# Optimal Feature-Based Market Segmentation and Pricing

Titing Cui, Michael L. Hamilton
Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15213,
tic54@pitt.edu, mhamilton@katz.pitt.edu

In this work, we study semi-personalized pricing strategies where a seller uses features about their customers to segment the market, and customers are offered segment-specific prices. In general, finding jointly optimal market segmentation and pricing policies is computationally intractable. In response, we study how to optimize and analyze feature-based market segmentation and pricing under the assumption that the seller has a trained (noisy) regression model mapping features to valuations. First, we establish novel hardness and approximation results in the case when model noise is independent. Second, in the common cases when the noise in the model is log-concave, we show the joint segmentation and pricing problem can be efficiently solved, and characterize a number of attractive structural properties of the optimal feature-based market segmentation and pricing. Finally, we conduct a case study using home mortgage data, and show that compared to heuristic segment-then-price approaches, our optimal feature-based market segmentation and pricing model can achieve nearly all of the available revenue with only a few segments.

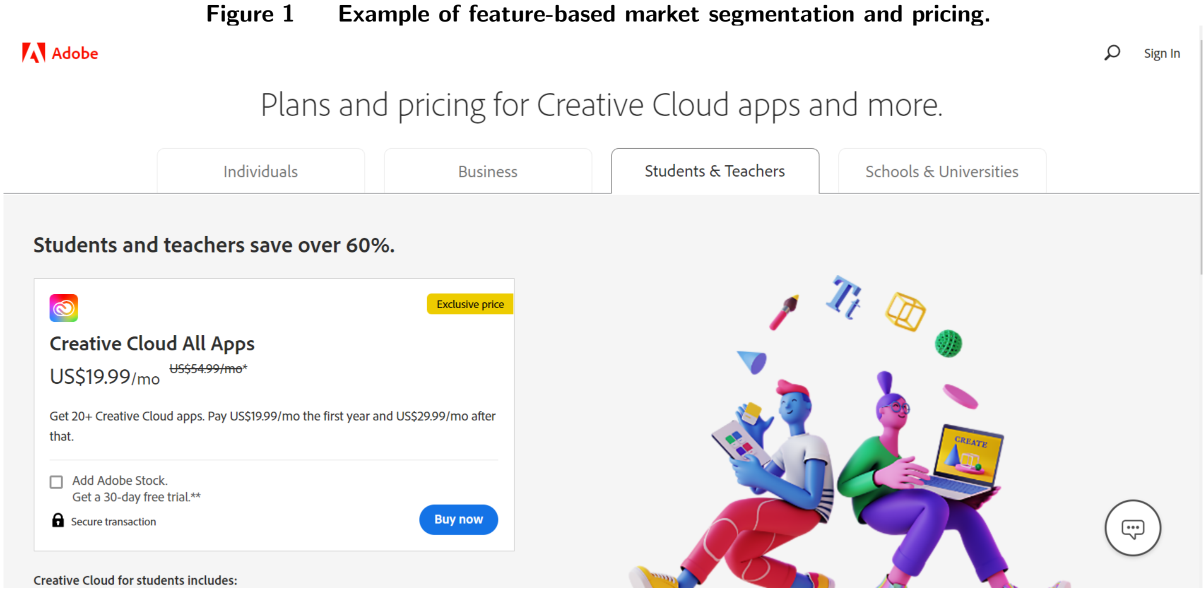
*Key words*: personalized pricing, market segmentation, third degree price discrimination, regression
*History*: June $30^{th}$, 2022.

---

## 1. Introduction

Third-degree price discrimination concerns the now ubiquitous practice of selling a good at different prices for different consumers (Varian 1989). For example, in the sale of proprietary software licenses, prices may differ based on whether the customer is a student, professional, or corporate user (see Fig. 1 for an example and Lehmann and Buxmann (2009) for an extended discussion). In insurance markets, firms gather extremely rich and nuanced feature information about their customers, ostensibly to estimate risk, but which is also leveraged to price discrimination on the basis of demographic and/or geographic information (Buzzacchi and Valletti 2005, Thomas 2012). In commercial markets facing walk-in customers, firms are comparatively limited in their information about their customers but can still profitably leverage feature information to price discriminate, for instance in movie theater ticket sales the customers' age (child, adult, senior) and the time of screening (weekend, matinee) can be used to issue semi-personalized discounts (Dubé et al. 2017).

**Figure 1 Example of feature-based market segmentation and pricing.**




*Note.* An example of feature-based market segmentation and pricing for Adobe Creative Cloud products (see `https://www.adobe.com/creativecloud/plans.html`). Here customers are segmented based on their attributes (i.e. student versus professional) and prices vary based on the segment customers are in, with the student rate being 60% of the professional rate (note discounts are enforced by requiring a valid .edu email). Further, note the the number of segments, $k$, is only four. It will be informative to think of $k$ as $\approx 4$ in this work.

Each of these markets vary in both the quality and descriptive power of the information they have about their customers, as well as in the operational difficulty of setting and changing their prices, and thus implementing price discrimination. When the information about customers in the market is of low quality or is largely censored, fully personalized price discrimination where each customer is charged a personalized price may be undesirable, however as mentioned above, that does not preclude the use of some modest price discrimination via market segmentation. In fact, even when information is richly textured and pricing is largely unconstrained by legal and/or operational considerations, a small static set of prices based on customer features is still often preferred to fully personalized pricing. A small set of prices and market segments is conceptually and operationally simple to implement, and a surprisingly small number of options is often sufficient to achieve strong revenue (Courty and Pagliero 2012). In this paper, we propose and study a general framework for implementing limited price discrimination capturing these trade-offs between predictive power and operational flexibility we term feature-based market segmentation and pricing (FBMSP).

Finding and optimizing generic market segmentation and pricing policies is a well-studied problem in industry with academic roots in operations research/management, marketing, economics, and computer science. However, archetypal formulations of the segmentation and pricing problem are well known to be intractably hard (Kleinberg et al. 1998, 2004). To deal with this hardness, much of the literature has taken a heuristic approach to the problem (Claycamp and Massy 1968, Assael and Roscoe Jr 1976, Chen 2001, Liu et al. 2010, Li and Qiu 2014), separating the segmentation and pricing components. Segmentation then pricing procedures use tools from unsupervised learning to first identify consumer segments/clusters with similar features, and then identify revenue optimal pricing for those chosen market segments. In this paper, we advance the study of segmentation and pricing by finding *jointly optimal* segmentation and pricing's under some realistic assumptions about how firms leverage feature information to predict customers' valuations. Specifically, in practice a firm's valuation model i.e. the model that maps features to valuation or a proxy for willingness-to-pay, is built using regression. Regression models come with their own theory and standard set of assumptions that we profitably utilize to study market segmentation and pricing as well. We show that by leveraging independence and log-concavity of residuals in the regression model, the revenue-maximizing feature-based market segmentation and pricing (FBMSP) enjoys a simple, intuitive structure and can be computed efficiently. Further, our structural results allow us to analyze optimal FBMSP and derive new insights operational and managerial insights about such policies, including guidance for choosing the number of segments and conditions for when they are near-optimal.

### 1.1. Our Contributions

To summarize our contributions:

1. We first study the algorithmic problem of finding the optimal FBMSP. In general, the problem is intractably hard, so we focus our attention on the case when valuations are predicted according to a regression model with independent residuals. We show that with no additional assumptions, while we can prove some promising structural properties (c.f. Lemma 1) and provide a $(1 - 1/e)$ approximation algorithm for the optimal segmentation and pricing (c.f. Remark 1), unfortunately finding the optimal FBMSP is still NP-Hard to compute (c.f. Theorem 1). However, when we further assume the residuals are log-concave, as is often the case in practice, we are able to evade our hardness result. Specifically, when residuals are independent and log-concave, we prove the optimal policy has a simple *interval* structure which allows us to compute in it cubic time via dynamic programming (Theorem 2).
2. We next turn our attention to analyzing the performance of optimal feature-based market segmentation and pricing. Specifically, we consider the practical operational question of how to choose the number of segments $k$ so as to guarantee minimal loss against a fully personalized pricing benchmark. We show three results that can help guide practitioners in choosing $k$. First, we show that an upper bound on the loss against personalized pricing can be achieved by simply examining the loss in the model, ignoring the noise term (c.f. Theorem 3). Next, we tightly upper bound the optimal rate at which FBMSP tends to personalized pricing as a function of the number of segments $k$ and some valuation parameters (c.f. Theorem 4). Finally, we show that the revenue of FBMSP is concave in $k$ (c.f. Theorem 5). Taken together, these three results allow a practitioner to use their regression model (without reference to the complicating error!) to find $k$ via a simple elbow method, and feel confident that the results of such a heuristic are provably close to optimal.
3. Finally in Section 5, we demonstrate our method on real housing loan data collected in Pennsylvania in 2020, and compare its performance against standard segment-then-price methodologies. We find our approach significantly outperforms heuristic methods, especially when the number of segments is small and the variation in the valuations comes primarily from variation in $\mu(\cdot)$, as opposed to variation from the prediction error $\epsilon$. We also note that the segmentations found by our approach are

qualitatively different than those in segment-then-price, with our approach quickly isolating key differences between groups whereas heuristic approaches can get bogged down in pointless discrimination between groups until it discovers the important differences for the revenue.

### 1.2. Literature Review

Our work is influenced by, and contributes to, several streams of literature across operations management, marketing, and computer science. Here we overview some of these streams and connect them to our work.

**Theory of Price Discrimination** There is a deep extensive literature on the theory of pricing discrimination beginning in economics and spanning operations management, marketing, and computer science. Much of the classic literature in this area (Schmalensee (1981), Narasimhan (1984), Katz (1984), Varian (1985), Shih et al. (1988), Bergemann et al. (2015), Cowan (2016), Xu and Dukes (2016)) focuses on the impact of price discrimination on social welfare, or the effects of price discrimination on the resultant equilibrium prices. In this paper, we investigate market segmentation and pricing exclusively from the perspective of a revenue-maximizing monopolist, focusing on computational/practical implementation of such policies.

Specifically, in the language of (Varian 1985) we study third degree price discrimination. In practice, firms engage most often in third-degree price discrimination, which can use additional information about consumer features to offer different prices to different implicit or segments in a multitude of ways (Su (2007), Jerath et al. (2010), Besbes and Lobel (2015), Chen et al. (2005), Cohen et al. (2017), Elmachtoub and Hamilton (2021)). Several papers have analyzed the value of such price-discrimination tactics compared to uniform pricing (Huang et al. (2019), Elmachtoub et al. (2021)). In contrast, we investigate the value of the optimal feature-based market segmentation and pricing in this paper, and compare these semi-personalized pricing's against a fully personalized benchmark.

**Regression Based Price Discrimination** In recent years, data-driven pricing strategies have become increasingly common (Chen et al. (2015), Ferreira et al. (2016), Shukla et al. (2019), Aouad et al. (2019), Elmachtoub et al. (2020), Niu et al. (2020), Biggs et al. (2021), Elmachtoub and Grigas (2022)). In these works, customers are offered a personalized price based on features that are predictive of their valuation of the product, especially by tree-based prescriptive approaches (Athey and Imbens (2016), Kallus (2017), Bertsimas et al. (2019), Biggs et al. (2021)). Unlike most data-driven pricing literature, here, we ignore how the regression model is found and instead take the prediction of customer's valuation as input, and analyze how it may be profitable leveraged to compute and analyze optimal FBMSP.

**Algorithms for Market Segmentation and Pricing** Our paper contributes to a line of literature studying the market segmentation and pricing from a algorithmic/computational complexity perspective. Indeed many models of joint market segmentation and pricing are known to be intractably hard to compute going back at least to the pioneering work of Kleinberg et al. (1998, 2004), restricting their applicability in practice. Often in marketing to evade these hardness results the segmentation and pricing decisions are made sequentially instead of being evaluated together (Dolgui and Proth (2010)), and at first blush it seems that Theorem 1 implies our model, for all the structure gained through independence, is ultimately no better. Fortunately, we will see for almost all regression models in practice our model makes jointly optimal segmentation and pricing tractable and well structured.

If the regression error is log-concave, as we assume in Section 3.2, computing the optimal feature-based segmentation is structurally similar to the *1D Clustering* problem for which dynamic programming approaches have been employed (see Gronlund et al. (2017) for a modern overview), and can be solved in polynomial time. Other algorithmic approaches for feature-based pricing can be seen in Cohen et al. (2016), Qiang and Bayati (2016), Javanmard and Nazerzadeh (2016), albeit in different models.

### 1.3. Paper outline

The remainder of this paper is organized as follows. In Section 2 we introduce our model for feature-based market segmentation and pricing and provide some preliminary structural results. In Section 3 we study the problem of computing the revenue-optimal FMBSP. In Section 4 we analyze the structure of revenue-optimal FBMSP and provide some theory to guide practitioners in choosing the number of segments, $k$. In Section 5 we demonstrate our approach on a well known Home Mortgage Disclosure Act dataset. Finally, in Section 6 we provide concluding remarks and highlight future directions.

## 2. Model and Preliminaries

We consider a revenue-maximizing seller offering a good in unlimited supply. For simplicity of presentation, we will assume the good is produced costlessly and so revenue and profit are equivalent (we note the model presented in this paper and all results do however easily extend to the case when each good has a per unit cost $c$). We will further assume each customer in the market is described by some feature vector $\mathbf{x}$ of their observable characteristics, and has some valuation for the good which depends on their feature vector, $V|\mathbf{x}$. The market characteristics as a whole can be described as a distribution over the feature vectors $\mathbf{X} \sim F_{\mathbf{X}}$, which is supported on some feature space $\mathcal{X} := \text{supp}(\mathbf{X})$. These features vectors can consist of any information about the customers, including demographic information like gender, household status, income etc.

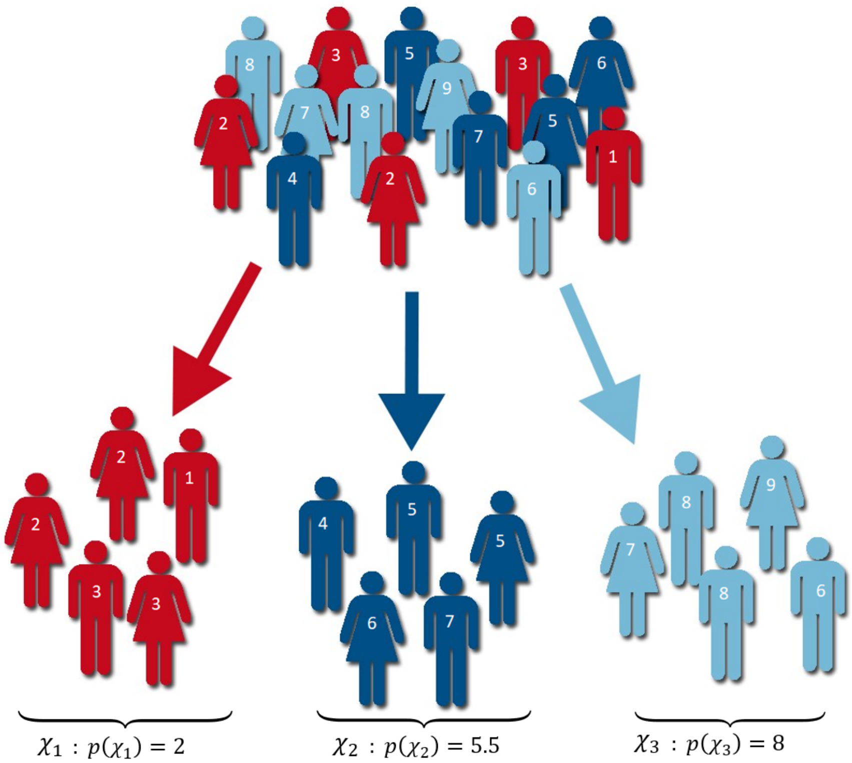
In line with modern practice, we model the seller as having trained some regression model $\mu : \mathcal{X} \to \mathbb{R}^+$ to predict a customers valuation for a good from their feature vector. We assume the regression model has residual error $\epsilon$ but is correct in expectation, so that the predicted valuation for a customer with features $\mathbf{x}$ is $\mu(\mathbf{x}) := \mathbb{E}[V|\mathbf{x}]$, and the valuation model is $V = \mu(\mathbf{X}) + \epsilon$. We will use $F$ to be the distribution of the valuations $V$, $F_{\mathbf{X}}$ to be the distributions of the feature vector, and $F_\epsilon$ to be the distribution of the error term $\epsilon$, and $f_{\mathbf{X}}$, $f_\epsilon$, and $f$ to be the densities, respectively. We will use $\overline{F}$ to denote the survival function, i.e., $\overline{F}(x) := 1 - F(x)$.

For a seller with a valuation model $\mu(\cdot)$, we will study the revenue achievable by selling strategies where the feature space of the market, $\mathcal{X}$, is partitioned into $k$ segments $\{\mathcal{X}_i\}_{i=1}^k$, $\cap \mathcal{X}_i = \emptyset$, $\cup \mathcal{X}_i = \mathcal{X}$, such that on each segment the seller offers a distinct price $p_\epsilon(\mathcal{X}_i)$. Now we are ready to define feature-based market segmentation and pricing strategies, which is the main object of this study.

**Feature-Based Market Segmentation and Pricing:** In feature-based market segmentation and pricing the seller partitions the feature space in into $k$ segments $\{\mathcal{X}_i\}_{i=1}^k$, and on each segment offers a single price $p(\mathcal{X}_i)$. The expected profit of such a segmentation is,

$$\mathcal{R}_{kXP}\left(\{\mathcal{X}_i\}_{i=1}^k, \{p_\epsilon(\mathcal{X}_i)\}_{i=1}^k\right) = \sum_{i=1}^k p_\epsilon(\mathcal{X}_i) \int_{\mathbf{x} \in \mathcal{X}_i} \Pr\left(\mu\left(\mathbf{x}\right) + \epsilon \geq p_\epsilon(\mathcal{X}_i)\right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \tag{1}$$

**Figure 2** **An example of FBMSP.**




*Note.* Depicted is an example of feature based market segmentation. For each customer some numeric prediction of their valuation is given. The feature space $\mathcal{X}$ consists of the color and gender of the customer, and the feature-based market segmentation leverages color (not necessarily optimally!) to sort them into three segments $\mathcal{X}_i$, $i \in [3]$, each with a distinct segment level price, $p(\mathcal{X}_i)$.

Given a segmentation it will often be convenient to think of the prices as the revenue optimal ones for that segment. To that end we denote the optimal price on segment $\mathcal{X}_i$ by $p_\epsilon(\mathcal{X}_i)$ i.e.,

$$p_\epsilon(\mathcal{X}_i) = \arg\max_p p \int_{\mathbf{x}\in\mathcal{X}_i} \Pr\left(\mu\left(\mathbf{x}\right) + \epsilon \geq p\right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

We will use $\mathcal{R}_{kXP} \equiv \max_{\mathcal{X}_1,\ldots,\mathcal{X}_k} \sum_{i=1}^k \mathcal{R}_{kXP}(\{\mathcal{X}_i\}_{i=1}^k, \{p_\epsilon(\mathcal{X}_i)\}_{i=1}^k)$ denote the optimal profit for a feature-based market segmentation and pricing strategy.

Note that our framework for feature-based market segmentation and pricing is very flexible, and captures many well studied models as a special case. For instance when $k = 1$, FBMSP is just the revenue generated by single static price for the good (sometimes referred to as the revenue of the monopoly price, or posted price, or single price). Similarly when the number of segments is very large, i.e. $k \to \infty$, FBMSP becomes the revenue of feature-based personalized pricing, where each customer receives an appropriately personalized price. The revenue of feature-based personalized pricing is a useful upper bound to compare with the revenue of optimal FBMSP for some fixed number of segments $k$. In Section 4 we consider the question of how large does $k$ need to be, in general, to approach the revenue of personalized pricing, with the hope that a reasonably small $k$ should suffice (see Elmachtoub et al. (2021) for a detailed discussion of when feature-based personalized pricing is provably close or far from the revenue of a single price).

Similarly when the error distribution $\epsilon$ is 0 almost surely (a.s.), our model represents the achievable revenue in the world of the prediction model $\mu$ without regard to the models potential error. We term this optimistic case *model market segmentation*, and in Section 4 will show that reasoning about the profit in a world of perfect prediction can provide a useful upper bound on the loss of FBMSP with error in practice.

### 2.1. Key assumptions and preliminaries.

As mentioned in the introduction, the optimal feature-based market segmentation and pricing is generally very hard to compute. To ensure tractability, in our work we will carry through the common regression assumption that the error $\epsilon$ is independent across features i.e. $\mathbf{X} \perp\!\!\!\perp \epsilon$. We consider this assumption to be quite mild, as it underlies many predictive models used in practice, including for example, the well known logit model where a customer's valuation is a linear combination of that customer's features, the offered price, and an idiosyncratic error following a logistic distribution which is *independent* of $\mathbf{X}$. Similar remarks hold for other regression-based models with independent errors. The upshot will be that this necessary assumption for regression is also quite harmonious with pricing and gives considerable structure and control for analyzing pricing models.

In the next section we will delve into the structure of optimal FBMSP, but first we will illustrate how the independence assumption smooths our problem for some related objectives. To this end, consider three auxiliary functions that will be helpful in analysis of FBMSP, and also are of independent interest.

$$\textbf{Price: } p_\epsilon(x) := \inf\{\arg\max_p p\overline{F}_\epsilon(p-x)\}, \quad \textbf{Margin: } \theta_\epsilon(x) := p_\epsilon(x) - x,$$
$$\textbf{Revenue: } \mathcal{R}_\epsilon(x) := \max_p p\Pr(x+\epsilon \geq p) = p_\epsilon(x)\overline{F}_\epsilon(\theta_\epsilon(x)).$$

The price, margin, and revenue functions all serve to model a seller pricing a good for a customer after having predicted their valuation as $x$, up to some stochastic error $\epsilon$. $p_\epsilon(x)$ is the optimal price to offer a customer with predicted valuation $x \in \mathbb{R}$, $\mathcal{R}_\epsilon(x)$ is the revenue of the optimal monopoly price when the valuation distribution is $x+\epsilon$, and $\theta_\epsilon(x)$ is the difference or *margin* between the predicted valuation $x$ and the offered price $p_\epsilon(x)$. Note, $p_\epsilon(x)$ is uniquely defined to be the minimum price that achieves the maximum revenue, such a minimum is necessary since for some distributions $\epsilon$, there may be many prices that maximize the revenue (for an extensive discussion on when the optimal price is unique, or equivalently when the revenue function is strictly unimodal, see Ziya et al. (2004)).

In the following lemma we summarize some of the structure we observe in these functions simply from assuming independence of residuals.

LEMMA 1 **(General Properties of $p_\epsilon(\cdot), \theta_\epsilon(\cdot), \mathcal{R}_\epsilon(\cdot)$).** *Suppose that $V = \mu(\mathbf{X}) + \epsilon$ where $\mathbf{X} \perp\!\!\!\perp \epsilon$, and $\mathbb{E}[\epsilon] = 0$. Then the following properties hold:*

*(a) $\theta_\epsilon(x)$ is an decreasing function.*

*(b) For any $0 < x_1 < x_2$, we have*

$$\overline{F}_\epsilon\left(\theta_\epsilon(x_1)\right)(x_2 - x_1) \le \mathcal{R}_\epsilon(x_2) - \mathcal{R}_\epsilon(x_1) \le \overline{F}_\epsilon\left(\theta_\epsilon(x_2)\right)(x_2 - x_1).$$

*Moreover, for all $x$ such that $p_\epsilon(x)$ is continuous (i.e. $p_\epsilon(x^-) = p_\epsilon(x^+)$), the derivative of $\mathcal{R}_\epsilon(x)$ exists and $\frac{d}{dx}\mathcal{R}_\epsilon(x) = \overline{F}_\epsilon(\theta_\epsilon(x))$.*

*(c) $\mathcal{R}_\epsilon(x)$ is increasing, continuous, and convex.*

Lemma 1 states that independence alone induces prices that result in a monotone increasing sales probabilities $\theta_\epsilon(x)$, and a convex revenue function $\mathcal{R}_\epsilon(x)$ with interpretable, bounded derivatives. All three parts of the lemma are proved by examining the induced optimal prices, $p_\epsilon(x)$, and noting that $p_\epsilon(x)$ cannot increase very quickly (i.e. super linearly). Unfortunately, our control is not perfect as $p_\epsilon(x)$ can otherwise be quite poorly behaved; there can be many optimal prices for a given valuation, and worse $p_\epsilon(x)$ can be discontinuous in $x$ at arbitrarily many points $x$, see Example EC.1 for an example. The jump discontinuities in $p_\epsilon(x)$ translate directly to non-differentiable points in the revenue function. As we will see in Section 3, the structure provided by independence is not quite enough to enable the efficient computation of the optimal FBMSP, however it will be critical in our analysis of such policies.

## 3. Computing Optimal Feature-Based Market Segmentation and Pricing

In this section we will study the problem of finding the jointly optimal FBMSP, culminating with conditions and an algorithm under which the optimal policy can be computed.

We will first show that when valuations are drawn from a regression model with general independent residuals, the optimal policy is NP-Hard to compute. We then identify that the hardness stems from some pathological segmentation properties, and define a natural property to characterize nice segmentations which we call *interval*. Our main positive result of this section is to show that when the residuals are log-concave the optimal segmentations are interval, and further, the optimal interval segmentations can be found in cubic time via dynamic programming. Thus, for realistic valuation models under standard regression assumptions, the jointly optimal segmentation and pricing can be directly computed instead of having to resort to heuristic segment-then-price approaches.

### 3.1. Hardness of FBMSP

To understand some of the difficulty of FBMSP, in this subsection we will review some standard hardness results and show that even under the assumption of independent residuals, the problem remains intractable. Our proof of this hardness result will yield guiding intuition for how an additional condition of log-concavity on the residuals should be computationally useful.

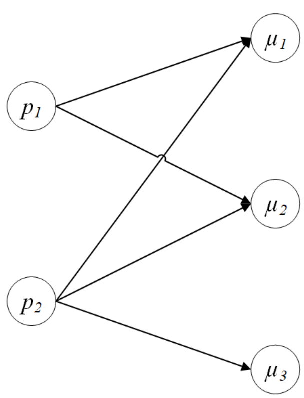
In general, the hardness of market segmentation and pricing problems typically follows from a reduction to hitting set (Kleinberg et al. 1998, 2004). The correspondence between the two problems often works as follows: imagine you have $n$ customers each with valuations described by an independent distribution $F_i$ such that $p\overline{F}_i(p)$ is maximized by some discrete set of optimal prices $S_i$. Then the market segmentation and pricing problem is simply to find a partition of customers in $k$ segments such that on each segment, the intersection of each customers optimal price set is non-empty. This implies then that there is an obvious optimal price for the each segment that clearly maximizes the revenue by construction, and so the hardness is merely to find the partition with such a property. This is exactly the difficulty in $k$-hitting set, and so follows the reduction.

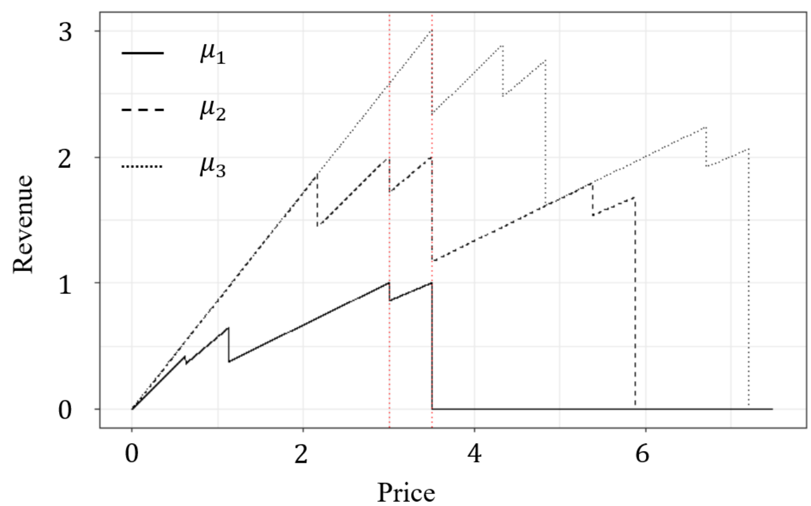
Note, in the above that the assumption of unique valuation distributions for each customer, $F_i$, is crucially important in setting up the correspondence with hitting set. In our model, where valuations are described by a common regression function with independent noise, it is no longer clear if such a construction is possible. That is, now for customer $i$ with features $\mathbf{x}_i$, their valuation distribution is $V|\mathbf{x}_i = \mu(\mathbf{x}_i) + \epsilon$. As we show in the following theorem, the problem remains NP-hard in this case, although the proof requires an significantly more intricate construction of the common error distribution $\epsilon$.

Theorem 1 **(Hardness of FBMSP)**. *Suppose that* $V=\mu(\mathbf{X})+\epsilon$, $\mathbf{X}\perp\!\!\!\perp\epsilon$, *and* $\mathbb{E}[\epsilon]=0$. *Then finding the optimal FBMSP policy is NP-hard.*

*Sketch of Proof of Theorem 1* The proof of Theorem 1 follows by reduction to hitting set, as in the general case. For every instance of the hitting set problem, we show that there exist estimations of customers' valuation and prediction error's, such that deciding if there are $k$ (or less) elements that hits all the subsets is equivalent to deciding whether there are $k$ segments and prices such that the total revenue is $\frac{n(n+1)}{2}$, where $n$ is the number of subsets in the hitting set problem (equivalently, the number of customers in the market). Our construction follows by designing a error distribution $\epsilon$ which results in a number claw like functions for each customer, that are then spread by translation to encode a set of optimal prices for each valuation level $\mu(\mathbf{x}_i)$. Fig. 3 gives an example of our hardness reduction for a small instance. □

**Figure 3 Example of hitting set and FBMSP.**




*Note.* In the left panel is a graph representation of a small instance of a hitting set problem. To illustrate the translation of a hitting set problem to a FBMSP problem, assume we have 3 customers with predicted valuation $\mu_1=3$, $\mu_2=\frac{43}{8}$, $\mu=\frac{161}{24}$. Further, let the estimation error $\epsilon$ be supported on $\left\{-\frac{161}{24},-\frac{77}{5},-\frac{19}{8},-\frac{15}{8},0,\frac{1}{2}\right\}$ with probability mass $\left\{\frac{1}{7},\frac{4}{21},\frac{2}{21},\frac{5}{21},\frac{1}{21},\frac{2}{7}\right\}$. Finally define $p_1=3$ and $p_2=3.5$. In the right panel we plot the revenue curves for each valuation. We can see that the revenue for each customer is maximized only at either $p_1$ or $p_2$ which represent the connections between price nodes and valuation nodes in left panel, *i.e.*, revenue from customer 3 is maximized at price $p_2=3.5$, revenues from customer 2 and 1 are both maximized at $p_1=3$ and $p_2=3.5$.

The NP-hardness of general FBMSP implies it's impossible to solve general FBMSP efficiently if P $\neq$ NP. This leaves us two options, either to look at approximate solutions for general FBMSP or to enrich the structure of our model by imposing additional assumptions. We briefly explore the former in Remark 1, but will focus mainly on the later.

REMARK 1. While an optimal policy for general error distribution cannot be found in polynomial time, we remark that a constant factor approximation to the optimal feature-based market segmentation is obtainable. Specifically, a $(1-1/e)$ factor approximate policy can be found in polynomial time since the objective function is positive valued, monotone, and submodular. We formalize this observation in Section D in the appendix. □

### 3.2. Feature-Based Market Segmentation and Pricing with Log-Concave Residuals

In the previous subsection we studied FBMSP where we allowed the underlying error in the valuation model to be arbitrary. Recall in Lemma 1 were able to characterize many things about the revenue function itself, but less about the structure of the segmentation and pricing policy. In Example EC.1 we showed that the price function may not vary smoothly with the predicted valuations, potentially dropping dramatically between similar customers. In part due to these pathological discontinuities in $p_\epsilon(x)$, FBMSP with general prediction error has a discrete quality to it that makes it difficult to optimize.

Further, in the proof Theorem 1 we constructed general error distributions such that they induce jagged, non-differentiable revenue functions which make FBMSP NP-Hard to solve. Of course, a quick examination of the Fig. 3 suggests that this error distribution is quite contrived, with carefully chosen spikes in it's density function to induce delicately structured overlap in the revenue curves. A natural question then to ask is, for suitably smooth erro distributions/revenue functions, is it still hard to compute the optimal FBMSP? In this subsection we will make one additional assumption about the error distribution that enforces a notion of smoothness, namely will assume the distribution is log-concave, a canonical assumption in the pricing and revenue management literature which often makes the revenue curve concave. Note many standard distributions are log-concave including normal, exponential, uniform distributions, etc.

DEFINITION 1 (LOG-CONCAVE ERROR). A random variable $\epsilon$ with density $f_\epsilon$ is **log-concave** if $\log(f_\epsilon(x))$ is a concave function.

To understand how log-concavity in the error function translates into tractability for FBMSP, we will begin with a definition which captures what market segmentations that are well structured. In particular, one could hope that a nice market segmentation would be one that groups together customers with similar predicted valuation for the good. Such segmentations are natural, easy to interpret as low/medium/high type segments, and as we will show, easy to optimize and analyze. We emphasize that not all optimal market segmentations are interval, certainly the ones induced by the construction in Theorem 1 are not, but also even simple error distributions can have this type of pathological structure as we demonstrate in Example EC.2. We will call segmentations that group together customers with similar valuations *interval*, and define them as follows.

DEFINITION 2 (INTERVAL SEGMENTATION). We will call an segmentation, $\{\mathcal{X}_i\}_1^k$, an **interval** segmentation if there exists real numbers $0 = s_0 < s_1 < \ldots < s_k < s_{k+1}$ such that each segment $\mathcal{X}_i$ can be written as $\mathcal{X}_i = \{\mathbf{x}|\mu(\mathbf{x}) \in [s_i, s_{i+1})\}$.

As it turns out, the smooth notion of error captured by log-concavity, and the intuitive structure of interval segmentations are harmonious notions. In following lemma we continue our study of optimal FMBSP by showing that log-concavity in the error removes any jump discontinuities from the price function, which in turn allows us to prove that the revenue optimal FBMSP is interval.

LEMMA 2 **(Properties of Log-Concave Error)**. *Suppose that $V = \mu(\mathbf{X}) + \epsilon$ where $\epsilon$ is log-concave with density $f_\epsilon$, $\mathbf{X} \perp\!\!\!\perp \epsilon$, and $\mathbb{E}[\epsilon] = 0$. Then,*

*(a) $p_\epsilon(x)$ is an increasing and continuous function.*

*(b) The optimal segmentation is interval.*

*(c) The price on each segment $p_\epsilon(\mathcal{X}_i) = p_\epsilon(\mu(\mathbf{x}))$ for some $\mathbf{x} \in \mathcal{X}_i$.*

Lemma 2 shows that, by assuming the error in the regression model is log-concave, all the previously mentioned pathologies vanish. First, we show that $p_\epsilon(x)$ becomes a strictly increasing function which, combined with Lemma 1, implies that the revenue function $\mathcal{R}_\epsilon(x)$ is differentiable everywhere, and it's derivative is simply the sale probability. We then show in (b) that the upshot of this additional smoothness for FBMSP is that the segmentation policy becomes interval. Moreover, the optimal price to offer on each segment is contained *in the segment*, as the optimal price for some feature vector inside the segment. This locality of the price and segment then enables fast computation of the optimal policy via dynamic programming, as we describe in the main theorem for this section.

THEOREM 2 **(Computing Feature-Based Market Segmentation)**. *Suppose that $V = \mu(\mathbf{X}) + \epsilon$ where $\epsilon$ is log-concave with density $f_\epsilon$, $\mathbf{X} \perp\!\!\!\perp \epsilon$, and $\mathbb{E}[\epsilon] = 0$. Let $n = |supp(\mu(\mathbf{X}))|$ and suppose $\mathcal{R}_\epsilon(x)$ for fixed $x$ can be computed in time $m_\epsilon$. Then the optimal feature-based market segmentation can be computed in $O(kn^3 m_\epsilon)$.*

Theorem 2 is our main result, and states that by leveraging the structural properties in Lemma 2 the optimal policy can be computed quickly and efficiently in terms of the size of the support of the regression model. Note, we assume $\mu(\mathbf{X})$ is finitely supported, we believe this is natural and corresponds to simply running your regression model over a sample of customers. Further we assume the running time to compute $\mathcal{R}_\epsilon(x)$ is bounded, again we believe this assumption is natural since when $\epsilon$ is log-concave, the revenue function $p\Pr(x+\epsilon \geq p)$ is concave in $p$ and can be computed simply by checking first order conditions. We also note for discrete log-concavity Saumard and Wellner (2014) of the error, our results continue hold without modification.

In this section, we have characterized when and how we can compute FBMSP optimally, in the subsequent sections we turn our attention to it efficacy as a revenue management strategy.

# 4. Analyzing Feature-Based Market Segmentation and Pricing

In the previous section, we studied how to compute the optimal FBMSP under some assumptions, for a given number of segments/prices $k$. We showed that while, in general, it is hard to do so, in the important and realistic case when error is log-concave, the optimal policy has an intuitive structure that allows for easy computation. In this section, we continue to build on the structural insights of the last section, and show that beyond just computation, optimal FBMSP inherits a number of attractive properties and performance guarantees that may help guide practitioners in implementing such policies, and particularly in deciding how to choose $k$.

## 4.1. FBMSP vs. Feature-Based Personalized Pricing

In this subsection, we study the relative gaps between the optimal FBMSP and the natural upper bound of feature-based personalized pricing, paying close attention to how this gap informs a good choice of $k$. As mentioned in the introduction, FBMSP closely resembles real-world data-driven semi-personalized pricing strategies where sellers are constrained in the number of the segments/prices they can offer. Specifically, in FBMSP the number of prices and segments is capped at $k$ whereas feature-based personalized pricing is equivalent to FBMSP when $k \to \infty$. In fact, for any market where valuations are distributed according $V = \mu(\mathbf{X}) + \epsilon$, the revenue of a seller implementing feature-based personalized pricing can be succinctly described as an expectation over the revenue function i.e., $\lim_{k\to\infty} \mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon} =$

$\mathbb{E}_{x\sim F_{\mathbf{X}}}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))]$, since the seller offers the optimal price for each context $\mathbf{x}$ which garners revenue $\mathcal{R}_\epsilon(\mu(\mathbf{x}))$.

Intuitively, a good choice of $k$ should be one that is not too large, so as to be implementable, but one that is still close to the maximum achieveable revenue, i.e., one that minimizes

$$E_{\mathbf{X}\sim F_{\mathbf{X}}}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon}. \tag{2}$$

One difficulty that may be encountered when attempting to choose $k$ to minimize Eq. (2) is that it is sensitive to the error distribution $\epsilon$, which may be hard to know precisely or require extensive market research to obtain. It may be preferable for an analyst attempting to choose $k$ to work with an upper bound on this difference that is agnostic to the true error distribution. Interestingly, by assuming there is no error an analyst can achieve precisely such an upper bound. In the following theorem we show that one can bound the loss between FBMSP and feature-based personalized pricing by examining the gap between the two policies when $\epsilon$ is assumed to be 0 a.s. We will refer to this loss as *model market loss*, since it depends only on $\mu(\mathbf{X})$ and not the underlying error distribution.

Theorem 3 **(Model Loss vs. True Loss)**. *Suppose* $V = \mu(\mathbf{X}) + \epsilon$, $\mathbf{X} \perp\!\!\!\perp \epsilon$, *and* $\mathbb{E}[\epsilon] = 0$. *Then,*

$$\underbrace{\mathbb{E}_{\mathbf{X}\sim F_{\mathbf{X}}}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon}}_{\textit{Actual Market Loss}} \leq \underbrace{\mathbb{E}_{X\sim F_{\mathbf{X}}}[\mu(X)] - \mathcal{R}_{kXP}^{\mu(\mathbf{X})}}_{\textit{Model Market Loss}} .$$

Theorem 3 gives a theoretical foundation through which an analyst can analyze the performance of FBMSP for various $k$ directly in the model without worrying about the particular form of the error distribution.

Further, we note that the proof of Theorem 3 is constructive, and implies a simple heuristic for setting feature-based market segmentation and pricing strategies when $\epsilon$ is not log-concave, or $\epsilon$ is unknown. In these instances, a seller can simply compute the optimal $k$-FBMSP letting $\epsilon$ be 0 a.s. In this situation, the optimal policy is interval and can be described by segmentation end points $\{s_i\}_{i=0}^k$ on the model market $\mu(\mathbf{X})$, which can be used to generate the segments $\mathcal{X}_i$. From those segments, since $\epsilon$ is either unknown or not tractable to work with computationally, the firm can instead perform price experimentation to learn the prices that maximize $p_\epsilon(\mathcal{X}_i)\Pr(s_i + \epsilon \geq p_\epsilon(\mathcal{X}_i))$, and offer that price on each segment. While both the partition into segments $\{\mathcal{X}_i\}_{i=1}^k$, and the prices offered on each segment $\{p_i\}_{i=1}^k$ may be sub-optimal, such a strategy is guaranteed to earn more than $\mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon} + \mathcal{R}_{kxP}^{\mu(\mathbf{X})} - \mathbb{E}_{X\sim F_{\mathbf{X}}}[\mu(X)]$ by rearranging Theorem 3.

Theorem 3 allows an analyst to search for a choice of $k$ without referring to the error distribution, a next natural question to ask is how long can it take for $\mathcal{R}_{kXP}$ to converge to the revenue of feature-based personalized pricing. In our next theorem we show this convergence is linear in $k$, and quite fast when the range of valuations is not too wide.

THEOREM 4 **(Bounded Loss with $k$ Segments)**. *Suppose $V = \mu(\mathbf{X}) + \epsilon$, $\mathbf{X} \perp\!\!\!\perp \epsilon$, and $\mathbb{E}[\epsilon] = 0$. Let $L = \mathcal{R}_\epsilon(x_L)$ and $U = \mathcal{R}_\epsilon(x_U)$, then*

$$\mathbb{E}_{\mathbf{X}\sim F_{\mathbf{X}}}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon} \leq \frac{U-L}{k}.$$

The proof of Theorem 4 constructs a (suboptimal) segmentation strategy by equally partitioning the quantile space. Interestingly, the dependence $O\left(\frac{1}{k}\right)$ appears typical for many valuation distributions. Intuitively, this behavior can be explained in the following way. As we segment into smaller pieces, any distribution with a smooth density appears locally uniform on each segment. Example EC.3 establishes that the convergence rate for a uniform matches Theorem 4 up to constant factors, suggesting that, at least for large $k$, the rate should also be approximately tight for many distributions.

### 4.2. Revenue Concavity in the Number of Segments

Theorem 3 and Theorem 4 give an analyst insight into how to handle the error when searching for $k$, and a bound on how large a $k$ may be necessary to achieve a desired level of revenue loss. In the final result of this section we will show that the analyst can search for this choice of $k$ incrementally. Specifically, in Theorem 5, we show that when the residual is log-concave, the revenue of FBMSP is concave in the number of segments.

THEOREM 5 **(Segmentation Convexity)**. *Suppose that $V = \mu(\mathbf{X}) + \epsilon$ where $\epsilon$ is log-concave with density $f_\epsilon$, $\mathbf{X} \perp\!\!\!\perp \epsilon$, and $\mathbb{E}[\epsilon] = 0$. Then $\{\mathcal{R}_{(k+1)XP} - \mathcal{R}_{kXP}\}_{k=1}^\infty$ is a non-increasing sequence.*

The proof idea of Theorem 5 is to construct a feasible k segmentation and pricing, using the optimal FBMSP of size $k-1$ and $k+1$ segmentation, such that the revenue of $k$ segmentation and pricing is at least the average of $k-1$ segmentation and $k+1$ segmentation.

Combining Theorems 3 and 4 and Theorem 5, an analyst may use the elbow method on the model market to find a small number of segments with particularly good performance.

As we will see in our numeric case study in Section 5, FBMSP converges to feature-based personalized pricing exceptionally quickly, and this best number of segments can be easily detected via an elbow method assuming no error in the model.

## 5. Case Study: Setting Mortgage Interest Rates

In Section 3, we showed how to find jointly optimal FBMSP when a firm has trained a regression-based valuation model with independent, log-concave residuals. Then in Section 4, we provided a set of results to aid in the analysis of FBMSP policies and guide the choice of $k$, the number of segments/prices. In this section, we perform a case study to highlight some features of our approach, which we compare and contrast with prominent heuristic-based approaches for segment-then-price. Specifically, using a real data set of home mortgage offers and acceptances in Pennsylvania in 2020, we build a probit regression model to predict the probability that an applicant will take a mortgage at a given the interest rate. Then, we transform the probit regression model into a model of customer valuation measured as the maximum interest rate they will accept. We then compare our optimal method for FBMSP with segment-then-price (STP) via a number of different simulations on the data set. All data and code for this section are publicly available at [**Blinded for Review**].

| Variable | Type | Description and Statistics |
|---|---|---|
| Action taken | Binary | The action taken on the covered loan or application<br>• 1 (accepted), Frequency = 11491, Percent = 77.0%<br>• 0 (rejected), Frequency = 3425, Percent = 23.0% |
| Interest rate | Continuous | The interest rate for the covered loan or application (%)<br>• Mean = 3.4%, Std = 0.9% |
| Income | Continuous | Applicant's gross annual income (in thousands of dollars)<br>• Mean = 110.08, Std = 94.76 |
| Derived race | Binary | Single aggregated race categorization derived from applicant race fields<br>• 1 (white), Frequency = 12724, Percent = 85.3%<br>• 0 (not white), Frequency = 2192, Percent = 14.7% |
| Derived gender | Binary | Single aggregated gender categorization derived from applicant gender fields<br>• 1 (joint), Frequency = 5799, Percent = 38.9%<br>• 0 (male or female), Frequency = 9117, Percent = 61.1% |

**Table 1** **Descriptions and summary statistics for explanatory variables.**

### 5.1. Description of data set

Our case study is based on a data set collected in accordance with the Home Mortgage Disclosure Act (HMDA)[1]. Specifically, we downloaded the data provided by all financial institutions in Pennsylvania who had offered a loan for the purpose of enabling a home purchase. The data set includes all the information for approved applications in 2020, including demographic information about the applicant, their income level, and the interest rate and loan amount the bank offered. After removing unsuitable rows, there were 14,916 approved applications in total, and 11,491 (77%) of approved applications resulted in a loan accepted at a bank offered interest rate. Table 1 summarizes the variables (feature of customers) we use in our case study.

As a preliminary, note we can think of the interest rate the bank offers on the loan as a take-it-or-leave-it price, and the customers choice whether or not to accept the loan as a decision to purchase or not purchase at a given price. Using the price variation in this data, we will estimate a customer valuation model so that we can train and evaluate market segmentation and pricing models. In the first step, we use a probit regression model to predict the probability that a customer will take the offered interest rate. Table 2 shows the coefficient estimates for the probit regression model, and Fig. EC.3 shows the prediction of the probability a customer will take the approved application for a given interest rate. We then transform our probit model into a linear valuation model of the form $V|\mathbf{x} = \mu(\mathbf{x}) + \mathcal{N}(0,\sigma)$, where $\mathbf{x}$ is a feature vector including the interest rate, income level of the customer, and demographic information, and $\mu(\mathbf{x}) = \mathbf{x}^t\beta$. Our transformation from probit regression model to linear valuation model follows Cameron and James (1987), for a short primer describing such transformations, see Section C in the appendix. All subsequent feature based segmentation and pricing policies will be based on this derived linear valuation model, $V|\mathbf{x} = \mu(\mathbf{x}) + \mathcal{N}(0,\sigma)$.

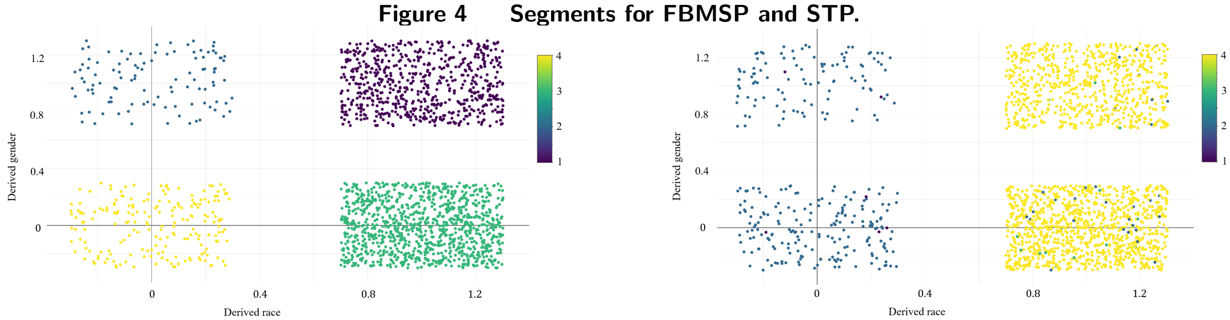
### 5.2. Comparison with Segment-then-Price

To access the real impact of our optimal FBMSP policies, we will compare against heuristic segment-then-price (STP) policies. For STP, we will segment customers using the popular $k$-mediods algorithm (Reynolds et al. 2006, Schubert and Rousseeuw 2019, 2021) with Gower distance (Gower 1966, 1967). The price optimization is then done over the found segments and can be computed in polynomial time for error with finite support.

[1] The HMDA website where the data is hosted is `https://ffiec.cfpb.gov/`

First, we look at the segments generated by segment-then-price and our FBMSP model. Fig. 4 shows that STP will group customers first based on differences in gender and/or race. Gender and race are certainly heterogeneous across our data set, however, these differences are not necessarily the distinctions that are revenue-maximizing to delineate. In comparison, optimal FBMSP will segment customers into different groups based on their valuations, which is only weakly correlated with gender/race in our data. Therefore, compared to STP, FBMSP will not only achieve better revenue but does so in an explainable way by grouping customers with similar predicted valuations, instead of merely similar demographic features which may have negative social or legal ramifications.

**Figure 4 Segments for FBMSP and STP.**




*Note.* Here we plot segments for STP and FBMSP when the number of segments $k = 4$. Since both derived gender and derived race are binary variables, we add some random noise to each point for clarity of presentation (without noise, all points of the same color would be on top of one another in the left panel). In the left panel, the segments are obtained using $k$-medoids algorithm. In the right panel, we use dynamic programming to do optimal FBMSP.

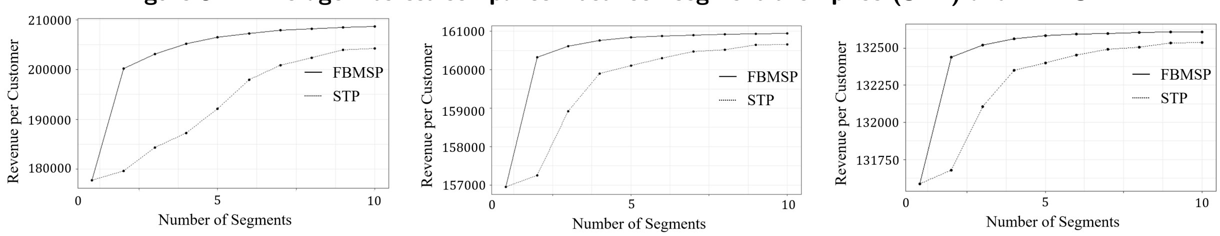
To compare the difference in revenue garnered by STP and FBMSP, we will examine the difference in total interest a customer will pay on average over the lifetime of the loan (i.e. average revenue per customer), where the interest is calculated using the standard fixed monthly payment formula (Capinski and Zastawniak (2003)). In Fig. 5, we plot the expected revenue the firm can get from one customer on average, against the number of segments. We first note that the revenue per customer is increasing for both FBMSP and

| Variable | Estimate | Std. Error | z value | $\Pr(>\lvert z\rvert)$ | Significance |
|---|---|---|---|---|---|
| (Intercept) | 3.8869 | 0.1126 | 34.5314 | $< 2.2 \times 10^{-16}$ | *** |
| Interest Rate | -0.8704 | 0.0277 | -31.4735 | $< 2.2 \times 10^{-16}$ | *** |
| Income | -0.0009 | 0.0002 | -4.1142 | $3.885 \times 10^{-05}$ | *** |
| Derived race | 0.4709 | 0.0545 | 8.6374 | $< 2.2 \times 10^{-16}$ | *** |
| Derived gender | 0.1721 | 0.0441 | 3.9003 | $9.606 \times 10^{-5}$ | *** |

**Table 2 Probit regression coefficients. Significance levels: ***: $< 0.001$, **: $< 0.01$, *: $< 0.05$.**

STP model. However for FBMSP, the revenue the firm can get from each customer on average, is concave in the number of segments, while the STP model is not. In our case study, STP often gets stuck at small choices of $k$ and requires a 3+ of segments before it can achieve strong revenue, whereas FBMSP is guaranteed to get the most revenue out of a small number of segments. We see the differences then between FBMSP and STP are most pronounced when only a small number of segments are used, which is precisely the case of interest in industry. We further note that for both models while smaller error in the prediction model will yield higher revenue, the gap between the two strategies is also more pronounced when the error is small, suggesting that for sophisticated firms with high quality feature data the benefits of FBMSP are even greater.

**Figure 5 Average interest comparison between segment-then-price (STP) and FBMSP.**




*Note.* Here we plot the average interest per customer for STP and FBMSP for different levels of prediction error in our valuation model. In the left panel, the standard deviation of prediction error is $\sigma = 0$, in the middle panel, the standard deviation of prediction error is $\sigma = 0.5$, in the right panel, the standard deviation of prediction error is $\sigma = 1$.

### 5.3. Finding the Optimal Number of Segments using Regression Model

One additional benefit of the concavity of optimal FBMSP is it enables us to easily choose the number of segments via the elbow method heuristic. The elbow method is the most commonly used heuristic for finding the optimal number of segments for unsupervised learning. The intuition is that one should choose a number of segments so that adding another segment doesn't give much better modeling of the data (see Bholowalia and Kumar (2014) for more discussion about the elbow method and its applications). To use the elbow method, one prerequisite is that the objective function is monotone in the number of segments. In general, the objective function, revenue per customer, is not necessarily even increasing for the STP. Unlike STP, in Theorem 5 we showed that the revenue is concave in the number of segments $k$. At some value for $k$, the revenue increases dramatically, and after that, it reaches a plateau, and increasing the number of segments does not dramatically increase revenue. In Fig. 5, for our FBMSP model, 2 or 3 is the elbow of the revenue per customer vs. $k$ plot, whereas for STP, the possible elbow is at 8 or 9, a prohibitively large number of segments in practice.

## 6. Conclusions

Increasingly rich consumer profiles and choice models enable retailers to personalize consumers at finer and finer levels. However, building such tools comes at an investment cost in the form of technology, data scientists, marketing, etc. Motivated by this trade-off, and by a desire to improve on common heuristic approaches, we provide a framework to compute and analyze semi-personalized, feature-based market segmentation and pricing policies under realistic assumptions about how firms predict the valuations of customers.

Specifically, we define and study the feature-based market segmentation and pricing problem, where sellers have trained a regression model to predict customers' valuations using their features. We first prove the computation of optimal feature-based market segmentation and pricing is NP-hard for independent residuals, and provide a $(1 - 1/e)$ approximation algorithm by leveraging the greedy algorithm in the submodular maximization problem. With an additional assumption of log-concavity on the prediction error, we show that the optimal policy is cubic-time solvable with dynamic programming.

Further, we analyze the properties of optimal feature-based market segmentation and pricing. We prove a number of results for the revenue of $k$ segmentation against a fully personalized pricing. We show that the loss of $k$ segmentation is upper bounded the loss without prediction error. We also show the revenue of optimal FBMSP is concave in the number of segments $k$, practitioners can find the most suitable $k$ by the simple elbow method, and without loss of much revenue.

Overall, our work seeks to deepen our understanding of semi-personalized pricing strategies, and demonstrate that they are computable, and effective when compared to complicated fully personalized pricing strategies. There are many interesting and important directions to consider for future work, we highlight three of them here. First, this paper assumes the production cost of the good is uniform over all segments. Follow-up work may consider heterogeneous production costs among different segments, and ask whether the optimal FBMSP in this case still uses interval segments when the residuals are log-concave. Second, it may also be interesting to consider the approximation ratio for interval segmentations facing general error distributions. Example EC.2 demonstrates that interval segmentation is not optimal for some error distributions, but how far it is from the optimal segmentation in the worst case is unknown. Finally, we assume the firm can charge customers in different segments any segment level price. In practice, the firm may only be able to offer a price menu for customers to choose from it. One may also consider models similar to FBMSP in the price menu setting.

## References

Aouad, Ali, Adam N Elmachtoub, Kris J Ferreira, Ryan McNellis. 2019. Market segmentation trees. *arXiv preprint arXiv:1906.01174* .

Assael, Henry, A Marvin Roscoe Jr. 1976. Approaches to market segmentation analysis. *Journal of Marketing* **40**(4) 67–76.

Athey, Susan, Guido Imbens. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* **113**(27) 7353–7360.

Bagnoli, Mark, Ted Bergstrom. 2005. Log-concave probability and its applications. *Economic theory* **26**(2) 445–469.

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

Bertsimas, Dimitris, Jack Dunn, Nishanth Mundru. 2019. Optimal prescriptive trees. *INFORMS Journal on Optimization* **1**(2) 164–183.

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

Bholowalia, Purnima, Arvind Kumar. 2014. Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications* **105**(9).

Biggs, Max, Wei Sun, Markus Ettl. 2021. Model distillation for revenue optimization: Interpretable personalized pricing. *International Conference on Machine Learning*. PMLR, 946–956.

Buzzacchi, Luigi, Tommaso M Valletti. 2005. Strategic price discrimination in compulsory insurance markets. *The Geneva Risk and Insurance Review* **30**(1) 71–97.

Cameron, Trudy A, Michelle D James. 1987. Estimating willingness to pay from survey data: an alternative pre-test-market evaluation procedure. *Journal of Marketing Research* **24**(4) 389–395.

Capinski, Marek, Tomasz Zastawniak. 2003. Mathematics for finance. *An Introduction* 118–124.

Chen, Fangruo. 2001. Market segmentation, advanced demand information, and supply chain performance. *Manufacturing & Service Operations Management* **3**(1) 53–67.

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

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

Claycamp, Henry J, William F Massy. 1968. A theory of market segmentation. *Journal of Marketing Research* **5**(4) 388–394.

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

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

Courty, Pascal, Mario Pagliero. 2012. The impact of price discrimination on revenue: Evidence from the concert industry. *Review of Economics and Statistics* **94**(1) 359–369.

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

Dolgui, Alexandre, Jean-Marie Proth. 2010. Pricing strategies and models. *Annual Reviews in Control* **34**(1) 101–110.

Dubé, Jean-Pierre, Zheng Fang, Nathan Fong, Xueming Luo. 2017. Competitive price targeting with smartphone coupons. *Marketing Science* **36**(6) 944–975.

Elmachtoub, Adam, Jason Cheuk Nam Liang, Ryan McNellis. 2020. Decision trees for decision-making under the predict-then-optimize framework. *International Conference on Machine Learning*. PMLR, 2858–2867.

Elmachtoub, Adam N, Paul Grigas. 2022. Smart predict, then optimize. *Management Science* **68**(1) 9–26.

Elmachtoub, Adam N, Vishal Gupta, Michael L Hamilton. 2021. The value of personalized pricing. *Management Science* .

Elmachtoub, Adam N, Michael L Hamilton. 2021. The power of opaque products in pricing. *Management Science* **67**(8) 4686–4702.

Ferreira, Kris Johnson, Bin Hong Alex Lee, David Simchi-Levi. 2016. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & service operations management* **18**(1) 69–88.

Gower, John C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**(3-4) 325–338.

Gower, John C. 1967. A comparison of some methods of cluster analysis. *Biometrics* 623–637.

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

Hart, Sergiu, Philip J Reny. 2015. Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics* **10**(3) 893–922.

Huang, Jiali, Ankur Mani, Zizhuo Wang. 2019. The value of price discrimination in large random networks. *Proceedings of the 2019 ACM Conference on Economics and Computation*. 243–244.

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

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

Kallus, Nathan. 2017. Recursive partitioning for personalization using observational data. *International conference on machine learning*. PMLR, 1789–1798.

Katz, Michael L. 1984. Price discrimination and monopolistic competition. *Econometrica: Journal of the Econometric Society* 1453–1471.

Kleinberg, Jon, Christos Papadimitriou, Prabhakar Raghavan. 1998. A microeconomic view of data mining. *Data mining and knowledge discovery* **2**(4) 311–324.

Kleinberg, Jon, Christos Papadimitriou, Prabhakar Raghavan. 2004. Segmentation problems. *Journal of the ACM (JACM)* **51**(2) 263–280.

Krause, Andreas, Daniel Golovin. 2014. Submodular function maximization. *Tractability* **3** 71–104.

Lehmann, Sonja, Peter Buxmann. 2009. Pricing strategies of software vendors. *Business & Information Systems Engineering* **1**(6) 452–462.

Li, Shujing, Jiaping Qiu. 2014. Financial product differentiation over the state space in the mutual fund industry. *Management Science* **60**(2) 508–520.

Liu, Ying, Sudha Ram, Robert F Lusch, Michael Brusco. 2010. Multicriterion market segmentation: a new model, implementation, and evaluation. *Marketing Science* **29**(5) 880–894.

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

Nemhauser, George L, Laurence A Wolsey, Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functionsi. *Mathematical programming* **14**(1) 265–294.

Niu, Chaoyue, Zhenzhe Zheng, Fan Wu, Shaojie Tang, Guihai Chen. 2020. Online pricing with reserve price constraint for personal data markets. *IEEE Transactions on Knowledge and Data Engineering* .

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

Reynolds, Alan P, Graeme Richards, Beatriz de la Iglesia, Victor J Rayward-Smith. 2006. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *Journal of Mathematical Modelling and Algorithms* **5**(4) 475–504.

Saumard, Adrien, Jon A Wellner. 2014. Log-concavity and strong log-concavity: a review. *Statistics surveys* **8** 45.

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

Schubert, Erich, Peter J Rousseeuw. 2019. Faster k-medoids clustering: improving the pam, clara, and clarans algorithms. *International conference on similarity search and applications*. Springer, 171–187.

Schubert, Erich, Peter J Rousseeuw. 2021. Fast and eager k-medoids clustering: O (k) runtime improvement of the pam, clara, and clarans algorithms. *Information Systems* **101** 101804.

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

Shukla, Naman, Arinbjörn Kolbeinsson, Ken Otwell, Lavanya Marla, Kartik Yellepeddi. 2019. Dynamic pricing for airline ancillaries with customer context. *Proceedings of the 25th ACM SIGKDD International Conference on knowledge discovery & data mining*. 2174–2182.

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

Thomas, R Guy. 2012. Non-risk price discrimination in insurance: market outcomes and public policy. *The Geneva Papers on Risk and Insurance-Issues and Practice* **37**(1) 27–46.

Varian, Hal. 1989. Price discrimination. R. Schmalensee, R. Willig, eds., *Handbook of Industrial Organization*, vol. 1, 1st ed., chap. 10. Elsevier, 597–654. URL `https://EconPapers.repec.org/RePEc:eee:indchp:1-10`.

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

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

Ziya, Serhan, Hayriye Ayhan, Robert D Foley. 2004. Relationships among three assumptions in revenue management. *Operations Research* **52**(5) 804–809.

## Appendix A: Omitted Examples

Example EC.1 ($p_\epsilon(x)$ can be discontinuous). Suppose $\mu(\mathbf{X}) \sim$ Uniform$[1,2]$ (or any other continuous distribution on [1,2]) and $\epsilon$ is either -.5 or .5 with probability $\frac{1}{2}$. Then for every $x \leq 1.5$, the optimal price is $p_\epsilon(x) = x + .5$, and $p_\epsilon(x) = x - .5$ otherwise. Thus at 1.5 $p_\epsilon(x)$ is discontinuous, and by Lemma 1 the revenue function $\mathcal{R}_\epsilon(x)$ is non-differentiable. □

Example EC.2 (Optimal Segmentations Need Not be Interval). In this example we give $\mu(\mathbf{X})$ and $\epsilon$ such that the optimal segmentation and pricing is non-interval. Specifically, for any number $k \geq 2$, assume $\mu(\mathbf{X})$ is uniformly distributed on the set $\{k, k+2, k+4, \cdots, 3k\}$, and $\epsilon$ is either $k$ or $-k$ with probability $\frac{1}{2}$, respectively. It can be computed without too much difficulty that the unique optimal $k$-market segmentation and pricing uses segments,

$$\mathcal{X}_1 = \{k, 3k\}, \mathcal{X}_2 = \{k+2\}, \cdots \mathcal{X}_k = \{k + 2(k-1)\},$$

with corresponding price for each segment,

$$p(\mathcal{X}_1) = 2k, p(\mathcal{X}_2) = 2k+2, \cdots, p(\mathcal{X}_k) = 2k + 2(k-1).$$

As the first segment is not interval, the optimal segmentation needs not to be interval for any $k$. □

Example EC.3 (Tightness of Theorem 4). Suppose the regression model has no error, i.e. $V|\mathbf{x} = \mu(\mathbf{x})$, and let $\mu(\mathbf{X}) \sim$ Uniform$[0,t]$ for some $t > 0$. Then, $\mathbb{E}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] = \mathbb{E}[V] = \frac{t}{2}$. To compute $\mathcal{R}_{kXP}$ for some $k$, note by Theorem 2 the optimal segmentation here is interval and further each segment can be described by a left and right endpoint in the space of predicted valuations. Let $0 = s_0, \ldots, s_{k+1} = t$ describe those segments (i.e. $\mathcal{X}_i = \{x | \mu(x) \in [s_i, s_{i+1})\}$ with corresponding prices $p_1, \ldots, p_k$. It is easy to see since $\epsilon = 0$, the optimal price and segmentations must satisfy $s_i = p_i$ for $i = 1, \ldots, k$.

Now, on segment $\mathcal{X}_i$, the conditional distribution of $V$ is still uniform, so the contribution of that segment to $\mathbb{E}[V]$ is $\frac{s_{i+1}+s_i}{2} \cdot \frac{s_{i+1}-s_i}{t}$ for all $i$, since only $\frac{s_{i+1}-s_i}{t}$ fraction of the market is in this interval. By contrast, for $i = 1, \ldots, k-1$, the $k$-market segmentation strategy on segment $i$ earns revenue $p(\mathcal{X}_i)\Pr(\mu(\mathbf{X}) \geq p(\mathcal{X}_i) | \mathcal{X} \in [s_i, s_{i+1})\Pr(\mathbf{X} \in [s_i, s_{i+1}) = s_i \frac{s_{i+1}-s_i}{t}$ since $p_i = s_i$ and thus, all customers in the segment buy. The difference in revenue is then

$$\begin{aligned}\mathbb{E}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP} &= \frac{s_1^2}{2t} + \sum_{i=1}^{k} \frac{s_{i+1}+s_i}{2} \cdot \frac{s_{i+1}-s_i}{t} - s_i \frac{s_{i+1}-s_i}{t} \\ &= \frac{s_1^2}{2t} + \frac{1}{2t}\sum_{i=1}^{k}(s_{i+1}-s_i)^2 = \frac{1}{2t}\sum_{i=0}^{k}(s_{i+1}-s_i)^2.\end{aligned}$$

By inspection, for a fixed $s_1$, the segmentation which minimizes this difference is equispaced, i.e., $s_i = s_{i-1} + \frac{t}{k+1}$ for $i = 1, \ldots, k$. Plugging in gives $\mathbb{E}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP} = \frac{t}{2(k+1)} = \frac{E[V]}{k+1}$. □

Example EC.4 (Cluster then price is not necessarily increasing in $k$.). asd □

## Appendix B: Omitted Proofs

### B.1. Omitted Proofs from Section 2

*Proof of Lemma 1.* **(a)** Fix some $\epsilon$ and positive real numbers $x_1$, $x_2$ such that $x_1 < x_2$ and recall $\theta_\epsilon(x) := p_\epsilon(x) - x$ is the difference between the price and $x$. Further recall $p_\epsilon(x_1), p_\epsilon(x_2)$ are prices that maximize $p\overline{F}_\epsilon(p - x_1)$ and $p\overline{F}_\epsilon(p - x_2)$ respectively. Thus, by optimality we have the following two inequalities

$$(x_1 + \theta_\epsilon(x_1))\,\overline{F}_\epsilon(\theta_\epsilon(x_1)) \geq (x_1 + \theta_\epsilon(x_2))\,\overline{F}_\epsilon(\theta_\epsilon(x_2)), \tag{EC.1}$$

$$(x_2 + \theta_\epsilon(x_2))\,\overline{F}_\epsilon(\theta_\epsilon(x_2)) \geq (x_2 + \theta_\epsilon(x_1))\,\overline{F}_\epsilon(\theta_\epsilon(x_1)). \tag{EC.2}$$

Rearranging the two inequalities yields,

$$\frac{x_1 + \theta_\epsilon(x_1)}{x_1 + \theta_\epsilon(x_2)} \geq \frac{\overline{F}_\epsilon\,(\theta_\epsilon(x_2))}{\overline{F}_\epsilon\,(\theta_\epsilon(x_1))} \geq \frac{x_2 + \theta_\epsilon(x_1)}{x_2 + \theta_\epsilon(x_2)}.$$

Consequently,

$$(x_1 + \theta_\epsilon(x_1))\,(x_2 + \theta_\epsilon(x_2)) \geq (x_1 + \theta_\epsilon(x_2))\,(x_2 + \theta_\epsilon(x_1))\,,$$

Simplifying the expression, we get

$$(x_2 - x_1)\,\theta_\epsilon(x_1) \geq (x_2 - x_1)\,\theta_\epsilon(x_2).$$

Finally, noting $x_2 - x_1 > 0$, the inequality is equivalent to $\theta_\epsilon(x_1) \geq \theta_\epsilon(x_2)$ and thus the margin monotone decreasing.

**(b)** Again, Fix some $\epsilon$ and positive real numbers $x_1$, $x_2$ such that $x_1 \leq x_2$. Then $x_1 + \epsilon \leq_{\text{st}} x_2 + \epsilon$ in the sense of first order stochastic dominance, and it is well known that stochastic dominance of the valuations implies $\mathcal{R}_\epsilon(x_1) \leq \mathcal{R}_\epsilon(x_2)$ (see for instance Hart and Reny (2015) for an extended discussion). Combining this observation with Eqs. (EC.1) and (EC.2) above yields,

$$\mathcal{R}_\epsilon(x_2) - \mathcal{R}_\epsilon(x_1) \geq (x_2 + \theta_\epsilon(x_1))\,\overline{F}_\epsilon(\theta_\epsilon(x_1)) - (x_1 + \theta_\epsilon(x_1))\,\overline{F}_\epsilon(\theta_\epsilon(x_1)) = (x_2 - x_1)\overline{F}_\epsilon(\theta_\epsilon(x_1)),$$
$$\mathcal{R}_\epsilon(x_2) - \mathcal{R}_\epsilon(x_1) \leq (x_2 + \theta_\epsilon(x_2))\,\overline{F}_\epsilon(\theta_\epsilon(x_2)) - (x_1 + \theta_\epsilon(x_2))\,\overline{F}_\epsilon(\theta_\epsilon(x_2)) = (x_2 - x_1)\overline{F}_\epsilon(\theta_\epsilon(x_2)).$$

Dividing through both sides by $x_2 - x_1$ gives,

$$\overline{F}_\epsilon(\theta_\epsilon(x_1)) \leq \frac{\mathcal{R}_\epsilon(x_2) - \mathcal{R}_\epsilon(x_1)}{x_2 - x_1} \leq \overline{F}_\epsilon(\theta_\epsilon(x_2)). \tag{EC.3}$$

When $p_\epsilon$ is continuous then $\theta_\epsilon$ is also continuous, and taking $x_1 \to x_2$ squeezes the derivative to be $\overline{F}_\epsilon(p_\epsilon(x) - x)$ as desired.

**(c)** $\mathcal{R}_\epsilon(x)$ was noted to be increasing in the proof of (b). Now to prove continuity, fix $\epsilon$ and positive real numbers $x_1$, $x_2$ such that $x_1 < x_2$. Then,

$$\mathcal{R}_\epsilon(x_1) \leq \mathcal{R}_\epsilon(x_2) = (x_2 + \theta_\epsilon(x_2) + (x_2 - x_1) - (x_2 - x_1))\,\overline{F}_\epsilon(\theta_\epsilon(x_2)) \leq \mathcal{R}_\epsilon(x_1) + (x_2 - x_1),$$

where the last inequality follows from distributing and applying Eq. (EC.2), and the fact that $\overline{F}_\epsilon(\cdot) \leq 1$. Taking $x_1 \to x_2$ gives us the continuity of $\mathcal{R}_\epsilon(x)$.

For convexity, again fix positive real numbers $x_1$, $x_2$ and also $\lambda \in (0,1)$. Then,

$$\begin{aligned}\mathcal{R}_\epsilon(x_1) &= p_\epsilon(x_1)\overline{F}_\epsilon(p_\epsilon(x_1)-x_1)\\ &\geq (p_\epsilon(\lambda x_1+(1-\lambda)x_2)+(1-\lambda)(x_1-x_2))\,\overline{F}_\epsilon(p_\epsilon(\lambda x_1+(1-\lambda)x_2)-(\lambda x_1+(1-\lambda)x_2)),\end{aligned}$$

where the inequality follows from noting that $p_\epsilon(x_1)$ is revenue optimal for $x_1+\epsilon$ and any other price can earn no more. Similarly,

$$\begin{aligned}\mathcal{R}_\epsilon(x_2) &= p_\epsilon(x_2)\overline{F}_\epsilon(p_\epsilon(x_2)-x_2)\\ &\geq (p_\epsilon(\lambda x_1+(1-\lambda)x_2)-\lambda(x_1-x_2))\,\overline{F}_\epsilon(p_\epsilon(\lambda x_1+(1-\lambda)x_2)-(\lambda x_1+(1-\lambda)x_2)).\end{aligned}$$

Combine the two inequalities above, we get

$$\lambda\mathcal{R}_\epsilon(x_1)+(1-\lambda)\mathcal{R}_\epsilon(x_2)\geq \mathcal{R}_\epsilon(\lambda x_1+(1-\lambda)x_2)),$$

which means $\mathcal{R}_\epsilon(x)$ is convex in $x$. □

### B.2. Omitted Proofs from Section 3

*Proof of Theorem 1.* We will prove the hardness by showing the Hitting set problem can be reduced to an instance of k feature-based market segmentation and pricing (kXP). Let $\mathcal{X}$ be the ground set of elements of size, $|\mathcal{X}|=m$, and let $\{H_i\}_{i=1}^n$ be a collection of subsets of $\mathcal{X}$. Consider the decision version of the hitting set problem, which asks whether there exists a subset of $\mathcal{X}^*\subset\mathcal{X}$, $|\mathcal{X}^*|\leq k$, such that $\mathcal{X}^*$ has non-empty intersection with each $H_i$. To build a corresponding $k$-market segmentation and pricing problem, suppose we have $n$ customers such that each customer's valuation is $\mu_i+\epsilon^2$, where $i=1,2,...,n$. Further, let $p_j=n+\frac{j-1}{m}$, for $j=1,2,...,m$. Then, let $\mu_1=p_1$, and $\mu_i=\mu_{i-1}+\frac{p_1}{2(i-1)}+\frac{p_m}{2i}$, for $i=2,...,n$. We will now construct an $\epsilon$ such that, for each $\mu_i$, $\mathcal{R}_\epsilon(\mu_i)$ is maximized at $p_j$ if and only if in the hitting set problem the subset $H_i$ contains element $x_j$.

We will construct $\epsilon$ so that it is supported on numbers of the form $p_i-\mu_j$. Before constructing $\epsilon$, note that $p_j$ is strictly increasing in $j$, and that $p_j-\mu_i<p_{j'}-\mu_i$ as long as $j'>j$. Further note $p_m-\mu_{i+1}<p_1-\mu_i$ since by the definition of $\mu_i$ and $\mu_{i+1}$, $\mu_{i+1}-\mu_i>\frac{p_1}{2n}+\frac{p_m}{2n}>1$ and since $p_m-p_1=\frac{m-1}{m}<1$. Let $t_{1,1}=p_1-\mu_n$, ..., $t_{j,i}=p_j-\mu_{n+1-i}$, ..., $t_{m,n}=p_m-\mu_1$, and let $t_{0,n}=-\mu_n$, and $t_{m,n+1}=p_m$. Thus, we have

$$t_{0,n}<t_{1,1}<t_{2,1}<\ldots<t_{m,i}<t_{1,i+1}<t_{2,i+1}<\ldots t_{m-1,n}<t_{m,n}<t_{m,n+1} \tag{EC.4}$$

Now we are ready to define the complementary cumulative distribution function (cCDF) of $\epsilon$. We will let $\epsilon$ be such $\overline{F}_\epsilon(t_{0,n})=1$, $\overline{F}_\epsilon(t_{m,n+1})=0$, and working backwards recursively from $\overline{F}_\epsilon(t_{m,n+1})$ as follows:

$$\overline{F}_\epsilon(t_{j,i})=\begin{cases}\frac{i}{p_j}, & \text{if } x_j\in H_i\\ \overline{F}_\epsilon(t_{j+1,i}), & \text{if } x_j\notin H_i \text{ and } j<m\\ \overline{F}_\epsilon(t_{1,i+1}) & \text{if } x_j\notin H_i \text{ and } j=m,\ i<n\\ 0 & \text{otherwise.}\end{cases}$$

Note this construction is well defined and is quadratically supported, an example what $\overline{F}_\epsilon$ looks like is provided in Fig. EC.4. Further, for any value $t$ such that $t_{j,i}\leq t<t_{j+1,i}$, $\overline{F}(t)=\overline{F}(t_{j,i})$. Now we need to

[2] Equivalently, $\mu(\mathbf{X})$ is uniformly supported on these valuations.

check that $\overline{F}_\epsilon$ is non-increasing and thus a properly defined cCDF, and also that $p_j\overline{F}_\epsilon(p_j-\mu_i)$ is revenue-maximizing only when $j$, $i$ are such that $x_j \in H_i$.

To the first point, since $p_j \geq n$ for all $j=1,2,...,m$, and $\{p_j\}$ is increasing, therefore, $\frac{i}{p_j} < \frac{i}{p_{j+1}}$. Then, to show $\overline{F}_\epsilon$ is non-increasing, we only need to show $\frac{i}{p_m} < \frac{i-1}{p_1}$. Note that

$$\begin{aligned}\frac{i}{p_m} - \frac{i-1}{p_1} &= \frac{ip_1-(i-1)p_m}{p_1p_m} \\ &= \frac{i(p_1-p_m)+p_m}{p_1p_m} > 0,\end{aligned}$$

where the inequality follows from the fact that $p_m - p_1 = \frac{m-1}{m} < 1$ for $1 \leq i \leq n$, and $p_m > n$. Thus $\overline{F}_\epsilon$ is non-increasing, i.e., $\overline{F}_\epsilon$ is a proper cumulative distribution function.

Next, we show that $p\overline{F}_\epsilon(p-\mu_i) = i$ iff $p = p_j$ and $x_j \in H_i$, and for all other prices the revenue $p\overline{F}_\epsilon(p-\mu_i)$ is strictly less than $i$. By the definition of $\overline{F}_\epsilon(p_j-\mu_i)$, if $x_j \in H_i$, $\overline{F}_\epsilon(p_j-\mu_i) = \frac{i}{p_j}$, consequently, $p_j\overline{F}_\epsilon(p_j-\mu_i) = i$. So now suppose price $p$ satisfy $\overline{F}_\epsilon(p-\mu_i) = \overline{F}_\epsilon(p_{j'}-\mu_{i'}) = \frac{i'}{p_{j'}}$. To simplify the discussion, we take the largest price $p$ such that $\overline{F}_\epsilon(p-\mu_i) = \overline{F}_\epsilon(p_{j'}-\mu_{i'})$, i.e., $p-\mu_i = p_{j'}-\mu_{i'}$, and by rearranging $p = p_{j'}-\mu_{i'}+\mu_i$. All other prices less than $p_{j'}-\mu_{i'}+\mu_i$ and which satisfies $\overline{F}_\epsilon(p-\mu_i) = \overline{F}_\epsilon(p_{j'}-\mu_{i'}) = \frac{i'}{p_{j'}}$ will give us less revenue. We want to show that $p\overline{F}_\epsilon(p-\mu_i) < i$, i.e.,

$$(p_{j'}-\mu_{i'}+\mu_i)\,\frac{i'}{p_{j'}} < i.$$

If $i' < i$, the inequality is the same as

$$\mu_i - \mu_{i'} \leq \frac{i-i'}{i'}p_{j'}.$$

By the definition of $p_j$ and $\mu_i$,

$$\mu_{i+1} - \mu_i = \frac{p_1}{2i} + \frac{p_m}{2(i+1)} < \frac{p_1}{i},$$

where the inequality comes from $\frac{p_m}{p_1} < \frac{n+1}{n}$. Therefore,

$$\mu_i - \mu_{i'} \leq \sum_{j=i'}^{i}\frac{p_1}{j} < \frac{i-i'}{i}p_1 < \frac{i-i'}{i}p_j.$$

Similarly, if $i' > i$, $p\overline{F}_\epsilon(p-\mu_i) < i$ is equivalent to

$$\mu_{i'} - \mu_i > \frac{i'-i}{i'}p_{j'}.$$

Now, by the definition of $p_j$ and $\mu_i$,

$$\mu_{i+1} - \mu_i = \frac{p_1}{2i} + \frac{p_m}{2(i+1)} > \frac{p_m}{i+1},$$

where the inequality comes from $\frac{p_m}{p_1} < \frac{n+1}{n}$. Therefore,

$$\mu_{i'} - \mu_i \geq \sum_{j=i'}^{i}\frac{p_m}{j+1} > \frac{i-i'}{i}p_m > \frac{i-i'}{i}p_j,$$

as desired.

Finally, to determine whether there exists a subset of $X^* \subset X$, $|X^*| \leq k$, such that $X^*$ has non-empty intersection with each $H_i$, it is equivalent to determine whether there is a $k$ market segmentation and pricing that yields revenue $1+2+3+\ldots+n = \frac{n(n+1)}{2}$. Since the hitting set problem NP-hard, thus FBMSP is also NP-hard. □

*Proof of Lemma 2.* **(a)** First we will show $p_\epsilon(x)$ is increasing. Since $f_\epsilon$ is log-concave, $\overline{F}_\epsilon$ is also log-concave (see Bagnoli and Bergstrom (2005) for an extensive overview of the transformation that preserve log-concavity). Further, $\frac{d}{dx}\log(\overline{F}_\epsilon(x)) = \frac{-f_\epsilon(x)}{\overline{F}_\epsilon}$ which by concavity implies the inverse hazard rate, $\frac{\overline{F}_\epsilon(x)}{f_\epsilon(x)}$, is decreasing in $x$. $p_\epsilon(\cdot)$ is thus unique, satisfies first order conditions for revenue optimality, $\frac{d}{dp}p\overline{F}_\epsilon(p-x)|_{p=p(x)} = 0$, and can be written as $p_\epsilon(x) = \frac{\overline{F}_\epsilon(p_\epsilon(x)-x)}{f_\epsilon(p_\epsilon(x)-x)}$, but by Lemma 1(a) $p_\epsilon(x) - x$ is decreasing thus $p_\epsilon(x)$ it must be an increasing function of $x$.

**(b)** Again note if $f(x)$ is log-concave, $\overline{F}_\epsilon$ is a log-concave function, thus it has Pólya frequency of order 2 (PF2), which is equivalent to that statement that, for any real numbers $x_1, x_2$, and $y_1, y_2$, such that $x_1 < x_2$ and $y_1 < y_2$, then $\frac{\overline{F}_\epsilon(x_1-y_2)}{\overline{F}_\epsilon(x_1-y_1)} \le \frac{\overline{F}_\epsilon(x_2-y_2)}{\overline{F}_\epsilon(x_2-y_1)}$ (see Saumard and Wellner (2014), Section 11).

Now, let $\{\mathcal{X}_i\}_1^k, \{p_\epsilon(\mathcal{X}_i)\}_1^k$ be the optimal segmentation and pricing and suppose WLOG that the prices are distinct. Further suppose the optimal segmentation was not interval, then there exists $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ such that $\mu(\mathbf{x}_1) < \mu(\mathbf{x}_2) < \mu(\mathbf{x}_3)$, but with $\mathbf{x}_1, \mathbf{x}_3 \in \mathcal{X}_i$, and $\mathbf{x}_2 \in \mathcal{X}_j$ for some $i \neq j$. Suppose $p_\epsilon(\mathcal{X}_i) < p_\epsilon(\mathcal{X}_j)$ (the opposite case when $p_\epsilon(\mathcal{X}_i) > p_\epsilon(\mathcal{X}_j)$ follows by an identical argument, swapping $\mathbf{x}_3$ with $\mathbf{x}_1$) and note by optimality of the segmentation,

$$p_\epsilon(\mathcal{X}_i)\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i) - \mu(\mathbf{x}_3)) > p_\epsilon(\mathcal{X}_j)\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_j) - \mu(\mathbf{x}_3)),$$
$$p_\epsilon(\mathcal{X}_j)\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_j) - \mu(\mathbf{x}_2)) > p_\epsilon(\mathcal{X}_i)\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i) - \mu(\mathbf{x}_2)).$$

Combining these two inequalities gives

$$\frac{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i) - \mu(\mathbf{x}_3))}{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_j) - \mu(\mathbf{x}_3))} > \frac{p_\epsilon(\mathcal{X}_j)}{p_\epsilon(\mathcal{X}_i)} > \frac{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i) - \mu(\mathbf{x}_2))}{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_j) - \mu(\mathbf{x}_2))}.$$

Which can be further rearranged to $\frac{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i)-\mu(\mathbf{x}_3))}{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i)-\mu(\mathbf{x}_2))} > \frac{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_j)-\mu(\mathbf{x}_3))}{\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_j)-\mu(\mathbf{x}_2))}$ which contradicts the PF2 property. Thus the optimal segmentation must be interval and $\mathcal{X}_i = \{\mathbf{x}|\mu(\mathbf{x}) \in [s_i, s_{i+1})\}$ for some real numbers $s_i < s_{i+1}$.

**(c)** To show $p_\epsilon(\mathcal{X}_i) = p_\epsilon(\mathbf{x})$ for some $\mathbf{x} \in \mathcal{X}_i$, let $\mathbf{x}' = \arg\min_{\mathbf{x}\in\mathcal{X}_i}\mu(\mathbf{x})$ and recall $p_\epsilon(\mathcal{X}_i) = \arg\max\int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} p\overline{F}_\epsilon(p-s)f(\mu^{-1}(s))ds$. Now suppose $p_\epsilon(\mathcal{X}_i) < \mu(\mathbf{x}')$. By log-concavity, each function $\mathcal{R}_\epsilon(\mu(\mathbf{x}), p) := p\overline{F}_\epsilon(p - \mu(\mathbf{x}))$ is unimodal, and thus increasing in $p$ for $p \le p_\epsilon(\mu(\mathbf{x}))$,

$$p_\epsilon(\mathcal{X}_i)\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i) - s) \le p_\epsilon(\mu(\mathbf{x}'))\overline{F}_\epsilon(p_\epsilon(\mu(\mathbf{x}')) - s),$$

for any $s \in [s_i, s_{i+1}]$, which implies

$$\int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} p_\epsilon(\mathcal{X}_i)\overline{F}_\epsilon(p_\epsilon(\mathcal{X}_i) - s)f(\mu^{-1}(s))ds \le \int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} p_\epsilon(\mu(\mathbf{x}'))\overline{F}_\epsilon(p_\epsilon(\mu(\mathbf{x}')) - s)f(\mu^{-1}(s))ds,$$

thus $p_\epsilon(\mathcal{X}_i) \ge p_\epsilon(\mathbf{x}')$. A symmetric argument similarly shows $p_\epsilon(\mathcal{X}_i) \le \arg\max_{\mathbf{x}\in\mathcal{X}_i} p_\epsilon(\mathbf{x})$. □

*Proof of Theorem 2* Suppose the firms prediction model $\mu(\mathbf{X})$ is supported on $n$ values $\{\mu_i\}_{i=1}^n$, occurring with probabilities $\{q_i\}_{i=1}^n$, where $\mu_1 \le \mu_2 \le \ldots \le \mu_n$, and let $|\text{supp}(\epsilon)| = m_\epsilon$. By Lemma 2, the optimal segmentation can be index by the sequence $\{s_i\}_{i=0}^k$ is contained in the support of $\mu(\mathbf{X})$. Let the optimal price for segment $[s_i, s_{i+1})$ be $p_\epsilon([s_i, s_{i+1}))$ i.e.

$$p_\epsilon([s_i, s_{i+1})) = \arg\max p_i \Pr(\mu(\mathbf{x}) + \epsilon \ge p_i|\mu(\mathbf{x}) \in [s_i, s_{i+1}))\Pr(\mu(\mathbf{x}) \in [s_i, s_{i+1})).$$

We wish to find $\{s_i\}_{i=0}^k \subset \{\mu_i\}_{i=1}^n$ that maximizes

$$\sum_{i=1}^k p_\epsilon([s_i, s_{i+1})) \Pr(\mu(\mathbf{x}) + \epsilon \geq p_\epsilon([s_i, s_{i+1})) \,|\, \mu(\mathbf{x}) \in [s_i, s_{i+1})) \Pr(\mu(\mathbf{x}) \in [s_i, s_{i+1})).$$

where the optimal price $p_\epsilon([s_i, s_{i+1}))$ for each segmentation can be calculated in time $O(nm_\epsilon)$ by simple search over the support. We now give a dynamic programming solution that uses time $O(kn^3m_\epsilon)$ and to populate a table of size $kn$. Define $D[n', k']$ as the optimal $k'$-market segmentation that considers only the $n'$ lowest predicted valuations $\{(\mu_i, q_i)\}_{i=1}^{n'}$, our goal is to compute $D[n, k]$ which is the revenue of the optimal FBMSP (the optimal policy can further be reconstructed by standard backward search). Our algorithm depends on the following observation: consider the optimal $k$-market segmentation and suppose $[s_k, s_{k+1}] = [\mu_{i_k}, \mu_n]$ defines the $k^{th}$ segment. If one considers the market without the customers in the $k^{th}$ segment, the remaining $k-1$ segments must be an optimal $(k-1)$-market segmentation on $\{(\mu_i, q_i)\}_{i=1}^{i_k - 1}$. Formally, we express this observation as the following recursion,

$$D[n', k'] = \max_{l \in [n'-1]} D[l, k'-1] + p_\epsilon([s_i, s_{i+1})) \sum_{i=l+1}^{n'} \Pr(\mu(\mathbf{x}) + \epsilon \geq p_\epsilon([s_i, s_{i+1})) \,|\, \mu(\mathbf{x}) \in [s_i, s_{i+1}))\, q_i, \quad \text{(EC.5)}$$

which states that the optimal $k'$-market segmentation on the lowest $n'$ valuations, is equal to some optimal $(k'-1)$-segmentation on a smaller market, plus the value of the $k'^{th}$ segment. Using Eq. (EC.5) we may populate a table of size $kn$, starting at $D[0, 0] = 0$, and computing column-wise. Each computation of $D[n', k']$ requires $O(n^2 m_\epsilon)$ operations, thus the table may be populated in $\mathrm{O}(kn^3 m_\epsilon)$ time. □

### B.3. Omitted Proofs from Section 4

*Proof of Theorem 3* Let $\{s_i\}_{i=0}^k \in \mathbb{R}^{k+1}$ be an optimal $k$-market segmentation for $\mathcal{R}_{kXP}^{\mu(\mathbf{X})}$. Consider the sub-optimal feature-based market segmentation which uses segments $\mathcal{X}_i = \{\mathbf{x} \,|\, \mu(\mathbf{x}) \in [s_i, s_{i+1})\}$ and prices $p_\epsilon(s_i)$. Note for all $\mathbf{x} \in \mathcal{X}_i$, $p_\epsilon(s_i) \Pr(\mu(\mathbf{x}) + \epsilon \geq p_\epsilon(s_i)) \geq \mathcal{R}_\epsilon(s_i)$ and thus summing over all segments

$$\mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon} \geq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x}) \in [s_i, s_{i+1})} p_\epsilon(s_i) \Pr(\mu(\mathbf{x}) + \epsilon \geq p_\epsilon(s_i))\, f_{\mathbf{X}}(\mathbf{x}) dx \geq \sum_{i=0}^{k-1} \mathcal{R}_\epsilon(s_i) \int_{\mu(\mathbf{x}) \in [s_i, s_{i+1})} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad \text{(EC.6)}$$

Now,

$$\begin{aligned}
\mathbb{E}_{\mathbf{X} \sim F_{\mathbf{X}}}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon} &\leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x}) \in [s_i, s_{i+1})} (\mathcal{R}_\epsilon(\mu(\mathbf{x})) - \mathcal{R}_\epsilon(s_i))\, f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} && \text{Eq. (EC.6)} \\
&\leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x}) \in [s_i, s_{i+1})} (\mu(\mathbf{x}) - s_i) \overline{F}_\epsilon(\theta_\epsilon(\mu(\mathbf{x}))) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} && \text{Lemma 1(b)} \\
&\leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x}) \in [s_i, s_{i+1})} (\mu(\mathbf{x}) - s_i) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \mathbb{E}_{X \sim F_{\mathbf{X}}}[\mu(\mathcal{X}_i)] - \mathcal{R}_{kXP}^{\mu(\mathcal{X}_i)}
\end{aligned}$$

as desired. □

*Proof of Theorem 4.* Let $L = \mathcal{R}_\epsilon(x_L)$ and $U = \mathcal{R}_\epsilon(x_U)$, and recalling the proof of Theorem 3 we have,

$$\mathbb{E}_{V \sim F}[\mathcal{R}_\epsilon(V)] - \mathcal{R}_{kXP}^V \leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x}) \in [s_i, s_{i+1})} (\mathcal{R}_\epsilon(\mu(\mathbf{x})) - \mathcal{R}_\epsilon(s_i))\, \overline{F}_\epsilon(\theta_\epsilon(s_i))\, f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Now instead we choose the sub-optimal interval segmentation such that it subdivides the quantile space of $F_{\mathbf{X}}$ into $k$ equal regions i.e., $\{s_i\}_{i=0}^{k}$ such that $\overline{F}(s_i) - \overline{F}(s_{i+1}) = \frac{1}{k}$ for $i = 0, ..., k-1$. Then

$$\begin{aligned}
\mathbb{E}_{\mu(\mathbf{X})\sim F}[\mathcal{R}_\epsilon(\mu(\mathbf{X}))] - \mathcal{R}_{kXP}^{\mu(\mathbf{X})+\epsilon} &\leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} (\mathcal{R}_\epsilon(x) - \mathcal{R}_\epsilon(s_i)) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\
&\leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} (\mu(\mathbf{x}) - s_i)\overline{F}_\epsilon(\theta_\epsilon(x)) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \qquad \text{Lemma 1(b)} \\
&\leq \sum_{i=0}^{k-1} \int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} (\mu(\mathbf{x}) - s_i) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\
&= \sum_{i=0}^{k} (s_{i+1} - s_i) \int_{\mu(\mathbf{x})\in[s_i,s_{i+1})} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\
&= \sum_{i=0}^{k-1} (s_{i+1} - s_i) \left(\overline{F}(s_i) - \overline{F}(s_{i+1})\right) \\
&= \frac{\sum_{i=0}^{k-1} (s_{i+1} - s_i)}{k} \\
&= \frac{U-L}{k},
\end{aligned}$$

where the third inequality follows $\overline{F}_\epsilon(\theta_\epsilon(s_i)) \leq 1$, the third equality comes from the choice of $\{x_i\}_{i=0}^{k}$, then summing $s_{i+1} - s_i$ we get the final equality as desired.

□

*Proof of Theorem 5.* Fix some $k \geq 2$, we will prove the rearranged inequality $\mathcal{R}_{(k-1)XP} + \mathcal{R}_{(k+1)XP} \leq 2\mathcal{R}_{kXP}$ by explicitly constructing feasible (but not necessarily optimal) size $k$ segmentations. Note, since $\epsilon$ is log-concave, by Lemma 2 the optimal segmentation is interval, and can be described by the sequence of numbers $\{s_i^k\}_{i=0}^{k}$ such that $\mathcal{X}_i^k = \{\mathbf{x} | \mu(\mathbf{x}) \in [s_i, s_{i+1})\}$. Further, let $\mathcal{S}_{k-1} := \{s_i^{k-1}\}_{i=0}^{k-1}$, $\mathcal{S}_k := \{s_{i=0}^k\}$ and $\mathcal{S}_{k+1} := \{s_{i=1}^{k+1}\}$ be the optimal segmentations of size $k-1$, $k$, and $k+1$, respectively, as described by segmentation endpoints.

Additionally, note that by the pigeonhole principle, there must be some interval from the $(k-1)$-segmentation which fully contains an interval of the $(k+1)$-segmentation as a subset. Our proof will consider four cases based on how that subset can arise. For the first three cases we will use the following simple fact that the if one segment subsumes another, it provides more revenue, i.e., if $\mathcal{X}_i \subset \mathcal{X}_j$ then

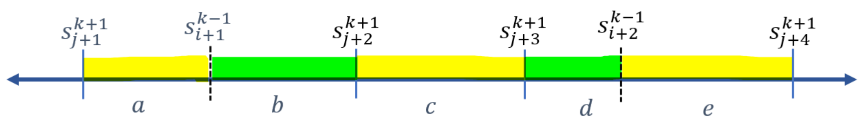
$$\max_p p \Pr(\mu(\mathbf{x}) + \epsilon \geq p | \mathbf{x} \in \mathcal{X}_i) \Pr(\mathcal{X}_i) \leq \max_p p \Pr(\mu(\mathbf{x}) + \epsilon \geq p | \mathbf{x} \in \mathcal{X}_j) \Pr(\mathcal{X}_j). \tag{EC.7}$$

*Case 1:* $s_2^{k+1} \leq s_1^k$. In this case the first segments of the optimal $(k+1)$-segmentation is a subset of the first segment of the $(k-1)$-segmentation. Consider two feasible $k$-segmentations $\mathcal{S}'_k = s_1^{k+1} \cup \mathcal{S}_{k-1}$ and $\mathcal{S}''_k = \mathcal{S}_{k+1} \setminus s_1^{k+1}$. Now note the combined revenue from $\mathcal{S}'_k$ and $\mathcal{S}''_k$ fully covers the revenue the revenue from $\mathcal{S}_{k-1}$ (as a subset of $\mathcal{S}'_k$) and $\mathcal{S}_{k+1}$ except for possibly on the first segment (all others are segments of the optimal $(k+1)$-segmentation are covered by $\mathcal{S}''_k$). Now, in the constructed segmentation $\mathcal{S}'_k$, the unaccounted for first segment has end points $[s_1^{k+1}, s_1^{k-1}]$ which by assumption contains the first segment of the $(k+1)$-segmentation $[s_1^{k+1}, s_2^{k+1}]$, and by Eq. (EC.7) implies $\mathcal{R}_{(k-1)XP} + \mathcal{R}_{(k+1)XP} \leq \mathcal{R}_{kXP}(\mathcal{S}'_k) + \mathcal{R}_{kXP}(\mathcal{S}''_k) \leq 2\mathcal{R}_{kXP}$.

*Case 2: $s_{k-1}^{k-1} \le s_k^{k+1}$.* In this case the last two segments of the optimal $(k+1)$-segmentation is a subset of the last segment of the $(k-1)$-segmentation. Consider the two feasible $k$-segmentations $\mathcal{S}'_k = \mathcal{S}_{k-1} \cup s_{k+1}^{k+1}$ and $\mathcal{S}''_k = \mathcal{S}_{k+1} \setminus s_{k+1}^{k+1}$. Now note the combined revenue from $\mathcal{S}'_k$ and $\mathcal{S}''_k$ fully covers the revenue from $\mathcal{S}_{k+1}$ (via $\mathcal{S}''_k$ and the last segment of $\mathcal{S}'_k$) and $\mathcal{S}_{k-1}$ except for possibly on the last segment (all others are segments of the optimal $(k+1)$-segmentation are covered by $\mathcal{S}''_k$). Now, in the constructed segmentation $\mathcal{S}'_k$, the unaccounted for first segments has end points $\left[s_{k-1}^{k-1}, s_{k+1}^{k+1}\right]$ which by assumption contains the last segment of the $(k+1)$-segmentation $\left[s_k^{k+1}, s_{k+1}^{k+1}\right]$, and by Eq. (EC.7) implies $\mathcal{R}_{(k-1)XP} + \mathcal{R}_{(k+1)XP} \le \mathcal{R}_{kXP}(\mathcal{S}'_k) + \mathcal{R}_{kXP}(\mathcal{S}''_k) \le 2\mathcal{R}_{kXP}$.

*Case 3: There exists $i, j$ such that $s_j^{k-1} \le s_{i-1}^{k+1}$ and $s_{i+1}^{k+1} \le s_{j+1}^{k-1}$.* In this case two consecutive segments of the optimal $(k+1)$-segmentation are a subset of some segment of the $(k-1)$-segmentation. Again, consider two feasible $k$-segmentations $\mathcal{S}'_k = s_i^{k+1} \cup \mathcal{S}_{k-1}$ and $\mathcal{S}''_k = \mathcal{S}_{k+1} \setminus s_i^{k+1}$. Now note the combined revenue from $\mathcal{S}'_k$ and $\mathcal{S}''_k$ fully covers the revenue the revenue from $\mathcal{S}_{k-1}$ (as a subset of $\mathcal{S}'_k$) and $\mathcal{S}_{k+1}$ except for possibly on the interval $[s_i^{k+1}, s_{i+1}^{k+1}]$ (all others are segments of the optimal $(k+1)$-segmentation are covered by $\mathcal{S}''_k$). Now, in the constructed segmentation $\mathcal{S}'_k$, the unaccounted for added segment has end points $[s_i^{k+1}, s_{j+1}^{k-1}]$ which by assumption contains the $i^{th}$ segment of the $(k+1)$-segmentation $[s_i^{k+1}, s_{i+1}^{k+1}]$, and by Eq. (EC.7) implies $\mathcal{R}_{(k-1)XP} + \mathcal{R}_{(k+1)XP} \le \mathcal{R}_{kXP}(\mathcal{S}'_k) + \mathcal{R}_{kXP}(\mathcal{S}''_k) \le 2\mathcal{R}_{kXP}$.

**Figure EC.1 Case 4 of Theorem 5.**




*Note.* Segment $\left[s_{j+2}^{k+1}, s_{j+3}^{k+1}\right)$ is the segment fully contained in the segment $\left[s_{i+1}^{k-1}, s_{i+2}^{k-1}\right)$ of $k-1$ segmentation. The segmentation points divide the region $\left[s_{j+1}^{k+1} l, s_{j+4}^{k+1}\right)$ into 5 small regions, recombine them will give us two new $k$-segmentaions.

*Case 4: Remaining case.* If the three previous cases do not hold, then by the pigeonhole principle, there must exists exactly one segment in $\mathcal{S}_{k+1}$ and segmentation points $s_{i+1}^{k-1}$, $s_{i+2}^{k-1}$, and $s_{j+1}^{k+1}$, $s_{j+2}^{k+1}$, $s_{j+3}^{k+1}$, $s_{j+4}^{k+1}$, such that

$$s_{j+1}^{k+1} < s_{i+1}^{k-1} < s_{j+2}^{k+1} < s_{j+3}^{k+1} < s_{i+2}^{k-1} < s_{j+4}^{k+1}.$$

Such an arrangement of segmentation points is shown in Fig. EC.1. Now consider the feasible $k$-segmentations induced by the $\mathcal{S}_{k-1}$ and $\mathcal{S}_{k+1}$. Let $\mathcal{S}'_k = \left(\mathcal{S}_{k-1} \setminus s_{i+2}^{k-1}\right) \cup s_{j+1}^{k+1} \cup s_{j+4}^{k+1}$ and $\mathcal{S}''_k = \left(\mathcal{S}^{k+1} \setminus \left(s_{j+3}^{k+1} \cup s_{j+4}^{k+1}\right)\right) \cup s_{i+2}^{k-1}$. The new arrangement of segmentation points is shown in Fig. EC.2. Like before, $\mathcal{S}'_k$ and $\mathcal{S}''_k$ fully covers the revenue the revenue from $\mathcal{S}_{k-1}$ and $\mathcal{S}_{k+1}$ by Eq. (EC.7), implies $\mathcal{R}_{(k-1)XP} + \mathcal{R}_{(k+1)XP} \le \mathcal{R}_{kXP}(\mathcal{S}'_k) + \mathcal{R}_{kXP}(\mathcal{S}''_k) \le 2\mathcal{R}_{kXP}$.

All the cases together complete the proof for the concavity of the revenue function it terms of $k$. $\square$

**Figure EC.2 New Feasible $k$-Segmentations in Case 4 of Theorem 5.**


*Note.* In the left panel, $a$ and $b$ forms a new segment in $\mathcal{S}_k'$, $c$ and $d$ form another new segment in $\mathcal{S}_k'$, $e$ is combined with the next segment on the right in $\mathcal{S}_{k-1}$. In the right panel, $b$ and $c$, $d$ and $e$ form two new segments of $\mathcal{S}_k''$, $a$ is combined with the previous segment on the left in $\mathcal{S}_{k+1}$.

## Appendix C: Transforming a Probit Regression Model into a Valuation Model

In this section we overview how to transform a prediction model for sales into a linear valuation model. Note, in reality we cannot observe customer's valuation for one product directly. Instead, we can see whether the customer will buy the product or not for the offered price $p$ (see Cameron and James (1987) for more details). Assume that the unobserved continuous dependent variable $Y$ is the customer's true valuation or willingness to pay (WTP) for the product. Further, the relation of $Y$ and customer's feature $X$ is

$$Y = \beta_0 + X\beta + \epsilon, \tag{EC.8}$$

where $\epsilon \sim N(0,\sigma)$. Customer $i$'s decision $I_i$ will be

$$I_i = \begin{cases} 1, & \text{if } Y_i \geq p_i, \\ 0, & \text{otherwise,} \end{cases} \tag{EC.9}$$

where $p_i$ is the price offered to customer $i$. Then, the probability that customer $i$ will buy the product is

$$\begin{aligned} \Pr(I_i = 1) = \Pr(Y_i \geq p) &= \Pr(+\epsilon_i \geq p) \\ &= Pr\left(\frac{\epsilon_i}{\sigma} \geq \frac{p - X_i\beta}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{p - \beta_0 - X_i\beta}{\sigma}\right), \end{aligned}$$

where $\Phi$ is the cumulative distribution function for the standard normal distribution. Using $p$ and $X$ as explanatory variable, the probit regression model is

$$\Pr(I = 1|X) = 1 - \Phi(\beta_0' + p\beta_p + X\beta').$$

Therefore, we can use the maximum likelihood estimator (MLE) of probit regression model to recover the regression model of customer's valuation, *i.e.*,

$$\hat{\beta}_0 = \frac{\hat{\beta}_0'}{\hat{\beta}_p}, \quad \hat{\beta} = \frac{\hat{\beta}'}{\hat{\beta}_p}, \quad \hat{\sigma} = \frac{1}{\hat{\beta}_p},$$

where $\hat{\beta}_0'$, $\hat{\beta}_p$, $\hat{\beta}'$ are the MLE of $\beta_0'$, $\beta_p$, $\beta'$. Further, the regression model of customer's valuation recovered from probit regression model is asymptotically unbiased if the price variance is large enough.

## Appendix D: Constant Factor Approximation for General Error

In this section we will describe how to obtain a 1-1/e approximation of the optimal FBMSP when the residuals are independent and follow an arbitrary distribution, as sketched in Remark 1.

$(1-1/e)$ **Approximation Algorithm**: Our polynomial time approximation algorithm will leverage the *submodularity* of the objective function for FBMSP, defined as follows:

Definition EC.1 (Submodularity). A set function $f: 2^V \to \mathbb{R}$ is submodular if for every $A, B \subseteq V$,

$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B).$$

An important subclass of submodular functions are those which are monotone, *i.e.*, functions for which enlarging the choice set cannot cause the function value to decrease.

Definition EC.2 (Monotonicity). A set function $f: 2^V \to \mathbb{R}$ is monotone if for every $A \subseteq B \subseteq V$, $f(A) \leq f(B)$.

We will show that the objective function for FBMSP can be expressed as a set function over prices which is monotone and submodular. Note that for $n$ customers with predicted valuations $\{x_i\}_{i=1}^n$ and for error distribution $\epsilon$ supported on $m$ points, there are at most $\mathrm{O}(nm)$ distinct possible valuation realizations. Further, any optimal price for a segment must correspond to one of these realizations (since if not, raising the price until it reaches a valuation in the support is strictly revenue improving). Thus the set of potential prices is a polynomially sized set equivalent to the set of potential realized valuations, and the revenue objective of FBMSP can be viewed as a set function over that set.

Specifically, if $f$ is the revenue function of FBMSP on price set $A$, it takes the form,

$$f(A) = \sum_{i=1}^{n} \max_{p \in A} p\overline{F}(p - \mu(x_i)).$$

Then expressed as a set function over the prices, optimal FBMSP is the solution to

$$\max_{|A| \leq k} \sum_{i=1}^{n} \max_{p \in A} p\overline{F}(p - \mu(x_i)).$$

The monotonicity of the revenue objective is easy to see since, by definition, enlarging the set of possible prices that can be used for a segment will keep at least the same revenue as for a smaller set of prices. The submodularity comes from the fact, any customer $i$ facing the prices in price set $A \cap B$ will result in less revenue than when facing the prices in price set $A$ or $B$, i.e.,

$$\max_{p \in A \cap B} p\overline{F}(p - \mu(x_i)) \leq \min\left\{\max_{p \in A} p\overline{F}(p - \mu(x_i)), \quad \max_{p \in B} p\overline{F}(p - \mu(x_i))\right\},$$

further,

$$\max_{p \in A \cup B} p\overline{F}(p - \mu(x_i)) = \max\left\{\max_{p \in A} p\overline{F}(p - \mu(x_i)), \quad \max_{p \in B} p\overline{F}(p - \mu(x_i))\right\}.$$

Note that

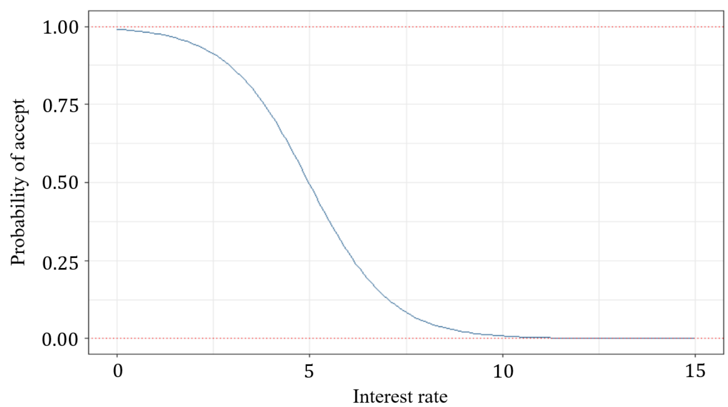
$$\begin{aligned}\max_{p \in A} p\overline{F}(p - \mu(x_i)) + \max_{p \in B} p\overline{F}(p - \mu(x_i)) = \min\left\{\max_{p \in A} p\overline{F}(p - \mu(x_i)), \quad \max_{p \in B} p\overline{F}(p - \mu(x_i))\right\} + \\ \max\left\{\max_{p \in A} p\overline{F}(p - \mu(x_i)), \quad \max_{p \in B} p\overline{F}(p - \mu(x_i))\right\},\end{aligned}$$

and thus combining these observations and summing over all customers proves submodularity of the objective function.

Note that maximizing positive monotone submodular functions maximization with cardinality constraints is NP-hard in general (see Krause and Golovin (2014)). The cardinality constraint in FBMSP is the number of segments (in some sense, the same as number of prices). Nemhauser et al. (1978) shows that a greedy algorithm can obtain an approximation guarantee of $(1 - 1/e)$ for class of monotone submodular functions maximizing with cardinality constraints. Since FBMSP problem is can be written as a problem of maximizing a monotone submodular function with cardinality constraints, it can be approximated at least within a factor of $(1 - 1/e)$ via the same greedy algorithm.

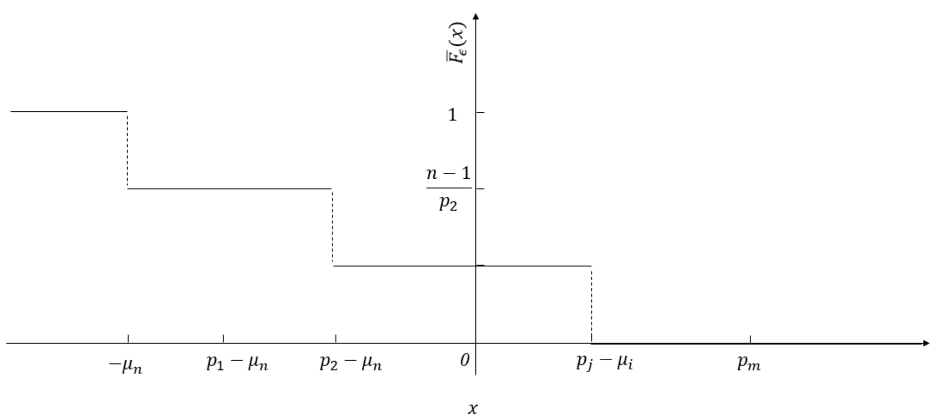
## Appendix E: Omitted Figures

**Figure EC.3 Prediction of the probit regression model.**




*Note.* Depicted is the output of a probit regression model to predict the probably of mortgage acceptance, our proxy for purchase in the loan setting. The model is trained using features in Table 2.

**Figure EC.4 An example of the error distribution $\overline{F}_\epsilon$, constructed for the proof of Theorem 1.**




*Note.* Depicted is an example of the cCDF $\overline{F}_\epsilon$ constructed to prove the hardness of FBMSP. Note on the $x$-axis are the valuation support points, and that the resultant error distribution is a step-function on these supports.