On the limitations of data-based price discrimination*

Haitian Xie[†] Ying Zhu[‡] Denis Shishkin[§]

July 24, 2024

Abstract

The classic third degree price discrimination (3PD) model requires the knowledge of the distribution of buyer valuations and the covariate to set the price conditioned on the covariate. In terms of generating revenue, the classic result shows that 3PD is at least as good as uniform pricing. What if the seller has to set a price based only on a sample of observations from the underlying distribution? Is it still obvious that the seller should engage in 3PD? This paper sheds light on these fundamental questions. In particular, the comparison of the revenue performance between 3PD and uniform pricing is ambiguous overall when prices are set based on samples. This finding is in the nature of statistical learning under uncertainty: a curse of dimensionality, but also other small sample complications.

Keywords: price discrimination, empirical revenue maximization, information theory, prior-independent pricing, optimal rate of convergence

JEL classification code: C14, C44, D42, D82

*Haitian Xie and Ying Zhu share the first authorship and are listed alphabetically. This paper supersedes a previously circulated draft by Xie and Zhu (https://arxiv.org/pdf/2204.12723v1.pdf). All three authors are grateful to the constructive comments from two anonymous reviewers at *Theoretical Economics*, and three anonymous reviewers and the meta reviewer at *ACM Economics and Computation*. The authors would also like to thank Dirk Bergemann, Songzi Du, Federico Echenique, Graham Elliott, Yannai Gonczarowski, Roger Gordon, Nima Haghpanah, Johannes Horner, Jonathan Libgober, Esfandiar Maasoumi, Maximilian Schaefer, Joel Sobel, Karl Schlag, Larry Samuelson, Yixiao Sun, J. Miguel Villas-Boas, and Joel Watson for valuable comments and discussions.

[†]Assistant professor at the Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University. The author is grateful to the UC San Diego Department of Economics where this project was developed during his doctoral studies. Xie is supported by the Fundamental Research Funds for the Central Universities at Peking University. Email: xht@gsm.pku.edu.cn.

[‡]Assistant professor at the Department of Economics, University of California San Diego. Email: yiz012@ucsd.edu. The author is grateful to the Society of Hellman Fellows at University of California and the Cowles Foundation at Yale University, and also thanks participants at her seminars.

[§]Assistant professor at the Department of Economics, University of California San Diego. Email: dshishkin@ucsd.edu.

1 Introduction

In the past few decades, the advances in the theory of mechanism design have been followed by a tremendous interest in its practical applications. At the same time, classic theoretical models typically make strong assumptions about the designer's knowledge of the environment which may lead the optimal mechanism to be sensitive to the details of the environment (which is sometimes referred to as the Wilson critique).[1]

Third degree price discrimination (3PD) requires an observable covariate value associated with the buyer valuation. To set the price conditioned on the covariate, the classic pricing model requires the knowledge of the distribution of buyer valuations and the covariate. In terms of generating revenue, the classic result shows that 3PD is at least as good as uniform pricing. What if the seller has only partial information about those distributions? Is it still obvious that the seller should engage in 3PD? Setting the optimal price for each observed value of the covariate may not "extrapolate" well to the unobserved covariate values, and yield a lower expected revenue than a uniform price. On the other hand, too little discrimination underutilizes the information contained in the covariate about buyer valuations. This paper is concerned with how much information the seller will need in order to make 3PD generate more revenue. Suppose a unit demand buyer with a privately-known valuation Y and a one-dimensional continuous covariate X drawn from a joint distribution $F_{Y,X}$, that is unknown to the seller. The continuous covariate X can be a single index or score that summarizes the relevant characteristics for pricing and marketing. Hartmann, Nair and Narayanan (2011) provide examples where marketing firms use a one-dimensional continuous score function of customer characteristics, past response histories, and features of the zip code, and casinos use a one-dimensional continuous score referred to as the average daily win.

While our seller is ignorant of $F_{Y,X}$, he/she does have access to a random sample of i.i.d. $\{Y_i, X_i\}_{i=1}^n$ drawn from $F_{Y,X}$. A natural strategy is to choose prices that optimize against the empirical distribution of $\{Y_i, X_i\}_{i=1}^n$. The K-markets empirical revenue maximization (ERM) divides the covariate space into K equal-length segments, and the optimal price based on the conditional empirical distribution for each segment is calculated. We show that when $K = \Theta(n^{1/4})$, the K-markets ERM strategy generates an expected revenue converging to that of the true distribution 3PD optimum at the rate $O(n^{-1/2})$. The 1-market ERM strategy is simply the (uniform) ERM strategy, which we show generates a revenue converging to that of the true-distribution uniform optimum at the rate $O(n^{-2/3})$. The K-markets ERM is just one possible strategy and one may wonder if a more sophisticated strategy

[1] In some cases, this leads to extreme or unrealistic results as in, e.g., Crémer and McLean (1988).

might provide faster convergence rates. In a sense, the answer is no. We show that these rates are asymptotically unimprovable for the worst case distributions of (Y, X) subject to some mild smoothness conditions. In other words, to guarantee a revenue deficiency of δ uniformly over a class of distributions, the necessary condition for the sample size is that $n = \Omega(\delta^{-2})$ in the 3PD problem and $n = \Omega(\delta^{-3/2})$ in the uniform pricing problem.

For sufficiently small δ, the K-markets ERM and the uniform ERM strategies are optimal on the growth requirements of the sample size, respectively; that is, $n = \Theta(\delta^{-2})$ in the 3PD problem and $n = \Theta(\delta^{-3/2})$ in the uniform pricing problem. To show this optimality result, we establish a lower bound for the revenue deficiency in *any* data-based pricing strategy relative to the true-distribution optimal strategy in the worst case (by considering the supremum over a class of joint distributions, $F_{Y,X}$, subject to some mild smoothness assumptions). In particular, data-based uniform pricing strategies are algorithms that depend on $\{Y_i\}_{i=1}^n$ only, and the true-distribution optimal strategy corresponds to the optimal uniform pricing strategy derived from F_Y. Similarly, data-based 3PD strategies are algorithms that depend on $\{Y_i, X_i\}_{i=1}^n$, and the true-distribution optimal strategy corresponds to the optimal 3PD strategy derived from $F_{Y,X}$. We show that the minimax revenue deficiency is $\Omega(n^{-2/3})$ and $\Omega(n^{-1/2})$ in the uniform and 3PD cases, respectively.

Our results highlight the following economic trade-off. When the seller has the access to a sample of i.i.d. $\{Y_i,\ X_i\}_{i=1}^n$, she can choose the K-markets ERM strategy that exploits both $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, or the uniform ERM strategy that ignores $\{X_i\}_{i=1}^n$ and exploits only $\{Y_i\}_{i=1}^n$. Inherently, the former is an algorithm trying to learn the $F_{Y,X}$-optimal pricing function $p(\cdot)$ while the latter is an algorithm trying to learn the F_Y-optimal (constant) pricing function. As a result of the curse from the *extra* dimensionality, the former is more demanding in the sample size than the latter. On the other hand, in terms of generating revenue, the true-distribution optimal 3PD strategy is at least as good as the true-distribution optimal uniform pricing strategy. This trade-off suggests that, even if X contains useful information about Y, the K-markets ERM strategy based on a random sample *can* be revenue inferior to the uniform ERM strategy when the sample size n is not large enough, and vice versa.

To verify these potential implications, we conduct several numerical studies. In particular, we calculate the revenues of the K-markets ERM and the uniform ERM strategies based on a real-world data set from eBay auctions and two simulated data sets. Our numerical results illustrate the aforementioned trade-off. When the sample size is small, the uniform ERM strategy can generate higher expected revenue than the K-markets ERM strategy. As the sample size grows, the K-market ERM strategy (the uniform ERM strategy) gets closer to the true-distribution optimal 3PD strategy (respectively, the true-distribution optimal uniform pricing strategy). The slower rate of convergence in the revenue from the K-markets ERM strategy (in contrast to the faster rate of convergence in the revenue from the uniform ERM strategy) is dominated by the benefit of price discrimination (based on $F_{Y,X}$) over uniform pricing (based on F_Y). Consequently, the revenue of the K-markets ERM strategy overtakes that of the uniform ERM strategy when the sample size becomes sufficiently large and X contains sufficient information about Y.

The key takeaways from this paper are summarized here. First, no sample-based 3PD strategy is able to escape from the curse of dimensionality, shown by our information theoretic lower bounds. Second, absent uncertainty regarding the underlying probability laws, third-degree price discrimination is at least as good as uniform pricing in generating revenue. In contrast, the comparison of the revenue performance between the K-markets ERM and the uniform ERM strategies is ambiguous overall. This finding is in the nature of statistical learning under uncertainty: a curse of dimensionality, but also other small sample complications.[2] Empirical revenue maximization is not free of these issues. Ultimately, this paper poses a challenging open question of whether there exist some $\underline{n} < \bar{n} < \infty$ such that for any $n \in [\underline{n}, \bar{n}]$ and distribution in the class defined in this paper, the K-markets ERM strategy (for any $K > 1$) is always revenue-inferior to the uniform ERM strategy.

1.1 Related literature

Complexity measures and information theoretic lower bounds. Information theoretic lower bounds and sample complexity are important notions in machine learning. Both aim to characterize learnability, i.e., how easy it is to learn an unknown object of interest (in our context, the true-distribution optimal 3PD strategy) from data where the uncertainty arises. Sample complexity derives the rate at which the sample size needs to grow to guarantee a desired learning accuracy. Information theoretic lower bound derives a lower bound as a function of the sample size on the learning error (in our context, the revenue deficiency) in the worst case. Sample complexity and information theoretic lower bounds are intrinsically tied to the complexity or size of the underlying function class of interest. Vapnik-Chervonenkis (VC) dimensions, shattering dimensions, and metric entropy (such as the cardinality of packing sets) are popular measures of complexity in machine learning. There have been a number of innovative applications of VC dimensions or shattering dimensions in economic theory and algorithmic economics. Together with the Probably Approximately

[2]Specifically, there exists a distribution F_Y where the revenue of the uniform ERM strategy is worse with two observations than with one; see Babaioff, Gonczarowski, Mansour and Moran (2018). We illustrate in Section 6 that this seemingly counter-intuitive result highlights the difficulty of establishing general comparative results with very small sample size and sheds some light on the comparison of the revenue performance of the K-markets ERM strategy with $K = 1$ vs $K = 2$ in the case of $n = 2$.

Correct (PAC) framework, they are used to study the complexity of the classes of demand and utility functions (Beigman and Vohra, 2006; Balcan, Daniely, Mehta, Urner and Vazirani, 2014), k-demand buyer's valuation (Zhang and Conitzer, 2020), theories of choices (Basu and Echenique, 2020), preference functions (Chambers, Echenique and Lambert, 2021, 2023), as well as the resulting learnability from data. VC dimension is useful for deriving sample complexity bounds concerning discrete function sets and finite dimensional vector spaces, and shattering dimension is useful for certain real functions.

From the theory of machine learning, when a class has infinite VC or shattering dimensions, this class is not PAC learnable. For example, a collection of sinusoids have subgraphs with infinite VC dimension. The max-min expected utility model with at least three states of the world has infinite VC dimension (Basu and Echenique, 2020). The class of demand functions has infinite shattering dimension (Beigman and Vohra, 2006). Nonetheless, the notion of "learnability" can be generalized using a different type of complexity analysis that gives rise to our information theoretic lower bound in the 3PD problem. This type of analysis is built upon the notion of packing sets, along with tools from information theory. In particular, packing sets are useful for studying classes with an infinite number of elements (see Kolmogorov and Tikhomirov (1959) and Wainwright (2019)). This is the case for our 3PD problem as we try to learn an optimal pricing function of the covariate (an infinitely-dimensional parameter) and bound the deficiency in the *expected* revenue, which concerns the entire pricing function at all covariate values.

Prior-independent mechanism design. Most of the classic monopoly pricing literature assumes a known distribution of valuations (and covariates).[3] More recently, some papers (e.g., those surveyed in Carroll, 2019) studied "prior"-independent mechanism design.[4] The main focus of that literature is on deriving a robustly optimal mechanism in the absence of both "prior" and data. In particular, Bergemann and Schlag (2008, 2011) derive the minimax-regret uniform pricing strategy in closed form; that is, the strategy that guarantees the smallest deficiency in revenue relative to the known distribution case. Like Bergemann and Schlag (2008, 2011), we study the revenue deficiencies, but in contrast, we assume the availability of data and focus on the (inevitable) information-theoretic limitations of any *data-based* pricing strategies and the achievability of the limitation.

This paper is inspired by the literature that studies approximately optimal "prior"-independent

[3] See also Segal (2003) for a study of optimal multi-unit auctions where the seller has a probabilistic belief about the valuation distribution of the i.i.d. buyers.

[4] Here, "prior" distribution refers to the seller's prior belief about buyers' valuations and is often taken to be the true distribution.

mechanism design, in particular monopoly pricing with a single buyer.[5] This literature assumes that the seller has access to a random sample of i.i.d. $\{Y_i\}_{i=1}^n$ drawn from F_Y and proposes variants of the uniform ERM strategy to derive the revenue guarantee in relation to that from the true-distribution optimal uniform pricing strategy. There are two types of analyses in this literature. The first one focuses on the guarantees for the specific case of $n = 1$ or $n = 2$ (Babaioff et al., 2018; Allouah, Bahamou and Besbes, 2023). The second one (e.g., Huang, Mansour and Roughgarden, 2018) establishes "sample complexity bounds" such that the uniform ERM variants achieve a $(1 - \epsilon)$ fraction guarantee when the sample size grows at a rate depending on ϵ, and also derives the rate at which the sample size needs to grow (as a function of ϵ) for any data-based uniform pricing strategies to obtain a given $(1 - \epsilon)$ fraction guarantee. Allouah, Bahamou and Besbes (2022) involves both types of analyses.

In this paper, we ask the related question, how fast the revenue deficiency decays as a function of n, and provide an answer using information-theoretic lower bounds (independent of algorithms) and upper bounds with respect to specific algorithms in the worst case scenarios.[6] The main difference with the majority of the data-based literature is that, we study third-degree price discrimination (3PD) with a continuous covariate and compare the revenue performance of data-based 3PD and uniform pricing strategies.

To understand why the 3PD problem in our context is more challenging than the uniform pricing problem, note that fundamentally, the latter tries to learn the constant optimal pricing function (a scalar parameter) while the former tries to learn an optimal pricing function of the covariate (an infinitely-dimensional parameter), where the deficiency in the *expected* revenue concerns the entire pricing function at all covariate values. Our framework allows us to tackle several challenging aspects of the 3PD problem, which might be difficult to analyze with the toolkit in the existing pricing literature. We describe one example below.

Somewhat related, Devanur, Huang and Psomas (2016) studies sample complexity of optimal pricing with "side information". In their "signals model" (Sections 5.1 and 5.3), there is a covariate (signal) $X \in [0, 1]$, and the seller can condition the data-based reserve price on the covariate. For the single-buyer case (which would correspond to our 3PD problem), they derive upper and lower sample

[5]There is a less related literature that studies optimal auctions; see, e.g., Cole and Roughgarden (2014); Dhangwatnotai, Roughgarden and Yan (2015); Fu, Immorlica, Lucier and Strack (2015); Guo, Huang and Zhang (2019); Fu, Haghpanah, Hartline and Kleinberg (2021).

[6]A large literature studies data-based auctions by focusing on guarantees for revenue deficiencies (instead of fractions), such as how the revenues from the data-based strategies converge in probability to the true-distribution benchmark, e.g., Baliga and Vohra (2003); Goldberg, Hartline, Karlin, Saks and Wright (2006); Gonçalves and Furtado (2024). This line of work does not consider the optimal rates of convergence or optimal sample size requirements.

complexity bounds. Importantly, they assume that the true joint distribution $F_{Y,X}$ has the following property: larger values of X are associated with larger values of Y in the sense of first-order stochastic dominance of conditional distributions. In contrast, our 3PD setup imposes no assumptions about the relationship between the covariate X and the valuation Y; meanwhile, our proposed K-market ERM strategy learns the relationship from the data. Moreover, our K-market ERM strategy attains the optimal rate of convergence in revenue deficiency (as described before), while the upper and lower bounds in Devanur et al. (2016) have different rates, and hence, the optimal sample size requirement is unclear.

2 Setup

The seller is selling an item to a buyer. Let $Y \in [0,1]$ be the valuation (i.e., willingness to pay) of the buyer, and X the covariate (such as a characteristic) associated with the buyer. The joint distribution of (Y, X) is denoted by $F_{Y,X}$. We assume that X is supported on a bounded interval, and without loss of generality, we take the interval to be $[0,1]$.[7]

Given a covariate value, the seller wants to set a price according to a mapping from the covariate to a set of prices. We use $\mathcal{D}$ to denote the set of all pricing functions:

$$\mathcal{D} \equiv \{p\colon [0,1] \to [0,1],\ \text{measurable}\}.$$

For a generic pricing strategy $p \in \mathcal{D}$, the price depends on the covariate value x. This scheme falls in the realm of third-degree price discrimination (3PD). Uniform pricing can be viewed as a special case where the price is the same for all covariate values. We use $\mathcal{U}$ to denote the set of all uniform pricing functions:

$$\mathcal{U} \equiv \{p \in \mathcal{D}\colon\ p \text{ is a constant function}\}.$$

To lighten the notation, we express $p \in \mathcal{U}$ as a scalar rather than a function for the uniform pricing problem.

Let $F_{Y|X}$ be the conditional CDF and f_X the marginal density function. Given a price $y \in [0,1]$

[7]The assumption that $Y, X \in [0,1]$ is made merely for simplicity. First of all, our results in Sections 3 and 4 hold for general bounded supports. Second, the precise knowledge of the support boundaries is unnecessary because they can be readily estimated using extremum order statistics. The estimator converges at a superconsistent rate of n^{-1} (see, e.g., Hirano and Porter, 2003), significantly faster than the convergence of revenue deficiency that we show in Section 3. Therefore, in our analysis, the estimation error resulting from the unknown support is negligible. We are grateful to a referee for raising this discussion.

and a covariate value $x \in [0,1]$, there are $1 - F_{Y|X}(p|x)$ buyers whose valuation is above the price. The revenue generated from these buyers is

$$r(y, x, F_{Y,X}) \equiv (1 - F_{Y|X}(y|x))y, \tag{1}$$

and the *expected* revenue for a pricing function p is

$$R(p, F_{Y,X}) \equiv \int_0^1 r(p(x), x, F_{Y,X}) f_X(x) dx.$$

In various places of the rest of the paper, we will slightly abuse the notation and denote $r(p,x) \equiv r(p(x),x)$ when p is a pricing function and also write $r(y,x) = r(y,x,F_{Y,X})$ for brevity when $F_{Y,X}$ is clear from the context. In the special case where the pricing strategy is uniform (i.e., $p \in \mathcal{U}$), the revenue only depends on the marginal distribution F_Y:

$$R(p, F_{Y,X}) = p\mathbb{P}(Y \geq p) = p(1 - F_Y(p)), p \in \mathcal{U}.$$

The true-distribution optimal 3PD strategy p_D^* is the one that maximizes the revenue:

$$R(p_D^*, F_{Y,X}) = \sup_{p \in \mathcal{D}} \int_0^1 r(p(x), x, F_{Y,X}) f_X(x) dx.$$

In a similar fashion, we denote p_U^* as the true-distribution optimal uniform pricing strategy such that

$$R(p_U^*, F_Y) = R(p_U^*, F_{Y,X}) = \sup_{p \in \mathcal{U}} p(1 - F_Y(p)).$$

Note that p_D^* depends on $F_{Y,X}$ and p_U^* depends on F_Y.

In terms of generating revenue, the classic pricing theory shows that 3PD is at least as good as uniform pricing when the joint distribution $F_{Y,X}$ is known to the seller. In this case, we can solve analytically or numerically for the optimal pricing strategies p_D^* and p_U^*. Since $\mathcal{U}$ is contained in $\mathcal{D}$, p_D^* must achieve a (weakly) better revenue than p_U^*. Intuitively, when Y is correlated with X, p_D^* utilizes the information in X.

Now suppose that the seller knows neither $F_{Y,X}$ nor F_Y, but instead observes a random sample of $data \equiv \{(Y_i, X_i), 1 \leq i \leq n\}$ drawn from $F_{Y,X}$, or $data_Y \equiv \{Y_i, 1 \leq i \leq n\}$ from F_Y, and wants to construct a pricing strategy based on the sample. The following assumption is used throughout this paper.

Assumption 1. *data and data$_Y$ consist of i.i.d. draws from $F_{Y,X}$ and F_Y, respectively.*

The following assumption is used to establish the results concerning our 3PD problem. Instead of a single known joint distribution $F_{Y,X}$, there is a class $\mathcal{F}$ of unknown distributions which are deemed possible and our data-based pricing strategies can be evaluated within this class. The functions in $\mathcal{F}$ satisfy several smoothness and regularity conditions stated below.

Assumption 2. *Any distribution function in the set $\mathcal{F}$ satisfies the following conditions.*

(i) (Lipschitz continuity) There exists $C_0 \in (0, \infty)$ such that, for any $y, y', x \in [0, 1]$, the conditional density $f_{Y|X}$ satisfies

$$|f_{Y|X}(y|x) - f_{Y|X}(y'|x)| \leq C_0|y - y'|.$$

(ii) (Strong concavity) There exists $C^* > 0$ such that the revenue function $r(y, x) \equiv y(1 - F_{Y|X}(y|x))$ is strictly concave with second-order derivative

$$-2f_{Y|X}(y|x) - y\frac{\partial}{\partial y}f_{Y|X}(y|x) \leq -C^*, \text{ a.e.} \tag{2}$$

(iii) (Interior solution) For each $x \in [0, 1]$, the optimal price is an interior solution; that is, $p^*_D(x; F_{Y,X}) \in (0, 1)$.

(iv) (Differentiability) The conditional distribution function $f_{Y|X}(y|x)$ is continuously differentiable in (x, y) in a neighborhood of the curve $\{(x, p^*_D(x; F_{Y,X})) : x \in [0, 1]\}$.

(v) (Boundedness) The functions

$$\left|2f_{Y|X}(y|x) + y\frac{\partial}{\partial y}f_{Y|X}(y|x)\right| \tag{3}$$

$$\text{and } \left|\frac{\partial}{\partial x}F_{Y|X}(y|x) + y\frac{\partial}{\partial x}f_{Y|X}(y|x)\right| \tag{4}$$

are bounded from above by $\overline{C} \in (0, \infty)$ a.e.

(vi) (Marginal density) The marginal density f_X is bounded from above by $\overline{C}' \in (0, \infty)$ and bounded away from zero; that is, $f_X \geq \underline{C} > 0$.

Part (i) requires the conditional density function to be sufficiently smooth. The partial derivative $\frac{\partial}{\partial y}f_{Y|X}(y|x)$ is well defined almost everywhere because $f_{Y|X}$ is Lipschitz continuous and hence absolutely continuous. Part (iii) ensures that the first-order condition holds for the optimal price. Part

(iv) ensures that the optimal pricing function $p_D^*(x; F_{Y,X})$ is sufficiently smooth in x. Part (v) requires the partial derivatives of the revenue to be bounded. Part (vi) ensures that the covariate does not take vanishing or dominating values.

Under part (ii), the optimal price is well defined. Part (ii) is a standard assumption in the optimal auctions/pricing literature also known as regularity (Myerson, 1981), which is a so-called "strong concavity" condition from machine learning theory. It is well known that any distribution F with the monotone hazard rate satisfies regularity.

Analogously, the following assumption is used to establish the results for the uniform pricing problem which concerns a class $\mathcal{F}^U$ of unknown marginal distributions that are deemed possible.

Assumption 3. *Let $\mathcal{F}^U$ be the set of marginal distributions such that any $F_Y \in \mathcal{F}^U$ satisfies parts (i), (ii), and (v)(3) of Assumption 2 with $f_{Y|X}(y|x)$ replaced by $f_Y(y)$. Moreover, the optimal price is an interior solution; that is, $p_U^*(F_Y) \in (0, 1)$. The distribution function $f_Y(y)$ is continuously differentiable in y in a neighborhood of $p_U^*(F_Y)$.*

Remark. *By defining $\mathcal{F}^U$ in the way above, note that the marginal distribution associated with any joint distribution satisfying (i), (ii) and (v)(3) of Assumption 2 satisfies the counterpart conditions in Assumption 3.*

Notation. For functions $f(n)$ and $g(n)$, we write $f(n) \gtrsim g(n)$ to mean that $f(n) = \Omega(g(n))$. Similarly, we write $f(n) \lesssim g(n)$ to mean that $f(n) = O(g(n))$. The notation $f(n) \asymp g(n)$ means that $f(n) = \Theta(g(n))$; that is, $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$. As a general rule for this paper, the various c and C constants denote positive universal constants that are independent of the sample size n, and may vary from place to place. For functions f and g, the unweighted L_2 norm (L_2 as the short form) $\|f - g\|_2 \equiv \left(\int_0^1 [f(x) - g(x)]^2 \, dx\right)^{\frac{1}{2}}$.

3 The K-markets ERM strategy

In this section, we propose the K-markets ERM strategy, and compare its revenue with that of the true-distribution optimal 3PD strategy. In particular, we provide upper bounds for the pointwise and expected revenue deficiency as a function of n. We also compare the revenue of the 1-market (uniform) ERM strategy with that of the true-distribution uniform optimum, and provide an upper bound on the revenue deficiency.

3.1 Price discrimination

We propose the *"K-markets"* ERM strategy for the data-based 3PD problem with a continuous covariate:

1. Divide the individuals into $K(\equiv K_n)$ markets by splitting the covariate space $[0,1]$ into K equally spaced intervals

$$I_k \equiv [(k-1)/K, k/K], k = 1, \dots, K.$$

2. For each market I_k, based on the empirical distribution of $\{Y_i \colon X_i \in I_k\}$,

$$\hat{F}_k(y) = \frac{1}{n_k} \sum_{i\colon X_i \in I_k} \mathbf{1}\{Y_i \le y, X_i \in I_k\} \tag{5}$$

where n_k is the cardinality of $\{i \colon X_i \in I_k\}$, solve for the optimal price $\hat{p}_{D,k}$ as follows,

$$\hat{p}_{D,k} \equiv \operatorname*{argmax}_{p\in[0,1]} p(1-\hat{F}_k(p)).$$

The resulting pricing function is a piece-wise function

$$\hat{p}_D(x; data) = \hat{p}_{D,k}, x \in I_k.$$

If the kth market does not contain any observation, then simply choose $\hat{p}_{D,k}$ to be any arbitrary number in $[0,1]$. Doing so has no impact on the asymptotic guarantee implied by the following theorem. For practical implementation, the desired choice may change from context to context, depending on the seller's specific knowledge about a buyer, and the related analysis would be beyond the scope of this paper.

Theorem 1. *Suppose Assumptions 1 and 2 hold. There exists a positive universal constant $c_1 \in (0,\infty)$ such that the following results hold.*[8]

(i) At a given covariate value x_0, the revenue generated by the K-markets ERM strategy $\hat{p}_D$ satisfies

$$\sup_{F_{Y,X}\in\mathcal{F}} \left(r(p_D^*, x_0) - \mathbb{E}_{F_{Y,X}}\left[r(\hat{p}_D(data), x_0)\right]\right) \lesssim 1/K^2 + (K/n)^{2/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K\right), \quad x_0 \in I_k,$$

[8]For example, the constant $c_1 = 1$ when $X \sim U[0,1]$.

where the expectation $\mathbb{E}_{F_{Y,X}}$ is taken with respect to data $\sim F_{Y,X}$ and K satisfies $\frac{c_1}{K} \leq \frac{1}{2}$; moreover,

$$(K/n)^{2/3} + 1/K^2 \asymp n^{-1/2} \text{ when } K \asymp n^{1/4},$$

in which case,

$$\sup_{F_{Y,X} \in \mathcal{F}} \left(r(p_D^*, x_0) - \mathbb{E}_{F_{Y,X}} \left[r(\hat{p}_D(data), x_0) \right] \right) \lesssim n^{-1/2}.$$

(ii) The expected revenue generated by the K-markets ERM strategy $\hat{p}_D$ satisfies

$$\sup_{F_{Y,X} \in \mathcal{F}} \left(R(p_D^*, F_{Y,X}) - \mathbb{E}_{F_{Y,X}} \left[R(\hat{p}_D(data), F_{Y,X}) \right] \right) \lesssim 1/K^2 + (K/n)^{2/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K \right)$$

where the expectation $\mathbb{E}_{F_{Y,X}}$ is taken with respect to data $\sim F_{Y,X}$ and K satisfies $\frac{c_1}{K} \leq \frac{1}{2}$; moreover,

$$(K/n)^{2/3} + 1/K^2 \asymp n^{-1/2} \text{ when } K \asymp n^{1/4},$$

in which case,

$$\sup_{F_{Y,X} \in \mathcal{F}} \left(R(p_D^*, F_{Y,X}) - \mathbb{E}_{F_{Y,X}} \left[R(\hat{p}_D(data), F_{Y,X}) \right] \right) \lesssim n^{-1/2}.$$

Remark. *The term $\exp\left(-\frac{nc_1^2}{8K^2} + \log K \right)$ is technical and comes from a binomial tail bound on n_k in (5); see (22) and the following derivation in the appendix for more detail. Suppose $8K^2 = n^{1-c} c_1^2$ with $c \in (0, 1)$ so that $\frac{nc_1^2}{8K^2} = n^c$ (for example, $c = \frac{1}{2}$ which gives $K \asymp n^{1/4}$ as in the theorem above). Then, there exists some positive universal constant $c_0 \in (0, \infty)$ such that*

$$\exp\left(-\frac{nc_1^2}{8K^2} + \log K \right) = \exp(-c_0 n^c) \text{ as } n \to \infty.$$

In this case, note that $\exp(-c_0 n^c) = o\left((K/n)^{2/3} \right)$ and the term $\exp\left(-\frac{nc_1^2}{8K^2} + \log K \right)$ can be dropped from the bounds in Theorem 1.

Note that having an upper bound on the supremum of the revenue deficiency immediately implies that this upper bound holds for every distribution $F_{Y,X} \in \mathcal{F}$. Moreover, the revenue of the K-markets ERM strategy is guaranteed to have a convergence rate no greater than the provided upper bound, in particular $n^{-1/2}$ when $K \asymp n^{1/4}$.

The interpretation of our results is as follows. The deficiency in revenues comes from two sources.

The first part $(K/n)^{2/3}$ is related to the "variance", which is due to the randomness of the sample, making $\hat{F}_k(\cdot)$ different from its expectation. The second part $1/K^2$ is related to the approximation error due to the fact that we set the same price for all covariate values in the market I_k. Note that more discrimination (larger K) increases the "variance" but reduces the approximation error, and selecting $K \asymp n^{1/4}$ minimizes the upper bound on revenue deficiency.

To show $(K/n)^{2/3}$, we use a peeling argument and other tools from empirical process theory (Alexander, 1987; van der Vaart and Wellner, 1996; van de Geer, 2000). Even though this toolkit is widely used in mathematical statistics and theoretical machine learning, to our knowledge, it has not been introduced to the data-based pricing literature. Showing $1/K^2$ requires controlling $|\tilde{p}_k - p_D^*(x_0)|$, where $\tilde{p}_k \equiv \mathrm{argmax}_{p\in[0,1]}\, p\mathbb{P}(Y > p, X \in I_k)$ and $x_0 \in I_k$. Using the implicit function theorem, we show that, (i) $p_D^*(x)$ is Lipschitz continuous on $[0,1]$, and (ii) $\tilde{p}_k$ is a weighted average of $p_D^*(x), x \in I_k$. These facts imply that $|\tilde{p}_k - p_D^*(x_0)|^s \lesssim 1/K^s$ for any fixed $s \geq 1$.

3.2 Uniform pricing

Based on the empirical distribution of $\{Y_i\}_{i=1}^n$

$$\hat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{Y_i \leq y\},$$

the uniform ERM strategy simply solves for the optimal price $\hat{p}_U$ as follows:

$$\hat{p}_U(data_Y) \equiv \underset{p\in[0,1]}{\mathrm{argmax}}\, p(1 - \hat{F}(p)).$$

We have the following result as a corollary of Theorem 1.

Corollary 1. *Let Assumptions 1 and 3 hold. The revenue generated by $\hat{p}_U$ satisfies*

$$\sup_{F_Y\in\mathcal{F}^U} \left(R(p_U^*, F_Y) - \mathbb{E}_{F_Y}\left[R(\hat{p}_U(data_Y), F_Y)\right]\right) \lesssim n^{-2/3}$$

where the expectation $\mathbb{E}_{F_Y}$ is taken with respect to $data_Y \sim F_Y$.

The 3PD ERM problem with a continuous covariate is more delicate than the uniform ERM problem. The latter does not involve a (continuous) covariate and hence incurs no approximation error. Contrasting Corollary 1 with Theorem 1, one can see that the only source of revenue deficiency in the uniform ERM strategy comes from the "variance".

3.3 Welfare analysis

From the perspective of a policy maker, it is also of interest to study the welfare under the specific pricing strategies in Sections 3.1 and 3.2. In this section, we derive the rate at which the welfare generated by these data-based pricing strategies converges to the welfare generated by their respective true-distribution optimal strategies.

We assume that there is no production cost for the seller, and there is no utility for the seller if the item is not sold. These assumptions are typically imposed in a benchmark model in the auction and pricing literature. For any pricing strategy $p \in \mathcal{D}$, its welfare can be written as

$$W(p, F_{Y,X}) \equiv \mathbb{E}_{F_{Y,X}}[Y\mathbf{1}\{Y > p(X)\}].$$

Theorem 2.

(i) Let Assumptions 1 and 2 hold. Take $K \asymp n^{1/4}$ in the "K-markets" ERM strategy. Then

$$\sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}_{F_{Y,X}} |W(\hat{p}_D(data), F_{Y,X}) - W(p_D^*, F_{Y,X})| \lesssim n^{-1/4}.$$

(ii) Let Assumptions 1 and 3 hold. Then

$$\sup_{F_Y \in \mathcal{F}^U} \mathbb{E}_{F_Y} |W(\hat{p}_U(data_Y), F_Y) - W(p_U^*, F_Y)| \lesssim n^{-1/3}.$$

4 Information-theoretic limitation of data-based pricing

The revenue deficiency in the K-markets ERM strategy and uniform ERM strategy in Section 3 is $O\left(n^{-\frac{2}{2+2}}\right)$ and $O\left(n^{-\frac{2}{2+1}}\right)$, respectively. Note the "2" and "1" in the second terms of the denominators of the exponents in these upper bounds, where the "$2 - 1 = 1$" difference is a result of the extra dimension from the covariate X in the 3PD problem. Without any lower bounds, the upper bounds alone are unable to confirm that the curse of the extra dimensionality necessarily exists and is unimprovable.

In this section, we establish lower bounds to show that *no 3PD strategy is able to escape the curse of the extra dimensionality* and hence the K-markets ERM strategy is not an exception. Our lower bounds also conclude the optimality of the convergence rates $n^{-1/2}$ and $n^{-2/3}$ from Section 3 within the respective realms of 3PD and uniform pricing. Therefore, the dependence of the extra dimension due to X in our 3PD problem cannot be improved. As discussed in the introduction, rate optimality speaks to the optimality or efficiency of the growth requirement of the sample size.

For the lower bounds, it makes little sense to consider a framework recommending the data-based pricing strategies that are only good for a single distribution. For any *fixed* joint distribution $F_{Y,X}$, there is always a trivial data-based pricing strategy: simply ignore the data and select the optimal pricing scheme given $F_{Y,X}$. For this particular distribution, the revenue deficiency is zero. However, such a pricing strategy may perform poorly under other distributions of (Y,X). One solution to circumvent this issue is to compute the worst revenue deficiency over the class $\mathcal{F}$ of possible distributions.

To be specific, we consider the minimax difference in the revenues at a given covariate value x_0 for 3PD,

$$\mathscr{R}_n^D(x_0;\mathcal{F}) \equiv \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \left(r(p_D^*, x_0, F_{Y,X}) - \mathbb{E}_{F_{Y,X}} \left[r(\check{p}_D(data), x_0, F_{Y,X}) \right] \right),$$

and the minimax difference in the *expected* revenues for 3PD,

$$\mathscr{R}_n^D(\mathcal{F}) \equiv \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \left(R(p_D^*, F_{Y,X}) - \mathbb{E}_{F_{Y,X}} [R(\check{p}_D(data), F_{Y,X})] \right),$$

where the expectation $\mathbb{E}_{F_{Y,X}}$ is taken with respect to $data \sim F_{Y,X}$ and $R(\cdot,\cdot)$ is defined in Section 2. In the definitions above, $\check{p}_D(data)$ is a pricing function in $\mathcal{D}$ and $\check{p}_D(x_0; data)$ corresponds to its value at a covariate $x_0 \in [0,1]$; moreover, $\check{\mathcal{D}}$ is the set of *all data-based* 3PD functions $\check{p}_D$.

Similarly, for uniform pricing, we consider

$$\mathscr{R}_n^U(\mathcal{F}^U) \equiv \inf_{\check{p}_U \in \check{\mathcal{U}}} \sup_{F_Y \in \mathcal{F}^U} \left(R(p_U^*, F_Y) - \mathbb{E}_{F_Y} [R(\check{p}_U(data_Y), F_Y)] \right),$$

where the expectation $\mathbb{E}_{F_Y}$ is taken with respect to $data_Y \sim F_Y$. In the definition above, $\check{p}_U(data_Y)$ is a uniform pricing function in $\mathcal{U}$ and $\check{p}_U(x_0; data_Y)$ corresponds to its value at a covariate $x_0 \in [0,1]$; moreover, $\check{\mathcal{U}}$ is the set of *all data-based* uniform pricing functions $\check{p}_U$.

In what follows, we derive a lower bound for $\mathscr{R}_n^D(x_0;\mathcal{F})$, $\mathscr{R}_n^D(\mathcal{F})$ and $\mathscr{R}_n^U(\mathcal{F}^U)$, respectively. These lower bounds are algorithm independent and reveal the fundamental information-theoretic limitation of data-based pricing strategies.

4.1 Price discrimination

The first theorem presents a lower bound for the revenue difference at a given covariate value x_0, between *any* data-based 3PD strategy and the true-distribution optimal 3PD strategy under the worst-case distribution by taking the supremum over $\mathcal{F}$.

Theorem 3 (Lower bounds for 3PD, deficiency in pointwise revenue). *Let Assumption 1 hold. For any $\mathcal{F}$ satisfying Assumption 2 with $C^* \in (0,2)$ in (2), the minimax difference in the revenues at a given covariate value x_0 is bounded from below as*

$$\mathscr{R}_n^D(x_0; \mathcal{F}) \gtrsim n^{-1/2}, \quad x_0 \in (0,1),$$

if $x_0 n^{1/4} \geq c'$ and $(1-x_0)n^{1/4} \geq c''$ for some positive universal constants c' and c'' (independent of n and x_0).

The second theorem presents a lower bound for the difference in *expected* revenues between *any* data-based 3PD strategy and the true-distribution optimal 3PD strategy under the worst-case distribution by taking the supremum over $\mathcal{F}$.

Theorem 4 (Lower bounds for 3PD, deficiency in expected revenue). *Let Assumption 1 hold. For any $\mathcal{F}$ satisfying Assumption 2 with $C^* \in (0,2)$ in (2), the minimax difference in the expected revenues is bounded from below as*

$$\mathscr{R}_n^D(\mathcal{F}) \gtrsim n^{-1/2}.$$

Remark. *By requiring $C^* \in (0,2)$ in the theorems above, we allow $r(y,x)$ associated with an $f_{Y|X}$ to have a second derivative bounded from above by a number smaller than or equal to -2. To motivate the use of $C^* \in (0,2)$, suppose $f_{Y|X} = f_Y$ (that is, the valuation and covariate are independent of each other) and f_Y is the uniform distribution on $[0,1]$, $U[0,1]$. In this case, the revenue function equals $R(y) = y(1-y)$, which is twice-differentiable with second-order derivative $R''(y) = -2$ for any $y \in [0,1]$. In our proof for the lower bounds, $U[0,1]$ is used as the benchmark distribution.*

Theorems 3 and 4 state that, there is an inevitable deficiency, $\Omega(n^{-1/2})$, in the revenue from *any* data-based 3PD strategy relative to the revenue from the true-distribution optimal 3PD strategy in the worst case by taking the supremum over $\mathcal{F}$.

Recalling Theorem 1 on the convergence rate $O(n^{-1/2})$ of the revenue from the K-markets ERM strategy, despite its simplicity, Theorems 3 and 4 imply that the revenue from this strategy achieves the optimal rate of convergence (as a function of n) to the revenue from the true-distribution optimal 3PD strategy uniformly over $\mathcal{F}$. In other words, more sophisticated pricing strategies (e.g., with partitioning the covariate space based on observed frequencies) cannot improve upon the K-market ERM algorithm asymptotically.

To prove the lower bounds, we convert the problem into a classification task that tries to distinguish between distributions that are sufficiently close to each other but yield significantly different optimal prices. This technique was used in Huang et al. (2018); there, the bound concerns data-based uniform pricing strategies, which only require constructing two distributions and simpler techniques. To establish the lower bound in Theorem 4, two distributions are far from being enough. The reason is that, unlike the uniform pricing problem where the optimal pricing function is a scalar parameter, the 3PD problem tries to learn an optimal pricing function of the covariate (an infinitely-dimensional parameter) and the deficiency in the *expected* revenue concerns the entire pricing function at all covariate values. The notion of packing sets in Kolmogorov and Tikhomirov (1959) and the Gilbert-Varshamov bound from coding theory are useful ingredients for proving Theorem 4. The most intricate part of the proof involves carefully constructing M conditional densities (where M grows with n) and bounding the separation between the optimal prices associated with these densities. The desired set of optimal prices in our proof is a packing set where the separation between elements is $\Omega(n^{-1/4})$ with respect to the unweighted L_2 norm, and the cardinality of this set is $\Omega(2^{n^{1/4}})$.

4.2 Uniform pricing

We have the following theorem for uniform pricing.

Theorem 5. *Let Assumption 1 hold. For any $\mathcal{F}^U$ satisfying the conditions in Assumption 3 with $C^* \in (0, 2)$ in (2), the minimax difference in the revenues is bounded from below as*

$$\mathscr{R}_n^U(\mathcal{F}^U) \gtrsim n^{-2/3}.$$

Theorem 5 states that there is an inevitable deficiency, $\Omega(n^{-2/3})$, in the revenue from *any* data-based uniform pricing strategy relative to the revenue from the true-distribution optimal uniform pricing strategy by taking the supremum over $\mathcal{F}^U$.

Recalling Corollary 1 on the convergence rate $O(n^{-2/3})$ of the 1-market ERM strategy, despite its simplicity, Theorem 5 implies that the revenue from this algorithm achieves the optimal rate of convergence (as a function of n) to the revenue from the true-distribution optimal uniform pricing strategy uniformly over $\mathcal{F}^U$.

4.3 Sketches of the proofs

To facilitate understanding, we start with a preliminary of the proof for Theorem 3 before laying out the preliminaries for Theorems 4 and 5.

4.3.1 Preliminary of the proof for Theorem 3

For Theorem 3, we first show that the minimax difference in price at a given covariate value x_0 is bounded from below as follows:

$$\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}_{F_{Y,X}} |\check{p}_D(x_0; data) - p_D^*(x_0; F_{Y,X})| \gtrsim n^{-1/4}, x_0 \in (0,1). \tag{6}$$

Using Taylor expansion type of arguments and condition (2), we can relate the revenue difference to the minimax squared difference in price at x_0:

$$\begin{aligned}\mathscr{R}_n^D(x_0; \mathcal{F}) &\gtrsim \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}_{F_{Y,X}} \left[|\check{p}_D(x_0; data) - p_D^*(x_0; F_{Y,X})|^2 \right] \\ &\geq \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \left\{ \mathbb{E}_{F_{Y,X}} \left[|\check{p}_D(x_0; data) - p_D^*(x_0; F_{Y,X})| \right] \right\}^2\end{aligned}$$

where the last line follows from the Jensen's inequality.

The derivation of the lower bound (6) can be reduced to a binary classification problem. In a binary classification problem, we have two distributions $F_{Y,X}^1, F_{Y,X}^2 \in \mathcal{F}$ whose optimal prices are separated by some number 2ε; that is,

$$|p_D^*(x_0; F_{Y,X}^{j'}) - p_D^*(x_0; F_{Y,X}^j)| \geq 2\varepsilon, \quad j, j' \in \{1,2\}. \tag{7}$$

A binary classification rule uses the data to decide whether the true distribution is $F_{Y,X}^1$ or $F_{Y,X}^2$. To relate the binary classification problem to the pricing problem, note that, given any pricing function $\check{p}_D$, we can use it to distinguish between $F_{Y,X}^1$ and $F_{Y,X}^2$ in the following way. Define the binary classification rule

$$\psi(data) = \arg\min_{j \in \{1,2\}} |p_D^*(x_0; F_{Y,X}^j) - \check{p}_D(x_0; data)|.$$

We claim that when the underlying distribution is $F^j_{Y,X}$, the decision rule ψ is correct if

$$|p^*_D(x_0; F^j_{Y,X}) - \check{p}_D(x_0; data)| < \varepsilon. \tag{8}$$

To see this, note that by the triangle inequality, (7) and (8) guarantee that

$$\begin{aligned}
&|p^*_D(x_0; F^{j'}_{Y,X}) - \check{p}_D(x_0; data)| \\
\geq &|p^*_D(x_0; F^{j'}_{Y,X}) - p^*_D(x_0; F^j_{Y,X})| - |p^*_D(x_0; F^j_{Y,X}) - \check{p}_D(x_0; data)| \\
> &2\varepsilon - \varepsilon = \varepsilon, \text{ where } j' \neq j, \ j, j' \in \{1, 2\}.
\end{aligned}$$

This implies that

$$\mathbb{P}_{F^j_{Y,X}}(\psi(data) \neq j) \leq \mathbb{P}_{F^j_{Y,X}}(|p^*_D(x_0; F^j_{Y,X}) - \check{p}_D(x_0; data)| \geq \varepsilon), \quad j = 1, 2.$$

Therefore, we can upper bound the average probability of mistakes in the binary classification problem as

$$\begin{aligned}
&\frac{1}{2}\mathbb{P}_{F^1_{Y,X}}(\psi(data) \neq 1) + \frac{1}{2}\mathbb{P}_{F^2_{Y,X}}(\psi(data) \neq 2) \\
\leq &\frac{1}{2}\mathbb{P}_{F^1_{Y,X}}(|p^*_D(x_0; F^1_{Y,X}) - \check{p}_D(x_0; data)| \geq \varepsilon) + \frac{1}{2}\mathbb{P}_{F^2_{Y,X}}(|p^*_D(x_0; F^2_{Y,X}) - \check{p}_D(x_0; data)| \geq \varepsilon) \\
\leq &\sup_{F_{Y,X} \in \mathcal{F}} \mathbb{P}_{F_{Y,X}}(|p^*_D(x_0; F_{Y,X}) - \check{p}_D(x_0; data)| \geq \varepsilon).
\end{aligned}$$

By the Markov inequality, we have

$$\begin{aligned}
&\sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}|\check{p}_D(x_0; data) - p^*_D(x_0; F_{Y,X})| \\
\geq &\varepsilon \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{P}\left(|\check{p}_D(x_0; data) - p^*_D(x_0; F_{Y,X})| \geq \varepsilon\right) \\
\geq &\varepsilon \left(\frac{1}{2}\mathbb{P}_{F^1_{Y,X}}(\psi(data) \neq 1) + \frac{1}{2}\mathbb{P}_{F^2_{Y,X}}(\psi(data) \neq 2)\right).
\end{aligned}$$

Finally, we take the infimum over all pricing strategies on the left-hand side (LHS), and the infimum over the induced set of binary decisions on the right-hand side (RHS). This leads to

$$\begin{aligned}
&\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}|\check{p}_D(x_0; data) - p^*_D(x_0; F_{Y,X})| \\
\geq &\varepsilon \inf_{\psi} \left(\frac{1}{2}\mathbb{P}_{F^1_{Y,X}}(\psi(data) \neq 1) + \frac{1}{2}\mathbb{P}_{F^2_{Y,X}}(\psi(data) \neq 2)\right).
\end{aligned} \tag{9}$$

The RHS of the above inequality consists of two parts: (1) ε, related to the separation between two optimal prices, and (2) the average probability of making a mistake in distinguishing the two distributions. To obtain a meaningful bound, we want to find two distributions $F^1_{Y,X}$ and $F^2_{Y,X}$ that are close to each other (hard to distinguish) but their optimal prices are sufficiently separated. We leave the details of the construction of such distributions to the *proof of Theorem 3* given in Appendix B.

4.3.2 Preliminary of the proof for Theorem 4

For Theorem 4, we first show that the minimax (unweighted) L_2-distance in price is bounded from below as follows:

$$\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}\|\check{p}_D(data) - p^*_D(F_{Y,X})\|_2^2 \gtrsim n^{-1/2}$$

where

$$\|\check{p}_D(data) - p^*_D(F_{Y,X})\|_2^2 = \int_0^1 |\check{p}_D(x; data) - p^*_D(x; F_{Y,X})|^2 dx.$$

Using Taylor expansion type of arguments and condition (2), we can relate the difference in the *expected* revenues to the minimax (unweighted) L_2-distance in price:

$$\mathscr{R}^D_n(\mathcal{F}) \gtrsim \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}_{F_{Y,X}}\|\check{p}_D(data) - p^*_D(F_{Y,X})\|_2^2$$

where the expectation $\mathbb{E}_{F_{Y,X}}$ is taken with respect to $data \sim F_{Y,X}$.

The object above concerns the entire pricing function $p^*_D(\cdot; F_{Y,X})$. As a result, bounding the RHS of the above inequality is more complicated than the previous one (6). In particular, we consider a multiple classification problem that tries to distinguish among M distributions, where M is a function of the sample size n. Similar as before, we want the optimal prices of these M distributions to be sufficiently separated. Similar derivations show that the lower bound of the revenue problem can be reduced to that of a multiple classification problem:

$$\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}_{F_{Y,X}}\|\check{p}_D(data) - p^*_D(F_{Y,X})\|_2^2 \geq \varepsilon^2 \inf_{\psi} \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_{F^j_{Y,X}}(\psi(data) \neq j), \tag{10}$$

where the infimum $\inf_\psi$ is taken over the set of all multiple decisions (with M choices). To proceed, we apply the Fano's inequality from information theory (Cover and Thomas, 2005). Fano's inequality gives a lower bound on the average probability of mistakes:[9]

$$\frac{1}{M}\sum_{j=1}^{M}\mathbb{P}_{F^j_{Y,X}}(\psi(data)\neq j)\geq 1-\frac{\sum_{j,j'=1}^{M}\mathrm{KL}(F^j_{Y,X}\|F^{j'}_{Y,X})/M^2+\log 2}{\log M}, \tag{11}$$

where $\mathrm{KL}(\cdot\|\cdot)$ denotes the Kullback-Leibler (KL) divergence between two distributions:

$$\mathrm{KL}(F_1\|F_2)\equiv\int f_1(y,x)\log\frac{f_1(y,x)}{f_2(y,x)}dydx.$$

To obtain a sharp bound based on the multiple classification problem, we want to find a set of distributions (where the cardinality M of the set is large enough) that are close enough to each other (small enough pairwise KL divergence) but their optimal prices are sufficiently separated. We leave the detailed proof to Appendix B. Our proof is based on a delicate construction of conditional densities along with an application of the Gilbert-Varshamov Lemma from coding theory. Specifically, we use the distribution $Y,X\sim U[0,1]$ with X independent of Y as the benchmark distribution and construct its perturbed versions with some correlation.

4.3.3 Preliminary of the proof for Theorem 5

Relative to the proofs in the case of 3PD, the proofs for the price- and revenue-deficiency lower bounds in uniform pricing are simpler. We first show that the minimax difference in price is bounded from below as follows:

$$\inf_{\check{p}_U\in\check{\mathcal{U}}}\sup_{F_Y\in\mathcal{F}^U}\mathbb{E}_{F_Y}|\check{p}_U(data_Y)-p_U^*|\gtrsim n^{-1/3}. \tag{12}$$

As previously, we can relate the revenue difference to the minimax squared difference in price:

$$\begin{aligned}\mathscr{R}_n^U(\mathcal{F}^U)&\gtrsim\inf_{\check{p}_U\in\check{\mathcal{U}}}\sup_{F_Y\in\mathcal{F}^U}\mathbb{E}_{F_Y}\left[|\check{p}_U(data_Y)-p_U^*|^2\right]\\&\geq\inf_{\check{p}_U\in\check{\mathcal{U}}}\sup_{F_Y\in\mathcal{F}^U}\left\{\mathbb{E}_{F_Y}\left[|\check{p}_U(data_Y)-p_U^*|\right]\right\}^2\end{aligned}$$

where the last line follows from the Jensen's inequlity. The derivation of (12) only requires constructing two distributions, similar to the approach discussed in Section 4.3.1.

[9] We do not present the Fano's inequality in its standard form as in Cover and Thomas (2005). Instead, we use a version from Wainwright (2019) that is more convenient for our purposes.

5 Numerical evidence

Sections 3 and 4 establish that the K-market ERM strategy achieves the optimal rates of convergence in revenue uniformly over a class of distributions. In this section, we turn to specific distributions and study the revenue performance of our K-markets ERM strategies in these cases. We present numerical evidence that supports the implications of our theoretical results. Specifically, we calculate the revenues of the pricing strategies proposed in Section 3 using real-world and simulated data. We describe the data in detail below.

Data. For the empirical study, we use an eBay auction data set (Jank and Shmueli, 2010). Because eBay uses a sealed-bid second-price auction format, the bid of each participant can serve as a proxy for an individual valuation of the object. In particular, we use the data on 194 7-day auctions for the new Palm Pilot M515 PDAs.[10] The data has 3,832 observations at the bid level, and each observation includes an auction id, a bid amount, a bidder id, and a bidder rating. Some bidders appear in the data set several times because either they revised their bid during an auction or participated in several auctions. To be consistent with our assumption of independent sampling, we analyze the data at the bidder level and use the highest bid of each bidder across all auctions she participated in as the one representing her valuation. This leaves 1,203 observations from which we draw samples of various sizes. For Y_i, we use the bid (as described above) of bidder i normalized to $[0, 1]$. For X_i in the 3PD case, we use bidder i's rating on eBay, which indicates the number of times sellers left feedback after a transaction with i.

For the simulation study, we let the marginal distribution of X be uniform on $[0, 1]$ and the CDF of Y conditional on $X = x$ be

$$F_{Y|X}(y|x) = y^{x+1}. \tag{13}$$

Implementation. For each type of data, we calculate (a Monte-Carlo approximation of) the expected revenue generated by the uniform ERM and the K-markets ERM strategies for various sample sizes as follows. First, fix n and K. Then, draw a sample $\{Y_i, X_i\}_{i=1}^n$ and, for each $k = 1, \ldots, K$, let

$$\text{market}_k \equiv \{Y_i : X_i \in I_k\}, \quad \hat{F}_k(t) \equiv \frac{|Y_i \in \text{market}_k : Y_i \le t|}{|\text{market}_k|}.$$

[10] Jank and Shmueli (2010) also provide data on Cartier wristwatches, Swarovski beads, and Xbox game consoles, but each of these data sets may pool various configurations or models of these products categories. Thus, we choose the data on the Palm Pilot M515 to minimize such variations.

Then, the empirical optimal price in the kth market is given by

$$\hat{p}_{D,k} \equiv \operatorname*{argmax}_{y \in [0,1]} y(1 - \hat{F}_k(y)) = \operatorname*{argmax}_{y \in \text{market}_k} y(1 - \hat{F}_k(y)),$$

where the second equality holds because $\hat{F}_k$ is a step function. Note that the uniform ERM strategy simply corresponds to the 1-market ERM strategy. When $K > 1$ and a drawn sample results in empty markets that contain no observations, we set the prices in those markets to one. Finally, we compute the revenue deficiency for the uniform ERM and K-markets ERM strategies (under $K \asymp n^{1/4}$).

Numerical findings. Figure 1 plots the expected revenue generated by the K-markets ERM strategy for $K \in \{1, \ldots, 5\}$ as a function of the sample size n (with $K = 1$ corresponding to the uniform ERM strategy). To facilitate the exposition, we use a logarithmic scale for the n-axis. For both types of data, one can see that for sufficiently small n, the K-markets revenue is *decreasing* in K. As n grows, the performance of higher K improves faster than that of lower K, and, for sufficiently large n, the K-markets revenue overtakes that with any lower K. This finding can be explained by the bound $(K/n)^{2/3} + 1/K^2$ in Theorem 1(ii), which implies that higher K (more discrimination) approximates the revenue generated by the $F_{Y,X}$-optimum better but incurs a larger "variance". When the sample size is small, a lower K can indeed be more beneficial.

Figure 1 also suggests that, even if X contains useful information about Y, the uniform ERM strategy may be revenue superior to any $K(> 1)$-markets ERM strategy when n is sufficiently small. Recall from Theorem 1 that the bound $(K/n)^{2/3} + 1/K^2$ is minimized at $K = n^{1/4}$, which gives $n^{-1/2}$, the optimal rate of convergence to the revenue generated by the $F_{Y,X}$-optimal 3PD strategy. This convergence rate is slower than $n^{-2/3}$, the optimal rate of convergence to the revenue generated by the F_Y-optimal uniform pricing strategy (c.f. Corollary 1). The slower convergence of the rate-optimal K-market ERM strategy can potentially dominate the revenue gain from price discrimination over without discrimination for small n.

Figure 2 illustrates the difference in the convergence rates of the uniform ERM and the K-markets ERM strategies to their respective theoretical benchmarks. In particular, we set $K = \frac{1}{5}\lfloor n^{1/4} \rfloor$ for the simulation study and $K = \max\{1, \lfloor 2n^{1/4} - 7 \rfloor\}$ for the empirical study. As predicted by the rate $n^{-1/2}$ in Theorem 1 and the rate $n^{-2/3}$ in Corollary 1, the revenue from the uniform ERM strategy is converging to the revenue from the F_Y-optimal uniform pricing strategy faster than the K-markets revenue to the revenue from the $F_{Y,X}$-optimal 3PD strategy.

Figure 3 exhibits the revenue under the K-markets ERM strategy for $K = 1, \ldots, 5$ and $n =$

$2, \ldots, 10^5$, in the case where X and Y are uniform on $[0, 1]$ and independent of each other. Not surprisingly, there is no benefit from price discrimination for revenue.

Figure 1: Revenue under uniform and K-markets ERM strategies

(a) Empirical data

(b) Data simulated from (13)

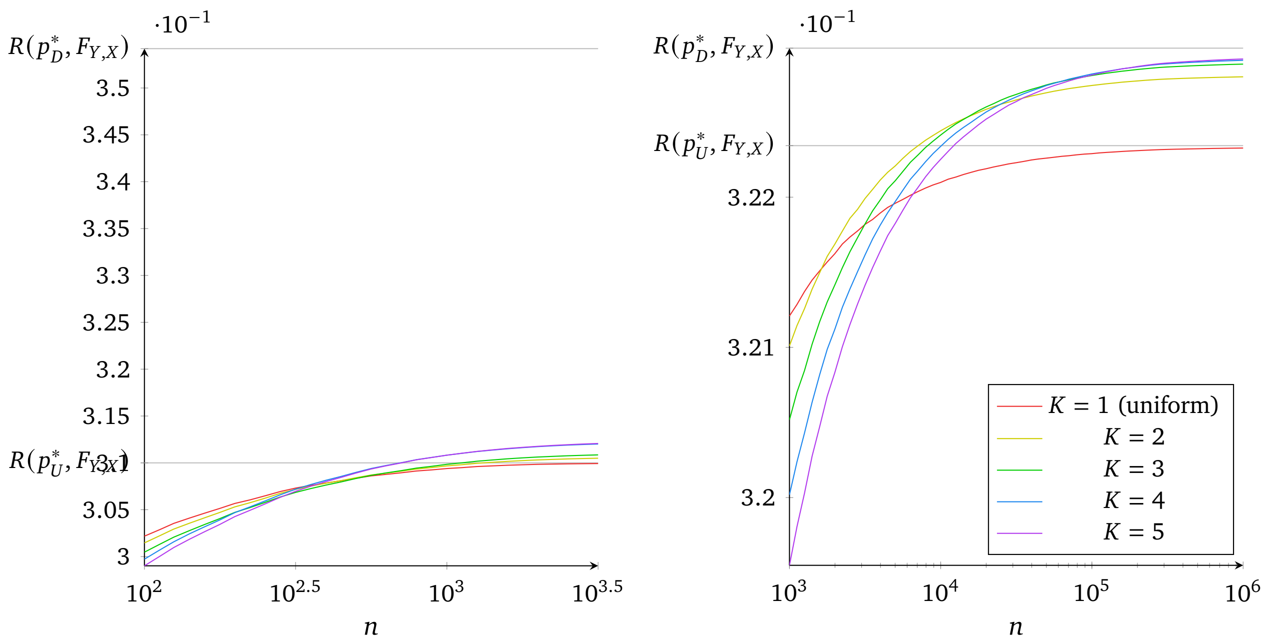


Figure 2: Data-based revenue deficiency under uniform and K-markets ERM strategies (with $K \asymp n^{1/4}$).

(a) Empirical data

(b) Data simulated from (13)

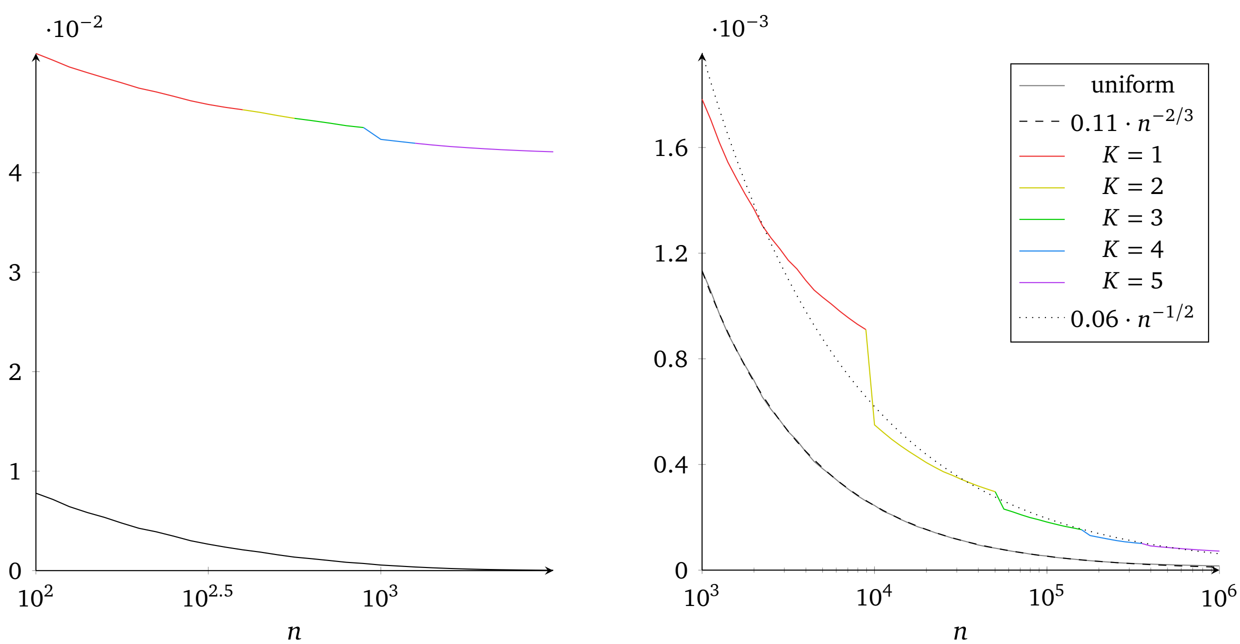
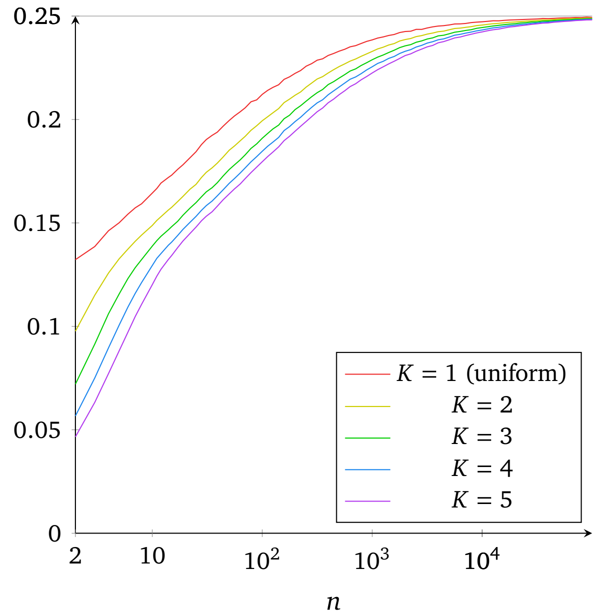


Figure 3: Uniform and K-markets revenue for the case of X and Y uniform on $[0, 1]$ and independent of each other.



6 Discussions

Recall that p_D^* is the true-distribution optimal 3PD strategy and $\hat{p}_D$ is the K-markets ERM strategy with $K = \Theta(n^{1/4})$ giving the best trade-off between the "variance" and approximation error as shown in Theorem 1; p_U^* is the true-distribution optimal uniform pricing strategy and $\hat{p}_U$ is the uniform ERM strategy. We can decompose the difference between the expected revenues generated respectively from $\hat{p}_D$ and $\hat{p}_U$ as follows:

$$\mathbb{E}[R(\hat{p}_D)] - \mathbb{E}[R(\hat{p}_U)] = -\underbrace{\left\{R(p_D^*) - \mathbb{E}[R(\hat{p}_D)]\right\}}_{A_1} + \underbrace{R(p_D^*) - R(p_U^*)}_{A_2} + \underbrace{R(p_U^*) - \mathbb{E}[R(\hat{p}_U)]}_{A_3}.$$

The first term $A_1 = \Theta\left(n^{-1/2}\right)$ under a worst-case distribution $F_{Y,X} \in \mathcal{F}$, and the third term $A_3 = O\left(n^{-2/3}\right)$ under F_Y, the marginal of $F_{Y,X}$. The second term $A_2 = \Theta(1)$ when X contains sufficient information about the valuation Y. Then, a sufficient condition for $\hat{p}_D$ to be revenue superior to $\hat{p}_U$ is that $n \to \infty$. In theory, this claim can be proved with the upper bounds in Section 3 and a different construction in the derivations of the lower bounds. Particularly, this new construction would first find a density $f_{Y,X}$ such that the revenue generated by the corresponding $f_{Y,X}$-optimal 3PD strategy is well separated from the revenue generated by the optimal uniform pricing strategy associated with f_Y, and then build a large enough class of perturbed versions of $f_{Y,X}$; finally we would bound the separation between the optimal prices associated with these densities, in a similar fashion as what is done in Appendix B. In the paper, to make the analysis tractable, we choose the distribution $Y, X \sim U[0,1]$ with X independent of Y as the benchmark distribution and construct its perturbed versions with some correlation.

A challenging open question is, can the condition on n be weakened to some finite number and if so, when? To answer this question, we would have to derive the universal constants in our bounds in meaningful forms. Unfortunately, due to the complexity of our problem, this exercise is infeasible under the existing techniques from mathematical statistics, probability theory, and information theory.

Our results suggest that it is more beneficial to engage in sample-based uniform pricing when X is independent of Y.[11] The fundamental reason lies in the proofs for Theorems 3 and 4: unless $n = \infty$, no strategies that exploit $\{(Y_i, X_i), 1 \le i \le n\}$ are able to distinguish with certainty the distribution $Y, X \sim U[0,1]$ with X independent of Y from its perturbed versions with some correlation (see the detailed constructions in Appendix B). The curse of dimensionality from exploiting the covariate X makes the convergence of 3PD strategies based on $\{(Y_i, X_i), 1 \le i \le n\}$ slower than that of the uniform pricing strategies based on $\{Y_i, 1 \le i \le n\}$.

Our upper and lower bounds together suggest the following possibility: even when the covariate X contains useful information about the valuation Y, the K-markets ERM strategy *can* be revenue inferior to the uniform ERM strategy in finite samples, due to the curse of dimensionality and slower convergence of the K-markets ERM strategy to its true-distribution optimal counterpart (and hence, a more stringent growth requirement of the sample size). Indeed, the numerical evidence in Section 5 confirms this possibility. But such an implication should be taken with caution in small samples.

Small sample complication. Given the pattern observed in our numerical studies, one might conjecture the following: there exists some $\bar{n} > 1$ such that when $n < \bar{n}$, the uniform ERM is always revenue-superior to the K-markets ERM (with $K > 1$). In what follows, we explain why this conjecture may not hold.

Specifically, in our language, Babaioff et al. (2018) construct a distribution F_Y such that the uniform ERM revenue under $n = 2$ is strictly smaller than the uniform ERM revenue under $n = 1$. This seemingly counterintuitive result highlights the difficulty of establishing general comparative results

[11] The information of independence is unknown to the seller. He/she can statistically test for the independence of Y and X from the data but any such tests would suffer from Type I and Type II errors.

with very small sample size. We now argue that this construction also sheds some light on the comparison of the revenue performance of the K-markets ERM strategy with $K = 1$ vs $K = 2$ in the case of $n = 2$.

To make this connection, we take X to be uniform on $[0, 1]$ and independent of Y, and assume that in the case $K = 2$, when one of the markets is empty, the price for this market is set at the same level as for the other market. Then, if $K = 2$ and both markets are non-empty, the revenue in each market equals the 1-market ERM revenue under $n = 1$. Otherwise, if both observations are in the same market, then the revenue equals the 1-market ERM revenue under $n = 2$. Therefore, the expected 2-markets ERM revenue with $n = 2$ is the average of the 1-market ERM revenue under $n = 1$ and $n = 2$ and hence strictly higher than the 1-market ERM revenue with $n = 2$ for a distribution F_Y exhibiting the property discussed in Babaioff et al. (2018).

More formally, let $R_{K,n}$ denote the expected revenue of the K-markets ERM strategy with a sample of size n. Then,

$$
\begin{aligned}
R_{2,2} &= \mathbb{E}_{data^{n=2} \sim F_{Y,X}}[R(\hat{p}_D(data), F_{Y,X})] \\
&= \frac{1}{2}\mathbb{E}_{data^{n=2} \sim F_{Y,X} | I_1 = \varnothing \text{ or } I_2 = \varnothing}[R(\hat{p}_D(data), F_{Y,X})] \\
&\quad + \frac{1}{2}\mathbb{E}_{data^{n=2} \sim F_{Y,X} | I_1 \neq \varnothing \text{ and } I_2 \neq \varnothing}[R(\hat{p}_D(data), F_{Y,X})] \\
&= \frac{1}{2}R_{1,2} + \frac{1}{2}R_{1,1}.
\end{aligned}
$$

Therefore, $R_{1,1} > R_{1,2}$ implies $R_{2,2} > R_{1,2}$.

Finally, we add the caveat that the construction in Babaioff et al. (2018) is based on an atomless approximation of the censored equal-revenue distribution $F_Y(y) = 1 - 1/y, y \in [1, \infty)$ which has a discontinuous density. However, it is straightforward to verify that the same property holds for the equal-revenue distribution truncated at any $y > 4$, which has a Lipschitz continuous and differentiable density. Moreover, the equal revenue distribution truncated at y and translated to the left by $t > 1/y$ (so that the support is $[1 - t, y - t]$) also has a Lipschitz continuous and differentiable density, the interior optimal price (in line with our assumptions), and satisfies the Babaioff et al. (2018) property, e.g., for $y = 4, t = 1/2$.

An open problem. To conclude, we would like to propose a challenging open problem: Do there exist some $3 \leq \underline{n} < \bar{n} < \infty$ such that for any $n \in [\underline{n}, \bar{n}]$ and distribution in $\mathcal{F}$, the K-markets ERM strategy (for any $K > 1$) is always revenue-inferior to the uniform ERM strategy?

A Proofs for upper bounds

To facilitate the presentation, we first give the proof for Corollary 1.

Proof of Corollary 1. Denote $\kappa' \equiv \inf_{p\in[0,1]} |R''(p)|/2 > 0$. By Taylor expansion, for any p,

$$R(p_U^*) - R(p) \geq \kappa'(p - p_U^*)^2.$$

Denote $\hat{R}(p) \equiv p(1-\hat{F}(p))$. Combining the inequality above with the basic inequality (i.e., $\hat{R}(\hat{p}_U) \geq \hat{R}(p_U^*)$), we have

$$\kappa'(\hat{p}_U - p_U^*)^2 \leq R(p_U^*) - R(\hat{p}_U) \leq R(p_U^*) - \hat{R}(p_U^*) - (R(\hat{p}_U) - \hat{R}(\hat{p}_U)). \tag{14}$$

For $\delta \in (0, p_U^*]$, define

$$\mathcal{G}_\delta \equiv \{y \mapsto p\mathbf{1}\{y \geq p\} - p_U^*\mathbf{1}\{y \geq p^*\} \colon p \in [p_U^* - \delta, p_U^* + \delta]\}$$

and

$$G_\delta(y) \equiv \begin{cases} 0, & \text{if } y < p_U^* - \delta, \\ p_U^*, & \text{if } p_U^* - \delta \leq y \leq p_U^* + \delta, \\ \delta, & \text{if } y > p_U^* + \delta. \end{cases}$$

Then G_δ is an envelope function of the class $\mathcal{G}_\delta$. The $L_2\,(P)$ −norm of G_δ is bounded by

$$\|G_\delta\|_{L_2(P)} = \left((p_U^*)^2\mathbb{P}(Y \in [p_U^* - \delta, p_U^* + \delta]) + \delta^2\mathbb{P}(Y > p_U^* + \delta)\right)^{1/2} \leq C\sqrt{\delta}.$$

As we argue in the proof of Lemma C.6, $\mathcal{G}_\delta$ is a VC-subgraph class, so we have

$$\mathbb{E}\sup_{g\in\mathcal{G}_\delta}\left|\frac{1}{n}\sum_{i=1}^n g(Y_i) - \mathbb{E}g(Y_i)\right| \leq C\sqrt{\delta/n}. \tag{15}$$

We derive the convergence rate of $\hat{p} - p^*$ via a peeling argument. Consider the following decomposition

$$\mathbb{P}\left(n^{1/3}|\hat{p}_U - p_U^*| > M\right) = \sum_{j=M+1}^{\infty} \mathbb{P}\left(n^{1/3}|\hat{p}_U - p_U^*| \in (j-1, j]\right).$$

For any $j \geq M+1$, we have

$$
\begin{aligned}
&\{|\hat{p}_U - p_U^*| \in ((j-1)n^{-1/3}, jn^{-1/3}]\} \\
=&\{|\hat{p}_U - p_U^*| > (j-1)n^{-1/3}, |\hat{p}_U - p_U^*| \leq jn^{-1/3}\} \\
\subset& \left\{R(p_U^*) - \hat{R}(p_U^*) - (R(\hat{p}_U) - \hat{R}(\hat{p}_U)) \geq \kappa'(j-1)^2 n^{-2/3}, |\hat{p}_U - p_U^*| \leq jn^{-1/3}\right\} \\
\subset& \left\{\Delta_{j,n} \geq \kappa'(j-1)^2 n^{-2/3}\right\},
\end{aligned}
$$

where the third line follows from (14), and $\Delta_{j,n}$ in the last line is defined as

$$
\Delta_{j,n} \equiv \sup_{g \in \mathcal{G}_{jn^{-1/3}}} \left| \frac{1}{n} \sum_{i=1}^{n} g(Y_i) - \mathbb{E} g(Y_i) \right|.
$$

Therefore,

$$
\mathbb{P}\left(|\hat{p}_U - p_U^*| \in ((j-1)n^{-1/3}, jn^{-1/3}]\right) \leq \mathbb{P}\left(\Delta_{j,n} \geq \kappa'(j-1)^2 n^{-2/3}\right).
$$

To bound the probability on the RHS of the above inequality, we use the concentration inequality given by Theorem 7.3 in Bousquet (2003), which is a version of Talagrand's (1996) inequality. The concentration inequality states that for all $t > 0$,

$$
\mathbb{P}\left(\Delta_{j,n} \geq \mathbb{E}\Delta_{j,n} + \sqrt{2t(\sigma^2 + 2\mathbb{E}\Delta_{j,n})/n} + t/(3n)\right) \leq \exp(-ct),
$$

for some universal constant $c > 0$, where

$$
\sigma^2 \equiv \sup_{g \in \mathcal{G}_{jn^{-1/3}}} \mathbb{E} g(Y_1)^2 \leq \|\mathcal{G}_{jn^{-1/3}}\|_{L_2}^2 \leq Cjn^{-1/3}.
$$

From (15), we have

$$
\mathbb{E}\Delta_{j,n} \leq C\sqrt{jn^{-1/3}/n} = C\sqrt{j}n^{-2/3}.
$$

By setting $t = \kappa' j^2$, we have

$$
\begin{aligned}
&\mathbb{E}\Delta_{j,n} + \sqrt{2t(\sigma^2 + 2\mathbb{E}\Delta_{j,n})/n} + t/(3n) \\
\leq& C\sqrt{j}n^{-2/3} + \sqrt{2\kappa' j^2(Cjn^{-1/3} + 2C\sqrt{j}n^{-2/3})/n} + \kappa' j^2/(3n) \\
\leq& C' j^{3/2} n^{-2/3} \leq C^*(j-1)^2 n^{-2/3},
\end{aligned}
$$

when j is large enough. Then we have

$$\mathbb{P}\left(\Delta_{j,n} \geq C^*(j-1)^2 n^{-2/3}\right) \leq \mathbb{P}\left(\Delta_{j,n} \geq Cjn^{-2/3}\right) \leq \exp(-c\kappa' j^2), \text{ for } j \text{ large.}$$

To summarize, we have shown that

$$\mathbb{P}\left(n^{1/3}|\hat{p}_U - p_U^*| > M\right) \leq \sum_{j=M+1}^{\infty} \exp(-C_1 j^2) \leq C_3 \exp(-C_2 M^2).$$

By integrating the tail probability, we have

$$\mathbb{E}|\hat{p}_U - p_U^*|^s \lesssim n^{-s/3}. \tag{16}$$

For revenue, we use the second-order Taylor expansion and obtain that

$$\mathbb{E}[R(p_U^*) - R(\hat{p}_U)] \leq \sup_p |R''(p)| \mathbb{E}(\hat{p}_U - p_U^*)^2 \lesssim n^{-2/3}.$$

□

Proof of Theorem 1. We introduce some notations. Let $\tilde{R}_k(p)$ denote the revenue collected from the kth market by charging price p; that is,

$$\begin{aligned}\tilde{R}_k(p) &\equiv p\mathbb{P}(Y > p, X \in I_k) \\ &= p \int_p^1 \int_{I_k} f_{Y|X}(y|x) f_X(x) dx dy.\end{aligned}$$

Denote $\tilde{p}_k \equiv \operatorname{argmax}_{p \in [0,1]} \tilde{R}_k(p)$ as the maximizer of $\tilde{R}_k$. The first- and second-order derivatives of $\tilde{R}_k$ are respectively

$$\begin{aligned}\tilde{R}_k'(p) &= \int_p^1 \int_{I_k} f_{Y|X}(y|x) f_X(x) dx dy - p \int_{I_k} f_{Y|X}(p|x) f_X(x) dx, \\ \tilde{R}_k''(p) &= \int_{I_k} \left(-2 f_{Y|X}(p|x) - p \frac{\partial}{\partial y} f_{Y|X}(p|x)\right) f_X(x) dx.\end{aligned}$$

By the Lipschitz continuity assumption, the second-order derivative $\tilde{R}_k''(p)$ exists for almost all $p \in [0, 1]$. Recall that

$$-2 f_{Y|X}(p|x) - p \frac{\partial}{\partial y} f_{Y|X}(p|x) \leq -C^*,$$

and f_X is bounded away from zero. Denote $2\kappa'' \equiv C^* \inf_{x\in[0,1]} f_X(x)$. Then

$$\tilde{R}_k''(p) \le -2\kappa'' \int_{I_k} dx = -2\kappa''/K$$

for almost all $p \in [0,1]$. By Lemma C.1, we have

$$\tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(p) = |\tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(p)| \ge \frac{\kappa''}{K}(\tilde{p}_k - p)^2, p \in [0,1]. \tag{17}$$

Note that $\tilde{p}_k$ is not the true optimal price under $F_{Y,X}$. We need to relate it to the true optimal price. Let $k(x_0)$ be such that $x_0 \in I_k$. Then by the triangle inequality, we can decompose the pricing difference into estimation error and approximation error:

$$\begin{aligned}\mathbb{E}|\hat{p}_D(x_0; data) - p_D^*(x_0)| &= \mathbb{E}|\hat{p}_{k(x_0)} - p_D^*(x_0)| \\ &\le \underbrace{\mathbb{E}|\hat{p}_{k(x_0)} - \tilde{p}_{k(x_0)}|}_{\text{Estimation error}} + \underbrace{|\tilde{p}_{k(x_0)} - p_D^*(x_0)|}_{\text{Approximation error}}. \end{aligned} \tag{18}$$

Estimation error. Denote $\hat{R}_k$ as the empirical counterpart of $\tilde{R}_k$; that is,

$$\hat{R}_k(p) \equiv \frac{p}{n_k} \sum_{i\in\{j:X_j\in I_k\}} \mathbf{1}\{Y_i > p, X_i \in I_k\}.$$

Recall that $\hat{p}_k$ is the maximizer of $\hat{R}_k$. The basic inequality (i.e., $\hat{R}_k(\hat{p}_k) \ge \hat{R}_k(\tilde{p}_k)$) gives that

$$\begin{aligned}\tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(\hat{p}) &= \tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(\hat{p}) - \hat{R}_k(\tilde{p}_k) + \hat{R}_k(\tilde{p}_k) \\ &\le \tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(\hat{p}) - \hat{R}_k(\tilde{p}_k) + \hat{R}_k(\hat{p}_k).\end{aligned} \tag{19}$$

Combining (17) and (19) yields

$$\frac{\kappa''}{K}(\tilde{p}_k - \hat{p}_k)^2 \le \tilde{R}_k(\tilde{p}_k) - \hat{R}_k(\tilde{p}_k) - (\tilde{R}_k(\hat{p}) - \hat{R}_k(\hat{p}_k)). \tag{20}$$

In each I_k ($k = 1, \ldots, K$), the optimal price is the same.

In what follows, $s = 1$ or $s = 2$. Conditioning on X_i where i falls in the kth market, the proof for Corollary 1, in particular, (16) yields

$$\mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s | X_i,\ i \in \{j : X_j \in I_k\}\right] \lesssim (1/n_k)^{s/3}. \tag{21}$$

By Assumption 2(vi), the ith observation falls into the kth market with probabilty $\frac{c_1}{K}$ and other markets with probability $\frac{K-c_1}{K}$. Let us consider the event $\mathcal{A}_q = \{n_k > qn\}$ where $q \in (0, \frac{c_1}{K})$. If K is large enough such that $\frac{c_1}{K} \leq \frac{1}{2}$, the classic binomial tail bound yields

$$\mathbb{P}(\mathcal{A}_q) > 1 - \exp(-n\mathrm{KL}(q||c_1/K)) \tag{22}$$

where

$$\mathrm{KL}(q||c_1/K) := q \log \frac{qK}{c_1} + (1-q) \log \frac{(1-q)K}{K-c_1}.$$

Therefore, we have

$$\mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s | X_i,\ i \in \{j : X_j \in I_k\}\right] \lesssim (qn)^{-s/3},\ \text{for a given } k,$$

with probability at least $1 - \exp(-n\mathrm{KL}(q||c_1/K))$. With a union bound, we also have

$$\mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s | X_i,\ i \in \{j : X_j \in I_k\}\right] \lesssim (qn)^{-s/3},\ \text{for all } k, \tag{23}$$

with probability at least $1 - \exp(-n\mathrm{KL}(q||c_1/K) + \log K)$.

Furthermore, we have

$$\mathrm{KL}(q||c_1/K) \geq \frac{1}{2}\left(\frac{c_1}{K} - q\right)^2. \tag{24}$$

We show a more general result

$$\mathrm{KL}(q||\alpha) =: g_q(\alpha) \geq \frac{(\alpha-q)^2}{2}$$

for any $q \in (0, \alpha)$. Because $g_q(\cdot)$ is twice differentiable and $g_q(q) = 0$, a second-order Taylor expansion gives

$$g_q(\alpha) = g_q^{'}(q)(\alpha - q) + \frac{g_q^{''}(t)}{2}(\alpha-q)^2$$

where $t \in [q, \alpha]$ and $g_q^{'}(t) = -\frac{q}{t} + \frac{1-q}{1-t}$. Note that $g_q^{'}(q) = 0$. Moreover, given $t \in (0,1)$ such that $\frac{1}{t^2} \geq 1$ and $\frac{1}{(1-t)^2} \geq 1$, we have

$$g_q^{''}(t) = \frac{q}{t^2} + \frac{1-q}{(1-t)^2} \geq 1.$$

As a consequence of (23) and (24), we have

$$\mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s | X_i,\ i \in \{j : X_j \in I_k\}\right] \lesssim (qn)^{-s/3},\ \text{for all } k, \tag{25}$$

with probability at least $1 - \exp\left(-n\left(\frac{c_1}{K} - q\right)^2/2 + \log K\right)$. Taking $q = \frac{c_1}{2K}$, (25) gives

$$\mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s | X_i,\ i \in \{j : X_j \in I_k\}\right] \lesssim (K/n)^{s/3},\ \text{for all } k, \tag{26}$$

with probability at least $1 - \exp\left(-\frac{nc_1^2}{8K^2} + \log K\right)$.

In view of (21), (23), (25) and (26), the source of uncertainty from the conditioning is solely from the statistics n_k. Using this fact, (26) as well as the fact that $\hat{p}_k$ and $\tilde{p}_k$ are bounded, we have

$$\begin{aligned}\mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s\right] &= \mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s 1\left\{n_k > \frac{nc_1}{2K}\right\}\right] + \mathbb{E}\left[|\hat{p}_k - \tilde{p}_k|^s 1\left\{n_k \le \frac{nc_1}{2K}\right\}\right] \\ &\lesssim (K/n)^{s/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K\right).\end{aligned}$$

Approximation error. The second term $|\tilde{p}_{k(x_0)} - p_D^*(x_0)|$ in (18) is deterministic and can be controlled by using the smoothness conditions. By definition, $p_D^*(x_0)$ satisfies the first-order condition

$$0 = \frac{\partial}{\partial p} r(p_D^*(x), x).$$

By the differentiability condition of $\mathcal{F}$, $\frac{\partial}{\partial p} r(p, x)$ is continuously differentiable in (p, x) in a neighborhood of $(p_D^*(x), x)$. By the strong concavity, $\frac{\partial^2}{\partial p^2} r(p_D^*(x), x)$ is non-zero. Then by the implicit function theorem, the function $p_D^*(x)$ is well-defined (uniquely determined by the first-order condition) and is differentiable. Its derivative is given as follows:

$$\frac{d}{dx} p_D^*(x) = -\frac{\frac{\partial^2}{\partial p \partial x} r(p_D^*(x), x)}{\frac{\partial^2}{\partial p^2} r(p_D^*(x), x)}.$$

By the strong concavity, the absolute value of $\frac{\partial^2}{\partial p^2} r(p, x)$ is bounded away from zero; also, the function $\left|\frac{\partial}{\partial x} F_{Y|X}(y|x) + y \frac{\partial}{\partial x} f_{Y|X}(y|x)\right|$ is bounded above by $\bar{C}$. This implies that $p_D^*(x)$ is Lipschitz continuous on $[0, 1]$. We use L_1 to denote the Lipschitz constant. By applying Taylor expansion to the first-order condition of $\tilde{p}_k$, we have

$$\begin{aligned}0 = \int_{I_k} \frac{\partial}{\partial p} r(\tilde{p}_k, x) f_X(x) dx = \int_{I_k} &\underbrace{\frac{\partial}{\partial p} r(p_D^*(x), x)}_{=0} f_X(x) dx \\ &+ \int_{I_k} \frac{\partial^2}{\partial p^2} r(\bar{p}(x), x)(\tilde{p}_k - p_D^*(x)) f_X(x) dx,\end{aligned}$$

for some $\bar{p}(x)$ between $\tilde{p}_k$ and $p_D^*(x)$. Rearranging terms shows that $\tilde{p}_k$ is a weighted average of

$p_D^*(x), x \in I_k$; that is,

$$\tilde{p}_k = \frac{\int_{I_k} \frac{\partial^2}{\partial p^2} r(\bar{p}(x), x) p_D^*(x) f_X(x) dx}{\int_{I_k} \frac{\partial^2}{\partial p^2} r(\bar{p}(x), x) f_X(x) dx}.$$

Since $p_D^*(x)$ is Lipschitz continuous, the triangle inequality implies that

$$|\tilde{p}_k - p_D^*(x_0)|^s \leq L_1^s / K^s, \text{ for any fixed } s \geq 1.$$

Therefore, we obtain the following upper bound

$$\mathbb{E}|\hat{p}_D(x_0; data) - p_D^*(x_0)|^2 \lesssim 1/K^2 + (K/n)^{2/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K\right).$$

By choosing $K \asymp n^{-1/4}$, the above bound becomes $n^{-1/4}$.

In addition, Lemma C.1(iii) gives that

$$\begin{aligned}
\mathbb{E}\left[r(p_D^*, x_0) - r(\hat{p}_D(data), x_0)\right] &\lesssim \mathbb{E}\left[|\hat{p}_D(x_0; data) - p_D^*(x_0)|^2\right] \\
&\lesssim 1/K^2 + (K/n)^{2/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K\right).
\end{aligned}$$

This proves part (i) of the theorem.

For part (ii), we want to bound the expected revenue difference. Consider the following decomposition:

$$\begin{aligned}
&R(p_D^*, F_{Y,X}) - R(\hat{p}_D(data), F_{Y,X}) \\
\leq &R(p_D^*, F_{Y,X}) - R(\tilde{p}, F_{Y,X}) + |R(\tilde{p}, F_{Y,X}) - R(\hat{p}_D(data), F_{Y,X})|.
\end{aligned}$$

The first term on the RHS is deterministic and can be bounded by using Lemma C.1(iii) as follows:

$$\begin{aligned}
&|R(p_D^*, F_{Y,X}) - R(\tilde{p}, F_{Y,X})| \\
\leq &\int_0^1 |r(p_D^*(x), x) - r(\tilde{p}(x), x)| f_X(x) dx \\
= &\sum_{k=1}^K \int_{I_k} |r(p_D^*(x), x) - r(\tilde{p}_k, x)| f_X(x) dx \\
\leq &\sum_{k=1}^K \int_{I_k} \frac{1}{2} \left| 2 f_{Y|X}(y|x) + y \frac{\partial}{\partial y} f_{Y|X}(y|x) \right| \sup_{y,x} |p_D^*(x) - \tilde{p}_k|^2 f_X(x) dx \\
\lesssim &1/K^2.
\end{aligned}$$

where we have used the first-order condition of p_D^*. For the second term, we have

$$R(\tilde{p}, F_{Y,X}) - R(\hat{p}_D(data), F_{Y,X}) = \sum_{k=1}^{K} \tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(\hat{p}_k).$$

This is because both $\hat{p}(data)$ and $\tilde{p}$ are constant within each I_k. Their revenues on I_k are reduced to $\tilde{R}_k$. Note that for every k, $\tilde{R}_k'(\tilde{p}_k) = 0$, and

$$\begin{aligned}|\tilde{R}_k''(p)| &\leq \int_{I_k} \left| 2 f_{Y|X}(y|x) + y \frac{\partial}{\partial y} f_{Y|X}(y|x) \right| f_X(x) dx \\ &\leq \frac{1}{K} \sup_{y,x} \left(\left| 2 f_{Y|X}(y|x) + y \frac{\partial}{\partial y} f_{Y|X}(y|x) \right| f_X(x) \right).\end{aligned}$$

Then Lemma C.1(iii) gives that

$$\tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(\hat{p}_k) \lesssim 1/K(\tilde{p}_k - \hat{p}_k)^2.$$

Hence, we have

$$\begin{aligned}\mathbb{E}|R(\tilde{p}, F_{Y,X}) - R(\hat{p}_D(data), F_{Y,X})| &\leq \sum_{k=1}^{K} \mathbb{E}|\tilde{R}_k(\tilde{p}_k) - \tilde{R}_k(\hat{p}_k)| \\ &\lesssim (K/n)^{2/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K \right).\end{aligned}$$

To summarize, we have shown that

$$R(p_D^*, F_{Y,X}) - R(\hat{p}_D(data), F_{Y,X}) \lesssim 1/K^2 + (K/n)^{2/3} + \exp\left(-\frac{nc_1^2}{8K^2} + \log K \right).$$

By choosing $K \asymp n^{-1/4}$, the above bound becomes $n^{-1/2}$. This proves part (ii) of the theorem. □

Proof of Theorem 2. For part (i), notice that the welfare can be written as a double integral

$$W(p, F_{Y,X}) = \int_0^1 \int_0^{p(x)} y f_{Y|X}(y|x) dy f_X(x) dx.$$

The function $y f_{Y|X}(y|x)$ is nonnegative and bounded for $y, x \in [0, 1]$. Then by the integral mean value theorem, we have

$$
\begin{aligned}
&\mathbb{E}|W(\hat{p}_D(data), F_{Y,X}) - W(p_D^*, F_{Y,X})| \\
=&\mathbb{E}\Big|\int_0^1 \int_{p_D^*(x)}^{\hat{p}_D(x;data)} y f_{Y|X}(y|x) dy f_X(x) dx\Big| \\
\leq& \sup_{y,x} |y f_{Y|X}(y|x)| \mathbb{E}\int_0^1 |\hat{p}_D(x; data) - p_D^*(x)| f_X(x) dx.
\end{aligned}
$$

The integral on the last line can be decomposed based on the K markets:

$$
\begin{aligned}
\mathbb{E}\int_0^1 |\hat{p}_D(x; data) - p_D^*(x)| f_X(x) dx &\leq \sum_{k=1}^K \int_{I_k} \left[\mathbb{E}|\hat{p}_D(x; data) - \tilde{p}_k| + |\tilde{p}_k - p_D^*(x)|\right] f_X(x) dx \\
&= \sum_{k=1}^K \mathbb{E}|\hat{p}_k - \tilde{p}_k|/K + \sum_{k=1}^K \int_{I_k} |\tilde{p}_k - p_D^*(x)| f_X(x) dx \\
&\lesssim (K/n)^{1/3} + 1/K + \exp\left(-\frac{nc_1^2}{8K^2} + \log K\right) \asymp n^{-1/4},
\end{aligned}
$$

where the last line follows from the proof of Theorem 1.

For part (ii), since p_U^* is a scalar, the welfare can be simplified to

$$
W(p_U^*, F_Y) = \int_0^{p_U^*} y f_Y(y) dy.
$$

Then we have

$$
\begin{aligned}
\mathbb{E}|W(\hat{p}_U(data_Y), F_Y) - W(p_U^*, F_Y)| &= \mathbb{E}\Big|\int_{p_U^*}^{\hat{p}_U(data_Y)} y f_Y(y) dy\Big| \\
&\leq \sup_y |y f_Y(y)| \mathbb{E}|\hat{p}_U(data_Y) - p_U^*| \\
&\lesssim n^{-1/3},
\end{aligned}
$$

where we have used Corollary 1 along with the fact that $y f_Y(y)$ is nonnegative and bounded for $y \in [0, 1]$. □

B Proofs for lower bounds

Proof of Theorem 3. For Theorem 3, we use Lemma C.4 to prove the lower bound. Define

$$\omega_D(\epsilon) \equiv \sup_{F_1, F_2 \in \mathcal{F}} \left\{ |p_D^*(x_0; F_1) - p_D^*(x_0; F_2)| : H(F_1 \| F_2) \leq \epsilon \right\}.$$

By Lemma C.4, we have

$$\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}_{F_{Y,X}} |\check{p}_D(x_0; data) - p_D^*(x_0)| \geq \frac{1}{8} \omega_D \left(1/(2\sqrt{n}) \right).$$

Therefore, we only need to find a lower bound for ω_D. Based on the explanation in Section 4.3.1, we want to construct two distributions that are hard to distinguish but their optimal prices are well-separated. We start by defining two perturbation functions. Let ϕ_Y be defined as

$$\phi_Y(t) \equiv \begin{cases} t+1, & t \in [-1, 0], \\ -t+1, & t \in [0, 2], \\ t-3, & t \in [2, 3], \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

Notice that ϕ_Y is Lipschitz continuous on $\mathbb{R}$. Let ϕ_X be defined as

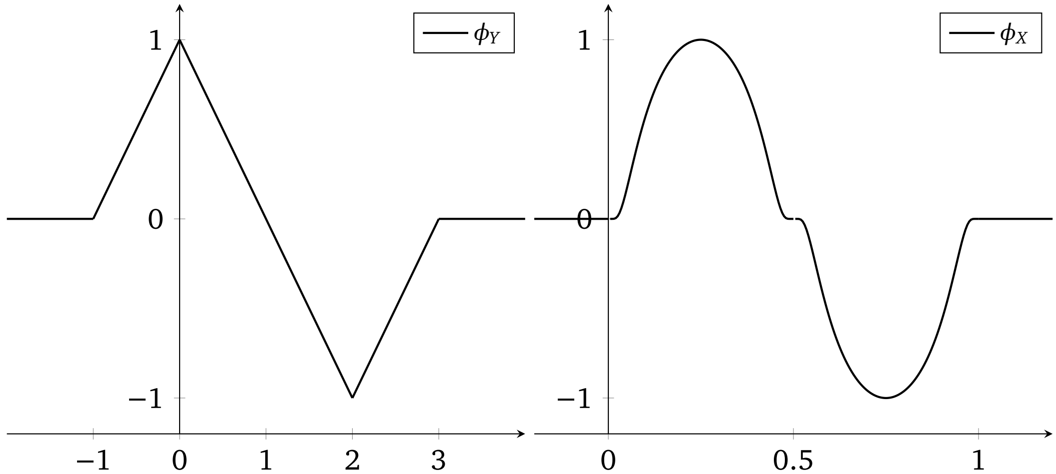
$$\phi_X(t) \equiv \begin{cases} e^{-(4t-1)^2/(1-(4t-1)^2)}, & t \in (0, 1/2), \\ -e^{-(4t-3)^2/(1-(4t-3)^2)}, & t \in (1/2, 1), \\ 0, & \text{otherwise.} \end{cases}$$

Notice that ϕ_X is infinitely differentiable on $\mathbb{R}$. We plot the two perturbation functions in Figure B.1.

Now we construct the two distributions. Let $\delta \in (0, 1/4)$ be a small number (that depends on n) to be specified later. Let a be any number in the interval $(0, 4 - 2C^*)$. Define the two conditional density functions of Y given X as

$$\begin{aligned} f_1(y|x) &\equiv 1, \\ f_2(y|x) &\equiv 1 + a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right)\phi_X\left(\frac{x - x_0}{\delta} + 1/4\right). \end{aligned} \tag{28}$$

Figure B.1: Perturbation functions ϕ_Y and ϕ_X.



We let the marginal distribution $f_X(x)$ of X be the uniform distribution on $[0, 1]$. Note that $f_1(y|x)$, $f_2(y|x)$, $f_1(y, x) = f_1(y|x)f_X(x)$, and $f_2(y, x) = f_2(y|x)f_X(x)$ are non-negative everywhere, with integrals over their respective entire spaces all equaling to 1.

The first task is to verify that the two distributions are indeed in the class $\mathcal{F}_\kappa$. For $C^* \in (0, 2)$, the first distribution is in $\mathcal{F}$ by Lemma C.2 and the fact that Y is independent of X. Given any $x \in [0, 1]$, we can treat the whole term $a\phi_X((x - x_0)/\delta + 1/4)$ as the coefficient b in Lemma C.3. Then the results of Lemma C.3 applies since $|\phi_X| \leq 1$. In particular, the revenue function at x is twice-differentiable a.e., the absolute value of the second-order partial derivative with respect to y is bounded, and is also bounded from below by C^*. The optimal price is an interior solution and is in the interior of a region on which the revenue function is twice-differentiable. Lastly, the absolute value of the partial derivative of $f_2(y|x)$ with respect to x is bounded. This ensures that the quantity $|\frac{\partial}{\partial x}F_{Y|X}(y|x) + y\frac{\partial}{\partial x}f_{Y|X}(y|x)|$ is bounded.

Next, we want to derive the Hellinger distance between the two joint densities

$$
\begin{aligned}
&f_1(y, x) = 1, \\
&f_2(y, x) = 1 + a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right)\phi_X\left(\frac{x - x_0}{\delta} + 1/4\right).
\end{aligned}
$$

Let $\Psi(t) \equiv \sqrt{1 + t}$. Its second-order derivate is bounded when $|t| < 1/2$; that is,

$$
\sup_{|t|<1/2} |\Psi''(t)| < C.
$$

We use H to denote the Hellinger distance:

$$H(f_1\|f_2)^2 \equiv \int_0^1 \left(\sqrt{f_1(y)} - \sqrt{f_2(y)}\right)^2 dy.$$

The Hellinger distance can be bounded as

$$\begin{aligned}
H^2(f_1\|f_2)/2 &= 1 - \int_0^1\int_0^1 \Psi\left(a\delta\phi_Y\left(\frac{y-1/2}{\delta}\right)\phi_X\left(\frac{x-x_0}{\delta}+1/4\right)\right)dydx \\
&= \int_0^1\int_0^1 \Psi(0) - \Psi\left(a\delta\phi_Y\left(\frac{y-1/2}{\delta}\right)\phi_X\left(\frac{x-x_0}{\delta}+1/4\right)\right)dydx \\
&\leq -a\Psi'(0)\delta\int_0^1\int_0^1 \phi_Y\left(\frac{y-1/2}{\delta}\right)\phi_X\left(\frac{x-x_0}{\delta}+1/4\right)dydx \\
&\quad + a^2C\delta^2\int_0^1\int_0^1 \phi_Y^2\left(\frac{y-1/2}{\delta}\right)\phi_X^2\left(\frac{x-x_0}{\delta}+1/4\right)dydx,
\end{aligned}$$

where we have applied the second-order Taylor expansion to obtain the last inequality. By the change of variables $u = (y-1/2)/\delta$ and $v = (x-x_0)/\delta + 1/4$, for sufficiently small $\delta \in (0,\ 1/2]$,

$$\int_0^1\int_0^1 \phi_Y\left(\frac{y-1/2}{\delta}\right)\phi_X\left(\frac{x-x_0}{\delta}+1/4\right)dydx = \delta^2\int_{-1}^1 \phi_Y(u)\,du\int_0^1 \phi_X(v)\,dv = 0, \tag{29}$$

and

$$\int_0^1\int_0^1 \phi_Y^2\left(\frac{y-1/2}{\delta}\right)\phi_X^2\left(\frac{x-x_0}{\delta}+1/4\right)dydx = \delta^2\int_{-1}^1 \phi_Y^2(u)\,du\int_0^1 \phi_X^2(v)\,dv \leq C\delta^2.$$

Therefore, the Hellinger distance is bounded as

$$H^2(f_1\|f_2) \lesssim \delta^4.$$

Now we take δ such that $\delta^4 \asymp 1/n$. Note that (29) holds when $\delta > 0$ is small enough such that $\delta \in (0,\ 1/2]$, $1/4 - x_0/\delta \leq 0$ and $(1-x_0)/\delta + 1/4 \geq 1$; that is, when $x_0 n^{1/4} \geq c'$ and $(1-x_0)n^{1/4} \geq c''$ for positive universal constants c' and c'' (independent of n and x_0). This ensures that $H^2(f_1\|f_2) \lesssim 1/n$. Then from Lemma C.4, we know that

$$\inf_{\check{p}_D \in \check{\mathcal{D}}}\sup_{F_{Y,X}\in\mathcal{F}} \mathbb{E}_{F_{Y,X}}|\check{p}_D(x_0; data) - p_D^*(x_0)| \gtrsim n^{-1/4}, x_0 \in (0,1).$$

For bounding the revenue, recall that the revenue achieved at the price p and covariate value x_0 is

$r(p,x_0) = \max_p p(1 - F_{Y|X}(p|x_0))$. By Lemma C.1, we have

$$r(p_D^*(x_0)) - r(\check{p}_D(x_0;data)) \geq \frac{C^*}{2}|p_D^*(x_0) - \check{p}_D(x_0;data)|^2.$$

As a result, we have

$$\begin{aligned}
&\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}[r(p_D^*, x_0) - r(\check{p}_D(data), x_0)] \\
\geq &\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}\left[\frac{C^*}{2}|p_D^*(x_0) - \check{p}_D(x_0;data)|^2\right] \\
\geq &\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \frac{C^*}{2}\left\{\mathbb{E}\left[|p_D^*(x_0) - \check{p}_D(x_0;data)|\right]\right\}^2 \gtrsim n^{-1/2}.
\end{aligned}$$

This proves Theorem 3. □

Proof of Theorem 4. To prove Theorem 4, we follow the explanation in Section 4.3.2 and use the Fano's inequality to bound the probability of mistakes in the multiple classification problem. Before solving the revenue problem, we first study the lower bound for the L_2-distance of pricing functions. For two pricing functions p_1 and p_2, we define the (unweighted) L_2-distance as

$$\|p_1 - p_2\|_2 \equiv \left(\int_0^1 |p_1(x) - p_2(x)|^2 dx\right)^{1/2}.$$

In part (i), we defined the perturbation on the X dimension at a fixed point x_0. Now we want to define a large set of perturbed distributions. Each of these distributions is perturbed in a small interval on the X dimension. Let $m \geq 8$ be a large number (depending on n) that we specify later. Let $\alpha \in \{0,1\}^m$ be a vector of length m; that is,

$$\alpha \equiv (\alpha_1, \ldots, \alpha_m), \text{ where } \alpha_j \in \{0,1\}, j = 1, \ldots, m.$$

We construct a set of conditional density functions indexed by α:

$$f_{Y|X}^{\alpha}(y|x) \equiv 1 + \frac{a}{m}\sum_{j=1}^{m} \alpha_j \phi_Y(m(y - 1/2))\, \phi_X(mx - (j-1)).$$

The marginal distribution of X is taken to be the uniform distribution on $[0,1]$, that is, $f_X \equiv \mathbf{1}_{[0,1]}$. We denote the joint distribution by $f_{Y,X}^{\alpha} \equiv f_{Y|X}^{\alpha} f_X$.

We briefly describe this construction of the conditional density. The unit interval $[0,1]$ is divided equally into m subintervals:

$$I_j \equiv [(j-1)/m, j/m], j = 1, \ldots, m.$$

For $x \in I_j$, if $\alpha_j = 0$, then the conditional density is 1. If $\alpha_j = 1$, then the conditional density

$$f^{\alpha}_{Y|X}(y|x) \equiv 1 + \frac{a}{m}\phi_Y(m(y-1/2))\,\phi_X(mx-(j-1)), x \in I_j.$$

By treating $1/m$ as the scalar δ in part (i), we can see that, for m large enough, each $f^{\alpha}_{Y,X}$ belongs to the set $\mathcal{F}_\kappa$.

From the set $\{f^{\alpha}_{Y,X} : \alpha \in \{0,1\}^m\}$, we want to pick out a large enough subset of distributions whose optimal price functions are well-separated. For this purpose, we use the Gilbert-Varshamov bound (Lemma 2.9, Chapter 2 Tsybakov, 2009). The Gilbert-Varshamov bound states that for $m \geq 8$, there exists a subset $\mathcal{A} \subset \{0,1\}^m$ with cardinality $M \equiv |\mathcal{A}| \geq 2^{m/8}$, and the pairwise rescaled Hamming distance between elements in this set is greater than $1/8$. That is,

$$\frac{1}{m}\sum_{j=1}^{m} \mathbf{1}\{\alpha_j \neq \alpha'_j\} \geq \frac{1}{8}, \text{ for any } \alpha, \alpha' \in \mathcal{A}.$$

Applying the Gilbert-Varshamov bound, we can show that for $\alpha, \alpha' \in \mathcal{A}$, the optimal pricing functions of $f^{\alpha}_{Y,X}$ and $f^{\alpha'}_{Y,X}$ are well-separated. Let p_α be the pricing function associated with $f^{\alpha}_{Y,X}$; that is,

$$p_\alpha(x) \equiv \operatorname*{argmax}_{p\in[0,1]} p(1-F^{\alpha}_{Y|X}(p|x)),$$

where $F^{\alpha}_{Y|X}(y|x)$ is the corresponding conditional cumulative distribution function. Note that $\alpha, \alpha' \in \mathcal{A}$ differ in at least $m/8$ positions. This means that $f^{\alpha}_{Y|X}$ and $f^{\alpha'}_{Y|X}$ differ in $m/8$ intervals. Suppose that I_j is such an interval, where $\alpha_j = 0$ and $\alpha'_j = 1$. We restrict our attention to a subset of this interval:

$$\tilde{I}_j \equiv \left[\frac{1}{6m} + \frac{j-1}{m}, \frac{1}{3m} + \frac{j-1}{m}\right] \subset I_j.$$

When $x \in \tilde{I}_j$, we have

$$mx - (j-1) \in [1/6, 1/3] \implies \phi_X(mx-(j-1)) \in [\phi_X(0), \phi_X(1/2)]. \tag{30}$$

By Lemma C.3 (where $b = a\phi_X(mx-(j-1)) > 0$, $\delta = 1/m$), the choice $a \in (0, 4-2\kappa)$, and the fact (30), if we fix $x \in \tilde{I}_j$, then $p_\alpha(x) = 1/2$ while

$$p_{\alpha'}(x) \le 1/2 - \frac{c}{m}\phi_X\left(mx - (j-1)\right) \le 1/2 - \frac{c\phi_X(1/6)}{m}, x \in \tilde{I}_j,$$

where $c > 0$ is a universal constant that does not depend on n.[12] This implies that

$$|p_\alpha(x) - p_{\alpha'}(x)| \gtrsim \frac{1}{m}, x \in \tilde{I}_j.$$

Therefore, on the interval I_j, the separation between p_α and $p_{\alpha'}$ is lower bounded as

$$\int_{I_j} |p_\alpha(x) - p_{\alpha'}(x)|^2 dx \gtrsim \int_{\tilde{I}_j} 1/m^2 dx = \frac{1}{6m} \times \frac{1}{m^2} \gtrsim 1/m^3.$$

By the Gilbert-Varshamov bound, there are at least $m/8$ such intervals. Therefore, we can lower bound the total separation by

$$\|p_1 - p_2\|_2 \gtrsim \left(m/8 \times 1/m^3\right)^{1/2} \gtrsim 1/m.$$

Next, we want to compute the KL divergence between $f^{\alpha}_{Y,X}$ and $f^{\alpha'}_{Y,X}$. Note that the term $\phi_X\left(mx - (j-1)\right)$ is non-zero only when $x \in I_j$. The KL divergence can therefore be treated as a sum of m integrals:

$$\mathrm{KL}(f^{\alpha}_{Y,X} \| f^{\alpha'}_{Y,X}) = \int_0^1 \int_0^1 f^{\alpha}_{Y,X}(y,x) \log \frac{f^{\alpha}_{Y,X}}{f^{\alpha'}_{Y,X}} dy dx = \sum_{j=1}^{m} E_j,$$

where

$$\begin{aligned} E_j \equiv \int_{I_j} \int_0^1 & \left(1 + \frac{a}{m}\alpha_j \phi_Y\left(m(y-1/2)\right)\phi_X\left(mx-(j-1)\right)\right) \\ & \times \log \frac{1 + \frac{a}{m}\alpha_j \phi_Y\left(m(y-1/2)\right)\phi_X\left(mx-(j-1)\right)}{1 + \frac{a}{m}\alpha'_j \phi_Y\left(m(y-1/2)\right)\phi_X\left(mx-(j-1)\right)} dy dx. \end{aligned}$$

Notice that when $\alpha_j = \alpha_{j'}$, $E_j = 0$. Therefore, we only need to consider the j's where $\alpha_j \ne \alpha'_j$. Denote $\Psi_1(t) = -\log(1+t)$ and $\Psi_2(t) = (1+t)\log(1+t)$. Then we can write E_j as

$$E_j = \begin{cases} \int_{I_j} \int_0^1 \Psi_1\left(\frac{a}{m}\phi_Y\left(m(y-1/2)\right)\phi_X\left(mx-(j-1)\right)\right) dy dx, & \text{if } \alpha_j = 0, \alpha'_j = 1, \\ \int_{I_j} \int_0^1 \Psi_2\left(\frac{a}{m}\phi_Y\left(m(y-1/2)\right)\phi_X\left(mx-(j-1)\right)\right) dy dx, & \text{if } \alpha_j = 1, \alpha'_j = 0. \end{cases}$$

[12]For example, c can be equal to $a/8$ according to Lemma C.3.

By the second-order Taylor expansion at zero, we have

$$\Psi_1(t) = -t + \frac{1}{2(1+t')^2} t^2,$$

for some t' between 0 and t. When $|t| \leq 1/4$,[13] we have

$$\Psi_1(t) \leq -t + Ct^2,$$

for some universal constant $C > 0$. Similarly, we can show that

$$\Psi_2(t) \leq t + Ct^2.$$

Applying these inequalities to E_j, we have

$$\begin{aligned} E_j \leq & \pm \int_{I_j} \int_0^1 \frac{a}{m} \phi_Y\left(m(y-1/2)\right) \phi_X\left(mx-(j-1)\right) dydx \\ & + C \int_{I_j} \int_0^1 \frac{a^2}{m^2} \phi_Y^2\left(m(y-1/2)\right) \phi_X^2\left(mx-(j-1)\right) dydx. \end{aligned}$$

Similar to the derivation in Part (i), we know that the first term on the RHS is zero. For the second term, we can apply change of variables $u = m(y-1/2)$ and $v = mx-(j-1)$ and obtain that

$$\begin{aligned} & \int_{I_j} \int_0^1 \phi_Y^2\left(m(y-1/2)\right) \phi_X^2\left(mx-(j-1)\right) dydx \\ = & \frac{1}{m^2} \int_0^1 \phi_X^2(v)\, dv \int_{-1}^3 \phi_Y^2(u)\, du \leq \frac{C'}{m^2} \end{aligned}$$

for some universal constant $C' > 0$. Putting the results results together, we know that $E_j \leq \frac{C}{m^4}$ for all j. Since there are m intervals, we can bound the KL divergence by

$$\mathrm{KL}(f_{Y,X}^{\alpha} \| f_{Y,X}^{\alpha'}) = \sum_{j=1}^m E_j \lesssim \frac{1}{m^3}.$$

This is the KL distance for a single observation. For the entire data set with n i.i.d. observations, the KL divergence is upper bounded by Cn/m^3.

[13]Later we show that m is chosen to be $c_0 n^{1/4}$ where $c_0 > 0$ is a universal constant. As a result, $|t| \leq 1/4$ is guaranteed as long as c_0 is sufficiently large.

Lastly, we can summarize our results into the Fano inequality presented in Lemma C.5. We have

$$\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}\|\check{p}_D(data) - p_D^*\|_2^2 \geq \frac{C_1}{m^2}\left(1 - \frac{C_2 n/m^3 + \log 2}{\log 2^{m/8}}\right)$$
$$\geq \frac{C_1}{m^2}\left(1 - \frac{C_2 n/m^3 + \log 2}{C_3 m}\right).$$

By choosing $m = c_0 n^{1/4}$ for a sufficiently large universal constant $c_0 > 0$, we can make the factor $\left(1 - \frac{C_2 n/m^3 + \log 2}{C_3 m}\right)$ stay above, say, $1/2$. Then we have

$$\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}\|\check{p}_D(data) - p_D^*\|_2^2 \gtrsim \frac{1}{m^2} \asymp n^{-1/2}.$$

So far we have derived the lower bound for the L_2-distance of pricing. Moving onto the revenue problem, recall that the revenue achieved at the price p and covariate value x is $r(p,x) = \max_p p(1 - F_{Y|X}(p|x))$. By Lemma C.1, we have

$$r(p_D^*, x) - r(\check{p}_D(data), x) \geq \frac{C^*}{2}|p_D^*(x) - \check{p}_D(x; data)|^2.$$

Since f_X is bounded away from zero, we have

$$\begin{aligned}
&\inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}[R(p_D^*) - R(\check{p}_D)] \\
&= \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}\left[\int_0^1 (r(p_D^*, x) - r(\check{p}_D, x)) f_X(x) dx\right] \\
&\geq \inf_{\check{p}_D \in \check{\mathcal{D}}} \sup_{F_{Y,X} \in \mathcal{F}} \mathbb{E}\left[\frac{C^*}{2}\left(\inf_{x \in [0,1]} f_X(x)\right)\int_0^1 |p_D^*(x) - \check{p}_D(x; data)|^2 dx\right] \gtrsim n^{-1/2}.
\end{aligned}$$

□

Proof of Theorem 5. We use Lemma C.4 to prove the lower bound for Theorem 5. Define

$$\omega_U(\epsilon) \equiv \sup_{F_1, F_2 \in \mathcal{F}^U} \left\{|p_U^*(F_1) - p_U^*(F_2)| : H(F_1\|F_2) \leq \epsilon\right\}.$$

Then by Lemma C.4, we have

$$\inf_{\check{p}_U \in \check{\mathcal{U}}} \sup_{F_Y \in \mathcal{F}^U} \mathbb{E}_{F_Y}|\check{p}_U(data_Y) - p_U^*| \geq \frac{1}{8}\omega_U\left(1/(2\sqrt{n})\right).$$

Therefore, we only need to find a lower bound for ω_U. The proof proceeds in three steps. In the first step, we construct two distributions and compute the separation between their optimal prices.

The second step bounds the Hellinger distance between these two distributions. The third step summarizes.

Step 1. We construct two distribution functions. The first distribution is the uniform distribution on the unit interval $[0, 1]$. We denote this density function as

$$f_1(y) = 1_{[0,1]}(y).$$

The distribution function is $F_1(y) = y$ on the support $[0, 1]$. The revenue function under this distribution is $R_1(p) = p(1 - p)$. The optimal price is

$$p_1 = \operatorname*{argmax}_{p\in[0,1]} R_1(p) = \operatorname*{argmax}_{p\in[0,1]} p - p^2 = 1/2.$$

The second distribution function is a small twist of the uniform distribution. We use the same perturbation function ϕ_Y defined in (27).

We apply a small perturbation to the uniform density. Let $\delta > 0$ be a small number (that depends on n) specified later. Let $a \in (0, 4 - 2C^*)$. The formula of the density f_2 is given by

$$f_2(y) \equiv 1 + a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right) = \begin{cases} 1, & \text{if } y \in [0, 1/2 - \delta), \\ ay + 1 - \frac{a}{2} + a\delta, & \text{if } y \in [1/2 - \delta, 1/2), \\ -ay + 1 + \frac{a}{2} + a\delta, & \text{if } y \in [1/2, 1/2 + 2\delta), \\ ay + 1 - \frac{a}{2} - 3a\delta, & \text{if } y \in [1/2 + 2\delta, 1/2 + 3\delta), \\ 1, & \text{if } y \in [1/2 + 3\delta, 1]. \end{cases}$$

We compare the two densities f_1 and f_2 in the following graph.

Denote the optimal price under f_2 by p_2. By Lemma C.3(ii), we have

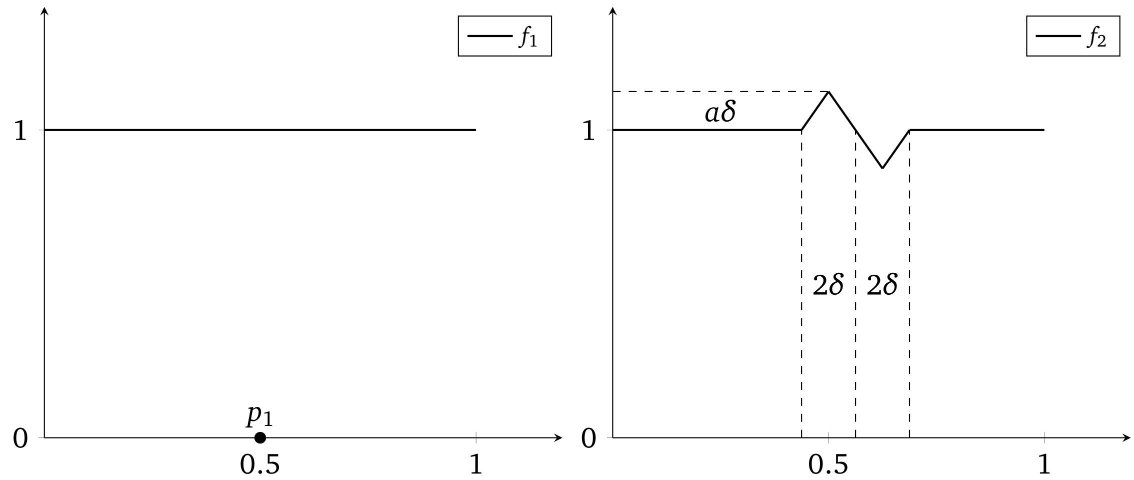
$$|p_2 - p_1| \geq a\delta/8$$

when δ is sufficiently small.

Step 2. We want to bound the Hellinger distance $H(F_1\|F_2)$. Define the function $\Psi(t) = \sqrt{1 + t}$. Its second-order derivative is bounded when $|t| < 1/2$; that is,

$$\sup_{|t|<1/2} |\Psi''(t)| \leq \frac{\sqrt{2}}{2}.$$

Figure B.2: Density functions f_1 and f_2.



Since $f_1(y) = 1$, we have

$$H(F_1 \| F_2)^2/2 = 1 - \int_0^1 \Psi\left(a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right)\right) dy$$
$$= \int_0^1 \Psi(0) - \Psi\left(a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right)\right) dy.$$

By the second-order Taylor expansion, we have

$$\Psi(0) - \Psi\left(a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right)\right)$$
$$\leq -\Psi'(0)a\delta\phi_Y\left(\frac{y - 1/2}{\delta}\right) + \frac{\sqrt{2}}{4}a^2\delta^2\phi_Y^2\left(\frac{y - 1/2}{\delta}\right).$$

By the construction of ϕ_Y, we have

$$\int_0^1 \phi_Y\left(\frac{y - 1/2}{\delta}\right) dy = 0.$$

By the change of variables $u = (y - 1/2)/\delta$, we have

$$\int_0^1 \phi_Y^2\left(\frac{y - 1/2}{\delta}\right) dy = \delta \int_{\mathbb{R}} \phi_Y^2(u)\, du \leq 4\delta \int_{-1}^0 (x + 1)^2 dx = \frac{4}{3}\delta.$$

Combining these results together, we obtain a bound on the Hellinger distance

$$H(F_1 \| F_2)^2 \leq \frac{2\sqrt{2}}{3}a^2\delta^3.$$

Step 3. By setting $\delta = c_0'(3/8\sqrt{2})^{1/3}a^{-2/3}n^{-1/3}$ for $c_0' \in (0, 1)$, we can ensure that $H(F_1 \| F_2) \leq$

$1/(2\sqrt{n})$. Previously, we assumed that $a\delta \le 1/2$ for the second-order Taylor expansion. This is true if c_0' is chosen to be sufficiently small. In this case, the separation between p_1 and p_2 is lower bounded as below:

$$|p_1 - p_2| \ge a\delta/8 = \frac{c_0'}{16}\left(\frac{3}{\sqrt{2}}\right)^{1/3}\left(\frac{a}{n}\right)^{1/3}.$$

By Lemma C.4, we have

$$\inf_{\check{p}_U \in \check{\mathcal{U}}} \sup_{F_Y \in \mathcal{F}^U} \mathbb{E}|\check{p}_U(data_Y) - p_U^*| \ge \frac{c_0'}{16}\left(\frac{3}{\sqrt{2}}\right)^{1/3}\left(\frac{a}{n}\right)^{1/3}.$$

Lastly, we want to lower bound the revenue. By Lemma C.1, we have

$$\begin{aligned}
\mathscr{R}_n^U(\mathcal{F}^U) &= \inf_{\check{p}_U \in \check{\mathcal{U}}} \sup_{F_Y \in \mathcal{F}^U} \mathbb{E}|R(\check{p}_U(data_Y), F_Y) - R(p_U^*, F_Y)| \\
&\ge \inf_{\check{p}_U \in \check{\mathcal{U}}} \sup_{F_Y \in \mathcal{F}^U} \mathbb{E}\left[\frac{C^*}{2}|\check{p}_U(data_Y) - p_U^*|^2\right] \\
&\ge \inf_{\check{p}_U \in \check{\mathcal{U}}} \sup_{F_Y \in \mathcal{F}^U} \frac{C^*}{2}\left\{\mathbb{E}\left[|\check{p}_U(data_Y) - p_U^*|\right]\right\}^2 \\
&\gtrsim \left(\frac{1}{n}\right)^{2/3}.
\end{aligned}$$

□

C Auxiliary Lemmas

Lemma C.1. *Let f be a function on $[0,1]$. Assume that f is differentiable and its derivative f' is Lipschitz continuous. Let z^* be a point in $[0,1]$ such that $f'(z^*) = 0$.*

(i) The derivative f' is a.e. differentiable on $[0,1]$.

(ii) Assume that there exists $\kappa_1 > 0$ such that $f''(z) \le -\kappa_1$ for almost all $z \in [0,1]$. Then, for any $z \in [0,1]$, we have

$$|f(z) - f(z^*)| \ge \frac{\kappa_1}{2}(z - z^*)^2.$$

(iii) Assume that there exists $\kappa_2 > 0$ such that $|f''(z)| \le \kappa_2$ for almost all $z \in [0,1]$. Then, for any

$z \in [0,1]$, *we have*

$$|f(z) - f(z^*)| \leq \frac{\kappa_2}{2}(z - z^*)^2.$$

Proof of Lemma C.1. For part (i), notice that a Lipschitz continuous function is absolutely continuous. By Theorem 3.35 in Chapter 3 of Folland (1999), we know that f' is differentiable a.e. with

$$f'(z_1) - f'(z_2) = \int_{z_2}^{z_1} f''(z)dz.$$

For part (ii), we can apply the fundamental theorem of calculus twice and obtain that

$$\begin{aligned} f(z) - f(z^*) &= \int_{z^*}^{z} f'(\tilde{z})d\tilde{z} \\ &= \int_{z^*}^{z} (f'(z_1) - f'(z^*))dz_1 \\ &= \int_{z^*}^{z} \int_{z^*}^{z_1} f''(z_2)dz_2dz_1 \\ &\leq -\kappa_1 \int_{z^*}^{z} \int_{z^*}^{z_1} dz_2dz_1, \end{aligned}$$

where in the second line we have used the assumption that $f'(z^*) = 0$, and in the last line we have used the assumption that $f''(z) \leq -\kappa_1$ for almost all $z \in [0,1]$. The double integral in the last line is equal to

$$\int_{z^*}^{z} \int_{z^*}^{z_1} dz_2dz_1 = \int_{z^*}^{z} (z_1 - z^*)dz_1 = \frac{(z - z^*)^2}{2}.$$

Therefore, we have

$$|f(z) - f(z^*)| \geq \frac{\kappa_1}{2}(z - z^*)^2.$$

Part (iii) can be proved analogously. □

Lemma C.2. *For the uniform distribution on* $[0,1]$*, the revenue function* $R(y) = y(1-y)$*. The revenue function is twice-differentiable with second-order derivative* $R''(y) = -2, y \in [0,1]$*. The optimal price is* $1/2$*.*

Proof of Lemma C.2. The proof is straightforward. □

Lemma C.3. *Recall the perturbation function ϕ_Y defined in (27). Consider the following density function*

$$f(y) \equiv 1 + b\delta\phi_Y\left(\frac{y-1/2}{\delta}\right) = \begin{cases} 1, & \textit{if } y \in [0, 1/2-\delta), \\ by + 1 - \frac{b}{2} + b\delta, & \textit{if } y \in [1/2-\delta, 1/2), \\ -by + 1 + \frac{b}{2} + b\delta, & \textit{if } y \in [1/2, 1/2+2\delta), \\ by + 1 - \frac{b}{2} - 3b\delta, & \textit{if } y \in [1/2+2\delta, 1/2+3\delta), \\ 1, & \textit{if } y \in [1/2+3\delta, 1], \\ 0, & \textit{otherwise.} \end{cases}$$

Denote F as the corresponding cumulative distribution function, $R(y) \equiv y(1-F(y))$ the revenue function, and $p^* \equiv \operatorname{argmax}_{y\in[0,1]} R(y)$ the optimal price. If $C^* \in (0,2)$, $|b| < 4-2C^*$, and $\delta > 0$ is sufficiently small, then the following statements hold.

(i) The density f is Lipschitz continuous.

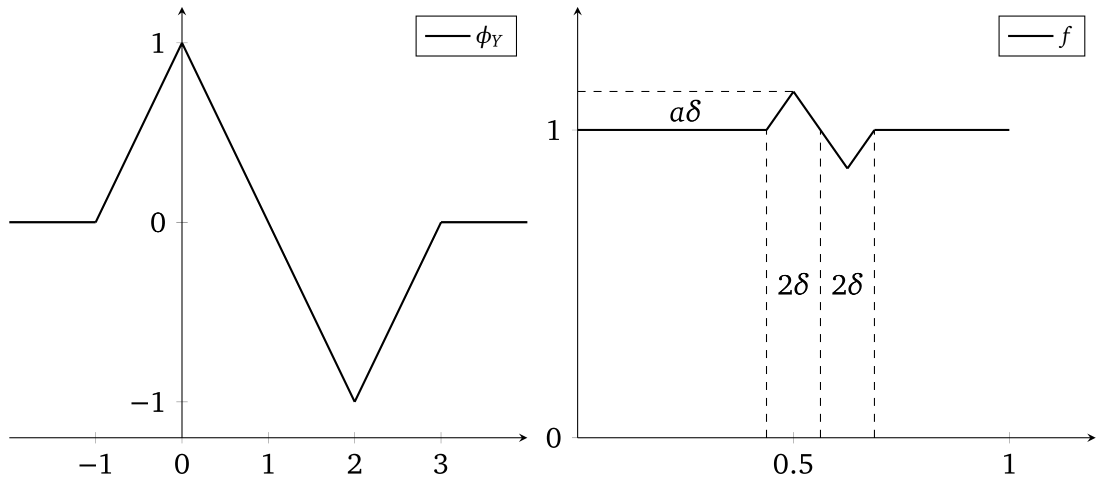
(ii) The revenue function is twice-differentiable a.e. The second-order derivative is bounded a.e. and satisfies that

$$-2f(y) - yf'(y) \geq -C^* \textit{ for almost all } y.$$

(iii) For $b > 0$, the optimal price $p^* \in (1/2-\delta, 1/2-b\delta/8)$. For $b < 0$, the optimal price $p^* \in (1/2-b\delta/8, 1/2+2\delta)$. For $b = 0$, the optimal price $p^* = 1/2$. In particular, p^* is always an interior solution, and f is always differentiable in a neighborhood of p^*.

Proof of Lemma C.3. For reference, we plot here the perturbation function ϕ_Y and the perturbed density f. Part (i) is straightforward. The density f is piecewise linear and hence Lipschitz continuous with Lipschitz constant b. To verify the strong concavity in part (ii), note that the corresponding revenue function R is continuously differentiable and twice-differentiable a.e. on the support $[0,1]$.

Figure C.1: Perturbation function and perturbed density.



Its second-order derivative

$$R''(y) = -2f(y) - yf'(y) = \begin{cases} -2, & \text{if } y \in [0, 1/2 - \delta], \\ -3by - 2 + b - 2b\delta, & \text{if } y \in [1/2 - \delta, 1/2], \\ 3by - 2 - b - 2b\delta, & \text{if } y \in [1/2, 1/2 + 2\delta], \\ -3by - 2 + b + 6b\delta, & \text{if } y \in [1/2 + 2\delta, 1/2 + 3\delta], \\ -2, & \text{if } y \in [1/2 + 3\delta, 1]. \end{cases}$$

We can see that R'' is piecewise linear and hence bounded a.e. We further show that R'' is bounded away from zero by κ. On the intervals $[0, 1/2 - \delta]$ and $[1/2 + 3\delta, 1]$, we have $R''(y) = -2 < -C^*$. We check the remaining three intervals one by one. On the interval $[1/2 - \delta, 1/2]$, the condition $|b| < 4 - 2C^*$ ensures that

$$b \geq 0 \implies R''(y) \leq R''(1/2 - \delta) = -b/2 - 2 + b\delta \leq -C^*,$$
$$b < 0 \implies R''(y) \leq R''(1/2) = -b/2 - 2 - 2b\delta \leq -C^*,$$

when δ is sufficiently small. On the interval $[1/2, 1/2 + 2\delta]$, we have

$$b \geq 0 \implies R''(y) \leq R''(1/2 + 2\delta) = b/2 - 2 + 4b\delta \leq -C^*,$$
$$b < 0 \implies R''(y) \leq R''(1/2) = b/2 - 2 - 2b\delta \leq -C^*,$$

when δ is sufficiently small. On the interval $[1/2+2\delta, 1/2+3\delta]$, we have

$$\begin{aligned} b \geq 0 &\implies R''(y) \leq R''(1/2+2\delta) = -b/2-2 < -C^*, \\ b < 0 &\implies R''(y) \leq R''(1/2+3\delta) = -b/2-2-3b\delta < -C^*, \end{aligned}$$

To summarize, we have shown that $R''(y) \leq -C^*$ a.e. on $[0,1]$ provided that $\delta > 0$ is sufficiently small.

For part (iii), we first consider the case $b > 0$. We only need to consider the interval $[1/2-\delta, 1/2]$. The reason will become clear later. The cumulative distribution function

$$F(y) = \frac{b}{2}y^2 + \left(1 - \frac{b}{2} + b\delta\right)\delta y + \frac{b}{2}\left(1/2-\delta\right)^2, y \in [1/2-\delta, 1/2].$$

The revenue function

$$R(y) = -\frac{b}{2}y^3 - \left(1 - \frac{b}{2} + b\delta\right)y^2 + \left(1 - \frac{b}{2}\left(1/2-\delta\right)^2\right)y, y \in [1/2-\delta, 1/2].$$

The marginal revenue

$$R'(y) = -\frac{3b}{2}y^2 - \left(2-b+2b\delta\right)y + 1 - \frac{b}{2}\left(1/2-\delta\right)^2, y \in [1/2-\delta, 1/2].$$

We evaluate the marginal revenue at two points $1/2-\delta$ and $1/2-\frac{b\delta}{8}$. When $y = 1/2-\delta$, the marginal revenue

$$R'\left(1/2-\delta\right) = \delta > 0.$$

When $y = 1/2 - b\delta/8$, the marginal revenue

$$R'\left(1/2 - \frac{b\delta}{8}\right) \approx \frac{b(b-4)}{16}\delta < 0,$$

where we have omitted higher order terms involving δ^2. Therefore, $R'\left(1/2-\frac{b\delta}{8}\right)$ is negative for sufficiently small δ. Since the marginal revenue R' is strictly decreasing on the entire domain $[0,1]$, we know that the only zero of R' (which is the optimal price p^*) is within the region $(1/2-\delta, 1/2-\frac{b\delta}{8})$. Within this region, the revenue is twice-differentiable everywhere.

Next, we consider the case $b < 0$. In this case, we only need to study the region $[1/2, 1/2+2\delta]$.

The cumulative distribution function

$$F(y) = -\frac{b}{2}y^2 + \left(1 + \frac{b}{2} + b\delta\right) y + \frac{b}{2}\delta^2 - \frac{b}{2}\delta - \frac{b}{8}, y \in [1/2, 1/2 + 2\delta] \,.$$

The revenue function

$$R(y) = y(1 - F(y)) = \frac{b}{2}y^3 - \left(1 + \frac{b}{2} + b\delta\right) y^2 + \left(1 + \frac{b}{8} - \frac{b}{2}\delta^2 + \frac{b}{2}\delta\right) y, y \in [1/2, 1/2 + 2\delta] \,.$$

The marginal revenue

$$R'(y) = \frac{3b}{2}y^2 - (2 + b + 2b\delta)y + \left(1 + \frac{b}{8} - \frac{b}{2}\delta^2 + \frac{b}{2}\delta\right), y \in [1/2, 1/2 + 2\delta] \,.$$

We evaluate the marginal revenue at two points $1/2 + \delta$ and $1/2 - \frac{b\delta}{8}$. When $y = 1/2 + \delta$, the marginal revenue

$$R'\left(1/2 + \delta\right) \approx -2\delta < 0,$$

where we have omitted higher order terms involving δ^2. When $y = 1/2 - b\delta/8$, the marginal revenue

$$R'\left(1/2 - \frac{b\delta}{8}\right) \approx \frac{b(b+4)}{16}\delta > 0,$$

where we have omitted higher order terms involving δ^2. Since the marginal revenue R' is strictly decreasing on the entire domain $[0, 1]$, we know that the only zero of R' (which is the optimal price p^*) is within the region $(1/2 - \frac{b\delta}{8}, 1/2 + \delta)$. Within this region, the revenue is twice-differentiable everywhere.

Lastly, when $b = 0$, the density function is constant, and Lemma C.2 shows that the optimal price is $1/2$. Therefore, regardless of the sign of b, the optimal price is always an interior solution, and is in the interior of a region on which the revenue function is twice-differentiable. □

Lemma C.4. *Take $x_0 \in [0, 1]$. Recall the following definition of $\omega_D(\epsilon)$ and $\omega_U(\epsilon)$:*

$$\omega_D(\epsilon) \equiv \sup_{F_1, F_2 \in \mathcal{F}} \left\{ |p_D^*(x_0; F_1) - p_D^*(x_0; F_2)| : H(F_1 \| F_2) \leq \epsilon \right\},$$

$$\omega_U(\epsilon) \equiv \sup_{F_1, F_2 \in \mathcal{F}^U} \left\{ |p_U^*(F_1) - p_U^*(F_2)| : H(F_1 \| F_2) \leq \epsilon \right\}.$$

Then

$$\inf_{\check{p}_D\in\check{\mathcal{D}}}\sup_{F_{Y,X}\in\mathcal{F}}\mathbb{E}_{F_{Y,X}}|\check{p}_D(x_0;data)-p_D^*(x_0;F_{Y,X})|\geq\frac{1}{8}\omega_D\left(1/(2\sqrt{n})\right),$$
$$\inf_{\check{p}_U\in\check{\mathcal{U}}}\sup_{F_Y\in\mathcal{F}^U}\mathbb{E}_{F_Y}|\check{p}_U(data_Y)-p_U^*(F_Y)|\geq\frac{1}{8}\omega_U\left(1/(2\sqrt{n})\right).$$

Proof of Lemma C.4. By treating $p_D^*(x_0;\cdot)$ and $p_U^*(\cdot)$ as functionals, the desired results directly follow from Corollary 15.6 (Le Cam for functionals) in Chapter 15 of Wainwright (2019). □

Lemma C.5. *Let $\{F_{Y,X}^j: 1\leq j\leq M\}\subset\mathcal{F}$ be such that*

$$\|p_D^*(F_{Y,X}^j)-p_D^*(F_{Y,X}^{j})\|_2\geq 2\delta,\, j\neq j'.$$

Then we have

$$\inf_{\check{p}_D\in\check{\mathcal{D}}}\sup_{F_{Y,X}\in\mathcal{F}}\mathbb{E}\|\check{p}_D(data)-p_D^*(F_{Y,X})\|_2^2\geq\delta^2\left(1-\frac{\sum_{j,j'=1}^M KL(F_{Y,X}^j\|F_{Y,X}^{j'})/M^2+\log 2}{\log M}\right)$$

Proof of Lemma C.5. The result follows from Proposition 15.12 (the Fano's inequality) and inequality (15.34) (convexity of the KL divergence) in Chapter 15 of Wainwright (2019), where Φ is taken to be the square function, ρ the L_2-distance, and θ the functional p_D^*. □

Lemma C.6. *Consider the following function class:*

$$\{(y,x)\mapsto(p\mathbf{1}\{y\geq p\}-\tilde{p}\mathbf{1}\{y\geq\tilde{p}\})\mathbf{1}\{x\in[k/K,(k+1)/K)\}: p\in[0,1]\}.$$

For any $\tilde{p}\in[0,1]$, $K\geq 1$, and $0\leq k\leq K-1$, the above class is a VC-subgraph with VC-dimension no greater than 2*.*

Proof of Lemma C.6. By Lemma 2.6.22 in Chapter 2 of van der Vaart and Wellner (1996), the class

$$\{(y,x)\mapsto p\mathbf{1}\{y\geq p\}: p\in[0,1]\}$$

is a VC-subgraph with VC-dimention no greater than 2.[14] The function $(y,x)\mapsto\tilde{p}\mathbf{1}\{y\geq\tilde{p}\}$ is a fixed function that does not depend on the index p. By the proof Lemma 2.6.18(v) in van der Vaart

[14]In the original statement of the lemma, the VC dimension is no greater than 3. This is because the definition of VC dimension in van der Vaart and Wellner (1996) is the smallest number n for which no set of n points is shattered. The definition we use in this paper is the largest number n that some set of n points is shattered.

and Wellner (1996), the class

$$\{(y,x)\mapsto p\mathbf{1}\{y\geq p\}-\tilde{p}\mathbf{1}\{y\geq\tilde{p}\}:\,p\in[0,1]\}$$

is a VC-subgraph with VC-dimention no greater than 2. Lastly, we multiply each function in the class by an indicator $\mathbf{1}\{x\in[k/K,(k+1)/K)\}$. This does not increase the VC-dimension. □

Lemma C.7. *Let $Z_1,\ldots,Z_n$ be an i.i.d. sequence of random variables from distribution P. Let $\mathcal{G}$ be a class of VC-subgraph functions with VC-dimension v and envelope function G. Assume that $\|G\|_{L_2(P)}<\infty$. Then we have*

$$\mathbb{E}\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}g(Z_i)-\mathbb{E}g(Z_i)\right|\leq 8\sqrt{2}\frac{\|G\|_{L_2(P)}}{\sqrt{n}}\left(\log(2C)+\log(v)+(\log(16)+3)v\right),$$

for some universal constant C, where the $L_2(P)$ norm $\|f-g\|_{L_2(P)}\equiv\left(\int_{\mathcal{X}}[f(x)-g(x)]^2\,\mathbb{P}(dx)\right)^{\frac{1}{2}}$.

Proof of Lemma C.7. This is a well-known result in the literature. We include it here for completeness. Let $N(\mathcal{G},L_2(Q),\tau)$ denote the covering number of $(\mathcal{G},L_2(Q))$. By Remark 3.5.5 in Chapter 3 of Giné and Nickl (2015), we know that

$$\mathbb{E}\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i)-\mathbb{E}g(X_i)\right|\leq 8\sqrt{2}\frac{\|G\|_{L_2(P)}}{\sqrt{n}}\int_0^1\sup_Q\sqrt{\log 2N(\mathcal{G},L_2(Q),\tau\|G\|_{L_2(Q)})}d\tau,$$

where the supremum is taken over all discrete probabilities with a finite number of atoms. By Theorem 2.6.7 in Chapter 2 of van der Vaart and Wellner (1996), we know that for any probability measure Q,

$$N(\mathcal{G},L_2(Q),\tau\|G\|_{L_2(Q)})\leq Cv(16e)^v(1/\tau)^{2v},$$

for some universal constant C. Therefore,

$$\int_0^1\sup_Q\sqrt{\log 2N(\mathcal{G},L_2(Q),\tau\|G\|_{L_2(Q)})}d\tau\leq\log(2C)+\log(v)+(\log(16)+3)v$$

Then the desired result follows. □

References

Alexander, Kenneth S, "Rates of growth and sample moduli for weighted empirical processes indexed by sets," *Probability Theory and Related Fields*, 1987, *75* (3), 379–423.

Allouah, Amine, Achraf Bahamou, and Omar Besbes, "Pricing with samples," *Operations Research*, 2022, *70* (2), 1088–1104.

— , — , **and** — , "Optimal pricing with a single point," *Management Science*, 2023.

Babaioff, Moshe, Yannai A Gonczarowski, Yishay Mansour, and Shay Moran, "Are two (samples) really better than one?," in "Proceedings of the 2018 ACM Conference on Economics and Computation" 2018, pp. 175–175.

Balcan, Maria-Florina, Amit Daniely, Ruta Mehta, Ruth Urner, and Vijay V Vazirani, "Learning economic parameters from revealed preferences," in "Web and Internet Economics: 10th International Conference, WINE 2014, Beijing, China, December 14-17, 2014. Proceedings 10" Springer 2014, pp. 338–353.

Baliga, Sandeep and Rakesh Vohra, "Market research and market design," *Advances in theoretical Economics*, 2003, *3* (1).

Basu, Pathikrit and Federico Echenique, "On the falsifiability and learnability of decision theories," *Theoretical Economics*, 2020, *15* (4), 1279–1305.

Beigman, Eyal and Rakesh Vohra, "Learning from revealed preference," in "Proceedings of the 7th ACM Conference on Electronic Commerce" 2006, pp. 36–42.

Bergemann, Dirk and Karl H. Schlag, "Pricing without Priors," *Journal of the European Economic Association*, 2008, *6* (2-3), 560–569.

— **and** — , "Robust Monopoly Pricing," *Journal of Economic Theory*, 2011, *146* (6), 2527–2543.

Bousquet, Olivier, "Concentration inequalities for sub-additive functions using the entropy method," in "Stochastic inequalities and applications," Springer, 2003, pp. 213–247.

Carroll, Gabriel, "Robustness in Mechanism Design and Contracting," *Annual Review of Economics*, 2019, *11* (1), 139–166.

Chambers, Christopher P, Federico Echenique, and Nicolas S Lambert, "Recovering preferences from finite data," *Econometrica*, 2021, *89* (4), 1633–1664.

— , — , **and** — , "Recovering utility," *arXiv preprint arXiv:2301.11492*, 2023.

Cole, Richard and Tim Roughgarden, "The sample complexity of revenue maximization," in "Proceedings of the forty-sixth annual ACM symposium on Theory of computing" 2014, pp. 243–252.

Cover, Thomas M. and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Ltd, 2005.

Crémer, Jacques and Richard P McLean, "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica: Journal of the Econometric Society*, 1988, pp. 1247–1257.

Devanur, Nikhil R, Zhiyi Huang, and Christos-Alexandros Psomas, "The sample complexity of auctions with side information," in "Proceedings of the forty-eighth annual ACM symposium on Theory of Computing" 2016, pp. 426–439.

Dhangwatnotai, Peerapong, Tim Roughgarden, and Qiqi Yan, "Revenue maximization with a single sample," *Games and Economic Behavior*, 2015, *91*, 318–333.

Folland, Gerald B, *Real analysis: modern techniques and their applications*, 2 ed., Vol. 40, New York, NY: John Wiley & Sons, 1999.

Fu, Hu, Nicole Immorlica, Brendan Lucier, and Philipp Strack, "Randomization beats second price as a prior-independent auction," in "Proceedings of the sixteenth ACM conference on economics and computation" 2015, pp. 323–323.

__ , Nima Haghpanah, Jason Hartline, and Robert Kleinberg, "Full surplus extraction from samples," *Journal of Economic Theory*, 2021, *193*, 105230.

Giné, Evarist and Richard Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2015.

Goldberg, Andrew V, Jason D Hartline, Anna R Karlin, Michael Saks, and Andrew Wright, "Competitive auctions," *Games and Economic Behavior*, 2006, *55* (2), 242–269.

Gonçalves, Duarte and Bruno A. Furtado, "Statistical Mechanism Design: Robust Pricing, Estimation, and Inference," *arXiv preprint arXiv:2405.17178*, 2024.

Guo, Chenghao, Zhiyi Huang, and Xinzhi Zhang, "Settling the sample complexity of single-parameter revenue maximization," in "Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing" 2019, pp. 662–673.

Hartmann, Wesley, Harikesh S Nair, and Sridhar Narayanan, "Identifying causal marketing mix effects using a regression discontinuity design," *Marketing Science*, 2011, *30* (6), 1079–1097.

Hirano, Keisuke and Jack R Porter, "Asymptotic efficiency in parametric structural models with parameter-dependent support," *Econometrica*, 2003, *71* (5), 1307–1338.

Huang, Zhiyi, Yishay Mansour, and Tim Roughgarden, "Making the most of your samples," *SIAM Journal on Computing*, 2018, *47* (3), 651–674.

Jank, Wolfgang and Galit Shmueli, *Modeling online auctions*, John Wiley & Sons, 2010.

Kolmogorov, Andrei Nikolaevich and Vladimir Mikhailovich Tikhomirov, "ε-entropy and ε-capacity of sets in function spaces," *Uspekhi Matematicheskikh Nauk*, 1959, *14* (2), 3–86.

Myerson, Roger B, "Optimal auction design," *Mathematics of operations research*, 1981, *6* (1), 58–73.

Segal, Ilya, "Optimal pricing mechanisms with unknown demand," *American Economic Review*, 2003, *93* (3), 509–529.

Talagrand, Michel, "New concentration inequalities in product spaces," *Inventiones mathematicae*, 1996, *126* (3), 505–563.

Tsybakov, Alexandre B., *Introduction to Nonparametric Estimation* Springer Series in Statistics, 1 ed., Springer, New York, NY, 2009.

van de Geer, Sara A, *Empirical Processes in M-estimation*, Vol. 6, Cambridge university press, 2000.

van der Vaart, Aad W and Jon A Wellner, *Weak Convergence and Empirical Processes*, New York, NY: Springer, 1996.

Wainwright, Martin J, *High-dimensional statistics: A non-asymptotic viewpoint*, Vol. 48, Cambridge University Press, 2019.

Zhang, Hanrui and Vincent Conitzer, "Learning the Valuations of a k-demand Agent," in "International Conference on Machine Learning" PMLR 2020, pp. 11066–11075.