Second-degree Price Discrimination: Theoretical Analysis, Experiment Design, and Empirical Estimation*

Soheil Ghili†, K. Sudhir‡, Nitish Jain§, Ankur Garg¶

March 3, 2025

Abstract

We propose an empirical model of second-degree price discrimination (2PD) that closely follows the theoretical literature on screening. Crucially, our model captures the covariance, across consumers, between the "baseline" willingness to pay (affecting all product versions) and the perceived differentiation between versions. While essential in determining the shape of the optimal 2PD mechanism, this covariance is challenging to identify in settings with aggregate data and a small number of products/versions, typical of 2PD environments. We develop an experimental framework that resolves these identification challenges. Our experimental-econometric methodology, hence, allows mechanism designers to empirically solve for optimal 2PD mechanisms in a wide range of applications where the differentiation between product versions is based on quality, quantity, timing of purchase, etc. We demonstrate the applicability of our framework by implementing it in the context of seat selection for flights in collaboration with an international airline.

*We thank Dirk Bergemann, Steven Berry, Eric Bradlow, Tat Chan, Judy Chevalier, Zhenling Jiang, Katja Siem, Jidong Zhou, and participants at various seminars and conferences for helpful comments. We are grateful to Akhil Dua for advice on and assistance with data collection. We thank Chinmay Ingalgavi, Mohammad Reza Nouri, and Wanxi Zhou for outstanding research assistance.

†Yale University, email address: soheil.ghili@yale.edu

‡Yale University, email address: k.sudhir@yale.edu

§London Business School, email address: njain@london.edu

¶Air India, email address: ankur1808@gmail.com

1 Introduction

Second-degree price discrimination ("2PD") is ubiquitous. Its common—and overlapping—forms range from quality-based tiering/versioning of products (e.g., Macbook Pro v.s. Macbook Air, or tiers of digital products such as cloud services or software), quantity discounts, varying prices based on purchase timing (e.g., a discount for early booking of accommodation), and product bundling. The theoretical literature on the optimal selling mechanism for a monopolist in 2PD settings is vast, especially in the case of product versioning and bundling. As a number of studies in the screening literature suggest, the monopolist's optimal mechanism often depends on certain properties of the joint distribution on consumers' valuations of the different products. Hence, to execute the optimal mechanism, a requirement is knowledge of this joint distribution by the designer/seller. Sellers in real-world 2PD settings, however, often lack such information. This can hinder the use of screening theory for 2PD decisions.

This paper's objective is to make the theory of screening empirically applicable by mechanism designers who do not apriori possess the required information on the joint distribution of the valuations. To this end, we accomplish three objectives. First, we develop a theory-driven empirical model of 2PD. Our model allows to replicate and build on results from the screening literature on the shape of the optimal mechanism. At the same time, the model has desirable computational properties. Second, we design an experiment that helps identify the joint distribution of values in our model under two key constraints common to applications of 2PD: (i) aggregate data and (ii) small number of products/versions. Third, we implement our experimental and empirical framework in the context of flight ancillary products, demonstrating its applicability to real-world 2PD policy design settings.

Our model considers a monopolist selling two products (or, versions of a product), although the key insights can generalized to higher numbers. To maintain generality of applications, we remain agnostic on the nature of the differentiation between the two products; it can be quality, quantity, time of purchase, or any set of features. Motivated by theoretical results from the literature, we not only allow for heterogeneity in valuations for each product, but also for different degrees of co-variation between the two. More specifically, our utility specification is a fixed-effects model with a constant and a "dummy" variable for one version. We choose the on-average-superior version as the non-omitted group. Here, the constant may be interpreted as the "baseline valuation," shifting the willingness to pay for both versions; whereas the dummy represents the differentiation value. The model has two crucial features: (i) Both the constant and the dummy vary at the individual level. That is, there are random effects on both objects. (ii) The co-variance between these random effects is flexible. This co-variance determines which consumers place a higher differentiation value between the baseline and the "premium" product: those with higher or lower baseline willingness to pay. The relevance of this co-variance to the optimal 2PD policy, as will be discussed in detail, is strongly motivated by the theory literature. Last, our utility model includes an individual-product level idiosyncratic error term which is meant to capture the idea that the horizontal component of the differentiation between the products.

We start by using our model to deliver theoretical results which replicate and build on some of the findings established in the literature. We present formal results in section 3. Here, we provide an informal discussion assuming both versions are produced at zero marginal cost. Our analysis shows at least three quantities play key roles in shaping the monopolist's optimal mechanism: (i) the variance across individuals in the baseline valuation (i.e., the variance of the random effect on the constant term), (ii) the covariance, across individuals, between the baseline and differentiation valuations, (iii) the relative

weights of vertical and horizontal differentiation between the two product versions. The effect of the first factor is relatively straightforward. A higher variance in the baseline valuation means consumers are more heterogeneous. This should increase the likelihood that a monopolist seller benefits from targeting them with multiple prices i.e., price discrimination. Nonetheless, directly charging different prices to different consumers (i.e., third-degree PD) is not feasible. If the covariance between the baseline valuation and the differentiation valuation is positive, then charging more for the "premium" product can indirectly help the monopolist charge more to those with higher baseline valuations. If this covariance is negative, however, the monopolist's optimal policy depends on the degree of horizontal differentiation between the products. First consider a case where the differentiation is fully vertical, i.e., the on-average-inferior version is found inferior by all consumers. In this case, a negative covariance between baseline and differentiation valuations implies that the monopolist finds it optimal to sell only the premium product; because the inferior version costs no less to produce, is not preferred by any consumer, and does not provide an instrument of screening. Next, consider a substantial degree of horizontal differentiation, i.e., the on-average-inferior version the preferred version by a non-trivial portion of the market. In this case, perhaps counter-intuitively, we show that the monopolist finds it optimal to charge more for the inferior product. The intuition is nuanced and is discussed in detail in section 3.2 along with examples. In short, the optimal selling mechanism involves selling the on-average-superior product to the majority of purchasing consumers, while pricing the other product targeted towards the minority who find it the more desirable product.

We also provide theory-driven empirical tests that help to identify the three aforementioned quantities. Formal results are provided in section 4. Here, we confine to an informal discussion. First, to identify the across consumers in the baseline valuation, we propose partial-market tests. In this test, the price of one product/version is changed exogenously, while the other price is kept fixed. The cross-substitution rate under this test can help identify the variance of the baseline valuation. Upon an increase in the price of one version, if a high portion of that version's consumers switch to the other version (as opposed to the outside option), it suggests a high variation in the baseline valuation. Informally speaking, this is because such a pattern would suggest that those who have a high willingness to pay for one product are likely to have higher willingnesses to pay for the other as well.

Our second test helps identify the covariance between the baseline valuation and the differentiation valuation. The literature on estimating the covariance between random effects in discrete choice models is scant. The few studies that do examine this assume the number of products in the data is substantially larger than the number of features with random effects. However, the possible number of versions in 2PD applications tends to be few, often as small as two. To our knowledge, our 2PD model is the first empirical setting in which the covariance between two random effects is central to the key policy of interest, rendering its credible estimation essential despite these challenges. We devise a total-market test in which the prices of both versions change by the same amount. By fixing the price difference between the versions, this test is designed to shut off cross substitution. As a result, any change in the composition of the demand in response to this test can only be attributed to taste differences between marginal and infra-marginal consumers. To illustrate, consider a total-market test uniformly discounting both prices. Suppose in response to this treatment, the demand for the on-average-inferior product, relative to the demand for the other version, increases. This change in the demand composition implies that the marginal consumers (i.e., those responding to the discount) have a lower differentiation value relative to the infra-marginal ones (i.e., those who would have purchased absent the discount). Given that marginal consumers also tend to have lower baseline valuations, this in turn implies a positive covariance

between the random effects. Not only does this test work with only two products, but also it does not require data on characteristics of individual consumers. Thus, it can be applied under restrictions that mechanism designers are expected to face in real settings. Measuring the third quantity of interest, i.e., the degree of horizontal differentiation relative to vertical, is simpler. It entails examining the demand ratio between the two versions under equal prices. If the ratio is close to zero, the differentiation is mostly vertical. If it is close to one, the differentiation is mostly horizontal. This does not require performing a separate test. If the prices of the products are close under the default condition (i.e., the control group), the ratio could be measured under the control group and/or the total market tests. Otherwise, it would suffice if the partial-market test equated them.

In section 5, we set up the full empirical model to be estimated using data. We argue that our model is one of random-coefficients discrete choice; hence, the procedure developed by BLP can be leveraged for estimation. This also means that characteristics (or fixed effects) for different markets can be incorporated into the utility model and interacted with version fixed effects at negligible cost to the computation. Furthermore, individual-level moments can be incorporated into the estimation procedure a la BLP 2004. Although it is feasible to implement these extensions, it is worth emphasizing that our framework allows a monopolist to empirically optimize a 2PD problem without market- or individual-level data. As such, this is the case that we focus on as our main specification.

To demonstrate the applicability of our experimental-empirical framework, we implement it in collaboration with an international airline based in Asia and describe the application in section 6. Our application is seat selection, and the differentiating feature between the two versions is timing of purchase. That is, we seat selection can be priced differently based on whether the passenger (who has purchased the flight ticket already), books the seat within or outside a three-day interval of the date of departure. Such an exercise is a common form of 2PD, especially used in the context of products that have a set consumption date and may be booked (e.g., airfare and its ancillaries, accommodation, concerts, etc.) Under the default condition, i.e., our control group, the airline was not offering an early purchase discount. As such, the default prices of the two versions are equal. We run (i) partial-market tests offering 30% and 60% discounts for early purchase only, and (ii) total-market tests offering 30% and 60% discounts uniformly on both products. We estimate our demand model and use it to study the optimal selling mechanism. In addition, we provide a discussion on the interpretability of our experiments. We show that a combined application of our propositions on optimal mechanism and our propositions on experiment design can often help guess the optimal mechanism solely based on observing the experiment results, and before estimating the demand model and conducting counterfactual analysis.

Although our seat-selection study demonstrates the applicability of our method, our experimental data exhibits only a subset of the patterns that could emerge in applications. As a result, in section 7, we supplement the experimental analysis with additional simulation analyses. We start examine simulated datasets that show a variety of patterns when it comes to the three quantities suggested to be key by theory: (i) the variance of the baseline valuation, (ii) the covariance between baseline and differentiation valuations, and (iii) the relative weights of vertical v.s. horizontal differentiation. Our analysis here takes a similar approach to how we studied the main experiment: in each simulation, we use our optimal-mechanism and experiment-design propositions to showcase how the statistics resulting from the experiment can be directly suggestive on optimal policy, even before model estimation takes place.

Our methodology can apply to a broad set of 2PD cases, helping mechanism designers solve for the

optimal pricing policy in the field. Underlying this applicability are three strengths of the framework: (i) The key parameters of our model are motivated by the theory of screening. (ii) Our experiment design helps to identify the key parameters–especially the covariance between baseline and differentiation valuations—under constraints that a 2PD design problem is likely to face in the field. (iii) The model can be cast as BLP, and hence, estimated efficiently.

2 Related Literature

Our paper contributes to the literature on 2PD by making a certain connection between the theoretical and empirical sides of the research on this topic. Specifically, we develop an empirical-experimental framework that allows a seller (mechanism designer) to make 2PD decisions in a theory-driven manner. Our framework achieves this goal by enabling the designer to estimate key (as suggested by theory) parameters of a 2PD model under data limitations inherent to 2PD environments.

2.1 Theoretical Literature

The theoretical literature on 2PD is vast. As a selling mechanism, 2PD is considered a mechanism of "screening." There is a notable difference between screening and some other common selling mechanisms such as auctions. In auctions, we know since Vickrey (1961) that under conditions that are expected to hold in a range of applied settings, the optimal selling mechanism does not depend on the exact shape of the distribution of the model primitives. In fact, in a wide range of settings, the revenue-equivalence theorem holds; and different formats of auctions yield the same expected payoff for the seller. In screening, however, knowledge of this distribution is necessary to solve for the optimal mechanism. The fact that the distribution over the primitives plays a more salient role in shaping the optimal screening mechanism–than it does in determining the optimal auction–can be verified by examining most screening papers (for a non-exhaustive set of examples, see Mussa and Rosen (1978); Maskin and Riley (1984); Rochet and Stole (2003, 2002); Adams and Yellen (1976); McAfee et al. (1989); Armstrong (1996, 2013, 2016); Bergemann et al. (2024)).[1] As a result, a full implementation of a screening mechanism should involve empirical tests (or analysis of observational data) to infer that distribution. In the theory literature, however, the focus is on characterizing the optimal mechanism assuming the distribution of the primitives is known. To the best of our knowledge, the only theory paper that proposes empirical tests for inference about the primitives of a screening problem is Yang et al. (2024).

Within the screening literature, there is a subset that is especially related to our study. This subset of the literature emphasizes that certain feature of the joint distribution of primitives has a key role in shaping the optimal mechanism: the co-variation between the willingness-to-pay type of a buyer and her preference for the differentiating feature between products. Examples of such papers–mostly applied to bundling or product-line design settings–are Salant (1989); Anderson and Dana Jr (2009); Haghpanah and Hartline (2021); Ghili (2023); Yang (2021, 2023). Our model of 2PD allows to replicate a version of the results in these papers which ensures a connection to the key insights they provide on optimal

[1] Of course, even for auctions, there are conditions about the primitives that can change the optimal mechanism (see, for instance Bergemann et al. (2019)). That said, the set of possible auction mechanisms is typically small (1st price, second-price, etc.). For screening, on the other hand, even a simple two-price mechanism can take infinitely many forms depending on the prices chosen for the product versions; and which of these forms is optimal depends on the distribution of model primitives.

screening. In addition, our empirical tests allow for the identification of the key parameters, especially the covariance parameter.

2.2 Empirical Literature

Empirical literature on 2PD also vast in quantitative marketing and economics. Examples range from dynamic pricing (Nair, 2007; Hendel and Nevo, 2013), to nonlinear pricing (Iyengar et al., 2008; Iyengar and Gupta, 2009; Iyengar and Jedidi, 2012; Nevo et al., 2016; Luo et al., 2018; Ghili and Yoon, 2023), version-based pricing (Verboven, 2002; McManus, 2007; Aryal et al., 2021), etc.

Our methodology contributes to this strand by providing a framework that enables mechanism designers to empirically implement 2PD models under constraints that are expected to be present in many applications. To this end, unlike a subset of the aforementioned papers which incorporate supply side moments into their estimation procedures (e.g., Luo et al. (2018); Aryal et al. (2021)), we do only use demand-side moments. In addition, motivated by the theoretical literature, our empirical model keeps flexible the covariance between the baseline and differentiation values and aims to estimate it. Some of the papers in the literature impose more structure by assuming independence between the two (e.g., Aryal et al. (2021)). Finally, to estimate the parameters of interest, we design and implement empirical tests that enable identification without a requirement for a large number of products or for detailed data on consumer observables.

Summary

In sum, our paper contributes to the literature by bridging between the theoretical and empirical studies of 2PD. More specifically, we develop a framework that (i) adopts a certain subset of 2PD models from the theory literature; and (ii) amends them with an experiment design that helps to empirically solve for the optimal selling mechanism under limitations likely to be faced by the mechanism designer in real-world applications. This makes our framework applicable in a wide range of 2PD settings such as quality-based discrimination, quantitiy-based discrimination, and discrimination based on the timing of purchase.

3 Theoretical Analysis

In this section, we build a theory model of second-degree price discrimination which will motivate our empirical analysis. The objective is to build a model that (i) is parsimonious, (ii) highlights the forces that, based on the microeconomics theory literature, play key roles in shaping price discrimination policies, and (iii) is easy to extend to an empirical model that can be estimated on data. We start with a model that allows vertical differentiation only between versions of a product. We then extend the model to allow for horizontal differentiation.

3.1 Vertically Differentiated Products

Consider two version of a product sold by a monopolist. The "better" or default version is indexed 0, whereas an inferior version is indexed -1. The inferiority of version -1 comes from lack of a feature that is expected to be on average desirable by consumers. Examples of such features are: extra memory in a smartphone, higher quality seats on a flight, timing flexibility on ticket purchase, etc.

Consider a market in which for any individual i, the utility from consuming version k of the product is described as follows:

$$u_{ik} = a - b \times p_k + \mu_i + \gamma_i \times \mathbb{1}_{k=0} \tag{1}$$

The utility from the outside option is normalized to zero. The term a is a real-valued constant, possibly negative. The term p_k represents the price of version k; and b is the price coefficient. The term μ_i shifts individual i's willingness to pay for both inside options. Given the presence of term a, the term μ_i can be assumed mean-zero without loss. We assume (with loss and only in this section) that μ_i is distributed uniformly on the interval $[-\sigma_\mu, \sigma_\mu]$ where $\sigma_\mu$is a non-negative real number. The term γ_i represents consumer i's willingness to pay to upgrade to version $k = 0$ from version $k = -1$. We assume γ_i is a function of μ_i given as follows:

$$\gamma_i = \gamma^0 + \gamma^1 \times \mu_i \tag{2}$$

Without loss of generality, we assume $\gamma^0 \geq 0$. We assume (with loss and only in this section) that the differentiation between the two products is only vertical. That is: $\forall i : \gamma_i \geq 0$, which would hold if and only if $|\gamma^1| \leq \frac{\gamma^0}{\sigma_\mu}$.

The firm to optimize its profit by choosing prices p_{-1} and p_0.

Next, we deliver two sets of results. We start from "Optimal Mechanism" results which. As the name suggests, these results characterize the optimal pricing mechanism under different conditions. Optimal mechanism results, hence, help to identify what parameters play key roles in shaping the optimal firm policy. Next, we shift the focus away from optimal mechanisms, and turn to "Demand Response" which have a different purpose. These results examine the responses of demand levels for the two products to a set of price treatments. Importantly, our demand response results provide predictions on how the values of key model parameters (as identified by the optimal mechanism results) shape demand patterns in response to the price treatments. Theses results, as we will discuss later, can be leveraged for effective experiment design and empirical estimation of the model.

Proposition 1 below characterizes this optimal decision making.

Proposition 1. *Consider a demand system given by equations 1 and 2. Supposed all marginal costs are zero. Under the firm's optimal pricing policy* (p^*_{-1}, p^*_0)*:*

- *The monopolist finds 2nd-degree price discrimination optimal if* $\gamma^1 > 0$ *and* $\frac{\gamma^0}{\gamma^1} < a < 3\sigma_\mu$*. In this case, the optimal prices are given by:*

$$p^*_{-1} = \frac{a + \sigma_\mu}{2b}, \qquad p^*_0 - p^*_{-1} = \frac{\frac{\gamma^0}{\gamma^1} + \sigma_\mu}{2b}$$

- *In all other conditions, the firm will find uniform prices optimal:*

$$p^*_{-1} = p^*_0 = \frac{\sigma_\mu(1+\gamma^1)+\gamma^0+a}{2b}$$

Proof. The proof of this proposition invokes the revelation principle and shows the above statements are true in the space of direct-revelation mechanisms. It may be found in appendix A.1.

Before turning to interpretations of the two scenarios described in proposition 1, an observation is worth making. Our setting assumes version $k = 0$ is vertically superior over version $k = -1$.[2] This version is considered of (weakly) higher-quality by *every* consumer in our model, while it is not more costly to produce. As a result, a socially efficient allocation of products would never allocate version $k = -1$ to any consumer. What is the economic intuition, then, for the observation that the monopolist sometimes sells the inferior version? The answer is price discrimination: a profit-maximizing seller can benefit from introducing the inferior version $k = -1$ and targeting its price to lower-WTP consumers, which in turn allows to price the superior version $k = 0$ with a primary focus on higher-WTP consumers. In other words, the monopolist uses the differentiating feature between the two product versions as an instrument of screening. Variants of this phenomenon have been studied in Mussa and Rosen (1978); Maskin and Riley (1984), and follow-up studies.

It should be noted, however, that it is not always optimal to leverage a particular product feature for price discrimination. Some recent studies (Anderson and Dana Jr, 2009; Haghpanah and Hartline, 2021; Ghili, 2023) provide theoretical characterizations on when and to what degree product features may be used for 2PD. In fact, part of the proof of proposition 1 leverages the main result in Ghili (2023).[3] An adaptation of that main result (which is proposed for a general bundling problem) to our setting would state the following: the monopolist would find it optimal to price discriminate if the optimal sold-alone quantity of version $k = -1$ surpasses that of $k = 0$. More specifically, denote by D^*_{-1} the volume of demand if the monopolist sells only $k = -1$ and prices it optimally. Define D^*_0 in a similar fashion. An adaptation of Ghili (2023) to our setting, then, holds that the monopolist should price discriminate (by selling both versions and setting $p^*_{-1} < p^*_0$) if and only if $D^*_{-1} > D^*_0$.

With this background, we next turn to interpreting result in proposition 1 by examining the role of each parameter in shaping the monopolist's optimal decision on whether or not to price discriminate.

- **The role of σ_μ:** A higher $\sigma_\mu$makes it more likely that price discrimination is optimal. This should be expected, as $\sigma_\mu$represents the magnitue of heterogeneity in consumers' WTP. The higher the consumers are differentiated in terms of their overall WTP for purchase, the more the monopolist's gain from targetting them with multiple prices.
- **The role of γ^1:** Although a high degree of WTP heterogeneity among consumers (captured through σ_μ) means the monopolist would have a higher incentive to price discriminate, an important next question in a 2PD context is whether the particular differentiation between the two versions is a "suitable instrument" to implement the discrimination. This is governed by the parameter γ^1. A positive γ^1 means consumers who have a higher baseline valuation shifted by μ_i, also have higher differentiation valuation shifted by γ_i. As a result, the monopolist may optimally decide to charge more for $k = 0$ as an instrument for charging more to those with higher μ_i. The same cannot be

[2] Recall that this assumption will be relaxed in the empirical section.

[3] Under some parameter values of our model, the main theorem in Ghili (2023) does not apply. As a result, the proof of our proposition 1 entails more than a direct application of Ghili (2023).

said when $\gamma^1 < 0$. In this case, a higher preference for version $k = 0$ does not positively co-vary with a higher baseline willingness to pay. Hence, the differentiation between the two versions may not be used as an instrument of screening.

The other two parameters are may be less directly related to 2PD. Hence, in order to interpret their impact on the monopolist's optimal mechanism, we leverage the main result from Ghili (2023).

- **The role of the constant a:** a higher constant a means a higher mean for the baseline valuation fo the products. This implies that the on-average-inferior version $k = -1$ may be sufficiently valuable for a large enough volume of consumer to make it a worthwhile sell for the monopolist to those who may have lower differentiation valuations. Per the intuition discussed above from Ghili (2023), this makes it more likely that the monopolist finds price discrimination optimal.
- **The role of the average differentiation valuation γ^0:** Perhaps counter-intuitively, a higher γ^0 means the monopolist is more likely to not find it optimal to charge two different prices for the product versions. This, too, however, can be justified in light of the result from Ghili (2023). A high γ^0 means that version $k = 0$ is superior by a margin wide enough to incentivize the monoplist ot sell this version to any consumer who would buy any version. Hence, a high γ^0 may incentivize the monopolist against price discrimination.

3.2 Vertically and Horizontally Differentiated Products

Our ultimate objective is to develop an empirical framework that can be applied to real-world settings. In applications of price discrimination, products are often differentiated not only vertically, but also horizontally. This often happens because advantageous product features come not only at extra monetary expenses to consumers, but also at extra non-monetary cost. As an example, Large Language Models (LLMs) typically come in multiple, vertically differentiated models. More advanced models have better reasoning capabilities but may take longer to run. As a result, some consumers (especially in business-to-business applications) may prefer to use less advanced versions such as Open AI's Chat GPT 4o for speed, even if each API call (i.e., instance use) is priced the same as a call on more advanced models such as Chat GPT o1. Our model, hence, would need to allow for varying degrees of vertical as well as horizontal differentiation.

There is also a more mechanical reason why we need to extend our model to horizontal differentiation. Our objective is empirical estimation. In empirical models of utility, idiosyncratic error terms are present; and error terms introduce horizontal differentiation. Utility equation 3 adds error terms to our original utility model from the previous subsection:

$$u_{ik} = a - b \times p_k + \mu_i + \gamma_i \times \mathbb{1}_{k=0} + \epsilon_{ik} \tag{3}$$

To see why this formulation introduces horizontal differentiation, observe that if the error terms are of extreme type I distribution,[4] then for any price combination (p_{-1}, p_0), there would be a positive measure of consumers who would purchase each version k.

[4] Any other mean-zero and symmetric distribution with unbounded domain would work as well.

We now turn to a key question: is the characterization in proposition 1 to the addition of error terms to the utility? Note that providing a full characterization of the optimal selling mechanism in this case is difficult; because the error terms ϵ_{ik} as as many dimensions to the screening problem as the number of products plus one. We are not aware of tools in the mechanism design literature that would allow one to characterize the optimal mechanism in the presence of such idiosyncratic terms.[5] Proposition 2, however, shows that a partial but useful characterization is attainable.

Proposition 2. *Consider a demand system given by equations 3 and 2. Assume that the error terms have an extreme type I distribution, and that marginal costs are zero. The following are true about the monopolist's optimal prices* (p^*_{-1}, p^*_0)*:*

- *If* $\gamma^1 < 0$*: then* $p^*_{-1} > p^*_0$*.*
- *If* $\gamma^1 > 0$*: then* $p^*_{-1} < p^*_0$*.*

The proof may be found in A.2 appendix. In words, this result indicates that the characterization proposition 1 is not robust to introducing horizontal differentiation. In fact corollary 1 below shows this non-robustness is the case even under arbitrarily negligible degrees of horizontal differentiation.

Corollary 1. *Proposition 2 continues to hold if the variance of the error terms is set to any positive real number, instead of the standard level of 1.*

The proof of this corollary is immediate and as based on the observation that the variance of the error terms is a normalization.

Note that the simplicity of the conditions in proposition 2 (i.e., the result that only γ^1 matters) is not due to the simplicity of the model. In fact the setting of proposition 2 is substantially more general than that in proposition 1: the addition of error terms, no functional form assumption on the distribution of μ_i, and no restriction on absolute value of γ^1. Yet the result is much simpler. This raises two questions: (i) what is the intuition behind the result in proposition 2? (ii) Does proposition 2 render the result in proposition 1 irrelevant to empirical analysis of price discrimination?

The result in proposition 2 and its corollary is counter-intuitive: The monopolist will charge more for $k = -1$ under $\gamma^1 < 0$, even if horizontal differentiation (measured by the variance of the error term) is arbitrarily small and the average vertical differentiation in favor of $k = 0$ (measured by the magnitude of $\gamma^0 > 0$) is arbitrarily large. In short, optimal pricing under $\gamma^1 < 0$ charges more for the "worse" product even if it is worse by a wide margin. To see why, fix any pricing policy (p_{-1}, p_0) and consider two possible levels $\mu^1 > \mu^2$ for the term μ_i. Denote by $s_k(p_{-1}, p_0|\mu^1)$ the share of individuals i with $\mu_i = \mu^1$ who purchase version k under prices (p_{-1}, p_0). Consider similar definitions for $s_k(p_{-1}, p_0|\mu^2)$. One can use simple logit demand formulas to show that under $\gamma^1 < 0$:

$$\frac{s_k(p_{-1}, p_0|\mu^1)}{s_k(p_0, p_0|\mu^1)} > \frac{s_k(p_{-1}, p_0|\mu^2)}{s_k(p_0, p_0|\mu^2)}$$

That is, as we examine higher WTP consumers (i.e., μ^1 as compared to μ^2), a higher share of those who purchase any version choose the "inferior" version $k = -1$. As a result, the population of those

[5] Rochet and Stole (2002) examine a setting with random participation and do characterize the optimal mechanism. However, in their setting, randomness only impacts participation, i.e., the IR constraint; and eventual empirical implementation is not part of their objective. In our setting, the error terms impact not only the IR constraints, but also the IC, which is why direct-mechanism tools in the screening literature do not help provide a characterization in the case of proposition 2.

who purchase $k = -1$ has higher μ_i levels than those who purchase $k = 0$ in the monotone-likelihood ratio (MLR) sense. This is what makes charging more for $K = -1$ profitable. Note that this remains true regardless of how large γ^0 is. A large γ^0 would imply that for all μ_i, consumer i is more likely than not to prefer $k = 0$ over $k = -1$. Whereas a negative γ^1 implies that those who prefer $k = -1$ over $k = 0$ tend to have higher μ_i than those who have the opposite preference. Both of these comparative statements can be true at the same time; but it is only that latter that shapes the ordinal comparison between p^*_{-1} and p^*_0.

Do proposition 2 and corollary 1 render the insights of proposition 1 irrelevant to empirical applications of screening and price discrimination? Our answer is no. As our simulation analyses in the empirical section of the paper will demonstrate, the insights from proposition 1 continue to "qualitatively" hold when the vertical aspect of the differentiation between product versions dominates the horizontal. In those cases, a negative γ^1 would technically imply $p^*_{-1} > p^*_0$, but the this optimal pricing policy performs only negligibly better than a uniform pricing policy that forces $p_0 = p_{-1}$ and chooses this common price optimally. The case of $\gamma^1 > 0$ is different: in markets where vertical differentiation dominates horizontal, we have $p^*_{-1} < p^*_0$ as both propositions 1 and 2 predict; and the respective performance improvement upon uniform pricing is non-negligible.

Based on this discussion, we do not see proposition 2 as an invalidation of proposition 1. Rather, we view it as a signal that a useful empirical model of 2nd-degree price discrimination should not confine to importing insights from the mechanism design and screening literature which, for tractability purposes, often restrict attention to vertically differentiated products.[6]

Note on marginal costs. Before turning to experiment design, it is worth discussing how our results on the shape of the optimal mechanism generalize to the case where instead of zero marginal costs, the two product versions have marginal costs of c_{-1} and c_0 respectively. Proposition 1 should extend in a straightforward fashion given that its proof mainly leverages results from Ghili (2023), results that allow for flexible marginal costs. Proposition 2 also extends naturally. One can show that the proposition continues to hold if we replace the prices by the margins, comparing $p^*_0 - c_0$ to $p^*_{-1} - c_{-1}$ instead of comparing p^*_0 to p^*_{-1}. Given that the equivalent of this result is not proven in a previous paper, we provide it in appendix B. In the main text, however, we focus on the case with zero marginal costs.

4 Experiment Design

We now turn to designing an experiment to empirically estimate our model. The first step is to ask what is the ideal dataset for estimating our demand model? The answer is a dataset in which consumers are exposed, at random, to all relevant pairs (p_{-1}, p_0), so that demand may be measured for all such combinations. Such ideal experiments are, however, likely infeasible in applications. Thus, it would be beneficial to provide minimal empirical tests that can help identify the key parameters of the model at the smallest possible cost to the mechanism designer. We turn to this task in this section.

[6] To the best of our knowledge, Rochet and Stole (2002) is the only paper that attempt to bring anything akin to error terms into the analysis of a multi-dimensional screening model. Even this paper, however, keeps the term fixed at the level of the individual and does not interact it with quality/quantity.

4.1 Theoretical Results

The analyses in the previous subsections suggest there are three important factors that can shape 2nd-degree price discrimination policies: (i) the degree of WTP heterogeneity among consumers, (ii) the degree to which the preference for higher quality versions is correlated with the baseline WTP, and (iii) the relative strengths of vertical and horizontal differentiation between the versions in the market. In this section, we ask what are the minimal price treatments that can help identify these parameters using *aggregate data* on demand.

4.1.1 "Partial Market Tests" and the identification of $\sigma_\mu$with aggregate data

Definition 1. *Consider a treatment that changes $P = (p_{-1}, p_0)$ to $\tilde{P}$ by increasing p_k to $\tilde{p}_k = p_k + r$ but keeping $p_{k'}$ fixed: $\tilde{p}_{k'} = p_{k'}$. Define the* ***cross substitution rate*** *$\psi_{k,k'}(r)$ as:*

$$\psi_{k,k'} := \frac{s_{k'}(\tilde{P}) - s_{k'}(P)}{s_k(P) - s_k(\tilde{P})}$$

In words, $\psi_{k,k'}$ shows what portion of consumers who abandon (purchase) product k in response to a increase (decrease) in p_k are those who would otherwise purchase k'.

Now we introduce a proposition that makes a connection between the parameter $\sigma_\mu$and a data pattern that can be directly observed under a partial-market test (i.e., a test that fixed the price of one product version and varies the other).

Proposition 3. ***Partial Market Test:*** *Under σ_μ, we have $\psi_{k,k'} = \frac{s_{k'}}{1-s_k}$. Under $\sigma_\mu > 0$, we have $\psi_{k,k'} > \frac{s_{k'}}{1-s_k}$.*

The proof may be found in the appendix. The intuition is the same as what can be found in the literature on nested and random-coefficient logit models: a high σ_μ, implies a high degree "vertical" heterogeneity among consumers in terms of their willingness to pay for either version. As a result, knowing that a consumer i has purchased version k makes it more likely (than under a low/negligible σ_μ) that this consumer would switch to the other version, as opposed to the outside option, upon an increase in p_k. This suggests that the substitution rate $\psi_{k,k'}$ may be useful in empirically identifying σ_μ.

4.1.2 "Total Market Tests" and the identification of γ^1 with aggregate data

Our utility model, described in eq. (3), can be interpreted as a random-coefficients discrete-choice model with two features: (i) it has product fixed effects, (ii) it has random effects on the contstant term and on the fixed effect for product $k = 0$, and the two are correlated. The parameter γ^1, which according to proposition 2 plays a key role in shaping the optimal mechanism, is proportional to the covariance between the two random effects.

To the best of our knowledge, the literature is sparse on discussions of what variation identifies correlated random effects in aggregate data. The most closely related reference is Gandhi and Houde (2023). The analysis by Gandhi and Houde (2023) suggests that characteristics should independently vary among products in the data, for the covariances among their respective random effects to be identified. This, in turn, suggests that a large number of products may be needed to identify the covariances. More

specifically, when estimating a model with independent random coefficients on four variables, Gandhi and Houde (2023) simulate a market with 15 products; but when turning to an analysis of correlated random effects, they increase the number of products to 50.[7]

Although the number of products is typically tens or hundreds in applications of random coefficients models to oligopolistic competition,[8] the same is typicall not the case in applications to second-degree price discrimination. In 2PD applications, the number of products, across all markets, is often as small as two, rendering the identification of the random-effects covariance chanllenging. Yet, as suggested by proposition 1 and proposition 2, the covariance between the random effect on the constant and that on the fixed effect for product $k = 0$ is essential to the shape of the optimal mechanism. This calls for an alternative approach to identifying this covariance in 2PD cases.

Our analysis in this section proposes a simple test which leads to a statistic that can be directly tied to the parameter γ^1.[9] We develop this test in the two-product environment of our model, though it can be extended to more products as well.

Definition 2. *Consider a uniform price change across products. Formally, fix a vector $\tilde{P} = (\tilde{p}_{-1}, \tilde{p}_0)$ of prices and consider price vectors $P = \tilde{P} + r$ where r is a scalar. Define the* ***relative response rate*** *$\phi_{k,k'}(r)$ as:*

$$\phi_{k,k'} := \frac{s_k(\tilde{P})}{s'_k(\tilde{P})} \Big/ \frac{s_k(P)}{s_{k'}(P)}$$

Intuitively, $\phi_{k,k'}$ shows how the semi-elasticities of demand compare to each other across products k, when all of them are subject to the same level of price chance, i.e., a price change that induces no cross-substitution. If $\phi_{k,k'}$ is larger than 1, it means that demand for product k is more responsive than demand for k' to uniform price changes.

Next we introduce a proposition that makes a connection between the sign of the parameter γ^1 and a data pattern that can be directly observed under a total-market test (i.e., a test that uniformly varies all prices, shutting off cross-substitution).

Proposition 4. ***(Total Market Test:)*** *Assume that the dataset is generated by our model and that $|\gamma^1| < 1$. Then, for any vector P of initial prices and any positive price change r, we have:*

$$\phi_{-1,0} \underset{(<)}{>} 1 \Rightarrow \gamma^1 \underset{(<)}{>} 0$$

The proof may be found in the A.3 appendix. The intuition is as follows: Given that a uniform price increase/decrease shuts off cross substitution, a change in the demand ratio between the two products in response to such a treatment can only be attributed to taste differences between marginal and infra-marginal buyers. To illustrate, suppose that compared to the control group, a treatment with uniform discounts shifts the demand ratio in favor of the "baseline" product $k = -1$. This means that the marginal buyers (i.e., those with lower constant μ_i) tend to also have a lower differentiation utility γ_i, implying a positive covariance between the random effects, hence a positive γ^1. A similar argument would

[7] The authors do not explicitly discuss whether they independently draw product characteristics for each of their 100 simulated markets. However, based on their Appendix B2, it appears that the characteristics are indeed redrawn for each simulated market. This raises the total number of products (among which feature level may independently vary) to 5000.

[8] For instance, the number of products was close to 1000 in Berry et al. (1995), and more than 20 in Nevo (2001). Neither of these papers estimates/reports off-diagonal terms in the variance-covariance matrix for the random effects.

[9] We are thankful to Steven Berry for insightful discussions on this topic.

hold for a negative γ^1. This difference-in-difference type of identification argument shares similarities with arguments describing how individual-level data may be leveraged to identify the nesting parameter in nested logit models. See Berry and Haile (2023)[10] for more details.[11]

To recap, as definition 2 and proposition 4 show, there is a simple test (i.e., uniformly changing all prices) that can generate a statistic which directly informs the sign of γ^1 regardless of the values of the other parameters. Importantly, this test does not require a large number of products.

Note that the combination of propositions 2 and 4 illustrates the versatility of our framework in connecting data patterns to optimal mechanisms, even before a model of demand is estimated. Corollary 2 formalizes this:

Corollary 2. *Suppose the conditions in propositions 2 and 4 hold. Then, we have:*

$$\phi_{-1,0} \underset{(<)}{>} 1 \Rightarrow p_0^* \underset{(<)}{>} p_{-1}^*$$

As this corollary shows, one can directly learn about the optimal pricing mechanism from the directly observable relative response rate ϕ under a total market test, without estimating a structural model of demand. Of course, the extent of such learning is limited; and to quantify the exact levels of the two optimal prices, an estimated model of demand would be required. Note that this corollary, too, will extend in a straightforward fashion to the case where each version k has a marginal cost of c_k: the only modification will be replacing the optimal prices p_k^* with optimal margins $p_k^* - c_k$ in the formula above.

4.1.3 The identification of the other parameters

Last, we turn to the identification of other parameters, especially the vertical-differentiation parameter γ^0.

Proposition 5. *Consider a uniform vector of prices: $\forall k, k' : p_k = p_{k'} = p \in \mathbb{R}$. Then, the following hold about the demands ratio $\frac{s_{-1}}{s_0}$: (i) $\partial \frac{s_{-1}}{s_0} \;/\; \partial \gamma^0 < 0$; (ii) $\lim_{\gamma^0 \to +\infty} \frac{s_{-1}}{s_0} = 0$; and (iii) if $\gamma^0 = 0$, then $\lim_{p \to -\infty} \frac{s_{-1}}{s_0} = 1$.*

The proof of this results is straightforward and is, hence, left to the reader. The intuitive content is that the demands ratio $\frac{s_{-1}}{s_0}$ under uniform prices is a good metric for assessing how much vertical differentiation there is between product versions. As item (iii) in the proposition suggests, if there are multiple vectors of uniform prices $p_k = p \, \forall k$, the lowest-price vector is the preferred one for examining $\frac{s_{-1}}{s_0}$ to gain intuition on the extent of vertical v.s. horizontal differentiation.

Propositions 3 through 5 help bridge between our theoretical analysis in the previous subsections and experiment design. More specifically, subsections 3.1 and 3.2 point to three key quantities that need to be considered when studying second-degree price discrimination: (i) σ_μ: vertical heterogeneity among consumers in their baseline willingness to pay, (ii) γ^1: the co-variation between baseline WTP μ_i and the differential WTP γ_i, and (iii) the level γ^0 of average vertical product differentiation. The results in this section suggest that to identify these key parameters, one could leverage data on (i) cross substitution rates $\psi_{-1,0}$, (ii) relative response rates $\phi_{-1,0}$, and (iii) shares ratio $\frac{s_{-1}}{s_0}$. We conjecture that in the case

[10] Berry and Haile (2023) is an older version of Berry and Haile (2024).
[11] We thank Steven Berry for pointing to this analogy.

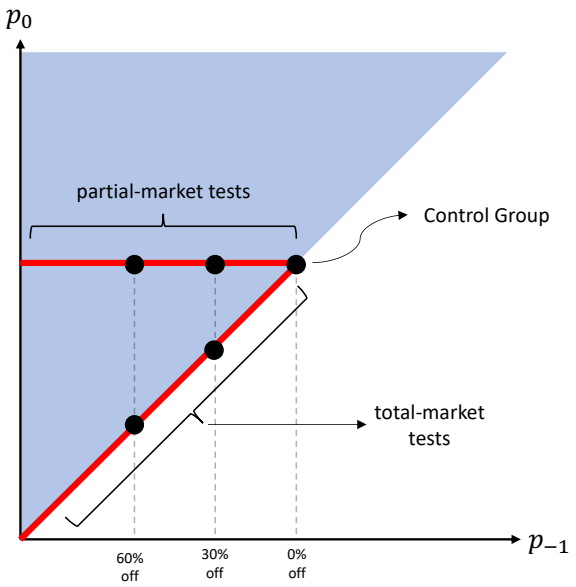


Figure 1: Schematic representation of the experiment design. On the two-dimensional space of product prices, the horizontal line captures partial-market tests, while the 45-degree line has total-market tests.

of each proposition, a stronger version of the result can be proven. For our specific purpose of obtaining ideas for experiment design, however, we find the current version adequate.

We next turn to an experiment design based on the results above.

4.2 Illustration of the design

Inspired by propositions 3 to 5, our experiment design consists of two types of testing: **total-market tests** and **partial-market tests**. A total-market test, as described previousely, is one in which all versions of the product are subject to the same level of price change. A partial-market test, on the other hand, subjects only one version of the product to a price modification.

Figure 1 schematically illustrates the experiment design. This figure shows the possible pricing policies (p_{-1}, p_0) in a two-dimensional space. One black dot is shows the prices under pre-experiment conditions, which would also coincide with price combinations under the control group. Price combinations on the horizontal red line to the left of the control group implement partial-market tests. Price combinations on the 45-degree line departing the control group implement total-market tests.

There are a number of features that are specific to the illustration in Figure 1 but not essential to our experiment design. First, the default (i.e., control-group) levels of p_{-1} and p_0 are equal in the figure. This is not an essential feature of the experiment. With any price combination as the control group, total-market tests would equally discount both prices[12] while partial tests would only discount one. Second, the fact that there are exactly two separate discount levels and they are at 30% and 60%

[12] Whether this "equal discounting" should be measured in relative, percentage, terms, or absolute, dollar, terms, depends on what one thinks is the best model of how price impacts utility. With a model where price enters the utility function linearly, such as our model in this paper, we propose a discount that is uniform across versions in absolute value. With a model in which it is the logarithm of price that enters the utility function linearly (such as the specification in Berry et al. (1995)), we propose a discount that is uniform in relative terms. Note that the absolute and relative schemes coincide if the prices of the product versions are uniform outside of the experiment (i.e., in the control group).

are not essential and may be modified to have only one, or more than two, discount levels. Third, our partial-market tests discount version $k = -1$. This is not essential; and one could alternatively discount $k = 0$ in applications where p_0 is non-trivially larger than p_{-1} in the control group. In those cases, we in fact recommend discounting version $k = 0$, for reasons discussed shortly.

Before turning to the empirical model, it is worth counting a number of practical consideration in deciding the details of the design. First, as Figure 1 shows, both partial- and total-market price variations are generated by a price decrease, as opposed to a price increase, relative to the contorl group. This is not econometrically essential but perhaps the only possibility from a practical standpoint. Price decreases could be framed as an "offer", while experimenting with price increases would face severe feasibility challenges. Second, we recommend at least two distinct price changes for each test. This should help alleviate potential placebo-effect concerns by creating varying degrees of treatment as opposed to just a binary treatment-control comparison. Third, if under the control group, the two prices are different, we propose that the partial-market test discount the more expensive product (i.e., version $k = 0$) and choose one of the discount level so that the two prices are equal. This should allow for an easier measurement of the degree of horizontal v.s. vertical differentiation. Finally, we recommend that the RCT be implemented at the market level if possible. If not, individual-level RCTs should also work. In fact, due to feasibility constraints, our own application of the framework is presented on an invididual-level RCT.

5 Empirical Model: Random-Coefficient Nested Logit

In this section, we extend the model described above to an empirical model that can be estimated using data on prices and quantities. The consumer utility equation for this empirical model is given as follows:

$$u_{imk} = \beta X_{mk} + \xi_{mk} + \mu_i + \gamma_i \mathbb{1}_{k=0} + \epsilon_{imk} \tag{4}$$

This equation closely resembles equation 3 with minor yet important modifications. First, we have an additional index m for market. If an individual-level randomized control trial partitions the consumers in one market into multiple treatment groups, we use different m indices for those groups. Next, instead of $a - b \times p_k$, we have a more general term βX_{mk}, where X_{mk} captures observable characteristics of product version k in market m. The most crucial such characteristic is price p_{mk}. The vector X_{mk} should also include a constant term. Finally, in addition to the individual-level idiosyncratic error term ϵ_{imk}, we have a market level term ξ_{mk} as well.

We now turn to describing the terms that, as motivated by the theoretical analysis in the previous section, capture key forces in characterizing the optimal price discrimination policy. These terms are μ_i and γ_i. The term γ_i can be interpreted as a random coefficient on the degree of differentiation individuals see between the two versions. The term μ_i can be interpreted as a "random coefficient on the constant term" in the utility function. Our model of the relationship between γ_i and μ_i follows a generalization of equation 2, described below:

$$\gamma_i = \gamma^0 Z_m + \gamma^1 \times \mu_i \tag{5}$$

where γ^0 is, unlike in equation 2, no longer a constant. It is, rather, a vector of coefficients on market characteristics Z_m.

Note that this assumption implies that the correlation between γ_i and μ_i is always within the set $\{-1, 0, +1\}$. In principle, one could model the joint distribution between μ_i and γ_i more generally and with a flexible variance co-variance structure. We opted for this simpler model, however, because it is more parsimonious and, at the same time, it does allow the co-variation (as opposed to correlation) between γ_i and μ_i to assume any real value, depending on the sign and magnitude of the parameter γ^1.

Observe that this empirical model, unlike the theoretical analysis in the previous section, we can no longer enforce the assumption that $|\gamma^1| < \frac{\gamma^0}{\sigma_\mu}$ given since these are parameters to be estimated from data (we did not need this assumption on proving 2 either). If this assumption is violated by the data, there will be a second source (in addition to ϵ_{imk}) of horizontal differentiation between the two versions. Also, observe that the assumption of $\gamma^0 \geq 0$, which was made without loss of generality before, may not be apriori enforced here (in the form of $\gamma^0 Z_m \geq 0$) either.

We finish this section by emphasizing the practical properties of the empirical demand model described by equations 4 and 2. This model has two features. First, it closely follows the model in the previous section, hence capturing forces essential in analyzing second-degree price discrimination. Second, this is a random-coefficients logit model a la Berry et al. (1995) (BLP henceforth). The combination of these two features implies that the computational machinery developed in the literature for BLP models can be used to empirically estimate models of second-degree price discrimination. Perhaps most crucial among such computational methods is the contraction-mapping-based nested-loop structure first introduced by Berry (1994) and adopted by Berry et al. (1995). This means that, in the estimation process, only the heterogeneity parameters governing μ_i and γ_i (i.e., the "outer-loop parameters") would need to be estimated by directly searching a multi-dimensional space. Vector β of parameters (i.e., "inner-loop parameters) would be estimated through an Ordinary Least Squares regression which can be computed in polynomial time in the number of parameters. This would allow the econometrician to specify as rich a characteristics vector X as the application warrants, without concerns about the computational complexity exponentially increasing with the parameters count. To sum up: absent such a connection to BLP, the empirical estimation of models of second-degree price discrimination, would suffer from a heightened risk of getting stuck at local optima (with moderate number of parameters) or a prohibitively long computation time (with large number of parameters).

For further clarity, we henceforth refer to our model as Random Coefficient Nested Logit or RCNL a la Grigolon and Verboven (2014), as opposed to random coefficients logit or BLP. This is because, as previously mentioned, one of our random-coefficients terms is μ_i which essentially constitutes random coefficients on the constant term. As Berry (1994) points out, this is mathematically equivalent to allowing for correlations in error terms within two nests of inside and outside options, which is akin to a nested logit model. As a result, RCNL is technically not a generalization of BLP. Nonetheless, we find the use of the term RCNL helpful in emphasizing the nature of the random coefficients we employ.

Next, we turn to an application of this experimental-structural framework.

6 Empirical Application: Airline Ancillary Products

Our setting is an international airline operating mainly in Asia. We may not name this airline due to a non-disclosure agreement. Moving forward, for convenience, we refer to this firm as "the airline". The product we study in our empirical application is seat selection. Each customer who has purchased a flight ticket may pay an extra amount in order to book a seat of their choosing as opposed to being assigned a random seat on the day of the flight. The dimension along which we vary the prices is timing: we study the firm's optimal policy on whether and to what degree it should discount booking a seat early, as opposed to closer to the departure date.

There are a few reasons why we chose the context of ancillary product such as seat selection, and purchase timing as the discrimination dimension, for the application for our econometric framework:

1. Once passengers have purchased flight tickets, the market for seat selection becomes essentially monopolistic. This is useful given that our framework focuses on price-discrimination under monopoly and relegates investigating the role of competition to future research.

2. The monopoly power possessed by airlines when it comes to ancillaries makes the ancillaries a market of substantial size. By some estimates, the total ancillaries revenue by eight key U.S. based airlines was \$4.2B/year (Sorensen, 2023).

3. In spite of their large market size, ancillaries are not the main source of revenue for airlines (they trail airfare). Hence, a total- and partial-market tests with nontrivial discounts levels (30% and 60%) are more feasible. These discount levels were necessary to give a degree of ex-ante assurance that the setting would help illustrate the working of our experimental framework.

4. As will be discussed shortly, summary statistics show that under no early-purchase discount, the share of early seat purchase is consistently less than that of late purchase but the ratio is consistently non-negligible. This suggests both vertical and horizontal differentiation are present when it comes to purchase timing. This makes timing a useful application, given theoretical results suggesting both dimensions of differentiation play roles in shaping the prices.

6.1 Data and Setting

Our data records each consumer who has obtained a flight ticket. In other words, the unit of observation is a flight-customer, or more simply, a booking. For each booking, we have information on the flight: origin and destination, domestic or international, length, time and date of flight. We also have data on the passenger: gender, age, number of co-travellers, and product type. Additionally, we observe whether a seat selection was purchased by the passenger, what "seat type" it was if one was indeed purchased, and at what time the purchase was made. Finally, we observe the pricing of seat selection for each passenger. The pricing can indeed vary passenger by passenger because of the experiment we ran in collaboration with the airline. Figure 2 summarizes some statistics about the population in our data.

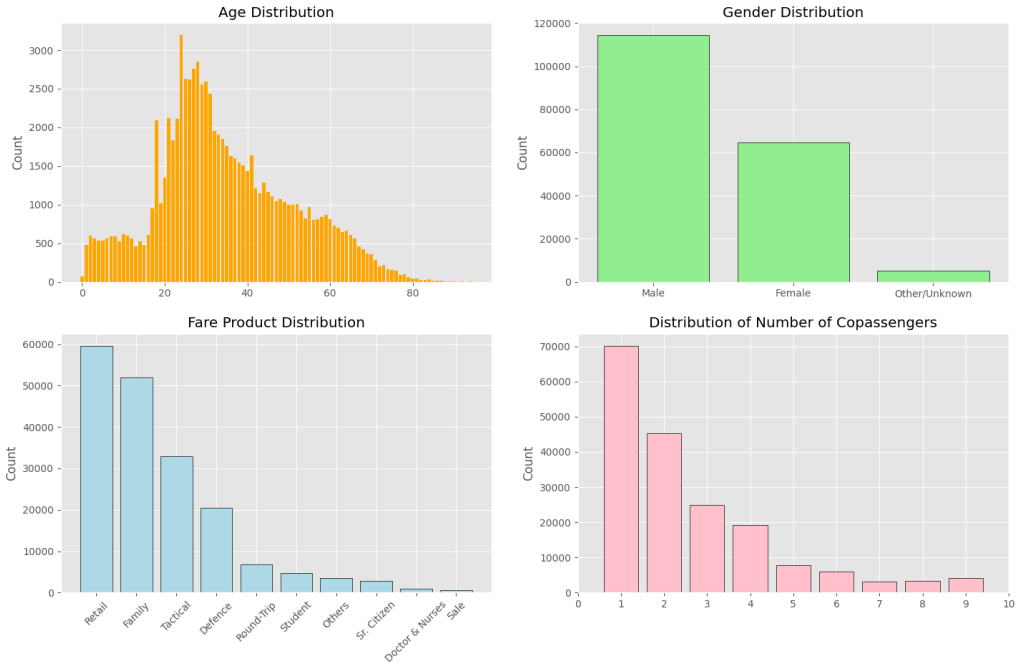


Figure 2: Demographic Data Summary (invalid ages discarded while making age bar plot)

The strategic questions by the airline are two-fold. First, whether they should offer an early-purchase discount for buying seat selection at least X number of days prior to the departure? And if so, how deep should the discount be? Second, would an early purchase discount allow them to potentially raise the non-discounted, "late purchase" price? And by how much? In other words, the instrument of discrimination is time of purchase, with early purchase being considered product $k = -1$ and late purchase product $k = 0$.

The experiment was run by dividing the passengers into two sub-populations based on how long before the departure they booked their flight. The first sub-population, which we term "D14+" are those who purchased the flight ticket more than 14 days prior to departure. The second group, which we call "D3-14" are those who booked the flight between 3 and 14 days before departure. We run the experiment that was visualized in fig. 1 on both sub-populations. In the main text of the paper, however, we discuss only the D3-14 group; because due to higher data quality, this is the group we chose for our main specification. Table and 1 depicts the early and late purchase rates by different treatment groups for D3-14.

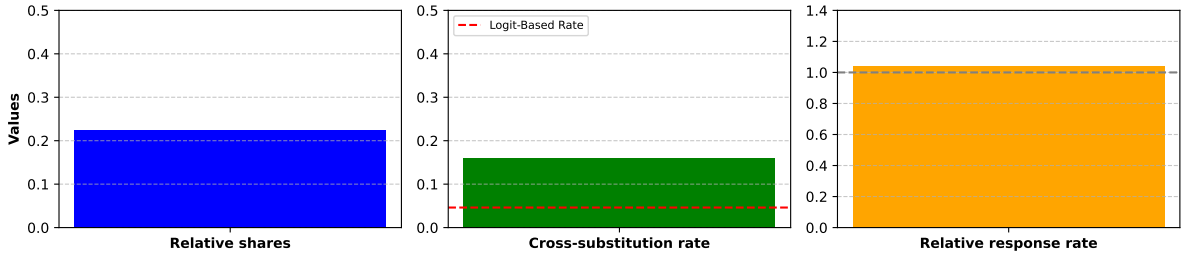


Figure 3: Visual representation of the three key summary statistics motivated by our theoretical analysis: (i) the relative shares between early and late purchase under equal prices, (ii) the cross-substitution rate under partial-market tests compared against the benchmark of "logit cross-substitution", and (iii) the relative-response rate under total-market tests, compared against the benchmark of 1.

Deadline	Discount	Count	Total Demand	Early dmnd (k=-1)	Late dmnd (k=0)
Yes	30%	12,856	5.78%	1.70%	4.10%
Yes	60%	12,772	6.36%	1.90%	4.40%
No	30%	14,669	6.24%	1.25%	4.98%
No	60%	14,615	8.05%	1.47%	6.58%
-	0%	16,487	5.61%	1.04%	4.57%

Table 1: Pre- and post deadline seat selection takeup rates (aggregate across seat types) for the D3-14 group.

A few patterns from these shares are worth discussing. First, and as expected, total demand increases as we offer (larger) blanket discounts. The total demand (measured in percentage of booking passengers that also reserve selected seats early or late) increases from 5.61% to 6.24% to 8.05% as we move from the control group to a treatment group with a blanket discount of 30% to a treatment group with a blanket discount of 60%. Next, we turn to the three statistics mentioned in section 4: (i) the share ratio between the two product versions under equal prices, (ii) the cross substitution rates under partial-market tests, and (iii) the relative response rates under total-market tests. All three of these statistics are visually represented in fig. 3 as well.

As figure 3 and table 1 show, the ratio between early shares and late shares under equal prices is less than one but always non-trivial, which is suggestive that (i) on average, there is vertical differentiation between late and early in favor of late, but (ii) there is also a non-negligible degree of horizontal differentiation. The presence of horizontal differentiation is suggestive that if an estimation of our model finds a negative relationship between μ_i and γ_i (captured by a negative γ^1), charging a higher price for early purchase may be optimal.

The average cross-substitution rate across partial-market tests, as shown by figure 3, is 0.18. This is above the logit-based rate of 0.05. As suggested by proposition 3, hence, one should expect a positive variance σ_μ for μ_i. By proposition 1, this suggests that the monopolist has an incentive to offer multiple prices and price discriminate. Whether discrimination based on time of purchase would be a suitable instrument of screening, however, is a question not possible to answer with cross-substitution rates only. We, hence, turn to the relative response rates.

The relative response rate under the total market test is approximately one. That is, as we uniformly

change both prices, the ratio between the demands for early and later purchase of the seat selection remains almost unchanged. This, by proposition 2, is suggestive that purchase timing is not a suitable instrument of discrimination, hence optimal pricing involves little to no discrimination.

6.2 Estimation Results, and Model Fit

Table 2 reports the estimation results. Note that the estimated parameters are in line with what was anticipated from figure 3. The parameter γ^0, i.e., the fixed effect on product $k = 0$, is positive. This implies a degree of vertical differentiation between the products, consistent with the left panel of figure 2. The parameter $\sigma_\mu$is also positive and of non-trivial magnitude, consistent with the implication by the cross-substitution rate from the middle panel of the figure. This, as previously discussed, is indicative of heterogeneity in consumers' willingnesses to pay for seat selection. As a result, the monopolist airline would find it optimal to use timing of purchase to price discriminate among these consumers if purchase timing is a suitable instrument of screening. By proposition 2, whether timing is indeed a suitable instrument depends on the parameter γ^1. Our estimation results show that, as expected from proposition 4 and the right panel of figure 3, the covariance parameter γ^1 is close to zero. This, as we will shortly examine, has implications for optimal pricing policy.

const	**Price**	$\boldsymbol{\gamma^0}$	$\boldsymbol{\sigma_\mu}$	$\boldsymbol{\gamma^1}$
−6.086	0.01094	1.220	3.005	0.039
(1.077)	(0.00169)	(0.804)	(0.877)	(0.243)

Table 2: Estimation Results for the D3-14 group.

Table 3 shows that our estimated model generates a good model fit to the real shares. Note that, due to incorporating the structural error terms ξ_{mk}, our model exactly matches the data. Hence, in generating the simulated shares, we set the ξ_{mk} terms to zero, in order to obtain a measure of the degree to which the "key forces in the model" were able to explain the shares, without help from the structural errors.

Deadline	Discount	Total Demand		Early dmnd (k=-1)		Late dmnd (k=0)	
		Data	Est.	Data	Est.	Data	Est.
Yes	30%	5.78%	5.79%	1.68%	1.46%	4.10%	4.33%
Yes	60%	6.36%	6.11%	1.93%	1.95%	4.42%	4.16%
No	30%	6.24%	6.65%	1.25%	1.30%	4.98%	5.34%
No	60%	8.05%	7.90%	1.47%	1.56%	6.58%	6.34%
-	0%	5.61%	5.55%	1.04%	1.08%	4.57%	4.47%

Table 3: Estimated vs. observed seat selection takeup rates

Next, we turn to analyzing the optimal pricing policy.

6.3 Counterfactual Analysis

In this section, we simulate the optimal pricing policy from the perspective of the firm. We compare the performance of that policy to two other regimes: (i) the pricing policy observed in the data, and (ii) the pricing policy that is optimal subject to $p_{-1} = p_0$, meaning an optimal policy subject to no price discrimination.

Table 4 summarizes the results. Recall that the prices are de-scaled so that they take the value of 100 under the control group. As a result of this de-scaling, the absolute values of the prices as well as those of the revenue, consumer welfare, and social welfare, are not informative. Comparisons among these numbers remains informative nonetheless.

pricing	p_{-1}	p_0	Share_{-1}	Share_0	revenue	consumer welfare	social welfare
Data	100.0	100.0	1.08%	4.47%	5.55	7.94	13.49
Optimal Mechanism	151.9	155.0	0.78%	3.16%	6.09	5.38	11.47
Optimal Constant	154.0	154.0	0.76%	3.19%	6.098	5.391	11.489

Table 4: Pricing policies under three regimes: (i) observed data, (ii) optimal mechanism, (iii) optimal mechanism subject to constant pricing over time i.e., $p_{-1} = p_0$

Two key lessons from the analysis. First, the optimal pricing mechanism for the monopolist raises the prices relative to the data/control group. This, unsurprisingly increases the revenue and lowers the consumer welfare. We find the impact on social welfare to also be negative.

The second lesson, and the one more relevant to our focus, is on price discrimination. The optimal pricing mechanism sets $p_0^* > p_{-1}^*$ but by a negligible margin. The ordinal comparison was to be expected from proposition 2 based on our estimate that $\gamma^1 > 0$. And the negligible margin could informally be suggested by the small magnitude of our estimated γ^1. Given that the optimal degree of price discrimination is minimal, it should also be unsurprising to observe that, as Table 4 shows, this pricing policy does not deliver meaningfully different numbers on revenue, social welfare, and consumer welfare, relative to a scenario in which the monoplist optimizes under the constraint that $p_{-1} = p_0$.

Discussion. We finish this application section with a brief review of how our theoretical results connect with empirical inference. Recall that proposition 4 helps interpret the relative response rates from our total market test (exhibited in the right panel of figure 3), suggesting that the co-variance parameter γ^1 should be close to zero. Also, recall that proposition 2 suggests that a small γ^1 implies a small gain from price discrimination. Put differently, propositions 2 and 4, together, suggest that a relative response rate of approximately 1 should imply negligible gains from price discrimination. This is an instance of how combining our optimal-mechanism results and demand-response results can provide insights into optimal policy direct from the data, even when a formal demand model has not been estimated.

7 Additional Simulations

Although the estimation and counterfactual analysis of pricing policy on the data from our application is informative of the applicability of our framework, it only illustrates a subset of possible scenarios to which the framework applies. As a result, we turn to additional, counterfactual, scenarios that should showcase a broader range of applications for our model. We examine a number of scenarios based on the three statistics we have emphasized throughout the paper: (i) demand ratio $\frac{s_{-1}}{s_0}$ under equal prices which helps measure the degree of vertical v.s. horizontal differentiation, (ii) the cross-substitution rates in response to partial market tests which sheds light on the degree $\sigma_\mu$of preference heterogeneity among consumers, and (iii) the relative response rates under total market tests which sheds light on the co-variance γ^1 between preference for the baseline version $k = -1$ and the premium version $k = 0$. In addition to these three factors, we also examine marginal-cost heterogeneity. Under each scenario constructed based on these factors, we connect key data patterns to values for parameter estimates, to implications for optimal pricing and consumer- and social-welfare. We believe these additional simulations (which effectively start the counterfactual analysis from a counterfactual dataset) are a useful tool for studies such as ours where the key contribution is the framework.

7.1 Scenario 1

Table 5 and figure 4 describe the demand levels under this scenario. There are simulated levels of demand for the control group, a partial market test of 30%, as well as a total market test of 30%. Table 5 summarizes the share levels. Figure 4 provides the three summary statistics discussed above.[13]

Market	**Share** ($k = -1$)	**Share** ($k = 0$)
Control Group	0.14%	44.84%
Partial market test ($k = -1$, 30%)	67.02%	8.69%
Total market test (30%)	2.96%	81.30%

Table 5: Demand patterns representing scenario 1

Our theoretical results, combined with figure 4, can help paint a picture of this market. The left panel of the figure suggests a substantial degree of vertical differentiation between the two versions $k = 0$ and $k = -1$ of the product. The middle panel suggests a high cross substitution rate under a partial-market test. As is also clear from table 5, when moving from the control group to the partial market test, version $k = -1$ gains additional share, a significant portion of which come out of the consumers of version $k = 0$. By proposition 3, this implies some heterogeneity $\sigma_\mu > 0$ among consumers. By proposition 1, such a heterogeneity potentially incentivizes the monopolist to offer multiple prices and screen the consumers based on their μ_i levels. The next question is: is the differentiation between versions $k = -1$ and $k = 0$ a suitable "instrument of screening"? That is, would product differentiation by offering $k = -1$ and $k = 0$ help the monopolist to profitably implement price discrimination? proposition 4 suggests the answer would be yes if the relative response rate under a total market test is above one. The right

[13]For the shares ratio $\frac{s_{-1}}{s_0}$ under equal prices, we display the shares ratio under the total market test. We make this choice although there were other options, such as $\frac{s_{-1}}{s_0}$ under the control group. Our choice was guided by proposition 5 which implies $\frac{s_{-1}}{s_0}$ under equal prices of smaller value provides a better approximation of vertical differentiation.

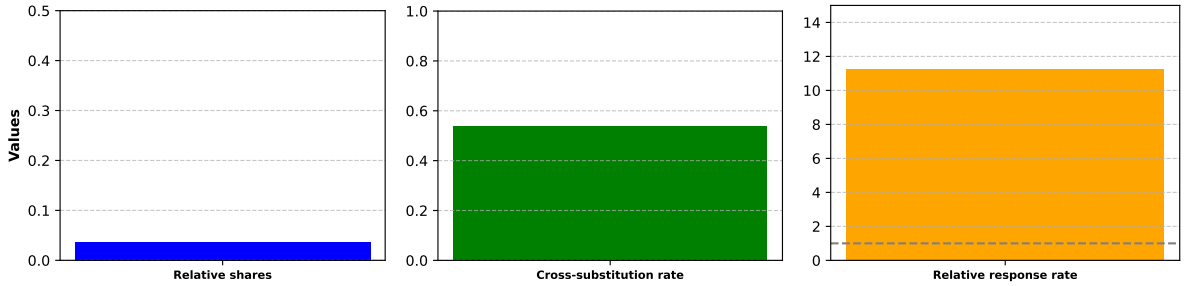


Figure 4: Relative shares, the cross substitution rate, and relative response rate, extracted from shares shown in table 5.

panel of figure 4 confirms that this is indeed the case. This can also be verified from table 5: when we move from the control group to the total market test, demand for both products $k = -1$ and $k = 0$ increases. Nonetheless, we observe a larger degree of relative demand increase for $k = -1$. This implies that marginal consumers–those with lower μ_i–tend to have a lower differential preference γ_i for product $k = 0$, relative to the infra-marginal consumers. That is, we should expect $\gamma^1 > 0$. In other words, preference for $k = 0$ positively varies with WTP and can, hence be used as a WTP proxy and leveraged for price discrimination.

What we expect about the estimation results, based on combining our propositions with the simulated shares above, is indeed confirmed when we calibarate a model to these data. Table 6 shows that we indeed fine a high γ^0, a non-trivial σ_μ, and a positive and non-trivial γ^1.

γ^0	σ_μ	γ^1
4.26	5.0	0.6

Table 6: Estimated parameters derived from shares shown in table 5

Finally, our simulations of optimal pricing policy also confirm what our propositions pointed to from raw data. We find that the optimal policy charges a p_0 that is by about 20% higher than p_{-1}. This generates a 4% higher profit relative to optimizing the prices subject to $p_{-1} = p_0$. Note that a 4% increase in profit is not a trivial number in the context of price discrimination. Especially in the case of 2nd-degree price discrimination, there are other studies that report similar numbers (e.g., Ghili and Yoon (2023)), and we are not aware of studies that report substantially higher gains. The impact of price discrimination on consumer welfare is, unlike the effect on profits, negative. But the overall effect on social welfare is positive.

pricing	p_{-1}	p_0	Share$_{-1}$	Share$_0$	revenue	consumer welfare	social welfare
Optimal Constant	74.0	74.0	2.13%	78.02%	59.36	25.75	85.10
Optimal Mechanism	64.7	78.0	32.74%	51.82%	61.71 (+4%)	24.14 (-6.3%)	85.85 (+0.9%)

Table 7: Pricing policies for scenario 1 under two regimes: (i) optimal mechanism subject to constant pricing over time i.e., $p_{-1} = p_0$, (ii) optimal mechanism

To summarize, this analysis with simulated data goes beyond our field experiment to further showcase the usefulness of the theoretical results, in conjunction with our experiment design, in tying empirical patterns to optimal policy, often even before a structural model is estimated.

7.2 Scenario 2

This scenario is summarized in table 8 and figure 5. These numbers portray a similar market to what we considered in the previous scenario, but with one major difference: the relative response rate shows that under a 30% total market discount, demand for the superior version $k = 0$ increases by a wider relative margin, compared to the demand for $k = -1$.

Market	**Share** ($k = -1$)	**Share** ($k = 0$)
Control Group	7.67%	41.49%
Partial market test ($k = -1$, 30%)	75.00%	4.95%
Total market test (30%)	8.57%	91.40%

Table 8: Demand patterns representing scenario 2

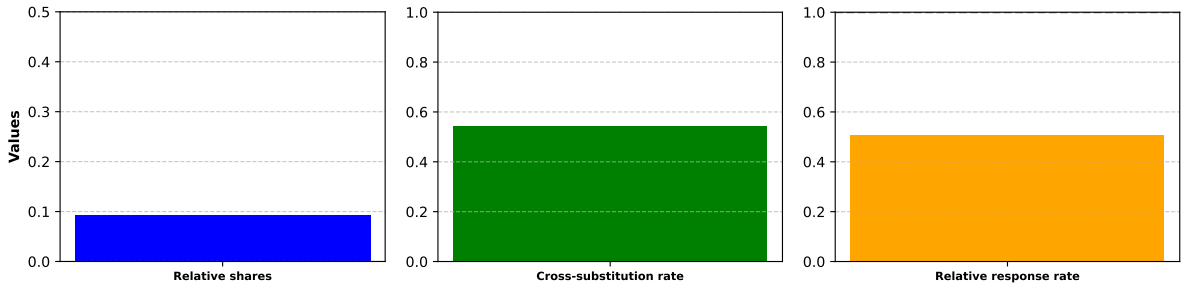


Figure 5: Relative shares, the cross substitution rate, and relative response rate, extracted from shares shown in table 8.

Similar to scenario 1, in this scenario we also expect a degree of preference heterogeneity $\sigma_\mu > 0$. This can be inferred combining proposition 3 with the high cross-substitution rate under the partial market test. This, once again, implies that the monopolist would have the incentive to use the versioning based

on $k = -1, 0$ for price discrimination, if the differentiation between the two version provided a suitable instrument of screening. Unlike scenario 1, however, in our current scenario, this differentiation is not a good instrument. As the relative response rate, combined with proposition 4, implies, we should expect a negative γ^1. That is, higher WTP consumers tend to have a lower differentiation willingness to pay for version $k = 0$.

What is the monopolist's optimal policy in this market? The high cross-substitution rate suggests a non-trivial preference heterogeneity $\sigma_\mu$according to proposition 3. In conjunction with this, proposition 1 suggests that the monopolist would want to offer multiple prices in order to capture a higher surplus. The next question is whether the versioning based on $k = -1, 0$ provides a suitable means of screening and capturing surplus. proposition 4, combined with the observed relative response rates result from the righ panel of figure 5, suggests that the answer is no. The co-variation parameter γ^1 is expected to be negative, according to this proposition. Table 9 below confirms these expectations:

γ^0	σ_μ	γ^1
5.2	4.6	-0.73

Table 9: Estimated parameters derived from shares shown in table 8

Note that propositions 1 and 2 may have different implications for optimal pricing in this scenario. proposition 2, which directly applies to the setting we are studying, implies that $p^*_{-1} > p^*_0$. proposition 1 which only allows for vertical differentiation between the two products, however, suggests that price discrimination is not profitable, i.e., the monopolist would not lose profit by setting $p_0 = p_{-1}$. Both of these results are reflected in table 10 . In line with proposition 2, we can indeed see that $p^*_{-1} > p^*_0$. However, given the large degree of vertical differentiation (corroborated by the left panel of figure 5), the implication from proposition 1 holds in approximation: unlike scenario 1 in which the optimal flexible policy delivered a 4% higher profit relative to the optimal uniform-price policy, here the improvement is only 0.2%. This is in line with our previous conjecture that when the differentiation between the products is primarily vertical, proposition 1 holds in approximation.

pricing	p_{-1}	p_0	Share$_{-1}$	Share$_0$	revenue	consumer welfare	social welfare
Optimal Constant	87.0	87.0	8.58%	87.24%	83.43	13.26	96.69
Optimal Mechanism	92.22	87.0	3.62%	92.21%	83.61 (+0.2%)	12.96 (-2.3%)	96.57 (-0.1%)

Table 10: Pricing policies for scenario 2 under two regimes: (i) optimal mechanism subject to constant pricing over time i.e., $p_{-1} = p_0$, (ii) optimal mechanism

Summary and other simulations. Our two simulation analyses above showcase how our theoretical results, in conjunction with our proposed tests, can help understand the fundamentals of a price-discrimination market even before estimating a structural model. In both cases, one can use our propositions to make conjectures about the parameter estimates and, the shape of the optimal policy, and the relative performance of the optimal policy, once compared against uniform pricing.

8 Discussion

8.1 Applicability

Our methodology helps to empirically address a set of screening problems with a focus on second-degree price discrimination. At the firm level, this methodology is applicable by any firm that offers multiple versions of a product and is able to carry out partial- and total-market tests. We expect digital and technology products to constitute a large subset of the possible applications. Cloud computing, LLM-based products, digital devices such as smartphones and laptops, are among such examples. In general, any setting in which a firm considers whether to "version" its products based on one or more features can benefit from our framework. Our total-market tests can be implemented using by issuing potential consumers discount offers that can be applied regardless of the product purchased. Partial-market tests can be implemented through discount offers that apply only to one product.

8.2 Future Research, Extensions, and Limitations

8.2.1 Future research

The empirical implementation of screening models is a difficult yet crucial task. This is especially the case when one's objective is to make the framework useful to the mechanism designer (e.g., monopolist), which requires avoiding the use moments that assume optimal policy on the part of the designer. We see our methodology as an early attempt at this objective. Additional advances, both on the mechanism design side and on the econometric side, should further help with this objecive.

On the theory side, we see two key areas. The first would be to extend the models of screening in the literature to settings in which the utility function includes idiosyncratic terms at the *individual-item* level. This is equivalent to the ϵ_{ik} terms in our specification. These terms help connect screening models to empirical applications because they help to rationalize a wide range of datasets which may not otherwise perfectly fit with the behavior predicted by a theoretical model model. We are not aware of any study that allows for such randomness at the ik in a general fashion. Our treatment of the terms ϵ_{ik} imposed an extreme type I distribution assumption. The closest other study that we are aware of is Rochet and Stole (2002) which introduces random participation, equivalent to having ϵ_{ik} only for the case where k is the outside option. This structure allows the idiosyncratic term to only impact the individual rationality (IR) constraint, rather than both IR and IC. We believe a general treatment of the idiosyncratic terms (i.e., for all ik, without functional form assumptions) will add significant value; but at the same time, it will complicate the screening problem, since it introduces disturbances to shape of the incentive compatibility (IC) constraints.

Another direction of future research in the theory of screening that would facilitate empirical applications is the development of demand-response results. To the best of our knowledge, the chief focus in theoretical literature on screening models is the characterization of optimal mechanisms. That is, results that characterize the IC and IR constraints as a function of the model primitives. We believe that supplementing these optimal-mechanism results with demand-response results will increase the likelihood of empirical follow-ups. As a reminder, by demand-response results, we mean those that predict how certain patterns of buyer behavior under a pre-determined (and not necessarily optimal) mechanisms can help infer certain features of the model primitives. propositions 3 and 4 provide illustrate this in

our specific context, but we believe demand-response results can add value to screening models more broadly.

On the empirical front, a key avenue for future research is the design of minimal experimental treatments for the identification of screening models without supply side moments. Another important direction would be extending our analysis of experimental price variation to a study of optimal instruments in an environment where prices may be endogenous. We conjecture that, in line with our tests, the suitable instruments would also need to be of two types: (i) instruments inducing variation in the differential value between the products, in a similar spirit to our partial-market tests and in line with differential-valuation instruments a la Gandhi and Houde (2023); (ii) instruments inducing exogenous variation in market participation without inducing cross-substitution, in a similar spirit to our total-market tests. We conjecture that with these two types of instruments, an empirical model of monopoly 2PD may be identified without supply-side moments and with aggregate non-experimental data. However, we leave a full analysis of this subject to future research.

More specific directions of future research on empirical implementation of 2PD models would entail a number of possible extensions to our model, which we turn to next.

8.2.2 Extensions to our model

There are a number of possible extensions to our model which we believe may be useful in some applications.

First, instead of the linear relationship $\gamma_i \equiv \gamma^0 + \gamma^1 \times \mu_i$, one can extend the relationship between the two random effects to allow for nonlinearity. Such an extension would lead to different functional forms for the marginal distribution of μ_i and that of γ_i. If such a model is to be empirically estimated using partial- and total-market tests, we expect that multiple levels of treatment are necessary. In particular, we expect that the shape of the relative response rate across different total-market treatment levels will help to identify the curvature of the function $\gamma(\mu_i)$. We conjecture that all of our propositions extend to this nonlinear case in a relatively straightforward fashion. Nevertheless, we leave a foral extension to future research.

Second, one can extend the model to capture a full joint distribution on μ_i and γ_i, compared to the current setting in which γ_i is modeled as a deterministic function of μ_i. Such an extension will allow for the correlation between γ_i and μ_i to assume any value in the interval $[-1, 1]$ whereas our current model only allows for $\{-1, 0, 1\}$ (note that our current model does allow for any co-variance level between the two random effects). Our initial analyses with simulated data suggests that this extension is of lesser relevance to the analysis of 2PD, relative to an extension of the relation between γ_i and μ_i to nonlinear functions.

A third direction in which our model can be extended would be by increasing the number of products beyond two. Such applications are relatively common, especially cases of three versions of a product. We expect our results to naturally extend to the cases of $K > 2$ products. That said, a formal analysis falls beyond the scope of this paper.

Finally, an important extension would be incorporation of oligopolistic competition and the implications it might have for firm-optimal and socially efficient screening. We expect this extension to be challenging for a number of reasons. Chief among those reasons is the presence of multiple equilibria

(see Ellison (2005) for more details).

9 Conclusion

This paper developed a model of second-degree price discrimination (2PD). Our model follows the recent theoretical literature on 2PD, thereby allowing to replicate and build on some of the relevant results from that literature on how the joint distribution of the model primitives impacts the optimal selling mechanism. Among other objects, it captures the role of the covariance, across consumers, between the baseline valuation (i.e., affecting WTP for all versions of the product) and the differentiation valuation. We argue that this parameter is difficult to credibly estimate when the number of product versions is small (an inherent trait of 2PD application), and when data on consumer characteristics that derive the covariance is not available to the mechanism designer. We then offer empirical testes that help estimate the parameters of the model under such constraints. Our partial-market tests fixe the price of one version of the product, varying the other price. Our total-market tests tests vary both prices uniformly, shutting off cross substitution between versions.

To demonstrate the applicability of our framework in real-world 2PD cases, we implement the experiment design and estimate the empirical model in the cotext of seat selection, in collaboration with an international airline based in Asia. The differentiating feature between the two versions of the product in our context is timing of purchase. We estimate the model and find that differentiating the prices based on timing of purchase has only a negligible impact on firm profit and consumer/social welfare. We also use this analysis to showcase the direct interpretability of our experiment design. We argue that the results from our partial- and total-market tests in this application are already suggestive that price discrimination would have little effect, even before a structural model of demand is estimated and optimal pricing policy analyzed.

We further expand on our interpretability analysis by studying a number of simulated datasets presented in the form of market shares for the two products under the control group, partial-market test, and total-market test. Each simulated dataset, hence, represents a scenario on how the market responds to the key tests. We argue that the propositions we developed when analyzing the optimal mechanisms, alongside our propositions for experiment design, can help interpret the results and form expectations on the shape of the optimal policy. We then estimate the model and solve for the optimal policy, demonstrating that those expectations are confirmed.

Although the differentiating feature between the product version in our application was purchase timing, the methodology applies more broadly. The framework applies in the case any form of differentiating factor between the products. We expect digital and technology products (e.g., cloud computing, Gen AI products, digital devices such as smartphones and laptops, etc) to constitute a large set of possible applications. The only requirement is that the two types of tests be feasible to implement.

There are a number of future directions in which our study may be extended. We argue that some such directions, such as extending the analysis to more than two product versions, are more straightforward; while other directions, such as extending the framework to oligopoly markets, would be more complex.

References

Adams, W. J. and J. L. Yellen (1976). Commodity bundling and the burden of monopoly. *The quarterly journal of economics*, 475–498.

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.

Armstrong, M. (2013). A more general theory of commodity bundling. *Journal of Economic Theory 148*(2), 448–472.

Armstrong, M. (2016). Nonlinear pricing. *Annual Review of Economics 8*, 583–614.

Aryal, G., C. Murry, and J. W. Williams (2021). Price discrimination in international airline markets. *arXiv preprint arXiv:2102.05751*.

Bergemann, D., B. Brooks, and S. Morris (2019). Revenue guarantee equivalence. *American Economic Review 109*(5), 1911–1929.

Bergemann, D., T. Heumann, and M. C. Wang (2024). A unified approach to second and third degree price discrimination. *arXiv preprint arXiv:2401.12366*.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T. and P. A. Haile (2023). Nonparametric identification of differentiated products demand using micro data. Technical report, Working Paper.

Berry, S. T. and P. A. Haile (2024). Nonparametric identification of differentiated products demand using micro data. *Econometrica 92*(4), 1135–1162.

Ellison, G. (2005). A model of add-on pricing. *The Quarterly Journal of Economics 120*(2), 585–637.

Gandhi, A. and J.-F. Houde (2023). Measuring substitution patterns in differentiated-products industries. *NBER Working paper* (w26375).

Ghili, S. (2023). A characterization for optimal bundling of products with nonadditive values. *American Economic Review: Insights 5*(3), 311–326.

Ghili, S. and R. Yoon (2023). An empirical analysis of optimal nonlinear pricing. *arXiv preprint arXiv:2302.11643*.

Grigolon, L. and F. Verboven (2014). Nested logit or random coefficients logit? a comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics 96*(5), 916–935.

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.

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.

McAfee, R. P., J. McMillan, and M. D. Whinston (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics 104*(2), 371–383.

McManus, B. (2007). Nonlinear pricing in an oligopoly market: The case of specialty coffee. *The RAND Journal of Economics 38*(2), 512–532.

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

Nair, H. (2007). Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. *Quantitative Marketing and Economics 5*, 239–292.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

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

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

Rochet, J.-C. and L. A. Stole (2003). The economics of multidimensional screening. *Econometric Society Monographs 35*, 150–197.

Salant, S. W. (1989). When is inducing self-selection suboptimal for a monopolist? *The Quarterly Journal of Economics 104*(2), 391–397.

Sorensen, J. (2023, May). Airlines assign big revenue priority to seat selection. Technical report. Created with Microsoft® Word 2016.

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

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance 16*(1), 8–37.

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

Yang, F. (2023). Nested bundling.

Yang, F., P. Dworczak, and M. Akbarpour (2024). Comparison of screening devices. *Available at SSRN 4456198*.

Appendices

A Proofs

A.1 Proof of proposition 1

We invoke the revelation principle and examine the space of direct mechanisms. In a direct mechanism, the monopolist commits to a selling mechanism which consists of a payment rule $t : \mathbb{R} \to \mathbb{R}$ and an allocation rule $x : \mathbb{R} \to \{-1, 0, \emptyset\}$. In words, the mechanism (t, x) can take a consumer type μ_i and, as a function of that, determine (i) which of the three options (version $k = -1$, version $k = 0$, or the outside option) she receives, and (ii) how much she should pay the monopolist. This mechanism is announced to all consumers. Then each consumer i "announces type $\tilde{\mu}_i$" which, in principle, may or may not coincide with her true type μ_i. A mechanism satisfies incentive compatibility (IC) if each consumer is weakly better off announcing her true type:

$$\forall \tilde{\mu}_i, \mu_i : u_{i,x(\mu_i)} - t(\mu_i) \geq u_{i,x(\tilde{\mu}_i)} - t(\tilde{\mu}_i)$$

A mechanism satisfies individual rationality (IR) if the payoff to being truthful is weakly larger than the outside option payoff (in our case, zero):

$$\forall \mu_i : u_{i,x(\mu_i)} - t(\mu_i) \geq 0$$

Applying the revelation principle to our context implies that the optimal posted pricing mechanism is equivalent to the optimal direct mechanism that satisfies IC and IR. Next, we turn to proving the statements of the theorem.

We first show that under $\gamma^1 < 0$, optimal pricing involves no price discrimination: $p_0^* = p_{-1}^*$. To see this, assume, on the contrary, that $p_0^* \neq p_{-1}^*$. There are two possibilities:

- $p_0 < p_{-1}$: This means that there are μ and $\tilde{\mu}$ such that: $x(\tilde{\mu}) = 0$, $x(\mu) = -1$ $p_0 = t(\tilde{\mu}) < t(\mu) = p_{-1}$. Such a mechanism violates the IC condition since μ would prefer the allocation for $\tilde{\mu}$ which entails both a better product and a smaller payment. Thus, $p_0 < p_{-1}$ cannot hold.
- $p_0 > p_{-1}$: Consider two consumer types μ and $\tilde{\mu}$ and assume $x(\tilde{\mu}) = 0$, $x(\mu) = -1$, $t(\tilde{\mu}) = p_0$, and $t(\mu) = p_{-1}$. That is, consumer $\tilde{\mu}$ is allocated $k = 0$ at price p_0 while μ is allocated $k = -1$ at price $p_{-1} < p_0$. Applying IC twice, we get:

$$u_{i,x(\mu)} - t(\mu) \geq u_{i,x(\tilde{\mu})} - t(\tilde{\mu})$$

and

$$u_{\tilde{i},x(\mu)} - t(\mu) \leq u_{\tilde{i},x(\tilde{\mu})} - t(\tilde{\mu})$$

where $\tilde{i}$ indexes consumer $\tilde{\mu}$. Replacing from eq. (1), eq. (2), and the allocation into the above inequalities, we get:

$$a + \mu - p_{-1} \geq a + \mu(1 + \gamma^1) + \gamma^0 - p_0$$

and

$$a + \tilde{\mu}(1 + \gamma^1) + \gamma^0 - p_0 \geq a + \tilde{\mu} - p_{-1}$$

Simplifying, we obtain:

$$p_0 - p_{-1} \geq \mu\gamma^1 + \gamma^0$$

$$\tilde{\mu}\gamma^1 + \gamma^0 \geq p_0 - p_{-1}$$

Adding the left sides together and further simplifying, we have:

$$\tilde{\mu}\gamma^1 \geq \mu\gamma^1$$

By $\gamma^1 < 0$, this means $\tilde{\mu} \leq \mu$. In words, the type allocated version $k = -1$ has to be the "higher" type. Next, note that by IR, we have:

$$a + \tilde{\mu}(1 + \gamma^1) + \gamma^0 - p_0 \geq 0$$

This, combined with $\tilde{\mu} \leq \mu$, yields:

$$a + \mu(1 + \gamma^1) + \gamma^0 - p_0 \geq 0$$

This latter inequality means that if the monopolist increased the price of $k = -1$ to a number at or above p_0, all types μ who before the price increase purchased $k = -1$, would now switch to purchasing $k = 0$ at $p_0 > p_{-1}$. This would raise the firm's profit strictly. As a result, price-discriminating with $p_0 > p_{-1}$, although possible to implement with IC and IR, would not be optimal from the perspective of the monopolist.

The above discussion completes the proof in the case of $\gamma^1 < 0$. We now turn to the case of $\gamma^1 \geq 0$. To complete the proof in this case, we closely follow the analysis in Ghili (2023). Note that under $\gamma^1 \geq 0$,

the setting without error terms, the type space of consumers exhibits a notion of monotonicity. More specifically, for any $\mu \geq \tilde{\mu}$, we have:

$$a + \mu \geq a + \tilde{\mu}$$

Also, by $\gamma^1 \geq 0$, we get:

$$\gamma^0 + \mu\gamma^1 \geq \gamma^0 + \tilde{\mu}\gamma^1$$

In words, consumer μ is both has a higher willingness to pay for $k = -1$ than does $\tilde{\mu}$ and has a higher willingness to pay for an upgrade from $k = -1$ to $k = 0$ than does $\tilde{\mu}$. In addition to monotonicity, the demand system also satisfies quasi-concavity: given that the utilities are distributed unifromly, the profit function is quasi-concave in prices.

By the monotonicity and quasi-concavity properties mentioned above, the main result from Ghili (2023) applies. According to this result, price discrimination is optimal if and only if the demand for version $k = -1$, if sold alone (i.e., no option to buy $k = -1$) and priced optimally, is strictly larger than the demand for version $k = 0$, if sold alone and priced optimally.

We skip the derivations but it is straightforward to show that if version $k = 0$ is sold on its own, the optimal price will be given by:

$$p_0^{**} = \frac{\sigma_\mu(1 + \gamma^1) + \gamma^0 + a}{2b}$$

and that the corresponding demand volume is:

$$D_0^{**} = \frac{\sigma_\mu(1 + \gamma^1) + \gamma^0 + a}{2(1 + \gamma^1)}$$

Likewise, it can be shown that the optimal sold-alone price and demand volume for $k = -1$ are, respectively, given by:

$$p_{-1}^{**} = \frac{\sigma_\mu + a}{2b}$$

and:

$$D_{-1}^{**} = \frac{\sigma_\mu + a}{2}$$

Ghili's condition for the optimality of price discrimination, hence, translates to $D_0^{**} < D_{-1}^{**}$. Replacing into this inequality from the above and simplifying, we get:

$$D_0^{**} < D_{-1}^{**} \Leftrightarrow a > \frac{\gamma^0}{\gamma^1} \tag{6}$$

Next, we turn to what the optimal prices (p_{-1}^*, p_0^*) are when price discrimination is optimal for the

monopolist. Ghili (2023) shows that the optimal price p^*_{-1} in this case will be the same as the optimal sold-alone price p^{**}_{-1}. He also shows that $p^* = p^{**}_{-1} + \Delta p^*$ where Δp^* is the optimal price that the monopolist would charge under a hypothetical where (i) all consumers are endowed with $k = -1$ and (ii) the monopolist sells "an upgrade to $k = 0$" and prices it optimally. Note that each consumer i's utility for such an upgrade will be given by:

$$u_{i,diff} = u_{i,0} - u_{i,-1} \equiv \gamma_i - b\Delta p \tag{7}$$

where $\gamma_i := \gamma^0 + \gamma^1 \mu_i$ is uniformly distributed on the interval $[\gamma^0 - \gamma^1\sigma_\mu, \gamma^0 + \gamma^1\sigma_\mu]$. This uniform distribution yields a linear demand curve, thereby allowing for a straightforward derivation of optimal pricing. We skip the derivation. The result is as follows:

$$\Delta p^* = \frac{\sigma_\mu + \frac{\gamma^0}{\gamma^1}}{2b}$$

Note that both p^*_{-1} and Δp^* agree with the claims made in the statement of the proposition.

Also note that the for the case where non-discrimination is optimal, the optimal price p^*_0 is the same as p^{**} derived above, which agrees with the statement of the proposition. This finishes the proof of the proposition. **Q.E.D.**

A.2 Proof of proposition 2

We only show that $p^*_{-1} > p^*_0$ if $\gamma^1 < 0$. The proofs of the other cases (i.e., $\gamma^1 = 0$ and $\gamma^1 > 0$) are identical. Also, for this proof, we will normalize the total market size to 1. This is without loss.

We take a contrapositive approach. We assume that $\gamma^1 < 0$ and, at the same time, $p^*_{-1} \leq p^*_0$; and we reach a contradiction.

Start by observing that the share of product k under arbitrary prices (p_{-1}, p_0), conditional on the set of consumers who have the same level of μ, is given by:

$$s_{k|\mu}(p_{-1}, p_0) = \frac{e^{a - bp_k + (\gamma^0 + \mu\gamma^1)\mathbb{1}_{k=0} + \mu}}{1 + \Sigma_{k'} e^{a - bp_{k'} + (\gamma^0 + \mu\gamma^1)\mathbb{1}_{k'=0} + \mu}} \tag{8}$$

The profit to the monopolist, $\pi(p_{-1}, p_0)$ is, hence, given by:

$$\pi(p_{-1}, p_0) = \Sigma_k \big(p_k \times \int_\mu s_{k|\mu}(p_{-1}, p_0) f_\mu d\mu \big) \tag{9}$$

Lemma 1. *At any pair of prices (p_{-1}, p_0), the partial derivative of the profit function with respect to price p_k of version k is given by:*

$$\forall k \in \{-1, 0\} : \frac{\partial \pi}{\partial p_k} = \int_\mu s_{k|\mu} \times \big(1 - b \times p_k \times (1 - s_{k|\mu}) + b \times p_{k'} \times s_{k'|\mu}\big) f_\mu d\mu \tag{10}$$

where $k' := -1 - k$ is the other version, and where the notation on the dependency of profits and shares has been suppressed for brevity.

Also, the sum of the derivatives of the profit with respect to all prices is given by:

$$\Sigma_k \frac{\partial \pi}{\partial p_k} = \Sigma_k\big(\int_\mu s_{k|\mu} f_\mu d\mu\big) - b\Sigma_k\big(p_k \times \int_\mu s_{k|\mu} \times s_{\emptyset|\mu} f_\mu\big) \tag{11}$$

where $s_\emptyset$ denotes the share of the outside option.

Proof. The proof of this lemma is tedious but straightforward and involves replacing from eq. (8) into eq. (9), differentiating, and re-arranging terms. It is left to the reader.

Now, observe that at the optimal prices (p^*_{-1}, p^*_0), the gradient of the profit function with respect to the vector of prices is zero. That is:

$$\Sigma_k \frac{\partial \pi}{\partial p_k}|_{(p^*_{-1}, p^*_0)} = 0 \tag{12}$$

We will use eq. (12) to show that $\frac{\partial \pi}{\partial p_{-1}}|_{(p^*_{-1}, p^*_0)} > 0$, which would contradict the optimality of (p^*_{-1}, p^*_0). Before we can do show this, we need additional lemmas.

Lemma 2. *Take any pair $\mu > \mu'$. Under any pair of price (p_{-1}, p_0), optimal or otherwise, the following are true when $\gamma^1 < 0$:*

- *If we denote by $s_\mu := \Sigma_k s_{k|\mu}$, we have:*

$$\frac{s_{-1|\mu}}{s_{-1|\mu'}} > \frac{s_\mu}{s_{\mu'}} > \frac{s_{0|\mu}}{s_{0|\mu'}}$$

- *If $\gamma^1 > -1$, then $s_\mu > s_{\mu'}$.*

Lemma 3. *At the optimal prices, we have:*

$$\int_\mu s_{-1|\mu} \times (1 - bp_{-1} \times s_{\emptyset|\mu}) f_\mu d\mu > 0.$$

We first use these lemmas to complete the proof of the main proposition by reaching the contradiction. We will then provide the proofs of these lemmas.

To reach a contradiction, we will show that at the optimal prices, we get $\frac{\partial \pi}{\partial p_{-1}} > 0$, which contradicts the optimality. To see why $\frac{\partial \pi}{\partial p_{-1}} > 0$ holds, observe that eq. (10) can be rearranged as:

$$\frac{\partial \pi}{\partial p_k} = \int_\mu s_{k|\mu} \times \big(1 - b \times p_k \times (1 - s_{k|\mu} - s_{k'|\mu})\big) f_\mu d\mu$$

$$+ \int_\mu s_{k|\mu} \times b \times \big(p_{k'} - p_k\big) \times s_{k'|\mu} f_\mu d\mu \tag{13}$$

Note that in this rearrangement, $\frac{\partial \pi}{\partial p_k}$ has two additive terms. Once we set $k = -1$ and observe that $1 - s_{k|\mu} - s_{k'|\mu}$ is by definition the same object as $s_{\emptyset|\mu}$ the first term in the expression for $\frac{\partial \pi}{\partial p_{-1}}$ collapses to the term that lemma 3 proves to be strictly positive. The second term is non-negative since it is an integral over a multiplication of four objects each of which is non-negative: shares $s_{-1|\mu}$ and $s_{0|\mu}$ are by

construction non-negative. So is the price coefficient b. As for $p_{k'} - p_k$, note that we are evaluating this term under the optimal prices and under $k = -1$. This term, hence, boils down to $p_0^* - p_{-1}^*$, which by our contrapositive assumption, is non-negative. As a result, $\frac{\partial \pi}{\partial p_{-1}}$ should be strictly positive when evaluated at optimal prices. This contradicts the optimality assumption, thereby proving the proposition assuming lemmas 2 and 3 are correct. We turn to proving these two lemmas next.

Proof of lemma 2. It is straightforward to derive the following expression for $\frac{s_{0,\mu}}{s_{-1,\mu}}$:

$$\frac{s_{0,\mu}}{s_{-1,\mu}} = e^{\gamma - b \times (p_0 - p_{-1})}$$

Thus, for any pair μ and μ', we get:

$$\frac{s_{0,\mu}}{s_{-1,\mu}} \Big/ \frac{s_{0,\mu'}}{s_{-1,\mu'}} = e^{\gamma - \gamma'} = e^{(\mu - \mu') \times \gamma^1}$$

As a result, if $\gamma^1 < 0$ and $\mu > \mu'$, we get:

$$\frac{s_{0,\mu}}{s_{-1,\mu}} \Big/ \frac{s_{0,\mu'}}{s_{-1,\mu'}} < 1$$

Re-arranging, we get:

$$\frac{s_{0,\mu}}{s_{0,\mu'}} < \frac{s_{-1,\mu}}{s_{-1,\mu'}}$$

which is part of the first statement of the lemma. Noticing that $s_\mu = s_{-1|\mu} + s_{0|\mu}$ and $s_{\mu'} = s_{-1|\mu'} + s_{0|\mu'}$, it is straightforward (hence left to the reader) to verify that the three-way comparison also holds:

$$\frac{s_{0,\mu}}{s_{0,\mu'}} < \frac{s_\mu}{s_{\mu'}} < \frac{s_{-1,\mu}}{s_{-1,\mu'}}$$

We now turn to the second statement of the lemma. With $\gamma^1 \geq -1$, one can verify that for any $\mu > \mu'$, we have:

$$a - bp_{-1} + \mu > a - bp_{-1} + \mu'$$

and

$$a - bp_0 + \mu + \gamma^0 + \gamma^1 \times \mu > a - bp_0 + \mu' + \gamma^1 \times \mu'$$

In other words: $\mathbb{E}u_{k|\mu} > \mathbb{E}u_{k|\mu'}$ for all k. As a result:

$$\Sigma_k e^{\mathbb{E}u_{k|\mu}} > \Sigma_k e^{\mathbb{E}u_{k|\mu'}}$$

$$\Rightarrow \frac{\Sigma_k e^{\mathbb{E}u_{k|\mu}}}{1+\Sigma_k e^{\mathbb{E}u_{k|\mu}}} > \frac{\Sigma_k e^{\mathbb{E}u_{k|\mu'}}}{1+\Sigma_k e^{\mathbb{E}u_{k|\mu'}}}$$

which means $s_\mu > s_{\mu'}$.

Both claims in the lemma have now been proven. ■

Proof of lemma 3. Start by observing that the optimality of (p^*_{-1}, p^*_0) implies $\Sigma_k \frac{\partial \pi}{\partial p_k} = 0$. Replacing from eq. (11), we obtain:

$$\Sigma_k\big(\int_\mu s_{k|\mu} f_\mu d\mu\big) - b\Sigma_k\big(p_k \times \int_\mu s_{k|\mu} \times s_{\emptyset|\mu} f_\mu\big) = 0$$

Observing that $p^*_0 \geq p^*_{-1}$ and replacing p^*_0 by p^*_{-1} in the above formula, we get:

$$\Sigma_k\big(\int_\mu s_{k|\mu} f_\mu d\mu\big) - b\Sigma_k\big(p_{-1} \times \int_\mu s_{k|\mu} \times s_{\emptyset|\mu} f_\mu\big) \geq 0$$

Further rearranging, we get:

$$\Sigma_k\big(\int_\mu s_{k|\mu} \times (1 - bp_{-1} \times s_{\emptyset|\mu}) f_\mu d\mu\big) \geq 0$$

Denoting $s_\mu := \Sigma_k s_{k|\mu}$, we get:

$$\int_\mu s_\mu \times (1 - bp_{-1} \times s_{\emptyset|\mu}) f_\mu d\mu \geq 0 \tag{14}$$

Note that eq. (14) shows a weak inequality which, if proven to be strict, would be the exact statement of the lemma. Thus, we next turn to the task of showing the inequality is strict.

Denote by $\tilde{\mu}$ a value of μ such that $1 - bp_{-1} \times s_{\emptyset|\tilde{\mu}} = 0$. If such $\tilde{\mu}$ cannot be chosen because $1 - bp_{-1} \times s_{\emptyset|\mu}$ is strictly positive for the entire support of μ, then set $\tilde{\mu}$ to be the largest number in the support. Then, for any $\mu > (<)\tilde{\mu}$, we have:

$$1 - bp_{-1} \times s_{\emptyset|\mu} > (<) 0 \tag{15}$$

and

$$\frac{s_{-1|\mu}}{s_\mu} \Big/ \frac{s_{-1|\tilde{\mu}}}{s_{\tilde{\mu}}} > (<) 1 \tag{16}$$

where the latter inequality directly follows from lemma 2. Note that eq. (15) and eq. (16), together, imply that for any μ we have:

$$1 - bp_{-1} \times s_{\emptyset|\mu} > (<) 0 \Rightarrow \frac{s_{-1|\mu}}{s_\mu} \Big/ \frac{s_{-1|\tilde{\mu}}}{s_{\tilde{\mu}}} > (<) 1 \tag{17}$$

Next, note that if we take eq. (14), multiply all the positive terms in the integral by a scalar strictly

larger than 1, and multiply all the negative terms in the integral by a scalar strictly smaller than 1, the resulting expression will strictly increase, which means it will be strictly positive. Based on eq. (17), one can do this by multiplying the integrand by $\frac{s_{-1|\mu}}{s_\mu} \;/\; \frac{s_{-1|\tilde{\mu}}}{s_{\tilde{\mu}}}$. This yields:

$$\int_\mu (\frac{s_{-1|\mu}}{s_\mu} \;/\; \frac{s_{-1|\tilde{\mu}}}{s_{\tilde{\mu}}}) \times s_\mu \times (1 - bp_{-1} \times s_{\emptyset|\mu}) f_\mu d\mu > 0$$

Canceling the s_μ terms and factoring the constant term $\frac{s_{-1|\tilde{\mu}}}{s_{\tilde{\mu}}}$ outside of the integral, we get:

$$\frac{s_{\tilde{\mu}}}{s_{-1|\tilde{\mu}}} \times \int_\mu s_{-1|\mu} \times (1 - bp_{-1} \times s_{\emptyset|\mu}) f_\mu d\mu > 0$$

By $\frac{s_{\tilde{\mu}}}{s_{-1|\tilde{\mu}}} \geq 0$, we get:

$$\int_\mu s_{-1|\mu} \times (1 - bp_{-1} \times s_{\emptyset|\mu}) f_\mu d\mu > 0$$

which completes the proof of the lemma. ■

With the proofs of these two lemmas at hand, the proof of the proposition is complete. **Q.E.D.**

A.3 Proofs of proposition 4

Similar to proposition 2, we prove this result for the case of $\gamma^1 < 0$. The $\gamma^1 > 0$ and $\gamma^1 = 0$ cases would be almost identical. In the rest of the proof, we only work with price vectors in the form of $(p_{-1} + r, p_0 + r)$ where (p_{-1}, p_0) is fixed an r varies. As a result, we suppress the notations on the former and write various objects as functions of the latter only.

Lemma 4. *As we vary r, the relative shares of the two products $k = -1$ and $k = 0$ remains unchanged. Formally, there is a function $\eta : \mathbb{R} \to \mathbb{R}$ such that for any $r \in \mathbb{R}$: and any μ, we have:*

$$\frac{s_{-1|\mu}(r)}{s_\mu(r)} = \eta(\mu)$$

This function is strictly increasing in μ when $\gamma^1 < 0$, constant when $\gamma^1 = 0$, and strictly decreasing when $\gamma^1 > 0$.

In other words, the $k = -1$ to $k = 0$ share ratio does not depend on price shifter r.

Proof. Writing out the expressions for the shares, one can show:

$$\frac{s_{-1|\mu}(r)}{s_\mu(r)} = \frac{1}{1 + e^{\left(b(p_{-1} - p_0) + \gamma^0 + \gamma^1 \times \mu\right)}}$$

Given that b, p_{-1}, p_0, γ^0, and γ^1 are all fixed, the lemma immediately follows. ■

Lemma 5. *Assume the function $\eta(\mu)$ is positive and strictly increasing in μ. Also take two atom-less probability distributions $w_1(\cdot)$ and $w_2(\cdot)$ over μ that have the same support $[\underline{\mu}, \bar{\mu}]$ and assume they are*

both strictly positive on this support. If w_1 dominates w_2 in a Monotone Likelihood Ratio (MLR) sense: $\forall \mu > \mu' : \frac{w_1(\mu)}{w_1(\mu')} > \frac{w_2(\mu)}{w_2(\mu')}$*, then we have:*

$$\int_{\mu} \eta(\mu) w_1(\mu) d\mu > \int_{\mu} \eta(\mu) w_2(\mu) d\mu$$

Proof. Given that w_1 and w_2 are probability distributions, we get: $\int_{\mu} w_1(\mu) d\mu = \int_{\mu} w_2(\mu) d\mu$. This, combined with the MLR relationship, implies that:

$$\exists \tilde{\mu} : \begin{cases} w_1(\mu) > w_2(\mu) & \forall \mu > \tilde{\mu} \\ w_1(\mu) < w_2(\mu) & \forall \mu < \tilde{\mu} \end{cases}$$

Denoting $\Delta w(\mu) := w_1(\mu) - w_2(\mu)$, we can write: $\int_{\mu} \Delta w(\mu) d\mu = 0$. Multiplying by $\eta(\tilde{\mu})$, we obtain:

$$\int_{\mu} \eta(\tilde{\mu}) \Delta w(\mu) d\mu = 0$$

$$\Rightarrow \int_{\mu=\underline{\mu}}^{\tilde{\mu}} \eta(\tilde{\mu}) \Delta w(\mu) d\mu + \int_{\mu=\tilde{\mu}}^{\bar{\mu}} \eta(\tilde{\mu}) \Delta w(\mu) d\mu = 0$$

Note that the first integral consists of strictly negative Δw terms while the second integral consists of strictly positive ones. As a result, if we discount the terms in the first integral by multiplying them by factors in $(0, 1)$ and amplify the terms in the second by multiplying its terms by factors in $(1, \infty)$, the expression will strictly increase. Next, observe that by η being positive and strictly increasing, the fraction $\frac{\eta(\mu)}{\eta(\tilde{\mu})}$ is in $(0, 1)$ for all $\mu < \tilde{\mu}$ and it is in $(1, \infty)$ for all $\mu > \tilde{\mu}$. Thus, we can write:

$$\int_{\mu=\underline{\mu}}^{\tilde{\mu}} \eta(\tilde{\mu}) \frac{\eta(\mu)}{\eta(\tilde{\mu})} \times \Delta w(\mu) d\mu + \int_{\mu=\tilde{\mu}}^{\bar{\mu}} \frac{\eta(\mu)}{\eta(\tilde{\mu})} \times \eta(\tilde{\mu}) \Delta w(\mu) d\mu > 0$$

Simplifying and re-arranging, we get:

$$\int_{\mu=\underline{\mu}}^{\tilde{\mu}} \eta(\mu) \times \Delta w(\mu) d\mu + \int_{\mu=\tilde{\mu}}^{\bar{\mu}} \eta(\mu) \times \eta(\tilde{\mu}) \Delta w(\mu) d\mu > 0$$

$$\Rightarrow \int_{\mu=\underline{\mu}}^{\bar{\mu}} \eta(\mu) \times \Delta w(\mu) d\mu > 0$$

$$\Rightarrow \int_{\mu=\underline{\mu}}^{\bar{\mu}} \eta(\mu) \times w_1(\mu) d\mu > \int_{\mu=\underline{\mu}}^{\bar{\mu}} \eta(\mu) \times w_2(\mu) d\mu$$

which is the statement of the lemma. Note that throughout the proof, we implicitly used the atomlessness assumption when we decomposed the integral on the entire support into two integrals.

The proof of the lemma is now complete. ■

Lemma 6. *For any $r > 0$, the following comparison holds relative to $r = 0$:*

$$\frac{s_{-1}(r)}{s(r)} > \frac{s_{-1}(0)}{s(0)}$$

where the term s denotes the total ($k = -1$ and $k = 0$) share.

Proof. For any r, the expression $\frac{s_{-1}(r)}{s(r)}$ can be expanded as:

$$\frac{s_{-1}(r)}{s(r)} = \frac{\int_\mu s_{-1|\mu}(r) f(\mu) d\mu}{\int_\mu s_\mu(r) f(\mu) d\mu}$$

By lemma 4 and the definition of the function η, we get:

$$\frac{s_{-1}(r)}{s(r)} = \frac{\int_\mu \eta(\mu) \times s_\mu(r) f(\mu) d\mu}{\int_\mu s_\mu(r) f(\mu) d\mu}$$

Bringing the denominator inside of the integral in the numerator yields:

$$\frac{s_{-1}(r)}{s(r)} = \int_\mu \eta(\mu) \times \frac{s_\mu(r) f(\mu)}{\int_{\tilde{\mu}} s_{\tilde{\mu}}(r) f(\tilde{\mu}) d\tilde{\mu}} d\mu$$

Next, denote $\frac{s_\mu(r) f(\mu)}{\int_{\tilde{\mu}} s_{\tilde{\mu}}(r) f(\tilde{\mu}) d\tilde{\mu}}$ by $w_r(\mu)$.

If we follow similar steps to the above for the $r = 0$ case and define the function $w_0(\mu)$ appropriately, then proving the lemma would be equivalent to proving the following:

$$\int_\mu \eta(\mu) w_r(\mu) d\mu > \int_\mu \eta(\mu) w_0(\mu) d\mu \tag{18}$$

If we show that w_0 and w_r are both probability distributions and that the latter dominates the former in the MLR sense, we can apply lemma 5 and proof the current lemma.

Verifying that these functions are probability distribution is straightforward. One could integrate over μ and show the outcome is 1:

$$\int_\mu w_r(\mu) d\mu = \int_\mu \frac{s_\mu(r) f(\mu)}{\int_{\tilde{\mu}} s_{\tilde{\mu}}(r) f(\tilde{\mu}) d\tilde{\mu}} d\mu$$

$$= \frac{\int_\mu s_\mu(r) f(\mu) d\mu}{\int_{\tilde{\mu}} s_{\tilde{\mu}}(r) f(\tilde{\mu}) d\tilde{\mu}} = 1$$

Clearly, the same applies when $r = 0$.

We next show that for any $\mu > \mu'$:

$$\frac{w_r(\mu)}{w_r(\mu')} > \frac{w_0(\mu)}{w_0(\mu')}$$

It is straightforward to verify that all $f(\mu)$ and $f(\mu')$ terms as well as the integrals in the denominators

of these functions cancel out and the above comparison becomes equivalent to the following:

$$\frac{s_\mu(r)}{s_{\mu'}(r)} > \frac{s_\mu(0)}{s_{\mu'}(0)}$$

or, equivalently:

$$\frac{s_\mu(r)}{s_\mu(0)} > \frac{s_{\mu'}(r)}{s_{\mu'}(0)}$$

If for any k we denote $v_k(\mu) := a - b \times p_k + \mu + \mathbb{1}_{k=0} \times (\gamma^0 + \gamma^1 \mu)$, (i.e., basically utility the expected utility for product k), one can show the following for any μ:

$$\frac{s_\mu(r)}{s_\mu(0)} = \frac{1 + \Sigma_k e^{v_k(\mu)}}{e^{br} + \Sigma_k e^{v_k(\mu)}}$$

Next, observe that by $\mu > \mu'$ and $|\gamma^1| < 1$, we have: $\forall k : v_k(\mu) > v_k(\mu')$. This, in turn, implies:

$$\Sigma_k e_k^v(\mu) > \Sigma_k e_k^v(\mu')$$

This latter inequality, combined with $e^{br} > r$ (which comes from $b > 0$ and $r > 0$), yields:

$$\frac{1 + \Sigma_k e^{v_k(\mu)}}{e^{br} + \Sigma_k e^{v_k(\mu)}} > \frac{1 + \Sigma_k e^{v_k(\mu')}}{e^{br} + \Sigma_k e^{v_k(\mu')}}$$

Thus, we have $\frac{s_\mu(r)}{s_\mu(0)} > \frac{s_{\mu'}(r)}{s_{\mu'}(0)}$, which means probability distribution w_r strictly MLR-dominates probability distribution w_0. Also, it is straightforward to see that these distributions have no atoms. Thus, lemma 5 applies to eq. (18) and the proof of the current lemma is now complete. ■.

With lemma 6 proven, the proposition immediately follows from the observation that $\forall r : s(r) := s_{-1}(r) + s_0(r)$:

$$\frac{s_{-1}(r)}{s(r)} > \frac{s_{-1}(0)}{s(0)}$$

$$\Rightarrow \frac{s_{-1}(r)}{s_0(r)} > \frac{s_{-1}(0)}{s_0(0)}$$

Q.E.D.

B Extension to marginal costs

In this section, we state and prove a version of proposition 2 that allows for nonzero and heterogeneous marginal costs across versions.

Proposition 6. *Suppose the conditions for proposition 2 hold. Also suppose the marginal costs for producing version k is given by $c_k \geq 0$. The following are true about the monopolist's optimal prices*

(p^*_{-1}, p^*_0)*:*

- *If* $\gamma^1 < 0$*: then* $p^*_{-1} - c_{-1} > p^*_0 - c_0$*.*
- *If* $\gamma^1 > 0$*: then* $p^*_{-1} - c_{-1} < p^*_0 - c_0$*.*

In words, proposition 6 shows that the only change one needs to make to the statement of proposition 2 under non-zero marginal costs is to replace prices with margins.

Proof. Consider an "auxiliary" statement of a different 2PD problem with the following changes relative to what we have in the statement of the proposition:

1. Marginal costs are zero
2. Instead of eqs. (2) and (3), the utility is described by the following equations:

$$u_{ik} = (a - b \times c_{-1}) - b \times p_k + \mu_i + \gamma_i \times \mathbb{1}_{k=0} + \epsilon_{ik} \tag{19}$$

and

$$\gamma_i = \left(\gamma^0 - b \times (c_0 - c_{-1})\right) + \gamma^1 \times \mu_i \tag{20}$$

In words, this new problem sets marginal costs back to zero but replaces the utility parameters a and γ_0, respectively with $a - b \times c_{-1}$ and $\gamma^0 - b \times (c_0 - c_{-1})$.

Lemma 7. *If a monopolist facing the auxiliary problem charges prices* (m_{-1}, m_0)*, she will have the same demand (for each version) and the same total profit as a monopolist who faces the original problem and charges prices* $(p_{-1}, p_0) = (m_{-1} + c_{-1}, m_0 + c_0)$*.*

Proof. To see why the demands are the same, observe that all the utilities are the same. In one replaces $m_k = p_k - c_k$ into the formulas eqs. (19) and (20) and expands, one can verify that they collapse to the original utility formulas eqs. (2) and (3). Next, observe that in addition to all demand levels, margins are also the same between the two problems. In both cases, the margins are m_{-1} and m_0. Thus, profits are the same as well. ■

Next, note that (i) proposition 2 is applicable to the auxiliary problem, and that (ii) the formulation of the auxiliary problem modified a and γ^0 but did not change γ^1. Applying proposition 2, hence, yields:

- If $\gamma^1 < 0$: then $m^*_{-1} > m^*_0$.
- If $\gamma^1 > 0$: then $m^*_{-1} < m^*_0$.

where m^*_{-1} and m^*_0 are, respectively, the optimal prices for versions $k = -1$ and $k = 0$ in the auxiliary problem. But given lemma 7, $m^*_{-1} > m^*_0$ is equivalent to $p^*_{-1} - c^*_{-1} > p^*_0 - c^*_0$, and the similar equivalences apply when $m^*_{-1} < m^*_0$ or $m^*_{-1} = m^*_0$. This completes the proof of proposition 6. **Q.E.D.**