The Value of Personalized Pricing

Adam N. Elmachtoub

Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, adam@ieor.columbia..edu

Vishal Gupta

Data Science and Operations, USC Marshall School of Business, Los Angeles, CA 90089, guptavis@usc.edu

Michael L. Hamilton

Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, mh3461@columbia.edu

Increased availability of high-quality customer information has fueled interest in personalized pricing strategies, i.e., strategies that predict an individual customer's valuation for a product and then offer a customized price tailored to that customer. While the appeal of personalized pricing is clear, it may also incur large costs in the form of market research, investment in information technology and analytics expertise, and branding risks. In light of these tradeoffs, our work studies the value of idealized personalized pricing over a spectrum of pricing strategies varying in pricing flexibility and prediction model accuracy.

We first provide tight, closed-form upper bounds on the ratio between the profits of an idealized personalized pricing strategy and a single price strategy. These bounds depend on simple statistics of the valuation distribution and shed light on the types of markets for which personalized pricing has the most potential. Next, we consider two stylized price discrimination strategies that isolate the key assumptions underlying idealized personalized pricing: *(i)* a k-market segmentation strategy where the firm knows all customer valuations precisely but can only charge customers one of k prices and *(ii)* a feature-based pricing strategy, where the firm can charge a continuum of prices, but no longer knows customer valuations precisely. For each strategy, we bound the ratio of idealized personalized pricing profits to the profits of that strategy. These bounds quantify the value of the operational capability of charging distinct prices and the value of additional predictive accuracy, respectively. Finally, we synthesize these results to study a more realistic personalization strategy in which the seller neither knows customer valuations precisely nor is able to offer a continuum of prices. Computational evidence suggests that our bounds are both qualitatively and quantitatively representative under several common valuation distributions.

Key words: price discrimination, personalization, market segmentation

History: Submitted October 2018.

1. Introduction

Over the last decade, increased availability of high-quality customer information has fueled interest in personalized pricing strategies. At a high-level, these strategies combine customer data with machine learning and optimization tools to predict an individual cus-

tomer's willingness to pay and then customize a price for that customer. This customized price can be delivered as a discount via a mobile application or other channel.

The appeal of personalized pricing is clear – If a seller could accurately predict individual customer valuations, then it could (in principle) charge each customer exactly their valuation, increasing profits and market penetration. Given this appeal, grocery chains (Clifford 2012), department stores (D'Innocenzio 2017), airlines (Tuttle 2013), and many other industries (Obama 2016) have begun experimenting with personalized pricing. Moreover, within the operations community, there has been a surge in research on how to practically and effectively implement personalized pricing stragies (e.g., Aydin and Ziya (2009), Phillips (2013), Bernstein et al. (2015), Chen et al. (2015), Ban and Keskin (2017)).

Unfortunately, implementing any form of price discrimination, including personalized pricing, may be costly and/or difficult. A firm would need to engage in price experimentation and market research, invest in information systems to store customer data, and build analytics expertise to transform these data into a personalized pricing strategy (see Arora et al. (2008) for an extensive discussion). Moreover, price discrimination tactics involve serious branding risks and potential customer ill-will, and, in some markets, may be of questionable legality. Finally, personalized pricing may impact competitors' (Zhang 2011) and manufacturers' (Liu and Zhang 2006) behavior.

In light of these tradeoffs, in this work we complement the existing operations literature on *how* to implement personalized pricing by quantifying *when* personalized pricing offers significant value. Specifically, for a single-product monopolist, we bound the profit ratio between *idealized personalized pricing (PP)*, i.e., charging each customer exactly their willingness to pay, and a spectrum of various simpler pricing strategies. The spectrum of strategies vary on the degree of pricing flexibility as well as prediction model accuracy. Thus, these bounds can guide managers in assessing the potential upside of the above tradeoffs, and provide a fundamental understanding of the value of offering more prices and of the value of reducing prediction error.

With full-information about the customer valuation distribution, computing the exact ratio between idealized personalized pricing over simpler pricing strategies is straightforward; there is no need for bounding. However, in our opinion, a firm not currently engaging in personalized pricing is unlikely to know the full valuation distribution. Indeed, it is not necessary to learn this distribution to price effectively (Besbes et al. 2010, Besbes and Zeevi

2015) and learning it may be difficult since real-world distributions are typically complex and irregular (see, e.g., Celis et al. (2014) for a discussion in an auction setting).

Consequently, we focus instead on parametric bounds that depend on a few statistics of the valuation distribution. On the one hand, we believe these statistics are more easily estimated by a seller not currently engaging in personalized pricing than the full valuation distribution. On the other hand, and perhaps more importantly, parametric bounds based on these statistics provide structural insights into the types of markets where the value of personalized pricing is potentially large. In particular, we leverage these structural insights to disentangle the contributions from increased operational flexibility (offering many distinct prices) and improved prediction accuracy (gathering additional data) in personalized pricing strategies.

More specifically, in the first part of the paper, we bound the profit ratio between idealized personalized pricing and posting a *single price (SP)* for all customers. We call this ratio the *value of personalized pricing over single-pricing*. Notice that idealized personalized pricing as we define it is often called first-degree price discrimination in the economics literature, and observe that it upper bounds the profit of any other price discrimination strategy. Thus, the value of personalized pricing over single-pricing also upper bounds the potential gains of any other price discrimination strategy over single pricing.

We prove bounds that are tight, closed-form and depend on three unitless statistics of the valuation distribution: (i) the *scale*, which is the ratio of the upper bound of the support to the mean, (ii) the *margin*, which we define as the margin of a unit sold at a price equal to the mean valuation, and (iii) the *coefficient of deviation*, which is the mean absolute deviation over twice the mean. Knowing these three quantities is equivalent to knowing the mean, support, and mean absolute deviation of the distribution. Our bounds are tight in the sense that we give an explicit valuation distribution for which the value of personalized pricing over single-pricing matches the bound. The precise form of the tight distribution depends on the relevant parameters, but consists of a mixture of Pareto and two-point distributions. These results generalize folklore results that the Pareto distribution (a.k.a. "equal-revenue" distribution) represents the worst-case for single-pricing. Perhaps surprisingly, we also find that our bound is maximal for intermediate values of the coefficient of deviation and approaches one as the coefficient deviation increases with all other parameters fixed.

Of course, idealized personalized pricing is not achievable in practice. It hinges on two assumptions: First, the monopolist has the ability to charge a potentially distinct price to each customer. Second, the monopolist is omniscient and can perfectly predict each customer's valuation. In the second part of this paper, we study price discrimination strategies that relax these two assumptions and more closely model personalized pricing strategies used in practice. To this end, we first compute the value of personalized pricing over two stylized price-discrimination strategies: k-market segmentation and feature-based pricing.

In the *k-market segmentation (kP)* strategy, we assume the monopolist is still omniscient, but can charge at most k distinct prices, relaxing the assumption of a continuum of prices. Thus, the value of personalized pricing over k-market segmentation quantifies the value of the operational capability of charging a continuum of prices, which is equivalent to the case where $k \to \infty$. Under a mild assumption, we show that this value is at most $1 + C\frac{1}{k}$, where C is an explicit constant depending on distributional parameters. We prove theoretically that this worst-case dependence on k is tight and provide numerical evidence that it is in fact typical of many distributions. This analysis yields a natural rule of thumb; to half the gap to the ideal personalized pricing profits, one needs to double the number of prices offered.

By contrast, in the *feature-based pricing (XP)* strategy, we assume the monopolist can in principle offer a continuum of prices, but is no longer omniscient. Rather, she observes a feature vector (sometimes called a context) for each customer which she can use to (imperfectly) predict the customer's valuation. Thus, the value of personalized pricing over feature-based pricing quantifies the value of additional information, i.e., a richer set of features that would enable perfect prediction. Leveraging our earlier results, we prove that this value is bounded by an explicit factor that depends on the coefficient of deviation of the error in the valuation prediction model. Thus, our bound quantifies the degree of prediction accuracy necessary to guarantee a certain percentage of profits. Again, we provide numerical evidence suggesting our worst-case analysis is qualitatively typical of many valuation distributions. Our result yields another natural rule of thumb; to half the gap to the ideal personalized pricing profits, one needs to quadruple the the prediction accuracy.

Finally, we use the above results on these stylized pricing strategies as building blocks to study a more realistic *feature-based market segmentation (kXP)* strategy . In this strategy,

Figure 1   Personalized Pricing Landscape
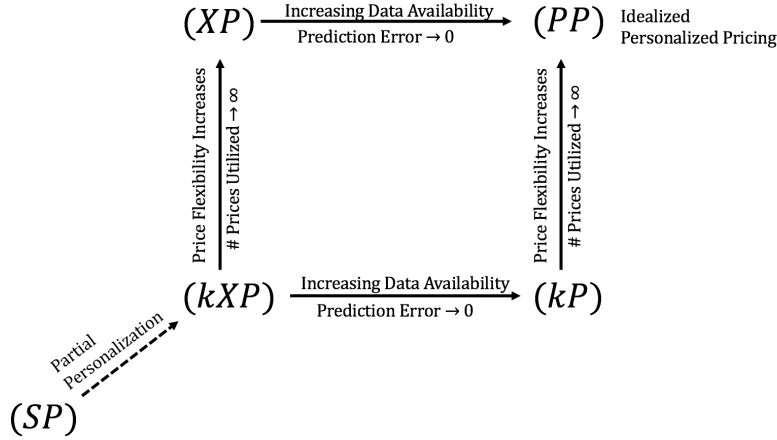
Note. We represent our pricing models in terms of their *price flexibility* (y-axis) and *prediction accuracy* (x-axis).

we assume the monopolist is neither omniscient nor operationally able to offer a continuum of prices. Rather, as in the feature-based pricing strategy, she observes a feature for each customer. Based on this feature, she then offers the customer one of k prices. Bounding the relative difference between idealized personalized pricing and feature-based market segmentation quantifies the impact of both limited price flexibility and prediction error on the profit of personalized pricing strategies. We believe that feature-based market segmentation also closely resembles data-driven price discrimination strategies commonly used in practice. Under mild assumptions, we show that one can decompose the value of personalized pricing over feature-based market segmentation by separately considering the profit loss from prediction inaccuracy and the profit loss from limited price flexibility on a related, "de-noised" market. These two losses can be analyzed directly using the previously discussed bounds. Importantly, our decomposition is constructive and yields an algorithm for generating a feature-based market segmentation strategy with a provable performance guarantee. In Fig. 1 we visualize the relations between all of our various pricing strategies.

   To summarize our contributions:

1. We prove closed-form, tight bounds for the value of personalized pricing over single-pricing when the scale, margin, and coefficient of deviation of the valuation distribution are known (c.f. Theorem 1). This bound is maximized for intermediate levels of market heterogeneity. We also provide corresponding valuation distributions which confirm these bounds are tight.

2. We prove closed-form bounds on the value of personalized pricing over k-market segmentation which are tight in their dependence on k, and describe a distribution-agnostic segmentation procedure that achieves this bound (c.f. Theorem 2). We provide numerical evidence that this worst-case dependence is in fact typical. Thus, the bound quantifies the operational value of being able to charge infinitely many prices over k prices.

3. We further prove closed-form bounds on the value of personalized pricing over feature-based pricing (c.f. Theorem 3). The bound gives an explicit relationship between the accuracy in predicting valuations and value of personalized pricing. Thus, these bound help quantify the value of additional consumer data.

4. We analyze the value of personalized pricing over feature-based market segmentation by synthesizing our results on k-market segmentation and feature-based pricing. We show that the feature-based market segmentation strategy's profit loss can be bounded by the sum of the profit loss from feature-based pricing and the profit loss of k-market segmentation on a related, "noiseless" market (c.f. Theorem 4). Thus, the decomposition yields a useful, rigorous paradigm for tuning personalized pricing strategies by explicitly identifying the impact on revenue due to limited price flexibility and prediction error.

1.1. Connections to Existing Literature

The study of price discrimination tactics has a long history in economics dating back at least to Robinson (1934). Historically, the economics literature has focused on how various forms of price discrimination affect social welfare (see, e.g., Narasimhan (1984), Schmalensee (1981), Varian (1985), Shih et al. (1988) or Bergemann et al. (2015), Cowan (2016), Xu and Dukes (2016) for more recent results). In contrast to these works, we take an operational perspective, focusing on the individual firms relative profits under first-degree price discrimination and other forms of pricing.

That said, we are not the first to study the value of personalized pricing over single pricing. Previous authors have also studied the value of personalized pricing over single pricing under different distributional assumptions. Barlow et al. (1963) prove that if the valuation distribution has monotone hazard rates, the value of personalized pricing is at most Euler's constant $e \approx 2.718$. Similarly, Tamuz (2013) prove that if the ratio of the geometric mean over the mean is at least $1 - \delta$, then the value of personalized pricing is at

most $(1 - 2^{\frac{4}{3}} \delta^{\frac{1}{3}})^{-1}$. Our single-pricing results differ from these existing results in two critical ways. First, our bounds are tight in the input parameters. Indeed, we show numerically they can be significantly stronger than these existing bounds. Second, since our bounds explicitly depend on simple statistics of the valuation distribution such as the scale and coefficient of deviation, we argue that is easier to use them to assess the effects of various operational decisions. In particular, the parametric dependence on the support and mean absolute deviation of the distribution directly enables us to study personalization strategies which segment markets and leverage features to reduce valuation uncertainty, respectively. It is less clear how to use existing techniques to develop similar results for such strategies.

As mentioned above, idealized personalized pricing (first-degree price discrimination) is an idealized strategy. In practice, firms implement some form of third-degree price discrimination. While approaches differ widely, most implicitly or explicitly leverage some form of market segmentation – either segment customers directly or incentivize them to self-segment, and then offer different prices to each segment. Indeed, the operations literature contains many examples of such strategies including intertemporal pricing (Su (2007), Besbes and Lobel (2015)), opaque selling (Jerath et al. (2010), Elmachtoub and Hamilton (2017)), rebates/promotions (Chen et al. (2005), Cohen et al. (2017)), markdown optimization (Caro and Gallien (2012), Özer and Zheng (2015)), product differentiation (Moorthy (1984), Choudhary et al. (2005)), dynamic pricing and learning (Cohen et al. (2016), Qiang and Bayati (2016), Javanmard and Nazerzadeh (2016)), and many others.

By contrast, the focus of our work is not on "how to price discriminate" but rather the value of price discrimination. Nonetheless, our bounds on the value of personalized pricing over feature-based market segmentation do provide insight into the above strategies. Since the strategy relaxes the two key assumptions of first-degree price discrimination, our bounds help establish guarantees on the performance of imperfect personalized pricing strategies. Perhaps more importantly, by characterizing the types of valuation distributions for which the value of personalized pricing is high, our bounds highlight the settings in which it is most important that the above strategies perform well and inform their analysis.

Finally, we contrast our work to several recent works that study how to set a single-price near-optimally given limited distribution information such as the support (Cohen et al. 2015), mean and variance (Chen et al. 2017, Azar et al. 2013), or a neighborhood containing the true valuation distribution (Bergemann and Schlag 2011). Indeed, these works support

our earlier claim that it is not generally necessary to learn the whole valuation distribution in order to price effectively, but are very different in perspective from our work.

2. Model and Preliminaries

We consider a profit-maximizing monopolist selling a product with per unit cost c. A random customer's valuation for the product is denoted by the non-negative random variable $V \sim F$. The mean valuation $E[V]$ is denoted by μ. Since it is never profitable to sell to customers with valuations less than c, assume without loss of generality, that $V \geq c$ almost surely. We consider a spectrum of five pricing strategies for the monopolist:

1) Single Pricing (SP): In the single pricing strategy, the monopolist offers the product to all customers at the same price p. Thus, the probability that a customer purchases is given by the complementary cumulative distribution function (ccdf) $\overline{F}(p) := 1 - F(p)$, and the seller's corresponding expected profit is $(p-c)\overline{F}(p)$. Let $\mathcal{R}_{SP}(F,c) := \max_p\{(p-c)\overline{F}(p)\}$ denote the seller's maximal expected profit under single-pricing.

2) k-Market Segmentation (kP): In the k-market segmentation strategy, the monopolist partitions the valuation space into $k + 1$ disjoint intervals $[s_0, s_1), [s_1, s_2), \ldots, [s_{k-1}, s_k), [s_k, s_{k+1})$ with $s_0 = c$ and assigns distinct prices $p_i \in [s_i, s_{i+1})$ to each segment $i \geq 1$. When a customer with valuation $V \in [s_i, s_{i+1})$ arrives, she is offered the product at price p_i. Given F and c, let $\mathcal{R}_{kP}(F, c, \mathbf{s}, \mathbf{p})$ denote the profit from this strategy with partition $\mathbf{s}$ and prices $\mathbf{p}$, and let $\mathcal{R}_{kP}(F, c) \equiv \max_{\mathbf{s}, \mathbf{p}} \mathcal{R}_{kP}(F, c, \mathbf{s}, \mathbf{p})$ denote the optimal profit for this strategy.

3) Feature-Based Pricing (XP): In the feature-based pricing strategy, the monopolist observes a feature vector $\mathbf{X}$ supported on $\mathcal{X}$ for each customer before offering a price, but does *not* directly observe her valuation V. Based on $\mathbf{X}$, she offers a customized price $p(\mathbf{X})$, and the customer purchases with probability $\Pr(V \geq p(\mathbf{X}) \mid \mathbf{X})$. Note that unlike k-market segmentation, if $\mathbf{X}$ is continuous, the seller can in principle offer a continuum of prices, one for each possible value of $\mathbf{X}$. Given a joint distribution $F_{\mathbf{X}V}$ of $(\mathbf{X}, V)$, let $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c) \equiv \max_{p(\cdot)} E\left[(p(\mathbf{X}) - c)\mathbb{I}(V \geq p(\mathbf{X}))\right]$ denote the optimal profit under feature-based pricing.

4) Feature-Based Market Segmentation (kXP): In the feature-based market-segmentation strategy, the monopolist observes a feature $\mathbf{X}$ for each customer, but again

does not directly observe her valuation V. Based on $\mathbf{X}$, she offers one of k prices, $p(\mathbf{X}) \in \{p_i\}_{i=1}^k$, and the customer purchases with probability $\Pr(V \geq p(\mathbf{X})|\mathbf{X})$. The monopolists choice of pricing function naturally induces a partition of the market into k segments $\mathcal{X}_i = \{\mathbf{x} \in \mathcal{X}|p(\mathbf{x}) = p_i\}$, and yields expected profit $\sum_{i=1}^k (p_i - c)\Pr(V \geq p_i|\mathbf{X} \in \mathcal{X}_i)\Pr(\mathbf{X} \in \mathcal{X}_i)$. Given a joint distribution $F_{\mathbf{X}V}$ of $(\mathbf{X}, V)$, let $\mathcal{R}_{kXP}(F_{\mathbf{X}V}, c) \equiv \max_{p_1,...,p_k,\mathcal{X}_1,...,\mathcal{X}_k} \sum_{i=1}^k (p_i - c)\Pr(V \geq p_i|\mathbf{X} \in \mathcal{X}_i)\Pr(\mathbf{X} \in \mathcal{X}_i)$ denote the optimal profit for this strategy.

5) Idealized Personalized Pricing (PP): In the idealized personalized pricing strategy, the monopolist can potentially offer a different price to each customer and has full knowledge of each customer's valuation. Since $V \geq c$, it is optimal to offer each customer precisely her valuation. Let $\mathcal{R}_{PP}(F, c) := \mu - c$ denote the seller's maximal expected profit under idealized personalized pricing.

By construction, $\mathcal{R}_{SP}(F, c) \leq \mathcal{R}_{kXP}(F_{\mathbf{X}V}, c) \leq \mathcal{R}_{kP}(F, c) \leq \mathcal{R}_{PP}(F, c)$ and $\mathcal{R}_{SP}(F, c) \leq \mathcal{R}_{kXP}(F, c) \leq \mathcal{R}_{XP}(F_{\mathbf{X}V}, c) \leq \mathcal{R}_{PP}(F, c)$. However, in general the ordering between $\mathcal{R}_{kP}(F, c)$ and $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c)$ is instance dependent. Given F and c, we define the *value of personalized pricing over single-pricing* as $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)}$. The value of personalized pricing over k-market segmentation, feature-based pricing and feature-based market segmentation are each defined similarly. When F, $F_{\mathbf{X}V}$, and c are clear from context, we sometimes omit them and write, e.g., $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$.
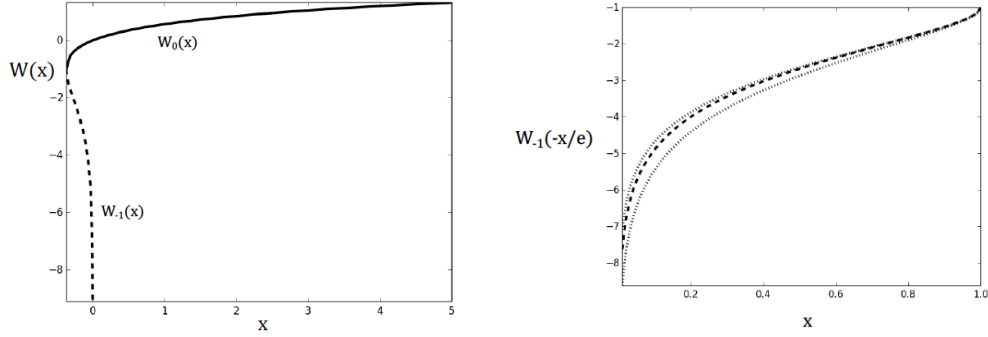
2.1. The Lambert-W Function

Many of our closed-form bounds involve $W_{-1}(\cdot)$, the negative branch of the *Lambert-W* function. Although the Lambert-W function is pervasive in mathematics, it is less common in the pricing literature. We refer the reader to Corless et al. (1993) for a thorough review of its properties and provide only a brief summary below.

Recall, the general (multi-valued) Lambert-W function $W(x)$, is defined as a solution to

$$W(x)e^{W(x)} = x.$$

When $x \in [-1/e, 0)$, this equation has two distinct real solutions. The branch $W_{-1}(\cdot)$ gives the solution that lies in $(-\infty, -1]$. The other branch $W_0(\cdot)$ gives the solution in $[-1, \infty)$, but will not be needed in our work. Both branches are illustrated in the left panel of Fig. 2.

Figure 2   The Lambert-W Function.

Note. The left panel shows the two real branches of the Lambert-W function, $W_0(\cdot)$ (solid black), and $W_{-1}(\cdot)$ (dashed). Our bounds depend upon the $W_{-1}(\cdot)$ branch (rescaled), as shown in right panel, and which can be upper and lower bounded via Chatzigeorgiou (2013) (dotted).

To build intuition, we encourage the reader to think of $W_{-1}(\cdot)$ as analogous to the natural logarithm, $\log(\cdot)$ Indeed, like $W_{-1}(x)$, $\log(x)$ is defined as a solution to an equation, namely,

$$e^{\log(x)} = x.$$

For a handful of values, both $W_{-1}(\cdot)$ and $\log(\cdot)$ can be evaluated exactly. For example, $W_{-1}(-1/e) = -1$, $\log(1) = 0$, and $\lim_{x \to 0} W_{-1}(x) = \lim_{x \to 0} \log(x) = -\infty$. For most values, however, both functions must be evaluated numerically. Fortunately, evaluating an expression using $W_{-1}(\cdot)$ is numerically no more difficult than evaluating a similar expression using $\log(\cdot)$.

Moreover, the natural logarithm provides simple bounds on $W_{-1}(\cdot)$. Indeed, Chatzigeorgiou (2013) proves that for $0 < x \leq 1$,

$$-1 - \sqrt{2\log(1/x)} - \log(1/x) \;\leq\; W_{-1}\left(-\frac{x}{e}\right) \;\leq\; -1 - \sqrt{2\log(1/x)} - \frac{2}{3}\log(1/x). \quad (1)$$

(Recall $W_{-1}(\cdot)$ is defined on $[-1/e, 0)$, so that this inequality spans its domain.) The right panel in Fig. 2 illustrates these bounds and shows they are quite tight.

3.   The Value of Personalized Pricing over Single Pricing

In this section, we provide tight bounds on the value of personalized pricing over single pricing using simple statistics of F. We begin by bounding the value of personalized pricing using the *scale* (S), and *margin* (M), defined respectively as:

$$S := \frac{\inf\{k \mid F(k) = 1\}}{\mu}, \quad M := 1 - \frac{c}{\mu}.$$

These two statistics are unitless and can be thought of as (rescaled) measurements of the maximal valuation and per unit cost. More specifically, S is the ratio of the largest valuation in the market to the average valuation. By construction, $S \geq 1$, and measures the maximal dispersion of valuations. By contrast, $M = \frac{\mu - c}{\mu} \in [0, 1]$, and can be interpreted as the margin of a unit sold at a price equal to the mean valuation,

Before stating our tight bound, we introduce a transformation that reduces the problem of bounding the value of personalization for a product with $c > 0$ and $\mu > 0$ to an equivalent problem with $c = 0$ and $\mu = 1$. This reduction is used repeatedly throughout the paper.

LEMMA 1 (**Reduction to Zero Costs and Unit Mean**). *Let $V \sim F$, and let the distribution of $\frac{1}{\mu - c}(V - c)$ be denoted by F_c. Then,*

$$\frac{\mathcal{R}_{PP}(F, c)}{\mathcal{R}_{SP}(F, c)} = \frac{\mathcal{R}_{PP}(F_c, 0)}{\mathcal{R}_{SP}(F_c, 0)}.$$

Moreover, if the scale and margin of F are S and M, respectively, then the mean, scale, and margin of F_c are $\mu_c = 1$, $S_c = \frac{S + M - 1}{M}$, and $M_c = 1$, respectively.

The key to the following bound is that $\mathcal{R}_{SP}(F, 0)$ directly yields a bound on the tail behavior of F. Indeed, for any price $p > 0$, $p\overline{F}(p) \leq \mathcal{R}_{SP}(F, 0)$ by definition, and thus $\overline{F}(p) \leq \mathcal{R}_{SP}(F, 0)/p$. We use this result repeatedly in what follows, terming it the *pricing inequality*:

$$\overline{F}(x) \leq \frac{\mathcal{R}_{SP}(F, 0)}{x}, \quad \forall x > 0. \qquad \text{(Pricing Inequality)}$$

This inequality drives the following lemma.

LEMMA 2 (**Bounding $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ using the Scale and Margin**). *For any F with scale S and margin M, we have*

$$\frac{\mathcal{R}_{PP}(F, c)}{\mathcal{R}_{SP}(F, c)} \leq -W_{-1}\left(\frac{-M}{e(S + M - 1)}\right).$$

Moreover, this bound is tight.

Proof. First, suppose $c = 0$ and $\mu = 1$. Then, $\mathcal{R}_{PP} = 1$ and $M = 1$. Since $\mu = 1$, $\overline{F}(S) = 0$, i.e., $0 \leq V \leq S$, a.s. Using the tail integral formula for expectation, we have that

$$\mathcal{R}_{PP} = \int_0^S \overline{F}(x)dx \qquad (2)$$

$$\leq \mathcal{R}_{SP} + \int_{\mathcal{R}_{SP}}^S \overline{F}(x)dx \qquad (0 \leq \mathcal{R}_{SP} \leq S) \qquad (3)$$

$$\leq \mathcal{R}_{SP} + \int_{\mathcal{R}_{SP}}^{S} \frac{\mathcal{R}_{SP}}{x} dx \qquad \text{(Pricing Inequality)} \qquad (4)$$

$$= \mathcal{R}_{SP} + \mathcal{R}_{SP} \log\left(S\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}\right) \qquad \text{(since } \mathcal{R}_{PP} = 1).$$

Rearranging this inequality yields

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq 1 + \log\left(S\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}\right). \qquad (5)$$

We next use properties of $W_{-1}(\cdot)$ to simplify Eq. (5). Exponentiating both sides yields,

$$e^{\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}} \leq eS\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \iff \frac{1}{eS} \leq \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}e^{-\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}} \iff \frac{-1}{eS} \geq -\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}e^{-\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}} \qquad (6)$$

Since $\frac{-1}{eS} \in [-1/e, 0)$ and the function $W_{-1}(\cdot)$ is non-increasing on this range, applying it to both sides of (6) and multiplying by -1 yields

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq -W_{-1}\left(\frac{-1}{eS}\right), \qquad (7)$$

which proves the bound when $c = 0$ and $\mu = 1$, since $M = 1$.

To prove tightness, it suffices to construct a nonnegative random variable $V \sim F$ with $\mu = 1$ and scale S, such that $\mathcal{R}_{SP}(F, 0) = \frac{-1}{W_{-1}\left(\frac{-1}{eS}\right)}$. For convenience, define $\alpha = \frac{-1}{W_{-1}\left(\frac{-1}{eS}\right)}$, and notice, by definition of $W_{-1}(\cdot)$,

$$-\frac{1}{Se} = -\frac{1}{\alpha}e^{-\frac{1}{\alpha}} \iff \frac{\alpha}{S} = e^{1-\frac{1}{\alpha}} \iff \log\left(\frac{\alpha}{S}\right) = 1 - \frac{1}{\alpha} \iff \frac{1}{\alpha} = 1 + \log\left(\frac{S}{\alpha}\right). \qquad (8)$$

Next consider a random variable with ccdf

$$\overline{F}_S(x) = \begin{cases} 1 & \text{if } x \in (0, \alpha] \\ \frac{\alpha}{x} & x \in (\alpha, S] \\ 0 & \text{otherwise.} \end{cases}$$

Observe that F_S has mean 1, since

$$\mu = \int_0^S \overline{F}_S(x)dx = \alpha + \alpha\log\left(\frac{S}{\alpha}\right) = \alpha\left(1 + \log\left(\frac{S}{\alpha}\right)\right) = 1, \qquad (9)$$

by Eq. (8). By inspection, F_S has scale S. Finally, for any $x \in (\alpha, S]$, $x\overline{F}_S(x) = \alpha$, and for any other x, $x\overline{F}_S(x) \leq \alpha$. Hence, $\mathcal{R}_{SP}(F, 0) = \alpha$, and, thus, the bound is tight for F_S.

For a general $c > 0$ and $\mu \neq 1$, use Lemma 1 to reduce to the case that $c = 0$, $\mu_c = 1$, $M_c = 1$, and $S_c = \frac{S+M-1}{M}$. Lemma 1 and Eq. (7) then imply that $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} = \frac{\mathcal{R}_{PP}(F_c,0)}{\mathcal{R}_{SP}(F_c,0)} \leq -W_{-1}\left(\frac{-1}{eS_c}\right)$. Replacing S_c proves the upper bound. Create a tight distribution by scaling F_{S_c} (defined above) by $\mu - c$ and shifting by c. □

Figure 3   Bound and tight distribution of Lemma 2.
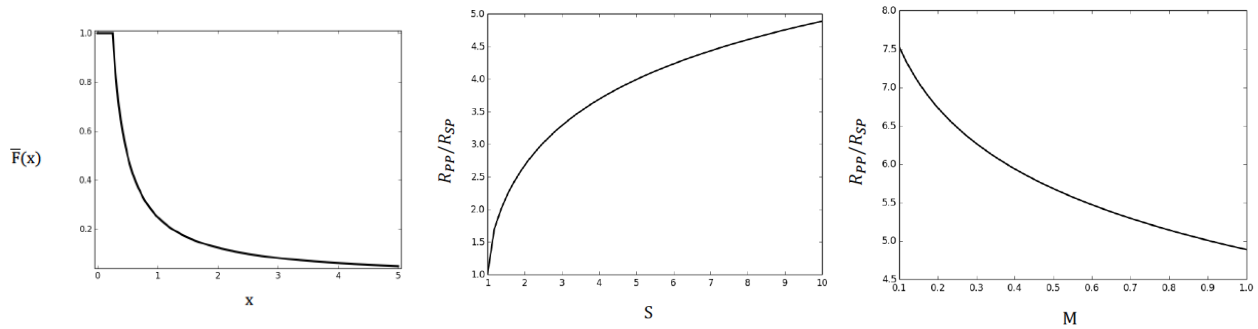
Note. The left panel shows the tight distribution of Lemma 2 when $M = 1$ and $S = 5$. The middle panel shows the bound in Lemma 2 when $M = 1$ and as S varies between 1 and 10. The right panel shows the bound in Lemma 2 when $S = 5$ and as M varies between 0.1 and 1.0.

The described tight distribution is a truncated Pareto distribution on $[\alpha, S]$ for some $\alpha \in [c, S]$, which satisfies $\overline{F}_S(x) \propto 1/x$ on its support (see left panel Fig. 3). In the auction literature, this distribution is sometimes called the "equal-revenue" distribution, since all prices in $[\alpha, S]$ yield the same single-pricing profit. Thus, one optimal pricing strategy for this distribution is to price at $p = \alpha$ and sell to *all* customers.

In the middle and right panels of Figure 3, we plot the bound of Lemma 2 versus M and S. Intuitively, as the scale increases, valuations become more dispersed and personalization offers greater potential value, as seen in the middle panel. On the other hand, increasing the margin with a fixed mean is equivalent to decreasing the cost per unit. As discussed above, an optimal single-pricing strategy has the same market share as idealized personalized pricing under the tight distribution. Thus, in the right panel, as margin increases, the profits of both idealized personalized pricing and single pricing increase at the same rate, and their relative ratio decreases. We stress that this behavior crucially depends on the properties of the tight distribution.

REMARK 1. Many of our subsequent proofs utilize techniques similar to the proof of Lemma 2. Consequently, we highlight some of its high-level features before proceeding. First, the proof is centered around an integral representation of a moment of V (in this case μ) in terms of the ccdf $\overline{F}$ (cf. Eq. (2)). The key step is to point-wise upper bound $\overline{F}(x)$ at each x. For $x \leq \mathcal{R}_{SP}$, the tightest bound possible is simply 1 (cf. Eq. (3)). For $x \geq \mathcal{R}_{SP}$, we use the Pricing Inequality (cf. Eq. (4)). The tight distribution is constructed by constructing a *valid* ccdf $\overline{F}$ that *simultaneously* makes each of these point-wise bounds tight. The remaining steps are simple algebraic manipulation. Thus, the three key elements

are an integral representation in terms of the ccdf, point-wise bounds on the ccdf, and identifying a single distribution which simultaneously matches all point-wise bounds. $\square$

3.1. Bounds Incorporating the Coefficient of Deviation

A drawback of Lemma 2 is that the bound becomes vacuous as the scale $S \to \infty$. The issue is that S, alone, cannot distinguish between markets where most customers have relatively similar valuations (which may be relatively low or high) and markets where customer valuations vary widely. We next provide more descriptive upper bounds on the value of personalized pricing by incorporating a measure of the market's heterogeneity, i.e., the typical dispersion in valuations. Specifically, we define the *coefficient of deviation* of F by

$$D := \frac{E[|V - \mu|]}{2\mu}.$$

By construction, $D \in [0, 1]$ since $E[|V - \mu|] \leq E[|V|] + \mu = 2\mu$ by the triangle inequality. Intuitively, D is the (rescaled) mean absolute deviation of V. Mean absolute deviation (or MAD) is a common measure of a random variable's dispersion, similar to standard deviation. Intuitively, when D is small, we expect most valuations to be close to μ, and, hence, the value of personalization to be small. By contrast, when D is large, we expect there to be larger dispersion in valuations, and, hence, the potential value of personalization to be much larger.

This intuition is not entirely correct as we shall see below. In fact, when D is *very* large and S is finite, there is a boundary effect; F is approximately a two-point distribution concentrated near c and μS, and single-pricing strategies are very effective. A single price can be used to capture the high valuation customers, while the low valuation customers are simply ignored since their potential profitability is near zero. Consequently, for very large D, the value of personalization is, in fact, low.

This qualitative description is formalized in Theorem 1 which bounds the value of personalization in terms of S, M, and D. The theorem partitions the space of markets into three distinct regimes depending on the magnitude of D and provides distinct bounds for each regime. Specifically, we define the three regimes by

    (L) Low Heterogeneity: $0 \leq D \leq \delta_L$

    (M) Medium Heterogeneity: $\delta_L \leq D \leq \delta_M$

    (H) High Heterogeneity: $\delta_M \leq D \leq \delta_H,$

where $\delta_L, \delta_M, \delta_H$ are constants that depend on M and S:

$$\delta_L := -\frac{M \log\left(\frac{S+M-1}{M}\right)}{W_{-1}\left(\frac{-1}{e^{\frac{S+M-1}{M}}}\right)}, \quad \delta_M := \frac{M \log\left(\frac{S+M-1}{M}\right)}{1 + \log\left(\frac{S+M-1}{M}\right)}, \quad \delta_H := \frac{M(S-1)}{S+M-1}.$$

The following lemma proves these regimes form a true partition:

LEMMA 3 (**Partitioning the Range of** D). *Given F with scale S and margin M, the coefficient of deviation of F satisfies $0 \leq D \leq \delta_H$. Moreover, $0 \leq \delta_L \leq \delta_M \leq \delta_H$.*

Equipped with Lemma 3, we can state Theorem 1, the main result of this section. We defer a proof until Section 3.2.

THEOREM 1 (**Bounding** $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ **using** D). *For any F with scale S, margin M, and coefficient of deviation D, we have the following:*

a) *If $0 \leq D \leq \delta_L$, then*

$$\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \leq \frac{-W_{-1}\left(\frac{\frac{D}{M}-1}{e}\right)}{1-\frac{D}{M}}. \qquad \text{(Low Heterogeneity)}$$

b) *If $\delta_L \leq D \leq \delta_M$, then*

$$\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \leq \frac{M \log\left(\frac{S+M-1}{M}\right)}{D}. \qquad \text{(Medium Heterogeneity)}$$

c) *If $\delta_M \leq D \leq \delta_H$, then*

$$\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \leq -W_{-1}\left(\frac{-1}{e(\frac{S+M-1}{M})(1-\frac{D}{M})}\right). \quad \text{(High Heterogeneity)}$$

Moreover, for any S, M, D there exists a valuation distribution F with scale S, margin M and coefficient of deviation D such that the corresponding bound is tight.

Theorem 1 gives a complete, closed-form upper bound on the value of personalized pricing for any distribution in terms of its scale, margin, and coefficient of deviation. The bound is defined piecewise, but is continuous (cf. Fig. 4). Note that the bound captures the intuition that the value of personalization increases as D increases for small to moderate D, but also captures the boundary behavior as D becomes very large. The maximal point in Fig. 4, at the transition between the low and medium regimes, corresponds exactly to the bound in Theorem 2. The bound is neither convex nor concave as a function of D.

Figure 4   Value of Personalized Pricing vs. the Coefficient of Deviation
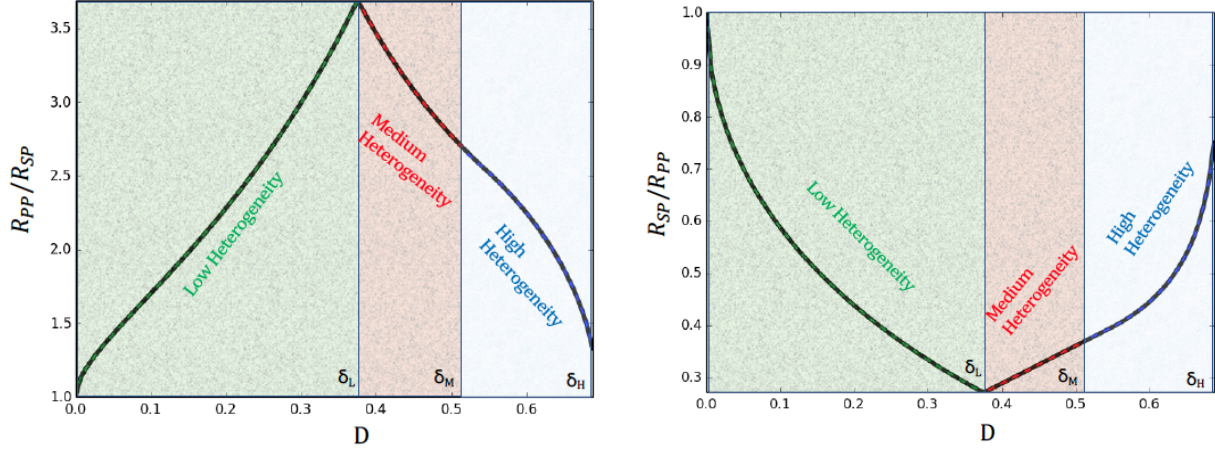
Note. The left panel plots the bound from Theorem 1 as a function of D with $S = 4$ and $M = .9$. The right panel plots the inverse of this bound, which we note is convex.

We also observe that our bound in Figure 4 can be significantly above or below e, the uniform bound proven for *monotone hazard rate* (MHR) distributions in Barlow et al. (1963) and Hartline et al. (2008). Further, although the value of personalized pricing can be infinite, our refined analysis characterizes precisely when classes of distributions lead to a low values of personalized pricing. Finally, if desired, the bound can easily be further upper-bounded using Eq. (1) to avoid the Lambert-W function, but at the cost of tightness.
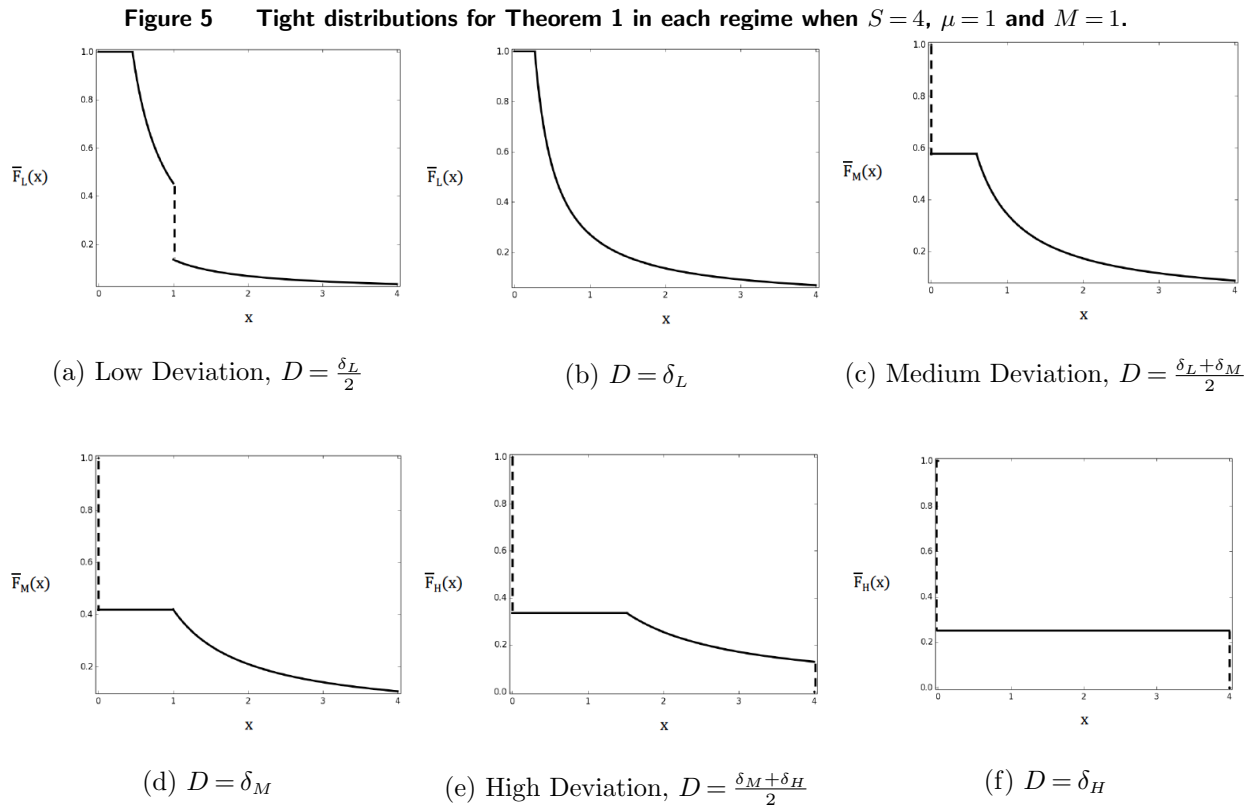
Single-Pricing Guarantee: An alternative interpretation of Theorem 1 is that the reciprocal of the bound is a tight guarantee on the performance of single-pricing relative to idealized personalized-pricing. In other words, the single-pricing strategy is guaranteed to earn at least the given percentage of the idealized personalized pricing profits. This perspective, i.e., interpreting single-pricing as an approximation to idealized personalized pricing, is common in the approximation algorithm literature.

We plot this guarantee, i.e., the reciprocal of the bound in Theorem 1, in the right panel of Fig. 4. Perhaps surprisingly, the reciprocal appears convex as a function of D. We prove this formally in Lemma 4 and leverage this observation later in Section 4.2.

LEMMA 4 **(Convexity of the Single-Pricing Guarantee)**. *For any S, M, and D, let $\alpha(S, M, D)$ denote the reciprocal of the bound on the value of personalized pricing in Theorem 1. Then $\alpha(S, M, D)$ is a convex function in D.*

Tight Distributions: Like Lemma 2, Theorem 1 is a tight bound. The distribution which achieves the bound depends on the regime but is not unique. See Fig. 5 for typical examples and Lemma 11 in the appendix for explicit formulas. In all three regimes, a worst-case distribution can be constructed from a mixture of a two-point distribution and truncated Pareto distributions; what differs between the regimes is the placement and sizes of these components. We show in the course of proving Theorem 1 that any price along the truncated Pareto section is an optimal price for the single-pricing strategy. These results generalize a folklore result from the auction literature that the Pareto distribution represents the worst-case valuation distribution (where S and D are unrestricted).

Figure 5   Tight distributions for Theorem 1 in each regime when $S = 4$, $\mu = 1$ and $M = 1$.

(a) Low Deviation, $D = \frac{\delta_L}{2}$ (b) $D = \delta_L$ (c) Medium Deviation, $D = \frac{\delta_L + \delta_M}{2}$

(d) $D = \delta_M$ (e) High Deviation, $D = \frac{\delta_M + \delta_H}{2}$ (f) $D = \delta_H$

Although the forms of the tight distributions differ by regime and are not unique, it is instructive to consider a class of them as a function of D and, in particular, study how they evolve as D increases with all other parameters are fixed. We focus on $c = 0$ and $\mu = 1$ as in Fig. 5.

- When $D = 0$, the bound of Theorem 1 is 1 and the unique tight distribution is a point-mass on μ. For $D > 0$ but small, this point mass stretches into two Pareto curves

above and below the mean. Every point along the Pareto curve below the mean is an optimal point at which to price, whereas no point along the Pareto curve above the mean is optimal (cf. Fig. 5a).

- As D grows towards δ_L, the Pareto curve above the mean rises to meet the curve below the mean. They join when $D = \delta_L$ as illustrated in Fig. 5b. Every point along the resulting single curve is an optimal price, and we recover the tight distribution of Lemma 2. This point is the transition between the low and medium regimes, and yields the greatest value of personalized pricing.

- As D grows past δ_L, boundary effects force mass to begin to pool at zero, and the single Pareto curve begins to shrink with the left most end point tending away from 0 and back towards μ. Again, every point along the Pareto curve is an optimal point at which to price (cf. Fig. 5c).

- When $D = \delta_M$, all mass below μ is contained in a point mass on 0. The Pareto curve extends from μ to μS, and pricing at any point along it is optimal (cf. Fig. 5d).

- As D grows past δ_M, boundary effects intensify and force mass to pool on both zero and μS. Past μ is an increasingly short, flat Pareto curve, along which every point is optimal (cf. Fig. 5e).

- Finally, the distribution converges to a two-point distribution on 0 and μv_{max} as illustrated in Fig. 5f.

Asymptotics Finally, from a theoretical point of view, one might seek to characterize the value of personalized pricing as D approaches its extreme values $D \to 0$ or $D \to \delta_H$. In particular, we will see in Section 4.2 that the first limit also provides insight into the performance of certain third-degree price discrimination tactics. These limits are below:

COROLLARY 1 (**Asymptotic Behavior**). *For any S, M, D, let $1/\alpha(D, M, S)$ denote the bound from Theorem 1. Then,*

a) *As $D \to 0$,*

$$\frac{1}{\alpha(S, M, D)} = 1 + \sqrt{2\frac{D}{M}} + O\left(\frac{D}{M}\right).$$

b) *As $D \to \delta_H$,*

$$\frac{1}{\alpha(S, M, D)} = 1 + \sqrt{2\frac{S + M - 1}{M}} \cdot \sqrt{\delta_H - \frac{D}{M}} + O\left(\delta_H - \frac{D}{M}\right).$$

In both cases, $\frac{1}{\alpha(S,M,D)}$ approaches its limit like the square root of the difference from the boundary.

3.2. Proof of Theorem 1

The proof of Theorem 1 treats each regime separately. Within each regime, we utilize the same basic technique as in Lemma 2. To that end, we first establish two integral representations of D in terms of $\overline{F}(x)$.

LEMMA 5 (**Integral Representations of** D). *For any F with scale S and margin M, the coefficient of deviation D satisfies*

$$D = \int_{M}^{S+M-1} \overline{F}(\mu x + c)dx = \int_{0}^{M} 1 - \overline{F}(\mu x + c)dx. \tag{10}$$

We now prove Theorem 1. For simplicity, we first consider the special case when $c = 0$ and $\mu = 1$. In this setting $\mathcal{R}_{PP} = \mu = 1$ and $M = 1$. We follow the general technique of Lemma 2. Starting with the second identity of Lemma 5,

$$D = \int_{0}^{1} 1 - \overline{F}(x)dx \geq \int_{0}^{\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}} 0 \ dx + \int_{\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}}^{1} 1 - \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{x} \ dx, \tag{11}$$

where we have pointwise upper bounded $\overline{F}(x)$ by 1 for $x \in [0, \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}]$ and used the Pricing Inequality for $x \in [\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}, 1]$. Evaluating the integrals yields,

$$D \geq \left(1 - \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}\right) + \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \log\left(\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}\right). \tag{12}$$

We next use properties of $W_{-1}(\cdot)$ to rewrite the inequality. For brevity, let $\alpha = \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}$. Then,

$$D \geq 1 - \alpha + \alpha \log(\alpha) \iff D - 1 \geq \alpha(\log(\alpha) - 1)$$
$$\iff \frac{D-1}{e} \geq e^{\log(\alpha)-1}(\log(\alpha) - 1) \qquad (\text{using } \alpha = e \cdot e^{\log(\alpha)-1}).$$

Since $D \in [0, 1]$, the right hand side is between $-1/e$ and 0. Applying $W_{-1}(\cdot)$ to both sides (and recalling this function is non-increasing) yields

$$W_{-1}\left(\frac{D-1}{e}\right) \leq \log(\alpha) - 1 \iff e \cdot e^{W_{-1}\left(\frac{D-1}{e}\right)} \leq \alpha \iff \frac{W_{-1}\left(\frac{D-1}{e}\right)}{D-1} \geq \frac{1}{\alpha} \tag{13}$$

$$\iff \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{W_{-1}\left(\frac{D-1}{e}\right)}{D-1}, \tag{14}$$

where the penultimate implication follows from the definition of $W_{-1}(\cdot)$, and the last line follows from the definition of α. We stress Eq. (14) holds for all D and coincides with the Low Heterogeneity bound when $c = 0$, $\mu = 1$.

Similarly, we can bound the ccdf in the first identity in Lemma 5 to yield an alternate bound. Specifically,

$$D = \int_1^S \overline{F}(x)dx \leq \int_1^S \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{dx}{x} = \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \log(S).$$

Rearranging yields,

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{\log(S)}{D}. \tag{15}$$

Again, we stress Eq. (15) holds for all D and coincides with the Medium Heterogeneity bound.

The High Heterogeneity bound can be derived similarly, using a different bounding of the ccdf which is tighter when D is large. We defer the details to the appendix and only state the result in Lemma 6 below.

LEMMA 6 (**High Heterogeneity Bound when $c=0$ and $\mu=1$**). *If $D > \delta_M$, then,*

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq -W_{-1}\left(\frac{-1}{eS(1-D)}\right). \tag{16}$$

To summarize, when $c = 0$, Eqs. (14) and (15) hold for all $0 \leq D \leq \delta_H$ and Eq. (16) holds for all $\delta_M \leq D \leq \delta_H$. These results are sufficient to prove that the bounds from the theorem are valid. For completeness, however, the next lemma further proves that in each regime, the bound for that regime is the strongest of the applicable bounds.

LEMMA 7 (**Strongest Bound by Regime**).

a) *The function*

$$D \mapsto \frac{-W_{-1}\left(-\frac{1-D}{e}\right)}{1-D} - \frac{\log(S)}{D},$$

    is negative for $D \in (0, \delta_L)$, is positive for $D \in (\delta_L, \delta_H]$, and has a unique root at $D = \delta_L$.

b) *The function*

$$D \mapsto \frac{\log(S)}{D} + W_{-1}\left(\frac{-1}{eS(1-D)}\right)$$

    has a unique root at $D = \delta_M$ and is non-negative for all $D \in [0, \delta_H]$.

A consequence of Lemma 7 is

- When $D \in [0, \delta_L]$, Eq. (14) dominates Eq. (15).
- When $D \in (\delta_L, \delta_M]$, Eq. (15) dominates Eq. (14).
- When $D \in (\delta_M, \delta_H]$, Eq. (16) dominates Eqs. (14) and (15).

This concludes the proof that the bounds are valid when $c = 0$ and $\mu = 1$.

For a general $c > 0$ and $\mu > 0$, we transform the problem to one in which $c = 0$ and $\mu = 1$ using Lemma 1 and apply the results from Eqs. (14) to (16) using the new S_c, M_c and D_c. Note, the coefficient of deviation D_c of F_c (as defined in Lemma 1) is related to D by $D_c = D/M$. Simplifying proves that the bounds are valid for general c and μ.

It only remains to establish that the bounds are tight. We use the same technique as in Lemma 2. Namely, in each regime, given S, M, D, and μ, we construct a ccdf that makes all pointwise bounds on the ccdf simultaneously. A difference from Lemma 2 is that the integral representations of D in the proof of Theorem 1 do not determine $\overline{F}$ over its whole domain $[0, S\mu]$; they only span $[0, \mu]$, or $[\mu, S]$ depending on the regime. This introduces some freedom in constructing the ccdf on the remaining segment and causes the tight distributions to be non-unique. Nonetheless, since these constructions follow the proof of Lemma 2 closely, we defer the details to Lemma 11 in the appendix for brevity. □

4. From Third-Degree to First-Degree Price Discrimination

As mentioned in the introduction, idealized personalized pricing is a strategy that hinges on two assumptions: 1) the firm can charge potentially distinct prices to every customer and 2) the firm is omniscient. In this section, we analyze how each of these assumptions contributes to the value of personalized pricing. In particular, we compute the value of personalized pricing over k-market segmentation and feature-based pricing. We then use both of these results to bound the value of personalized pricing over feature-based market segmentation. Our bounds yield insight into how these strategies "converge" to idealized personalized pricing as $k \to \infty$ or predictive accuracy increases. Said another way, they quantify both the value of the operational capability to charge a continuum of prices and the value of additional predictive accuracy.

4.1. Market Segmentation

In the section, we study the value of personalized pricing over *k-market segmentation*. From a practical point of view, k-market segmentation approximates settings in which the monopolist's ability to predict customer valuations is good, but her ability to charge different prices to different customers is limited. For instance, the monopolist may be constrained to only offer 10%, 20%, or 30%-off coupons (rather than a continuum of prices), but can identify the valuation of a customer accurately enough to place them in one of these buckets.

From a theoretical point of view, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{kP}}$ quantifies the benefit of an *operational* capabililty – the ability to offer a continuum of prices rather than a finite set. Intuitively, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{kP}} \to 1$ as $k \to \infty$. We will be most interested in the rate at which this convergence occurs. Intuitively, this rate characterizes how many segments one must use to guarantee a given percentage of idealized personalized pricing profits.

We first establish a simple lemma on the structure of the optimal segmentation.

LEMMA 8 **(Structure of Optimal Segmentation).** *There exists an optimal segmentation $s_0, \ldots, s_{k+1}$ and pricing $p_1, \ldots, p_k$ for $\mathcal{R}_{kP}$ such that $s_i = p_i$ for $i = 1, \ldots, k$.*

Proof. If $s_i < p_i$, then increasing s_i to p_i does not affect revenue in segment $[s_i, s_{i+1})$ and can only increase revenue in segment $[s_{i-1}, s_i)$. $\qquad\square$

Using this simple observation, we can explicitly compute the value of personalized pricing over k-market segmentation for uniform random variables.

LEMMA 9 **(k-Market Segmentation and Uniform Valuations).** *Let $V \sim F$ be a uniform random variable supported on $[0, t]$. Then, $\frac{\mathcal{R}_{PP(F,0)}}{\mathcal{R}_{kP(F,0)}} = 1 + \frac{1}{k}$.*

The first part of Theorem 2 below proves that in the worst-case, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{kP}}$ is also $1 + \tilde{O}\left(\frac{1}{k}\right)$ which matches the uniform case (up to logarithmic factors). The second part shows that with a mild assumption on F, i.e., its support is well-separated from c, we can additionally drop these logarithmic factors. Thus, in light of Lemma 9 the worst-case rate of convergence of Theorem 2 is essentially tight.

THEOREM 2 **(Idealized Personalized Pricing vs. k-Market Segmentation).**
*For any valuation distribution F with scale S and margin M, and for any $k \in \mathbb{N}$,*
  a) If F has coefficient of deviation D, then

$$\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{kP}(F,c)} \le 1 + \frac{1 + \log\left(\frac{S+M-1}{M}\left(1 + \frac{D}{M}(k+1)\right)\right)}{k} = 1 + \tilde{O}\left(\frac{1}{k}\right).$$

  b) If there is some $\delta > 0$ such that $\overline{F}(c + \delta(\mu - c)) = 1$, then

$$\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{kP}(F,c)} \le 1 + \frac{\log\left(\frac{S}{\delta}\right)}{k}.$$

The proof of Theorem 2 constructs a (suboptimal) segmentation strategy by geometrically partitioning the valuation space. In each segment, we then leverage our previous bounds

Figure 6 **Convergence of** $\mathcal{R}_{PP}/\mathcal{R}_{kP} - 1$.

(a) Beta   (b) Truncated Exponential   (c) Truncated Normal
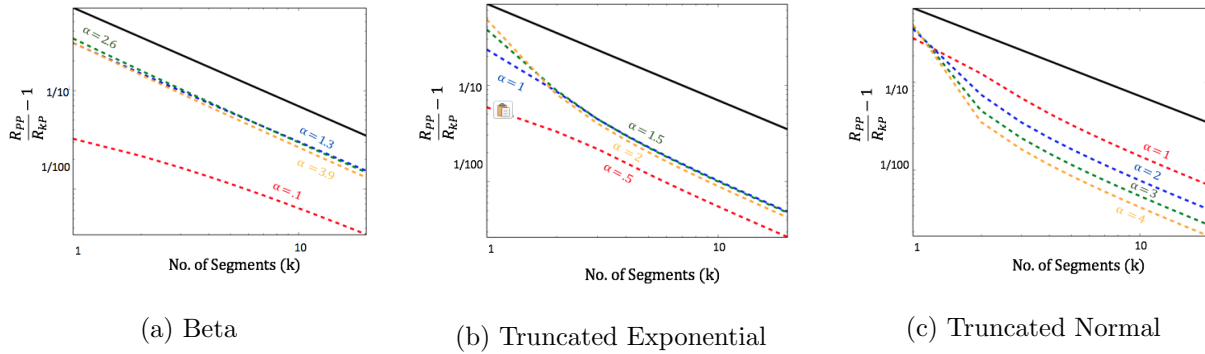
Note. Illustrates the decreasing benefit of idealized personalized pricing over optimally segmenting into k groups for particular distributions (dotted lines), and our distribution-agnostic bound (solid line). Note the log-scales.

on the value of personalized pricing over single-pricing to bound the profit. As in Lemma 8, the first segment requires special treatment.

Interestingly, the dependence $1 + O\left(\frac{1}{k}\right)$ appears typical for many distributions. In Figure 6, we plot the exact ratio $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{kP}} - 1$ for three different distributions, as well as our bound from Theorem 2 with $c = 0$ and $\delta = \frac{1}{\mu}$. Specifically, the first panel considers shifted Beta distributions, i.e., $V \sim \mathrm{Beta}(\alpha, 3) + 1$ for $\alpha = 0.1, 1.325, 2.55, 3.775, 5.0$. The second panel considers truncated exponential distributions, i.e., $V \sim \max(\min(\mathrm{Exp}(\alpha), 2), 1)$ for $\alpha = 0.5, 1.0, 1.5, 2.0$. Finally, the third panel considers truncated normal distributions, i.e., $V \sim \max(\min(\mathrm{Norm}(1, \alpha), 2), 1)$, for $\alpha = 1, 2, 3, 4$. Note the log-scales.

In each case, the dependence on k appears similar, and matches the dependence in Theorem 2. Intuitively, this behavior can be explained in the following way. As we segment into smaller pieces, any distribution with a continuous density appears locally uniform on each segment. Example 9 establishes that the convergence rate for a uniform matches Theorem 2 up to constant factors, suggesting that, at least for large k, the rate should also be approximately tight for many distributions.

Further, Theorem 2 can be used prescriptively to determine the number of segments necessary to guarantee a specific percentage of idealized personalized pricing profits. Namely, a monopolist seeking to guarantee $1 - \beta$-fraction of the idealized personalized pricing profits needs to use $k \geq (1/\beta - 1)\log(\frac{S}{\delta}) = O(1/\beta)$ by part (b) of Theorem 2. While this upper bound does not provide a tight analysis, Fig. 6 suggests the dependence on k is approximately tight for small β, i.e, to halve the relative gap to idealized personalized pricing, one needs twice as many segments.

Finally, we note that when F is given and discretely supported on N points, the optimal k-market segmentation strategy can be efficiently computed via a simple dynamic programming algorithm. We defer the details to Appendix B and shall return to this observation in Section 4.3.

4.2. Feature-Based Pricing

In this section, we study the value of personalized pricing over *feature-based pricing*. From a practical point of view, feature-based pricing approximates a host of third-degree price discrimination strategies in common use. For example, student discounts are a form of feature-based pricing where $\mathbf{X}$ is a binary indicating that the customer is a student. More generally, in online retailing settings, sellers often have access to rich contextual information for each customer from her cookies, such as demographics, browsing history, etc., that can be used to personalize the offered price via a custom coupon.

Clearly, if one can perfectly predict V from $\mathbf{X}$, feature-based pricing is equivalent to idealized personalized pricing. Typically, however, $\mathbf{X}$ is not rich enough to predict V perfectly, entailing some loss in profits. Thus, from a theoretical point of view, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{XP}}$ quantifies the benefits of additional *information*, i.e., the benefit of observing a richer set of features that enable perfect prediction. We will be most interested in the rate at which $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{XP}} \to 1$ as the information in $\mathbf{X}$ increases. Loosely speaking, this rate describes the predictive accuracy needed from a model to guarantee a given percentage of idealized personalized pricing profits.

Formally, we assume that the seller has trained a prediction model using historical data such that for any realization of $\mathbf{X}$, the seller knows the conditional distribution $V \mid \mathbf{X} \sim F_{V \mid \mathbf{X}}$. Let $\mu(\mathbf{X}) \equiv E[V \mid \mathbf{X}]$, and define the residual ϵ of the model by $V = \mu(\mathbf{X}) + \epsilon$. Note, by construction, $E[\epsilon \mid \mathbf{X}] = 0$ almost surely.

As an example, suppose the valuations follow the well-known logit model, i.e., a customer's valuation is a linear combination of that customer's features, the offered price and an idiosyncratic error with a logistic distribution. For this model, the conditional distribution is known precisely, and

$$\mathbb{P}(V - c \geq p \mid \mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 p + \boldsymbol{\beta}^\top \mathbf{X})}}.$$

Our assumption is that the seller has learned the coefficients $\beta_0, \boldsymbol{\beta}$.

A first, perhaps obvious, observation is that given $\mathbf{X}$, it is not optimal to price at $E[V \,|\, \mathbf{X}]$. To the contrary, one should price at the optimal price for the conditional distribution $F_{V|\mathbf{X}}$. This essentially proves Lemma 10.

LEMMA 10 **(Relating Feature-Based Pricing and Single-Pricing)**. *For any joint distribution $F_{\mathbf{X}V}$, we have $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c) = E[\mathcal{R}_{SP}(F_{V|\mathbf{X}}, c)]$.*

In Theorem 3, we use this observation in conjunction with our previous bounds on $\mathcal{R}_{SP}$ to bound the value of idealized personalized pricing over feature-based pricing under mild assumptions on the form of the valuation distribution.

THEOREM 3 **(Idealized Personalized Pricing vs. Feature-Based Pricing)**. *Suppose that $V = E[V \mid \mathbf{X}] + \epsilon$ where the residual ϵ satisfies $E\left[|\epsilon| \mid \mathbf{X}\right] = E\left[|\epsilon|\right]$. Suppose further that there exists δ with $0 < \delta < 1$ such that $\mu(\mathbf{X}) \geq \frac{c}{1-\delta}$ almost surely. Then,*

$$\frac{\mathcal{R}_{PP}(F, c)}{\mathcal{R}_{XP}(F_{\mathbf{X}V}, c)} \leq \frac{1}{\alpha\left(\bar{S}, M, \frac{E[|\epsilon|]}{2\mu}\right)},$$

where $\alpha(S, M, D)$ denotes the reciprocal of bound in Theorem 1 and $\bar{S} = \frac{M(S-1)+\delta}{\delta}$.

Proof. We let $F_{V|\mathbf{X}}$ denote the conditional distribution of $V \mid \mathbf{X}$, with corresponding mean $\mu(\mathbf{X})$, scale $S(\mathbf{X})$, margin $M(\mathbf{X})$, and coefficient of deviation $D(\mathbf{X})$. By assumption, $M(\mathbf{X}) = 1 - \frac{c}{\mu(\mathbf{X})} \geq \delta$ almost surely, and

$$D(\mathbf{X}) = \frac{E\left[|V - \mu(\mathbf{X})| \mid \mathbf{X}\right]}{2\mu(\mathbf{X})} = \frac{E\left[|\epsilon| \mid \mathbf{X}\right]}{2\mu(\mathbf{X})} = \frac{E\left[|\epsilon|\right]}{2\mu(\mathbf{X})},$$

where the last equality follows from the assumption on ϵ. From Lemma 10, we have that

$$
\begin{aligned}
\mathcal{R}_{XP} &= E\left[\mathcal{R}_{SP}(F_{V|\mathbf{X}}, c)\right] \\
&\geq E\left[\mathcal{R}_{PP}(F_{V|\mathbf{X}}, c) \cdot \alpha(S(\mathbf{X}),\ M(\mathbf{X}),\ D(\mathbf{X}))\right] && \text{(Theorem 1)} \\
&= E\left[(\mu(\mathbf{X}) - c) \cdot \alpha\left(\frac{S(\mathbf{X}) + M(\mathbf{X}) - 1}{M(\mathbf{X})},\ 1,\ \frac{D(\mathbf{X})}{M(\mathbf{X})}\right)\right] && \text{(Lemma 1)} \\
&= E\left[(\mu(\mathbf{X}) - c) \cdot \alpha\left(\frac{S(\mathbf{X}) + M(\mathbf{X}) - 1}{M(\mathbf{X})},\ 1,\ \frac{E[|\epsilon|]}{2(\mu(\mathbf{X}) - c)}\right)\right] \\
&\geq E\left[(\mu(\mathbf{X}) - c) \cdot \alpha\left(\frac{S + \delta - 1}{\delta},\ 1,\ \frac{E[|\epsilon|]}{2(\mu(\mathbf{X}) - c)}\right)\right]. && \text{(17)}
\end{aligned}
$$

The last inequality follows because $\alpha(S, M, D)$ is non-increasing in S, $S(\mathbf{X}) \leq S$, and the function $M \mapsto \frac{S+M-1}{M}$ is decreasing in M for $S > 1$. Since $M(\mathbf{X}) \geq \delta$, the bound follows.

By Lemma 4, $D \mapsto \alpha(\frac{S+\delta-1}{\delta}, 1, D)$ is convex. Hence, we claim that $y \mapsto y\alpha\left(\frac{S+\delta-1}{\delta}, 1, \frac{E[|\epsilon|]}{2y}\right)$ is also convex whenever $y \geq 0$. Indeed, the second derivative of this function is

$$\frac{(E[|\epsilon|])^2 \cdot \frac{\partial^2}{\partial D^2}\alpha\left(\frac{S+\delta-1}{\delta}, \ 1, \ \frac{E[|\epsilon|]}{2y}\right)}{4y^3},$$

which is non-negative since $\frac{\partial^2}{\partial D^2}\alpha(\frac{S+\delta-1}{\delta}, 1, D) \geq 0$ for all D. Letting $y \to \mu(\mathbf{X}) - c$, we recognize that the right-hand side of Eq. (17) is an expectation of a convex function of $\mu(\mathbf{X}) - c$, and hence, by Jensen's inequality,

$$\mathcal{R}_{XP} \geq (\mu - c) \cdot \alpha\left(\frac{S+\delta-1}{\delta}, \ 1, \ \frac{E[|\epsilon|]}{2(\mu-c)}\right) = (\mu - c)\alpha\left(\frac{M(S-1)+\delta}{\delta}, \ M, \ \frac{E[|\epsilon|]}{2\mu}\right),$$

where the equality follows from Lemma 1. This proves the result. $\square$

Interestingly, when the coefficient of deviation of V is in the 'low heterogeneity' regime, the bound in Theorem 3 has same form as the one in Theorem 1, except that the MAD of V is replaced by the MAD of the residual noise. (The scale does not appear in either bound.) This implies that the value of additional feature information in this regime can be directly measured by how much the residual MAD is reduced.

We consider the assumption in Theorem 3 that $E[|\epsilon| \mid \mathbf{X}] = E[|\epsilon|]$ to be quite mild. This assumption underlies many predictive models used in practice, including the logit model described above. Indeed, for the logit model, ϵ is a (centered) logistic random variable, *independent* of $\mathbf{X}$, so the above assumption holds directly. Similar remarks hold for other regression-based models with independent errors.

Similarly, we also consider assumption that $\mu(\mathbf{X}) \geq \frac{c}{1-\delta}$ almost surely to be mild. It holds, e.g, whenever $\overline{F}(c + \delta(\mu - c)) = 1$, i.e., the effective support of V is well-separated from c, mirroring Theorem 2. We stress that when D is in the low-heterogeneity regime, $\alpha(S, M, D)$ does not depend on S. Thus, even if δ is very small (causing $\overline{S}$ to be very large), the above bound is unaffected.

Intuitively, one can think of ϵ as the residual in the non-parametric regression $V = \mu(\mathbf{X}) + \epsilon$. If $\mathbf{X}$ is very informative for V, we expect ϵ, and hence, $\frac{E[|\epsilon|]}{2\mu}$ to be small. As we acquire more features, $E[|\epsilon|] \to 0$ and valuations can be predicted more accurately from these features. Simultaneously, the value of personalized pricing over feature-based pricing tends to 1. In sum, Theorem 3 provides a simple formula for benchmarking the quality of a predictive model for pricing and understanding the benefit of additional features.

We next provide numerical experiments to demonstrate that the bound in Theorem 3 is reasonably accurate in terms of magnitude and shape, as a function of the coefficient of deviation of the residual ϵ. Specifically, we generate valuations according the model

$$V = 10 + \sum_{i=1}^{10} X_i + \epsilon \tag{18}$$

where each $X_i \sim N(0,1)$, and ϵ is either a (centered) Logistic, Gumbel, or (shifted) Exponential distribution with standard deviation of 1 and mean zero.

In our experiment, we suppose the seller only knows the first k features, and thus $\mu(X_1, \ldots, X_k) = 10 + \sum_{i=1}^{k} X_i$. The corresponding error term, ϵ_k, has distribution of $X_{k+1} + \ldots + X_{10} + \epsilon$. In Fig. 7, we plot the actual value of personalized pricing as a function of the number of features available to the seller and the bound from Theorem 3. Observe the bound is quite illustrative in both magnitude and shape.

Figure 7   **Convergence of $\mathcal{R}_{PP}/\mathcal{R}_{XP}$.**
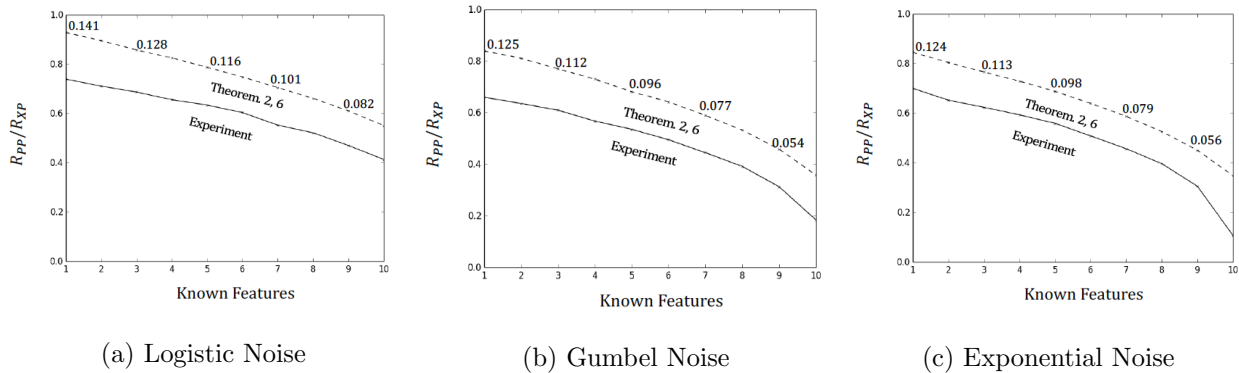
(a) Logistic Noise   (b) Gumbel Noise   (c) Exponential Noise

Note. Illustrates the decreasing benefit of idealized personalized pricing over feature-based pricing as the number of incorporated model features increases. The numbers above the curve denote the scaled MAD of the unexplained noise, $\frac{E[|\epsilon|]}{2\mu}$, for every other k.

Finally, Theorem 3 can be used prescriptively to determine the accuracy of a predictive model needed to guarantee a given percentage of idealized personalized-pricing profits. In particular, if a monopolist seeks $(1-\beta)$-fraction of the idealized personalized pricing profits, it suffices to construct a predictive model with enough features so that $\alpha(\bar{S}, M, \frac{E[|\epsilon|]}{2\mu}) \geq 1 - \beta$. By Corollary 1, for small β, this amounts to $\frac{E[|\epsilon|]}{2\mu} < \frac{\beta^2}{2(1-\beta)^2} = O(\beta^2)$. Although this analysis is based on an upper-bound, Fig. 7 suggests the general dependence on $E[|\epsilon|]$ is correct, i.e., to halve the gap (between $\mathcal{R}_{PP}$ and $\mathcal{R}_{XP}$), one needs 4 times the predictive accuracy for small β.

4.3. Feature-Based Market Segmentation

In this section, we study the value of personalized pricing over feature-based market segmentation. Feature-based market segmentation closely resembles real-world data-driven personalization strategies where sellers are both constrained by the number of the prices they can offer, and must predict customer valuations from data. In this way, feature-based market segmentation synthesizes the two previous models of personalized pricing discussed in this section. Formally, feature-based market segmentation is equivalent to feature-based pricing with the restriction that the seller can offer only k distinct prices. As in the previous section, we assume the seller has learned the conditional distribution $F_{V|\mathbf{X}}$ from data, and we let $\mu(\mathbf{X}) = E[V \mid \mathbf{X}]$ and define the residual ϵ by $V = \mu(\mathbf{X}) + \epsilon$.

We bound the value of personalized pricing over feature-based market segmentation by separately considering the loss from limited price flexibility and the loss from the prediction error in the valuation model. We measure *loss* as the difference in profit between a personalized pricing strategy ((kP), (XP), or (kXP)) and idealized personalized pricing. The following theorem states that one can bound the loss of feature-based market segmentation by the loss of two more powerful strategies: feature-based pricing, and k-market segmentation on a noiseless market.

THEOREM 4 **(Idealized Personalized Pricing vs. Feature-Based Market Segmentation)**. *As above, let $V = \mu(\mathbf{X}) + \epsilon$, and suppose $\mathbf{X}$ and ϵ are independent. Then*

$$\underbrace{\mathcal{R}_{PP}(F,c) - \mathcal{R}_{kXP}(F_{\mathbf{X}V},c)}_{\text{Loss of kXP}} \leq \underbrace{\mathcal{R}_{PP}(F,c) - \mathcal{R}_{kP}(F_{\mu(\mathbf{X})},c)}_{\text{Loss of kP on noiseless market}} + \underbrace{\mathcal{R}_{PP}(F,c) - \mathcal{R}_{XP}(F_{\mathbf{X}V},c)}_{\text{Loss of XP}}.$$

If $\mathcal{R}_{kP}(F_{\mu(\mathbf{X})}) + \mathcal{R}_{XP}(F_{\mathbf{X}V},c) > \mathcal{R}_{PP}(F,c)$, this implies

$$\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{kXP}(F_{\mathbf{X}V},c)} \leq \frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{kP}(F_{\mu(\mathbf{X})},c) + \mathcal{R}_{XP}(F_{\mathbf{X}V},c) - \mathcal{R}_{PP}(F,c)}.$$

Proof. We shall prove that

$$\mathcal{R}_{kP}(F_{\mu(\mathbf{X})},c) - \mathcal{R}_{kXP}(F_{\mathbf{X}V},c) \leq \mathcal{R}_{PP}(F_{\mu(\mathbf{X})},c) - \mathcal{R}_{XP}(F_{\mathbf{X}V},c). \tag{19}$$

Note that $\mathcal{R}_{PP}(F_{\mu(\mathbf{X})},c) = \mathcal{R}_{PP}(F,c)$ so that Eq. (19) implies the two inequalities above by rearranging. In fact, we will prove that the portion of profits earned by each strategy for a fixed context $\mathbf{x} \in \mathcal{X}$ satisfies Eq. (19).

To that end, let $\{s_i\}_{i=0}^k \in \mathbb{R}^{k+1}$ be an optimal k-market segmentation for $\mathcal{R}_{kP}(F_{\mu(\mathbf{X})})$ of the form described in Lemma 8 where $s_0 = c$ and $s_{k+1} = \infty$. Partition the feature space, $\mathcal{X}_i := \{\mathbf{x} \in \mathcal{X} \,|\, \mu(\mathbf{x}) \in [s_i, s_{i+1})\}$ for $i \in 0, 1, \ldots, k$. Let $p(\cdot) : \mathcal{X} \to \mathbb{R}$ denote the optimal pricing function for $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c)$. Finally, let $\mathbf{x} \in \mathcal{X}_i$ be some fixed realization of $\mathbf{X}$ for some $i = 0, \ldots, k$.

When $\mathbf{X} = \mathbf{x}$, the XP strategy earns

$$\mathbf{x}\text{-Contribution to XP} \;=\; (p(\mathbf{x}) - c)\mathbb{P}\left\{\mu(\mathbf{x}) + \epsilon \geq p(\mathbf{x})\right\},$$

where we have used independence to drop the conditioning on $\mathbf{X} = \mathbf{x}$.

Similarly, when $\mathbf{X} = \mathbf{x}$, the kP strategy earns

$$\mathbf{x}\text{-Contribution to kP} \;=\; (s_i - c)$$

on $F_{\mu(\mathbf{X})}$, since by Lemma 8 all customers in the segment buy at price s_i.

Next, we lower bound the $\mathbf{x}$-Contribution to kXP by considering a feasible feature-based segmentation strategy. Let $\mathbf{x}_i \in \arg\min_{\mathbf{x} \in \mathcal{X}_i} \mu(\mathbf{X})$, and consider offering price $p(\mathbf{x}_i)$ to every customer in $\mathcal{X}_i$. When $\mathbf{X} = \mathbf{x}$, this strategy earns at most

$$(p(\mathbf{x}_i) - c)\mathbb{P}\left\{\mu(\mathbf{x}) + \epsilon \geq p(\mathbf{x}_i)\right\} \;\geq\; (p(\mathbf{x}_i) - c)\mathbb{P}\left\{\mu(\mathbf{x}_i) + \epsilon \geq p(\mathbf{x}_i)\right\},$$

since $\mu(\mathbf{x}) \geq \mu(\mathbf{x}_i)$ for all $\mathbf{x} \in \mathcal{X}_i$. We again use independence to drop conditioning on $\mathbf{X} = \mathbf{x}$. On the other hand, by Lemma 10, $p(\mathbf{x}_i)$ is the optimal single price when $\mathbf{X} = \mathbf{x}_i$, so that pricing at $p(\mathbf{x}) + \mu(\mathbf{x}_i) - \mu(\mathbf{x})$ must earn less profit, i.e.,

$$(p(\mathbf{x}_i) - c)\,\mathbb{P}\left\{\mu(\mathbf{x}_i) + \epsilon \geq p(\mathbf{x}_i)\right\} \;\geq\; (p(\mathbf{x}) + \mu(\mathbf{x}_i) - \mu(\mathbf{x}) - c)\mathbb{P}\left\{\mu(\mathbf{x}_i) + \epsilon \geq p(\mathbf{x}) + \mu(\mathbf{x}_i) - \mu(\mathbf{x})\right\}$$

$$= (p(\mathbf{x}) + \mu(\mathbf{x}_i) - \mu(\mathbf{x}) - c)\mathbb{P}\left\{\mu(\mathbf{x}) + \epsilon \geq p(\mathbf{x})\right\},$$

where now we use independence to drop the conditioning $\mathbf{X} = \mathbf{x}_i$ throughout. Combining these two inequalities shows

$$\mathbf{x}\text{-Contribution to kXP} \;\geq\; (p(\mathbf{x}) + \mu(\mathbf{x}_i) - \mu(\mathbf{x}) - c)\mathbb{P}\left\{\mu(\mathbf{x}) + \epsilon \geq p(\mathbf{x})\right\}.$$

Now combine these contributions as in Eq. (19),

$$\mathbf{x}\text{-Contribution to XP} + \mathbf{x}\text{-Contribution to kP} - \mathbf{x}\text{-Contribution to kXP}$$

$$\leq (s_i - c) + (\mu(\mathbf{x}) - \mu(\mathbf{x}_i))\mathbb{P}\left\{\mu(\mathbf{x}) + \epsilon \geq p(\mathbf{x})\right\}$$

$$\leq (s_i - c) + (\mu(\mathbf{x}) - \mu(\mathbf{x}_i))$$

$$\leq \mu(\mathbf{x}) - c,$$

where the last line follows because $\mu(\mathbf{x}_i) \in [s_i, s_{i+1})$. Recall $\mathbf{x}$ was chosen arbitrarily. Averaging this inequality over realizations of $\mathbf{X}$ yields Eq. (19) to complete the proof. □

Unlike Theorem 3, Theorem 4 requires that ϵ is independent of $\mathbf{X}$. This is a stronger assumption, but, as discussed previously, is satisfied by many common valuation models, such as the logit model.

We emphasize that Theorem 4 is tight when the firm can offer an infinite number of prices, or when the prediction error of the valuation model vanishes. To see this note that for any market F, when a firm has infinite price flexibility $\lim_{k\to\infty} \mathcal{R}_{kXP}(F_{\mathbf{X}V}, c) = \mathcal{R}_{XP}(F_{\mathbf{X}V}, c)$ and $\lim_{k\to\infty} \mathcal{R}_{kP}(F_{\mu(\mathbf{X})}) = \mathcal{R}_{PP}(F_{\mu(\mathbf{x})}, c) = \mathcal{R}_{PP}(F, c)$. When the firm's valuation model has perfect prediction accuracy, i.e. $\epsilon \sim 0$, $\mathcal{R}_{kXP}(F_{\mathbf{X}V}, c) = \mathcal{R}_{kP}(F_{\mu(\mathbf{X})})$ and $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c) = \mathcal{R}_{PP}(F, c)$. In general Theorem 4 forms a theoretical framework through which a firm can analyze the performance of personalized pricing strategies. In particular, the bound neatly decouples the profit loss from limited price flexibility, through analyzing (kP) strategies on the noiseless market $F_{\mu(\mathbf{X})}$, and the prediction error through analyzing the (XP) strategy.

Operationally, a monopolist may use Theorem 4 to guide the tuning of personalized pricing strategies. Consider a firm that learns a valuation model $\mu(\cdot)$ from customer data. Suppose $\mu(\mathbf{X})$ is supported on n points (one can imagine that $\mu(\cdot)$ is trained on n sampled data points), then the optimal (kP) pricing strategy on the noiseless market $F_{\mu(\mathbf{X})}$ can be computed in time $O(kn^2)$ by dynamic programming, using the algorithm in Appendix B. Thus $\mathcal{R}_{kP}(F_{\mu(\mathbf{X})})$ can be computed exactly by a firm after training the model. Since valuations are not observable, the distribution of ϵ is not obtainable. Instead, we can use the basic statistics of ϵ (typically given as output after training the prediction model) and then bound $\mathcal{R}_{XP}$ using Theorem 3 and Theorem 1. With an exact computation of $\mathcal{R}_{kP}(F_{\mu(\mathbf{X})})$ and a bound on $\mathcal{R}_{XP}$, the firm can apply Theorem 4 to study the performance of the feature-based market segmentation strategy. This allows the seller to reason about where exactly the inefficiency is arising, and decide whether to increase the number of prices/segments, or invest in additional data gathering to reduce the prediction error, or both.

Finally, we note that the proof Theorem 4 is constructive, and implies a heuristic for setting feature-based market segmentation strategies: compute the optimal k-market segmentation $\{s_i\}_{i=1}^{k+1}$ for the noiseless market $F_{\mu(\mathbf{X})}$, use it to generate the segments $\mathcal{X}_i$, then

perform price experimentation to learn the prices that maximize $p_i \Pr(s_i + \epsilon \geq p_i)$, and offer that price on each segment. While both the partition into segments $\{\mathcal{X}_i\}_{i=1}^k$, and the prices offered on each segment $\{p_i\}_{i=1}^k$ may be sub-optimal, such a strategy is guaranteed to earn more than $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c) + \mathcal{R}_{kP}(F_{\mu(\mathbf{X})}, c) - \mathcal{R}_{PP}(F, c)$ by Theorem 4.

5. Conclusions

Increasingly rich consumer profiles and choice models enable retailers to personalize to consumers at finer and finer levels. However, building such tools comes at an investment cost in the form of technology, data scientists, marketing, etc. Motivated by this trade-off, we provide a framework to quantify the benefits of personalized pricing in terms of the features of the underlying market. In particular, we exactly characterized the value of personalized pricing over posting a single price for all customers in terms of the scale, coefficient of deviation, and margin of the valuation distribution in closed-form.

Using our closed-form bound, we are also able to bound the value of personalized pricing over certain third-degree price discrimination tactics that more closely mirror current practice. Specifically, we first provide an order optimal bound on the value of personalized pricing over k-market segmentation. Intuitively, this bound quantifies the benefit of the operational ability to set a continuum of prices rather than k fixed prices. We then provide a bound on the value of personalized pricing over feature-based pricing strategies. Intuitively, this second bound quantifies the benefit of obtaining additional market information or improving one's predictive model. Finally we leveraged these two bounds to study the performance of feature based market segmentation, a strategy which closely models popular, data-driven personalized pricing strategies.

Overall, we believe that our results provide a rigorous foundation for analyzing pricing strategies in the context of personalization. Our results can be used both by researchers attempting to design algorithms for personalized pricing, as well as by managers seeking to implement or improve their pricing strategies.

References

Arora, Neeraj, Xavier Dreze, Anindya Ghose, James D Hess, Raghuram Iyengar, Bing Jing, Yogesh Joshi, V Kumar, Nicholas Lurie, Scott Neslin, S Sajeesh, Syam Niladri Su, Meng, Jacquelyn Thomas, John Zhang. 2008. Putting one-to-one marketing to work: Personalization, customization, and choice. *Marketing Letters* **19**(3-4) 305.

Aydin, Goker, Serhan Ziya. 2009. Personalized dynamic pricing of limited inventories. *Operations Research* **57**(6) 1523–1531.

Azar, Pablo, Constantinos Daskalakis, Silvio Micali, S Matthew Weinberg. 2013. Optimal and efficient parametric auctions. *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 596–604.

Ban, Gah-Yi, N Bora Keskin. 2017. Personalized dynamic pricing with machine learning. *Available at SSRN: https://ssrn.com/abstract=2972985* .

Barlow, Richard E, Albert W Marshall, Frank Proschan. 1963. Properties of probability distributions with monotone hazard rate. *The Annals of Mathematical Statistics* 375–389.

Bergemann, Dirk, Benjamin Brooks, Stephen Morris. 2015. The limits of price discrimination. *The American Economic Review* **105**(3) 921–957.

Bergemann, Dirk, Karl Schlag. 2011. Robust monopoly pricing. *Journal of Economic Theory* **146**(6) 2527–2543.

Bernstein, Fernando, A Gürhan Kök, Lei Xie. 2015. Dynamic assortment customization with limited inventories. *Manufacturing & Service Operations Management* **17**(4) 538–553.

Besbes, Omar, Ilan Lobel. 2015. Intertemporal price discrimination: Structure and computation of optimal policies. *Management Science* **61**(1) 92–110.

Besbes, Omar, Robert Phillips, Assaf Zeevi. 2010. Testing the validity of a demand model: An operations perspective. *Manufacturing & Service Operations Management* **12**(1) 162–183.

Besbes, Omar, Assaf Zeevi. 2015. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science* **61**(4) 723–739.

Caro, Felipe, Jérémie Gallien. 2012. Clearance pricing optimization for a fast-fashion retailer. *Operations Research* **60**(6) 1404–1422.

Celis, L Elisa, Gregory Lewis, Markus Mobius, Hamid Nazerzadeh. 2014. Buy-it-now or take-a-chance: Price discrimination through randomized auctions. *Management Science* **60**(12) 2927–2948.

Chatzigeorgiou, I. 2013. Bounds on the lambert function and their application to the outage analysis of user cooperation. *IEEE Communications Letters*. IEEE, 1505–1508.

Chen, Hongqiao, Ming Hu, Georgia Perakis. 2017. Distribution-free pricing. *Available at SSRN: https://ssrn.com/abstract=3090002* .

Chen, Xi, Zachary Owen, Clark Pixton, David Simchi-Levi. 2015. A statistical learning approach to personalization in revenue management. *Available at SSRN: https://ssrn.com/abstract=2579462* .

Chen, Yuxin, Sridhar Moorthy, Z John Zhang. 2005. Research note-price discrimination after the purchase: Rebates as state-dependent discounts. *Management Science* **51**(7) 1131–1140.

Choudhary, Vidyanand, Anindya Ghose, Tridas Mukhopadhyay, Uday Rajan. 2005. Personalized pricing and quality differentiation. *Management Science* **51**(7) 1120–1130.

Clifford, Stephanie. 2012. Shopper alert: Price may drop for you alone. *The New York Times* URL `http://www.nytimes.com/2012/08/10/business/supermarkets-try-customizing-prices-for-shoppers.html`.

Cohen, Maxime C, Ngai-Hang Zachary Leung, Kiran Panchamgam, Georgia Perakis, Anthony Smith. 2017. The impact of linear optimization on promotion planning. *Operations Research* **65**(2) 446–468.

Cohen, Maxime C., Ilan Lobel, Renato Paes Leme. 2016. Feature-based dynamic pricing. *Available at SSRN: httpps://ssrn.com/abstract=2737045* .

Cohen, Maxime C., Georgia Perakis, Robert S. Pindyck. 2015. Pricing with limited knowledge of demand. *Available at SSRN: https://ssrn.com/abstract=2673810* .

Corless, R.M., G.H. Gonnet, D.E.G. Hare, D.J. Jeffery, D.E. Knuth. 1993. On the lambert w function. *Advances in Computational Mathematics* .

Cowan, Simon. 2016. Welfare-increasing third-degree price discrimination. *The RAND Journal of Economics* **47**(2) 326–340.

D'Innocenzio, Anne. 2017. Neiman marcus focuses on exclusives, personalized offers; ends merger talks. *USA Today* URL `https://www.usatoday.com/story/money/business/2017/06/13/neiman-marcus-focuses-exclusives-personalized-offers/102814962/`.

Elmachtoub, Adam N., Michael L. Hamilton. 2017. The power of opaque products in pricing. *Available at SSRN: https://ssrn.com/abstract=3025944* .

Gronlund, Allan, Kasper Larsen, Alexander Mathiasen, Jesper Nielsen, Stefan Schneider, Mingzhou Song. 2017. Fast exact k-means, k-medians and bregman divergence clustering in 1d. *ArXiv e-prints: arXiv:1701.07204* .

Hartline, Jason, Vahab Mirrokni, Mukund Sundararajan. 2008. Optimal marketing strategies over social networks. *Proceedings of the 17th International Conference on World Wide Web*. ACM, 189–198.

Javanmard, Adel, Hamid Nazerzadeh. 2016. Dynamic pricing in high-dimensions. *Available at SSRN: https://ssrn.com/abstract=2855843* .

Jerath, Kinshuk, Serguei Netessine, Senthil K Veeraraghavan. 2010. Revenue management with strategic customers: Last-minute selling and opaque selling. *Management Science* **56**(3) 430–448.

Liu, Yunchuan, Z John Zhang. 2006. The benefits of personalized pricing in a channel. *Marketing Science* **25**(1) 97–105.

Moorthy, K Sridhar. 1984. Market segmentation, self-selection, and product line design. *Marketing Science* **3**(4) 288–307.

Narasimhan, Chakravarthi. 1984. A price discrimination theory of coupons. *Marketing Science* **3**(2) 128–147.

Obama, Administration. 2016. Big data and differential pricing. *White House Report* URL `https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf/`.

Özer, Özalp, Yanchong Zheng. 2015. Markdown or everyday low price? the role of behavioral motives. *Management Science* **62**(2) 326–346.

Phillips, Robert. 2013. Optimizing prices for consumer credit. *Journal of Revenue and Pricing Management* **12**(4) 360–377.

Qiang, Sheng, Mohsen Bayati. 2016. Dynamic pricing with demand covariates. *Available at SSRN: https://ssrn.com/abstract=2765257* .

Robinson, J. 1934. The economics of imperfect competition. *Journal of Political Economy* **42**(2) 249–259.

Schmalensee, Richard. 1981. Output and welfare implications of monopolistic third-degree price discrimination. *The American Economic Review* **71**(1) 242–247.

Shih, Jun-ji, Chao-cheng Mai, Jung-chao Liu. 1988. A general analysis of the output effect under third-degree price discrimination. *The Economic Journal* **98**(389) 149–158.

Su, Xuanming. 2007. Intertemporal pricing with strategic customer behavior. *Management Science* **53**(5) 726–741.

Tamuz, Omer. 2013. A lower bound on seller revenue in single buyer monopoly auctions. *Operations Research Letters* **41**(5) 474–476.

Tuttle, Brad. 2013. Flight prices to get personal? airfares could vary depending on who is traveling. *Time* URL `http://business.time.com/2013/03/05/flight-prices-to-get-personal-airfares-could-vary-depending-on-who-is-traveling/`.

Varian, Hal R. 1985. Price discrimination and social welfare. *The American Economic Review* **75**(4) 870–875.

Xu, Zibin, Anthony J Dukes. 2016. Price discrimination in a market with uninformed consumer preferences. *Available at SSRN: https://ssrn.com/abstract=2777081* .

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

Appendix A: Omitted Proofs

A.1. Omitted Proofs from Section 3

Proof of Lemma 1. First note the profit from personalized pricing under valuation distribution F is $\mathcal{R}_{PP}(F,c) = E[V] - c = \mu - c$ and under F_c is $\mathcal{R}_{PP}(F_c,0) = E[\frac{1}{\mu-c}(V-c)] - 0 = 1$. Hence, it suffices to show that $\mathcal{R}_{SP}(F,c) = (\mu-c)\mathcal{R}_{SP}(F_c,0)$ to prove the first statement. Observe that

$$
\begin{aligned}
\mathcal{R}_{SP}(F,c) &= \max_p (p-c)\Pr(V \ge p) \\
&= \max_p (p-c)\Pr\left(\frac{V-c}{\mu-c} \ge \frac{p-c}{\mu-c}\right) \\
&= \max_q (\mu-c)q\Pr\left(\frac{V-c}{\mu-c} \ge q\right) \qquad \text{(Making the substitution } \frac{p-c}{\mu-c} \to q) \\
&= (\mu-c)\mathcal{R}_{SP}(F_c,0).
\end{aligned}
$$

For the last statement of the theorem, note that $\mu_c = E[\frac{1}{\mu-c}(V-c)] = 1$, $M_c = 1 - 0/\mu_c = 1$, and

$$
S_c = \frac{\inf\{k \mid F_c(k) = 1\}}{\mu_c} = \frac{\frac{1}{\mu-c}(\inf\{k \mid F(k) = 1\} - c)}{1} = \frac{\mu}{\mu-c}\left(\frac{\inf\{k \mid F(k) = 1\}}{\mu} - \frac{c}{\mu}\right) = \frac{S-1+M}{M}.
$$

This completes the proof. $\qquad \square$

A.2. Omitted Proofs from Section 3.1.

We first provide proofs for the lemmas necessary to complete the proof of Theorem 1. We then prove the remaining corollaries and extensions.

Proof of Lemma 3. Consider the case when $c = 0$ and $\mu = 1$, which implies that $M = 1$. We first prove that $D \le \delta_H$ and that there exists an F whose coefficient of deviation is exactly δ_H. To this end, consider an arbitrary random variable V, and define the new random variable $\overline{V}$ with two-point support

$$
\overline{V} = \begin{cases} E[V|V \le 1] & \text{with probability } \Pr(V \le 1) \\ E[V|V > 1] & \text{with probability } \Pr(V > 1). \end{cases}
$$

By construction, $E[\overline{V}] = E[V] = 1$. Furthermore,

$$
\begin{aligned}
E\Big[|V-1|\Big] &= E\Big[|V-1| \mid V \le 1\Big]\Pr(V \le 1) + E\Big[|V-1| \mid V > 1\Big]\Pr(V > 1) \\
&= E\Big[1-V \mid V \le 1\Big]\Pr(V \le 1) + E\Big[V-1 \mid V > 1\Big]\Pr(V > 1) \\
&= \Big(1 - E\Big[V \mid V \le 1\Big]\Big)\Pr(V \le 1) + \Big(E\Big[V \mid V > 1\Big] - 1\Big)\Pr(V > 1) \\
&= E\Big[|\overline{V}-1|\Big],
\end{aligned}
$$

i.e., both V and $\overline{V}$ have the same coefficient of deviation. Thus, to find a distribution with maximal coefficient of deviation, it suffices to consider two-point distributions.

We compute such a distribution explicitly via the following optimization problem:

$$
\begin{aligned}
\frac{1}{2}\max_{x,y,q} \quad & q(1-x) + (1-q)(y-1) \\
\text{s.t.} \quad & qx + (1-q)y = 1 \\
& 0 \le x \le 1 \le y \le S, \quad 0 \le q \le 1,
\end{aligned}
$$

where the objective is the coefficient of deviation of a distribution with mass q at $x < 1$ and mass $1 - q$ at $y > 1$. The constraint ensures that the mean is 1. In particular, this constraint implies $q = \frac{y-1}{y-x}$ for any feasible solution, whereby the objective simplifies to $\frac{(1-x)(2y-1)}{y-x}$. This function is decreasing in x, whereby the optimal solution is $x^* = 0$, $y^* = S$ and $q^* = \frac{S-1}{S}$ with optimal value $\frac{S-1}{S}$. Note $\frac{S-1}{S} = \delta_H$ since $M = 1$.

Next we show $0 \le \delta_L \le \delta_M \le \delta_H$. Notice that $\delta_L = \frac{\log(S)}{-W_{-1}(-\frac{1}{eS})}$ is the ratio of two positive terms. Thus, it is positive. To show $\delta_L \le \delta_M$, note that, since $S \ge 1$,

$$1 + \log(S) \ge 1 = \frac{e^{1+\log(S)}}{eS},$$

which, after rearranging, implies

$$-(1 + \log(S)) e^{-(1+\log(S))} \le \frac{-1}{eS}.$$

Applying $W_{-1}(\cdot)$ to both sides and noting this function is decreasing shows

$$-(1 + \log(S)) \ge W_{-1}\left(\frac{-1}{eS}\right),$$

which implies

$$\delta_L = \frac{\log(S)}{-W_{-1}\left(\frac{-1}{eS}\right)} \le \frac{\log(S)}{1 + \log(S)} = \delta_M,$$

as was to be shown.

To show $\delta_M \le \delta_H$, observe that since $S \ge 1$, $0 \le \log(S) \le S - 1$, which implies that

$$\delta_M = \frac{\log(S)}{1 + \log(S)} \le \frac{S-1}{1 + (S-1)} = \delta_H,$$

since $x \mapsto \frac{x}{1+x}$ is an increasing function for $x \ge 0$. This completes the proof in the case $c = 0$ and $\mu = 1$.

For general $c > 0$ and $\mu > 0$, first apply Lemma 1 to obtain an instance with zero cost and unit mean with corresponding parameters D_c, S_c, and M_c. From the previous arguments, we have that $0 \le D_c \le \frac{S_c-1}{S_c}$ and $0 \le \frac{\log(S_c)}{W_{-1}\left(\frac{-1}{eS_c}\right)} \le \frac{\log(S_c)}{1+\log(S_c)} \le \frac{S_c-1}{S_c}$. Transform back to the original parameters to prove the lemma, noting that $D_c = \frac{D}{M}$ and $S_c = \frac{S+M-1}{M}$. $\qquad\square$

Proof of Lemma 5.   Let $V \sim F$ and note,

$$0 = E[V - \mu] = E[(V - \mu)^+] - E[(\mu - V)^+] \implies E[(V - \mu)^+] = E[(\mu - V)^+].$$

Moreover, $E[|V - \mu|] = E[(V - \mu)^+] + E[(\mu - V)^+]$, hence, combining with the above yields $E[|V - \mu|] = 2E[(V - \mu)^+] = 2E[(\mu - V)^+]$. We use these two identities to re-express D. From the first equality and the tail integral formula for expectation,

$$D = \frac{1}{\mu}E[(V - \mu)^+] = \frac{1}{\mu}\int_0^\infty \Pr\left((V - \mu)^+ \ge t\right) dt = \frac{1}{\mu}\int_0^{\mu(S-1)} \Pr\left(V \ge \mu + t\right) dt = \int_M^{S+M-1} \overline{F}(\mu x + c)dx,$$

where the last line follows from the change of variables $\mu + t \to \mu x + c$. Similarly, using second equality and the tail integral formula for expectation,

$$D = \frac{1}{\mu}E[(\mu - V)^+] = \frac{1}{\mu}\int_0^\infty \Pr((\mu - V)^+ > t)dt = \frac{1}{\mu}\int_0^{\mu-c} \Pr(V \le \mu - t)dt = \int_0^M F(\mu x + c)dx$$

where the last line follows from the change of variables $\mu - t \to \mu x + c$. $\qquad\square$

Proof of Lemma 6. We follow the same strategy as previous bounds. Note that when the coefficient of deviation is high, the probability that V is "close" to 1 is low, since $\mu = 1$. Formally, we claim that

$$\Pr(V \geq t) \leq 1 - D \quad \forall t \in (1, S) \tag{20}$$

To prove the claim, note that $D = E[(1-V)^+] \leq \Pr(V \leq 1)$, where the equality is Lemma 5 and the inequality uses $(1-V)^+ \leq 1$. Rearranging proves $\Pr(V \geq 1) \leq 1 - D$, which in turn implies Eq. (20).

We use this inequality when pointwise bounding our integral representation. Specifically, for any $1 \leq t_0 \leq S$, we have

$$D = \int_1^S \Pr(V > t)dt \qquad \text{(Lemma 5)}$$

$$= \int_1^{t_0} \Pr(V > t) + \int_{t_0}^S \Pr(V > t)dt$$

$$\leq \int_1^{t_0} (1-D)dt + \int_{t_0}^S \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{dt}{t} \qquad \text{(Eq. (20) and Pricing Inequality)}$$

$$= (t_0 - 1)(1-D) + \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \log\left(\frac{S}{t_0}\right) \tag{21}$$

Minimizing over t_0 yields $t_0 = \max\left\{1, \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{(1-D)}\right\}$. We next argue that $D \geq \delta_M$ implies $1 \leq \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{(1-D)}$, so that the unique minimizer is $t_0 = \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{(1-D)}$.

Recall by Eq. (15) $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{\log(S)}{D}$ for all values of D and, in particular, we have that for $D \in [\delta_M, \delta_H]$,

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{\log(S)}{D} \leq \frac{\log(S)}{\delta_M} = 1 + \log(S).$$

Further $D \geq \delta_M = \frac{\log(S)}{1 + \log(S)}$ implies that $1 + \log(S) \leq \frac{1}{1-D}$. Combining shows

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{1}{1-D} \iff 1 \leq \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{(1-D)},$$

which confirms that $t_0 = \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{(1-D)}$ is the unique minimizer.

Plugging in this value $t_0 = \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{(1-D)}$ into Eq. (21) yields:

$$1 \leq \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} + \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \log\left(\frac{S(1-D)}{\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}}\right)$$

We next use properties of the Lambert-W function to simplify this equation. For notational convenience define $\alpha = \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}$. Then,

$$1 \leq \alpha + \alpha \log\left(\frac{S(1-D)}{\alpha}\right) \iff 1 \leq \alpha(1 + \log(S(1-D)) - \log(\alpha)) \tag{22}$$

$$\iff -1 \geq \alpha(\log(\alpha) - \log(eS(1-D)))$$

Note $\alpha = e^{\log(\alpha)} = e^{\log(\alpha) - \log(eS(1-D))} \cdot e \cdot S(1-D)$. Substituting above proves

$$\frac{-1}{eS(1-D)} \geq e^{\log(\alpha) - \log(eS(1-D))}(\log(\alpha) - \log(eS(1-D))).$$

The left hand side is between $-1/e$ and 0 by inspection. The function $W_{-1}(\cdot)$ is non-increasing on this range, so that applying $W_{-1}(\cdot)$ to both sides yields

$$W_{-1}\left(\frac{-1}{eS(1-D)}\right) \leq \log(\alpha) - \log(eS(1-D)) \iff \alpha \geq eS(1-D) \cdot e^{W_{-1}\left(\frac{-1}{eS(1-D)}\right)} \tag{23}$$

$$\iff \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq -\frac{-1}{eS(1-D)} e^{-W_{-1}\left(\frac{-1}{eS(1-D)}\right)}.$$

Finally, from the definition of W_{-1},

$$\frac{-1}{eS(1-D)} = W_{-1}\left(\frac{-1}{eS(1-D)}\right) e^{W_{-1}\left(\frac{-1}{eS(1-D)}\right)},$$

which we use to simplify the last inequality to obtain $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq -W_{-1}\left(\frac{-1}{eS(1-D)}\right).$ □

Proof of Lemma 7. First consider part a). Recalling that $-W_{-1}(-1/e) = 1$, we confirm directly that the given function is negative as $D \downarrow 0$ since it is continuous. Notice further that $-W_{-1}(\cdot)$ is an increasing function (c.f. Fig. 2), whereby $\frac{-W_{-1}\left(-\frac{1-D}{e}\right)}{1-D}$ is an increasing function, while $\log(S)/D$ is a decreasing function. It follows that the given function has a unique root, and it suffices to show this root is δ_L to complete the proof. To this end, write,

$$-\frac{W_{-1}\left(-\frac{1-D}{e}\right)}{1-D} = \frac{\log(S)}{D} \iff W_{-1}\left(-\frac{1-D}{e}\right) = \log\left(S^{\frac{D-1}{D}}\right)$$

$$\iff -\frac{1-D}{e} = \log\left(S^{\frac{D-1}{D}}\right)\cdot\exp\left(\log\left(S^{\frac{D-1}{D}}\right)\right) \quad \text{(definition of Lambert-}W)$$

$$\iff -\frac{1}{eS} = S^{\frac{-1}{D}}\cdot\frac{-\log(S)}{D} \quad \text{(simplifying)}$$

$$\iff -\frac{1}{eS} = \exp\left(-\frac{\log(S)}{D}\right)\cdot\frac{-\log(S)}{D} \quad \text{(using } S^{\frac{-1}{D}} = \exp\left(-\frac{\log(S)}{D}\right))$$

$$\iff W_{-1}\left(-\frac{1}{eS}\right) = -\frac{\log(S)}{D} \quad \text{(Applying } W_{-1}(\cdot))$$

$$\iff D = -\frac{\log(S)}{W_{-1}\left(-\frac{1}{eS}\right)} = \delta_L.$$

This completes the proof of part a).

To prove part b), first observe that

$$W_{-1}\left(-\frac{1}{eS(1-D)}\right) \geq -\frac{\log(S)}{D} \iff -\frac{1}{eS(1-D)} \leq -\frac{\log(S)}{D}\exp\left(-\frac{\log(S)}{D}\right),$$

because the function $y \mapsto ye^y$ is the inverse of $W_{-1}(\cdot)$ and is non-increasing on the domain of $W_{-1}(\cdot)$, i.e., $[-1/e, 0)$. Simplifying the righthand inequality yields,

$$\frac{-1}{e} \leq \log\left(S^{\frac{D-1}{D}}\right)\cdot S^{\frac{D-1}{D}}.$$

Now make the substitution $\log\left(S^{\frac{D-1}{D}}\right) \to y$ so this last inequality is equivalent to $\frac{-1}{e} \leq ye^y$. One can confirm by differentiation that $y \mapsto ye^y$ has a unique minimizer at $y = -1$, and, thus, this last inequality holds for all y. This proves the function defined in part b) is nonnegative everywhere. Moreover, it has a root at $y = 1$ which corresponds to $\log\left(S^{\frac{D-1}{D}}\right) = -1$. Simplifying shows this condition is equivalent to $D = \log(S)/(1+\log(S)) = \delta_M$, as was to be proven.   □

We next explicitly describe the distributions which make Theorem 1 tight. By Lemma 1, it suffices to consider the case where $c = 0$ and $\mu = 1$. The general case can be handled by scaling and shifting the below tight distributions:

LEMMA 11 **(Tight distributions).**

a) *Suppose $D \in [0, \delta_L]$, and let $\alpha_L = \left(\frac{W_{-1}\left(\frac{D-1}{e}\right)}{D-1} \right)^{-1}$. Then, there is a random variable V with ccdf*

$$\overline{F}_L(x) = \begin{cases} 1 & \text{if } 0 \leq x < \alpha_L \\ \frac{\alpha_L}{x} & \text{if } \alpha_L \leq x \leq 1 \\ \frac{D}{\log(S)x} & \text{if } 1 < x \leq S \\ 0 & \text{otherwise,} \end{cases}$$ (Tight CCDF, Low Heterogeneity)

and this random variable has scale S, coefficient of deviation D, and mean 1 and satisfies Eq. (14) with equality.

b) *Suppose $D \in [\delta_L, \delta_M]$, and let $\alpha_M = \frac{D}{\log(S)}$. Then, there is a random variable V with ccdf*

$$\overline{F}_M(x) = \begin{cases} 1 & \text{if } x = 0, \\ \frac{\alpha_M}{e} S^{\frac{1}{D}-1} & \text{if } x \in (0, \, eS^{1-\frac{1}{D}}) \\ \frac{\alpha_M}{x} & \text{if } x \in [eS^{1-\frac{1}{D}}, S] \\ 0 & \text{otherwise,} \end{cases}$$ (Tight CCDF, Medium Heterogeneity)

and this random variable has scale S, coefficient of deviation D, and mean 1 and satisfies Eq. (15) with equality.

c) *Suppose $D \in [\delta_M, \delta_H]$, and let $\alpha_H := \left(-W_{-1}\left(\frac{-1}{eS(1-D)} \right) \right)^{-1}$. Then, there is a random variable V with ccdf*

$$\overline{F}_H(x) = \begin{cases} 1 & \text{if } x = 0, \\ 1 - D & \text{if } x \in (0, \frac{\alpha_H}{1-D}] \\ \frac{\alpha_H}{x} & \text{if } x \in (\frac{\alpha_H}{1-D}, S) \\ 0 & \text{otherwise,} \end{cases}$$ (Tight CCDF, High Heterogeneity)

and this random variable has scale S, coefficient of deviation D, and mean 1 and satisfies Eq. (16) with equality.

Proof of Lemma 11. Intuitively, $\overline{F}_L$, $\overline{F}_M$, and $\overline{F}_H$ each make all the pointwise bounds on the ccdf the integral representation of D used in the proofs of Eqs. (14) to (16) tight, simultaneously. Thus, they will make the overall bound tight.

To prove the lemma formally, we will prove that $\overline{F}_L$, $\overline{F}_M$ and $\overline{F}_H$ are valid ccdfs, each with mean 1, scale S, and coefficient of deviation D, and that $\mathcal{R}_{SP}(F_L, 0) = \alpha_L$, $\mathcal{R}_{SP}(F_M, 0) = \alpha_M$ and $\mathcal{R}_{PP}(F_H, 0) = \alpha_H$, respectively. The lemma then follows directly from the definition of α_L, α_M and α_H since $\mathcal{R}_{PP}(F_L, 0) = \mathcal{R}_{PP}(F_M, 0) = \mathcal{R}_{PP}(F_H, 0) = \mu = 1$.

a) *(Low Heterogeneity)* Note that replacing α by α_L and the inequality by equality in Eq. (13) and then following the implications backwards proves that α_L satisfies

$$D = 1 - \alpha_L + \alpha_L \log(\alpha_L).$$

We next prove $\overline{F}_L$ is a valid ccdf. By inspection, we need only prove $\overline{F}_L$ is non-increasing, i.e., that $\alpha_L \geq D/\log(S) \iff 1/\alpha_L \leq \log(S)/D$. This inequality follows directly from Lemma 7 since $D \in [0, \delta_L]$, and the lefthand side is low-heterogeneity bound while the right side is the medium heterogeneity bound. This proves $\overline{F}_L$ is valid.

Next, write

$$\int_0^\infty \overline{F}_L(t)dt = \int_0^1 \overline{F}_L(x)dx + \int_1^S \overline{F}_L(x)dx = \alpha_L - \alpha_L \log(\alpha_L) + D = 1,$$

where the last equality uses the identity proven above for α_L. Thus, $\overline{F}_L$ has mean 1. By Lemma 5, its coefficient of deviation is

$$\int_0^1 1 - \overline{F}_L(x)dx = \int_0^{\alpha_L} 0dx + \int_{\alpha_L}^1 1 - \frac{\alpha_L}{x} \, dx = 1 - \alpha_L + \alpha_L \log(\alpha_L) = D, \tag{24}$$

again using the identify for α_L. By inspection, it has scale S.

Finally, any price $x \in [\alpha_L, 1]$ earns profit α_L, while any price $x \in [0, \alpha_L)$ earns profit strictly less than α_L. Any price $x \in (1, S]$ earns profit $D/\log(S)$ which is at most α_L as we noted when proving that $\overline{F}_L$ is valid. Thus, $\mathcal{R}_{SP}(F_L, 0) = \alpha_L$, which proves that a random variable V with ccdf $\overline{F}_L$ will satisfy Eq. (14) with equality.

b) *(Medium Heterogeneity)* To prove that $\overline{F}_M$ is a valid ccdf, it suffices to show that $eS^{1-\frac{1}{D}} \leq S$, which is equivalent to $1 \geq \frac{D}{\log(S)}$. Rewrite this last inequality as $\frac{1}{\alpha_M} \geq 1$, and recall from Step 1 of the proof of Theorem 1 that $\frac{1}{\alpha_M}$ is an upper bound on the value of personalization and, thus, must be at least 1.

Next, write

$$\int_0^\infty \overline{F}_M(x)dx = \int_0^{eS^{1-\frac{1}{D}}} \overline{F}_M(x)dx + \int_{eS^{1-\frac{1}{D}}}^S \overline{F}_M(x)dx = \alpha_M + \alpha_M \log\left(\frac{S}{eS^{1-\frac{1}{D}}}\right) = 1,$$

where the last equality uses the definition of α_M. It follows that $\overline{F}_M$ has mean 1, and, by inspection, scale S. Write,

$$\int_1^S \overline{F}_M(x)dx = \alpha_M \log S = D,$$

to conclude from Lemma 5 that $\overline{F}_M$ has coefficient of deviation D. Finally, observe that any price $x \in [eS^{1-\frac{1}{D}}, S]$ earns profit α_M, while any other price earns strictly less profit. Thus, $\mathcal{R}_{SP}(F_M, 0) = \alpha_M$, completing this part of the lemma.

c) *(High Heterogeneity)* To prove $\overline{F}_H$ is a valid ccdf, it suffices to show that $\alpha_H/(1-D) \leq S$. Note that by Lemma 6, $1/\alpha_H$ is an upper-bound on the value of personalization, whereby α_H is necessarily at most 1. Moreover, for the Lambert-W function defining α_H to be well-defined, we must have that $\frac{1}{S(1-D)} \leq 1$ which implies $S(1-D) \geq 1$. Thus, $\alpha_H \leq 1 \leq S(1-D)$ which implies that $\alpha_H/(1-D) \leq S$ and that $\overline{F}_H$ is a valid ccdf.

Next write,

$$\int_0^S \overline{F}(x)dx = \int_0^{\frac{\alpha_H}{1-D}} (1-D)dx + \int_{\frac{\alpha_H}{1-D}}^S \frac{\alpha_H}{x}dx$$

$$= \alpha_H + \alpha_H \log\left(\frac{S}{\alpha_H}(1-D)\right). \tag{25}$$

We claim this last quantity equals 1. Indeed, from the definition of $W_{-1}(\cdot)$, $\alpha_H = eS(1-D) \cdot e^{W_{-1}\left(\frac{-1}{eS(1-D)}\right)}$. Then, replace α by α_H and the inequality by equality in Eq. (23) and follow the implications backwards to Eq. (22), proving the claim. Thus, $\overline{F}_H$ has mean 1, and, by inspection, has scale S.

To compute its coefficient of deviation, we first claim that $\alpha_H/(1-D) \geq 1$. Indeed, recall that

$$D \geq \delta_M = \frac{\log(S)}{1+\log(S)} \iff \log(S) \leq \frac{D}{1-D} \iff \frac{\log(S)}{D} \leq \frac{1}{1-D}.$$

It follows that

$$\frac{\alpha_H}{1-D} \geq \alpha_H \frac{\log(S)}{D} = \frac{\alpha_H}{\alpha_M} \geq 1,$$

where the last inequality follows from Lemma 7. Now compute

$$\int_0^1 1 - \overline{F}_H(x)dx = D,$$

whereby $\overline{F}_H$ has coefficient of deviation D by Lemma 5.

$$\int_0^1 1 - \overline{F}(x)dx = D.$$

It remains to check that $\mathcal{R}_{SP}(F,0) = \alpha_H$, which we verify directly by observing that any price $x \in [\frac{\alpha_H}{1-D}, S]$ obtains profit α_H any any other price obtains profit no more than α_H. $\quad\square$

Proof of Lemma 4. Let us fix S and M, and define $\alpha(D) := \alpha(S, M, D)$. Fix any D_1, D_2, with $0 \leq D_1 \leq D_2 \leq \delta_H$, and any $t \in [0,1]$. We will show that $\alpha(tD_1 + (1-t)D_2) \leq t\alpha(D_1) + (1-t)\alpha(D_2)$ to prove the theorem.

By Theorem 1, there exists random variables $V_1 \sim F_1$ and $V_2 \sim F_2$ each with scale S and margin M such that the coefficient of deviation of F_1 is D_1, the coefficient of deviation of F_2 is D_2, $\alpha(D_1) = \frac{\mathcal{R}_{SP}(F_1,c)}{\mathcal{R}_{PP}(F_1,c)}$ and $\alpha(D_2) = \frac{\mathcal{R}_{SP}(F_2,c)}{\mathcal{R}_{PP}(F_2,c)}$.

Since both V_1 and V_2 have the same margin and cost, they also have the same mean $\mu = \frac{c}{1-M}$. Take X to be a Bernoulli random variable with parameter t, and let $\tilde{V} \equiv XV_1 + (1-X)V_2$ where X, V_1, V_2 are sampled independently. Note that $\tilde{V}$ has mean μ, margin M, and scale S. Furthermore, the coefficient of deviation of $\tilde{V}$ is

$$\begin{aligned}
\tilde{D} &= \frac{1}{2\mu}\left(E\Big[|XV_1 + (1-X)V_2 - \mu|\Big]\right) \\
&= \Pr(X=1) \cdot \frac{1}{2\mu}E\Big[|V_1 - \mu|\Big] + \Pr(X=0) \cdot \frac{1}{2\mu}E\Big[|V_2 - \mu|\Big] \\
&= tD_1 + (1-t)D_2.
\end{aligned} \tag{26}$$

To conclude the proof, write

$$\begin{aligned}
t\alpha(D_1) + (1-t)\alpha(D_2) &= t\frac{\mathcal{R}_{SP}(F_1,c)}{\mathcal{R}_{PP}(F_1,c)} + (1-t)\frac{\mathcal{R}_{SP}(F_2,c)}{\mathcal{R}_{PP}(F_2,c)} \\
&= \frac{t\mathcal{R}_{SP}(F_1,c) + (1-t)\mathcal{R}_{SP}(F_2,c)}{\mathcal{R}_{PP}(\tilde{F},c)} \\
&\geq \frac{\mathcal{R}_{SP}(\tilde{F},c)}{\mathcal{R}_{PP}(\tilde{F},c)} \\
&\geq \alpha(\tilde{D}) \\
&= \alpha(tD_1 + (1-t)D_2).
\end{aligned}$$

The first equation follows from the definitions of F_1 and F_2. The second equation follows from the fact that the personalized pricing strategy yields $\mu - c$ for F_1, F_2, and $\tilde{F}$. The first inequality follows from the fact

that the optimal single price for $\tilde{V}$ yields revenue of at most $\mathcal{R}_{SP}(F_1, c)$ for the market corresponding to V_1 and at most $\mathcal{R}_{SP}(F_2, c)$ for the market corresponding to V_2. The second inequality follows Theorem 1. The last equality follows from Eq. (26). $\square$

Proof of Corollary 1.   Note that Eq. (1) shows that

$$W_{-1}\left(-\frac{x}{e}\right) = 1 + \sqrt{2\log(1/x)} + O(\log(1/x)) \quad \text{as } x \to 1.$$

Substituting this expression into the bounds in the low heterogeneity and high heterogeneity regimes proves the result. $\square$

A.3. Omitted Proofs from Section 4

Proof of Lemma 9.   By inspection, $\mathcal{R}_{PP} = \frac{t}{2}$. To compute $\mathcal{R}_{kP}$, consider an optimal segmentation $s_0, \ldots, s_{k+1}$ with corresponding prices $p_1, \ldots p_k$. (Recall $s_0 = c = 0$, $s_{k+1} = t$, and $p_i \in [s_i, s_{i+1}), i \geq 1$.) By Lemma 8, $s_i = p_i$ for $i = 1, \ldots, k$.

Now, on segment $[s_i, s_{i+1})$, the conditional distribution of V is uniform, so the personalized pricing strategy earns profit $\frac{s_{i+1}+s_i}{2} \cdot \frac{s_{i+1}-s_i}{t}$ for all i, since only $\frac{s_{i+1}-s_i}{t}$ fraction of the market is in this interval. By contrast, for $i = 1, \ldots, k-1$, the k-market segmentation strategy earns revenue $s_i \frac{s_{i+1}-s_i}{t}$ since $p_i = s_i$ and thus, all customers in the segment buy. The difference in revenue between the two strategies is then

$$\mathcal{R}_{PP}(F, 0) - \mathcal{R}_{kP}(F, 0, \mathbf{s}, \mathbf{p}) = \frac{s_1^2}{2t} + \sum_{i=1}^{k} \frac{s_{i+1}+s_i}{2} \cdot \frac{s_{i+1}-s_i}{t} - s_i \frac{s_{i+1}-s_i}{t}$$

$$= \frac{s_1^2}{2t} + \frac{1}{2t}\sum_{i=1}^{k}(s_{i+1}-s_i)^2 = \frac{1}{2t}\sum_{i=0}^{k}(s_{i+1}-s_i)^2.$$

The segmentation which maximizes $\mathcal{R}_{kP}(F, 0, \mathbf{s}, \mathbf{p})$ also minimizes this difference. By inspection, for a fixed s_1, the optimal segmentation is equispaced, i.e., $s_i = s_{i-1} + \frac{t}{k+1}$ for $i = 1, \ldots, k$. Thus $\mathcal{R}_{kP} = \sum_{i=1}^{k} \frac{it}{k+1} \frac{1}{k+1} = \frac{t}{2}\frac{k}{k+1}$ Consequently, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{kP}} = 1 + 1/k$. $\square$

Proof of Theorem 2.   We prove the second part of the theorem first.

(b) We first consider the case where $V \sim F$ has $\mu = 1$ and $c = 0$. Consider a partition $\delta = s_0 < s_1 < \ldots s_k < s_{k+1} = S$. Let $V_i \sim F_i$ denote the random variable V conditional on the event $V \in [s_i, s_{i+1})$, i.e., $\Pr(V_i \leq t) \equiv \Pr\left(V \leq t \mid s_i \leq V \leq s_{i+1}\right)$. Further, let $q_i = \Pr(s_i \leq V \leq s_{i+1})$ be the market share of the i-th segment, S_i be the scale of V_i and $\mathcal{R}_{SP}(F_i, 0) = \max_p p\overline{F_i}(p)$.

From Eq. (5) in the proof of Lemma 2 we have that for any $\gamma \leq \mathcal{R}_{SP}(F_i, 0)$,

$$\frac{\mathcal{R}_{PP}(F_i, 0)}{\mathcal{R}_{SP}(F_i, 0)} \leq 1 + \log\left(\frac{E[V_i]S_i}{\gamma}\right) \tag{27}$$

We will apply Eq. (27) to each F_i with the trivial lower bound s_i. Notice that since $V_i \leq s_{i+1}$ almost surely, $S_i E[V_i] \leq x_i$.

Then, since $\mu = 1$,

$$1 = \sum_{i=0}^{k} q_i E[V_i] = \sum_{i=0}^{k} q_i \mathcal{R}_{PP}(F_i, 0)$$

$$\leq \sum_{i=0}^{k} q_i \left(1 + \log\left(\frac{s_{i+1}}{s_i}\right)\right) \mathcal{R}_{SP}(F_i, 0) \qquad\qquad (\text{since } S_i E[V_i] \leq s_{i+1})$$

$$\leq \left(\sum_{i=0}^{k} q_i \mathcal{R}_{SP}(F_i, 0)\right) \max_{i=0,\dots,k} \left(1 + \log\left(\frac{s_{i+1}}{s_i}\right)\right)$$

$$\leq \mathcal{R}_{kP}(F, 0) \max_{i=0,\dots,k} \left(1 + \log\left(\frac{s_{i+1}}{s_i}\right)\right)$$

where the last line follows because partitioning at the s_i and offering prices $p^i \in \arg\max_{p \geq s_i} p\overline{F_i}(p)$ is a feasible segmentation policy. We minimize this last bound by setting $s_i = (\delta)^{\frac{k-i}{k}}(S)^{\frac{i}{k}}$ which implies $\frac{s_{i+1}}{s_i} = \left(\frac{S}{\delta}\right)^{1/k}$. This choice of s_i yields

$$\mathcal{R}_{PP}(F, 0) \leq \mathcal{R}_{kP}(F, 0) \left(1 + \frac{\log\left(\frac{S}{\delta}\right)}{k}\right). \tag{28}$$

For the general case note that for the transformation in Lemma 1, one can prove that $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{kP}(F,c)} = \frac{\mathcal{R}_{PP}(F_c,0)}{\mathcal{R}_{kP}(F_c,0)}$ by considering each segment separately and applying an argument analogous to Lemma 1. Thus, given an F with arbitrary mean and $c > 0$, first transform to F_c and apply the above result. Substituting the original parameters proves the second part of theorem.

We now use the previous result to prove the first part of the theorem.

(a) First consider the case where $\mu = 1$ and $c = 0$. We prove the bound by separating the distribution into a small lower component near 0 and an upper component. We bound the lower component of F in terms of D and bound the upper component by applying (a). Fix some $\Delta > 1$ which we shall select later. To bound the lower tail, note

$$D = E[(1-V)^+] \geq E\left[(1-V)^+ \mathcal{I}\left(V \leq \frac{1}{\Delta}\right)\right] \geq E\left[\left(1 - \frac{1}{\Delta}\right)\mathbb{I}\left(V \leq \frac{1}{\Delta}\right)\right] = \left(1 - \frac{1}{\Delta}\right)\Pr\left(V \leq \frac{1}{\Delta}\right).$$

Rearranging yields $\Pr(V \leq 1/\Delta) \leq \frac{\Delta D}{\Delta - 1}$. This further implies that

$$E\left[V\mathbb{I}\left(V \leq \frac{1}{\Delta}\right)\right] \leq \frac{1}{\Delta}\Pr(V \leq 1/\Delta) \leq \frac{D}{\Delta - 1}. \tag{29}$$

Now by splitting the expectation,

$$1 = \mathcal{R}_{PP}(F, 0) = E\left[V\mathbb{I}\left(V \leq \frac{1}{\Delta}\right)\right] + E\left[V\mathbb{I}\left(V \geq \frac{1}{\Delta}\right)\right]$$

$$\leq \frac{D}{\Delta - 1} + E\left[V\mathbb{I}\left(V \geq \frac{1}{\Delta}\right)\right] \qquad\qquad (\text{using Eq. (29)})$$

$$= \frac{D}{\Delta - 1} + E[V_\Delta]\Pr(V \geq 1/\Delta)$$

where V_Δ is the conditional distribution of V given that $V \geq 1/\Delta$, i.e., $\Pr(V_\Delta \geq t) = \Pr(V \geq t | V \geq 1/\Delta)$.

Note that $E[V_\Delta] \geq 1$ and that V_Δ has scale $S/E[V_\Delta] \leq S$. Most importantly, $\Pr(V_\Delta \geq \frac{1}{\Delta}) = 1$, so that we can apply part (b) to upper bound the expectation yielding

$$\mathcal{R}_{PP}(F, 0) \leq \frac{D}{\Delta - 1} + \left(1 + \frac{\log(S\Delta)}{k}\right)R_{kP}(F_\Delta, 0)\Pr(V \geq 1/\Delta).$$

Finally, letting $\mathbf{s}^*$ be the optimal segmentation for $\mathcal{R}_{kP}(F_\Delta, 0)$ and $\mathbf{p}^*$ be the corresponding prices. Define the function

$$r(v) = \begin{cases} p_i^* & \text{if } v \in [s_i^*, s_{i+1}^*) \text{ and } v \geq p_i^*, \quad i = 1, \dots, k \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$
\begin{aligned}
\mathcal{R}_{kP}(F,0) &\geq \mathcal{R}_{k,P}(F,0,\mathbf{s}^*,\mathbf{p}^*) \\
&= E[r(V)] \\
&= E\left[r(V) \mid V \leq \frac{1}{\Delta}\right] \Pr\left(V \leq \frac{1}{\Delta}\right) + E\left[r(V) \mid V \geq \frac{1}{\Delta}\right] \Pr\left(V \geq \frac{1}{\Delta}\right) \\
&= E\left[r(V) \mid V \leq \frac{1}{\Delta}\right] \Pr\left(V \leq \frac{1}{\Delta}\right) + R_{kP}(V_\Delta) \Pr\left(V \geq \frac{1}{\Delta}\right) \\
&\geq R_{kP}(V_\Delta) \Pr\left(V \geq \frac{1}{\Delta}\right)
\end{aligned}
$$

Plugging in above yields,

$$
\mathcal{R}_{PP}(F,0) \leq \frac{D}{\Delta - 1} + \left(1 + \frac{\log(S\Delta)}{k}\right) R_{kP}(F,0).
$$

Letting $\Delta = 1 + D(k+1)$ and rearranging yields

$$
\frac{\mathcal{R}_{PP}}{\mathcal{R}_{kP}} \leq \frac{1 + \frac{\log(S\Delta)}{k}}{1 - \frac{D}{\Delta - 1}} \leq 1 + \frac{1}{k} + \frac{\log(S + SD(k+1))}{k} = 1 + O\left(\frac{\log(k)}{k}\right).
$$

For a general $c > 0$ and $\mu \neq 1$, apply the transformation of Lemma 1. As in the previous part, note that $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{kP}(F,c)} = \frac{\mathcal{R}_{PP}(F_c,0)}{\mathcal{R}_{kP}(F_c,0)}$. Apply the result of the previous part and then make the appropriate substitutions.
$\square$

Appendix B: A Dynamic Programming Algorithm for Computing (kP)

In this section, we describe an efficient dynamic programming algorithm for computing the optimal k-market segmentation when the valuation distribution is known precisely and discretely supported on n values. One should compare this algorithm to the distribution-agnostic procedure given in Theorem 2 when the valuation distribution is not known precisely.

Structurally, computing the optimal k-market segmentation is extremely similar to the *1D Clustering* problem for which dynamic programming approaches have been employed (see Gronlund et al. (2017) for a modern overview). Formally, suppose V is supported on n values $\{v_i\}_{i=1}^n$, occurring with probabilities $\{q_i\}_{i=1}^n$. Without loss of generality suppose the values are indexed from low to high, i.e., $v_i \leq v_{i+1}$ for all i. By an argument identical to Lemma 8 the optimal segmentation $\{s_i\}_{i=0}^k$ is contained in the support of V, we wish to find $\{s_i\}_{i=0}^k \subset \{v_i\}_{i=1}^n$ that maximizes

$$
\sum_{i=1}^k s_i \Pr\left(V \in [s_i, s_{i+1})\right).
$$

We give a dynamic programming solution that uses time $O(kn^2)$. Define $D[m,j]$ as the optimal j-market segmentation that considers the m lowest points $\{(v_i, q_i)\}_{i=1}^m$, our goal is to compute $D[n,k]$. Our algorithm depends on the following observation: consider the optimal k-market segmentation and suppose $[v_{i_k}, v_n]$ defines the kth segment. If one considers the market without the customers in the kth segment, the remaining $k - 1$ segments must be an optimal $(k-1)$-market segmentation on $\{(v_i, q_i)\}_{i=1}^{i_k - 1}$. Formally we express this observation as the following recursion,

$$
D[m,j] = \max_{l \in [m-1]} D[l, j-1] + v_{l+1} \sum_{i=l+1}^m q_i, \tag{30}
$$

which states that the optimal j-market segmentation on the lowest m valuations, is equal to some optimal $(j-1)$-segmentation on a smaller market, plus the value of the jth segment. Using Eq. (30) we may populate a table of size kn, starting at $D[0,0]=0$, and computing column-wise. Each computation of $D[m,j]$ requires $O(n)$ operations, thus the table may be populated in $\mathrm{O}(kn^2)$ time.