Optimal Pricing and Introduction Timing of Technology Upgrades in Subscription-Based Services

Ian A. Kash*, Peter B. Key , and Spyros I. Zoumpoulis†

*Computer Science, University of Illinois at Chicago

†Decision Sciences, INSEAD

August 10, 2023

Abstract

In the context of subscription-based services, many technologies improve over time and service providers can provide increasingly powerful service upgrades to their customers, but at a launching cost, and the expense of the sales of existing products. We propose a model of technology upgrades and characterize the optimal pricing and timing of technology introductions for a service provider who price-discriminates among customers based on their upgrade experience, in the face of customers who are averse to switching to improved offerings.

We first characterize optimal discriminatory pricing for the infinite horizon pricing problem with fixed introduction times. We reduce the optimal pricing problem to a tractable optimization problem and propose an efficient algorithm for solving it. Our algorithm computes optimal discriminatory prices within a fraction of a second, even for large problem instances.

We then show that periodic introduction times, combined with optimal pricing, enjoy optimality guarantees. In particular, we first show that as long as the introduction intervals are constrained to be non-increasing, it is optimal to have periodic introductions after an initial warm-up phase. When allowing general introduction intervals, we show that periodic introduction intervals after some time are optimal in a more restricted sense. Numerical experiments suggest that it is generally optimal to have periodic introductions after an initial warm-up phase.

Finally, we focus on a setting in which the firm does not price-discriminate based on customers' experience. We show both analytically and numerically that in the non-discriminatory setting, a simple policy of Myerson (i.e., myopic) pricing and periodic introductions enjoys good performance guarantees.

1 Introduction

Many technologies that fuel subscription-based services develop and improve over time, often driven by improvements in the underlying components. For example, mobile phones improve with better screen, processor, and battery technology; software suites such as Microsoft Office and Adobe Creative Cloud advance as new features are added to individual applications; subscription services such as Netflix upgrade their offerings providing higher tiers of service — in the case of Netflix, higher quality of video, higher allowance of number of devices to stream on, etc.; and cloud service providers offer upgraded virtual machines with better performance, when the processors, memory, and expansions cards used in the underlying servers improve.

Service providers face several trade-offs as they seek to make the best use of improved technology. On one hand, the improved technologies are more valuable to customers and command a higher price. On the other hand, there is a cost to develop and launch a new version of a product. Further, the new version of the product competes with existing versions. Thus, the provider faces two questions. First, when should a new version be introduced? Second, how should it be priced, taking into account both the versions that currently exist and the ones that will be introduced in the future?

Existing versions of products and new ones compete in the market in two related ways. First, new customers may be faced with a choice of versions. Second, existing customers have to decide whether to stay with the version they are currently using or switch to a new one. We study a model where the technology in question is rented or otherwise paid for over time (which is the case for at least some customers of all of the previously mentioned products), so that there is no explicit monetary cost when switching other than paying the higher rental price for the improved version.

However, this does not mean that switching is frictionless or costless to the customer. Adopting the new version may prove costly for a variety of reasons, such as redesigning the customer's technology to make use of the new version (e.g., updating business processes to use the new feature of a software suite), downtime during the transition (e.g., needing to transfer the number, settings, and applications to the new phone, or idle time during the updating to the new version of a software suite, or needing to shut down a cloud application so it can be relaunched on new hardware), and human costs associated with users' familiarity with a particular version (e.g., retraining on a new version of a subscription-based service, or simply inertia). Indeed, there is 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 virtual machine 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).[1]

Another salient feature of subscription-based services is price discrimination. Modern subscription services are increasingly personalized to individual subscribers, and as a result discriminatory pricing is ever-present in the marketplace. Specifically, the subscription-based model allows firms to keep track of their customers and to reward existing customers with discounts and special offers on upgrades, as an incentive to upgrade to the improved service. For example, discounts for existing customers for technology upgrades are commonplace in mobile plan services (Verizon, 2021; Vodafone, 2021), home entertainment subscription services (Sky, 2021), and specialized software (PRO Landscape, 2021; Arturia, 2019) and hardware (Ableton, 2021) suites.

In this context, the service provider has to choose from a wide range of possible policies for how to price and time the introduction of new technology versions. Our main contribution is to characterize the optimal discriminatory pricing policy for the service provider and to provide an efficient algorithm for computing optimal discriminatory prices. We also show that it is generally optimal to introduce new technologies in periodic intervals after some time[2]. This policy produces a marketplace where new customers always select the newest and best offering, while existing customers may stick with older versions due to switching costs.

In more detail, we propose a model of technology upgrades featuring discriminatory pricing, in the face of customers who are averse to switching to improved offerings (Section 3). We model technology introduction as a discrete-time process over an infinite time horizon, with future rewards discounted. The quality of a new version is assumed to grow linearly with the introduction time. New customers arrive at each time period and stay for a fixed number of time periods. The provider seeks to maximize expected discounted profit, with her decisions being when to introduce new

[1] We provide further justification for modeling switching costs in Section 2.4.

[2] Periodic introductions have been noticeable in the contexts we are considering. For example, Apple has adopted a cadence of yearly updates of its phones, with alternating major and minor updates.

technology classes and how to price them. The provider can price-discriminate among customers and charge a customer a different price according to how many technology introductions that customer has switched to. Customers act as simple utility maximizers, and choose the technology offering that is most valuable to them; if they are existing customers, they incur a switching cost for switching technologies.

Our first set of results is on developing an efficient algorithm to solve the infinite horizon optimal discriminatory pricing problem with fixed introduction times (Section 4). We first reduce the optimal pricing problem to a tractable optimization problem. We then identify key properties of the optimal solution, which lead naturally to an algorithm for optimal discriminatory pricing. When introduction intervals are non-increasing these prices are particularly easy to compute and take the form of discounts offered to existing customers. The proposed algorithm computes optimal discriminatory prices within a fraction of a second, even for large problem instances.

Our second set of results shows that generally periodic introduction times enjoy optimality guarantees (Section 5). In particular, we first show that, as long as the introduction intervals are constrained to be non-increasing, it is optimal to have periodic introductions after an initial "warm-up" phase. We characterize the optimal period as a function of the details of the setting. We then allow general introduction intervals and show that periodic introduction intervals after some time are optimal in a more restricted sense. Finally, we show numerical experiments that suggest that, when allowing for general introduction intervals, it is optimal to have periodic introduction intervals after an initial warm-up phase of non-increasing introduction intervals.

For our third set of results, we focus on settings in subscription-based services in which the firm does not price-discriminate among customers (Section 6). We show both analytically and numerically that in the non-discriminatory setting, a simple policy of Myerson (i.e., myopic) pricing and periodic introductions enjoys good performance guarantees.

Finally, we discuss the importance of our modelling assumptions to our results with respect to the substantive context of subscription-based services, as well as extensions of our assumptions (Section 7).

In the next section we review the relevant literature on introduction of improved product generations, motivate some of our modelling assumptions, and summarize our contribution.

2 Literature Review and Contribution

We first review works that study the optimal decision making of a firm which launches a new product or technology, or successive generations thereof. Policies for introduction of improved product generations 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. Furthermore, we specifically review works on introduction of new product generations that focus on price discrimination based on customers' purchase history. We then discuss some of our modelling assumptions and state our contribution, in the light of the related literature.

2.1 Introduction of improved product generations

Operations management. 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) construct a decision model to solve the firm's introduction timing problem, and 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. Lobel et al. (2016) 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.

Several other works in operations management have dealt with launch policies for (generations of) new products. Barriola and de Albéniz (2019) model product renewal with a focus on endogenizing the firm's decision for obsolescence, by allowing the firm to choose the decay rate of the consumer's utility for a product. Cohen et al. (1996) 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. 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; they conclude that it is not wise for profit-maximizing firms to arbitrarily shorten product life cycle for the sake of competition. Casadesus-Masanell and Yoffie (2007) study competitive interactions between Intel and Microsoft through a duopoly model between producers of complementary products and 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. Araman and Caldentey (2016) consider a seller who has the ability to first test the market and gather demand information through crowdvoting, before deciding whether or not and when to launch a new product.

Marketing. 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. Gordon (2009) develops a model of consumer product replacement behavior using data from the PC processor industry.

Economics. 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.

2.2 Product upgrades and price discrimination

Our work relates closely to the strand of the literature on introduction of new product generations that studies price discrimination based on customers' purchase history (Acquisti and Varian, 2005; Fudenberg and Villas-Boas, 2006; Li and Jain, 2016; Jing, 2017; Cosguner et al., 2017), and in particular in the form of discounts to existing customers on upgrades (Fudenberg and Tirole, 1998). This form of price discrimination that uses information about the consumers' past purchases to offer different prices (and/or products) to consumers with different purchase histories is oftentimes referred to as behavior-based price discrimination.

A few papers on this topic deal specifically with software technology upgrades. Mehra et al. (2012) allow the software vendor to offer discounts on upgrades both to existing customers of a competitor, and to its own existing customers and, similarly to us, they recognize switching costs as an important aspect of technology upgrade adoption. Jia et al. (2018) analyze the profitability of a selling and a leasing model and different price discrimination strategies, including strategies that differentiate based on consumers' past purchase behavior. Bala and Carr (2009) analyze the optimality of upgrade pricing by characterizing the relationship between magnitude of product improvement and the equilibrium pricing structure, particularly in the context of user upgrade costs.

The insight that a firm can keep track of its former customers and price-discriminate based on customers' previous purchase behavior has been applied in contexts other than software technologies as well. Penmetsa et al. (2015) study past purchase behavior-based price discrimination in the general context of subscription markets. Ray et al. (2005) consider trade-in rebates to existing customers, as an incentive to replace their product with a new one, in the context of remanufacturable products.

2.3 Revenue maximization in service provision

A recent strand of the operations management literature focuses on the revenue maximization problem of subscription-based service providers. We give three examples of recent work. Borgs et al. (2014) 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 computing 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. Gao et al. (2019) 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.

2.4 Modelling assumptions and related literature

We next discuss some of our modelling assumptions in the light of the related literature.

A crucial aspect of our model is that customers are averse to upgrading to improved versions of the provided technology. In the introduction, we identified three sources of switching costs: (a) costs associated to redesigning and reengineering the customer's business processes, (b) costs associated to downtime and disruption during the transition, and (c) human costs associated with customer inertia. While the first two can be readily understood in some contexts, including mobile phones, software suites, and subscription services, we now provide justification for customer inertia. Indeed, there is evidence, documented in the marketing and information systems literatures, that a consumer's past purchase decisions can create inertia in the context of technology adoption. Huang

(2019) finds empirically that learning a technology by doing builds up consumer human capital, whose non-transferability results in switching costs. Zhu et al. (2006) analyze whether switching costs are significant barriers to entry of a new standard. They determine that experience with older standards may create switching costs and make it difficult to shift to potentially better standards, a phenomenon called "excess inertia" in technology change. Oren and Rothkopf (1984) propose a model for new product planning that accounts for customer inertia, and describe specific details of the model used in a system developed for market analysis of high speed nonimpact computer printers. Finally, other than behavioral and cognitive forces, the organizations literature highlights the role of an organization's identity in explaining inertia when faced with new technologies (Tripsas, 2009).

Our model assumes myopic customers who make their decisions based on the current service offerings, as opposed to forward-looking customers who take into account beliefs about future offerings.[3] 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 (2018) 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. (2019) demonstrate that for a broad class of customer utility models, static prices surprisingly continue to remain asymptotically optimal, and

[3] In Section 7 we discuss another sense in which our customers are myopic, as well as the importance of the myopia assumption to our results.

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 (2020), 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 (2020) 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 (2020) proposes an efficient dynamic programming algorithm for finding optimal pricing policies for arbitrary joint distributions of patience levels and valuations.

2.5 Our contribution

We summarize the key points of our contribution in light of the related literature. First, we propose a model for technology upgrades in the context of subscription-based services. Two important features of our model are (i) the switching cost for the customers who upgrade to improved offerings; and (ii) price discrimination based on customers' upgrade history. We recognize these as key features of modern subscription services markets. Second, we characterize optimal discriminatory pricing for the service provider and provide an efficient algorithm for retrieving optimal prices. Third, we show that policies with periodic introduction times after an initial "warm-up" phase, combined with optimal pricing, enjoy optimality guarantees. Fourth, in the setting where the service provider does not price-discriminate, we show that a simple policy of Myerson (i.e., myopic) pricing and periodic introductions enjoys good performance guarantees.

3 A Model of Technology Upgrades with Price Discrimination

Time is discrete with an infinite horizon. At each time period t, a single service provider can introduce up to one new technology class. We generally assume that technology classes, once introduced, remain available for customers to choose thereafter.[4] We assume an unlimited capacity of units for all introduced technology classes.[5]

At each period, a unit mass of new customers arrives. Customers stay in the system for $d \geq 2$ 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 unit of a technology class that was introduced in period t. A customer who has switched m times to an upgraded technology class as an existing customer incurs cost $x_{j,m}$ per time period for using technology class j. Any customer incurs a switching cost $c > 0$ whenever they switch to a different technology class.

The decisions of the provider are when to introduce new technology classes and how to price them. The provider incurs provisioning cost $C > 0$ for introducing a new technology class. We assume that the provider can price-discriminate among customers, and charge a customer a different price according to how many technology introductions that customer has switched to. Thus, for the jth technology class that is introduced, the provider charges price $x_{j,0}$ for newly arriving customers; price $x_{j,1}$ for customers who have already used one previous technology class before j and for whom technology class j is the first introduction they switch to as existing customers; et cetera; and price $x_{j,d-1}$ for customers who have already used $d-1$ previous technology classes before j and for whom technology j will be the $(d-1)$th introduction they switch to as existing customers. We refer to the number of times a customer has switched to a better technology class as the customer's upgrade experience. We take n to be the maximum possible customer upgrade experience, and we have $n \leq d-1$. We denote a policy for the provider by

$$\pi = \Big((s_0 = 0, x_0 = 0), (s_1, x_{1,0}), (s_2, x_{2,0}, x_{2,1}), \ldots, (s_j, x_{j,0}, \ldots, x_{j,\min(j-1,n)}), \ldots\Big), \tag{1}$$

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

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

which specifies a (possibly infinite) sequence of tuples of introduction time and prices, where we use s_i to denote the time of the ith introduction. For notational convenience, we sometimes summarize the introduction times using vector $\mathbf{s}$; we summarize the corresponding pricing decisions with $\mathbf{x_s}$; and we denote the resulting policy as $\pi(\mathbf{s}, \mathbf{x_s})$.

We next define the customers' decision problem, starting with our assumptions. We assume that customers make decisions myopically, in two senses: we restrict to customers who (i) base their decisions on the current service offerings, rather than on the basis of beliefs about future service offerings; and (ii) make their decision only maximizing their utility in the current period, rather than over their remaining lifetime.[6] We also impose restrictions on possible upgrades for the customers. First, we prohibit any "jump" upgrades, i.e., any upgrades from technology class $j-k$ to j for $k>1$. That is, if a customer does not choose a technology upgrade, then she cannot switch to any of the subsequent upgrades until the end of her customer lifetime. This restriction is necessary to preclude non-monotone allocation rules, which limits the number of cases which need to be considered and allows us to use connections between monotonicity and incentive compatibility to derive results on the optimal pricing. Furthermore, we assume that customers can only upgrade when they were previously already using the latest technology; as we discuss in Section 7 this rules out policies which use unreasonable price discrimination.

Assumption 1. The only possible upgrades for customers are to the current technology from the previous technology.

As an example to motivate this assumption, consider software suites where a user can license a particular version and then use that version forever, or choose the subscription option which allows the user to upgrade as new versions come out.

We assume that each customer at each time period can use a single unit from an available technology class of her choice, or she can opt out. A *newly arriving* customer of type θ at time t simply chooses her preferred quality, $q_1 \in \{0, 1, ...\}$, among the introduced technology classes, so her choice at time t is

$$q_1(\pi, t, \theta) = \underset{i \text{ s.t. } s_i \leq t}{\arg\max}\, \theta \cdot s_i - x_{i,0}. \tag{2}$$

[6] We discuss the myopia assumption and its importance to our results in more detail in Section 7.

Note that $q_1 = 0$ encodes the customer opting out. We assume that customers who don't get service in the current period are not available as customers in future periods.

An *existing* customer in her ℓ-th period, with $2 \le \ell \le d$, will either stay with her previous choice, or pay the switching cost to adopt a new technology(if there is one), so her choice at time t is given as follows:

$$q_\ell(\pi, t, \theta) = \begin{cases} 0, & \text{if } t \le s_1, \\ q_{\ell-1}(\pi, t-1, \theta) + 1, & \text{if } t > s_1 \text{ and } s_{q_{\ell-1}(\pi,t-1,\theta)+1} = t \text{ and} \\ & \quad \theta \cdot s_{q_{\ell-1}(\pi,t-1,\theta)} - x_{q_{\ell-1}(\pi,t-1,\theta), m_{\ell-1}(\pi,t-1,\theta)} \\ & \quad \le \theta \cdot s_{q_{\ell-1}(\pi,t-1,\theta)+1} - x_{q_{\ell-1}(\pi,t-1,\theta)+1, m_{\ell-1}(\pi,t-1,\theta)+1} - c, \\ q_{\ell-1}(\pi, t-1, \theta), & \text{otherwise.} \end{cases} \tag{3}$$

where for $t \ge s_1$ and $1 \le \ell \le \min(t - s_1 + 1, d)$,

$$m_\ell(\pi, t, \theta) := \begin{cases} 0, & \text{if } \ell = 1 \\ m_{\ell-1}(\pi, t-1, \theta) + 1, & \text{if } q_\ell(\pi, t, \theta) > q_{\ell-1}(\pi, t-1, \theta) \\ m_{\ell-1}(\pi, t-1, \theta), & \text{if } q_\ell(\pi, t, \theta) = q_{\ell-1}(\pi, t-1, \theta) \end{cases} \tag{4}$$

tracks the customer's upgrade experience: this is the number of upgrades that a customer of type θ, who is in her ℓ-th period at time t, has switched to as an existing customer under policy π.

In particular, we define $q_\ell(\pi, t, \theta) = 0$ for $t \le s_1$ and $\ell = 2, \ldots, d$. We note that there is an inherent asymmetry between the first $d-1$ periods, counting from the first introduction at time $t = s_1$, and subsequent periods. In period $t = s_1$, only new customers can choose the new technology and necessarily $q_2(\pi, s_1, \theta) = \ldots = q_d(\pi, s_1, \theta) = 0$. In general, in period $t = s_1 + t'$ with $0 \le t' \le d-2$, we have $q_{t'+2}(\pi, t, \theta) = \ldots = q_d(\pi, t, \theta) = 0$. We assume without loss of generality that in case of ties in these definitions, the customer chooses the latest technology class. Specifically, we assume that a new customer who is indifferent between opting out and buying will buy.

We next define the expected revenue and utility of a policy π. The expected revenue of policy π at time t is

$$\text{Revenue}(\pi, t) = \int (x_{q_1(\pi,t,\theta), m_1(\pi,t,\theta)} + x_{q_2(\pi,t,\theta), m_2(\pi,t,\theta)} + \ldots + x_{q_d(\pi,t,\theta), m_d(\pi,t,\theta)}) f(\theta) d\theta.$$

The provider discounts future utility at a rate of $\delta \in (0,1)$ 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

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

where we write $t \in \pi$ as shorthand for t being an introduction time in policy π, i.e., for the existence of some introduction $(s_j, x_{j,0}, \ldots, x_{j,n}) \in \pi$ such that $s_j = t$, with $j \geq 1$. The total cost of policy

$$\pi = \Big((s_0 = 0, x_0 = 0), (s_1, x_{1,0}), (s_2, x_{2,0}, x_{2,1}), \ldots, (s_j, x_{j,0}, \ldots, x_{j,\min(j-1,n)}), \ldots\Big)$$

is

$$\text{Cost}(\pi) = \sum_{t=1}^{\infty} \delta^t Cost(\pi, t) = C \sum_{j \geq 1} \delta^{s_j}. \tag{5}$$

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 utility

$$U(\pi) = \text{Revenue}(\pi) - \text{Cost}(\pi) = \sum_{t=1}^{\infty} \delta^t U(\pi, t). \tag{6}$$

We finally present the provider's optimization problem. The *infinite horizon introduction time and discriminatory pricing problem* is to find a policy that maximizes expected utility, i.e., to find π^* such that

$$\pi^* \in \arg\max_{\pi} U(\pi). \tag{7}$$

We are also interested in a variation of the optimization problem where the introduction times are considered fixed, and we optimize the revenue over the prices. The *infinite horizon discriminatory pricing problem with fixed introduction times* $\mathbf{s}$ is to find prices that maximize the expected revenue given the introduction times $\mathbf{s}$, i.e., to find $\mathbf{x}_{\mathbf{s}}^*$ such that

$$\mathbf{x}_{\mathbf{s}}^* \in \arg\max_{\mathbf{x}_{\mathbf{s}}} \text{Revenue}(\pi(\mathbf{s}, \mathbf{x}_{\mathbf{s}})). \tag{8}$$

4 Optimal Pricing Policy

In this section we characterize the optimal pricing scheme for the infinite horizon discriminatory pricing problem with fixed introduction times. That is, we consider a fixed set of introduction times, and characterize the optimal discriminatory pricing to maximize the expected revenue. We first reduce the optimal pricing problem to a tractable optimization problem (Section 4.2). We then identify key properties of the optimal solution, which lead naturally to an algorithm for optimal discriminatory pricing (Section 4.3). We finally provide numerical evidence for the efficiency of our pricing algorithm (Section 4.4).

4.1 Preliminaries

We start with some definitions and an assumption. We first define p^* as

$$p^* := \arg\max_{\theta \geq 0} (1 - F(\theta)) \cdot \theta.$$

We refer to *Myerson pricing* as the pricing that sets the price at introduction time s_j to $x_j = s_j p^*$, with $j = 1, 2, \ldots$. This is the price that maximizes the one-period expected revenue, assuming the newly introduced technology class is the only product being offered.

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

Assumption 2. The support of the density f is the interval $[0, \zeta]$ (or $[0, \zeta)$ if $\zeta = \infty$) and the function $\frac{1-F(p)}{f(p)}$ is monotonically decreasing on this support.

Assumption 2 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 $k \geq 1$).

We denote the virtual valuation function by

$$v(\theta) := \theta - \frac{1 - F(\theta)}{f(\theta)}$$

and define its inverse (extended to include values too large or small to be in the image of v) as[7]

$$v^{-1}(\gamma) := \min\left(\inf\{\theta \mid v(\theta) \geq \gamma\}, \zeta\right). \tag{9}$$

4.2 The optimization problem

We first argue that the infinite horizon discriminatory pricing problem with fixed introduction times can be decomposed into smaller problems that are decoupled.

We fix $j \geq 1$. We focus on the problem of optimizing revenue from customers who arrive in periods $s_j, \ldots, s_{j+1} - 1$, which we denote with $\text{Revenue}_{[s_j, s_{j+1})}$:

$$\text{Revenue}_{[s_j,s_{j+1})} := \sum_{t=s_j}^{s_{j+1}+d-2} \delta^t \int \left(\sum_{\ell=\max(1,t-s_{j+1}+2)}^{\min(t-s_j+1,d)} x_{q_\ell(\pi,t,\theta), m_\ell(\pi,t,\theta)} \right) f(\theta) d\theta. \tag{10}$$

These customers all face the same set of available technology classes as new customers and, if still in the system, face the same prices when considering whether to upgrade. By Assumption 1, customers can only upgrade if they first adopt the latest technology, and the only relevant prices are those specified in the following observation.

Observation 1. *Revenue*$_{[s_j,s_{j+1})}$ *is a function of prices* $x_{1,0}, x_{2,0}, \ldots, x_{j,0}, x_{j+1,1}, x_{j+2,2}, \ldots, x_{j+n,n}$*, and no other prices.*

We show that with optimal prices newly arriving customers will always choose the latest technology (or nothing), which means that prices of older technology classes targeted to new customers, $x_{i,0}$ for $i < j$, are not relevant. Apart from this initial decision, the only future options that will be offered to these customers under our discriminatory policy are those with prices of the form $x_{j+i,i}$. Therefore, maximizing Revenue(π) given fixed introduction times $s_1, s_2, \ldots$ reduces to maximizing $\text{Revenue}_{[s_j,s_{j+1})}$ separately for each j.

[7] We write $\inf \emptyset = +\infty$.

We next focus on the problem of maximizing $\text{Revenue}_{[s_j,s_{j+1})}$ for a fixed j. We reduce the search for optimal prices to an optimization problem over thresholds θ_i^j for the customer type at which customers switch to more advanced technologies.

Lemma 1. *For fixed $j \geq 1$, consider the optimization problem*

$$\max_{x_{1,0},x_{2,0},\ldots,x_{j,0},x_{j+1,1},x_{j+2,2},\ldots,x_{j+n,n}} \mathit{Revenue}_{[s_j,s_{j+1})}\left(x_{1,0},x_{2,0},\ldots,x_{j,0},x_{j+1,1},x_{j+2,2},\ldots,x_{j+n,n}\right) \tag{11}$$

and the optimization problem

$$\begin{aligned} \max_{\theta_j^j,\theta_{j+1}^j,\ldots,\theta_{j+n}^j} & A_j^j\left(1-F(\theta_j^j)\right)s_j\theta_j^j+\sum_{i=1}^{n}A_{j+i}^j\left(1-F(\theta_{j+i}^j)\right)\left((s_{j+i}-s_{j+i-1})\theta_{j+i}^j-c\right) \\ & s.t.\ p^*=\theta_j^j\leq\theta_{j+1}^j\leq\ldots\leq\theta_{j+n}^j. \end{aligned} \tag{12}$$

There exist constants $A_j^j, A_{j+1}^j, \ldots, A_{j+n}^j$ such that the two optimization problems have equal optimal objective values. The optimal prices for problem (11) *and the optimal thresholds for problem* (12) *are related as follows:*

$$\begin{aligned} x_{i,0}^* &= s_ip^*, & 1\leq i\leq j \\ x_{j+i,i}^* &= x_{j+i-1,i-1}^*+(s_{j+i}-s_{j+i-1})\theta_{j+i}^{j*}-c, & 1\leq i\leq n \end{aligned} \tag{13}$$

All proofs are deferred to the Appendix.

Intuitively, the formulation of problem (12) optimizes over the choice of the minimum value for the customer type θ_i^j that upgrades at the ith introduction. This relies on Assumption 1 to avoid having these cutoffs depend on the pattern of previous upgrades. In turn, the assumption imposes the monotonicity constraint to ensure that only prior upgraders accept future upgrades. The prices then follow to ensure that customers whose type is exactly the chosen cutoff are indifferent about whether to upgrade. In particular, $\theta_j^j = p^*$ because p^* is defined to optimize $(1-F(\theta))\,\theta$. New customers then prefer buying technology class j to buying nothing if $s_j(\theta-p^*)\geq 0$, or $\theta\geq p^*$. New customers prefer technology $j\geq 2$ 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 class, or opt out.

Having reduced optimizing prices to solving optimization problem (12), we now examine its solution. We first address the unconstrained version of the problem and show that it is straightforward to solve. This turns out to solve the full problem in an important special case. Building on the insights from the unconstrained version, we then address the solution of the full problem.

4.2.1 Solving the unconstrained problem

The objective of (12) is nicely separable: there is a separate summand for each θ_i^j, for $j+1 \le i \le j+n$. To optimize $\left(1 - F(\theta_i^j)\right)\left((s_i - s_{i-1})\theta_i^j - c\right)$ we take the derivative with respect to θ_i^j yielding

$$\begin{aligned}&\left(-f(\theta_i^j)\right)\left((s_i - s_{i-1})\theta_i^j - c\right) + \left(1 - F(\theta_i^j)\right)(s_i - s_{i-1})\\ &= (s_i - s_{i-1})\left(1 - F(\theta_i^j) - f(\theta_i^j)\theta_i^j\right) + f(\theta_i^j)c.\end{aligned}$$

The derivative is non-negative if and only if

$$\theta_i^j - \frac{1 - F(\theta_i^j)}{f(\theta_i^j)} \le \frac{c}{s_i - s_{i-1}}. \tag{14}$$

The left hand side is monotone increasing by Assumption 2, meaning that the summand is quasi-concave and maximized by

$$\theta_i^{j,FOC} := v^{-1}\left(\frac{c}{s_i - s_{i-1}}\right). \tag{15}$$

An important implication is that for $n \ge 2$, the optimal solution to problem (12), which we denote θ_j^{j*}, θ_{j+1}^{j*}, $\ldots$, θ_{j+n}^{j*}, satisfies $\left(\theta_{j+1}^{j*}, \ldots, \theta_{j+n}^{j*}\right) = \left(\theta_{j+1}^{j,FOC}, \ldots, \theta_{j+n}^{j,FOC}\right)$ if and only if[8] introduction intervals are non-increasing, i.e.,

$$s_i - s_{i-1} \ge s_{i+1} - s_i, \qquad j+1 \le i \le j+n-1. \tag{16}$$

We return to this class of introduction times, which notably includes policies that introduce with a fixed period, in Section 5, where we argue that good polices are largely of this form.

This class of policies suffices to guarantee that optimal prices take the form of a discount. By the definition of p^*, $v(\theta_i^j) = 0$ for $\theta_i^j = p^*$, and v is increasing at a rate of at least 1 by Assumption 2.

[8]For $n = 1$, the summation in the second summand of (12) has only one term, and $\theta_{j+1}^{j*} = \theta_{j+1}^{j,FOC}$ is satisfied without requiring non-increasing intervals. The case $n = 0$ is trivial. In the sequel, the main case of interest is $n \ge 2$.

Thus we have $p^* \leq \theta_i^{j,FOC} \leq p^* + c/(s_i - s_{i-1})$ for $j+1 \leq i \leq j+n$. Combining this with (13) yields the following observation, which confirms that the pricing for upgraders is in fact a discount.

Observation 2. *If introduction intervals are non-increasing then the optimal prices satisfy*

$$x^*_{j+i,i} \leq x^*_{j+i,0}, \qquad 1 \leq i \leq n.$$

This observation is not true for general introduction times, meaning that in some cases optimal prices for upgrading customers are higher than prices charged to new customers. Working an example requires our upcoming analysis of the constrained optimization problem, so we defer it to Appendix C.

4.2.2 Pricing when introduction intervals are not non-increasing

In the general case there may be some index i, with $j+1 \leq i \leq j+n$, for which $\theta_i^{j*} \neq \theta_i^{j,FOC}$: that is, the unconstrained optimum $\theta_i^{j,FOC}$ is not chosen at the optimal solution. Then, because each of the summands in the objective of problem (12) is quasiconcave, at least one of the following two monotonicity constraints must be binding at the optimal solution: $\theta_{i-1}^j \leq \theta_i^j$, $\theta_i^j \leq \theta_{i+1}^j$. In the sequel, if at the optimal solution we have that $\theta_i^{j*} = \theta_{i+1}^{j*}$ for some i, then we say that thresholds $\theta_i^j, \theta_{i+1}^j$ are "lumped".

If we know which thresholds are lumped, problem (12) can be solved in the same manner as in the unconstrained case. Assume that at the optimal solution, the following thresholds are lumped: $\theta_{i'}^j, \theta_{i'+1}^j, \ldots, \theta_{i''}^j$, with $j+1 \leq i' < i'' \leq j+n$. That is, we have $\theta_{i'}^{j*} = \theta_{i'+1}^{j*} = \ldots = \theta_{i''}^{j*}$, while the remaining θ_i^{j*}'s, with $j+1 \leq i \leq j+n$, are not equal to $\theta_{i'}^{j*} = \ldots = \theta_{i''}^{j*}$. Then the optimal $\theta_{i''}^{j*}$ maximizes the joint term

$$\Big(1 - F(\theta_{i''}^j)\Big) \sum_{i=i'}^{i''} A_i^j \,\Big((s_i - s_{i-1})\,\theta_{i''}^j - c\Big), \tag{17}$$

and therefore satisfies the first-order condition[9]

$$\theta_{i''}^{j*} = v^{-1}\left(\frac{c\sum_{i=i'}^{i''} A_i^j}{\sum_{i=i'}^{i''} A_i^j\,(s_i - s_{i-1})}\right). \tag{18}$$

The challenge for optimization is that there are exponentially many ways in which the thresholds θ_i^j can be lumped. Therefore, we next present an algorithm that efficiently identifies how to optimally lump the thresholds θ_i^j.

4.3 An algorithm for optimal discriminatory pricing

In developing our algorithm we first identify two key properties of the optimal lumping and then specify an algorithm which naturally follows.

4.3.1 Properties of the optimal lumping

Throughout the discussion, we fix introduction index j and focus on optimizing $\text{Revenue}_{[s_j,s_{j+1})}$. Recall that the θ_i^j are monotone. Thus, we formally represent a *lumping* as a subset of the following set of equality constraints: $\{\theta_i^j = \theta_{i+1}^j \mid j+1 \le i < j+n\}$. That is, a lumping is fully defined by which of these $n-1$ constraints are required to be satisfied. Given a lumping $\mathcal{L}^j$ and an index i with $j+1 \le i \le j+n$, we define

$$\mathcal{L}^j(i) := \{\ell \in \{j+1,\ldots,j+n\} : \mathcal{L}^j \vdash \theta_\ell^j = \theta_i^j\} \tag{19}$$

(where $\vdash$ denotes logical entailment) to be the set of all indices whose corresponding thresholds are lumped together with θ_i^j in lumping $\mathcal{L}^j$. In the sequel, we suppress the superscript j from notation $\mathcal{L}^j$ for simplicity. We use $\theta_{\mathcal{L}(i)}^{j,FOC}$ to denote the solution to the first-order condition

$$\theta_{\mathcal{L}(i)}^{j,FOC} := v^{-1}\left(\frac{c\sum_{\ell\in\mathcal{L}(i)} A_\ell^j}{\sum_{\ell\in\mathcal{L}(i)} A_\ell^j\,(s_\ell - s_{\ell-1})}\right), \tag{20}$$

which is a translation of (18) to our lumping formalism.

[9] Comparing the condition in (18) against the simple first-order condition in (15), we note that (18) collapses to (15) when no θ_i^j's are lumped together.

Our first key property essentially states that when lumping terms together, the maximizer of the new joint term lies somewhere "in the middle" of the maximizers of the terms that make up the lumping. This is a direct consequence of a quasiconcavity property implied by Assumption 2 and applies even if some of these terms have already been lumped together, which is why the statement of the lemma includes two lumpings rather than one.

Lemma 2. *Let lumpings $\mathcal{L}, \mathcal{L}'$ satisfy $\mathcal{L} \subseteq \mathcal{L}'$. Then*

$$\min_{\iota \in \mathcal{L}'(i)} \theta^{j,FOC}_{\mathcal{L}(\iota)} \leq \theta^{j,FOC}_{\mathcal{L}'(i)} \leq \max_{\iota \in \mathcal{L}'(i)} \theta^{j,FOC}_{\mathcal{L}(\iota)}, \qquad \forall j+1 \leq i \leq j+n.$$

This leads to our second key property, which establishes a condition under which two terms must be lumped together. Intuitively, if we find the largest θ^j_i which is out of order then by Lemma 2 there is no way to fix this without adding it to the lumping. In the statement of the lemma, we denote an optimal lumping for optimization problem (12) by $\mathcal{L}^*$.

Lemma 3. *Assume that a lumping $\mathcal{L}$ satisfies $\mathcal{L} \subseteq \mathcal{L}^*$. Let $k^* := \arg\max_{k \in S} \theta^{j,FOC}_{\mathcal{L}(j+k)}$ where $S = \{k : \exists k' \text{ s.t. } 1 \leq k < k' \leq n \wedge \theta^{j,FOC}_{\mathcal{L}(j+k)} > \theta^{j,FOC}_{\mathcal{L}(j+k')}\}$, with $k^* := -\infty$ if $S = \emptyset$. Then if $k^* > -\infty$, it holds that at the optimal solution $\theta^{j*}_{j+k^*} = \theta^{j*}_{j+k^*+1}$.*

We now interpret Lemma 3. Start with a lumping $\mathcal{L}$ that has a subset of the constraints in $\mathcal{L}^*$: that is, all the θ^j_i's that are lumped together in $\mathcal{L}$, are also lumped together in the optimal solution. Identify $\theta^j_{j+k^*}$: the θ^j_i with the largest $\theta^{j,FOC}_{\mathcal{L}(i)}$ which is out of the desired order $\theta^j_{\mathcal{L}(j+1)} \leq \ldots \leq \theta^j_{\mathcal{L}(j+n)}$. The lemma says that such $\theta^j_{j+k^*}$, if it exists, is lumped with $\theta^j_{j+k^*+1}$ at the optimal solution.

4.3.2 The algorithm

Lemma 3 motivates naturally the following algorithm for identifying an optimal lumping: (i) start with a lumping that is a subset of the optimal lumping; (ii) identify the largest first-order condition solution that is out of the desired order, $\theta^j_{j+k^*}$; (iii) update the lumping by adding the constraint $\theta^j_{j+k^*} = \theta^j_{j+k^*+1}$; (iv) repeat, starting from the updated lumping.

Having fixed an introduction index j, the OptimalLumping algorithm (Algorithm 1) executes this idea in order to produce an optimal lumping of the upgrade thresholds θ^j_i's for optimizing

Revenue$_{[s_j,s_{j+1})}$, i.e., the revenue from customers who arrive in periods $s_j, \dots, s_{j+1}-1$. Rather than repeatedly doing calculations of the form $\theta_i^j = v^{-1}(\gamma_i^j)$, the algorithm operates directly on the right-hand side terms γ_i^j's of first-order conditions of the form $\theta_i^j - \frac{1-F(\theta_i^j)}{f(\theta_i^j)} = \gamma_i^j$, and not on thresholds θ_i^j's. Because $v^{-1}(\gamma)$ is non-decreasing in γ by Assumption 2, the two are equivalent.[10]

As its output, the algorithm produces a set of n terms (indexed $\gamma_{j+1}^j, \dots, \gamma_{j+n}^j$), which are then fed into the DiscriminatoryPricing algorithm (Algorithm 2) as the right-hand side terms for first-order conditions of the form $\theta_i^j - \frac{1-F(\theta_i^j)}{f(\theta_i^j)} = \gamma_i^j$. In particular, the DiscriminatoryPricing algorithm calls the OptimalLumping routine and solves the first-order condition equations to produce the optimal upgrade thresholds θ_i^{j*}'s, which it then converts to optimal prices using (13).

Algorithm 1 OptimalLumping Algorithm

Input: Introduction times $(s_1, s_2, \dots,)$, current introduction j, maximum customer upgrade experience n, switching cost c, objective coefficients A_i^j

$\mathcal{L} \leftarrow \emptyset$
for $i = j+1, \dots, j+n$ **do**
  $\gamma_i^j \leftarrow \frac{c}{s_i - s_{i-1}}$
end for
$S \leftarrow \{k : \exists k' \ s.t.\ 1 \le k < k' \le n \wedge \gamma_{j+k}^j > \gamma_{j+k'}^j\}$

$$k^* \leftarrow \begin{cases} -\infty, & \text{if } S = \emptyset \\ \arg\max_{k \in S} \gamma_{j+k}^j, & \text{otherwise} \end{cases}$$

while $k^* > 0$ **do**
  $\mathcal{L} \leftarrow \mathcal{L} \bigcup \{\theta_{j+k^*}^j = \theta_{j+k^*+1}^j\}$
  for $\iota \in \mathcal{L}(j+k^*)$ **do**
    $\gamma_\iota^j \leftarrow \frac{c \sum_{\ell \in \mathcal{L}(j+k^*)} A_\ell^j}{\sum_{\ell \in \mathcal{L}(j+k^*)} A_\ell^j (s_\ell - s_{\ell-1})}$
  end for
  $S \leftarrow \{k : \exists k' \ s.t.\ 1 \le k < k' \le n \wedge \gamma_{j+k}^j > \gamma_{j+k'}^j\}$

$$k^* \leftarrow \begin{cases} -\infty, & \text{if } S = \emptyset \\ \arg\max_{k \in S} \gamma_{j+k}^j, & \text{otherwise} \end{cases}$$

end while
return $\gamma_{j+1}^j, \dots, \gamma_{j+n}^j$

[10] Interestingly, this means that the optimal lumping is independent of the virtual valuation function v (and thus the customer type distribution F).

Algorithm 2 DiscriminatoryPricing Algorithm

Input: Introduction times $(s_1, s_2, \dots,)$, maximum customer upgrade experience n, customer type distribution f, switching cost c, discount factor δ, customer lifetime d

for $j = 1, \dots,$ **do**
  Calculate $A_i^j, i = j+1, \dots, j+n$ according to (38)
  $\gamma_{j+1}^j, \dots, \gamma_{j+n}^j \leftarrow \text{OptimalLumping}\left((s_1, s_2, \dots,), j, n, c, A_i^j\right)$
  $x_{j,0} \leftarrow s_j p^*$
  for $i = 1, \dots, n$ **do**
    $\theta_{j+i}^j \leftarrow v^{-1}\left(\gamma_{j+i}^j\right)$
    Set price $x_{j+i,i} \leftarrow x_{j+i-1,i-1} + (s_{j+i} - s_{j+i-1})\,\theta_{j+i}^j - c$
  end for
end for
return Prices $(x_{j,0}, x_{j+1,1}, \dots, x_{j+n,n})_{j=1,\dots}$

When an iteration of the OptimalLumping algorithm lumps together some thresholds θ_i^j, then these θ_i^j's are truly lumped together in the optimal solution, by Lemma 3. This, in turn, implies that the DiscriminatoryPricing algorithm correctly produces optimal prices for the infinite horizon discriminatory pricing problem with fixed introduction times.

Theorem 1. *The DiscriminatoryPricing algorithm produces optimal prices for the infinite horizon discriminatory pricing problem with fixed introduction times given by* (8)*.*

4.4 Efficient computation of optimal prices

We note that the OptimalLumping Algorithm (Algorithm 1) has time complexity that grows as $\mathcal{O}(n^2)$, where n is the maximum customer upgrade experience. In turn, the DiscriminatoryPricing algorithm (Algorithm 2) has linear time complexity in the number of introductions (indexed by j). We provide an implementation of these algorithms in order to compute optimal prices for given introduction times. The proposed implementation is practical, as it can compute optimal prices within a fraction of a second even for large problem instances.

We provide some numerical results to show that the proposed implementation computes optimal prices efficiently. We fix the discount rate δ and the switching cost c, and vary the customer lifetime, the end of the horizon, as well as the given introduction times. We run 1000 simulations for each distinct setting. In each simulation we randomly generate introduction intervals, with introductions

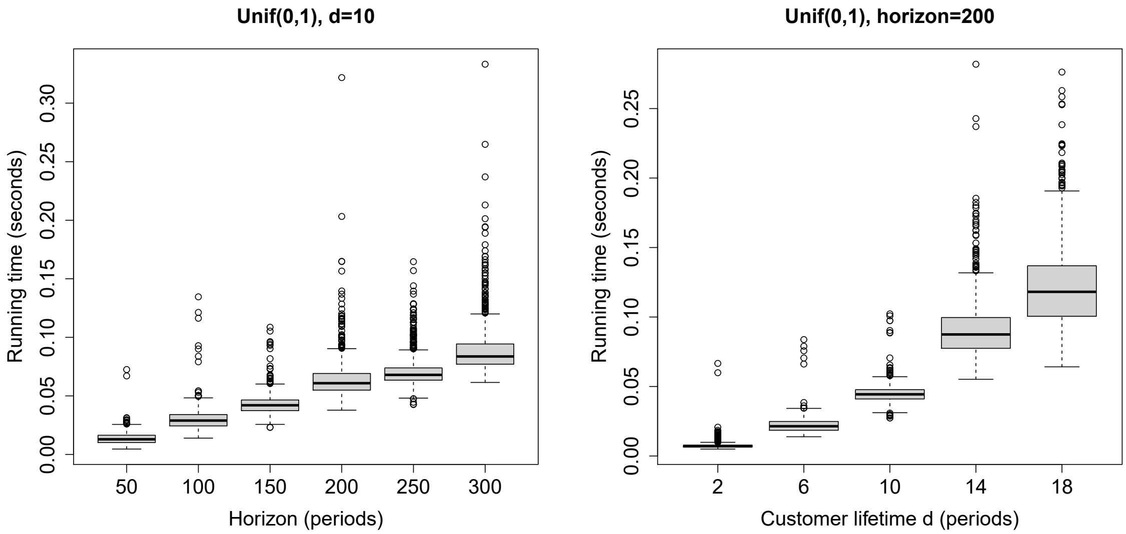

Figure 1: Box plots for the running time required to compute optimal prices for the entire horizon, over 1000 simulations, against the length of the horizon and customer lifetime d, for switching cost $c = 0.5$, for discount rate $\delta = 0.9$, for the uniform distribution on $[0, 1]$. For the left subfigure, d is set to 10; for the right subfigure, the horizon is set to 200 periods. The experiments were run using R version 4.3 on a 2.5 GHz Intel Core i7 processor.

up to 20 periods apart, up to the end of the horizon. For each simulation we keep track of the running time for computing the optimal prices for *all* technology classes within the horizon, i.e., the running time of the DiscriminatoryPricing algorithm when truncated at the end of horizon. Figure 1 summarizes running times using box plots that capture the outcomes of each set of 1000 simulations. Our optimal pricing algorithm can compute an optimal set of introduction and upgrade prices for the entire horizon in a fraction of a second, even for large horizons and for large values of the customer lifetime d.

5 Optimal Introduction Times

Having characterized optimal discriminatory pricing given fixed introduction times, in this section we focus on optimal introduction times. Overall, we show that policies with periodic introduction times after an initial “warm-up” phase, combined with optimal pricing, enjoy optimality guarantees. First, we show the optimality of periodic introduction intervals after an initial warm-up phase, when the technology introductions are constrained to have non-increasing intervals. We characterize the optimal period as a function of the details of the setting (customer type distribution, switching cost, provisioning cost, discount factor). Secondly, when we allow general introduction intervals, we show that periodic introduction intervals after some time are optimal in a more restricted sense. Finally, we detail numerical experiments that suggest that, when allowing for general introduction intervals, it is optimal to have non-increasing introduction intervals during a warm-up phase, and periodic introduction intervals after that phase, with a possible reduction of the length of the intervals between the warm-up and the continuation.

5.1 Optimality of periodic introductions under non-increasing introduction intervals

In this subsection we assume that introduction intervals are non-increasing and show that, after the initial warm-up phase, periodic introductions are optimal. The initial warm-up phase corresponds to the early stages of the process, when the mass of customers in the system is still building up (because we are less than $d-1$ periods after the first introduction, and therefore less than a mass d of customers have arrived since and including the first introduction time). We show that once this phase has ended, periodic introductions from then on are optimal.

Specifically, we assume

Assumption 3. Introduction intervals are non-increasing; that is, $s_i - s_{i-1} \geq s_{i+1} - s_i$ for all $i \geq 2$.

We also assume that all prices are calculated optimally given the introduction times, hence any policy is fully characterized by the introduction times.[11]

We now state our theorem on optimal introduction times.

Theorem 2. *Assume that technology introductions are restricted to have non-increasing intervals (Assumption 3). Then there is an optimal policy with the following property: all introductions made at or after time $t = s_1 + d - 1$ are periodic, with each introduction a constant interval from its previous introduction.*

We highlight the main intuition underlying this result. To do so, we first introduce some notation. Given a policy π with introduction times (s_i) and a fixed $j \geq 1$, define $\pi_{s_j} = ((s_0 = 0, x_0 =$

[11] An optimal policy in this restricted policy space exists, i.e., there is some π^* that introduces at times $(s_1^*, s_2^*, \ldots)$ such that $U(\pi^*) = \sup_\pi U(\pi)$, where $U(\pi^*)$ is finite.

$0), (s_1, \mathbf{x}_1^*), \ldots, (s_j, \mathbf{x}_j^*))$ to be the finite (truncated) policy that has introduction times $(s_1, ..., s_j)$, where s_j is its last introduction, and where the prices $\mathbf{x}_i^* = \left(x_{i,0}^*, \ldots, x_{i,\min(i-1,n)}^*\right)$, $i = 1, \ldots, j$, are optimal prices given the introduction times. Assuming non-increasing introduction intervals, we express the excess utility of policy $\pi_{s_{j+1}}$ (which makes one more introduction, at some time s_{j+1}) over policy π_{s_j} as

$$U(\pi_{s_{j+1}}) - U(\pi_{s_j}) = \delta^{s_{j+1}} \left(g(s_{j+1} - s_j) - C\right), \tag{21}$$

for some function g that captures the difference in the revenue. Function g depends on introduction times only through the difference $s_{j+1} - s_j$, and also depends on the details of the problem setting (customer type distribution F, customer lifetime d, switching cost c, discount factor δ.) The utility of a policy π can be thought of in terms of the excess utility of the policy over the utility $U(\pi_{s_j})$:

$$U(\pi) = U(\pi_{s_j}) + \underbrace{\sum_{k=j+1}^{\infty} \delta^{s_k} \left(g(s_k - s_{k-1}) - C\right)}_{\text{excess utility}}. \tag{22}$$

We show that this excess utility is maximized for periodic introduction intervals.

We note that when customers switch to upgraded technologies as existing customers at most once, i.e., $n = 1$, then a stronger version of Theorem 2 holds: periodic introductions after time $s_1 + d - 1$ are optimal, without restricting to non-increasing introductions.

5.1.1 Characterizing the optimal period

We now characterize the optimal period. The proof of Theorem 2 shows that the excess revenue of policy $\pi_{s_{j+1}}$ over policy π_{s_j}, shown in (21), can be written in terms of the function

$$g(z) := \frac{1-\delta^d}{(1-\delta)^2} \left(1 - F(p^*)\right) z p^* + \frac{\delta^d + d(1-\delta) - 1}{(1-\delta)^2} \left(1 - F(\theta^*(z))\right) \left(z\theta^*(z) - c\right), \quad z \in \mathbb{N}^+. \tag{23}$$

where $\theta^*(z)$ denotes the solution to the first-order condition

$$\theta^*(z) := v^{-1}\left(\frac{c}{z}\right). \tag{24}$$

We also define

$$h(z) := g(z) - C, \quad z \in \mathbb{N}^+. \tag{25}$$

The following corollary of Theorem 2 characterizes the optimal periodicity after the initial warm-up phase.

Corollary 1. *Assume that technology introductions are restricted to have non-increasing intervals (Assumption 3). Then there is an optimal policy with the following property: all introductions made at or after time* $t = s_1 + d - 1$ *are at a time which is an interval of*

$$T^* \in \arg\max_{z \in \mathbb{N}^+} \frac{\delta^z}{1-\delta^z} h(z) \tag{26}$$

after their previous introduction.

We make some observations related to the characterization of T^*. First, $h(z)$ is naturally extended from the positive integers to the positive reals. We note that $\frac{\delta^z}{1-\delta^z} h(z)$ is bounded and therefore has a finite maximum, because function $h(z)$ is bounded above by a linear function of z for fixed δ, d, c, C and distribution F.[12] Lastly, at a critical point z^* where the derivative of $\frac{\delta^z}{1-\delta^z} h(z)$ equals zero, we have that

$$\log \delta \cdot h(z^*) + (1 - \delta^{z^*}) h'(z^*) = 0. \tag{27}$$

More details on the derivatives of $\frac{\delta^z}{1-\delta^z} h(z)$ are given in Appendix I.

The example of the uniform distribution on $[0, 1]$**.** To illustrate the behavior of the optimal periodicity as a function of the problem parameters, we provide further details for the specific case when F is the uniform distribution on $[0, 1]$. Then[13] we have $p^* = 1/2$ and $\theta^*(z) =$

[12]We also remark that if the set $\arg\max_{z \in \mathbb{N}^+} \frac{\delta^z}{1-\delta^z} h(z)$ has more than one element, then choosing any of the elements in the set as the period would be optimal. However, setting the introduction intervals to alternate between different elements of the set does not guarantee optimality (in particular, cycling through the elements of the set would break the assumption of non-increasing intervals).

[13]We provide more details for the $U[0, 1]$ case in Appendix I. For convenience, we repeat here the expression

$$\begin{aligned} h_{U[0,1]}(z) &= g_{U[0,1]}(z) - C \\ &= \begin{cases} \frac{1-\delta^d}{(1-\delta)^2} \cdot \frac{z}{4} + \frac{\delta^d + d(1-\delta) - 1}{(1-\delta)^2} \cdot \frac{(z-c)^2}{4z} - C & \text{if } z > c \\ \frac{1-\delta^d}{(1-\delta)^2} \cdot \frac{z}{4} - C & \text{if } z \leq c. \end{cases} \end{aligned}$$

$\min\left(\frac{1}{2}\left(\frac{c}{z}+1\right), 1\right)$. Notice that for $z \leq c$, $\theta^*(z) = 1$, meaning customers never upgrade: they just stick with the latest technology they were offered at their arrival time.

In general, $\frac{\delta^z}{1-\delta^z} h_{U[0,1]}(z)$ can be multi-modal. For $z \leq c$, $\frac{\delta^z}{1-\delta^z} h_{U[0,1]}(z)$ is either decreasing or has a local maximum at $z^* < c$ satisfying (27). Similarly for $z > c$, $\frac{\delta^z}{1-\delta^z} h_{U[0,1]}(z)$ can be decreasing or have a local maximum at $z^* > c$ satisfying (27). As a consequence, we find that three possible scenarios can play out, depending on the parameter values for d, c, C, δ. We describe these scenarios qualitatively below, and provide specific examples in Appendix I and Figure 4 therein.

1. $T^* = 1$, where introductions are made as frequently as possible. This happens, for example, when c, C are both small. In general, for small provisioning cost C there is an incentive to introduce frequently. If c is also small enough, then it is optimal to put $T^* = 1$ (i.e., introduce at every period), where almost all customers would always upgrade.

   Furthermore, if C is sufficiently small and c large, then it can still be optimal to introduce as frequently as possible ($T^* = 1$), but now $T^* < c$, hence no customer ever upgrades.

2. $1 < T^* \leq c$. No customer upgrades, but now introductions are delayed to offset introduction costs. This happens, for example, when C is small but larger than 1, c is large (larger than C), and δ is neither too large nor too small. The local maximum for $z \leq c$ ($T^* < c$) exceeds that for $z > c$.

3. $T^* > \max(c, 1)$. Some customers upgrade, and introductions are delayed to offset introduction costs. In this case the local maximum for $z > c$ dominates that for $z \leq c$, and hence some customers upgrade.

   $T^* > c \geq 1$ can happen, for instance, when $C \geq c \geq 1$: the provisioning cost to the provider is greater than or equal to the switching cost to the customer.

   Interestingly, even for quite small provisioning cost C, the optimal period T^* can be greater than 1, i.e., it may be best not to introduce as frequently as possible, and we have $T^* > 1 \geq c$.

We note that we can have a situation where different values of the period, $T_1^* < c < T_2^*$ yield near identical values of the function $\frac{\delta^z}{1-\delta^z} h(z)$, where the corresponding real-valued local maxima

z_1^*, z_2^* yield near identical values (Figure 5 in Appendix I).

We see similar behavior when the customer type has a beta distribution (with shape parameters larger than 1). When the distribution F has infinite support, such as for the exponential or gamma distributions, then some customers will always upgrade.

5.2 Optimality of periodic introductions under general introduction intervals

If we allow general introduction intervals (e.g., increasing intervals), we can demonstrate that periodic intervals after some time are optimal in a more restricted sense. Recall that, given a policy π with introduction times (s_i), π_{s_j} refers to a policy that has introduction times $(s_1, ..., s_j)$ (and none after that) and optimal prices given these introductions.

Theorem 3. *Fix $j \geq 2$ and an arbitrary sequence of introduction times $s_1, ..., s_j$ with $s_j \geq s_1 + d - 1$. Fix time $\tilde{t}$ such that $\tilde{t} \geq s_j + d - 1$ and $\tilde{t} = s_j + mT^*$ for some m. Consider the policy $\hat{\pi}$ with introduction times*

$$(\hat{s}_1 = s_1, ..., \hat{s}_{j-1} = s_{j-1}, \hat{s}_j = s_j, \hat{s}_{j+1} = s_j + T^*,$$
$$\hat{s}_{j+2} = s_j + 2T^*, \ldots, \hat{s}_{j+m} = \tilde{t} = s_j + mT^*, \hat{s}_{j+m+1} = s_j + (m+1)T^*, \ldots)$$

and optimal pricing. Then

$$U(\hat{\pi}) - U(\hat{\pi}_{\tilde{t}}) \geq U(\pi') - U(\pi'_{\tilde{t}}) \tag{28}$$

for any policy π' with introduction times $(s'_1 = s_1, \ldots, s'_{j-1} = s_{j-1}, s'_j = s_j, s'_{j+1}, \ldots, s'_{k-1}, s'_k = \tilde{t}, s'_{k+1}, \ldots)$ for some $k > j$, and optimal pricing. Furthermore, the policy $\hat{\pi}$ satisfies

$$\hat{\pi} \in \arg\max_{\pi} \sum_{t=\tilde{t}}^{\infty} \delta^t U(\pi, t),$$

where the supremum is over all policies π which have the first j introduction times fixed at $(s_1, \ldots, s_j)$ and a later introduction at time $\tilde{t}$.

The interpretation of Theorem 3 is that, having fixed the first j introduction times and also an introduction at time $\tilde{t}$ at least $d-1$ periods after the jth introduction time, a policy with periodic introductions starting from the jth introduction time optimizes the total discounted utility from time $\tilde{t}$ onwards. In particular, take any policy π' with the first j introductions $(s_1, \ldots, s_j)$ fixed, that also makes an introduction at time $\tilde{t}$, where $\tilde{t}$ is at least $d-1$ periods after s_j. Consider the policy $\hat{\pi}$ that starts with introductions at $(s_1, \ldots, s_j)$ and is periodic starting at time s_j with period T^*, including having an introduction at time $\tilde{t}$. Compared to π', policy $\hat{\pi}$ has (i) at least as large additional utility over the truncated policy that stops introducing at introduction time $\tilde{t}$; and (ii) at least as large utility from time $\tilde{t}$ onward.

Note that this doesn't imply that $\hat{\pi}$ is optimal from time s_1 onwards (given initial introductions $(s_1, \ldots, s_j)$); it only implies that $\hat{\pi}$ is optimal for time $\tilde{t}$ onward. Theorem 3 doesn't rule out the existence of a policy $\bar{\pi}$ for which $U(\bar{\pi}) - U(\bar{\pi}_{s_j}) > U(\hat{\pi}) - U(\hat{\pi}_{s_j})$, as it says nothing about the utility between times s_j and $s_j + d - 2$.

The restriction that the policy $\hat{\pi}$ has an introduction at time $s_j + mT^*$ for some m is essentially without loss of generality. Fixing some arbitrary $\tilde{t}$ with $\tilde{t} \geq s_j + d - 1$, for any comparison policy π' with introductions at times $(s'_1 = s_1, \ldots, s'_j = s_j, s'_{j+1}, \ldots, s'_k = \tilde{t}, \ldots)$ for some k, we can redefine $\hat{\pi}$ to be any policy that has non-increasing introductions starting from period s_j, an introduction at period $\tilde{t}$, and then periodic introductions from $\tilde{t}$ onward with period T^*. In particular, we state the following result, whose proof mirrors the proof of Theorem 3:

Corollary 2. *Fix $j \geq 2$ and introduction times s_1 and s_j with $s_j \geq s_1 + d - 1$. Fix time $\tilde{t}$ such that $\tilde{t} \geq s_j + d - 1$. Fix $m \geq 1$ and consider the policy $\hat{\pi}$ with introduction times $(\hat{s}_1 = s_1, \hat{s}_2, \ldots, \hat{s}_{j-1}, \hat{s}_j = s_j, \hat{s}_{j+1}, \ldots, \hat{s}_{j+m-1}, \hat{s}_{j+m} = \tilde{t}, s_{j+m+1} = \tilde{t} + T^*, s_{j+m+2} = \tilde{t} + 2T^*, \ldots)$, where the intervals of introduction times $(\hat{s}_j, \hat{s}_{j+1}, \ldots, \hat{s}_{j+m})$ are non-increasing, and pricing is optimal. Then*

$$U(\hat{\pi}) - U(\hat{\pi}_{\tilde{t}}) \geq U(\pi') - U(\pi'_{\tilde{t}}) \tag{29}$$

for any policy π' with introduction times $(s'_1 = s_1, s'_2, \ldots, s'_{k-1}, s'_k = \tilde{t}, s'_{k+1}, \ldots)$ which has a first introduction at s_1 and an introduction at $s'_k = \tilde{t}$ for some $k \geq 2$, and optimal pricing. Furthermore, *the policy $\hat{\pi}$ satisfies*

$$\hat{\pi} \in \arg\max_{\pi} \sum_{t=\tilde{t}}^{\infty} \delta^t U(\pi, t),$$

where the supremum is over all polices that have an introduction at s_1 and a later introduction at time $\tilde{t}$.

The interpretation of Corollary 2 is that, having fixed the first introduction time and a time at least $2d-2$ periods after the first introduction, a policy with periodic introductions from that latter time on optimizes the total discounted utility from that time onward. In particular, take any policy π' with its first introduction s_1 fixed and also an introduction at time $\tilde{t}$, where $\tilde{t}$ is at least $2d-2$ periods after starting time s_1. Now consider the policy $\hat{\pi}$ which starts at time s_1, has non-increasing introductions for at least $d-1$ periods before $\tilde{t}$, and then is periodic starting at $\tilde{t}$ with period T^*. Compared to π', policy $\hat{\pi}$ has (i) at least as large additional utility over the truncated policy that stops introducing at introduction time $\tilde{t}$, and (ii) at least as large utility from time $\tilde{t}$ onward.

Note that this doesn't imply that the policy $\hat{\pi}$ is optimal from time s_1 onwards, only that it is optimal for time $\tilde{t}$ onwards. Neither does it imply that given an arbitrary set of fixed introductions up to time $\tilde{t}$, that it is then optimal to have periodic introductions from time $\tilde{t}$ on. Instead, the result is about the optimality, from time $\tilde{t}$ onwards, of a policy that has non-increasing introductions for at least $d-1$ periods before time $\tilde{t}$, and then periodic introductions starting at $\tilde{t}$.

5.3 Numerical experiments

We have shown that periodic introductions after a warm-up phase are optimal under non-increasing introduction intervals. We have also shown a more restricted optimality guarantee for periodic introduction intervals when allowing for general introduction intervals. In this subsection, we present numerical experiments in order to answer two questions: (i) Is there an introduction policy that can outperform periodic introductions? (ii) How should the provider best time introductions in the warm-up phase?

5.3.1 Experimental setup

We refer to a combination of customer type distribution F, customer lifetime d, provisioning cost C, switching cost c, and discount rate δ, as an (experimental) problem instance. For each problem instance, we sample patterns of introduction times according to a set of introduction pattern generating schemes. These sampling schemes are the following: random introduction intervals (*Random*); periodic introductions (*Per*); non-increasing introduction intervals (*NI*); non-decreasing introduction intervals (*ND*); non-increasing introduction intervals in the warm-up, periodic introductions after (*NI-Per*); non-increasing introductions in the warm-up, non-increasing introductions after (*NI-NI*); non-decreasing introduction intervals in the warm-up, periodic introductions after (*ND-Per*); non-decreasing introduction intervals in the warm-up, non-decreasing introduction intervals after (*ND-ND*); and introduction intervals from grid search (*Grid* and *Log-Grid*). The details of the pattern generating schemes are provided in Appendix K. We fix the first introduction as $s_1 = 1$ in all the generated introduction patterns.

For a given experimental problem instance, for each introduction pattern generating scheme, we retain the introduction pattern that achieves the highest utility out of all the generated introduction patterns from that scheme. We use the *NI-Per* scheme as the benchmark against which we compare the utilities achieved by the other schemes. The end of horizon is set at 200, meaning we only calculate utility accumulated until period 200. We provide results for the uniform distribution on $[0, 1]$, while the results for 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 $\theta = 0.25$, are similar.

5.3.2 Results

For a given problem instance and a given introduction pattern generating scheme, we are interested in the utility ratio

$$\frac{\text{best utility achieved by scheme}}{\text{best utility achieved by } \textit{NI-Per}}. \tag{30}$$

For a given introduction pattern we calculate the generated utility as follows. We first optimize the pricing for the given introduction times using the DiscriminatoryPricing algorithm (Algorithm 2). We then use the optimized prices to calculate the ensuing revenue by (10). We use the introduction times to calculate the ensuing cost by (5).[14] Finally we calculate the utility by (6).

Figure 2 summarizes the results for all the described introduction pattern generating schemes, over several problem instances, for a customer type distribution that is uniform on $[0, 1]$. The figure shows a box plot for the comparison of each introduction pattern generating scheme against the benchmark *NI-Per* scheme. Each box plot summarizes the values for the utility ratio in (30) for the respective scheme, across the considered experimental problem instances.

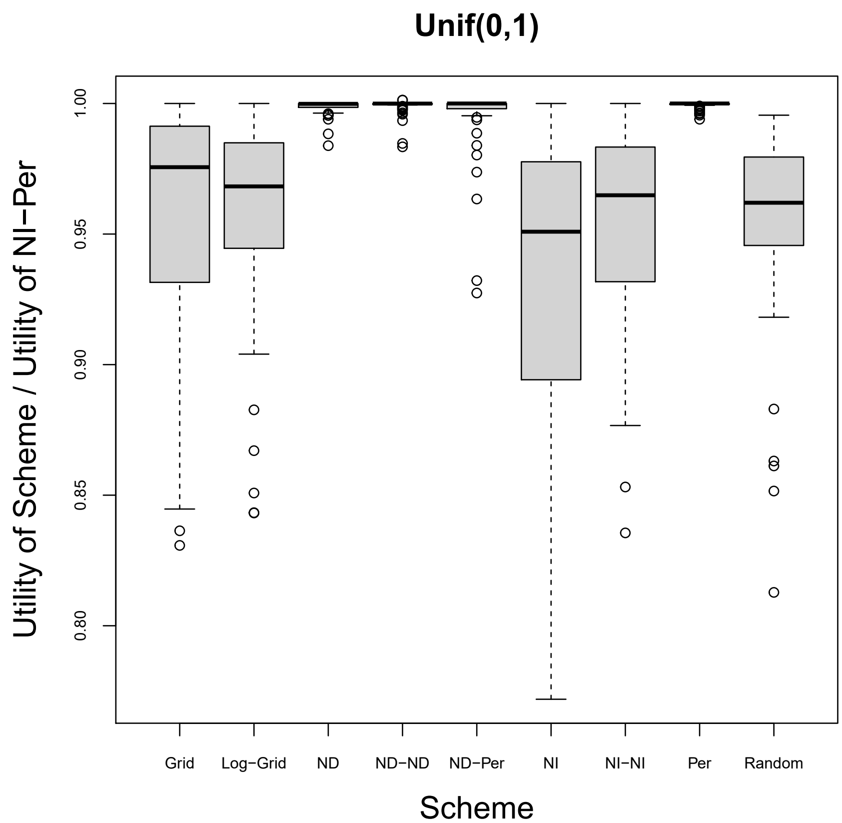

Figure 2: Box plots for the ratio between the best utility achieved by each introduction pattern generating scheme and the best utility achieved by the *NI-Per* benchmark, across different experimental problem instances. Each box plot summarizes 60 utility ratio values: as many as the combinations of three values for customer lifetime d (6,10,14), four values for provisioning cost C (1, 5, 10, 20), and five values for switching cost c (0.5, 1, 2, 5, 10). The discount rate is fixed at $\delta = 0.9$ and the customer type distribution F is the uniform distribution on $[0, 1]$.

[14] For both the revenue and the cost calculation, we truncate time at the end of horizon, which in our experiments is set at 200.

Essentially no pattern generating scheme achieves a better utility than the best policy with periodic introductions after the warm-up phase, for any problem instance.[15] Two of the schemes that don't impose periodicity have the exact same introduction pattern as the best policy with periodic introductions after the warm-up, and therefore achieve a utility ratio against *NI-Per* equal to one, for some experimental instances (10 out of 60 for *Grid*, eight out of 60 for *Log-Grid*). The *NI-Per* scheme outperforms these two schemes for all other instances.

Four other introduction pattern generating schemes that don't impose periodicity have a utility ratio against the best *NI-Per* pattern that is only slightly smaller than one:

$$1 - 10^{-9} \leq \frac{\text{best utility achieved by scheme}}{\text{best utility achieved by } \textit{NI-Per}} < 1,$$

for some experimental problem instances (seven instances out of 60 for *ND-ND*; six instances for *ND*; and three instances for each of *NI* and *NI-NI*). Nevertheless, for the vast majority of experimental problem instances, the best *NI-Per* introduction pattern does better than the best patterns from these schemes by more than a small margin.[16]

Having established that it is optimal to have periodic introduction intervals after the warm-up in the scenarios covered by our experimental setup, we now look closer at the optimal introduction intervals to use *during* the warm-up. The concrete question of interest here is the following: out of the three schemes that impose periodic introductions after the warm-up (*NI-Per*, *Per*, and *ND-Per*), which one attains the highest utility? Our numerical results show that in general it is optimal to have non-increasing introduction intervals in the warm-up phase.[17] In 43 out of 60 experimental

[15]For 11 generating scheme-problem instance tuplets, out of 540 total, the utility ratio in (30) is actually slightly larger than one, exceeding one by a decimal that has its first non-zero at the 10th decimal digit or after. For these tuplets, the earliest discrepancy compared to the best identified *NI-Per* introduction pattern occurs late in the horizon (at period 169 at the earliest, and mostly at or after period 190). For these 11 cases, the utility ratio is greater than one due to end of horizon effects.

For two other scheme-problem instance tuplets, the utility ratio is around 1.0013. These tuplets are $\{\textit{ND-ND}, d = 14, c = 0.5, C = 20\}$ and $\{\textit{ND-ND}, d = 14, c = 1, C = 10\}$. For each of these there is an *NI-Per*-consistent introduction pattern that beats the corresponding *ND-ND* pattern, and that was not identified by our *NI-Per* scheme in the reported experiment. Recall that the *NI-Per* scheme searches the policy space through sampling, and therefore it may miss the optimal introduction pattern.

[16]As a side note, we observe that in general the best *ND* patterns do well, and in particular generally better than the best *NI* patterns. A main reason for this observation is that the shape of the marginal revenue for the particular examples considered here make it such that introducing too late hurts less than introducing too soon; many *NI* sampled patterns randomly choose to introduce too soon and get stuck with short intervals forever, which is bad, while the consequences for similar behavior by *ND* patterns are less severe.

[17]In detail, we report the following results for the comparison between the utilities of the best introduction pattern

problem instances, the best *NI-Per* introduction pattern is identical to the best periodic (*Per*) pattern, and therefore has the same periodicity in the warm-up phase as it does in the continuation. However, in the remaining 17 instances the best *NI-Per* introduction pattern has at least one introduction interval that is shorter than its previous interval — i.e., it has at least one instance of decreasing intervals. For these 17 latter parameter settings, the best *NI-Per* introduction pattern, which has at least one instance of decreasing intervals in the warm-up phase and periodic intervals afterwards, attains strictly higher utility than the best periodic introduction pattern, which introduces at constant intervals throughout. Furthermore, in 36 out of 60 experimental problem instances, the best *NI-Per* introduction pattern is identical to the best *ND-Per* pattern — and in 24 out of these 36 instances, the best *NI-Per* and *ND-Per* introduction patterns coincide with the best fully periodic (*Per*) pattern. However, in 22 instances the best *NI-Per* introduction pattern attains strictly higher utility than the best *ND-Per* introduction pattern. Lastly, there are two instances where the best *ND-Per* introduction pattern achieves slightly higher utility than the best *NI-Per* and *Per* patterns. That this may happen for some parameter settings we attribute to the discreteness of introduction times in our model.

5.3.3 Discussion

These results shed light on the optimal introduction timing for the general discriminatory pricing problem, i.e., without assuming structural constraints for the introduction times. They suggest that no introduction timing policy can outperform a policy that introduces periodically after an initial warm-up phase. These results also suggest that, in general, it is optimal to have non-increasing introduction intervals in the warm-up phase. For some parameter settings, having decreasing introduction intervals in the warm-up phase is strictly better than having periodic intervals throughout.

We provide some intuition about the optimal introduction timing suggested by our numerical

of each of the *NI-Per*, *Per*, and *ND-Per* schemes:

- *NI-Per* = *Per* = *ND-Per* for 24 out of 60 experimental problem instances
- *NI-Per* = *Per* > *ND-Per* for 18 instances
- *NI-Per* = *ND-Per* > *Per* for 12 instances
- *NI-Per* > *Per*, *ND-Per* for four instances.

In the remaining two experimental problem instances, the best *ND-Per* introduction pattern actually dominates the best *NI-Per* pattern. These instances are $\{d = 10, c = 1, C = 1\}$, for which *ND-Per* > *NI-Per* > *Per*; and $\{d = 14, c = 2, C = 1\}$, for which *ND-Per* > *NI-Per* = *Per*.

results, starting by explaining why it is optimal to have periodic introductions after a warm-up phase. When comparing a policy π_{s_j}, which makes j technology introductions and none after that, to a policy $\pi_{s_{j+1}}$, which makes one additional introduction at s_{j+1}, we can show (Lemma 5 in Appendix J) that the excess utility from the additional introduction is upper bounded by

$$U(\pi_{s_{j+1}}) - U(\pi_{s_j}) \leq \delta^{s_{j+1}} \left(g(s_{j+1} - s_j) - C\right), \tag{31}$$

where g is the revenue difference function we characterized when assuming non-increasing introduction intervals (Section 5.1). Periodic introductions achieve that upper bound in the case where we start from a sequence of non-increasing introduction intervals, and are optimal, by Theorem 2.

We next explain why it may be best to have decreasing intervals in the warm-up phase. During the warm-up phase the mass of customers in the system has not yet reached its steady state level of d customers, and is still building up. Before the system has "warmed up", the provider is more reluctant to introduce than when there is already a mass of d customers in the system, because the introduction costs are the same, yet the benefits are lower. However, this effect wears off as the mass of customers in the system accumulates, particularly for low values of the switching cost. As a result, for some parameter settings it is best to start introducing at intervals that are longer than the identified optimal period T^* from (26) during the warm-up; and to then introduce at periodic intervals that are T^* periods apart. On the other hand, a policy that is constrained to introduce at constant intervals throughout will have to choose between using period T^* throughout, which is optimal for the continuation, but not great for the warm-up; and using a period that is slightly higher than T^* throughout, ensuring a good start, but performing suboptimally in the continuation. A well tuned *NI-Per* introduction pattern can achieve the best of both worlds.

Finally, we argue why it is not easy to show that non-increasing introductions in the warm-up phase are optimal for some parameter settings, even conditional on optimal introduction intervals thereafter. In the warm-up phase, i.e., before a mass of d customers has been built in the system, we can generalize the argument of Section 5.1 and Lemma 5 to show that the utility difference between a policy that terminates at s_j and the related policy which has an additional introduction at s_{j+1}

is upper bounded by

$$U(\pi_{s_{j+1}}) - U(\pi_{s_j}) \le \delta^{s_{j+1}} \left(G(s_j - s_1, s_{j+1} - s_j) - C \right), \tag{32}$$

for some function $G(x, y) \le g(y)$ that captures the exact difference in the revenue under the assumption of non-increasing introduction intervals (even when there are less than d customers in the system). For $s_j - s_1 \ge d - 1$, $G(s_j - s_1, s_{j+1} - s_j) = g(s_{j+1} - s_j)$. However, proving that it is optimal for $(s_1, \ldots, s_j)$ to have non-increasing intervals is not easy, even when conditioning on optimal periodic introductions subsequent to s_j. $G(x, y)$ is now a bivariate function, and we are seeking the arguments that maximize a sum of terms, similar to (22), each of which has form $\delta^{s_{i+1}}(G(s_i - s_1, s_{i+1} - s_i) - C)$. Even when the customer type distribution F is uniform, the analysis is not straightforward, as can be seen in the complexities that the simpler function $\delta^{s_{i+1}}(g(s_{i+1} - s_i) - C)$ exhibits and that were illustrated in Section 5.1.[18]

6 The Non-Discriminatory Pricing Setting

We have proposed an algorithm for identifying optimal discriminatory prices given fixed introduction times; and have shown that, under some conditions, periodic introductions after an initial warm-up phase are optimal for the infinite horizon introduction time and discriminatory pricing problem. In this section we provide results for the setting where the firm does not price-discriminate based on customer experience. In the non-discriminatory setting, we show that, under some conditions, a simple policy of myopic pricing and periodic introductions enjoys good performance guarantees.

In a setting with non-discriminatory pricing, the price of every technology class is the same for all customers, regardless of how many technology classes a customer has switched to. Using our notation for the discriminatory pricing setting in Section 3, this imposes the following constraints:

$$x_{j,0} = x_{j,1} = \ldots = x_{j,min(j-1,n)}, \quad \forall j = 1, \ldots \tag{33}$$

[18]As an aside, in the special case where c and all optimal introduction intervals $z = s_j - s_{j-1}$ are such that $c/z \ge \zeta$, then for any customer type distribution F no customer upgrades, i.e., $F(\theta^*(z)) = 1$, implying $G(x, y) = g(y)$ and that g is linear. It is then optimal to make all introduction intervals periodic with the same period T^*, including those in the warm-up phase.

In the non-discriminatory setting, we denote the price of technology class j simply by x_j.

Adding the constraints (33) when solving the infinite horizon discriminatory pricing problem with fixed introduction times (8) makes the optimization problem difficult. Being able to decouple the problems of optimizing $\text{Revenue}_{[s_j, s_{j+1})}$, i.e., the revenue from customers who arrive in periods $s_j, s_j + 1, \ldots, s_{j+1} - 1$, across different introductions $j = 1, \ldots$, is what leads to a tractable optimization problem for the discriminatory setting (Section 4). However, the constraints (33) couple these problems together. For this reason, in the non-discriminatory setting we focus on characterizing the optimal single-period revenue.

For the results in this section we assume that consecutive introductions are at least $d - 1$ periods apart. Because each customer stays in the system for d periods, this implies that each customer can experience at most one introduction as an existing customer.[19]

Assumption 4. For every $i \geq 2$, we have that $s_i - s_{i-1} \geq d - 1$.

In the remainder of the section, we first show that under a linear pricing rule, which subsumes Myerson pricing, there is no loss of optimality with a periodic schedule of introductions (Section 6.1). We then characterize optimal pricing for a single period (Section 6.2). Third, we give a formal statement for the performance guarantees of Myerson pricing (Section 6.3), and provide two different bounds on the approximation ratio in terms of the type distribution (Section 6.4). Although the focus in this section is on the non-discriminatory pricing setting, our results also speak to the power of simple prices in at least a subset of the discriminatory pricing setting (where the provider can price-discriminate between customers who were present as existing customers for the most recent introduction and customers who arrived after it), beyond what our results in Section 4 show. Finally, we show numerically that our analytical bounds for Myerson pricing provide strong guarantees, and that in reality Myerson pricing is often some orders of magnitude closer to optimal than our bounds suggest (Appendix Q).

[19] The results in this section can in principle be extended to allowing customers to experience multiple introductions as existing customers during their customer lifetime. Such analysis relies on characterizing optimal pricing for a single period for the setting where the provider can price-discriminate between customers depending on how many introductions they have experienced as existing customers. This analysis would be similar to the analysis for optimal discriminatory pricing in Section 4, but focused on a single period rather than the entire slice $\text{Revenue}_{[s_j, s_{j+1})}$.

6.1 Results on linear pricing

We provide a summary of our results on linear pricing here, and include all statements and details in Appendix L.

We first analyze a simple, natural pricing policy: charge a price which is linear in the quality of the technology class. We show this has several nice properties. First, with linear pricing all newly arriving customers will select the latest quality. Second, when restricting to linear pricing, it is optimal to have periodic introductions after the first introduction. A particularly interesting special case is one where the linear prices are chosen to be optimal for each technology class as if it were the only item offered for sale, as per Myerson's approach. Third, we show that 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.

6.2 Optimal pricing for a single period

We provide a summary of our results on optimal pricing for a single period here, and include all statements and details in Appendix N.

The objective of the service provider in our model is to maximize *infinite-horizon* revenue and utility. In this subsection we provide optimal pricing results for the *single-period* problem. These results are useful for the infinite-horizon problem, because we bound the infinite-horizon revenue of a policy, and the infinite-horizon competitive ratio between policies, using the single-period revenue.

We consider the set of all non-discriminatory policies with a particular pattern of introductions.

Definition 1. For a set of introduction times $\mathbf{s} = (s_0, s_1, s_2, \ldots)$, we denote by $\Pi(\mathbf{s})$ the set of all policies with these introduction times. That is, we define

$$\Pi(\mathbf{s}) := \{\pi' = \big((s'_0, x'_0), (s'_1, x'_1), (s'_2, x'_2), \ldots\big) \mid s'_i = s_i \ \forall i\}.$$

We use the following trivial upper bound, for which the revenue in each time period is optimized separately.

Observation 3. *Let introduction times* $\mathbf{s}$ *be given. Then, for all* $\delta > 0$ *we have*

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

The proposed upper bound calculates the revenue in the case where the provider is 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.

For maximizing revenue at a particular time, our first observation is that to maximize revenue from customers who have not seen an introduction as existing customers (i.e., revenue from customers who entered the system at the time of or after the most recent introduction), the optimal policy simply uses Myerson prices (Lemma 6).

This leaves the question of how to set prices to maximize revenue at a particular time from customers who have experienced an introduction as existing customers during their lifetime. These are the customers who entered the system before the most recent introduction. We show that it is optimal to set all but the most recent introduction prices to the Myerson price; furthermore, we can put a lower and an upper bound on the optimal price for the most recent introduction (Lemma 7).

We have identified two classes of customers that are in the system at a particular time: customers who have arrived since the most recent introduction, and customers who arrived before the most recent introduction. We have argued that the policies for maximizing revenue from each of these two classes of customers agree that Myerson pricing should be used for all but the most recent introduction, but disagree on what the price of the most recent introduction should be. In particular, both policies 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. We show (Theorem 4) that the optimal policy for maximizing combined revenue over both classes of customers at a particular time will also be of the same form, intuitively with a compromise over what the price of the current introduction should be. This follows from a quasiconcavity property implied by Assumption 2.

6.3 Performance guarantees of Myerson pricing

We provide a summary of our results on performance guarantees of Myerson pricing here, and include all statements and details in Appendix O.

Having proved our result on optimal one-period pricing, we can now combine it with our results on linear pricing (Section 6.1) to give a precise sense in which Myerson pricing with periodic introductions enjoys good performance guarantees for the non-discriminatory setting. In particular, optimizing U is a bicriterion problem: we want to simultaneously maximize revenue while minimizing cost. While we do not achieve a bounded approximation ratio to U, we can simultaneously approximate these two objective functions. Such bounds are common in bicriteria settings, where an algorithm is an (α, β) approximation if its result is simultaneously an α approximation to the first objective[20] and a β approximation to the second (Ravi et al., 2001; Iyer and Bilmes, 2013).

We make the approximation ratio precise (Corollary 3), and we interpret that result next. Profit has two parts—revenue and cost. For revenue, our result shows that, given a set of introduction times, pricing à la Myerson guarantees a revenue close to the optimal revenue. We characterize analytically how close the revenue of Myerson pricing is to the optimal revenue in the next subsection. In Appendix Q, we show that in simulations the approximation is substantially tighter than guaranteed by our analytical bounds. For cost, note that once we have fixed introduction times the cost is also fixed, thus clearly the Myerson policy achieves that same cost.

As is common in bicriteria settings, such bicriteria approximations do not provide a bounded approximation ratio to the combined objective U. However, since one of the approximation ratios is 1, we achieve the quite strong guarantee that, for whatever introduction strategy the provider chooses, she can capture most of the revenue by using Myerson pricing, while keeping the cost fixed. Furthermore, we can guarantee that she can always do weakly better than that by keeping Myerson pricing and switching to the optimal choice of periodic introductions (which in particular guarantees that U will be non-negative).

[20]When maximizing (minimizing), we say an algorithm is an α approximation to an objective, with $\alpha \geq 1$ ($\alpha \leq 1$), if the best objective value attainable, divided by the value of the objective function that the algorithm obtains, is at most (at least) α.

6.4 Bounding the competitive ratio of Myerson pricing

We now use the above characterization of the approximation ratio to provide our performance guarantee for Myerson pricing. We provide a summary of the results here, and include all statements and details in Appendix P.

Corollary 3 implies that we can bound our approximation ratio by bounding the competitive ratio of Myerson pricing for the one-period revenue. We provide two such bounds in terms of the customer type distribution F and the introduction times sequence (s_i) (Proposition 3, Corollary 4). Our first bound is directly in terms of F but worsens with increasing customer lifetime d, while our second bound requires the derivative of F and an additional optimization to make the bound concrete, but improves with increasing d. Our bounds show that Myerson pricing is approximately optimal when switching costs for the customers who upgrade are small or large. Intuitively, with small switching costs all customers act essentially like new ones, while with large ones few customers will switch so only the new ones are relevant when considering pricing.

Putting together the characterization of the approximation ratio (Corollary 3) with the bounds for the competitive ratio yields the main performance guarantee for Myerson pricing (Theorem 5).

7 Discussion

As technology improves over time, service providers have the ability to offer more powerful products and services, which are more valuable to customers. At the same time, introduction of a new technology class comes at a cost for development and launching, and at the expense of the sales of existing classes. We have presented a model of improved technology introductions for subscription services markets that addresses this trade-off, considering a service provider who price-discriminates based on customers' upgrade experience, in the face of customers who are averse to upgrading to improved offerings. The decision problem for the provider is when to introduce a new technology class and how to price it in order to maximize total profit, taking into account (discounted) future rewards. We have reduced the optimal pricing problem to a tractable optimization problem and developed an efficient algorithm for solving it. We have also shown that periodic introduction times after an initial "warm-up" phase enjoy optimality guarantees.

We conclude by discussing our various modelling assumptions, and their importance to our results, with respect to the substantive context of subscription-based services.

7.1 Discussion of modelling assumptions

We restrict to myopic customers, where myopia is meant in two senses: first, customers base their decisions on the current service offerings, rather than on the basis of beliefs about future service offerings; and secondly, customers make their decision only maximizing their utility in the current period, rather than over their remaining lifetime.

We first comment on the first sense of myopia: we restrict to customers who make their decisions based on the currently offered technologies, as opposed to forward-looking customers who take into account future introductions. This avoids introducing a separate Bayesian belief framework for customers. In general, a forward-looking customer may choose not to upgrade when a new technology class is introduced, but upgrade on the next introduction. Our Assumption 1 doesn't allow such upgrading. In our model we also do not allow for patient customers, that is customers who do not purchase, yet are willing to wait (and stay in the system) and buy later. Therefore, precluding patient customers, the only thing that changes with forward-looking customers in our setting is that it is possible that they will buy a unit of a new technology class with negative utility in their first period, if an introduction in a subsequent 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.

We now comment on the second sense in which our customers are myopic: we assume customers maximize their utility in the current period, as opposed to maximizing their utility over the anticipated usage time. Relatedly, we model a customer's switching cost as being assessed against a single-period revenue gain, and not amortized over a customer's anticipated remaining lifetime. For example, this reflects a setting where the upgrade generates a capital cost or labor cost incurred only at the time when the upgrade occurs, and offset against the single-period revenue gain, consistent with a single-period accrual basis of accounting. If customers make decisions maximizing their utility over their remaining lifetime, then their decisions would change. Such behavior can be incorporated into our model, but at a complexity cost: the state space necessarily depends upon the remaining lifetime of a customer. While the same general principle of optimizing over each slice separately when pricing applies, the optimization problem is more complex. One heuristic approach would be to apply our algorithm from Section 4 using an adjusted c reflecting an "average" amortized switching cost. It is unclear whether results on the optimality of periodic introductions carry over, since the dependence of the state space on the remaining customer lifetime makes the revenue difference function g multidimensional rather than univariate.

The details of the switching cost model play an important role in the analysis. In our model, the switching cost c is a constant, regardless of the technology levels that a customer upgrades from and to. In an alternative specification, the switching cost can be modeled as being related to the technology difference, e.g., as $c \cdot (s_j - s_{j-1})$ for a transition where a customer upgrades from technology class $j-1$ to technology class j. This switching cost specification would simplify the analysis significantly: upgrades happen at a single θ threshold for the customer type, and difficulties with increasing introduction intervals disappear, leading to a quite simpler pricing algorithm. Note that this specification implies that an upgrade from technology A to technology B, followed by an upgrade from technology B to technology C, would generate total switching costs equal to a single upgrade from technology A to technology C (when ignoring discounting). This modeling is thus quite different from our main specification. Out of the three sources of switching costs that we identified in the introduction, costs related to redesigning the customer's business processes may be well modelled with this alternative specification; whereas downtime costs and customer inertia costs may be better modelled with our main switching cost specification. If the described alternative specification is considered as an extreme point in the space of switching cost models, one can also consider switching cost models that lie in between our main specification and the alternative specification described here.

We assumed that customers can only upgrade if they were already using the latest technology (Assumption 1). This partially captures, for example, the model of various subscription software suites, including for productivity (e.g., Office), music creation and editing (e.g., Pro Tools), technical computing (e.g., Mathematica), etc.[21] For tractability, some restriction on possible upgrades seems

[21] However, our framework does not capture some of the details of software suite offerings. For example, oftentimes the web versions of a software suite update immediately and the customer has no control. On the installed versions, oftentimes the user can avoid upgrading, at least for a while, but when she does upgrade, in most cases she will be

necessary as otherwise the number of upgrade patterns to be considered is exponential. A weaker assumption would be to permit customers to upgrade one technology level at a time, even if they are not currently using the latest technology. This opens the possibility, for example, of customers buying some technology other than the latest and then upgrading over time. Much of our analysis in Section 4 carries over, but it is no longer the case that it is always optimal to set prices so that new customers buy the latest technology. The reason is that this opens another avenue for price discrimination where a high initial price for the latest technology causes some customers with lower values to choose an older technology. But now that the market is segmented between high- and low-valued customers, the pricing of future upgrade discounts can exploit this. As such pricing schemes rely heavily on our assumption of myopia and seem unrealistic in practice, we prefer our stronger assumption on possible upgrades.

We lastly discuss some of our other modelling assumptions. Our assumption that technology classes remain available once introduced is not crucial, given that our results show that with optimal prices, all new customers choose the latest class anyways. We also make the assumption that customers can choose up to one unit of a technology class in any given period, meaning we do not allow using multiple units in a period. This modeling assumption fails to capture cases, such as in cloud computing, where some degree of distribution or parallelization is needed. For example, in the context of cloud computing, consider a customer who minimizes purchasing cost subject to achieving a level of computational power, which is known to her in advance. Such a customer may naturally consider the option of purchasing more units of an older and weaker virtual machine class, rather than one unit of the newest and stronger virtual machine class. Although our model fails to capture such cases, which is a limitation, our assumption is natural in many other contexts of interest, including subscription software suites, subscription services, mobile phones, and others. Finally, 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

forced to the latest version, and won't be able to install the old version any more on a new machine. Upgrading to the latest version after having avoided some upgrades is a "jump" upgrade and violates our assumption on possible upgrades. Furthermore, in this case the price the user pays does not depend on the user's upgrade history.

periodicity, but they will not affect the general shape of our results on optimal pricing (Section 4) .

References

Ableton (2021), "Upgrade FAQ." https://help.ableton.com/hc/en-us/articles/206407004-Upgrade-FAQ. Accessed: 2021-05-10.

Acquisti, Alessandro and Hal R. Varian (2005), "Conditioning prices on purchase history." *Marketing Science*, 24, 367–381, URL https://doi.org/10.1287/mksc.1040.0103.

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

Arturia (2019), "Arturia announce huge software deals for Black Friday." https://www.arturia.com/company/press/press-releases/arturia-announce-huge-software-deals-for-black-friday. Accessed: 2021-05-10.

Bagnoli, Mark and Ted Bergstrom (1989), "Log-concave probability and its applications." URL http://econ.ucsb.edu/~tedb/Theory/delta.pdf. Subsequent version published in *Economic Theory*, 2005.

Bala, Ram and Scott Carr (2009), "Pricing software upgrades: The role of product improvement and user costs." *Production and Operations Management*, 18, 560–580, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1937-5956.2009.01030.x.

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.

Barriola, Xabier and Victor Martínez de Albéniz (2019), "The race for product renewal." *Working paper*.

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 (2018), "Robust dynamic pricing with strategic customers." *Mathematics of Operations Research*, 43, 1119–1142, URL `https://doi.org/10.1287/moor.2017.0897`.

Chen, Yiwei, Vivek F. Farias, and Nikolaos K. Trichakis (2019), "On the efficacy of static prices for revenue management in the face of strategic customers." *Management Science, Forthcoming.*

Chen, Yiwei and Ming Hu (2020), "Pricing and matching with forward-looking buyers and sellers." *Manufacturing & Service Operations Management*, 22, 717–734, URL `https://doi.org/10.1287/msom.2018.0769`.

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`.

Cosguner, Koray, Tat Y. Chan, and P. B. (Seethu) Seetharaman (2017), "Behavioral price discrimination in the presence of switching costs." *Marketing Science*, 36, 426–435, URL https://doi.org/10.1287/mksc.2016.1024.

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.

Fudenberg, Drew and Jean Tirole (1998), "Upgrades, tradeins, and buybacks." *The RAND Journal of Economics*, 29, 235–258, URL http://www.jstor.org/stable/2555887.

Fudenberg, Drew and J Miguel Villas-Boas (2006), "Behavior-based price discrimination and customer recognition." *Handbook on economics and information systems*, 1, 377–436.

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 (2019), "When fixed price meets priority auctions: Competing firms with different pricing and service rules." *Stochastic Systems*, 9, 47–80, URL https://doi.org/10.1287/stsy.2018.0021.

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.

Gordon, Brett R. (2009), "A dynamic model of consumer replacement cycles in the pc processor industry." *Marketing Science*, 28, 846–867, URL https://doi.org/10.1287/mksc.1080.0448.

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.

Huang, Yufeng (2019), "Learning by doing and the demand for advanced products." *Marketing Science*, 38, 107–128, URL https://doi.org/10.1287/mksc.2018.1118.

Iyer, Rishabh K and Jeff A Bilmes (2013), "Submodular optimization with submodular cover and submodular knapsack constraints." In *Advances in Neural Information Processing Systems*, 2436–2444.

Jia, Kunhao, Xiuwu Liao, and Juan Feng (2018), "Selling or leasing? dynamic pricing of software with upgrades." *European Journal of Operational Research*, 266, 1044–1061, URL https://www.sciencedirect.com/science/article/pii/S037722171730992X.

Jing, Bing (2017), "Behavior-based pricing, production efficiency, and quality differentiation." *Management Science*, 63, 2365–2376, URL https://doi.org/10.1287/mnsc.2016.2463.

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.

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.

Li, Krista J. and Sanjay Jain (2016), "Behavior-based pricing: An analysis of the impact of peer-induced fairness." *Management Science*, 62, 2705–2721, URL https://doi.org/10.1287/mnsc.2015.2265.

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 (2020), "Technical note — dynamic pricing with heterogeneous patience levels." *Operations Research*, 68, 1038–1046, URL https://doi.org/10.1287/opre.2019.1951.

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`.

Mehra, Amit, Ram Bala, and Ramesh Sankaranarayanan (2012), "Competitive behavior-based price discrimination for software upgrades." *Information Systems Research*, 23, 60–74, URL `https://pubsonline.informs.org/doi/abs/10.1287/isre.1100.0291`.

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

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`.

Oren, Shmuel S. and Michael H. Rothkopf (1984), "A market dynamics model for new industrial products and its application." *Marketing Science*, 3, 247–265, URL `https://doi.org/10.1287/mksc.3.3.247`.

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`.

Penmetsa, Nabita, Esther Gal-Or, and Jerrold May (2015), "Dynamic pricing of new services in subscription markets." *Production and Operations Management*, 24, 896–916, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.12317.

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.

PRO Landscape (2021), "Existing customers upgrades & copies." https://prolandscape.com/en/buy-now/existing-customers/. Accessed: 2021-05-10.

Puterman, Martin L (2014), *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Ravi, R, Madhav V Marathe, SS Ravi, Daniel J Rosenkrantz, and Harry B Hunt III (2001), "Approximation algorithms for degree-constrained minimum-cost network-design problems." *Algorithmica*, 31, 58–78.

Ray, Saibal, Tamer Boyaci, and Necati Aras (2005), "Optimal prices and trade-in rebates for durable, remanufacturable products." *Manufacturing & Service Operations Management*, 7, 208–228, URL https://doi.org/10.1287/msom.1050.0080.

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

Sky (2021), "Sky deals for existing customers – upgrade offers and deals." https://www.skyoffersnewcustomers.co.uk/sky-deals-for-existing-customers/. Accessed: 2021-05-10.

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.

Tripsas, Mary (2009), "Technology, identity, and inertia through the lens of "the digital photography company"." *Organization Science*, 20, 441–460, URL https://doi.org/10.1287/orsc.1080.0419.

Verizon (2021), "Upgrade and get up to $440 off iPhone 12." `https://www.verizon.com/deals/upgrades/`. Accessed: 2021-05-10.

Vodafone (2021), "Existing customer deals for business." `https://www.vodafone.co.uk/business/deals/existing-customer-deals`. Accessed: 2021-05-10.

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`.

Zhu, Kevin, Kenneth L. Kraemer, Vijay Gurbaxani, and Sean Xin Xu (2006), "Migration to open-standard interorganizational systems: Network effects, switching costs, and path dependency." *MIS Quarterly*, 30, 515–539, URL `http://www.jstor.org/stable/25148771`.

A Analysis of Virtual Machine Series Launches

Every virtual machine (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 3.

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.

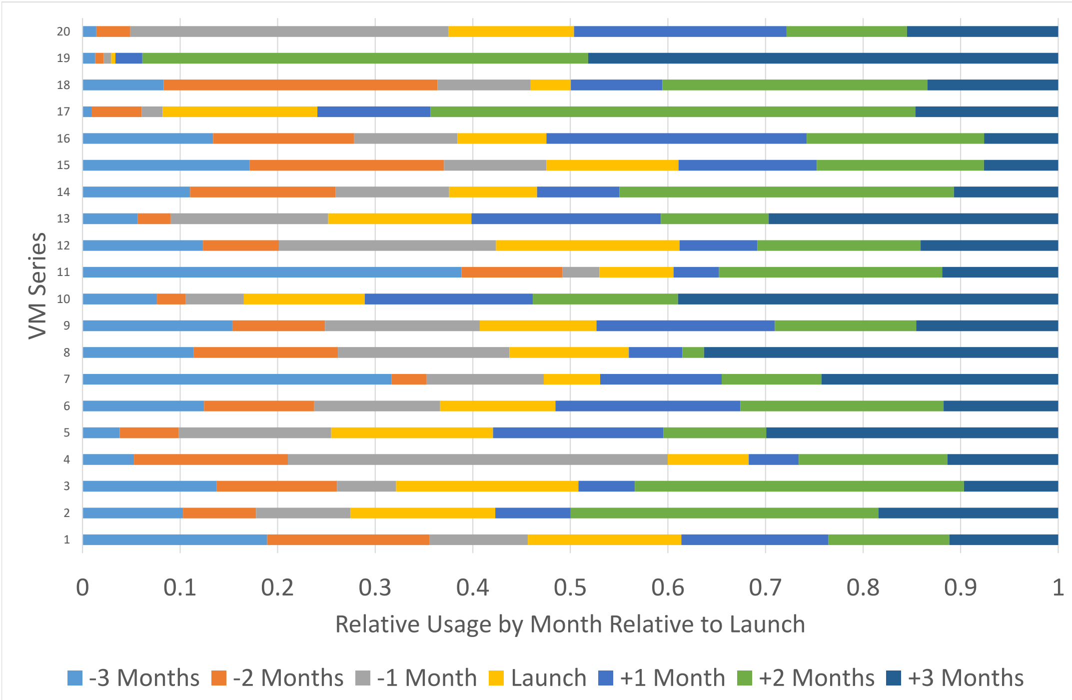

Figure 3: Relative usage of twenty different VM series on Azure by subscriptions created the specified number of months relative to their launch month.

B Proof of Lemma 1

We begin by considering optimal pricing under a simpler version of the problem, where newly arriving customers are constrained to either buy the jth technology or nothing (so in particular they cannot buy technology class 1 through $j-1$). We subsequently show that this behavior is consistent with optimal pricing.

Under this assumption, a newly arriving customer will buy technology j if

$$\theta s_j - x_{j,0} \geq 0$$

and customers will upgrade to technology $j+i+1$ when it becomes available if

$$\theta s_{j+i+1} - x_{j+i+1,i+1} - c \geq \theta s_{j+i} - x_{j+i,i}.$$

Given prices $x_{j,0}, x_{j+1,1} \ldots, x_{j+n,n}$, we can compute the minimum types that choose to buy and upgrade as, respectively,

$$\theta_j^j = \frac{x_{j,0}}{s_j} \tag{34}$$

and

$$\theta_{j+i+1}^j = \max\left(\frac{x_{j+i+1,i+1} - x_{j+i,i} + c}{s_{j+i+1} - s_{j+i}}, \theta_{j+i}^j\right). \tag{35}$$

The max in (35) follows from Assumption 1 as a given type can only upgrade if it also adopted the most recent prior technology. In particular, this means that the θ_{j+i}^j are monotone non-decreasing in i.

At optimal prices, we must have

$$\theta_{j+i+1}^{j*} = \frac{x_{j+i+1,i+1}^* - x_{j+i,i}^* + c}{s_{j+i+1} - s_{j+i}}, \quad 0 \leq i \leq n-1, \tag{36}$$

which we refer to as the prices being "non-wasteful". Suppose for contradiction that

$$\frac{x_{j+i+1,i+1}^* - x_{j+i,i}^* + c}{s_{j+i+1} - s_{j+i}} < \theta_{j+i}^{j*}.$$

Consider the effects of increasing $x_{j+i+1,i+1}$ so that equality holds while holding all other prices fixed. This has no effect on which types will upgrade, and increases revenue. The other effect of this price increase is that it may lower the threshold at which customers adopt technology $j+i+2$, which again increases revenue, because $x_{j+i+2,i+2}$ as well as all subsequent prices (if this leads to further adoptions) were fixed. Thus making prices non-wasteful can only increase revenue and so optimal prices are non-wasteful. Solving (34) and (36) for prices yields (13).

We can write the revenue

$$\text{Revenue}_{[s_j,s_{j+1})} := \sum_{t=s_j}^{s_{j+1}+d-2} \delta^t \int \left(\sum_{\ell=\max(1,t-s_{j+1}+2)}^{\min(t-s_j+1,d)} x_{q_\ell(\pi,t,\theta),m_\ell(\pi,t,\theta)}\right) f(\theta) d\theta$$

in terms of the θ^j_{j+i} and prices as

$$\sum_{t'=s_j}^{s_{j+1}-1}\sum_{t=t'}^{t'+d-1}\delta^t x_{j,0}(1-F(\theta^j_j)) + \sum_{i=1}^{n}\sum_{t'=s_j}^{s_{j+1}-1}\sum_{t=s_{j+i}}^{t'+d-1}\delta^t(x_{j+i,i}-x_{j+i-1,i-1})(1-F(\theta^j_{j+i})).$$

We explain the last expression. The summation over t' is to sum over the times when the customers from this slice arrive: that could be in periods $s_j, s_j+1, \ldots, s_{j+1}-1$. The summation over t is to sum over the times when the revenue is being accumulated, starting at s_{j+i} (t' for the first part of the expression) and going through the last period during which that customer is in the system: $t'+d-1$. The term with the price difference between the current and the previous technology comes from telescoping: as we sum over i, we end up with the price for the current technology.

Substituting the thresholds for the prices yields

$$\sum_{t'=s_j}^{s_{j+1}-1}\sum_{t=t'}^{t'+d-1}\delta^t s_j\theta^j_j(1-F(\theta^j_j)) + \sum_{i=1}^{n}\sum_{t'=s_j}^{s_{j+1}-1}\sum_{t=s_{j+i}}^{t'+d-1}\delta^t((s_{j+i}-s_{j+i-i})\theta^j_{j+i}-c)(1-F(\theta^j_{j+i})).$$

Finally, letting

$$A^j_j = \sum_{t'=s_j}^{s_{j+1}-1}\sum_{t=t'}^{t'+d-1}\delta^t \tag{37}$$

and

$$A^j_{j+i} = \sum_{t'=s_j}^{s_{j+1}-1}\sum_{t=s_{j+i}}^{t'+d-1}\delta^t, \qquad i=1,\ldots,n \tag{38}$$

yields the desired form

$$A^j_j s_j\theta^j_j(1-F(\theta^j_j)) + \sum_{i=1}^{n} A^j_{j+i}((s_{j+i}-s_{j+i-i})\theta^j_{j+i}-c)(1-F(\theta^j_{j+i})).$$

The optimal choice of θ^j_j maximizes $A^j_j s_j\theta^j_j(1-F(\theta^j_j))$, making the optimal choice $\theta^{j*}_j = p^*$.

Next, we deal with the possibility of newly arrived customers buying some older technology than j, paying a lower price and never upgrading. We have seen that, neglecting this possibility, the optimal prices to set for new customers are $x^*_{i,0} = s_i p^*$ for all i. With these prices, the customer only prefers technology class i over j if $\theta s_j - s_j p^* < \theta s_i - s_i p^*$, or $s_j < s_i$. Thus newly arrived customers always buy the newest technology (or nothing), even if they are allowed to buy older technologies, as long as $x_{i,0} = s_i p^*$ for all i. This shows that under these prices, new customers behave as described in our analysis even if they have the option of choosing an old technology.

Finally, it remains to be shown that it is not profitable to set prices differently so that some new customers do buy older technologies. However, as the initial decision is a single-parameter setting, Myerson (1981) shows that the optimal solution is to sell the latest technology to all customers with a non-negative virtual valuation, which is exactly what the prices we derived do.

C Example where Optimal Prices Are Higher for Existing Customers

The following example shows how increasing introduction intervals can lead to optimal prices that charge existing customers higher prices than new customers.

Example 1. Suppose d is large and $s_{j+2} = s_{j+1} + d - 2$. By taking d large and $s_{j+2} - s_j >> s_{j+1} - s_j$ we can make the constants from (12), provided in (38), satisfy $A^j_{j+1} >> A^j_{j+2}$. Concretely, fixing $s_{j+1} = s_j + 1$, we have by (38) that $A^j_{j+1} = \sum_{t=s_{j+1}}^{s_{j+2}} \delta^t$ and $A^j_{j+2} = \delta^{s_{j+2}}$. In particular, $A^j_{j+2} \leq A^j_{j+1}\delta^{(s_{j+2}-s_{j+1})}$. Essentially, the price $x_{j+1,1}$ matters for an arbitrarily long period of time while $x_{j+2,2}$ matters for a single period.

As $s_{j+2} - s_{j+1} > s_{j+1} - s_j$, θ^j_{j+1} and θ^j_{j+2} are lumped at the optimal solution and satisfy

$$\theta^{j*}_{\{j+1,j+2\}} = v^{-1}\left(\frac{c(A^j_{j+1} + A^j_{j+2})}{(s_{j+1} - s_j)A^j_{j+1} + (s_{j+2} - s_{j+1})A^j_{j+2}}\right) \approx v^{-1}\left(\frac{c}{s_{j+1} - s_j}\right) = \theta^{j,FOC}_{j+1},$$

as A^j_{j+2} is exponentially smaller than A^j_{j+1}.

Taking customer types to be uniformly distributed on $[0, 1]$ (so that $v(\theta) = 2\theta - 1$) and applying the equations for prices (13) gives

$$
\begin{aligned}
x_{j,0} &= s_j p^* \\
x_{j+1,1} &\approx x_{j,0} + (s_{j+1} - s_j) v^{-1} \left(\frac{c}{s_{j+1} - s_j} \right) - c \\
x_{j+2,2} &\approx x_{j+1,1} + (s_{j+2} - s_{j+1}) v^{-1} \left(\frac{c}{s_{j+1} - s_j} \right) - c \\
&= s_j p^* + (s_{j+2} - s_j) \left(p^* + \frac{c/2}{s_{j+1} - s_j} \right) - 2c \\
&= s_{j+2} p^* + \left(\frac{s_{j+2} - s_j}{s_{j+1} - s_j} \right) c/2 - 2c
\end{aligned}
$$

Since we took $s_{j+2} - s_j >> s_{j+1} - s_j$, we have $x_{j+2,2} > s_{j+2} p^*$ meaning the price to existing customers who upgrade is higher than that charged to new customers.

D Proof of Lemma 2

Proof. For lumping $\mathcal{L}$, each of the joint terms

$$\Big(1 - F(\theta_i^j)\Big) \sum_{\ell \in \mathcal{L}(i)} A_\ell^j \Big((s_\ell - s_{\ell-1}) \, \theta_i^j - c \Big), \quad j + 1 \le i \le j + n \tag{39}$$

of the objective function of optimization problem (12) is quasiconcave in its argument θ_i^j. To see this, note that Assumption 2 is equivalent to the log-concavity of $1 - F(\theta)$ as a function of θ, because in general log-concavity of a function g is equivalent to g'/g being monotonically decreasing (Bagnoli and Bergstrom, 1989, Remark 1). Note also that the term $\sum_{\ell \in \mathcal{L}(i)} A_\ell^j \Big((s_\ell - s_{\ell-1}) \, \theta_i^j - c \Big)$ is log-concave in θ_i^j, because it can be written as $a \cdot \theta_i^j + b$, where a, b are constants, and a linear function of θ_i^j is log-concave in θ_i^j. Then (39) is log-concave as the product of log-concave functions. Because log-concavity implies quasiconcavity, it follows that (39) is quasiconcave as a function of θ_i^j.

Fix i, with $j+1 \leq i \leq j+n$. $\theta^{j,FOC}_{\mathcal{L}'(i)}$ maximizes the joint term

$$\Big(1 - F(\theta^j_i)\Big) \sum_{\ell \in \mathcal{L}'(i)} A^j_\ell \Big((s_\ell - s_{\ell-1})\,\theta^j_i - c\Big). \tag{40}$$

Because each of the terms

$$\Big(1 - F(\theta^j_\iota)\Big) \sum_{\ell \in \mathcal{L}(\iota)} A^j_\ell \Big((s_\ell - s_{\ell-1})\,\theta^j_\iota - c\Big), \quad \iota \in \mathcal{L}'(i) \tag{41}$$

is quasiconcave, and $\mathcal{L}(\iota) \subseteq \mathcal{L}'(i)$ for $\mathcal{L} \subseteq \mathcal{L}'$ and $\iota \in \mathcal{L}'(i)$, it follows that any θ^j_i with $\theta^j_i < \min_{\iota \in \mathcal{L}'(i)} \theta^{j,FOC}_{\mathcal{L}(\iota)}$ does not maximize (40); and neither does any θ^j_i with $\theta^j_i > \max_{\iota \in \mathcal{L}'(i)} \theta^{j,FOC}_{\mathcal{L}(\iota)}$. $\square$

E Proof of Lemma 3

Proof. Per the statement of the lemma, we assume that $\mathcal{L} \subseteq \mathcal{L}^*$ and $k^* > -\infty$. This means that $\left(\theta^{j,FOC}_{\mathcal{L}(j+1)}, \theta^{j,FOC}_{\mathcal{L}(j+2)}, \ldots, \theta^{j,FOC}_{\mathcal{L}(j+n)}\right)$ is not a feasible solution for optimization problem (12) because $\theta^{j,FOC}_{\mathcal{L}(j+k^*)} > \theta^{j,FOC}_{\mathcal{L}(j+k^*+1)}$.

Let $k' = \min\left\{k : \theta^{j,FOC}_{\mathcal{L}(j+k)} > \theta^{j,FOC}_{\mathcal{L}(j+k^*)}\right\}$. By definition, $j+k^*$ is the index of the largest out of order θ^j_i, so if $\left\{k : \theta^{j,FOC}_{\mathcal{L}(j+k)} > \theta^{j,FOC}_{\mathcal{L}(j+k^*)}\right\} \neq \emptyset$ we have $\theta^{j,FOC}_{\mathcal{L}(j+k')} \leq \ldots \leq \theta^{j,FOC}_{\mathcal{L}(j+n)}$. By the definition of k', $\theta^j_{j+k'-1} = \theta^j_{j+k'} \notin \mathcal{L}$. We claim that $\theta^j_{j+k'-1} = \theta^j_{j+k'} \notin \mathcal{L}^*$ either. Suppose for contradiction that it is and let $\mathcal{L}' = \mathcal{L}^* - \{\theta^j_{j+k'-1} = \theta^j_{j+k'}\}$. $\mathcal{L} \subseteq \mathcal{L}'$, so applying Lemma 2 with $i = j+k'-1$ we can conclude that $\theta^{j,FOC}_{\mathcal{L}'(j+k'-1)} \leq \max_{1\leq k\leq k'-1} \theta^{j,FOC}_{\mathcal{L}(j+k)} < \theta^{j,FOC}_{\mathcal{L}(j+k')} \leq \theta^{j,FOC}_{\mathcal{L}'(j+k')}$, where the last inequality is again by Lemma 2 but with $i = j+k'$. Because by assumption $\theta^j_{j+k'-1} = \theta^j_{j+k'} \in \mathcal{L}^*$ while $\theta^j_{j+k'-1} = \theta^j_{j+k'} \notin \mathcal{L}'$, the constraint $\theta^j_{j+k'-1} \leq \theta^j_{j+k'}$ is violated in $\mathcal{L}'$. Therefore we have $\theta^{j,FOC}_{\mathcal{L}'(j+k')} < \theta^{j,FOC}_{\mathcal{L}'(j+k'-1)}$, contradicting the above.

Assume for the purposes of contradiction that $\theta^j_{j+k^*}$ and $\theta^j_{j+k^*+1}$ are not lumped together in the optimal solution, i.e., that $\theta^j_{j+k^*} = \theta^j_{j+k^*+1} \notin \mathcal{L}^*$. Then we can write

$$\begin{aligned}
\theta^{j*}_{j+k^*+1} &= \theta^{j,FOC}_{\mathcal{L}^*(j+k^*+1)} \\
&\leq \max_{\iota \in \mathcal{L}^*(j+k^*+1)} \theta^{j,FOC}_{\mathcal{L}(\iota)} \\
&\leq \max_{k \in [k^*+1,\ldots,k'-1]} \theta^{j,FOC}_{\mathcal{L}(j+k)} \\
&< \theta^{j,FOC}_{\mathcal{L}(j+k^*)},
\end{aligned} \tag{42}$$

where the equality is definitional for optimal lumping $\mathcal{L}^*$, the first inequality follows by the right-hand side bound of Lemma 2 applied for lumpings $\mathcal{L} \subseteq \mathcal{L}^*$, the second inequality follows because $\theta^j_{j+k'-1} = \theta^j_{j+k'} \notin \mathcal{L}^*$ and by our assumption that $\theta^j_{j+k^*} = \theta^j_{j+k^*+1} \notin \mathcal{L}^*$, and the third inequality follows because we have assumed $\theta^j_{j+k^*} = \theta^j_{j+k^*+1} \notin \mathcal{L}^*$ and because $\theta^{j,FOC}_{\mathcal{L}(j+k^*)}$ is the largest θ^j_i in lumping $\mathcal{L}$ for $i < j + k'$.

Since by assumption $\theta^j_{j+k^*}$ and $\theta^j_{j+k^*+1}$ are not lumped together in the optimal solution, it must be that $\theta^{j*}_{j+k^*} < \theta^{j*}_{j+k^*+1}$. This means that no upper bound constraint binds on $\theta^j_{j+k^*}$ in the optimal solution. Therefore, a small increase for $\theta^j_{j+k^*}$ would be feasible. If it were the case that $\theta^{j,FOC}_{\mathcal{L}^*(j+k^*)} < \theta^{j,FOC}_{\mathcal{L}(j+k^*)}$, then such an increase would also lead to higher objective value, because $\mathcal{L} \subseteq \mathcal{L}^*$ and by quasiconcavity. But then $\theta^{j,FOC}_{\mathcal{L}^*(j+k^*)}$ would not be optimal. Therefore, we have

$$\theta^{j,FOC}_{\mathcal{L}(j+k^*)} \leq \theta^{j,FOC}_{\mathcal{L}^*(j+k^*)}. \tag{43}$$

Combining (42) with (43) yields $\theta^{j*}_{j+k^*+1} < \theta^{j*}_{j+k^*}$ at the optimal solution, which is infeasible, leading to a contradiction. □

F Proof of Theorem 1

Proof. For a fixed $j \geq 1$, the OptimalLumping algorithm produces a set of right-hand side terms γ^j_i's. The DiscriminatoryPricing algorithm calls the OptimalLumping algorithm as a subroutine, and sets $\theta^j_i \leftarrow v^{-1}(\gamma^j_i)$ for $j+1 \leq i \leq j+n$. Because function $v^{-1}(\cdot)$ is non-decreasing, an ordering of γ^j_i's corresponds to the same ordering of the θ^j_i's, possibly with some more ties. In this proof, as in the rest of Section 4, we keep referring to a lumping as a set of equality constraints on the thresholds θ_i^j, although our implementation of the algorithm in Algorithm 1 operates on the γ_i^j's.

We first show that for given $j \geq 1$, the OptimalLumping algorithm correctly identifies an optimal lumping for $\text{Revenue}_{[s_j, s_{j+1})}$.

To show correctness of the OptimalLumping algorithm (Algorithm 1), we show the following claim: At each iteration, the OptimalLumping algorithm results in a lumping that is a subset of an optimal lumping $\mathcal{L}^*$ for problem (12). We proceed to prove this claim by induction on the iterations of the algorithm.

First, we show the base case. At the zero-th iteration of the OptimalLumping algorithm, the algorithm proposes no equality constraints for different thresholds θ_i^j's. This is the "no lumping", which is a lumping that is vacuously a subset of $\mathcal{L}^*$.

We proceed to show the inductive step. Given a lumping $\mathcal{L}$, we denote by $OLiter(\mathcal{L})$ the lumping that results from one iteration of the OptimalLumping algorithm on lumping $\mathcal{L}$. The inductive step claims that if for a lumping it holds that $\mathcal{L} \subseteq \mathcal{L}^*$, then $OLiter(\mathcal{L}) \subseteq \mathcal{L}^*$.

Recall that an iteration of the OptimalLumping algorithm, when applied on lumping $\mathcal{L}$, identifies the highest $\theta_{\mathcal{L}(i)}^{j,FOC}$ that is out of the desired order, which we denote $\theta_{\mathcal{L}(j+k^*)}^{j,FOC}$. Then the algorithm adds to the lumping the constraint $\theta_{j+k^*}^j = \theta_{j+k^*+1}^j$. By Lemma 3, the resulting lumping $OLiter(\mathcal{L})$ satisfies $OLiter(\mathcal{L}) \subseteq \mathcal{L}^*$.

We have shown that in each iteration, the OptimalLumping algorithm only lumps together terms that are truly lumped together in the optimal solution. Also, when lumping terms together, i.e., when imposing more equality constraints, the objective value of problem (12) decreases. Furthermore, the OptimalLumping algorithm terminates with a lumping that corresponds to a feasible solution to problem (12). It follows that, for fixed $j \geq 1$, the OptimalLumping algorithm identifies an optimal lumping, and that the DiscriminatoryPricing algorithm produces optimal prices $\left(x_{j,0}^*, x_{j+1,1}^*, \ldots, x_{j+n,n}^*\right)$ for $\text{Revenue}_{[s_j, s_{j+1})}$.

By producing optimal prices for each revenue slice $\text{Revenue}_{[s_j, s_{j+1})}$, $j = 1, \ldots$, the DiscriminatoryPricing algorithm produces optimal prices for the infinite horizon discriminatory pricing problem with fixed introduction times.

$\square$

G Proof of Theorem 2

Proof. Fix $j \geq 1$. Let $\pi_{s_j} = ((s_0 = 0, x_0 = 0), (s_1, \mathbf{x}_1^*), \ldots, (s_j, \mathbf{x}_j^*))$ be the finite policy that has introduction times $(s_1, ..., s_j)$, where s_j is the last introduction, and where the prices $\mathbf{x}_i^* = \left(x_{i,0}^*, \ldots, x_{i,min(i-1,n)}^*\right)$, $i = 1, \ldots, j$, are optimal prices given these introduction times. We take j large enough so that $s_{j+1} \geq s_1 + d - 1$. Under our assumptions, the prices are derived from expression (13) and the first order condition (15), hence

$$x_{i,m}^* = \begin{cases} s_i p^*, & \text{if } m = 0 \\ x_{i-1,m-1}^* + (s_i - s_{i-1})\theta_i^{FOC} - c, & \text{if } 0 < m \leq n \end{cases} \tag{44}$$

where θ_i^{FOC} is given by

$$\theta_i^{FOC} = v^{-1}\left(\frac{c}{s_i - s_{i-1}}\right), \tag{45}$$

for $1 \leq i \leq j$. Note that θ_i^{FOC} doesn't depend upon m, and we write

$$\theta_i^* = \theta_i^{FOC} = v^{-1}\left(\frac{c}{s_i - s_{i-1}}\right). \tag{46}$$

We will calculate the additional revenue generated by introducing at time s_{j+1}, by comparing policies $\pi_{s_{j+1}}$ and π_{s_j}. We first look at the revenue of policy π_{s_j}. We fix a time $t > \min(s_j + 1, s_1 + d - 1)$, such that there is a mass d of customers in the system, and we break Revenue(π_{s_j}, t) into four parts.

1. Revenue from newly arriving customers at time t who buy technology class j because $\theta s_j - s_j p^* \geq 0 \iff \theta \geq p^*$. This expected revenue is $(1 - F(p^*))s_j p^*$.

2. Revenue from customers who arrive at or after s_j, before t, and within the last $d-1$ periods before t, and who buy technology class j. There are $\min(t - s_j, d-1)$ such periods, making the expected revenue for this part $\min(t - s_j, d-1)(1 - F(p^*))s_j p^*$.

3. Revenue from customers who arrive before s_j and upgrade from technology class $j-1$ to class

j at time s_j, because their type satisfies $\theta \geq \theta_j^*$. There are $(d - \min(t - s_j + 1, d)) \left(1 - F(\theta_j^*)\right)$ such customers, and such a customer pays price $x_{j,m}^*$, conditional on the number of upgrades m she has switched to as an existing customer.

4. Revenue from customers who arrive prior to t, and didn't upgrade to technology class j, or use a technology class older than j.

Notice that the first three revenue parts are from customers who use technology j at time t.

We next consider policy $\pi_{s_{j+1}}$. We take time $t = s_{j+1} (> s_j)$, and focus on the additional revenue of policy $\pi_{s_{j+1}}$ over policy π_{s_j}, that is, $\text{Revenue}(\pi_{s_{j+1}}, t) - \text{Revenue}(\pi_{s_j}, t)$.

1. Newly arriving customers at t who purchased technology class j as new customers under policy π_{s_j}, can now purchase $j+1$ instead, at a price $s_{j+1}p^*$, generating revenue $(1 - F(p^*))s_{j+1}p^*$, i.e., an additional revenue of
$$(1 - F(p^*))(s_{j+1} - s_j)p^*.$$

2. Customers who purchased technology class j as new customers at times prior to $t = s_{j+1}$ under policy π_{s_j}, can now upgrade to class $j+1$ at price $x_{j+1,1}^*$, as long as their type satisfies $\theta \geq \theta_{j+1}^*$. There are $\min(t - s_j, d-1)(1 - F(\theta_{j+1}^*))$ such customers, and they generate additional revenue (over policy π_{s_j}) of
$$\min(t - s_j, d-1) \left(1 - F(\theta_{j+1}^*)\right) \left((s_{j+1} - s_j)\theta_{j+1}^* - c\right).$$

3. Customers who upgraded to technology class j and are still in the system at $t = s_{j+1}$, can now further upgrade to technology class $j+1$, provided their type satisfies $\theta \geq \theta_{j+1}^*$. There are $(d - \min(t - s_j + 1, d))(1 - F(\theta_{j+1}^*))$ such customers, and customers who previously paid $x_{j,m}^*$ now pay upgrade prices $x_{j+1,m+1}^*$, for $m = 2, ..., n$. They generate additional revenue of
$$\begin{aligned}&(d - \min(t - s_j + 1, d)) \left(1 - F(\theta_{j+1}^*)\right) \left((s_{j+1} - s_j)\theta_{j+1}^* - c\right)\\ &= (d - 1 - \min(t - s_j, d-1)) \left(1 - F(\theta_{j+1}^*)\right) \left((s_{j+1} - s_j)\theta_{j+1}^* - c\right)\end{aligned}$$

4. The revenue from customers[22] who don't upgrade to technology class $j+1$ from j, or who stay with existing technologies, is the same under both policies π_{s_j} and $\pi_{s_{j+1}}$.

At time $t = s_{j+1}$, we can thus write

$$\begin{aligned}
&\text{Revenue}(\pi_{s_{j+1}}, t) - \text{Revenue}(\pi_{s_j}, t) \\
&\quad = (1 - F(p^*))(s_{j+1} - s_j)p^* + (d-1)(1 - F(\theta^*_{j+1}))\Big((s_{j+1} - s_j)\theta^*_{j+1} - c\Big)
\end{aligned} \tag{47}$$

Similarly, for $t = s_{j+1} + k, k \geq 0$, we can write

$$\begin{aligned}
&\text{Revenue}(\pi_{s_{j+1}}, t) - \text{Revenue}(\pi_{s_j}, t) \\
&= \min(k+1, d)\,(1 - F(p^*))\,(s_{j+1} - s_j)p^* + (d - \min(k+1, d))\Big(1 - F(\theta^*_{j+1})\Big)\Big((s_{j+1} - s_j)\theta^*_{j+1} - c\Big) \\
&= \min(k+1, d)\,(1 - F(p^*))\,\Delta s_{j+1} p^* \\
&\quad + (d - \min(k+1, d))\left[1 - F\left(v^{-1}\left(\frac{c}{\Delta s_{j+1}}\right)\right)\right]\left[\Delta s_{j+1} v^{-1}\left(\frac{c}{\Delta s_{j+1}}\right) - c\right],
\end{aligned} \tag{48}$$

where we define $\Delta s_{j+1} := s_{j+1} - s_j$.

We note that this revenue difference is a function of $s_{j+1} - s_j$. Multiplying (48) by the discount factor δ^t and summing over t for $t \geq s_{j+1}$, as well as noting that we can write $\delta^t = \delta^{s_{j+1}}\delta^k$ for $t = s_{j+1} + k$, implicitly defines a function $g(\cdot)$ on the positive integers, such that

$$\sum_{t=s_{j+1}}^{\infty} \delta^t \Big(\text{Revenue}(\pi_{s_{j+1}}, t) - \text{Revenue}(\pi_{s_j}, t)\Big) = \delta^{s_{j+1}} g(s_{j+1} - s_j),$$

where $g(\cdot)$ depends on d, F, c and δ but not on the introduction times $(s_1, \ldots, s_j, s_{j+1})$. Explicitly,

[22] Note that this includes both group 4 customers from policy π_{s_j}, and group 2 and group 3 customers from policy π_{s_j} who don't upgrade to technology class $j+1$.

since[23] $\sum_{k=0}^{\infty} \delta^k \min(k+1, d) = \frac{1-\delta^d}{(1-\delta)^2}$, we define

$$g(z) := \frac{1-\delta^d}{(1-\delta)^2} (1 - F(p^*))\, zp^* + \frac{\delta^d + d(1-\delta) - 1}{(1-\delta)^2} (1 - F(\theta^*(z)))\, (z\theta^*(z) - c)\,, \tag{49}$$

where $\theta^*(z) := v^{-1}\left(\frac{c}{z}\right)$, see (9).

Since $\text{Revenue}(\pi_{s_{j+1}}, t) = \text{Revenue}(\pi_{s_j}, t)$ for $t < s_{j+1}$, we can write

$$\sum_{t=s_1}^{\infty} \delta^t \Big(\text{Revenue}(\pi_{s_{j+1}}, t) - \text{Revenue}(\pi_{s_j}, t)\Big) = \delta^{s_{j+1}} g(s_{j+1} - s_j). \tag{50}$$

We recall that the utility of a policy π is $U(\pi) = \sum_{t=s_1}^{\infty} \delta^t \left(\text{Revenue}(\pi, t) - \text{Cost}(\pi, t)\right)$, hence we have

$$U(\pi_{s_{j+1}}) - U(\pi_{s_j}) = \delta^{s_{j+1}} \left(g(s_{j+1} - s_j) - C\right). \tag{51}$$

Pick an arbitrary policy $\pi = ((s_0 = 0, x_0 = 0), (s_i, \mathbf{x}_i^*)_{i=1}^{\infty})$ which has non-increasing introduction intervals, and uses optimal prices. For fixed j such that $s_{j+1} \geq s_1 + d - 1$, we can write the utility of policy π as

$$U(\pi) = U(\pi_{s_j}) + \left[\sum_{k=j+1}^{\infty} U(\pi_{s_k}) - U(\pi_{s_{k-1}})\right] \tag{52}$$

$$= U(\pi_{s_j}) + \sum_{k=j+1}^{\infty} \delta^{s_k} \left(g(s_k - s_{k-1}) - C\right), \tag{53}$$

[23] We explain this algebraic step in detail:

$$\begin{aligned}
\sum_{k=0}^{\infty} \delta^k \min(k+1, d) &= 1 + 2\delta + 3\delta^2 + \ldots + (d-1)\delta^{d-2} + d \sum_{k=d-1}^{\infty} \delta^k \\
&= 1 + 2\delta + 3\delta^2 + \ldots + (d-1)\delta^{d-2} + d\frac{\delta^{d-1}}{1-\delta} \\
&= \frac{(1-\delta)^2 + 2\delta(1-\delta)^2 + \ldots + (d-1)\delta^{d-2}(1-\delta)^2 + d\delta^{d-1}(1-\delta)}{(1-\delta)^2} \\
&= \frac{(1-\delta)\big[(1-\delta) + 2\delta(1-\delta) + \ldots + (d-1)\delta^{d-2}(1-\delta) + d\delta^{d-1}\big]}{(1-\delta)^2} \\
&= \frac{(1-\delta)\big[1 + \delta + \delta^2 + \ldots + \delta^{d-1}\big]}{(1-\delta)^2} \\
&= \frac{1-\delta^d}{(1-\delta)^2}
\end{aligned}$$

where π_{s_j} is the finite policy defined above, using the same introduction times s_i as π for $i \leq j$, but that has j as the last introduction. The second summand in (52) is a telescopic series, whose individual terms are bounded and decreasing, because g is an increasing function by Lemma 4 below, while by assumption $\Delta s_k = s_k - s_{k-1}$ are non-increasing with k.

We next introduce some definitions and notation. For any policy, we assume that prices are calculated optimally given the introduction times. We first define an operator that shifts all introduction times of a policy by a constant:

Definition 2. Fix a positive integer K. Given a policy π that introduces at times $s_0 = 0, s_1, s_2, \ldots$, we define $\mathcal{T}_K(\pi)$ to be the policy that introduces at times $s'_0 = 0, s'_1 = s_1 + K, s'_2 = s_2 + K, \ldots$.

Given a policy π that introduces at times $(s_0 = 0, s_1, \ldots, s_k, s_{k+1}, \ldots)$, and finite policy π_{s_k}, we denote by $\pi - \pi_{s_k}$ the policy with introduction times $(s'_1 = s_{k+1}, s'_2 = s_{k+2}, \ldots)$. That is, $\pi - \pi_{s_k}$ is the policy that uses introductions from π, with its first introduction at time s_{k+1}.

Let π^* denote a policy that is optimal among the policies with non-increasing introduction intervals. We denote its introduction times by (s_i^*), and its prices are calculated optimally given the introduction times. Fix j, with $s_{j+1}^* \geq s_1^* + d - 1$. Consider a policy π with non-increasing introduction intervals, which has the first j introductions timed optimally, and denote its introductions times by $\left(s_1^*, s_2^*, \ldots, s_j^*, s_{j+1}, s_{j+2}, \ldots\right)$, where we restrict $s_{j+1} \leq s_{j+1}^*$.

Consider now the policy $\tilde{\pi}$ that starts with introduction times $\left(s_1^*, s_2^*, \ldots, s_j^*\right)$, and then uses the introduction times of policy $\mathcal{T}_{\Delta s_{j+1}^*}(\pi - \pi^*_{s_{j-1}^*})$, where $\Delta s_{j+1}^* = s_{j+1}^* - s_j^*$. That is, the introduction times of policy $\tilde{\pi}$ are

$$\left(\tilde{s}_1 = s_1^*, \ldots, \tilde{s}_j = s_j^*, \tilde{s}_{j+1} = s_{j+1}^*, \tilde{s}_{j+2} = s_{j+1} + \Delta s_{j+1}^*, \tilde{s}_{j+3} = s_{j+2} + \Delta s_{j+1}^*, \ldots\right), \tag{54}$$

and its prices are calculated optimally given the introduction times. To denote the concatenation of the two sets of introduction times that make up $\tilde{\pi}$, we use notation

$$\tilde{\pi} = \left(\pi^*_{s_j^*}, \mathcal{T}_{\Delta s_{j+1}^*}(\pi - \pi^*_{s_{j-1}^*})\right).$$

Note that $\tilde{\pi}$ has non-increasing introduction intervals, as

$$s_2^* - s_1^* \geq s_3^* - s_2^* \ldots \geq s_{j+1}^* - s_j^* \geq s_{j+1} - s_j^* \geq s_{j+2} - s_{j+1} \geq s_{j+3} - s_{j+2} \ldots$$

We define

$$h(z) = g(z) - C. \tag{55}$$

Then we can write

$$U(\tilde{\pi}) = U(\pi^*_{s_j^*}) + \delta^{s_{j+1}^*} h(s_{j+1}^* - s_j^*) + \delta^{\Delta s_{j+1}^*} \delta^{s_{j+1}} h(s_{j+1} - s_j^*) + \delta^{\Delta s_{j+1}^*} \sum_{k=j+2}^{\infty} \delta^{s_k} h(s_k - s_{k-1}) \tag{56}$$

$$= U\left(\pi^*_{s_{j+1}^*}\right) + \delta^{\Delta s_{j+1}^*} \left[U(\pi) - U(\pi^*_{s_j^*})\right], \tag{57}$$

where the first line follows from (53); and the second line follows because $U(\pi^*_{s_j^*}) + \delta^{s_{j+1}^*} h(s_{j+1}^* - s_j^*) = U\left(\pi^*_{s_{j+1}^*}\right)$ by (51), while the rest of the summands in the right-hand side of (56) sum up to $\delta^{\Delta s_{j+1}^*} \left[U(\pi) - U(\pi^*_{s_j^*})\right]$ by (53).

Note that setting $\pi = \pi^*$ optimizes (57), because for fixed $s_1^*, \ldots, s_j^*, s_{j+1}^*$, optimizing (57) reduces to optimizing $U(\pi)$. Now any policy with non-increasing introduction intervals which has the first $j+1$ introduction times set optimally ($s_i = s_i^*$ for $i \leq j+1$) can be written as $\left(\pi^*_{s_j^*}, \mathcal{T}_{\Delta s_{j+1}^*}(\pi - \pi^*_{s_{j-1}^*})\right)$ for some policy π with non-increasing introduction intervals which has the first j introduction times set optimally ($s_i = s_i^*, i \leq j$).[24]

[24] We detail this construction for completeness. Start with policy π' that introduces at times $\left(s_1' = s_1^*, \ldots, s_j' = s_j^*, s_{j+1}' = s_{j+1}^*, s_{j+2}', s_{j+3}', \ldots\right)$, and has non-increasing introduction intervals:

$$s_2^* - s_1^* \geq \ldots \geq s_j^* - s_{j-1}^* \geq s_{j+1}^* - s_j^* \geq s_{j+2}' - s_{j+1}^* \geq s_{j+3}' - s_{j+2}' \geq \ldots$$

Then policy π that introduces at times $\left(s_1 = s_1^*, \ldots, s_j = s_j^*, s_{j+1} = s_{j+2}' - \Delta s_{j+1}^*, s_{j+2} = s_{j+3}' - \Delta s_{j+1}^*, \ldots\right)$ also has non-increasing introduction intervals:

$$s_2^* - s_1^* \geq \ldots \geq s_j^* - s_{j-1}^* \geq s_{j+2}' - s_{j+1}^* \geq s_{j+3}' - s_{j+2}' \geq \ldots,$$

and satisfies that $\left(\pi^*_{s_j^*}, \mathcal{T}_{\Delta s_{j+1}^*}(\pi - \pi^*_{s_{j-1}^*})\right) = \pi'$.

Hence, defining $\tilde{\pi}^* := \left(\pi^*_{s^*_j}, \mathcal{T}_{\Delta s^*_j}(\pi^* - \pi^*_{s^*_{j-1}})\right)$, we have

$$\begin{aligned} U(\tilde{\pi}^*) &\geq \sup_{s_{j+2}, s_{j+3}\dots} U\left(\pi : s_1 = s^*_1, \dots, s_{j+1} = s^*_{j+1}, s_{j+2}, s_{j+3}\dots\right) \\ &= U(\pi^*) \end{aligned}$$

This shows that policy $\tilde{\pi}^*$, which introduces at times

$$\left(\tilde{s}^*_1 = s^*_1, \dots, \tilde{s}^*_j = s^*_j, \tilde{s}^*_{j+1} = s^*_{j+1}, \tilde{s}^*_{j+2} = s^*_{j+1} + \Delta s^*_{j+1}, \tilde{s}^*_{j+3} = s^*_{j+2} + \Delta s^*_{j+1}, \dots\right),$$

and calculates prices optimally given these introduction times, is also optimal: $\tilde{\pi}^* \in \arg\max_\pi U(\pi)$. Note that policy $\tilde{\pi}^*$ has $\tilde{s}^*_{j+1} - \tilde{s}^*_j = \tilde{s}^*_{j+2} - \tilde{s}^*_{j+1} = \Delta s^*_{j+1}$.

Define recursively[25]

$$\tilde{\pi}^{*(k)} = \begin{cases} \pi^*, & \text{if } k = 0 \\ \left(\pi^*_{s^*_j}, \mathcal{T}_{\Delta s^*_j}(\tilde{\pi}^{*(k-1)} - \pi^*_{s^*_{j-1}})\right), & \text{if } k \geq 1. \end{cases} \tag{58}$$

For any fixed $k > 0$, policy $\tilde{\pi}^{*(k)}$, which introduces at times

$$\begin{gathered} \left(\tilde{s}^{*(k)}_1 = s^*_1, \dots, \tilde{s}^{*(k)}_j = s^*_j, \tilde{s}^{*(k)}_{j+1} = s^*_{j+1}, \tilde{s}^{*(k)}_{j+2} = s^*_{j+1} + \Delta s^*_{j+1}, \tilde{s}^{*(k)}_{j+3} = s^*_{j+1} + 2\Delta s^*_{j+1}, \dots, \right. \\ \left. \tilde{s}^{*(k)}_{j+k+1} = s^*_{j+1} + k\Delta s^*_{j+1}, \tilde{s}^{*(k)}_{j+k+2} = s^*_{j+2} + k\Delta s^*_{j+1}, \tilde{s}^{*(k)}_{j+k+3} = s^*_{j+3} + k\Delta s^*_{j+1}, \dots\right), \end{gathered}$$

and calculates prices optimally given these introduction times, is also optimal: $\tilde{\pi}^{*(k)} \in \arg\max_\pi U(\pi)$. Note that policy $\tilde{\pi}^{*(k)}$ has introduction intervals

$$\tilde{s}^{*(k)}_{j+1} - \tilde{s}^{*(k)}_j = \tilde{s}^{*(k)}_{j+2} - \tilde{s}^{*(k)}_{j+1} = \dots = \tilde{s}^{*(k)}_{j+k+1} - \tilde{s}^{*(k)}_{j+k} = \Delta s^*_{j+1}.$$

In the final step of the proof we will show that the sequence of policies defined in (58) converges to a policy that is periodic after time s^*_j and is optimal. Remember that we restrict to policies with non-increasing introduction intervals and that use optimal pricing given the introduction times.

[25] We have $\tilde{\pi}^{*(1)} = \tilde{\pi}^*$.

Therefore a policy is characterized by its introduction times.

For the remainder of the proof, we formally define the introduction times of a policy as an infinite binary sequence. We use $\boldsymbol{\sigma}^{\pi}$ to denote the binary sequence for the introduction times of a policy π which introduces at times $\mathbf{s}$ and that uses optimal pricing. In particular, $\sigma_t^{\pi} = 1$ ($\sigma_t^{\pi} = 0$) means that policy π makes (does not make) an introduction in period t.[26]

We next define a metric space. Consider the set of all infinite binary sequences: $\{0,1\}^{\mathbb{N}}$. On that set, we define the following metric:

$$d(\mathbf{x}, \mathbf{y}) = 2^{-\max\{n:\ x_i = y_i\ \ \forall i \leq n\}}. \tag{59}$$

Consider the sequence $\{\boldsymbol{\sigma}^{\tilde{\pi}^{*(k)}}\}$, where $\tilde{\pi}^{*(k)}$ is the policy defined in (58) for a fixed k. Denote by $\tilde{\pi}^{*(\infty)}$ the policy with introduction times

$$\Big(\tilde{s}_1^{*(\infty)} = s_1^*, \ldots, \tilde{s}_j^{*(\infty)} = s_j^*, \tilde{s}_{j+1}^{*(\infty)} = s_{j+1}^*, \tilde{s}_{j+2}^{*(\infty)} = s_{j+1}^* + \Delta s_{j+1}^*, \tilde{s}_{j+3}^{*(\infty)} = s_{j+1}^* + 2\Delta s_{j+1}^*, \ldots,$$
$$\tilde{s}_{j+k+1}^{*(\infty)} = s_{j+1}^* + k\Delta s_{j+1}^*, \ldots\Big),$$

i.e., that has periodic introduction intervals after time s_j^* ad infinitum, and that has optimal prices.

We can see that for every $\epsilon > 0$ there is an integer N such that $k \geq N$ implies that

$$d\left(\boldsymbol{\sigma}^{\tilde{\pi}^{*(k)}}, \boldsymbol{\sigma}^{\tilde{\pi}^{*(\infty)}}\right) < \epsilon.$$

Therefore, the sequence $\{\boldsymbol{\sigma}^{\tilde{\pi}^{*(k)}}\}$ converges to $\boldsymbol{\sigma}^{\tilde{\pi}^{*(\infty)}}$ in the defined metric space; we write

$$\lim_{k \to \infty} \boldsymbol{\sigma}^{\tilde{\pi}^{*(k)}} = \boldsymbol{\sigma}^{\tilde{\pi}^{*(\infty)}}.$$

We next argue that the utility U is a continuous mapping from the set of infinite binary sequences with the metric defined in (59), into $\mathbb{R}$ with the usual metric $d(x, y) = |x - y|$. Take two policies π_1, π_2 with corresponding introduction times and optimal prices given by their respective introduction times. Take time ν to be the latest time such that the two policies are identical up to that time,

[26] For policy π with introduction times $\mathbf{s}$, $s_j = t$ is equivalent to: (i) $\sigma_t^{\pi} = 1$, and (ii) $\sum_{i=1}^{t} \sigma_i^{\pi} = j$.

i.e., $d(\boldsymbol{\sigma}^{\pi_1}, \boldsymbol{\sigma}^{\pi_2}) = 2^{-\nu}$. Then we can bound the difference in the utilities as follows:

$$\begin{aligned}
|U(\pi_1) - U(\pi_2)| &\leq \sum_{t=\nu+1}^{\infty} \delta^t d\,(1-F(p^*))\,tp^* + C \sum_{t=\nu+1}^{\infty} \delta^t \\
&= d\,(1-F(p^*))\,p^* \cdot \sum_{t=\nu+1}^{\infty} t\delta^t + C \sum_{t=\nu+1}^{\infty} \delta^t \\
&= d\,(1-F(p^*))\,p^* \cdot \frac{\delta^{\nu+1}}{1-\delta}\left(\nu + \frac{1}{1-\delta}\right) + C\frac{\delta^{\nu+1}}{1-\delta} \\
&= \frac{\delta^{\nu+1}}{1-\delta}\left(d\,(1-F(p^*))\,p^*\left(\nu + \frac{1}{1-\delta}\right) + C\right)
\end{aligned}$$

where the last step follows after some algebra.[27] This bound can become arbitrarily small by setting a large enough ν, showing that U is uniformly continuous with the respect to the defined metric spaces.[28]

Having argued that U is uniformly continuous and therefore continuous with respect to the defined metric spaces, we can write that $U(\tilde{\pi}^{*(\infty)}) = \lim_{k\to\infty} U(\tilde{\pi}^{*(k)})$. Because for each k, the policy $\tilde{\pi}^{*(k)}$ is optimal, $\tilde{\pi}^{*(\infty)}$ is also optimal. Policy $\tilde{\pi}^{*(\infty)}$ has periodic introductions starting at its jth introduction, with $\tilde{s}^{*(\infty)}_{j+1} \geq \tilde{s}^{*(\infty)}_1 + d - 1$, and the statement follows.

□

Lemma 4. *$g(\cdot)$ is an increasing function.*

[27]We detail here the algebra for the last step:

$$\begin{aligned}
\sum_{t=\nu+1}^{\infty} t\delta^t &= \sum_{t'=1}^{\infty}(t'+\nu)\delta^{t'+\nu} \\
&= \delta^\nu\left(\nu\sum_{t'=1}^{\infty}\delta^{t'} + \sum_{t'=1}^{\infty} t'\delta^{t'}\right) \\
&= \delta^\nu\left(\frac{\nu\delta}{1-\delta} + \frac{\delta}{(1-\delta)^2}\right) \\
&= \frac{\delta^{\nu+1}}{1-\delta}\left(\nu + \frac{1}{1-\delta}\right)
\end{aligned}$$

[28]In fact, it can be shown that the defined metric space of infinite binary sequences is compact, and therefore the concepts of continuity and uniform continuity are equivalent for mappings from that metric space.

Proof. For $F(\theta^*(z)) < 1$, we have

$$
\begin{aligned}
\frac{d}{dz}\left(1 - F(\theta^*(z))\right)\left(z\theta^*(z) - c\right) &= -f(\theta^*(z))\frac{d\theta^*(z)}{dz}\left(z\theta^*(z) - c\right) + \left(1 - F(\theta^*(z))\right)\left(\theta^*(z) + z\frac{d\theta^*(z)}{dz}\right) \\
&= \left(1 - F(\theta^*(z))\right)\theta^*(z) + \frac{d\theta^*(z)}{dz}\Big[\left(1 - F(\theta^*(z))\right)z - f(\theta^*(z))\left(z\theta^*(z) - c\right)\Big] \\
&= \left(1 - F(\theta^*(z))\right)\theta^*(z), \qquad (60) \\
&\geq 0,
\end{aligned}
$$

where (60) follows because $\Big[\left(1 - F(\theta^*(z))\right)z - f(\theta^*(z))\left(z\theta^*(z) - c\right)\Big] = 0$ by the definition of $\theta^*(z) := v^{-1}\left(\frac{c}{z}\right)$. If $F(\theta^*(z)) = 1$, then we have that $\left(1 - F(\theta^*(z))\right)\left(z\theta^*(z) - c\right) = 0$.

Overall, by the definition of function g in (49), it follows that

$$
g'(z) = \frac{1 - \delta^d}{(1-\delta)^2}\left(1 - F(p^*)\right)p^* + \frac{\delta^d + d(1-\delta) - 1}{(1-\delta)^2}\left(1 - F(\theta^*(z))\right)\theta^*(z) > 0,
$$

because[29] $\delta^d + d(1-\delta) - 1 > 0$. Therefore g is increasing.

□

H Proof of Corollary 1

Proof. Using the notation in the proof of Theorem 2, and repeating (53), we write

$$
\begin{aligned}
U(\pi) &= U(\pi_{s_j}) + \sum_{k=j+1}^{\infty} \delta^{s_k}\left(g(s_k - s_{k-1}) - C\right) \\
&= U(\pi_{s_j}) + \sum_{k=j+1}^{\infty} \delta^{s_k}\left(h(s_k - s_{k-1})\right)
\end{aligned}
$$

for a policy π with non-increasing introduction intervals, and for j such that $s_{j+1} \geq s_1 + d - 1$. We apply this to policy π_T^*: a policy with introduction times $\left(s_1^*, \ldots, s_j^*, s_j^* + T, s_j^* + 2T, \ldots\right)$, which has introduction intervals of length T starting at time s_j^*, where (s_i^*) are the introduction times of

[29]To see this, observe that

$$
\delta^d + d(1-\delta) - 1 > 0 \iff d > \frac{1-\delta^d}{1-\delta} \iff d > 1 + \delta + \delta^2 + \ldots + \delta^{d-1},
$$

which holds as long as $0 \leq \delta < 1$ and integer $d > 1$.

a policy that is optimal among the policies with non-increasing introduction intervals. We get

$$U(\pi_T^*) = U(\pi_{s_j^*}^*) + \delta^{s_j^*} \sum_{k=1}^{\infty} \delta^{kT} h(T) \tag{61}$$

$$= U(\pi_{s_j^*}^*) + \delta^{s_j^*} \frac{\delta^T}{1-\delta^T} h(T). \tag{62}$$

By Theorem 2, there is an optimal policy with the following property: all introductions made at or after time $t = s_1 + d - 1$ are periodic, with each introduction a constant interval from its previous introduction. Denoting its period by T^*, it follows that T^* satisfies

$$T^* \in \arg\max_{T \in \mathbb{N}^+} \frac{\delta^T}{1-\delta^T} h(T). \tag{63}$$

We note that $\frac{\delta^T}{1-\delta^T} h(T)$ is bounded, because function $h(z)$ is bounded above by a linear function of z for fixed δ, d, c, C and distribution F.

□

I Details of Characterizing the Optimal Period and of the Uniform Distribution Example

We have

$$\frac{d}{dz}\left(\frac{\delta^z}{1-\delta^z} h(z)\right) = \frac{\delta^z}{(1-\delta^z)^2}\left(\log\delta \; h(z) + (1-\delta^z) h'(z)\right)$$

and a turning point z^* satisfies the necessary condition

$$\log\delta \; h(z^*) + (1-\delta^{z^*}) h'(z^*) = 0. \tag{64}$$

We calculate the second derivative to be

$$\frac{d^2}{dz^2}\left(\frac{\delta^z}{1-\delta^z} h(z)\right) = \frac{\delta^z \left((1-\delta^z)\left((1-\delta^z)\, h''(z) + 2\log\delta \; h'(z)\right) + (1+\delta^z)\log^2(\delta) h(z)\right)}{(1-\delta^z)^3}, \tag{65}$$

hence at turning point $z = z^*$, using (64) to replace

$$(1+\delta^z)\log^2(\delta)h(z) = -\left(1+\delta^z\right)\left(1-\delta^z\right)\log\delta\ h'(z)$$

and simplifying, we have that

$$\frac{d^2}{dz^2}\left(\frac{\delta^z}{1-\delta^z}h(z)\right) = \frac{\delta^z}{1-\delta^z}\left(\log\delta\ h'(z) + h''(z)\right). \tag{66}$$

We restrict to the uniform distribution on $[0,1]$. When f is the density of the uniform distribution on $[0,1]$, then $\zeta = 1$, $p^* = 1/2$, $\theta^*(z) = \min\left(\frac{1}{2}\left(\frac{c}{z}+1\right), 1\right)$, and explicitly,

$$h_{U[0,1]}(z) = \begin{cases} \frac{1-\delta^d}{(1-\delta)^2}\cdot\frac{z}{4} + \frac{\delta^d + d(1-\delta)-1}{(1-\delta)^2}\cdot\frac{(z-c)^2}{4z} - C & \text{if } z > c \\ \frac{1-\delta^d}{(1-\delta)^2}\cdot\frac{z}{4} - C & \text{if } z \le c. \end{cases}$$

We can naturally extend h to take values on the positive reals. $h(z)$ is upper bounded by a linear function, implying $\frac{\delta^z}{1-\delta^z}h(z)$ has a finite maximum at z^*. Note that for $0 \le \delta < 1$ and $d > 1$, we have $\delta^d + d(1-\delta) - 1 > 0$.

There are potentially two local maxima of $\frac{\delta^z}{1-\delta^z}h(z)$, one in $z \le c$ and one in $z > c$, where the local maximum can be a turning point or the left-hand boundary of the interval.

For $z \le c$, it is straightforward to see that the right-hand side of (66) is negative, since $h'' = 0$. Hence for $z \le c$ a turning point is a local maximum. It is also possible to show that $\frac{\delta^z}{1-\delta^z}h(z)$ can decrease over the interval $z \le c$.

For $z > c$, analytical results are no longer straightforward. For example, showing a turning point that is a local maximum analytically requires $z >> c$, since there can be a local minimum in $z > c$. However, it is simple to numerically evaluate the function and describe its qualitative behavior.

Specific examples of parameter settings: We next provide some specific examples of parameter settings for the different scenarios discussed for the uniform distribution on $[0,1]$ in Section 5.1.1.

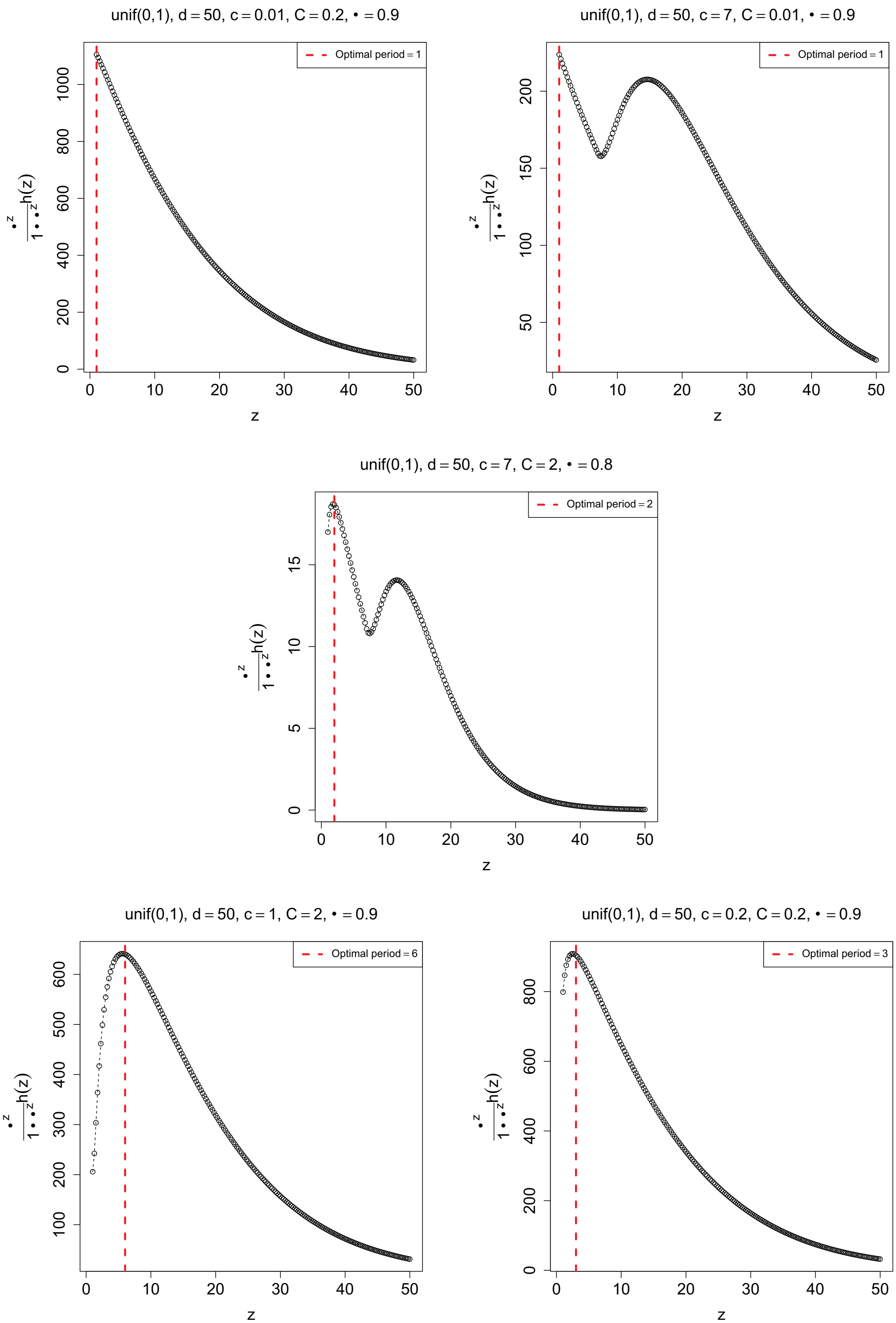

Figure 4: Examples that illustrate the behavior of the optimal periodicity T^* across different cases for the parameters d, c, C, δ. The red vertical dashed line indicates the optimal period T^*.

- $T^* = 1$ (Figure 4, top row).
  - Example with small c — almost all customers always upgrade: $d = 50, c = 0.01, C = 0.2, 0 < \delta < 1$.
  - Example with large c — no customer upgrades: $d = 50, c = 7, C = 0.01, \delta = 0.9$. In this case $\frac{\delta^z}{1-\delta^z}h(z)$ decreases for $z \leq c$, then increases after a local minimum, and has a local maximum for $z^* = 14.6$ ($T^* = 15$). The value at the local maximum however is less than the value at $z = 1$.
- $1 < T^* \leq c$ (Figure 4, middle row). No customer ever upgrades. For $d = 50, c = 7, C = 2, \delta = 0.8$, the optimal period is $T^* = 2$.
- $T^* > \max(c, 1)$ (Figure 4, bottom row). Some customers upgrade.
  - For $d = 50, c = 1, C = 2$, the optimal period increases with δ, from $T^* = 2$ for small δ to $T^* = 9$ as δ approaches 1.
  - For $d = 50, c = C = 0.2, \delta = 0.9$, the optimal period is $T^* = 3$. Despite the small provisioning cost C, the optimal period is not 1, i.e., it is best to not introduce in every period.
- Two different values of the period that are local maxima, one smaller and the other larger than the switching cost c, can yield near identical values of the function $\frac{\delta^z}{1-\delta^z}h(z)$. Figure 5 illustrates.
  - For $d = 50, c = 7, C = 5$, at $\delta = 0.82$ the optimal strategy is to set $T^* = 3$, where no customer upgrades.
  - For $d = 50, c = 7, C = 5$, at $\delta = 0.83$, the optimal period is $T^* = 12$, yielding $\theta^* = 0.79$, where users with θ larger than 0.79 will upgrade.

    At $\delta = 0.82$, the difference in the value of the function $\frac{\delta^z}{1-\delta^z}h(z)$ between the two policies (introduce with period 3 or with period 12) is 1.3%, and at $\delta = 0.83$ the difference is 0.1%.[30]

[30]The value of the function $\frac{\delta^z}{1-\delta^z}h(z)$ is equal for the two policies (introduce with period 3 or with period 12) for δ around 0.8293.

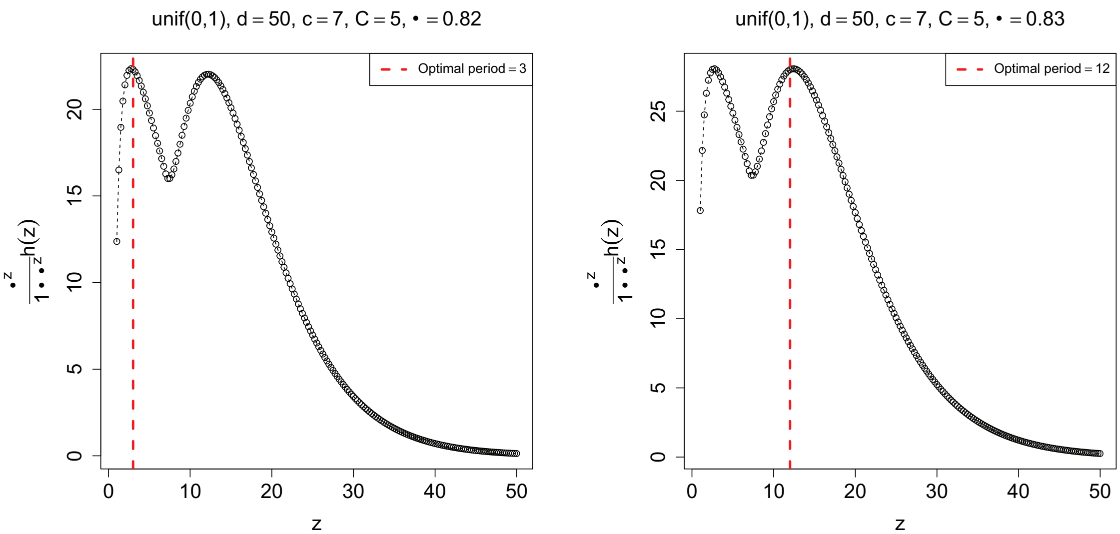

Figure 5: The case when two different values of the period that are local maxima, one smaller and the other larger than the switching cost c, can yield near identical values of the function $\frac{\delta^z}{1-\delta^z}h(z)$. The red vertical dashed line indicates the optimal period T^*.

J Proof of Theorem 3

We first state and prove the following lemma.

Lemma 5. *For any policy π with introduction times $(s_1, s_2, \ldots,)$ that uses optimal pricing, for $j \geq 1$ we have*

$$U(\pi_{s_{j+1}}) - U(\pi_{s_j}) \leq \delta^{s_{j+1}} \left(g(s_{j+1} - s_j) - C\right). \tag{67}$$

Proof. At time $t \geq s_{j+1}$, under policy $\pi_{s_{j+1}}$, users who purchase technologies can be classified as newly arrived, upgraders from technology s_j, or those who stick with existing technology s_j or previous technologies. Only the first two types of customers generate more revenue at t compared to policy π_{s_j}. The incremental revenue per customer for each of these two types is, respectively, $(1 - F(p^*))\,(s_{j+1} - s_j)p^*$ and $\left(1 - F\left(\theta^{(i)*}_{j+1}\right)\right)\left((s_{j+1} - s_j)\theta^{(i)*}_{j+1} - c\right)$, where $\theta^{(i)*}_{j+1}$ is the optimal upgrade threshold at time s_{j+1} for a user for whom technology $j+1$ is the i^{th} upgrade, $i = 0, 1, \ldots, n$. But

$$\left(1 - F\left(\theta^{(i)*}_{j+1}\right)\right)\left((s_{j+1} - s_j)\theta^{(i)*}_{j+1} - c\right) \leq \left(1 - F(\theta^{FOC}_{j+1})\right)\left((s_{j+1} - s_j)\theta^{FOC}_{j+1} - c\right)$$

where $\theta^{FOC}_{j+1} := v^{-1}\left(\frac{c}{s_{j+1}-s_j}\right)$ is the value of the threshold θ that maximizes $(1 - F(\theta))\left((s_{j+1} - s_j)\theta - c\right)$.

Because $g(s_{j+1} - s_j)$ is defined in (49) to use θ_{j+1}^{FOC}, the result follows. □

We now prove Theorem 3.

Proof. Consider a policy π with introduction times $(s'_1, \dots, s'_j, s'_{j+1}, \dots)$, and the derived policy $\pi_{s'_j}$ with the last introduction at s'_j. We have

$$U(\pi) = U(\pi_{s'_j}) + \left[\sum_{i=j+1}^{\infty} U(\pi_{s'_i}) - U(\pi_{s'_{i-1}}) \right],$$

which, combined with Lemma 5, implies

$$U(\pi) - U(\pi_{s'_j}) \leq \sum_{i=j+1}^{\infty} \delta^{s'_i} \left(g(s'_i - s'_{i-1}) - C \right) \tag{68}$$

$$= \sum_{i=j+1}^{\infty} \delta^{s'_i} h(s'_i - s'_{i-1}). \tag{69}$$

By the arguments in the proof of Theorem 2 and the proof of Corollary 1, the right-hand side is maximized by having $s'_i - s'_{i-1} = T^*$ for $i \geq j+1$.

Let $\hat{\pi}$ denote the policy with introduction times $(\hat{s}_1 = s_1, ..., \hat{s}_{j-1} = s_{j-1}, \hat{s}_j = s_j, \hat{s}_{j+1} = s_j + T^*, \hat{s}_{j+2} = s_j + 2T^*, \dots, \hat{s}_{j+m} = \tilde{t} = s_j + mT^*, \dots)$, and optimal pricing. Then we have

$$U(\hat{\pi}) - U(\hat{\pi}_{\hat{s}_{j+m}}) = \sum_{i=j+m+1}^{\infty} \delta^{\hat{s}_i} h(T^*) \tag{70}$$

$$= \delta^{\hat{s}_{j+m}} \sum_{i=1}^{\infty} \delta^{iT^*} h(T^*) \tag{71}$$

$$\geq \sum_{i=k+1}^{\infty} \delta^{s'_i} h(s'_i - s'_{i-1}) \tag{72}$$

$$\geq U(\pi') - U(\pi'_{s'_k}) \tag{73}$$

for any policy π' with introductions at times $(s'_1 = s_1, \dots, s'_j = s_j, s'_{j+1}, \dots, s'_k = \tilde{t}, s'_{k+1}, \dots)$ and optimal pricing that has its k-th introduction at $\tilde{t}$ for some $k \geq j$. The first equality follows because all customers present at time $t \geq \tilde{t}$ have arrived within the last d periods, when all the introduction times were periodic and hence non-increasing under $\hat{\pi}$. The last inequality follows from (69).

Consider a policy π with its first j introduction times fixed at given $(s_1, \ldots, s_j)$, and the rest denoted by $s_{j+1}, \ldots$. For $t \geq s_j + d - 1$, policy π_{s_j} just depends upon s_j (as under π_{s_j} new arrivals are offered technology s_j at price $s_j p^*$, and no one upgrades). We have that (48) holds exactly, and therefore

$$\sum_{t=s_{j+m}}^{\infty} \delta^t \left(U(\pi, t) - U(\pi_{s_j}, t) \right) = \sum_{i=j+m+1}^{\infty} \delta^{s_i} \left(g(s_i - s_{i-1}) - C \right). \tag{74}$$

By the arguments in the proof of Theorem 2, $\hat{\pi}$ maximizes the right-hand side. This implies that

$$\hat{\pi} \in \arg\max_{\pi} \sum_{t=\tilde{t}}^{\infty} \delta^t U(\pi, t),$$

where the supremum is taken over all policies π which have the first j introductions $(s_1, \ldots, s_j)$ fixed. $\square$

K Details of the Experimental Setup for Experiments on Periodicity

We describe in detail the schemes we use to generate introduction patterns.

- Random introduction intervals (*Random*): We generate 1000 random introduction patterns. For each pattern, each introduction interval is drawn uniformly at random and i.i.d. from the discrete uniform distribution whose support is the set $\{1, 2, \ldots, 2d\}$.
- Periodic introductions (*Per*): All introduction intervals are set to be equal. For each d, we produce periodic introduction patterns for all values of the period between 1 and $2d$.
- Non-increasing introduction intervals (*NI*): We randomly generate 1000 random introduction patterns with introduction intervals that are non-increasing: $s_i - s_{i-1} \geq s_{i+1} - s_i$, for $i \geq 2$. To do this we first sample the first interval $first\_interval$ uniformly at random from range $[1, 2d]$; we then sample independently sufficiently many intervals uniformly from the range $[1, first\_interval]$ and sort them in non-increasing order.[31] We only retain introduction patterns that are not periodic.

[31] We follow a similar process for sampling monotonic intervals for other schemes.

- Non-decreasing introduction intervals (*ND*): We randomly generate 1000 random introduction patterns with introduction intervals that are non-decreasing: $s_i - s_{i-1} \le s_{i+1} - s_i$, for $i \ge 2$. We only retain introduction patterns that are not periodic.

- Non-increasing introduction intervals in the warm-up; periodic introductions after (*NI-Per*): We randomly generate 1000 introduction patterns with introduction intervals that are non-increasing during the warm-up phase, and periodic after. Consistently with Theorem 2, periodic intervals are imposed starting with the first interval whose end falls at or after time $s_1 + d - 1$. We use this scheme as the benchmark against which we compare the utilities achieved by the other schemes.

- Non-increasing introduction intervals in the warm-up; non-increasing introduction intervals after (*NI-NI*): We randomly generate 1000 introduction patterns with introduction intervals that are non-increasing during the warm-up phase, and non-increasing also in the continuation. We build two different patterns of non-increasing introduction intervals: one for the warm-up and one for the continuation. That is, we allow the transition from the warm-up phase to the continuation to violate the monotonicity of the introduction differences. We stop building the warm-up when adding one more interval in the warm-up would result in reaching at or after time $s_1 + d - 1$ — we don't add that last interval in the warm-up, and start the continuation instead. We only retain introduction patterns that are not periodic after the warm-up phase.

- Non-decreasing introduction intervals in the warm-up; periodic introductions after (*ND-Per*): We randomly generate 1000 introduction patterns with introduction intervals that are non-decreasing during the warm-up phase, and periodic after. Consistently with Theorem 2, periodic intervals are imposed starting with the first interval whose end falls at or after time $s_1 + d - 1$.

- Non-decreasing introduction intervals in the warm-up; non-decreasing introduction intervals after (*ND-ND*): We randomly generate 1000 introduction patterns with introduction intervals that are non-decreasing during the warm-up phase, and non-decreasing after. We build two different patterns of non-decreasing introduction intervals: one for the warm-up and one for the continuation. That is, we allow the transition from the warm-up phase to the continuation to violate the monotonicity of the introduction differences. We stop building the warm-up when adding one more interval in the warm-up would result in reaching at or after time $s_1 + d - 1$ — we don't add that last interval in the warm-up, and start the continuation instead. We only retain introduction patterns that are not periodic after the warm-up phase.

- Introduction intervals from grid search (*Grid*): We pick a total of 2000 introduction patterns over the space of all possible patterns; we next describe how. Having fixed $s_1 = 1$, there are 2^{H-1} possible introduction patterns until time H: in each period $2, 3, \ldots, H$, one has the choice of introducing or not introducing. To ensure numerical precision, we first focus only on the first 53 periods following $s_1 = 1$. We first pick 1000 integers, equally spaced, from the set $\{0, 1, 2, \ldots, 2^{53} - 1\}$. These 1000 integers are the set

$$\left\{ \left\lceil \frac{n}{999} \left(2^{53} - 1\right) \right\rceil : n = 0, 1, \ldots, 999 \right\}.$$

  We convert each picked integer to two binary numbers; each binary number is a binary vector, so that a "1" represents an introduction at that time, whereas a "0" means no introduction. We next detail the two ways to produce a binary vector from a picked integer:

  - Implement a binary vector that grows from right to left — i.e, the rightmost bit corresponds to 2^0. We then add a 1 followed by as many 0's as necessary in the beginning (left) of each vector to ensure each vector has length 54, representing an introduction pattern in the first 54 periods.
  - Implement a binary vector that grows from left to right — i.e, the leftmost bit corresponds to 2^0. We then add a 1 in the beginning (left), and as many 0's as necessary at the end (right) of each vector to ensure each vector has length 54, representing an introduction pattern in the first 54 periods.

  For each binary vector, we repeat the derived introduction pattern as many times as necessary to reach to a specified multiple of the end of the horizon.

- Introduction intervals from grid search on the exponent (*Log-Grid*): We pick a total of 2000 introduction patterns over the space of all possible patterns, this time doing a grid search on the *exponent*. The process for *Log-Grid* mirrors the process described above for *Grid*, with the only difference that the 1000 initially picked integers are the set

$$\left\{\left\lceil 2^{\left(\frac{n}{999}\cdot 53\right)}-1\right\rceil : n=0,1,\ldots,999\right\}.$$

L Linear Pricing in the Non-Discriminatory Pricing Setting

In this section we analyze a simple, natural pricing policy: charge a price which is linear in the quality of the technology class. We show this has several nice properties. First, with linear pricing all newly arriving customers will select the latest quality. Second, with linear pricing the optimal policy has a periodic pattern of introductions. A particularly interesting special case is one where the linear prices are chosen to be optimal for each technology class as if it were the only item offered for sale, as per Myerson's approach. Third, we show that 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.

Formally, we define *linear pricing with base price* $p > 0$ to be as follows: set the price at introduction time s_j to $x_j = s_j p$, with $j = 1, 2, \ldots$. Assuming linear pricing with base price $p > 0$, *new* customers prefer buying technology $j \geq 1$ to buying nothing if $s_j(\theta - p) \geq 0$, or $\theta \geq p$, so the set of customers willing to buy each technology is the same. Customers prefer technology $j \geq 2$ 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. *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 $\theta \geq p + c/(s_j - s_{j-1})$.

We next define p^* as the price that maximizes the single-item expected revenue in one period:

$$p^* := \arg\max_p (1 - F(p)) \cdot p.$$

We refer to *Myerson pricing* as the special case of linear pricing with base price p^*. Myerson pricing, which sells to the set of customers with $\theta \geq p^*$, is optimal for new customers.

Assuming a policy π_p which follows linear pricing with base price p, we write down the revenue for the provider at time t with $s_j \leq t < s_{j+1}, j \geq 2$:

$$\begin{aligned}
\text{Revenue}(\pi_p, t) = &\min(t - s_j + 1, d)\,(1 - F(p))\, s_j p \\
&+ \max\left(d - (t - s_j + 1), 0\right) \left\{ \left(1 - F\left(p + \frac{c}{s_j - s_{j-1}}\right)\right) s_j p \right. \\
&\left. + \left(F\left(p + \frac{c}{s_j - s_{j-1}}\right) - F(p)\right) s_{j-1} p \right\}. \qquad (75)
\end{aligned}$$

The first summand is the revenue from customers who arrive at or after period s_j. These customers buy technology j as long as $\theta \geq p$. The term $(1 - F(p))\, s_j p$ is the per-period expected revenue for customers who arrive at or after s_j, and within the last $d-1$ periods before, or at, period t. There are $\min(t - s_j + 1, d)$ such terms.

The second summand is the revenue from customers who arrive before period s_j. Some of them switch to technology j at period s_j; these are the customers for whom $\theta \geq p + c/(s_j - s_{j-1})$ (first term inside the curly brackets). Some of them stick to technology $j-1$; these are the customers for whom $p \leq \theta < p + c/(s_j - s_{j-1})$ (second term inside the curly brackets). The same terms apply to all customers who arrive in each of the periods before s_j, and within the last $d-1$ periods before period t. There are $d - \min(t - s_j + 1, d) = \max\left(d - (t - s_j + 1), 0\right)$ such periods.

For $j = 1$, we can write

$$\text{Revenue}(\pi_p, t) = \min(t - s_1 + 1, d)\,(1 - F(p))\, s_1 p. \qquad (76)$$

For the special case of periods when an introduction occurs, i.e., $t = s_j$ for some $j \geq 1$, expression (75) becomes

$$\text{Revenue}(\pi_p, s_j) = \begin{cases} (1 - F(p))\, s_j p + (d-1) \left\{ \left(1 - F\left(p + \frac{c}{s_j - s_{j-1}}\right)\right) s_j p \right. & \\ \quad \left. + \left(F\left(p + \frac{c}{s_j - s_{j-1}}\right) - F(p)\right) s_{j-1} p \right\}, & j \geq 2 \\ (1 - F(p))\, s_1 p, & j = 1. \end{cases} \qquad (77)$$

L.1 Under linear pricing, periodic introductions are optimal

Our first result shows that there exists a policy that is optimal within the class of policies that use linear 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 linear pricing, periodic introductions are optimal after the first introduction. In particular, this applies to Myerson pricing.*

The proof is given in Appendix M.

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 $\delta^s C < \text{Revenue}((0,0),(s,sp))$, which implies

$$
\begin{aligned}
C &< \frac{1}{\delta^s}\left\{\delta^s\left(1-F(p)\right)sp + \sum_{t=s+1}^{s+d-2}\delta^t(t-s+1)\left(1-F(p)\right)sp + \sum_{t=s+d-1}^{\infty}\delta^t d\left(1-F(p)\right)sp\right\}\\
&= \frac{1}{\delta^s}\delta^s\left(1+2\delta+3\delta^2+\ldots+(d-1)\delta^{d-2}+\frac{d\delta^{d-1}}{1-\delta}\right)\left(1-F(p)\right)sp\\
&= \left(1+2\delta+3\delta^2+\ldots+(d-1)\delta^{d-2}+\frac{d\delta^{d-1}}{1-\delta}\right)\left(1-F(p)\right)sp.
\end{aligned}
$$

Hence, given a base price level p, 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. In the rest of Section 6 we show that Myerson pricing guarantees a bounded approximation ratio to both revenue and cost for the infinite-horizon problem, having fixed the introduction times.

L.2 Under periodic introductions, Myerson pricing is optimal for one-period revenue in the limit of many introductions

We have shown that under Myerson pricing, periodic introductions are optimal. In the rest of Section 6 we will show the near-optimality of Myerson pricing even with arbitrary introduction times. Together, these results yield the insight that a simple policy combining Myerson pricing with periodic introductions is effective.

Before we show this, we first argue for the efficacy of Myerson pricing from a different angle: Myerson pricing is effective when introductions are periodic. Under periodic introduction times, an alternate pricing scheme may have revenue gains over Myerson pricing early in the horizon; however such gains vanish after sufficiently many introductions.

In particular, shading prices down from the Myerson levels 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 introduction times as π, but Myerson pricing. Then we have*

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

Proof. Fix the periodicity of introductions $\tau > 0$ and the introduction times. Using the alternative expression (78), the revenue of policy π_M, which uses Myerson pricing, at period s_j can be written as

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

for $j \geq 2$. 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 $d\left(1 - F(p^*)\right) s_j p^*$, so

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

□

The key insight is that the potential gains from alternate prices can be bounded in terms of the length of the periodicity used by policy π, independent of introduction index j. As a result, assuming periodic introductions, the potential additional revenue of any pricing 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 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 for one-period revenue, after sufficient introductions.

M Proof of Proposition 1

Proof. Let $\pi_p^* = ((s_0 = 0, x_0 = 0), (s_1 = s_1^*, x_1 = s_1^* p), (s_2 = s_2^*, x_2 = s_2^* p), (s_3 = s_3^*, x_3 = s_3^* p), \ldots)$ denote a policy which is optimal among those policies using linear pricing with base price p. We first show that such a policy exists.

For an arbitrary policy using linear pricing with base price p,

$$\pi_p = ((s_0 = 0, x_0 = 0), (s_1, x_1 = s_1 p), (s_2, x_2 = s_2 p), (s_3, x_3 = s_3 p), \ldots),$$

we have that

$$U(\pi_p) = \text{Revenue}(\pi_p) - \text{Cost}(\pi_p) = \sum_{t=s_1}^{\infty} \delta^t \left(\text{Revenue}(\pi_p, t) - \text{Cost}(\pi_p, t)\right),$$

i.e., revenue accumulated less costs incurred during period s_1 and subsequent periods. Optimal policy π_p^* must optimize $U(\pi_p)$.

For policy π_p, we can write Equation (75) as

$$\begin{aligned}\text{Revenue}(\pi_p, t) =& \min(t - s_j + 1, d)\,(1 - F(p))\, s_j p \\ &+ (d - \min(t - s_j + 1, d)) \Bigg\{ (1 - F(p))\, s_j p \\ &\qquad - \left(F\left(p + \frac{c}{s_j - s_{j-1}}\right) - F(p)\right)(s_j - s_{j-1})p \Bigg\} \\ =& d\,(1 - F(p))\, s_j p - (d - \min(t - s_j + 1, d)) \left(F\left(p + \frac{c}{s_j - s_{j-1}}\right) - F(p)\right)(s_j - s_{j-1})p.\end{aligned} \tag{78}$$

Notice that the first term of Equation (78) depends only on the introduction times through s_j, while the second term only depends on the introduction times through s_j and the difference $s_j - s_{j-1}$. Therefore, we can represent the problem of choosing introduction times as a Markov Decision Process whose states are a pair consisting of the current time and the previous introduction time, while the actions are to either introduce or not in the current period. As this MDP has a countably infinite set of states and a finite set of actions, it has a (deterministic) optimal policy (Puterman, 2014, Thm 6.2.10). Since the state transitions are deterministic, such an optimal policy for the MDP induces an optimal policy for our problem.

Given the structure of expression (78), we show how to take any optimal policy π_p^* and construct a slightly different policy that is also optimal. As a first step, we define policy mapping $\mathcal{T}_k(\cdot)$, where k is a positive integer.

Definition 3. Fix a positive integer k. Given a policy $\pi_p = ((s_0 = 0, x_0 = 0), (s_1, x_1 = s_1 p), (s_2, x_2 = s_2 p), (s_3, x_3 = s_3 p), \dots)$ that uses linear pricing with base price p, and introduces at times $(s_1, s_2, s_3, \dots)$, we define $\mathcal{T}_k(\pi_p)$ to be the following policy:

$$\begin{aligned}\mathcal{T}_k(\pi_p) \coloneqq& ((s'_0 = 0, x'_0 = 0), (s'_1 = s_1 + k, x'_1 = s'_1 p), \\ &\qquad (s'_2 = s_2 + k, x'_2 = s'_2 p), (s'_3 = s_3 + k, x'_3 = s'_3 p), \dots).\end{aligned}$$

Since mapping $\mathcal{T}_k(\pi_p)$ delays each introduction by the same constant, it affects the timing and revenue from the first term of expression (78): introduction times have been changed, which changes the prices customers pay. But it only changes the timing from the second term of expression (78): the loss due to customers not switching to a newly introduced technology only depends on the differences in introduction times, which we have not changed, although we have shifted the periods in which these losses occur later.

We first write

$$\begin{aligned}
U(\pi_p) &= \sum_{t=s_1}^{\infty} \delta^t \left(\text{Revenue}(\pi_p, t) - \text{Cost}(\pi_p, t)\right) \\
&= \delta^{s_1} \left(1 - F(p)\right) s_1 p + \sum_{t=s_1+1}^{s_1+d-2} \delta^t (t - s_1 + 1) \left(1 - F(p)\right) s_1 p + \sum_{t=s_1+d-1}^{s_2-1} \delta^t d \left(1 - F(p)\right) s_1 p \quad (79) \\
&\quad + \delta^{s_2} \left\{ d \left(1 - F(p)\right) s_2 p - (d-1) \left(F\left(p + \frac{c}{s_2 - s_1}\right) - F(p) \right) (s_2 - s_1) p \right\} \quad (80) \\
&\quad\quad + \sum_{t=s_2+1}^{s_3-1} \delta^t d \left(1 - F(p)\right) s_2 p - \sum_{t=s_2+1}^{s_2+d-2} \delta^t \left(d - (t - s_2 + 1)\right) \left(F\left(p + \frac{c}{s_2 - s_1}\right) - F(p) \right) (s_2 - s_1) p \quad (81) \\
&\quad + \delta^{s_3} \left\{ d \left(1 - F(p)\right) s_3 p - (d-1) \left(F\left(p + \frac{c}{s_3 - s_2}\right) - F(p) \right) (s_3 - s_2) p \right\} \\
&\quad\quad + \sum_{t=s_3+1}^{s_4-1} \delta^t d \left(1 - F(p)\right) s_3 p - \sum_{t=s_3+1}^{s_3+d-2} \delta^t \left(d - (t - s_3 + 1)\right) \left(F\left(p + \frac{c}{s_3 - s_2}\right) - F(p) \right) (s_3 - s_2) p \\
&\quad + \ldots \\
&\quad - C \left(\delta^{s_1} + \delta^{s_2} + \delta^{s_3} + \ldots\right). \quad (82)
\end{aligned}$$

Line (79) corresponds to the revenue accumulated during periods $s_1, \ldots, s_2{-}1$. The term $\delta^{s_1} \left(1 - F(p)\right) s_1 p$ is the revenue from customers who arrive in period s_1 and buy technology class 1. The term $\sum_{t=s_1+1}^{s_1+d-2} \delta^t (t - s_1 + 1) \left(1 - F(p)\right) s_1 p$ is the revenue accumulated during periods $s_1{+}1, \ldots, s_1{+}d{-}2$ from both new and existing customers. Notice that the mass of existing customers builds up during this time, as a unit mass of new customers arrives in each period, while customers who buy the technology do not leave the system yet. The term $\sum_{t=s_1+d-1}^{s_2-1} \delta^t d \left(1 - F(p)\right) s_1 p$ is the revenue accumulated during periods $s_1 + d - 1, \ldots, s_2 - 1$ from both new and existing customers. Notice that during these periods, and for the continuation of the infinite horizon, the expected mass of customers who yield revenue remains stable.

Lines (80), (81) correspond to the revenue accumulated during periods $s_2, \ldots, s_3 - 1$. In particular, the term in line (80) corresponds to the revenue accumulated during the period of the second introduction, as per Equation (78) for $s_j = 2$; and the term in line (81) corresponds to the revenue accumulated during the non-introduction periods $s_2+1,\ldots,s_3-1$. Line (82) accounts for the provisioning costs from the introductions.

We then write the utility of policy $\mathcal{T}_k(\pi_p)$, for a fixed positive integer k:

$$\begin{aligned}
U(\mathcal{T}_k(\pi_p)) &= \sum_{t=s_1}^{\infty} \delta^t \left(\text{Revenue}(\mathcal{T}_k(\pi_p),t) - \text{Cost}(\mathcal{T}_k(\pi_p),t)\right) \\
&= \delta^{s_1+k}\left(1-F(p)\right)(s_1+k)p + \sum_{t=s_1+k+1}^{s_1+k+d-2} \delta^t (t-(s_1+k)+1)\left(1-F(p)\right)(s_1+k)p + \sum_{t=s_1+k+d-1}^{s_2+k-1} \delta^t d\left(1-F(p)\right)(s_1+k)p \\
&\quad +\delta^{s_2+k}\left\{d\left(1-F(p)\right)(s_2+k)\,p-(d-1)\left(F\left(p+\frac{c}{s_2-s_1}\right)-F(p)\right)(s_2-s_1)p\right\} \\
&\qquad + \sum_{t=s_2+k+1}^{s_3+k-1} \delta^t d\left(1-F(p)\right)(s_2+k)\,p \\
&\qquad - \sum_{t=s_2+k+1}^{s_2+k+d-2} \delta^t \left(d-(t-(s_2+k)+1)\right)\left(F\left(p+\frac{c}{s_2-s_1}\right)-F(p)\right)(s_2-s_1)p \\
&\quad +\delta^{s_3+k}\left\{d\left(1-F(p)\right)(s_3+k)\,p-(d-1)\left(F\left(p+\frac{c}{s_3-s_2}\right)-F(p)\right)(s_3-s_2)p\right\} \\
&\qquad + \sum_{t=s_3+k+1}^{s_4+k-1} \delta^t d\left(1-F(p)\right)(s_3+k)\,p \\
&\qquad - \sum_{t=s_3+k+1}^{s_3+k+d-2} \delta^t \left(d-(t-(s_3+k)+1)\right)\left(F\left(p+\frac{c}{s_3-s_2}\right)-F(p)\right)(s_3-s_2)p \\
&\quad +\ldots \\
&\quad -C\left(\delta^{s_1+k}+\delta^{s_2+k}+\delta^{s_3+k}+\ldots\right). \qquad (83)
\end{aligned}$$

Thus we have

$$
\begin{aligned}
U(\mathcal{T}_k(\pi_p)) =& \delta^k U(\pi_p) \\
& + \delta^{s_1+k}(1-F(p))kp + \sum_{t=s_1+k+1}^{s_1+k+d-2} \delta^t (t-(s_1+k)+1)(1-F(p))kp \\
& \quad + \sum_{t=s_1+k+d-1}^{s_2+k-1} \delta^t d(1-F(p))kp \\
& + \delta^{s_2+k} d\left(1-F(p)\right)kp + \sum_{s_2+k+1}^{s_3+k-1} \delta^t d\left(1-F(p)\right)kp \\
& + \delta^{s_3+k} d\left(1-F(p)\right)kp + \sum_{s_3+k+1}^{s_4+k-1} \delta^t d\left(1-F(p)\right)kp \\
& + \dots \\
=& \delta^k \left[U(\pi_p) + \delta^{s_1}\left(1+2\delta+3\delta^2+\dots+(d-1)\delta^{d-2}\right)\left(1-F(p)\right)kp + \sum_{s_1+d-1}^{\infty} \delta^t d\left(1-F(p)\right)kp \right] \\
=& \delta^k \left[U(\pi_p) + \left(\delta^{s_1}\left(1+2\delta+3\delta^2+\dots+(d-1)\delta^{d-2}\right) + d\frac{\delta^{s_1+d-1}}{1-\delta} \right)\left(1-F(p)\right)kp \right].
\end{aligned}
\tag{84}
$$

This is because shifting the introductions k periods later discounts all revenue and costs by δ^k, and the remaining terms capture the changes to the first term of (78), with the $t = s_1+k, s_1+k+1, \dots, s_1+k+d-2$ terms handled differently from the rest, due to the expected mass of revenue-generating customers still building up at those times.

Every policy using linear pricing with base price p, whose first introduction is s_2^*, can be written as $\mathcal{T}_{k=s_2^*-s_1^*}(\pi_p)$ for some policy π_p that uses linear pricing with base price p, and whose first introduction is s_1^*.[32] Thus by Equation (84) and the optimality of π_p^*, $\mathcal{T}_{s_2^*-s_1^*}(\pi_p^*)$ is optimal among all policies using linear pricing with base price p whose first introduction time is s_2^*.

We now define policy mapping $\mathcal{S}_k(\cdot)$, where k is a positive integer.

Definition 4. Fix a positive integer k. Given a policy $\pi_p = ((s_0 = 0, x_0 = 0), (s_1, x_1 = s_1 p), (s_2, x_2 = s_2 p), (s_3, x_3 = s_3 p), \dots)$ that uses linear pricing with base price p, and introduces at times $(s_1, s_2, s_3, \dots)$,

[32] We detail this construction, for completeness. We take all considered policies to use linear pricing with base price p. Start with policy π_p' that introduces at times $(s_1' = s_2^*, s_2', s_3', \dots)$. Then policy π_p that introduces at times $(s_1 = s_1^*, s_2 = s_2' - (s_2^* - s_1^*), s_3 = s_3' - (s_2^* - s_1^*), \dots)$ satisfies that $\mathcal{T}_{s_2^*-s_1^*}(\pi_p) = \pi_p'$.

we define $\mathcal{S}_k(\pi_p)$ to be the following policy:

$$\begin{aligned}\mathcal{S}_k(\pi_p) :=& ((s_0' = 0, x_0' = 0), (s_1' = s_1, x_1' = s_1'p), (s_2' = s_1 + k, x_2' = s_2'p),\\ & (s_3' = s_2 + k, x_3' = s_3'p), (s_4' = s_3 + k, x_4' = s_4'p), \dots).\end{aligned}$$

That is, $\mathcal{S}_k(\pi_p)$ uses the same introductions as $\mathcal{T}_k(\pi_p)$, but additionally introduces in period s_1.

Since $\pi_p^* = ((s_0 = 0, x_0 = 0), (s_1 = s_1^*, x_1 = s_1^*p), (s_2 = s_2^*, x_2 = s_2^*p), (s_3 = s_3^*, x_3 = s_3^*p), \dots)$ is an optimal policy among those policies using linear pricing with base price p, it follows that policy

$$\begin{aligned}\mathcal{S}_{k=s_2^*-s_1^*}(\pi_p^*) :=& ((0,0), (s_1^*, s_1^*p), (s_2^*, s_2^*p), (s_2^* + (s_2^* - s_1^*), (s_2^* + (s_2^* - s_1^*))\, p),\\ & (s_3^* + (s_2^* - s_1^*), (s_3^* + (s_2^* - s_1^*))\, p), \dots)\end{aligned}$$

is also optimal. We prove this by contradiction.

Assume $\mathcal{S}_{s_2^*-s_1^*}(\pi_p^*)$ is not optimal. Then $U(\pi_p^*) > U(\mathcal{S}_{s_2^*-s_1^*}(\pi_p^*))$. Because $U(\pi_p^*, t) = U\left(\mathcal{S}_{s_2^*-s_1^*}(\pi_p^*), t\right)$ for $t = s_1^*, \dots, s_2^*$, it then follows that policy π_p^* is superior when looking at the sum of discounted utilities starting at time s_2^*:

$$\sum_{t=s_2^*}^{\infty} \delta^t U(\pi_p^*, t) > \sum_{t=s_2^*}^{\infty} \delta^t U\left(\mathcal{S}_{s_2^*-s_1^*}(\pi_p^*), t\right). \tag{85}$$

We can subtract from both sides of the inequality the revenues accumulated at time s_2^* or after from customers who arrive in periods $s_1^*, s_1^*+1, \dots, s_2^*-1$. By Equation (75), these terms can be written as

$$\sum_{t=s_2^*}^{s_2^*+d-2} \delta^t \left(d - (t - s_2^* + 1)\right) \left\{ \left(1 - F\left(p + \frac{c}{s_2^* - s_1^*}\right)\right) s_2^*p + \left(F\left(p + \frac{c}{s_2^* - s_1^*}\right) - F(p)\right) s_1^*p \right\}. \tag{86}$$

These terms will be equal for both policies π_p^* and $\mathcal{S}_{s_2^*-s_1^*}(\pi_p^*)$, because the two policies are identical for the customers who arrive in periods $s_1^*, s_1^*+1, \dots, s_2^*-1$.

After subtracting expression (86) from both sides of inequality (85), we recognize the right-hand side as the utility of policy $\mathcal{T}_{s_2^*-s_1^*}(\pi_p^*)$, which first introduces at time s_2^*. So we have shown that the policy whose first introduction is at s_2^* and from then on follows policy π_p^*, has greater utility than policy $\mathcal{T}_{s_2^*-s_1^*}(\pi_p^*)$. This means that $\mathcal{T}_{s_2^*-s_1^*}(\pi_p^*)$ is not optimal among all policies using linear pricing with base price p and whose first introduction time is s_2^*. This is a contradiction.

Since $\mathcal{S}_{s_2^*-s_1^*}$ preserves optimality, any number of applications of it does so as well. Since each application shifts the schedule of introductions by $s_2^* - s_1^*$, a simple induction argument shows that $\mathcal{S}^{(j)}_{s_2^*-s_1^*}(\pi_p^*)$, i.e., applying the $\mathcal{S}_{s_2^*-s_1^*}$ operator j times, results in an optimal policy whose $j+1$ introductions after the first each come after $s_2^* - s_1^*$ periods from the previous introduction. That is, it is periodic with period length $s_2^* - s_1^*$ for the $j+1$ introductions after the first.

This allows us to define a sequence of optimal policies $\{\pi_p^{*(j)}\}$, such that the jth element of the sequence, $\pi_p^{*(j)} := \mathcal{S}^{(j)}_{s_2^*-s_1^*}(\pi_p^*)$, is periodic through the $j+1$th introduction after the first. This sequence converges to the infinitely periodic policy with first introduction s_1^* and period $s_2^*-s_1^*$, in the metric space defined in the proof of Theorem 2 in Appendix G. Similarly to the proof of Theorem 2, we can show that U is a continuous mapping with respect to the defined metric spaces. As U is continuous and each policy in the sequence is optimal, the periodic policy with first introduction s_1^* and period $s_2^* - s_1^*$ is also optimal. □

N Details of Optimal Pricing Results for a Single Period in the Non-Discriminatory Pricing Setting

Lemma 6. *Let introduction times* $\mathbf{s}$ *and time* t *be given with* $s_j \leq t < s_{j+1}$ *for some introduction* $j \geq 1$*, and* π_M *be a policy that uses introduction times* $\mathbf{s}$ *and Myerson pricing. Then*

$$\pi_M \in \arg\max_{\pi\in\Pi(\mathbf{s})} \int \left(\sum_{\ell=1}^{\min(t-s_j+1,d)} x_{q_\ell(\pi,t,\theta)} \right) f(\theta) d\theta.$$

Proof. All the terms in this sum correspond to customers who arrived at the time of or after the most recent introduction, and therefore none of them faced a switching cost when deciding which technology to adopt. In the spirit of Myerson's argument, we know that the optimal policy would be to offer all such customers only the latest technology at the Myerson price. But we know from Section L that all these customers will choose the latest technology regardless for any linear pricing scheme. Thus, the given policy is optimal. □

Lemma 7. *Let introduction times* $\mathbf{s}$ *and time* t *be given with* $s_j \le t < s_j + d - 1$ *for some introduction* $j \ge 1$*. There exists a policy* $\pi' \in \Pi(\mathbf{s})$ *that maximizes*

$$\int \left(\sum_{\ell = t - s_j + 2}^{d} x_{q_\ell(\pi, t, \theta)} \right) f(\theta) d\theta$$

among all policies $\pi \in \Pi(\mathbf{s})$*, and that uses pricing* $x_i = s_i p^*$ *for* $i < j$ *and* $\max(s_{j-1}p^*, s_j p^* - c) \le x_j \le s_j p^*$*.*

Proof. For $j = 1$, $q_\ell(\pi, t, \theta) = 0$ for $\ell \ge t - s_1 + 2$, regardless of the choice of π, so optimality holds vacuously.

Fix $j \ge 2$. We first focus on finding an optimal setting of prices for the expected revenue in period t that comes from the existing customer who is in her period ℓ at time t,

$$\int x_{q_\ell(\pi, t, \theta)} f(\theta) d\theta,$$

where $t - s_j + 2 \le \ell \le d$. (This customer arrived in the system at time $t - \ell + 1$.)

In the spirit of Myerson's argument, we know that all allocation rules achievable by pricing are incentive compatible and thus monotone, so we can optimize over them instead. In particular, with a finite menu of technology classes, the monotone allocation function is piecewise constant: customers who do not buy get an allocation of 0, those who do get some technology 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:

$$a(\theta) = \begin{cases} 0 & \text{if} \quad \theta < \theta_1 \\ i & \text{if} \quad \theta_i \le \theta < \theta_{i+1}, \quad 1 \le i \le j - 1 \\ j & \text{if} \quad \theta_j \le \theta. \end{cases} \tag{87}$$

Fix a and the resulting policy $\pi'(a)$, and let $I_{\theta_j}(\theta) = 1$ if $\theta \ge \theta_j$ 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 is in her period ℓ at time t is

$$x_{q_\ell(\pi'(a),t,\theta)} = s_{a(\theta)}\theta - \int_0^\theta s_{a(\theta')}d\theta' - I_{\theta_j}(\theta)c. \tag{88}$$

This makes the expected revenue $\int x_{q_\ell(\pi'(a),t,\theta)} f(\theta)d\theta$ equal to

$$\begin{aligned}
&\int_0^{\theta_1} x_{q_\ell(\pi'(a),t,\theta)} f(\theta)d\theta + \sum_{i=1}^{j-1}\int_{\theta_i}^{\theta_{i+1}} x_{q_\ell(\pi'(a),t,\theta)} f(\theta)d\theta + \int_{\theta_j}^{\infty} x_{q_\ell(\pi'(a),t,\theta)} f(\theta)d\theta \\
&= \sum_{i=1}^{j-1}\int_{\theta_i}^{\theta_{i+1}} \left(s_i\theta - s_i(\theta-\theta_i) - \sum_{i'=1}^{i-1} s_{i'}(\theta_{i'+1}-\theta_{i'}) \right) f(\theta)d\theta \\
&\quad + \int_{\theta_j}^{\infty} \left(s_j\theta - s_j(\theta-\theta_j) - \sum_{i'=1}^{j-1} s_{i'}(\theta_{i'+1}-\theta_{i'}) - c \right) f(\theta)d\theta \\
&= \sum_{i=1}^{j-1}\int_{\theta_i}^{\theta_{i+1}} s_i\theta_i f(\theta)d\theta - \sum_{i=1}^{j-1}\int_{\theta_{i+1}}^{\infty} \left(s_i(\theta_{i+1}-\theta_i)\right) f(\theta)d\theta + \int_{\theta_j}^{\infty} \left(s_j\theta_j - c\right) f(\theta)d\theta \\
&= \left(\sum_{i=1}^{j-1} \left(F(\theta_{i+1}) - F(\theta_i)\right) s_i\theta_i - \left(1 - F(\theta_{i+1})\right) s_i(\theta_{i+1}-\theta_i) \right) + \left(1 - F(\theta_j)\right)\left(s_j\theta_j - c\right) \\
&= \left(\sum_{i=1}^{j-1} \left(1 - F(\theta_i)\right) s_i\theta_i - \left(1 - F(\theta_{i+1})\right) s_i\theta_{i+1} \right) + \left(1 - F(\theta_j)\right)\left(s_j\theta_j - c\right) \\
&= \left(\sum_{i=1}^{j-1} \left(1 - F(\theta_i)\right) (s_i - s_{i-1})\theta_i \right) + \left(1 - F(\theta_j)\right)\left((s_j - s_{j-1})\theta_j - c\right).
\end{aligned} \tag{89}$$

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 $i < j$. This implies that it is optimal to set $x_i = s_i p^*$ 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

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

or

$$\left(s_j - s_{j-1}\right)\left(1 - F(\theta_j) - f(\theta_j)\theta_j\right) + f(\theta_j)c. \tag{90}$$

The first order condition can then be rewritten as

$$(s_j - s_{j-1})\left(\theta_j - \frac{1 - F(\theta_j)}{f(\theta_j)}\right) = c. \tag{91}$$

By the definition of p^*, the left hand side of Equation (91) is exactly 0 for $\theta_j = p^*$, and is increasing in θ_j by Assumption 2. 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 $s_j\theta_j - x_j - c = s_{j-1}\theta_j - x_{j-1}$, therefore we have

$$x_j = x_{j-1} + (s_j - s_{j-1})\theta_j - c. \tag{92}$$

We observe that for $x_{j-1} = s_{j-1}p^*$, $\theta \geq p^*$ implies $x_j \geq s_j p^* - c$.

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. Therefore, we have $x_j \geq s_{j-1}p^*$.

To obtain an upper bound on x_j, we rewrite the first order condition in (91) as

$$\begin{aligned} \theta_j &= \frac{1 - F(\theta_j)}{f(\theta_j)} + \frac{c}{s_j - s_{j-1}} \\ &\leq \frac{1 - F(p^*)}{f(p^*)} + \frac{c}{s_j - s_{j-1}} \\ &= p^* + \frac{c}{s_j - s_{j-1}}, \end{aligned}$$

where the inequality follows because $\theta_j \geq p^*$ and by Assumption 2. We observe that, by Equation (92), and for $x_{j-1} = s_{j-1}p^*$, $\theta \leq p^* + \frac{c}{s_j - s_{j-1}}$ implies $x_j \leq s_j p^*$.

Note that nothing in the above analysis is specific to the choice of ℓ, the tenure of the customer at time t, as long as $t - s_j + 2 \leq \ell \leq d$. Therefore, there exists a policy $\pi' \in \Pi(\mathbf{s})$ that maximizes $\int x_{q_\ell(\pi,t,\theta)} f(\theta) d\theta$ for each ℓ such that $t - s_j + 2 \leq \ell \leq d$, and that uses pricing $x_i = s_i p^*$ for $i < j$ and $\max(s_{j-1}p^*, s_j p^* - c) \leq x_j \leq s_j p^*$. The result follows. $\square$

Lemma 6 and Lemma 7 show that the policy that maximizes revenue from customers who arrived at or after the most recent introduction, and the policy that does for those who arrived before the most recent introduction, agree that Myerson pricing should be used for all but the most recent introduction, but disagree on what the price of the most recent introduction should be. As a step towards showing that the optimal prices for the combined revenue share this structure (intuitively with a compromise over what the price of the most recent introduction should be), we first give an explicit characterization of the revenue from such policies.

As a further step toward our goal of providing an upper bound to the optimal revenue, we consider an expansion of the set of policies to allow separate prices to be offered to customers, depending on whether they were already existing customers at the time of the most recent introduction. Such a discriminatory strategy would offer a discount to the customers who arrived before the most recent introduction, as an incentive to upgrade. Assuming $s_j \leq t < s_j + d - 1$ so that both types of customers exist[33], the (expected) revenue of such a discriminatory strategy employing policy π_n for customers arriving since the most recent introduction, and policy π_e for customers who arrived before it, at time t, is

$$\text{RevenueD}(\pi_n, \pi_e, t) = \int \left(x_{q_1(\pi_n,t,\theta)} + \ldots + x_{q_{t-s_j+1}(\pi_n,t,\theta)} + x_{q_{t-s_j+2}(\pi_e,t,\theta)} + \ldots + x_{q_d(\pi_e,t,\theta)} \right) f(\theta) d\theta.$$

The following lemma gives our characterization of the revenue of both types of policies.

Lemma 8. *Let introduction times* $\mathbf{s}$ *and time* t *be given with* $s_j \leq t < s_{j+1}$ *for some introduction* $j \geq 1$*. Consider policy* $\pi \in \Pi(\mathbf{s})$ *that uses prices* $x_i = s_i p^*$ *for* $i < j$ *and* $x_j = x$*, with* $\max(s_{j-1}p^*, s_j p^* - c) \leq x \leq s_j p^*$*. Then*

$$\mathit{Revenue}(\pi, t) = \begin{cases} \max(d - (t - s_j + 1), 0) \left\{ (1 - F(p^*)) \, s_{j-1} p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*) \right\} & \\ \qquad + (\min(t - s_j + 1, d)) \left(1 - F\left(\frac{x}{s_j}\right)\right) x, & j \geq 2 \\ (\min(t - s_1 + 1, d)) \left(1 - F\left(\frac{x}{s_1}\right)\right) x, & j = 1. \end{cases} \tag{93}$$

[33] A customer who arrives right before time s_j, i.e., at time $s_j - 1$, stays in the system until period $s_j + d - 2$. Therefore, during times t such that $s_j \leq t < s_j + d - 1$, both customers who have arrived since the most recent introduction, and customers who arrived before the most recent introduction exist in the system.

Consider also a discriminatory strategy that uses policies $\pi_n, \pi_e \in \Pi(\mathbf{s})$ *with prices* $x_i = s_i p^*$ *for* $i < j$, $x_j = x_n$ *for customers who arrive at or after introduction* j, *and* $x_j = x_e$ *for customers who arrive before introduction* j, *with* $\max(s_{j-1}p^*, s_j p^* - c) \leq x_n, x_e \leq s_j p^*$. *Then*

$$RevenueD(\pi_n, \pi_e, t) = \begin{cases} \max(d-(t-s_j+1),0)\left\{(1-F(p^*))\, s_{j-1}p^* + \left(1-F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x_e - s_{j-1}p^*)\right\} & \\ \quad +(\min(t-s_j+1,d))\left(1-F\left(\frac{x_n}{s_j}\right)\right)x_n, & j \geq 2 \\ (\min(t-s_1+1,d))\left(1-F\left(\frac{x_n}{s_1}\right)\right)x_n, & j = 1. \end{cases} \tag{94}$$

Proof. For $j \geq 2$, we show that

$$\begin{aligned} \text{Revenue}(\pi, t) &= \min(t-s_j+1,d)\left(1-F\left(\frac{x}{s_j}\right)\right)x \\ &\quad + \max(d-(t-s_j+1),0)\cdot\left[\left(1-F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)x\right. \\ &\quad \left. + \left(F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right) - F(p^*)\right)s_{j-1}p^*\right]. \end{aligned} \tag{95}$$

We first explain the summand in the first row of Equation (95). This summand corresponds to revenue from customers who arrive at or after period s_j. The term $\left(1-F\left(\frac{x}{s_j}\right)\right)x$ is the expected revenue accumulated at time t from a customer who arrives at or after period s_j and buys the new technology class as long as $\theta s_j - x \geq 0 \iff \theta \geq \frac{x}{s_j}$. Notice that this customer would buy technology class $k < j$ instead of technology class j if $s_k(\theta - p^*) \geq 0 \iff \theta \geq p^*$ and $\theta s_j - x < s_k(\theta - p^*) \iff \theta < \frac{x - s_k p^*}{s_j - s_k}$. Since $\frac{x - s_k p^*}{s_j - s_k} \leq \frac{s_j p^* - s_k p^*}{s_j - s_k} = p^*$, the two cannot happen at the same time. We record this revenue term for all customers who arrive at or after s_j, and within the last $d-1$ periods before, or at, period t. There are $\min(t-s_j+1,d)$ such terms.

We next explain the summand in the second and third rows of Equation (95). This summand corresponds to revenue from customers who arrive before period s_j. The term

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

is the expected revenue accumulated at time t from a customer who arrives before period s_j and switches to the new technology class introduced at time s_j, because $\theta s_j - x - c \geq s_{j-1}(\theta - p^*) \iff \theta \geq \frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}$. Notice that $\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}} \geq \frac{s_j p^* - c - s_{j-1}p^* + c}{s_j - s_{j-1}} = p^*$, therefore as long as $\theta \geq \frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}$, this customer buys technology class $j-1$ when she arrives and doesn't opt out, because $\theta \geq p^* \iff s_{j-1}(\theta - p^*) \geq 0$.

The term

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

is the expected revenue accumulated at time t from a customer who arrives before period s_j, and does not switch to the new technology class j at time s_j, because $\theta s_j - x - c < s_{j-1}(\theta - p^*) \iff \theta < \frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}$, while she buys technology class $j-1$ when she arrives, because $s_{j-1}(\theta - p^*) \geq 0 \iff \theta \geq p^*$.

We record these revenue terms for all customers who arrive in each of the periods before s_j, and within the last $d-1$ periods before period t. There are $d - \min(t - s_j + 1, d) = \max(d - (t - s_j + 1), 0)$ such periods.

By rewriting

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

the result in Equation (93) for $j \geq 2$ follows. The result for the $j \geq 1$ case follows by only taking into account the summand in the first row of Equation (95).

The results for $\mathrm{RevenueD}(\pi_n, \pi_e, t)$ follow analogously. A similar argument shows that, for $j \geq 2$,

$$\begin{aligned}
\mathrm{RevenueD}(\pi_n, \pi_e, t) \quad = \quad & \min(t - s_j + 1, d)\left(1 - F\left(\frac{x_n}{s_j}\right)\right) x_n \\
& + \max(d - (t - s_j + 1), 0) \cdot \left[\left(1 - F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right) x_e \right. \\
& \left. + \left(F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right) - F(p^*)\right) s_{j-1}p^*\right], \qquad (96)
\end{aligned}$$

from which the result in Equation (94) for $j \geq 2$ follows. The result for the $j \geq 1$ case follows by only taking into account the summand in the first row of Equation (96). $\square$

Theorem 4. *Let introduction times* $\mathbf{s}$ *and time* t *be given with* $s_j \leq t < s_{j+1}$ *for some introduction* $j \geq 1$*. 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* (93)*.*

Proof. By definition, we have

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

For $j = 1$, all the non-zero terms correspond to customers who arrived since the most recent introduction. Therefore, by Lemma 6, the given form is optimal. Similarly, if $j \geq 2$ and $t \geq s_j + d - 1$, then again all the non-zero terms correspond to customers who arrived since the most recent introduction, and the given form is optimal.

Fix $j \geq 2$ and take t such that $s_j \leq t < s_j + d - 1$. The essence of our proof is that our assumption makes both the sum of terms representing revenue from customers who arrived since the most recent introduction (i.e., terms covered by Lemma 6) and the sum of terms representing revenue from customers who arrived before the most recent introduction (i.e., terms covered by Lemma 7) 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 begin with the terms from Lemma 7. Fix ℓ with $t - s_j + 2 \leq \ell \leq d$ and consider the term $\int x_{q_\ell(\pi,t,\theta)} f(\theta) d\theta$. Let policy π' have pricing $x_i = s_i p^*$ for $i < j$ and $x_j = x$. In the proof of Lemma 7, we argued that the threshold θ_j at which customers switch to technology j solves

$$s_j \theta_j - x_j - c = s_{j-1} \theta_j - x_{j-1}, \tag{97}$$

from which it follows that

$$\theta_j = \frac{x_j - x_{j-1} + c}{s_j - s_{j-1}}. \tag{98}$$

We rewrite Equation (89) in the proof of Lemma 7, using (97) and (98), as follows:

$$\int x_{q_\ell(\pi',t,\theta)} f(\theta) d\theta = (1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*), \tag{99}$$

and therefore

$$\begin{aligned} &\int \left(\sum_{i=t-s_j+2}^{d} x_{q_i(\pi,t,\theta)}\right) f(\theta) d\theta \\ &\quad = (d - (t - s_j + 1)) \cdot \left\{(1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*)\right\}. \end{aligned} \tag{100}$$

By Lemma 7, the maximizing x is at least $\max(s_{j-1}p^*, s_j p^* - c)$ and at most $s_j p^*$. Furthermore, Assumption 2 is equivalent to the log-concavity of $1 - F(x)$ as a function of x, because in general log-concavity of a function g is equivalent to g'/g being monotonically decreasing (Bagnoli and Bergstrom, 1989, Remark 1). In turn, term $\left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)$ is log-concave as a function of x, and so is term $\left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*)$ as product of log-concave functions. Because log-concavity implies quasiconcavity, the latter term is quasiconcave on $[s_{j-1}p^*, \infty)$, and so is the entire right-hand side of (100), because a non-decreasing function of a quasiconcave function remains quasiconcave.

We next turn to the terms covered by Lemma 6, i.e.,

$$\int \left(\sum_{i=1}^{t-s_j+1} x_{q_i(\pi,t,\theta)}\right) f(\theta) d\theta.$$

Note that each term $\int x_{q_\ell(\pi,t,\theta)} f(\theta) d\theta$, with $1 \le \ell \le t - s_j + 1$, is identical to term $\int x_{q_\ell(\pi,t,\theta)} f(\theta) d\theta$, with $t - s_j + 2 \le \ell \le d$, if $c = 0$. Our analysis above only assumed that $c \ge 0$. Thus the same analysis, *mutatis mutandis*, shows that the desired properties hold for the sum of those terms as well.

Therefore both terms

$$\int \left(\sum_{i=1}^{t-s_j+1} x_{q_i(\pi,t,\theta)}\right) f(\theta) d\theta$$

and

$$\int \left(\sum_{i=t-s_j+2}^{d} x_{q_i(\pi,t,\theta)} \right) f(\theta) d\theta$$

are quasiconcave and, by Lemma 6 and Lemma 7, have their maxima on the interval

$$\left[\max(s_{j-1}p^*, s_j p^* - c), s_j p^*\right],$$

meaning their combined maximum is on this interval as well. For the final part of the theorem, note that this means we are optimizing over policies of the form contemplated by Lemma 8. $\square$

O Details of Performance Guarantees of Myerson Pricing in the Non-Discriminatory Pricing Setting

Before stating our approximation ratio, we first give a lemma which shows that the single period where Myerson pricing performs worst relative to optimal pricing is one of the introduction periods.

Lemma 9. *Let introduction times* $\mathbf{s}$ *and time* t *be given with* $s_j \leq t < s_{j+1}$ *for some introduction* $j \geq 1$. *Let* π_t^* *maximize* (93) *for period* t, π_j^* *do so for period* s_j, *and* π_M *use Myerson pricing. Then*

$$\frac{\textit{Revenue}(\pi_t^*, t)}{\textit{Revenue}(\pi_M, t)} \leq \frac{\textit{Revenue}(\pi_j^*, s_j)}{\textit{Revenue}(\pi_M, s_j)}.$$

Similarly let $\pi_{t,n}^*$ *and* $\pi_{t,e}^*$ *maximize* (94) *for period* t, *and* $\pi_{j,n}^*$ *and* $\pi_{j,e}^*$ *do so for period* s_j. *Then*

$$\frac{\textit{RevenueD}(\pi_{t,n}^*, \pi_{t,e}^*, t)}{\textit{Revenue}(\pi_M, t)} \leq \frac{\textit{RevenueD}(\pi_{j,n}^*, \pi_{j,e}^*, s_j)}{\textit{Revenue}(\pi_M, s_j)}.$$

Proof. If $j = 1$ or $t \geq s_j + d - 1$, then Myerson pricing is trivially optimal and both ratios are 1.

For $j \geq 2$ and $t < s_j + d - 1$, let x_t^* be the price for the jth introduction used by policy π_t^*, and x_j^* be the price for the jth introduction used by policy π_j^*. For brevity, let $g(x) = (1 - F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x - s_{j-1}p^*)$ and $h(x) = \left(1 - F\left(\frac{x}{s_j}\right)\right) x$. Then

$$\begin{aligned}
\frac{\text{Revenue}(\pi_t^*, t)}{\text{Revenue}(\pi_M, t)} &= \frac{(d-(t-s_j+1))g(x_t^*)+(t-s_j+1)h(x_t^*)}{(d-(t-s_j+1))g(s_jp^*)+(t-s_j+1)h(s_jp^*)} \\
&= \frac{\frac{d-(t-s_j+1)}{d-1}\left((d-1)g(x_t^*)+h(x_t^*)\right)+\left(t-s_j+1-\frac{d-(t-s_j+1)}{d-1}\right)h(x_t^*)}{\frac{d-(t-s_j+1)}{d-1}\left((d-1)g(s_jp^*)+h(s_jp^*)\right)+\left(t-s_j+1-\frac{d-(t-s_j+1)}{d-1}\right)h(s_jp^*)} \\
&\leq \frac{(d-1)g(x_t^*)+h(x_t^*)}{(d-1)g(s_jp^*)+h(s_jp^*)} \\
&= \frac{\text{Revenue}(\pi_t^*, s_j)}{\text{Revenue}(\pi_M, s_j)} \\
&\leq \frac{\text{Revenue}(\pi_j^*, s_j)}{\text{Revenue}(\pi_M, s_j)}
\end{aligned}$$

The first inequality follows because the term $\left(t-s_j+1-\frac{d-(t-s_j+1)}{d-1}\right)$ is non-negative for $s_j \leq t < s_j+d-1$, and s_jp^* optimizes h, which allows us to rewrite the ratio without the second summand of the numerator and the denominator. The third equality follows from Lemma 8. The second inequality follows because π_j^* is defined to optimize revenue at s_j.

The same argument applies, *mutatis mutandis*, to discriminatory pricing. $\square$

Lemma 9 allows us to focus on a special case of Equations (93) and (94), where $t = s_j$. As our subsequent results all focus on this special case, we give it its own notation.

Definition 5. We define

$$Rev_j(x) \coloneqq \begin{cases} (d-1)\left\{(1-F(p^*))\,s_{j-1}p^* + \left(1-F\left(\frac{x-s_{j-1}p^*+c}{s_j-s_{j-1}}\right)\right)(x-s_{j-1}p^*)\right\} & \\ \qquad + \left(1-F\left(\frac{x}{s_j}\right)\right)x, & j \geq 2 \\ \left(1-F\left(\frac{x}{s_1}\right)\right)x, & j = 1. \end{cases} \tag{101}$$

and

$$
RevD_j(x_n, x_e) \coloneqq \begin{cases} (d-1)\left\{(1-F(p^*))\, s_{j-1}p^* + \left(1 - F\left(\frac{x_e - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x_e - s_{j-1}p^*)\right\} & \\ \qquad + \left(1 - F\left(\frac{x_n}{s_j}\right)\right) x_n, & j \geq 2 \\ \left(1 - F\left(\frac{x_n}{s_1}\right)\right) x_n, & j = 1. \end{cases} \tag{102}
$$

We remark that the revenue under Myerson pricing at introduction period s_j is given by Equation (77) (for $p = p^*$) and Equation (101) (for $x = s_j p^*$), which are equivalent: $\text{Revenue}(\pi_M, s_j) = Rev_j(s_j p^*)$, where we let π_M be the policy that uses Myerson pricing and introduction times $\mathbf{s}$.

We can now state a corollary making the approximation ratio precise.

Corollary 3. *Let introduction times* $\mathbf{s}$ *be given. Let* $\pi_M \in \Pi(\mathbf{s})$ *be a policy that uses Myerson pricing and uses introduction times* $\mathbf{s}$*. Then, for every policy* $\pi \in \Pi(\mathbf{s})$*, we have that* π_M *is a* $\left(\max_j \frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)}, 1\right)$ *approximation to the revenue of and cost of* π *respectively, where* x_j^* *is the value* x *that maximizes* $Rev_j(x)$*. Similarly, for every pair of policies* $\pi_n, \pi_e \in \Pi(\mathbf{s})$*, we have that* π_M *is a* $\left(\max_j \frac{RevD_j(x_{j,n}^*, x_{j,e}^*)}{Rev_j(s_j p^*)}, 1\right)$ *approximation to the revenue of and cost of discriminatory policy* (π_n, π_e)*, where* $x_{j,n}^*$ *and* $x_{j,e}^*$ *are the values of* x_n *and* x_e *respectively that maximize* $RevD_j(x_n, x_e)$*. Furthermore, denoting by* π_M^* *the policy that is optimal among those that use Myerson pricing and periodic introductions (after the first introduction), we have* $U(\pi_M) \leq U(\pi_M^*)$*.*

Proof. Since $\pi, \pi_M \in \Pi(\mathbf{s})$, they have the same cost. For revenue, we can write

$$
\begin{aligned}
\frac{\text{Revenue}(\pi)}{\text{Revenue}(\pi_M)} &\leq \frac{\sum_t \delta^t \max_{\pi \in \Pi(\mathbf{s})} \text{Revenue}(\pi, t)}{\sum_t \delta^t \text{Revenue}(\pi_M, t)} \\
&\leq \max_t \frac{\max_{\pi \in \Pi(\mathbf{s})} \text{Revenue}(\pi, t)}{\text{Revenue}(\pi_M, t)} \\
&\leq \max_j \frac{\max_{\pi \in \Pi(\mathbf{s})} \text{Revenue}(\pi, s_j)}{\text{Revenue}(\pi_M, s_j)} \\
&= \max_j \frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)}
\end{aligned}
$$

The first line follows from Observation 3, the third by Lemma 9, and the fourth by Theorem 4 and by the observation that $\text{Revenue}(\pi_M, s_j) = Rev_j(s_j p^*)$. The proof for the discriminatory case is the same, *mutatis mutandis*. The last part of the statement follows directly from Proposition 1. $\square$

P Details of Bounding the Competitive Ratio of Myerson Pricing in the Non-Discriminatory Pricing Setting

Corollary 3 implies that we can bound our approximation ratio by bounding the competitive ratio of *Rev* or *RevD*. In the remainder of the section we provide two such bounds in terms of the distribution F. Our first bound is directly in terms of F but worsens with increasing customer lifetime d, while our second bound requires the derivative of F and an additional optimization to make the bound concrete, but improves with increasing d.

Proposition 3. *Let introduction number $j \geq 2$ be given. Let x_j^* be the value that maximizes $Rev_j(x)$ and $x_{j,n}^* = s_j p^*, x_{j,e}^*$ be the values that maximize $RevD_j(x_n, x_e)$. It holds that*

$$\frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)} \leq \frac{RevD_j(x_{j,n}^*, x_{j,e}^*)}{Rev_j(s_j p^*)} \leq 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}. \tag{103}$$

For the first introduction, we can write

$$\frac{Rev_1(x_1^*)}{Rev_1(s_j p^*)} = \frac{RevD_1(x_{1,n}^*, x_{1,e}^*)}{Rev_1(s_j p^*)} = 1.$$

Proof. The result for the first introduction is trivial.

For $j \geq 2$, we have

$$
\begin{aligned}
\frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)} &\leq \frac{RevD_j(s_j p^*, x_{j,e}^*)}{Rev_j(s_j p^*)} \\
&= 1 + \frac{RevD_j(s_j p^*, x_{j,e}^*) - Rev_j(s_j p^*)}{Rev_j(s_j p^*)} \\
&= 1 + \frac{(d-1)\left[\left(1 - F\left(\frac{x_{j,e}^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(x_{j,e}^* - s_{j-1}p^*) - \left(1 - F\left(p^* + \frac{c}{s_j - s_{j-1}}\right)\right)(s_j - s_{j-1})p^*\right]}{Rev_j(s_j p^*)} \\
&\leq 1 + \frac{(d-1)\left(F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{x_{j,e}^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(s_j - s_{j-1})p^*}{Rev_j(s_j p^*)} \\
&\leq 1 + \frac{(d-1)\left(F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{x_{j,e}^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)\right)(s_j - s_{j-1})p^*}{(d-1)\left(1 - F(p^*)\right)s_{j-1}p^* + \left(1 - F(p^*)\right)s_j p^*} \\
&= 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{x_{j,e}^* - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)}{1 - F(p^*)} \\
&\leq 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{\max(s_{j-1}p^*, s_j p^* - c) - s_{j-1}p^* + c}{s_j - s_{j-1}}\right)}{1 - F(p^*)} \\
&\leq 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}.
\end{aligned}
$$

The first inequality follows because $s_j p^*$ is the optimal price for new customers and we define x_e^* to be optimal for existing ones, the second equality follows from the definitions in (101) and (102), the second inequality because $x_{j,e}^* \leq s_j p^*$ by Lemma 7, the third inequality because, by (101), $Rev_j(s_j p^*) \geq (d-1)\left(1 - F(p^*)\right)s_{j-1}p^* + \left(1 - F(p^*)\right)s_j p^*$, and the fourth inequality follows from the lower bound on $x_{j,e}^*$ by Lemma 7. □

Several remarks are in order. First, the bound of Proposition 3 deteriorates with increasing customer lifetime d.[34] Second, the right-hand side of Equation (103) can be simplified by observing that

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

[34]Formally, if we define $g(d) = \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j}$, we have $g'(d) = \frac{s_j(s_j - s_{j-1})}{\left((d-1)s_{j-1} + s_j\right)^2} > 0$.

Third, the right-hand side of Equation (103) is close to 1 for large or small c. This is aligned with intuition. For switching cost c close to zero, an existing customer is likely 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.

Fourth, for a loose upper bound, the right-hand side of Equation (103) can be bounded by

$$\begin{aligned}
&\leq 1 + \frac{(d-1)(s_j - s_{j-1})}{s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)} \\
&\leq 1 + (d-1) \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}.
\end{aligned}$$

Fifth, a tighter bound to the right-hand side of Equation (103) can be obtained if an upper bound in the time difference between consecutive introductions can be assumed. For example, assuming periodic introductions every $d-1$ periods, with $s_0 = 0$ and $s_j = j \cdot (d-1)$, we can write

$$\frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} = \frac{(d-1)^2}{(d-1)(j-1)(d-1) + j(d-1)} = \frac{d-1}{d(j-1)+1},$$

which shows that Myerson pricing gets arbitrarily close to the optimal policy after sufficient introductions. Note that this is consistent with Proposition 2 and the discussion in Section L.2.

As a corollary to Proposition 3, we propose the following result, which provides an upper bound for the competitive ratio of Myerson pricing that does not deteriorate with the customer lifetime d.

Corollary 4. *Let introduction number $j \geq 2$ be given. Let x_j^* be the value that maximizes $Rev_j(x)$ and $x_{j,n}^* = s_j p^*, x_{j,e}^*$ be the values that maximize $RevD_j(x_n, x_e)$. It holds that*

$$\frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)} \leq \frac{RevD_j(x_{j,n}^*, x_{j,e}^*)}{Rev_j(s_j p^*)} < 1 + \frac{c}{d(j-2)+1} \cdot \frac{f(\hat{p})}{p^* f(p^*)}, \tag{104}$$

for some $\hat{p} \in \left(\max\left(p^*, \frac{c}{s_j - s_{j-1}}\right), p^* + \frac{c}{s_j - s_{j-1}}\right)$.

Proof. If $p^* \geq \frac{c}{s_j - s_{j-1}}$, we have

$$\begin{aligned}
\frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)} \leq \frac{RevD_j(x_{j,n}^*, x_{j,e}^*)}{Rev_j(s_j p^*)} &\leq 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F(p^*)}{1 - F(p^*)} \\
&= 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot f(\hat{p}) \cdot \frac{c}{s_j - s_{j-1}} \cdot \frac{1}{1 - F(p^*)} \\
&= 1 + \frac{c(d-1)}{(d-1)s_{j-1} + s_j} \cdot \frac{f(\hat{p})}{p^* f(p^*)} \\
&< 1 + \frac{c}{d(j-2) + 1} \cdot \frac{f(\hat{p})}{p^* f(p^*)},
\end{aligned}$$

where the second line follows by the mean value theorem on function F with $\hat{p} \in \left(p^*, p^* + \frac{c}{s_j - s_{j-1}}\right)$, the third line follows by the definition of p^*, and the fourth line is using that introductions are at least $d-1$ periods apart (after the first introduction), thus $s_{j-1} > (j-2)(d-1)$ and $s_j > (j-1)(d-1)$.

If $p^* < \frac{c}{s_j - s_{j-1}}$, then we have

$$\begin{aligned}
\frac{Rev_j(x_j^*)}{Rev_j(s_j p^*)} \leq \frac{RevD_j(x_{j,n}^*, x_{j,e}^*)}{Rev_j(s_j p^*)} &\leq 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - F\left(\frac{c}{s_j - s_{j-1}}\right)}{1 - F(p^*)} \\
&= 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{f(\hat{p}) \cdot p^*}{1 - F(p^*)} \\
&< 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot f(\hat{p}) \cdot \frac{c}{s_j - s_{j-1}} \cdot \frac{1}{1 - F(p^*)} \\
&\leq 1 + \frac{c}{d(j-2) + 1} \cdot \frac{f(\hat{p})}{p^* f(p^*)},
\end{aligned}$$

where the second line follows by the mean value theorem on function F with $\hat{p} \in \left(\frac{c}{s_j - s_{j-1}}, p^* + \frac{c}{s_j - s_{j-1}}\right)$, and in the third line we have used $p^* < \frac{c}{s_j - s_{j-1}}$. □

Putting together Corollary 3 and Proposition 3, we can state the main performance guarantee attained by our analysis for Myerson pricing in the following theorem.[35]

Theorem 5. *Let introduction times* $\mathbf{s}$ *be given. Let* $\pi_M \in \Pi(\mathbf{s})$ *be a policy that uses Myerson pricing and uses introduction times* $\mathbf{s}$*. Then, for every policy* $\pi \in \Pi(\mathbf{s})$*, we have that* π_M *is a*

$$\left(\max_{j \geq 2} 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}, 1\right)$$

[35] We can similarly state a performance guarantee by putting together Corollary 3 and Corollary 4.

approximation to the revenue of and cost of π respectively. Similarly, for every pair of policies $\pi_n, \pi_e \in \Pi(\mathbf{s})$, we have that π_M is a

$$\left(\max_{j\geq 2} 1 + \frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)}, 1\right)$$

approximation to the revenue of and cost of discriminatory policy (π_n, π_e). Furthermore, denoting by π_M^* the policy that is optimal among those that use Myerson pricing and periodic introductions (after the first introduction), we have $U(\pi_M) \leq U(\pi_M^*)$.

Q Numerical Illustrations for Myerson Pricing in the Non-Discriminatory Pricing Setting

We now illustrate our results with a variety of distributions for customer type θ. We show that our bounds for Myerson pricing from Section 6 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, after sufficiently many introductions, 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 Q.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[36], which was proposed in Theorem 4, over the Myerson

[36]Remember from Section 6 and in particular Observation 3 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.

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 customers depending on whether they were already existing customers at the time of the most recent introduction. Again, we show that Myerson pricing is close to optimal.

Q.1 Examining our analytical bounds

Although the bound of Proposition 3 deteriorates with the length d of customer lifetime, we show that it is a meaningful upper bound after sufficiently many introductions, even for quite a long customer lifetime. Figure 6 shows the upper bound on the gain ratio $\frac{Rev_j(x_j^*) - Rev_j(s_j p^*)}{Rev_j(s_j p^*)}$ from Proposition 3, which is

$$\frac{(d-1)(s_j - s_{j-1})}{(d-1)s_{j-1} + s_j} \cdot \frac{F\left(p^* + \frac{c}{s_j - s_{j-1}}\right) - \max\left(F(p^*), F\left(\frac{c}{s_j - s_{j-1}}\right)\right)}{1 - F(p^*)},$$

against the switching cost c, for customer lifetime $d = 14$. We show the upper 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 switching costs, while it raises the possibility that there is room for substantial improvement over the Myerson pricing for intermediate switching costs.[37]

We also note that, rather than applying the general bound from Proposition 3, we can directly calculate the left-hand side of Equation (103) in Proposition 3, and subtract 1 to recover the gain ratio $\frac{Rev_j(x_j^*) - Rev_j(s_j p^*)}{Rev_j(s_j p^*)}$. Figure 7 plots the gain ratio for customer lifetime $d = 14$, 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 explained before, the optimal revenue for a single introduction period is an upper bound on the real revenue in

[37] We note that it is also possible to derive bounds for specific distributions from first principles. For example, we can derive a bound of $(d-1)^2/(2d)$ for the case of the uniform distribution on $[0,1]$.

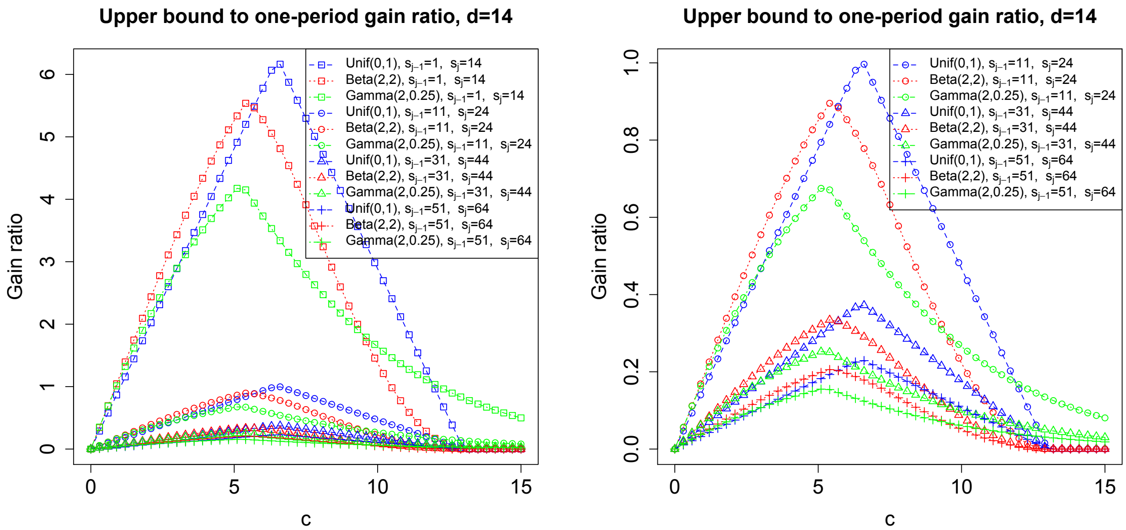

Figure 6: The right-hand side of Equation (103) in Proposition 3, reduced by 1, against the switching cost c, for customer lifetime $d = 14$, for selected pairs of introduction times s_{j-1}, s_j, for the uniform distribution on $[0, 1]$, the beta distribution with shape parameters $\alpha = \beta = 2$, and the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The right subfigure zooms in, plotting the upper bound for the same introduction times as the left subfigure except for the early introduction times.

that period under the optimal pricing policy, and therefore the gain ratio $\frac{Rev_j(x_j^*) - Rev_j(s_j p^*)}{Rev_j(s_j p^*)}$ is an upper bound to the gain ratio in that period under the optimal pricing policy for the real problem. The gains over Myerson pricing are less than 25% after sufficiently many introductions, even for intermediate switching costs — a bound substantially tighter than the one given by the right-hand side of Equation (103) in Proposition 3. We note that this bound is calculated using quite a long[38] customer lifetime $d = 14$, and that the gain ratio bound is smaller for shorter customer lifetimes.

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 $Rev_j'(s_j p^*) = 0$.

Q.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, customer lifetime d, switching cost c, and discount rate δ. In each simulation, a set of 50 introduction times is randomly generated,

[38]Remember that one period in our model corresponds in practice to the time interval after which the provider would revisit the decision of launching a new technology class or not.

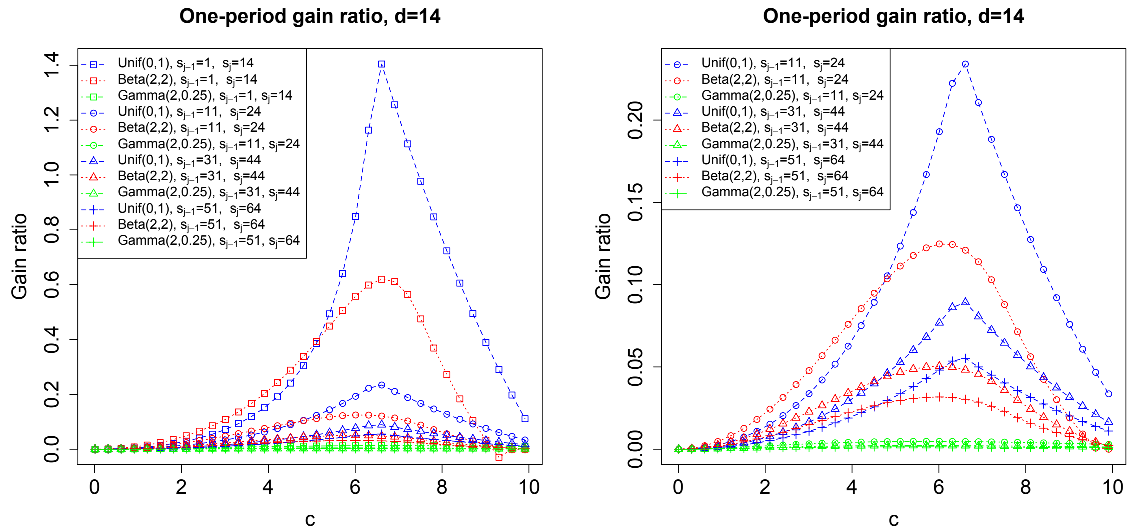

Figure 7: The left-hand side of Equation (103) in Proposition 3, reduced by 1, against the switching cost c, for customer lifetime $d = 14$, for selected pairs of introduction times s_{j-1}, s_j, for the uniform distribution on $[0, 1]$, the beta distribution with shape parameters $\alpha = \beta = 2$, and the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The right subfigure zooms in, plotting the gain ratio for the same introduction times as the left subfigure except for the early introduction times.

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 been optimized). Stop when no introduction can have an improvement ratio above 10^{-30}. Because of the arguments in Rosen (1965), 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 8 plots the average of the gain ratio of the optimal total revenue over Myerson total revenue,

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

over 100 simulations, against the switching cost c, for different values of the customer lifetime d and the discount rate δ, for the uniform distribution on $[0, 1]$.

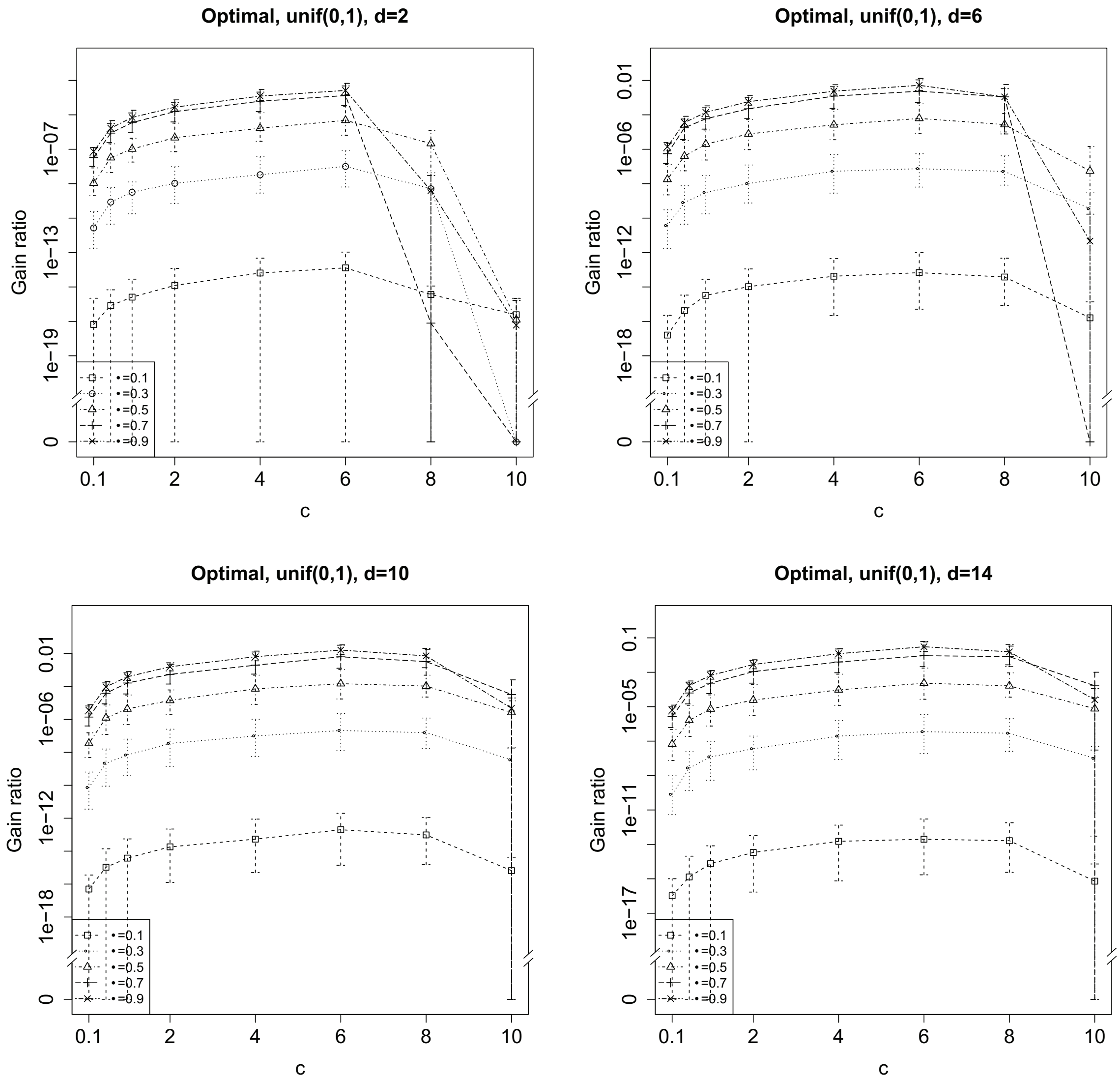

Figure 8: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c, for customer lifetime $d = 2, 6, 10, 14$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the uniform distribution on $[0, 1]$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

Holding the switching cost c constant, the higher the δ, the larger 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 less patient. 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 to behave as if she were a new customer, and chooses her preferred quality among the available ones,

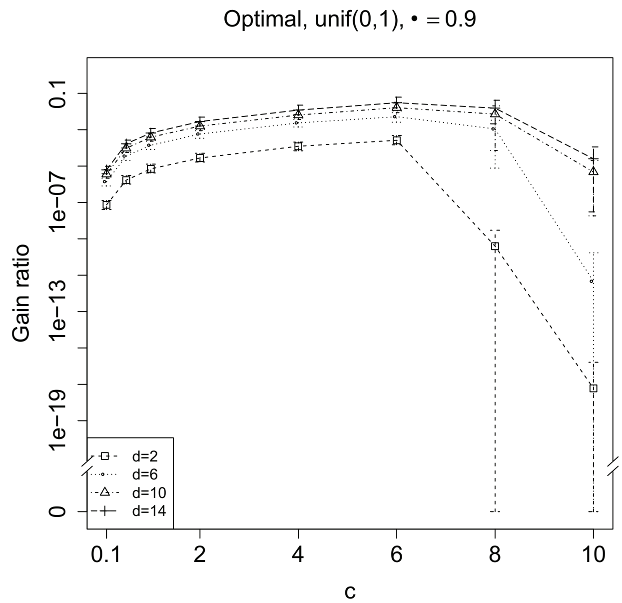

Figure 9: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c for customer lifetime $d = 2, 6, 10, 14$, for discount rate $\delta = 0.9$, for the uniform distribution on $[0, 1]$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

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.

Figure 9 shows that the gain ratio of the optimal total revenue over the Myerson total revenue increases with higher customer lifetime d, yet stays small even for quite long customer lifetimes, and even for intermediate switching costs. We note that the gain ratio is in the order of 0.01 for $d = 14$ in the worst case for the switching cost.

We also look at the full histogram of the total revenue gain ratio over 100 simulations for different values of d, 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 maximum gain ratio, $\max_j \frac{Rev_j(x_j^*) - \text{Revenue}(\pi_M, s_j)}{\text{Revenue}(\pi_M, s_j)}$, over 100 simulations. In Figure 10 we show the histograms for a reasonably high value for the customer lifetime d, and we note that the gain ratio values are smaller for shorter customer lifetimes. Figure 10 shows the effect of varying the

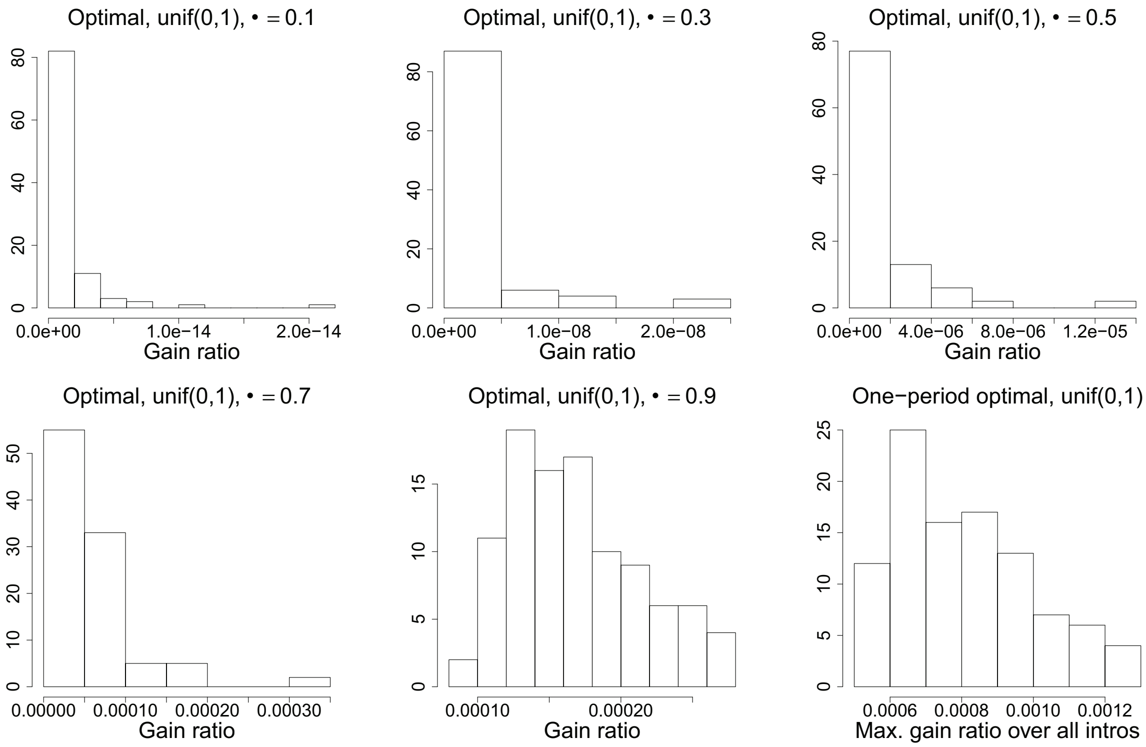

Figure 10: Histograms of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, for the uniform distribution on $[0, 1]$, for customer lifetime $d = 14$, switching cost $c = 0.5$, and discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$. Also, a histogram of the gain ratio of the optimal revenue for a single introduction period over Myerson revenue, for the introduction that attains the maximum gain ratio, over 100 simulations.

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 our worst case analytical bounds for the uniform distribution on $[0, 1]$ are loose even relative to this relaxation. Figure 11 shows again that in most instances Myerson pricing is essentially optimal, if instead we vary the switching cost c.

Figures 12 and 13 show experiments for the beta distribution, while Figures 14 and 15 show experiments for the gamma distribution. The results for the beta and gamma distributions are consistent with our results for the uniform distribution.

We present experiments for the discriminatory setting in Appendix R. Both in the non-discriminatory

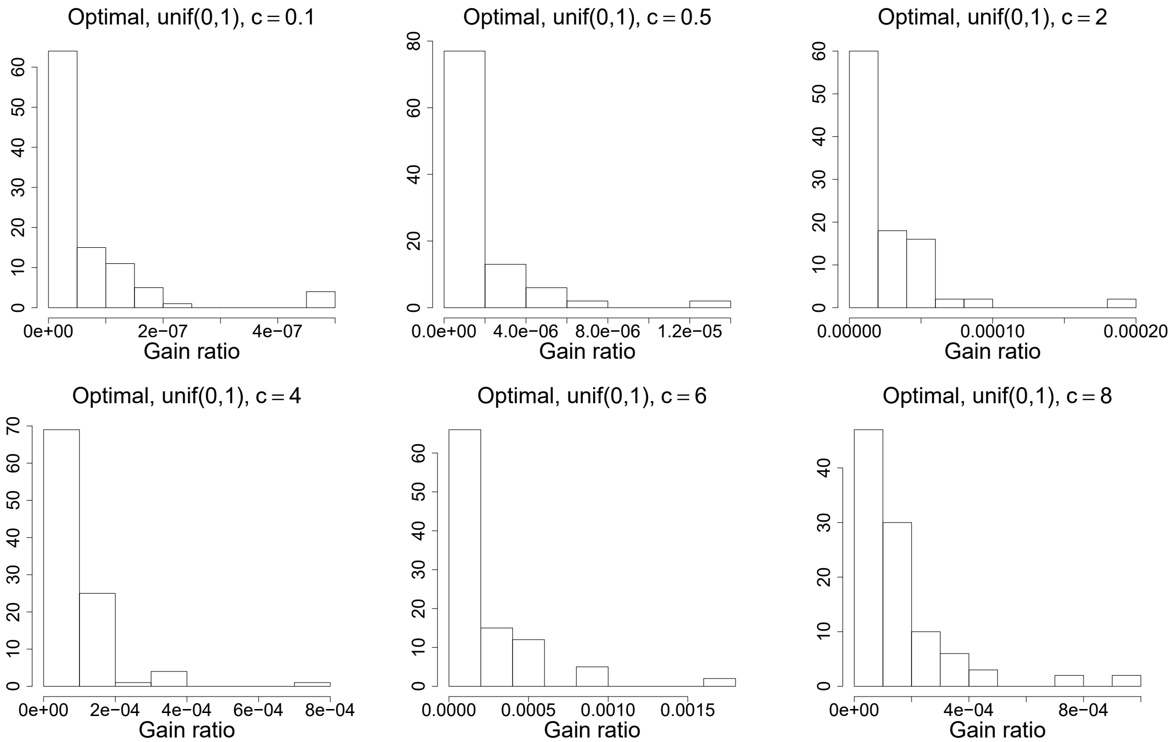

Figure 11: Histograms of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, for the uniform distribution on $[0, 1]$, for customer lifetime $d = 14$, discount rate $\delta = 0.5$, and for switching cost $c = 0.1, 0.5, 2, 4, 6, 8$.

and in the discriminatory setting, the experiments show that the gain ratio of optimal total revenue over Myerson total revenue is several orders of magnitude smaller than our theoretical bounds suggest.

R Numerical Experiments for Myerson Pricing in the Discriminatory Pricing Setting

We focus on the total revenue under a discriminatory strategy that offers separate prices to customers, depending on whether they were already existing customers at the time of the most recent introduction. In particular, at time t such that $s_j \leq t < s_{j+1}$ for some introduction $j \geq 1$, the provider offers technology class j at price $x_{j,n}$ to customers who arrived at time s_j or after, and at price $x_{j,e}$ to customers who arrived before time s_j. As argued in Appendix N, the optimal revenue in this discriminatory setting is an upper bound of the optimal revenue in the non-discriminatory

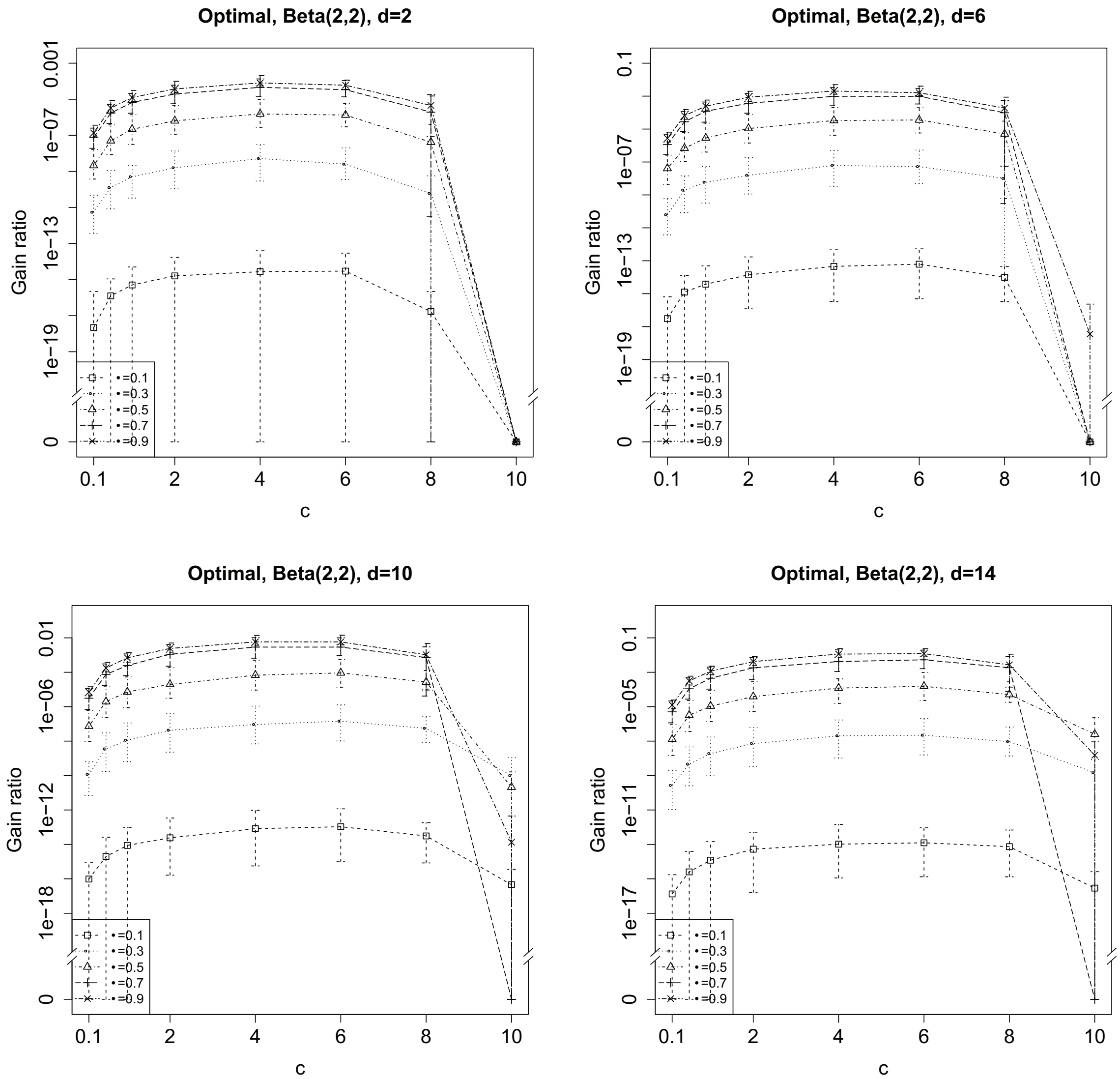

Figure 12: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c, for customer lifetime $d = 2, 6, 10, 14$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the beta distribution with shape parameters $\alpha = \beta = 2$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

setting.

Before presenting numerical experiments for the setting with discriminatory pricing, we first look at the corresponding revenue optimization problem. Fix introduction times $s_0 = 0, s_1, \ldots$ and

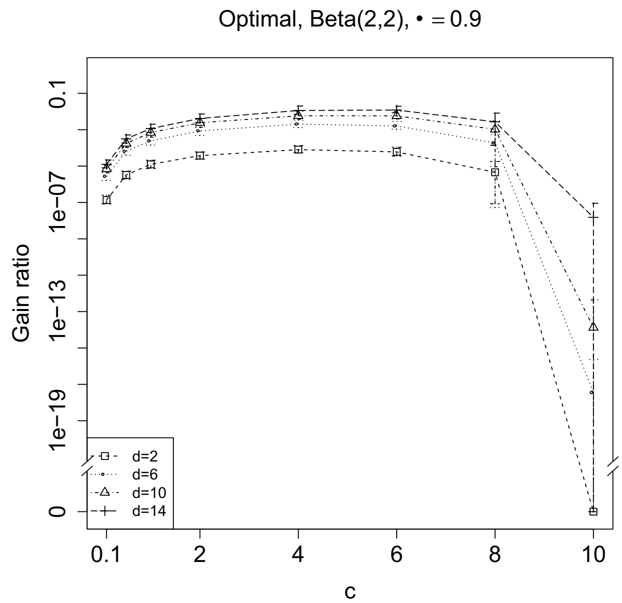

Figure 13: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c for customer lifetime $d = 2, 6, 10, 14$, for discount rate $\delta = 0.9$, for the beta distribution with shape parameters $\alpha = \beta = 2$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

introduction index j. We write down all the terms of total revenue that include $x_{j,n}$ or $x_{j+1,e}$:

$$\begin{aligned}&\sum_{t=s_j}^{s_{j+1}-1} \delta^t \cdot \min(t - s_j + 1, d)\left(1 - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} \\ &+ \sum_{t=s_{j+1}}^{s_{j+2}-1} \delta^t \cdot \max(d - (t - s_{j+1} + 1), 0)\left\{\left(1 - F\left(\max\left(\frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}, \frac{x_{j,n}}{s_j}\right)\right)\right) \cdot x_{j+1,e}\right. \\ &\qquad\qquad \left. + \left(F\left(\frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}\right) - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} \cdot \mathbb{1}_{\frac{x_{j,n}}{s_j} \leq \frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}}\right\},\end{aligned} \tag{105}$$

where the first summand is the revenue accumulated in periods $s_j, \ldots, s_{j+1} - 1$ from customers who arrive at or after period s_j and buy technology class j, and the second summad is the revenue accumulated in periods $s_{j+1}, \ldots, s_{j+2} - 1$ from customers who arrive before period s_{j+1}. In the second summand, the first term inside the curly brackets is the revenue from customers who switch to technology class $j + 1$ in period s_{j+1}, and the second term inside the curly brackets is the revenue from customers who do not switch to technology class $j + 1$ in period s_{j+1}.

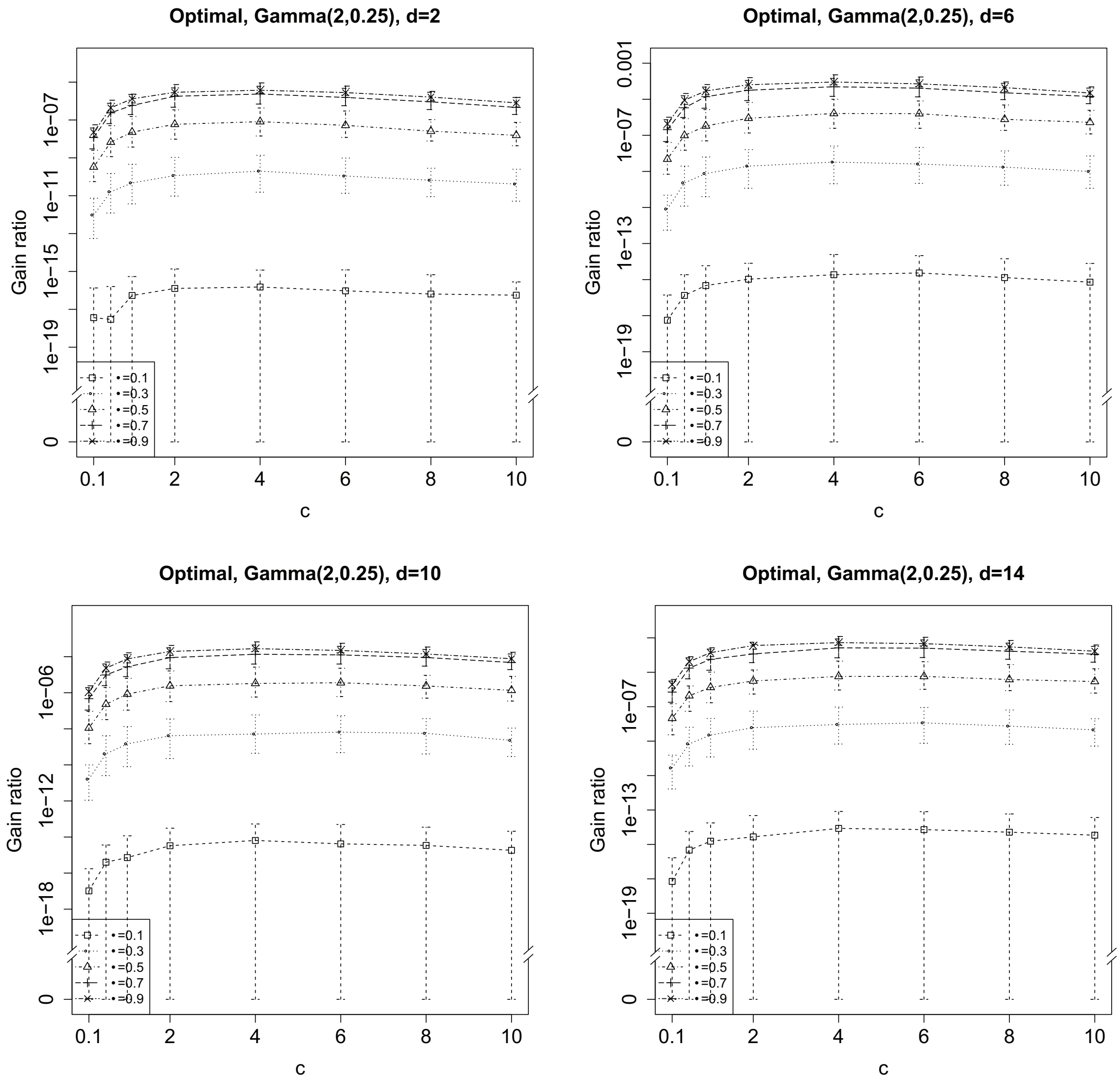

Figure 14: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c, for customer lifetime $d = 2, 6, 10, 14$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

Assuming $\frac{x_{j,n}}{s_j} \leq \frac{x_{j+1,e}-x_{j,n}+c}{s_{j+1}-s_j}$, the second summand of (105) can be rewritten as

$$\sum_{t=s_{j+1}}^{s_{j+2}-1} \delta^t \cdot \max(d-(t-s_{j+1}+1), 0) \left\{ \left(1 - F\left(\frac{x_{j,n}}{s_j}\right)\right) \cdot x_{j,n} + \left(1 - F\left(\frac{x_{j+1,e} - x_{j,n} + c}{s_{j+1} - s_j}\right)\right) \cdot (x_{j+1,e} - x_{j,n}) \right\}. \tag{106}$$

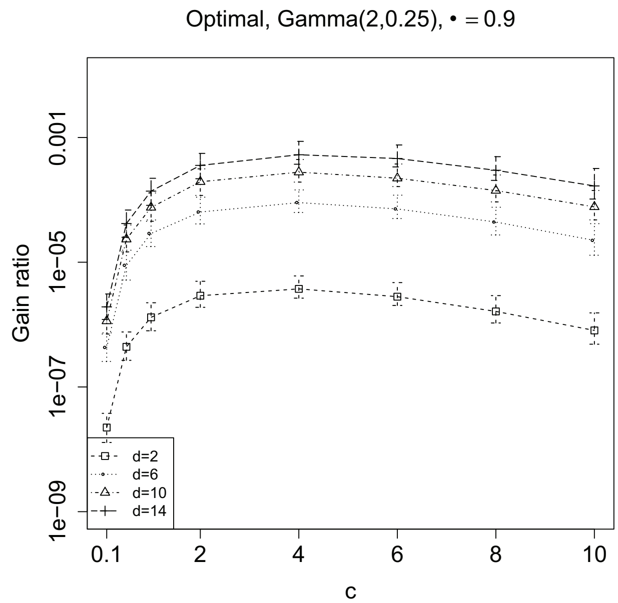

Figure 15: The average of the gain ratio of optimal total revenue over Myerson total revenue, over 100 simulations, against the switching cost c for customer lifetime $d = 2, 6, 10, 14$, for discount rate $\delta = 0.9$, for the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

Notice that the Myerson pricing $x_{j,n} = s_j p^*$ optimizes the first summand of (105), as well as the first term inside the curly brackets in (106). 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 (102). 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 16 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 customer lifetime d and the discount rate δ, for the uniform distribution on $[0, 1]$. The same patterns are observed as in the non-discriminatory

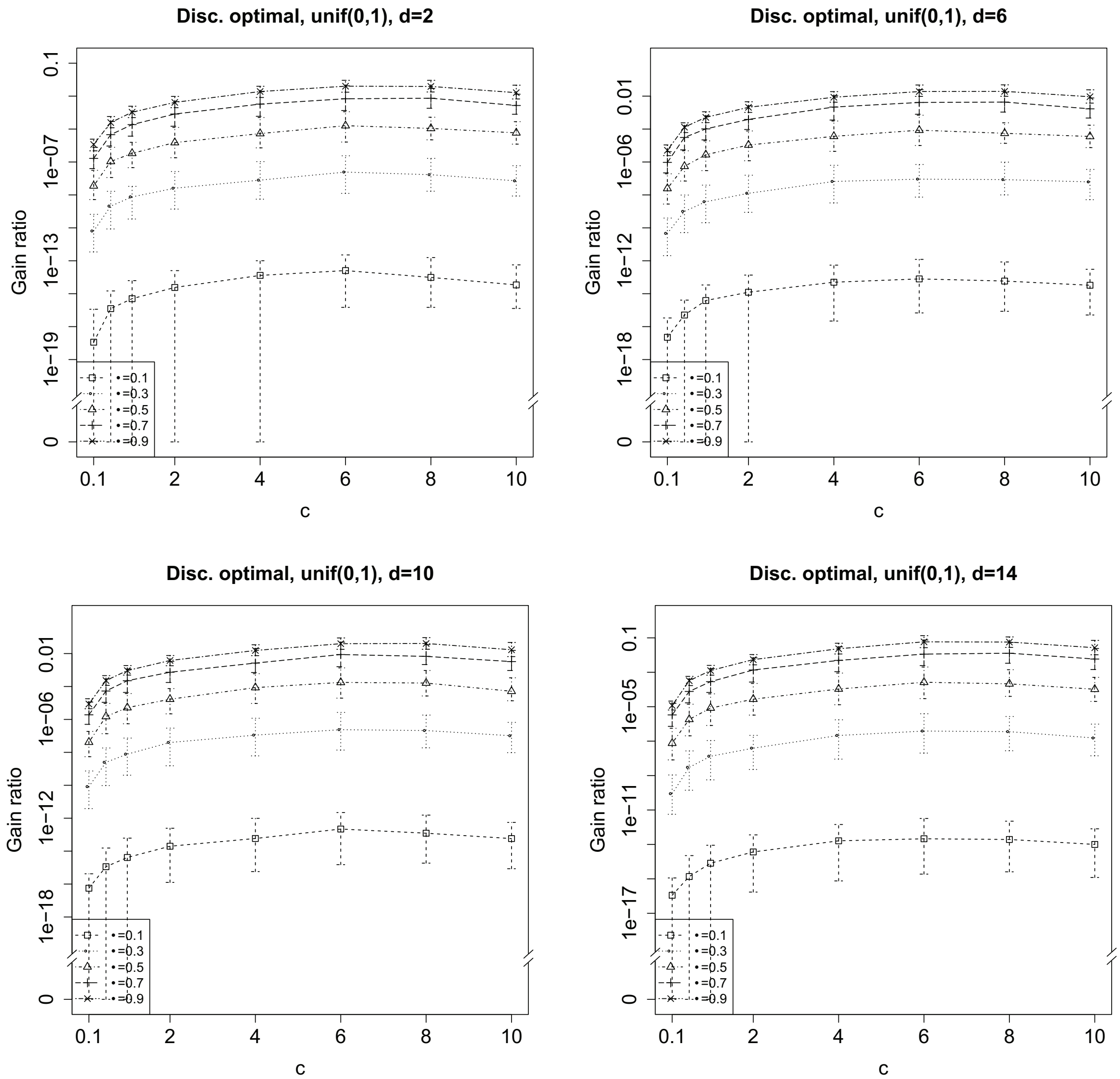

Figure 16: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c, for customer lifetime $d = 2, 6, 10, 14$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the uniform distribution on $[0, 1]$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

setting, with the difference that the values of the gain ratio are now generally larger by one order of magnitude or more.

Figure 17 shows that the gain ratio of the optimal total revenue over the Myerson total revenue increases with higher customer lifetime d, yet stays small even for quite long customer lifetimes, and even for intermediate switching costs. We note that the gain ratio is less than 0.1 for $d = 14$ in the worst case for the switching cost.

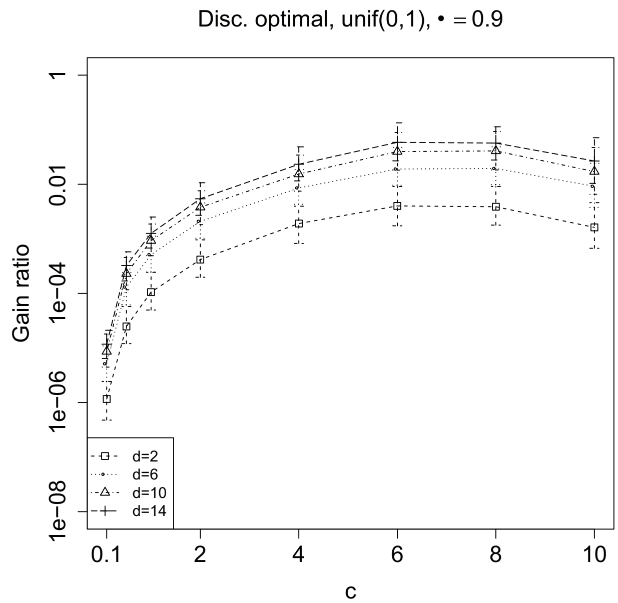

Figure 17: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c for customer lifetime $d = 2, 6, 10, 14$, for discount rate $\delta = 0.9$, for the uniform distribution on $[0, 1]$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

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 d, 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 maximum 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. In Figure 18 we show the histograms for a reasonably high value for the customer lifetime d, and we note that the gain ratio values for lower values of d are smaller. Figure 18 shows the effect of varying the discount rate δ in detail. Similarly to the non-discriminatory 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 non-discriminatory setting. Figure 19 shows again that in most instances Myerson pricing is essentially optimal, if instead we vary the switching cost c.

Figures 20 and 21 show experiments for the beta distribution, while Figures 22 and 23 show experiments for the gamma distribution, in the discriminatory setting. The results are consistent

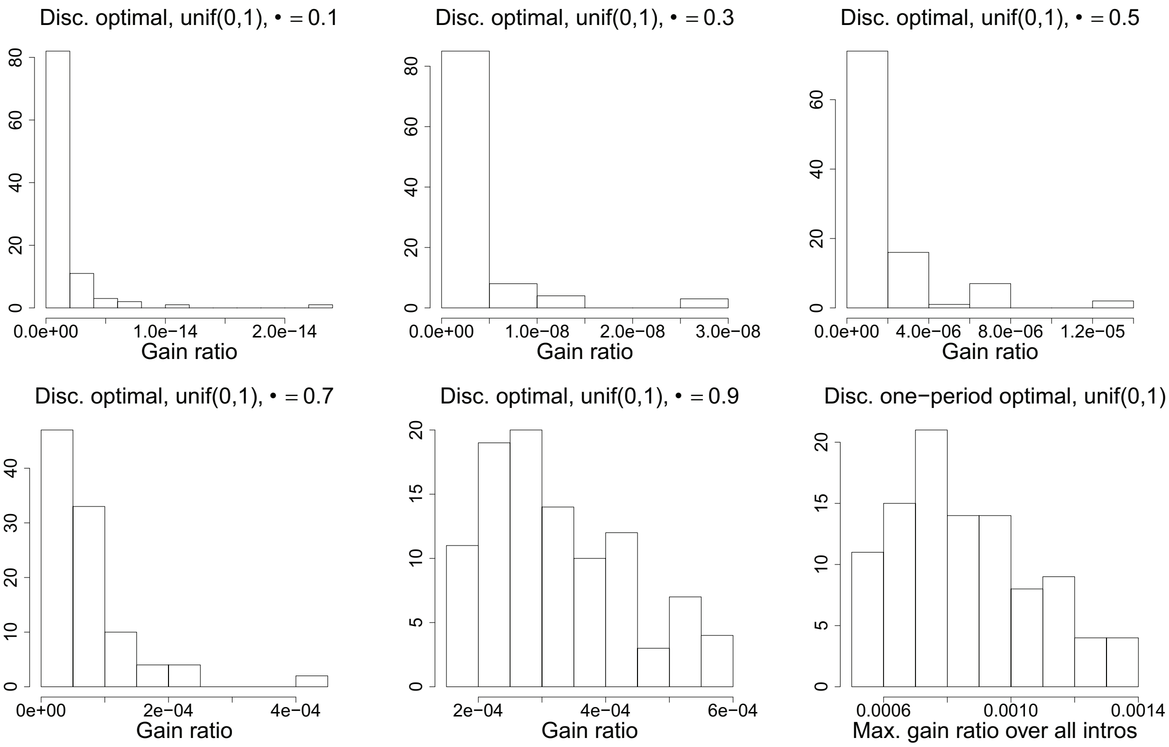

Figure 18: Histograms of the gain ratio of optimal total revenue over Myerson total revenue in the discriminatory setting, over 100 simulations, for the uniform distribution on $[0, 1]$, for customer lifetime $d = 14$, switching cost $c = 0.5$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$. 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 maximum gain ratio, in the discriminatory setting, over 100 simulations.

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 non-discriminatory setting, but still small in absolute terms.

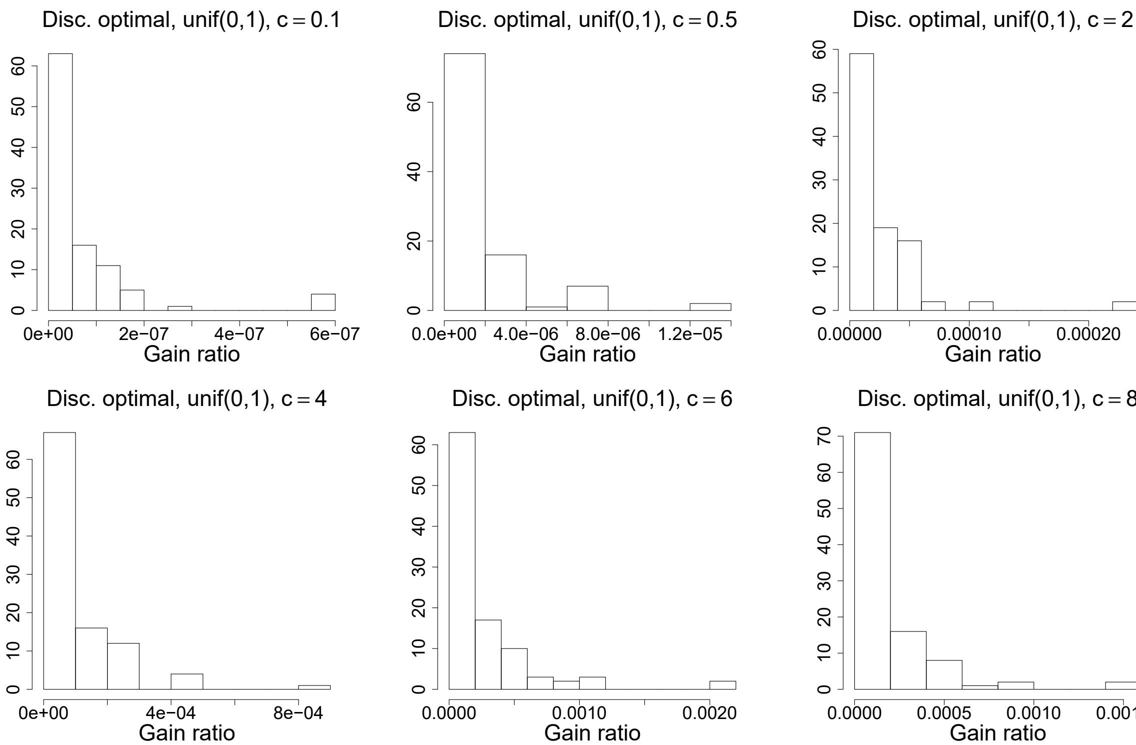

Figure 19: Histograms of the gain ratio of optimal total revenue over Myerson total revenue in the discriminatory setting, over 100 simulations, for the uniform distribution on $[0, 1]$, for customer lifetime $d = 14$, discount rate $\delta = 0.5$, and for switching cost $c = 0.1, 0.5, 2, 4, 6, 8$.

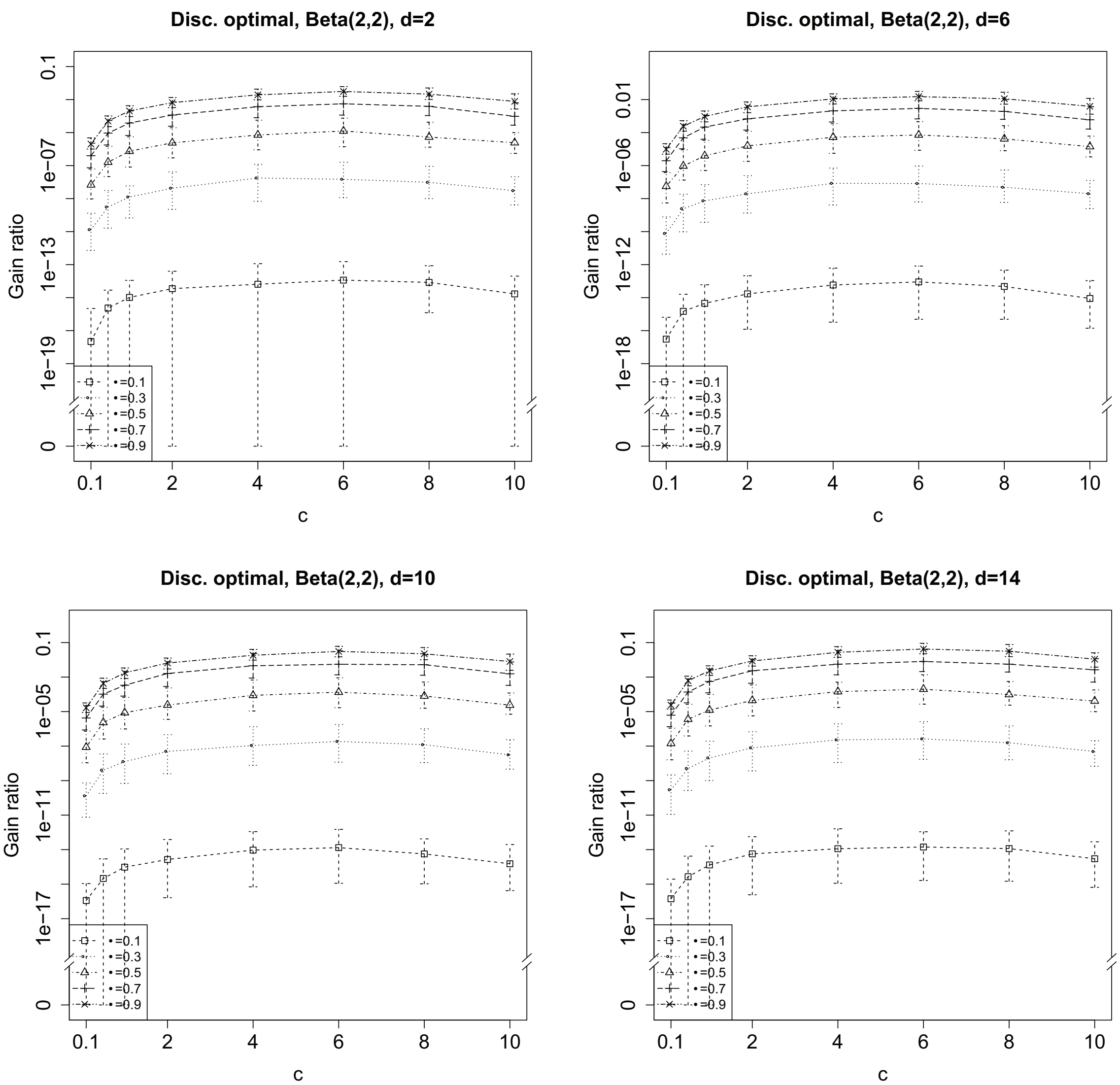

Figure 20: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c, for customer lifetime $d = 2, 6, 10, 14$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the beta distribution with shape parameters $\alpha = \beta = 2$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

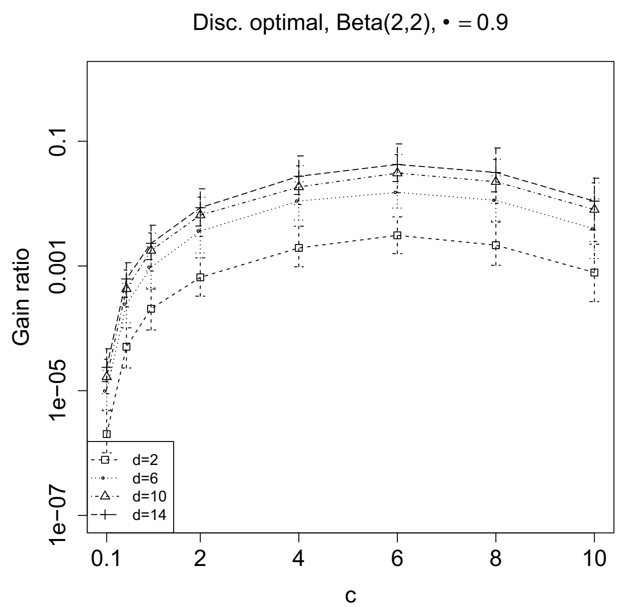

Figure 21: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c for customer lifetime $d = 2, 6, 10, 14$, for discount rate $\delta = 0.9$, for the beta distribution with shape parameters $\alpha = \beta = 2$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

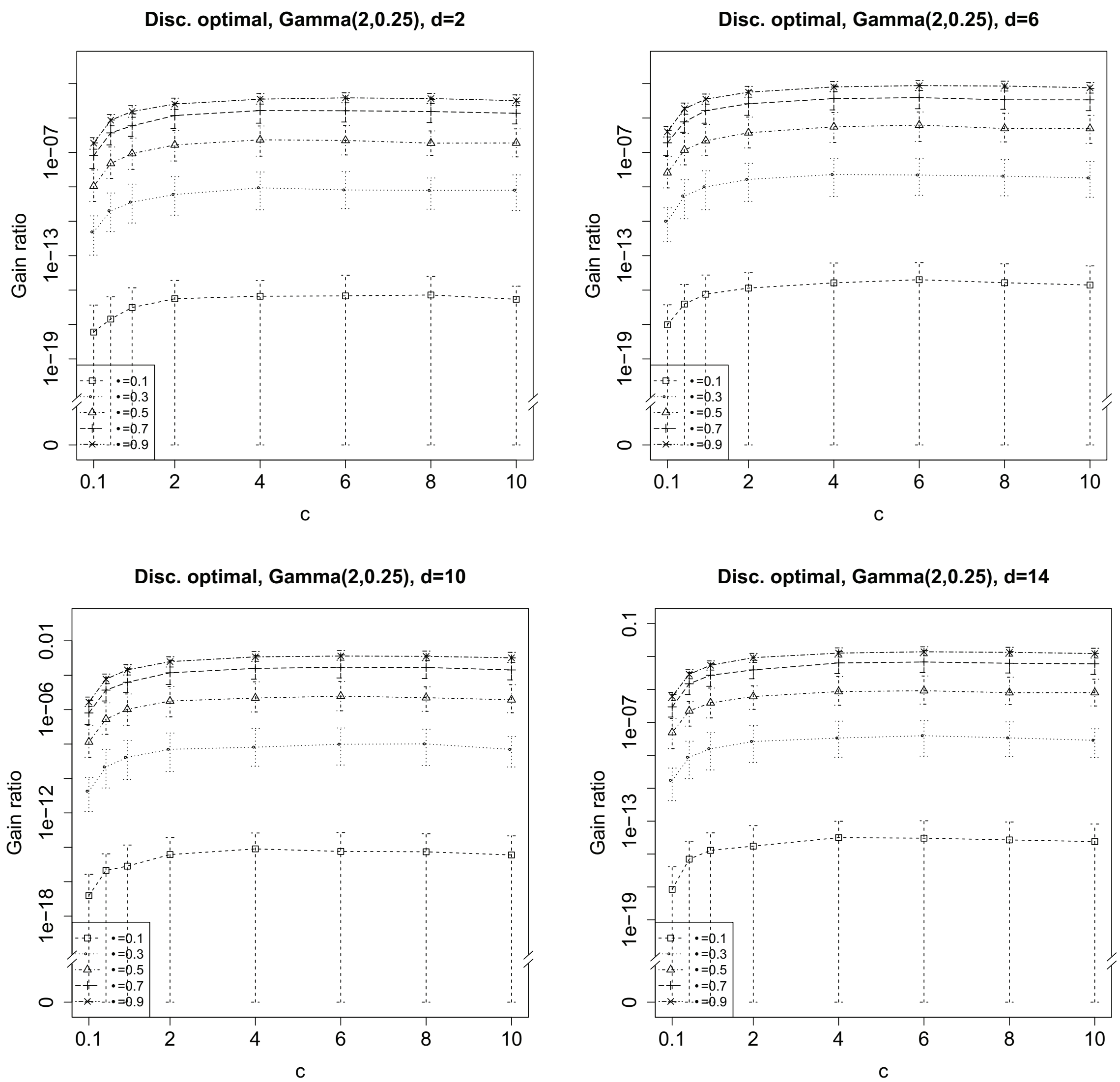

Figure 22: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c, for customer lifetime $d = 2, 6, 10, 14$, and for discount rate $\delta = 0.1, 0.3, 0.5, 0.7, 0.9$, for the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.

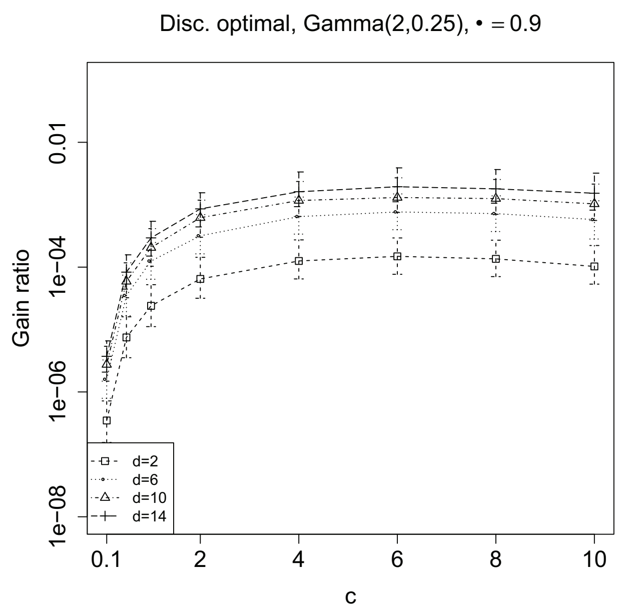

Figure 23: The average of the gain ratio of optimal total revenue in the discriminatory setting over Myerson total revenue, over 100 simulations, against the switching cost c for customer lifetime $d = 2, 6, 10, 14$, for discount rate $\delta = 0.9$, for the gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.25$. The whiskers indicate the minimum and maximum gain ratio over 100 simulations for each setting.