An Empirical Analysis of Optimal Nonlinear Pricing*

Soheil Ghili†, Youngkyun "Russ" Yoon‡

February 7, 2023

Abstract

In "continuous choice" settings, consumers decide not only on whether to purchase a product, but also on how much to purchase. As a result, firms should optimize a full price schedule rather than a single price point. This paper provides a methodology to empirically estimate the optimal schedule under multi-dimensional consumer heterogeneity. We apply our method to novel data from an educational-services firm that contains purchase-size information not only for deals that materialized, but also for potential deals that eventually failed. We show that the optimal second-degree price discrimination (i.e., optimal nonlinear tariff) improves the firm's profit upon linear pricing by about 7.9%. That said, this second-degree price discrimination scheme only recovers 7.4% of the gap between the profitability of linear pricing (i.e., no price discrimination) and that of infeasible first degree price discrimination. We also conduct several further counterfactual analyses (i) comparing the role of demand- v.s. cost-side factors in shaping the optimal price schedule, (ii) examining third-degree price discrimination, and (iii) empirically quantifying the magnitude by which incentive-compatibility constraints impact the optimal pricing and profits.

*We thank Saman Ghili for his advice on optimization methods and literature. We also thank Yufeng Huang, Yewon Kim, Tesary Lin, Wenting Yu, and various conference and seminar participants for their helpful comments. Ghili acknowledges financial support from the Yale Center for Customer Insights. We thank Wanxi Zhou for outstanding research assistance. All errors are our own.

†Yale University. Email: soheil.ghili@yale.edu

‡Massachusetts Institute of Technology. Email: ykyoon@mit.edu

1 Introduction

"Continuous choice" products are increasingly common. For such products (unlike in discrete choice environments,) each consumer decides not only on whether to purchase, but also on how much. Traditional B2C examples of such products are cell-phone plans and utility. More recently, many B2B products (e.g., cloud services, SaaS products, etc) are of this form. In continuous choice settings, the firm's pricing problem goes beyond finding an optimal price point; it entails optimizing a full pricing schedule.

The objective of this paper is to provide an empirical framework for optimizing a nonlinear price schedule. In developing this framework, we contribute to the literature on multiple fronts: (i) the data, (ii) the model, and (iii) the analyses run and lessons learned.

We leverage a novel dataset from a firm whose main product is leadership workshops for employees of companies. Our data records not only deals that succeeded, but also sale efforts that started but did not lead to a transaction. Crucially, for unsuccessful sale efforts, the data records, among other things, the *would-be number of workshops that the potential customer was considering buying.* It is worth noting that even though we are not aware of academic papers that use a dataset comprehensively recording detailed features of unsuccessful deals, such datasets are increasingly collected and maintained by firms. This gives rise to a critical question: how can one leverage such data for optimal nonlinear pricing using a model that (i) is relatively simple to empirically implement and, at the same time, (ii) flexibly captures some of the key economic forces that the mechanism design literature finds relevant to the design of optimal nonlinear contracts? We expect such a model to be directly applicable by firms that seek to optimize their pricing schedules using structural econometric tools.

Our model is one of continuous-choice demand with multi-dimensional heterogeneity across customers. It is built to capture a key insight from the economic theory literature on multi-dimensional screening. More specifically, we develop a demand model that allows for flexibility in the joint distribution of a customer's "size" (i.e., how many units of the product the customer needs) and her price sensitivity. The correlation across customers between these two quantities (or more broadly, the joint distribution of them) is critical for determining the shape of the optimal nonlinear pricing scheme. To illustrate, if "larger" customers are more price sensitive, then flatter tariffs are more profitable whereas if "smaller" ones have a higher sensitivity to price, then steeper schedules are recommended. This is closely related to results from recent multi-dimensional-screening/2nd-dgree-disrimination literature on how the optimal tariff is impacted by the shape of the joint distribution between price sensitivity and taste/need for quality/quantity (Anderson and Dana Jr, 2009; Haghpanah and Hartline, 2021; Ghili, 2022; Yang, 2021). Nevertheless, we are not aware of empirical work that directly models customer size in the context of optimal nonlinear pricing.

We estimate this model on our dataset. The key object of interest in estimation is the joint distribution over customer size and price sensitivity. What makes it possible to flexibly recover this distribution is, roughly, the variation across deal sizes in percent acceptance rate. To informally illustrate, if a significantly higher fraction of larger deals fail relative to smaller deals, then large customers are on average more price sensitive than smaller customers. More complex joint distributions may also be recovered if, for instance, medium-sized deals are more likely to be unsuccessful compared to both larger and smaller deals. The key feature of our data (i.e., observing the intended sizes of eventually unsuccessful deals) gives us access to this necessary variation in deal success rates across sizes, thereby enabling us

to estimate the joint distribution of interest. We use our estimation and price-schedule optimization methodology to simulate the optimal nonlinear schedule (i.e., optimal 2nd degree price discrimination) and assess its effects on profit and welfare. We find that the optimal schedule lowers the prices for larger deal sizes, relative to the observed prices charged by the firm. The cost-side reason for this is that a portion of production costs is incurred per customer rather than per unit. The demand side reasons are more complex and will be discussed in detail later in the paper; but a simplified intuition is that a higher portion of larger deals fail relative to smaller ones, even though the observed pricing strategy of the firm already involved some volume discounting. To give a sense of the magnitude of our results, we find that it is best to offer about 18% lower per-unit fees to customers purchasing more than 100 units relative to those who purchase less than five. The discount is around 9% for deals of sizes 10-19 relative to those less than five. The optimal nonlinear pricing schedule delivers a 7.9% higher profit relative to optimal linear pricing. In addition, we find that the optimal second-degree schedule covers 7.4% of the profitability-gap between optimal linear pricing and optimal first-degree price discrimination. This stands in contrast to third-degree price discrimination which has been shown to closely approximate the profitability of first degree discrimination (see Dubé and Misra (2017)). Also, we find that the optimal second degree price discrimination leads to an almost 1% increase in consumer welfare relative to linear pricing (consumer welfare for first-degree discrimination is, by construction, zero). In addition, nonlinear pricing (i.e., second degree price discrimination) increases total social welfare relative to linear pricing by about 5.5%.

Finally, we conduct a several additional counterfactual analyses that both produce useful insights and demonstrate the ability of our framework to generate similar insights in other settings: (i) We empirically compare the role of cost side factors to those of demand side ones in determining the shape of the optimal price schedule. We show that the nonlinearities in the optimal schedule are more heavily shaped by the heterogeneity in demand than by nonlinear cost function. We also document the interaction between the two factors: the presence of production costs moderates the effect of demand heterogeneity on the optimal contract. (ii) We study the role of "incentive compatibility constraints". More specifically, we quantify the profit from a "local" pricing schedule in which the per-unit price for each size-range is optimized only considering the per-unit willingnesses to pay by customers belonging that same size group. Under counterfactual demand scenarios in which average willingness to pay per unit varies substantially across customers of different sizes sizes, we show that such "locally optimal" price schedules lead to significant losses. This is because customers belonging to a size-group that faces a high marginal price are incentivized to adjust their purchase sizes to take advantage of substantially lower per-unit prices offered to other size groups. We show that the "globally optimal" contract tends to moderate the variation of the per-unit prices across sizes in order to mitigate the loss arising from such incentives by customers. (iii) We examine the effectiveness of 3rd-degree price discrimination, as well as that of combining 2nd and 3rd degree discrimination approaches with each other. We only find small to moderate profit improvement from introducing 3rd degree discrimination in our context.

The rest of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data and setting and provides some summary statistics. Section 4 presents the model. Section 5 discusses the estimation procedure, identification, and estimation results. Section 6 presents the optimal nonlinear price schedule and discusses in detail how it compares to a linear schedule and to first-degree price discrimination. Section 7 conducts further counterfactual analysis. 8 discusses the general applicability of our method beyond our specific setting, as well as some caveats and avenues for future research. Section 9 concludes.

2 Related Literature

This paper sits at the intersection of two strands of the literature on theory and empirics of nonlinear pricing. On the theoretical end, there exists a large literature on "screening," with one of its major applications/interpretations being nonlinear pricing. Many papers in this domain, such as seminal work by Mussa and Rosen (1978) and Maskin and Riley (1984), focus on uni-dimensional consumer type spaces. The literature on multi-dimensional screening (e.g., Wilson (1993); Armstrong (1996); Laffont et al. (1987); Rochet and Stole (2002); Carroll (2017)) relaxes this assumption. The present paper relates to this literature by being one of the first studies to directly analyze optimal screening and second-degree price discrimination in an empirical setting. We empirically optimize a full price schedule for a firm that, when deciding the charge amount for a given quantity/quality level q, tries not only to maximize the profit made from those customers who purchase at q, but also to incentivize as many customers as possible to purchase more profitable quantities q'.

Though it may not be clear in the first glance, our emphasis on correlation between size and value makes this paper related to the literature on bundling of products with non-additive values. In that literature, there are results stating that bundling a set of set of products is recommended if consumers who see a higher willingness-to-pay consumers see a lower degree of complementarity among products (see Anderson and Dana Jr (2009); Haghpanah and Hartline (2021); Ghili (2022) among others). This has parallels to our intuition that flatter contracts are optimal when higher value customers tend to be the smaller ones.

On the empirical side, there have been a number of studies on nonlinear tariffs. Most commonly studied applications have been electricity and cell-phone plans. Some of these papers examine a one-dimensional type space (Luo et al., 2018) whereas others study mutli-dimensional type spaces (e.g., Nevo et al. (2016); Reiss and White (2005)). We have three points of differentiation relative to these studies. First, we are explicitly after capturing "customer size" as one dimension along which there can be heterogeneity. The literature does not directly study this. The closest paper to us on this front is Reiss and White (2005) which studies nonlinear pricing in the electricity market and examines heterogeneity across consumers in the number and types of electric home-appliances they own. Second, and relatedly, we bring novel data on the would-be sizes of unsuccessful deals which plays a key role in being able to estimate the model. Finally, this literature tends to take the regulator's perspective and study how a number of various nonlinear pricing schemes fare against each other in terms of consumer welfare (see Nevo et al. (2016) for instance).[1] In contrast, this paper is one of the few empirical studies of nonlinear pricing to take the firm's perspective and provide a method to optimize the nonlinear tariff (other examples of firm-level studies are Narayanan et al. (2007); Iyengar et al. (2008); Iyengar and Jedidi (2012)).

Our work is also related to the empirical literature that quantifies the efficiency of price discrimination strategies. For example, Dubé and Misra (2017) quantify that a sufficiently fine-grained third-degree price discrimination strategy can replicate the profitability of first degree price discrimination in the context of their study. Our paper is complementary in that it quantifies the same effect for second degree price discrimination, and is one of the few to do so, especially in the context of nonlinear pricing (for other examples of empirical quantification of the effects of second-degree discrimination, see Hendel and Nevo (2013); Draganska and Jain (2006); Verboven (2002); Iyengar and Gupta (2009); Kadiyali et al. (1996);

[1] Some studies, such as Luo et al. (2018), even assume that the observed pricing scheme is optimal and use its optimality condition as a supply-side moment to back out marginal costs.

Leslie (2004). For a survey of this literature, see Chan et al. (2009)). We show that at least in our context, second degree price discrimination recovers only a small portion of the profitability gap between first degree price discrimination and no discrimination at all.

3 Data, Setting, and Descriptive Statistics

We study pricing by LifeLabs Learning, a New York based company offering workshops to employees of its business clients. Lifelabs serves customers within and outside of the United States. The workshops are on leadership and other business-related skills. As of 2021, LifeLabs does not directly approach potential customers and its marketing activities are mainly based on word of mouth. When a potential client reaches out to LifeLabs, a conversation about a potential deal begins. The time window of our data from the company encompasses 2020 and 2021. The most important aspects of each deal in our data are quantity (the number of workshops to be delivered by LifeLabs to the company) and the total price. The price is determined based on a pre-set schedule as a function of quantity.

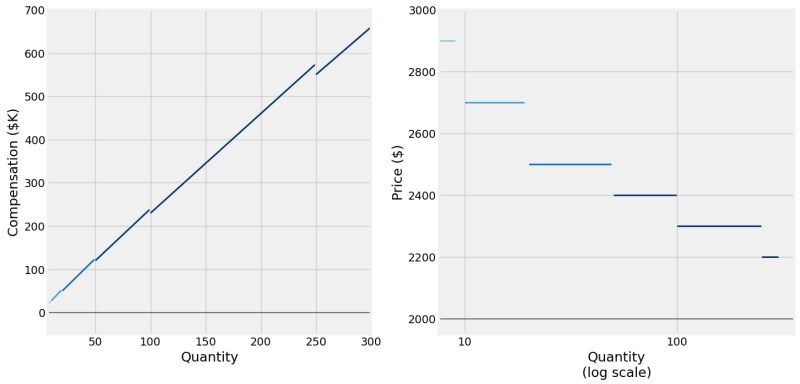


Figure 1: LifeLabs Price Schedule. The left panel shows the total charge as a function of quantity while the right panel depicts the marginal price. Note that the schedule is approximately concave. If it seems convex in the left panel, it is because of the log scale on the horizontal axis.

Figure 1 depicts LifeLabs' current price schedule. For each deal, aside from the quantity and the total price, we also observe whether the deal eventually succeeded. In other words, our data provides information on (potential) price and quantity not only for actual transactions, but also for potential potential ones that did not actualize. As we will discuss later on, this is key to identifying how per-workshop valuation across firms correlates with the sizes of their needs.

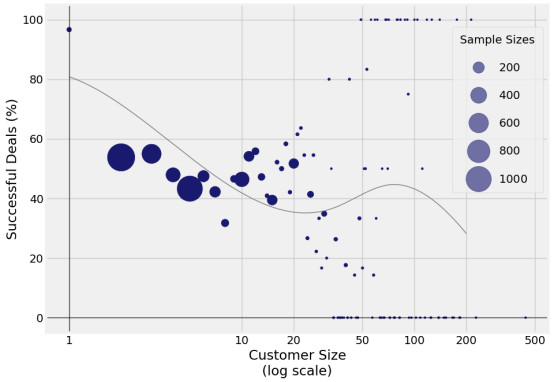


Figure 2: Percent Successful vs Deal Size

Figure 2 plots percent successful against deal size. As can be seen from this figure, a higher percentage of deals fail as we look at larger sizes; and this is in spite of the fact that for those deal sizes the per-workshop price is cheaper. This is suggestive that LifeLabs might want to further sharpen its volume discount policy in order to increase profitability. Of course a structural analysis encompassing both demand- and cost-side data would be necessary before one can (i) determine whether such a strategy is indeed recommended to LifeLabs and (ii) quantify the extent of it.

Figure 3 presents two histograms. The left panel simply shows the counts of deals (successful or not) of different size groups. The right panel depicts the total revenue (in $M) from each such size group. As these two panels together depict, larger deals are substantially less frequent than smaller ones. Nevertheless, they contribute meaningfully to the total firm revenue because each large deal contributes substantially more than each small one.

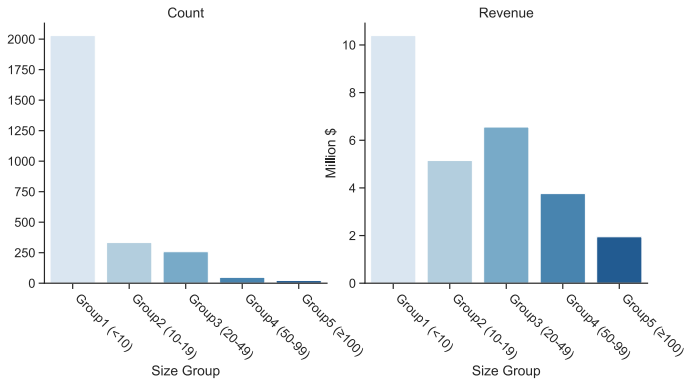
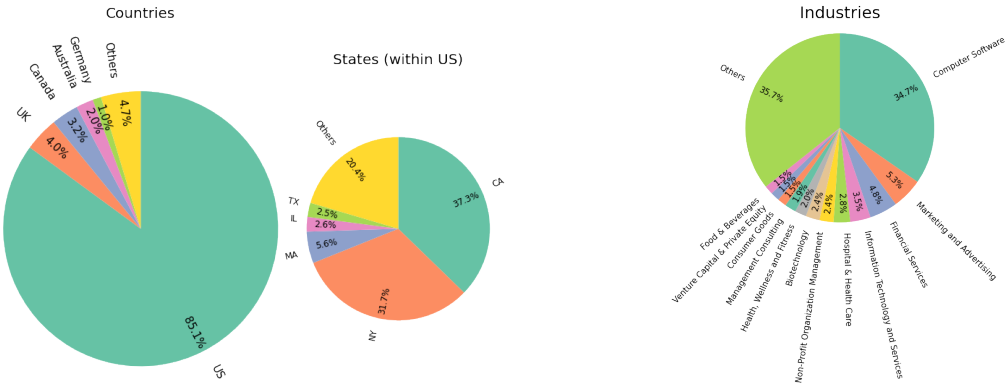
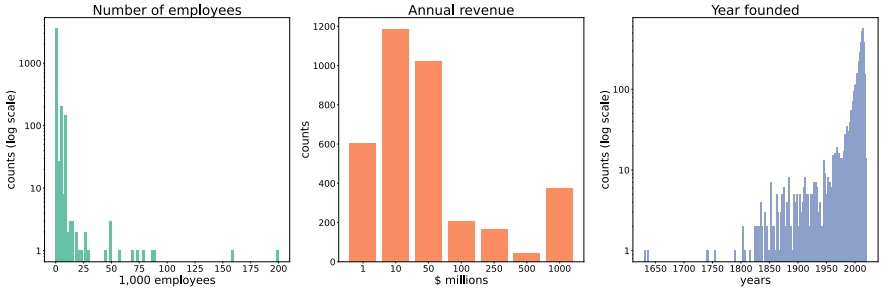


Figure 3: Count and Revenue by Groups



(a) By countries and By states

(b) By industries



(c) Histograms

Figure 4: Summary stats on customer characteristics

In addition to the deals data, we also possess a company-level dataset which provides information on different client characteristics. Among those are number of employees, a rough measure of annual revenue ($< \$1M$, $\$1 - 10M$, $\$10 - 50M$, $\$50 - 100M$, $\$100 - 1,000M$,) geographical location (country if abroad, city if in the U.S. or Canada,) industry, year founded, and some behavioral characteristics that we may not publicly share. Figure 4 provide summary statistics on deals and customer characteristics. Customers come from various industries with "Computer Software" being the most common. Although most customers are U.S. based, there is a non-trivial international demand. Finally, customers are mostly companies with fewer than 10,000 employees, $50M or less in annual revenue, and founded after 1950.

In addition to purchase data, we were provided with costs data and the company's estimates of its fixed and variable costs which we will use in our nonlinear tariff optimization procedure.

4 Model

Here, we seek to construct a model that allows us to capture the notion central to our problem: a flexible joint distribution over customers' "size" on the one hand and their willingness to pay on the other. Below, we describe different ingredients of this model.

Value functions. For each year t, each potential customer i decides whether and how many workshops it wants to purchase from the company. We directly model each customer i based on two aforementioned parameters: "size" $\bar{q}_i$ captures how many workshops customer i needs at most, and "valuation" v_i denotes the customer's willingness to pay for each workshop. Formally, customer i's willingness to

pay for q workshops is given by:

$$V_i(q) = v_i \times \min(q, \bar{q}_i) \tag{1}$$

In words, the customer values each workshop at v_i until it has received $\bar{q}_i$ ones, at which point it will no longer value additional workshops. This formulation has been used before in the literature to model demand for similar products to what we are considering, such as cloud computing services, etc (Devanur et al., 2020).

In spite of its parsimony, our formulation for value functions $V_i(\cdot)$ has the advantage that it captures exactly the two dimensions along which we seek to model heterogeneity across customers: size and value. Figure 5 should help illustrate this matter. This figure uses two simple examples to show how our parsimonious value function captures a key economic force. Each example describes a market with only two customers $i = 1, 2$. As the figure depicts, when the customer with a larger size q_i has a lower valuation v_i, a concave (i.e., flattening) contract is optimal. On the other hand, when the larger customer also has a higher per-unit valuation, then a convex (steepening) schedule is best. For a robustness analysis to smooth (as opposed to piece-wise linear) value functions, see appendix C.

Finally, Figure 5 demonstrates how our formulation, in spite of its simplicity, relaxes a restrictive assumption: As the bottom-right panel shows, $V_1(\cdot) - V_2(\cdot)$ is non-monotonic in q and has two roots. This relaxes the "single-crossing" assumption in much of the theory on nonlinear pricing (e.g., Maskin and Riley (1984); Mussa and Rosen (1978) and a long line of uni-dimensioanl-screening literature following them) and some of the empirical studies (e.g., Luo et al. (2018)).

Joint distribution of parameter values. In addition to the reassurance that our parsimonious approach to modeling value functions $V_i(\cdot)$ adequately captures a crucial economic force, Figure 5 has another implication. It suggests that the joint distribution $f(\cdot, \cdot)$ over $(\bar{q}_i, v_i)$ across costumers i is critical to our purposes in this study. The importance of the relationship between a "baseline willingness-to-pay" and "preference for quality/quantity" is congruent with theoretical and empirical research on price discrimination, market design, and multi-dimensional screening. Instances of such theoretical papers are Anderson and Dana Jr (2009); Haghpanah and Hartline (2021); Ghili (2022); Yang (2021). For an example of empirical papers emphasizing the importance of estimating joint distribution of multi-dimensional heterogeneity for market design, see Derdenger and Kumar (2013); and for a paper emphasizing the importance of the flexibility of this joint-distribution estimation, see Goldberg (2021).

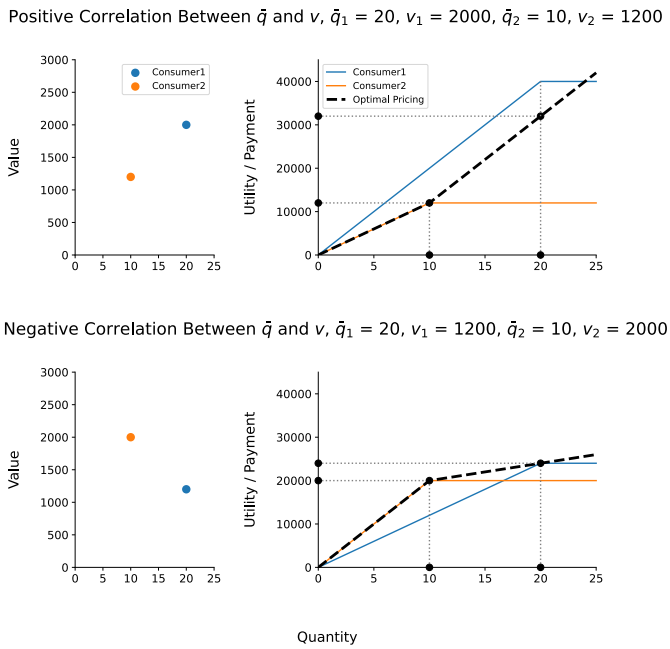


Figure 5: A two-customer illustration for the role of the size-value relationship on the shape of optimal pricing strategy.

As a result, unlike with the form of the value functions $V_i(\cdot)$, we will not be parsimonious about the modeling of $f(\cdot, \cdot)$. We will, rather, estimate this object flexibly. This flexibility, as will be demonstrated in next sections, will allow us to consider cases that are relevant to the design of the optimal pricing schedule but are more complex than a simple "positive or negative correlation" between size $\bar{q}$ and value v. An example is a scenario in which "mid-size" customers have on average lower valuations v_i compared to both smaller and larger customers.

Nonlinear price schedule. Denote the nonlinear tariff using function $P(q)$. This simply means any customer who purchases q units of the product will have to pay the total amount of $P(q)$ dollars. Without loss of generality, we restrict attention to price schedules that satisfy $P(0) = 0$.

Timeline. We assume a simple timeline for the model. In stage 1, all N potential customers learn about their $\bar{q}_i$ but not v_i. In stage 2, all potential customers engage in *costless* exchange with the seller to learn about the product. At the end of this stage, each customer i learns its v_i and the firm learns $\bar{q}_i$. In stage 3, potential customers make decisions on how many units of the product (if any) to purchase.

Figure 6 depicts the model timeline graphically. Note that the sequential revelation of $\bar{q}_i$ and v_i to the buyer does not have a central role in how the purchase decisions are made: the buyer learns both of these parameters before stage 3 where it makes the purchase decision. The purpose of setting up this model timeline is, hence, *not* to model incomplete or asymmetric information. But rather, the purpose is to construct the most succinct possible micro-foundation that can explain why those customers i who do not end up purchasing take the time to communicate their $\bar{q}_i$ to the seller. The simple answer provided by this model is that it is only at the end of this very communication that they learn their v_i is too low to justify any purchase.

To sum up, the importance of the timeline in Figure 6 is that it provides a framework that allows to

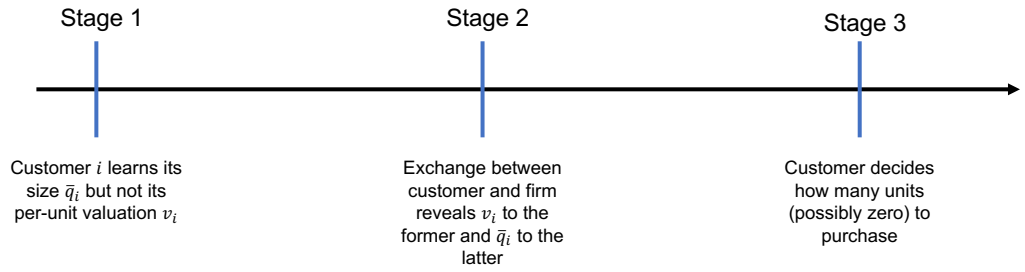


Figure 6: The timeline by which the seller and the customer learn and the customer makes purchase decision.

interpret and structurally analyze datasets that record deal-size information for unsuccesful deals. This simple setup is also helpful in clarifying the directions in which the model can be extended, a discussion of which is provided in section 8.

Customer purchase decisions and firm profit. We now turn to quantifying the purchase decisions. Each consumer's net value for q workshops will be given by: $V_i(q) - P(q)$. Thus, customer i's purchase decision q_i will solve the following optimization problem:

$$q_i = q^*(P|v_i, \bar{q}_i) \coloneqq \arg\max_{q \geq 0} V_i(q) - P(q) \tag{2}$$

Seller profit under pricing strategy $P(\cdot)$ is given by:

$$\pi(P) = N \times \int_{v,\bar{q}} P\big(q^*(P|v,\bar{q})\big) - c_1 \times \mathbf{1}_{q^*(P|v,\bar{q})>0} - c_2 \times \big(q^*(P|v,\bar{q})\big) f(v,\bar{q}) dv d\bar{q} \tag{3}$$

In words, the profit is given by revenue net of costs, integrated over customer types. Note that in this continuous-choice setting, the cost function is more complex than it would be under discrete choice. In particular, marginal cost has two components: c_1 is the per-customer component of costs which would be incurred once any customer decides to buy a positive amount. On the other hand, c_2 is the per-unit cost and would be incurred with every additional workshop provided to a customer.

The firm problem is to find the price schedule $P(\cdot)$ to maximize profit:

$$P^* = \arg\max_{P(\cdot)} \pi(P) \tag{4}$$

We next turn to the estimation procedure and identification.

5 Estimation

The object we seek to estimate is the joint probability distribution $f(\cdot,\cdot)$ over values v and sizes $\bar{q}$. We do this in two steps. We start by estimating the marginal distribution $f_{\bar{Q}}(\cdot)$ for size $\bar{q}$, and then move on to estimate the conditional distribution $f_{V|\bar{Q}}(\cdot)$ over values v.

For the first step, we take advantage of a feature of a model, alongside an approximation.

Lemma 1. *If price schedule $P(\cdot)$ is strictly increasing and concave, then for any customer type $(v, \bar{q})$, we have $q^*(P|v, \bar{q}) \in \{0, \bar{q}\}$.*

The proof of this lemma is relegated to the appendix. This lemma simply says that under a weakly concave price schedule, any customer will either buy no workshops or exactly as much as its size $\bar{q}$.

Due to gradually decreasing marginal prices (as depicted by figure 1), we treat the observed price schedule in our data is approximately concave. As a result, we assume, for each customer i who purchased q_i unites, that $q_i = \bar{q}_i$. Also, for those who did not purchase, we assume the q_i size in our data which they were considering purchasing was indeed their size $\bar{q}_i$. To sum up, we estimate the marginal distribution $f_{\bar{Q}}(\cdot)$ by equating it with the distribution of observed q^* amounts (across both successful and unsuccessful deals).

In the appendix, we perform a robustness check to our concavity approximation by removing the few data points for which the observed q_i cannot be equal to $\bar{q}_i$ (due to the discontinuities in the schedule). The results change only negligibly. Also note that the marginal distribution $f_{\bar{Q}}(\cdot)$ can still be estimated without a concave (or approximately) price schedule, though in that case some parametric assumptions may be necessary.

With our estimation for $f_{\bar{Q}}(\cdot)$ at hand, we next turn to estimating $f_{V|\bar{Q}}(\cdot)$. We use the following model:

$$v_{it} = \beta \times X_i + \alpha_t + \gamma_{\tilde{q}} + \epsilon_{it} \tag{5}$$

In this equation, i denotes the customer and t represents the year. Also, X_i is observable customer characteristics, β is a vector of coefficients determining the weights of different customer characteristics, and α_t represents yearly fixed effects. Also, in order to directly relate $\bar{q}$ to v beyond what could be provided by observables, we allow "size-group" fixed effects $\gamma_{\tilde{q}}$ which captures fixed effects at the level of $\tilde{q}$, a dummy that captures whithin which of the three intervals $[1, 20), [20, 50)$ or $[50, \infty)$ customer i's size $\bar{q}$ falls. These size-group fixed effects have a critical role in our model's ability to capture the relationship between $\bar{q}_i$ and v_i across customers i. Finally, $\epsilon_{it} \overset{iid}{\sim} logistic(0, \sigma)$. That is, ϵ is the mean-zero error term with a logistic distribution. Note that we do not normalize the standard deviation of ϵ_{it} to 1. This is because the price coefficient in the customers' net value formula has already been normalized to -1.

We estimate the model in equation 5 using a MLE approach. The likelihood function is as follows:

$$\mathcal{L}(\beta, \alpha, \gamma) = \Pi_{it} \text{Prob}\big(s_{it} = \mathbf{1}_{q^*(P|v_{it}, \bar{q}_{it}) > 0}\big) \times \text{Prob}\big(v_{it} \geq 0\big) \tag{6}$$

where s_{it} is the observable binary variable denoting whether deal it was successful. This likelihood function gives the probability that the predicted deal success matches with the observed one. It also adds a term that would punish for predicting negative values for workshops.

Identification. Of course, overall, the entire variation in the data identifies all of the variables of interest. But an informal description of our intuition for what mainly identifies what would still be useful. The nonlinear function β is identified by the variation in firm characteristics. Year fixed effects are identified by the variation in deal success across years for similarly looking firms negotiating similarly sized deals. Size bucket fixed effects are identified by the differential rates of deal success across different

deal sizes. The standard deviation σ of the error term is identified in part by the price variation within the price schedule. Recall that the price schedule has finer variation in it than our size bucket fixed effects. Thus, within each size bucket there is some variation in price. If σ is too small, the price variation will cause large variation in deal success within each size bucket. If σ is too large, then all deal success rates will be close to each other and close to 0.5. As a result, the deal success rate variation in the data will help pin down σ. In addition, if σ is too large, the probability that each v_{it} is negative will approach 0.5, which is also punished by the likelihood function.

5.1 Estimation Results

There are both cost parameters and demand parameters in our model. This section describes how we estimate the demand-side parameters using the procedure described before and how we directly calibrate the cost-side ones based on data from the company.

5.1.1 Demand Side Parameters

As mentioned before, the objective here is to estimate the joint distribution $f(\cdot,\cdot)$ over $\bar{q}$ and v as flexibly as possible. We directly estimate $f_{\bar{Q}}(\cdot)$ directly off of the data and estimate $f_{V|\bar{Q}}(\cdot)$ by finding the parameter values (β, α, γ) in equation 5 for v_{it} that maximize the likelihood function in equation 6. The MLE results are presented in Table 1. As for what company characteristics are included in vector X from equation 5, Aside from the year fixed effects and size-group fixed effects which seemed to be the most explanatory regressors, two features proved to have the most predictive power on whether a deal would happen, conditional on α and γ. One was age of the customer (as a firm) and the other was a behavioral characteristic which we term "feature 1".[2] Interestingly, features such as number of employees, revenue, location, or industry were not highly predictive and were left out of our analysis. One could in principle include them but we would expect the change that would cause to our results to be negligible.

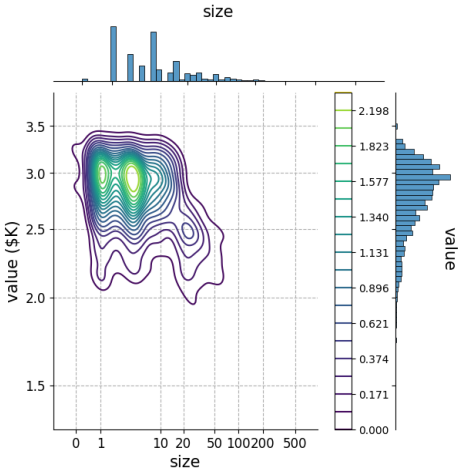
Log Feature 1 β_1	*Log Firm Age* β_{age}	*Time* α_{2021}	*Intercept* γ_{small}	*Mid Size* γ_{medium}	*Large Size* γ_{big}	*Scale* σ	*Negative Log-Likelihood*
180.45	-43.29	80.59	2,226.74	-549.94	-666.16	388.12	3076.80
(9.77)	(13.41)	(24.10)	(52.42)	(43.29)	(80.17)	(9.53)	

Notes: Regression results for equation 5. Constrained optimization with bound on positive scale parameter(or precision parameter) was implemented using *L-BFGS-B* method with stopping tolerance of 10^{-15}.

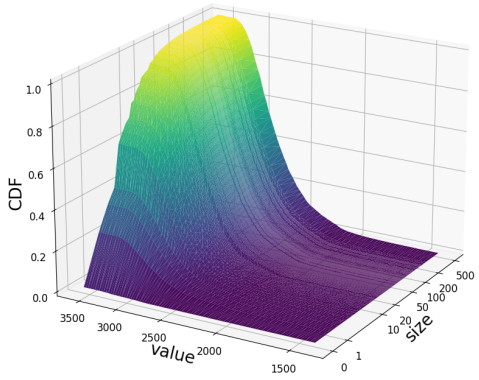
Table 1: Maximum Likelihood Estimates for the parameters describing $f_{V|\bar{Q}}(\cdot)$ according to equation 5. Bootstrapped standard errors are shown in parentheses.

As the results from this table suggest, both mid-size and large customers have, on average, smaller per-unit willingnesses to pay–respectively by \$550 and \$666 per unit–for the company's product relative to smaller customers (again, recall that by "size" we do not mean the size of the customer as a firm. We mean the size of their need for the product the company sells to them).

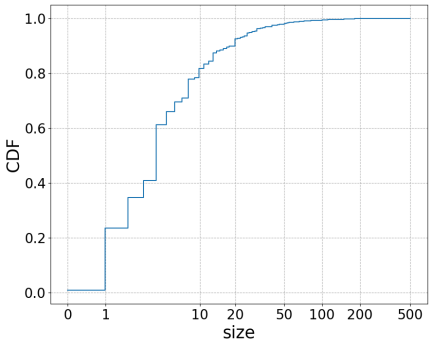
[2] We cannot disclose the nature of Feature 1 due to non-disclosure agreement.



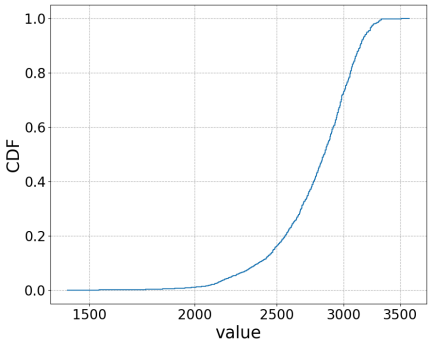
(a) Joint PDF of Size and Value



(b) Joint CDF of Size and Value



(c) Marginal CDF of Size



(d) Marginal CDF of Value

Notes: For visualization *Size* and *Value* used in generating pdfs and cdfs are log transformed. (a) is the joint pdf contour plot of size and value, with marginal histograms of value on the right and size on the top. (b) is the joint cdf of size and value in 3D plot, lighter shade meaning higher value of cdf. (c) and (d) are marginal cdf of size and value respectively.

Figure 7: Estimated Distributions for $\widehat{v}_{it}$ and $ln\,\bar{q}$

With both $f_{V|\bar{Q}}(\cdot)$ and $f_{\bar{Q}}(\cdot)$ in hand, the joint distribution $f(\cdot,\cdot)$ is recovered. Figure 7 visually presents this joint distribution as well as each of the marginal distributions. Each distribution can be found both in the density form and in the cumulative form by examining different panels of the figure. The top-left panel of this figure, confirms the suggestive evidence in 2 that the per-unit willingness to pay seems to be smaller for customers with larger needs. As a result , it seems natural to expect the company to want to offer lower per-unit prices for larger deals. The question is "by how much," which our estimates of the demand parameters, as well as cost estimates provided below, help us quantify.

5.1.2 Cost Parameters

We have direct data on the breakdown of cost side parameters from our conversations with the company. Based on these conversations, we arrive at the following figures: c_1 ≈\$3,490/customer and c_2 ≈\$748/workshop.

Figure 8 plots the average cost (i.e., $\frac{c_1}{q} + c_2$) as a function of quantity q. This figure shows that, similar to demand side parameters, our cost side analysis suggests that the company's optimal nonlinear

tariff will lower the per-unit price for larger deals. We now turn to quantifying the exact shape of this optimal nonlinear contract.

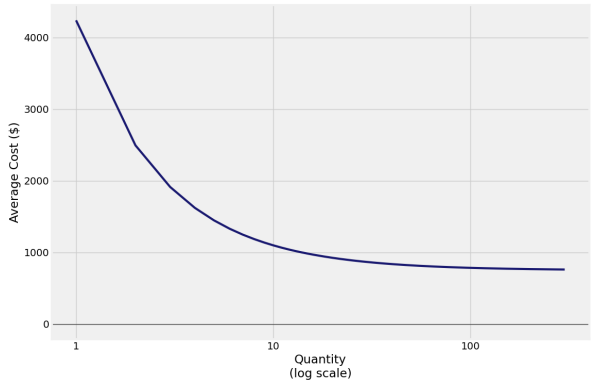


Figure 8: Average Cost Curve

5.2 Model Fit

Figure 9 describes goodness of fit for our estimated model. It compares multiple quantities measured directly on the data to their counterparts generated by the model. In particular, it examines deal purchase rate, total revenue, total cost, and total profit (revenue net of costs) by three size groups. As can be seen from the figure, our model fits the data very closely.

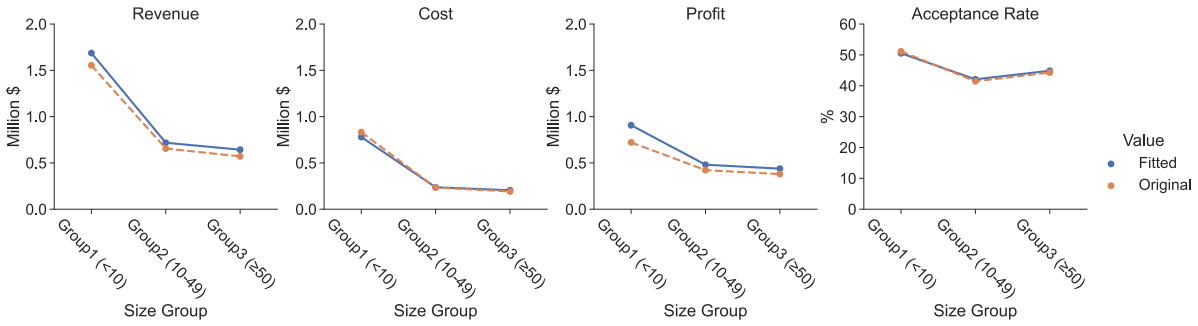


Figure 9: Model Prediction VS Original Data

One reason for the close fit is the size-group fixed effects $\gamma_{\tilde{q}}$ in our specification of v_i in equation 5. If we extend the model fit analysis to finer size groups, the fit will still be strong but will be slightly less accurate (especially for acceptance rates). See Figure 10.

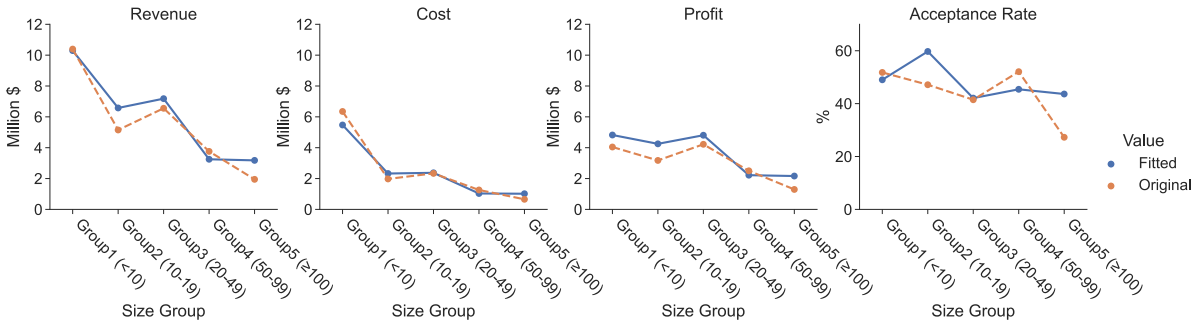


Figure 10: Model Prediction VS Original Data (5 Groups)

6 Optimal Nonlinear Pricing Scheme

With an estimated model of demand and costs in hand that closely fits the data, we are now ready to solve the optimization problem in equation 4 and arrive at the profit-maximizing schedule. In this section, we discuss three topics. First, we provide an overview of our optimization method. Then we we present the main results on the optimal schedule. Finally, we move to discussing this optimal schedule's profitability, and consumer- and social-welfare implications.

We need to start with a parameterization for the price schedule $P(\cdot)$. The main parameterization we work with is similar to LifeLabs' current pricing strategy: it consists of a few linear segments where the continuation of each of them would pass through the origin. Formally, consider intervals $I_k = [q_k, q_{k+1})$ where $k = 1, 2, ...$ and $q_1 = 0$. We allow price schedule P to take the form $P(q) = p_k \times q, \forall q \in I_k$, where p_k values are constants. In our application, we consider I_1 through I_5 to be, respectively, $[0, 10), [10, 20), [20, 50), [50, 100), [100, \infty)$. Note this parameterization restricts the space of all possible schedules from an infinite dimensional object to a 5-dimensional one. As a result, with some abuse of notation, we sometimes refer to the function $P(\cdot)$ as a vector $P = (p_1, ..., p_5)$.

Note that we could in principle restrict this space in alternative ways. For instance, we could consider a piecewise linear and continuous schedule using the five intervals above. We in fact do examine this continuous alternative initially (see next subsection). But given that the collection of segments whose contiuations pass through the origin yielded a higher profit than the optimal continuous schedule, we focus for the most part on this latter form of pricing strategy. With the parameterization structure in place, we next turn to our method on how to optimize the tariff.

6.1 Optimization Method

Though mechanism design theory does provide tractable methods to compute optimal nonlinear tariffs, all of those methods are devised under assumptions that seldom hold empirically.[3] As a result, we turn to numerical approaches to find the optimal schedule.

Our problem has some features that are crucial in determining what method is used for optimization. First, the number of dimensions (i.e., five) is non-trivial but not too large. Second, we seek to find the

[3]For instance, methods provided in Mussa and Rosen (1978); Maskin and Riley (1984) and similar papers all rely on a "single-crossing" condition which make the customer heterogeneity one-dimensional, an assumption that is clearly violated in our context (for instance in the left panel of Figure 5).

global maximum as opposed to a local one. Third, each instance of computing the objective function (i.e., the profit) is costly. Fourth, the objective function is expected to behave non-smoothly, and there are no guarantees on concavity/linearity features that one could leverage.

Under the above circumstances, the optimization literature recommends the use of Bayesian methods (see Frazier (2018a,b) for an overview). We found, however, that a multi-dimensional variant of bisection search outperforms a Bayesian approach in the sense of delivering a higher objective-function value in a shorter amount of time. In this subsection, we briefly describe our method. In Appendix D, we also comparatively analyze multiple alternatives (a variant of Bayesian Optimization and multiple versions of gradient descent) and provide a summary on which approach we think is the most appropriate one under different problem settings.

In a nutshell, our method starts with a 5-dimensional grid of all the five possible prices $(p_1, ..., p_5)$ described in the previous subsection. Each grid is of size d, which implies that we initially evaluate the profit function π under d^5 possible values for vector P. The values for each p_k ($k \in \{1, ..., 5\}$) in this initial grid are chosen to be equi-distant points on the interval $[0, 5000]$. In other words, the lower and upper bounds in iteration-1 of the algorithm for each p_k are given by $\underline{p}_k^1 = 0$ and $\bar{p}_k^1 = 5000$. Denote the length of the interval on the k-th dimension by $l_k^1 = \bar{p}_k^1 - \underline{p}_k^1$.[4]

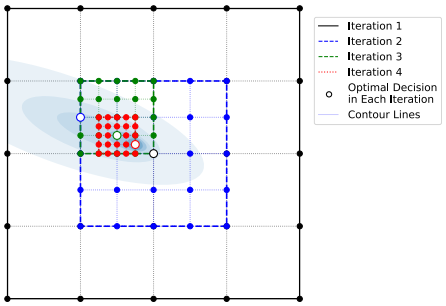
We then proceed with an iterative process. In each iteration t, we first find the optimal price vector $P^{t*} = (p_1^{t*}, ..., p_5^{t*})$ among the d^5 candidates in our price grid. Next, we form a new grid (for the next iteration) by "zooming in" on P^{t*}. More precisely, the new grid is constructed based on P^{t*} and is a zoom factor $z \in (0, 1)$ in the following manner:

$$(\underline{p}_k^{t+1}, \bar{p}_k^{t+1}) = \begin{cases} (\underline{p}_k^t, \underline{p}_k^t + l_k^t), & \text{if } p_k^{t*} - z\frac{l_k^t}{2} < \underline{p}_k^{t+1} \\ (\bar{p}_k^t - l_k^t, \bar{p}_k^t), & \text{if } p_k^{t*} + z\frac{l_k^t}{2} > \bar{p}_k^{t+1} \\ (p_k^{t*} - z\frac{l_k^t}{2}, p_k^{t*} + z\frac{l_k^t}{2}), & \text{otherwise} \end{cases} \tag{7}$$

In words, the new bounds $\underline{p}_k^{t+1}$ and $\bar{p}_k^{t+1}$ form an interval of length $z \times l_k^t$ centered around p_k^{t*}, unless this interval itself falls partially outside of the old bounds $\underline{p}_k^t$ and $\bar{p}_k^t$ (in which case the new interval is moved until it falls just within the old one). We then iterate and construct smaller and smaller intervals until all we reach t such that all l_k^t are smaller than a pre-determined threshold ε, at which point the algorithm stops.

Though we are not aware of a systematic analysis of the properties and performance of this grid-bisection approach, we are aware that variants of it have been previously used in different fields. One example is Yin et al. (2020), who also provide a visual illustration of their variation. Following their illustration, Figure 11 schematically describes our grid-binary optimization procedure.

[4]Note that by our choices of the $\underline{p}_k^1$ and $\bar{p}_k^1$ values, all l_k^1 are equal to each other; but this need not be the case generally.



Notes: (1) We are using $d = 2$ for demonstration. In the paper, $d = 5$. (2) In each loop t, the length of the binary search interval for each dimension k shrinks by $z = 0.5$.

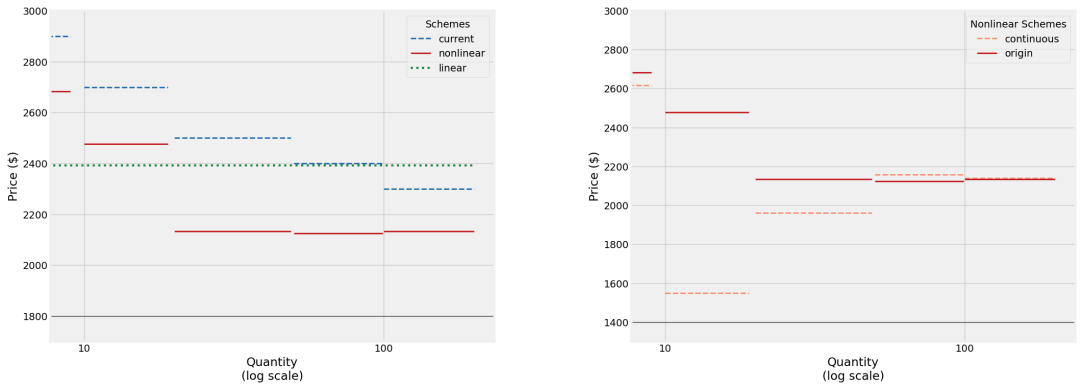
Figure 11: Visual illustration of the grid-bisection optimization method.

The algorithm involves d^5 instances of computing the profit function π in each iteration. It approximately takes $\frac{\log(l_k^1)-\log(\varepsilon)}{-\log(z)}$ iterations for the algorithm to stop. We choose $d = 5$ and $z = \frac{1}{2}$ for our application.

With this grid-bisection optimization method in hand, we next turn to using the method to recover the optimal price schedule.

6.2 Optimal Price Schedule

Figure 12 compares multiple pricing schemes: (i) the current price schedule by LifeLabs, (ii) the optimal linear price schedule, and (iii) the optimal nonliner price schedule. Notably, as mentioned in the previous subsection, we solve for the optimal nonlinear schedule in two ways: first we find the optimal schedule among those that charge fixed per-unit prices within each of the five intervals mentioned above. Second, we find the optimal schedule among those that charge a fixed per-unit *incremental* rate within each of these five intervals. In other words, the first type of nonlinear schedule consists of five segments where the continuation of each of them passes through the origin; whereas the second type consists of five segments that are connected together and form a continuous function.



(a) Marginal Price

(b) Marginal Price: nonlinear price schemes

Notes: Panel (a) shows the marginal prices for *current*, *continuous nonlinear*, *via origin nonlinear*, and *linear* price schedules. (b) is the comparison of marginal prices between *continuous* and *via origin* nonlinear price schedules. Note the horizontal axis is on the log scale.

Figure 12: Pricing Schemes

Panel (a) provides a comparison across different contracts by plotting marginal prices as a function of deal size (for all contracts except the continuous one.) As can be seen from this panel, the optimal nonlinear schedule charges more than the optimal linear one for small deals and less for larger deals. This should not be surprising: the optimization process for a linear schedule tries to strike a balance between the higher prices needed for small deals and the lower ones needed for larger deals.

Panel (b) attempts to provide a comparison between marginal prices of the optimal continuous schedule and those of the optimal schedule with segments that would pass through origin.[5] Next section shows that in terms of profitability, the two schedules indeed behave similarly with the discrete one slightly outperforming the continuous one.

6.3 Profit and Welfare Analysis

Table 2 shows how different pricing schemes fare against one another with respect to a number of measures. In addition to the current pricing by the firm, three other pricing schemes are examined. The first scheme is first-degree price discrimination in which the firm tailors pricing toward each individual customer. In this regime, the firm will sell to any customer i with $v_i \times \bar{q}_i \geq c_1 + c_2 \times \bar{q}_i$, charging exactly $v_i \times \bar{q}_i$. The second regime is linear pricing, meaning the firm is only allowed to charge tariffs in the form $P(q) \equiv p \times q$, choosing p optimally. Finally, the third tariff is the optimal nonlinear pricing scheme $P^*(\cdot)$ which was described shortly before. Both the oprimal linear price and the optimal nonlinear schedule were shown in figure 12. The three measures on which the above pricing schemes are compared against each other are (i) firm profit, (ii) consumer surplus, and (iii) social surplus.

[5]Though some of the marginal prices are far from each other, this is in fact an indication of overall similarity. To illustrate, in the optimal continuous schedule, the segment for deals of sizes 10 to 19 would have a higher starting point relative to a similarly sloped segment for those sizes in a schedule with segments passing through origin (this is because the former simply starts form the end of the segment for smaller deals). As a result, a lower slope in this case for the continuous schedule compensates for the higher starting point.

Scheme	*Revenue* *($M)*	change (%)	*Profit* *($M)*	change (%)	*Consumer* *Welfare* *($M)*	change (%)	*Social* *Welfare* *($M)*	change (%)
current	30.48	-	18.26	-	6.69	-	24.96	-
1^{st} *degree*	61.99	+103.38%	38.78	+112.38%	0	−100.00%	38.78	+55.37%
linear	34.86	+14.37%	18.71	+ 2.46%	10.10	+ 50.97%	28.81	+15.42%
nonlinear								
continuous	36.81	+20.77%	20.13	+ 10.24%	11.11	+ 66.07%	31.23	+25.12%
origin	36.16	+18.64%	20.19	+ 10.57%	10.20	+ 52.47%	30.39	+21.75%

Table 2: Profit and Welfare Analysis

As can be seen from the table, the optimal linear pricing strategy delivers almost as much profit as the current nonlinear pricing scheme used by the firm. The optimal nonlinear pricing scheme improves profitability by about 7.9%. The profitability has been computed both for the optimal nonlinear schedule with segments passing through origin and for the optimal continuous schedule. As the table shows, the former slightly outperforms the latter. As a result of this, for the rest of the paper, we work with optimal nonlinear schedules with segments passing through origin.

Another question that table 2 helps answer is about the comparison between first- and second-degree price discrimination. If first degree price discrimination were possible, it would more than double the profit. This means that optimal second degree price discrimination recovers only about 7.4% of the profitability gap between linear pricing on the one hand and first degree price discrimination on the other.

Aside from profit levels, Table 2 also compares different pricing schemes with regards to their effects on consumer- and social-surplus. Compared to linear pricing, nonlinear pricing increases consumer welfare hurts some consumers and benefits others.[6] On the front of social surplus, nonlinear pricing outperforms linear pricing by about 5.5%. This is because the benefit of nonlinear pricing to the company outweighs its net harm to consumers.

Segment-by-segment analysis. In addition to the aggregate analysis presented by Table 2, it is worth conducting a segment-by-segment analysis of how different pricing policies compare against one another. Figure 13 presents such segment-based results. There are two general lessons from this segment-based analysis.

[6]Note that this empirical result is not general. In theory, the welfare effects of price discrimination could go either way. See Schmalensee (1981) or Varian (1985) for instance.

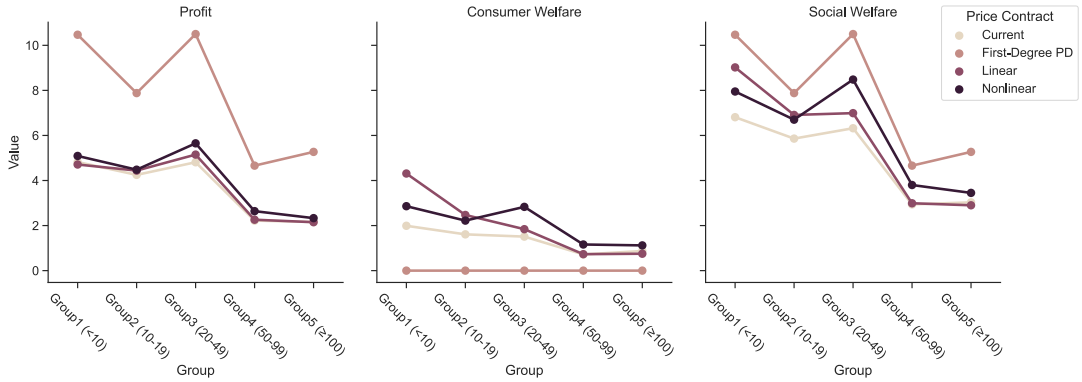


Figure 13: Visualization of Profit and Welfare Analysis by Groups

	Group1 (<10)	Group2 (10-19)	Group3 (20-49)	Group4 (50-99)	Group5 (≥100)	Total
			Profit			
Current	4.82	4.25	4.81	2.22	2.16	18.26
First-Degree PD	10.47	7.88	10.50	4.66	5.27	38.78
Linear	4.71	4.44	5.15	2.26	2.15	18.71
Nonlinear	5.09	4.48	5.65	2.64	2.33	20.19
			Consumer Welfare			
Current	1.99	1.61	1.51	0.72	0.87	6.70
First-Degree PD	0.00	0.00	0.00	0.00	0.00	0.00
Linear	4.31	2.47	1.84	0.73	0.75	10.10
Nonlinear	2.86	2.22	2.83	1.16	1.12	10.19
			Social Welfare			
Current	6.81	5.86	6.32	2.94	3.03	24.96
First-Degree PD	10.47	7.88	10.50	4.66	5.27	38.78
Linear	9.02	6.91	6.99	2.99	2.90	28.81
Nonlinear	7.95	6.70	8.48	3.80	3.45	30.38

Table 3: Profit and Welfare Analysis by Groups

First, and unsurprisingly, 1st-degree price discrimination dominates all other methods both on the front of firm profit and on the front of social welfare. Its high performance on social welfare is because this approach, by construction, allows a transaction to take place if and only if it is socially efficient (i.e., if $v_i\bar{q}_i \geq c_1 + c_2\bar{q}_i$). Given the wide gap between the profitability of this method and other ones, it is natural to expect that the firm would benefit from a customer-by-customer negotiation approach as opposed to posted pricing; because even though this approach is unlikely to replicate the exact profit from 1st-degree price discrimination, replicating even a fraction of it would outperform nonlinear pricing. Based on the most up-to-date knowledge that we have from the firm, this is the approach they are taking.

The second lesson from the segment-based analysis concerns the comparison between linear and nonlinear schedules. Nonlinear pricing generates lower consumer welfare for small-size customer than does linear pricing. This is reversed for larger customers. This should not be surprising. By finding the per-unit price that is, loosely speaking, "optimal on average", linear pricing benefits segments that

received higher prices under nonlinear pricing and hurts those who received lower prices under the nonlinear schedule. With respect to profitability, nonlinear pricing fully dominates linear pricing. This, too, should not be surprising given that nonlinear pricing allows the firm to tailor the charges toward respective size groups.[7]

7 Further Counterfactual Analyses

The main purposes of this paper were (i) to develop a method for optimizing a fully nonlinear price schedule and (ii) to demonstrate its relevance by empirically implementing it in the context of one application. We believe, however, there are additional insights from our analysis that are worth noting. In this section, we conduct further counterfactual analysis in order to arrive at insights which could be useful to practitioners even if they do not fully implement our estimation and/or optimization procedure.

7.1 The Role of Incentive Compatibility Constraints

As we explained in section 6, we *jointly* optimize the five per-unit prices $p_1, ..., p_5$. A key question is: can we optimize these prices separately instead? That is, can we take each price p_k for $k = 1, ..., 5$ and optimize it *only for the market of customers within the relevant range I_k of sizes*? Formally, define the "local profit function" $\pi_k(p)$ for $p \in \mathbb{R}$ to mean the profit to the firm if (i) the set of potential customer consisted only of those with $\bar{q}_i \in I_k$ and (ii) the firm charged the linear price schedule of $P(q) = p \times q$:

$$\pi_k(p) = N_k \times \int_{v,\bar{q}} p \times \left(q^*(p|v,\bar{q})\right) - c_1 \times \mathbf{1}_{q^*(p|v,\bar{q})>0} - c_2 \times \left(q^*(p|v,\bar{q})\right) f(v,\bar{q}|\bar{q} \in I_k) dv d\bar{q} \tag{8}$$

where N_k is the total count of all potential customers i with $\bar{q}_i \in I_k$; and with a slight abuse of notation, $q^*(p|v,\bar{q})$ is the amount purchased by customer with size $\bar{q}$ and value v under the linear price schedule of $P(q) \equiv p \times q$.

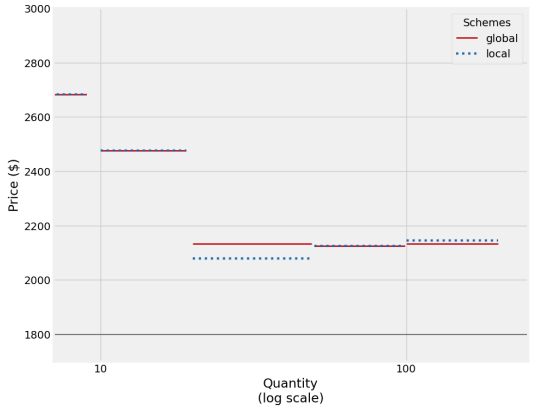
Denote by $\tilde{p}_k$ the "locally optimal price for size group I_k". That is:

$$\tilde{p}_k = \arg\max_{p \in \mathbb{R}} \pi_k(p) \tag{9}$$

We can now formally state what it means to "optimize prices $p_1, ..., p_5$ separately". It means charging the schedule $\tilde{P} = (\tilde{p}_k)_{k=1,...,5}$ instead of $P^* = (p_k^*)_{k=1,...,5}$. Likewise, a formal way of asking "what would be the consequences of optimizing prices separately" would be to ask "how closely does $\pi(\tilde{P})$ approximate $\pi(P^*)$?" This is the key question that this subsection investigates. This is an important question both conceptually (are different size groups effectively "separate" markets?) and computationally (can we avoid the difficult joint optimization problem and simply optimize one-dimensional objects?).

By the definition of P^* as the optimal schedule, it has to be that $\pi(\tilde{P}) \leq \pi(P^*)$. The reason why

[7] As we will see later, however, this intuition is incomplete. Optimal nonlinear pricing entails more than choosing the right price for each segment. As emphasized in a long line of theoretical research on screening, nonlinear pricing also requires careful attention to whether customers in each segment respond to relative prices by purchasing a different amount that was "meant for them."



(a) Marginal Price

Scheme	*Revenue ($M/y)*	*Profit ($M/y)*
local	36.16	20.06
global	36.16	20.19

(b) Scheme Performance

Figure 14: Local and Global Schemes

this inequality may be strict is "incentive compatibility" constraints that are ignored when prices are separately optimized. To illustrate, optimizing separately ignores the fact that if $\tilde{p}_3$ is substantially smaller than $\tilde{p}_4$, then some customers i of size-group 4 (i.e., $\bar{q}_i \in I_4$) who would purchase under $\tilde{p}_4$ absent other options, might take the opportunity presented by the wide gap between $\tilde{p}_3$ and $\tilde{p}_4$ and reduce their purchase sizes. Similarly, if $\tilde{p}_3$ is substantially lower than $\tilde{p}_2$, customers of size group 2 might respond by increasing their purchase sizes, paying an overall lower total price, and imposing a higher cost of production to the firm. Put differently, selling to these size groups separately-which, if feasible, would yield a total profit of $\Sigma_k \pi_k(\tilde{p}_k)$– is *not incentive compatible*.

As a long line of research on mechanism design suggests, it would not be surprising to see that ignoring incentive compatibility results in strictly positive losses: $\pi(P^*) - \pi(\tilde{P}) > 0$. The empirical question, however, is how significant is the loss? Figure 14 answers this question by plotting the locally and globally optimal pricing schemes and reporting their overall profitabilities.

As the figure shows, the "locally optimized" schedule $\tilde{P}(\cdot)$ turns out to approximate the profitability of the "globally optimized" $P^*(\cdot)$ reasonably well in this application. As shown in panel (a), the two schedules look similar. Also, as shown in panel (b), the yearly profitability of the locally optimized nonlinear pricing schedule $\tilde{P}(\cdot)$ is \$20.06M/y. Unsurprisingly, this is higher than the profitability of the optimal linear schedule but lower than that of the globally optimized nonlinear one. This local optimization delivers approximately 99% of the profitability of the global optimum ($\frac{20.06}{20.19} \approx 0.99$), thereby recovering 91% of the profitability gap between this global optimum and the best linear schedule ($\frac{20.06-18.71}{20.19-18.71} \approx 0.91$).

If the above insight is generalizable, then firms may be able to achieve nearly optimal profitability using technically and computationally cheaper means. We will delve deeper into the generalizability of this insight in subsection 7.4 and try to sketch out conditions under which, unlike the current situation, "local optimization" is risky. In that section, we also move beyond a comparison of profit levels, and provide a detailed discussion of the ways in which the *shapes* of locally and globally optimal contracts differ from one another.

7.2 Demand- v.s. Cost-Side Factors

The fact that our optimal price schedule involves lower per-unit prices for larger deals seems consistent with both demand-side cost-side estimates. On the demand side, as Table 1 suggests, customers with medium and large sizes $\bar{q}_i$ tend to have lower valuations v_i. On the cost-side, as Figure 8 suggests, average cost to sell deals of larger sizes is substantially lower than that for smaller deals. An empirical question is, hence, which of these two factors is a more important reason behind the shape of our optimal price schedule?

Figure 15 helps answer this question. In the left panel, we compare the optimal price schedule (solid blue lines) to the optimal schedule under the counterfactual scenario of no costs. The right panel considers a counterfactual scenario in which size fixed-effects in the formulation for valuation v_i are homogenized to a proper weighted average. As the figure shows, the downward trend of per-unit prices in deal size is qualitatively preserved when we "shut off" either factor. Thus both demand and cost factors seem to have a role in the shape of the optimal contract.

That said, in terms of magnitude, demand side factors seem to have a larger role. To see this, observe that the counterfactual optimal schedule on the left panel of Figure 15 has a much wider range of prices compared to the one on the right. For an intuitive understanding of why this is the case, note that costs affect the shape of the optimal schedule through *two* channels. The first channel is what was discussed so far: deal sizes with higher average costs tend to have higher per-unit prices in the optimal contract. The second channel has to do directly with the existence of costs rather than heterogeneity in them: when there are marginal costs, the optimal price schedule is in part guided by how to cover those costs; and this leaves the seller with less freedom to shape the price schedule based on demand-side factors. To sum up: when costs are assumed away (green dashed lines in the left panel), not only will their direct effect on prices will be gone, but so will their moderating effect on the role of demand.

With this high-level comparison between cost- and demand-side factors completed, we next turn to a deeper analysis of each side.

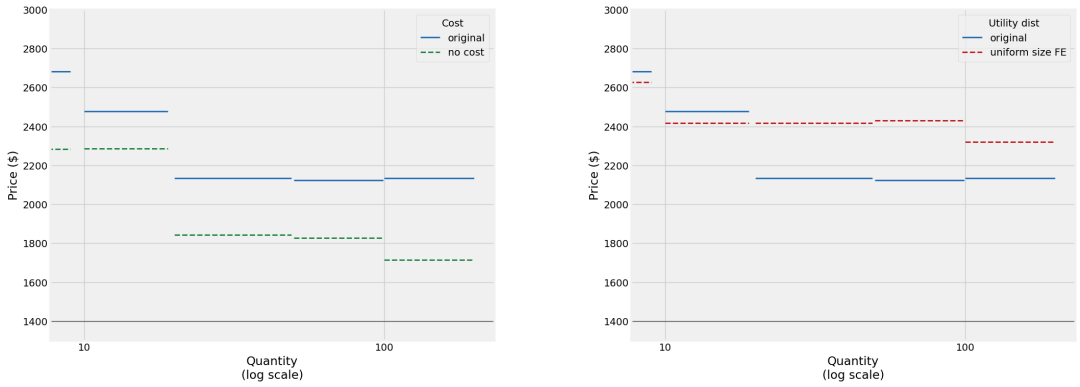
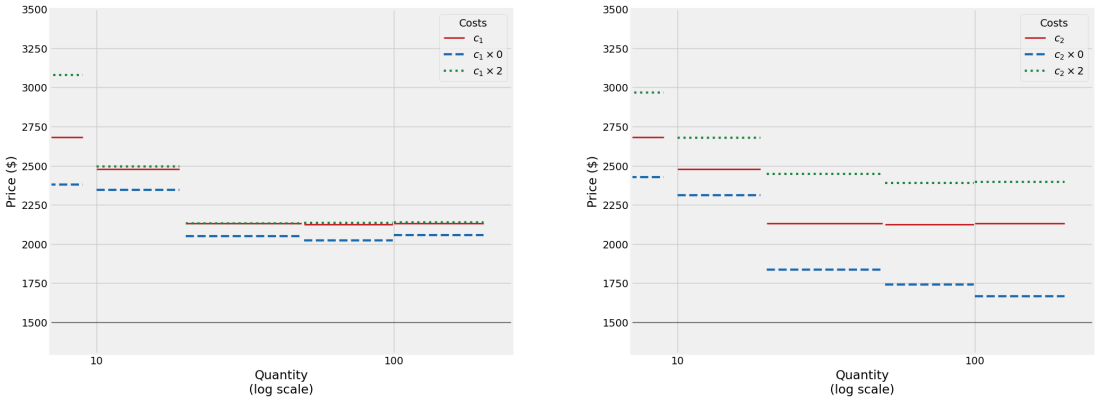


Figure 15: Counterfactual optimal price schedule under no cost (left panel) and uniform size-fixed-effects in value (right panel).

7.3 Detailed Analysis of Cost-Side Factors

In this section, we examine the way in which cost parameters c_1 and c_2 shape the optimal price schedule. Recall that c_1 ≈\$3,490/customer and c_2 ≈\$748/workshop. Figure 16 depicts the optimal contracts as these parameters change. Panel (a) shows the optimal contract for a range of c_1 values while panel (b) does the same for c_2.



(a) Different c_1s

(b) Different c_2s

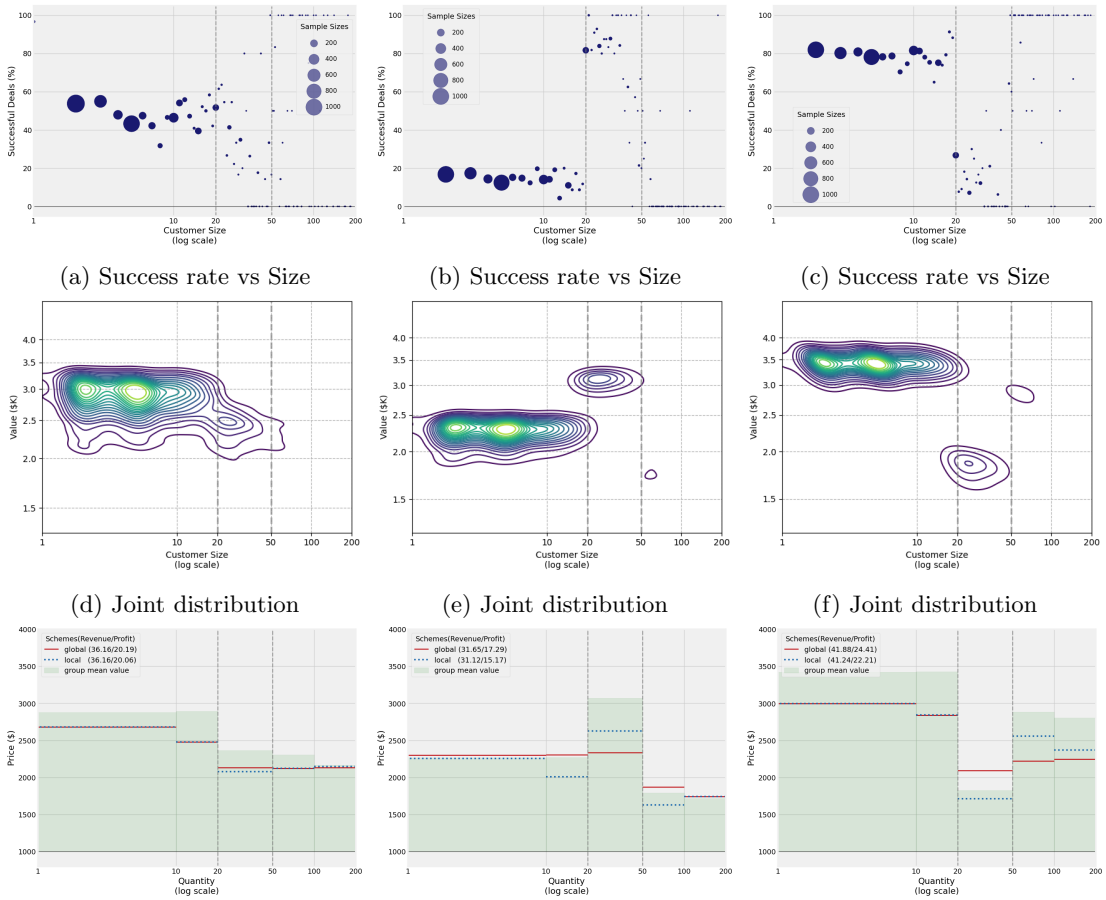
Figure 16: Price Schemes under Different Cost Parameters

As evident by panel (a) of this figure, and perhaps unsurprisingly, changes in c_1 mostly impact the marginal price for smaller size deals. This is because the larger the deal size, the smaller the customer-level fixed cost will be as a fraction of the total cost associated with the customer. It is worth noting that even with $c_1 = 0$, the optimal contract roughly preserves its shape. This indicates that demand side factors (i.e., the shape of the joint distribution $f(\cdot, \cdot)$) have a critical role in shaping the optimal price schedule. To see why this is true, note that the cost-side factor for the decreasing marginal prices is c_1. Thus, if this shape is preserved without c_1, it has to be attributed to the demand side.

The effect of c_2 on the optimal pricing schedule is shown in panel (b) of the figure. Prices increases across the board as the per-unit cost increases, and the pass-through is more or less uniform and around $\frac{1}{3}$. Observe that the direction in which c_2 affects the shape of the optimal schedule is in contrast to that of c_1. The larger the marginal cost c_2, the closer the optimal schedule is to a linear contract. On the other hand, with small c_2, the seller has more freedom to tailor the schedule based on how customers of different sizes value the product differently.

7.4 Detailed Analysis of Demand-Side Factors (and a Revisit of the Role of Incentives)

In this section, we dive deeper into the role of demand-side factors in shaping the optimal contract. To state the bottom-line result first, our analysis has two main messages. First, it is the entire shape of the joint distribution $f(\cdot, \cdot)$ that matters for optimal pricing, as opposed to a simple correlation between size $\bar{q}$ and value v across customers i. Second, under some shapes of the joint distribution $f(\cdot, \cdot)$, ignoring incentive compatibility (i.e., approximating the "global" optimal contract with the "local" one as previously discussed) may bring about substantial losses.



(a) Success rate vs Size (b) Success rate vs Size (c) Success rate vs Size

(d) Joint distribution (e) Joint distribution (f) Joint distribution

(g) Price schemes (h) Price schemes (i) Price schemes

Figure 17: Analysis of optimal pricing under three scenarios: Left column: original data; middle column: counterfactual data with higher acceptance rates for mid-size deals compared to small and large ones; right column: counterfactual data with lower acceptance rates for mid-size deals compared to small and large ones. For details of counterfactual data simulation, refer to appendix E.

We find it more illustrative to start the counterfactual analysis in this section from the data rather than from (the later stage of) the estimated model. In one counterfactual, we modify our original dataset so that mid-size deals ($q \in [20, 50)$, equivalent to I_3) have a meaningfully higher acceptance rate than small ($q < 20$, equivalent to $I_1 \cup I_2$) and large ($q \geq 50$, equivalent to $I_4 \cup I_5$) ones. In a second counterfactual, we do the opposite. For each counterfactual, we re-estimate the model (i.e., the joint distribution $f(\cdot, \cdot)$), and re-compute both the global and local optimal price schedules. The two columns of Figure 17 depict these two counterfactuals. We now turn to analyzing these results based on the two major lessons mentioned above.

Correlation	**Left col of Fig 17**	**Mid col of Fig 17**	**Right col of Fig 17**
Success & $\bar{q}$	-0.06^{***}	0.13^{***}	-0.13^{***}
$\hat{v}$ & $\bar{q}$	-0.38^{***}	0.04^{***}	-0.44^{***}

Table 4: Correlation between deal success and size, (as well as value and size) for each of the three scenarios of Figure 17

Importance of the shape of $f(\cdot,\cdot)$ beyond simple correlation: As can be seen from panels (d), (e), and (f) of Figure 17, the flexibility of the shape of $f(\cdot,\cdot)$ in our model allows us to capture the nature of the relationship between deal accept/reject outcome and deal size in a comprehensive manner. All of the estimated distributions in Figure 17 have flexible forms and multiple local peaks. Such flexibility is not feasible to capture with more restrictive models of $f(\cdot,\cdot)$ or with simple measures such as correlation. See Table 4 for illustration: the correlation between v_i and $\bar{q}_i$ across i in the original model, left column of Figure 17, and right column of Figure 17, are -0.38 and -0.44 respectively. Although these correlations are fairly similar to one another, the shapes of the estimated distribution and the optimal contracts are meaningfully different. There are at least two reasons why such simple correlational measures cannot replace a flexible structure for $f(\cdot,\cdot)$.

First, as can be seen from Figure 17 the relationship between v and $\bar{q}$ need not be monotone. And non-monotone relationships that are fundamentally different from one another (such as U-shaped and inverse U-shaped) may look similar when assessed using a linear model (e.g., correlation).

Second, and more crucially, the importance of one sample in the estimation procedure might differ from its importance with respect to the counterfactual policy analysis of interest. For instance, consider the smaller local peaks in the distribution $f(\cdot,\cdot)$ in panels (e) and (f) of Figure 17. They belong to mid- and large-size deals. Note that these peaks are substantially lower compared to the corresponding peak(s) for small-size deals. This is simply because $f(\cdot,\cdot)$ is a probability distribution and there are far fewer mid- and large-size samples in the data compared to smaller ones. As a result, if we impose a restrictive model of $f(\cdot,\cdot)$ that does not adequately separate the estimation of v across sizes, the weight of these fewer samples will be dwarfed in the estimation procedure by substantially more frequent small-size deals. This, in turn, would bias our estimation of average v for these less frequent deal sizes. Such a bias would be detrimental to our counterfactual analysis. This is because as depicted by Figure 3 earlier in the paper, larger deals, although much less frequent, have a meaningful role in shaping the revenue and profitability.

In sum, we find the analysis summarized by Figure 17 to be in support of our modeling choices on where to be parsimonious (shape of $V_i(\cdot)$) and where to be flexible (shape of $f(\cdot,\cdot)$).

Role of incentive compatibility constraints: The bottom panels of Figure 17 depict not only the "globally optimal" schedule P^* for each demand scenario, but also the "locally optimal" schedule $\tilde{P}$ as defined in section 7.1. There are two broad lessons to learn from this figure.

The first lesson is that under substantially different valuations–put differently, under substantial heterogeneity in acceptance rates–across size groups, the profitability gap between the local and global optima may be wide. This is because the locally optimal marginal prices end up being far from each other, exacerbating the loss from the lack of incentive compatibility. To see this, note that $\frac{\pi(P^*)-\pi(\tilde{P})}{\pi(P^*)}$ is equal to 1% is equal to 1% in the original data, whereas it is 11% for the demand scenario in the middle column and around 9% in the right-most column (the profit and revenue levels are posted on the top left corner of bottom panels in Figure 17).

The second important lesson is about how the *shapes* of the two contracts compare to one another. As can be seen from the bottom panels of the figure, the globally optimal contract seems to cover a shorter range of prices. Take the second column of the figure for illustration. In general, relative to the locally optimal contract $\tilde{P}$, the (globally) optimal contract P^* seems to *moderate* the price variation by size. Take the second column of the figure for illustration. Here, p_3^* is substantially smaller than $\tilde{p}_3$,

even though mid-size customers (i.e., those with $\bar{q}_i \in I_3$) have much larger average valuations v_i relative to other sizes. The (globally) optimal contract P^* does this moderation in order to discourage some of the mid-size customer from "fleeing to other sizes". For the same reason, P^* increases the prices in adjacent size ranges in order to make them less attractive to incoming mid-size customers (or, to make more profits off of the mid-size customers who change their purchase sizes anyway). Conceptually similar arguments (but in part in the opposite direction) hold for the right column of the figure.

Table 5 should shed more light on the matter. In this table, we compare the profitabilities of P^* and $\tilde{P}$ *from different size-groups*. In particular, we break down $\pi(P^*) - \pi(\tilde{P})$ for customers that are small ($\bar{q}_i < 20$), mid-size ($\bar{q}_i \in [20, 50)$), and large ($\bar{q}_i > 50$).

Scenario	**Customer Size Group**	**Total Diff Profit**	**Avg Diff in Purchased Units**
Mid col of Fig 17	small (<20)	−\$0.35M	−0.46
	medium (20-49)	+\$2.57M	+0.36
	large (≥50)	−\$0.10M	−8.08
Right col of Fig 17	small (<20)	+\$0.85M	−0.36
	medium (20-49)	−\$0.59M	−6.35
	large (≥50)	+\$1.93M	+9.10

Table 5: global vs local by size segment

As the table shows, for the demand system in the middle column of Figure 17, P^* delivers \$2.57M/y more profit from medium-size customers, relative to $\tilde{P}$. This happens because in the vector P^*, the element p_3^* is only moderately larger than the other elements, as opposed to the wider gap between $\tilde{p}_3$ and other elements of $\tilde{P}$. This prevents many mid-size customers from "flocking" to cheaper quantities, thereby boosting the profitability from those customers. Also helpful toward this objective is the fact that p_2^* and p_4^* are increased (relative to their $\tilde{p}$ counterparts); this either further prevents mid-size customers from adjusting purchase sizes, or helps make a higher profit off of those mid-size customers who do adjust anyway. Of course this means that the tuning of p_2^* and p_4^* is done in part with the goal of taming mid-size customers' behavior; which means those prices are locally sub-optimal. Thus, it should not be surprising that P^* delivers less profit than $\tilde{P}$ from small and large customers, by the amounts of \$0.35M/y and \$0.10M/y respectively.

Similarly, as Table 5 shows, P^* makes \$0.59M/y less profit than $\tilde{P}$ from mid-size customers under the demand system in the right-most column of Figure 17. This is because globally optimal P^* charges a substantially higher per-unit price for mid-size deals than mid-size customers are willing to pay. Though this leads to losses from those customers, it prevents smaller and larger customers from taking advantage of low rates in the mid-range. Alongside this, P^* also charges lower prices outside of mid-size deals to further discourage customers of other sizes from moving. As a result, the profitability from those sizes, which is compromised under $\tilde{P}$ due to not accounting for incentive compatibility issues (i.e., large and small customers taking advantage of mid-size rates), are partially protected under P^*.

To recap, compared to the locally optimal schedule, the globally optimal one seems to charge prices that vary less significantly with size.

7.5 Third Degree Price Discrimination

In this section, we study third-degree price discrimination. In particular, we consider two kinds: (i) third-degree price discrimination based on customer sizes, and (ii) combining second-degree price discrimination (based on size) with third-degree discrimination (based on other observables).

7.5.1 Third-degree discrimination based on customer size

In the discussion of incentive compatibility constraints, we introduced the "locally optimal" contract $\tilde{P}$. We argued this price schedule is "naive" in that it fails to anticipate the ability by each customer of size $\bar{q}_i \in I_k$ to tune its purchase size in order to take advantage of other, lower, prices $p_{k'}$. In this section, we ask "what if such anticipation is indeed correct?" In other words, what if the seller is able to forces customers of each size group k to only pay p_k per unit no matter how many unites q they purchase? This means the firm can third-degree price discriminate based on size $\bar{q}$ and need not worry about incentive compatibility. Under these conditions, $\tilde{P}_k$ would indeed become the true optimal price schedule. The profit here, will be equal to $\Sigma_k \pi_k(\tilde{p}_k)$ where "local" profit functions $\pi_k(\cdot)$ are defined as in equation 8.

Note that due to the relaxation of the incentive constraints, the third-degree-discrimination profit $\Sigma_k \pi_k(\tilde{p}_k)$ is larger than $\pi(\tilde{P})$ and, likely, also than the second-degree-discrimination profit $\pi(P^*)$. Table 6 empirically analyzes these profits under the three data scenarios described in Figure 17 (recall that one scenario is the original data and the other two are counterfactual datasets, modifying the demand system).

	Original	**(i)**	**(ii)**
Global Optimal Pricing	20.19	17.28	24.41
Local Optimal Pricing	20.06	15.18	22.21
	(-0.64%)	(-12.15%)	(-9.01%)
Third-Degree PD (by Size)	20.29	18.54	25.62
	(+0.50%)	(+7.29%)	(+4.96%)

Notes: Columns (i) & (ii) indicate profits for Figure 17 middle and right columns. (i) is produced from the scenario with the highest success rate in the median group (15%/87%/14%) while (ii) is from the scenario with the lowest success rate in the median group (80%/14%/80%).

Table 6: Profits from Global and Local Optimal Pricing and Third-Degree PD by Size

As the left column of this table shows, size-based third degree price discrimination delivers little to no extra profitability above second-degree discrimination (only by 0.5%). This should not be too surprising given that the main advantage of size-based third-degree discrimination is the relaxation of incentive-compatibility constraints. But as first reported in figure 14 and repeated in the middle row of table 6, in the original dataset the profit impact of incentive compatibility constraints was rather slim (around 0.64%). Looking at alternative data scenarios (mid and right columns in Table 6) confirm our intuition: the benefit of third-degree price discrimination based on size is large (small) if the loss from charging the locally optimal contract instead of the global one is large (small). Formally, there seems to be a strong and positive association between $\pi(P^*) - \pi(\tilde{P})$ on the one hand and $\left(\Sigma_k \pi_k(\tilde{p}_k)\right) - \pi(P^*)$ on the other.

7.5.2 Combining second and third degree discrimination

Another approach to third degree price discrimination would be to combine it with second degree. More precisely, the firm can rank customers based on their βX_i from equation 5 in a descending order and create equally-sized groups $j \in \{1, ..., J\}$. The firm can then offer each group j a separate optimal price schedule $P^{*j}(\cdot)$. Given the specification of β, this means customers who were established earlier, or those with higher amounts of behavioral feature 1 will be put in higher-ranked bins and face higher prices.[8]. Figure 18 shows the optimal strategy if the firm divides all the customers into $J = 2$ groups based on ranked $X_i\beta$ values. The original (i.e., one-group) optimal schedule P^* has also been plotted for comparison. As can be seen from the figure, the higher-willingness-to-pay group $j = 1$ gets charged an additional \$200 or more per workshop (across different sizes) relative to the lower-willingness-to-pay group $j = 2$.

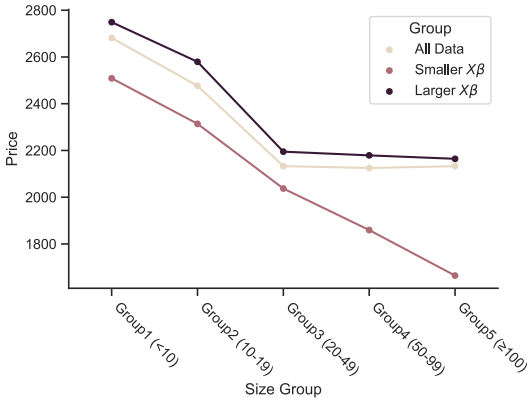


Figure 18: Optimal price schedules P^{*1} and P^{*2} when customers i are divided into $J = 2$ groups based on ranked $X_i\beta$ values

We close this section by a discussion of the profitability of third-degree price discrimination. Figure 19 plots the profitability of the optimal pricing as a function of the "extent of third degree discrimination J". Under purely second-degree discrimination ($J = 1$), total profit is $\$20.19M/y$ whereas under substantial third-degree discrimination $J = 9$, the profitability is around $\$21.06M/y$. This profit surpasses the second-degree-discrimination profit by only 4.31%. As a result, third-degree discrimination, though non-trivially useful when deployed alone or in conjunction with second-degree, does not seem to generate substantial extra profitability above sole second-degree discrimination.

[8] Of course many other customer characteristics can be incorporated into this. But as mentioned before, in our specific contexts, many potentially relelvant features turned out to be of little impact empirically.

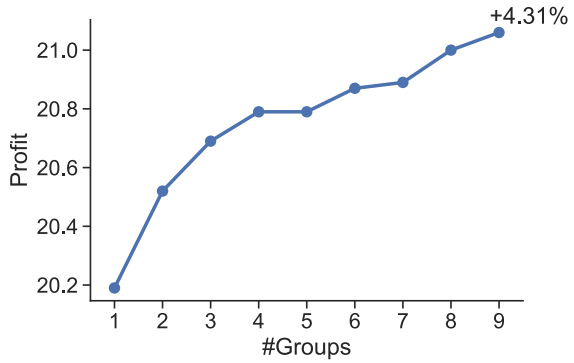


Figure 19: Total Profit from optimal 2nd+3rd degree price discrimination as a function of the number of groups J

We do not find the less-than-stellar performance of third-degree price discrimination in our context surprising. This is because as mentioned before in the process of variable selection for the modeling of v (see equation 5), we found that many potentially relevant variables (especially industry fixed effects) have little explanatory power. We do not expect this empirical observation to be generalizable. In other contexts where observable customer characteristics are highly predictive of purchasing power (see Dubé and Misra (2017) for instance), third degree price discrimination may even closely approximate the profitability of first-degree discrimination.

8 Discussion

This section provides discussions on broad level applications of our framework as well as caveats and future research.

8.1 General Applicability

The main purpose of this paper was to provide an empirical methodology to optimize their nonlinear pricing strategy. The key feature of our data that made this possible was detailed information (including the size of the deal) not only on deals that succeeded, but also on those that failed. Even though to the best of our knowledge the use of a dataset with this feature is novel in scholarly work on pricing, we believe such data is increasingly collected and maintained by firms. As a result, our framework can be readily applied by all such firms as long as they are currently employing a nonlinear price schedule that provides variation in marginal prices encountered by different consumer types. Even if a firm currently uses a linear pricing scheme, a small-scale experiment can provide the necessary variation and allow our methodology to be implemented.

In the specific case of this study, our results did lead to a change in LifeLabs' pricing strategy that is qualitatively in line with our recommendation, though not fully coinciding with it. More specifically, they reduced their prices as our study recommends; but did so only for larger deals.[9] See Figure 20.

[9]Firms adjusting their pricing strategies in the direction of (but not exactly coinciding with) recommendations from academic studies is not unprecedented. An instance is Dubé and Misra (2017) and the impact of their study on Ziprecruiter's pricing.

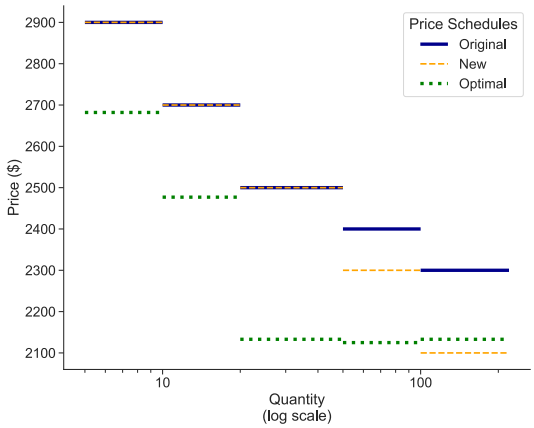


Figure 20: Comparison among: Lifelabs' original pricing, optimal schedule P^*, and Lifelabs' adjusted prices based on the recommendation of P^*.

8.2 Limitations and Future Research

We discuss two avenues for future research here. Of course this list is non-exhaustive.

8.2.1 Allowing for smoothly diminishing rates of return to customers

Our formulation of gross value is $V_i(q) \equiv v_i \times \min(q, \bar{q}_i)$. Though this captures the two key dimensions of heterogeneity that we are focused on in this paper (i.e., size and value,) it does abstract away from smoothly diminishing rate of return. That is, the marginal value for another unit of the product is initially v_i and it abruptly reduces to zero. A more general formulation would allow for smoothly diminishing returns to scale on the part of the buyer. We do examine a form of such an extension in appendix C, and show that our results do not change drastically as a result of this extension.

That said, in our extension in that appendix, we take the estimated distribution $f(\cdot, \cdot)$ over $\bar{q}_i$ from our main analysis, and then add another parameter to the value functions in order to capture the degree of smoothness. A natural next step for future research would be to incorporate this smoothness parameter in the estimation stage rather than add it after the initial, non-smooth, model has been estimated. It can be expected (and we showcase in the appendix) that in the smooth model, the purchase size need not coincide with $\bar{q}_i$ even under a concave contract. Similarly, it is less reasonable to assume that the recorded would-be size for unsuccessful deals corresponds to $\bar{q}_i$.[10] As a result a new estimation process would be necessary in order to incorporate smoothness.

8.2.2 Endogenous set of potential customers

Second, we assume that the set of deal-talks is exogenous. Put in the terminology of our timeline in Figure 6, we assume that participating in a conversation with the seller, which leads to a record of $\bar{q}_i$, is costless and is, hence, done by any firm could benefit from the product (i.e., any firm with $\bar{q}_i > 0$). Assumptions similar to this have previously been made in the discrete choice literature (see Cohen et al.

[10] One possibility would be to assume this quantity represents the number of units that the buyer would expect to buy conditional on its uncertainty being resolved in a way that would make non-zero purchase optimal.

(2016) and Dubé and Misra (2017) for instance.) Nevertheless, this assumption is not ideal, neither in the discrete-choice setting nor in our setting. It could be that the firm more aggressively markets its products to some types of customers, depending (among other things) on what price those customers are more likely to see. Also, it could be that potential customers decide whether to start talking to the firm and learning about its products if they find it likely enough that they eventually buy. Both of these possibilities can be modeled using a positive (and perhaps heterogeneous) cost of communicating with the seller.

In spite of these issues, we believe our results are reasonable. This is because we know (through interviews with the managers of the firm) that most customers reach out to the firm rather than the other way around. As a result, the effect of differential marketing activities should not be large. Last, and perhaps more importantly, potential customers who did not initiate a conversation are likely those with lower v values. As a result, missing them from the analysis would *overestimate* the overall optimal prices. Given that our current recommendation to the company is to lower their prices, our recommendation would only be strengthened if we were able to observe potential customers who did not start a conversation about a deal.

9 Conclusion

This paper empirically analyzed optimal nonlinear pricing. We proposed a model of demand in continuous choice settings which captures the notion of "customer size." We estimated the joint distribution of customer size and customer per-unit willingness to pay by leveraging a novel dataset that records information (including prospective deal size) not only for successful deals but also for unsuccessful ones. We then used the estimated model to solve for the optimal nonlinear tariff.

We find that optimal nonlinear pricing improves upon the profitability of optimal linear pricing by about 7.9%. Nevertheless, this second-degree price discrimination method recovers only about 7.4% of the profitability gap between linear pricing and first-degree price discrimination. We also find that second degree price discrimination improves consumer welfare by about 1% and social welfare by about 5.5%.

We conducted further counterfactual analyses in order to generate general insights above and beyond our specific application. Among other analyses, we (i) examined the role of cost side v.s. demand side factors in shaping the optimal contract, (ii) quantified the magnitude of the profit impact of incentive compatibility constraints, and (iii) simulated the profitability of combining second and third degree price discrimination strategies.

We believe our analysis can be broadly applicable by firms that seek to optimize a nonlinear pricing tariff. A pre-requisite for the possibility to implement our method would be maintenance of sufficiently detailed data not only on successful deals but also on unsuccessful ones.

The main direction in which our analysis may be extended would be adding smoothness to the value functions by customers. Another potential future direction would be to extend the analysis from a monopolistic setting to an oligopolistic one.

References

Anderson, E. T. and J. D. Dana Jr (2009). When is price discrimination profitable? *Management Science 55*(6), 980–989.

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica: Journal of the Econometric Society*, 51–75.

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica 85*(2), 453–488.

Chan, T., V. Kadiyali, and P. Xiao (2009). Structural models of pricing. In *Handbook of pricing research in marketing*, pp. 108–131. Edward Elgar Publishing.

Cohen, P., R. Hahn, J. Hall, S. Levitt, and R. Metcalfe (2016). Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research.

Derdenger, T. and V. Kumar (2013). The dynamic effects of bundling as a product strategy. *Marketing Science 32*(6), 827–859.

Devanur, N. R., N. Haghpanah, and A. Psomas (2020). Optimal multi-unit mechanisms with private demands. *Games and Economic Behavior 121*, 482–505.

Draganska, M. and D. C. Jain (2006). Consumer preferences and product-line pricing strategies: An empirical analysis. *Marketing science 25*(2), 164–174.

Dubé, J.-P. and S. Misra (2017). Scalable price targeting. Technical report, National Bureau of Economic Research.

Frazier, P. I. (2018a). Bayesian optimization. In *Recent advances in optimization and modeling of contemporary problems*, pp. 255–278. Informs.

Frazier, P. I. (2018b). A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Ghili, S. (2022). A characterization for optimal bundling of products with non-additive values. *arXiv preprint arXiv:2101.11532*.

Goldberg, S. (2021). Designing monitoring programs.

Haghpanah, N. and J. Hartline (2021). When is pure bundling optimal? *The Review of Economic Studies 88*(3), 1127–1156.

Hendel, I. and A. Nevo (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review 103*(7), 2722–2751.

Iyengar, R. and S. Gupta (2009). Nonlinear pricing. In *Handbook of pricing research in marketing*, pp. 355–383. Edward Elgar Publishing.

Iyengar, R. and K. Jedidi (2012). A conjoint model of quantity discounts. *Marketing Science 31*(2), 334–350.

Iyengar, R., K. Jedidi, and R. Kohli (2008). A conjoint approach to multipart pricing. *Journal of Marketing Research 45*(2), 195–210.

Kadiyali, V., N. J. Vilcassim, and P. K. Chintagunta (1996). Empirical analysis of competitive product line pricing decisions: lead, follow, or move together? *Journal of Business*, 459–487.

Laffont, J.-J., E. Maskin, and J.-C. Rochet (1987). Optimal nonlinear pricing with two-dimensional characteristics. *Information, Incentives and Economic Mechanisms*, 256–266.

Leslie, P. (2004). Price discrimination in broadway theater. *RAND Journal of Economics*, 520–541.

Luo, Y., I. Perrigne, and Q. Vuong (2018). Structural analysis of nonlinear pricing. *Journal of Political Economy 126*(6), 2523–2568.

Maskin, E. and J. Riley (1984). Monopoly with incomplete information. *The RAND Journal of Economics 15*(2), 171–196.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic theory 18*(2), 301–317.

Narayanan, S., P. K. Chintagunta, and E. J. Miravete (2007). The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quantitative Marketing and economics 5*, 1–34.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The computer journal 7*(4), 308–313.

Nevo, A., J. L. Turner, and J. W. Williams (2016). Usage-based pricing and demand for residential broadband. *Econometrica 84*(2), 411–443.

Reiss, P. C. and M. W. White (2005). Household electricity demand, revisited. *The Review of Economic Studies 72*(3), 853–883.

Rochet, J.-C. and L. A. Stole (2002). Nonlinear pricing with random participation. *The Review of Economic Studies 69*(1), 277–311.

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

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

Verboven, F. (2002). Quality-based price discrimination and tax incidence: evidence from gasoline and diesel cars. *RAND Journal of Economics*, 275–297.

Wilson, R. B. (1993). *Nonlinear pricing*. Oxford University Press on Demand.

Yang, F. (2021). Costly multidimensional screening. *arXiv preprint arXiv:2109.00487*.

Yin, L., H. Zhang, X. Zhou, X. Yuan, S. Zhao, X. Li, and X. Liu (2020). Kaml: improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome biology 21*(1), 1–22.

Appendix

This appendix provides multiple complementary analyses to the paper. Section A provides the proof to lemma 1. Section B analyzes the robustness of the results to our approximation whereby we treated the observed price schedule by LifeLabs as concave. In another robustness check, section C studies the optimal price schedule and its profitability and welfare under the assumption that the gross valuation functions $V_i(\cdot)$ are smooth rather than piecewise linear. Section D compares our grid-bisection optimization method to our alternatives and provides recommendations on algorithm choice for researchers studying similar problems. Section E describes the process by which our counterfactual data (visually depicted in figure 17 of section 7) are generated.

A Proof of Lemma 1

As a reminder:

$$q^*(P|\bar{q}, v) := \arg\max_{q \geq 0} \big[V(q|\bar{q}, v) - P(q)\big]$$

where $V(q|\bar{q}, v) := \min(q, \bar{q}) \times v$.

Note that $V(q|\bar{q}, v)$ is constant in q for $q \geq \bar{q}$ whereas $P(q)$ is strictly increasing. Thus, no quantity $q > \bar{q}$ can be in the arg max. As a result, we can rewrite:

$$q^*(P|\bar{q}, v) := \arg\max_{q \in [0, \bar{q}]} \big[V(q|\bar{q}, v) - P(q)\big]$$

But within the $[0, \bar{q}]$ interval, the value function can be written as: $V(q|\bar{q}, v) := q \times v$ which yields:

$$q^*(P|\bar{q}, v) := \arg\max_{q \in [0, \bar{q}]} \big[q \times v - P(q)\big]$$

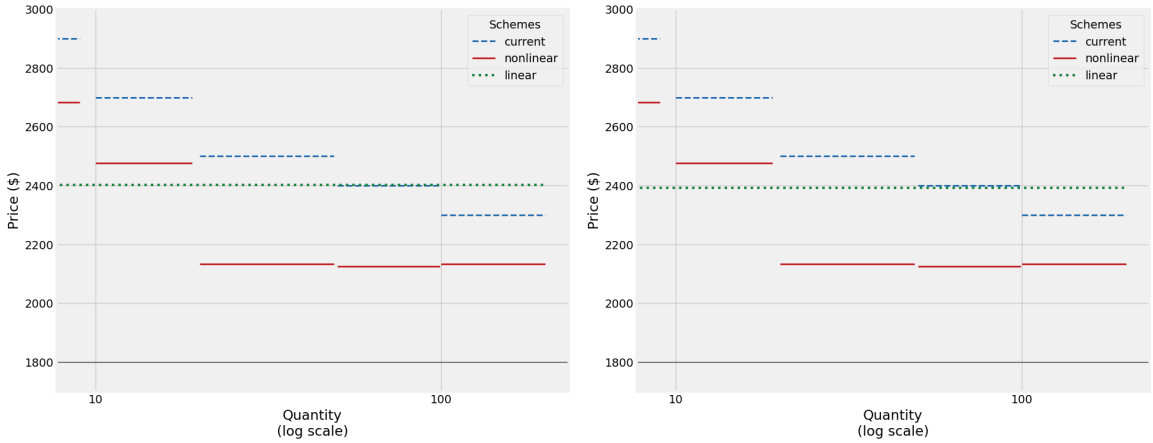
Note that by strict concavity of $P(q)$ and linearity of $q \times v$, the function $q \times v - P(q)$ is strictly convex. This implies that its global maximum on the interval has to be an extreme point of the interval. That is: $q^*(P|\bar{q}, v) \subset \{0, \bar{q}\}$. **Q.E.D.**

B Robustness to the treatment of observed schedule as increasing and concave

In estimating the model, we leveraged Lemma 1 which stated that the observed q for each customer i would have to equate $\bar{q}_i$ if the observed price schedule $P(\cdot)$ charged by the firm is weakly concave. As Figure 1 depicts, however, the observed price schedule is not concave. It does have a decreasing slope but it also have some discontinuous downward jumps. In this appendix, we show that the schedule is indeed approximately concave. That is, these jumps are sufficiently small for the result to be robust to them.

More precisely, we re-estimate the model and the optimal pricing analysis without using the data points i for whom the observed q_i may be different from $\bar{q}_i$ due to non-concavities in the price schedule. Those are basically all data points that fall on the "dips" of the observed schedule, i.e., those observations i with q_i such that $\exists q < q_i$ with $P(q) > P(q_i)$, where $P(\cdot)$ is the observed price schdeule.

This new sample is slightly smaller than the original one (2,585 datapoints instead of 2,686). Based on this sample, the optimal price schedule and the profitability measures are re-computed and depicted in figure 21 and table 7.



(a) Marginal Prices (restricted sample)

(b) Marginal Prices (full sample)

Figure 21: Optimal price schedules under restricted and full sample. As can be seen from this figure, differences between the two are negligible.

Scheme	*Revenue* *($M)*	change (%)	*Profit* *($M)*	change (%)	*Consumer* *Welfare* *($M)*	change (%)	*Social* *Welfare* *($M)*	change (%)
current	27.85	-	16.52	-	6.09	-	22.61	-
1st degree	56.03	+101.18%	34.87	+111.08%	0	−100.00%	34.87	+54.22%
linear	31.83	+14.29%	16.82	+ 1.82%	9.26	+ 52.05%	26.08	+15.35%
nonlinear	32.75	+17.59%	18.11	+ 9.62%	9.14	+ 50.08%	27.26	+20.57%

Table 7: Profit and Welfare Analysis

As can be seen from comparing these results to the corresponding ones from the main text, there is little difference caused by the few points for the observed purchase size and customer size differ.[11] This is especially the case for the optimal price schedule, where the differences are not even large enough to be discernible when visually plotted.

[11] Perhaps the most notable difference is that profit and revenue levels are generally slightly lower here compared to the results from the main text. This should not be surprising given that by removing some buyers from the sample, we are also removing the profit and revenue they generate for the seller.

C An analysis of smooth value functions

As mentioned in the main text of the paper, we make the simplifying assumption that value functions take the form $V_i(q) \equiv v_i \times \min(q, \bar{q}_i)$. In this section, we explore smoothing these value functions to examine the effects of such smoothness on the shape of the optimal schedule as well as the welfare effects.

Formally, we modify the value function from $V_i(q) \equiv v_i \times \min(q, \bar{q}_i)$ to:

$$V_i(q) \equiv v_i \times \gamma(q, \bar{q}_i) \tag{10}$$

where the function $\gamma(\cdot, \cdot)$ provides a "fuzzy" or smooth form of taking the minimum between its two arguments. More formally, for some b and m:

$$f(\bar{q}, q) \equiv (\bar{q}^{-b} + mq^{-b})^b \tag{11}$$

To illustrate the working of this function, figure 22 depicts the behavior of $\gamma(\bar{q}, q)$ as a function of q when $\bar{q} = 200$, $m = 0.73$, and $b = 10$. As can be seen from this figure, γ tends to $\min(\bar{q}, q)$ as $b \to \infty$. Also, γ tends to 0 as $b \to^+ 0$. m determines how early the smoothed function converges to $\bar{q}$. Parameter b measures the degree of smoothness of the function. Given any value of b parameter m are calibrated so that the difference of resulting smoothed minimum $\gamma(\bar{q}, q) - \min(\bar{q}, q)$ integrates to zero over q. Larger values for b would mean less smooth functions (and would require m closer to 1 for the difference between the functions to integrate to zero).

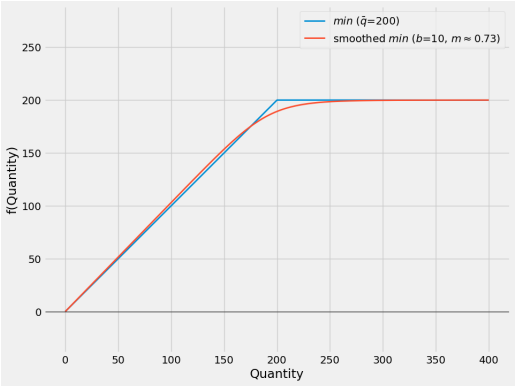
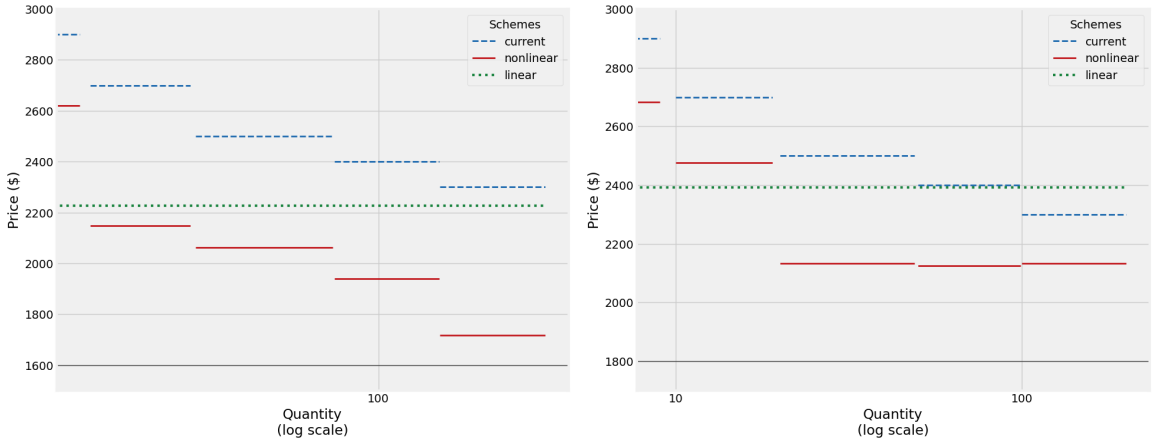


Figure 22: *min* and smoothed *min* comparison

We next produce the optimal price schedule and welfare results using the smooth min function γ described above (under $m = 0.73$, and $b = 10$) instead of the non-smooth min function.

Figure 22 compares the optimal linear and nonlinear schedules under the two settings. As can be seen from this figure, the general shape of the optimal nonlinear contract is preserved under the alternative assumption on the smoothness of the value functions.



(a) Marginal Prices (smooth minimum function) (b) Marginal Prices (non-smooth minimum function)

Figure 23: Optimal price schedules under smooth and non-smooth (original) minimum functions.

Additionally, table 8 describes the analysis of profits and welfare under smooth value functions. As can be seen in this table, these results too are fairly robust to smoothness.

Scheme	*Revenue* *($M)*	change (%)	*Profit* *($M)*	change (%)	*Consumer* *Welfare* *($M)*	change (%)	*Social* *Welfare* *($M)*	change (%)
current	26.46	-	15.23	-	5.87	-	21.09	-
1^{st} *degree*	61.99	+134.28%	38.78	+154.63%	0	−100.00%	38.78	+83.88%
linear	32.49	+22.79%	16.26	+ 6.76%	10.48	+ 78.53%	26.74	+26.79%
nonlinear origin	32.70	+23.58%	16.90	+ 10.97%	10.19	+ 73.59%	27.09	+28.45%

Table 8: Profit and Welfare Analysis Under Smooth Value Functions

D Comparison between our optimization method and other alternatives

This appendix briefly compares the grid-bisection optimization method used in our paper to other alternatives we considered. In particular, we examine the following three approaches: 1) Bayesian Optimization a la Frazier (2018a,b), 2) Gradient method with the initial candidate for P^* chosen from an initial grid of 5^d possible $P \in [0, 3000]^d$ vectors, and 3) A simplex-based direct search method a la Nelder and Mead (1965) with the initial candidate for P^* chosen from an initial grid of 10^d possible $P \in [0, 3000]^d$ vectors. We do this comparison in three scenarios: (i) solving for the optimal linear price (d=1), (ii) solving for the optimal price when we have three marginal costs for small (i.e., $\bar{q}_i \in I_1 \cup I_2$), medium-size (i.e., $\bar{q}_i \in I_3$), and large (i.e., $\bar{q}_i \in I_4 \cup I_5$)customers, i.e., $d = 3$, and finally (iii) the optimal price P^* in our main 2nd-degree price discrimination setting which was five-dimensional $d = 5$. Table 9 presents the results. It presents both the objective value (i.e., profit) and runtime of each algorithm. Note that in addition to the methods whose performances are reported in this table, we examined a number Gradient-method alternatives, such as "conjugate gradient", "BFGS", "Truncated newton algorithm",

and "Powell algorithm". However, due to weak performance, we do not report them here.

Method		Marginal Price ($)	Profit ($M)	Time (sec)
Grid-Bisection	linear	[2392]	18.71	7
	nonlinear 3	[2584 2132 2136]	20.07	276
	nonlinear 5	[2682 2477 2133 2125 2133]	20.19	1243
Bayesian Optimization	linear	[2392]	18.71	25
	nonlinear 3	[2578 2134 2132]	20.06	342
	nonlinear 5	[2696 2456 2098 2079 1810]	20.07	1194
Nelder-Mead 5	linear	[2390]	18.71	3
	nonlinear 3	[2590 2077 2074]	20.08	14
	nonlinear 5	[2694 2456 2072 2034 1835]	20.13	157
Nelder-Mead 10	linear	[2390]	18.71	3
	nonlinear 3	[2585 2133 2130]	20.08	55
	nonlinear 5	[2656 2443 2077 2076 2075]	20.16	3006

Table 9: Comparative analysis of the performances of different optimization algorithms

Discussion and recommendations: Based on the results presented in table 9, the grid-bisection approach outperforms consistently outperforms other candidates with respect to the objective function. It also outperforms the Bayesian method on the front of the algorithm run-time. The comparison between Grid-Bisection and the Nelder Mead simplex approach (Nelder and Mead, 1965) is more nuanced. In the case of $d = 5$, which is the main case we analyze, Grid-Binary takes shorter to run than Nelder-Mead with an initial grid of 10^5 but takes much longer if compared instead to Nelder-Mead with an initial grid of size 5^5.

Given the above summary, we recommend that Grid-Bisection be used for similar empirical multi-dimensional-screening problems if (i) precision is of utmost importance, and (ii) the number of dimensions is not too high (e.g., $d = 5$). For higher numbers of dimensions (e.g., $d = 10$ or larger), we conjecture that the Nelder-Mead approach with an initial grid that is not too large will make the best trade-off between accuracy and runtime.

E Details on data simulation

This section provides details on how the counterfactual datasets for the middle and right-most columns of Figure 17 were produced.

To produce the counterfactual data in the middle column, we alter the original data in a way that increases the deal acceptance rate in the for mid-size customers (i.e., $\bar{q}_i \in I_3$) while decreasing the acceptance rates for smaller ($\bar{q}_i \in I_1 \cup I_2$) and larger ($\bar{q}_i \in I_4 \cup I_5$) ones. More specifically, for each row of the data (i.e., each it combination), we alter the observed value s_{it} for deal success status with probability p (i.e., according to a random draw from a Bernoulli distribution with parameter p). If the Bernoulli draw is 1, (that is, if we are set to alter s_{it}), then we alter it to 1 if customer i is mid-size and to 0 otherwise. If the Bernoulli draw is 0, however, we do not alter s_{it}.

The procedure for generating the counterfactual data for the right-most column of the figure is similar, except that the s_{it} values for mid-size deals are altered to 0 and those for small and large deals

are altered to 1. Another difference between the generating processes for the middle and right-most columns of Figure 17 is that we set $p = 0.7$ for the former and $p = 0.6$ for the latter.

The algorithm below formally describes the process.

ALGORITHM 1: Counterfactual data simulation process

Input: Demand data with the number of samples N and for some predetermined probability $p \in (0, 1)$

Output: Demand data with modified deal outcomes

Note $Group_i$: size group(small/medium/large) of sample i

$Deal\ Success_i$: binary deal success variable of sample i

Obtain N draws from Bernoulli distribution with p and let it $\mathbf{d} \in \{0, 1\}^N$

if $\mathbf{d}_i == 1$ **then**

  if $Group_i == medium$ **then**

    $Deal\ Success_i = 1$

  else

    $Deal\ Success_i = 0$

  end

else

  pass

end
