The Value of Personalized Pricing

Adam N. Elmachtoub
Department of Industrial Engineering and Operations Research & Data Science Institute, 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
Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, mhamilton@katz.pitt.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 trade-offs, our work studies the value of personalized pricing strategies over a simple single price strategy.

We first provide closed-form lower and upper bounds on the ratio between the profits of an idealized personalized pricing strategy (first-degree price discrimination) and a single price strategy. Our bounds depend on simple statistics of the valuation distribution and shed light on the types of markets for which personalized pricing has little or significant potential value. Second, we consider a feature-based pricing model where customer valuations can be estimated from observed features. We show how to transform our aforementioned bounds into lower and upper bounds on the value of feature-based pricing over single pricing. Finally, we demonstrate how to obtain sharper bounds by incorporating additional information about the valuation distribution (moments or shape constraints) by solving tractable linear optimization problems.

Key words: price discrimination, personalization, market segmentation

1. Introduction

Over the last decade, increased availability of 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 customer's willingness to pay and then customize a price for that customer. This customized price is often delivered as a discount to a universal, posted price 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 strategies (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 provide various upper and lower bounds on the profit ratio between personalized pricing and a simple single price strategy. We consider two different strategies: *(i) idealized personalized pricing (PP)*, i.e., charging each customer exactly their willingness to pay, and *(ii) feature-based personalized pricing (XP)*, i.e., charging each customer a price based on their observed feature data. For both personalization strategies, we benchmark the profit against the simple *single price (SP)* strategy that offers one price uniformly to all customers. The bounds we develop on the profit ratios between personalized pricing and single pricing can guide managers in assessing the upside of personalized pricing in potential markets. For example, in settings where an upper bound is close to one, we know that *any* form of price discrimination necessarily has limited value, while in settings where a lower bound is far from one, we are guaranteed the value of personalized pricing is significant.

With full-information about the customer valuation distribution, computing the exact ratio between personalized pricing over single pricing 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. For example, in data-poor settings, managers may be able to estimate simple statistics such as the mean based on domain knowledge or comparable products, but may find it impossible to accurately specify an entire distribution. Even in data-rich settings, no non-parametric density estimator using n data points converges in mean-integrated squared error (MISE) at rate at faster than $O(n^{-4/5})$, while a simple sample moment converges to its true moment in mean-squared error at a rate of $O(n^{-1})$ (Van der Vaart 2000, Chapt. 24). 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 or minimal. These structural insights can guide practitioners weighing the benefits of price discrimination for a particular market against the aforementioned drawbacks.

More specifically, in the first part of the paper, we prove upper and lower bounds on the profit ratio between idealized personalized pricing and 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. We prove upper and lower bounds that are tight, closed-form, and depend on simple properties of the valuation distribution. Specifically, our upper bounds depend on three unit-less 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 upper 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. Perhaps surprisingly, we also find that our upper bound is maximal for intermediate values of the coefficient of deviation and approaches one as the coefficient deviation increases with all other parameters fixed.

We complement our upper bounds with novel lower bounds that depend on the coefficient of deviation and mild shape assumptions on the valuation distribution such as i) unimodality or ii) unimodality and symmetry. We also show that without any shape assumptions, no non-trivial lower bound is theoretically possible. To the best of our knowledge, our lower bounds yield the first provable separation between personalized pricing and single price strategies for a generic class of distributions. Indeed, our lower bounds provide precise conditions for when increased heterogeneity in the market guarantees increased value in personalized pricing. Together our bounds yield strong conditions for identifying which markets are ripe for personalized pricing and which are well-served by a single price.

Idealized personalized pricing is not implementable in practice as it assumes the monopolist can perfectly predict each customer's valuation. Hence, we also study an alternate pricing strategy that we call feature-based pricing, where the seller observes a feature vector (sometimes called a context) for each customer which the seller can use to (imperfectly) predict the customer's valuation and offer a custom price. This strategy more closely resembles price discrimination strategies implemented in practice. We prove a novel theorem that relates lower and upper bounds on the profit ratio of feature-based pricing over single pricing to the profit ratio of idealized personalized pricing over single pricing (discussed above). The relationship between these two ratios is driven by the degree to which the observable contexts are informative for the unknown customer valuation, as measured by the size of the residual error when predicting valuations. More specifically, our bounds depend on the mean absolute deviation of this residual error. Our bounds make precise the intuition that when the contexts are very informative, feature-based pricing performs comparably to first-degree price discrimination, but when contexts are uninformative, feature-based pricing offers little benefit over single-pricing. Moreover, our bounds show how one can decompose the value of feature-based pricing strategies into the potential benefits of perfect personalization and the losses from less than perfectly informative features.

In the last part of our paper, we then show how to generalize our work beyond dependency on the coefficient of deviation. Specifically, we provide an algorithmic procedure to compute essentially tight bounds on the value of idealized personalized pricing over single pricing given any generalized moment of the valuation distribution, such as the variance or quantile information. The key ideas leverage infinite-dimensional linear optimization duality and a careful discretization argument to generate a tractable optimization formulation suitable for off-the-shelf software. We show that when using variance (coefficient of variation), our bounds have the same insights and structure as the ones derived in closed-form for the case of coefficient of deviation.

To summarize our contributions:

1. We prove closed-form, tight upper bounds for the value of idealized personalized pricing over single-pricing when the scale, margin, and coefficient of deviation of the valuation distribution are known (cf. Theorems 1 and 2). When these upper bounds are small, this suggests the value of any personalized pricing strategy is rather limited.
2. We prove closed-form lower bounds on the value of idealized personalized pricing that rely on necessary shape assumptions such as unimodality or unimodality and symmetry (cf. Theorem 3). In the latter case, our bound is tight for any specified coefficient of deviation. Our lower bounds provide guarantees on how much increased value personalized pricing can provide as a function of the market heterogeneity.
3. We then consider the more practical feature-based pricing, and generate lower and upper bounds on its value in comparison to the ideal case and single pricing (cf. Theorems 4 and 5). These bounds make explicit the relationship between the informational value of the features, and the value of feature-based pricing in a market. The proof fundamentally utilizes the previously derived bounds in the ideal case.
4. Finally, we provide a general methodology for computing essentially tight upper and lower bounds on the value of personalized pricing over single pricing when additional or different moment information is known about the valuation distribution. Our methodology also allows for shape assumptions such as unimodality without losing computational tractability (cf. Theorems 6, 7, and 8).

In the interest of reproducibility, open-source code for computing all of our bounds and reproducing all of our plots is available at `https://github.com/vgupta1/VoPP_OptProblems`.

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

Electronic copy available at: https://ssrn.com/abstract=3127719

operational perspective, focusing on the individual firms relative profits under first-degree price discrimination and other forms of 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 (MHR), the value of personalized pricing is at most $e \approx 2.718$. MHR is a technical and arguably unintuitive assumption on the distribution. In experiments, we show this bound is generally loose even when the assumption is satisfied (c.f. Fig. 3). Tamuz (2013) shows that if the ratio of the geometric mean over the mean of the valuation distribution is at least $1-\delta$, then the value of personalized pricing is at most $(1-2^{\frac{4}{3}}\delta^{\frac{1}{3}})^{-1}$, while Medina and Vassilvitskii (2017), shows the value of personalized pricing over single pricing is at most $4.78 + 2\log(1+C^2)$, where C is the coefficient of variation of the valuation distribution. These two bounds are not tight in dependence on δ and C, respectively. By contrast, our analogous upper bounds rely on coefficient of deviation and are proven to be tight for all possible values. We also stress that these existing results all pertain to *upper* bounds on the value of personalized pricing. To the best of our knowledge, we are the first to develop lower bounds for the value of personalized pricing over single-pricing and the first to develop bounds on the value of feature-based pricing over single-pricing.

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 such as the feature-based pricing strategy we consider. Indeed, the operations literature contains many examples of (implicit or explicit) third-degree price discrimination 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. Our results shed insight into on when the value of such price discrimination tactics may be high and worth pursuing, and when the value may be low and not worthwhile. Huang et al. (2019) also studies the value of personalized pricing, but in a social network. There, all customers are identical except for their position in the network, and the proven bounds are asymptotic in the size of the (random) graph.

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 $\mathbb{E}[V]$ is denoted by μ. For mathematical brevity, we shall assume V has countably many point masses. For convenience, we shall also define $\overline{F}(p) := \mathbb{P}(V \geq p)$, which is the probability that a customer shall purchase a product if priced at p. [1] 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 three 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 $\overline{F}(p)$, 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) Feature-Based Pricing (XP): In the feature-based pricing strategy, the monopolist observes a feature vector $\mathbf{X}$ for each customer before offering a price, but does *not* directly observe their valuation V. Based on $\mathbf{X}$, the seller offers a customized price $p(\mathbf{X})$, and the customer purchases with probability $\mathbb{P}(V \geq p(\mathbf{X}) \mid \mathbf{X})$. Given a joint distribution $F_{\mathbf{X}V}$ of $(\mathbf{X}, V)$, let $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c) := \max_{p(\cdot)} \mathbb{E}\left[(p(\mathbf{X}) - c)\mathbb{I}(V \geq p(\mathbf{X}))\right]$ denote the optimal profit under feature-based pricing.

3) 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

[1] It is traditional to assume that if a customer values a product exactly at the price, then a purchase is made. $\overline{F}(\cdot)$ thus includes the $\mathbb{P}(V = p)$, and is not the complementary CDF of V. Note however that since V has countable many point masses, that $\int_{x_1}^{x_2} \overline{F}(t)dt = \int_{x_1}^{x_2} \mathbb{P}(V > t)dt$ for any $x_1 < x_2$.

Electronic copy available at: https://ssrn.com/abstract=3127719

their valuation and, thus, the total revenue earned is $\mathbb{E}[V]=\mu$. Let $\mathcal{R}_{PP}(F,c):=\mu-c$ denote the seller's maximal expected profit under idealized personalized pricing.

By construction, $\mathcal{R}_{SP}(F,c)\le\mathcal{R}_{XP}(F_{\mathbf{X}V},c)\le\mathcal{R}_{PP}(F,c)$. Given F and c, we define the *value of idealized personalized pricing over single-pricing* as $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)}$. The value of feature-based pricing over single-pricing is 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

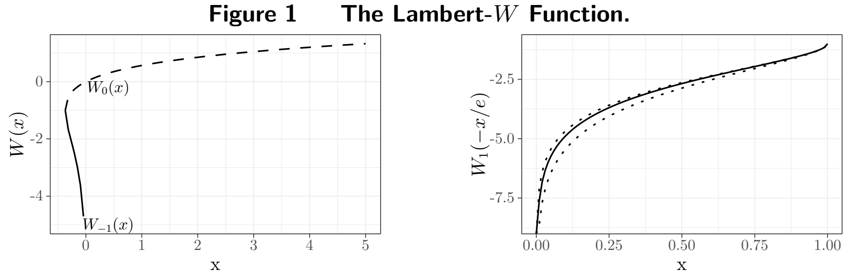
Many of our closed-form bounds involve $W_{-1}(\cdot)$, the negative branch of the *Lambert-W* function. Although the Lambert-W function is pervasive in mathematics, it is less common in the pricing literature. We refer the reader to Corless et al. (1993) for a thorough review of its properties and provide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

Figure 1 The Lambert-W Function.



Note. The left panel shows the two real branches of the Lambert-W function, $W_0(\cdot)$ (dashed black), and $W_{-1}(\cdot)$ (solid). 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, numerically evaluating $W_{-1}(\cdot)$ is no more difficult than evaluating $\log(\cdot)$.

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

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

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

3. The Value of Idealized Personalized Pricing over Single Pricing

In this section, we provide upper and lower bounds on the value of idealized personalized pricing over single pricing using simple statistics and/or shape assumptions of the valuation distribution F. The statistics we shall consider are *scale* (S), *margin* (M), and *coefficient of deviation* (D) defined respectively as

$$S := \frac{\inf\{k \mid F(k) = 1\}}{\mu}, \quad M := 1 - \frac{c}{\mu}, \quad D := \frac{\mathbb{E}[|V - \mu|]}{2\mu}.$$

These three statistics are unit-less and can be thought of as (rescaled) measurements of the maximal valuation, per unit cost, and mean absolute deviation. 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. We stress that S might be infinite when valuations are unbounded, and, indeed, all of our closed-form bounds below will still be valid in this setting. 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. Finally, by construction, $D \in [0, 1]$ since $\mathbb{E}[|V - \mu|] \leq E[|V|] + \mu = 2\mu$ by the triangle inequality. Note D is the (rescaled) mean absolute deviation of V. Mean absolute deviation (MAD) is a common measure of a random variable's dispersion, similar to standard deviation. Intuitively, D measures the overall level of heterogeneity in the market.

Next, 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 $V_c := \frac{1}{\mu - c}(V - c)$ be denoted by F_c. Then,*

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

Electronic copy available at: https://ssrn.com/abstract=3127719

Moreover, if the scale, margin, and coefficient of deviation of F are S, M and D, respectively, then the mean, scale, margin, and coefficient of deviation of F_c (with no marginal cost) are $\mu_c = 1$, $S_c = \frac{S+M-1}{M}$, $M_c = 1$, and $D_c = \frac{D}{M}$, respectively.

We sometimes refer to $V_c \sim F_c$ as the standardized valuation distribution.

3.1. A First Upper Bound

We begin by first providing an upper bound on $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ using only the scale S and margin M. The key to the bound is that $\mathcal{R}_{SP}(F,0)$ directly yields a bound on the tail behavior of F. Indeed, for any price $p > 0$, $p\overline{F}(p) \le \mathcal{R}_{SP}(F,0)$ by definition, and thus $\overline{F}(p) \le \mathcal{R}_{SP}(F,0)/p$. We use this result repeatedly below, terming it the *pricing inequality*:

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

This inequality drives Theorem 1 below.

THEOREM 1 **(Upper Bounding $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ using S and M).** *For any F with scale $S > 1$ and margin $M > 0$, we have*

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

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., $0 \le V \le S$, a.s. Using the tail integral formula for expectation, we have that

$$\begin{aligned}
\mathcal{R}_{PP} &= \int_0^S \overline{F}(x)dx && (2)\\
&\le \mathcal{R}_{SP} + \int_{\mathcal{R}_{SP}}^S \overline{F}(x)dx \qquad (0 \le \mathcal{R}_{SP} \le S) && (3)\\
&\le \mathcal{R}_{SP} + \int_{\mathcal{R}_{SP}}^S \frac{\mathcal{R}_{SP}}{x}dx \qquad \text{(Pricing Inequality)} && (4)\\
&= \mathcal{R}_{SP} + \mathcal{R}_{SP}\log\left(S\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}\right) \qquad (\text{since } \mathcal{R}_{PP} = 1).
\end{aligned}$$

Rearranging this inequality yields

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

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

$$e^{\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}} \le eS\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \iff \frac{1}{eS} \le \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}e^{-\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}} \iff \frac{-1}{eS} \ge -\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}e^{-\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}} \tag{6}$$

Electronic copy available at: https://ssrn.com/abstract=3127719

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

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

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 $W_{-1}(\cdot)$,

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

Next consider a random variable with

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

Observe that F_S has mean 1, since

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

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.

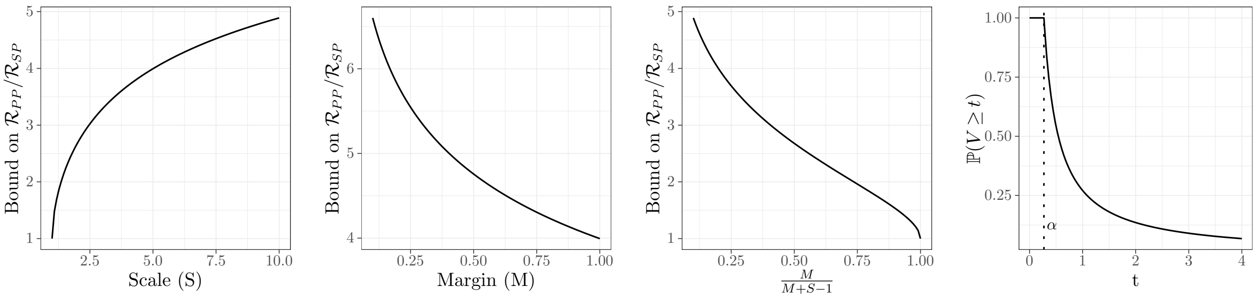
For a general $c>0$ and $\mu \neq 1$, use Lemma 1 to transform to a standardized valuation distribution with $c=0$, $\mu_c=1$, $M_c=1$, and $S_c = \frac{S+M-1}{M}$. Lemma 1 and Eq. (7) then imply that $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} = \frac{\mathcal{R}_{PP}(F_c,0)}{\mathcal{R}_{SP}(F_c,0)} \leq -W_{-1}\left(\frac{-1}{eS_c}\right)$. 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. $\square$

The described tight distribution in the proof 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 rightmost panel of Fig. 2). 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 first three panels of Figure 2, we plot the bound of Theorem 1 as a function of S, M, and the fraction $\frac{M}{S+M-1}$, since the bound only depends on this ratio. Intuitively, as the scale increases, valuations become more dispersed and personalization offers greater

Electronic copy available at: https://ssrn.com/abstract=3127719

Figure 2 Bounds and Tight Distribution from Theorem 1.



Note. The first panel shows the bound in Thm. 1 when $M = 1$ and as S varies from 1 and 10. The second panel shows the bound in Thm. 1 when $S = 5$ and as M varies from 0.1 and 1.0. The third panel shows the bound in Thm. 1 as $\frac{M}{M+S-1}$ varies from 0.1 and 1.0. The fourth panel shows the tight distribution of Thm. 1 when $M = 1$ and $S = 5$.

potential value, as seen in the first panel. On the other hand, increasing the margin with a fixed mean is equivalent to decreasing the cost per unit. As discussed above, under the tight distribution, an optimal single-pricing strategy is to price at $p = \alpha$, which has the same market share as idealized personalized pricing. Thus, in the second 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 Theorem 1. 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 $\overline{F}$ (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* CDF 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. □

3.2. Upper Bound Incorporating the Coefficient of Deviation

A drawback of Theorem 1 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 sharper upper bounds on the value of idealized personalized pricing by incorporating a measure of the market's heterogeneity, i.e., the coefficient of deviation D.

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 larger dispersion in valuations, and, hence, the potential value of personalization to be 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 2 which upper 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 2 **(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 2, we can state Theorem 2, the main upper bound of this section.

THEOREM 2 **(Upper Bounding $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ using S, M, and D)**. *For any F with scale $S > 1$, margin $M > 0$, and coefficient of deviation D, we have the following:*

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

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

Electronic copy available at: https://ssrn.com/abstract=3127719

b) If $\delta_L \le D \le \delta_M$, then

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

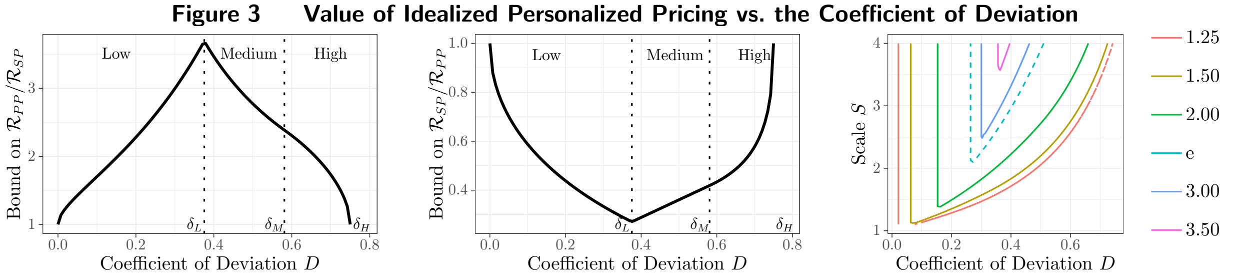
c) If $\delta_M \le D \le \delta_H$, then

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

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 2 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. 3). 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. Recall that since $\mathcal{R}_{PP}$ upperbounds the value of *any* price-discrimination strategy, when D is either very small or very large and the bound is close to 1, Theorem 2 suggests that there is limited benefits to *any* price-discrimination strategy.

Figure 3 Value of Idealized Personalized Pricing vs. the Coefficient of Deviation



Note. The left panel plots the bound from Theorem 2 as a function of D with $S=4$ and $M=1$. The middle panel plots the inverse of this bound, which we note is convex. The right panel shows Theorem 2 as a surface plot, where D ranges over $[0,1]$, and S ranges over $[1.1,4]$. The dashed contour is the uniform bound for MHR distributions, $e \approx 2.718$, from Barlow et al. (1963) and Hartline et al. (2008).

Finally, the maximal point in Fig. 3, at the transition between the low and medium regimes, corresponds to the bound in Theorem 1. Moreover, when S is infinity, $\delta_L = \delta_M = \delta_H = 1$ and the low heterogeneity bound (which does not depend on S) always pertains.

We also observe that our bound in the right panel of Figure 3 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). In summary, although the value of personalized pricing can be large in some settings, our refined analysis based on D characterizes precisely markets necessarily have a low potential value of personalized pricing.

Finally, for simplicity, in Corollary 1 below we use Eq. (1) to rewrite our bounds without the Lambert-W function. The resulting bounds are arguably simpler, but no longer tight.

COROLLARY 1 **(Simpler Upper Bound on $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$)**. *For any F with scale $S>1$, margin $M>0$, and coefficient of deviation D, we have the following:*

a) If $0 \le D \le \delta_L$, then $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \le \frac{1+\sqrt{2\log\left(\frac{1}{1-\frac{D}{M}}\right)}+\log\left(\frac{1}{1-\frac{D}{M}}\right)}{1-\frac{D}{M}}$.

b) If $\delta_L \le D \le \delta_M$, then $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \le \frac{M\log\left(\frac{S+M-1}{M}\right)}{D}$.

c) If $\delta_M \le D \le \delta_H$, then $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \le 1+\sqrt{2\log\left(\frac{S+M-1}{M}(1-\frac{D}{M})\right)}+\log\left(\frac{S+M-1}{M}(1-\frac{D}{M})\right)$.

Single-Pricing Guarantee: An alternative interpretation of Theorem 2 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 2, in the middle panel of Fig. 3. Perhaps surprisingly, the reciprocal is convex as a function of D (our original function was neither convex nor concave). We prove this formally in Lemma 3.

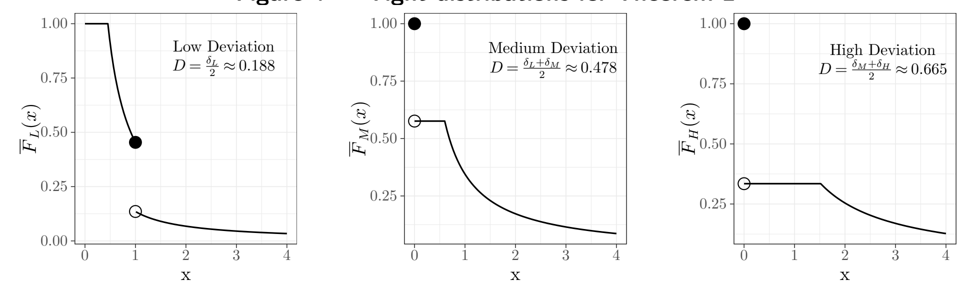
LEMMA 3 **(Convexity of the Single-Pricing Guarantee)**. *For any S, M,and D, let $\alpha(S,M,D)$ denote the reciprocal of the bound on the value of personalized pricing in Theorem 2. Then $\alpha(S,M,D)$ is a convex function in D.*

Tight Distributions: Like Theorem 1, Theorem 2 is a tight bound. The distribution which achieves the bound depends on the regime but is not unique. See Fig. 4 for typical examples and Lemma EC.4 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 2 that any price along the truncated Pareto section is an optimal price for the single-pricing strategy. These results

Electronic copy available at: https://ssrn.com/abstract=3127719

generalize a folklore result that the Pareto distribution represents the worst-case valuation distribution when S and D are unrestricted to the case where these values are known.

Figure 4 **Tight distributions for Theorem 2**



Note. $S = 4$, $\mu = 1$ and $M = 1$. Distribution varies by regime and is non-unique. See Lemma EC.4 for explicit formulas.

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 4.1 that the first limit also provides insight into the performance of certain third-degree price discrimination tactics. These limits are below:

COROLLARY 2 **(Asymptotic Behavior)**. *For any S, M, D, let $1/\alpha(D, M, S)$ denote the bound from Theorem 2. 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 $D \to \delta_H$,

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

3.3. Lower Bounds on the Value of Personalized Pricing

In this subsection, we complement our upper bounds on the value of personalized pricing with closed-form lower bounds. Such lower bounds are helpful in identifying when personalized pricing strategies can guarantee increased revenues. Unfortunately, when only S, M, and D are given, no non-trivial lower bound exists, i.e., no lower bound strictly greater than 1 can be proven. Consider the following two-point distribution in Example 1 where $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} = 1$ for any S, M, and D.

EXAMPLE 1 ($\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} = 1$ FOR BERNOULLI DISTRIBUTIONS). Given S, M, and D, recall that $D \le \frac{M(S-1)}{S+M-1}$ from Lemma Lemma 2. Define the two point distribution

$$V = \begin{cases} 1 - M & \text{with probability } \frac{D}{M} \\ \frac{D(1-M)-M}{D-M} & \text{with probability } 1 - \frac{D}{M}. \end{cases}$$

One can confirm directly that $\mathbb{E}[V]=1$, the margin of V is M and the coefficient of deviation of V is D. Furthermore, $D \leq \frac{M(S-1)}{S+M-1} \implies S \geq \frac{D(1-M)-M}{D-M}$ so that the scale of V is at most S. Finally, one can confirm that pricing at $\frac{D(1-M)-M}{D-M}$ earns a profit of $M=1-c=\mathcal{R}_{PP}(F,c)$. Hence, $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)}=1$. □

To avoid these pathological two-point distributions, we require additional assumptions about the distribution's *shape*. We study two such assumptions below.

DEFINITION 1. A random variable V is *unimodal* with mode m if $\overline{F}(t) := \mathbb{P}(V \geq t)$ is a concave function on $(-\infty, m]$ and convex on (m, ∞).

DEFINITION 2. A random variable V is *symmetric* about point m if $\mathbb{P}(V \in [m-x, m]) = \mathbb{P}(V \in [m, m+x])$ for all $x \geq 0$.

These two definitions generalize the usual notions definitions of unimodality and symmetry for random variables that admit densities to allow for point masses.

We utilize these shape assumptions to prove non-trivial, parametric lower bounds on the value of personalized pricing over single-pricing in Theorem 3. To the best of our knowledge, these bounds are the first results of their kind, proving strict separation between the revenue of idealized personalized pricing and a single price strategy for a general class of distributions based on the level of heterogeneity in the market. The bounds describe markets where one is guaranteed that personalized pricing improves upon single-pricing.

THEOREM 3 **(Lower Bounding $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ using D)**. *Consider a valuation distribution $V \sim F$, with margin $M > 0$ and coefficient of deviation D.*

a) If V is unimodal and symmetric, then $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \geq \frac{1}{1-2\frac{D}{M}}$. Moreover, for every value of $\frac{D}{M}$ there exists a unimodal and symmetric distribution such that this bound is tight.

b) If V is unimodal and

- $0 \leq \frac{D}{M} \leq \frac{1}{3}$, *then* $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \geq \frac{1}{1-\frac{D}{M}}$.
- $\frac{1}{3} \leq \frac{D}{M} \leq 1$, *then* $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)} \geq \frac{8\frac{D}{M}}{\left(1+\frac{D}{M}\right)^2}$.

Moreover, if $\frac{D}{M}=0$, this bound is tight, and, as $\frac{D}{M}$ tends to 1, there exists a family of unimodal valuation distributions such that this bound is asymptotically tight.

Theorem 3 gives optimal (near-optimal), closed-form lower bound on the value of personalized pricing for any symmetric & unimodal (unimodal) distribution in terms of its margin and coefficient of deviation. We prove part (a) of Theorem 3 below. The proof of part (b) is similar, and we relegate it to Appendix A.3 for brevity.

Proof of Theorem 3(a). First suppose $c=0$ and $\mu=1$. Symmetry implies that the mode equals μ (which equals 1) and $1 \leq S \leq 2$. Moreover, by Lemma EC.5 in Appendix A.3, we have $D \leq 0.25$. Now, consider two cases based on the optimal single-price p^*.

Case 1: $p^* > 1$**.** Define the function $\overline{G}(x) = \overline{F}(x)$ for all $x \in (1,2]$, and $\overline{G}(1) := \lim_{t\downarrow 1} \overline{F}(t)$. These functions agree everywhere except perhaps at 1 if V has a point mass at 1. Moreover, by unimodality, $\overline{G}(x)$ is convex on $[1,2]$.

Next, by symmetry about 1, $G(1) = \lim_{t\downarrow 1} \overline{F}(t) \leq \frac{1}{2}$ and since $S \leq 2$, $\overline{G}(2) = \overline{F}(2) = 0$. In particular, this implies $p^* \leq 2$. Thus, writing p^* as a convex combination,

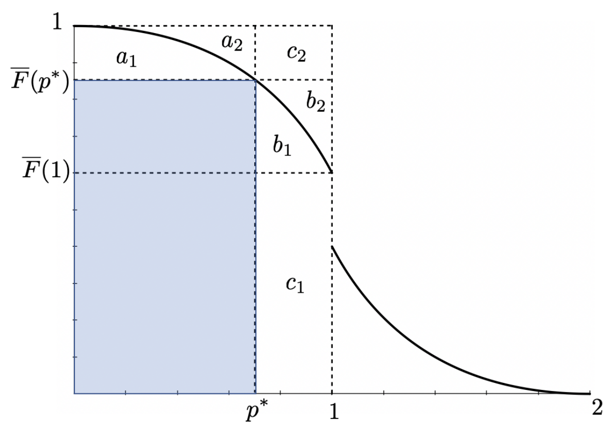
$$\overline{G}(p^*) \;=\; \overline{G}\big((2-p^*)\cdot 1 + (p^*-1)\cdot 2\big) \;\leq\; (2-p^*)\overline{G}(1) + (p^*-1)\overline{G}(2) \;\leq\; (2-p^*)\cdot \frac{1}{2}.$$

Hence,

$$\mathcal{R}_{SP} \;=\; p^*\overline{F}(p^*) \;=\; p^*\overline{G}(p^*) \;\leq\; p^*(2-p^*)/2 \;\leq\; \max_x \, x(2-x)/2 \;=\; \frac{1}{2}.$$

Finally, since $D \leq .25$, $\mathcal{R}_{SP} \leq 1/2 \leq 1-2D$, and $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \frac{1}{1-2D}$.

Figure 5 Geometric Proof of Theorem 3(a).



Note. The revenue of a single pricing using price $p^* < 1$ (shaded rectangle) is depicted relative to the area under a symmetric unimodal cCDF (solid line). The proof relates this rectangle to the area of regions a_1, a_2, b_1, b_2, c_1, and c_2.

Case 2: $p^* \leq 1$**.** Referring to Fig. 5 note that $\mathcal{R}_{SP} = p^*\overline{F}(p^*)$ is the area of the shaded rectangle. Re-express this quantity as the area of the unit-square (dashed rectangle in figure) minus the area of the regions a_1, a_2, b_1, b_2, c_1, and c_2. Formally,

$$\mathcal{R}_{SP} = 1 - \text{Area}(a_1 \cup b_1 \cup c_1) - \text{Area}(a_2 \cup b_2 \cup c_2),$$

because the regions are disjoint.

Next, by unimodality, $\overline{F}$ is concave on $[0,1]$, hence $\text{Area}(a_1) \geq \text{Area}(a_2)$ and $\text{Area}(b_1) \geq \text{Area}(b_2)$. Moreover, by symmetry, $\lim_{t \uparrow 1} \overline{F}(t) \geq \frac{1}{2}$, hence $\text{Area}(c_1) \geq \text{Area}(c_2)$, and, in sum, $\text{Area}(a_1 \cup b_1 \cup c_1) \geq \text{Area}(a_2 \cup b_2 \cup c_2)$. Substituting above shows $\mathcal{R}_{SP} \leq 1 - 2\text{Area}(a_2 \cup b_2 \cup c_2)$. Finally, by Lemma EC.1, $D = 1 - \int_0^1 \overline{F}(x)dx$. Referring to Fig. 5, this shows that $D = \text{Area}(a_2 \cup b_2 \cup c_3)$. Substituting above shows $\mathcal{R}_{SP} \leq 1 - 2D$, which implies $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \frac{1}{1-2D}$.

To show the bound is tight we construct a distribution that is a mixture of a point mass at 1 and a uniform random variable on $[0,2]$, namely,

$$V_0 = \begin{cases} 1 & \text{with probability } 1 - 4D \\ \text{Unif}[0,2] & \text{with probability } 4D \end{cases}$$

By inspection, V_0 is symmetric, unimodal, satisfies $\mathbb{E}[V_0] = 1$ and $\frac{\mathbb{E}[|V_0 - 1|]}{2} = D$ and pricing at 1 earns revenue $1 - 2D$. Hence, for V_0, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} = \frac{1}{1-2D}$. This completes the proof for standardized valuation distributions.

For general c and μ, apply Lemma 1 to transform to a standardized distribution $V_c \sim F_c$. From above, the value of personalized pricing for V_c is at least $\frac{1}{1-2D_c}$. Replace $D_c = D/M$ to prove the bound, and scale V_0 by $(\mu - c)$ and shift by c to form a tight distribution. □

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

As mentioned in the introduction, idealized personalized pricing is unachievable in practice. Here we study a more realistic form of personalized pricing termed feature-based pricing.

4.1. Feature-Based Pricing

In feature-based pricing, the seller predicts the customer valuation V from a set of observed customer features, $\mathbf{X}$. 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 binary and indicates if the customer is a student. More generally, in online retailing settings, sellers often have access to rich contextual information for each customer from their 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.

Formally, let the random variable $\mu(\mathbf{X}) := \mathbb{E}[V \mid \mathbf{X}]$ and define the residual $\epsilon := V - \mathbb{E}[V \mid \mathbf{X}]$. By construction, $\mathbb{E}[\epsilon \mid \mathbf{X}] = 0$ almost surely, i.e., the noise term always has conditional mean 0. More importantly, when $\mathbf{X}$ is very informative for V, we expect ϵ to be "small". In this sense, the size of ϵ measures the degree to which $\boldsymbol{X}$ can be used to predict V. Intuitively, one might think of ϵ as the residual in a non-parametric regression of V on $\boldsymbol{X}$.

A first, perhaps obvious, observation is that given $\mathbf{X}$, it is not optimal to price at $\mathbb{E}[V \big| \mathbf{X}]$. To the contrary, one should price at the optimal price for the conditional distribution $F_{V|\mathbf{X}}$. We encapsulate this observation in Lemma 4.

Lemma 4 **(Relating Feature-Based Pricing and Single-Pricing)**. *For any joint distribution $F_{\mathbf{X}V}$, we have $\mathcal{R}_{XP}(F_{\mathbf{X}V}, c) = \mathbb{E}[\mathcal{R}_{SP}(F_{V|\mathbf{X}}, c)]$.*

The main results of this section are bounds on the ratio between feature-based pricing ($\mathcal{R}_{XP}$) and a single pricing strategy ($\mathcal{R}_{SP}$) that depend explicitly on the degree to which $\boldsymbol{X}$ is informative for V as measured by the size of the residual ϵ (more specifically, $\frac{\mathbb{E}[|\epsilon|]}{2\mu}$). To this end, we first bound the ratio between $\mathcal{R}_{XP}$ and $\mathcal{R}_{PP}$ in terms of the magnitude of the residual noise ϵ. For convenience, we define $D_\epsilon := \frac{\mathbb{E}[|\epsilon|]}{2\mu}$.

Theorem 4 **(Feature-Based Pricing vs. Idealized Personalized Pricing)**. *Suppose that $V = \mu(\mathbf{X}) + \epsilon$ where the residual ϵ is independent of $\mathbf{X}$ and let $D_\epsilon = \frac{\mathbb{E}[|\epsilon|]}{2\mu}$.*

a) Then,

$$\frac{\mathcal{R}_{PP}(F, c)}{\mathcal{R}_{XP}(F_{\mathbf{X}V}, c)} \leq \frac{1}{\alpha\left(S, M, D_\epsilon\right)},$$

where $\alpha(S, M, D)$ denotes the reciprocal of the bound in Theorem 2.

b) If, additionally, ϵ is unimodal and symmetric, then,

$$\frac{\mathcal{R}_{PP}(F, c)}{\mathcal{R}_{XP}(F_{\mathbf{X}V}, c)} \geq \frac{1}{1 - \frac{2D_\epsilon}{M}}.$$

Notice that when $\boldsymbol{X}$ is very informative for V, D_ϵ is small, and thus the first part of Theorem 4 implies $\mathcal{R}_{PP}$ offers limited benefits over $\mathcal{R}_{XP}$. Correspondingly, when $\boldsymbol{X}$ does not contain much information about V, the second part guarantees (idealized) personalized pricing earns significantly more than feature-based pricing under some additional assumptions. We stress that it is common in statistics to assume the noise ϵ is Gaussian, which is significantly stronger than assuming ϵ is unimodal and symmetric.

We leverage Theorem 4 to bound $\frac{\mathcal{R}_{XP}}{\mathcal{R}_{SP}} = \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \cdot \frac{\mathcal{R}_{XP}}{\mathcal{R}_{PP}}$ by bounding the second term.

THEOREM 5 **(Feature-Based Pricing vs. Single Pricing)**. *Suppose $V = \mu(\mathbf{X}) + \epsilon$ with ϵ independent of $\mathbf{X}$. Let $D_\epsilon = \frac{\mathbb{E}[|\epsilon|]}{2\mu}$.*

a) Then,

$$\frac{\mathcal{R}_{XP}(F,c)}{\mathcal{R}_{SP}(F,c)} \geq \frac{1-\frac{D_\epsilon}{M}}{-W_{-1}\left(\frac{\frac{D_\epsilon}{M}-1}{e}\right)} \cdot \frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)}.$$

b) If, additionally, ϵ is unimodal and symmetric, then

$$\frac{\mathcal{R}_{XP}(F,c)}{\mathcal{R}_{SP}(F,c)} \leq \left(1-\frac{2D_\epsilon}{M}\right)\cdot \frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)}.$$

The proof above is immediate. Note we have used the (looser) low-heterogeneity bound of Theorem 2 in place of $\alpha(S,M,D_\epsilon)$. As noted in the proof of Theorem 2, this bound pertains to all D and is strongest when D is small. Since we expect one to be interested in feature-based pricing mostly in settings with relatively informative features $\boldsymbol{X}$, we state the bound with this simpler constant. Moreover, we have used Theorem 3(a) to form the upper bound which requires symmetry of ϵ. With minor modifications, one can instead use Theorem 3(b) which does not require symmetry but increases the constant beyond $\frac{1}{1-\frac{2D_\epsilon}{M}}$.

Intuitively, Theorem 5 decomposes the benefits of feature-based pricing into those stemming from pure price discrimination and those (losses) stemming from prediction error. From a theoretical point of view, this result highlights that the value of personalized pricing ($\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$) is *the* fundamental mathematical quantity for study. Indeed, using Theorem 5, we can plug-in *any* bounds on $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ and obtain corresponding bounds on $\frac{\mathcal{R}_{XP}}{\mathcal{R}_{SP}}$. These include the bounds developed in Section 3 above and the bounds developed in Section 5 below. Although we focus on feature-based pricing in this paper, we also suspect that $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$ may be a primitive "building block" when studying other forms price discrimination.

From a more practical point of view, Theorem 5 allows a seller who is currently using a single-pricing strategy and considering switching to a feature-based pricing strategy to assess the potential benefits of the switch. The key issue is the informativeness (as measured by D_ϵ) of the features $\boldsymbol{X}$ that the seller currently has or hopes to obtain. If these features are not sufficiently informative, the second part of the theorem shows there is little value to the switch. On the other hand, if one intends to collect additional features on the customers, Theorem 5 also indicates how informative those features must be to guarantee a desired fraction of idealized personalized-pricing profits. From Theorem 5(a), we see that to be guaranteed to halve the relative performance gap between personalized pricing and feature-based pricing, one needs to reduce the size of ϵ by a factor of 4. Loosely,

Electronic copy available at: https://ssrn.com/abstract=3127719

this corresponds to collecting features $\boldsymbol{X}$ which allow one to predict V four times more accurately. [2]

5. Bounds Based upon General Moments

In Section 3 we developed upper and lower bounds on the ratio $\frac{\mathcal{R}_{PP}(F,c)}{\mathcal{R}_{SP}(F,c)}$ based upon the coefficient of deviation. Although the coefficient of deviation enjoys properties that make it amenable to closed-form analysis, in principle, any statistic might be used. In this section, we show how to compute upper and lower bounds on the value of personalized pricing over single-pricing for other natural statistics and when imposing various shape constraints (such as unimodality) on F. Furthermore, via Theorem 5, these bounds can naturally be transformed into bounds on the value of feature-based pricing over single-pricing.

Specifically, we assume F satisfies a single moment constraint of the form $\mathbb{E}[h(V)] = \mu_h$ for some known, fixed function $h(\cdot)$ and constant μ_h. For example, when $h(v) = \frac{|v-\mu|}{2\mu}$ and $\mu_h = D$, this constraint requires the coefficient of deviation of F to be D, generalizing our analysis from the previous section. On the other hand, when $h(v) = \frac{(v-\mu)^2}{\mu^2}$ and $\mu_h = C^2$, this constraint requires the coefficient of variation of F to be C (cf. Section 5.5 below). Two other noteworthy examples are i) when $h(v) = -\log(v/\mu)$ and $\mu_h = -\log(B/\mu)$, which constrains the geometric mean of F to be B i.e., $\exp(\mathbb{E}[\log(V)]) = B$ and ii) when $h(v) = \mathbb{I}(v \geq \hat{p}\mu)$ and $\mu_h = q$, which asserts that a fraction q of the market purchases at price $\hat{p}\mu$, e.g., an incumbent price that has been used historically.

Notice that if $\mathbb{E}[h(V)] = \mu_h$, then $\mathbb{E}[\overline{h}(V_c)] = 0$ where $\overline{h}(t) := h((\mu - c)t + c)) - \mu_h$ and V_c is the standardized valuation distribution of Lemma 1. Hence, to bound the value of personalized pricing with a general moment constraint defined by h, it suffices to bound the value of personalized pricing for a standardized valuation distribution satisfying a moment constraint defined by $\overline{h}$. Thus, without loss of generality in what follows, we assume $\mathbb{E}[V] = 1$, $M = 1$ and $\mu_h = 0$. Finally, with some loss of generality, we assume throughout this section that $S < \infty$ as it simplifies many of our formulations.[3]

Numerical examples illustrating each of the below bounds for the coefficient of deviation and the coefficient of variation can be found in Sections 5.4 and 5.5 below.

[2] More specifically, using Corollary 2, we can rewrite Theorem 5(a)) as $\frac{\mathcal{R}_{XP}}{\mathcal{R}_{SP}} \geq \left(1 - \sqrt{2D_\epsilon/M} + o(\sqrt{D_\epsilon/M})\right)\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}$, where we have used the fact that $(1 + \sqrt{2D_\epsilon/M} + o(\sqrt{D_\epsilon/M}))^{-1} = 1 - \sqrt{2D_\epsilon/M} + o(\sqrt{D_\epsilon/M})$. Rearranging shows $\frac{\mathcal{R}_{PP} - \mathcal{R}_{XP}}{\mathcal{R}_{PP}} \leq \sqrt{2D_\epsilon/M} + o(\sqrt{D_\epsilon})$. Hence reducing D_ϵ by a factor of 4 halves the relative performance gap.

[3] The case of $S = \infty$ can be handled with similar techniques, albeit somewhat more tedious calculations.

5.1. Upper Bounds Based upon General Moments

The key idea of our approach is to formulate an optimization problem over probability measures that satisfy the given moment constraint to compute the value of personalized pricing. Similar ideas have been used in the literature to develop generalized Chebyshev inequalities (Bertsimas and Popescu (2005), Popescu (2005)). Consider the problem

$$\begin{aligned} z^* := \inf_{y, d\mathbb{P}_v} \quad & y \\ \text{s.t.} \quad & \int_0^S d\mathbb{P}_v = 1, \quad d\mathbb{P}_v \geq 0, \quad \forall v \in [0,S] \\ & \int_0^S v d\mathbb{P}_v = 1, \quad \int_0^S h(v) d\mathbb{P}_v = 0, \quad y \geq p \int_0^S \mathbb{I}(v \geq p) d\mathbb{P}_v, \quad \forall p \in [0,S]. \end{aligned} \tag{9}$$

The decision variables above are $\mathbb{P}_v$, which represents the distribution of V, and y, which represents the single-pricing profit. The first two constraints ensure that $\mathbb{P}_v$ is a valid probability measure. The next two constraints ensures the mean of $\mathbb{P}_v$ is 1 and $\mathbb{P}_v$ satisfies the moment constraint. Finally, the last (infinite) family of constraints ensures that y is at least the revenue achieved by pricing at p for any $p \in [0,S]$. At optimality, y will equal the optimal single price revenue. Therefore, $1/z^*$ is a tight upper bound on the value of personalized pricing.

Unfortunately, since Problem (9) has both an infinite number of variables $\mathbb{P}_v$ for $v \in [0,S]$ and an infinite number of constraints (indexed by p), it is not clear how to solve it. A first thought might be to discretize Eq. (9) by restricting $\mathbb{P}_v$ to have (fixed) finite, discrete support and only enforcing the semi-infinite constraint on some grid. The resulting value, however, is *not* a valid lower bound on $\mathcal{R}_{SP}$, and, hence, its reciprocal does not upper bound the value of personalized-pricing.

Theorem 6 below provides an alternate approach by discretizing the dual of (9) which does yield a valid bound. See Appendix A.6 for details.

THEOREM 6 **(Upper Bounding VoPP for General Moments)**. *Let F be any valuation distribution with scale S, margin $M = 1$ and mean $\mu = 1$ that satisfies $\mathbb{E}[h(v)] = 0$ for a fixed, known $h(\cdot)$. Let $0 = p_0 < p_1 < \ldots < p_{N-1} < p_N = S$ be a discretization of the interval $[0,S]$ and define*

$$z_N^* := \max_{\theta, \boldsymbol{\lambda}, \mathbf{Q}} \quad \theta + \lambda_1 \tag{10}$$

Electronic copy available at: https://ssrn.com/abstract=3127719

$$s.t. \quad \sum_{j=0}^{N} Q_j = 1, \quad \boldsymbol{Q} \geq \boldsymbol{0}, \quad \theta + \lambda_1 S + \lambda_2 h(S) \leq \sum_{j=0}^{N} p_j Q_j \ ,$$

$$\theta + \lambda_1 v + \lambda_2 h(v) \leq \sum_{j=0}^{k-1} p_j Q_j, \qquad \forall v \in [p_{k-1}, p_k), \quad k = 1, \ldots, N.$$

Then, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq 1/z^* \leq 1/z_N^*$.

Remark 2 (Computational Tractability). Although Problem (10) still contains semi-infinite constraints indexed by v, we argue that this formulation is both theoretically and practically tractable for many $h(\cdot)$. Such formulations are well-studied in the robust optimization literature (Ben-Tal and Nemirovski 2000, Ben-Tal et al. 2015). For certain special classes of $h(\cdot)$, it is often possible to leverage classical results to give an explicit, convex reformulation of these semi-infinite constraints in terms of a finite number of variables and constraints. These reformulations can then be passed directly to off-the-shelf solvers (see below for examples).

For general $h(\cdot)$ that might not admit a simple reformulation, Problem (10) is still computationally tractable if one can efficiently find an optimizer of

$$\max_{v \in [p_{k-1}, p_k]} \lambda_1 v + \lambda_2 h(v) \tag{11}$$

for every k and every λ_1, λ_2. Such a subroutine can be used with constraint-generation to solve the Problem (10) as a linear optimization problem efficiently (see Bertsimas et al. (2016) for details). Fortunately, for many $h(\cdot)$, an optimizer is often available in closed-form (see Propositions 1 and 2 below). □

5.2. Upper Bounds Based upon General Moments under Unimodality

We next compute upper bounds on the value of personalized pricing under a general moment constraint *and* assuming F is unimodal with mode m. As was seen in Section 3.3, such shape constraints can often significantly strengthen the bound.[4]

We adapt our argument in Theorem 6 by leveraging (Popescu 2005, Lemma 4.2) which shows that any m-unimodal distribution can be represented as a (potentially continuous)

[4] We focus on the case of unimodality as it seems most relevant for pricing applications, however, our techniques can be applied to other shape constraints that describe a convex class of distributions, e.g., symmetric distributions, by leveraging the appropriate representation theorems from Popescu (2005).

Electronic copy available at: https://ssrn.com/abstract=3127719

mixture of uniform distributions supported on $[t,m]$ for $t<m$, uniform distributions supported on $[m,t]$ for $t>m$, and a dirac distribution at m. Note that if V follows a uniform distribution on $[t,m]$ or $[m,t]$, then $\mathbb{E}[V]=(t+m)/2$, and

$$\mathbb{P}(V\geq p):=G(p,m,t):=\begin{cases}0 & \text{if } p>\max(m,t)\\ 1 & \text{if } p<\min(m,t)\\ \mathbb{I}(m\geq p) & \text{if } m=t,\\ \frac{\max(m,t)-p}{|m-t|} & \text{otherwise.}\end{cases} \qquad \mathbb{E}[h(V)]:=H(t):=\begin{cases}\frac{1}{m-t}\int_t^m h & \text{if } m\neq t,\\ h(m) & \text{otherwise.}\end{cases} \tag{12}$$

Using these observations, we write the analogue of Eq. (9) when V is m-unimodal as

$$\begin{aligned} z^{*,m}:=\inf_{y,\mathbb{M}} \quad & y \\ \text{s.t.} \quad & \int_0^S d\mathbb{M}_t=1, \quad d\mathbb{M}_t\geq 0, \quad \int_0^S \frac{t+m}{2}d\mathbb{M}_t=1, \quad \int_0^S H(t)d\mathbb{M}_t=0, \\ & y\geq p\int_0^S G(p,m,t)d\mathbb{M}_t, \quad \forall p\in[0,S]. \end{aligned} \tag{13}$$

Here $\mathbb{M}_t$ is the mixing distribution over the requisite uniform distributions (including a dirac distribution at m), and the constraints ensure the corresponding mixture distribution satisfies the requisite moments, similar to Problem 9. Following an argument similar to Theorem 6, we prove an upper bound in Theorem 7 below.

Theorem 7 **(Upper Bounding VoPP under Unimodality)**. *Let F be any unimodal valuation distribution with scale S, margin $M=1$, mean $\mu=1$, and mode m that satisfies $\mathbb{E}[h(v)]=0$ for a fixed, known $h(\cdot)$. Let $0=p_0<p_1<\ldots<p_N=S$, be a discretization of $[0,S]$ such that $p_{j^*}=m$ for some j^*, and let $z_N^{*,m}$ denote the optimal value of*

$$\begin{aligned} \sup_{\theta,\boldsymbol{\lambda},\boldsymbol{Q}} \quad & \theta+\lambda_1(2-m) \\ s.t. \quad & \boldsymbol{Q}\geq 0, \quad \sum_{j=0}^N Q_j=1, \\ & \theta+\lambda_1 m+\lambda_2 h(m) \leq \sum_{j=0}^{j^*} p_jQ_j, \\ & \theta(m-t)+\lambda_1 t(m-t)+\lambda_2\int_t^m h \leq \sum_{j=0}^{k} p_jQ_j(m-t)+\sum_{j=k+1}^{j^*} p_jQ_j(m-p_j), \quad \forall t\in[p_k,p_{k+1}), \quad k=0,\ldots,j^*-1, \\ & \theta(t-m)+\lambda_1 t(t-m)-\lambda_2\int_t^m h \leq \sum_{j=0}^{j^*} p_jQ_j(t-m)+\sum_{j=j^*+1}^{k} p_jQ_j(t-p_j), \quad \forall t\in(p_k,p_{k+1}], \quad k=j^*+1,\ldots,N-1. \end{aligned} \tag{14}$$

Then, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}=1/z^{*,m}\leq 1/z_N^{*,m}$.

Electronic copy available at: https://ssrn.com/abstract=3127719

Remark 3 (Tractability). Similar to Problem 10, we can solve Problem 14 as a linear optimization with constraint generation if we can efficiently identify optimizers of

$$\max_{t\in[p_k,p_{k+1}]} at+bt^2+c\int_t^m h(s)ds \tag{15}$$

for every $a,b,c\in\mathbb{R}$ and k. Notice, optimizers are either one of the end points p_k,p_{k+1} or (by differentiating) a solution to $a+2bt=c\cdot h(t)$. This equation can often be solved either in closed-form or by bisection (see Propositions 1 and 2 below). □

5.3. Lower Bounds Based upon General Moments under Unimodality

We next complement the upper bounds of the previous section by lower bounds. For many moment functions $h(\cdot)$, we can adapt the argument underlying Example 1 to construct a two-point distribution satisfying the given moment constraint for which the value of personalized pricing over single pricing is 1. Consequently, we focus below on the cases where V is unimodal with mode m to derive non-trivial bounds.

Recall any such distribution can be written as a mixture of uniform distributions on $[t,m]$ for $t<m$, uniform distributions on $[m,t]$ for $t>m$ and a dirac distribution at m (c.f. (Popescu 2005, Lemma 4.2)). Then, using the function $G(p,m,t)$ and $H(t)$ defined in Eq. (12), we argue that pricing at p earns at most

$$\begin{aligned} r^m(p) := \sup_{d\mathbb{M}_t} \quad & \int_0^S pG(p,m,t)d\mathbb{M}_t \\ \text{s.t.} \quad & \int_0^S d\mathbb{M}_t = 1, \quad d\mathbb{M}_t \geq 0, \quad \int_0^S \frac{t+m}{2} d\mathbb{M}_t = 1, \quad \int_0^S H(t)d\mathbb{M}_t = 0. \end{aligned}$$

As in Problem (13), $\mathbb{M}_t$ describes the relevant mixing distribution, and the constraints ensure the resulting mixture distribution is a valid distribution, has mean 1, and satisfies the moment constraint $\mathbb{E}[h(V)]=0$. Unlike in the previous section, the objective here maximizes the single pricing profit for pricing at p. Thus, the value of personalized pricing over single pricing satisfies $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \frac{1}{\max_{p\in[0,S]} r(p)}$.

By combining a duality argument with a careful discretization of the prices, we can lower bound the value of personalization. Since the techniques and results are quite similar to those in the previous section, we simply summarize the main result and relegate the precise formulations and proofs to Appendix A.7.

Electronic copy available at: https://ssrn.com/abstract=3127719

Theorem 8 **(Lower Bounding VoPP under Unimodality)**. *Let F be any unimodal valuation distribution with scale S, margin $M = 1$, mean $\mu = 1$, and mode m that satisfies $\mathbb{E}[h(v)] = 0$ for a fixed, known $h(\cdot)$. Fix any $0 < \delta < 1$. Then, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq r_\delta^{*,m}$ where $r_\delta^{*,m}$ is non-increasing in δ and tight in the limit $\delta \to 0$.*

Moreover, $r_\delta^{*,m}$ can be evaluated by solving $N := 1 + \frac{\log(S/\delta)}{\log(1+\delta)}$ optimization problems. Each of these N problems can be solved efficiently as a linear optimization problems if for any $a, b, c, \in \mathbb{R}$ and $[l, u] \subseteq [0, S]$, one can efficiently identify an optimizer for each of

$$\min_{t \in [l,u]} \quad aH(t) + bt \qquad \text{and} \qquad \min_{t \in [l,u]} at^2 + bt + c \int_t^m h. \tag{16}$$

See Propositions 1 and 2 below for examples where (16) can be solved efficiently.

We next illustrate the above theorems with specific examples, showing that each of the above optimization problems is computationally tractable for common functions $h(\cdot)$.

5.4. Bounds Using Shape Constraints and Coefficient of Deviation

We first consider the case where the coefficient of deviation is specified. Let $h(t) = M|t-1|/2 - D$ and V_c be the standardized valuation distribution. Then, $\mathbb{E}[h(V_c)] = 0$ if and only if the coefficient of deviation of V is D. Furthermore, we argue that for this $h(\cdot)$, the optimization problems from Theorems 7 and 8 are tractable and solvable using off-the-shelf software. We summarize the key results in the following Proposition 1, and relegate the precise formulations and details to Appendix A.7.

Proposition 1 **(Tractability of VoPP Optimizations for Coefficient of Deviation)**. *Suppose $h(t) = M|t-1|/2 - D$. Then,*

a) For each $a, b, c \in \mathbb{R}$ and k, an optimizer to Eq. (15) can be found in closed-form.

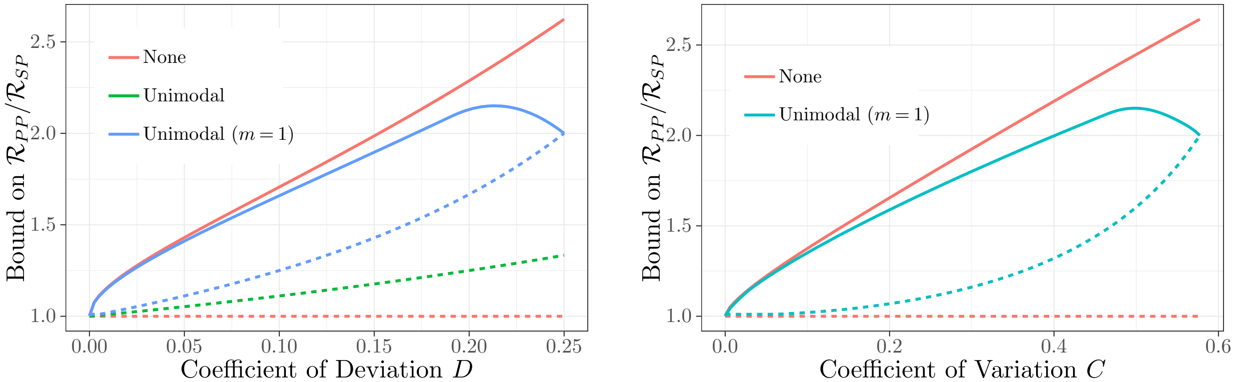
b) For any $a, b, c \in \mathbb{R}$ and $[l, u] \subseteq [0, S]$, optimizers to the two problems in Eq. (16) can be found by bisection search and in closed-form, respectively.

In other words, the problems in Theorems 7 and 8 can each be solved efficiently as a linear optimization with constraint generation.

Remark 4. We note that for the special case of $h(t) = M|t-1|/2 - D$, Theorem 6 is superceded by the closed-form bound Theorem 2, and, hence, omitted above. □

Figure 6 uses Proposition 1 to compare the upper and lower bounds on the value of personalized pricing from Theorems 2, 3, 7 and 8 for a specific example. When $S = 2$, the maximal deviation achievable by any unimodal distribution is only .25 (achieved by a

Figure 6 Upper and Lower Bounds on $\mathcal{R}_{PP}/\mathcal{R}_{SP}$ Given Various Shape Constraints



Note. In both panels $\mu=1$, $M=1$, $S=2$. We plot the upper and lower bounds from Theorems 2, 3, 7 and 8 assuming various shape constraints. Solid lines represent upper bounds while dotted lines represent lower bounds. In the left panel, we vary the coefficient of deviation D. In the right, we vary the coefficient of variation C.

uniform distribution), not 1. Even within this range, however, we see that adding additional a priori shape constraints tightens the bounds substantially for intermediate values of heterogeneity. The gap between the "Unimodal" curve (computed from Theorem 3 Part b)) and the "Unimodal ($m=1$)" (computed from Theorem 8) largely stems from the fact that Theorem 3 is agnostic to both S and the location of the mode. We stress that the two "Unimodal ($m=1$)" curves (blue) are tight by construction.

5.5. Bounds Using Shape Constraints and Coefficient of Variation

As a second example, we consider the case where the coefficient of variation C is specified, i.e., $\sqrt{\mathbb{E}[(V-\mu)^2]}/\mu = C$. Let $h(t) = M^2(t-1)^2 - C^2$, and V_c be the standardized valuation distribution. Then $\mathbb{E}[h(V_c)] = 0$ if and only if the coefficient of variation of V is C. For this $h(\cdot)$, the optimization problems from Theorems 6 to 8 are again tractable and solvable using off-the-shelf software. We summarize below, relegating details to Appendix A.7.

Proposition 2 **(Tractability of VoPP Optimizations for Coefficient of Variation).** *Suppose $h(t) = M^2(t-1)^2 - C^2$. Then,*

a) Problem (10) can be solved explicitly as a (finite) convex second order cone problem.

b) For each $a, b, c \in \mathbb{R}$ and k, an optimizer to Eq. (15) can be found in closed-form.

c) For any $a, b, c, l, u \in \mathbb{R}$, optimizers to the two problems in Eq. (16) can be found by bisection search and in closed-form, respectively.

In other words, the problems in Theorems 6 to 8 are each computationally tractable.

Figure 6 uses Proposition 2 to illustrate our various bounds. Note, again, that the maximal C achievable by a unimodal distribution is significantly smaller than a general distribution. At the maximal C, all unimodal bounds are tight and coincide, with the tight distribution being a uniform distribution.

6. Conclusions

Increasingly rich consumer profiles enable retailers to price discriminate among customers at finer and finer granularity for increased profits. However, such price discrimination strategies entail upfront investment costs in the form of information technology, analytics expertise, and market research. 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 bounds, we are also able to bound the value of certain third-degree price discrimination tactics that more closely mirror current practice. Specifically, we show how to transform our previous bounds on idealized personalized pricing into more practical bounds on the value of feature-based pricing over single price strategies. We also show how to incorporate alternative moment information for sharper bounds by solving tractable optimization problems.

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.

Acknowledgments

We greatly appreciate the feedback from the three anonymous reviews, the AE, and the DE, all of whom helped improve the paper signifantly.

References

Arora, N., X. Dreze, A. Ghose, J. D. Hess, R. Iyengar, B. Jing, Y. Joshi, V. Kumar, N. Lurie, S. Neslin, S. Sajeesh, M. Su, N. Syam, J. Thomas, J. Zhang. 2008. Putting one-to-one marketing to work: Personalization, customization, and choice. *Marketing Letters* **19**(3-4) 305.

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

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

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

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

Ben-Tal, A., D. Den Hertog. 2014. Hidden conic quadratic representation of some nonconvex quadratic optimization problems. *Mathematical Programming* **143**(1-2) 1–29.

Ben-Tal, A., D. Den Hertog, J.-P. Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* **149**(1-2) 265–299.

Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming* **88**(3) 411–424.

Bergemann, D., B. Brooks, S. Morris. 2015. The limits of price discrimination. *The American Economic Review* **105**(3) 921–957.

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

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

Bertsimas, D., I. Dunning, M. Lubin. 2016. Reformulation versus cutting-planes for robust optimization. *Computational Management Science* **13**(2) 195–217.

Bertsimas, D., I. Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* **15**(3) 780–804.

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

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

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

Boyd, S., L. Vandenberghe. 2004. *Convex optimization*. Cambridge University Press.

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

Celis, L. E., G. Lewis, M. Mobius, H. 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, H., M. Hu, G. Perakis. 2017. Distribution-free pricing. *Available at SSRN: https://ssrn.com/abstract=3090002* .

Chen, X., Z. Owen, C. Pixton, D. Simchi-Levi. 2015. A statistical learning approach to personalization in revenue management. *Available at SSRN: https://ssrn.com/abstract=2579462* .

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

Choudhary, V., A. Ghose, T. Mukhopadhyay, U. Rajan. 2005. Personalized pricing and quality differentiation. *Management Science* **51**(7) 1120–1130.

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

Cohen, M. C., N. Z. Leung, K. Panchamgam, G. Perakis, A. Smith. 2017. The impact of linear optimization on promotion planning. *Operations Research* **65**(2) 446–468.

Cohen, M. C., I. Lobel, R. Paes Leme. 2016. Feature-based dynamic pricing. *Available at SSRN: https://ssrn.com/abstract=2737045* .

Cohen, M. C., G. Perakis, R. 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, S. 2016. Welfare-increasing third-degree price discrimination. *The RAND Journal of Economics* **47**(2) 326–340.

D'Innocenzio, A. 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, A. N., M. L. Hamilton. 2017. The power of opaque products in pricing. *Available at SSRN: https://ssrn.com/abstract=3025944* .

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

Huang, J., A. Mani, Z. Wang. 2019. The value of price discrimination in large random networks. *Available at SSRN 3368458* .

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

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

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

Medina, A. M., S. Vassilvitskii. 2017. Revenue optimization with approximate bid predictions. *Advances in Neural Information Processing Systems, 31*. Curran Associates Inc., 1856–1864.

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

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

Obama, Administration. 2016. Big data and differential pricing. *White House Report* .

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

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

Popescu, I. 2005. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Math. Oper. Res.* **30**(3) 632–657. doi:10.1287/moor.1040.0137. URL `http://dx.doi.org/10.1287/moor.1040.0137`.

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

Shapiro, A. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135–165.

Shih, J., C. Mai, J. Liu. 1988. A general analysis of the output effect under third-degree price discrimination. *The Economic Journal* **98**(389) 149–158.

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

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

Tuttle, B. 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/`.

Van der Vaart, A. W. 2000. *Asymptotic Statistics*, vol. 3. Cambridge University Press.

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

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

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

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Online Appendix: The Value of Personalized Pricing

Appendix A: Omitted Proofs

A.1. Proof of Theorem 2

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

Lemma EC.1 **(Integral Representations of D).** *For any F with scale S and margin M, the coefficient of deviation D satisfies*

$$D \;=\; \int_M^{S+M-1} \overline{F}(\mu x + c)dx \;=\; \int_0^M 1 - \overline{F}(\mu x + c)dx. \tag{EC.1}$$

We now prove Theorem 2. 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 Theorem 1. Starting with the second identity of Lemma EC.1,

$$D = \int_0^1 1 - \overline{F}(x)dx \geq \int_0^{\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}} 0 \; dx + \int_{\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}}^1 1 - \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{1}{x} \; dx, \tag{EC.2}$$

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,

$$D \geq \left(1 - \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}\right) + \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \log\left(\frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}}\right). \tag{EC.3}$$

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

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

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

$$W_{-1}\left(\frac{D-1}{e}\right) \leq \log(\alpha) - 1 \quad\iff\quad e \cdot e^{W_{-1}\left(\frac{D-1}{e}\right)} \leq \alpha \quad\iff\quad \frac{W_{-1}\left(\frac{D-1}{e}\right)}{D-1} \geq \frac{1}{\alpha} \tag{EC.4}$$

$$\iff \quad \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{W_{-1}\left(\frac{D-1}{e}\right)}{D-1}, \tag{EC.5}$$

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. (EC.5) 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 EC.1 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,

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{\log(S)}{D}. \tag{EC.6}$$

Again, Eq. (EC.6) 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 next subsection and only state the result in Lemma EC.2 below.

Electronic copy available at: https://ssrn.com/abstract=3127719

Lemma EC.2 **(High Heterogeneity Bound when $c=0$ and $\mu=1$).** *If $D>\delta_M$, then,*

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq -W_{-1}\left(\frac{-1}{eS(1-D)}\right). \tag{EC.7}$$

To summarize, Eqs. (EC.5) and (EC.6) hold for all $0 \leq D \leq \delta_H$ and Eq. (EC.7) 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 EC.3 **(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 EC.3 is

- When $D \in [0,\delta_L]$, Eq. (EC.5) dominates Eq. (EC.6).
- When $D \in (\delta_L,\delta_M]$, Eq. (EC.6) dominates Eq. (EC.5).
- When $D \in (\delta_M,\delta_H]$, Eq. (EC.7) dominates Eqs. (EC.5) and (EC.6).

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. (EC.5) to (EC.7) using the new S_c, M_c and D_c. 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 Theorem 1. 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 Theorem 1 is that the integral representations of D in the proof of Theorem 2 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. We defer the details to Lemma EC.4 in the appendix for brevity. $\square$

A.2. Omitted Details, Proofs, and Lemmas for Theorem 2

We now provide proofs for the lemmas necessary to complete the proof of Theorem 2.

Proof of Lemma 1. First note the profit from personalized pricing under valuation distribution F is $\mathcal{R}_{PP}(F,c)=\mathbb{E}[V]-c=\mu-c$ and under F_c is $\mathcal{R}_{PP}(F_c,0)=\mathbb{E}[\frac{1}{\mu-c}(V-c)]-0=1$. 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

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

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

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

and

$$D_c=\frac{\mathbb{E}[|V_c-\mu_c|]}{2\mu_c}=\frac{\mathbb{E}\left[|\frac{V-c}{\mu-c}-1|\right]}{2}=\frac{\mathbb{E}\left[|V-c-(\mu-c)|\right]/\mu}{2(\mu-c)/\mu}=\frac{D}{M}.$$

This completes the proof. $\square$

Proof of Lemma 2. Consider the case when $c=0$ and $\mu=1$, which implies that $M=1$. We first prove that $D\leq\delta_H$ 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

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

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

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

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:

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

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

Electronic copy available at: https://ssrn.com/abstract=3127719

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

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

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

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

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 $S\ge1$, $0\le\log(S)\le S-1$, 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 $x\ge0$. 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 $S_c=\frac{S+M-1}{M}$. $\square$

Proof of Lemma EC.1. Let $V\sim F$ and note,

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

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

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

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,

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

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

Proof of Lemma EC.2. We follow the same strategy as previous two regimes 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

$$\mathbb{P}(V \geq t) \leq 1 - D \quad \forall t \in (1, S) \tag{EC.8}$$

To prove the claim, note that $D = E[(1-V)^+] \leq \mathbb{P}(V \leq 1)$, where the equality is Lemma EC.1 and the inequality uses $(1-V)^+ \leq 1$. Rearranging proves $\mathbb{P}(V \geq 1) \leq 1 - D$, which in turn implies Eq. (EC.8).

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

$$\begin{aligned} D &= \int_1^S \mathbb{P}(V > t) dt && \text{(Lemma EC.1)} \\ &= \int_1^{t_0} \mathbb{P}(V > t) + \int_{t_0}^S \mathbb{P}(V > t) dt \\ &\leq \int_1^{t_0} (1 - D) dt + \int_{t_0}^S \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \frac{dt}{t} && \text{(Eq. (EC.8) and Pricing Inequality)} \\ &= (t_0 - 1)(1 - D) + \frac{\mathcal{R}_{SP}}{\mathcal{R}_{PP}} \log\left(\frac{S}{t_0}\right) && \text{(EC.9)} \end{aligned}$$

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. (EC.6) $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq \frac{\log(S)}{D}$ 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. (EC.9) 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,

$$\begin{aligned} 1 \leq \alpha + \alpha \log\left(\frac{S(1-D)}{\alpha}\right) &\iff 1 \leq \alpha(1 + \log(S(1-D)) - \log(\alpha)) \\ &\iff -1 \geq \alpha(\log(\alpha) - \log(eS(1-D))) \end{aligned} \tag{EC.10}$$

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

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

Electronic copy available at: https://ssrn.com/abstract=3127719

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

$$W_{-1}\left(\frac{-1}{eS(1-D)}\right) \leq \log(\alpha) - \log(eS(1-D)) \iff \alpha \geq eS(1-D)\cdot e^{W_{-1}\left(\frac{-1}{eS(1-D)}\right)} \tag{EC.11}$$
$$\iff \frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \leq -\frac{-1}{eS(1-D)} e^{-W_{-1}\left(\frac{-1}{eS(1-D)}\right)}.$$

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 EC.3. 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 (cf. Fig. 1), 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,

$$\begin{aligned}
-\frac{W_{-1}\left(-\frac{1-D}{e}\right)}{1-D} = \frac{\log(S)}{D} &\iff W_{-1}\left(-\frac{1-D}{e}\right) = \log\left(S^{\frac{D-1}{D}}\right) \\
&\iff -\frac{1-D}{e} = \log\left(S^{\frac{D-1}{D}}\right)\cdot\exp\left(\log\left(S^{\frac{D-1}{D}}\right)\right) && \text{(definition of Lambert-}W\text{)} \\
&\iff -\frac{1}{eS} = S^{\frac{-1}{D}}\cdot\frac{-\log(S)}{D} && \text{(simplifying)} \\
&\iff -\frac{1}{eS} = \exp\left(-\frac{\log(S)}{D}\right)\cdot\frac{-\log(S)}{D} && \left(\text{using } S^{\frac{-1}{D}} = \exp\left(-\frac{\log(S)}{D}\right)\right) \\
&\iff W_{-1}\left(-\frac{1}{eS}\right) = -\frac{\log(S)}{D} && \text{(Applying } W_{-1}(\cdot)\text{)} \\
&\iff D = -\frac{\log(S)}{W_{-1}\left(-\frac{1}{eS}\right)} = \delta_L.
\end{aligned}$$

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 2 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 EC.4 **(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

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

and this random variable has scale S, coefficient of deviation D, and mean 1 and satisfies Eq. (EC.5) *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

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

and this random variable has scale S, coefficient of deviation D, and mean 1 and satisfies Eq. (EC.6) *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

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

and this random variable has scale S, coefficient of deviation D, and mean 1 and satisfies Eq. (EC.7) *with equality.*

Proof of Lemma EC.4. 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. (EC.5) to (EC.7) 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. (EC.4) 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 $\alpha_L \ge D/\log(S) \iff 1/\alpha_L \le \log(S)/D$. This inequality follows directly from Lemma EC.3 since $D\in[0,\delta_L]$,

Electronic copy available at: https://ssrn.com/abstract=3127719

and the left-hand side is low-heterogeneity bound while the right side is the medium heterogeneity bound. This proves $\overline{F}_L$ is valid.

Next, write

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

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

$$\int_0^1 1-\overline{F}_L(x)dx = \int_0^{\alpha_L} 0dx + \int_{\alpha_L}^1 1-\frac{\alpha_L}{x}\, dx = 1-\alpha_L+\alpha_L\log(\alpha_L) = D, \tag{EC.12}$$

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. (EC.5) with equality.

b) (Medium Heterogeneity) To prove that $\overline{F}_M$ is a valid cCDF, it suffices to show that $eS^{1-\frac{1}{D}}\le S$, which is equivalent to $1\ge\frac{D}{\log(S)}$. Rewrite this last inequality as $\frac{1}{\alpha_M}\ge 1$, and recall from Step 1 of the proof of Theorem 2 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 EC.1 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 $\alpha_H/(1-D)\le S$. Note that by Lemma EC.2, $1/\alpha_H$ 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 $\frac{1}{S(1-D)}\le 1$ which implies $S(1-D)\ge 1$. Thus, $\alpha_H\le 1\le S(1-D)$ which implies that $\alpha_H/(1-D)\le S$ and that $\overline{F}_H$ is a valid cCDF.

Next write,

$$\begin{aligned}\int_0^S \overline{F}(x)dx &= \int_0^{\frac{\alpha_H}{1-D}}(1-D)dx + \int_{\frac{\alpha_H}{1-D}}^S \frac{\alpha_H}{x}dx\\ &= \alpha_H + \alpha_H\log\left(\frac{S}{\alpha_H}(1-D)\right).\end{aligned} \tag{EC.13}$$

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. (EC.11) and follow the implications backwards to Eq. (EC.10), 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 EC.3. Now compute

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

whereby $\overline{F}_H$ has coefficient of deviation D by Lemma EC.1.

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

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. $\square$

A.3. Proof of Theorem 3

Part (a) of the theorem was proven in the main text, except for the following lemma:

LEMMA EC.5 **(Maximum Deviation for Symmetric, Unimodal Distributions)**. *Suppose $V \sim F$ is symmetric, unimodal and supported on $[0,S]$ with mean $\mu = 1$. Then the mean absolute deviation of V is at most $\frac{1}{4}$. Moreover, this bound is tight for uniform random variable on $[0,2]$.*

Proof. Note by unimodality, V may have at most one point mass, located at 1. Define the function $G(x) = F(x)$ for $x \in [0,1)$ and $G(1) \equiv \lim_{t\uparrow 1} F(x)$. Note, since V is unimodal, $F(x)$ and $G(x)$ are convex on $[0,1]$.

Now, by Lemma EC.1,

$$D = \int_0^1 F(x)dx = \int_0^1 G(x)dx,$$

since the two functions differ only at one point. Then, by convexity

$$D \leq \int_0^1 xG(0) + (1-x)G(1)dx = G(1)\int_0^1 (1-x)dx = \frac{1}{2}G(1),$$

where the first equality uses $G(0) = F(0) = 0$. Finally, by symmetry, $G(1) := \lim_{t\uparrow 1} F(x) \leq \frac{1}{2}$, whereby $D \leq .25$. The tightness for the uniform is immediate. $\square$

Next we prove part (b).

Proof of Theorem 3(b). First consider a standardized valuation distribution where $c = 0$ and $\mu = 1$. Fix F, D, let m be the mode of F, and suppose p^* is the revenue maximizing single price. The proof will proceed in four cases depending on the sizes of m and p^*.

(Case 1: $m \geq 1$, $p^* \leq 1$) By Lemma EC.1, $1 - D = \int_0^1 \overline{F}(x)dx$. Thus,

$$\mathcal{R}_{SP} = p^*\overline{F}(p^*) \leq \int_0^{p^*} \overline{F}(x)dx \leq \int_0^1 \overline{F}(x)dx = 1 - D,$$

Electronic copy available at: https://ssrn.com/abstract=3509952

where the first inequality follows since $\overline{F}$ is decreasing and the second inequality follows because $p^* \le 1$. This implies $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \ge \frac{1}{1-D}$.

(Case 2: $m \ge 1$, $p^* > 1$) Since $m \ge 1$, $\overline{F}(x)$ is concave on $[0,1]$. Hence, for any $x \in [0,1)$, $\overline{F}(x) \ge (1-x)\overline{F}(0) + x\overline{F}(1) = (1-x) + x\overline{F}(1) = 1 - x(1-\overline{F}(1))$.

Thus, by Lemma EC.1,

$$D \;=\; 1 - \int_0^1 \overline{F}(x)dx \;\le\; 1 - \int_0^1 \big(1-(1-\overline{F}(1))x\big)dx \;=\; \frac{1-\overline{F}(1)}{2}$$

and, hence,

$$\overline{F}(1) \;\le\; 1-2D. \tag{EC.14}$$

Now since $p^* > 1$,

$$\begin{aligned}
\mathcal{R}_{SP} &= \overline{F}(p^*) + (p^*-1)\overline{F}(p^*) \\
&= \overline{F}(p^*) + (p^*-1)\overline{F}(p^*) \\
&\le\; \overline{F}(p^*) + \int_1^{p^*} \overline{F}(x)dx && (\overline{F}(x) \text{ is decreasing}) \\
&\le\; \overline{F}(1) + D && (p^* > 1 \text{ and Lemma EC.1 }) \\
&\le (1-2D) + D && (Eq.\ \text{(EC.14)}) \\
&= 1-D.
\end{aligned}$$

Thus in this case $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \ge \frac{1}{1-D}$.

(Case 3: $m \le 1$, $p^* \le m$) Much like in Case 1, since $p^* \le m \le 1$ it follows that

$$\mathcal{R}_{SP}(F) \;=\; p^*\overline{F}(p^*) \;\le\; \int_0^{p^*} \overline{F}(x)dx \;\le\; \int_0^1 \overline{F}(x) = 1-D,$$

which implies $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \ge \frac{1}{1-D}$.

(Case 4: $m \le 1$, $p^* > m$) Let $l(x) = \overline{F}(p^*) - f(p^*)\,(x-p^*)$ be the tangent line of $\overline{F}(x)$ at p^*. This line equals 0 at $p^* + \frac{\overline{F}(p^*)}{f(p^*)}$. Since p^* is an optimal price, it satisfies the first order condition $\frac{d}{dp}p\overline{F}(p) = \overline{F}(p) - pf(p) = 0$. Thus $\frac{\overline{F}(p^*)}{f(p^*)} = p^*$ and the root of $l(x)$ is actually $p^* + \frac{\overline{F}(p^*)}{f(p^*)} = 2p^*$.

Thus, $l(x)$ passes through the points $\{(m, l(m)), (2p^*, 0)\}$, and we may equivalently rewrite $l(x) = \frac{2l(m)p^* - l(m)x}{2p^*-m}$. Hence, we also have the identity $\overline{F}(p^*) = l(p^*) = \frac{l(m)p^*}{2p^*-m}$.

Now define the parameter $\lambda := \int_0^m \overline{F}(x)dx$. The proof will proceed in two additional sub-cases depending on the size of λ.

(Sub-case 4(a): $\lambda \ge \frac{2}{3}$) Notice, because $p^* > m$, we have $p^* < 2p^* - m$, which implies that $\frac{(p^*)^2}{2p^*-m} < 2p^* - m$. Thus, we can upper bound $\mathcal{R}_{SP}$ by

$$\begin{aligned}
\mathcal{R}_{SP} \;=\; p^*\overline{F}(p^*) &\;=\; \frac{l(m)}{2p^*-m}\,(p^*)^2 \\
&\;\le\; l(m)(2p^*-m) \\
&\;=\; 2\int_m^{2p^*} l(x)dx && \left(\int_m^{2p^*} l(x)dx = l(m)\frac{2p^*-m}{2}\right) \\
&\;\le\; 2\int_m^{\infty} \overline{F}(x)dx && (l(x) \le \overline{F}(x) \text{ for } x \in (m,\infty)) \\
&\;=\; 2(1-\lambda).
\end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=3127719

Finally, since $\lambda \geq 2/3$, it follows that $\mathcal{R}_{SP} \leq 2(1-\lambda) \leq \frac{2}{3} \leq \lambda$. Thus, in this subcase $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \frac{1}{\lambda}$. Since $m \leq 1$, $\lambda \leq \int_0^1 \overline{F}(x)dx = 1-D$, and it follows that $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \frac{1}{1-D}$.

(Sub-case 4(b): $\lambda \leq \frac{2}{3}$**)**

Write $\mathcal{R}_{SP}(F)$ as the sum before the mode and after the mode

$$\mathcal{R}_{SP}(F) = m\overline{F}(p^*) + (p^* - m)\,\overline{F}(p^*). \tag{EC.15}$$

The first term on the right hand side of Eq. (EC.15) is bounded by

$$\begin{aligned}
m\overline{F}(p^*) &= m\overline{F}(m)\frac{m\overline{F}(p^*)}{m\overline{F}(m)} \\
&\leq \lambda \frac{m\overline{F}(p^*)}{m\overline{F}(m)} && \left(\text{since } \overline{F}(x) \text{ is decreasing } \implies m\overline{F}(m) \leq \int_0^m \overline{F}(x)dx\right) \\
&\leq \lambda \frac{l(p^*)}{l(m)} && \big(\text{ since } l(m) \leq \overline{F}(m) \text{ and } l(p^*) = \overline{F}(p^*)\big) \\
&= \lambda \frac{p^*}{2p^* - m},
\end{aligned} \tag{EC.16}$$

The second term on the right hand side of Eq. (EC.15) is bounded by

$$\begin{aligned}
(p^* - m)\,\overline{F}(p^*) &= l(m)\left(p^* - \frac{m}{2}\right)\frac{(p^* - m)\,\overline{F}(p^*)}{l(m)(p^* - \frac{m}{2})} \\
&= \left(\int_m^{2p^*} l(x)dx\right)\frac{(p^* - m)\,\overline{F}(p^*)}{l(m)(p^* - \frac{m}{2})} && \left(l(m)(p^* - \frac{m}{2}) = \int_m^{2p^*} l(x)dx\right) \\
&\leq (1-\lambda)\frac{(p^* - m)\,\overline{F}(p^*)}{l(m)(p^* - \frac{m}{2})} && \left(\int_m^{2p^*} l(x) \leq \int_m^{\infty} \overline{F}(x)dx = 1-\lambda\right) \\
&= (1-\lambda)\frac{2p^*\,(p^* - m)}{(2p^* - m)^2}. && \left(\overline{F}(p^*) = \frac{l(m)p^*}{2p^* - m}\right)
\end{aligned} \tag{EC.17}$$

Thus we can upper bound $\mathcal{R}_{SP}$ by combining Eqs. (EC.16) and (EC.17)

$$\mathcal{R}_{SP}(F) \ \leq \ \lambda\frac{p^*}{2p^* - m} + (1-\lambda)\,\frac{2p^*\,(p^* - m)}{(2p^* - m)^2} \ = \ \frac{p^*(2p^* + m(\lambda - 2))}{(2p^* - m)^2} \ \leq \ \max_{p \geq m} \frac{p(2p + m(\lambda - 2))}{(2p - m)^2}$$

This last optimization problem is differentiable in p. As $p \to \infty$, the objective tends to $\frac{1}{2}$. At $p = m$, the objective becomes λ. There is one critical point obtained by differentiation at $p = \frac{m(2-\lambda)}{2\lambda} > m$ since $\lambda \leq 2/3$. At the critical point, the objective is $\frac{(2-\lambda)^2}{8(1-\lambda)}$. For $0 \leq \lambda \leq 2/3$, this value always exceeds λ and $\frac{1}{2}$, and hence this is the optimum.

Thus in this subcase $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \frac{8(1-\lambda)}{(2-\lambda)^2}$. Further, since $m \leq 1$, $\lambda \leq \int_0^1 \overline{F}(x)dx = 1-D$ and it follows that $\frac{8D}{(1+D)^2}$.

Combining all cases and sub-cases gives $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \geq \min\{\frac{1}{1-D}, \frac{8D}{(1+D)^2}\}$ which yields the desired bound.

The tightness at $D = 0$ is immediate since the only feasible distribution is a point mass at m and $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} = 1$.

To prove the asymptotic tightness as $D \to 1$, we construct a family of unimodal distributions $\{V_\delta\}$ indexed by δ such that as $\delta \to 0$, the coefficient of deviation of V_δ tends to 1, and the value of personalized pricing tends to 2. Namely,

$$V_\delta = \begin{cases} \text{Unif}[0, \delta] & \text{with probability } 1 - \delta \\ \text{Unif}[\delta, \frac{2}{\delta} - 1] & \text{with probability } \delta \end{cases}$$

Electronic copy available at: https://ssrn.com/abstract=3127719

By inspection, each distribution in this family is unimodal with mode δ and $\mathbb{E}[V]=1$. To see the deviation tends to 1 as δ tends to 0 consider the lower uniform component of V_δ i.e. $\lim_{\delta\to 0^+}\mathbb{E}[|V_\delta-1|]\geq \lim_{\delta\to 0^+}(1-\delta)(1-\delta/2)=1$. Furthermore, pricing at $\frac{1}{\delta}-\frac{1-\delta}{2}$ earns revenue $(\frac{1}{\delta}-\frac{1-\delta}{2})\mathbb{P}\left(V_\delta\geq \frac{1}{\delta}-\frac{1-\delta}{2}\right)=(\frac{1}{\delta}-\frac{1-\delta}{2})\frac{\delta}{2}$. Taking the limit as δ tends to 0 yields revenue $1/2$ and thus value of personalized pricing of 2 matching the above lower bound. $\square$

A.4. Other Omitted Proofs from Section 3

Proof of Lemma 3. 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 2, 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

$$\begin{aligned}\tilde{D}&=\frac{1}{2\mu}\left(\mathbb{E}\Big[|XV_1+(1-X)V_2-\mu|\Big]\right)\\&=\mathbb{P}(X=1)\cdot\frac{1}{2\mu}\mathbb{E}\Big[|V_1-\mu|\Big]+\mathbb{P}(X=0)\cdot\frac{1}{2\mu}\mathbb{E}\Big[|V_2-\mu|\Big]\\&=tD_1+(1-t)D_2.\end{aligned}\tag{EC.18}$$

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 2. The last equality follows from Eq. (EC.18). $\square$

Proof of Corollary 2. Note that Eq. (1) shows that

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

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

Electronic copy available at: https://ssrn.com/abstract=3127719

A.5. Omitted Proofs from Section 4

Proof of Theorem 4.

Part a) Using Lemma 4, write $\mathcal{R}_{XP} = \max_{p(\cdot)} \mathbb{E}[p(\mathbf{X})\mathbb{I}(\mu(\mathbf{X}) + \epsilon \geq p(\mathbf{X})]$, where the maximization is taken over all (measurable) functions of the features representing the pricing policy. We lower bound this quantity by constructing a feasible pricing policy. Let $p_0 \in \arg\max_{p\geq 0} \; p\mathbb{P}(\mu + \epsilon \geq p)$, where $\mu = \mathbb{E}[V]$. We consider the feasible pricing policy that offers price $p_0 + \mu(\mathbf{X}) - \mu$ to a customer with features $\mathbf{X}$. Then,

$$\begin{aligned}
\mathcal{R}_{XP} &\geq \mathbb{E}\left[(p_0 + \mu(\mathbf{X}) - \mu)\mathbb{I}(\mu(\mathbf{X}) + \epsilon \geq p_0 + \mu(\mathbf{X}) - \mu)\right] \\
&= \mathbb{E}\left[(p_0 + \mu(\mathbf{X}) - \mu)\mathbb{I}(\mu + \epsilon \geq p_0)\right] \\
&= \mathbb{E}\left[p_0\mathbb{I}(\mu + \epsilon \geq p_0)\right] + \mathbb{E}\left[\mu(\mathbf{X}) - \mu)\mathbb{I}(\mu + \epsilon \geq p_0)\right].
\end{aligned}$$

The first expectation equals $\mathcal{R}_{SP}(F_{\mu+\epsilon}, c)$ by choice of p_0. By independence, the second expectation is $\mathbb{E}\left[\mu(\mathbf{X}) - \mu)\right]\mathbb{P}(\mu + \epsilon \geq p_0) = 0$ since $\mathbb{E}[\mu(\mathbf{X})] = \mu$. Thus, we have shown $\mathcal{R}_{XP}(F, c) \geq \mathcal{R}_{SP}(F_{\mu+\epsilon}, c)$.

Finally, applying Theorem 2 to the random variable $\mu + \epsilon$, we can bound $\mathcal{R}_{SP}(F_{\mu+\epsilon}, c) \;\geq\; (\mu - c) \cdot \alpha(S_{\mu+\epsilon}, M, D_\epsilon)$, where $S_{\mu+\epsilon}$ is the scale of $\mu + \epsilon$. Notice, by independence of ϵ and $\mathbf{X}$, $S_{\mu+\epsilon} \leq S$. Hence we further lower bound this quantity by $(\mu - c)\alpha(S, M, D_\epsilon)$ to complete the first part.

Part b) Write

$$\mathcal{R}_{XP}(F, c) = \mathbb{E}[\mathcal{R}_{SP}(F_{V|\mathbf{X}}, c)] = \int_c^\infty \mathcal{R}_{SP}(F_{t+\epsilon}, c) f_{\mu(\mathbf{X})}(t)dt \tag{EC.19}$$

where we have used the fact that $F_{V|\mathbf{X}} = F_{\mu(X)+\epsilon}$ because $\mathbf{X}$ and ϵ are independent. Now applying Theorem 3(a) to the random variable $t + \epsilon$ yields,

$$\mathcal{R}_{SP}(F_{t+\epsilon}, c) \;\leq\; (t - c)\left(1 - 2\frac{\frac{\mathbb{E}[|\epsilon|]}{2t}}{\frac{t-c}{t}}\right) \;=\; (t - c)\left(1 - \frac{\mathbb{E}[|\epsilon|]}{(t - c)}\right) \;=\; t - c - \mathbb{E}[|\epsilon|]$$

. Substituting into the integral above shows

$$\mathcal{R}_{XP}(F, c) \;\leq\; \int_c^\infty (t - c - \mathbb{E}[|\epsilon|])\, f_{\mu(\mathbf{X})}(t)dt \;=\; (\mu - c) - \mathbb{E}[|\epsilon|] \;=\; (\mu - c) \cdot \left(1 - \frac{D_\epsilon}{M}\right).$$

Noting $\mathcal{R}_{PP} = \mu - c$ and rearranging completes the proof. □

A.6. Omitted Proofs from Sections 5.1 and 5.2.

Proof of Theorem 6. Following Shapiro (2001), the dual to Eq. (9) is

$$\begin{aligned}
\sup_{\theta, \boldsymbol{\lambda}, d\mathbb{Q}_p} \quad & \theta + \lambda_1 \\
\text{s.t.} \quad & \int_0^S d\mathbb{Q}_p = 1, \quad d\mathbb{Q}_p \geq 0, \\
& \theta + \lambda_1 v + \lambda_2 h(v) - \int_0^S p\mathbb{I}(v \geq p)d\mathbb{Q}_p \leq 0, \quad \forall v \in [0, S].
\end{aligned} \tag{EC.20}$$

Here, $\mathbb{Q}_p$ is a probability measure defined on $p \in [0, S]$. By weak-duality, any feasible solution to problem (EC.20) yields a valid lower bound to $\mathcal{R}_{SP}$. To form such a feasible solution to (EC.20), we constrain $\mathbb{Q}_p$ to

Electronic copy available at: https://ssrn.com/abstract=3127719

be supported only on $\{p_0,\dots,p_N\}$ and denote the corresponding point masses as $Q_0,Q_1,\dots,Q_N$. Then, the value of (EC.20) is at least

$$\begin{aligned} z_N^* := \max_{\theta,\boldsymbol{\lambda},\mathbf{Q}} \quad & \theta+\lambda_1 \\ \text{s.t.} \quad & \sum_{j=0}^{N} Q_j = 1, \qquad \boldsymbol{Q}\geq \mathbf{0} \\ & \theta+\lambda_1 v+\lambda_2 h(v)-\sum_{j=0}^{N} p_j \mathbb{I}(v\geq p_j)Q_j \leq 0, \qquad \forall v\in[0,S]\ . \end{aligned} \tag{EC.21}$$

Notice that the sum of indicators in Eq. (EC.21) is constant over $v\in[p_{k-1},p_k)$. Thus, we can rewrite this constraint of Eq. (EC.21) as $N+1$ separate sets of constraints:

$$\begin{aligned} &\theta+\lambda_1 v+\lambda_2 h(v)\leq \sum_{j=0}^{k-1} p_j Q_j, \quad \forall v\in[p_{k-1},p_k),\ k=1,\dots,N, \\ &\theta+\lambda_1 S+\lambda_2 h(S)\leq \sum_{j=0}^{N} p_j Q_j\ . \end{aligned}$$

Replacing Eq. (EC.21) with these $N+1$ sets of constraints completes the proof. □

Proof of Theorem 7. Using the fact that $\int_0^S d\mathbb{M}_t = 1$, we can replace the constraint $\int_0^S \frac{t+m}{2} d\mathbb{M}_t = 1$ by the constraint $\int_0^S t d\mathbb{M}_t = 2-m$. Then, following Shapiro (2001) the dual to Eq. (13) with this constraint replaced is

$$\sup_{\theta,\boldsymbol{\lambda},\mathbb{Q}} \quad \theta+\lambda_1(2-m) \tag{EC.22}$$

$$\text{s.t.} \quad \theta+\lambda_1 t+\lambda_2 H(t)\leq \int_0^S pG(p,m,t)d\mathbb{Q}_p \quad \forall t\in[0,S], \tag{EC.23}$$

$$d\mathbb{Q}_p\geq 0, \quad \int_0^S d\mathbb{Q}_p = 1.$$

Again, by weak duality, any feasible solution to this problem lower-bounds $z^{*,m}$. We restrict $\mathbb{Q}$ to discrete distributions supported on the given discretization over p, and denote the corresponding point masses as $Q_0,Q_1,\dots,Q_N$. The last two constraints then become $\boldsymbol{Q}\geq\mathbf{0}$ and $\sum_{j=0}^N Q_j=1$.

Constraint (EC.23) can also be written as the following three (families) of constraints

$$\theta+\lambda_1 t+\lambda_2 H(t)\leq \sum_{j=0}^{N} p_j G(p_j,m,t)Q_j \quad \forall t\in[p_k,p_{k+1}) \quad k=0,\dots,j^*-1, \tag{EC.24a}$$

$$\theta+\lambda_1 m+\lambda_2 H(m)\leq \sum_{j=0}^{N} p_j G(p_j,m,m)Q_j, \tag{EC.24b}$$

$$\theta+\lambda_1 t+\lambda_2 H(t)\leq \sum_{j=0}^{N} p_j G(p_j,m,t)Q_j \quad \forall t\in(p_k,p_{k+1}] \quad k=j^*+1,\dots,N-1. \tag{EC.24c}$$

These three cases correspond to whether t is less than, equal to, or greater than the mode. We further simplify these constraints:

Electronic copy available at: https://ssrn.com/abstract=3127719

Consider Eq. (EC.24a), fix some k and note that necessarily $p_k \le t < p_{k+1} < m$. Split the sum as

$$\sum_{j=0}^{k} p_j G(p_j, m, t) Q_j + \sum_{j=k+1}^{j^*} p_j G(p_j, m, t) Q_j + \sum_{j=j^*+1}^{N} p_j Q_j G(p_j, m, t)$$

In the first sum, $p_j \le t$, in the second sum, $t < p_j \le m$, and in the third sum, $m < p_j$. Consequently, by the definition of $G(\cdot)$, we can rewrite these three sums as

$$\sum_{j=0}^{k} p_j Q_j + \sum_{j=k+1}^{j^*} p_j \frac{m - p_j}{m - t} Q_j$$

Plugging this expression back into Eq. (EC.24a) and multiplying through by $(m - t)$ yields,

$$\theta(m-t) + \lambda_1 t(m-t) + \lambda_2 \int_t^m h \le \sum_{j=0}^{k} p_j Q_j (m-t) + \sum_{j=k+1}^{j^*} p_j Q_j (m - p_j), \quad \forall t \in [p_k, p_{k+1}) \quad k = 0, \ldots, j^* - 1, \tag{EC.25}$$

Next consider Eq. (EC.24b) and use that $H(m) = h(m)$ and $G(p, m, m) = \mathbb{I}(m \ge p)$ to rewrite it as

$$\theta + \lambda_1 m + \lambda_2 h(m) \le \sum_{j=0}^{j^*} p_j Q_j. \tag{EC.26}$$

Finally consider Eq. (EC.24c), fix some k and note that necessarily $m < p_k < t \le p_{k+1}$. Split the sum as

$$\sum_{j=0}^{j^*} p_j G(p_j, m, t) Q_j + \sum_{j=j^*+1}^{k} p_j G(p_j, m, t) Q_j + \sum_{j=k+1}^{N} p_j G(p_j, m, t) Q_j.$$

In the first sum, $p_j \le m$, in the second sum, $m < p_j < t$, and in the third sum, $t \le p_j$. Hence, by the definition of $G(\cdot)$, we can rewrite these three sums as

$$\sum_{j=0}^{j^*} p_j Q_j + \sum_{j=j^*+1}^{k} p_j \frac{t - p_j}{t - m} Q_j.$$

Plugging this expression back into Eq. (EC.24c) and multiplying through by $(t - m)$yields

$$\theta(t-m) + \lambda_1 t(t-m) - \lambda_2 \int_t^m h \le \sum_{j=0}^{j^*} p_j Q_j (t-m) + \sum_{j=j^*+1}^{k} p_j Q_j (t - p_j), \quad \forall t \in (p_k, p_{k+1}] \quad k = j^* + 1, \ldots, N - 1. \tag{EC.27}$$

Combining completes the proof. □

A.7. Omitted Proofs from Section 5.3 to Section 5.5.

Recall we have shown in the main text that when V_c is unimodal, $\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}} \ge \frac{1}{\max_{p \in [0,S]} r(p)}$. Unfortunately, this maximization is not concave. Hence, to form a bound, we discretize the price space. The next lemma quantifies the error induced from such a procedure.

Lemma EC.6 **(Error from Geometric Price Ladder)**. *Fix $0 < \delta < 1$ and let $N = \lceil 1 + \frac{\log(S/\delta)}{\log(1+\delta)} \rceil$. Let $p_0 = 0$, $p_1 = \delta$, $p_N = S$ and $p_j = \delta(1+\delta)^{j-1}$ for $j = 2, \ldots, N-1$, so that $\{p_j\}_{j=0}^N$ discretize the interval $[0, S]$. Define $r_\delta^* := \max_{j: 0 \le j \le N} p_j \mathbb{P}(V_c \ge p_j)$. Then, for any standardized valuation distribution,*

$$r_\delta^* \ \le\ \mathcal{R}_{SP}(F_c, 0) \ \le\ \max(\delta,\ (1+\delta) r_\delta^*).$$

Electronic copy available at: https://ssrn.com/abstract=3127719

Proof of Lemma EC.6. The first inequality follows because the price ladder restricts the feasible region and hence reduces the possible single-pricing revenue. For the second, let p^* be the optimal single price and let k be such that $p_k \le p^* \le p_{k+1}$. We consider two cases: If $k=0$, then $\mathcal{R}_{SP}(F_c,0)=p^*\mathbb{P}(V_c\ge p^*)\le p_1=\delta$. Alternatively, if $k\ge 1$, then,

$$\mathcal{R}_{SP}(F_c,0)=p^*\mathbb{P}(V_c\ge p^*)\le p_{k+1}\mathbb{P}(V_c\ge p_k)\le(1+\delta)p_k\mathbb{P}(V_c\ge p_k)$$

Combining yields the lemma. □

Notice by inspection, the error from this discretization decreases as $\delta\to 0$ and is tight in the limit. We next leverage Lemma EC.6 together with duality to prove Theorem 8.

Proof of Theorem 8. Consider the geometric price ladder described in Lemma EC.6. By that lemma, we have

$$\frac{\mathcal{R}_{PP}}{\mathcal{R}_{SP}}\ge\frac{1}{\max\left(\delta,\ (1+\delta)\max_{j:0\le j\le N} r^m(p_j)\right)}$$

This bound clearly improves as $\delta\to 0$ and is tight in the limit. Thus it only remains to prove that $r^m(p_j)$ can be evaluated as an optimization problem for each j.

Since $\int_0^S d\mathbb{M}_t=1$, we can replace the constraint $\int_0^S \frac{t+m}{2}d\mathbb{M}_t=1$ by $\int_0^S t d\mathbb{M}_t=2-m$ in the definition of $r^m(p_j)$. Then, by duality, we have

$$\begin{aligned} r^m(p_j)=\inf_{\theta,\boldsymbol{\lambda}}\quad & \theta+\lambda_2(2-m)\\ \text{s.t.}\quad & \theta+\lambda_1H(t)+\lambda_2 t\ \ge\ p_jG(p_j,m,t)\quad t\in[0,S]. \end{aligned}$$

We consider three cases based on the value of p_j:

Case i) $p_j<m$. Separate the semi-infinite constraint into two constraints depending on whether $t\in[0,p_j]$, $t\in(p_j,S]$ and use the definition of $G(p_j,m,t)$ to write it as

$$\begin{aligned} \theta+\lambda_1H(t)+\lambda_2 t\ &\ge\ p_j\left(\frac{m-p_j}{m-t}\right)\quad t\in[0,p_j].\\ \theta+\lambda_1H(t)+\lambda_2 t\ &\ge\ p_j\qquad t\in[p_j,S]. \end{aligned}$$

Multiply the first of these constraints through by $m-t>0$ and combine to obtain the optimization problem

$$\begin{aligned} r^m(p_j)=\inf_{\theta,\boldsymbol{\lambda}}\quad & \theta+\lambda_2(2-m)\\ \text{s.t.}\quad & \theta(m-t)+\lambda_1\int_t^m h+\lambda_2 t(m-t)\ \ge\ p_j(m-p_j)\quad\forall t\in[0,p_j],\\ & \theta+\lambda_1H(t)+\lambda_2 t\ \ge\ p_j\quad\forall t\in[p_j,S]. \end{aligned}$$

We can solve this optimization problem efficiently as a linear optimization with constraint generation if we can efficiently identify an optimizer for each of

$$\max_{t\in[0,p_j]} -\theta t+\lambda_1\int_t^m h+\lambda_2 t(m-t),\qquad\text{and}\qquad\max_{t\in[p_j,S]}\lambda_1H(t)+\lambda_2 t.$$

Both of these are special cases of Eq. (16).

Case ii) $p_j = m$. In this case we separate the semi-infinite constraint into two constraints depending on whether $t \in [0,m)$ or $t \in [m,S]$, and use the definition of $G(m,m,t)$ to write

$$\begin{aligned} \theta + \lambda_1 H(t) + \lambda_2 t \geq 0 \quad \forall t \in [0,m] \\ \theta + \lambda_1 H(t) + \lambda_2 t \geq m \quad \forall t \in [m,S], \end{aligned}$$

where we have used continuity to close the half-open interval. Substituting above yields the optimization problem

$$\begin{aligned} r^m(m) = \inf_{\theta,\boldsymbol{\lambda}} \quad & \theta + \lambda_2(2-m) \\ \text{s.t.} \quad & \theta + \lambda_1 H(t) + \lambda_2 t \geq 0 \quad \forall t \in [0,m] \\ & \theta + \lambda_1 H(t) + \lambda_2 t \geq m \quad \forall t \in [m,S], \end{aligned}$$

Notice we can solve this problem as a linear optimization with constraint generation if we can efficiently identify an optimizer for each of

$$\max_{t\in[0,m]} \ \lambda_1 H(t) + \lambda_2 t \quad \text{ and } \quad \max_{t\in[m,S]} \ \lambda_1 H(t) + \lambda_2 t,$$

both of which are special cases of Eq. (16).

Case iii) $p_j > m$ We now consider two cases depending on whether $t \in [0,p_j]$ or $t \in (p_j, S]$. Again, split the semi-infinite constraint and use the definition of $G(p_j,m,t)$ to write

$$\begin{aligned} \theta + \lambda_1 H(t) + \lambda_2 t \ \geq 0 \quad \forall t \in [0,p_j] \\ \theta + \lambda_1 H(t) + \lambda_2 t \geq p_j \left(\frac{t-p_j}{t-m} \right) \quad \forall t \in [p_j,S]. \end{aligned}$$

Multiply the second constraint through by $t - m > 0$, and combine to show that

$$\begin{aligned} r^m(p_j) = \inf_{\theta,\boldsymbol{\lambda}} \quad & \theta + \lambda_2(2-m) \\ \text{s.t.} \quad & \theta + \lambda_1 H(t) + \lambda_2 t \ \geq 0 \quad \forall t \in [0,p_j] \\ & \theta(t-m) - \lambda_1 \int_t^m h + \lambda_2 t(t-m) \geq p_j(t-p_j) \quad \forall t \in [p_j,S]. \end{aligned}$$

We can solve this optimization efficiently as a linear optimization with constraint generation whenever we can identify an optimizer for each of

$$\max_{t\in[0,p_j]} \lambda_1 H(t) + \lambda_2 t \quad \text{ and } \quad \max_{t\in[p_j,S]} \theta t - \lambda_1 \int_t^m h + \lambda_2 t(t-m) - p_j t,$$

both of which are special cases of Eq. (16).

These three cases thus complete the proof. $\square$

Proof of Proposition 1.

Part a): As mentioned in the main text, we consider critical points of Eq. (15), i.e., solutions to $\frac{a}{c} + \frac{2b}{c}t = \frac{M}{2}|t-1| - D$. We first seek roots where $t \leq 1$. There is at most one such root, given by $t_{\leq 1} \equiv \frac{-2a-2D+cM}{4b+cM}$, but only if this value is less than equal to 1. Otherwise, there is no root less than 1. We next seek roots for

$t \geq 1$. Again, there is at most one such root, given by $t_{\geq 1} \equiv \frac{-2a-2D-cM}{4b-cM}$, but only if this value is great than or equal to 1. Otherwise there is no root greater than one.

In summary, an optimizer is one of p_k, p_{k-1}, $t_{\leq 1}$ (if $t_{\leq 1} \leq 1$) or $t_{\geq 1}$ (if $t_{\geq 1} \geq 1$), and can be identified by simply checking the feasibility and comparing these (at most) 4 values.

Part b): Consider the first of the two optimization problems. Notice that if V is uniform on $[t,m]$ with $t \leq m$ or on $[m,t]$ with $t \geq m$, we can write $V = t + (m-t)\xi$ with $\xi \sim \text{Uniform}[0,1]$. Hence, we can rewrite $H(t) = \mathbb{E}[h(t + (m-t)\xi)]$. Since $h(\cdot)$ is convex, it follows that this function is convex in t; $h(t+(m-t)\xi)$ is the composition of a convex and affine function, and expectations preserve convexity.

We conclude that if $a \leq 0$, the first optimization problem is the minimization of a concave function, and the optimum occurs at an end point $\{l,u\}$. If $a > 0$, then it is the minimization of a convex function. The optimum occurs either at an end point $\{l,u\}$, or else at t^* solving $aH'(t) + b = 0$. Such a t^* can be obtained by bisection search.

For the second optimization problem, the optimum occurs either at an endpoint $\{l,u\}$ or else at a critical point solving $2at + b = ch(t)$. If the critical point is less than or equal to 1, this is a linear equation with at most one root at $t^*_{\leq} = \frac{-2b-2cD+cM}{4a+cM}$. If the critical point is greater than 1, this is again a linear equation with at most one root at $t^*_{\geq} = \frac{-2b-2cD-cM}{4a-cM}$. Hence, the optimum of the second optimization problem occurs at one of at most 4 points: l, u, $t^*_{\leq}$ (if this quantity is less than or equal to 1 and in $[l,u]$) or $t^*_{\geq}$ (if this quantity is greater than 1 and in $[l,u]$). We can simply compare these 4 points to identify an optimizer. □

Proof of Proposition 2.

Part a): Since $h(\cdot)$ is continuous, it suffices to reformulate the semi-infinite constraint

$$\lambda_1 v + \lambda_2 h(v) \leq \sum_{j=0}^{k-1} p_j Q_j - \theta, \quad \forall v \in [p_{k-1}, p_k],$$

for each k. Notice that depending on the sign of λ_2, this constraint may or may not be convex in v for a fixed $\lambda_1, \lambda_2, \boldsymbol{Q}$.

Since $v \in [p_{k-1}, p_k] \iff (v - p_{k-1})(v - p_k) \leq 0$, we can use the definition of $h(\cdot)$ to rewrite the k^{th} constraint as

$$\theta - \sum_{j=0}^{k} p_j Q_j - \lambda_2 C^2 \leq \min_{v:(v-p_{k-1})(v-p_k)\leq 0} -\lambda_1 v - \lambda_2 M^2 (v-1)^2.$$

The (possibly non-convex) minimization on the right is an example of a quadratic optimization problem in which quadratic forms in the objective and in the constraint are simultaneously diagonalizable. Such problems were studied in Ben-Tal and Den Hertog (2014) which shows they can be equivalently written as convex, second order cone problems. Indeed, applying the key result of that paper shows the k^{th} constraint is equivalent to the constraints

$$\begin{aligned}
(y_k + \lambda_2 M^2) p_{k-1} p_k - x \geq \theta - \sum_{j=0}^{k} p_j Q_j + \lambda_2 (M^2 - C^2) \\
4 y_k x_k \geq z_k^2 \\
z_k = 2\lambda_2 M^2 - \lambda_1 - (y_k + \lambda_2 M^2)(p_{k-1} + p_k) \\
x_k, y_k \geq 0, \quad y + \lambda_2 M^2 \geq 0,
\end{aligned} \tag{EC.28}$$

Electronic copy available at: https://ssrn.com/abstract=3509952

with the auxiliary variables x_k, y_k, z_k. This formulation is always convex (Constraint (EC.28) is a rotated second-order cone constraint Boyd and Vandenberghe (2004)). Performing this transformation for each of the semi-infinite constraints yields a (convex) second order cone representation, proving the theorem.

Part b): Again, an optimizer of Eq. (15) occurs either at endpoint p_k, p_{k-1}, or else at a critical point, i.e., a solution to $\frac{a}{c} + \frac{2b}{c}t = M^2(t-1)^2 - C^2$. This equation has two roots, given by $\frac{b+cM^2 \pm \sqrt{b^2+c(a+2b+cC^2)M^2}}{cM^2}$. These roots can only be optimizers of Eq. (15) if they lie within $[p_k, p_{k+1}]$, which can be checked explicitly. Hence there are at most 4 possible optimizers, and we can identify an optimizer in closed form by comparing their objective values.

Part c): Consider the first of the two optimization problems. The same convexity argument that applied in the case of Theorem 1 applies here unchanged. Hence, when $a \leq 0$, an optimum occurs at an end point of $\{l, u\}$. If $a > 0$, then an optimum occurs either at this end point or else the unique solution t^* to $aH'(t^*) + b = 0$ which can be obtained by bisection.

For the second optimization problem, the optimum occurs either at an endpoint $\{l, u\}$ or else at a critical point solving $2at + b = ch(t)$. This function is a quadratic equation with precisely two roots,

$$t_{\pm} = \frac{a - cM^2 \pm \sqrt{a^2 + c(-2a + b - C^2)M^2}}{cM^2}.$$

Hence, the optimum of the second optimization problem occurs at one of at most 4 points: l, u, t^*_+ or t^*_-. We can simply check feasibility and compare these 4 points to identify an optimizer. $\square$