_{q_1(\pi,t,\theta)} + x_{q_2(\pi,t,\theta)}) f(\theta) d\theta.$$

Optimizing this in general is difficult because the price set for the jth introduction, x_j, affects periods after s_j. Instead, we use the following trivial upper bound, for which the revenue in each time period is optimized separately.

Observation 1. *Let introduction times* $\mathbf{s}$ *be given. Then we have*

$$\max_{\pi \in \Pi(\mathbf{s})} \sum_t Revenue(\pi, t) \leq \sum_t \max_{\pi \in \Pi(\mathbf{s})} Revenue(\pi, t).$$

The proposed upper bound calculates the revenue in the case where the provider were allowed to pick new prices, at each time t, for introductions that have happened already. Clearly the optimal revenue at time t in this case is an upper bound of the real revenue at time t under the optimal pricing policy.

Note that, as we saw in deriving Equation (2), the optimal policy to select for non-introduction periods simply uses Myerson prices. This just leaves the periods when a new VM class is introduced. For these periods, we know that Myerson prices are the optimal setting of prices x_{q_1}, but not necessarily of prices x_{q_2}. The following lemma captures the relevant facts about the optimization over prices x_{q_2}. In particular, it is optimal to set all prior introduction prices to the Myerson price. Furthermore, we can put a lower and an upper bound on the optimal price for the current introduction.

Before we state the lemma, we state a "niceness" assumption on F, in particular Myerson's regularity condition of monotone hazard rate, which our analysis requires.

Assumption 1. The function $\frac{1-F(p)}{f(p)}$ is monotonically decreasing.

Assumption 1 is common in the literature and satisfied by a number of common distributions, including uniform, normal, exponential, beta (with shape parameters $a \geq 1, b \geq 1$), and gamma (with shape parameter $k \geq 1$).

Lemma 1. *Assume F satisfies Assumption 1 and let introduction times* $\mathbf{s}$ *and time t be given with $s_j = t$ for some introduction j. There exists a policy $\pi' \in \Pi(\mathbf{s})$ that maximizes $\int x_{q_2(\pi,t,\theta)} f(\theta) d\theta$ among all policies $\pi \in \Pi(\mathbf{s})$, and that uses pricing $x_i = s_i p^*$ for $i < j$ and* $\max(s_{j-1}p^*, s_j p^* - c) \leq x_j \leq s_j p^*$.

We use the lemma to characterize the competitive ratio of Myerson pricing, i.e., the ratio between the optimal revenue and the Myerson revenue. As previously mentioned, we first provide an upper bound of the optimal revenue by allowing a separate optimal policy to be chosen for each time period. In turn, we now upper bound this upper bound by expanding the set of policies to allow separate policies to be applied to new and existing customers. Such a discriminatory strategy would separate new and existing customers at introduction times, offering a discount to existing customers as an incentive to upgrade. The (expected) revenue of such a discriminatory strategy employing policy π_1 for newly arriving customers and policy π_2 for existing customers at time t is

$$\text{RevenueD}(\pi_1, \pi_2, t) = \int \left(x_{q_1(\pi_1,t,\theta)} + x_{q_2(\pi_2,t,\theta)} \right) f(\theta) d\theta.$$

Having fixed the introduction times, and per the characterization in Lemma 1, the optimal choices of π_1, π_2 for the revenue of a single introduction period have a particular restricted form: they charge Myerson prices for previous introductions and a price from within a restricted range for the current introduction. We provide the following expressions for the revenue at a period when a new VM class is introduced under this class of policies for both the original and the discriminatory setting.

Lemma 2. *Let introduction times* $\mathbf{s}$ *and introduction number* j *be given. Consider policy* $\pi \in \Pi(\mathbf{s})$ *that uses prices* $x_i = s_i p^*$ *for* $i < j$ *and* $x_j = x$*, with* $\max(s_{j-1}p^*, s_j p^* - c) \le x \le s_j p^*$*. The revenue for the provider of policy* π *at time* s_j*, which we denote* $Rev_j(x)$*, is*

$$Rev_j(x) = (1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*) + \left(1 - F\left(\frac{x}{s_j}\right)\right) x. \quad (3)$$

Consider also a discriminatory strategy that uses policies $\pi_1, \pi_2 \in \Pi(\mathbf{s})$ *with prices* $x_i = s_i p^*$ *for* $i < j$*,* $x_j = x_n$ *for new customers in period* s_j*, and* $x_j = x_e$ *for existing customers in period* s_j*, with* $\max(s_{j-1}p^*, s_j p^* - c) \le x_n, x_e \le s_j p^*$*. The revenue for the provider of policies* π_1, π_2 *at time* s_j*, which we denote* $RevD_j(x_n, x_e)$*, is*

$$RevD_j(x_n, x_e) = (1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x_e - s_{j-1}p^*) + \left(1 - F\left(\frac{x_n}{s_j}\right)\right) x_n. \quad (4)$$

We provide an upper bound to the competitive ratio of Myerson pricing for policies of this form. We note that the revenue under Myerson pricing at introduction period s_j is given by Equations (1) and (3), which are equivalent: $\text{Revenue}(\pi_M, s_j) = Rev_j(s_j p^*)$.

Proposition 3. *Assume F satisfies Assumption 1 and let x^* be the value that maximizes $Rev_j(x)$ and $x_n^* = s_j p^*, x_e^*$ be the values that maximize $RevD_j(x_n, x_e)$. It holds that*

$$\frac{Rev_j(x^*)}{Rev_j(s_j p^*)} \leq \frac{RevD_j(x_n^*, x_e^*)}{Rev_j(s_j p^*)} \leq 1 + \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}. \tag{5}$$

The right-hand side is reduced to the trivial bound of 2 by observing that $F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right) \leq 1 - F(p^*)$. Note also that the right hand side is close to 1 for large or small c.

We have argued that the optimal policy for maximizing revenue of a particular introduction period from new customers or existing customers in isolation will be of the following restricted form: charge Myerson prices for previous introductions and a price from within a restricted range for the current introduction. Therefore, the optimal policy for the discriminatory setting will also be of the same form. Our final result shows that the optimal policy for maximizing combined revenue of a particular introduction period from new and existing customers, for the original setting, will also be of the same form. This follows from a quasiconcavity property implied by Assumption 1.

Theorem 1. *Assume F satisfies Assumption 1 and let introduction times $\mathbf{s}$ and time t be given with $s_j = t$ for some introduction j. There exists a policy $\pi' \in \Pi(\mathbf{s})$ that maximizes Revenue(π, t) among all policies $\pi \in \Pi(\mathbf{s})$, and that uses pricing $x_i = s_i p^*$ for $i < j$ and $\max(s_{j-1}p^*, s_j p^* - c) \leq x_j \leq s_j p^*$. Furthermore, the price x_j of this policy π' can be determined as the maximizer x^* of $Rev_j(x)$.*

Theorem 1 allows us to understand how the optimization of revenue from existing customers, $\int x_{q_2} f(\theta) d\theta$, affects the optimization of revenue from new customers, $\int x_{q_1} f(\theta) d\theta$. In particular, we know from Section 4 that Myerson prices optimize revenue from new customers, and at those prices all arriving customers select the newest VM class. Optimizing revenue from existing, as opposed to new, customers only affects the price of the newest VM class, by reducing it from the Myerson benchmark. Thus, as long as the revenue from new customers and the revenue from existing customers behave well as functions of prices x_{q_1} and x_{q_2}, which our assumption ensures, it is still the case that under the optimal policy for revenue at a particular time, all new customers choose the latest VM class; they may just pay a price for it lower than the Myerson benchmark.

6 Numerical Illustrations

We now illustrate our results with a variety of distributions for customer type θ. We show that our bounds for Myerson pricing from Section 5 can provide strong guarantees for natural families of distributions. We also show numerically that in reality Myerson pricing is often some orders of magnitude closer to optimal than our bounds suggest.

We first examine our analytical bounds for Myerson pricing from Proposition 3 and show that they are tight not only for small and large values of the switching cost c, but also for intermediate ones, for a variety of distributions for customer type θ.

We then run simulations and, for fixed introduction times, calculate the gain ratio of the optimal total revenue (estimated by a best-response updating algorithm we propose in Section 6.2), throughout the horizon, over Myerson total revenue:

$$\frac{\text{optimal revenue} - \text{Revenue}(\pi_M)}{\text{Revenue}(\pi_M)}.$$

We show that the gain ratio is small for a variety of distributions for customer type θ, thus showing that Myerson pricing is near optimal in many cases. We also look at the gain ratio of the optimal revenue for a single introduction period[4] proposed in Theorem 1 over the Myerson revenue in that introduction period, reporting the maximum gain ratio over all introductions,

$$\max_j \frac{Rev_j(x^*) - \text{Revenue}(\pi_M, s_j)}{\text{Revenue}(\pi_M, s_j)}.$$

Finally, we also run numerical experiments for the setting with discriminatory pricing, which separates new and existing customers at introduction times. Again, we show that Myerson pricing is close to optimal.

6.1 Examining our analytical bounds

Figure 1 shows the upper bound on the gain ratio $\frac{Rev_j(x^*)-Rev_j(s_jp^*)}{Rev_j(s_jp^*)}$ from Proposition 3, which is $\frac{F\left(p^*+\frac{c}{s_j-s_{j-1}}\right)-\max\left(F(p^*),F\left(\frac{c}{s_j-s_{j-1}}\right)\right)}{1-F(p^*)}$, against the scaled switching cost $\frac{c}{s_j-s_{j-1}}$. We show the upper

[4] Remember from Section 5 and in particular Observation 1 that the optimal revenue for a single introduction period is an upper bound of the real revenue in that period under the optimal policy, having fixed introduction times.

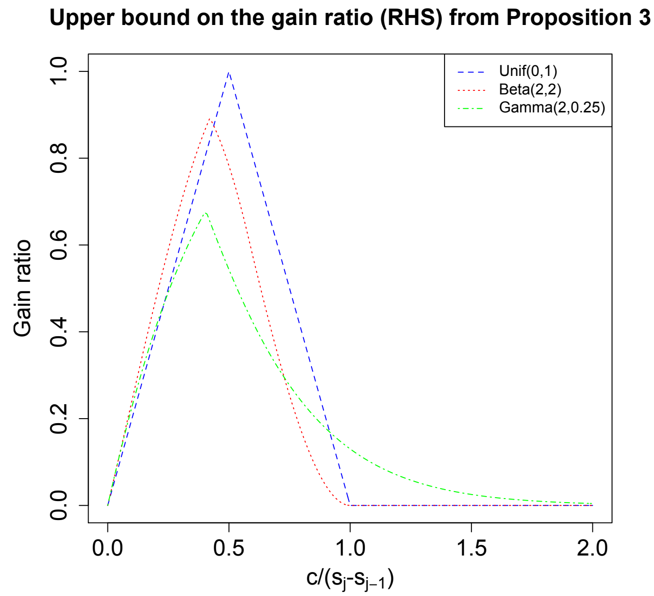

Figure 1: The right-hand side of Equation (5) in Proposition 3, reduced by 1, against the scaled switching cost $\frac{c}{s_j - s_{j-1}}$, for the uniform distribution on $[0, 1]$, the beta distribution with shape parameters $\alpha = \beta = 2$, and the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$.

bound for the uniform distribution on $[0, 1]$, the beta distribution (p.d.f. $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$) with shape parameters $\alpha = \beta = 2$, and the gamma distribution (p.d.f. $f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$) with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The figure illustrates that Myerson pricing is close to optimal for large or small scaled switching costs, while it raises the possibility that there is room for substantial improvement over the Myerson pricing for intermediate scaled switching costs.

We note that it is possible to derive tighter bounds for specific distributions. In the Appendix, we derive a tighter bound of 1/4 for the Uniform(0,1) case.

We also note that, rather than applying the general bound from Proposition 3, we can directly calculate the left-hand side of Equation (5) in Proposition 3, and subtract 1 to recover the gain ratio $\frac{Rev_j(x^*) - Rev_j(s_j p^*)}{Rev_j(s_j p^*)}$. However, this now depends on the three parameters c, s_j, and s_{j-1} separately. Figure 2 plots the gain ratio for selected pairs of s_j and s_{j-1}, for the uniform distribution on $[0, 1]$, the beta distribution with shape parameters $\alpha = \beta = 2$, and the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. As mentioned before, the optimal revenue for a single introduction period is an upper bound on the real revenue in that period under the optimal pricing policy, and therefore the gain ratio $\frac{Rev_j(x^*) - Rev_j(s_j p^*)}{Rev_j(s_j p^*)}$ is an upper bound to the gain ratio in

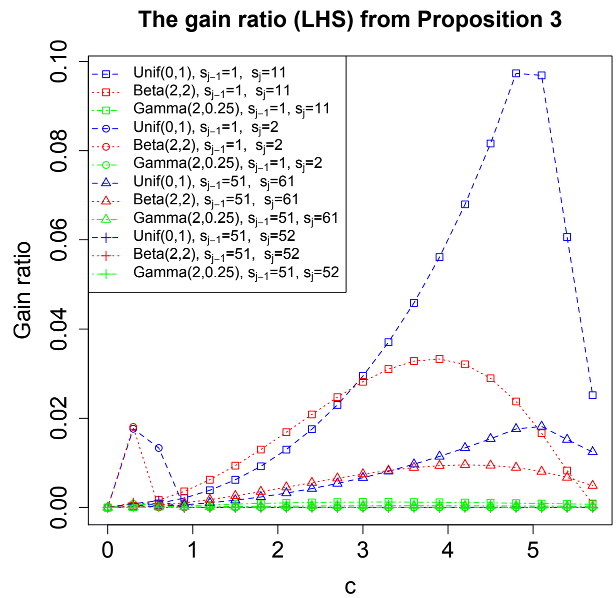

Figure 2: The left-hand side of Equation (5) in Proposition 3, reduced by 1, against the switching cost c, for selected pairs of introduction times s_{j-1}, s_j, for the uniform distribution on $[0,1]$, the beta distribution with shape parameters $\alpha = \beta = 2$, and the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$.

that period under the optimal pricing policy for the real problem. The gains over Myerson pricing are less than 10% — a bound substantially tighter than the one given by the right-hand side of Equation (5) in Proposition 3.

We can apply this approach to the exponential distribution to show that, for that distribution, Myerson pricing is in fact optimal. This can be verified by observing that $Rev_j'(s_j p^*) = 0$.

6.2 Numerical experiments

We now turn from analyzing a single introduction time in isolation to analyzing a full policy. To do so, we run 100 simulations for each combination of distribution f, switching cost c, and discount rate δ. In each simulation, a set of 50 introduction times is randomly generated, with introductions up to 15 periods apart.

We calculate the optimal pricing given the set of introduction times using the following best-response updating algorithm: initialize prices, then optimize the price of each introduction given the prices for the preceding and the subsequent introduction in the previous iteration, and proceed through all the introductions (looping back to the first introduction after the last introduction has

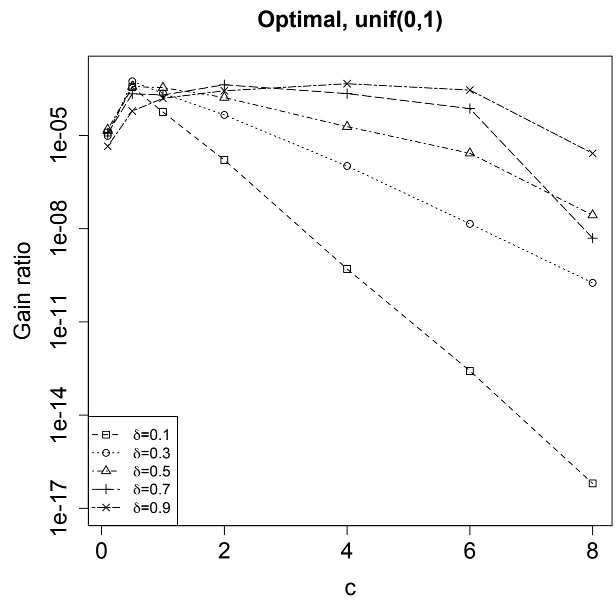

Figure 3: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c, for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the uniform distribution on $[0, 1]$.

been optimized). Stop when no introduction can have an improvement ratio above 10^{-30}. Because of the arguments in Rosen (1965), and as shown in the Appendix, there is a unique revenue maximizing price vector, and if our proposed updating algorithm converges (which it always did), it converges to the unique optimal pricing.

Figure 3 plots the average of the gain ratio of the optimal total revenue over Myerson total revenue,

$$\frac{\text{optimal revenue} - \text{Revenue}(\pi_M)}{\text{Revenue}(\pi_M)},$$

against the switching cost c, for different values of the discount rate δ, for the uniform distribution on $[0, 1]$.

Holding the switching cost c constant, if c is small, then the higher the discount rate δ, the smaller the gain ratio of the optimal total revenue over the Myerson total revenue. The near-optimality of Myerson pricing, that is, is more pronounced as the provider becomes more patient. If c is large, then the higher the δ, the larger the gain ratio of the optimal total revenue over the Myerson total revenue. Note that these are ratios and that the absolute gain is small for small δ.

Holding the discount rate δ fixed, for switching cost c close to zero, an existing customer is likely

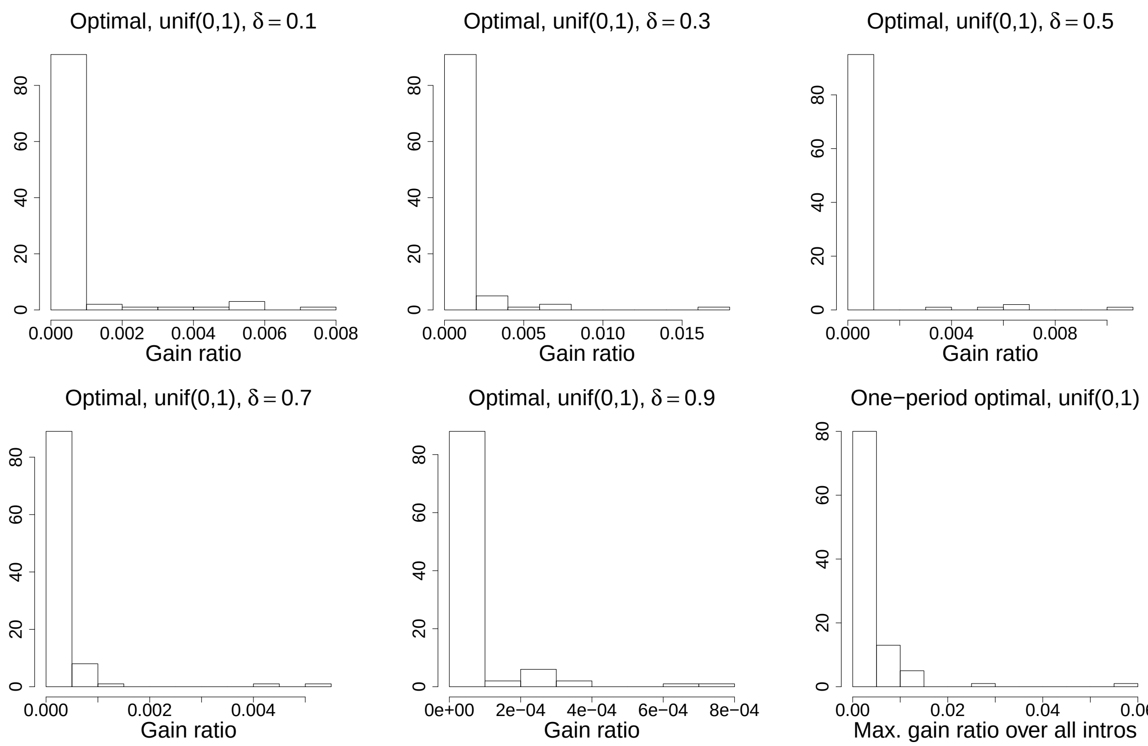

Figure 4: Histograms of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, for the uniform distribution on $[0, 1]$, for switching cost $c = 0.5$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$. Also, a histogram of the gain ratio of the optimal revenue for a single introduction period over Myerson revenue, for the introduction that attains the best gain ratio, over 100 simulations.

to behave as if she were a new customer, and chooses her preferred quality among the available ones, thus Myerson pricing is close to optimal. For large c, a customer is not likely to switch to a new technology, and again Myerson pricing is close to optimal. Therefore, the gain ratio of the optimal total revenue over the Myerson total revenue becomes smaller as the switching cost becomes very large or very small.

We also look at the full histogram of the total revenue gain ratio over 100 simulations for different values of c and δ, along with the histogram of the gain ratio of the optimal revenue for a single introduction period over the Myerson revenue in that period, for the introduction that attains the best gain ratio, $\max_j \frac{Rev_j(x^*) - \text{Revenue}(\pi_M, s_j)}{\text{Revenue}(\pi_M, s_j)}$, over 100 simulations. Figure 4 shows the effect of varying the discount rate δ in detail. The overall trend is consistent with the averages, so the main additional takeaway from the full histogram is that in most instances Myerson pricing is essentially optimal. The histogram of the gain ratio of the optimal revenue for a single introduction period shows that this relaxation is often loose by an order of magnitude or more. We note that

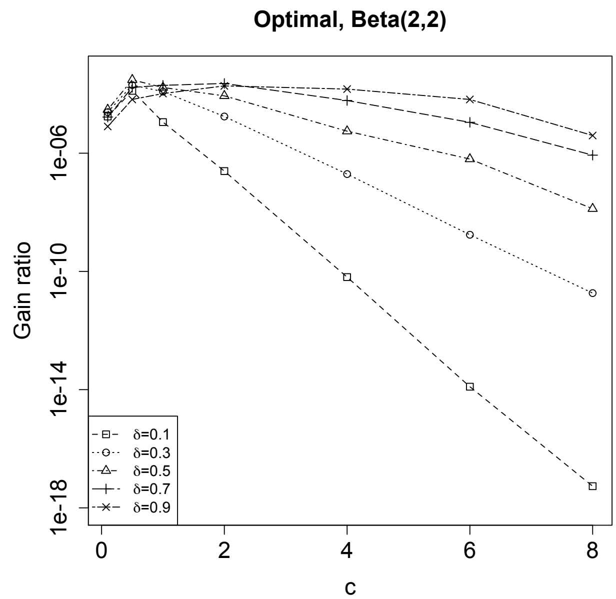

Figure 5: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c, for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the beta distribution with shape parameters $\alpha = \beta = 2$.

our worst case analytical bound of $1/4$ for the uniform distribution on $[0, 1]$ is loose even relative to this relaxation. Figure 12 in the Appendix shows again that in most instances Myerson pricing is essentially optimal, if instead we vary the switching cost c.

Figures 5 and 6 show experiments for the beta distribution and the gamma distribution. The results for the beta and gamma distributions are consistent with our results for the uniform distribution.

6.3 Numerical experiments for discriminatory pricing

We focus on the total revenue under a discriminatory strategy, which separates new and existing customers at introduction times, offering a discount to existing customers as an incentive to upgrade. In particular, at time s_j when VM class j is introduced, the provider offers VM class j at price $x_{j,n}$ to customers arriving at time s_j, and at price $x_{j,e}$ to customers already in the system. As argued in Section 5, the optimal revenue in this "discriminatory" setting is an upper bound of the optimal revenue in the original (real) setting. Furthermore, such strategies are potentially interesting in practice, although we are not aware of existing instances of their use in the cloud computing services market.

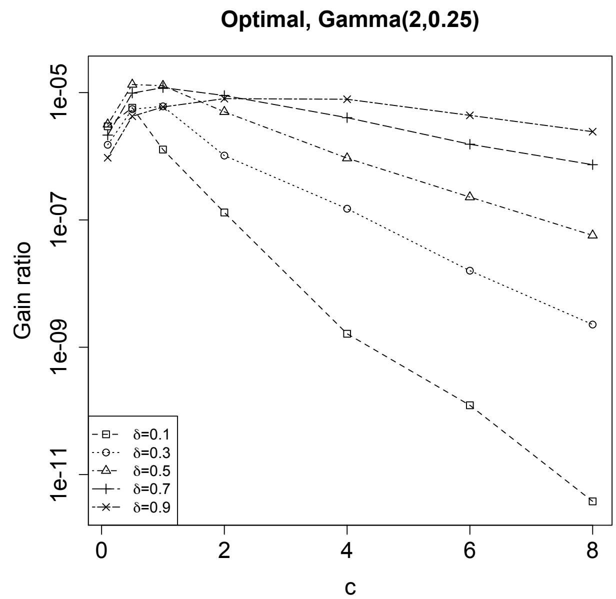

Figure 6: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c, for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$.

Before presenting numerical experiments for the setting with discriminatory pricing, we first look at the corresponding revenue optimization problem. Fix introduction times $s_0, s_1, \ldots$ and introduction index j. We write down all the terms of total revenue that include $x_{j,n}$ or $x_{j+1,e}$:

$$\begin{aligned}
&\delta^{s_j} \cdot \left(1 - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} \\
&+ 2 \cdot \frac{\delta^{s_j+1} - \delta^{s_{j+1}}}{1-\delta} \cdot \left(1 - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} \\
&+ \delta^{s_{j+1}} \cdot \left(1 - F\left(\max\left(\frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}, \frac{x_{j,n}}{s_j}\right)\right)\right) \cdot x_{j+1,e} \\
&+ \delta^{s_{j+1}} \cdot \left(F\left(\frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}\right) - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} \cdot \mathbb{1}_{\frac{x_{j,n}}{s_j} \le \frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}},
\end{aligned} \tag{6}$$

where the first term is the revenue accumulated in period s_j from new customers who arrive in period s_j and buy VM class j, the second term is the revenue accumulated during the periods in between introductions j and $j+1$, the third term is the revenue accumulated in period s_{j+1} from customers who arrive in period $s_{j+1} - 1$ and switch to VM class $j+1$ in period s_{j+1}, and the fourth term is the revenue accumulated in period s_{j+1} from customers who arrive in period $s_{j+1} - 1$ and do not switch to VM class $j+1$ in period s_{j+1}.

Assuming $\frac{x_{j,n}}{s_j} \leq \frac{x_{j+1,e}-x_{j,n}+c}{s_{j+1}-s_j}$, the last two terms of (6) can be jointly rewritten as

$$\delta^{s_{j+1}} \cdot \left[\left(1 - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} + \left(1 - F\left(\frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}\right)\right) \cdot (x_{j+1,e} - x_{j,n}) \right]. \tag{7}$$

Notice that the Myerson pricing $x_{j,n} = s_j p^*$ optimizes the first term and the second term of (6), as well as the first term of (7). So overall, an optimal setting for the total revenue in the discriminatory setting is to set $x_{j,n}$ to the Myerson pricing, and then optimize the term $\left(1 - F\left(\frac{x_{j+1,e}-x_{j,n}+c}{s_{j+1}-s_j}\right)\right) \cdot (x_{j+1,e} - x_{j,n})$ over $x_{j+1,e}$. This matches exactly the prices $x_{j,n}, x_{j+1,e}$ that would be set by optimizing $RevD_j$ and $RevD_{j+1}$ in (4). Therefore, optimizing the total revenue throughout the horizon in the discriminatory setting can be conveniently decomposed into optimizing the revenue $RevD_j$ per introduction period, for all introductions j.

Numerical experiments. Figure 7 plots the average of the gain ratio of the optimal total revenue over Myerson total revenue,

$$\frac{\text{optimal discriminatory revenue} - \text{Revenue}(\pi_M)}{\text{Revenue}(\pi_M)},$$

against the switching cost c, for different values of the discount rate δ, for the uniform distribution on $[0,1]$. The same patterns are observed as in the original setting, with the difference that the values of the gain ratio are now larger by some orders of magnitude.

We again also look at the full histogram of the total revenue gain ratio in the discriminatory setting over 100 simulations for different values of c and δ, along with the histogram of the gain ratio of the optimal revenue for a single introduction period over the Myerson revenue in that period, for the introduction that attains the best gain ratio, $\max_j \frac{RevD_j(s_j p^*, x_e^*) - \text{Revenue}(\pi_M, s_j)}{\text{Revenue}(\pi_M, s_j)}$, in the discriminatory setting, over 100 simulations. Figure 8 shows the effect of varying the discount rate δ in detail. Similarly to the original setting, the overall trend is consistent with the averages, so the main additional takeaway from the full histogram is that in most instances Myerson pricing is essentially optimal. Notice that the values of the gain ratio as indicated in the histograms are larger in the discriminatory setting than in the original setting. Figure 13 in the Appendix shows again that in most instances Myerson pricing is essentially optimal, if instead we vary the switching

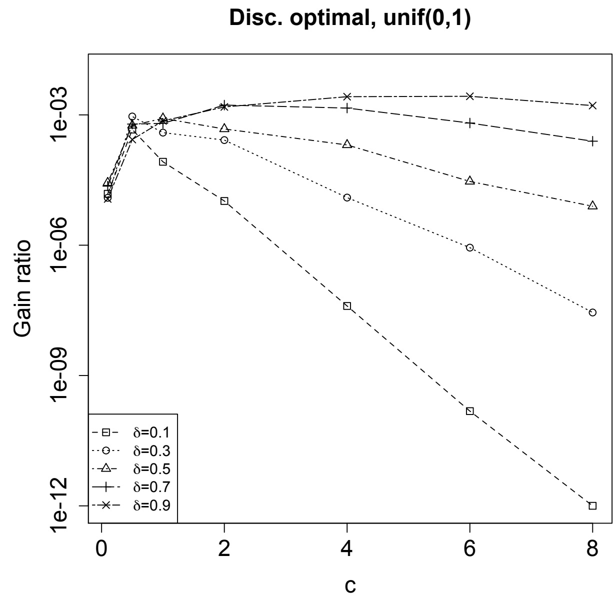

Figure 7: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c, for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the uniform distribution on $[0, 1]$.

cost c.

Figures 9 and 10 show experiments for the beta distribution and the gamma distribution in the discriminatory setting. The results are consistent with our results for the uniform distribution, and the values of the gain ratio in the discriminatory setting are again larger than the values of the gain ratio in the original setting, but still small in absolute terms.

7 Discussion

As technology improves over time, cloud service providers have the ability to offer more powerful VMs, which are more valuable to customers. At the same time, introduction of a new VM class comes at a cost for development and launching, and the new VM class competes with existing classes. We have presented a model of new product introductions for the cloud services market that addresses this trade-off, in the face of customers who are averse to upgrading to improved offerings. The decision problem for the cloud service provider is when to introduce a new VM class and how to price it in order to maximize total revenue, taking into account (discounted) future rewards.

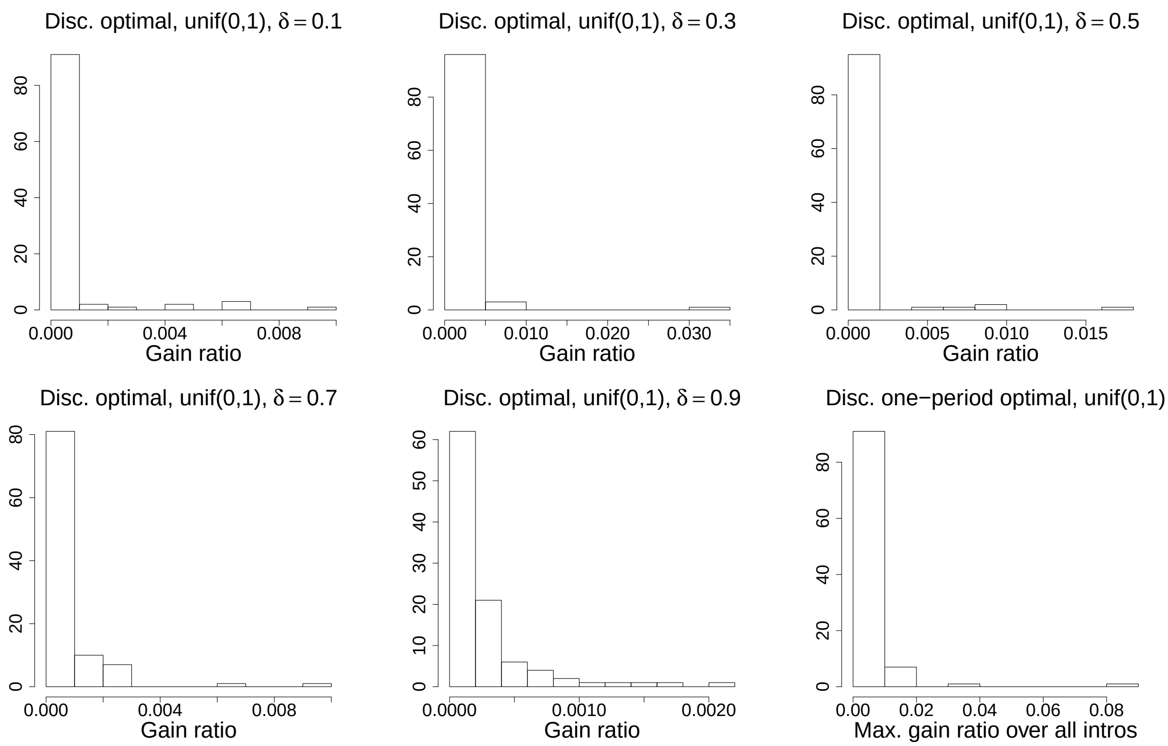

Figure 8: Histograms of the gain ratio of optimal total revenue over Myerson total revenue in the discriminatory setting, over 100 simulations, for the uniform distribution on $[0, 1]$, for switching cost $c = 0.5$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$. Also, a histogram of the gain ratio of the optimal revenue for a single introduction period over Myerson revenue, for the introduction that attains the best gain ratio, in the discriminatory setting, over 100 simulations.

In our analysis, we have shown that a surprisingly simple policy is approximately optimal: new VM classes are introduced on a periodic schedule, and each is priced as if it were the only product being offered (Myerson pricing). We first show that under a Myerson pricing rule, there is no loss of optimality with a periodic schedule of introductions. We also show that under periodic introductions, the potential additional revenue of any pricing policy over Myerson pricing decays to zero after sufficiently many introductions.

We then show that, given arbitrary fixed introduction times, Myerson pricing is approximately optimal. We characterize the prices that achieve optimal revenue in a single introduction period, and provide a bound for the competitive ratio of Myerson pricing over the optimal single-period pricing. This bound shows that Myerson pricing is approximately optimal when switching costs are small or large. Overall, combined with our first result, this implies that Myerson pricing with periodic introductions is approximately optimal.

Following our analysis, we examine our analytical bounds for Myerson pricing and show they

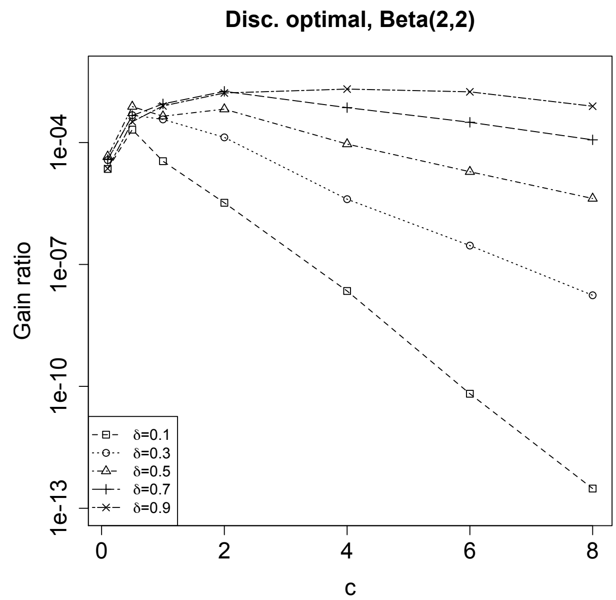

Figure 9: The average of the gain ratio of optimal total revenue over Myerson total revenue in the discriminatory setting, over 100 simulations, against the switching cost c, for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the beta distribution with shape parameters $\alpha = \beta = 2$.

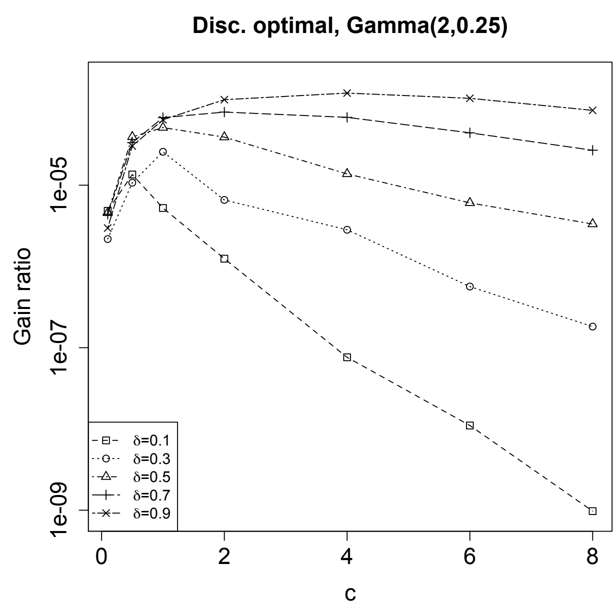

Figure 10: The average of the gain ratio of optimal total revenue over Myerson total revenue in the discriminatory setting, over 100 simulations, against the switching cost c, for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$.

can provide strong guarantees for all values of the switching cost, for several natural distributions. Furthermore, we run simulations where we numerically compute optimal prices, rather than using our bounds, and we find that Myerson pricing is often several orders of magnitude closer to optimal than our analytical bounds suggest.

We conclude by discussing our various model assumptions and their importance to our results. Our assumption that VM classes remain available once introduced is not crucial, given that our results for both Myerson pricing and our upper bound show that all new customers choose the latest VM class anyways. Furthermore, our upper bounds apply even in the case of discriminatory pricing, in which setting this assumption is without loss of generality if we can set arbitrary prices for new customers: we can always set prices sufficiently high for old VM classes that they will never be chosen.

Our results extend to the case where the number of customers arriving at each period is stochastic rather than a unit mass, as long as the expected arrival rate is constant through time and the policy is decided a priori rather than adaptively based on the state of the system. Time-varying expected arrival rates will affect the relative weights between current and future customers and the periodicity, but they will not affect the general shape of our results on the near-optimality of Myerson pricing. Similarly, we can extend our results to the case where customers stay in the system for a lifespan that is longer than two periods, and one that may be stochastic.

Our assumption of linear growth of the quality of the offered services with time is important only for proving that a policy with periodic introductions is optimal under Myerson pricing. Otherwise, the optimal policy is periodic in terms of the magnitude of the improvement, rather than time.

We restrict to myopic customers, as opposed to forward-looking customers. We now argue that this is a mild assumption given the structure of the optimal policies. What might a forward-looking customer do? Whichever VM class is myopically optimal for a customer in a period, will be again in the next period, assuming no new introduction in the next period. If we allow for a lifespan that is longer than two periods, then a customer may choose not to upgrade when a new VM class is introduced, but upgrade on the next introduction. However, with a lifespan of two periods, this is not possible. Note also that in our model we do not allow for patient customers, who are willing to wait (and stay in the system) before buying in the next period. Therefore, the only thing that changes with forward-looking customers in our setting is that it is possible that they will buy a VM of a new VM class with negative utility in their first period, if an introduction in the next period generates sufficient positive utility. This is a small effect and also not well aligned with common intuition about what customers actually do in practice.

Finally, one feature of our model that does not match reality is that in our model prices go up over time, whereas in the cloud computing market prices are going down over time. One cause of this is competition, which our model excludes. Another is that we have assumed that there is a cost for developing and introducing new products, but no change in that cost over time. In practice the cost of providing VMs is going down over time due to improvements in hardware and software that manages it, as well as due to economies of scale. We leave modelling these issues to future work.

Acknowledgments

The authors thank Victor Araman, Ilan Lobel and Jurgen Mihm for useful discussions.

References

Abhishek, Vineet, Ian A. Kash, and Peter Key (2012), "Fixed and Market Pricing for Cloud Services." *ArXiv e-prints*.

Anselmi, Jonatha, Danilo Ardagna, John C. S. Lui, Adam Wierman, Yunjian Xu, and Zichao Yang (2014), "The economics of the cloud: Price competition and congestion." *SIGecom Exch.*, 13, 58–63, URL `http://doi.acm.org/10.1145/2692375.2692380`.

Araman, Victor F. and Rene Caldentey (2016), "Crowdvoting the timing of new product introduction." Available at SSRN: `https://ssrn.com/abstract=2723515`.

Bagnoli, Mark and Ted Bergstrom (2005), "Log-concave probability and its applications." *Economic Theory*, 26, 445–469, URL `http://dx.doi.org/10.1007/s00199-004-0514-4`.

Balcer, Yves and Steven A Lippman (1984), "Technological expectations and adoption of improved technology." *Journal of Economic Theory*, 34, 292 – 318, URL `//www.sciencedirect.com/science/article/pii/0022053184901467`.

Barr, Jeff (2015), "EC2 Instance History." `https://aws.amazon.com/blogs/aws/ec2-instance-history/`. Accessed: 2018-02-14.

Bass, Frank M. (1969), "A new product growth for model consumer durables." *Management Science*, 15, 215–227, URL `http://dx.doi.org/10.1287/mnsc.15.5.215`.

Bayus, Barry L. (1992), "The dynamic pricing of next generation consumer durables." *Marketing Science*, 11, 251–265, URL `http://dx.doi.org/10.1287/mksc.11.3.251`.

Besbes, Omar and Ilan Lobel (2015), "Intertemporal price discrimination: Structure and computation of optimal policies." *Management Science*, 61, 92–110, URL `https://doi.org/10.1287/mnsc.2014.2049`.

Borgs, Christian, Ozan Candogan, Jennifer Chayes, Ilan Lobel, and Hamid Nazerzadeh (2014), "Optimal multiperiod pricing with service guarantees." *Management Science*, 60, 1792–1811, URL `http://dx.doi.org/10.1287/mnsc.2013.1839`.

Caldentey, René, Ying Liu, and Ilan Lobel (2017), "Intertemporal pricing under minimax regret." *Operations Research*, 65, 104–129, URL `https://doi.org/10.1287/opre.2016.1548`.

Casadesus-Masanell, Ramon and David B. Yoffie (2007), "Wintel: Cooperation and conflict." *Management Science*, 53, 584–598, URL `https://doi.org/10.1287/mnsc.1060.0672`.

Key: Casadesus-Masanell2007

Annotation: doi: 10.1287/mnsc.1060.0672

Chen, Yiwei and Vivek F. Farias (2015), "Robust dynamic pricing with strategic customers." In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, 777–777, ACM, New York, NY, USA, URL `http://doi.acm.org/10.1145/2764468.2764530`.

Chen, Yiwei, Vivek F. Farias, and Nikolaos K. Trichakis (2017), "On the efficacy of static prices for revenue management in the face of strategic customers." *Working Paper*.

Chen, Yiwei and Ming Hu (2017), "Pricing and matching with forward-looking buyers and sellers." *Working Paper*.

Cohen, Morris A., Jehoshua Eliasberg, and Teck-Hua Ho (1996), "New product development: The performance and time-to-market tradeoff." *Management Science*, 42, 173–186, URL `http://dx.doi.org/10.1287/mnsc.42.2.173`.

Farzin, Y.H., K.J.M. Huisman, and P.M. Kort (1998), "Optimal timing of technology adoption." *Journal of Economic Dynamics and Control*, 22, 779 – 799, URL http://www.sciencedirect.com/science/article/pii/S0165188997000973.

Gallego, Guillermo and Garrett van Ryzin (1994), "Optimal dynamic pricing of inventories with stochastic demand over finite horizons." *Management Science*, 40, 999–1020, URL https://doi.org/10.1287/mnsc.40.8.999.

Gao, Jiayang, Krishnamurthy Iyer, and Huseyin Topaloglu (2017), "When fixed price meets priority auctions: Competing firms with different pricing and service rules." Available at SSRN: https://ssrn.com/abstract=2778810.

Goettler, Ronald L. and Brett R. Gordon (2011), "Does AMD Spur Intel to Innovate More?" *Journal of Political Economy*, 119, 1141 – 1200, URL https://ideas.repec.org/a/ucp/jpolec/doi10.1086-664615.html.

Gowrisankaran, Gautam and Marc Rysman (2012), "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy*, 120, 1173 – 1219, URL https://ideas.repec.org/a/ucp/jpolec/doi10.1086-669540.html.

Kash, Ian A. and Peter B. Key (2016), "Pricing the cloud." *IEEE Internet Computing*, 20, 36–43.

Kilcioglu, Cinar and Costis Maglaras (2015), "Revenue maximization for cloud computing services." *SIGMETRICS Perform. Eval. Rev.*, 43, 76–76, URL http://doi.acm.org/10.1145/2847220.2847245.

Kilcioglu, Cinar and Justin M. Rao (2016), "Competition on price and quality in cloud computing." In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, 1123–1132, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, URL http://dx.doi.org/10.1145/2872427.2883043.

Klastorin, Ted and Weiyu Tsai (2004), "New product introduction: Timing, design, and pricing." *Manufacturing & Service Operations Management*, 6, 302–320, URL http://dx.doi.org/10.1287/msom.1040.0050.

Krankel, R. Mark, Izak Duenyas, and Roman Kapuscinski (2006), "Timing successive product introductions with demand diffusion and stochastic technology improvement." *Manufacturing & Service Operations Management*, 8, 119–135, URL `http://dx.doi.org/10.1287/msom.1060.0102`.

Kumar, Sunil and Jayashankar M. Swaminathan (2003), "Diffusion of innovations under supply constraints." *Operations Research*, 51, 866–879, URL `https://doi.org/10.1287/opre.51.6.866.24918`.

Liu, Yan and William L. Cooper (2015), "Optimal dynamic pricing with patient customers." *Operations Research*, 63, 1307–1319, URL `https://doi.org/10.1287/opre.2015.1438`.

Lobel, Ilan (2017), "Dynamic pricing with heterogeneous patience levels." *Working Paper*.

Lobel, Ilan, Jigar Patel, Gustavo Vulcano, and Jiawei Zhang (2016), "Optimizing product launches in the presence of strategic consumers." *Management Science*, 62, 1778–1799.

Mahajan, Vijay and Eitan Muller (1996), "Timing, diffusion, and substitution of successive generations of technological innovations: The {IBM} mainframe case." *Technological Forecasting and Social Change*, 51, 109 – 132, URL `http://www.sciencedirect.com/science/article/pii/0040162595002251`.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1990), "New product diffusion models in marketing: A review and directions for research." *Journal of Marketing*, 54, 1–26, URL `http://www.jstor.org/stable/1252170`.

Myerson, Roger B. (1981), "Optimal auction design." *Math. Oper. Res.*, 6, 58–73.

Nair, J., V. G. Subramanian, and A. Wierman (2014), "On competitive provisioning of ad-supported cloud services." In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 137–139.

Norton, John A. and Frank M. Bass (1987), "A diffusion theory model of adoption and substitution for successive generations of high-technology products." *Management Science*, 33, 1069–1086, URL `http://dx.doi.org/10.1287/mnsc.33.9.1069`.

Pae, Jae H. and Donald R. Lehmann (2003), "Multigeneration innovation diffusion: The impact of intergeneration time." *Journal of the Academy of Marketing Science*, 31, 36, URL `http://dx.doi.org/10.1177/0092070302238601`.

Paulson Gjerde, Kathy A., Susan A. Slotnick, and Matthew J. Sobel (2002), "New product innovation with multiple features and technology constraints." *Manage. Sci.*, 48, 1268–1284, URL `http://dx.doi.org/10.1287/mnsc.48.10.1268.270`.

Plambeck, Erica and Qiong Wang (2009), "Effects of e-waste regulation on new product introduction." *Management Science*, 55, 333–347, URL `http://dx.doi.org/10.1287/mnsc.1080.0970`.

Rosen, J. B. (1965), "Existence and uniqueness of equilibrium points for concave n-person games." *Econometrica*, 33, 520–534.

Stremersch, Stefan, Eitan Muller, and Renana Peres (2010), "Does new product growth accelerate across technology generations?" *Marketing Letters*, 21, 103–120, URL `http://dx.doi.org/10.1007/s11002-009-9095-0`.

Wilson, Lynn O. and John A. Norton (1989), "Optimal entry timing for a product line extension." *Marketing Science*, 8, 1–17, URL `http://dx.doi.org/10.1287/mksc.8.1.1`.

A Analysis of VM Series Launches

Every VM used on Microsoft Azure is part of a series (such as “Av2” or “NC”) which describes the features (such as relative amount of memory or availability of a graphics card) that are associated with it. Within each series, there are typically multiple different sizes of VM. We analyze the launch of new series by Azure, ignoring size distinctions since all sizes are launched simultaneously. Recall that our model assumes that customers are averse to switching away from a series of VM they are already using when a new series launches. Here we provide evidence for this modelling assumption based on a dataset consisting of a snapshot of all active VMs on Azure at a particular point in time.

We note that while our dataset shows the set of currently running VMs, we lack the larger context in which a given VM is being used. For example, a customer may be using multiple VMs to run a service, and this service may automatically launch and terminate VMs over time. Or a customer may have built a piece of software that launches a VM when run and terminates it when the task is complete. So even a VM that was recently created may be a part of some long-standing system. The switching cost in our model captures the cost of changing this underlying system, so what we would really like to analyze is the date this system was created. Of course, that date is not available.

Each VM running on Azure is associated with an account known as a subscription. As a proxy for the creation date of the system, we use the creation month of the subscription. This is an imperfect proxy for a number of reasons: a subscription could be repurposed or used for multiple systems created at different times; a new subscription might be created for an existing system for administrative reasons; a single system can span multiple subscriptions. Nevertheless, it is reasonable to assume that creation time of the system and creation time of the subscription are correlated. We show that subscriptions created before a VM series launches have less of a tendency to use VMs of that series at the time of the snapshot compared to subscriptions that are created after the VM series launches. We interpret this as evidence of customers’ aversion to upgrading, and justification for the switching cost in our model.

From our snapshot of all active VMs on Azure we computed the number of VMs for each (series, subscription creation month) pair. There is substantial variation in the number of subscriptions created each month as well as grown over time. Therefore, for each subscription creation month, and each series, we calculate the following fraction:

$$\frac{\text{number of VMs } \textit{of the series} \text{ from subscriptions created that month that are used at time of snapshot}}{\text{total number of VMs from subscriptions created that month that are used at time of snapshot}}.$$

For the twenty series for which we had adequate data and could identify the month in which they were launched, we calculated these fractions for each of seven months: from three months before the launch, to three months after it. For each of these twenty series, we then summed these fractions across the seven months to get to total relative usage over this seven month period, and plotted what fraction of this total is associated with each of the seven months in Figure 11.

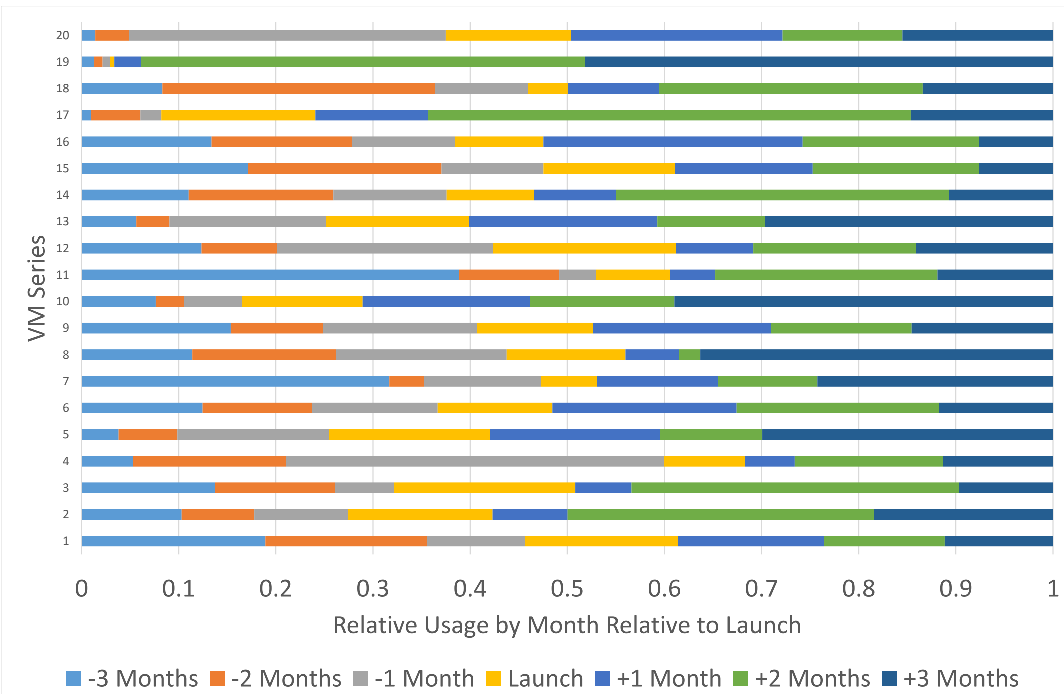

Figure 11: Relative usage of twenty different VM series on Azure by subscriptions created the specified number of months relative to their launch month.

While there is considerable variation among the series, the yellow bar, representing the launch month, is typically shifted to the left of 0.5, indicating that subscriptions created after the launch are typically more likely to use the series at the time of the snapshot than subscriptions created before the launch. On average, relative usage among subscriptions from the three months after the launch is 50% higher than from the three months before the launch, suggesting a substantial switching cost effect.

B Proof of Proposition 1

Proof. Let $\pi_M^* = ((s_0 = 0, x_0 = 0), (s_1 = s_1^*, x_1 = s_1^* p^*), (s_2 = s_2^*, x_2 = s_2^* p^*), (s_3 = s_3^*, x_3 = s_3^* p^*), \ldots)$ denote an optimal policy which uses Myerson pricing.

First, since π_M^* is an optimal policy, introduction times $(s_1^*, s_2^*, s_3^*, s_4^*, \ldots)$ optimize $U(\pi_M)=$ Revenue$(\pi_M)-$Cost$(\pi_M)= \sum_{t=s_1}^{\infty} \delta^t \left(\text{Revenue}(\pi_M, t) - \text{Cost}(\pi_M, t)\right)$, i.e., revenue accumulated less costs incurred during period s_1 and subsequent periods. Second, for an arbitrary policy using Myerson pricing $\pi_M = (s_0 = 0, x_0 = 0), (s_1, x_1 = s_1 p^*), (s_2, x_2 = s_2 p^*), (s_3, x_3 = s_3 p^*), \ldots)$ we can write (1) as

$$\begin{aligned} \text{Revenue}(\pi_M, s_j) =& 2\left(1 - F(p^*)\right) p^* s_j \\ & - \left(F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F(p^*)\right) p^* \left(s_j - s_{j-1}\right)\right). \end{aligned} \tag{8}$$

Notice that the first term depends only on the introduction times through s_j while the second second term only depends on the introduction times through the difference $s_j - s_{j-1}$. Consider the policy π_M' which instead uses introduction times $(s_1' = s_1 + (s_2 - s_1) = s_2, s_2' = s_2 + (s_2 - s_1) = 2s_2 - s_1, s_3' = s_3 + (s_2 - s_1), \ldots)$. Since it delays each introduction by the same constant, it affects the timing and revenue from the first term but only the timing from the second term. Thus we have

$$\begin{aligned} U(\pi_M') =& \delta^{s_2 - s_1} \left(\delta^{s_1}(1 - F(p^*))p^*(s_2 - s_1) + \left(\sum_{t=s_1+1}^{\infty} \delta^t 2(1 - F(p^*))p^*(s_2 - s_1)\right) + U(\pi_M)\right) \\ =& \delta^{s_2 - s_1} \left(\delta^{s_1} \frac{1+\delta}{1-\delta}(1 - F(p^*))p^*(s_2 - s_1) + U(\pi_M)\right). \end{aligned}$$

Every policy whose first introduction is s_2^* can be written as π_M' for some π_M whose first introduction is s_1^*. Thus by the optimality of π_M^*, ${\pi_M^*}'$ is optimal among all policies whose first introduction time is s_2^*. But then by our observations about (8), the policy that first introduces at s_1^* and then uses all the introductions in ${\pi^*_M}'$, i.e. the policy $\pi^{(3)}_M = (s_1 = s^*_1, s_2 = s^*_2, s_3 = s^*_2 + (s^*_2 - s^*_1), s_4 = s^*_3 + (s^*_2 - s^*_1), \ldots)$, is also optimal. Notice that policy $\pi^{(3)}_M$ is periodic through the 3rd introduction.

By repeating this construction we can create a sequence of optimal policies $\pi^{(j)}_M$ which are periodic through the jth introduction. Periodicity follows, with period $s^*_2 - s^*_1$. □

C Proof of Proposition 2

Proof. Fix the periodicity of introductions $\tau > 0$ and the introduction times. Using the alternative expression (8), the revenue of policy π_M, which uses Myerson pricing, at period s_j can be written as

$$\text{Revenue}(\pi_M, s_j) = 2\left(1 - F(p^*)\right) p^* s_j - \left(F\left(p^* + \frac{c}{\tau}\right) - F\left(p^*\right)\right) \cdot p^* \tau.$$

Note that the first term is linear in s_j, while the second term is constant with respect to s_j. By the optimality of p^*, an upper bound on the possible revenue of any policy at time period s_j is $2\left(1 - F(p^*)\right) p^* s_j$, so

$$\lim_{j\to\infty} \frac{\text{Revenue}(\pi, s_j)}{\text{Revenue}(\pi_M, s_j)} \le \lim_{j\to\infty} \frac{2\left(1 - F(p^*)\right) p^* s_j}{2\left(1 - F(p^*)\right) p^* s_j - \left(F\left(p^* + \frac{c}{\tau}\right) - F\left(p^*\right)\right) \cdot p^* \tau} = 1.$$

□

D Proof of Lemma 1

Proof. In the spirit of Myerson's argument, we know that all allocation rules achievable by pricing are monotone, so we can optimize over them instead. In particular, with a finite menu of VM classes, the monotone allocation function is piecewise constant: customers who do not buy get an allocation of 0, those who do get some VM class i. By monotonicity, we just need to choose the thresholds $\theta_1, \ldots, \theta_j$ where the transitions occur. Fixing these, we get an allocation function:

$$a(\theta) = \begin{cases} 0 & \text{if} \quad \theta < \theta_1 \\ i & \text{if} \quad \theta_i \leq \theta < \theta_{i+1}, \quad i \leq 1 \leq j-1 \\ j & \text{if} \quad \theta_j \leq \theta. \end{cases} \tag{9}$$

Fix a and the resulting policy $\pi'(a)$, and let $I_{\theta_j}(\theta) = 1$ if $\theta \geq \theta_j$ be an indicator for agents who do switch (and thus pay the cost of c). Then the payment the provider gets from a customer of type θ who arrived in the prior period is

$$x_{q2(\pi'(a),t,\theta)} = s_{a(\theta)}\theta - \int_0^\theta s_{a(\theta')} d\theta' - I_{\theta_j}(\theta)c. \tag{10}$$

This makes the expected revenue from a customer who arrived in the prior period $\int x_{q2(\pi'(a),t,\theta)} f(\theta) d\theta$ equal to

$$\begin{aligned}
&\int_0^{\theta_1} x_{q2(\pi'(a),t,\theta)} f(\theta)d\theta + \sum_{i=1}^{j-1} \int_{\theta_i}^{\theta_{i+1}} x_{q2(\pi'(a),t,\theta)} f(\theta)d\theta + \int_{\theta_j}^{\infty} x_{q2(\pi'(a),t,\theta)} f(\theta)d\theta \\
&= \sum_{i=1}^{j-1} \int_{\theta_i}^{\theta_{i+1}} \left(s_i\theta - s_i(\theta - \theta_i) - \sum_{i'=1}^{i-1} s_{i'}(\theta_{i'+1} - \theta_{i'}) \right) f(\theta)d\theta \\
&\quad + \int_{\theta_j}^{\infty} \left(s_j\theta - s_j(\theta - \theta_j) - \sum_{i'=1}^{j-1} s_{i'}(\theta_{i'+1} - \theta_{i'}) - c \right) f(\theta)d\theta \\
&= \sum_{i=1}^{j-1} \int_{\theta_i}^{\theta_{i+1}} s_i\theta_i f(\theta)d\theta - \sum_{i=1}^{j-1} \int_{\theta_{i+1}}^{\infty} \left(s_i(\theta_{i+1} - \theta_i)\right) f(\theta)d\theta + \int_{\theta_j}^{\infty} (s_j\theta_j - c) f(\theta)d\theta \\
&= \left(\sum_{i=1}^{j-1} \left(F(\theta_{i+1}) - F(\theta_i)\right) s_i\theta_i - \left(1 - F(\theta_{i+1})\right) s_i(\theta_{i+1} - \theta_i) \right) + \left(1 - F(\theta_j)\right)(s_j\theta_j - c) \\
&= \left(\sum_{i=1}^{j-1} \left(1 - F(\theta_i)\right) s_i\theta_i - \left(1 - F(\theta_{i+1})\right) s_i\theta_{i+1} \right) + \left(1 - F(\theta_j)\right)(s_j\theta_j - c) \\
&= \left(\sum_{i=1}^{j-1} \left(1 - F(\theta_i)\right)(s_i - s_{i-1})\theta_i \right) + \left(1 - F(\theta_j)\right)\left((s_j - s_{j-1})\theta_j - c\right).
\end{aligned} \tag{11}$$

Each summand in the summation of terms $i = 1, \ldots, j-1$, is, up to a constant multiplier, exactly what p^* is defined to optimize, so it is optimal to set $\theta_i = p^*$ for $i < j$. The term after the summation can be optimized using a first order condition. Taking the derivative with respect to θ_j

yields

$$(-f(\theta_j))\left((s_j - s_{j-1})\theta_j - c\right) + (1 - F(\theta_j))\,(s_j - s_{j-1}),$$

or

$$(s_j - s_{j-1})\,(1 - F(\theta_j) - f(\theta_j)\theta_j) + f(\theta_j)c. \tag{12}$$

The first order condition can then be rewritten as

$$(s_j - s_{j-1})\left(\theta_j - \frac{1 - F(\theta_j)}{f(\theta_j)}\right) = c. \tag{13}$$

By the definition of p^*, the left hand side of Equation (13) is exactly 0 for $\theta_j = p^*$, and is increasing in θ_j by Assumption 1. Thus the optimal solution satisfies $\theta_j \geq p^*$ and so our separate optimization of each θ_i does produce a monotone allocation rule.

We wish to turn $\theta_j \geq p^*$ into a lower bound on x_j. The threshold θ_j at which customers switch to technology j solves $s_j\theta_j - x_j - c = s_{j-1}\theta_j - x_{j-1}$. We observe that for $x_{j-1} = s_{j-1}p^*$ and $x_j = s_j p^* - c$, we have $\theta_j = p^*$. Therefore the optimal pricing satisfies $x_j \geq s_j p^* - c$.

Furthermore, a customer can only switch to technology j if she has already bought technology $j-1$, so any choice with $x_j < s_{j-1}p^*$ is dominated by $x_j = s_{j-1}p^*$, because in the latter case, customers that switch to the new technology pay strictly more than in the former case.

We have concluded that optimizing $\int x_{q_2(\pi',t,\theta)} f(\theta)d\theta$ boils down to optimizing over a threshold θ_j, as it is optimal to set $x_i = s_i p^*$ for $i < j$. For arbitrary $x_j \geq x_{j-1}$, the threshold θ_j at which existing customers switch to technology j solves

$$s_j\theta_j - x_j - c = s_{j-1}\theta_j - x_{j-1}, \tag{14}$$

or

$$\theta_j = \frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}. \tag{15}$$

Let policy π' have pricing $x_i = s_i p^*$ for $i < j$ and $x_j = x$. Then, using Equations (11), (14) and (15), we can write

$$\int x_{q_2(\pi',t,\theta)} f(\theta)d\theta = (1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*). \tag{16}$$

The first order condition is

$$1-F\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)-f\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)\frac{x-s_{j-1}p^*}{s_j-s_{j-1}}=0,$$

or

$$\frac{x-s_{j-1}p^*}{s_j-s_{j-1}}-\frac{1-F\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)}{f\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)}=0.$$

Note that if $c=0$, the left hand side equals 0 for $x=s_jp^*$. Also, note that the left hand side is increasing in both x and c by Assumption 1. Therefore the maximizing x for $c\geq 0$ is at most s_jp^*. □

E Proof of Lemma 2

Proof. We have

$$\begin{aligned} Rev_j(x) &= \left(1-F\left(\frac{x}{s_j}\right)\right)x+\left(1-F\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)\right)x \\ &\quad +\left(F\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)-F(p^*)\right)s_{j-1}p^*. \end{aligned}$$

The first summand is the revenue from the customers who arrive at period s_j and buys the new VM class as long as $\theta s_j - x\geq 0 \Longleftrightarrow \theta\geq\frac{x}{s_j}$. Notice that this customer would buy VM class $k<j$ instead of VM class j if $s_k(\theta-p^*)\geq 0\Longleftrightarrow\theta\geq p^*$ and $\theta s_j-x<s_k(\theta-p^*)\Longleftrightarrow\theta<\frac{x-s_kp^*}{s_j-s_k}$. Since $\frac{x-s_kp^*}{s_j-s_k}\leq\frac{s_jp^*-s_kp^*}{s_j-s_k}=p^*$, the two cannot happen at the same time.

The second summand is the revenue from the customers who arrive at period s_j-1 and switches to the new VM class introduced at time s_j, because $\theta s_j-x-c\geq s_{j-1}(\theta-p^*)\Longleftrightarrow\theta\geq\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}$. Notice that $\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\geq\frac{s_jp^*-c-s_{j-1}p^*+c}{s_j-s_{j-1}}=p^*$, therefore as long as $\theta\geq\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}$, this customer buys VM class $j-1$ in period s_j-1 and doesn't opt out, because $\theta\geq p^*\Longleftrightarrow s_{j-1}(\theta-p^*)\geq 0$.

The third summand is the revenue from the customers who arrive at period s_j-1 and do not switch to the new VM class j, because $\theta s_j-x-c<s_{j-1}(\theta-p^*)\Longleftrightarrow\theta<\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}$, while she buys VM class $j-1$ in period s_j-1, because $s_{j-1}(\theta-p^*)\geq 0\Longleftrightarrow\theta\geq p^*$.

Equation (3) follows.

We similarly have

$$
\begin{aligned}
RevD_j(x_n, x_e) \quad &= \quad \left(1 - F\left(\frac{x_n}{s_j}\right)\right) x_n + \left(1 - F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right) x_e \\
&\quad + \left(F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right) - F(p^*)\right) s_{j-1}p^*,
\end{aligned}
$$

from which Equation (4) follows. □

F Proof of Proposition 3

Proof. Defining x_2^* to be the optimal price for existing customers, we have

$$
\begin{aligned}
\frac{Rev_j(x^*)}{Rev_j(s_jp^*)} &\le \frac{RevD_j(s_jp^*, x_2^*)}{Rev_j(s_jp^*)} \\
&= 1 + \frac{RevD_j(s_jp^*, x_2^*) - Rev_j(s_jp^*)}{Rev_j(s_jp^*)} \\
&= 1 + \frac{\left(1 - F\left(\frac{x_2^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x_2^* - s_{j-1}p^*) - \left(1 - F\left(p^* + \frac{c}{s_j - s_{j-1}}\right)\right)(s_j - s_{j-1})p^*}{Rev_j(s_jp^*)} \\
&\le 1 + \frac{\left(F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{x_2^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(s_j - s_{j-1})p^*}{Rev_j(s_jp^*)} \\
&\le 1 + \frac{\left(F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{x_2^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(s_j - s_{j-1})p^*}{(1 - F(p^*))\, s_jp^*} \\
&\le 1 + \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{x_2^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)}{1 - F(p^*)} \\
&\le 1 + \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{\max(s_{j-1}p^*, s_jp^* - c) - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)}{1 - F(p^*)} \\
&= 1 + \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}.
\end{aligned}
$$

The first inequality follows because s_jp^* is the optimal price for new customers and we define x_2^* to be optimal for existing ones, the second because $x_2^* \le s_jp^*$ by Lemma 1, the third because, by the definition of Rev_j, $Rev_j(s_jp^*) \ge (1 - F(p^*))\, s_jp^*$, and the fifth by the lower bound on x_2^* from Lemma 1. □

G Proof of Theorem 1

Proof. By definition, we have

$$\text{Revenue}(\pi', t) = \int (x_{q_1(\pi',t,\theta)} + x_{q_2(\pi',t,\theta)}) f(\theta) d\theta.$$

The essence of our proof is that our assumption makes both terms quasiconcave. While the sum of two quasiconcave functions is not necessarily quasiconcave, for univariate quasiconcave functions it holds that if there is an interval that contains the maxima of both functions, then their sum is also maximized in that interval. We establish this for the interval $(\max(s_{j-1}p^*, s_j p^* - c), s_j p^*)$.

We begin with the second term. Let policy π' have pricing $x_i = s_i p^*$ for $i < j$ and $x_j = x$. Then, as in the proof of Lemma 1, we can write

$$\int x_{q_2(\pi',t,\theta)} f(\theta) d\theta = (1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*). \tag{17}$$

By Lemma 1, the maximizing x is at least $\max(s_{j-1}p^*, s_j p^* - c)$ and at most $s_j p^*$. Furthermore, Assumption 1 is equivalent to the log concavity of $1 - F(x)$ as a function of x, which in turn implies Equation (17) is a log concave function of x, and thus quasiconcave on $[s_{j-1}p^*, \infty)$ (see Bagnoli and Bergstrom, 2005).

Turning to the first term, note that q_1 and q_2 are identical if $c = 0$, and our analysis only assumed that $c \geq 0$. Thus the exact same analysis shows that the desired properties hold for x_{q_1} as well.

Finally, when taking into account the first term (i.e., new customers), the optimal price is the one that optimizes

$$(1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*) + \left(1 - F\left(\frac{x}{s_j}\right)\right) x = Rev_j(x)$$

over x, where we add the term corresponding to revenue extracted from new customers to the right hand side of Equation (17). $\square$

H A 1/4 Bound for Uniform(0,1)

The derivative of $Rev_j(x_j)$ in Equation (3) is

$$1 - \frac{x_j - s_{j-1}p^* + c}{s_j - s_{j-1}} - \frac{1}{s_j - s_{j-1}}(x_j - s_{j-1}p^*) + 1 - \frac{x_j}{s_j} - \frac{1}{s_j}x_j,$$

or

$$2 - 2\frac{x_j - s_{j-1}p^*}{s_j - s_{j-1}} - 2\frac{x_j}{s_j} - \frac{c}{s_j - s_{j-1}}.$$

This is clearly decreasing in x_j and plugging in $x_j = s_j p^* = s_j/2$ yields $-c/(s_j - s_{j-1})$.

The second derivative with respect to x_j is

$$-\frac{2}{s_j - s_{j-1}} - \frac{2}{s_j}.$$

This means the difference in price between the Myerson price and the optimal price is at most

$$\frac{\frac{c}{s_j - s_{j-1}}}{\frac{2}{s_j} + \frac{2}{s_j - s_{j-1}}} = \frac{s_j c}{4s_j - 2s_{j-1}}.$$

Multiplying these gives a bound on the gain of

$$\frac{s_j c}{4s_j - 2s_{j-1}} \cdot \frac{c}{s_j - s_{j-1}} = \frac{s_j c^2}{(4s_j - 2s_{j-1})(s_j - s_{j-1})}. \tag{18}$$

Plugging Myerson pricing into (3) and specializing to the uniform case gives a Myerson revenue at time s_j of

$$\begin{aligned}
&(1-p^*)s_{j-1}p^* + \left(1 - \frac{s_j p^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)(s_j p^* - s_{j-1}p^*) + (1-p^*)s_j p^* \\
&= s_{j-1}p^* - s_{j-1}(p^*)^2 + s_j p^* - s_{j-1}p^* - s_j(p^*)^2 + s_{j-1}(p^*)^2 - cp^* + s_j p^* - s_j(p^*)^2 \\
&= 2s_j p^* - 2s_j(p^*)^2 - cp^* \\
&= \frac{s_j - c}{2}.
\end{aligned}$$

We now bound the ratio of the revenue gain to the Myerson revenue, assuming that $c \leq (s_j - s_{j-1})/2$ (beyond this c, no customer will switch under Myerson pricing, and it can be directly verified that gains from non-Myerson pricing are small):

$$\frac{\frac{s_j c^2}{(4s_j - 2s_{j-1})(s_j - s_{j-1})}}{\frac{s_j - c}{2}} = \frac{2s_j c^2}{(4s_j - 2s_{j-1})(s_j - s_{j-1})(s_j - c)}$$
$$\leq \frac{2s_j(s_j - s_{j-1})^2}{(4s_j - 2s_{j-1})(s_j - s_{j-1})\,(4s_j - 2(s_j - s_{j-1}))}$$
$$= \frac{s_j(s_j - s_{j-1})}{(4s_j - 2s_{j-1})(s_j + s_{j-1})} \tag{19}$$

This is decreasing in s_{j-1}, so we take $s_{j-1} = 0$. The expression then simplifies to 1/4.

I Omitted Figures

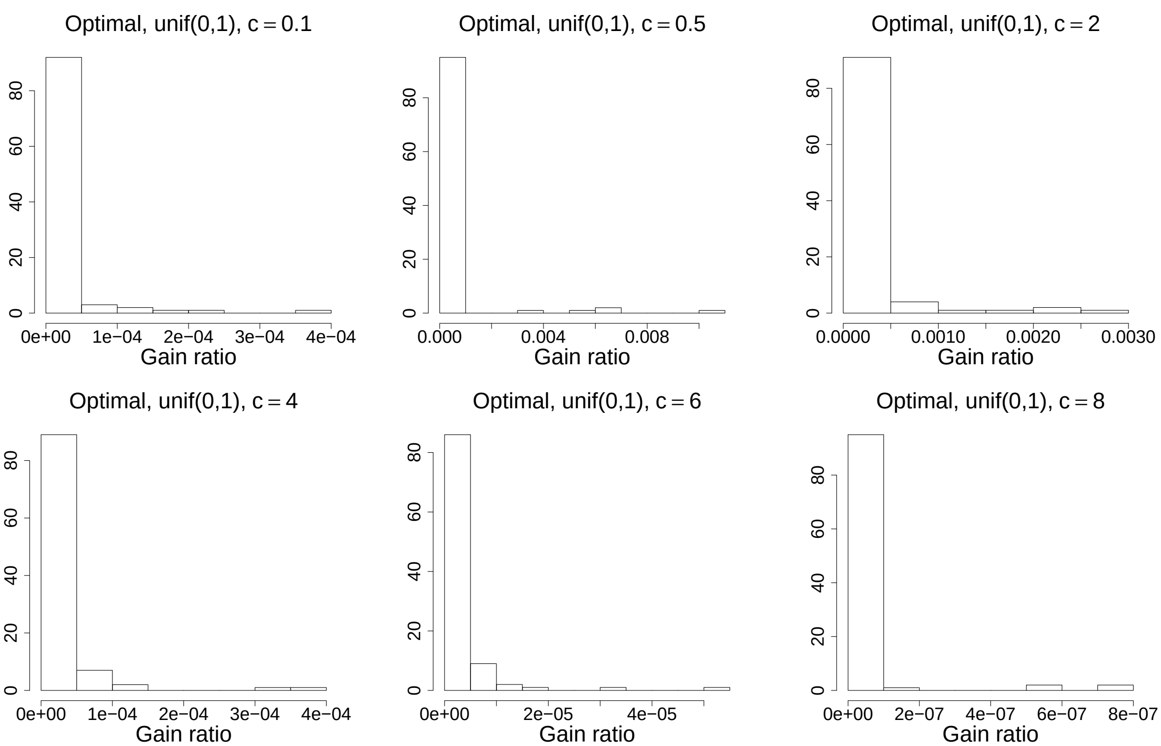

Figure 12: Histograms of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, for the uniform distribution on $[0, 1]$, for discount rate $\delta = 0.5$, and for switching cost $c = 0.1, 0.5, 2, 4, 6, 8$.

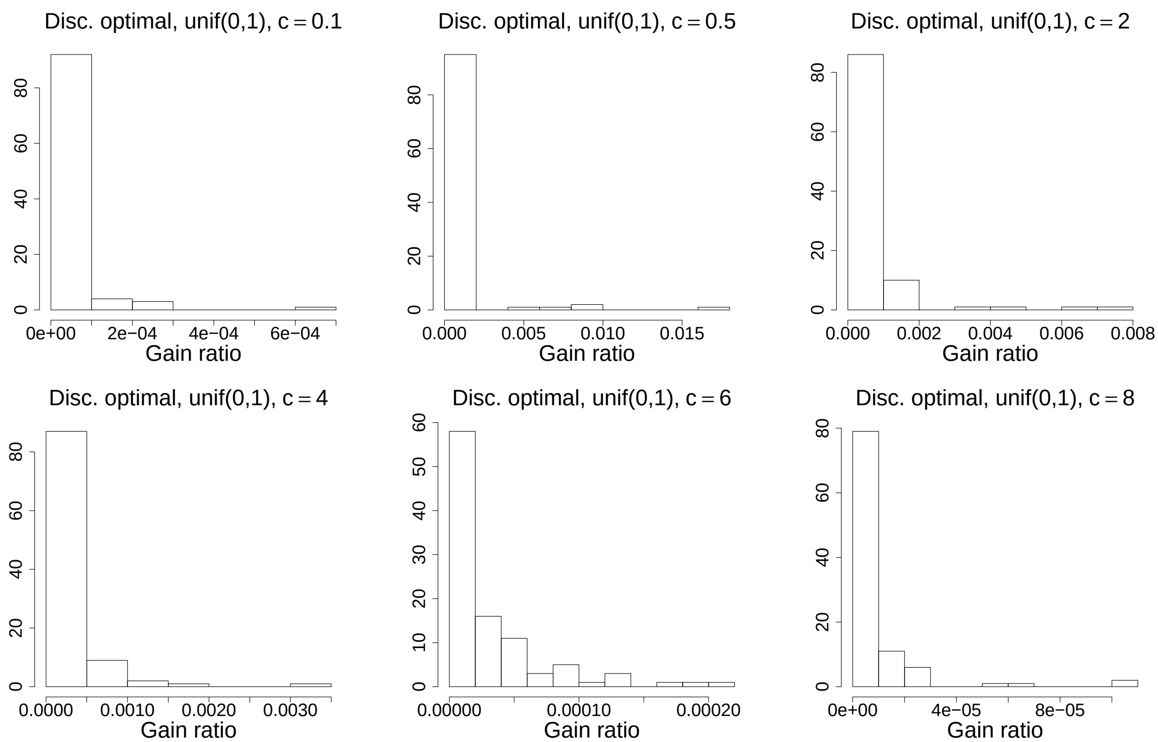

Figure 13: Histograms of the gain ratio of optimal total revenue over Myerson total revenue in the discriminatory setting, over 100 simulations, for the uniform distribution on $[0, 1]$, for discount rate $\delta = 0.5$, and for switching cost $c = 0.1, 0.5, 2, 4, 6, 8$.

J Unique Maximum for Revenue and Convergence of Best-Response Updating

We express the revenue of the provider as a function of the prices $x_1, x_2, \ldots$, given introduction times $s_1, s_2, \ldots$.

We make the following additional assumptions:

Assumption 2. Function $p(1 - F(p))$ is strictly concave for $p > 0$.

Assumption 3. For all introductions $j \geq 2$, we have that $x_j \geq \frac{s_j}{s_{j-1}} x_{j-1} - c$.

Both of these assumptions provide sufficient rather than necessary conditions for the subsequent analysis. In terms of applying this analysis to our numerical bounds, Assumption 2 is satisfied for the uniform distribution we use. It is not satisfied for the beta and gamma distributions we use, because at unreasonably high prices both become convex. However, it does hold for a large range of plausible prices, which is sufficient in practice. Similarly, it is possible to construct prices which violate Assumption 3, but such prices are typically unreasonable in practice and we did not observe them in calculating our numerical bounds.

To simplify notation, we denote by R_t the provider's discounted revenue accumulated during time period t (i.e., due to customers who arrived in periods $t-1$ and t). We have

$$R_{s_1}(x_1) = -\delta^{s_1} C + \delta^{s_1} \mathbb{P}\left(\theta \geq \frac{x_1}{s_1}\right) \cdot x_1, \tag{20}$$

and for $j \geq 2$,

$$\begin{aligned} R_{s_j}(x_{j-1}, x_j) = &- \delta^{s_j} C \\ &+ \delta^{s_j} \cdot \left[\mathbb{P}\left(\theta \geq \frac{x_j}{s_j}\right) \cdot x_j \right. \\ &+ \mathbb{P}\left(\theta \geq \max\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}, \frac{x_{j-1}}{s_{j-1}}\right)\right) \cdot x_j \\ &\left. + \mathbb{P}\left(\frac{x_{j-1}}{s_{j-1}} \leq \theta < \frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right) \cdot x_{j-1} \cdot \mathbb{1}\left(\frac{x_{j-1}}{s_{j-1}} \leq \frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\right] \\ = &\quad \delta^{s_j} \cdot \left[\left(1 - F\left(\frac{x_j}{s_j}\right)\right) \cdot x_j + \left(1 - F\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\right) \cdot x_j \right. \\ &\left. + \left(F\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right) - F\left(\frac{x_{j-1}}{s_{j-1}}\right)\right) \cdot x_{j-1}\right], \end{aligned} \tag{21}$$

where the first line in (21) captures the introduction cost to the provider, the second line captures the revenue due to customers who arrive in period s_j, the third line captures the revenue due to customers who arrive in period $s_j - 1$ and switch in period s_j, and the fourth line captures the revenue due to customers who arrive in period $s_j - 1$ and do not switch in period s_j. Notice that because of Assumption 3, we have $\frac{x_{j-1}}{s_{j-1}} \leq \frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}$ necessarily.

The discounted revenue accumulated during periods $s_j + 1, \ldots, s_{j+1} - 1$, preceding the $j+1$-th introduction, is given by

$$\begin{aligned} R_{s_j+1}(x_j) + R_{s_j+2}(x_j) + \ldots + R_{s_{j+1}-1}(x_j) &= \sum_{t=1}^{s_{j+1}-s_j-1} \delta^{s_j+t} \cdot 2 \cdot \mathbb{P}\left(\theta \geq \frac{x_j}{s_j}\right) \cdot x_j \\ &= 2 \cdot \frac{\delta^{s_j+1} - \delta^{s_{j+1}}}{1-\delta} \cdot \left(1 - F\left(\frac{x_j}{s_j}\right)\right) \cdot x_j. \end{aligned} \tag{22}$$

Fix an integer n. Assuming the horizon stops when the n-th VM class is introduced, i.e., exactly at period s_n, we denote the total revenue TR for the provider, given introduction times $s_1, s_2, \ldots, s_n$, by $TR_{s_1,\ldots,s_n}(x_1, \ldots, x_n)$, so that

$$\begin{aligned}
TR_{s_1,\ldots,s_n}(x_1, \ldots, x_n) =& R_{s_1}(x_1) + R_{s_1+1}(x_1) + \ldots + R_{s_2-1}(x_1) \\
&+ R_{s_2}(x_1, x_2) + R_{s_2+1}(x_2) + \ldots + R_{s_3-1}(x_2) \\
&\vdots \\
&+ R_{s_{n-1}}(x_{n-2}, x_{n-1}) + R_{s_{n-1}+1}(x_{n-1}) + \ldots + R_{s_n-1}(x_{n-1}) \\
&+ R_{s_n}(x_{n-1}, x_n).
\end{aligned}$$

In what follows, we show that total revenue $TR_{s_1,\ldots,s_n}(x_1, \ldots, x_n)$ has a unique maximum with respect to prices $x_1, \ldots, x_n$, and that an algorithm based on best-response updating would converge to it.

J.1 Revenue has a unique maximum

We define[5]

$$\phi_1(x_1, x_2) = R_{s_1}(x_1) + R_{s_1+1}(x_1) + \ldots + R_{s_2-1}(x_1) + R_{s_2}(x_1, x_2),$$

and, for $2 \leq j \leq n-1$,

$$\phi_j(x_{j-1}, x_j, x_{j+1}) = R_{s_j}(x_{j-1}, x_j) + R_{s_j+1}(x_j) + \ldots + R_{s_{j+1}-1}(x_j) + R_{s_{j+1}}(x_j, x_{j+1}),$$

and

$$\phi_n(x_{n-1}, x_n) = R_{s_n}(x_{n-1}, x_n).$$

Notice that

$$\frac{\partial TR_{s_1,\ldots,s_n}}{\partial x_j} = \frac{\partial \phi_j}{\partial x_j}, \qquad 1 \leq j \leq n. \tag{23}$$

We can consider function ϕ_i as the payoff of player i in a n-person game, where player i's strategy is x_i. We will use Theorem 2 in Rosen (1965) to show that there is a unique equilibrium

[5] We sometimes use instead the notation $\phi_j(\mathbf{x})$, where $\mathbf{x}$ is a $n \times 1$ price vector. This notation specifies all prices rather than only those used in the calculation of ϕ_j for each j.

point $\mathbf{x}^0 = (x_1^0, \ldots, x_n^0)$ satisfying

$$
\begin{aligned}
\phi_1(x_1^0, x_2^0) &= \max_{y_1} \phi_1(y_1, x_2^0), && (24)\\
\phi_j(x_{j-1}^0, x_j^0, x_{j+1}^0) &= \max_{y_j} \phi_j(x_{j-1}^0, y_j, x_{j+1}^0), \quad 2 \le j \le n-1, && (25)\\
\phi_n(x_{n-1}^0, x_n^0) &= \max_{y_n} \phi_n(x_{n-1}^0, y_n), && (26)
\end{aligned}
$$

Any equilibrium point, i.e., any price vector satisfying (24), (25), and (26), maximizes ϕ_j for all j (holding x_{-j} fixed), and because of (23), it also maximizes (locally) the total revenue $TR_{s_1,\ldots,s_n}$. Conversely, any (local) maximizer of total revenue $TR_{s_1,\ldots,s_n}$ (locally) optimizes ϕ_j for all j (holding the x_{-j} fixed). If all local maxima of each ϕ_j are global maxima, this is an equilibrium point of the proposed game. Therefore, uniqueness of equilibrium of the game plus strict concavity of each ϕ_j implies uniqueness of the maximum for the total revenue.

First, we show that each ϕ_j is strictly concave in x_j. We have

$$
\begin{aligned}
\frac{\partial^2 \phi_j}{\partial x_j^2} &= \sum_{\ell=s_j}^{s_{j+1}} \frac{\partial^2 R_{s_\ell}}{\partial x_j^2}, \qquad 1 \le j \le n-1,\\
\frac{\partial^2 \phi_n}{\partial x_n^2} &= \frac{\partial^2 R_{s_n}}{\partial x_n^2}.
\end{aligned}
$$

The first (partial) derivative of revenue accumulated during a period with an introduction of new technology is

$$\frac{dR_1}{dx_1} = \delta^{s_1} \cdot \left(-f\left(\frac{x_1}{s_1}\right) \frac{x_1}{s_1} + 1 - F\left(\frac{x_1}{s_1}\right)\right),$$

and for $2 \le j \le n$,

$$
\begin{aligned}
\frac{\partial R_{s_j}}{\partial x_j} = \delta^{s_j} \cdot \Bigg[& -f\left(\frac{x_j}{s_j}\right)\frac{x_j}{s_j} + 1 - F\left(\frac{x_j}{s_j}\right) - f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_j}{s_j - s_{j-1}}\\
& +1 - F\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right) + f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_{j-1}}{s_j - s_{j-1}}\Bigg].
\end{aligned}
$$

The second partial derivative is

$$\frac{d^2 R_1}{dx_1^2} = \delta^{s_1} \cdot \left(-f'\left(\frac{x_1}{s_1}\right)\frac{x_1}{s_1^2} - 2f\left(\frac{x_1}{s_1}\right)\frac{1}{s_1}\right), \tag{27}$$

and for $2 \le j \le n$,

$$\frac{\partial^2 R_{s_j}}{\partial x_j^2} = \delta^{s_j} \cdot \left[-f'\left(\frac{x_j}{s_j}\right)\frac{x_j}{s_j^2} - 2f\left(\frac{x_j}{s_j}\right)\frac{1}{s_j} - f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_j}{(s_j - s_{j-1})^2} \right.$$
$$\left. -2f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{1}{s_j - s_{j-1}} + f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_{j-1}}{(s_j - s_{j-1})^2} \right]. \tag{28}$$

The second derivative of revenue accumulated during periods in between introductions is, for $1 \le j \le n-1$,

$$\frac{d^2(R_{s_j+1} + R_{s_j+2} + \ldots + R_{s_{j+1}-1})}{dx_j^2} = 2 \cdot \frac{\delta^{s_j+1} - \delta^{s_{j+1}}}{1-\delta} \cdot \left(-f'\left(\frac{x_j}{s_j}\right)\frac{x_j}{s_j^2} - 2f\left(\frac{x_j}{s_j}\right)\frac{1}{s_j} \right). \tag{29}$$

The second partial derivative of revenue accumulated during the subsequent period right after a new VM class is introduced is, for $1 \le j \le n-1$,

$$\frac{\partial^2 R_{s_{j+1}}}{\partial x_j^2} = \delta^{s_{j+1}} \cdot \left[-f'\left(\frac{x_{j+1} - x_j + c}{s_{j+1} - s_j}\right)\frac{x_{j+1}}{(s_{j+1} - s_j)^2} + f'\left(\frac{x_{j+1} - x_j + c}{s_{j+1} - s_j}\right)\frac{x_j}{(s_{j+1} - s_j)^2} \right.$$
$$\left. -2f\left(\frac{x_{j+1} - x_j + c}{s_{j+1} - s_j}\right)\frac{1}{s_{j+1} - s_j} - f'\left(\frac{x_j}{s_j}\right)\frac{x_j}{s_j^2} - 2f\left(\frac{x_j}{s_j}\right)\frac{1}{s_j} \right]. \tag{30}$$

By Assumption 2, we have

$$-f'(x) \cdot x - 2f(x) < 0 \quad \text{for } x > 0. \tag{31}$$

It follows that the expressions in (27) and (29) are negative. Similarly, the sum of the first two terms in (28), which is also the sum of the last two terms in (30), i.e., $-f'\left(\frac{x_j}{s_j}\right)\frac{x_j}{s_j^2} - 2f\left(\frac{x_j}{s_j}\right)\frac{1}{s_j}$, is also negative. The sum of the remaining (i.e., the last three) terms in (28) is negative, because

$$\begin{aligned}
&-f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_j}{(s_j - s_{j-1})^2} - 2f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{1}{s_j - s_{j-1}} + f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_{j-1}}{(s_j - s_{j-1})^2} \\
&\quad = -f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_j - x_{j-1}}{(s_j - s_{j-1})^2} - 2f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{1}{s_j - s_{j-1}} \\
&\quad \leq \frac{2}{\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}} f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{x_j - x_{j-1}}{(s_j - s_{j-1})^2} - 2f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{1}{s_j - s_{j-1}} \\
&\quad = 2 \cdot f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right)\frac{1}{s_j - s_{j-1}}\left(\frac{x_j - x_{j-1}}{x_j - x_{j-1} + c} - 1\right) \\
&\quad < 0,
\end{aligned} \tag{32}$$

where the first inequality follows from (31). Similarly, the sum of the remaining (i.e., the first three) terms in (30) is negative.

Having shown that ϕ_j is strictly concave in x_j for all j, it follows that

$$\left(x_j^{(1)} - x_j^{(2)}\right) \cdot \left(\left.\frac{\partial \phi_j}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}^{(2)}} - \left.\frac{\partial \phi_j}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}^{(1)}}\right) > 0, \quad \text{for all } \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \text{ s.t. } x_j^{(1)} \neq x_j^{(2)}, \text{ for all } j,$$

because $\frac{\partial \phi_j}{\partial x_j}$ is strictly decreasing in x_j. It follows that

$$\sum_{j=1}^{n} \left(x_j^{(1)} - x_j^{(2)}\right) \cdot \left.\frac{\partial \phi_j}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}^{(2)}} + \left(x_j^{(2)} - x_j^{(1)}\right) \cdot \left.\frac{\partial \phi_j}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}^{(1)}} > 0, \quad \text{for all } \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \text{ s.t. } \mathbf{x}^{(1)} \neq \mathbf{x}^{(2)}.$$

We recognize this as the definition of diagonal strict concavity (Definition in Section 3, Rosen, 1965) for function $\sum_{j=1}^{n} \phi_j(\mathbf{x})$. By Theorem 2 in Rosen (1965), it follows that there exists a unique equilibrium point $\mathbf{x}^0$ satisfying (24), (25), (26).

J.2 Best-response updating converges to the unique maximizer of revenue

Following Rosen (1965), for vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{r} = (r_1, \ldots, r_n)$, we define

$$g(\mathbf{x}, \mathbf{r}) = \begin{pmatrix} r_1 \frac{\partial \phi_1(\mathbf{x})}{\partial x_1} \\ r_2 \frac{\partial \phi_2(\mathbf{x})}{\partial x_2} \\ \vdots \\ r_n \frac{\partial \phi_n(\mathbf{x})}{\partial x_n} \end{pmatrix}.$$

We will show that the symmetric matrix $G(\mathbf{x}, \mathbf{r}) + G'(\mathbf{x}, \mathbf{r})$ is negative definite for some $\mathbf{r}$, where $G(\mathbf{x}, \mathbf{r})$ is the Jacobian of $g(\mathbf{x}, \mathbf{r})$ with respect to $\mathbf{x}$:

$$G(\mathbf{x}, \mathbf{r}) = \begin{pmatrix} r_1 \frac{\partial^2 \phi_1}{\partial x_1^2} & r_1 \frac{\partial^2 \phi_1}{\partial x_2 \partial x_1} & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ r_2 \frac{\partial^2 \phi_2}{\partial x_1 \partial x_2} & r_2 \frac{\partial^2 \phi_2}{\partial x_2^2} & r_2 \frac{\partial^2 \phi_2}{\partial x_3 \partial x_2} & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & r_3 \frac{\partial^2 \phi_3}{\partial x_2 \partial x_3} & r_2 \frac{\partial^2 \phi_3}{\partial x_3^2} & r_3 \frac{\partial^2 \phi_3}{\partial x_4 \partial x_3} & 0 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & r_{n-1} \frac{\partial^2 \phi_{n-1}}{\partial x_{n-2} \partial x_{n-1}} & r_{n-1} \frac{\partial^2 \phi_{n-1}}{\partial x_{n-1}^2} & r_{n-1} \frac{\partial^2 \phi_{n-1}}{\partial x_n \partial x_{n-1}} \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & r_n \frac{\partial^2 \phi_n}{\partial x_{n-1} \partial x_n} & r_n \frac{\partial^2 \phi_n}{\partial x_n^2} \end{pmatrix}. \tag{33}$$

We notice that for $r_i = \frac{1}{2}$ for all i, $G\left(\mathbf{x}, \frac{\mathbf{1}}{\mathbf{2}}\right) + G'\left(\mathbf{x}, \frac{\mathbf{1}}{\mathbf{2}}\right)$ is

$$\begin{pmatrix} \frac{\partial^2 \phi_1}{\partial x_1^2} & \frac{1}{2} \frac{\partial^2 \phi_1}{\partial x_2 \partial x_1} + \frac{1}{2} \frac{\partial^2 \phi_2}{\partial x_1 \partial x_2} & 0 & 0 & 0 & \ldots \\ \frac{1}{2} \frac{\partial^2 \phi_1}{\partial x_2 \partial x_1} + \frac{1}{2} \frac{\partial^2 \phi_2}{\partial x_1 \partial x_2} & \frac{\partial^2 \phi_2}{\partial x_2^2} & \frac{1}{2} \frac{\partial^2 \phi_2}{\partial x_3 \partial x_2} + \frac{1}{2} \frac{\partial^2 \phi_3}{\partial x_2 \partial x_3} & 0 & 0 & \ldots \\ 0 & \frac{1}{2} \frac{\partial^2 \phi_2}{\partial x_3 \partial x_2} + \frac{1}{2} \frac{\partial^2 \phi_3}{\partial x_2 \partial x_3} & \frac{\partial^2 \phi_3}{\partial x_3^2} & \frac{1}{2} \frac{\partial^2 \phi_3}{\partial x_4 \partial x_3} + \frac{1}{2} \frac{\partial^2 \phi_4}{\partial x_3 \partial x_4} & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where the last three rows are

$$\begin{pmatrix} 0 & \ldots & 0 & \frac{1}{2} \frac{\partial^2 \phi_{n-3}}{\partial x_{n-2} \partial x_{n-3}} + \frac{1}{2} \frac{\partial^2 \phi_{n-2}}{\partial x_{n-3} \partial x_{n-2}} & \frac{\partial^2 \phi_{n-2}}{\partial x_{n-2}^2} & \frac{1}{2} \frac{\partial^2 \phi_{n-2}}{\partial x_{n-1} \partial x_{n-2}} + \frac{1}{2} \frac{\partial^2 \phi_{n-1}}{\partial x_{n-2} \partial x_{n-1}} & 0 \\ 0 & \ldots & 0 & 0 & \frac{1}{2} \frac{\partial^2 \phi_{n-2}}{\partial x_{n-1} \partial x_{n-2}} + \frac{1}{2} \frac{\partial^2 \phi_{n-1}}{\partial x_{n-2} \partial x_{n-1}} & \frac{\partial^2 \phi_{n-1}}{\partial x_{n-1}^2} & \frac{1}{2} \frac{\partial^2 \phi_{n-1}}{\partial x_n \partial x_{n-1}} + \frac{1}{2} \frac{\partial^2 \phi_n}{\partial x_{n-1} \partial x_n} \\ 0 & \ldots & 0 & 0 & 0 & \frac{1}{2} \frac{\partial^2 \phi_{n-1}}{\partial x_n \partial x_{n-1}} + \frac{1}{2} \frac{\partial^2 \phi_n}{\partial x_{n-1} \partial x_n} & \frac{\partial^2 \phi_n}{\partial x_n^2} \end{pmatrix}.$$

This can be written as

$$\begin{pmatrix} \frac{\partial^2(R_{s_1}+R_{s_1+1}+\ldots+R_{s_2})}{\partial x_1^2} & \frac{\partial^2 R_{s_2}}{\partial x_1 \partial x_2} & 0 & 0 & 0 & \ldots \\ \frac{\partial^2 R_{s_2}}{\partial x_1 \partial x_2} & \frac{\partial^2(R_{s_2}+R_{s_2+1}+\ldots+R_{s_3})}{\partial x_2^2} & \frac{\partial^2 R_{s_3}}{\partial x_2 \partial x_3} & 0 & 0 & \ldots \\ 0 & \frac{\partial^2 R_{s_3}}{\partial x_2 \partial x_3} & \frac{\partial^2(R_{s_3}+R_{s_3+1}+\ldots+R_{s_4})}{\partial x_3^2} & \frac{\partial^2 R_{s_4}}{\partial x_3 \partial x_4} & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{34}$$

where the last three rows are

$$\begin{pmatrix} 0 & \ldots & 0 & \frac{\partial^2 R_{s_{n-2}}}{\partial x_{n-3} \partial x_{n-2}} & \frac{\partial^2(R_{s_{n-2}}+R_{s_{n-2}+1}+\ldots+R_{s_{n-1}})}{\partial x_{n-2}^2} & \frac{\partial^2 R_{s_{n-1}}}{\partial x_{n-2} \partial x_{n-1}} & 0 \\ 0 & \ldots & 0 & 0 & \frac{\partial^2 R_{s_{n-1}}}{\partial x_{n-2} \partial x_{n-1}} & \frac{\partial^2(R_{s_{n-1}}+R_{s_{n-1}+1}+\ldots+R_{s_n})}{\partial x_{n-1}^2} & \frac{\partial^2 R_{s_n}}{\partial x_{n-1} \partial x_n} \\ 0 & \ldots & 0 & 0 & 0 & \frac{\partial^2 R_{s_n}}{\partial x_{n-1} \partial x_n} & \frac{\partial^2 R_{s_n}}{\partial x_n^2} \end{pmatrix}.$$

We recognize (34) as the Hessian H of total revenue $TR_{s_1,\ldots,s_n}$.

We show that the negation of the Hessian H is a strictly diagonally dominant[6] matrix with positive diagonal entries, and therefore that it is positive definite. It then follows that $G\left(\mathbf{x}, \frac{\mathbf{1}}{\mathbf{2}}\right) + G'\left(\mathbf{x}, \frac{\mathbf{1}}{\mathbf{2}}\right)$ is negative definite.

We first compute the derivative

$$\begin{aligned} \left|\frac{\partial^2 R_{s_j}}{\partial x_{j-1} \partial x_j}\right| = \delta^{s_j} \cdot \Bigg[& f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right) \frac{x_j}{(s_j - s_{j-1})^2} + 2f\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right) \frac{1}{s_j - s_{j-1}} \\ & -f'\left(\frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}\right) \frac{x_{j-1}}{(s_j - s_{j-1})^2} \Bigg], \end{aligned} \tag{35}$$

for $2 \leq j \leq n$. Notice that the right hand side of Equation (35) is positive, by Equation (32).

For $2 \leq j \leq n-1$, we want to compare the magnitude of the jth diagonal entry of the Hessian $|H_{jj}|$, to the sum of magnitudes of the non-diagonal entries in the same row j, $|H_{j,j-1}| + |H_{j,j+1}|$.

[6] A square matrix A is strictly diagonally dominant if $|a_{ii}| > \sum_{i \neq j} |a_{ij}|$, for all i, where a_{ij} denotes the entry in the ith row and jth column of A.

We have

$$
\begin{aligned}
|H_{j,j-1}|+|H_{j,j+1}| &= \left|\frac{\partial^2 R_{s_j}}{\partial x_{j-1}\partial x_j}\right| + \left|\frac{\partial^2 R_{s_{j+1}}}{\partial x_j \partial x_{j+1}}\right| \\
&= \delta^{s_j}\cdot\left[f'\left(\frac{x_j - x_{j-1}+c}{s_j - s_{j-1}}\right)\frac{x_j}{(s_j-s_{j-1})^2} + 2f\left(\frac{x_j - x_{j-1}+c}{s_j - s_{j-1}}\right)\frac{1}{s_j - s_{j-1}}\right. \\
&\qquad\qquad \left. - f'\left(\frac{x_j - x_{j-1}+c}{s_j - s_{j-1}}\right)\frac{x_{j-1}}{(s_j-s_{j-1})^2}\right] \\
&+ \delta^{s_{j+1}}\cdot\left[f'\left(\frac{x_{j+1} - x_j+c}{s_{j+1} - s_j}\right)\frac{x_{j+1}}{(s_{j+1}-s_j)^2} + 2f\left(\frac{x_{j+1} - x_j+c}{s_{j+1} - s_j}\right)\frac{1}{s_{j+1} - s_j}\right. \\
&\qquad\qquad \left. - f'\left(\frac{x_{j+1} - x_j+c}{s_{j+1} - s_j}\right)\frac{x_j}{(s_{j+1}-s_j)^2}\right], \qquad (36)
\end{aligned}
$$

while

$$
\begin{aligned}
|H_{jj}| &= \left|\frac{\partial^2 R_{s_j}}{\partial x_j^2} + \frac{\partial^2\left(R_{s_j+1}+\ldots+R_{s_{j+1}-1}\right)}{\partial x_j^2} + \frac{\partial^2 R_{s_{j+1}}}{\partial x_j^2}\right| \\
&= \left|\frac{\partial^2 R_{s_j}}{\partial x_j^2}\right| + \left|\frac{\partial^2\left(R_{s_j+1}+\ldots+R_{s_{j+1}-1}\right)}{\partial x_j^2}\right| + \left|\frac{\partial^2 R_{s_{j+1}}}{\partial x_j^2}\right|, \qquad (37)
\end{aligned}
$$

where the second equality follows because the expressions on the right hand side of each of Equations (28), (29), (30) are negative. By comparing Equation (36) to Equations (28), (29), and (30), it follows that $|H_{jj}| > |H_{j,j-1}| + |H_{j,j+1}|$.

It is also clear that

$$
\begin{aligned}
|H_{1,1}| &= \left|\frac{\partial^2 R_{s_1}}{\partial x_1^2} + \frac{\partial^2\left(R_{s_1+1}+\ldots+R_{s_2-1}\right)}{\partial x_1^2} + \frac{\partial^2 R_{s_2}}{\partial x_1^2}\right| \\
&> \left|\frac{\partial^2 R_{s_2}}{\partial x_1^2}\right| \\
&> \left|\frac{\partial^2 R_{s_2}}{\partial x_1 \partial x_2}\right| \\
&= |H_{1,2}|,
\end{aligned}
$$

where the second inequality follows by comparing Equations (30) for $j=1$ and (35) for $j=2$.

Also, we have that

$$\begin{aligned}|H_{n,n}| &= \left|\frac{\partial^2 R_{s_n}}{\partial x_n^2}\right| \\ &> \left|\frac{\partial^2 R_{s_n}}{\partial x_{n-1}\partial x_n}\right| \\ &= |H_{n,n-1}|,\end{aligned}$$

where the inequality follows by comparing Equations (28) and (35) for $j = n$.

Having shown that the Hessian $G\left(\mathbf{x}, \mathbf{\frac{1}{2}}\right) + G'\left(\mathbf{x}, \mathbf{\frac{1}{2}}\right)$ is negative definite, we can show, for some algorithms based on best-response updating, that the solution converges to the unique equilibrium $\mathbf{x}^0$. For example, we can employ Theorem 9 in Rosen (1965).